Grenze International Journal of Engineering and Technology, January Issue

Gesture Controlled Physical Therapy System for Upper Body using Mediapipe and OpenCV

Mangesh Balpande[1], Raj Nishad[2], Uday Lund[3], Rohit Patil[4], Prafull Patil[5] and Sagarkumar Badjate[6]
[1-6]Department of Information Technology, SVKM IOT, Dhule, India
Email: mangesh.balpande111@gmail.com, rajrnishad11@gmail.com, udaypunjabi3@gmail.com, rohitpatil200528@gmail.com, prafullpatil357@gmail.com, sagar.badjate@gmail.com

***Abstract*— Using this project we are introducing a gesture-based gaming system which will help people in their physiotherapy and physical rehabilitation. There are many tools which can be used for physiotherapy but they are either difficult to use or they lack in motivating the patients, which results in not total participation from the people. To overcome these issues we have developed a system which use technologies like Python, OpenCV and Media-pipe which detects the upper body gestures from video-input. Then these gestures are converted to keyboard inputs or commands, which makes it possible for users to control the games without traditional devices. This project aims to be useful to a wide range of needs, by making the physiotherapy rehabilitation more playful and enjoyable.**

***Keywords*— Accuracy, Gesture recognition system, Mediapipe, OpenCV, Physiotherapy, Video feed.**

I. INTRODUCTION

People who have problems in moving because of any injury, nerve problems or from any other accidents, for people like this physical rehabilitation is very important so that they can recover form their injuries and move again like they used to before. But the systems and devices which are available right now are not very useful and motivating for the people and some traditional devices are very much hard to use and not fun especially for people who already find rehabilitation as a difficult task.

But in todays time there are emerging systems where the people or patients can control things just by moving their upper-body without touching anything which is very helpful and efficient. These systems even help hospitals in their own way since the patients does not have to touch any equipment which leads to no spread of any contagious disease.

This project is about a gaming system which lets people do their physical rehabilitation in a fun and enjoyable way which keeps them motivated. This project is built using Mediapipe, Open-CV, and Python. In this project, first we capture the video frame by frame and then process the video by filtering out any unwanted noise and then finding the landmarks form the frames and connecting it to that specific function, OpenCV helps make the video clean and Mediapipe helps to figure out what gestures are needed. After the gestures are recognized then they are connected to their respective keyboard presses to control the game. This project contains 5 games which have their own specific upper-body gestures which helps people with different rehabilitation needs do their own specific exercises.

Grenze ID: 01.GIJET.12.1.513

The unique thing about this project is that it is fun and enjoyable unlike other traditional systems and devices which are not so motivating and not fun which may lead the patient to find the rehabilitation boring and avoid doing exercises which leads to longer time to recover. By using this system the patients can combine fun with their rehabilitation exercises.

II. Related work

Systems which are based on FPGA they enable real time recognition static hand gestures using features like area, perimeter, and thumb detection which are shape-based detection features. These systems using illumination compensation and skin color segmentation achieve 94% accuracy, the only downfall of these systems is that they can only be used for predefined hand positions and gestures. whereas Kinect based systems help students to enable physical activity and learning using motion detection by PANEGA, which is scalable and flexible. Even with these benefits the system needs expert help for efficient use [1][2].

Gesture-based control for security systems, multimedia and associating gestures with different actions is done with depth cameras with Domatiq Intelligent Building Systems. It helps user for improving feedback for efficient usability and ensuring the accuracy among different users and other settings remains a challenge. TANGO:H is a gamified platform for rehabilitation which personalizes different exercises for different users based on their profiles and skills and their interaction history which in turn improves engagement from people. But its Kinect dependence limits accessibility in settings of low
resource [3][4].

MathMazing integrates Kinect for arithmetic learning in maze game using gesture-based movements. It planned to use cost effective webcams instead of Kinect which will lead to more specific gestures and improving accessibility by incorporating multiple difficulty levels. But there are certain challenges such as accuracy of gestures with webcams and maintaining gameplay robust mechanism. Similarly, another game which uses upper-body movements which is Take Drunkard for gameplay, controlling characters through facial movements. However problems like motion sickness and fatigue of players persist [5][6].

Hand gestures are recognize, detect and segment using Image processing systems for interaction between human and computer using Haar-like features for Camshift and classification for tracking purpose. Another research paper blob-based hand recognition combined with face recognition for file management and scrolling using gestures. It provides security by only allowing authenticated users but struggles in colorful backgrounds and bad lighting conditions [7][8].

Genetic algorithms and Otsu thresholding are used for recognition using a wall display system in which we use bare-hand gestures for interaction with system. It achieves 93.35% of recognition rate by classifying gestures using finger count which uses convex hull methods. Similarly, a system which enables static hand gesture recognition with better accuracy using a two-pipeline architecture which combines hand bounding boxes with keypoint detection [9][10].

An exergame is a system which combines webcam gesture recognition with ergo-bike for real time control of the game. It achieves user independent recognition of gestures using a Bag-of-Features model and SVM. Similarly, a model for detecting hands and camshaft which uses Haar-like features that are vision-based for reliable and efficient gesture recognition for browsing on servers and gaming. It has its own limitations which are background noise and lighting changes [11][12].

Another system which works with combination of ergo-bike with physical exercise with webcam help for recognition of gesture for the real time control of the game. Another model using Haar-like features which uses vision based system and camshift which lets us improve the flexibility and its applications for recognition of the gestures. It also faces the same problems which are bad lighting conditions and noise in background [13][14].

A system which uses input from a single camera which is used for real-time recognition of the gesture which focuses on problems like hand rotation and occlusion. For identifying the key hand edges it uses a cylindrical modeling for fingers and a convolution operator which helps us in accurate and precise posture segmentation. But even though it helps in effective real-time applications, it needs or assumes a simple background and good lighting conditions for good and optimal performance [15] [16].

III. Proposed Work

A. System Overview

The Proposed system is Gesture Controlled Gaming System use to make physiotherapy interactive. It uses the Python, OpenCV and Mediapipe to capture the upper body movements. These movements or gestures are then

used to control different therapy games such as maze navigation and handling objects. It is make to peoples with mobility limitations to use it , It offers scalable solution that can be used for both home and clinic to support rehabilitation.
In Fig. 1, this system shows that it lets users control apps or games using upper-body gestures, so they don’t need to use their hands or controllers. A camera watches the user’s movements and the frames get processed one by one using OpenCV. Then, MediaPipe checks these frames with a model that’s already been trained to recognize gestures and key points. When the system finds a gesture, it sends it to Python, which turns it into a keyboard or mouse input, so the user can control things without a normal controller.

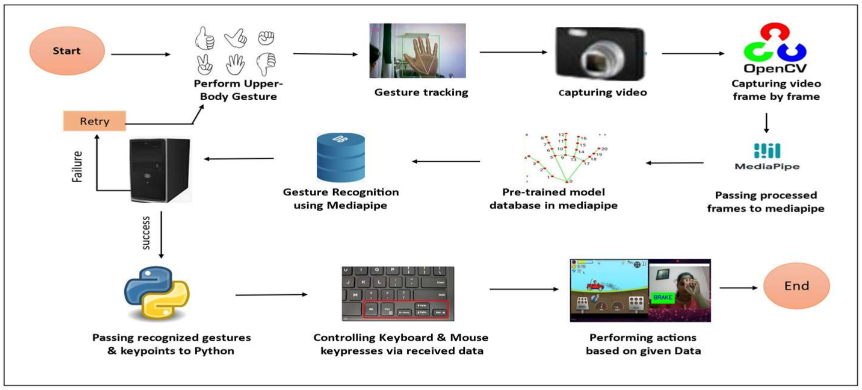

Fig.1: System Architecture

If the system can’t figure out the gesture, it will ask the user to try again, making sure it’s done right. By mixing gesture tracking, video processing, and recognizing these movements, the system makes it easier for all kinds of people to use. The retry option makes it more reliable, and the trained models make sure the system always recognizes gestures correctly.

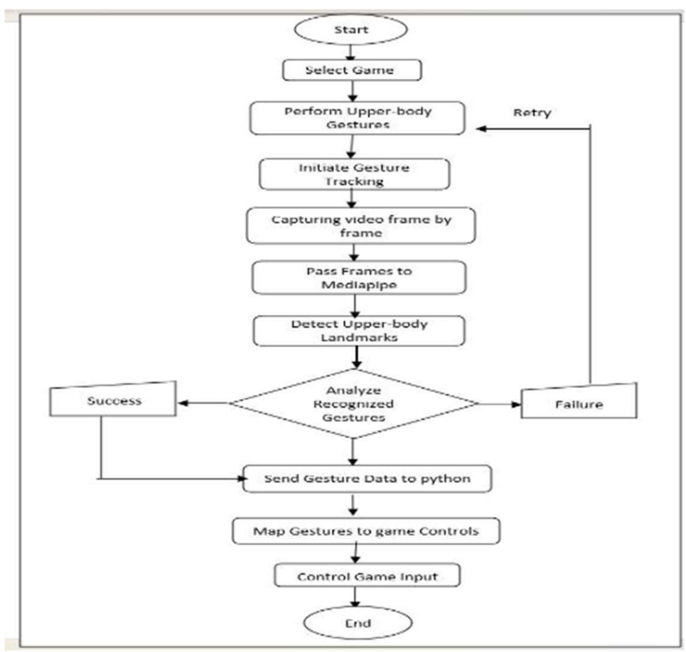

Fig.2: System Flowchart

In Fig. 2, you can see the flowchart that shows how this gesture-controlled system works. First, the user opens up the app and picks the game they wanna play. After they choose the game, they start doing upper-body gestures, and the webcam records them in real-time. Then, the app uses MediaPipe to process these video frames and

figure out where the key body parts are. The system looks at these gestures and checks if they match any pre-set patterns. If it doesn't get it right, it asks the user to try again, making sure the input is correct before moving ahead.
Once the gesture is recognized correctly, the system sends that data to Python, and it gets mapped to controls for the game. This way, the user can play the game without needing any physical controller.

B. Methodology

Python is the main brain here, controlling how data moves between OpenCV and Mediapipe. OpenCV grabs live video, cleans it by removing noise, and sends it to Mediapipe to figure out the gestures.
The process works like this: Python starts everything, OpenCV catches the video, Mediapipe finds the gestures, and then Python turns those gestures into actions. This whole setup is quick and works smooth, making therapy easy and fun for users. By mixing tech and games together, the system makes recovery exciting and motivating.
We pick Mediapipe because it makes gesture finding easy, saves lots of time, and works without much trouble. It keeps tracking gestures correctly, making tools like therapy stuff more trustable and simple to use.

Algorithm

- Input: Real-Time Video Feed

The system starts with a real-time video feed V, which is represented as a sequence of frames:

$$V = \{F_1, F_2, F_3, \ldots, F_n\} \quad \ldots\ldots(1)$$

Where V is the video stream, and F_i is the *i*th frame captured by the webcam.

- Hand Landmark Detection

For each frame F_i, the MediaPipe model M is applied to extract hand landmarks L:

$$L = M(Fi) \ldots\ldots(2)$$

Here, $L = \{(x_1, y_1, z_1), (x_2, y_2, z_2), \ldots,(x_{21}, y_{21}, z_{21})\}$, where (x_j, y_j, z_j) are the normalized 3D coordinates of the jth landmark.

- Gesture Recognition

A gesture recognition function G is defined to map hand landmarks L to a gesture g:

$$g = G(L) \ldots\ldots(3)$$

Here, $g \in \{g_1, g_2, \ldots, g_k\}$ represents a predefined set of gestures, such as swipe, fist, or point.
The recognition involves following process:
Distance Calculation
To calculate the Euclidean distance between two landmarks i and j the formula is follows:

$$d_{ij} = \sqrt{(xi - xj)2 + (yi - yj)2 + (zi - zj)2} \quad \ldots\ldots.(4)$$

Angle Calculation
To calculate the angle $\theta_{i,j,k}$ formed by three landmarks I, j, k, the formula is:

$$\theta i,j,k = \arccos\left(\frac{Vij * Vjk}{||Vij||*||Vjk||}\right) \ldots\ldots.(5)$$

Here, v_{ij} represents the vector between points i and j.

- Gesture-to-Action Mapping

The recognized gesture g is mapped to a game action a using the function A:

$$a = A(g) \quad \ldots\ldots(6)$$

Actions include jump, turn, accelerate, and other gameplay movements.

- Backend Communication

Once an action a is determined, it is sent to the game engine using a REST API:

$$R(a) \rightarrow \text{Game Logic} \ldots\ldots(7)$$

Where R is the function that communications with the backend system.

- Game State Update

The game state S is updated based on the action a using the following functions:

$$S_{t+1} = F(S_t, a) \text{ ......(8)}$$

Here, S_t is the game state at time t, and F is the game logic function that updates the state.

➢ Real-Time Feedback

Once the game state S_{t+1} is updated, it is sent to the frontend for rendering:

$$UI = H(S_{t+1}) \text{ ......(9)}$$

Where H maps the game state S_{t+1} to the user interface for display.

Summary Equation

The system overall functionality can be summarized with the following equations:

$$UI = H(F(R(A(G(M(Fi)))))) \text{ ......(10)}$$

Where:

- ➢ F_i: Input video frame.
- ➢ M: MediaPipe for hand landmark detection.
- ➢ G: Gesture recognition process.
- ➢ A: Gesture-to-action mapping.
- ➢ R: Backend communication.
- ➢ F: Game state update function.
- ➢ H: Frontend rendering function.

IV. Results and Discussion

The Gesture Controlled Physical Therapy System is developed for the user engagement and for there enjoyment. In this system there five different games and that games can be operate using five different upper-gestures using this system an individual can refresh there mind and body. The people who need physiotherapy they can easily use this system for there rehabilitation goals. This system was analyzed with respect to hits (detected gestures) and miss (undetected gestures) of gestures, accuracy of gestures and with existing system it provides better performance in all the metrics.

Fig. 3(a): Head Gesture – Down

Fig. 3(b): Head Gesture – Up

In this we added the head gesture to interact with game and that gesture recognized by the system and connected to the keyboard keys. The fig.3(a) refer to the tilt head in downward direction and that assign to down arrow key while the tilt head in upward direction assign to up arrow key. Every time user perform correct gesture and it recognized by the system it will provide correct key press. this can be used for enjoyment purpose or for the physiotherapy where the people who need therapy for the head movement can easily interact with it.

Fig. 4(a): Left Shoulder Raised

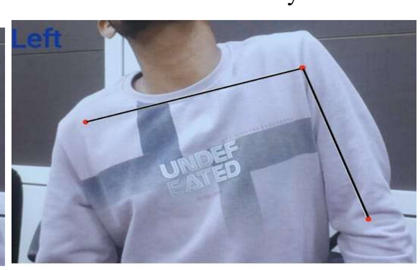

Fig. 4(b): Right Shoulder Raised

Fig 4 is about the shoulders gestures using this gesture the user can easily interact with the system or play the game. The Fig 4(a) shows left shoulder lift gesture and it connected to the left arrow key while fig 4(b) is about the right shoulder lift and it connected to the right arrow key every time when user perform shoulder gesture or actions the system recognized it and press the perfect key for that particular gesture which has been assigned to it.

In Fig. 5 we have added the mouth gesture to control or manage the game with gesture based interaction without the physical contact with the device or system. In fig. 5(a) it describes the open-mouth gesture, when this gesture is recognized the up arrow key will be pressed while on the other hand fig. 5(b) describe closed mouth gesture, it refers to no keypress when the mouth is closed. This gesture helps the individual for facial muscle.

Fig. 5(a): Mouth Open Gesture
Fig. 5(b): Mouth Closed Gesture

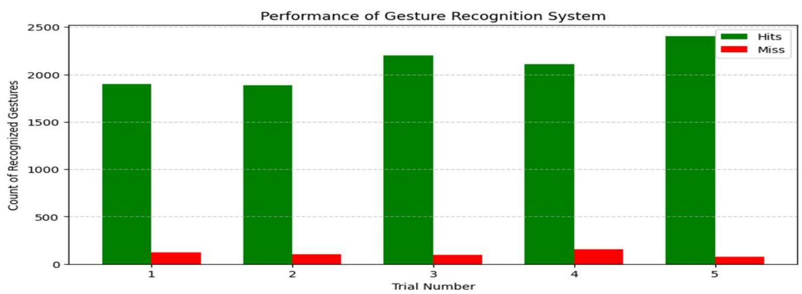

Fig. 6: Performance of Gesture Recognition System

Fig. 6 presents the performance of gesture recognition system on the basis of hits and miss. The green bar represents number of hits while red bar represents number of miss. The number of hits is always greater that the number of miss. It shows that system is very efficient at gesture recognition that leads to better performance. In first trial hits goes up around 2000 and miss are under 180. Also, in other four trials hits are for more better that the miss which provides high performance and better experience for user. The individual who needs physiotherapy they can use the system easily and it helps them in effective and accurate way for their treatment and in enjoyment.

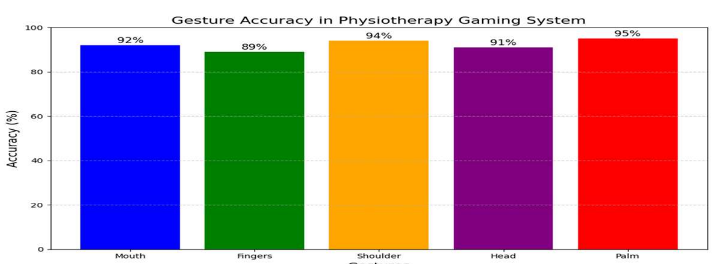

Fig. 7: Accuracy of Different Gestures

Fig. 7 illustrates the accuracy of each gesture, in this gesture control gaming system we used five different gestures – head, fingers, mouth, palm and shoulder. This system provides better performance and gesture tracking capabilities which gives high accuracy for all distinct gestures. In this palm has the highest of 95% accuracy on the other hand mouth and fingers gestures have 92% and 89% accuracy. Also, for shoulder and head gesture at 94% and 91%.

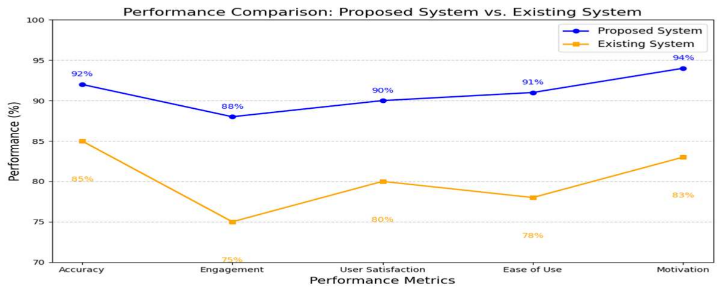

Fig. 8: Comparison of Proposed System vs Existing System

Figure 8 describe the performance comparison between the proposed system and existing system. The blue line represents the proposed or new system while yellow line represents existing system as shown in figure we use different metrics for comparison between proposed and existing system like accuracy, engagement, user satisfaction, ease of use and motivation. The proposed system far more better than the existing system, it maintains consistency high performance than the existing system, as shown in comparison graph ease of use is 91% for proposed system while for existing system that is 78% also other metrics are perform better in proposed system.

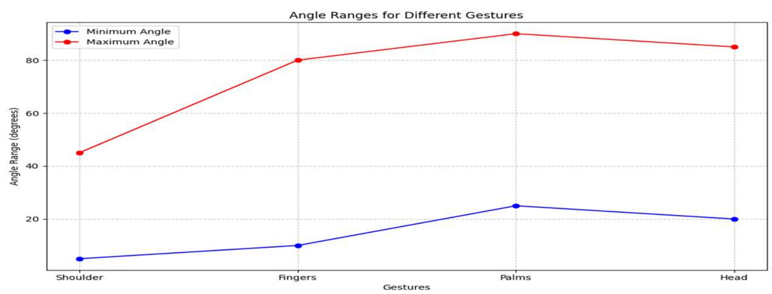

Fig. 9: Accuracy for Different Angles

As shown in Fig. 9, the figure represents range of angles for four different gestures like head, palms, fingers and shoulder with their minimum and maximum angles range. It shows in which angle the system gives accurate result. The blue line represents the minimum angle and on the other hand, the red line demonstrates the maximum angle. The above graph describe, range of different gestures to get accurate result different gestures.

V. Conclusion

In this Gesture Controlled system, the system is combination of different upper-body gestures like head up and down movement, shoulder tilt, fingers and mouth actions for playing different games with different upper-body gesture. This system gives hands free experience to play games using gestures. This system can also help in physiotherapy, the people who facing challenges to make movements of their hand, shoulders and other upper-body parts can easily treated while engaging with this system. This system helps in easy and effective recovery of an individual also it makes it more interesting and less boring for the user. In this system we have added five different games and that games can be played using five different upper-body gestures, the system gives accurate results with in the less time of period. It converts the gestures into game moves in less time without time delay. By using this system people can refresh the mind and body. There are some difficulties in system to capture gesture like background light and camera angle. The system can be improved by adding different games, gestures, levels and live feedback to the user.

References

[1] A. Gupta, V. K. Sehrawat, and M. Khosla, “FPGA based Real Time Human Hand Gesture Recognition System,” Procedia Technology, vol. 6, pp. 98–107, 2012, doi: 10.1016/j.protcy.2012.10.013.

[2] H. S. S. Blas, A. S. Mendes, D. H. de la Iglesia, L. A. Silva, and G. V. González, “A Multiagent Platform for promoting physical activity and learning through interactive educational games using the depth camera recognition system,” Entertain Comput, vol. 49, Mar. 2024, doi: 10.1016/j.entcom.2023.100629.

[3] M. Denkowski, K. Dmitruk, and Ł. Sadkowski, "Building automation control system driven by gestures," in IFAC PapersOnLine, Jun. 2015, pp. 246–251. doi: 10.1016/j.ifacol.2015.07.041.
[4] C. S. González-González, P. A. Toledo-Delgado, V. Muñoz-Cruz, and P. V. Torres-Carrion, "Serious games for rehabilitation: Gestural interaction in personalized gamified exercises through a recommender system," J Biomed Inform, vol. 97, Sep. 2019, doi: 10.1016/j.jbi.2019.103266.
[5] N. Gosalia, P. Jain, I. Shah, A. R. Joshi, N. Katre, and S. Sahasrabudhe, "3D gesture-recognition based animation game," in Procedia Computer Science, Elsevier B.V., 2015, pp. 712–717. doi: 10.1016/j.procs.2015.03.138.
[6] M. Ilves, Y. Gizatdinova, V. Surakka, and E. Vankka, "Head movement and facial expressions as game input," Entertain Comput, vol. 5, no. 3, pp. 147–156, 2014, doi: 10.1016/j.entcom.2014.04.005.
[7] S. S. Rautaray and A. Agrawal, "Real Time Gesture Recognition System for Interaction in Dynamic Environment," Procedia Technology, vol. 4, pp. 595–599, 2012, doi: 10.1016/j.protcy.2012.05.095.
[8] N. Lalithamani, "Gesture Control Using Single Camera for PC," in Physics Procedia, Elsevier B.V., 2016, pp. 146–152. doi: 10.1016/j.procs.2016.02.024.
[9] R. Zahra et al., "Camera-based interactive wall display using hand gesture recognition," Intelligent Systems with Applications, vol. 19, Sep. 2023, doi: 10.1016/j.iswa.2023.200262.
[10] T. L. Dang, S. D. Tran, T. H. Nguyen, S. Kim, and N. Monet, "An improved hand gesture recognition system using keypoints and hand bounding boxes," Array, vol. 16, Dec. 2022, doi: 10.1016/j.array.2022.100251.
[11] H. Mohyuddin, S. K. R. Moosavi, M. H. Zafar, and F. Sanfilippo, "A comprehensive framework for hand gesture recognition using hybrid-metaheuristic algorithms and deep learning models," Array, vol. 19, Sep. 2023, doi: 10.1016/j.array.2023.100317.
[12] K. Gao et al., "Challenges and solutions for vision-based hand gesture interpretation: A review," Computer Vision and Image Understanding, vol. 248, Nov. 2024, doi: 10.1016/j.cviu.2024.104095.
[13] T. Vuletic et al., "A novel user-based gesture vocabulary for conceptual design," International Journal of Human Computer Studies, vol. 150, Jun. 2021, doi: 10.1016/j.ijhcs.2021.102609.
[14] N. H. Dardas, J. M. Silva, and A. El Saddik, "Target-shooting exergame with a hand gesture control," Multimed Tools Appl, vol. 70, no. 3, pp. 2211–2233, 2014, doi: 10.1007/s11042-012-1236-4.
[15] Y. Zhou, G. Jiang, and Y. Lin, "A novel finger and hand pose estimation technique for real-time hand gesture recognition," Pattern Recognit, vol. 49, pp. 102–114, Jan. 2016, doi: 10.1016/j.patcog.2015.07.014.
[16] Bhushan Chaudhari, Durwa Patil, Hemant Patil, Rupal Patil, Jui Gawali, "Fruit Defect Inspection System Using Image Processing and IoT Framework", IEEE International Conference on Emerging Smart Computing and Informatics (ESCI), March 2023