Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

SignTalk
A Deep Learning-Based System for Real-Time Sign Language Recognition & Voice Generation

Gandu Lavanya
Malla Reddy University
Hyderabad, India

Jaligam Srilekha
Malla Reddy University
Hyderabad, India

Kare Ranjith Kumar
Malla Reddy University
Hyderabad, India

Kelim Praveen
Malla Reddy University
Hyderabad, India

Burra Manoj Kumar
Malla Reddy University
Hyderabad, India

Abstract*—Sign language is an important and expressive method of communication used by individuals who are deaf or hard of hearing. However, communication difficulties often arise when interacting with people who do not understand sign language. To overcome this issue, this research presents SignTalk AI, a real-time sign language recognition system that converts hand gestures into readable text and spoken output using artificial intelligence and computer vision techniques. The proposed system utilizes a Convolutional Neural Network (CNN) trained on the American Sign Language (ASL) Alphabet Dataset to recognize hand gestures captured through a webcam. Video frames are processed continuously, and the region containing the hand gesture is extracted and prepared before being passed to the trained model for classification. To improve prediction stability, the system uses a frame-based prediction smoothing method along with confidence threshold filtering. In addition, MediaPipe hand detection is integrated so that predictions are generated only when a hand is detected in the frame, which helps reduce noise and improves system reliability. The recognized gestures are displayed as text and can also be converted into speech using text-to-speech technology. A web application built with Flask and OpenCV provides real-time visualization of gestures, prediction tracking, sentence generation, and voice output. Experimental results show that the system performs with high recognition accuracy and stable real-time performance under normal conditions. This solution can assist speech- and hearing-impaired individuals in communication and can be extended in the future to support continuous sign interpretation and mobile-based applications.

***Index Terms*—Sign Language Recognition, Deep Learning, CNN, Computer Vision, ASL.**

I. Introduction

Human interaction strongly depends on effective communication. However, people who have hearing or speech disabilities may experience communication challenges when interacting with individuals who do not understand sign language. Sign language serves as an essential communication medium for the deaf and hard-of- hearing community, but the lack of widespread knowledge of this language creates barriers in everyday interactions.

Advances in artificial intelligence and computer vision have led to the development of automated systems capable of analyzing visual patterns… Deep learning approaches, especially Convolutional Neural Networks (CNNs), have shown strong performance in image classification and gesture recognition tasks. These technologies enable the development of systems that can identify sign language gestures and convert them into understandable information in real time. This study introduces SignTalk AI, a real-time sign language recognition system designed to translate hand gestures into text and speech. The system uses a CNN model trained on the ASL Alphabet Dataset to identify gestures captured through a webcam. Each video frame is processed to isolate the hand region, which is then analyzed by the model to determine the corresponding letter. Once the gesture is recognized, the system displays the output as text and optionally converts it into speech using text-to-speech technology. Furthermore, the system includes a web-based interface that enables users to observe real-time gesture recognition results, track predictions, and form sentences using the detected characters.

A. Background of Sign Language Communication

Sign language is a visual form of communication that conveys meaning using hand gestures, facial expressions, and body movements. American Sign Language (ASL) is widely used by the deaf community and includes gestures representing the alphabet, numbers, and words. Despite its effectiveness, communication between sign language users and non-signers can be difficult because many people are not familiar with sign language. As a result, intelligent systems that can automatically recognise sign language gestures can significantly improve accessibility and communication.

B. Challenges in Sign Language Recognition

Developing an accurate sign language recognition system involves several challenges. Variations in hand shapes, gesture

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

orientation, and lighting conditions can affect recognition of accuracy. Additionally, some gestures appear visually similar, which makes classification more complex. Real-time systems must also process video frames quickly while maintaining high prediction accuracy. These challenges require robust machine learning models and efficient image processing techniques.

C. Role of Artificial Intelligence in Gesture Recognition

Artificial intelligence, particularly deep learning, has significantly improved the ability of computers to recognize complex visual patterns. Convolutional Neural Networks automatically learn features from image data, making them highly effective for gesture recognition tasks. When trained on large datasets of labeled images, CNN models can accurately classify different hand gestures and enable real-time recognition systems using standard cameras.

D. Objectives of the Proposed System

This study aims to develop an intelligent system capable of recognizing sign language gestures in real time and converting them into text and speech. The specific objectives are:

- To develop a CNN-based model for recognizing ASL alphabet gestures.
- To implement real-time gesture recognition using webcam input.
- To improve prediction reliability using confidence filtering and stabilization techniques.
- To design a web-based interface for displaying predictions and building sentences.
- To provide text-to-speech output to assist communication.

II. LITERATURE REVIEW

Sign language recognition has attracted significant research attention due to its potential to improve communication for individuals with hearing and speech impairments. Over time, different approaches have been studied and developed, ranging from sensor-based systems to modern deep learning techniques. Earlier studies primarily relied on specialized hardware devices to capture hand movements, while recent research focuses on vision-based methods that utilize cameras and artificial intelligence algorithms for gesture recognition.

A. Traditional Methods for Sign Language Recognition

Early sign language recognition systems commonly used sensor-based technologies such as data gloves, motion sensors, and wearable devices to capture hand movements and finger positions. These systems relied on sensors to detect gesture coordinates and convert them into interpretable signals. Although sensor-based methods provided relatively accurate gesture tracking, they required specialized hardware, making them expensive and less practical for everyday use.

In addition, the requirement of wearing gloves or sensors limited user comfort and reduced the accessibility of such systems. These limitations encouraged researchers to explore alternative approaches that rely on camera-based gesture recognition, which is more convenient and cost-effective.

B. Deep Learning Approaches for Gesture Recognition

With advances in artificial intelligence, deep learning techniques have become widely used for gesture recognition tasks. In particular, Convolutional Neural Networks (CNNs) have demonstrated excellent performance in image classification and pattern recognition applications. CNN models can extract important features from images without manual intervention.

Several studies have applied CNN-based models for recognizing hand gestures in sign language datasets. These models are capable of identifying complex visual patterns such as finger positions and hand shapes with high accuracy. Deep learning approaches have significantly improved the performance of sign language recognition systems, especially when large datasets are available for training.

C. Vision-Based Sign Language Recognition Systems

Camera-based sign language recognition systems capture hand gestures using visual input devices and process them using computer vision techniques. These systems typically involve several steps, including image acquisition, preprocessing, feature extraction, and gesture classification. Recent research has integrated deep learning models with real-time video processing to achieve efficient gesture recognition.

Vision-based systems are more accessible than sensor-based systems because they do not require specialized equipment. Instead, they rely on standard cameras such as webcams or smartphone cameras. Additionally, frameworks such as OpenCV and MediaPipe have made it easier to implement real-time hand detection and gesture recognition.

D. Limitations of Existing Systems

Despite significant progress in sign language recognition, several limitations remain in existing systems. Many systems struggle with variations in lighting conditions, background noise, and hand orientations, which can affect recognition of accuracy. Additionally, some approaches focus only on static gestures and do not effectively handle continuous sign language interpretation.

Another limitation is the lack of real-time performance in some systems, especially when complex models are used. Achieving a balance between recognition accuracy and processing speed remains a challenge. Therefore, developing efficient systems that provide accurate and stable gesture recognition in real-time environments continues to be an important area of research.

III. PROPOSED SYSTEM

The proposed system, SignTalk AI, is designed to recognize sign language gestures in real time and convert them into readable text and speech. The system integrates computer vision techniques with a Convolutional Neural Network (CNN) model to identify hand gestures corresponding to the alphabet of American Sign Language (ASL). Live video frames captured through a webcam are processed to detect hand gestures and classify them using the trained deep learning model. The recognized gestures are displayed as text and can also be

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

converted into speech using text-to-speech technology. A web-based interface enables real-time visualization of predictions and sentence formation.

A. System Overview

The system captures live video input using a webcam and processes each frame to detect hand gestures. A hand detection module identifies the Region of Interest (ROI) containing the gesture, which is then resized and normalized before being passed to the CNN model. The trained model predicts the corresponding ASL alphabet along with a confidence score. To improve prediction stability, the system evaluates multiple consecutive frames before confirming the detected gesture. The predicted letters are displayed on the interface and can be combined to form words or sentences.

B. Key Features of the Proposed System

The developed system contains several key features:

- Real-Time Gesture Recognition using webcam input
- CNN-Based Classification for accurate gesture prediction
- tem Hand Detection to improve gesture localization
- Prediction Stabilization for reliable recognition results
- Web-Based Interface for real-time visualization
- Sentence Formation and Text-to-Speech Output

C. Advantages of the Proposed Approach

The proposed system offers several advantages compared to traditional methods. It eliminates the need for specialized hardware such as sensor gloves by using a camera-based approach. The integration of deep learning improves recognition accuracy, while prediction stabilization ensures consistent results. Additionally, the system provides real-time performance and includes a user-friendly interface with speech output, making it a practical assistive communication tool.

IV. SYSTEM ARCHITECTURE

The architecture of the proposed SignTalk AI system is designed to recognize sign language gestures in real time using computer vision and deep learning techniques. The system processes video input captured from a webcam and converts hand gestures into text and speech. The architecture consists of several interconnected modules, including video capture, hand detection, image preprocessing, gesture classification using a Convolutional Neural Network (CNN), and output generation.

The system operates in a sequential pipeline where each module performs a specific task to ensure accurate and efficient gesture recognition.

Initially, video frames are captured through the webcam and analyzed to detect the presence of a hand. Once the hand region is identified, the relevant portion of the frame is extracted and processed before being passed to the trained deep learning model for classification. The predicted output is then displayed on the interface and can also be converted into speech.

The proposed SignTalk AI system follows a sequential pipeline for real-time sign language recognition. Initially, video frames are captured using a webcam and processed to detect the presence of a hand using MediaPipe. Once a hand is detected, the region containing the gesture is extracted and preprocessed before being provided as input to the trained Convolutional Neural Network (CNN) model. The CNN model classifies the gesture and predicts the corresponding sign language alphabet with a confidence score. The predicted letters are then used to build words or sentences, which can be converted into speech using text-to-speech technology. This architecture enables efficient and real-time gesture recognition for assistive communication.

A. Overall System Architecture

The overall architecture of the proposed system integrates multiple components that work together to perform gesture recognition. The first step of the system involves capturing live video input through a webcam.

A hand detection module identifies whether a hand is present within the frame. When a hand is detected, the system extracts the Region of Interest (ROI) containing the gesture. The extracted gesture image is then resized and normalized to match the input requirements of the trained CNN model.

The deep learning model processes the image and predicts the corresponding American Sign Language alphabet along with a confidence score. The recognized gesture is displayed on the web interface, and the predicted letters can be combined to form words or sentences. Additionally, the system supports text-to-speech conversion to generate audible output.

B. Data Flow of the Recognition System

The data of the proposed recognition system follows a structured sequence of processing steps. First, the webcam continuously captures video frames that are passed to the hand detection module.

The system then checks whether a hand is present in the frame. If a hand is detected, the system extracts the relevant region containing the gesture.

The extracted image undergoes preprocessing steps such as resizing and normalization. The processed image is then provided as input to the CNN model, which performs gesture classification. The model generates prediction probabilities for all gesture classes, and the gesture with the highest confidence score is selected as the final prediction.

To ensure consistent predictions, the system uses a stabilization mechanism that analyzes multiple consecutive frames before confirming a gesture. Once confirmed, the predicted letter is displayed on the interface and added to the sentence builder module.

C. Real-Time Gesture Processing Pipeline

The real-time gesture processing pipeline enables the system to recognize gestures efficiently while maintaining stable predictions. The pipeline begins with capturing live video frames and detecting the hand region within each frame. After extracting the region of interest, the image is preprocessed and passed to the CNN model for classification.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

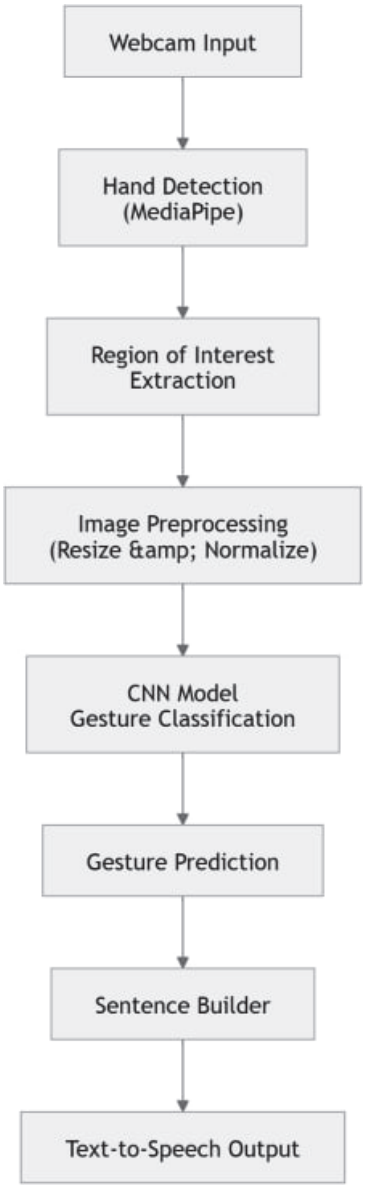


Fig. 1. System Architecture and Real-Time Processing Pipeline of SignTalk

The model predicts the corresponding gesture, and the system evaluates the confidence level of the prediction. A buffering mechanism stores predictions from several consecutive frames and selects the most frequent result to reduce noise and incorrect predictions. The final recognized gesture is then displayed on the web interface and optionally converted into speech. This pipeline ensures that the system operates efficiently in real- time environments while maintaining high recognition accuracy and stable outputs.

V. DATASET DESCRIPTION

The effectiveness of a sign language recognition system is highly influenced by the quality and variety of the dataset used for training. In this research, the American Sign Language (ASL) Alphabet Dataset is used to train the Convolutional Neural Network (CNN) model for gesture classification. The dataset contains labeled images representing different ASL alphabet gestures captured under controlled conditions. These images provide sufficient variation in hand shapes and orientations, enabling the model to learn the visual features required for accurate gesture recognition.

A. ASL Alphabet Dataset

- The ASL Alphabet Dataset consists of images representing hand gestures corresponding to the letters of the English alphabet used in American Sign Language. Each gesture represents a specific alphabet sign, allowing the system to translate hand movements into text.
- The dataset used in this study contains 29 classes, including the 26 alphabet gestures (A–Z) and three additional classes: “space,” “delete,” and “nothing.” These additional classes help the system handle real-time gesture interaction by allowing users to insert spaces between words or remove incorrect letters.
- The dataset includes thousands of images collected from different individuals, ensuring diversity in hand shapes and gesture positions. This diversity improves the model’s ability to generalize and recognize gestures from different users.

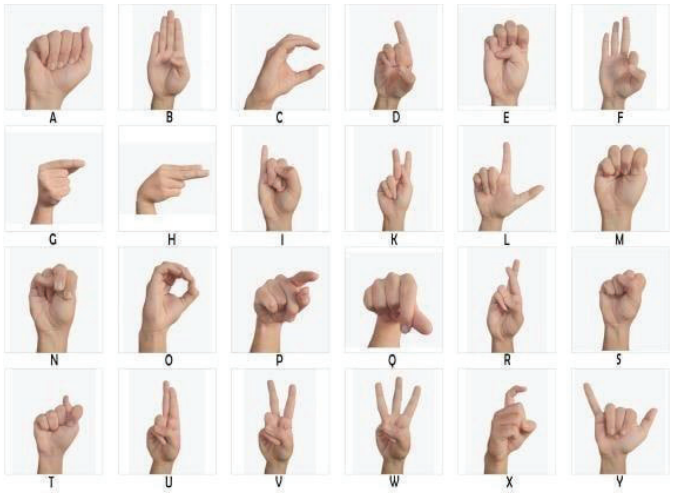


Fig. 2. Sample images from the ASL Alphabet Dataset representing different hand gesture classes.

B. Dataset Structure

The dataset is organized into separate folders for each gesture class. Each folder contains multiple images representing the corresponding ASL alphabet gesture. The images are labeled according to their gesture category, which allows the deep learning model to learn the relationship between the image features and the corresponding alphabet.

The dataset is divided into two main subsets: a training dataset and a testing dataset. The training dataset is used to train the CNN model by allowing it to learn visual patterns from labeled gesture images. The testing dataset is used to

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

evaluate the performance of the trained model and measure its ability to correctly recognize unseen gesture images. This structured organization of the dataset simplifies the process of loading and preprocessing the data during model training.

C. Data Preprocessing

Before training the CNN model, the dataset images undergo several preprocessing steps to ensure consistent input format and improve model performance. First, all images are resized to a fixed resolution of 64 × 64 pixels, which matches the input requirements of the CNN model. Resizing the images reduces computational complexity and ensures uniformity across all input samples.

Next, pixel values are normalized by scaling them to a range between 0 and 1. This normalization step improves the stability of the training process and helps the neural network learn more effectively. Additionally, the images are converted into numerical arrays so that they can be processed by the deep learning framework.

These preprocessing steps ensure that the dataset is properly prepared for training the CNN model used in the proposed sign language recognition system.

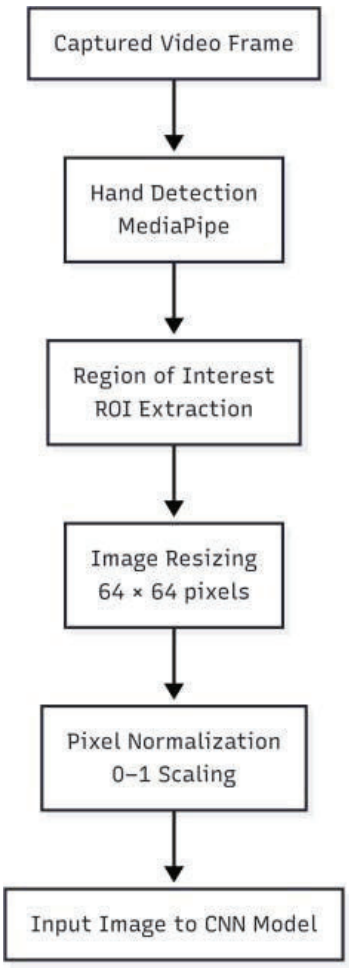


Fig. 3. Image preprocessing pipeline used to prepare gesture images before being provided as input to the CNN model.

VI. METHODOLOGY

The SignTalk AI system is developed using both computer vision techniques and deep learning approaches to recognize gestures in real time. The methodology consists of several sequential stages, including image acquisition, hand detection, gesture extraction, preprocessing, gesture classification using a Convolutional Neural Network (CNN), and prediction refinement. Each stage contributes to achieving accurate and consistent gesture recognition.

A. Image Acquisition Using Webcam

The first step of the recognition process involves capturing live video frames using a webcam. The webcam continuously records the user's hand gestures and sends each frame to the system for further processing. These frames serve as the raw input for the gesture recognition pipeline. Capturing frames in real time allows the system to detect and classify gestures dynamically as the user performs them.

B. Hand Detection Using MediaPipe

After capturing the video frame, the system utilizes MediaPipe, a computer vision framework, to detect the presence of a hand in the frame. MediaPipe identifies hand landmarks and ensures that gesture recognition is performed only when a hand is detected. This step helps eliminate unnecessary predictions and improves the reliability of the system by focusing only on frames containing hand gestures.

C. Region of Interest Extraction

Once a hand is detected, the system extracts the Region of Interest (ROI) from the captured frame. The ROI represents the area containing the hand gesture. By isolating the gesture region, the system removes irrelevant background information and focuses on the hand shape. This improves classification accuracy and reduces computational complexity.

D. Image Preprocessing

Before the gesture image is provided to the deep learning model, several preprocessing steps are applied. The extracted ROI is resized to a fixed resolution of 64 × 64 pixels, ensuring consistency across all input images. Pixel values are then normalized to a range between 0 and 1 to improve model stability and training efficiency. These preprocessing steps ensure that the input image matches the format expected by the CNN model.

E. Convolutional Neural Network Model

The core component of the system is the Convolutional Neural Network (CNN) model used for gesture classification. CNNs are well suited for image recognition tasks because they automatically learn hierarchical visual features from input images. The model analyzes the processed gesture image and predicts the corresponding sign language alphabet.

The CNN architecture consists of multiple convolutional layers for feature extraction, pooling layers for dimensionality reduction, and fully connected layers for classification. The

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

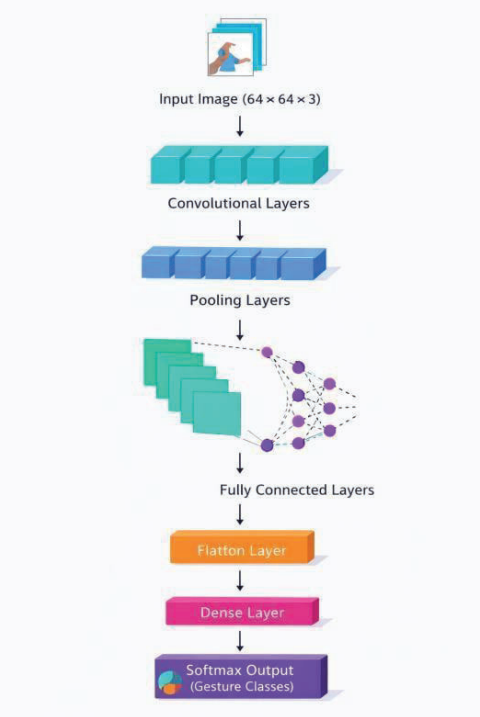
final output layer uses a softmax activation function to generate probability scores for each gesture class.The gesture with the highest score is selected as the final prediction.

F. Prediction Stabilization Technique

To improve the stability of gesture recognition, the system employs a prediction stabilization technique. Instead of relying on predictions from a single frame, the system evaluates predictions across multiple consecutive frames. A buffering mechanism stores recent predictions and selects the most frequently occurring gesture as the final output. This approach reduces noise and prevents incorrect predictions caused by sudden hand movements or temporary misclassifications.

G. Confidence Threshold Filtering

The system also applies confidence threshold filtering to ensure reliable predictions. Each prediction generated by the CNN model is associated with a confidence score. The system accepts a prediction only if the confidence value exceeds a predefined threshold. Predictions with low confidence are ignored, which helps reduce errors and improves overall recognition accuracy.



VII. IMPLEMENTATION

The proposed sign language recognition system was implemented using Python and various computer vision and deep learning libraries. The implementation consists of four major components: model training, real-time gesture recognition, web interface development, and sentence generation with text-to-speech output. These components collectively support real-time sign language interpretation.

A. Model Training Process

The gesture recognition model was trained using the ASL Alphabet Dataset, which contains images representing hand gestures corresponding to different sign language alphabets. The dataset includes multiple samples for each gesture class, enabling the model to learn variations in hand orientation and lighting conditions.

During training, the images were resized to 64 × 64 pixels and normalized to improve training efficiency. A Convolutional Neural Network (CNN) was used to learn spatial features from gesture images. The network was trained for 10 epochs, and the trained model was saved in HDF5 format (.h5) for later use in the real-time recognition system. After training, the model achieved high classification performance, enabling accurate recognition of sign language alphabets during real-time testing.

B. Real-Time Gesture Recognition Module

The real-time recognition module processes live video captured through the system webcam. Each video frame is analyzed using MediaPipe, which detects hand landmarks and confirms the presence of a hand gesture.

Once a hand is detected, a Region of Interest (ROI) is extracted from the frame. The extracted image is resized and normalized before being passed to the trained CNN model. The model predicts the corresponding gesture class and generates a confidence score indicating the reliability of the prediction.

To improve prediction stability, the system evaluates gesture predictions across multiple frames and selects the most frequent result. This approach reduces noise and improves the reliability of real- time gesture recognition.

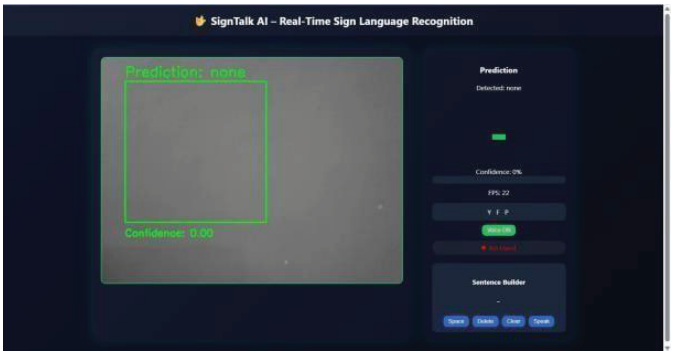


Fig. 4. Web interface displaying real-time gesture detection, prediction results, and sentence formation.

C. Web Interface Development Using Flask

A web-based user interface was developed using the Flask web framework to display the gesture recognition results. The interface provides real-time video streaming, gesture predictions, confidence scores, and system status indicators. The Flask backend processes camera frames and communicates prediction results to the front-end interface through API endpoints. The interface dynamically updates prediction

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

information without refreshing the page, enabling smooth real-time interaction.

D. Sentence Formation and Text-to-Speech Module

To make the system more practical for communication, a sentence formation module was implemented. Instead of displaying individual letters, the system collects recognized characters and gradually forms words or sentences. Users can manage the generated text using controls such as space, delete, and clear options. Once a complete word or sentence is formed, the system converts the text into speech using a text-to-speech engine. This functionality allows sign language gestures to be translated into audible speech, enabling communication between sign language users and non-sign language speakers.

VIII. EXPERIMENTAL RESULTS AND DISCUSSION

This section presents the experimental results obtained from training the proposed CNN model and evaluating its performance in real-time sign language recognition. The evaluation includes training performance, real- time prediction behavior, accuracy analysis, and overall system performance.

A. Model Training Performance

The Convolutional Neural Network (CNN) model was trained using the ASL Alphabet Dataset, which contains labeled images representing different American Sign Language hand gestures. During training, the dataset was divided into training and validation sets to evaluate the model's learning capability.

The model was trained for 10 epochs, allowing it to learn discriminative features of different gesture classes. Image preprocessing techniques such as resizing and normalization were applied to ensure consistent input dimensions and improve training stability.

The training results demonstrated strong model performance, achieving a training accuracy of 99.21% with a loss value of 0.0267. These results indicate that the CNN model successfully learned meaningful gesture patterns from the dataset.

B. Real-Time Recognition Performance

After training, the CNN model was integrated with a real-time gesture recognition pipeline using OpenCV and MediaPipe. The system captures video frames from the webcam, detects the presence of a hand, extracts the gesture region, and performs classification using the trained model.

The system is capable of recognizing gestures in real time with minimal delay. The predicted gesture along with its confidence score is displayed on the interface.

C. Accuracy Analysis

The high training accuracy achieved by the CNN model indicates that the model effectively learns the visual patterns associated with different sign language gestures. However, real-world performance can vary depending on environmental factors such as lighting conditions, background complexity, and hand positioning.

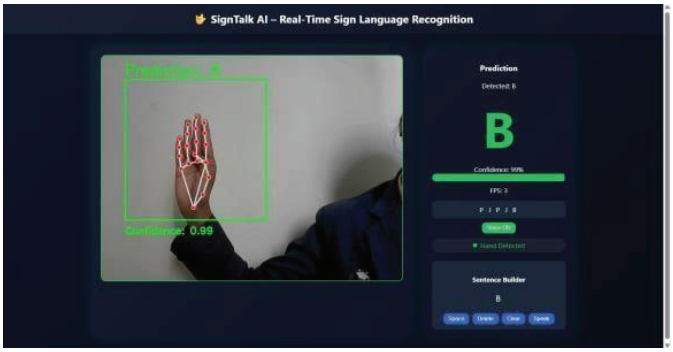


Fig. 5. Real-time gesture recognition showing the detected hand gesture and predicted alphabet.

To improve recognition reliability, the system incorporates additional mechanisms such as prediction stabilization and confidence threshold filtering. These techniques help reduce incorrect predictions and improve consistency across consecutive frames.

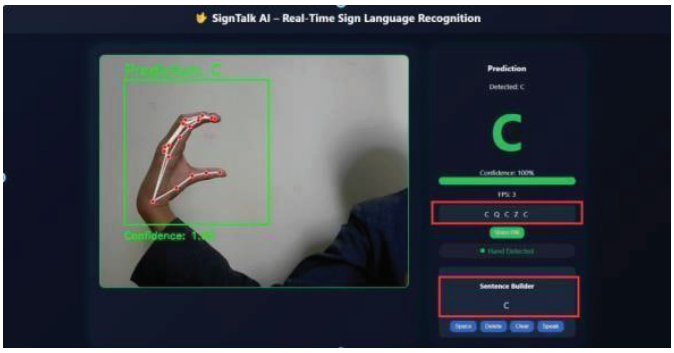


Fig. 6. Example of prediction stabilization where multiple frame predictions are used to determine the final gesture output.

D. System Performance Evaluation

The overall system performance was evaluated based on usability and responsiveness in real-time conditions. The proposed system processes video frames continuously and performs gesture recognition with minimal latency, enabling smooth user interaction.

The web interface provides additional functionalities such as sentence formation and text- to-speech conversion, allowing users to combine recognized gestures into meaningful text and convert them into audible speech.

IX. ADVANTAGES

The proposed SignTalk AI system offers several advantages compared to traditional sign language recognition approaches. One of the primary advantages is real-time gesture recognition, which allows the system to interpret sign language gestures instantly using live webcam input. This enables smooth and

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

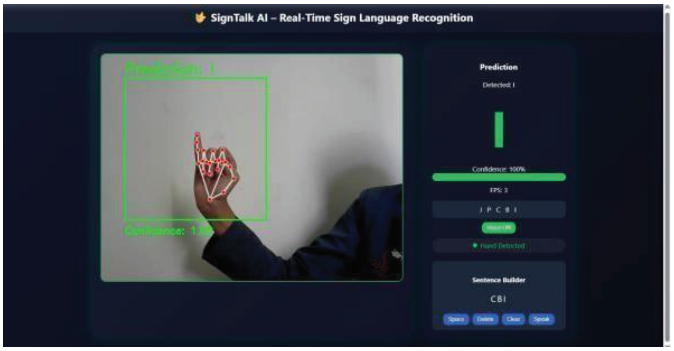


Fig. 7. Web interface of the proposed system showing gesture recognition results and sentence formation modules.

interactive communication without the need for pre-recorded videos.

Another important advantage is the use of deep learning using a Convolutional Neural Network (CNN). The CNN model automatically learns meaningful visual features from gesture images, which improves classification accuracy and reduces the need for manual feature extraction.

The system also utilizes MediaPipe hand detection, ensuring that gesture recognition only occurs when a hand is visible in the frame. This minimizes incorrect predictions and enhances the reliability of the recognition process.

Additionally, the system includes a sentence formation module, which allows individual recognized letters to be combined into meaningful words and sentences.. This greatly increases the system's usability for actual communication.

Finally, the system provides voice output through text-to-speech conversion, translating gestures into spoken language. This feature helps bridge the communication gap between sign language users and those who do not understand sign language.

FUTURE WORK

Although the SignTalk AI system shows promising results for real-time sign language recognition, there are several areas that could be improved in future research.

One possible enhancement is word-level or sentence-level recognition, where the system identifies complete words or phrases instead of individual gestures. This would significantly improve the speed and usability of communication.

Another potential development is the creation of a mobile application that runs the system on smartphones using optimized deep learning models such as TensorFlow Lite. This would increase the system's portability and make it more accessible for everyday use.

Finally, the system provides voice output through text-to-speech conversion, allowing gestures to be translated into spoken language. This feature helps bridge the communication gap between sign language users and those who do not understand sign language.

CONCLUSION

This paper introduces SignTalk AI, a real-time sign language recognition system that combines computer vision and deep learning methods to convert hand gestures into text and speech.The system uses a Convolutional Neural Network (CNN) trained on the ASL Alphabet Dataset to accurately classify sign language gestures captured via a webcam.

The proposed system integrates MediaPipe-based hand detection, region of interest extraction, image preprocessing, and CNN-based gesture classification to enable real-time gesture recognition. Additional improvements such as prediction stabilization, confidence threshold filtering, and sentence formation enhance the system's reliability and usability.

Experimental results show that the model achieves high training accuracy and effective real-time performance, enabling seamless gesture recognition and communication assistance.. The developed web interface further improves usability by displaying prediction results, confidence scores, and allowing users to create sentences that can be converted into speech.

Overall, the proposed system offers a practical and efficient solution for improving communication between sign language users and individuals who do not understand sign language.

REFERENCES

[1] Starner, T., Weaver, J., & Pentland, A. (1998). Real-time American Sign Language recognition using desk and wearable computer-based video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12), 1371–1375.

[2] Pigou, L., Dieleman, S., Kindermans, P., & Schrauwen, B. (2015). Sign language recognition using convolutional neural networks. *European Conference on Computer Vision Workshops*, 511–517.

[3] Molchanov, P., Gupta, S., Kim, K., & Kautz, J. (2015). Hand gesture recognition with 3D convolutional neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 284–293.

[4] Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations (ICLR)*.

[5] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

[6] Zhang, F., Bazarevsky, V., Vakunov, A., Tkachenka, A., Sung, G., Chang, C., & Grundmann, M. (2020). MediaPipe Hands: On-device real-time hand tracking. *arXiv preprint arXiv:2006.10214*.

[7] Oyedotun, O. K., & Khashman, A. (2017). Deep learning in vision-based static hand gesture recognition. *Neural Computing and Applications*, 28(12), 3941–3951.

[8] Jalal, A., Quaid, S., & Kim, K. (2018). A Survey on Sensor-based and Vision-based Human Activity Recognition: Challenges and Opportunities. *International Conference on Applied Sciences and Technology (IBCAST)*, 521–526.

[9] Das, A., Gawde, S., Suratwala, K., & Kalbande, D. (2019). Sign Language Recognition Using Deep Learning on Custom Dataset. *International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 121–126.

[10] Mavi, A. (2021). ASL Alphabet Dataset: A dataset for American Sign Language Alphabet Recognition. *arXiv preprint arXiv:2112.00034*.