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 10, June 2026

Impact Factor: 8.2

Intelligent Touchless Human-Computer Interaction Using Gesture Recognition and Voice Commands

Prof. Urmila Bhadange, Gauri Palde, Atharva Chavan, Swapnil Pawar

Department of Artificial Intelligence and Data Science

Pune Vidhyarti Griha's College of Engineering & S.S. Dhamankar Institute of Management, Nashik

Abstract: *The growing demand for touchless and intelligent computing systems has increased the importance of advanced Human-Computer Interaction (HCI) technologies. This paper presents a real-time AI-powered gesture recognition and voice-controlled automation system that enables users to interact with computers without traditional input devices such as keyboards and mice. The proposed system utilizes MediaPipe-based hand landmark detection, OpenCV image processing, and machine learning techniques to recognize hand gestures and convert them into computer control actions. In addition, a voice assistant module is integrated to execute voice commands for system operations and application control.The developed system supports various functionalities including cursor movement, click operations, scrolling, volume control, screenshot capture, and voice-based task execution. Experimental evaluation was conducted under different operating conditions to assess recognition accuracy, response time, and system reliability. The results demonstrate efficient real-time performance, accurate gesture classification, and smooth user interaction with minimal latency. The integration of gesture recognition and voice assistance provides a user-friendly, cost-effective, and accessible solution for touchless computing environments. The proposed approach has potential applications in smart homes, healthcare systems, educational platforms, virtual reality environments, and assistive technologies.*

Keywords: Hand Gesture Recognition, Computer Vision, Human-Computer Interaction (HCI), Voice Command System

I. INTRODUCTION

Human-Computer Interaction (HCI) plays a vital role in enabling effective communication between users and computer systems. Traditional input devices such as keyboards and mice have been widely used for decades; however, they require physical contact and may limit flexibility in modern computing environments. With recent advancements in Artificial Intelligence (AI), Computer Vision, and Machine Learning, touchless interaction methods have gained significant attention for providing more natural, efficient, and accessible user experiences.Gesture recognition technology allows computers to interpret human hand movements and convert them into meaningful commands. By using a webcam and computer vision algorithms, users can perform various operations such as cursor movement, clicking, scrolling, and media control without touching any hardware device. Similarly, voice recognition technology enables users to interact with systems through spoken commands, improving convenience and accessibility.The integration of gesture recognition and voice assistance has opened new possibilities for smart and contactless computing. Such systems are particularly useful in healthcare environments, educational institutions, smart homes, industrial automation, virtual reality applications, and assistive technologies for individuals with physical disabilities. The increasing demand for hygienic and touch-free interfaces has further accelerated research in this domain.This paper presents an AI-powered gesture recognition and voice-controlled automation system designed to provide real-time computer interaction. The system utilizes MediaPipe for hand landmark detection, OpenCV for image processing, and machine learning algorithms for gesture classification. A voice assistant module is integrated to execute system-level commands and automate routine tasks through speech input. Unlike many existing studies that primarily focus on system design, this work emphasizes practical implementation and experimental evaluation. The developed system was tested under different operating conditions to analyze gesture recognition accuracy, response time, reliability, and overall performance. The experimental results demonstrate the effectiveness of combining gesture-based control and voice commands into a unified platform for intelligent human-computer interaction. The primary objective of this research is to develop a cost-effective, user-friendly, and efficient touchless interaction system that enhances accessibility and improves the overall computing experience. By integrating computer vision, machine learning, and speech recognition technologies, the proposed system contributes to the advancement of next-generation intelligent interaction environments.

Key Contributions

* Developed an AI-powered touchless human-computer interaction system that combines hand gesture recognition and voice-controlled automation within a single platform.
* Implemented real-time hand landmark detection using MediaPipe and computer vision techniques for accurate gesture tracking and cursor control.
* Integrated a voice assistant module capable of executing system-level commands, application control, web searches, and multimedia operations through speech input.
* Conducted experimental evaluation under different operating conditions to analyze gesture recognition accuracy, response time, and system reliability.
* Performed performance analysis of commonly used gestures such as cursor movement, click, scroll, drag, volume control, and screenshot capture.
* Demonstrated low-latency and efficient real-time performance using standard hardware without requiring specialized sensors or external devices.
* Improved accessibility and user convenience by providing a cost-effective, contactless interaction solution suitable for healthcare, education, smart homes, and assistive technology applications.
* Presented experimental results and output analysis that validate the effectiveness of integrating gesture recognition and voice assistance for intelligent computing environments.

II. REVIEW OF LITERATURE

A. Media Pipe-Based Virtual Mouse System

Researchers developed a virtual mouse system using MediaPipe and OpenCV to control cursor movement through hand gestures. The system successfully detected hand landmarks in real time and performed operations such as cursor movement and clicking. However, the system mainly focused on gesture control and did not include voice-based interaction.

B. Hand Gesture Recognition Using Machine Learning

This study applied machine learning algorithms such as KNN, SVM, and MLP for gesture classification. The proposed approach improved recognition accuracy and enabled multiple gesture-based operations. However, the performance depended heavily on training data quality and environmental conditions.

C. Voice Assistant for Computer Automation

Researchers designed a voice-controlled assistant capable of executing commands such as opening applications, web searching, and controlling system settings. Speech recognition and text-to-speech technologies were used to improve user interaction. The system provided convenience but lacked gesture-based control functionality.

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 10, June 2026 Impact Factor: 8.2

D. AI-Based Human-Computer Interaction System

This research combined artificial intelligence and computer vision techniques to enhance human-computer interaction. The system enabled touchless operation and improved accessibility for users. However, detailed performance evaluation and experimental analysis were limited.

Research Gap:

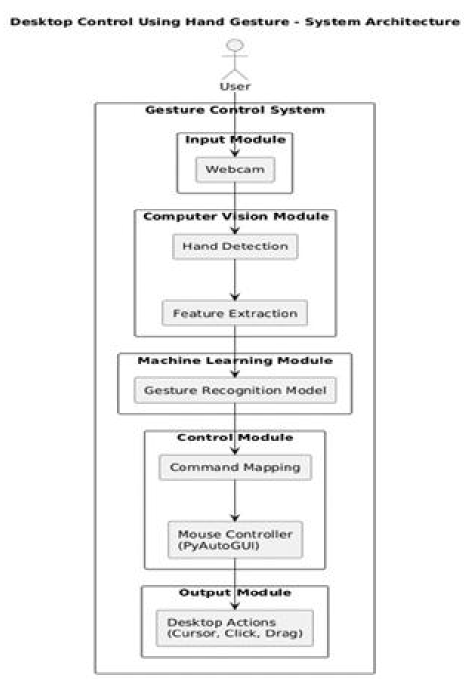
Most existing studies focus either on gesture recognition or voice-controlled automation separately. Limited work has been carried out on integrating both technologies into a single platform while performing comprehensive performance analysis. Therefore, the proposed work develops an AI-powered gesture and voice-controlled system and evaluates its real-time performance, accuracy, and reliability..

III. SYSTEM ARCHITECTURE AND DESIGN

The proposed system integrates computer vision, machine learning, and voice recognition technologies to provide a touchless human-computer interaction platform. The architecture is designed to recognize hand gestures, process voice commands, and execute corresponding system operations in real time. The system ensures smooth interaction, high accuracy, and efficient performance using standard hardware resources.

A. Architecture Overview

The architecture consists of a webcam-based image acquisition unit, a hand landmark detection module, a gesture classification engine, a voice assistant module, and a system automation layer. The webcam continuously captures live video frames, which are processed using MediaPipe for hand landmark extraction. The extracted features are analyzed by a machine learning classifier to identify user gestures. Simultaneously, the voice assistant receives spoken commands through a microphone and converts them into executable actions. The recognized gestures and voice commands are then translated into system operations such as cursor control, clicking, scrolling, volume adjustment, and application management.



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 10, June 2026
Impact Factor: 8.2

B. Role-Based Module Design

The image acquisition module is responsible for capturing real-time video streams from the webcam. The hand tracking module detects and tracks hand landmarks, generating coordinate-based feature data for gesture analysis. The gesture recognition module interprets these features and classifies user gestures using machine learning techniques. The voice assistant module processes speech input and converts spoken instructions into actionable commands. The automation module serves as the execution layer, performing system-level operations through software automation libraries. Together, these modules operate in coordination to provide a seamless and intelligent user experience.

C. Technology Stack

Component	Technology Used
Programming Language	Python 3.10
Computer Vision Framework	OpenCV
Hand Tracking Module	MediaPipe
Hands Gesture Classification	Scikit-Learn (MLP Classifier)
System Automation	PyAutoGUI
Voice Recognition	Speech Recognition
Text-to-Speech Engine	Pyttsx3
Input Devices	Webcam and Microphone
Development Environment	Visual Studio Code
Operating System	Windows 10/11

IV. SYSTEM ANALYSIS

System analysis is the process of examining existing methods and identifying their limitations in order to develop a more efficient and reliable solution. The proposed AI-powered gesture recognition and voice-controlled automation system was designed after analyzing the challenges associated with traditional input devices and existing touchless interaction systems. The analysis focuses on improving usability, accessibility, accuracy, and real-time performance.

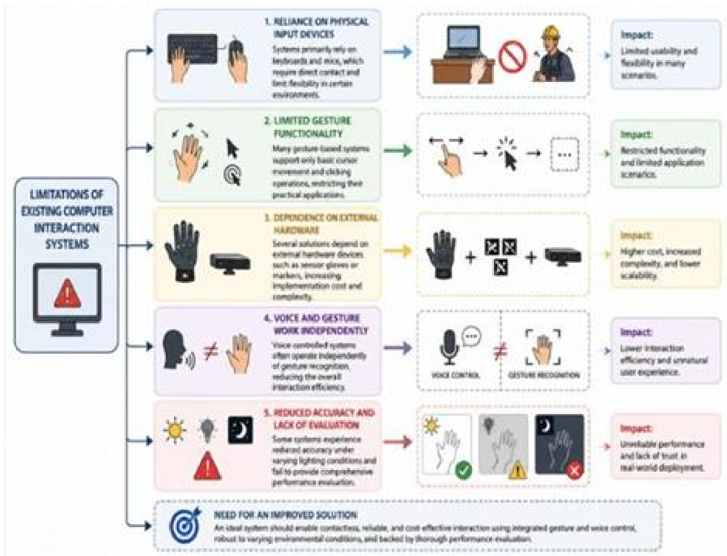
A. Limitations of Existing Systems

Existing computer interaction systems primarily rely on physical input devices such as keyboards and mice, which require direct contact and limit flexibility in certain environments. Many gesture-based systems support only basic cursor movement and clicking operations, restricting their practical applications. Several existing solutions depend on external hardware devices such as sensor gloves or markers, increasing implementation cost and complexity. Voice-controlled systems often operate independently of gesture recognition, reducing the overall interaction efficiency. Additionally, some systems experience reduced accuracy under varying lighting conditions and fail to provide comprehensive performance evaluation.

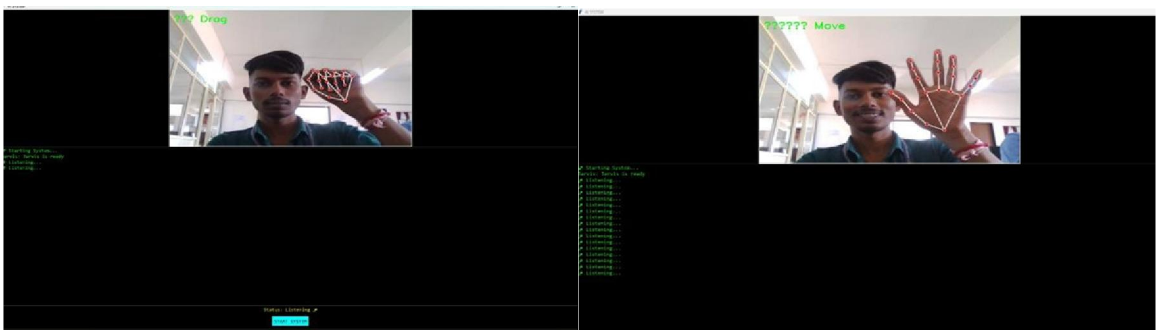
B. Proposed System Analysis

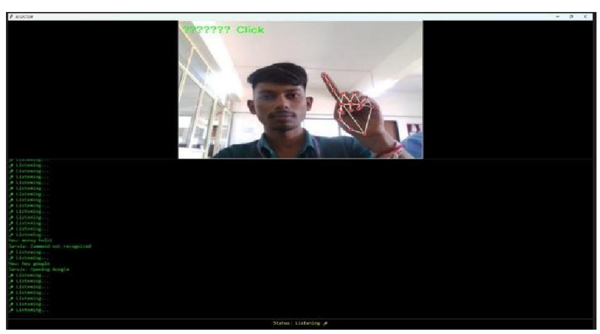
The proposed system integrates gesture recognition and voice-controlled automation into a unified platform for intelligent human-computer interaction. MediaPipe-based hand tracking enables accurate detection of hand landmarks without requiring external sensors or wearable devices. Machine learning techniques are employed to classify user gestures and perform real-time actions such as cursor movement, clicking, scrolling, volume adjustment, and screenshot capture. The integrated voice assistant processes spoken commands and executes system-level operations efficiently. The combination of gesture and voice interaction enhances user convenience, improves accessibility, and provides a cost-effective touchless computing solution. Experimental evaluation demonstrates reliable performance, high recognition accuracy, and smooth real-time operation across different usage conditions.

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 10, June 2026
Impact Factor: 8.2



B. System result





V. CONCLUSION

This paper presented an AI-powered gesture recognition and voice-controlled automation system for touchless human-computer interaction. The proposed system successfully integrates computer vision, machine learning, and speech recognition technologies to provide an efficient and user-friendly alternative to traditional input devices. By utilizing MediaPipe for hand landmark detection, OpenCV for image processing, and machine learning techniques for gesture classification, the system enables real-time execution of various computer operations through hand gestures. The integration of a voice assistant further enhances system functionality by allowing users to perform system-level tasks through spoken commands. Experimental evaluation demonstrated reliable gesture recognition, smooth cursor control, accurate command execution, and low response time under different operating conditions. The results indicate that the proposed approach provides an effective, cost-efficient, and accessible solution for intelligent touchless interaction.The developed system has potential applications in healthcare, education, smart homes, virtual reality, industrial automation, and assistive technologies. Future enhancements may focus on dynamic gesture recognition, multi-hand tracking, deep learning-based classification models, and advanced voice processing capabilities to further improve system accuracy and user experience

VI. ACKNOWLEDGMENT

The authors express their sincere gratitude to Prof. U. B. Bhandage, Project Guide, Department of Artificial Intelligence and Data Science, Pune Vidhyarthi Griha's College of Engineering, Nashik, for her invaluable guidance, consistent encouragement, and patient mentorship throughout this research. The authors likewise thank Prof. H. S. Borse, Project Co-ordinator, for her continued cooperation and institutional support. Special acknowledgment is due to Prof. S. G. Chordiya, Head of the Department of Artificial Intelligence and Data Science, for providing the necessary facilities and academic environment. The authors also thank all classmates and peers whose constructive discussions and iterative feedback substantially enriched the development and evaluation of this system.

REFERENCES

[1]. W. Zhang, X. Li, and Y. Chen, "Dynamic Hand Gesture Recognition Based on Deep Learning Networks," IEEE/CAA Journal of Automatica Sinica, vol. 8, no. 5, pp. 1023–1035, 2021.

[2]. M. Kowdiki, S. Maheshwari, and A. Verma, "Automatic Hand Gesture Recognition Using Hybrid Meta-Heuristic Techniques," Journal of Visual Communication and Image Representation, vol. 72, pp. 102–115, 2021.

[3]. K. C. Arthanari, A. P. Srivastava, and M. Gayathri, "Hand Gesture Recognition for Human-Computer Interaction Using Computer Vision," International Conference on Smart Computing, 2023.

[4]. J. Shin, A. S. M. Miah, and M. H. Kabir, "A Methodological Review of Hand Gesture Recognition Across Data Modalities," arXiv Preprint arXiv:2408.05436, 2024.

[5]. C. Cui, Y. Liu, and Z. Wang, "Deep Vision-Based Real-Time Hand Gesture Recognition: A Review," IEEE Access, vol. 13, pp. 1–15, 2025.

[6]. E. Fertl, T. Müller, and S. Wagner, "Hand Gesture Recognition on Edge Devices: Sensor and Vision-Based Approaches," Sensors, vol. 25, no. 6, pp. 1687, 2025.

[7]. M. Linardakis and I. Varlamis, "Survey on Hand Gesture Recognition from Visual Input," IEEE Access, 2025.

[8]. B. Sinha, "A Comprehensive Survey on Hand Gesture Recognition Using Deep Learning for Electronic Device Interaction," TechRxiv Preprint, 2025.

[9]. Google, "MediaPipe Hands: Real-Time Hand Tracking," Available: https://developers.google.com/mediapipe OpenCV Organization, "OpenCV Documentation," Available: https://opencv.org