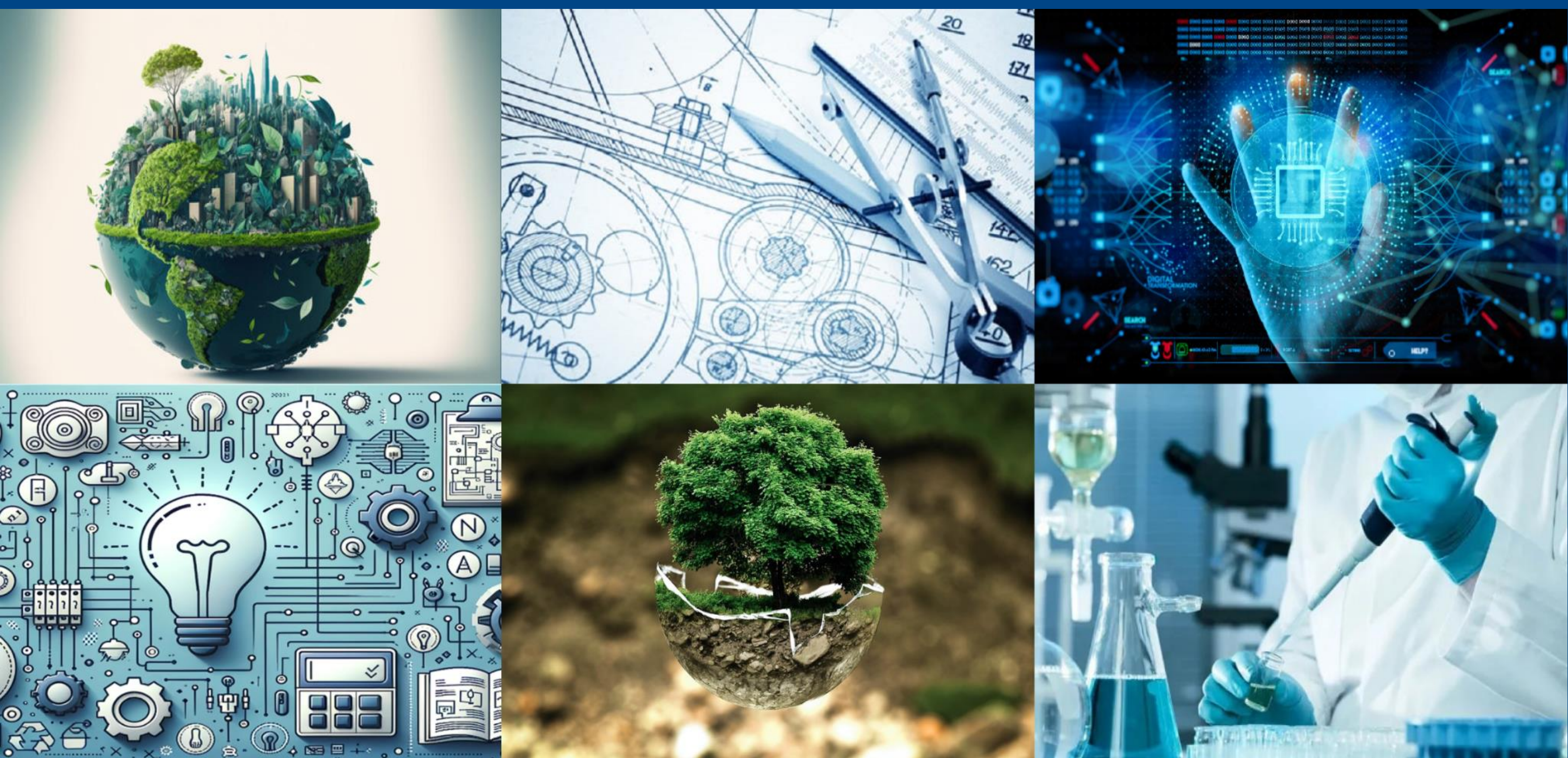
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 4, April 2025

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

A Multimodal System for Human-Computer Interaction

Maria Sobana S, Krishna Veni M, Mirsha Shelyn J, Nagarani V O

Assistant Professor, Dept. of CSE, K.L.N. College of Engineering Sivagangai, Tamil Nadu, India

UG Student, Dept. of CSE, K.L.N. College of Engineering, Sivagangai, Tamil Nadu, India

UG Student, Dept. of CSE, K.L.N. College of Engineering, Sivagangai, Tamil Nadu, India

UG Student, Dept. of CSE, K.L.N. College of Engineering, Sivagangai, Tamil Nadu, India

ABSTRACT: The primary objective of this system is to develop a virtual mouse using hand gesture recognition, enabling users to interact with computers without physical contact. It utilizes MediaPipe and OpenCV to detect and process hand gestures, performing mouse operations such as cursor movement, clicking, scrolling, and drag-and-drop. The system also integrates voice commands through SpeechRecognition and pyttsx3, allowing users to conduct Google searches, retrieve date/time, control system modes, and receive voice feedback. To enhance accessibility for differently-abled individuals, it includes eye movement tracking and a virtual keyboard as alternative input methods. Designed to be cost-effective and user-friendly, the system operates with a standard webcam, eliminating the need for specialized hardware. Its contactless interaction offers high responsiveness and accuracy, making it ideal for various real-world applications like education, healthcare, and assistive technologies. By combining machine learning, computer vision, and speech processing, the system delivers an intelligent, seamless human-computer interaction experience tailored to modern accessibility needs.

I. INTRODUCTION

Gesture-controlled systems provide a contactless and efficient way to interact with computers, enhancing both accessibility and user convenience. These systems replace traditional input devices like the mouse and keyboard with intuitive, natural interactions using hand gestures, voice commands, and eye movements. The proposed system uses MediaPipe and OpenCV to achieve real-time hand gesture recognition, allowing users to perform essential mouse operations such as cursor movement, clicking, scrolling, and dragging. Voice command integration, powered by libraries such as SpeechRecognition and pyttsx3, enables users to execute tasks like Google searches, retrieving the current date and time, and controlling system functions such as sleep and wake. Eye movement tracking further enriches the interaction by allowing gaze-based navigation and selection, providing an alternative for users with limited hand mobility. This multimodal approach ensures a flexible and adaptive user interface, especially beneficial for differently-abled individuals. Utilizing a standard webcam, the system offers a cost-effective and user-friendly solution that supports seamless real-time operation, promoting inclusive and intelligent human-computer interaction in various environments such as healthcare, education, and assistive technology.

II.LITERATURE REVIEW

[1] Rohan Nerurkar, Amit Kale, "Hand Gesture Controlled Virtual Mouse Using Computer Vision," International Research Journal of Engineering and Technology, Vol. 07, Issue 04, pp. 3055-3059,April 2020.
This paper presents a hand gesture-controlled virtual mouse system using computer vision to enable intuitive, hardware-free interaction. A standard webcam captures real-time hand movements, and OpenCV algorithms like contour detection and convex hull are used to identify gestures. Actions such as cursor movement, clicking, and dragging are linked to specific finger motions, eliminating the need for extra hardware like gloves or sensors. The system ensures smooth and low-latency interactions, offering over 90% accuracy for basic mouse functions. Optimized for speed and simplicity, it provides a reliable, cost-effective alternative to traditional input devices.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[2] Shubham Gupta, Deepanshu Sharma, "Virtual Mouse Using Hand Gestures," IEEE International Conference on Inventive Computation Technologies (ICICT), 2022.

This paper introduces a virtual mouse system that uses Convolutional Neural Networks (CNNs) to detect and classify hand gestures for mouse operations like movement, clicking, and scrolling. Gesture stabilization techniques are employed to reduce jitter and ensure smooth control. Designed for dynamic environments, the system performs well under different lighting and backgrounds. Achieving up to 95% gesture recognition accuracy, it offers a precise, intelligent, and hardware-free alternative to traditional input devices. Its adaptability and consistent performance showcase the effectiveness of deep learning in enhancing human-computer interaction..

[3] S. Ghosh, R. Kumar, "Smart Virtual Mouse Using Hand Gestures Based on MediaPipe," International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), IEEE, 2023.

This paper presents a smart virtual mouse system using Google's MediaPipe for real-time hand tracking and gesture recognition. By detecting 21 hand landmarks, it maps gestures to mouse functions like clicking and scrolling with high accuracy. MediaPipe ensures low computational load and fast response, making the system efficient and hardware-free. Tested on Windows and Linux, it shows strong performance and adaptability in dynamic environments. The study highlights how MediaPipe enables intuitive, low-latency human-computer interaction.

[4] T. B. Pujari, S. R. Dandge, "Human Computer Interaction Using Virtual Mouse Based on Hand Gesture Recognition," International Journal of Computer Sciences and Engineering (IJCSE), Vol. 7, Issue 6, pp. 199–204, June 2019.

This paper presents a gesture-based virtual mouse system that enhances human-computer interaction by replacing traditional input devices with hand gestures. Using HSV color segmentation and convexity defects, it identifies finger positions for basic mouse functions like movement and clicking. The rule-based approach ensures low computational load and real-time performance without machine learning. Tested in various lighting conditions, the system performs best indoors and is especially beneficial for users with physical impairments, offering an accessible and cost-effective computing solution.

III.EXISTING SYSTEM

The existing system requires users to wear colored markers or gloves for gesture recognition, making it less intuitive and inconvenient for natural interaction. Its performance is significantly affected by background color similarity or poor lighting, which can lead to inaccurate tracking and false detections. Additionally, the system supports only hand gestures and lacks integration of voice commands or alternative input modes, thereby limiting its accessibility and flexibility. It primarily relies on static hand positions and does not support real-time gesture sequence tracking, which restricts dynamic interactions such as swiping or dragging. Furthermore, the implementation is not optimized for cross-platform compatibility and may exhibit reduced performance or functionality across different hardware and operating system environments.

IV.PROPOSED SYSTEM

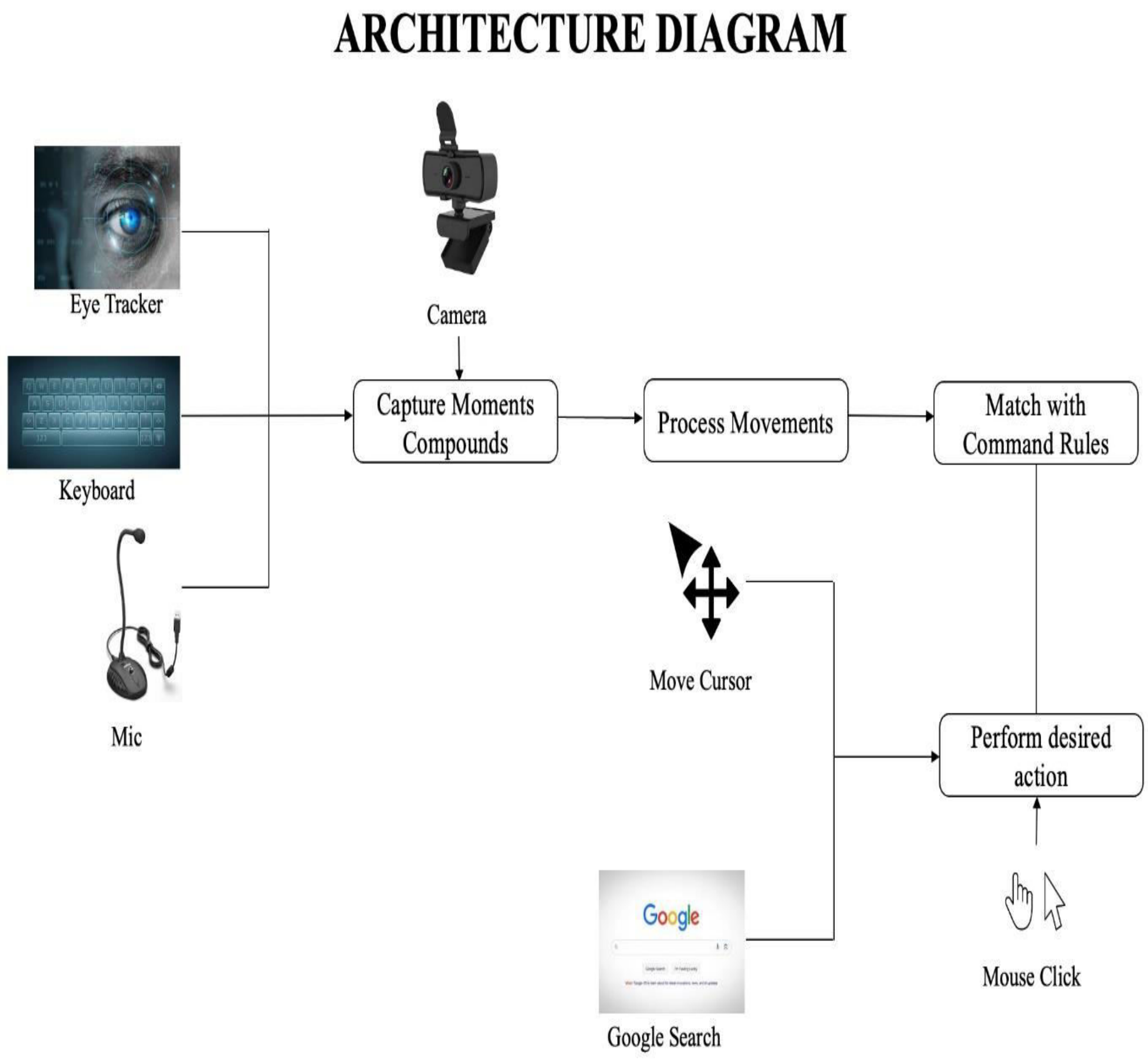
The proposed system integrates both voice and gesture recognition to enable hands-free control, offering accessibility and convenience to users with limited mobility. It supports essential system functions such as checking the current date and time, performing Google searches, and finding locations using voice commands for quick and efficient interaction. Gesture-based mouse control includes operations like cursor movement, left, right, and double clicks, scrolling, and drag-and-drop, significantly enhancing the user experience. Additionally, eye movement tracking allows users to interact with the system by simply looking at specific areas on the screen, providing an alternative and intuitive input method. Features like volume control and multiple item selection through gestures further add to the system's versatility, making it suitable for both general use and environments that require assistive technologies.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

V. ARCHITECTURE DIAGRAM

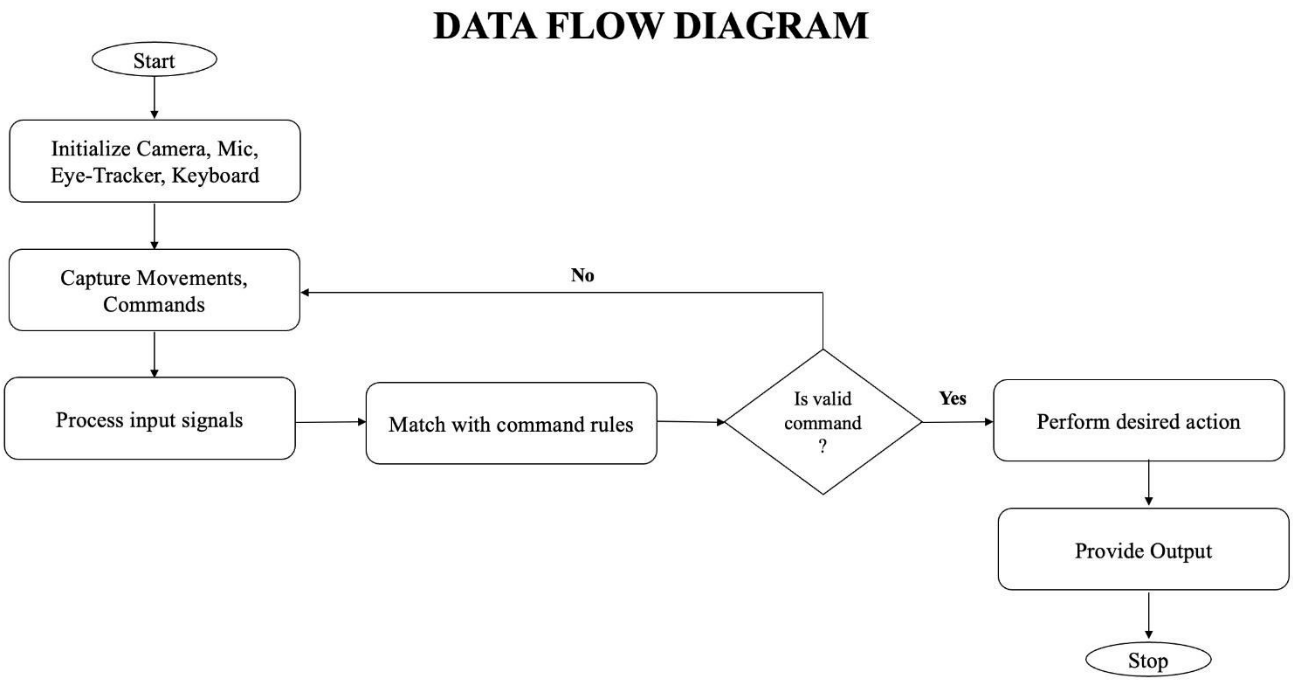


The architecture of the Virtual Mouse System is designed to enable hands-free interaction with a computer through multiple input modes including voice, gestures, eye movements, and traditional keyboard input. It begins with the Input Devices Layer, where devices like the microphone, camera, eye tracker, and keyboard collect raw user inputs. These inputs are passed to the Signal Acquisition Layer, which captures voice commands, tracks hand gestures, detects eye movements, and reads keyboard strokes to convert them into usable signals. The next stage is the Preprocessing Layer, where the raw data is cleaned and refined through noise filtering, frame segmentation, and feature extraction to isolate the important elements needed for accurate recognition. The system then moves to the Recognition and Classification Layer, where advanced algorithms or machine learning techniques are used to recognize gestures, match voice commands, and map eye gaze directions. Once the input is recognized, the Action Execution Layer translates these into specific system-level tasks such as cursor movement, clicking, dragging, volume control, performing Google searches, or executing sleep/wake functions. Finally, the Output Interface Layer provides feedback to the user via screen display and voice responses, confirming that the desired action has been performed. This layered architecture ensures seamless interaction between the user and the system, enabling real-time and intuitive control without the need for physical devices like a mouse.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VI.DATA FLOW DIAGRAM



The system begins by launching and preparing all necessary modules to capture user commands. This includes initializing input devices such as the camera, microphone, eye tracker, and keyboard. The camera is used to detect hand gestures, the microphone captures voice commands, the eye tracker monitors gaze movements, and the keyboard serves as an alternative input method if required. Once initialized, the system begins capturing real-time data from the user, including gestures, voice inputs, and eye movement patterns. These inputs act as control signals for the system. The collected data is then processed using specialized technologies computer vision for gestures, speech recognition for voice commands, and gaze-tracking algorithms for eye movements transforming them into meaningful signals. These signals are matched against a predefined set of command rules that map each gesture, voice instruction, or gaze behavior to a corresponding system action such as moving the cursor, clicking, scrolling, or launching applications. The system checks whether the interpreted input matches a valid command. If a valid command is found, the system proceeds to execute it. Otherwise, it continues capturing additional input to ensure accuracy. When a valid command is executed, the output is reflected in the user interface for instance, moving the cursor, opening an application, or registering a click. Feedback may be provided to confirm the action. The system continues this loop until the user ends the session or interaction ceases, at which point the system stops its operation.

VII.MODULES

- Hand Detection
- Gesture Recognition
- Voice Command
- Eye Movement Detection
- Virtual Keyboard

HAND DETECTION

MediaPipe's advanced Hand Tracking model captures hand landmarks with high accuracy in real-time, enabling the system to continuously monitor hand movements with minimal delay. The system maps the coordinates of each finger's

© 2025 IJMRSET | Volume 8, Issue 4, April 2025 |DOI:10.15680/IJMRSET.2025.0804451

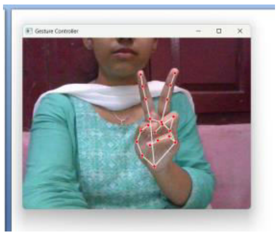
ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

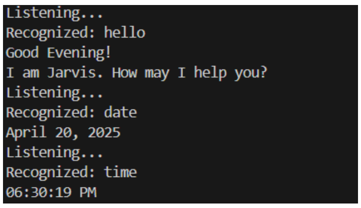
tip and base using the detected landmarks, helping determine the structure and orientation of the fingers. By comparing the vertical positions of finger tips with their corresponding lower joints, it accurately interprets gestures such as an open hand, closed fist, or specific finger signals. Each hand is represented using 21 distinct points from the wrist to the fingertips, which are essential for defining gesture patterns and tracking movements. The system supports detection and tracking of either one or both hands simultaneously using only a standard webcam, making the setup simple, cost-effective, and accessible.

GESTURE RECOGNITION



The system compares the current finger configurations against a library of predefined gesture patterns to trigger corresponding computer actions automatically. Common mouse functions such as left click, right click, double click, and precise cursor movement are executed based on specific hand gestures. Scrolling through documents or web pages, dragging files, and selecting multiple items are enabled through intuitive gestures for seamless control. Volume control is managed by measuring the distance between specific fingertips, allowing users to increase or decrease volume with natural hand motions. The gesture detection runs continuously in real-time, ensuring fluid and responsive user interaction without noticeable lag.

VOICE COMMAND



The system listens to voice input through the microphone using the SpeechRecognition library, accurately recognizing spoken commands. It converts the recognized speech into text and checks it against predefined commands to trigger appropriate system-level actions, such as announcing the time or launching applications. The pyttsx3 library is used to provide audible feedback by converting system responses into natural-sounding speech, enhancing user interaction and confirmation. Voice commands automate tasks like time announcements, application launches, and system sleep mode, improving efficiency and user experience. The voice control system allows users with mobility challenges to interact hands-free, providing an inclusive solution that doesn't rely on traditional input devices.

EYE MOVEMENT DETECTION

Haar Cascade classifiers in OpenCV detect the eye region in real-time by recognizing specific features, enabling accurate eye tracking. The system estimates the user's gaze direction by monitoring the pupil's position and the iris' relative location. Blink detection and dwell time can trigger actions like selecting items or moving the cursor. The

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

system allows hands-free control by moving the cursor or selecting items based on eye focus or blinking. Eye tracking is especially useful for users with limited mobility, enhancing accessibility and enabling interaction without physical touch.

VIRTUAL KEYBOARD

The system translates finger gestures, like pointing or tapping, into key presses on a virtual keyboard, enabling typing without a physical device. It tracks finger positions and maps them to specific keys on a virtual keyboard, allowing accurate gesture-based typing. Finger gestures replace the need for a physical keyboard, offering a seamless input method for those with limited mobility. Users can type by pointing or tapping in the air, with the system converting these actions into text, providing a hands-free typing experience. Voice commands combined with finger gestures offer flexible input, allowing users to type or control devices with both methods.

VIII.CONCLUSION

The system demonstrates a touch-free, intuitive method of human-computer interaction by using computer vision techniques to capture real-time hand movements through a webcam. By recognizing specific hand gestures, the technology can control the mouse pointer and perform essential functions like clicking, dragging, and scrolling. This innovation reduces dependency on physical input devices, promoting hygiene and accessibility, especially in environments where hands-free control is desired. It showcases the potential of integrating machine learning and computer vision to create intelligent systems that replicate natural human behavior. The approach lays the foundation for further development of gesture-based interfaces in areas like gaming, augmented reality, virtual reality, and assistive technology.

REFERENCES

[1] Abir Sen, Tapas Kumar Mishra, Ratnakar Dash, "Deep learning based Hand gesture recognition system and design of a Human-Machine Interface," arXiv preprint arXiv:2207.03112, 2022.
[2] Agarwal, A., & Sharma, R., "Hand Gesture Controlled Virtual Mouse Using Artificial Intelligence," International Journal of Advance Research and Innovative Ideas in Education, vol. 9, no. 1, pp. 1234–1240, 2023.
[3] Chaudhary, S., & Patel, M., "Virtual Mouse Using Hand Gesture Recognition," International Journal of Scientific Research in Engineering and Management, vol. 7, no. 2, pp. 89–95, 2023.
[4] Frederic Magoules, Qinmeng Zou, "A Novel Contactless Human Machine Interface based on Machine Learning," arXiv preprint arXiv:1907.04390, 2019.
[5] Katia Lupinetti, Andrea Ranieri, Franca Giannini, Marina Monti, "3D dynamic hand gestures recognition using the Leap Motion sensor and convolutional neural networks," arXiv preprint arXiv:2003.01450, 2020.