GymPro: A Pose-Based Gym Workout Analysis and Feedback System

Lasana Sanketh Subasinghe
Department of Computer Science and Engineering
University of Moratuwa, Sri Lanka
lasana.21@cse.mrt.ac.lk

S. M. C. R. Siriwardhana
Department of Computer Science and Engineering
University of Moratuwa, Sri Lanka
chirantha.21@cse.mrt.ac.lk

Anjitha Siyambalagaspitiya
Department of Computer Science and Engineering
University of Moratuwa, Sri Lanka
anjitha.21@cse.mrt.ac.lk

***Keywords*—Pose Estimation, Deep Learning, Gym Workout Analysis, Computer Vision, Fitness Application.**

I. Introduction

Maintaining proper exercise form is crucial for injury prevention and workout effectiveness. However, many gym-goers lack access to personal trainers. GymPro addresses this issue by utilizing deep learning and pose estimation to analyze workout movements in real-time and provide instant corrective feedback.

II. Literiture Review

A. AI-based Fitness Applications

Commercial posture-detection systems employ a variety of sensing and computational approaches, including Naked Labs' 3D smart mirror[1], which combines a rotating scanner with proprietary posture-tracking algorithms; Microsoft Kinect[2], which uses depth-based skeletal tracking for motion analysis; Tempo Studio's AI-driven 3D camera system[3] for real-time exercise feedback; and DARI Motion's markerless depth-camera-based system[4] for biomechanical assessment. While these technologies demonstrate the feasibility of posture analysis in controlled or clinical settings with an extensive cost, their adaptability to general fitness environments remains constrained by challenges in real-time processing, dynamic movement occlusion, and exercise-specific modeling.

B. Pose Estimation Techniques

Recent advances in computer vision have led to the development of various landmark detection models for posture analysis, each employing distinct technical approaches. OpenPose[5] utilizes a multi-stage CNN architecture with Part Affinity Fields to associate body parts across multiple persons, enabling robust multi-person tracking but with significant computational overhead. MediaPipe Pose[6] implements an efficient heatmap-based CNN pipeline optimized for real-time performance on edge devices, offering both 2D and 3D landmark estimation capabilities. Among these, MoveNet[7] distinguishes itself as a particularly mobile-friendly solution; its single-shot CNN architecture and extreme quantization enable quick inference times on modern smartphones while maintaining competitive accuracy, making it uniquely suitable for real-time fitness applications. However, as noted by Zhang et al. (2021) and TensorFlow's benchmarking studies [8] these models still face fundamental limitations in occlusion handling and biomechanical precision compared to clinical-grade motion capture systems, particularly for complex, dynamic movements in uncontrolled environments.

III. Methodology

An abstraction of the methodology implemented to achieve this functionality can be shown in these steps.

1) The workout video is processed by a landmark detection model to extract body landmark coordinates.
2) Exercise-specific features are computed from these landmark points.
3) The features are fed into a posture classification model.
4) Feedback is generated based on the model's output.

A. Landmark Detection Model(Movenet)

MoveNet is a bottom-up pose estimation model that uses heatmaps to precisely detect human keypoints. Its architecture has two main parts: a feature extractor and multiple prediction heads. It's based on CenterNet's[9] prediction approach but includes key improvements that make it faster and more accurate. The feature extractor in MoveNet is MobileNetV2[11] with an attached feature pyramid network (FPN), which allows for a high resolution (output stride 4), semantically rich feature map output. Due to the requirement of fast, accurate predictions, MoveNet lightning has been selected as the landmark detection model of GymPro application.

B. Posture analysis model

To classify posture from predicted landmarks, fully connected neural networks were used. Separate models were trained for each workout, enabling the use of workout-specific features and improving accuracy by capturing unique movement patterns.

https://doi.org/10.31705/ADScAI.2025.54

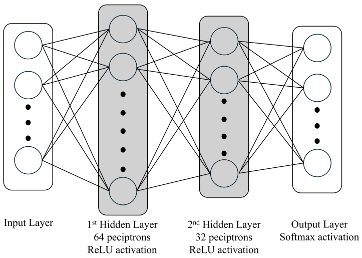

Fig. 1. Posture analysis neural network skeleton

TABLE I
LAYER-WISE NODE COUNT AND TOTAL PARAMETERS

Model	Input	Layer 1	Layer 2	Output	Params
Bicep Curl	8	64	32	4	2.5K
Lat Pull Down	8	64	32	4	2.7K
Plank	5	64	32	3	2.5K
Pushup	10	64	32	3	2.8K
Squatting	7	64	32	4	2.7K

IV. FEATURE SELECTION

The use of raw landmark coordinates from the MoveNet Lightning pose estimation model presents significant challenges for posture classification. These coordinates exhibit high sensitivity to subject orientation, camera distance, and frame positioning, introducing intersession and inter subject variability that compromises model generalization. Moreover, the heterogeneous movement dynamics across different exercises limits the effectiveness of a unified model trained on absolute coordinate values.

To address these limitations, we employ derived features based on inter-landmark angular relationships and normalized distances, which provide scale, translation, and rotation invariance. This feature engineering approach enhances robustness across varying capture conditions. Preliminary experiments suggest these relational features yield improved discriminative capability compared to raw coordinate-based approaches.

V. RESULTS AND EVALUATION

The costs of posture analysis models are calculated using categorical cross-entropy loss function, which is well-suited for multi-class classification tasks. This loss function measures the dissimilarity between predicted probabilities and ground truth labels, with the goal of minimizing it to improve posture classification accuracy.

The categorical cross-entropy loss $\mathcal{L}$ is defined as:

$$\mathcal{L} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \tag{1}$$

C = total classes, y_i = ground truth label for class i, $\hat{y}_i$ = predicted probability for class i. Training and testing accuracies and confusion matrices are shown in below figures(Table II, Fig 2).

VI. WEAKNESSES IN IMPLEMENTATION

- **Lack of Data Sources:** There is a scarcity of reliable, diverse datasets to help train the model, which affects its performance in real-world scenarios.

TABLE II
MODEL ACCURACY FOR DIFFERENT EXERCISE MODELS

Exercise	Train Accuracy	Validation Accuracy
Bicep Curl	0.9865	0.9802
Lat Pull Down	0.9423	0.9426
Plank	0.9891	0.9474
Pushup	0.9895	0.9901
Squatting	0.9501	0.9576

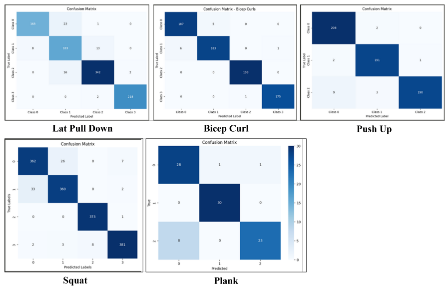

Fig. 2. Model-wise confusion matrices

VII. ENHANCEMENTS FOR FUTURE WORK

- **Expert Guidance:** Collaboration with fitness and healthcare experts is essential for building reliable, domain-specific applications.
- **User Validation:** Formal user studies can evaluate system effectiveness and guide improvements.
- **New Domains:** Extending the system to areas like rehabilitation or sports training can enhance its impact.

VIII. CONCLUSION

Current commercial systems show the promise of AI in posture analysis but are limited by cost and hardware dependency. This highlights the need for lightweight, real-time solutions. Our work addresses this through vision-based, mobile-friendly approaches for everyday fitness.

REFERENCES

[1] Intel RealSense, "Naked Home Body Scanner Based on Intel RealSense," 2018. [Online].

[2] A. Mentiplay *et al.*, "Evaluation of the Microsoft Kinect skeletal versus depth data in computing gait parameters," in *Proc. IEEE EMBC*, Milan, Italy, 2015, pp. 5897–5900.

[3] Analog Devices, "3D Time of Flight Technology Enables Tempo At-Home Fitness Platform," [Online].

[4] DARI Motion, "Proprietary Technology for Human Motion Analysis," [Online].

[5] Z. Cao *et al.*, "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 172–186, Jan. 2021.

[6] Google AI Edge, "Pose Landmark Detection Guide," [Online]

[7] K. LeViet and Y. Chen, "Pose estimation and classification on edge devices with MoveNet and TensorFlow Lite," *TensorFlow Blog*, Aug. 2021.

[8] Z. Zhang *et al.*, "Fusing wearable IMUs with multi-view images for human pose estimation: A geometric approach," in *Proc. CVPR*, 2020, pp. 2197–2206.

[9] K. Duan *et al.*, "CenterNet: Keypoint triplets for object detection," in *Proc. ICCV*, 2019, pp. 6568–6577.

[10] M. Sandler *et al.*, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. CVPR*, 2018, pp. 4510–4520.