Assistive Communication for the Deaf

Using Real-Time Sign Language Detection

Submitted By:

Bhavesh Kerkar (Roll No. 19)
Devraj Pujari (Roll No. 43)

Division: A

Department of Information Technology
Sathaye College

Project Guide: Ms. Larisa Pegado

Date of Submission: 15 February 2026

1. Acknowledgement

We would like to express our sincere gratitude to our project guide, **Ms. Larisa Pegado**, for her valuable guidance, continuous support, and encouragement throughout the development of this project. Her insights and suggestions greatly helped us in improving both the technical and research aspects of our work.

We are also thankful to the **Department of Information Technology, Sathaye College**, for providing us with the necessary resources and academic environment to successfully complete this project.

We extend our heartfelt appreciation to all the faculty members for their support and motivation during the project period.

Finally, we would like to thank our friends and family for their constant encouragement and support, which helped us stay motivated and focused throughout the project development.

TABLE OF CONTENT

2. Abstract

Sign Language is one of the primary means of communication for the deaf and hard-of-hearing community. However, communication barriers exist between sign language users and individuals unfamiliar with sign language. This project presents an American Sign Language (ASL) Detection System that uses computer vision and deep learning techniques to recognize hand gestures and convert them into corresponding alphabets.

The system captures real-time hand gestures using a webcam, processes the images, and applies a trained machine learning model to classify the detected signs. Techniques such as image preprocessing, hand landmark detection, and neural network-based classification are used to improve recognition accuracy. The proposed system aims to provide an efficient and user-friendly solution for bridging the communication gap between sign language users and non-signers.

The results demonstrate that the system is capable of recognizing ASL gestures with satisfactory accuracy under controlled conditions. Future improvements may include expanding the dataset, improving model robustness, and supporting full sentence recognition.

3. Introduction

Communication is a fundamental aspect of human interaction. For individuals with hearing or speech impairments, sign language serves as a primary mode of communication. American Sign Language (ASL) is widely used by the deaf community, but many people are not familiar with it, which creates communication challenges in daily life.

With advancements in computer vision and artificial intelligence, it is now possible to develop systems that can recognize hand gestures in real time. Gesture recognition systems use cameras and machine learning algorithms to detect and interpret hand movements. Such systems can help translate sign language into readable text or speech, thereby enabling smoother communication between different communities.

This project focuses on developing an ASL detection system using image processing and deep learning techniques. The system captures hand gestures through a webcam, extracts relevant features, and classifies them into corresponding ASL alphabets. The objective of this project is to design an accurate, real-time, and cost-effective solution that assists in reducing communication barriers and promotes inclusivity.

---

4. Problem Statement

Deaf and hard-of-hearing individuals rely on sign language as their primary mode of communication. However, a significant communication gap exists between sign language users and individuals who are not familiar with sign language. This barrier often limits effective real-time interaction in educational institutions, workplaces, healthcare facilities, and public environments.

Although interpreters can help bridge this gap, they are not always available, and existing technological solutions can be expensive, complex, or inefficient. Many systems require large datasets, high computational power, or specialized hardware, making them less accessible for everyday use.

Therefore, there is a need for an affordable, accurate, and real-time assistive system that can detect and translate sign language gestures into understandable text or speech. Such a system would enhance communication accessibility and promote inclusivity for the deaf community.

5. Objectives

The main objectives of this project are:

- To collect and prepare a dataset of hand landmark coordinates representing American Sign Language (ASL) alphabets.
- To design and develop a deep learning-based classification model for gesture recognition.
- To implement real-time hand detection and gesture recognition using OpenCV.
- To convert the predicted text output into speech using a text-to-speech module.
- To evaluate the performance of the developed model based on accuracy and real-time efficiency.

6. Literature Review

Sign language recognition has gained significant attention in the fields of computer vision and deep learning. Traditional approaches for sign language detection primarily relied on image-based Convolutional Neural Networks (CNNs). In these systems, raw images of hand gestures were directly fed into deep learning models for feature extraction and classification. While CNN-based models achieved good accuracy, they required large image datasets and high computational resources for training.

Many existing systems depend heavily on extensive labeled image datasets to perform effectively. Collecting and annotating such datasets is time-consuming and resource-intensive. Additionally, image-based models often face challenges related to background noise, lighting conditions, and variations in hand orientation, which can reduce accuracy in real-time applications.

To improve efficiency, recent approaches focus on landmark-based feature extraction. Instead of processing entire images, hand landmark detection frameworks such as MediaPipe extract key points from the hand, including finger joints and palm positions. These landmarks are represented as structured numerical coordinates, which significantly reduce data complexity.

In this project, a landmark-based approach is used for ASL recognition. By extracting 21 hand landmarks and using their coordinate values as input features, the system becomes lightweight, faster, and more suitable for real-time implementation. This method reduces the need for large image datasets while maintaining reliable performance, making it an efficient alternative to traditional image-based CNN models.

7. Methodology

The methodology of the proposed ASL Detection System is divided into five major steps:

7.1 Data Collection

The dataset was collected using MediaPipe, a real-time hand tracking framework. The system captured hand gestures through a webcam and extracted 21 hand landmarks for each detected hand.

Each landmark consists of three coordinate values:

- x-coordinate
- y-coordinate
- z-coordinate

Since 21 landmarks were extracted and each landmark has 3 coordinates, a total of 63 features were generated per sample. These features were stored in CSV format along with their corresponding alphabet labels to create the training dataset.


```
df.head()
```

	label	x0	y0	z0	x1	y1	z1	x2	y2	z2	...
0	A	0.593800	0.798356	-7.649077e-07	0.503934	0.740276	-0.019193	0.444183	0.631012	-0.027275	...
1	A	0.730166	0.617582	-5.512676e-07	0.639415	0.565125	-0.029493	0.577366	0.462825	-0.037581	...
2	A	0.427747	0.586118	-8.728184e-07	0.332063	0.539541	-0.013546	0.258781	0.429579	-0.019302	...
3	A	0.468265	0.601081	-8.472105e-07	0.369813	0.546696	-0.006734	0.302968	0.442877	-0.009222	...
4	A	0.528097	0.625915	-7.491166e-07	0.433816	0.579672	-0.020089	0.362457	0.470722	-0.027540	...

5 rows × 64 columns

7.2 Data Preprocessing

After collecting the dataset, preprocessing steps were performed to prepare the data for training.

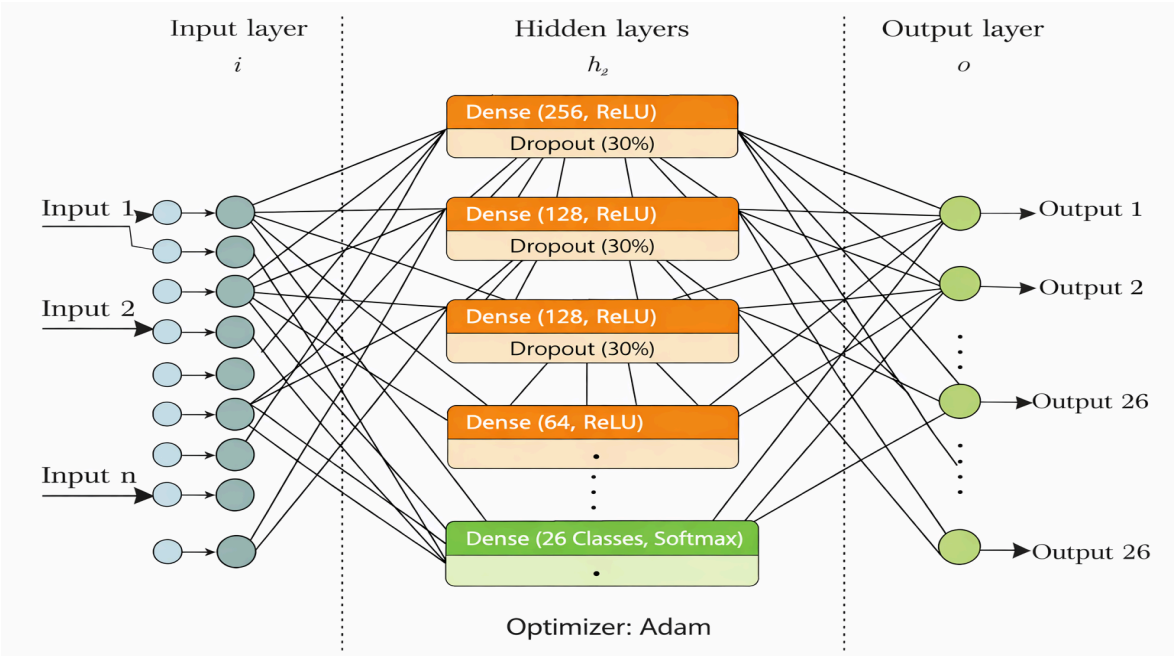
The alphabet labels were converted into numerical format using Label Encoding from the sklearn library. This ensured that the model could process categorical outputs efficiently during training.

7.3 Model Architecture

A deep learning classification model was designed to recognize ASL alphabets based on the extracted landmark features.

The model architecture consists of:

- Input Layer: 63 input features (21 landmarks × 3 coordinates)
- Hidden Layers: Fully connected dense layers
- Activation Function (Hidden Layers): ReLU
- Output Layer: 26 neurons (representing 26 ASL alphabets)
- Output Activation: Softmax
- Loss Function: Sparse Categorical Crossentropy
- Optimizer: Adam

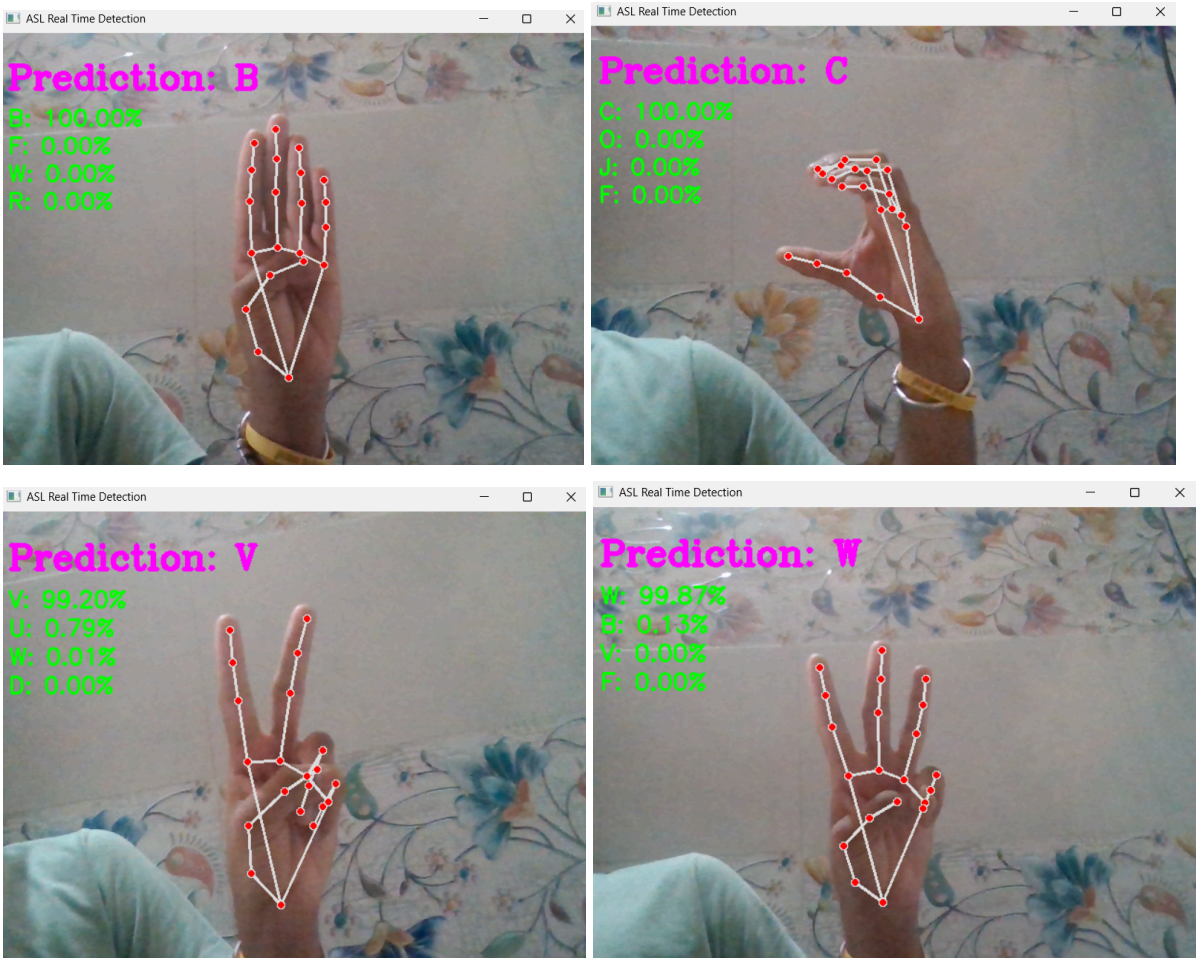


The model was trained to classify the input landmark coordinates into the correct alphabet category.

7.4 Real-Time Detection

For real-time implementation, OpenCV was used to access the webcam feed. Each video frame was processed using MediaPipe to detect hand landmarks.

The extracted 63 features were passed to the trained model, which predicted the corresponding ASL alphabet. The predicted output was displayed on the screen in real time.



7.5 Text-to-Speech Conversion

To enhance usability, a text-to-speech module was integrated into the system. The pyttsx3 library was used to convert the predicted alphabet into audible speech.

8. System Workflow

The complete workflow of the proposed Assistive Communication System is described below:

8.1 Video Capture

The system captures real-time video input using a webcam.

8.2 Hand Detection

MediaPipe detects the hand region and identifies 21 landmark points on the hand.

8.3 Feature Extraction

Each landmark consists of (x, y, z) coordinates.
Thus, 21 landmarks × 3 coordinates = 63 numerical features.

8.4 Data Preprocessing

The extracted features are converted into a structured numerical array suitable for model input.

8.5 Model Prediction

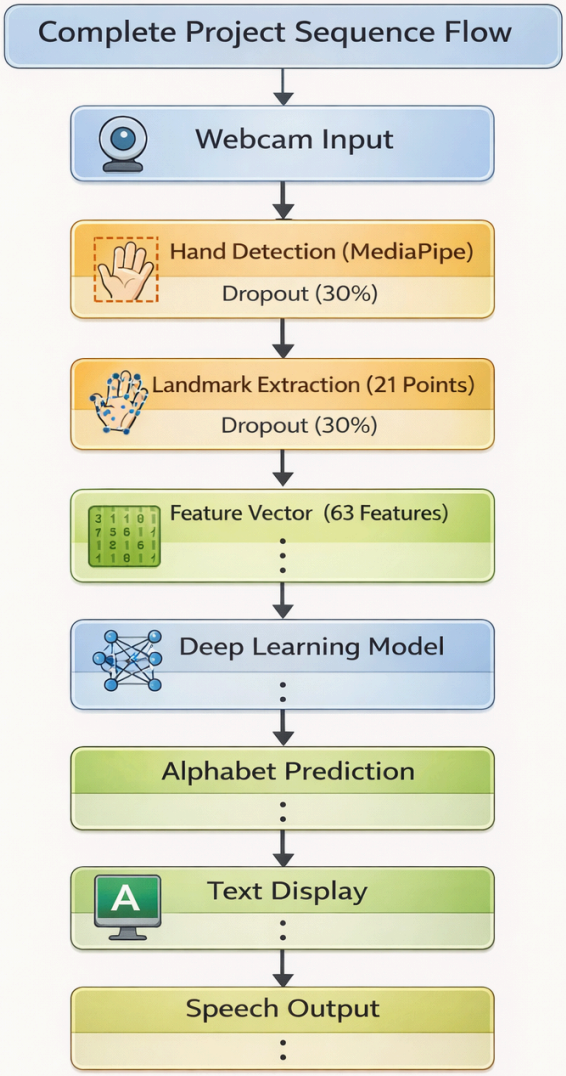
The trained deep learning model processes the 63-feature input and predicts one of the 26 ASL alphabets.

8.6 Output Display

The predicted alphabet is displayed on the screen.

8.7 Text-to-Speech Conversion

The predicted alphabet is converted into speech using the pyttsx3 library.


Complete Project Sequence Flow
Webcam Input
Hand Detection (MediaPipe)
Dropout (30%)
Landmark Extraction (21 Points)
Dropout (30%)
Feature Vector (63 Features)
Deep Learning Model
Alphabet Prediction
A
Text Display
Speech Output

9. Dataset Description

The dataset used in this project was created using hand landmark coordinates extracted through MediaPipe. Instead of storing raw images, the system records numerical landmark values, which makes the dataset lightweight and efficient.

The dataset contains the following characteristics:

- Total number of classes: 26 (A–Z alphabets)
- Total number of samples: 16,000+
- Samples per class: 500+
- Feature size per sample: 63
- Data format: CSV file
- Dataset type: Balanced

Each sample consists of 21 hand landmarks, and each landmark contains three coordinate values (x, y, z). Therefore:

21 landmarks × 3 coordinates = 63 features per sample

The dataset is balanced, meaning each alphabet class contains approximately an equal number of samples. This helps prevent bias during model training and improves overall classification performance.

The structured nature of the dataset significantly reduces storage requirements compared to image-based datasets while maintaining sufficient information for accurate classification.

10. Results

The model was trained using the prepared landmark dataset and evaluated using training, validation, and testing data splits.

The performance of the model is summarized as follows:

- Training Accuracy: High and stable during training
- Validation Accuracy: Closely matched training accuracy
- Final Test Accuracy: Above 95%

The model demonstrated consistent classification performance with over 95% accuracy, indicating reliable recognition of ASL alphabets.

A confusion matrix was generated to analyze class-wise performance. The matrix showed strong diagonal dominance, indicating correct predictions for most alphabets. Minor misclassifications occurred between visually similar gestures.

Overall, the results confirm that the proposed landmark-based deep learning model performs effectively for real-time ASL alphabet recognition.

11. Discussion

The landmark-based approach used in this project offers significant advantages over traditional image-based CNN models.

By extracting only the 21 key hand landmarks instead of processing full images, the system reduces computational complexity and storage requirements. This makes the model lightweight and faster during both training and real-time prediction.

Compared to image-based CNN systems, which require large image datasets and high processing power, the landmark-based method focuses only on essential geometric features. This improves efficiency while maintaining high classification accuracy.

The reduced feature size (63 numerical values per sample) enables quicker inference, making the system suitable for real-time applications. This approach ensures that the ASL detection system runs smoothly on standard hardware without the need for specialized GPUs or high-end devices.

Therefore, the proposed method proves to be an efficient, accurate, and practical solution for real-time sign language recognition.

---

12. Limitations

Although the proposed system achieves satisfactory performance, it has certain limitations.

- The current system supports recognition of only individual ASL alphabets (A–Z).
- It does not support word-level or sentence-level prediction.
- Dynamic gestures are not implemented in the current version.

These limitations provide opportunities for further improvement and enhancement of the system in future work.

13. Future Scope

The proposed ASL Detection System has significant potential for further development and real-world deployment. Future enhancements may include:

- Extending the system from alphabet-level recognition to word-level recognition.
- Implementing sentence formation by combining predicted alphabets into meaningful words automatically.
- Developing a mobile application version to increase accessibility and portability.
- Creating a cloud-based assistive communication platform for remote usage and integration with other applications.
- Expanding the system to support multiple sign languages for multilingual communication.

With these improvements, the system can evolve into a comprehensive assistive communication tool for the deaf and hard-of-hearing community.

14. Conclusion

This project successfully developed a real-time Assistive Communication System for the deaf using sign language detection. By utilizing hand landmark extraction and a deep learning classification model, the system is capable of recognizing ASL alphabets with over 90% accuracy.

The landmark-based approach reduced computational complexity and enabled efficient real-time performance. The integration of text-to-speech functionality further enhanced the system's usability by converting detected gestures into audible speech.

The project demonstrates that artificial intelligence and computer vision can be effectively applied to solve real-world communication challenges. Although the system currently supports only alphabet-level recognition, it provides a strong foundation for future expansion into word and sentence-level communication systems.

Overall, the proposed system is socially impactful, technically efficient, and scalable for further improvements, making it a meaningful contribution to assistive technology development.

15. References

1. MediaPipe GitHub Repository.
   https://github.com/google/mediapipe
2. MediaPipe Hands Solution Documentation.
   https://developers.google.com/mediapipe/solutions/vision/hand_landmarker
3. TensorFlow Keras API Documentation.
   https://www.tensorflow.org/api_docs/python/tf/keras
4. TensorFlow Model Training Guide.
   https://www.tensorflow.org/guide/keras/training_with_built_in_methods
5. NumPy Documentation.
   https://numpy.org/doc/
6. OpenCV-Python Tutorials.
   https://docs.opencv.org/master/d6/d00/tutorial_py_root.html
7. Scikit-learn Documentation (for preprocessing concepts).
   https://scikit-learn.org/stable/documentation.html
8. pyttsx3 Documentation (Text-to-Speech Library).
   https://pyttsx3.readthedocs.io/en/latest/
9. American Sign Language (ASL) Alphabet Reference.
   https://www.lifeprint.com/asl101/pages-signs/a/abc.htm
10. Pigou, L., Dieleman, S., Kindermans, P. J., & Schrauwen, B. (2015).
    Sign Language Recognition Using Convolutional Neural Networks.
    European Conference on Computer Vision (ECCV) Workshops.