International Journal of Computer Sciences and Engineering
Vol.12, Issue.3, pp.30-36, March 2024
ISSN: 2347-2693 (Online)
Available online at: www.ijcseonline.org

Review Article

Evolution in Real-Time Automated systems for Personalized Exercise Guidance and Monitoring

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: 01/Feb/2024; **Accepted**: 04/Mar/2024; **Published**: 31/Mar/2024. **DOI**: https://doi.org/10.26438/ijcse/v12i3.3036

Abstract: This comprehensive review delves into the dynamic realm of AI-driven fitness assistance and robotic navigation, exploring the evolving challenges and advancements in human pose estimation, fitness assessment, and user engagement during workout sessions. The surveyed studies employ diverse methodologies, spanning from real-time exercise pose identification using OpenCV and MediaPipe to innovative applications like sound localization and deep learning. The paper also explores the integration of robotics in fitness assistance, showcasing systems for social support and personalized workout recommendations. Furthermore, it investigates advancements in robotic navigation, employing both complex and simplified approaches to seamlessly integrate into workout scenarios. This integration aims to provide in-depth workout analysis and accurate guidance to users while autonomously navigating the environment. The convergence of computer vision, machine learning, image processing, and the Internet of Things emerges as a pivotal approach, offering a holistic solution for immersive fitness experiences in both home and gym settings.

Keywords: Autonomous Robot Navigation, Human Pose Estimation, Mediapipe, Computer Vision, Deep Learning

1. Introduction

Physical fitness represents a fundamental condition of health and overall well-being, encompassing the ability to engage in various sports, work-related tasks, and everyday activities. 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].

The importance of correct posture during physical training cannot be overstated. It enhances the effectiveness of your movements and reduces the risk of injury. Incorrect posture can result in discomfort, muscle tension, and possibly long-term damage to joints and muscles. For those who don't have access to a personal trainer, it can be difficult to ensure they are exercising correctly. This is where AI can assist, offering real-time feedback on posture during workouts. The concept of using AI for workout assistance, provided by numerous digital fitness apps and trainers, is now well-established. Recently, interactive home exercise systems have been developed using deep learning technologies for human posture estimation and action recognition. The ability of these systems to give precise, instant feedback on exercise posture depends on the strength of its intelligent posture prediction feature, which uses computer vision and machine learning to recognize different body parts and make predictions by analyzing the formed angles.

However, such systems for home workouts have had much difficulty in mobility due to the non-flexibility of the device used for monitoring purpose and also struggles with user comfort, accessibility, and other technical challenges as these systems are integrated into mobile phones or immobile devices which lack efficiency. To tackle this concern, in this paper, we present rigorous research and suitable techniques for an autonomous robot to provide real-time monitoring and guidance during workout sessions, enhancing the effectiveness and safety of physical fitness activities.

2. Current Methodologies

2.1 Common Frameworks and Approach to Pose Estimation

Pose estimation is a critical component within the realm of vision-based machine learning, focusing on the identification and extraction of significant body landmarks. These landmarks are essentially pivotal joints that enable movement and are integral for control and muscle engagement. This domain utilizes advanced computational techniques to pinpoint key landmarks on the human body, facilitating a myriad of applications ranging from activity recognition to enhancing interactive gaming experiences. In Python, several frameworks such as OpenPose, PoseNet, BlazePose by Mediapipe, and MoveNet offer specialized capabilities for pose estimation. Each of these frameworks has its own set of features, advantages, and considerations that make them suitable for different use cases. For instance, OpenPose offers robust multi-person detection, PoseNet provides a good balance between speed and accuracy, BlazePose by Mediapipe is known for its efficiency on mobile devices, and MoveNet shines with its speed and accuracy in real-time applications.

Table 1. Frameworks with their basic specifications

Frameworks	Year of release	Approach	body landmarks detection count
OpenPose	2017	Bottom-up	17
PoseNet	2017	Top-down	17
MoveNet	2021	Bottom-up	17
BlazePose	2019	Top-down	33

Given that many exercise guidance applications operate on web or embedded systems, they rely on efficient hardware to handle numerous calculations in real-time. This processing capability is essential for delivering the necessary information to analyze movements and offer feedback effectively. MoveNet, a model rooted in the MobileNetV2 architecture, stands out for its ability to predict the 2D locations of human joints from RGB images. It offers two distinct variants: Lightning and Thunder. The Lightning variant prioritizes speed, making it ideal for real-time inference without compromising performance. On the other hand, the Thunder variant prioritizes accuracy, albeit at a slightly reduced speed. The model produces a skeletal representation as its inference output, delineating the human body's joint positions. BlazePose, an integral component of Mediapipe's Pose solution, operates through a two-stage pipeline. Firstly, the BlazePose detector identifies regions of interest (ROIs) corresponding to individuals within the image. Subsequently, BlazePose GHUM 3D calculates both 2D and 3D keypoints, with the latter referencing the hip center. Similar to MoveNet, BlazePose offers Lite, Heavy, and Full variants, each striking a balance between speed and accuracy. PoseNet, the chosen version for this study, leverages the MobileNetV1 architecture to predict 2D joint locations from images. It can detect up to five persons simultaneously, although our focus lies on single-person detection for comparative analysis. The output of PoseNet comprises a skeletal representation, delineating the keypoints for the observed individual. These models, with their respective variants and capabilities, form the cornerstone of pose estimation research, offering nuanced solutions tailored to diverse application requirements [30].

The procedural framework for pose estimation typically starts by feeding a video stream into the pose detection module. This step is essential because it sets the stage for identifying key body landmarks. Once the video stream is enriched with landmark data, it undergoes processing using OpenCV. During this stage, the system calculates angles between different joints, which is crucial for accurately identifying posture. After calculating angles, the system moves on to tasks such as recognizing posture and counting repetitions for specific exercises. Finally, the system provides feedback to the user based on the analysis of pose estimation and the subsequent calculations performed. This feedback is tailored according to the needs of the application and the main aim, which is to provide guidance to the user [2], [12].

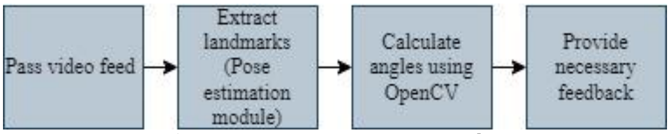

Figure 1. Basic pose estimation pipeline

2.2 Approaches in handling Natural language based interactions

Natural language has become crucial for exchanging information and facilitating interactions, emerging as the most intuitive mode of communication. Across a wide array of applications and systems, researchers have increasingly embraced the integration of voice communication to enhance the seamless interaction between humans and machines. This NLP-based approach is prominently observed in various technologies such as voice assistants, chatbots, and human-robot interactions employed in robot programming [25].

In the context of human-computer interaction (HCI), several challenges emerge that influence the design and implementation of computer systems aimed at facilitating effective communication and interaction between humans and machines. One significant challenge lies in aligning language and ontology between users and computer systems. Achieving agreement at both the linguistic and conceptual levels is essential for ensuring clear communication and mutual understanding. However, variations in language use and individual perceptions of reality can complicate this alignment process. Moreover, as computer systems, particularly autonomous mobile robots, operate in dynamic environments where they interact with other robots, people, and objects, ensuring safety and adaptability becomes crucial. Designing systems that can adjust decisions sensibly and safely based on environmental challenges and system capabilities presents a formidable challenge in HCI.

Natural Language Processing (NLP) encompasses a diverse array of methodologies tailored to analyze and comprehend human language within digital systems and applications. These approaches span from rule-based systems, which rely on predefined linguistic rules and patterns, to statistical models that employ machine learning algorithms for text categorization and pattern recognition, and finally to deep learning techniques leveraging neural networks for context understanding and language generation. Rule-based NLP systems are adept at tasks such as information extraction and chatbot responses, where explicit rules govern language processing. In contrast, statistical NLP models excel in sentiment analysis and spam detection by learning patterns from labeled data. Deep learning NLP techniques, powered by deep neural networks, have revolutionized tasks like language translation and text summarization by capturing intricate linguistic structures. Hybrid NLP methods integrate elements from multiple approaches, offering enhanced performance and versatility across various applications. In robotics, NLP facilitates seamless human-robot interaction, enabling robots to comprehend and execute verbal instructions effectively. Through the integration of these diverse NLP methodologies, researchers aim to advance the capabilities of language processing systems, driving innovation in fields ranging from communication technologies to intelligent automation.

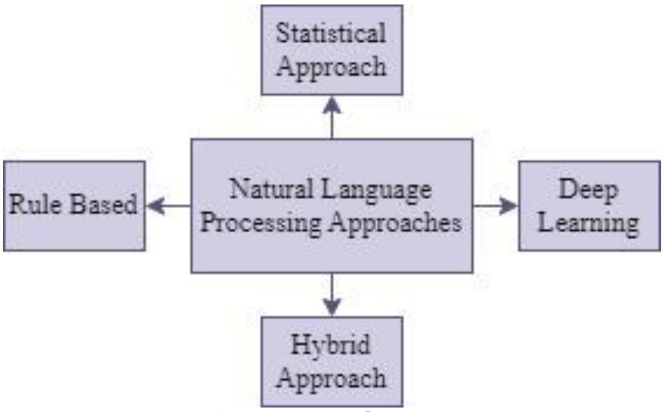

Figure 2. Approaches to NLP

One straightforward approach that leverages Natural Language Robot Programming (NLRP) to enhance human-computer interaction involves integrating a dependency parser, particularly implemented using spaCy, with autonomous algorithms. This setup facilitates the analysis of natural language commands issued by human users. Through this process, the syntactic structure of sentences is extracted, generating a spatial description clause (SDC) containing event, object, place, and path components. Subsequently, these components are utilized to generate action commands for the robot controller, enabling real-time task execution based on human speech inputs. Additionally, the system incorporates various levels of command, including task-level, operational-space level, joint-space level, and action level commands, to provide a versatile and intuitive interface for controlling the robotic system within the broader framework of human-computer interaction. [25].

2.3 Exercise Classification Approaches

The advancements in Machine Learning (ML) and Computer Vision (CV) have notably boosted activity recognition across different scenarios, with exercise recognition being a significant application of Human Activity Recognition (HAR). Two primary methods are used in exercise recognition tasks: sensor-based and machine learning model-based activity recognition.

Sensor-based activity recognition is a crucial field of research. It uses data from a variety of sensors to automatically detect and classify various activities or movements performed by individuals. The applications for this approach span health monitoring, sports analysis, human-computer interaction (HCI), and assistive technologies. Systems for sensor-based activity recognition employ sensors like Inertial Measurement Units (IMUs), Electromyography (EMG) sensors, pressure sensors, and Heart Rate Monitors (HRMs) to gather relevant data. Key characteristics for activity recognition are then represented by extracting features from the raw sensor data, which includes statistical measures, frequency-domain features, and time-domain features. Supervised machine learning algorithms, such as Support Vector Machines (SVM), Random Forests, k-Nearest Neighbors (k-NN), Convolutional Neural Networks (CNNs), or Recurrent Neural Networks (RNNs), are trained on labeled sensor data. The model's performance is assessed using separate test data, based on factors such as accuracy, precision, recall, and F1-score. After deployment, the model continuously processes sensor data in real-time for activity recognition, accounting for computational efficiency, power consumption, and adaptability to environmental changes. Regular refinement and adaptation of these systems are necessary for enhancing accuracy and usability in practical settings.

Exercise recognition, a part of the HAR process, involves four main steps: data acquisition, pre-processing, model training, and performance evaluation. The first step involves selecting an appropriate HAR device based on the target application, taking into account aspects like privacy and computational cost. The collected data may contain noise or unwanted signals, which are removed in the second step through methods like low-pass or high-pass filters and regional segmentation. The third step trains the HAR model using Machine Learning (ML) or Deep Learning (DL) techniques. ML-based methods are suitable for hand-crafted features, while DL frameworks are preferred for automated feature extraction. DL also allows for knowledge reusability and handling large datasets, aiding in the creation of hybrid models that can identify spatial and temporal features. The final step tests the HAR model's efficiency by applying it to real-world data, which may differ based on physical factors like age and physique. A successful HAR model should maintain performance independent of such factors, ensuring robustness across different applications [31].

Machine learning models for activity recognition differ from sensor-based methods as they train directly on data without specific sensor inputs. This approach primarily depends on computer vision techniques, using images or videos of individuals performing exercises. Deep learning models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), are adept at independently identifying features and patterns from visual data. By training on labeled datasets containing diverse images or videos of various exercises, researchers can develop robust classifiers that accurately distinguish different activities based solely on visual cues. This method allows for the creation of advanced recognition systems that comprehend exercise movements with high accuracy, removing the need for specialized sensors. As a result, it has broad applications in fields like fitness tracking, sports analysis, and healthcare monitoring.

2.4 Exercise Rep Counting Approaches

Exercise repetition counting using sensors involves classifying and tracking movements during workouts using wearable or equipment-embedded sensors like accelerometers, gyroscopes, and magnetometers. These sensors collect motion data which is further interpreted using machine learning techniques, like Support Vector Machines or Random Forests, that are trained on labeled datasets to identify specific exercise motions and count repetitions. Another strategy, the criteria-based approach, uses predefined rules or thresholds to detect repetitions from sensor data, such as identifying peaks in accelerometer data corresponding to repetitions. This approach is versatile, providing real-time feedback and customizable to various exercises, making it ideal for fitness tracking and personalized workout schedules. On the other hand, machine learning approach for exercise repetition counting involves analyzing images or video feeds of individuals exercising to identify and count repetitions. This method uses computer vision techniques to train machine learning models, like Convolutional Neural Networks, on labeled datasets of images or videos showcasing different exercises and their repetitions. The trained model then processes new video feeds or images, recognizing and counting repetitions. This approach is scalable and flexible, capable of recognizing a variety of exercises and adapting to different environments.

Moreover, angle calculation approach for exercise repetition counting uses OpenCV, a computer vision library, in conjunction with a pose estimation module to calculate joint angles from body landmarks during exercise movements. This method captures videos of individuals exercising, and the pose estimation module identifies key body landmarks and joints. These landmarks are then used to calculate joint angles specific to the performed exercise. Changes in joint angles over time are used to detect repetitions, with set thresholds or criteria marking the completion of each repetition. This approach, which offers real-time repetition counts and insights into exercise form and technique, is valuable for fitness coaching and rehabilitation programs.

2.4 Robot Navigation

The field of robot navigation is centered on several key concepts that enable a robot to move autonomously within its environment. These include localization, or the use of sensors such as GPS, odometry, and visual odometry to determine the robot's position; mapping, which uses sensor data to create environmental representations; path planning for creating routes from the robot's current position to a specific goal without colliding with obstacles; obstacle avoidance to prevent collision with detected obstacles; simultaneous localization and mapping (SLAM) for creating maps while also localizing the robot within those maps; and sensor fusion, which combines data from multiple sensors for more precise estimations. These foundational principles allow for the creation of robust navigation systems, enabling robots to safely and effectively navigate a variety of real-world situations, from indoor settings to outdoor landscapes, and are critical to progressing research and applications in robotics.

For obstacle detection and avoidance, a method is used in autonomous robotic vehicles that combines sensor technology, data analysis, and decision-making algorithms. The vehicle is equipped with ultrasonic sensors facing various directions that continuously emit signals and calculate their reflection when they hit an obstacle. When an obstacle is detected within a predefined distance, typically 15 cm, the built-in microcontroller activates the obstacle detection feature. It evaluates sensor data from the front, left, and right to determine the best path. By comparing the distances to obstacles in different directions, the microcontroller can intelligently direct the vehicle to avoid potential collisions. This decision-making process allows the vehicle to steer towards the path with the most space, thereby reducing the likelihood of crashes. This technique, which combines sensor technology, real-time data analysis, and vehicle control, enables the autonomous vehicle to safely navigate its surroundings while avoiding obstacles. This approach underscores the importance of integrating hardware and software components to ensure efficient and reliable obstacle avoidance in autonomous robotic systems [29].

Creating a strong, reliable Multi-sensor Fusion Robot Navigation System requires several crucial steps. This starts with integrating different sensors like LiDAR, cameras, and IMUs to gather comprehensive data about the environment. Then, data fusion methods, such as Kalman filtering or Bayesian inference, are applied to combine the information from various sensors, thereby increasing the system's accuracy and dependability [29]. Afterward, a mapping algorithm like SLAM is used to continuously create and update a model of the robot's environment. Path planning algorithms, such as A* or D* search, are then used to create the best possible paths, taking into account both sensor data and any environmental limitations.

3. Related Work

G. Taware et al. [3] initiated a research project introducing an application tailored to recognize and analyze users' exercise poses, count designated repetitions, and offer comprehensive feedback for posture improvement. The application, developed using JavaScript, Node.js, and libraries like OpenCV and MediaPipe, employs a two-step tracking machine learning approach. This involves using trackers to locate events or movements in live video and predicting critical moments in the target area based on input from recent videos.

S. Kardam et al. [4] proposed a system addressing color format challenges in image processing for compatibility between OpenCV and MediaPipe. The system begins by recoloring the image, performs detections using a given model, and then reverts the color format back to BGR for rendering with OpenCV.

G. Dsouza et al. [10] improved upon existing systems by incorporating deep learning and convolutional neural networks for human pose estimation. Their approach involves training the model with diverse sample data, creating a trained model with distinct identity coordinates for each body part. The

COCO and MPII model designation is applied, emphasizing the importance of an increased sample dataset for improved system performance. The system also features a user-friendly graphical interface developed using the Flask framework.

Ce Zheng et al. [12] proposed a model categorized into kinematic, planar, and volumetric types for real-time human pose detection in a two-dimensional space. Their method involves cropping input images to ensure each area contains only one person. The study highlights challenges in recognizing overlapping images and the reliance on motion-capturing systems for precise posture descriptions in a three-dimensional context.

Zell et al. [14] addressed the challenge of assessing hidden information in 2D video data, introducing a unified model for enhanced three-dimensional and spatial rendering and physical modeling simultaneously. Their factorization approach decomposes joint data into camera motion, base poses, and mixing coefficients, followed by a physical model projection. The paper also explores inner forces (joint torques) crucial for biomechanical studies, employing a data-driven statistical approach for direct 3D inference from monocular images.

V. S. P. Bhamidipati et al. [2] introduced a novel real-time posture estimation system which inculcates both Mediapipe and OpenCV to create a comprehensive human posture assessment solution. Initially, the the video stream of the user with their natural poses are captured by the camera and subsequently MediaPipe is employed to identify key landmarks on the human body. These recognized landmarks are passed onto the computer vision's open source library called OpenCV, where angle calculations are performed to evaluate and ascertain the user's posture. Users receive instant evaluations on their posture and are presented with actionable recommendations for improvement through the system. It has a robust performance under various luminosity conditions and is resilient to background interference and applicable across various exercise routines. Leveraging this system for real-time feedback can assist users in refining their posture and technique, thereby minimizing the likelihood of potential injuries.

Fasola et al. [5] developed a comprehensive System for Social Assistance through Robotics. This innovative system is designed to engage users during their exercise sessions by utilizing real-time recognition of hand movements, thus bringing a whole new dimension to physical fitness practices. The primary goal of this ingenious system is to provide detailed instructions on the correct exercise procedures, carry out an accurate assessment of the user's performance, and offer continuous encouragement and motivation. This is achieved through a specialized focus on recognizing the user's arm postures, which have been meticulously optimized for a streamlined exercise arrangement. This optimization ensures that users are not only performing their exercises correctly but also maximising their benefits from each session. The system serves as a virtual fitness assistant, guiding and supporting users on their journey towards improved physical health and wellness.

T. T. Tran et al. [18] enhanced fitness assistance and guidance systems by introducing a recommendation system (RS) to enhance the current fitness assistance and guidance system. This RS provides personalized workout recommendations for both newcomers and experienced users. To forecast appropriate workouts for beginners, they utilized neural networks and logistic regression techniques. Furthermore, they developed an agent with reinforcement learning capabilities using the Soar architecture to aid users in choosing workouts tailored to their specific conditions. The authors posit that the patterns of volunteers act as reference data for predicting and providing workout recommendations to newcomers through the fitness assistance system during their initial experience. User profiles are established through input data that includes various parameters like individual characteristics, exercise preferences, and a measure of maximal strength used in weight training. The suggestions presented in the output comprise exercise weight, repetitions, and rest intervals for each recommended set.

6. Conclusion

In conclusion, the intersection of technology and fitness has led to significant advancements in ensuring safer and more effective workout experiences. From sophisticated pose estimation systems leveraging deep learning algorithms to real-time feedback mechanisms and personalized workout recommendations, the landscape of fitness assistance continues to evolve rapidly. The integration of artificial intelligence, computer vision, and natural language processing not only enhances the monitoring and guidance during exercise sessions but also fosters a deeper level of engagement and motivation among users. As these technologies continue to mature, the potential for improving overall health and well-being through autonomous fitness assistance becomes increasingly promising. Moving forward, continued research and innovation in this field hold the key to unlocking even greater benefits for individuals seeking to lead healthier and more active lifestyles.

Conflict of Interest

The authors declare no conflicts of interest with any entities or individuals regarding the subject matter of this paper. No financial or non-financial support has been received from any parties connected to this review's content. Our findings and opinions are derived from an independent and objective evaluation of the available research and data.

Funding Source

The study was carried out independently without any specific grant support from public, commercial, or non-profit funding agencies. It was conducted without external financial aid, ensuring the research's neutrality and objectivity.

Author's Contribution

Author-1 led the project's conceptualization and compiled the literature review and assisted in manuscript development. Author-2 developed the manuscript and identified methodologies and defined processes. Author-3 assisted in refining and editing the manuscript, detailing underlying techniques and algorithms from research papers. Author-4 collected insights and provided analysis of research findings. Author-5 supervised the research and provided overall guidance.

References

[1] Štajer, Valdemar, Milovanović, Ivana, Todorović, Nikola, Ranisavljev, Marijana, Pišot, Saša, and Drid, Patrik. "Let's (Tik) Talk About Fitness Trends." Frontiers in Public Health, Vol. **10**, pp. **899949, 2022.**

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

[3] G. Taware, R. Kharat, P. Dhende, P. Jondhalekar, and R. Agrawal, "AI-Based Workout Assistant and Fitness Guide," in Proceedings of the 2022 6th International Conference On Computing, Communication, Control And Automation (ICCUBEA), Pune, India, pp. **1-4, 2022.**

[4] S. Kardam and S. Maggu, "AI Personal Trainer using OpenCV and Python," International Journal of Computer Science, Vol. **9**, Issue **12**, pp. **2425-2427, 2021.**

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

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

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

[8] F. Haugg, M. Elgendi, and C. Menon, "GRGB rPPG: An Efficient Low-Complexity Remote Photoplethysmography-Based Algorithm for Heart Rate Estimation," Bioengineering, vol. **10**, no. **2**, pp. **243, 2023.**

[9] G. Dsouza, D. Maurya, and A. Patel, "Smart gym trainer using Human pose estimation," in Proceedings of the 2020 IEEE International Conference for Innovation in Technology (INOCON), Bangalore, India, pp. **1-4, 2020.**

[10] L. Xu et al., "ZoomNAS: Searching for Whole-Body Human Pose Estimation in the Wild," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. **45**, no. **4**, pp. **5296-5313, 2023.**

[11] Ce Zheng, Wenhan Wu, Taojiannan Yang, Sijie Zhu, Chen Chen, Ruixu Liu, Ju Shen, Nasser Kehtarnavaz, and Mubarak Shah, "Deep Learning-Based Human Pose Estimation: A Survey," **2020.**

[12] Chen, Steven & Yang, Richard, "Pose Trainer: Correcting Exercise Posture using Pose Estimation," **2018.**

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

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

[15] N. Faujdar, S. Saraswat, and S. Sharma, "Human Pose Estimation using Artificial Intelligence with Virtual Gym Tracker," in Proceedings of the 2023 6th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, pp. **1-5, 2023.**

[16] H. V. R. Podduturi, C. Varla, K. R. Gopaldinne, N. Bhukya, R. K. Reddy Nallagondu, and G. S. Bapiraju, "Smart Trainer using OpenCV," in Proceedings of the 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA), Uttarakhand, India, pp. **477-480, 2023.**

[17] T. T. Tran, J. W. Choi, C. Van Dang, G. SuPark, J. Y. Baek, and J. W. Kim, "Recommender System with Artificial Intelligence for Fitness Assistance System," in Proceedings of the 2018 15th International Conference on Ubiquitous Robots (UR), Honolulu, HI, USA, pp. **489-492, 2018.**

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

[19] X. Li, M. Zhang, J. Gu, and Z. Zhang, "Fitness Action Counting Based on MediaPipe," in Proceedings of the 2022 15th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Beijing, China, pp. **1-7, 2022.**

[20] R. Achkar, R. Geagea, H. Mehio, and W. Kmeish, "SmartCoach personal gym trainer: An Adaptive Modified Backpropagation approach," in Proceedings of the 2016 IEEE International Multidisciplinary Conference on Engineering Technology (IMCET), Beirut, Lebanon, pp. **218-223, 2016.**

[21] K. B. Lee and R. A. Grice, "The Design and Development of User Interfaces for Voice Application in Mobile Devices," in Proceedings of the 2006 IEEE International Professional Communication Conference, Saragota Springs, NY, USA, pp. **308-320, 2006.**

[22] A. Chaudhry, M. Batra, P. Gupta, S. Lamba, and S. Gupta, "Arduino Based Voice Controlled Robot," in Proceedings of the 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, pp. **415-417, 2019.**

[23] S. Chakraborty, N. De, D. Marak, M. Borah, S. Paul, and V. Majhi, "Voice Controlled Robotic Car Using Mobile Application," in Proceedings of the 2021 6th International Conference on Signal Processing, Computing and Control (ISPCC), Solan, India, pp. **1-5, 2021.**

[24] F. Salih and M. S. A. Omer, "Raspberry Pi as a Video Server," in Proceedings of the 2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), Khartoum, Sudan, pp. **1-4, 2018.**

[25] Muhammad Khan, Max Kenney, Jack Painter, Disha Kamale, Riza Batista-Navarro, and Amir Ghalamzan, "Natural Language Robot Programming: NLP integrated with autonomous robotic grasping," **2023.**

[26] C. Wang, et al., "Autonomous mobile robot navigation in uneven and unstructured indoor environments," in Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vancouver, BC, Canada, pp. **109-116, 2017.**

[27] K. Yokoyama and K. Morioka, "Autonomous Mobile Robot with Simple Navigation System Based on Deep Reinforcement Learning and a Monocular Camera," in Proceedings of the 2020 IEEE/SICE International Symposium on System Integration (SII), Honolulu, HI, USA, pp. **525-530, 2020.**

[28] Y. Sun, J. Yang, D. Zhao, and S. Li, "Personal Care Robot Navigation System Based on Multi-sensor Fusion," in Proceedings of the 2021 IEEE International Conference on Intelligence and Safety for Robotics (ISR), Tokoname, Japan, pp. **408-412, 2021.**

[29] Goswami, Shankha & Sahoo, Sushil Kumar. "Design of a Robotic Vehicle to Avoid Obstacle Using Arduino Microcontroller and Ultrasonic Sensor". Journal of Advances in Transportation Engineering, **2023.**

[30] Cristina Bolaños, Jesús Fernández-Bermejo, Javier Dorado, Henry Agustín, Félix Jesús Villanueva, María José Santofimia. A Comparative Analysis of Pose Estimation Models as Enablers for a Smart-Mirror Physical Rehabilitation System. Procedia Computer Science, Vol. **207**, pp. **2536-2545, 2022.**

[31] Gupta, N., Gupta, S.K., Pathak, R.K. et al. "Human Activity Recognition in Artificial Intelligence Framework: A Narrative Review." Artificial Intelligence Review, Vol. **55**, pp. **4755-4808, 2022.**

AUTHORS PROFILE

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

Yash Wadekar is presently enrolled in PES Modern College of Engineering, Pune, which is connected to Savitribai Phule Pune University, to pursue a Bachelor of Engineering in Artificial Intelligence and Data Science. His areas of interest and competence are in AI-driven applications, IOT systems, backend architecture development, and Linux development.

Aditya Ladawa is pursuing his B.E in Artificial Intelligence and Data Science from PES Modern College of Engineering, Pune, which is affiliated with Savitribai Phule Pune University. His domain of expertise include backend systems, computer vision and NLP integrated system development, Big data and Data Science applications and also Generative AI.

Pratik Khopade is pursuing his B.E in Artificial Intelligence and Data Science from PES Modern College of Engineering, Pune, affiliated with Savitribai Phule Pune University. His domain of interest and skills lies in 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. At PES Modern College of Engineering in Pune, she is currently employed as a professor and the head of the department of artificial intelligence and data science. Artificial Intelligence, Swarm Intelligence, Databases & Data Mining, and Image Processing are the topics of her research interests. With more than 40 research papers published in highly regarded journals and conferences worldwide, Dr. Pandit has made a substantial impact on her field. More than fifteen master's level dissertations have been supervised by her. The International Association of Engineers (IAENG), the Institute of Electrical and Electronics Engineers (IEEE), and the Computer Society of India (CSI) are just a few of the academic and scientific societies that Dr. Pandit actively participates in.

