Sign Language Detection for Dumb and Deaf People Using Machine Learning

Sonali Nilesh Kelapure[1], Dr. Parminder Kaur Dhingra[2]

[1] Ph.D Scholar, Department of CSE, JNEC, MGM University.

[2] Professor and Director, Department of CSE, JNEC, MGM University.

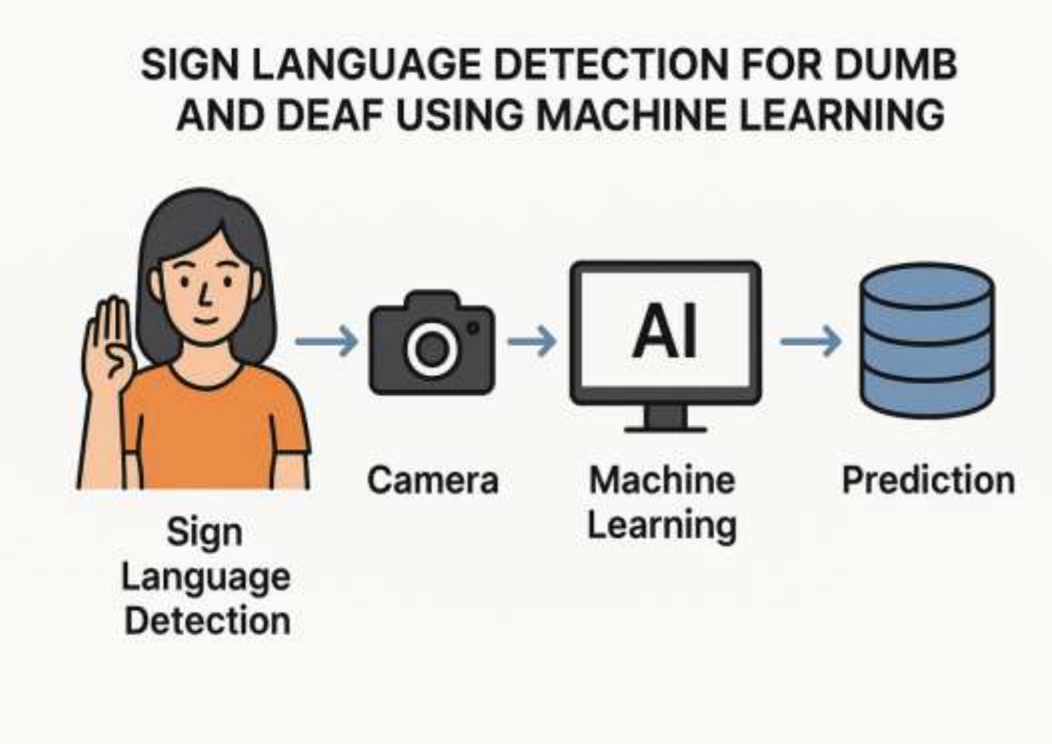
Abstract:

Communication is a fundamental human need, yet millions of individuals who are deaf or mute face significant barriers in expressing themselves due to the lack of understanding of sign language by the general population. This project presents a machine learning-based system designed to bridge this communication gap by detecting and translating sign language gestures into readable text and audible speech in real time. The proposed solution leverages computer vision and deep learning—specifically, a hybrid Convolution Neural Network (CNN) and Long Short-Term Memory (LSTM) model—to recognize static and dynamic hand gestures captured via webcam. A custom data-set of hand signs is used for training, and real-time prepossessing techniques such as background filtering and hand segmentation are applied for improved accuracy. Once a gesture is detected, it is translated into a corresponding word or sentence, and converted into voice output using a text-to-speech module. This system not only enhances communication for the deaf and mute but also paves the way for more inclusive human-computer Interaction systems.

Keywords: Deaf and Mute Communication, Gesture Recognition, Machine Learning, Real-time Detection; Assistive Technology, Text-to-Speech.

Early Approaches: Sensor-Based Systems

Initially, sign language recognition relied on sensor-based gloves (e.g., Data Glove), which captured finger joint movements using accelerometers and gyroscopes. These systems provided high accuracy but were expensive, uncomfortable, and not scalable for widespread use. *Gary Grimes et al.,(1991)*



Vision-Based Approaches

To overcome these limitations, researchers began developing vision-based systems using cameras to detect hand gestures. Starner.,Pentland.(1997) introduced real-time recognition of American Sign Language using Hidden Markov Models (HMMs) and video sequences. These systems laid the groundwork for modern deep learning approaches by emphasizing temporal sequence recognition.

Deep Learning Models

With the rise of Convolutional Neural Networks (CNNs), image-based gesture recognition improved significantly. CNNs enabled automatic feature extraction of hand shapes and positions from images.

- Simonyan & Zisserman (2014) introduced deeper CNN architectures (VGGNet) which improved accuracy in gesture classification.
- However, since sign language includes movement sequences, CNNs alone were insufficient.

To capture temporal dependencies, Recurrent Neural Networks (RNNs) and especially Long Short-Term Memory (LSTM) networks were combined with CNNs:

- Pigou et al. (2015) proposed a CNN-LSTM hybrid for continuous gesture recognition, demonstrating high accuracy in sign classification from video frames.

Real-Time Systems and Preprocessing

Open CV and Media Pipe are widely used for real-time hand detection and segmentation, enabling lightweight models suitable for webcams and mobile devices. Real-time systems typically include preprocessing stages like resizing, hand cropping, and frame normalization for model efficiency.

Audio Output and TTS Integration

To make systems more user-friendly for hearing users, Text-to-Speech (TTS) libraries such as gTTS (Google Text-to-Speech) and pyttsx3 are integrated for spoken output. This helps translate the sign language into audible language, making the system accessible and inclusive.

System Architecture for Real-Time Sign Language Detection:

1. Webcam Input – Captures video of sign gestures.
2. Preprocessing – Hand segmentation, frame resizing, normalization.
3. CNN Layers – Extract spatial features (hand position, shape).
4. LSTM Layers – Capture sequence of gestures over time.
5. Output Prediction – Classify as text/label.
6. TTS Engine – Converts text to voice output.

Summary of Research Gaps

Gap	Observations
Limited Sentence-Level Prediction	Most systems focus on word or alphabet-level detection.
Language Limitation	Few models are trained on Indian Sign Language (ISL).
Real-Time Integration	High accuracy systems are often not real-time.
Non-Manual Signs	Facial expressions, body posture often ignored.

Contribution of Present Work

This project builds a real-time, CNN-LSTM-based gesture recognition system, capable of interpreting static and dynamic sign language gestures and converting them into both text and speech, tailored for daily conversational use.

Proposed System: Sign Language Detection Using CNN + LSTM

1. Input Acquisition

Tool: Camera / Webcam

Output: Real-time video stream capturing hand/gesture movements.

2. Preprocessing Module

Operations:

Resize frames

Normalize images

Background subtraction (optional)

Frame sequencing

Output: Cleaned and shaped video frames sequence

3. Feature Extraction with CNN

CNN Model: Custom CNN / MobileNet / ResNet (lightweight preferred for real-time)

Purpose: Extract spatial features (hand shape, posture)

Output: Feature vectors per frame

Temporal Pattern Learning with LSTM

LSTM Model: Bidirectional/Stacked LSTM

Input: Sequence of CNN-extracted features

Purpose: Capture time-dependent gestures (like a full word or sentence)

Output: Predicted gesture/word label

Classification Layer

Model: Softmax / Fully Connected Dense Layer

Purpose: Map temporal features to sign class

Output: Final predicted label (e.g., "Hello", "Thank You")

Output Interface

Options:

Text Display

Voice Output using Text-to-Speech (TTS)

Tools: pyttsx3, gTTS, or browser TTS API

System Flow Diagram

Camera Input

↓

Frame Preprocessing

↓

CNN - Spatial Feature Extraction

↓

LSTM - Temporal Pattern Learning

↓

Dense Layer + Softmax Classification

↓

Predicted Sign Label

↓

Text Display / Voice Output

Technologies to Use

Language: Python

Deep Learning: TensorFlow / Keras / PyTorch

Real-Time Input: OpenCV

Deployment: Google Colab / Flask WebApp / Streamlit

TTS: pyttsx3 / gTTS

4. Comparative Summary of Methods

Approach	Pros	Cons
Sensor-based Gloves	High precision	Expensive, uncomfortable
Static Image CNN	Accurate for single sign	Cannot detect motion
CNN + LSTM	Effective for sequences	Needs large data, computationally expensive
MediaPipe + SVM	Lightweight, fast	Less accurate for complex signs
CNN + TTS	Complete solution	Needs fine-tuning for accent/speech

5. Conclusion

Machine learning has significantly enhanced sign language detection by providing robust techniques for recognizing both static and dynamic gestures. With advancements in deep learning and real-time computer vision, it is possible to build real-time systems to help bridge the communication gap for the deaf and mute. Future research should focus on creating large multilingual sign datasets, improving accuracy in real-time scenarios, and deploying efficient edge models.

6. References

1) Arpita Halder.,Akshit Tayade. (2021). Real-time Vernacular Sign Language Recognition using MediaPipe and Machine Learning. International Journal of Research Publication and Reviews. 2(5). 9-17.

2) Urav D.,Aasmi T.,Mahek U.,Shreya S. (2024) . Deep Learning-Enabled Smart Glove for Real-Time Sign Language Translation. J. Electrical Systems 20-10. 4874-4882.

3) Grimes *(1983), "Data Glove: Real-Time Gesture Input Device." US patent.* 4,414,537

4) Ian G.,Yoshua B., Aaron C. (2016). "Deep Learning", MIT Press. Genet Program Evolvable. 19 .305–307

5) Hochreiter S., Schmidhuber J., (1997). "Long Short-Term Memory". *Neural Computation*. 9 (8). 1735-1780.

6) Jens F.,Christoph S,Thomas H.(2014). RWTH-PHOENIX-Weather: A Large Vocabulary Sign Language Recognition and Translation Corpus. Conference: International Conference on Language Resources and Evaluation. 3785-3789.

7) Jungpil.,Abusalemas.,Kotas. (2023). Dynamic Korean Sign Language Recognition Using Pose Estimation Based and Attention-based Neural

8) Network, IEEE.

9) Melanshia I M.,Leena S.(2025). A comprehensive survey on recent advances and challenges in sign language recognition systems. Discov Artif Intell. 63-025-00629-7.

10) Muneeb U R., Fawad A., Muhammad K. (2021). Dynamic Hand Gesture Recognition Using 3D-CNN and LSTM Networks. Computers, Materials & Continua. Tech science press. 70(3)

11) Sharma R.(2022). "Static Sign Language Recognition Using CNN". *Journal of Computer Vision*.

12) Surya C.,Jeevanantham P.,Raja M.,Venkatesan P. (2025). Railway Foreign Object Detection System: A Novel Approach Using Machine Learning. International Journal of creative research thoughts. 13(5). 430-435.

13) Thad S.,Alex P.,(1997) Real time americal sign language recognition from video using hidden markov models. *Kluwer Academic Publishers. 227-243*

14) Wang, L. et al. (2023). "Dynamic Gesture Recognition Using CNN-LSTM Model". *IEEE Transactions on Multimedia*.

.