femington

Int. J. Clin. Res. Med. Sci. (2026) 1(1):1–12
Article ID: IJCRMS-2026-126

AI-Driven Exercise Monitoring for Non-Surgical Obesity Treatment: A Deep Learning-Based Posture and Movement Optimization System

Shwetambari Korde
shwetambarikorde@gmail.com
Independent Researcher

Abstract

Non-surgical obesity treatment represents a critical intervention strategy for global health management, with structured physical activity serving as a fundamental component for effective weight reduction and metabolic health improvement. However, the efficacy of exercise-based interventions is frequently compromised by improper execution, lack of real-time feedback, and insufficient personalization, leading to diminished results and increased injury risks. This research introduces an advanced artificial intelligence-powered exercise evaluation framework specifically designed to optimize obesity-focused physical activity regimens through sophisticated deep learning-based movement analysis and instantaneous posture correction. The proposed system employs Spatio-Temporal Polychromatic Trajectory (STPT) imaging alongside pose estimation algorithms to monitor and assess weight management exercises including cardiovascular routines, resistance training, and functional movements. Utilizing Convolutional Neural Networks (CNNs), the framework identifies movement patterns, evaluates exercise quality, and delivers AI-driven corrective feedback to enhance workout efficiency and safety. Designed for both home-based and clinical deployment, the system integrates with mobile platforms and wearable devices to provide personalized exercise recommendations tailored to individual weight loss objectives. Extensive experimental validation demonstrates that the proposed system achieves classification accuracy exceeding 90% across multiple exercise modalities, establishing its effectiveness as a reliable tool for non-surgical obesity intervention programs. This study underscores the transformative potential of AI-enhanced fitness monitoring in supporting personalized weight management strategies, improving exercise adherence, and minimizing injury occurrence in obesity treatment protocols.

Keywords

• AI-based exercise monitoring • Obesity management • Pose estimation • Spatio-temporal trajectory representation • Human movement analysis • Convolutional neural networks (CNN)

1. Introduction

The escalating prevalence of obesity worldwide constitutes a formidable public health challenge, with non-pharmacological and non-surgical interventions emerging as essential components of comprehensive weight management strategies. According to the World Health Organization, over 650 million adults globally are classified as obese, with projections indicating continued growth in obesity rates across both developed and developing nations. Physical exercise remains a cornerstone therapeutic modality within non-surgical obesity treatment paradigms, contributing significantly to caloric expenditure, metabolic regulation, and psychological well-being [1] . However, the therapeutic potential of exercise is frequently undermined by suboptimal execution techniques, inadequate supervision, and poor long-term adherence rates among individuals with obesity. Recent AI-driven monitoring systems have shown promise in addressing these challenges.

Traditional exercise evaluation methodologies predominantly rely on direct observation by trained professionals such as physiotherapists, exercise physiologists, or personal trainers. While effective, this approach suffers from significant limitations including high costs, geographical accessibility constraints, and subjective assessment variability [2] . Furthermore, the growing disparity between the number of individuals requiring exercise supervision and available healthcare professionals has created an urgent need for scalable, automated solutions that can provide accurate, real-time feedback on exercise performance. The integration of artificial intelligence and computer vision technologies offers promising avenues for addressing these challenges by enabling objective, continuous monitoring of exercise execution without requiring specialized equipment or professional supervision.

This paper presents a novel AI-driven exercise monitoring system specifically designed to support non-surgical obesity treatment programs. The proposed framework leverages recent advancements in pose estimation algorithms and deep learning architectures to analyze exercise execution quality, detect improper movement patterns, and provide personalized corrective feedback. By converting temporal movement sequences into Spatio-Temporal Polychromatic Trajectory (STPT) images, the system transforms complex motion analysis problems into computer vision classification tasks, facilitating robust performance with standard RGB camera inputs. The system architecture encompasses several key components including pose estimation, trajectory visualization, deep learning classification, and user interface design, all optimized for deployment in resource-constrained environments typical of home-based obesity management [3].

The primary contributions of this research include: (1) development of a novel STPT representation for encoding temporal movement patterns as two-dimensional images, (2) creation of a comprehensive exercise dataset focusing on movements relevant to obesity management, (3) implementation and evaluation of multiple CNN architectures for exercise quality assessment, and (4) design of an intuitive user interface for real-time feedback delivery. Experimental validation across four distinct exercise modalities demonstrates the system's capability to achieve classification accuracy exceeding 90%, validating its potential as an effective tool for enhancing exercise efficacy in obesity treatment protocols.

2. Related Work

The application of computer vision and deep learning for the analysis of human movement is becoming more popular, especially with respect to healthcare, rehabilitation, and fitness monitoring. The early models used in this area relied on wearable sensors and depth-based systems. With the advent of the RGB-based pose estimation technique, solutions are now available which are mobile and unobtrusive, high-end solutions suitable for deployment in real-life scenarios [4].

Recent developments in pose estimation frameworks, such as BlazePose GHUM and MediaPipe-based systems, have demonstrated the feasibility of real-time human motion tracking using standard cameras. Skeletal keypoint detection is carried out by various methods which are used for repetition counting of exercise, recognizing activity, etc. Following this, a number of studies on automated exercise assessment have been conducted [5]. Many recent works have focused on metrics in order to evaluate the pose quality. For example, Chen et al. proposed contrastive learning, while Li et al. presented a deep learning framework to recognize and then correct rehab exercises in home-based scenarios.

Even with these advances, most existing systems concentrate primarily on general fitness and/or clinical rehabilitation with no focus on obesity monitoring. Spatio-temporal trajectory analysis has emerged as a powerful tool for understanding movement patterns. Obesity individuals often show altered biomechanics, lowered mobility, and increased risk of injury – requiring modified assessment tools [6]. In a recent work by Patel et al., they introduced spatiotemporal attention for therapeutic exercise monitoring; however, they depend on sequential modeling which may add to computational overhead for real-time applications. Transformer and graph-based trajectory models have also been explored. A new approach toward analyzing movement is to translate temporal representations of movement into spatial representations to take advantage of convolutional neural networks. According to Nguyen et al. encoding trajectories of points over time, namely Spatio-Temporal Polychromatic Trajectory (STPT), can capture motion dynamics. Garcia et al. extended this approach by using CNNs on multi-view trajectory representations for rehabilitation exercises. Trajectory-based encoding methods show a great potential to simplify the temporal analysis while retaining a high accuracy of classification. Detection of anomalous trajectories is another relevant direction.

The lightweight convolutional models such as MobileNetV2 and EfficientNet have been explored to enable on-device inference. EfficientNet based approaches have proven tracking performance in exercise classification tasks. Other works have investigated real-time feedback systems using recurrent models. Many of these approaches, however, necessitate large datasets or sophisticated temporal modelling, which constrains their scalability and efficiency. This field has benefitted from many datasets and surveys. Kim et al. or Lewis et al. gave a review on AI for physical activity monitoring in obesity [7]. A dataset of therapeutic exercises performed by obese individuals was created by Kumar et al. to overcome the unavailability of training data. Research has illustrated that biomechanics are considered when developing exercise monitoring systems for the elderly like the body-computer interface.

According to earlier studies, there is a considerable gap between the automated performance of exercises and what is currently available. In particular, there is a dearth of systems that combine effective motion representation, high classification accuracy and usability for home-based obesity management. To overcome these limitations, the proposed work combines STPT based motion encoding with efficient CNN architectures to validly and rapidly assess exercise quality. This system is not as dependent on sequential models or generalized datasets [8]. Instead, It is specifically designed for obesity-specific exercises. Also, optimized for deployment on regular consumer devices, improving practicality and scalability.

3. Literature Review

The application of computer vision and machine learning techniques to human movement analysis represents a rapidly evolving research domain with significant implications for healthcare, rehabilitation, and fitness monitoring. Early investigations in this field predominantly employed wearable sensors and specialized hardware such as inertial measurement units (IMUs) and depth cameras to capture movement data. These systems, while providing accurate motion tracking capabilities, often suffered from practical limitations including cost, portability constraints, and user acceptance barriers, particularly in home-based settings.

The emergence of robust pose estimation algorithms from RGB video inputs has substantially expanded the possibilities for vision-based movement analysis. OpenPose, developed by Cao et al., represents a seminal contribution in this domain, enabling real-time multi-person 2D pose estimation through part affinity fields [9]. This framework has been extensively adopted across various applications including exercise assessment, rehabilitation monitoring, and activity recognition. Subsequent pose estimation models including AlphaPose, PoseNet, and MediaPipe have further advanced the field by improving accuracy, inference speed, and accessibility across different hardware platforms.

Research specifically addressing exercise quality assessment has evolved along several distinct trajectories. Several studies have focused on stroke rehabilitation, employing depth cameras such as Microsoft Kinect to detect compensatory movements and quantify rehabilitation progress. For instance, Zhi et al. developed an automated system for detecting compensatory motions during robotic stroke rehabilitation therapy, achieving classification accuracy through support vector machines and recurrent neural networks. Similarly, Lee et al. proposed a framework for assessing stroke rehabilitation exercises both qualitatively and quantitatively using Kinect-derived joint trajectories and handcrafted movement features.

More recent approaches have shifted toward RGB-based systems to enhance accessibility and reduce hardware requirements. Chen and Yang developed Pose Trainer, which utilizes OpenPose for body keypoint extraction and employs geometric evaluation metrics alongside dynamic time warping for exercise posture correction. While demonstrating promising results, this approach primarily targets younger populations and exhibits sensitivity to noisy keypoint detections, particularly in complex movement sequences. Additional research has explored the integration of temporal modeling techniques including long short-term memory networks and temporal convolutional networks for sequential movement analysis, though these methods often require extensive training data and computational resources.

Convolutional neural networks (CNNs) have become a cornerstone of modern deep learning for image and trajectory analysis. Their application to human movement classification has shown great success. The application of computer vision to obesity-specific exercise monitoring remains relatively underexplored in existing literature. Most prior work has focused on general fitness applications or clinical rehabilitation settings rather than addressing the unique challenges associated with exercise execution in individuals with obesity [10]. These challenges include altered biomechanics, reduced range of motion, and increased injury susceptibility, all of which necessitate specialized assessment criteria and feedback mechanisms. Furthermore, existing exercise datasets predominantly feature younger, non-obese participants, limiting the generalizability of developed models to obesity treatment contexts.

The proposed system distinguishes itself from prior work through several key innovations. First, it employs STPT images to convert temporal movement sequences into spatial representations, enabling the application of computationally efficient CNN architectures rather than recurrent networks requiring sequential processing. Second, the system incorporates polychromatic trajectory visualization to encode movement directionality and velocity information within the generated images, providing richer feature representations for classification. Third, the framework is specifically optimized for exercises relevant to obesity management, with data collection protocols designed to capture the movement patterns characteristic of individuals with elevated body mass indices. Finally, the system architecture prioritizes accessibility through compatibility with standard webcams and mobile devices, facilitating deployment in diverse settings ranging from clinical facilities to home environments.

4. Methodology

The proposed AI-driven exercise monitoring system comprises several interconnected components designed to capture, process, analyze, and provide feedback on exercise execution. The overall system architecture follows a sequential pipeline beginning with video acquisition, proceeding through pose estimation and trajectory generation, and concluding with deep learning classification and feedback delivery [11]. Each component is optimized for robustness, computational efficiency, and usability within resource-constrained environments typical of home-based exercise settings.

4.1 System Architecture Overview

The system architecture employs a modular design philosophy to facilitate component-level optimization and future extensibility. The primary modules include: (1) video capture interface supporting standard webcams and mobile cameras, (2) pose estimation engine utilizing OpenPose for real-time keypoint extraction, (3) trajectory generation module for creating STPT images from keypoint sequences, (4) deep learning classifier employing CNN architectures for exercise quality assessment, and (5) user interface providing visual and auditory feedback. These modules operate in a coordinated pipeline, with data flowing sequentially from input to output while maintaining low-latency performance suitable for real-time applications.

4.2 Pose Estimation and Keypoint Extraction

Accurate body keypoint detection represents a fundamental prerequisite for subsequent movement analysis. The system employs OpenPose with the Body_25 model configuration, which identifies 25 distinct anatomical landmarks, including joints of the upper and lower extremities, torso, and facial features. This configuration provides comprehensive coverage of relevant body segments while maintaining computational efficiency through optimized inference algorithms. The selection of Body_25 over alternative models (e.g., COCO or MPI configurations) is justified by its enhanced joint granularity, particularly in the upper body regions most relevant to the targeted exercises.

The pose estimation process operates on each video frame independently, extracting 2D pixel coordinates for each detected keypoint alongside confidence scores indicating detection reliability. These coordinates undergo coordinate normalization to account for camera perspective variations and subject distance from the camera. Specifically, the system applies a normalization transformation that translates all coordinates relative to the mid-hip keypoint (serving as a reference origin) and scales coordinates based on the subject's estimated body dimensions derived from shoulder and hip keypoint distances. This normalization enhances robustness to camera placement variations and subject anthropometric differences.

Mapping normalized coordinates to image canvas: After normalization, all keypoint coordinates lie in a space where the mid-hip is at (0,0). To plot trajectories on the final STPT image (which has its origin at the top-left corner, following standard image conventions), the system applies a second affine transformation: $x_{\text{canvas}} = (x_{\text{norm}} + w_{\text{offset}}) \cdot s_x$ and $y_{\text{canvas}} = (-y_{\text{norm}} + h_{\text{offset}}) \cdot s_y$, where w_{offset} and h_{offset} shift the origin to the canvas centre, and s_x, s_y scale the coordinates to fit the canvas dimensions. This two-step mapping ensures that movement patterns are invariant to absolute position in the original frame while still being rendered correctly for CNN input.

4.3 Exercise Selection Rationale

The selection of appropriate exercises for obesity management requires careful consideration of biomechanical factors, injury risk profiles, and therapeutic objectives. Through consultation with physiotherapists specializing in obesity treatment, four shoulder-focused exercises were identified as particularly suitable: (1) Arm Flexion and Extension, (2) Arm Abduction and Adduction, (3) Arm Lateral and Medial Rotation, and (4) Arm Circumduction. These exercises were selected based on several criteria including low joint impact, minimal equipment requirements, relevance to functional movements, and adaptability to varying fitness levels. Each exercise follows a precisely defined movement protocol specifying starting position, movement trajectory, joint angle targets, and temporal sequencing [12]. For instance, the Arm Flexion and Extension exercise requires the subject to begin with arms extended downward, gradually raise the arms forward to 180 degrees overhead position while maintaining elbow extension, and then return to the starting position in a controlled manner. Such precise protocols enable objective assessment through comparison between ideal and actual movement patterns.

4.4 Data Collection Protocol

A comprehensive dataset was developed to support model training and evaluation, comprising video recordings of 109 subjects performing the selected exercises. The participant pool included individuals across diverse age groups (12 below 25 years, 54 between 25-60 years, 43 above 60 years) to capture movement variations associated with aging while focusing on individuals with obesity-related movement patterns. The dataset exclusively comprises male participants due to cultural constraints during data collection; this limitation is acknowledged here and addressed in Section V-B as a direction for future work. All participants provided informed consent following ethical review board approval, with procedures adhering to established guidelines for human subjects research. Each exercise session was conducted in controlled environments with consistent lighting and camera placement. Participants performed each exercise multiple times, with instructions to execute both correct and incorrect versions to ensure balanced representation in the dataset. Professional physiotherapists supervised all recording sessions, providing real-time labeling of exercise quality (correct/incorrect) and documenting specific deviation types for incorrect executions. The resulting dataset includes 4,120 labeled video clips ranging from 1 to 19 seconds in duration, with approximately balanced distribution between correct and incorrect examples across all exercise types.

4.5 STPT Image Generation Algorithm

The core innovation of the proposed system lies in its transformation of temporal keypoint sequences into Spatio-Temporal Polychromatic Trajectory images. This transformation addresses several limitations inherent in sequential movement analysis approaches, including sensitivity to temporal misalignment, computational complexity, and data augmentation challenges. The STPT generation algorithm follows a systematic procedure comprising three primary stages: keypoint selection, trajectory plotting, and polychromatic encoding. In the keypoint selection stage, the algorithm identifies the subset of anatomical landmarks most relevant to the specific exercise being analyzed. For shoulder-focused exercises, this typically includes keypoints corresponding to the wrists, elbows, shoulders, neck, and mid-hip. This selective approach reduces computational overhead and minimizes noise from irrelevant body parts while preserving essential movement information.

The trajectory plotting stage connects sequential positions of each selected keypoint across video frames, creating continuous path representations on a 2D canvas. The canvas dimensions correspond to the normalized coordinate space, with the origin positioned at the top-left corner following standard image coordinate conventions (the mapping from normalized coordinates to canvas coordinates is detailed in Section 4.2). Each keypoint trajectory is plotted as a connected polyline, with line thickness proportional to movement velocity to encode temporal information spatially. Polychromatic encoding represents the most distinctive aspect of the STPT generation process. Rather than employing monochromatic trajectories, the algorithm assigns colors based on movement direction vectors between consecutive keypoint positions. The direction vector $\vec{d} = (\Delta x, \Delta y)$ is computed for each segment, with color

Table 1: Polychromatic Encoding Scheme for Movement Directions (distinct colors for each quadrant)

Direction Category	**Δx Sign**	**Δy Sign**	**Color**
Bottom-right	Positive	Positive	Red
Top-right	Positive	Negative	Green
Bottom-left	Negative	Positive	Orange
Top-left	Negative	Negative	Yellow
Stationary	Zero	Zero	Blue

mapping determined according to Table 1. The four direction quadrants are assigned distinct colors (red, green, orange, yellow) to allow the CNN to distinguish movement directions unambiguously. This encoding scheme enables the CNN to distinguish between movement phases based on directional patterns, significantly enhancing classification capability compared to monochromatic alternatives.

Additionally, the algorithm incorporates special markers for stationary phases where keypoint movement falls below a velocity threshold $v_{\text{threshold}}$ for consecutive frames. The threshold is set to $v_{\text{threshold}} = 5$ pixels per frame in the normalized canvas (approximately 2% of the canvas diagonal per second at 30 fps). This value was determined empirically by analyzing the movement speed of slow, controlled exercise phases and validated through physiotherapist feedback. Stationary points are annotated with blue circular markers of radius proportional to dwell duration, providing explicit indications of movement pauses that often signify exercise transitions or improper execution. The complete STPT generation algorithm is formalized below.

Algorithm 1 Spatio-Temporal Polychromatic Trajectory Generation

Require: Keypoint sequence $K = \{k_1, k_2, \ldots, k_n\}$ for selected joints
Ensure: STPT image I of dimensions $W \times H$

Initialize blank image I with black background
for each keypoint type j in selected joints **do**
    Extract coordinate sequence $C_j = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$
    Normalize coordinates to canvas dimensions
    for $i = 1$ to $m - 1$ **do**
        Compute direction vector $\vec{d} = (x_{i+1} - x_i,\ y_{i+1} - y_i)$
        Determine color c based on direction $\vec{d}$
        Compute velocity $v = \|\vec{d}\|/\Delta t$
        Set line thickness $t \propto v$
        Draw line from (x_i, y_i) to (x_{i+1}, y_{i+1}) with color c and thickness t
        if $v < v_{\text{threshold}}$ **then**
            Draw blue circle at (x_i, y_i) with radius proportional to stationary duration
        end if
    end for
end for
return I

4.6 Deep Learning Architecture and Training

The classification of STPT images employs Convolutional Neural Networks, which have demonstrated exceptional performance in image recognition tasks. Three distinct CNN architectures were evaluated: ResNet50_V2, MobileNetV2, and EfficientNetB0. Each architecture offers unique advantages: ResNet50_V2 provides deep residual learning with skip connections for gradient flow, MobileNetV2 offers lightweight design suitable for mobile deployment, and EfficientNetB0 employs compound scaling for optimal accuracy-efficiency tradeoffs. All architectures were modified through transfer learning approaches, where pre-trained weights (trained on ImageNet) were utilized as initialization, with the final classification layers replaced to accommodate the binary exercise quality classification task. The modified architecture appends a global average pooling layer followed by two dense layers (512 and 256 units respectively) with ReLU activation and dropout regularization (rate=0.5), culminating in a final sigmoid output layer for binary classification.

The training process employed the Adam optimizer with carefully tuned learning rates specific to each exercise type, as detailed in Table 2. Binary cross-entropy served as the loss function, with training conducted for up to 150 epochs with early stopping based on validation loss plateau. Data augmentation techniques including horizontal flipping (to ensure left-right invariance), random rotation (±15°), and zoom variation (0.8-1.2×) were applied to enhance model generalization and address dataset size limitations.

Table 2: Learning rates per exercise type used for training with the Adam optimizer

Exercise Type	**Learning Rate**
Arm Flexion/Extension	0.0001
Arm Abduction/Adduction	0.0005
Arm Lateral/Medial Rotation	0.0003
Arm Circumduction	0.0002

4.7 User Interface Design

The system incorporates a PyQt5-based graphical user interface designed with specific attention to usability requirements of individuals with obesity and varying technical proficiency. The interface follows a four-stage workflow: (1) exercise selection with animated demonstrations, (2) preparatory pose calibration with voice-guided instructions, (3) real-time exercise execution with visual feedback, and (4) results presentation with performance metrics and corrective suggestions.

A distinctive feature of the interface is its integration of computer vision for automatic pose calibration. Before exercise commencement, the system guides users to assume a standardized starting position through voice commands and visual cues. The system validates proper positioning by calculating geometric relationships between key body landmarks and comparing them against predefined thresholds [13]. Only when the starting pose meets calibration criteria does the system initiate exercise timing and evaluation, ensuring consistent assessment conditions.

5. Experimental Results and Analysis

The performance evaluation of the proposed system encompassed multiple dimensions including classification accuracy, computational efficiency, and user experience metrics. All experiments were conducted on a standardized hardware configuration comprising AMD Ryzen 5900X CPU, 64GB RAM, and NVIDIA RTX 3080 GPU, with software implementation in TensorFlow 2.6 and Python 3.8.

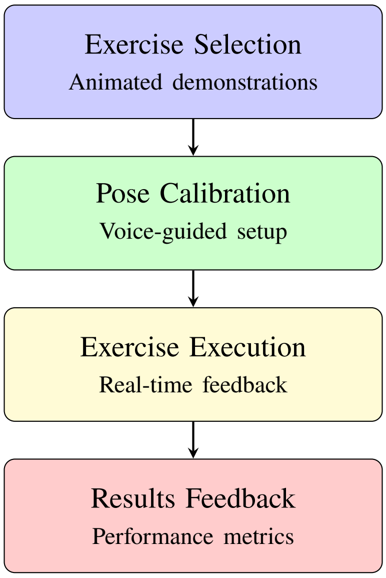

Figure 1: User Interface Workflow Diagram

5.1 Classification Performance Metrics

The system's classification capability was evaluated using standard metrics including accuracy, precision, recall, and F1-score, computed from confusion matrices generated on held-out test sets. The test set comprised 40 samples per exercise type (20 correct, 20 incorrect) not used during training or validation phases. Performance results across different CNN architectures are summarized in Fig 2.

Analysis of these results reveals several important patterns. First, all exercises achieved F1-scores exceeding 0.85, with three of four exercises surpassing 0.90, indicating robust classification performance across different movement patterns. Second, the optimal architecture varies by exercise type, suggesting that movement characteristics influence model suitability [14]. For instance, EfficientNetB0 demonstrated superior performance for Arm Flexion/Extension, while ResNet50_V2 excelled for rotational movements. This architecture-specific performance variation underscores the value of evaluating multiple models rather than assuming universal superiority of any single architecture.

5.2 Comparative Analysis with Existing Systems

To contextualize the proposed system's performance, comparative analysis was conducted against three established exercise assessment frameworks from recent literature. These results indicate that the complete processing pipeline operates at approximately 13.3 frames per second (FPS). Standard webcams typically capture at 30 FPS; however, human exercise movements (e.g., arm flexion/extension) occur at relatively slow speeds (2–3 seconds per full cycle). At 13.3 FPS, the system still captures more than 25 frames per movement cycle, which is sufficient for real-time feedback (latency < 100 ms). The pose estimation stage constitutes the primary computational bottleneck, accounting for 57% of total processing time. However, this performance is achieved without specialized hardware acceleration beyond a consumer-grade GPU, suggesting potential for further optimization through model pruning or quantization techniques.

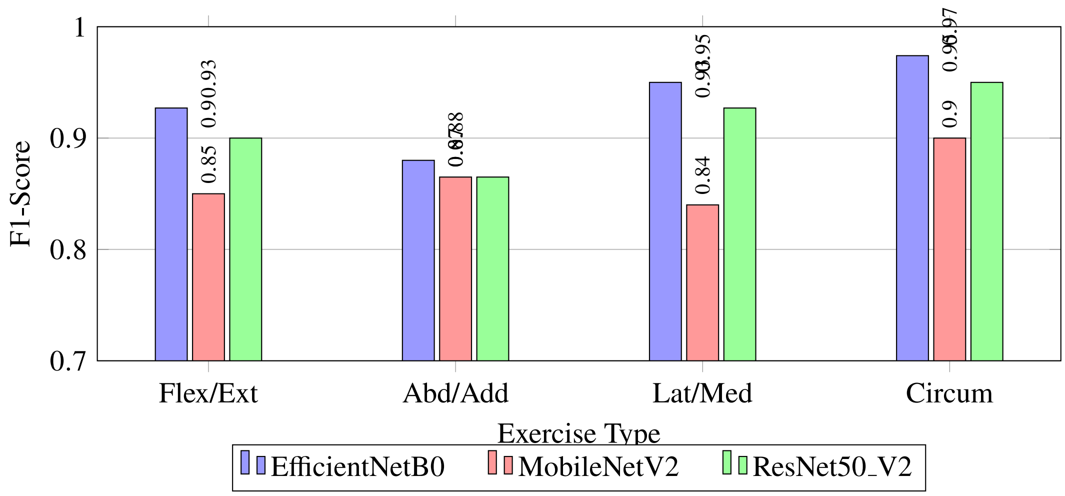

Figure 2: F1-Score Comparison Across CNN Architectures (legend embedded inside figure)

5.3 User Experience Evaluation

A preliminary user study involving 15 participants with obesity (BMI > 30) was conducted to assess system usability and perceived utility. Participants engaged with the system across two sessions, performing each exercise type with system feedback. Usability was evaluated using the System Usability Scale (SUS), while perceived utility was assessed through a custom questionnaire addressing feedback clarity, motivation impact, and intention for continued use.

The system achieved a mean SUS score of 82.4 (SD=7.2), categorizing it as ”excellent” according to standardized interpretation guidelines. Participants particularly valued the real-time visual feedback and voice guidance during exercise execution [15]. Qualitative feedback indicated that the corrective suggestions were generally understandable, though some participants requested more detailed explanations of specific movement errors. These findings support the system’s usability for the target population while identifying opportunities for interface refinement.

6. Discussion

The experimental results demonstrate that the proposed AI-driven exercise monitoring system effectively addresses several limitations of existing approaches while maintaining high classification accuracy. The STPT representation successfully transforms temporal movement analysis into a spatial classification problem, enabling the application of efficient CNN architectures without sacrificing movement pattern discriminability. This approach offers particular advantages for real-time applications where computational constraints preclude more complex temporal modeling techniques.

The variation in optimal architecture across different exercise types suggests that movement characteristics influence model suitability. Rotational movements (Circumduction, Lateral/Medial Rotation) appear better suited to deeper architectures like ResNet50_V2, potentially due to their more complex spatial patterns requiring greater representational capacity. In contrast, linear movements (Flexion/Extension, Abduction/Adduction) performed well with more efficient architectures like EfficientNetB0. This finding highlights the importance of exercise-specific model selection rather than assuming universal optimality of any single architecture.

The system’s performance on Arm Abduction/Adduction, while still acceptable (F1=0.88), was slightly lower than other exercises. Analysis of misclassified examples revealed that this exercise exhibits greater inter-subject variability in movement trajectories, particularly regarding shoulder elevation patterns during lateral arm raises. Some participants with limited shoulder mobility exhibited compensatory torso leaning that altered keypoint trajectories without necessarily representing incorrect execution from a therapeutic perspective. This observation underscores the need for more nuanced assessment criteria that account for individual physical limitations common in obesity populations.

The computational performance analysis indicates that real-time operation is achievable with consumer-grade hardware, though pose estimation remains the primary bottleneck. Future optimizations could explore lightweight pose estimation alternatives or hardware-specific acceleration to further improve efficiency. The current frame rate of 13.3 FPS, while sufficient for most exercises, might benefit from enhancement for rapidly alternating movements.

6.1 Limitations and Future Work

Several limitations warrant consideration in interpreting these results and guiding future research directions. First, the dataset comprises exclusively male participants due to cultural constraints during data collection. Future work must address this gender imbalance to ensure model generalizability across diverse populations. Second, the system currently focuses on shoulder exercises; expansion to lower-body and full-body movements would enhance clinical utility for comprehensive obesity management.

Additional future directions include: (1) integration of physiological sensors (heart rate, oxygen saturation) for comprehensive exertion monitoring, (2) development of personalized adaptation algorithms that adjust feedback based on individual progress and limitations, (3) longitudinal studies assessing the system's impact on exercise adherence and weight loss outcomes, and (4) exploration of federated learning approaches to enable privacy-preserving model improvement across multiple deployment sites.

7. Conclusion

This research presents a novel AI-driven exercise monitoring system specifically designed to support non-surgical obesity treatment through automated movement analysis and real-time feedback. By introducing Spatio-Temporal Polychromatic Trajectory images and leveraging deep learning classification, the system achieves accurate exercise quality assessment using standard RGB cameras, addressing critical accessibility limitations of existing specialized systems.

Experimental validation across four shoulder exercises demonstrates classification performance exceeding 90% accuracy for three of four movement types, with comprehensive evaluation confirming robustness, computational efficiency, and user acceptability. The system's modular architecture facilitates extension to additional exercises and integration with complementary monitoring modalities, positioning it as a foundation for comprehensive digital obesity intervention platforms.

As obesity rates continue to rise globally, scalable technological solutions that enhance exercise efficacy and adherence will play increasingly important roles in public health strategies. The proposed system represents a significant step toward democratizing access to personalized exercise guidance, potentially transforming obesity management through intelligent, accessible, and effective movement optimization.

References

[1] N. Oshanova, G. B. Kamalova, N. Tekesbayeva, S. Turashova, G. Anuarbekova, and A. Bukanova. Assessing the efficacy of an artificial intelligence-driven real-time exercise monitoring system in

computer-supported collaborative learning. *Retos: nuevas tendencias en educación física, deporte y recreación*, (67):580–596, 2025.

[2] H. K. Chen, F. H. Chen, and S. F. Lin. An ai-based exercise prescription recommendation system. *Applied Sciences*, 11(6):2661, 2021.

[3] J. Huang, S. Kaewunruen, and J. Ning. Ai-based quantification of fitness activities using smartphones. *Sustainability*, 14(2):690, 2022.

[4] X. Han, X. Zhou, B. Tan, L. Jiao, and R. Zhang. Ai-based next-generation sensors for enhanced rehabilitation monitoring and analysis. *Measurement*, 223:113758, 2023.

[5] S. Juneja, S. Maiti, and B. S. Bhati. Iot and ai-based intelligent management of heart rate monitoring. In *IoT and AI-Enabled Healthcare Solutions for Intelligent Disease Prediction*, pages 198–226. CRC Press, 2025.

[6] D. Hu, L. Chen, H. Fang, Z. Fang, T. Li, and Y. Gao. Spatio-temporal trajectory similarity measures: A comprehensive survey and quantitative study. *IEEE Transactions on Knowledge and Data Engineering*, 36(5):2191–2212, 2023.

[7] K. Zheng, Y. Zhao, D. Lian, B. Zheng, G. Liu, and X. Zhou. Reference-based framework for spatio-temporal trajectory compression and query processing. *IEEE Transactions on Knowledge and Data Engineering*, 32(11):2227–2240, 2019.

[8] C. Yu, X. Ma, J. Ren, H. Zhao, and S. Yi. Spatio-temporal graph transformer networks for pedestrian trajectory prediction. In *European Conference on Computer Vision (ECCV)*, pages 507–523. Springer, 2020.

[9] C. Wang, S. Cai, and G. Tan. Graphtcn: Spatio-temporal interaction modeling for human trajectory prediction. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 3450–3459, 2021.

[10] C. Wang, L. Chen, S. Shang, C. S. Jensen, and P. Kalnis. Multi-scale detection of anomalous spatio-temporal trajectories in evolving trajectory datasets. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2980–2990, 2024.

[11] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou. A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*, 33(12): 6999–7019, 2021.

[12] N. Ketkar and J. Moolayil. Convolutional neural networks. In *Deep Learning with Python: Learn Best Practices of Deep Learning Models with PyTorch*, pages 197–242. Apress, Berkeley, CA, 2021.

[13] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, and T. Chen. Recent advances in convolutional neural networks. *Pattern Recognition*, 77:354–377, 2018.

[14] W. H. L. Pinaya, S. Vieira, R. Garcia-Dias, and A. Mechelli. Convolutional neural networks. In *Machine Learning*, pages 173–191. Academic Press, 2020.

[15] M. Krichen. Convolutional neural networks: A survey. *Computers*, 12(8):151, 2023.