

# Real-Time Multi-Hand Gesture Recognition and Control System Using MediaPipe and Machine Learning

B Benny Joseph<sup>1</sup>, Dr. A. Damodaram<sup>2</sup>

<sup>1</sup>M.Tech Scholar, Dept. of CSE, JNTUH UCESTH, Hyderabad, India

<sup>2</sup>Senior Professor, Dept. of CSE, JNTUH UCESTH, Hyderabad, India

\*\*\*

**Abstract** - Human-computer interaction is evolving toward more natural and contactless methods of communication. Hand gesture recognition has emerged as an effective technology for enabling natural and contactless human-computer interaction. This paper presents a real-time multi-hand gesture recognition and control system using MediaPipe and Support Vector Machine (SVM). The proposed framework detects and tracks hand landmarks from live video streams and classifies gestures based on the extracted features. A custom dataset was created by capturing various hand gestures under different lighting conditions, backgrounds, and hand orientations to improve the robustness and generalization of the model. The system supports both single-hand and multi-hand gestures and translates recognized gestures into system control commands. These commands include mouse movement, left click, right click, scrolling, volume control, media playback control, muting, and system locking. MediaPipe is utilized for efficient hand landmark detection, while the SVM classifier is employed to accurately identify gesture patterns in real time. The integration of gesture recognition with system control provides a convenient and intuitive hands-free user experience. Experimental results demonstrate that the proposed system achieves high recognition accuracy and low response time across different operating conditions. The framework is simple, cost-effective, and suitable for practical applications in smart automation, assistive technologies, robotics, and virtual environments. The study highlights the effectiveness of combining MediaPipe and SVM for developing reliable and efficient real-time multi-hand gesture recognition systems.

**Key Words:** Gesture Recognition, Multi-Hand Tracking, MediaPipe, Machine Learning, Support Vector Machine (SVM), Human-Computer Interaction, OpenCV, Smart Automation.

## 1. INTRODUCTION

The rapid growth of computer vision and artificial intelligence has significantly improved the way humans interact with digital systems. Traditional input devices such as keyboards, mice, and touchscreens require physical contact, which may not always be convenient or efficient. As a result, gesture recognition has emerged as an effective alternative for enabling natural and contactless human-computer interaction. By interpreting hand movements and gestures, computers can understand

user intentions and execute corresponding actions in real time.

Hand gesture recognition has found applications in various domains, including smart home automation, robotics, virtual reality, gaming, healthcare, and assistive technologies. Most existing systems focus on recognizing gestures performed by a single hand. Although these approaches provide acceptable performance, they offer limited interaction capabilities and cannot effectively support complex commands that require the coordinated use of both hands.

To address these limitations, this paper proposes a real-time multi-hand gesture recognition and control system using MediaPipe and machine learning techniques. The system detects and tracks landmarks from both hands simultaneously and extracts meaningful features for gesture classification. A custom dataset is used to train the classification model, enabling accurate recognition of multiple gesture combinations. The recognized gestures are then mapped to control commands for various applications.

The proposed framework is implemented using Python, OpenCV, MediaPipe, and Scikit-learn. Experimental results demonstrate high recognition accuracy and real-time performance under different operating conditions. The developed system provides a flexible, efficient, and user-friendly solution for gesture-based control, making it suitable for next-generation interactive and intelligent systems.

## 2. LITERATURE SURVEY

Gesture recognition has become an active area of research due to its importance in creating natural and intuitive human-computer interaction systems. Several researchers have proposed different techniques for recognizing hand gestures using computer vision and machine learning algorithms.

Early gesture recognition systems primarily relied on wearable devices such as data gloves and sensors to capture hand movements. Although these methods provided accurate results, they were often expensive, uncomfortable, and unsuitable for everyday use. With the advancement of computer vision, camera-based gesture

recognition systems gained popularity because they eliminated the need for specialized hardware.

Recent studies have utilized deep learning and machine learning techniques for hand gesture recognition. Convolutional Neural Networks (CNNs) have been widely used for extracting visual features from hand images and achieving high classification accuracy. However, deep learning models generally require large datasets and significant computational resources. To overcome these challenges, researchers have adopted MediaPipe, a lightweight framework capable of real-time hand landmark detection and tracking. Combined with machine learning algorithms such as Support Vector Machines (SVM), Random Forests, and K-Nearest Neighbors (KNN), MediaPipe-based systems have demonstrated efficient and accurate gesture recognition.

Despite these advancements, most existing systems focus only on single-hand static gestures and provide limited interaction capabilities. They often struggle with recognizing simultaneous gestures from both hands and may experience performance degradation under varying environmental conditions. Furthermore, many systems lack gesture-based control functionalities for practical applications.

The proposed work addresses these limitations by introducing a real-time multi-hand gesture recognition and control framework that supports dual-hand interaction, improved gesture classification, and enhanced usability across diverse application domains.

### 3. PROBLEM STATEMENT

Gesture-based interaction systems have gained significant attention as an alternative to traditional input devices such as keyboards, mice, and touchscreens. However, most existing gesture recognition systems are designed to recognize only single-hand gestures, limiting the range of commands and interactions that can be performed. These systems often struggle to interpret complex gestures involving both hands, thereby reducing their effectiveness in real-world applications.

Additionally, many existing solutions focus solely on gesture recognition without providing direct control functionalities for devices or applications. Their performance may also be affected by variations in lighting conditions, background complexity, hand orientation, and user differences. Such limitations reduce the reliability and usability of gesture-based interfaces in practical environments.

There is a need for an efficient and robust system capable of recognizing multiple hand gestures simultaneously while maintaining high accuracy and real-time

performance. The system should be able to process gestures from both hands, classify them accurately, and translate them into meaningful control commands for applications such as smart device automation, robotics, virtual environments, and assistive technologies.

Therefore, the problem addressed in this work is the development of a real-time multi-hand gesture recognition and control framework using MediaPipe and machine learning techniques. The proposed system aims to overcome the limitations of existing single-hand approaches by enabling dual-hand gesture recognition, improving interaction capabilities, enhancing recognition accuracy, and providing an effective hands-free control mechanism suitable for various real-world applications.

### 3.1 System Architecture

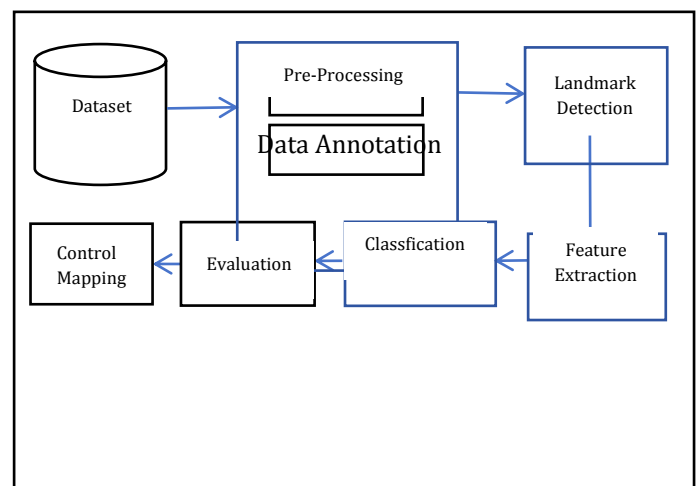


Fig 1: System Architecture

### 3.2 Dataset Description

The dataset used in this work was created by capturing real-time hand gesture images and videos using a webcam. It includes both single-hand and multi-hand gestures collected under different lighting conditions, backgrounds, and hand orientations to improve recognition accuracy and system robustness. Gesture classes such as Thumbs Up, Bird, OK, Not OK, Pointing Up, Pointing Down, Look, Power, Luck, Pointing Right, Pointng Left, Stop, peace and various two-hand gesture combinations were included in the dataset.

In addition to gesture recognition, the system supports gesture-based computer control operations. The collected dataset contains gestures mapped to actions such as Index Finger – Move Mouse, Two Fingers – Left Click, Three Fingers – Right Click, Four Fingers – Scroll Up, Open Palm – Scroll Down, Thumb – Volume Up, Pinky – Volume Down, Thumb + Index – Play/Pause, Thumb + Pinky – Mute, and

Fist – Lock PC. The labeled gesture samples were used to train and test the machine learning model, enabling accurate real-time recognition and system control.

architecture ensures efficient processing, reliable gesture recognition, and seamless human–computer interaction.

### 3.4 Results

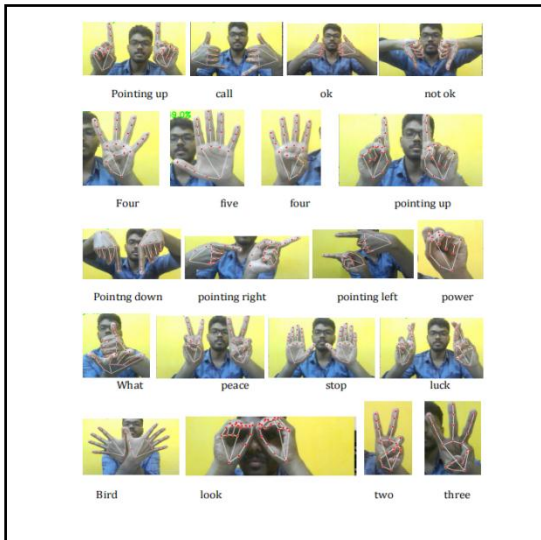


Fig -2: Dataset

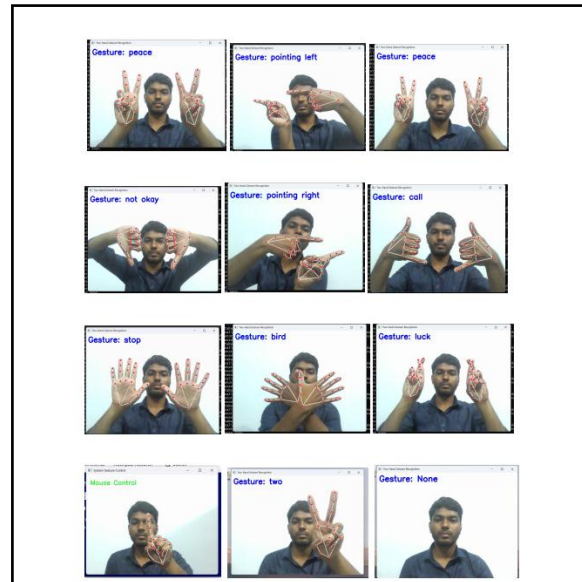


Fig - 3: Results

### 3.3 Module Building

The proposed multi-hand gesture recognition and control system is divided into several functional modules that work together to achieve accurate real-time gesture detection and system control. The Data Acquisition Module captures live video streams using a webcam and collects both single-hand and multi-hand gesture samples for training and testing purposes. The captured frames are then passed to the Pre-processing Module, where OpenCV techniques are applied for image enhancement, noise reduction, and normalization to improve the quality and consistency of the input data.

The processed frames are forwarded to the Feature Extraction Module, where MediaPipe detects hand landmarks and extracts key joint coordinates from each hand. These landmark features are used by the Gesture Classification Module, which employs a machine learning algorithm such as Support Vector Machine (SVM) to identify and classify different hand gestures in real time. Once a gesture is recognized, the Control and Integration Module maps it to a predefined system command. The framework supports various operations, including mouse movement, left click, right click, scrolling, volume control, media playback control, muting, and system locking through intuitive hand gestures.

Finally, the Performance Evaluation Module assesses the system based on recognition accuracy, response time, and robustness under varying lighting conditions, backgrounds, and hand orientations. This modular

Table 1: Performance Evaluation of the Proposed Multi-Hand Gesture Recognition System

Performance Metric	Value
Training Accuracy	98.4%
Testing Accuracy	96.8%
Precision	96.5%
Recall	96.2%
F1-Score	96.3%
Average Response Time	0.12 sec
Gesture Detection Rate	97.1%
False Positive Rate	2.4%

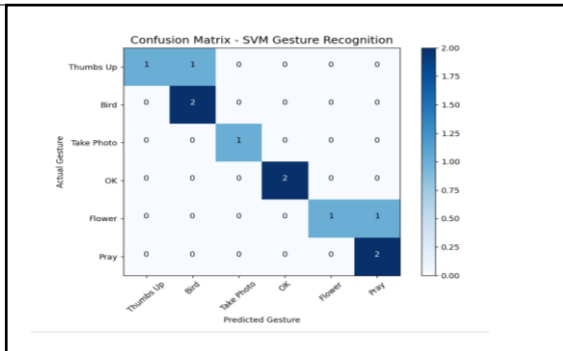


Fig - 4 : Confusion Matrix

#### 4. CONCLUSION

This paper presented a real-time multi-hand gesture recognition and control system that enables intuitive and contactless human-computer interaction. The proposed framework combines MediaPipe for accurate hand landmark detection with the Support Vector Machine (SVM) algorithm for gesture classification. By recognizing both single-hand and multi-hand gestures, the system provides a more flexible and efficient interaction mechanism compared to conventional gesture recognition approaches.

A custom dataset containing various hand gestures was created and used to train and evaluate the SVM classifier. The recognized gestures were successfully mapped to system control operations such as mouse movement, clicking, scrolling, volume adjustment, media playback control, and system locking. Experimental results demonstrated that the SVM-based model achieved high classification accuracy while maintaining real-time performance under different lighting conditions, backgrounds, and hand orientations.

The proposed system offers a reliable, cost-effective, and user-friendly solution for gesture-based control applications. Its ability to perform accurate gesture recognition and seamless system control makes it suitable for applications in smart automation, assistive technologies, robotics, and virtual environments. Future work can focus on expanding the gesture set, incorporating dynamic gesture recognition, and improving system performance in more challenging real-world scenarios. The results confirm that the combination of MediaPipe and SVM provides an effective framework for real-time multi-hand gesture recognition and control.

#### REFERENCES

[1] Wahane, A., et al. (2022). "Real-Time Sign Language Recognition Using Deep Learning Techniques." IEEE 7th International Conference for Convergence in Technology (I2CT), pp. 1-5.

[2] Panneer Selvam, E., and Sornam, M. (2022). "Recent Advances in Sign Language Recognition Using Deep Learning Techniques." 6th International Conference on Trends in Electronics and Informatics (ICOEI), IEEE, pp. 1261-1265.

[3] Mistry, P., Jotaniya, V., Patel, P., Patel, N., and Hasan, M. (2021). "Indian Sign Language Recognition Using Deep Learning." International Conference on Artificial Intelligence and Machine Vision (AIMV), IEEE, pp. 1-6.

[4] Naveed, M., Quratulain, Q., and Shaukat, A. (2021). "Comparison of GLCM-Based Hand Gesture Recognition Systems Using Multiple Classifiers." IEEE International Conference on Robotics and Automation in Industry (ICRAI), pp. 1-5.

[5] Wong, W. K., Juwono, F. H., and Khoo, B. T. T. (2021). "Multi-Features Capacitive Hand Gesture Recognition Sensor: A Machine Learning Approach." IEEE Sensors Journal, 21(6), 8441-8450. DOI: 10.1109/JSEN.2021.3049273.