Journal of the ACS Advances in Computer Science
Volume 15, Issue 1, 2024

Virtual Keyboard-Mouse in Real-Time using Hand Gesture and Voice Assistant

Dr/ Shaimaa Othman[1], Dr/ Hany Maher Sayed Lala [2], Yosr Mansour[1]

[1]Higher Institute of Computers and Information Technology, Computer Depart., El. Shorouk Academy, Cairo, Egypt

[2]Department of Mathematics, Faculty of Science, Al-Azhar, Cairo, Egypt

Email: dr.Shimaa.osman@sha.edu.eg, hmsayed@azhar.edu.eg , yosr.mansour @sha.edu.eg

Abstract

The advancement of computer applications has increasingly facilitated everyday tasks, with recent innovations focusing on voice assistants and virtual input devices. This technology is particularly beneficial for individuals with mobility challenges or in situations where direct manual computer interaction is limited. Leveraging computer vision and artificial intelligence, these applications can interpret visual data, such as human movements, and make decisions to execute corresponding commands. This research incorporates a voice assistant, virtual mouse, and virtual keyboard to enhance accessibility and usability, especially for individuals with physical disabilities or those who prefer alternative input methods. Using Python, MediaPipe, and OpenCV, the application effectively processes and interprets user gestures, providing a responsive and efficient computing experience. MediaPipe's features specifically contribute to the model's precision, optimizing hand tracking and gesture recognition for AI-driven tasks. Users can control the computer cursor through various hand gestures, type on a virtual keyboard using colored caps or tapes and perform essential actions like left-clicking and dragging items. This integrated solution aims to improve productivity, making computers more accessible and enhancing the overall digital experience for users. The fusion of AI and computer vision in such applications continues to drive forward innovative and inclusive computing solutions, promising a future of greater accessibility and convenience in human-computer interaction.

1. INTRODUCTION

In recent years, the subject of human-computer interaction has seen amazing advancements aimed at improving the user experience and accessibility. The advent of smart technologies and natural user interfaces has revolutionized how humans interact with machines. Traditional input devices, such as keyboards and mice, while efficient, have limitations in dynamic environments requiring hands-free operation or where physical space is constrained. These limitations have spurred the development of innovative human-computer interaction (HCI) methods that leverage natural gestures and voice commands. This research proposes a virtual keyboard-mouse system that combines real-time hand gesture recognition with a voice assistant, creating a seamless, intuitive, and accessible interaction platform.

The primary motivation for this work stems from the increasing demand for more inclusive and adaptable interaction solutions. Conventional input methods can be challenging for individuals with physical

DOI: 10.21608/ASC.2025.333048.1025

disabilities or in environments where touch-free operations are critical, such as healthcare, manufacturing, and remote work. Integrating gesture recognition and voice control offers a compelling alternative, enabling users to interact with digital systems in a manner that feels natural and requires minimal learning. This can lead to increased productivity, efficiency, and user satisfaction.

This research aims to explore developing a method for interacting with computers using hand gestures and voice commands, by which the researchers hope to provide an inclusive and accessible computing experience for a wider range of users.

By incorporating hand gestures and voice assistant technology, this research seeks to offer an alternative and intuitive method of controlling computers and devices. Hand gestures can provide a natural and expressive means of interaction, allowing users to perform various actions by simply moving their hands. Moreover, integrating voice assistant capabilities enables users to issue commands and control devices through voice commands, making the interaction process even more seamless and efficient.

Furthermore, the research paper aims to investigate the real-time capabilities of the proposed system. Real-time interaction is crucial in various domains, such as gaming, virtual reality, and augmented reality. By enabling real-time functionality, the researchers aim to enhance the user experience and enable seamless integration of the virtual keyboard-mouse system into various applications and scenarios.

The objectives of this research are twofold. Firstly, to design and develop a virtual keyboard and mouse system that can accurately interpret hand gestures and convert them into corresponding actions, such as cursor movement and clicking. This requires the creation of algorithms and models capable of recognizing and analyzing various hand gestures in real-time. Secondly, to integrate a voice assistant component that can understand and execute voice commands related to device control. To achieve this, speech recognition and natural language processing technologies must be implemented.

The motivations for the research on virtual keyboard-mouse using hand gesture and voice assistant revolve around improving usability, accessibility, natural interaction, and real-time capabilities of human-computer interaction. By addressing these motivations, the research aims to contribute to the advancement of human-computer interaction technologies and provide users with a more intuitive and efficient way of interacting with computers.

The rest of this paper is organized as follows: literature review is presented in Section 2. The system implementation is explained in Section 3, and the results are discussed in Section 4. Finally, the conclusion is drawn in Section 5.

2. LITERATURE REVIEW

Recently, interest in developing natural interactions between computers and humans has grown significantly. Numerous studies have explored universal computing's human-computer interface. Among these, the vision-based interface technique stands out for its ability to capture motion information from video input without requiring costly equipment. Consequently, this method is regarded as an effective tool for designing human-computer interfaces. This section conducts a comprehensive review of existing literature, highlighting the transformative impact of AI on computer-human interaction and its diverse applications.

2.1 Hand gesture

Hand gestures are performed by making specific hand movements that are recognized by a hand gesture recognition is typically done using computer vision techniques and machine learning algorithms [16]. First, the hand region is detected and tracked in real-time video using computer vision techniques like background subtraction, skin color segmentation, or a combination of both.

Many practical applications, such as virtual mice, remote controllers, sign-language recognition, and immersive gaming technologies, use fingertip detection. Therefore, one of the primary objectives of vision-based technology over the past few decades, particularly with conventional red-green-blue (RGB) cameras, has been virtual mouse control via fingertip recognition from pictures [1].

Numerous RGB-D sensor types, such the Kinect V2, or VicoVR [2], among others, can provide body tracking. Among them, the low-cost, CPU-free Kinect V2 is becoming increasingly popular these days. Convolutional neural network (CNN)-based RGB-D image-based systems have lately demonstrated exceptional performance in HCI [3,4]. However, for model rendering and assessment, these systems need high-performance CPUs and a bigger dataset.

Utilizing fingertip sensing and RGB-D pictures is considered a novel virtual mouse technique. Without using a mouse, gloves, or markers, the user interacted with the computer by moving their fingertip in front of a camera. The method showed off not just extremely precise gesture estimations but also useful applications [1].

2.2 Gesture Recognition Techniques

Gesture recognition forms the backbone of virtual keyboards and virtual mice operated by hand gestures [15]. Researchers have explored diverse methodologies, including computer vision and deep learning algorithms, to enhance accuracy [17]. Noteworthy studies include Lin et al.'s work, employing a combination of CNNs and LSTM networks for real-time hand gesture recognition [6]. Similarly, Zhang et al. proposed a method that integrates depth information for precise gesture recognition [5].

Shankar et al. have introduced two experiments. One involves controlling objects using hand motions, while the other involves identifying the symbols we write in the air. In the first, a ground-breaking method for manipulating the mouse cursor is introduced. By fusing motion modelling, pattern recognition, and machine learning, this technique perfectly supports a range of mouse movements. The second research explores air-writing recognition, a novel technique that involves drawing letters and numbers by hand, in conjunction with the first [12].

2.3 Virtual Mouse using Hand Gestures

Virtual mouse systems, such as the one explored in the study [7], represent a significant leap in HCI. By utilizing hand gestures, users can control the computer cursor intuitively. These systems commonly employ depth-sensing cameras and machine learning algorithms. The research underscores the precision and responsiveness achieved by mapping hand gestures to cursor movements, laying the foundation for seamless interaction.

Shetty et al. constructed a virtual mouse system using color detection [11].They used webcam for detecting mouse cursor movement and click events using OpenCV built-in functions. A mouse driver, written in java, is required as well. This system fails to perform well in rough background.

In the study [8], a method for hand gesture identification is presented, and computer vision techniques are used to construct a virtual mouse and keyboard with hand gesture recognition. To realistically control the computer, full keyboard functions, mouse pointer movement, and click events are incorporated. All the inputs taken into consideration have their recognition and response rates computed and shown in the findings. Comparing the accuracy of the proposed strategy to other cutting-edge algorithms reveals that it performs better with a 95% accuracy.

A multidimensional approach to HCI may be seen in the convergence of hand gestures and voice commands as shown in [9]. Users may interact with computers virtually by fusing several modalities, which improves conventional user interfaces. The study emphasizes the connection between movements and voice, demonstrating how subtle gestures combined with context-aware voice instructions allow for complex and flexible interactions.

2.4 Virtual Keyboard using Hand Gestures

A finger recognition and gesture based augmented keyboard system is introduced in [13]. The system was developed using OpenCV libraries and Python. Palm detection is used for typing on the augmented keyboard. Virtual Keyboard performs based on the movement of the finger.

Phursule et al. developed an optical mouse and a virtual keyboard based on hand and face gestures using computer vision. The system enables the computer's mouse or pointer to move according to gesture movements, with specific gestures performing right and left clicks. Additionally, various gestures can execute keyboard functions. The project relies solely on a camera for hardware and uses Python programming within the Anaconda platform [14].

3. SYSTEM IMPLEMENTATION

The aim of this paper is to implement a computer application that uses alternative methods of keyboard and mouse cursor control for rehabilitation purposes, allowing stroke patients to recover from. Therefore, we propose a new keyboard and mouse cursor control system based on visual and color recognition techniques that utilize hand gestures recorded by a webcam. A system, which incorporates cutting-edge computer vision, machine learning, and natural language processing technologies. By interpreting hand gestures collected by a camera and processing spoken commands through a voice assistant.

Modules partitions

First, as we declared in the architecture of our project in Figure 1, we'll be using Pyqt5 to design and activate the interface of the application to make it usable on Desktop systems using Python programming language.

The figure represents a system architecture for an integrated application that combines a voice assistant, virtual keyboard, virtual mouse, and camera functionality, all within a Windows application environment. The diagram outlines how these components interact with each other, leveraging deep learning models and cloud-based platforms for enhanced functionality and training.

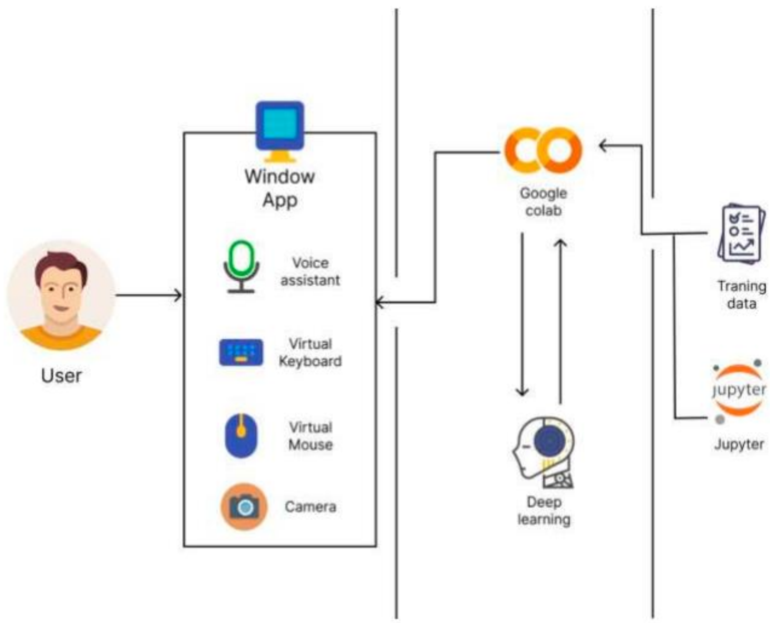

Figure 1 project Architecture

The application implementation consists of two main modules Hand Gesture and Voice Assistant.

3.1 Hand Gesture Module:

Creating a virtual keyboard using hand gestures involves using computer vision techniques to detect and recognize hand gestures in real-time. Using MediaPipe is an open-source framework developed by Google that provides a comprehensive solution for building real-time multimedia processing pipelines.

a. Capture Video Stream, Process Frames and Recognize Hand Gestures

In this step, you capture frames from the webcam, convert them to RGB (as MediaPipe requires RGB input), and process the frames to detect hands using the MediaPipe Hands model.

b. BlazePalm Detector

- We train a palm detector instead of a hand detector to localize hand landmarks in real-time from a single image or video frame.
- The model is based on a two-stage approach (hand region and hand landmarks)

c. Hand Landmark Model

- After running palm detection over the whole image, our subsequent hand landmark model performs precise landmark localization of 21.
- The model has three outputs.
    - 21 hand landmarks
    - A hand flag indicating the probability of hand presence in the input image.
    - A binary classification of handedness, e.g., left, or right hand.

d. Recognize Gestures

In this step, you analyze the positions of hand landmarks to recognize specific gestures. Implement the logic for recognizing gestures like thumbs up, peace sign, etc.

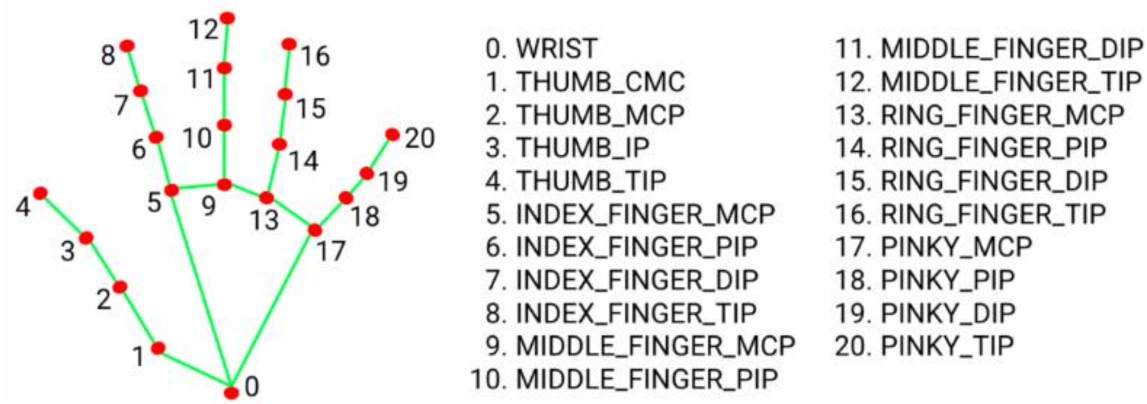

Figure 2 hand landmark

Following the successful development of our hand gesture recognition system, we implemented this technology into four specialized sub-modules to enhance user interaction and control:

3.1.1 Virtual Keyboard algorithm

System must capture images via camera. The system provides the user with typing by hand gestures such as the hardware keyboard. Here I used point number 12 to define an indicator and point No. 8 and 7 in the action process.

- The system must allow the user to "click" on the key.
- The system must display the correct output for the key entered by the user.
- Functions provided by the system to the user:
    - I. Letters from A to Z
    - II. Numbers from 0 to 9
    - III. Enter button.
    - IV. Delete button.
    - V. Spacebar
    - VI. Capitalization

3.1.2 Virtual mouse algorithm

- Use the camera to detect hand gestures through the 21 points on the hand.
- Know if the finger is closed (0) or open (1).
- By using points 8, 12, and 20, I can control the mouse.
- And control the action of the mouse by closing and opening points 8 and 12.
- The system provides control of the mouse by hand gestures and meets the needs of the user, such as the hardware mouse.
- Functions provided by the system to the user:
  i. Click on the left mouse button.
  ii. Click on the right mouse button.
  iii. Move the mouse pointer right, left, down and up That's like a hardware mouse)

3.1.3 Media control player

- Use the camera to detect hand gestures through the 21 points on the hand.
- Stop Video by points 4, 8, 12, 16 and 20.
- Video Plus Five Sec by points 8
- Video Minus Five Sec by points 8, 12
- Increase video volume by points 8, 12 and 16
- decrease video volume by points 8, 12, 16 and 20

3.1.4 Scroll

- Uses gestures to facilitate scrolling through documents and web pages, providing a seamless browsing experience.
- Using virtual mouse can be scroll in pc by points 8, 12 and 16.

3.2 Voice Assistant Module

- Allows access to the computer's mic when the user is talking.
- The user's voice is converted into text and processing is done, and then the required commands are executed.
- The system must provide the user with the ability to carry out the command requested by the user and talk to the computer by voice.
- Functions provided by the system:
  a. Internet searches
  b. Search inside the device for what the user wants, whether it is an audio file, video or text.
  c. Date >today-yesterday -last week -last year
  d. Time
  e. Weather
  f. Open google &YouTube.
  g. Write notes.
  h. Weather state

i. Send e-mails.

4. RESULTS & TESTING

In this paper, we comprehensively evaluated the performance and accuracy of various interactive models, including a voice assistant, a virtual keyboard, and a virtual mouse model. Our testing approach involved assessing a total of 15 voice assistant functions, each tested five times to ensure consistency. Similarly, we evaluated the virtual keyboard input model's ability to recognize and process commands for typing letters and numbers, along with control functions, with each function undergoing ten trials. Additionally, we assessed the virtual mouse model's performance in executing actions, also using ten trials for each function. Overall, our tests aimed to identify strengths and areas for improvement in these models, ensuring their reliability and effectiveness in practical applications.

4.1Voice assistant model test

This evaluation assessed voice assistant functions for accuracy, with core tasks like browser navigation, date retrieval, and virtual device activation achieving 100% success. Weather updates, speech-to-text, and note-taking functions performed well, though with slight inconsistencies, resulting in an 80% accuracy. Email and file search functions were mostly accurate, with minor issues in email generation. Overall, core functions excelled, while areas like speech recognition need refinement.

Name of function	input	output	task	No.of trails	Right output	percent age
open_goo gle	Voice	No	Open the google search at the browser	5	5	100 %
open_yout ube	Voice	No	Open the YouTube at the browser	5	5	100 %
speak	text	Voice	Turn the text into voice	5	4	80 %
today	Voice	Date (text)	Give you the date of the current day	5	5	100 %
yesterday	Voice	Date (text)	Give you the date of the previous day	5	5	100 %
last_year / last_week	Voice	Date (text)	Give you the date of the previous year /week from the current day	5	5	100 %
time	Voice	time (text)	Give you the time now	5	5	100 %
Weather	Voice	text	Give you the temperature degree and the state of the weather now	5	4	80 %
search	Voice	no	Search for the input text on google engine and	5	5	100 %

			display the results it in the browser			
rec_audio	Voice	text	Turn the voice into text	5	4	80 %
Make_note	Voice		It takes a voice command and turn them into txt file	5	4	80 %
E_mail	Voice and text (Voice receiver)	No	It takes a voice commands and turn them into email and send it to specific user(receiver)	5	4	80 %
file	email Voice (file that you want to find)	List of matched files	It searches in the whole hard desk to find the matched files	5	5	100 %
virtual_mouse	Voice	No	It opens the virtual mouse program	5	5	100 %
virtual_keyboard	Voice	No	It opens the virtual keyboard program	5	5	100 %

Table 1 Voice assistant test

4.2Virtual Keyboard model test

The Virtual Keyboard model test showed high accuracy across key functions. Capital letters, small letters, and enter commands achieved 100% success, while numbers, backspace, caps lock, and space functions demonstrated minor inconsistencies, scoring between 80-90%. Overall, the model effectively performed core typing tasks with strong reliability.

Name of function	input	output	task	Num of trails	Right output	percentage
Capital Letters	Video in real-time	A-Z	Write Sentences	10	10	100%
Small-Letters	Video in real-time	a-z	Write Sentences	10	10	100%
Numbers	Video in real-time	0-9	Write Numbers	10	8	80%
Backspace	Video in real-time	Backspace	Delete last character	10	9	90%
Caps lock	Video in real-time	Caps lock	Convert the letters	10	9	90%

Enter	Video in real-time	Enter	Press Enter Buttons	10	10	100%
Space	Video in real-time	Space	Make Space	10	9	90%

Table 2 Virtual Keyboard test

4.3Virtual Mouse model test

The Virtual Mouse model test demonstrated strong performance, with right-click, left-click, and stop-pointer functions achieving 100% accuracy. Double-click, scroll, and pointer movement functions showed minor inconsistencies, scoring between 80-90%. Overall, the model reliably executed core mouse actions with high precision.

Name of function	Input	output	task	Num of trails	Right output	percentage
Right-Click	Video in real-time	Right-Click	Make Right click	10	10	100%
Left-Click	Video in real-time	Left-Click	Make Left Click	10	10	100%
Double-Click	Video in real-time	Double-Click	Make Double-Click	10	8	80%
Scroll-UP	Video in real-time	Scroll-Up	Make Scroll Up In Files & PDF	10	9	90%
Scroll-down	Video in real-time	Scroll-Down	Make Scroll Down in Files & PDF	10	9	90%
Stop-Pointer	Video in real-time	NO	Give you the date of the per	10	10	100%
Move-Pointer	Video in real-time	Mouse Move	Make the mouse move	10	9	90%

Table 3 Virtual Mouse test

5. Conclusion and Future Work

This research proposed a cutting-edge framework for a real-time virtual keyboard and mouse system that combines hand gesture recognition and voice assistant technology, offering an intuitive and accessible method for HCI. The system employs computer vision for recognizing gestures and natural language processing to handle voice commands, ensuring smooth and effective user control over digital interfaces.

The study outlined the primary motivations for this work, emphasizing the need for enhanced HCI techniques and addressing the shortcomings of conventional input devices. The research focused on developing a system that integrates gesture recognition with voice-based controls, aiming to advance efficient and user-friendly interaction methods.

Experimental results demonstrated the system's remarkable accuracy in gesture recognition and its prompt responsiveness to voice commands in diverse user scenarios. While the voice assistant model performed well in its core functionalities, its speech recognition capabilities require further enhancement. In comparison, the virtual keyboard model proved highly reliable for fundamental typing tasks, and the virtual mouse consistently delivered precise execution of essential mouse functions. These findings underscore the potential of multimodal input approaches to improve HCI flexibility and accessibility.

The study's primary contribution lies in merging gesture and voice-based interaction into a unified framework, overcoming the limitations of traditional input devices and single-modality systems. By enabling hands-free operation and reducing dependence on physical hardware, this system holds promise for applications in accessibility, gaming, virtual reality, and remote work. Additionally, its real-time adaptability across diverse contexts offers valuable insights for developers and researchers designing more intuitive and inclusive HCI systems.

Future research could expand this framework by exploring more complex gestures, multi-user capabilities, and adaptive machine learning models to enhance scalability and accuracy. Further advancements, such as integrating the system into wearable devices or augmented reality platforms, could unlock novel real-world applications. Ultimately, this research contributes to the evolution of HCI technologies, paving the way for a more inclusive and seamless digital experience.

References

[1] Tran, DS., Ho, NH., Yang, HJ. et al. Real-time virtual mouse system using RGB-D images and fingertip detection. Multimed Tools Appl **80**, 10473–10490 (2021).
[2] Abhilash S S, Lisho Thomas, NWCC (2018) Virtual Mouse Using Hand Gesture. International Research Journal of Engineering and Technology (IRJET)
[3]Jiang D, Li G, Sun Y, Kong J, Tao B (2019) Gesture recognition based on skeletonization algorithm and CNN with ASL database. Multimed Tools Appl 78:29953–29970
[4] Ge L, Liang H, Yuan J, Thalmann D (2018) Robust 3D hand pose estimation from single depth images using multi-view CNNs. IEEE Trans Image Process 27:4422–4436

[5] Li Y, Zhang P. Static hand gesture recognition based on hierarchical decision and classification of finger features. Sci Prog. 2022 Jan-Mar;105(1):368504221086362. doi: 10.1177/00368504221086362. PMID: 35296188; PMCID: PMC10358564.
[6] Guo, Lin & Zongxing, Lu & Yao, Ligang. (2021). Human-Machine Interaction Sensing Technology Based on Hand Gesture Recognition: A Review. IEEE Transactions on Human-Machine Systems. 51. 300-309. 10.1109/THMS.2021.3086003.
[7] R. Matlani, R. Dadlani, S. Dumbre, S. Mishra and A. Tewari, "Virtual Mouse using Hand Gestures," *2021 International Conference on Technological Advancements and Innovations (ICTAI)*, Tashkent, Uzbekistan, 2021, pp. 340-345, doi: 10.1109/ICTAI53825.2021.9673251.
[8] 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.
[9] Sharma, M., Akilesh, G., Vishwaa, V. S., Sharon Femi, P., & Kala, A. (2022). Virtually controlling computers using hand gesture and voice commands. Journal of Current Research in Engineering and Science, 5(17).
[10] Gupta, D., Hossain, E., Hossain, M. S., Hossain, M. S., & Andersson, K. (2021). An interactive computer system with gesture-based mouse and keyboard. In Intelligent Computing and Optimization: Proceedings of the 3rd International Conference on Intelligent Computing and Optimization 2020 (ICO 2020) (pp. 894-906). Springer International Publishing.
[11] Shetty, S., Yadav, S., Upadhyay, R., Bodade, V.: Virtual mouse using colour detection (2016)
[12] Shankar, A., Bondia, A., Rani, R., Jaiswal, G., & Sharma, A. (2024, January). Gesture-Controlled Virtual Mouse and Finger Air Writing. In 2024 14th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 370-375). IEEE.
[13] Jagannathan, M. J., Surya, M., BT, A. M., & Poovaraghavan, R. J. (2018). Finger recognition and gesture based augmented keyboard.
[14] Phursule, R. N., Kakade, G. Y., Koul, A., & Bhasin, S. (2023, August). Virtual Mouse and Gesture Based Keyboard. In 2023 7th International Conference on Computing, Communication, Control And Automation (ICCUBEA) (pp. 1-4). IEEE.
[15] Yadav, K. S., Anish Monsley, K., & Laskar, R. H. (2023). Gesture objects detection and tracking for virtual text entry keyboard interface. Multimedia Tools and Applications, 82(4), 5317-5342.
[16] Godbole, A., Gondke, V., & Devadkar, K. K. (2022, December). Hand Gesture Identification and Voice Command Based Hardware Reduction. In 2022 IEEE Pune Section International Conference (PuneCon) (pp. 1-6). IEEE.
[17] Patel, U., Rupani, S., Saini, V., & Tan, X. (2022, November). Gesture Recognition Using MediaPipe for Online Realtime Gameplay. In 2022 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT) (pp. 223-229). IEEE.