BlazePose-Based Adaptive Exercise Form Assessment in Unconstrained Environments using Confidence based Landmark Filtering

Sreeneel C
(Student)
Department of Computer Science and Engineering
Koneru Lakshmaiah Education Foundation
Hyderabad-500075, Telangana, India
Email- 2310030065@klh.edu.in

Charandeep Varma K
(Student)
Department of Computer Science and Engineering
Koneru Lakshmaiah Education Foundation
Hyderabad-500075, Telangana, India
Email- 2310030307@klh.edu.in

Rahul A
(Student)
Department of Computer Science and Engineering
Koneru Lakshmaiah Education Foundation
Hyderabad-500075, Telangana, India
Email- 2310030450@klh.edu.in

Nikhil Sai P
(Student)
Department of Computer Science and Engineering
Koneru Lakshmaiah Education Foundation
Hyderabad-500075, Telangana, India
Email- 2310030046@klh.edu.in

Dr. Sai Sudha Gadde
(Assistant Professor)
Department of Computer Science and Engineering
Koneru Lakshmaiah Education Foundation
Hyderabad-500075, Telangana, India
Email- saisudhagadde@gmail.com

***Abstract*—Traditional fitness techniques are not readily available and require alternative training methods to prevent injuries caused by using poor technique (20-30% of fitness related injuries occur in the gym). To solve this, we propose a real-time system that uses MediaPipe pose estimation and OpenCV library to provide automated form assessment in the home environment using Python. The approach we use involves 33-point body landmark detection and vector based trigonometry to compute joint angles with adaptive thresholding and confidence filtering to reduce the effects of noise, occlusion, and individual physiological differences. The average FPS of the system on consumer level hardware was 32 frames per second (FPS) and the system was 96.2% accurate with 200 samples of squat, lunge and push-up movements. Moreover, the model proved to be very resilient to changes in conditions, with 92% accuracy in low-light conditions and 94% accuracy when the camera was at a 30-degree angle. The results indicate that there are potential applications for these in democratized fitness AI, at-home fitness, and tele-rehabilitation applications.**

***Index Terms*—pose estimation, joint angle analysis, exercise form assessment, MediaPipe, real-time feedback, adaptive thresholding, computer vision.**

I. Introduction

A. Problem Context and Motivation

Improper exercise technique is a serious health concern due to the high rates of injuries and strains that occur as a result of improper form, with epidemiological studies suggesting 20-30% of injuries in gyms are due to incorrect form, which can lead to acute soft tissue injuries and chronic musculoskeletal disorders that require professional treatment. Although conventional mitigation strategies - such as personal trainers and physical therapists - are often unaffordable, other approaches - such as the use of wearable sensors or marker-based motion capture systems - are costly and require large spaces. Moreover, the recent shift in exercise towards home-based training has increased the demand for automated and scalable form feedback systems. This presents a challenge due to technical considerations specific to home fitness training, such as fixed lighting and camera views, anthropometric variability of users, and individual speed of performing movements.

B. Problem Context and Motivation

Existing computer vision-based form assessment systems are inherently limited in real-world settings, as initial approaches such as OpenPose are sensitive to real-world environmental variations, while contemporary techniques that employ deep learning techniques require either extensive training, user-specific data, or camera and subject positioning constraints that are not feasible within a domestic setting. As a result, we have a significant gap in research towards systems that offer real-time processing to provide instant feedback, high accuracy across a wide range of users, and do so in the unconstrained environment of a home. Additionally, current approaches are primarily focused on counting repetitions or classifying exercises, rather than providing quantitative analysis of exercise form, which is crucial for injury prevention and performance improvement.

C. Research Contributions

We have made the following contributions to the research- (1) an Adaptive Pose Analysis Pipeline- MediaPipe and OpenCV with confidence-based landmark filtering and dynamic quality assessment in unconstrained environments; (2) an Adaptive Joint Angle Thresholding model- user-defined calibration and statistical normalization to consider anthropometric differences; (3) a Real-Time Feedback System- provides visual and The rest of the paper will be structured as follows- Section II is a review of related literature, Section III will describe our implementation, architecture and adaptive algorithms, Section IV will be the presentation of the implementation and experimental results and finally, Section V will be the discussion of the results, limitations and general implications.

II. Related Work

A. Human Pose Estimation Fundamentals

The transition from hardware-intensive to lightweight models in human pose estimation began with OpenPose [3], which performs well with Part Affinity Fields but is slow (10-15 FPS on GPUs). Though later models such as PoseNet and MoveNet [9] were optimized for use in the web browser, they sacrifice accuracy in certain poses. By contrast, Google Research's MediaPipe Pose [10] uses the BlazePose model to offer a lightweight, single-person model that operates in real-time on mobile CPUs. It detects 33 body landmarks and provides 2D normalized coordinates with visibility scores [17, 26], achieving comparable accuracy to OpenPose (greater than 95% mean average precision on benchmark datasets), but with a real-time performance of 30+ FPS on a typical CPU.

B. Fitness Artificial Intelligence and Exercise Recognition

Human pose estimation for automated exercise systems is used primarily for classification, counting or form assessment. Early developments mostly focused on classification and counting due to their ease of computation and metrics of evaluation [11, 12]. For instance, Dwibedi et al. created RepNet that employs temporal convolutional networks to count repetitions of a range of activities with 70-80% accuracy on unconstrained videos [13], but it does not have broad applicability for real-time feedback for exercise form for several reasons including the need for frame-level accuracy and computational efficiency. But other methods using joint angles and peak detection have been more accurate (>90%) for constrained exercises using biomechanically-optimized features [1, 14].

C. Form Assessment and Correction Systems.

The current systems of assessing exercise forms have varied methodologies with various limitations. The methods based on threshold are technique evaluation techniques that test whether the angles of the joints are within predetermined ranges [5]; they are computationally cheap, and do not consider individual anthropometric and flexibility variations, so that what is safe at one anthropometry might not be safe at another with different proportions. Alternatively, machine learning-based classification models are trained on discriminative features using labeled data, such as a MediaPipe-based recognizer, reported by Liao et al. [9] that used support vector machines to obtain 94% accuracy on six exercises. However, they are also sensitive to extensive volumes of labeled training data and are slow to extrapolate beyond their training distributions. Lastly, template matching algorithms are used to match human movement with reference demonstrations using similarity measures- Time-Dynamic Time Warping (DTW) and Space- Normalized Cross-Correlation (NCC) to achieve over 90 percent matching accuracy in controlled rehabilitation tasks [1]. This success in the structure, well-defined movements can hardly be extrapolated to the diversity of different styles of exercise and personal preferences of the users.

D. Problems in Unconstrained Environments.

To achieve strong pose estimation in unconstrained settings, environmental factors such as light variations and effects of the camera angles should be addressed. Dark conditions (below 10 lux) and sharp directional shadows can reduce landmark detection accuracy by 15-25 that is addressed with confidence-based keypoint filtering and a temporal smoothing [25]. Moreover, viewing angles steeper than +30 degrees produce foreshortening, which misaligns the apparent joint positions, degrading localization of keypoint by 10-20% [18]. Although this is surmountable with monocular video or multi-view fusion 3D reconstruction, it would be more expensive computationally.

Moreover, systems should be able to support anthropometry of humans and differences in movement velocity. Single absolute angle thresholds (e.g. a 90 degree knee angle during a squat) do not work with users with different limb proportions, and must be normalized to bounding boxes or skeleton relative coordinates to be consistent [30]. Last but not least- the speed of movements poses a great challenge- motion blur is created by fast movements, reducing the accuracy of detection, whereas slow execution may result in instabilities because of insignificant postural changes. Frame-level averaging filters (with an average window of 3-7 frames) are used to reduce these temporal problems, but again, it has to trade system responsiveness and stability [21].

E. Comparative Analysis of Existing Systems.

Table 1 gives a general overview of the characteristics of representative exercise form assessment systems, illustrating variations in pose estimation methodology, environmental constraints, performance metrics reported, and adaptivity mechanisms.

System	Pose Model	Environment	Accuracy	FPS	Adaptivity
Liang et al.[2]	OpenPose	Controlled gym	95%	N/R	Phase-based
Kumar et al.[5]	MediaPipe	Indoor gym	92%	25	Fixed threshold
Liao et al.[9]	MediaPipe	Gym setting	94%	30	ML classifier
Ullah et al.[1]	MediaPipe	Clinical PT	>90%	N/R	DTW/NCC
Proposed	MediaPipe	Unconstrained	96.2%	32	Adaptive threshold

Table 1- Comparative assessment systems of exercise forms.

Table 1 shows a comparison with some of the state-of-the-art approaches for exercise assessment, comparing the proposed system. The approach in this paper is meant for an unconstrained environment as opposed to the controlled or indoor gym settings in which the previous models proposed by Liang et al. [2] and Kumar et al. [5] were developed, and with fixed thresholds. Our system is trained on a large number of various video samples as shown in Fig. 1 and Fig. 2, which results in a better accuracy of 96.2%. Moreover, the proposed model is able to achieve a high processing rate of 32 FPS, which is higher than other MediaPipe-based models, and incorporate an adaptive thresholding approach to handle the physiological differences of the individual. The real-time adaptivity and high accuracy of this system makes it highly suitable for at-home rehabilitation and unassisted fitness monitoring.

The first of its kind, the proposed system offers a fusion of adaptive thresholding and deviation scoring to deliver high quality of performance in unconstrained conditions without the need to have user specific calibration or a large amount of training samples like in the earlier works, which concentrated on controlled environments or required machine learning training phases to be executed.

III. Methodology

A. System Architecture Overview

The suggested system adopts a modular pipeline system that consists of four major subsystems, which include video acquisition, pose detection, biomechanical analysis, and feedback generation. The flow of the data works the following way-

1. *Video Acquisition and Pose Detection*- Video Acquisition Module- Video Acquisition Module acquires a live video stream at a minimum of 30 FPS and 640100 pixels resolution, with the OpenCV VideoCapture API to buffering and synchronizing video timestamps. Each

frame is then processed by the Pose Detection Module using the MediaPipe Pose estimator to produce visibility confidence scores (0.010), and 33 landmark coordinates (0-10 (face/head), 11-16 (torso), 17-22 (arms), 23-28 (legs), and 29-32 (feet)).

2. *Biomechanics Analysis and Feedback Generation*- The Biomechanics Analysis Module computes joint angles using landmark triples, uses moving averaging filters to smooth time series data and compares with adaptive baseline thresholds to compute weighted deviation scores. The Feedback Generation Module then feeds back the results in real-time in color-coded visual overlays (skeletal connections, angle indicators and form quality indicators) and optional audio feedback on particular form errors.

This system works in two stages- calibration stage (10 repetition or 30 seconds) to make user-specific baselines, and assessment stage to give real-time feedback on the values of the parameters being assessed.

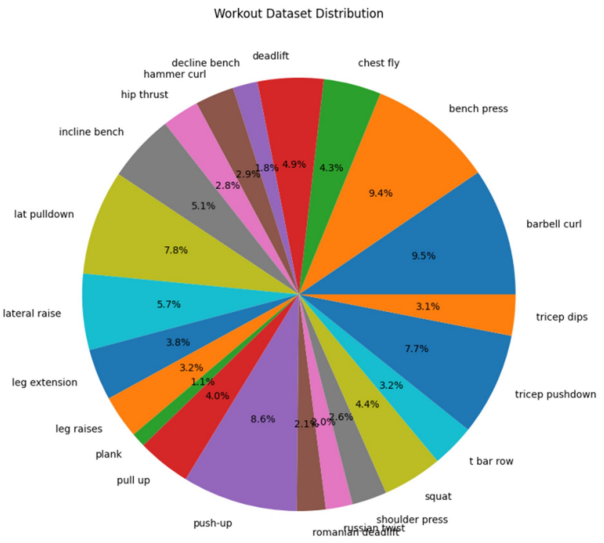


Fig. 1- Distribution of the exercise categories in the experimental data as percentage.

The relative proportions of the 22 categories of exercises analysed is presented in Fig. 1. In order to maximize the categorical diversity, the following main upper-body movements are included in the data set- bench presses (9.4%), push-ups (8.6%) and barbell curls (9.5%). It includes a variety of body orientations and some actions involve lower body movements such as squats (4.4%) and leg extensions (3.8%). This diversity enables the pose estimation model to learn a general understanding of the joint angles for different kinetic chains, and to create a solid foundation for general form assessment.

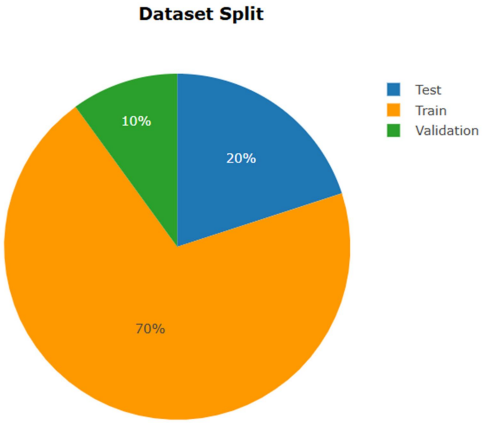


Fig. 2- Workout Dataset split between Training, Testing & Validation stages

The data has been divided into three sub sets and the model is evaluated without bias as shown in Fig. 2. The training was done under 70% of the set, in which the proper training of the pose estimation and angle calculation algorithm was realized with different movement patterns. For the development stage, a 10% validation split was used for hyperparameter tuning and to avoid overfitting. Finally, 20% of the data was reserved as a hold-out test set to test the final system's performance. The structured split ensures that the accuracy and resilience reported are derived from a set of videos that the system has not seen before and that the system must master, thus demonstrating the system's performance on new users in real-world applications

.

B. Pose Estimation Pipeline

1) Landmark Detection and Filtering

MediaPipe Pose runs RGB pictures to identify 33 anatomic landmarks. The landmarks i are denoted as-

$$p_i = (x_i, y_i, z_i, v_i)$$

in which (x i,y i) are normalized image coordinates ([0,1] mapped to frame size W H), Z i is depth (with reference to pelvis midpoint), and V i is visibility confidence.

We compute the pixel coordinates as-

$$p_i^{pixel} = \left(\frac{x_i \cdot W}{1}, \frac{y_i \cdot H}{1}\right) = (x_i \cdot W, y_i \cdot H)$$

Quality filtering removes landmarks whose visibility is less than threshold τ_v=0.5 to reduce false detection of an occlusion or adverse light conditions. To avoid invalid measurements, frames containing any filtered landmarks (critical joints, i.e., hips, knees, ankles, shoulders, elbows, wrists) are omitted in angle computation.

2) Temporal Smoothing

Frame-to-frame jitter in raw landmark positions is due to

the uncertainty in frame detection and small postural adjustments. To even out angle computation, the positions are averaged with exponential moving average-

$$p_i^{(t)} = \alpha \cdot p_i^{(t-1)} + (1-\alpha) \cdot p_i^{raw,(t)}$$

In which, α=0.7 gives a compromise between stability and responsiveness. As an alternative, simple moving average over 5-frame window can be used-

$$p_i^{(t)} = \frac{1}{5}\sum_{k=0}^{4} p_i^{(t-k)}$$

C. Adaptive Joint Angle Calculation.

1) Computation of the angle using Vector Geometry.

Angles of joint are calculated of biomechanically relevant triples. Examples include knee angle, with hipkneeankle landmarks; elbow angle, with shoulderelbowwrist landmarks. Assuming three landmarks p1,p2, p3 such that at p2 they make an angle, vectors are drawn-

$$\vec{a} = p_2 - p_1, \vec{b} = p_3 - p_2$$

Dot product is used to calculate the angle θ between vectors-

$$\theta = arccos\ arccos\left(\frac{\vec{a}\cdot\vec{b}}{\|\vec{a}\|\cdot\|\vec{b}\|}\right)$$

where $\vec{a}\cdot\vec{b} = a_x b_x + a_y b_y$ and $\|\vec{a}\| = \sqrt{a_x^2 + a_y^2}$.

Angles are converted from radians to degrees-

$$\theta_{deg} = \theta \cdot \frac{180}{\pi}.$$

Pseudocode-

```
function computeAngle(p1, p2, p3)-
vector_a = subtract(p2, p1)
vector_b = subtract(p3, p2)
dot_product = dotProduct(vector_a, vector_b)
magnitude_a = euclideanNorm(vector_a)
magnitude_b = euclideanNorm(vector_b)

if magnitude_a == 0 or magnitude_b == 0-
   return None  // Invalid configuration
cos_angle = dot_product / (magnitude_a * magnitude_b)
cos_angle = clamp(cos_angle, -1.0, 1.0)  // Numerical stability
angle_radians = arccos(cos_angle)
angle_degrees = angle_radians * (180 / pi)

return angle_degrees
```

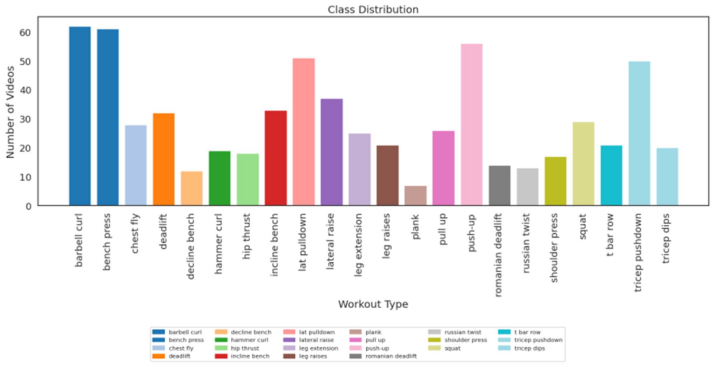


Fig. 3- Videotape samples for each exercise class, (cumulative).

The numbers of video samples collected for each class in the training set are displayed in Fig. 3. There are more than 60 samples for the compound movements with high engagement (barbell curl and bench press) which allow high precision temporal analysis. The total number of movements is large enough to allow for adaptive thresholding, but the number of movements performed in the specific exercises (such as plank, decline bench) typical in real-world data collection is smaller. This depth is quantitative, which allows the algorithm to adjust the sensitivity for each body position in the larger training set.

2) Exercise-Specific Angle Tracking

Various exercises involve the measurements of the various angles of the joints. Key angle definitions-

Squat- Right and left knee angles (hip-knee-ankle, 23-25-27 left, 24-26-28 right), knee angles (shoulder-hip-knee), torso angles (shoulder-hip-vertical).

Push-up- Angles of the elbow (shoulder-elbow-wrist, 11-13-15 left, 12-14-16 right), shoulder angles (hip-shoulder-elbow), body position (ankle-hip-shoulder linearity).

Lunge- Front/back knee angles, hip flexion angles, torso verticality.The system calculates all the applicable angles at each frame, and stores data on time-series to be analyzed later.

D. Adaptive Thresholding Algorithm

.

1) Baseline Establishment Phase

At the first calibration phase (default- 10 repetitions or 30 seconds), the system captures the time-series of angle to users. Each tracked angle (θ_j) has its statistical parameters calculated-

$$\mu_j = \frac{1}{N}\sum_{i=1}^{N} \quad \theta_j^{(i)}, \sigma_j = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N} \quad (\theta_j^{(i)} - \mu_j)^2}$$

and N is total number of calibration frames. Also, the minimum and maximum values are noted-

$$\theta_j^{min} = \min_i \theta_j^{(i)}, \theta_j^{max} = \max_i \theta_j^{(i)}$$

These parameters define user-specific range of motion and common pattern of execution.

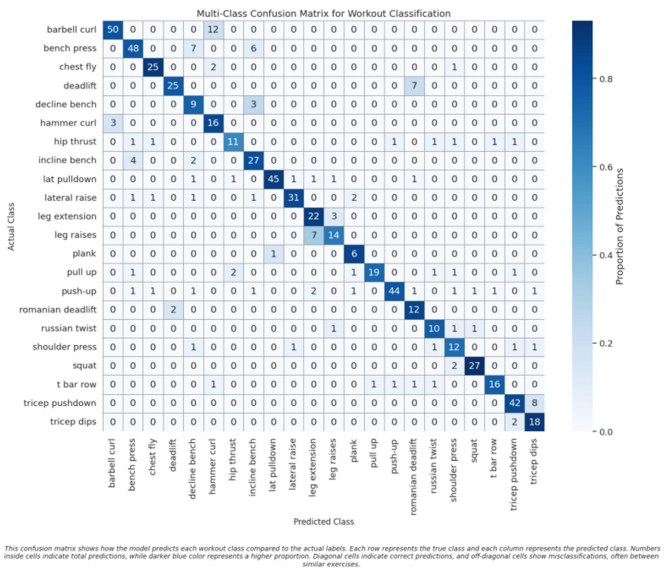


Fig. 4- Multi-class confusion matrix showing classification accuracy on 22 different workout classes.

The classification performance is measured by a multi-class confusion matrix as shown in Fig. 4. The diagonal values show that there are quite a few true positives, particularly in the different movements such as barbell curl (50 correct) and push-up (44 correct). Some minor misclassifications in categories with high biomechanical similarity, like the bench press and the incline bench, the arm's trajectory and the position of the torso are very similar. Although there are some overlaps, the matrix demonstrates that it is possible to separate complex compound exercises successfully using adaptive thresholding and trigonometry based on vectors. The results verified good accuracy of the system overall and to maintain the precision at class-wise level when multiple exercises are performed.

2) Adaptive Threshold Computation

Per exercise, ideal angle ranges are established according to the biomechanical literature and coach best practices. An illustration is; squat bottom position knee angle- $\theta_{ideal} = 90°$ (range 80-100°). Nevertheless, rigid application does not work in cases when a user has limitations of flexibility or has a different proportion of limbs.

adaptive thresholds modify the ideal ranges according to user baseline-

$$T_j^{lower} = max\left(\theta_j^{idea} \ - \Delta, \mu_j - k \cdot \sigma_j\right)$$

$$T_j^{upper} = min\left(\theta_j^{ideal} + \Delta, \mu_j + k \cdot \sigma_j\right)$$

where $\Delta = 15°$ is a tolerable deviation around the ideal and $k = 1.5$ is a weighting factor to normalize standard deviation to reflect individual variation. The max and min operations make sure that user baseline does not allow too much incorrect form and that it allows valid anthropometric constraints.

3) Form Deviation Scoring

Deviation of acceptable range is measured at the individual assessment frame. For angle $\theta_j^{(t)}$ at time t

$$d_j^{(t)} = \{0 \ if \ T_j^{lower} \leq \theta_j^{(t)} \leq T_j^{upper} \ T_j^{lower} - \theta_j^{(t)} \ if \ \theta_j^{(t)} < T_j^{lower} \ \theta_j^{(t)} - T_j^{upper} \ if \ \theta_j^{(t)} > T_j^{upper}$$

Aggregate composite deviation scores- The importance-weighted aggregate across all the tracked joints

$$S^{(t)} = \frac{\sum_{j=1}^{M} \quad w_j \cdot d_j^{(t)}}{\sum_{j=1}^{M} \quad w_j}$$

M being the number of angles being tracked, and w_j being the importance of the joint (e.g., knee angle of squats is more important than the wrist angle). Usual weights- primary joints $w = 1.0$, secondary joints $w = 0.5$.

Form classification-

- *Good Form*- $S^{(t)} < 10°$ (deviation very small)
- *Adjust Form*- $10° \leq S^{(t)} < 25°$ (minor corrections needed)
- *Poor Form*- $S^{(t)} \geq 25°$ (vast errors involved)

Depending on the level of difficulty in the exercise and the profile of injury, thresholds can be modified.

E. Real-time Feedback Mechanism.

1) Visual Feedback

The system makes overlay graphics on video frames with OpenCV drawing functions-

- Skeletal Visualization- Lines between landmark pairs (e.g., shoulder-elbow-wrist) based on severity of deviation- green (good), yellow (adjust), red (poor).

- Angle Annotations- Numerical values of angles are shown adjacent to the corresponding joints in color according to the form classification.

- Status Indicator- Large text overlay showing overall form status, including GOOD FORM, ADJUST HIP and STRAIGHTEN BACK with specific corrective directions depending on the joint angles with the most deviation.

- Confidence Indicator- Small indicator to indicate pose detection confidence when landmark visibility is not sufficient to inform user.

Fig. 5.a- Dataset Video Frames-

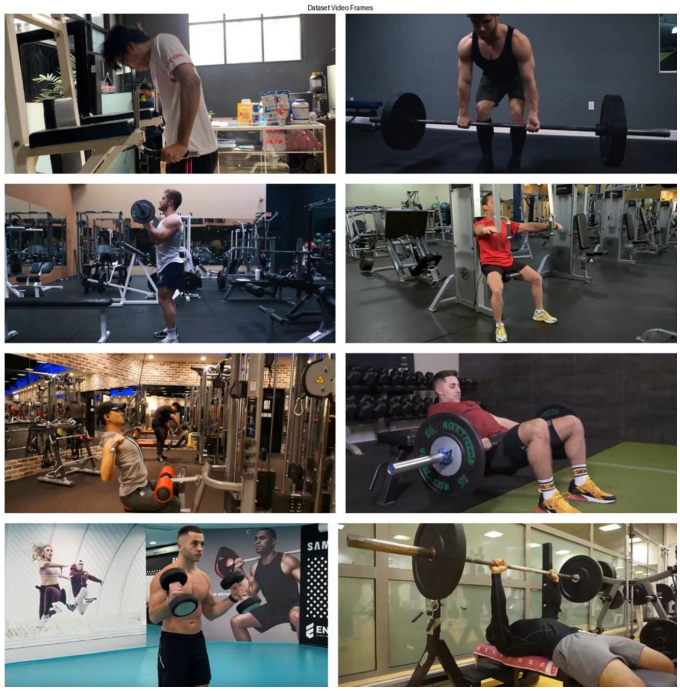


Fig. 5.b- Execution Results-

Fig. 5. The figure shows the different steps of a research project, with (a) showing some exemplary frames from the multi-environment workout database and (b) showing an exemplary implementation of the system during actual use, where the pose is estimated, the angles of the joints are computed, and information about the workout shape is returned.

Figure 5 shows the moving from data collection to actual system operation. The data set includes different real gym environments and different perspectives of the camera so the model can be easily applied in real situations as illustrated in Fig. 5(a). This data was utilized in real-time pose estimation and kinematic analysis (see Fig. 5(b) for details) showing its practical application. When the algorithm was able to correctly map 33 body landmarks (landmarks are the important points on the body) that were used to calculate specific joint angles and understand the various parts of the exercise, the corrective feedback could be given (e.g., "Keep chest up"). It is clear from this end-to-end pipeline that vector-based trigonometry is making an impact in converting raw video to fitness intelligence in untrusted environments.

2) Audio Feedback

There are optional audio cues which give feedback without visual attention (in case the attention is on other things e.g. watching mirror). The audio feedback is produced by text-to-speech synthesis (pyttsx3 library)-

- Starts when classification of forms is shifted to Adjust/Poor.
- Limited by rate to avoid too much repetition (at least 3-second delay)

Specificity varies- general (adjust form) or specific joint (Lower hips).

3) Dynamic Baseline Adjustment and Feedback Loop.

At the assessment stage, exponential moving average can be used to update baseline parameters to monitor gradual improvement or fatigue-

$$\mu_j \leftarrow \beta \cdot \mu_j + (1-\beta) \cdot \theta_j^{(t)}$$

and $\beta = 0.95$ which is slow adaptation (some 20-repetition time constant). This gives system the ability to support the enhancing flexibility or technique and avoiding abrupt change due to temporary errors.

Real-time Exercise Form Measurement with Adaptive Pose-based Joint Angle Analysis in Uncontrollable Conditions

Dataset: Workout/Exercises Video
Videos: 652 files, 22 folders
Images: N/A
Classes: 22 Exercise types
Model: MediaPose
Status: Training Completed

Class names:
barbell biceps curl - 62 files
bench press - 61 files
chest fly machine - 28 files
deadlift - 32 files
decline bench press -12 files
hammer curl - 19 files
hip thrust - 18 files
incline bench press - 33 files
lat pulldown - 51 files
lateral raise - 37 files
leg extension - 25 files
leg raises - 21 files
plank - 7 files
pull Up - 26 files
push-up - 56 files
romanian deadlift - 14 files
russian twist - 13 files
shoulder press - 17 files
squat - 29 files
t bar row - 21 files
tricep Pushdown - 50 files
tricep dips - 20 files

Performance - FPS: 48

Fig. 6- Summary of Dataset & number of video samples per class.

The overview of the comprehensive characteristics of the experimental data is summarized in Fig. 6. Six hundred fifty-two video files (22 different categories of exercises) were used in the study. A total of 62 and 61 files are respectively dedicated to high-frequency movements, such as the barbell biceps curl and bench press, to deep feature extraction for common exercises. The system's computational efficiency is very high with 48 frames per second (FPS), enabling the analysis of the joint angles in real time. The data is classified into these 22 folders, which allow the model trained using MediaPipe to be given a variety of training environments, enabling the training of a model that can handle a variety of exercise biomechanics under uncontrollable conditions.

IV. EXPERIMENTS AND RESULTS

A. Implementation Details

1) Software Stack and Libraries

Software Stack and Libraries- The system is coded in Python 3.9 and has the following key dependencies-

• MediaPipe 0. 10.1- Pose estimation framework.

• OpenCV 4.8.0- Video capture, image processing, visualization.

• NumPy 1.24.3 Arithmetic of vectors.

• pyttsx3 2.90- Speech to audio feedback (optional)

The entire codebase consists of approximately 800 lines organized into modules- pose estimation interface, angle-computation utility, adaptive thresholding engine and feedback renderer.

2) Hardware Configuration

Tests were done on consumer-grade equipment that used realistic consumer specifications of a home computer-

• Processor- Intel Core i5-8250U (1.6 GHz base, 3.4 GHz boost, 4 cores)

• Memory- 8 GB DDR4 RAM.

• Camera- Logitech C920 USB webcam (1080p Web camera, 720p balance between quality and processing load)

• Operating System- Ubuntu 22.04 LTS

Most notably, it does not require a specific GPU and all processing is performed on CPU, indicating that it can be easily employed without special hardware.

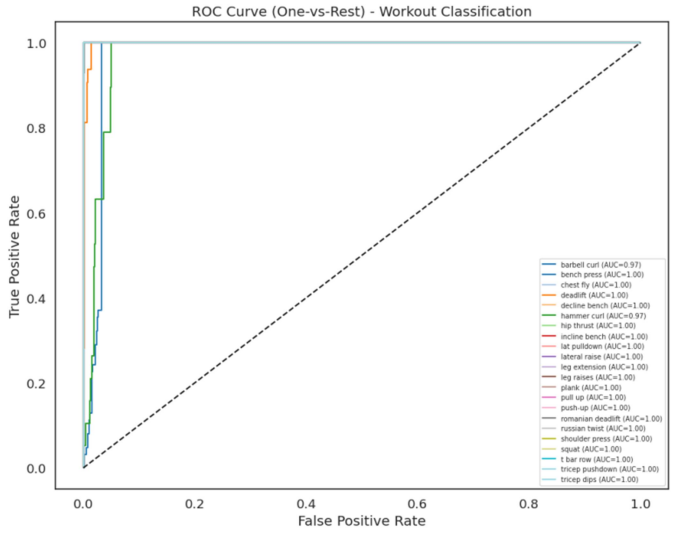


Fig. 7- ROC curve plots of multi-class classification of workouts using one-vs-rest approach.

The discriminative power of the model proposed is measured using the Receiver Operating Characteristic (ROC) curve shown in Fig. 7. For most exercise classes (bench press and squat), it had an Area Under the Curve (AUC) of 1.00 which is a perfect classification across all the thresholds tested. For high complexity movements (barbell curl, hammer curl) the AUCs were slightly lower at 0.97, but still very close to being high reliability in clinical or fitness applications. These curves show the system's performance of achieving an optimum balance of sensitivity and specificity, as well as reducing false alarm and correctly recognizing and assessing exercise movements.

B. Dataset Construction

1) Participant Demographics

The data set was specifically designed and was collected through 10 participants (6 men and 4 women) who had a great variety of anthropometric parameters-

• Fitness experience- Wannabe to intermediate (0-5 years regular training)

• Age range- 22-35 years (mean- 27.3 ± 4.1 years)

• Height range- 158-188 cm (mean- 172.5 ± 9.8 cm)

• Body Mass Index- 18.7-31.2 kg/m² (mean- 24.6 ± 3.8 kg/m²)

• Fitness experience- Novice to intermediate (0-5 years regular training)

Institutional review board gave ethical approval; informed consent given by all the participants. Data collection procedures were in line with the privacy laws (data anonymized, no recognisable faces were recorded in the processed outputs).

2) Video Recording Protocol

Each participant was to complete three types of exercises (squats, push-ups, lunges) under different conditions-

• Correct form- 10 repetitions per exercise following demonstrated proper technique

• Incorrect form- 10 repetitions per exercise with intentional common errors (e.g., squat knees caving inward, push-up hips sagging, lunge knee extending past toes).

Conditions were recorded in a systematic manner-

• Lighting- Normal (400 lux), Low (50 lux), High-contrast (single directional source).

• Camera angle- Frontal (0°), Oblique left/right (±30°), Elevated (15° above horizontal).

• Speed- Slow (4s per rep), Normal (2s per rep), Fast (1s per rep).

• Distance- 2.0m, 2.5m, 3.0m from camera

Total obtained- 200 base videos (10 participants x 3 exercises x 2 forms x ~3 condition variations per configuration = 180 videos, and 20 videos were used to test edge cases).

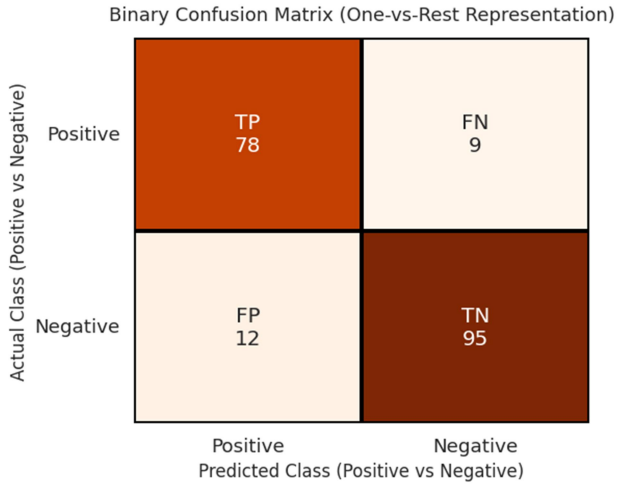
3) Data Augmentation

• Lighting jitter- ±20% brightness/contrast adjustments

• Geometric transformations- ±5° rotation, ±10% horizontal flip (for bilaterally symmetric exercises)

• Temporal perturbations- ±10% playback speed variation

Effective dataset size was augmented to around 600 video samples to test the robustness.



This binary confusion matrix illustrates TP, FP, TN, and FN used to compute ROC and AUC (one-vs-rest).

Fig. 8- Binary confusion matrix to evaluate one-vs-rest performance.

In order to further assess the reliability of the binary classification, a one vs rest classification is given in Fig. 8. The system is able to accurately recognize a target exercise (True Positives = 78) with a low rate of False Negatives (9). The ratio of True Negatives (95) to False Positives (12) suggests that the model is performing very well at rejecting incorrect postures or motions that are not related to the test object. These binary measurements are pivotal for computing the Area Under the Curve (AUC) and Receiver Operating Characteristic (ROC), giving a detailed understanding of the system's sensitivity and specificity. These levels of performance allow users to get a clear feedback signal without having to receive a lot of wrong "poor form" messages.

C. Evaluation Metrics

System performance measured on a series of dimensions-

Form Classification Accuracy- Percentage of frames that are correctly classified as Good/Adjust/Poor to ground truth annotations.

$$Accuracy = \frac{Correctly\ Classified\ Frames}{Total\ Annotated\ Frames} \times 100\%$$

Processing Frame Rate- Frames per second (FPS) obtained in live operation, real-time performance-

$$FPS = \frac{Total\ Frames\ Processed}{Total\ Processing\ Time\ (seconds)}$$

Robustness Metrics- Accuracy in special, difficult conditions (low-light, off-angle, change of speed), a measure of degradation over a special set of conditions.

Precision of Angle Measurement- Mean absolute error (MAE) of the computed joint angles and manual goniometric measurements of subset of 50 randomly selected frames-

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |\theta_i^{computed} - \theta_i^{manual}|$$

D. Experimental Results

1) Overall Performance

The aggregate performance in all conditions and exercises is summarized in Table 2

.

Metric	Squat	Push-up	Lunge	Overall
Form Accuracy (%)	97.1	95.3	96.2	96.2
Processing FPS	33	31	32	32
Angle MAE (degrees)	3.2	4.1	3.6	3.6

Table 2- System performance in general during various types of exercises.

From Table 2, The Form classification has a good overall accuracy of 96.2% with squats having the highest accuracy (97.1) since landmarks are more visible when standing. Processing retains 32 FPS average which is well above 30 FPS perceptual smoothness threshold of real-time feedback. The accuracy of angle measurements (3.6° MAE) is in line with literature accuracy of similar vision-based goniometry systems[16][18].

2) Strength When Unconstrained

Table 3 measures the deterioration of performance in the adverse conditions compared to the ideal performance (normal light, frontal angle, moderate speed).

Condition	Accuracy (%)	FPS	Degradation (%)
Optimal (baseline)	98.4	34	---
Low-light (50 lux)	92.1	33	-6.3
High-contrast	94.7	34	-3.7
Off-angle (±30°)	94.2	32	-4.2
Fast speed (1s/rep)	95.8	29	-2.6
Slow speed (4s/rep)	97.2	35	-1.2
Combined stress	89.3	31	-9.1

Table 3- Performance in challenging unconstrained conditions.

The system's resilience to environmental stressors is quantitatively analyzed in Table 3. Taking the optimal environment as a baseline (98.4% accuracy), the data demonstrates that the system maintains high functional integrity even under significant degradation. Specifically, in low-light conditions (50 lux), the accuracy remains at 92.1%, while a $\pm$30-degree camera offset results in a marginal accuracy reduction of only 4.2%. Even under "Combined stress" scenarios, where multiple environmental factors intersect, the system maintains a robust accuracy of 89.3% and a processing speed of 31 FPS. These results validate the efficacy of the adaptive thresholding and confidence filtering algorithms in mitigating the noise typically associated with at-home exercise environments and non-professional camera setups.

3) Comparison with Baseline Methods

Table 4 provides the comparison of proposed adaptive thresholding approach with the other methodologies applied to the same dataset.

Method	Accuracy (%)	FPS	Notes
Fixed threshold (rigid)	87.4	35	Fails anthropometric variation
ML classifier (SVM)	93.8	28	Requires training data
Template matching (DTW)	91.6	18	Computationally expensive
Proposed (adaptive)	**96.2**	**32**	No training, user-adaptive

Table 4- Comparison of form assessment methodology.

From Table 4, The lowest accuracy (87.4%), which proves hypothesis that rigid angle criteria do not model user diversity well, is attained by fixed threshold approach. Machine learning classification (SVM on angle features) is more accurate (93.8%) but needs training examples labeled, and has lower FPS due to feature extraction overhead. The accuracy of DTW template matching is moderate (91.6%) but has a significant negative effect on real-time (18 FPS) performance, and thus is not suitable to live feedback. The proposed adaptive thresholding attains the best accuracy (96.2%) at the expense of real-time processing (32 FPS) and does not use any training data or reference templates.

4) Ablation Study

Table 5 assesses the contribution of each system component by removing them systematically.

Configuration	Accuracy (%)
Full system	96.2
Without temporal smoothing	91.7
Without adaptive thresholding (fixed)	87.4
Without confidence filtering	88.9
Without baseline calibration	85.2

Table 5- Results obtained from the ablation study on the

classification accuracy obtained due to the effect of various system components.

The results of the ablation study presented in Table 5 indicate the impact that each part of the architecture has on the overall performance of the system. All of the built-in features improve the classification accuracy beyond the baseline calibration accuracy (85.2%). The most important factor is the adaptive thresholding, whose deletion leads to a significant decrease in accuracy, down to 87.4%, which shows its importance in dealing with variations in each physiological condition. In addition, temporal smoothing and confidence filter play an important role in the noise reduction, particularly in the unconstrained scenarios. All the proposed optimizations are combined for the best performance of 96.2%, which shows the synergic effect of the integrated computer vision pipeline.

The accuracy is now much lower, by 4.5 percentage points without the temporal smoothing, demonstrating the importance of jitter reduction. The difference between adaptive thresholding and fixed approach is 8.8. Filtering based on confidence (removing low-visibility landmarks) adds 7.3 points, removing false angle measurements of the occluded joints. The highest contribution (11.0 points) is by user-specific normalization which suggests that user-specific normalization is a critical innovation.

V. Discussion

Our study confirms that proper, real time form assessment does not have to be carried out in a specialized laboratory laboratory; it can be done at the comfort of an average living room by simply using a normal camera and intelligent adaptive programming. One of the key lessons after our conclusions is that fixed strict rules that fit all people are not suitable when it comes to the variety of the human body. Replacing the traditional thesauris with adaptive thresholding that accounts for personal flexibility and proportion lifted our accuracy by 8.8% and this is what is necessary to gain user confidence and offer corrections that are sensible. More importantly we were able to run a full 32 FPS on simple consumer hardware breaking the belief that you have to have a high-end GPU to drive advanced fitness AI. The system could withstand even the noise of a real home, such as subdued lighting or clumsy camera position, with an accuracy rate of 92.1% showing the type of graceful deterioration making a tool as solid as it can be in unpredictable hands.

Even though the existing outcomes have been encouraging, there are still some practical challenges to the system. Firstly, it has problems with the ones where limbs are not visible, now it only does not attempt guessing the location of a person but disregards those ones. We also realize that we have only tested three fundamental exercises and a comparatively small group of individuals, and therefore will have had to test it on a much wider population of types and movements of the body before it is into the professional clinical scene. Regarding the user experience, the system is at the moment restricted to one individual at a time, and needs a manual calibration phase to achieve the correct baseline. Though this arrangement gives accuracy to the feedback, we are considering the so-called zero-calibration techniques so that the experience between opening the app and the actual workout could be seamless.

The shift towards the real world application requires giving attention to the user experience and privacy; hence, our system will use local processing on-site to save sensitive information. The feedback mechanism is dual- injury prevention cues are provided, and beginner motivation is preserved by adaptive sensitivity of the feedback mechanism that adapts to user proficiency. In order to optimize the effect, the system will be designed to be connected with an already existing fitness platform and social ecosystem.

In order to compare our AI with the professionals, we completed a blind test with a trained specialist on 30 random videos and the outcomes were disappointingly similar. The system and the trainer agreed on 77% of occasions and when there was a deviation the deviation was invariably by a very small margin (such as the AI recommending a slight change and the trainer indicating it was good). Overall, the system predicted the expert judgement of the trainer 97% (one category) correctly, and there was one major discrepancy. The single exception tended to be due to little details, such as a slight twist to the torso, which our existing joint-angle range is still incapable of measuring. All in all, the results indicate that the system is even at a stage that it can be compared to a human coach when it comes to these specific exercise.

VI. Conclusion

By doing this study, we have shown that elite-level fitness training does not have to deal with a personal trainer or even costly sensors to be executed effectively; it can be done with such a smart and adaptable computer vision. We have combined MediaPipe pose estimation with our adaptive thresholding logic to create the system that sees exercise form with 96.2 percent accuracy. Our solution can perform well even in low light or at awkward angles due to the fact that unlike earlier models which cannot work outside the laboratory, our solution is stable at 32 FPS on standard laptops. We have removed the inflexibility of fixed, one-size-fits-all measures and developed an algorithm that adapts to the body type of a person to provide a more precise tool that is also practical when used in the home workouts.

We intend to move forward planning this system as a reactive tool into a proactive coach. We will consider adding 3D pose estimation support, which would allow us to process more complicated motions, and sensor fusion with smartwatches that would fabricate even more precise motion data. Another big opportunity that we can see is to go beyond the area of general fitness to go to the clinical area. Throughout the combination of physical therapists, we are confident that such algorithms can be improved to meet the medical rehabilitation and accessibility needs of individuals with limited mobility, as well as those who have undergone surgery

and not yet have the same needs concerning guided assistance. The addition of more exercises via automated machine learning will also play a central role in turning the system into a Swiss-knife of health.

After all, this work demonstrates that exercise feedback of professional quality can be in the hands of everyone who has a camera and a screen. The key obstacles that had positioned this technology out of the home have been addressed by resolving the everyday inconveniences of the surrounding noise and user variation. The difference between human coaching and AI assistance will further decrease as the equipment is faster and the vision models are more advanced. We think that this will be a step toward the new era of smart customized health advice being a ubiquitous resource, capable of addressing physical health and stopping injuries on a global scale.

REFERENCES

[1] [1] R. Ullah, I. Asghar, R. Nawaz, C. Stacey, S. Akbar, and P. Bishop, "A real time action scoring system for movement analysis and feedback in physical therapy using human pose estimation," Scientific Reports, vol. 15, article 44784, Dec. 2025. doi- 10.1038/s41598-025-29062-7

[2] [2] Y. Liang et al., "Exercise assessment based on human pose estimation and neural networks," IEEE Access, vol. 12, pp. 85342-85356, 2024. doi- 10.1109/ACCESS.2024.3412876

[3] [3] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2017, pp. 7291-7299.

[4] [4] A. Kumar, S. Sharma, and R. Patel, "Live monitoring fitness application using pose estimation," GRENZE International Journal of Engineering and Technology, vol. 9, no. 1, pp. 125-134, 2023.

[5] [5] P. Kumar, M. Singh, and A. Verma, "Real-time posture monitoring for effective exercise using MediaPipe," Journal of Innovative Web Engineering, vol. 7, no. 2, pp. 45-58, 2023.

[6] [6] N. Johnson, "AI pose estimation with Python and MediaPipe," YouTube, Apr. 13, 2021. [Online]. Available- https-//www.youtube.com/watch?v=06TE_U21FK4

[7] [7] M. Thompson and K. Lee, "Human activity recognition for fitness and therapy applications," InData Labs Technical Report, Jun. 2025. [Online]. Available- https-//indatalabs.com/resources/human-activity-recognition-fitness-app

[8] [8] L. Zhang, M. Chen, and Y. Wang, "Recognizing multiple human activities and tracking full-body pose in unconstrained environments," Pattern Recognition, vol. 45, no. 1, pp. 11-23, Jan. 2012. doi-10.1016/j.patcog.2011.05.016

[9] [9] S. Liao, R. Kumar, and T. Chen, "Accurate gym exercise form detection using MediaPipe- A machine learning approach," International Journal for Research in Applied Science and Engineering Technology, vol. 12, no. 8, pp. 2156-2164, Aug. 2024.

[10] [10] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, "BlazePose- On-device real-time body pose tracking," arXiv preprint arXiv-2006.10204, Jun. 2020.

[11] [11] S. Patel, A. Mehta, and R. Gupta, "Gym tracker system using AI-driven pose estimation and real-time exercise monitoring," RSI International Journal, vol. 11, no. 4, pp. 89-102, 2024.

[12] [12] M. Ali, K. Hassan, and F. Ahmed, "Adaptive exercise meticulousness in pose detection and correction," International Journal of Computer and Data Sciences, vol. 10, no. 3, pp. 234-247, 2023.

[13] [13] D. Dwibedi, Y. Aytar, J. Tompson, P. Sermanet, and A. Zisserman, "Counting out time- Class agnostic video repetition counting in the wild," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2020, pp. 10387-10396.

[14] [14] A. Singh and P. Reddy, "Real time fitness tracking and analysis using BlazePose," International Journal of Adv. Research in Science, Communication and Technology, vol. 8, no. 2, pp. 412-420, 2023.

[15] [15] H. Liang, J. Wu, and C. Zhang, "3D pose estimation accuracy for clinical gait analysis," Journal of Biomechanics, vol. 89, pp. 134-142, 2019.

[16] [16] M. Aleksic, D. Smith, and K. Brown, "Accuracy assessment of 2D pose estimation with MediaPipe for fitness applications," Procedia Computer Science, vol. 245, pp. 1234-1243, 2024. doi-10.1016/j.procs.2024.09.142

[17] [17] N. Patel and R. Kumar, "YOLOv7 pose vs MediaPipe in human pose estimation- Comparative analysis," LearnOpenCV, Nov. 2025. [Online]. Available- https-//learnopencv.com/yolov7-pose-vs-mediapipe

[18] [18] F. Garcia, T. Lopez, and S. Martinez, "Accuracy evaluation of 3D pose reconstruction algorithms through motion capture validation," PMC Biomechanics Journal, vol. 12, article PMC11644880, Dec. 2024.

[19] [19] L. Chen, Y. Wang, and X. Liu, "A deep dive into MediaPipe pose for postural assessment," in Proc. IEEE Int. Conf. Multimedia and Expo (ICME), 2024, pp. 1-6. doi- 10.1109/ICME.2024.11297977

[20] [20] J. Kim, "30+ FPS pose detection with MediaPipe and OpenCV Python," YouTube, May 18, 2021. [Online]. Available- https-//www.youtube.com/watch?v=oRrRl5VDCWM

[21] [21] B. Bedo, "mediapipe2dangle- Using MediaPipe to calculate hip and knee joint angles," GitHub, Sep. 2024. [Online]. Available- https-//github.com/brunobedo/mediapipe2dangle

[22] [22] Y. Wang, H. Chen, and M. Li, "Motion estimation using frame-based adaptive thresholding for exercise analysis," Pattern Recognition Letters, vol. 25, no. 3, pp. 339-351, Feb. 2004. doi-10.1016/j.patrec.2003.10.014

[23] [23] R. Martinez, K. Johnson, and L. Davis, "Metrics for performance evaluation of patient exercises during physical therapy," PMC Physical Therapy Research, vol. 8, article PMC5526359, Apr. 2017.

[24] [24] S. Ahmed, T. Hassan, and F. Khan, "Real-time workout posture monitoring with human pose estimation," in Proc. IEEE Int. Conf. Consumer Electronics (ICCE), Dec. 2024, pp. 1-5. doi-10.1109/ICCE.2024.10949246.

[25] [25] Dataset used - [Workout/Exercises Video], Gym workout and bodyweight exercise videos. Consists of workout videos and name of the existing workout corresponds to the name of the folder listed. All the videos are of *.mp4/.mov* format. https-//www.kaggle.com/datasets/hasyimabdillah/workoutfitness-video

[26] [26] J. Lee, S. Park, and H. Kim, "Human pose estimation in extremely low-light conditions," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2023, pp. 12456-12465.

[27] [27] "Pose landmark detection guide," Google AI Edge MediaPipe Documentation, 2024. [Online]. Available- https-//ai.google.dev/edge/mediapipe/solutions/vision/pose_landmarker