International Journal of Computer Sciences and Engineering
Vol.**12**, Issue.**8**, pp.**01-09**, August **2024**
ISSN: 2347-2693 (Online)
Available online at: www.ijcseonline.org

ISSN: 2347-2693 (E)

Research Article

Workout Monitoring Robot: A Robotic Approach for Real-Time Workout Monitoring and Guidance

Shreyas Walke[1*], Yash Wadekar[2], Aditya Ladawa[3*], Pratik Khopade[4], Shraddha V. Pandit[5]

[1,2,3,4,5]Dept. of Artificial Intelligence and Data Science, PES's Modern College of Engineering, India

**Corresponding Author: adityaladawa11@gmail.com*

Received: 23/Jun/2024; **Accepted**: 25/Jul/2024; **Published**: 31/Aug/2024. **DOI**: https://doi.org/10.26438/ijcse/v12i8.19

Abstract: In this research paper, we present the development and implementation of a cutting-edge Workout Monitoring Robot designed to monitor and guide user's exercises with its capability of pose estimation and correction and natural language interactions that revolutionize the way individuals engage in exercise routines. Whereas the current research that part-take in similar activities have had much difficulty in flexibility and ease of interaction. This study focuses on enhancing the effectiveness and safety of physical fitness activities by imposing advanced technologies including human pose estimation, autonomous robot navigation and a sophisticated human-computer interface driven by NLP. This research attempts to open the door to a new era of smart and responsive workout assistance, ultimately improving health and well-being.

Keywords: Human Pose Estimation, Remote Photoplethysmography, Autonomous Robot Navigation, Natural Language Robot Programming

1. Introduction

Physical fitness is a crucial aspect of health and overall well-being, reflecting the ability to participate in sports, perform work-related tasks, and carry out daily activities effectively. The increasing global focus on health awareness and the desire for an active way of life suggest that the fitness sector will experience ongoing expansion and evolution in the foreseeable future. Owing to this growth, many novice people are kicking off the fitness enthusiast journey. While this is highly beneficial, it also comes with certain kinds of risks. The number of injuries had increased by 8.89 percent last year with regards to normal exercise as well as exercise equipment [1].

Proper posture during physical training is essential for maximizing movement efficiency and reducing the risk of injury. Improper posture may lead to discomfort, muscle strain, and, in some cases, lasting harm to your joints and muscles [2]. For individuals who don't have access to a personal trainer, maintaining correct form during a workout can pose a challenge. Here, AI assistance can play an important role in helping people live healthy by providing instant feedback on their posture during exercise. With so many digital fitness applications and digital trainers already on the market, the term "AI powered approach for workout assistance" is no longer a new concept. Recently, many interactive systems for home exercise have emerged that use deep learning technologies such as human posture estimation, action recognition, etc. The system's capacity to provide accurate, immediate feedback on exercise posture relies on the robustness of its Intelligent Posture Prediction feature. This technology leverages computer vision and machine learning to identify distinct body elements and generate predictions by analyzing the angles created by these reference points. An example of at-home workout by an interaction system is shown in Figure 1.

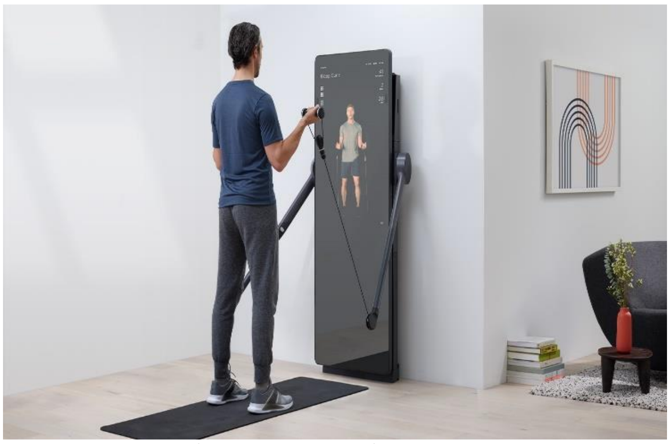

Figure 1. Example of an existing system for at home workout using mirror [9].

However, such systems for at-home workouts have faced significant challenges related to mobility, particularly due to the inflexibility of the monitoring devices. They also struggle with issues of user comfort, accessibility, and various technical challenges, especially when integrated into mobile phones or stationary devices, which can limit efficiency. To address these issues, this paper presents comprehensive research and suitable techniques for an autonomous robot designed to offer real-time monitoring and guidance during workout sessions. This advancement aims to enhance the effectiveness and safety of physical fitness activities.

2. Related Work

A research project introduced an application crafted to recognize and discern the user's exercise poses, tally the designated exercise repetitions, and offer personalized, comprehensive feedback on enhancing the user's posture. This application was developed using JavaScript, Node.js, and various libraries, including OpenCV and MediaPipe. This strategy uses a two-step tracking machine learning approach. It uses trackers to locate events or movements in live video and then uses input from recent videos to predict critical moments in the target area [3].

Another study proposed a system in which Shwetank Kardam first recolored the image because OpenCV renders the RGB image to BGR color format, but for MediaPipe to work, it needs to convert the BGR image back to RGB. It then prints the detections of the given model. Lastly, the color format is changed back to BGR format as OpenCV runs on BGR format, allowing it to render the detections [4].

To improve upon the previous system, Grandel Dsouza used deep learning combined with convolutional neural networks trained with various sample data to achieve human pose estimation. Different images depicting various body parts from multiple angles are input into the system, resulting in the creation of a trained model. The model is trained to ensure that each body part within it has a distinct and unique identity coordinate. This trained model, which includes twenty-five thousand images and forty thousand people, uses the COCO (Common Object in Context) and MPII model designations. It should be noted that increasing the size of the sample dataset used for training the system leads to improved system performance. A user-friendly graphical interface was also pursued in this system, using an open-source Python framework called Flask [10].

Lumin Xu and his colleagues noted the absence of datasets containing comprehensive annotations for the entire human body. Previous methods had to combine independently trained deep models from various datasets, covering the human face, hands, and body. Navigating imbalances in the dataset posed significant challenges, compounded by the increasing complexity of the model's structure. To address this issue, they introduced COCO-Whole Body, which extends the COCO dataset by including annotations for the entire body. They employed a single-network model named Zoom Net, designed to accommodate scale differences within various sections of an individual's body by considering the body's structure. On the proposed COCO-Whole Body dataset, Zoom Net demonstrated superior performance compared to competing methods. This robust pre-training dataset, COCO-Whole Body, has versatile applications, including hand key point estimation and facial landmark identification. It also enables the training of intricate models from scratch, specifically for predicting poses of the entire body, without relying on pre-existing models or frameworks [11].

To achieve real-time detection, Ce Zheng and colleagues proposed a model categorized into kinematic, planar, and volumetric domains. For estimating human poses in a two-dimensional space, the initial step involves cropping the input image to ensure each cropped area contains only one person. Within the realm of analyzing human posture using deep learning methods tailored for individual-focused pipelines, two primary categories emerge: regression approaches and heatmap-based techniques. Regression approaches employ complete frameworks to understand the relationship between input images and the identification of joints or attributes of the human body. Conversely, heatmap-based algorithms aim to predict and anticipate the estimated locations of various body parts and joints, with supervision provided by heatmap representations. This model relies on motion-capturing systems, which are essential for obtaining precise posture descriptions in a three-dimensional (3D) context but can be challenging to set up in random settings. Additionally, in pose estimation within two-dimensional approaches, detectors for individuals may struggle with recognizing the edges of heavily overlapping images [12].

Zell et al. addresses the challenge of assessing hidden information in 2D video data of human motion, including force and torque characteristics. Existing research has typically focused on either 3D motion reconstruction or physical modeling separately, with limited integration of both aspects. The paper introduces a unified model that enhances both three-dimensional pose estimation and spatial rendering, alongside physical modeling simultaneously. Extracting three-dimensional human poses from two-dimensional images is inherently challenging due to the ambiguous nature of the task and the ill-posed problem it presents. Current methods often rely on priors and learned subspaces but come with inherent limitations. The paper proposes a factorization approach to decompose joint data into camera motion, base poses, and mixing coefficients, followed by a physical model projection to infer plausible human motion. The paper also explores inner forces such as joint torques crucial for biomechanical studies, leveraging a data-driven statistical approach for direct 3D inference from monocular images [14].

In 2023, a new real-time posture estimation system was developed, integrating MediaPipe and OpenCV to provide a comprehensive human posture assessment solution. This system starts by recording the user's video stream and uses MediaPipe to identify key landmarks on the human body. These landmarks are then processed with OpenCV to calculate angles, which are used to evaluate and determine the user's posture.. Users receive instant feedback on their posture and actionable recommendations for improvement through the system. The system demonstrates robust performance across various lighting conditions and is resilient to background interference, applicable across a wide range of exercise routines. Utilizing this system for real-time feedback can assist users in refining their posture and technique, thereby reducing the risk of potential injuries [2].

In 2010, a System for Social Assistance through Robotics was developed for engaging individuals during exercise sessions, providing instructions, assessment, encouragement, and motivation. The system incorporates a vision module capable of real-time recognition of the user's hand movements, with minimal environmental requirements and no additional demands on the user. The methodology involves creating an arm pose recognition system that simplifies exercise setup [5].

Further advancements were made to these systems by T.T. Tran et al. [18], where they introduced a recommendation system (RS) to enhance the current fitness assistance and guidance systems. This RS provides personalized workout recommendations tailored for both newcomers and experienced users. To forecast suitable workouts for beginners, they employed neural networks and logistic regression techniques. Additionally, they developed an agent with reinforcement learning capabilities using the Soar architecture to assist users in selecting workouts tailored to their specific conditions. The authors utilized patterns from volunteer data as reference points for predicting and providing workout recommendations to newcomers through the fitness assistance system during their initial experiences. User profiles are established using input data that includes various parameters such as individual characteristics, exercise preferences, and measurements of maximal strength used in weight training. The recommendations generated include details such as exercise weight, repetitions, and rest intervals for each recommended set.

3. Methodology

In this section, we present the overall configuration of our studied system aimed at at-home workout as well as gymnasium service while navigating to the desired pose. The proposed methodology works with the comprehensive integration of hardware and software components to monitor and guide users in their workouts. It leverages sensory inputs from microphones and cameras, processes data, while offering clear communication and feedback through speakers and an LCD screen.

3.1 Hardware platform

The system is constructed using a variety of components to ensure robust functionality and efficiency. It includes a Raspberry Pi 5 with 8GB of RAM, serving as the main processing unit. For additional processing, particularly for navigation and movement tasks, an Arduino UNO board is incorporated. Visual outputs are managed through a 7-inch LCD display with a resolution of 800x480 pixels. The system features a standard webcam for capturing images and videos. For mobility, the design includes four dual-shaft BO Gear Motors and two servo motors, all controlled by two motor drivers to effectively manage the operation of these motors. Power requirements are handled by a 12V 5200mAh battery and a 10000mAh power bank for Raspberry Pi 5 and Arduino UNO, respectively. Additionally, a buck converter is used for efficient power conversion, since the Raspberry Pi 5 requires a steady supply of 5V and 5A.

Table 1. Hardware platform and robot configuration.

Component	Description	Quantity
Raspberry Pi 5	8GB Ram	1
Arduino	UNO Board	1
LCD Display	7-inch, 800 x 480	1
Camera	Webcam	1
BO Gear Motor	OLatus Dual Shaft	4
Servo Motor	Tower Pro MG995	2
Motor Driver	L 298N	2
Battery	12V, 5200 mAh	1
Power Bank	10000 mAh	1
Buck Converter	DC-DC XY3606 Power Converter	1

3.2 System Design and Architecture

Figure 2 presents the block diagram of the general system. The proposed system comprises eight modules: natural language robot programming, image processing, navigation system, user identification and dashboard, pose estimation, pulse rate monitoring, file storage system, and output and feedback mechanism. Each module is intricately interconnected to ensure smooth interaction and effective results.

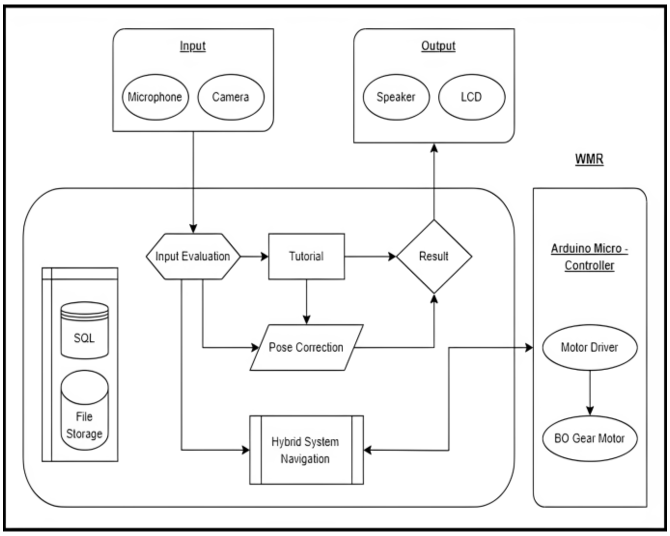

Figure 2. Modular design.

1) Input Evaluation: The system employs user-defined wake words for robot activation. Voice commands are captured by microphones, while real-time video data from cameras assists in locating the user's position, aiding navigation, user identification, and access to the dashboard through a face recognition module. The approach utilizes OpenCV to capture the user's face and the `face_recognition` Python module for authentication to access the dashboard. This system is user-friendly for both new and existing users. For existing users, the `face_recognition` module captures their facial data and matches it with the data in the database, ensuring accurate user identification. For new users, an information window is provided to fill along with the user's photograph, which is then saved in the database. Subsequently, during login, the system performs facial data matching for new users.

2) Navigation System: The navigation system for robot movement is structured in three phases:
a) Object Detection: The first phase involves implementing the MobileNet SSD V2 (Single Shot MultiBox Detector) object detection model, which leverages the COCO v2 (Common Objects in Context) dataset. This model is adept at identifying and categorizing a diverse array of objects within the camera's field of view. While the MobileNet SSD V2 model supports the detection of multiple object classes, our system specifically prioritizes the 'human' class. This focus on human detection is critical for the system's operation, as it enables the robot to recognize and track individuals effectively. By concentrating on human detection, the system enhances the robot's ability to navigate and interact within environments where human presence is a key factor, thereby ensuring more precise and context-aware navigation.

b) Object Tracking: Building upon the results of object detection, the object tracking model is implemented to monitor the identified humans continuously. Among the detected objects, those labeled as 'person' are singled out, and bounding box coordinates are extracted for each detected human: ***x_min***, ***y_min***, ***x_max, y_max***. Using these coordinates, the width and height of the bounding box are calculated to determine the center of the object. The horizontal and vertical centers of the object are calculated as follows:

$$\text{obj x center} = x_{min} + \frac{x_{max} - x_{min}}{2} \tag{1}$$

$$\text{obj y center} = y_{min} + \frac{y_{max} - y_{min}}{2} \tag{2}$$

The deviation of the object's center from the frame's horizontal center (assumed to be at 0.5 in normalized coordinates) is calculated as:

$$x_{deviation} = 0.5 - \text{obj x center} \tag{3}$$

This deviation serves as the basis for determining the direction and movement of the robot. Depending on the deviation, decisions are made on whether to move forward, turn left, or turn right:
- If , $|x_{deviation}| < 0.1$ the robot moves forward, provided the detected person is not too close to the bottom of the camera frame.
- If $|x_{deviation}| > 0.1$, indicating the person is to the right of the center, the robot turns left.
- Conversely, if $|x_{deviation}| < -0.1$, showing the person is to the left of the center, the robot turns right.

The duration and speed of these turns are dynamically adjusted based on the magnitude of $x_{deviation}$

c) Motor Control from Arduino: In this phase, GPIO pins of the Arduino are controlled from the Raspberry Pi to adjust the motors' speed and direction using PWM (Pulse Width Modulation) signals. The Arduino receives instructions from the Raspberry Pi to execute precise motor actions according to the calculated deviations. A feedback loop is incorporated to continuously refine the robot's trajectory based on the person's position in subsequent frames. This ensures that the robot maintains an accurate and responsive interaction with its environment, enhancing both safety and efficiency in navigation.

3) User Dashboard: The user dashboard serves as an interface facilitating user interaction with the system through natural language robot programming (NLRP). It offers users a range of functionalities, including user identification and access to robot utilities such as bio-signal recognition, exercise guidance, and monitoring. Users can also retrieve their past records through the dashboard interface. Communication between the user and the robot is managed by the speech module of NLRP. The comprehension of natural language commands relies on a dependency parser, with spaCy's implementation being particularly utilized. Dependency parsing involves the automatic extraction of the syntactic structure of a sentence in the form of a tree, achieved by recognizing categorized binary relationships among its constituent words. Extracted user commands by spaCy are then mapped with a bag of instructions. For instance, commands such as "count the reps" or "show my previous sessions" trigger the respective events, such as invoking the "count_rep" method [26].

4) Pose Correction: The robot employs a comprehensive approach utilizing both Mediapipe and OpenCV to monitor and guide users during exercises effectively. Initially, it captures the user's video feed, extracting crucial body landmarks using the Mediapipe framework. Subsequently, OpenCV is employed to calculate precise angles between significant joints such as the shoulder, elbow, knee, and wrist. These calculated angles serve as the basis for assessing the correctness of exercises, comparing them against a trained module dedicated to exercise form evaluation. Feedback is generated based on predefined evaluation metrics, offering guidance and correction to the user in real-time. The training data for the exercise correctness module is derived from a dataset containing pose landmarks extracted from various exercise videos. Each video in the dataset is labeled with the correct or incorrect exercise form. This labeled data is then converted into a .csv file format for further processing. Several machine learning models are trained on this dataset and evaluated using different metrics to identify the most effective model. The selected model is utilized for inference purposes during real-time exercise monitoring. During the detection phase, real-time video data is extracted using

Mediapipe landmarks. Calculations are performed on these landmarks to update the repetitions counter and detect errors in exercise form as they occur.

By combining angle-defined detection with machine learning-based classification, error detection while performing reps is significantly enhanced. This hybrid strategy provides more control over recognizing repetitions performed and errors detected, ensuring precise and effective guidance to the user during exercise sessions.

5) Storage System: The robot's storage system serves as a vital repository for housing various user-related data and interactions, comprising two distinct components: the user database and the file storage system.

a) User Database: The user database employs JSON format as its underlying structure, facilitating efficient storage and retrieval of user information. This system stores a wide array of data, including details of user sessions, general user profiles, and progress tracking metrics. Leveraging the simplicity and flexibility of JSON, user data is organized in a structured format, enabling easy access and manipulation. Structured queries enable seamless retrieval of past session details, user profiles, and progress calculations, enhancing the overall user experience.

b) File Storage: In addition to the user database, the file storage system plays a crucial role in providing supplementary support to users through the storage of video tutorials. These tutorials are stored within local directories in the memory card mounted on Raspberry Pi 5, allowing for easy management and access. Users can utilize natural language processing (NLP) commands to retrieve specific tutorials as needed. Upon request, these tutorials can be displayed on the robot's LCD screen, offering users convenient access to instructional content for further assistance and guidance.

c) Visualization in Dashboard: The dashboard also includes visualization capabilities to provide users with insightful representations of their workout data which included rep counts and error count. By leveraging JSON data from the user database, the dashboard displays visualizations using Python libraries like Pandas and Matplotlib. Users can easily track their progress, analyze trends, and make informed decisions based on the visualized data. This integration of visualization enhances the user experience by offering a comprehensive overview of their interactions and progress within the system.

6) Bio-signal recognition: Utilize computer vision and signal processing to monitor heart rate remotely via a webcam. OpenCV and cvzone libraries for image processing and face detection have been utilized. These libraries facilitate real-time video capture and processing. The cvzone 'FaceDetector' module detects faces within the video frames, which are crucial for isolating the regions of interest for pulse rate monitoring. Each detected face is resized, and a Gaussian pyramid reduction is applied to emphasize temporal changes rather than spatial details. This reduced data is then processed to build a time-series signal representing color intensity changes over time, which correlates with the blood flow. The signal undergoes transformation into the frequency domain through Fast Fourier Transform (FFT). A bandpass filter isolates frequencies between 1.0 and 2.0 Hz, corresponding to typical human heart rates. The dominant frequency component is identified, from which the heart rate in beats per minute (BPM) is calculated. The system outputs real-time heart rate measurements displayed alongside the video feed.

4. Implementation and Results

This section presents the outcomes achieved by combining various methods in developing a robot system designed to assist users during workouts. Each subsection describes achievements in user authentication and dashboard access, object detection and tracking, workout monitoring and guidance, real-time heart rate monitoring via rPPG, data storage, and hardware integration involving Raspberry Pi 5 and Arduino Uno.

i) User Authentication and Dashboard: Upon successful facial recognition, users gain direct access to their personal dashboard. This dashboard includes workout tutorials, progress graphs, and other helpful information, making it easy for users to engage and keep track of their progress.

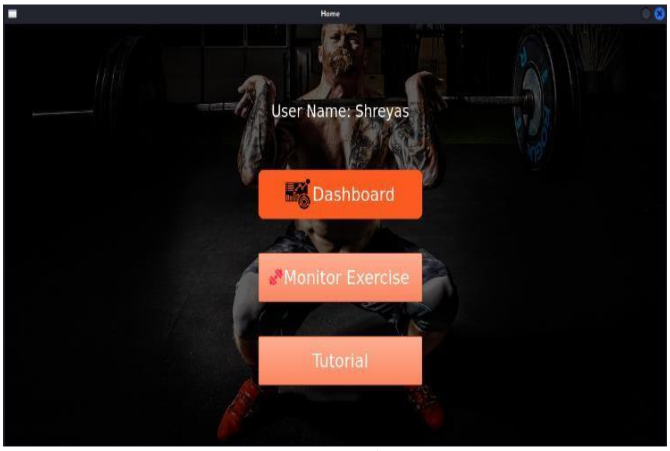

Figure 3. User dashboard after authentication.

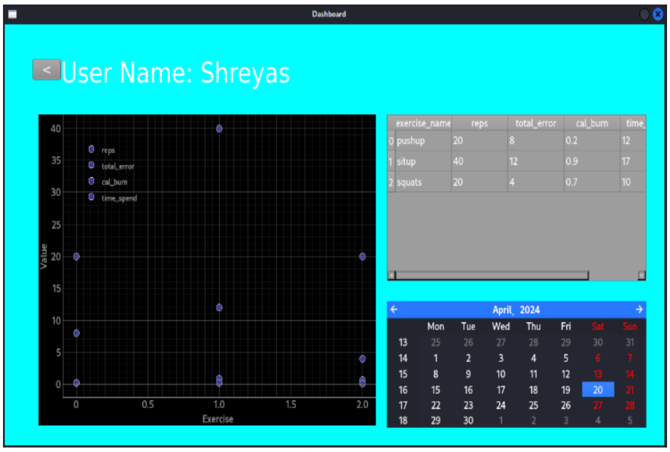

Figure 4. User dashboard for workout history and statistics.

ii) Object Detection and Tracking: The system utilizes a combination of MobileNet SSD, OpenCV, and MediaPipe, along with hardware such as a camera module and servo motors, to detect and track users. Algorithms like Mean Shift and Cam Shift ensure accurate tracking as users move within the robot's view.

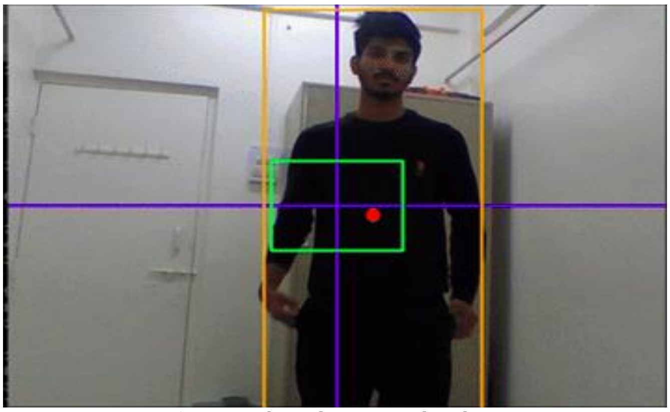

Figure 5. Object detection and tracking.

iii) Workout Monitoring and Guidance: OpenCV captures video feeds, and MediaPipe identifies human body landmarks to monitor workouts. Simple algorithms and machine learning models analyze these landmarks to detect errors, offering real-time guidance to users and enhancing the effectiveness of their exercises.

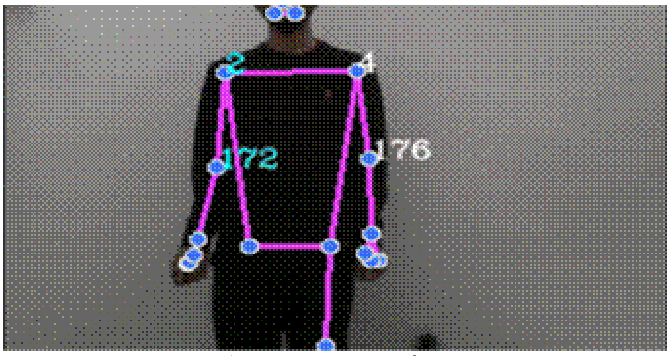

Figure 6. Monitoring phase.

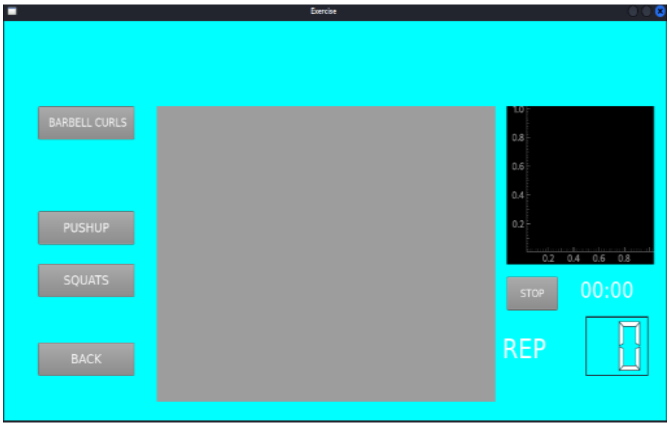

Figure 7. Monitoring window.

iv) rPPG (Remote Photoplethysmography): Real-time heart rate monitoring via rPPG is seamlessly integrated into the system. Users can view their heart rate alongside the video feed, providing immediate feedback on their physical responses during exercise.

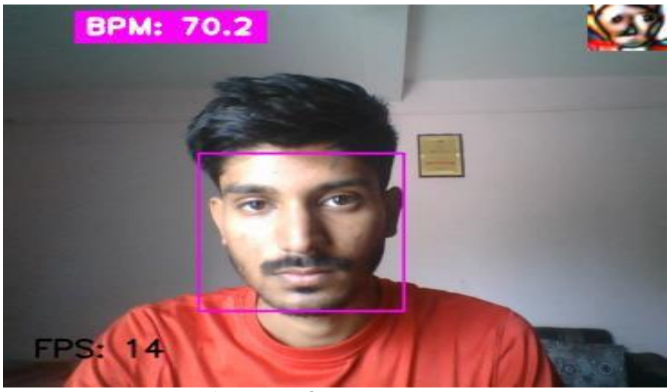

Figure 8. Pulse rate monitoring

v) Storage: User progress data, including workout details logged by date and time, is stored in JSON format. This structured storage method allows for efficient tracking and analysis of user performance over time.

vi) Raspberry Pi 5 Integration with Hardware: The Raspberry Pi 5 serves as the main computing unit for vision-based tasks. Its integration with hardware components is illustrated through circuit diagrams, showcasing the system's reliability and scalability in managing computational demands.

vii) Arduino Uno Integration with Hardware: Arduino Uno enables robot movement and seamlessly integrates with hardware components essential for locomotion. Detailed circuit diagrams highlight the connections between hardware elements, ensuring smooth operation of the robotic system.

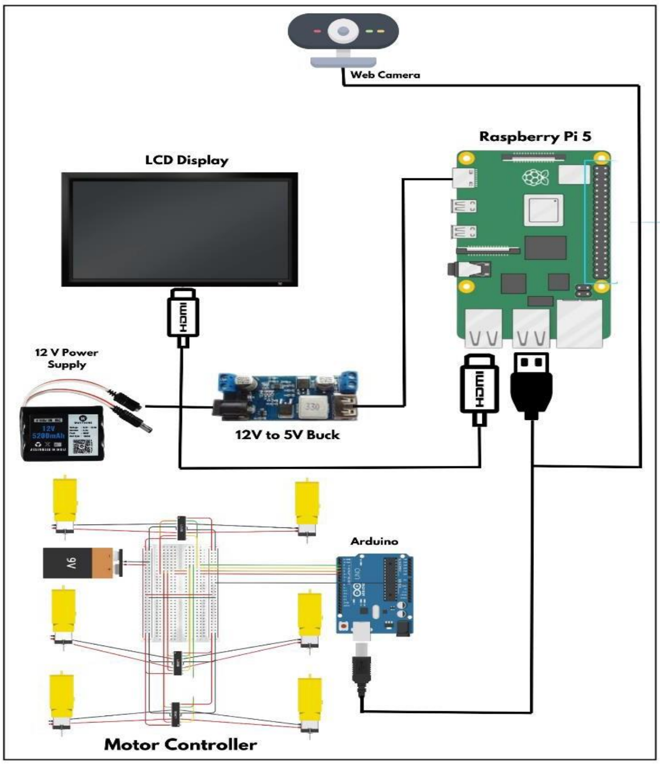

Figure 9. Hardware connections diagram

Figure 10. Workout monitoring robot.

viii) Performance: Table 2 depicts the time taken to carry out object detection and tracking and summaries the robot performance. Capturing a camera frame takes approximately 5 milliseconds. Performing inference and obtaining a prediction require about 90 milliseconds. Tracking a "person" takes an additional 5 milliseconds. The total average time for the entire process is 100 milliseconds, resulting in an average frame rate of 9.0 frames per second (FPS).

Table 2. Latency and Processing Time for Object Detection and Tracking

Process name	Time required (ms)
Average camera frame capture time	5.0
Perform inference and obtain prediction	90.0
Track person	5.0
Average total time	100.0
Average FPS	9.0 (FPS)

5. Conclusion

In conclusion, this research presents a significant advancement in fitness technology with the development of a real-time workout monitoring and guidance robot. The system integrates sophisticated hardware and software components, utilizing microphone and camera inputs to provide seamless user interaction, real-time feedback, and instructions. The output through speakers and an LCD screen enhances communication, improving the overall user experience. Arduino technology is employed for autonomous navigation and obstacle avoidance, which enhances the robot's efficiency and ensures smooth operation during workouts. On the software side, MySQL and file storage manage user records and instructional materials, while advanced algorithms facilitate, navigation, and pose estimation for exercise error analysis. This approach establishes a robust framework for advancing fitness technology, offering a more personalized and interactive method for at-home workouts and contributing to improved health and well-being. For future work, several areas are identified for further exploration. Optimization of object detection algorithms could improve efficiency and accuracy, potentially through model quantization and hardware acceleration. Developing adaptive object tracking algorithms could enhance robustness by adjusting to changes in object appearance and motion. Additionally, integrating multi-modal data from sensors like depth cameras or inertial measurement units (IMUs) could refine object detection and tracking capabilities in challenging conditions. Exploring domain-specific applications, such as in autonomous vehicles or healthcare, and developing user-friendly interfaces for practical integration are also valuable avenues for future research. Performance evaluations and benchmarking will be essential to assess and refine these advancements. This project provides a strong foundation for ongoing innovation in computer vision and fitness technology, opening avenues for further research and broader application across various industries.

Conflict of Interest

The authors affirm that there are no conflicts of interest with any entities or individuals concerning the subject matter discussed in this paper. No financial or non-financial support has been received from any parties associated with the content of this review. Our conclusions and viewpoints are based on an impartial and objective assessment of the existing research and data.

Funding Source

This study was conducted independently without specific grant support from public, commercial, or non-profit funding organizations. It was carried out without external financial assistance, ensuring the neutrality and objectivity of the research.

Authors' Contributions

Each author made significant contributions to this research. Author 1 was responsible for robot navigation, movement, and hardware integration. Author 2 managed backend development and the integration of modules and hardware. Author 3 developed the workout analysis algorithms and machine learning models used to identify and detect errors during workouts. Author 4 provided substantial assistance with model training and the development of data preprocessing pipelines. Author 5 played a crucial role in supervising and guiding the project throughout its duration. All authors participated in the development, review, and editing of the manuscript.

Acknowledgements

The completion of this paper and the underlying research would not have been achievable without the outstanding guidance of our project mentor, Prof. Dr. Mrs. Shraddha Pandit. She generously devoted her time, expertise, and perspectives to us. Her active involvement in our research was crucial to the accomplishment of this review paper, and we express our heartfelt appreciation for her invaluable contribution.

References

[1] V. Štajer, I.M. Milovanovic, N. Todorovic, M. Ranisavljev, S. Pišot, P. Drid, "Let's (Tik) Talk About Fitness Trends," Frontiers in Public Health, Vol.**10**, pp.**899-949, 2022.** DOI: 10.3389/fpubh.2022.899949.

[2] V.S.P. Bhamidipati, I. Saxena, D. Saisanthiya, M. Retnadhas, "Robust Intelligent Posture Estimation for an AI Gym Trainer using Mediapipe and OpenCV," 2023 International Conference on Networking and Communications (ICNWC), Chennai, India, pp.**1-7, 2023**. DOI: 10.1109/ICNWC57852.2023.10127264.

[3] G. Taware, R. Kharat, P. Dhende, P. Jondhalekar, R. Agrawal, "AI-Based Workout Assistant and Fitness Guide," 2022 6th International Conference on Computing, Communication, Control, and Automation (ICCUBEA), Pune, India, pp.**1-4, 2022.** DOI: 10.1109/ICCUBEA54992.2022.10010733.

[4] S. Kardam, S. Maggu, "AI Personal Trainer using OpenCV and Python," International Journal of Advanced Research in Engineering and Science, Vol.**9**, Issue.**12, 2021.**

[5] J. Fasola, M.J. Mataric, "Robot exercise instructor: A socially assistive robot system to monitor and encourage physical exercise for the elderly," 19th International Symposium on Robot and Human Interactive Communication, Viareggio, Italy, pp.**416-421, 2010.** DOI: 10.1109/ROMAN.2010.5598658

[6] K. Enoksson, B. Zhou, "Sound following robot," KTH Royal Institute of Technology, Stockholm, Sweden, June **2017.**

[7] Jitesh, "AI Robot - Human Following Robot using TensorFlow Lite on Raspberry Pi," **2021.**

[8] F. Haugg, M. Elgendi, C. Menon, "GRGB rPPG: An Efficient Low-Complexity Remote Photoplethysmography-Based Algorithm for Heart Rate Estimation," Bioengineering, Vol.**10**, No.2, pp.**243, 2023.** DOI: 10.3390/bioengineering10020243.

[9] J. Greenblatt, "AI-Powered Smart Mirror," April **2020.**

[10] G. Dsouza, D. Maurya, A. Patel, "Smart gym trainer using Human pose estimation," 2020 IEEE International Conference for Innovation in Technology (INOCON), Bangalore, India, pp.**1-4, 2020.** DOI: 10.1109/INOCON50539.2020.9298212.

[11] L. Xu, S. Jin, W. Liu, C. Qian, W. Ouyang, P. Luo, X. Wang, "ZoomNAS: Searching for Whole-Body Human Pose Estimation in the Wild," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.**45**, No.**4**, pp.**5296-5313, 2023.** DOI: 10.1109/TPAMI.2022.3197352.

[12] C. Zheng, W. Wu, T. Yang, S. Zhu, C. Chen, R. Liu, J. Shen, N. Kehtarnavaz, M. Shah, "Deep Learning-Based Human Pose Estimation: A Survey," **2020.**

[13] S. Chen, R. Yang, "Pose Trainer: Correcting Exercise Posture using Pose Estimation," **2018.**

[14] P. Zell, B. Wandt, B. Rosenhahn, "Joint 3D human motion capture and physical analysis from monocular videos," in Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp.**17–26, 2017.**

[15] A. Flores, B. Hall, L. Carter, M. Lanum, R. Narahari, G. Goodman, "Verum Fitness: An AI Powered Mobile Fitness Safety and Improvement Application," 2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI), Washington, DC, USA, pp.**980-984, 2021.** DOI: 10.1109/ICTAI52525.2021.00156.

[16] N. Faujdar, S. Saraswat, S. Sharma, "Human Pose Estimation using Artificial Intelligence with Virtual Gym Tracker," 2023 6th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, pp.**1-5, 2023**. DOI: 10.1109/ISCON57294.2023.10112064.

[17] H.V.R. Podduturi, C. Varla, K.R. Gopaldinne, N. Bhukya, R.K. Reddy Nallagondu, G.S. Bapiraju, "Smart Trainer using OpenCV," 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA), Uttarakhand, India, pp.**477-480, 2023.** DOI: 10.1109/ICIDCA56705.2023.10099780.

[18] T.T. Tran, J.W. Choi, C. Van Dang, G. SuPark, J.Y. Baek, J.W. Kim, "Recommender System with Artificial Intelligence for Fitness Assistance System," 2018 15th International Conference on Ubiquitous Robots (UR), Honolulu, HI, USA, pp.**489-492, 2018.** DOI: 10.1109/URAI.2018.8441895.

[19] R. Bi, D. Gao, X. Zeng, Q. Zhu, "LAZIER: A Virtual Fitness Coach Based on AI Technology," 2022 IEEE 5th International Conference on Information Systems and Computer Aided Education (ICISCAE), Dalian, China, pp.**207-212, 2022.** DOI: 10.1109/ICISCAE55891.2022.9927664.

[20] X. Li, M. Zhang, J. Gu, Z. Zhang, "Fitness Action Counting Based on MediaPipe," 2022 15th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Beijing, China, pp.**1-7, 2022.** DOI: 10.1109/CISPBMEI56279.2022.9980337.

[21] R. Achkar, R. Geagea, H. Mehio, W. Kmeish, "SmartCoach personal gym trainer: An Adaptive Modified Backpropagation approach," 2016 IEEE International Multidisciplinary Conference on Engineering Technology (IMCET), Beirut, Lebanon, pp.**218-223, 2016.** DOI: 10.1109/IMCET.2016.7777455.

[22] K.B. Lee, R.A. Grice, "The Design and Development of User Interfaces for Voice Application in Mobile Devices," 2006 IEEE International Professional Communication Conference, Saratoga Springs, NY, USA, pp.**308-320, 2006.** DOI: 10.1109/IPCC.2006.320364.

[23] A. Chaudhry, M. Batra, P. Gupta, S. Lamba, S. Gupta, "Arduino Based Voice Controlled Robot," 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, pp.**415-417, 2019.** DOI: 10.1109/ICCCIS48478.2019.8974532.

[24] S. Chakraborty, N. De, D. Marak, M. Borah, S. Paul, V. Majhi, "Voice Controlled Robotic Car Using Mobile Application," 2021 6th International Conference on Signal Processing, Computing and Control (ISPCC), Solan, India, pp.**1-5, 2021.** DOI: 10.1109/ISPCC53510.2021.9609396.

[25] F. Salih, M.S.A. Omer, "Raspberry pi as a Video Server," 2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), Khartoum, Sudan, pp.**1-4, 2018.** DOI: 10.1109/ICCCEEE.2018.8515817.

[26] M. Arshad Khan, M. Kenney, J. Painter, D. Kamale, R. BatistaNavarro, A. Ghalamzan-E, "Natural Language Robot Programming: NLP integrated with autonomous robotic grasping," arXiv e-prints, **2023**. DOI: 10.48550/arXiv.2304.02993.

[27] Anand John, Divyakant Meva, "A Comparative Study of Various Object Detection Algorithms and Performance Analysis", International Journal of Computer Sciences and Engineering, Vol.**8**, Issue.**10**, pp.**158-163, 2020.**

[28] N. Raviteja, M. Lavanya, S. Sangeetha, "An Overview on Object Detection and Recognition", International Journal of Computer Sciences and Engineering, Vol.**8**, Issue.**2**, pp.**42-45, 2020.**

AUTHORS PROFILE

Shreyas Walke has completed a Bachelor of Engineering degree in Artificial Intelligence and Data Science at PES Modern College of Engineering, Pune, affiliated with Savitribai Phule Pune University. His areas of proficiency and interest include developing computer vision applications powered by AI, integrating AI with embedded systems, and developing games using the Unity engine.

Yash Wadekar has completed a Bachelor of Engineering in Artificial Intelligence and Data Science at PES Modern College of Engineering, Pune, affiliated with Savitribai Phule Pune University. His areas of interest and expertise include AI-driven applications, IoT systems, backend architecture development, and Linux development.

Aditya Ladawa has completed his B.E. in Artificial Intelligence and Data Science from PES Modern College of Engineering, Pune, affiliated with Savitribai Phule Pune University. His areas of expertise include backend systems, computer vision and NLP integrated system development, big data and data science applications, and generative AI.

Pratik Khopade has completed his B.E. in Artificial Intelligence and Data Science from PES Modern College of Engineering, Pune, affiliated with Savitribai Phule Pune University. His areas of interest and expertise include computer vision-based applications and Java development.

Dr. Shraddha Viraj Pandit graduated from Rajiv Gandhi Technological University in Bhopal, India, with a Ph.D. in Computer Science and Engineering. She is currently a professor and the head of the Department of Artificial Intelligence and Data Science at PES Modern College of Engineering in Pune. Her research interests include Artificial Intelligence, Swarm Intelligence, Databases & Data Mining, and Image Processing. Dr. Pandit has made a significant impact in her field with over 40 research papers published in esteemed journals and conferences worldwide. She has supervised more than fifteen master's level dissertations. Dr. Pandit is an active member of several academic and scientific societies, including the International Association of Engineers (IAENG), the Institute of Electrical and Electronics Engineers (IEEE), and the Computer Society of India (CSI).

