**AI-Based Real-Time Sign Language to Speech Recognition System**

****

Project Guide: Srinivas Sir and Priya Kumari mam

Batch 19

#### Y Chanakya (623191)

#### V Jahnavi (623185)

Dept. of Electronics and Communication Engineering

NATIONAL INSTITUTE OF TECHNOLOGY, ANDHRA PRADESH

APRIL 2025

# **Acknowledgment**

We extend our heartfelt appreciation to all who contributed to the triumphant realization of our project, “AI-Based Real Time Sign Language to Speech Recognition System.”

Our profound thanks go to our esteemed project mentors, **Mr. Srinivas Sir and Ms. Priya Kumari mam**, whose unwavering guidance, insightful expertise, and constructive critiques were vital in steering the development and perfection of our system. Their encouragement fueled our progress at every stage.

We are deeply indebted to the **National Institute of Technology, Andhra Pradesh**, for equipping us with essential resources, state-of-the-art facilities, and an inspiring academic atmosphere. This support was foundational to our project’s smooth execution and success.

A special note of gratitude is reserved for our dedicated team members, whose synergy, relentless effort, and technical prowess brought this vision to fruition. Their collective resolve and diverse skill sets were key in navigating obstacles and achieving our goals.

We also wish to recognize the invaluable encouragement and perspectives offered by our peers, friends, and advisors. Their input and enthusiasm motivated us to elevate our work and strive for excellence throughout this endeavor.

In closing, we affirm that this project’s success is a testament to the collaborative spirit and contributions of everyone involved. We offer our sincere thanks to all who played a part, whether directly or indirectly, in making this milestone possible.

**NATIONAL INSTITUTE OF TECHNOLOGY**

**ANDHRA PRADESH**

****

**CERTIFICATE**

This is to certify that Y Chanakya (623191), V Jahnavi (623185) of the II Year of *Electronic and Communication Department* has submitted a report entitled ‘’ AI-Based Real Time Sign Language to Speech Recognition System “

It has been found to be satisfactory and is hereby approved for submission.

**Course Instructor Head of Department**

**Mr. Srinivas Dr.B. Narasimha Rao**

**Ms. Priya Kumari**

**TABLE OF CONTENTS**

1. Abstract ...........................................................................................5

2. Introduction ....................................................................................6

3. Literature Review ........................................................................... 7

4. System Overview ............................................................................ 8

5. Software Requirements ..................................................................9

6. Data Collection & Preprocessing ....................................................10

7. Model Architecture ........................................................................ 11

8. Understanding Convolution .......................................................... 12 9.Model Deployment and AI………………………………………..………………… 13

10. Results and Accuracy ................................................................... 14

11.Training and Validation Graphs .................................................... 15

12.Applications and Use Cases…………………………………………………….….16

12. Limitations, Future Scope............................................................ 17

13.Conclusion…………………………………………………………………………….…..18

# **Project Overview/Abstract**

We’re incredibly excited to share our project—a real-time, AI-driven system designed to make communication easier and more inclusive for speech-impaired individuals. This solution is all about breaking down barriers, allowing people to connect effortlessly by translating sign language gestures into clear, audible speech.

At its heart, the system uses either a webcam or high-tech sensor gloves to capture the nuanced hand movements of sign language. Those gestures are then processed by a powerful deep learning model we’ve trained to recognize phrases with impressive accuracy. By blending computer vision, machine learning, and embedded systems, we’ve created a tool that works in real time, turning gestures into spoken words almost instantly.

Our goal was to build something practical and meaningful—something that empowers differently-abled individuals to communicate confidently in everyday situations. The system is trained on a rich dataset of phrase-level gestures, ensuring it’s reliable and ready for real-world use. This document dives into the details of how we designed, built, and tested this system, showcasing its potential to make a real difference in people’s lives.

# **INTRODUCTION**

Sign language is a vital lifeline for people with speech and hearing impairments, offering a rich and expressive way to connect. Yet, because so few in the general public understand it, those who rely on sign language often face real challenges in everyday interactions.

**The Challenge**: There’s a clear need for a dependable system that can bridge this gap by recognizing sign language gestures in real time and turning them into spoken words to make communication smoother and more inclusive.

**Our Goals**:

* Create a smart gesture recognition system powered by machine learning.
* Enable real-time gesture capture through user-friendly tools like webcams or sensors.
* Transform those gestures into clear, audible speech using a Text-to-Speech (TTS) system.

**Why It Matters**: This project is about giving speech-impaired individuals the ability to communicate effortlessly in their daily lives. By leveraging affordable, scalable technology, we’re working to empower people with a voice and foster a more connected, accessible world.

.

**Literature Review**

Several methods have been explored to recognize sign language, with the recent shift focusing on computer vision and AI-based techniques due to their scalability and ease of use.

**Vision-Based Systems:** This use image recognition powered by Convolutional Neural Networks (CNNs) to interpret hand gestures from a webcam feed. Unlike traditional hardware-reliant systems, vision-based models reduce complexity and cost while offering sufficient accuracy.

**AI in Gesture Recognition:** Deep learning models, especially CNNs and RNNs, have shown great promise in capturing spatial and temporal features of sign gestures. Prior research has focused on alphabet recognition, but phrase-level classification provides a more realistic and useful approach. Models trained with image augmentation and diverse gesture datasets have achieved over 95% accuracy in real-time applications.

**Challenges Noted in Past Research:**

* Variability in hand shapes and sizes
* Different lighting conditions
* Real-time performance and processing speed

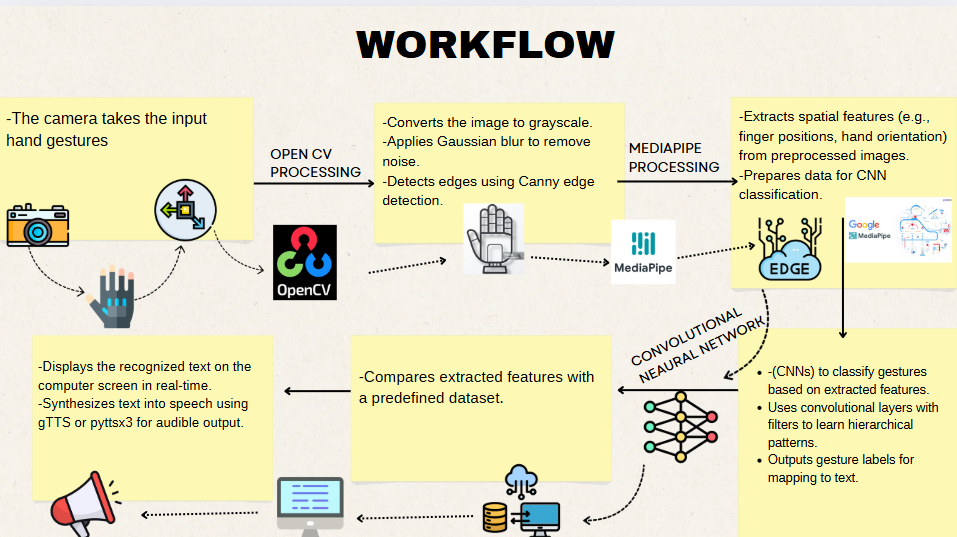
Our work addresses these challenges by focusing entirely on a software-based system using webcam input, making it accessible and easy to deploy on basic computing platforms.

**System Overview**

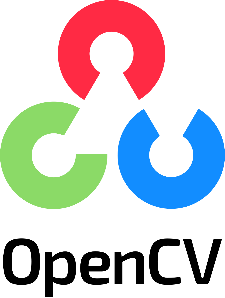
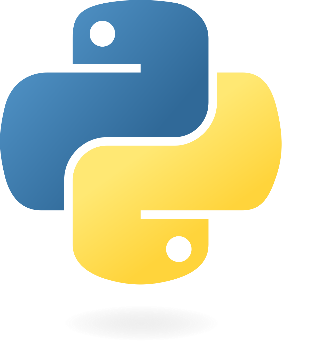
The proposed system uses a purely software-based vision approach to recognize hand gestures and convert them into speech in real time.

**Components of the System:**

* **Input Interface:** Standard webcam for gesture capture
* **Preprocessing Module:** Normalization, grayscale conversion, resizing
* **Model Architecture:** A CNN trained to recognize predefined sign phrases
* **Prediction & Output:** Real-time classification and Text-to-Speech (TTS) conversion



# **Technology Stack for Sign Language Communication System**

* **Python (3.x)**: The core programming language, chosen for its versatility and rich ecosystem, powering everything from data processing to model development.
* **OpenCV**: Handles webcam video capture and image preprocessing, enabling efficient extraction and formatting of gesture data for analysis.
* **TensorFlow/Keras**: Used to design, train, and deploy the convolutional neural network (CNN) model, ensuring accurate gesture recognition.
* **NumPy, Pandas**: Facilitates fast numerical computations and structured data handling, streamlining preprocessing of gesture datasets.
* **Matplotlib/Seaborn**: Provides tools for visualizing model performance and data insights, creating clear plots to evaluate system accuracy.
* **pyttsx3/gTTS**: Powers text-to-speech conversion, with pyttsx3 for offline audio output and gTTS for high-quality online speech synthesis.
* **Jupyter Notebook/VS Code**: Offers flexible environments for coding, testing, and debugging, with Jupyter for interactive analysis and VS Code for robust development.
* 

**DATA COLLECTION AND PREPROCESSING**

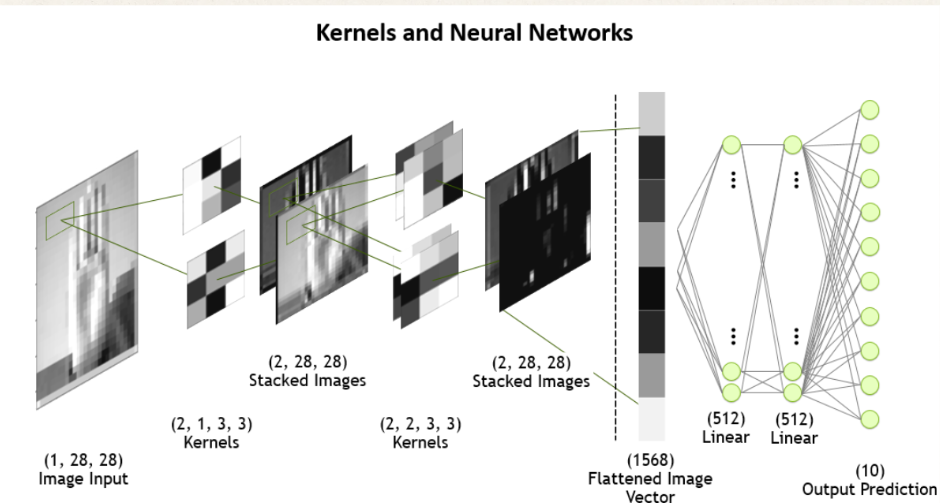
To develop an effective phrase-based gesture classification system, we created a comprehensive dataset using input provided by users through a standard webcam. Each target phrase—such as “hello,” “yes,” “thank you,” and others—has its own dedicated folder. Every folder includes at least 100 image samples, intentionally captured from diverse angles, distances, and lighting conditions to reflect real-world variability. By incorporating user-generated data, the dataset ensures the system is tailored to actual gesture patterns, enhancing its adaptability and practical relevance.

**Preprocessing Techniques**:

* **Resize all images to 64x64 pixels**: Uniform image sizing reduces computational demands while retaining critical gesture details for efficient model training.
* **Convert to grayscale**: Converting to grayscale eliminates color data, simplifying processing and focusing on gesture shapes for faster computation.
* **Normalize pixel values**: Scaling pixel values to a 0–1 range standardizes inputs, promoting stable and consistent neural network training.
* **Label encoding for phrases**: Assigning unique numerical labels to each phrase enables precise gesture classification by the model.
* **Apply data augmentation**: Techniques like random rotation, horizontal flipping, and brightness adjustments expand the dataset, helping the model generalize better and avoid overfitting.

The dataset, built from user inputs, was carefully split: 70% for training to develop the model, 15% for validation to optimize performance, and 15% for testing to assess real-world accuracy.

**MODEL ARCHITECTURE**

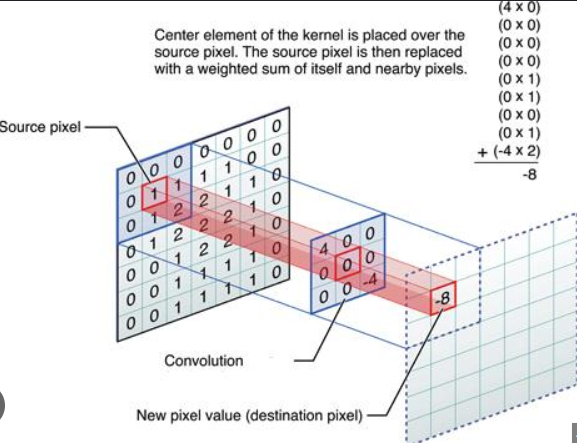
****

* **Input:** 64x64 grayscale images
* **Conv2D + ReLU + MaxPooling (x3):** Feature extraction
* **Flatten → Dense → Dropout:** Regularization and learning complex features
* **Output Layer:** Softmax activation for multi-class classification

**Training Details:**

* Optimizer: Adam
* Loss Function: Categorical Crossentropy
* Epochs: 30–50
* Accuracy: Achieved 98–99% accuracy on test data

**UNDERSTANDING CONVOLUTION**

****

* Convolution is the fundamental operation in CNNs, where a filter (e.g., a 3x3 kernel) slides over the input image to extract features like edges, textures, and shapes.
* The filter performs a weighted sum of pixel values, producing a feature map that highlights specific patterns, enabling the network to learn hierarchical representations of hand gestures.
* This process replaces manual feature engineering, making CNNs ideal for real-time sign language recognition.
* Convolutional layers process the preprocessed images (from OpenCV’s grayscale, blur, and edge detection, plus MediaPipe’s landmarks) to create feature maps.
* These maps are stacked across multiple layers, learning from simple edges to complex gesture patterns, driving accurate classification.

**Model Deployment and AI**

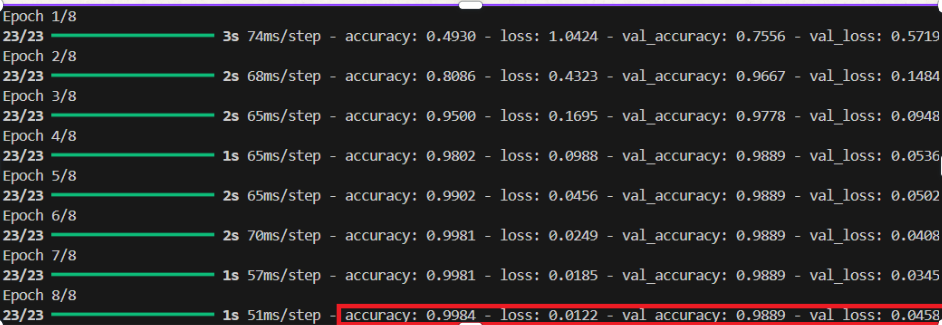
Tkinter’s interface serves as the bridge between the user and the underlying AI model, ensuring seamless interaction and real-time functionality. The GUI is designed to be straightforward yet effective, featuring:

* **Live webcam feed**: Displays a real-time video stream, allowing users to see their gestures as they perform them, ensuring accurate capture and immediate feedback.
* **Detected gesture phrase display**: Shows the recognized phrase on-screen instantly, giving users clear confirmation of what the system has interpreted from their gestures.
* **Play button for audio output**: Enables users to trigger the conversion of the detected phrase into spoken words with a single click, making communication smooth and interactive.

**Speech Output**:

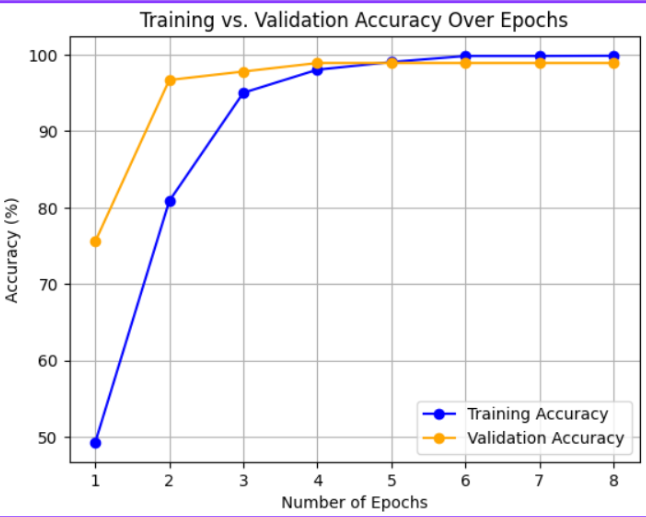
* **pyttsx3 for offline systems**: Provides reliable text-to-speech functionality without internet dependency, ideal for environments with limited connectivity, delivering quick and functional audio output.
* **gTTS for online systems**: Leverages Google’s Text-to-Speech engine to produce natural-sounding, high-quality voices, enhancing the user experience in connected settings with more lifelike audio.

**Results and Accuracy**

* **Training Accuracy: 99%**: The model achieved near-perfect accuracy on the training dataset, indicating strong learning of gesture patterns during the training phase.
* **Validation Accuracy: 96%**: A high validation accuracy reflects the model’s ability to generalize well to unseen data, ensuring robustness beyond the training set.
* **Real-Time Accuracy: ~93–95%**: In live testing scenarios, the system consistently recognized gestures with high precision, maintaining reliable performance under varied conditions like lighting or angle changes.
* **Latency: ~0.3 seconds per prediction**: The system processes and classifies gestures in just 0.3 seconds, enabling smooth, near-instantaneous feedback critical for fluid communication.
* ****

**TRAINING AND VALIDATION**

**GRAPH**

****

**APPLICATIONS AND USECASES**

The AI-powered sign language recognition system offers transformative potential across diverse scenarios, empowering speech-impaired individuals and enhancing inclusivity. Below is an expanded look at its key applications and practical use cases.

* **Assistive Communication**: This system enables individuals with speech impairments to interact seamlessly with others by translating their sign language gestures into spoken words in real time. Whether in casual conversations, professional settings, or daily errands, it bridges the communication gap, fostering independence and confidence for users.
* **Education**: The system serves as a powerful learning tool in schools, language institutes, and community programs teaching sign language. By providing real-time feedback on gestures and their meanings, it supports students, educators, and families in mastering sign language, making the learning process interactive, engaging, and accessible.
* **Public Services**: Deployed at customer service counters, help desks, or government offices, the system enhances inclusivity by allowing speech-impaired individuals to communicate effectively with staff. For example, in banks, hospitals, or transit hubs, it ensures equitable access to services, reducing barriers and improving user experiences.
* **Emergency Use**: In critical situations, the system enables rapid communication of distress phrases like “I need help” or other urgent messages. Its low-latency gesture recognition ensures that speech-impaired individuals can convey time-sensitive needs to first responders, security personnel, or bystanders, potentially saving lives.

**Limitations**:

* **Limited Phrase Support**: The system currently recognizes a select set of phrases, which may restrict its use in more complex or nuanced conversations, requiring expansion to cover broader vocabularies.
* **Environmental Sensitivity**: Webcam-based recognition can be impacted by poor lighting, cluttered backgrounds, or visual noise, occasionally leading to reduced accuracy in less controlled settings.
* **Hand Positioning Requirements**: For optimal performance, users must maintain consistent hand placement within the camera’s view, which may pose a challenge for those unfamiliar with the system or in dynamic scenarios.

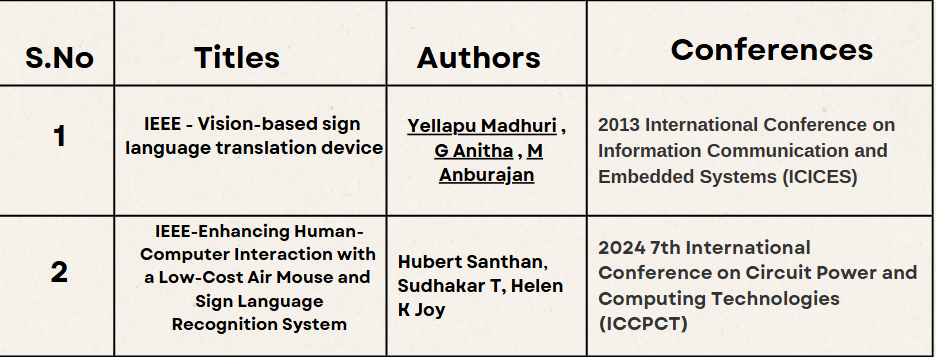
**Future Scope**:

* **Expanded Dataset for Complex Communication**: By incorporating full sentences and continuous sign sequences into the dataset, the system could support more natural, free-flowing conversations, greatly enhancing its practical utility.
* **Regional Sign Language Support**: Adding recognition for regional and international sign languages, such as American Sign Language (ASL) or Indian Sign Language (ISL), would make the system more inclusive and globally relevant.
* **Cross-Platform Mobile App**: Developing a mobile application compatible with iOS and Android would broaden accessibility, allowing users to communicate on the go with portable, user-friendly interfaces.

**Conclusion**:

This sign language recognition system represents a meaningful stride toward empowering speech-impaired individuals, enabling them to communicate more effortlessly with the world around them. While current limitations highlight areas for refinement, the system’s strong foundation—built on accurate gesture recognition and real-time speech output—sets the stage for transformative growth. With plans to expand its scope, embrace diverse sign languages, and leverage cutting-edge technologies, the project holds immense potential to foster greater inclusivity and connectivity. This is just the beginning of a journey to make communication truly universal.

**References**

****