iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: XII Month of publication: December 2025

DOI: https://doi.org/10.22214/ijraset.2025.76572

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

HCI with Gesture, Eye Tracking and Voice

Basava Darshan B S[1], Dr. Sankhya N Nayak[2], Benakesh S[3], Bhagath Mohan[4], Chethan J M[5]
[1, 3, 4, 5]*UG students,* [2] *Assoc. Professor, Department of Computer Science and Engineering, JNNCE, Visvesvaraya Technological University, Karnataka, India*

Abstract: The mouse, as an invention by Human-Computer Interaction (HCI) technology, is truly impressive. Still, wireless mouse or Bluetooth mouse uses things such as a battery and a mini dongle which means it isn't truly device-free. In the system being suggested, this limitation can be addressed using computer vision and a camera to sense hand movements and their tips. The system uses a machine learning algorithm within its algorithm. With the hand gestures in place, the computer can be managed virtually and lets you left click, right click, scroll and move the cursor without a physical mouse. The technique uses deep learning to discover the hands in the video. So, the system is intended to decrease the risk of pandemics spread by lessening interaction and not requiring any additional devices to manage computer
Keywords: Human-Computer Interaction, Computer vision, Machine Learning, Pandemics, Deep Learning.

I. INTRODUCTION

Technology is shaping more of our everyday lives these days. Many different types of computer. Many new technologies are gaining traction all over the globe. AI helps to automate many different roles that many humans struggle with or are unable to do themselves. They are they are getting stronger and better and may transform a lot of fields. aspects of things we do. We connect with computers with help of devices like mice. The mouse is a tool for connecting with a GUI through pointing, scrolling and moving around on the screen etc. When dealing with complicated computer or laptop tasks, it's helpful to use a mouse or touchpad. Take up a lot of time, especially if we are moving a hardware mouse everywhere we go damaged sometimes. With a gesture controlled virtual mouse, you can control your computer and interact with it. Easy to use with your hands and through voice directions. Information technology tools can be combined with There is hardly any direct contact. Static can be used to oversee every input and output operation in most cases. and moving your hands together with a voice assistant.

II. LITERATURE REVIEW

Existing virtual input systems demonstrate promising accuracy in controlled settings, but they suffer from lighting/background sensitivity, require markers or specialized sensors, and are typically limited to a single modality (mouse or keyboard or a specific application). There is a lack of integrated, low-latency multimodal systems which combine marker-less hand gestures, eye gaze for coarse cursor positioning, and voice commands for high-level actions while maintaining usability and minimal calibration. This project aims to design and implement a real-time Gesture-Controlled Virtual Mouse and Keyboard integrated with Eye-Controlled Cursor and Voice Bot, addressing robustness to illumination and background clutter, minimizing user instrumentation and calibration, and evaluating the system across diverse user conditions. The outcome will be an accessible, multi-modal desktop control framework with quantified performance.

III. PROBLEM STATEMENT

Traditional human–computer interaction still depends heavily on physical devices such as mice, keyboards, and touchscreens, which limit accessibility for individuals with motor impairments, create hygiene risks in shared environments, and fail to support natural multi-modal communication. Existing virtual input solutions provide only single-modality control and often require specialized hardware, markers, wearables, or perform poorly under varied lighting and backgrounds, making them impractical for real-world usage. To address these challenges, this project proposes an integrated, real-time multimodal HCI system that combines gesture recognition for virtual mouse and keyboard control, eye-tracking for cursor movement, and voice commands for high-level actions, using only a standard webcam without additional sensors. By enabling touch-less, intuitive, low-latency interaction, the solution aims to overcome accessibility barriers, reduce physical contact, support diverse user conditions, and provide a more natural, hygienic, and inclusive method for interacting with computers in healthcare, education, public spaces, and assistive technology environments

IV. PROPOSED METHODOLOGY

This study is concentrated on constructing an intuitive and efficient gesture control system for the improvement of user and computer interaction. Based on the context, the approach used in the system encompasses several stages, each designed to ensure effective gesture recognition and system control.

A. Initialization of Libraries and Webcam

The implementation begins by importing essential libraries such as OpenCV for video processing and MediaPipe for detecting hand landmarks. The webcam is initialized using OpenCV's Video Capture to enable real-time video capture. For detecting and tracking hand landmarks efficiently MediaPipe Hands is set up.

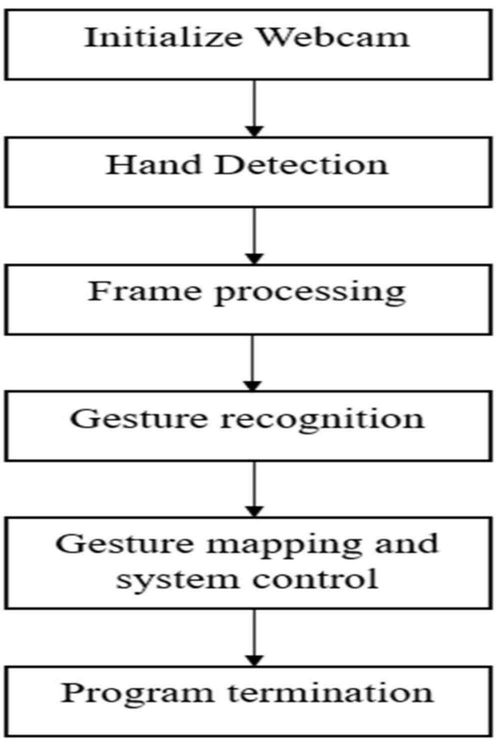


Fig. 1 System Architecture

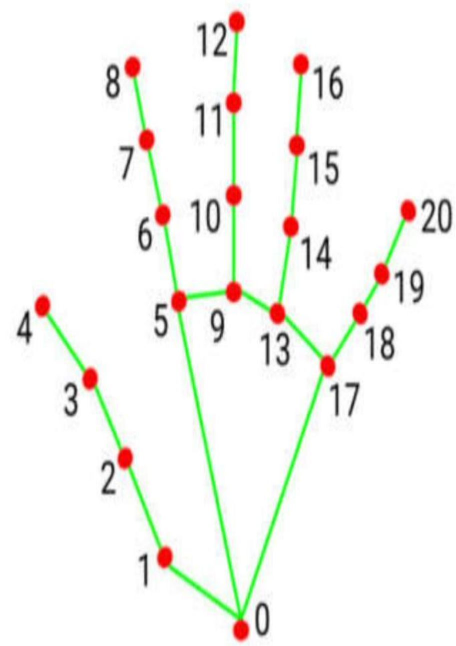


Fig. 2 Hand landmarks detected by MediaPipe

B. Frame Processing and Detection of Hand Landmark

Captured frames from the webcam are preprocessed by converting them into an RGB format suitable for MediaPipe. The MediaPipe Hands module analyzes each frame and detect if there is a presence of hand and identify key landmarks such as fingertips, knuckles, and the wrist. These landmarks are extracted and serve as the foundation for gesture recognition.

Here’s an overview of what such an image would depict:

21 Key Points (Landmarks): Each hand has 21 landmarks detected by Mediapipe, which include points on the tips, joints, and base of the fingers and the palm.

Landmark Positions (from base to tip):
Thumb: 4 landmarks
Index Finger: 4 landmarks
Middle Finger: 4 landmarks
Ring Finger: 4 landmarks
Pinky Finger: 4 landmarks
Palm: 1 landmark (center)

C. Gesture Recognition Based on Hand

Landmarks Using the detected landmarks, distances between specific points on the hand are calculated to determine finger positions. Gestures like pinching, pointing are identified based on these positions. Detected gestures are pinch, index and pinky, v gesture (victory), mid gesture (middle finger), index finger gesture, two finger closed gesture.

0. WRIST
1. THUMB_CMC
2. THUMB_MCP
3. THUMB_IP
4. THUMB_TIP
5. INDEX_FINGER_MCP
6. INDEX_FINGER_PIP
7. INDEX_FINGER_DIP
8. INDEX_FINGER_TIP
9. MIDDLE_FINGER_MCP
10. MIDDLE_FINGER_PIP
11. MIDDLE_FINGER_DIP
12. MIDDLE_FINGER_TIP
13. RING_FINGER_MCP
14. RING_FINGER_PIP
15. RING_FINGER_DIP
16. RING_FINGER_TIP
17. PINKY_MCP
18. PINKY_PIP
19. PINKY_DIP
20. PINKY_TIP

Fig. 3 Hand Landmark points

GESTURE	BINARY	DECIMAL	DESCRIPTION
FIST	00000	0	ALL FINGERS CLOSED
INDEX	01000	8	ONLY INDEX FINGER IS OPENED
FIRST2	01100	12	INDEX + MIDDLE OPEN
V_GEST	01100	33	V-SHAPE (WIDE SPREAD)
PALM	11111	31	ALL FINGERS OPEN
PINCH_MAJOR	CUSTOM	35	THUMB – INDEX PINCH (RIGHT)
PINCH_MINOR	CUSTOM	36	THUMB-INDEX PINCH (LEFT)
TWO_FINGER_CLOSED	CUSTOM	34	INDEX + MIDDLE CLOSE

Fig. 4 Mapping of hand gestures to Binary and Decimal codes

D. Gesture mapping and system control

Predefined gestures are mapped to system actions such as clicks, navigating, or adjusting system settings. Identified gestures are mapped to specific system actions:

1) Pinch Gesture (Right hand): This gesture controls system brightness and volume. When the index finger and thumb come close together (pinching), the system interprets it as a command to adjust brightness or volume. If the movement is horizontal then it is brightness control, if the movement is vertical then it is volume control.
2) Pinch Gesture (Left hand): This gesture controls the scrolling functionality. When the left finger's index finger and thumb come close together (pinching), the system interprets it as a command to perform scroll up or scroll down. If the movement is in upward direction then it triggers scroll up, if the movement is in downward direction then it triggers scroll down.
3) V-Gesture (Victory): Moves the cursor when the middle finger and index finger are straightened in a V-shape, the system moves the cursor to the specified position. Department of CS&E, JNNCE | 27 HCI with gesture, eye tracking and voice
4) Mid Gesture (Middle Finger): Single click action. This gesture is mapped for performing left click at the cursor's position.
5) Index Finger Gesture: Performs right-click action. When the index finger is extended, it triggers a right-click action.
6) Two Finger Closed Gesture: Performs double-click action. When the middle and index fingers are positioned close together, the system interprets it as a double-click command.

The table outlines the gesture-to-action mapping implemented in the gesture-controlled virtual mouse system. Each hand gesture is detected using computer vision techniques and mapped to specific mouse operations through the PyAutoGUI library. The V_GEST (victory/peace sign) enables cursor movement by tracking finger coordinates in real-time, while individual finger gestures trigger distinct mouse events: FIST for click-and-drag operations, MID (middle finger) for left-click, and INDEX for right-click functionality. Advanced gestures such as TWO_FINGER_CLOSED initiate double-click actions, and pinch gestures provide additional controls: PINCH_MINOR for scrolling and PINCH_MAJOR for system-level adjustments like volume and brightness through API calls. This intuitive gesture vocabulary allows users to perform comprehensive mouse operations hands-free, enhancing accessibility and providing a touch less interaction paradigm.

V. RESULTS AND ANALYSIS

This section presents the comprehensive evaluation and analysis of the integrated multimodal system, encompassing the gesture-controlled virtual mouse, virtual keyboard, eye-controlled mouse, and voice-bot. The performance of each module has been tested and assessed based on accuracy, response time, reliability, and user experience. The results validate the effectiveness of combining multiple interaction modalities to create an intuitive and accessible human-computer interface.

The figure 5 presents four functionalities: Gesture-Controlled Virtual Mouse, Virtual keyboard, Eye-Controlled Virtual Mouse and a Chat-Bot (Proton) in a unified interface.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

The figure 6 represents the V-shape hand gesture (Index finger, middle finger) which is responsible to trigger the mouse movement function.

Fig. 5 Final Outlook of GUI

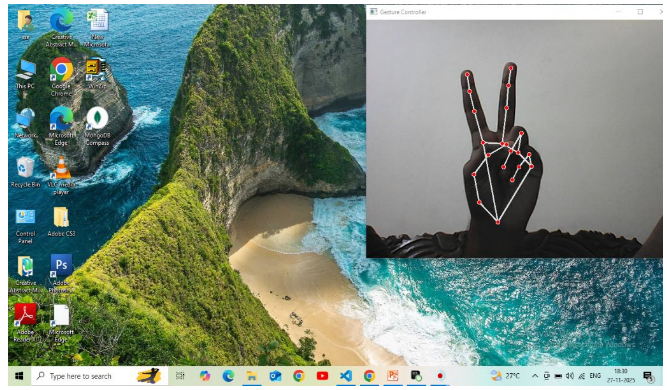


Fig. 6 Mouse movement gesture (V-Shape)

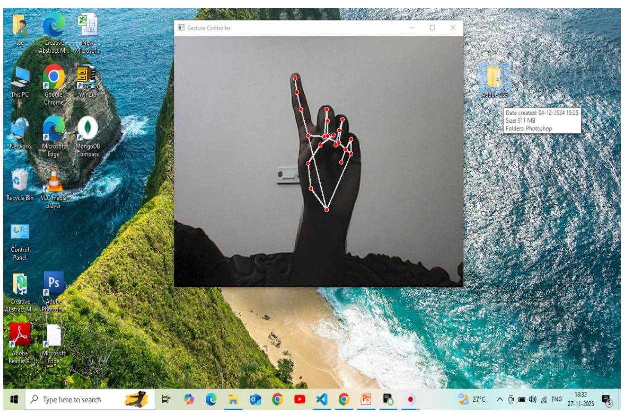


Fig. 7 Right click gesture (Middle finger)

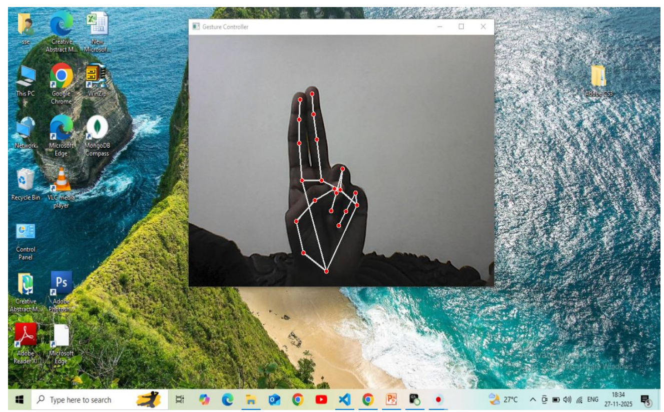


Fig. 8 Double click gesture (Index figure, middle finger)

The figure 7 represents the right click gesture (Middle finger) which triggers the right click function and figure 8 represents the Double-Click gesture (no space between index and middle finger) which triggers double click action.

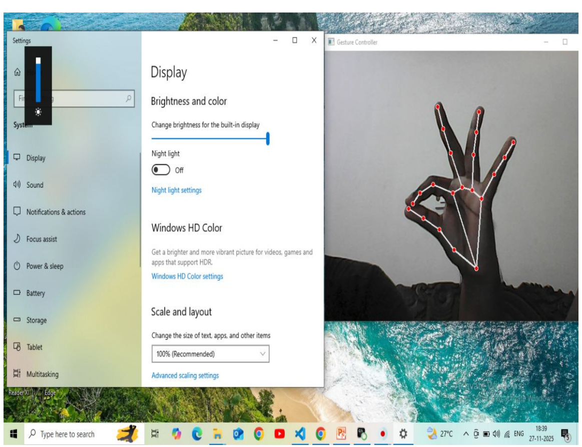


Fig. 9 System Brightness control

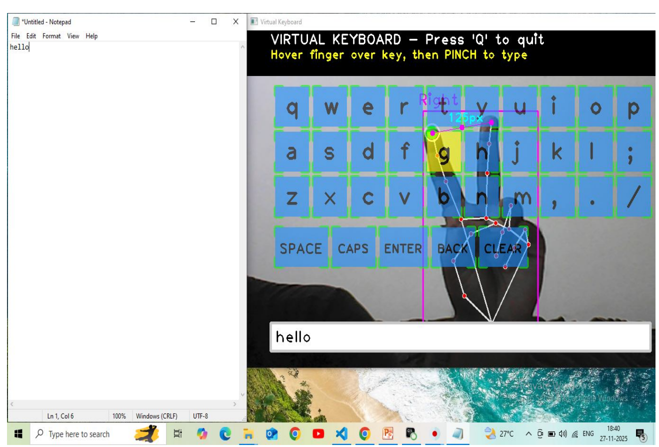


Fig. 10 Virtual Keyboard layout interface

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

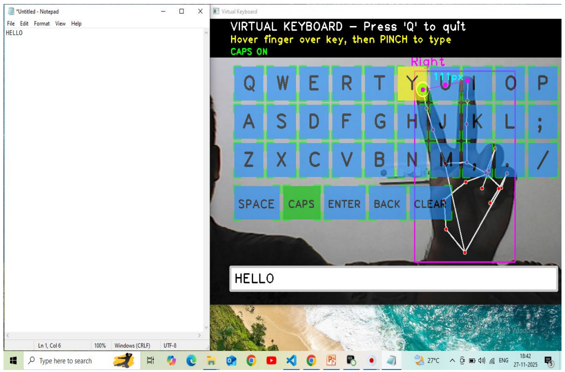


Fig. 11 Caps letter input

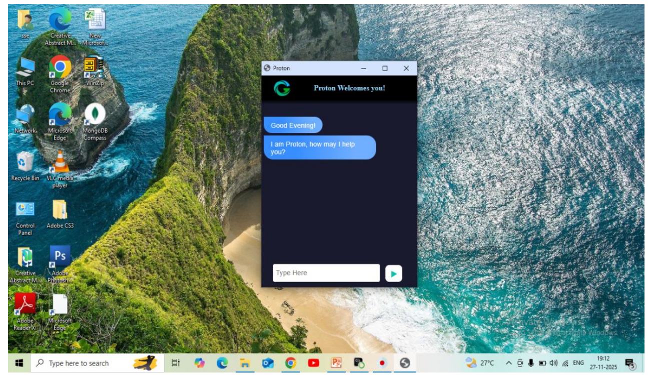


Fig. 12 Chat-Bot GUI

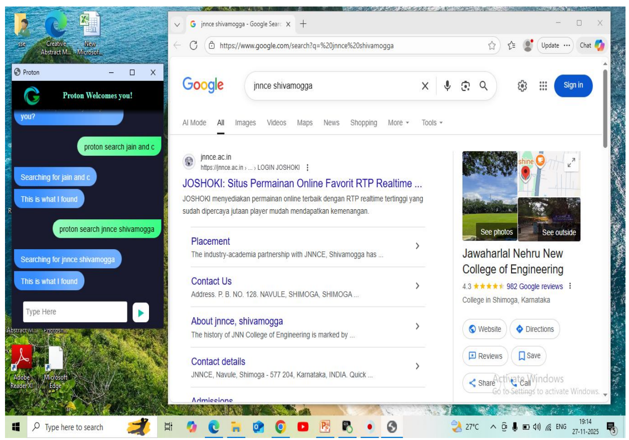


Fig. 13 Google search through Chat-Bot

The figure 9 displays the system brightness control through the same gesture but in this hand movement should be moved in horizontal direction.

The figure 10 represents the virtual keyboard layout interface, displaying an alphabetical (QWERTY) arrangement of keys that users interact with through mid-air finger movements. The interface features a grid-based design enabling touch less text input by hovering and selecting characters through hand gestures captured via webcam, eliminating physical contact requirements.

The figure 11 caps letters are being entered by manoeuvring hand on particular letter boxes.

The figure12 represents the graphical user interface of the Chat-Bot, where in the users can also provide commands in interface beside voice input.

The figure 13 depicts the facilitation of Google search by voice input command through Chat-Bot allowing users to interact intuitively.

This chapter presents the comprehensive results of the integrated system combining gesture-controlled virtual mouse, virtual keyboard, eye-controlled mouse, and chat-bot functionalities. The results demonstrate the successful implementation of various hand gestures for executing mouse operations and keyboard actions, showcasing the system's ability to recognize and respond to distinct gesture patterns. Additionally, the eye gaze tracking mechanism is illustrated, highlighting its accuracy in cursor control through eye movements. The chat-bot interface results reveal the system's capacity to process and execute voice or text commands effectively. Together, these results validate the feasibility and functionality of a multi-modal human-computer interaction system that offers users flexible and accessible control methods.

VI. CONCLUSION AND FUTURE SCOPE

This project presents a more intuitive and touch-free way of interacting with computers using simple hand gestures. By combining computer vision with hand-tracking technology, it eliminates the need for physical peripherals, making digital interaction more accessible and hygienic. The system works in real time and offers decent accuracy, showcasing the potential of gesture-based interfaces in daily computing. It lays the groundwork for more natural and immersive human-computer interaction in the future.

This system has broad real-world impact across multiple domains, including healthcare and medical rehabilitation, where AI-powered multimodal interfaces can help patients with neurological disorders, motor disabilities, and age-related limitations regain independence through touchless computer interaction; extended reality (XR) and spatial computing, enabling natural, hands-free control in VR/AR environments for gaming, training, virtual collaboration, and design; smart home and IoT ecosystems, where gesture, voice, and eye-based commands can serve as a universal controller for seamless, hygienic, and intuitive device management; accessibility-as-a-service platforms, offering cloud-based, customizable interfaces that promote inclusive technology access in education, workplaces, and public spaces; automotive and in-vehicle systems, enhancing safety and usability by allowing drivers and passengers to control infotainment, navigation, and vehicle functions without distraction; industrial automation and manufacturing, where touchless gesture control improves efficiency and safety in hazardous, sterile, or glove-restricted environments; and advanced biometric authentication and security, leveraging eye tracking and gesture recognition for secure, contactless user verification across financial, industrial, and high-security applications.

REFERENCES

[1] T. Moh R. Chitra, A. M. Anusha Bamini, D. Brindha, A. P. Reddy, N. Kezia and G. Beulah, "Hand Gesture Recognition using Shape-based Image Features for Music Controller," 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2022, pp. 1540-1544, doi: 10.1109/ICOSEC54921.2022.9952146.

[2] K. Aggarwal and A. Arora, "An Approach to Control the PC with Hand Gesture Recognition using Computer Vision Technique," 2022 9th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2022, pp. 760-764, doi: 10.23919/INDIACom54597.2022.9763282

[3] Zhi-heng, Wang & Jiang-tao, Cao & Liu, Jinguo & Zi-qi, Zhao. (2017). Design of human-computer. interaction control system based on hand-gesture recognition. 143-147. 10.1109/YAC.2017.7967394.

[4] Chengwei S. R. Chowdhury, S. Pathak and M. D. A. Praveena, "Gesture Recognition Based Virtual Mouse and Keyboard," 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), Tirunelveli, India, 2020, pp. 585-589, doi: 10.1109/ICOEI48184.2020.9143016.

[5] Ladwane, Avantika & Bhandeker, Sharayu & Dhaske, Suraksha & Ugalmugale, Dipali & Rane, Mugdha. (2023). OPERATING VIRTUAL MOUSE AND KEYBOARD USING GESTURE RECOGNITION. 10. 2349-5162.

[6] K. H. Shibly, S. Kumar Dey, M. A. Islam and S. Iftekhar Showrav, "Design and Development of Hand Gesture Based Virtual Mouse," 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), Dhaka, Bangladesh, 2019, pp. 1-5, doi: 10.1109/ICASERT.2019.8934612.

[7] M. Vidya, S. Vineela, P. Sathish and A. S. Reddy, "Gesture-Based Control of Presentation Slides using OpenCV," 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), Trichy, India, 2023, pp. 1786-1791, doi: 10.1109/ICAISS58487.2023.10250520.

[8] J. Hossain Gourob, S. Raxit and A. Hasan, "A Robotic Hand: Controlled With Vision Based Hand Gesture Recognition System," 2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI), Rajshahi, Bangladesh, 2021, pp. 1-4, doi: 10.1109/ACMI53878.2021.9528192.

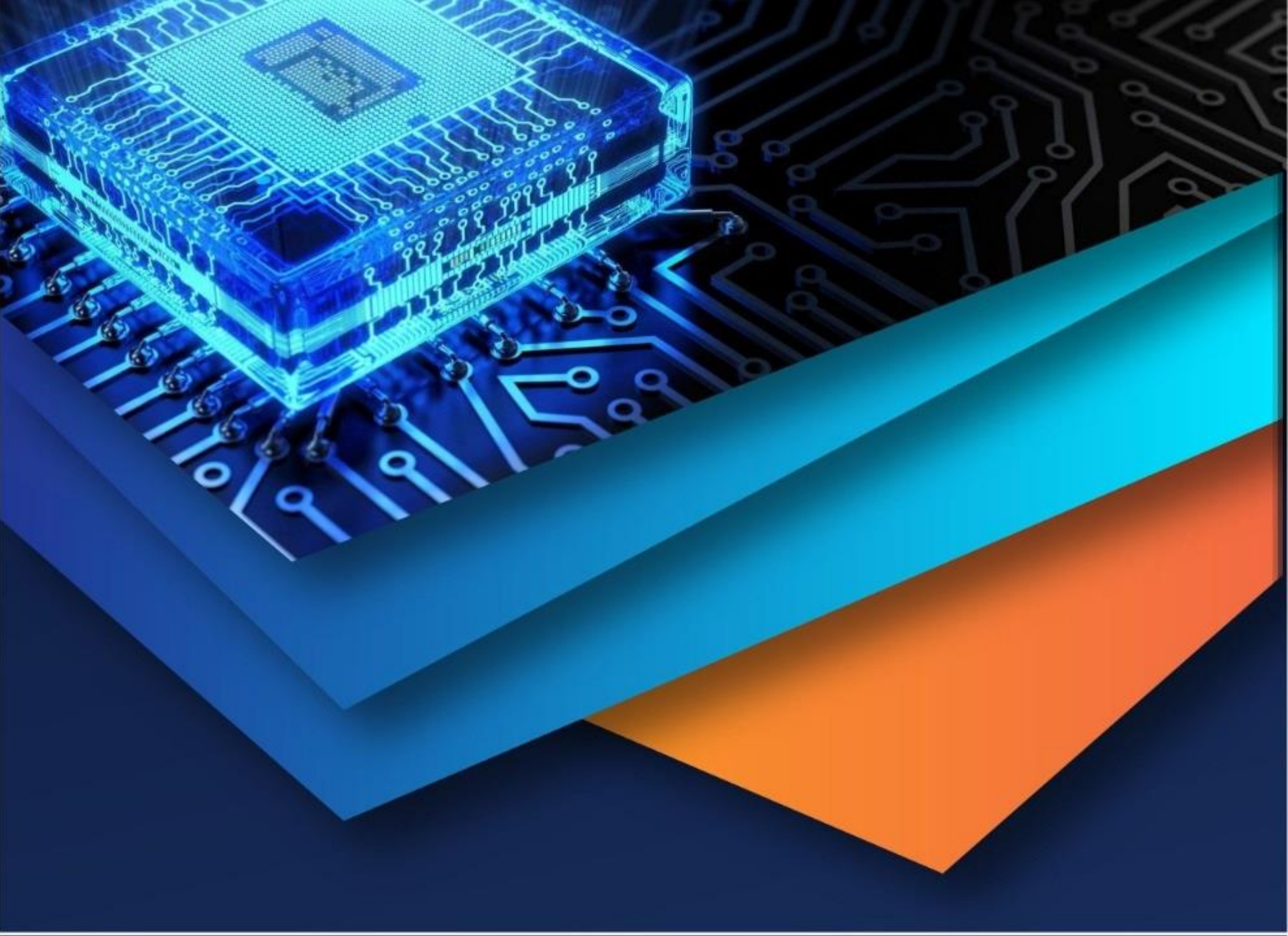
[9] Shibly, Kabid & Dey, Samrat & Islam, Md. Aminul & Showrav, Shahriar. (2019). Design and Development of Hand Gesture Based Virtual Mouse. 10.1109/ICASERT.2019.8934612.

[10] Kotti, J. ., Padmaja, B., & Deepa, D. (2024). Enhancing Gesture-Controlled Virtual Mouse and Virtual Keyboard Using AI Techniques. Journal of Mobile Multimedia, 20(02), 437–494.

[11] Y. Zhao, X. Ren, C. Lian, K. Han, L. Xin and W. J. Li, "Mouse on a Ring: A Mouse Action Scheme Based on IMU and Multi-Level Decision Algorithm," in IEEE Sensors Journal, vol. 21, no. 18, pp. 20512-20520, 15 Sept.15, 2021, doi: 10.1109/JSEN.2021.3096847.

[12] N. Bhuvaneswary, M. Sathish, S. S. Abhayankar, K. S. V. Barath and P. S. N. Reddy, "Real Time Eye Blink Recognition for Virtual Mouse Using Raspberry Pi," 2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), Bengaluru, India, 2025, pp. 972-975, doi: 10.1109/IDCIOT64235.2025.10915003.

[13] L, Sai & L, Kiran. (2024). A1 Voice Assistant Using Python and API. 1-6. 10.1109/ADICS58448.2024.10533536.



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)