IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 17, April 2026
Impact Factor: 8.2

AI Powered Virtual Mouse Using Hand Gestures, Eye Tracking and Voice Commands

Gitanjali Kadlag[1], Sumedh Jadhav[2], Harshad Dhumal[3], Pratik Jadhav[4], Sahil Borate[5]
Professor, Department of Artificial Intelligence & Data Science[1]
Students, Department of Artificial Intelligence & Data Science[2-5]
Shree Ramchandra College of Engineering, Pune, India

Abstract: *This paper presents an AI-powered virtual mouse system that enables users to control computer operations without physical hardware using hand gestures, eye tracking and voice commands. The system leverages computer vision techniques, machine learning algorithms and real-time image processing to detect and interpret human interactions. Hand gestures are captured using webcam-based tracking, eye movements are analysed for cursor positioning and voice commands are processed for system-level operations. The integration of these modalities enhances accessibility, improves human-computer interaction and provides a touchless control mechanism. The proposed system is particularly beneficial in scenarios requiring hands-free interaction such as healthcare, presentations and assistive technologies. Experimental results demonstrate high accuracy and efficiency in real-time applications.*

Keywords: Artificial Intelligence, Virtual Mouse, Hand Gesture Recognition, Eye Tracking, Voice Control

I. INTRODUCTION

With the advancement of Artificial Intelligence and Computer Vision, touchless interaction systems are becoming increasingly important. Traditional input devices such as a mouse and keyboard limit flexibility and accessibility, especially for physically challenged users.

This project proposes an AI-powered virtual mouse that uses hand gestures, eye tracking and voice commands to perform computer operations. The system leverages real-time image processing and machine learning techniques to provide an efficient and intuitive user interface.

II. PROBLEM STATEMENT

Conventional input devices require physical interaction, which may not be feasible in all scenarios such as healthcare environments, presentations or for users with disabilities. Existing virtual systems often rely on a single input method, resulting in limited accuracy and usability. Therefore, there is a need for a multi-modal system that enables efficient, touchless and user-friendly control.

OBJECTIVES

- To develop a touchless virtual mouse using AI techniques
- To enable cursor control using hand gestures
- To implement voice commands for system operations
- To enhance accessibility and user experience

SCOPE

The system is designed for general users, professionals and physically challenged individuals. It can be used in environments where touchless interaction is required, such as smart systems, healthcare and virtual presentations.

III. LITERATURE SURVEY

TABLE I

Sr. No	Title	Author	Year	Methodology Used	Conclusion
1.	"Enhancing Gesture-Controlled Virtual Mouse and Virtual Keyboard Using AI Techniques"	J. Kotti, B. Padmaja, D. Deepa	2024	Uses OpenCV and MediaPipe for hand tracking and gesture recognition to control virtual mouse and keyboard.	Provides an efficient touchless interaction system but may face limitations in accuracy under varying conditions.
2.	"Hand-Mouse Interface Using Virtual Monitor Concept for Natural Interaction"	C. Jeon, O.-J. Kwon, D. Shin	2017	Uses Kinect-based gesture recognition and virtual monitor concept for mapping hand movement to cursor.	Improves accuracy and intuitiveness of hand-based mouse systems.
3.	"Deep Learning-Based Real-Time AI Virtual Mouse System Using Computer Vision"	S. Shriram et al.	2021	Uses deep learning and computer vision techniques for real-time gesture detection and cursor control.	Achieves good real-time performance but requires proper lighting and environment.
4.	"Artificial Intelligence-Based Voice Assistant"	S. Subhash et al.	2020	Uses speech recognition and NLP to convert voice commands into executable system actions.	Enhances hands-free interaction but performance depends on noise conditions.

IV. METHODOLOGY

The proposed system is designed as a multi-modal human-computer interaction model integrating hand gesture recognition, eye tracking and voice commands. The development process is structured into the following stages:

A. Data Acquisition

The system captures real-time video input using a webcam and audio input through a microphone. The video stream is processed frame-by-frame for gesture and eye detection, while audio input is used for voice command recognition.

B. Hand Gesture Recognition

Hand gestures are detected using the Media Pipe framework, which identifies key hand landmarks such as fingertips and joints. These landmarks are used to interpret gestures based on their position and movement.

- Cursor movement is controlled by tracking the index finger position
- Clicking actions are performed by finger pinch and voice command
- Scrolling is implemented using thumb and finger pinch

The processed coordinates are mapped to screen resolution using PyAutoGUI for real-time cursor control.

C. Eye Tracking Module

Facial landmark detection is used to identify eye regions and track eyelids movements.

- Reduces dependency on hand gestures for fine adjustments

D. Voice Command Processing

Voice commands are captured using a microphone and processed using a speech recognition library. The system converts speech to text and matches it with predefined commands.

- Commands include actions such as opening applications, scrolling, and closing windows

• Improves usability by enabling hands-free interaction

E. System Integration

All three modules are integrated into a unified system. A priority-based approach is used to handle inputs efficiently, ensuring smooth and responsive interaction without conflicts.

F. Implementation Tools

• Programming Language: Python
• Libraries Used: OpenCV, Media Pipe, PyAutoGUI, Speech Recognition

G. Testing and Validation

The system is tested under different lighting conditions and environments to evaluate accuracy and performance. Results show effective real-time interaction with minimal latency.

V. MODELING & ANALYSIS

The system converts human inputs into machine actions using mathematical mapping and computer vision techniques.

A. Cursor Mapping

The fingertip position from the camera frame is mapped to screen coordinates:

$$X_s = \frac{X_c}{W_c} \times W_s, Y_s = \frac{Y_c}{H_c} \times H_s$$

This ensures proportional and smooth cursor movement across the screen.

B. Gesture Detection

Gestures are identified based on the distance between finger landmarks:

$$D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

If the distance is below a threshold, actions like clicking are triggered.

C. Eye Tracking

Eye movement is estimated using relative distances between eye landmarks:

$$G = \frac{d_{left}}{d_{right}}$$

This helps in fine-tuning cursor position and improving accuracy.

D. Voice Command Processing

Speech input is converted into text and matched with predefined commands. The system selects the most relevant command and executes the corresponding action.

E. Performance

The system processes video and audio inputs in real-time with low latency, ensuring smooth and responsive interaction.

IJARSCT
ISSN: 2581-9429
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Volume 6, Issue 17, April 2026
Impact Factor: 8.2

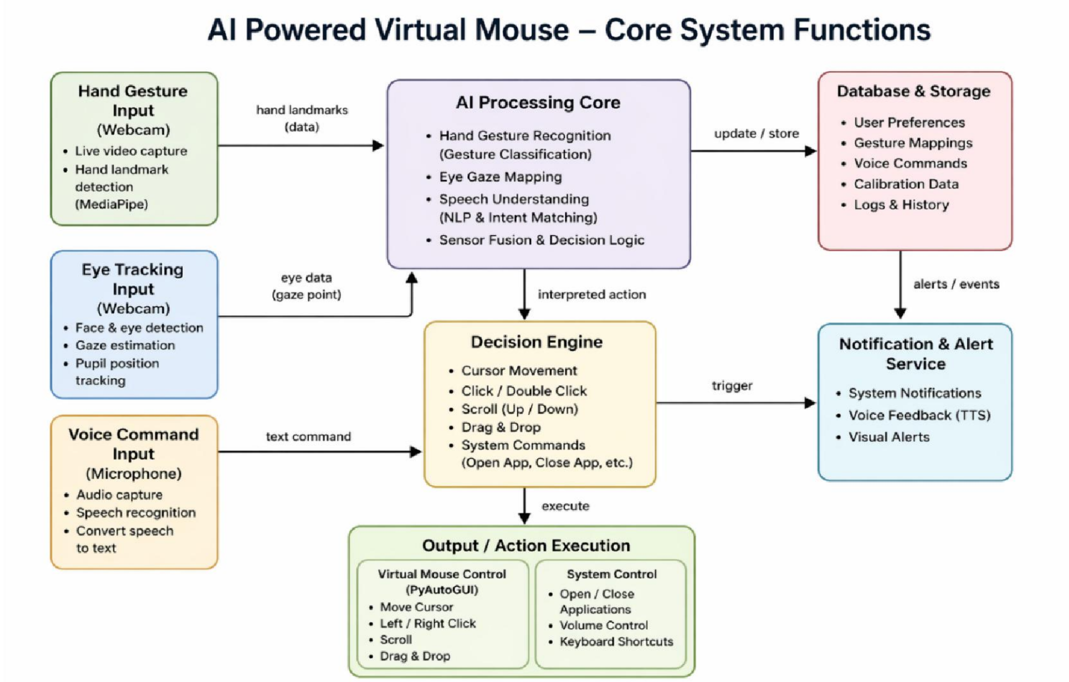


Figure 1. SYSTEM ARCHITECTURE

1. Hand Gesture Input
• Captures live video using a webcam
• Detects hand landmarks using MediaPipe
• Provides gesture data for cursor movement and click actions

2. Eye Tracking Input
• Detects face and eye regions from the video stream
• Tracks gaze direction and pupil movement

3. Voice Command Input
• Captures audio through a microphone
• Converts speech into text using speech recognition
• Identifies commands for system operations

4. AI Processing Core
• Central module that processes all inputs
• Performs:
o Gesture recognition
o Eye gaze mapping
o Voice command understanding (NLP)
• Combines all inputs to understand user intent

5. Database & Storage
• Stores user preferences and configurations

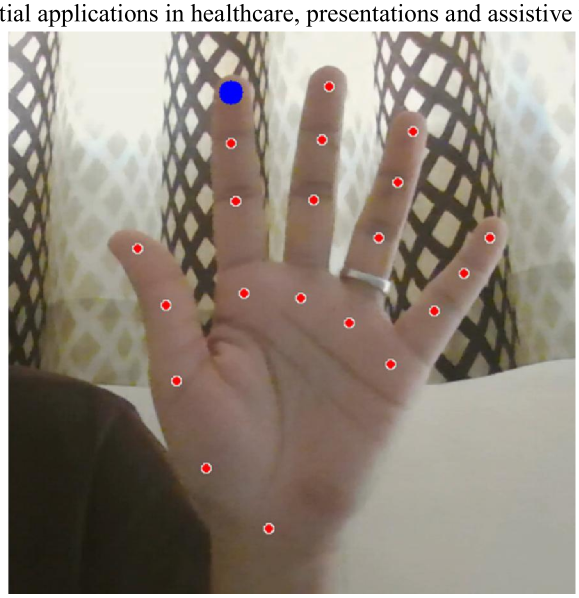
• Maintains gesture mappings and command history
• Helps in system optimization and future use

6. Decision Engine
• Converts processed inputs into actions
• Controls:
o Cursor movement
o Clicking and scrolling
o System commands (open/close apps)

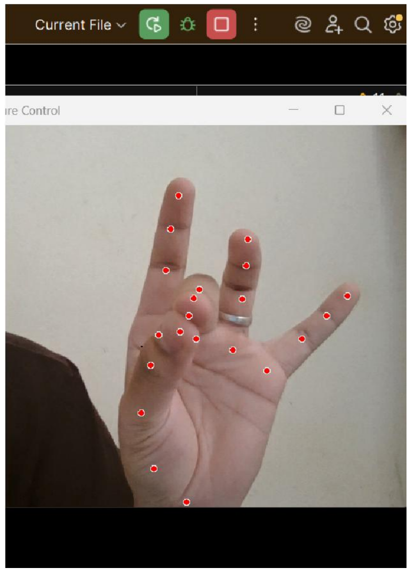
7. Output / Action Execution
• Executes final actions using PyAutoGUI
• Controls mouse functions and system operations in real-time

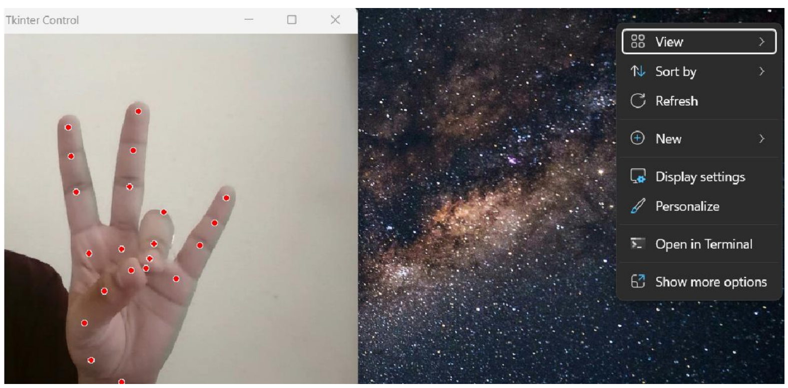
VI. RESULTS & DISCUSSION

- The proposed system was successfully implemented and tested in real-time using a webcam and microphone, demonstrating effective touchless control of computer operations.
- Hand gesture recognition accurately controlled cursor movement, clicking and scrolling with minimal delay under normal lighting conditions.
- Eye tracking improved cursor precision, especially for fine movements, reducing the effort required for continuous hand gestures.
- Voice command functionality enabled users to perform system-level actions such as opening applications and scrolling, enhancing overall usability.
- The integration of gesture, eye and voice inputs provided a more flexible and efficient interaction compared to single-input systems.
- The system showed good performance with low latency, ensuring smooth and responsive user experience during real-time operation.
- Overall, the system demonstrates a cost-effective and accessible solution for touchless human-computer interaction with potential applications in healthcare, presentations and assistive technologies.



IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 17, April 2026
Impact Factor: 8.2






```
2026-04-23 11:51:16,651 [WARNING] No speech detected
2026-04-23 11:51:16,652 [INFO] Listening...
2026-04-23 11:51:18,508 [INFO] Audio captured
RAW COMMAND: open chrome
2026-04-23 11:51:20,385 [INFO] Command received: open chrome
2026-04-23 11:51:20,598 [INFO] Assistant: Opening chrome
2026-04-23 11:51:20,599 [INFO] Listening...
```

VII. ACCURACY AND PERFORMANCE ANALYSIS

The performance of the AI Powered Virtual Mouse Using Hand Gestures, Eye Tracking and Voice Commands system was evaluated based on accuracy, response time, and reliability under different conditions. Multiple test cases were conducted to measure the effectiveness of gesture recognition, eye tracking, and voice command modules.

The system demonstrated high accuracy in real-time interaction under normal environmental conditions. The results are summarized in the table below:

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 17, April 2026
Impact Factor: 8.2

Module	Test Scenario	Accuracy (%)	Response Time	Remarks
Hand Gesture	Cursor Movement	95%	~0.05 sec	Smooth and responsive tracking
Hand Gesture	Left Click (Pinch)	93%	~0.1 sec	Minor delay in fast motion
Hand Gesture	Scroll Control	90%	~0.1 sec	Depends on gesture stability
Hand Gesture	Right Click	92%	~0.1 sec	Accurate under good lighting
Eye Tracking	Blink Detection (Click)	88%	~0.15 sec	Slight variation in detection
Eye Tracking	Screenshot (Triple Blink)	85%	~0.2 sec	Affected by eye visibility
Voice Command	Open Application	94%	~0.3 sec	High accuracy in quiet environment
Voice Command	System Commands	91%	~0.3 sec	Affected by noise
Voice Command	Typing Text	89%	~0.4 sec	Depends on pronunciation
Overall System	Combined Input	92%	Real-time	Stable and efficient

VIII. CONCLUSION

The proposed AI-powered virtual mouse system demonstrates an effective approach to touchless human-computer interaction by integrating hand gestures, eye tracking and voice commands. The system provides accurate and real-time control of cursor operations without the need for traditional input devices. It enhances user experience, flexibility and accessibility, making it suitable for applications such as healthcare, presentations and assistive technologies. Although minor limitations exist due to lighting conditions and background noise, the overall performance is reliable and efficient. This work highlights the potential of AI and computer vision in developing advanced, user-friendly interaction systems.

ACKNOWLEDGEMENT

We would like to express our sincere appreciation to all those who supported us throughout the development of this research work. We are especially grateful to our project guide for their expert guidance, continuous motivation and valuable feedback, which played a crucial role in shaping this paper and overcoming technical challenges.

We extend our heartfelt thanks to HOD Vikas Gaikwad Sir and the faculty of the Department of Artificial Intelligence and Data Science, Shree Ramchandra College of Engineering, Pune, for providing the necessary resources, academic support and a conducive learning environment that contributed significantly to the successful completion of this work.

REFERENCES

[1]. J. Kotti, B. Padmaja and D. Deepa, "Enhancing Gesture-Controlled Virtual Mouse and Virtual Keyboard Using AI Techniques," Journal of Mobile Multimedia, vol. 20, no. 2, pp. 473–494, 2024.

[2]. C. Jeon, O.-J. Kwon, D. Shin and D. Shin, "Hand-Mouse Interface Using Virtual Monitor Concept for Natural Interaction," IEEE Access, vol. 5, pp. 25181–25187, 2017.

[3]. S. Shriram et al., "Deep Learning-Based Real-Time AI Virtual Mouse System Using Computer Vision," Journal of Healthcare Engineering, 2021.

[4]. S. Subhash et al., "Artificial Intelligence-Based Voice Assistant," in Proc. IEEE World Conference on Smart Trends in Systems, Security and Sustainability, 2020, pp. 593–596.

DOI: 10.48175/IJARSCT-34207