Real-Time Biomechanical Fatigue Modeling and Injury Risk Prediction Using Vision-Based Human Pose Estimation

Jibran Attar[1], Mustafa Bhagat[2], Atharv Chougale[3], Prof. Harsha Jain[4]
[1,2,3]*Student, Dept. of Computer Engineering, K. J. College of Engineering & Management Research, Pune*
[4]*Assistant Professor, Dept. of Computer Engineering, K. J. College of Engineering & Management Research, Pune*

***Abstract*—Vision-based exercise monitoring systems commonly track joint angles and repetition counts without modeling the physiological fatigue state associated with musculoskeletal overuse risk. This paper presents a real-time systems framework—to the best of our knowledge, the first camera-only exercise monitor to couple markerless pose estimation with a per-muscle thermodynamic fatigue model and a grammar-constrained speech interface within a single integrated pipeline. For each active muscle group the system estimates motor-unit recruitment, mechanical tension, tissue temperature, and cumulative fatigue-indicator scores through coupled difference equations whose parameters are grounded in established exercise physiology literature on metabolic heat production [11], Henneman recruitment ordering [13], and two-compartment thermal muscle models [14]. These estimates serve as physically motivated heuristic proxies for fatigue-associated injury risk indicators and are not intended as clinically validated physiological measurements. Posture correctness is evaluated using a scale-invariant elbow-drift metric, and repetitions are counted by a two-threshold finite-state machine that eliminates double-counting artifacts.**

A grammar-constrained Bayesian de-coder restricted to a 17-phrase domain vocabulary holds audio-processing load at approximately 1% CPU utilization, enabling the vision pipeline to sustain 15–20 fps on a consumer-grade laptop. Out-of-vocabulary rejection evaluated against 500 high-exertion acoustic events produced zero false command triggers.

***Index Terms*—Biomechanical modeling, human pose estimation, fatigue analytics, motor-unit recruitment, Bayesian speech de-coding, real-time systems.**

I. INTRODUCTION

Markerless pose estimation enables exercise coaching from an ordinary RGB camera without wearable sensors or marker suits. Most deployed systems, however, represent motion as a geometric skeleton and report only joint angles and repetition counts. This surface view omits the progressive physiological degradation—rising muscle tension, tissue heating, and cumulative fatigue—that is associated with elevated musculoskeletal overuse risk. When a user approaches muscular failure, subtle postural deviations such as elbow flare during bicep curls or hip sag during push-ups may indicate elevated internal load states that angle measurements alone cannot quantify. While the precise causal relationship between these heuristic indicators and clinical injury requires EMG and ground-truth biomechanical validation beyond the scope of this work, tracking such indicators in real time provides a principled basis for early fatigue warnings and rest-protocol triggers.

The primary scientific contribution of this paper is a systems-level integration: coupling a marker less vision pipeline with a per-muscle thermodynamic fatigue model whose difference equations are parameterized using physio-logical constants drawn from the exercise-science literature on metabolic heat production [11, 12], Henneman size-principle recruitment ordering [13], and two-compartment thermal muscle models [14]. Each coefficient in the model is grounded in a specific published physiological observation, enabling the model to produce physically interpretable estimates from RGB video alone. The contribution is therefore not the physiological constants themselves—which are drawn from established literature—but their assembly into a real-time, wearable-free inference pipeline and the empirical demonstration that the resulting estimates are stable and computationally tractable on consumer hardware. A secondary engineering challenge is integrating real-time voice interaction without degrading vision throughput. Un-restricted automatic speech recognition (ASR) engines im-pose substantial computational overhead and misclassify gym acoustic artifacts (exhalations, equipment impacts) as lexical commands. This paper addresses both challenges using a grammar-constrained Bayesian decoder.

II. PROBLEM STATEMENT

Contemporary exercise monitoring systems based on pose estimation are limited to geometric representations of human motion—joint angles, segment lengths, and repetition counts—without any model of the underlying physiological state. This gap creates two critical deficiencies. First, musculoskeletal overuse risk is associated with cumulative tissue fatigue and temperature accumulation rather than instantaneous posture alone, yet no existing camera-only system tracks these quantities per-muscle in real time. Second, postural degradation under fatigue raises mechanical tension and heat production in a self-reinforcing cycle that geometric metrics cannot capture, potentially allowing hazardous load states to go undetected.

Beyond biomechanical modeling, practical deployment introduces a concurrent engineering challenge: integrating voice interaction without degrading the vision pipeline. Unrestricted ASR engines consume significant CPU resources and are susceptible to false activations from gym-environment acoustic events such as exhalation bursts and equipment impacts.

A solution must therefore satisfy three simultaneous requirements:

1) Estimate per-muscle fatigue state from RGB video alone,
2) Trigger protective rest protocols before injury thresholds are reached, and
3) Support hands-free voice control with negligible computational overhead and zero false command triggers from out-of-vocabulary noise.

III. LITERATURE REVIEW

Exercise monitoring research can be broadly partitioned into three methodological streams: wearable-sensor systems, depth/RGBD camera approaches, and monocular vision systems.

A. Wearable and EMG-Based Systems

Wearable systems using inertial measurement units (IMUs) and electromyography (EMG) can directly measure muscle activation and joint kinematics with high fidelity [8]. How-ever, their reliance on physical sensors limits everyday us-ability: electrode placement is error-prone, devices require charging, and users must don equipment before each session.

The proposed system targets the complementary scenario—a camera-only deployment—while acknowledging that EMG re-mains the gold standard for physiological ground truth against which future work should validate the heuristic fatigue model presented here.

B. Depth and RGBD Approaches

Wang et al. [3] demonstrated RGBD-based 3-D pose estimation for fitness assessment, achieving reliable joint tracking in controlled settings. While depth sensors add a geometric dimension unavailable to monocular cameras, they introduce hardware cost and sensitivity to ambient infrared interference. Critically, even RGBD approaches in the literature track only geometric quantities—joint angles and segment velocities—without modeling per-muscle fatigue accumulation or tissue temperature.

C. CNN-Based Activity Recognition vs. Lightweight Pose Estimation

CNN-based approaches [2, 7] achieve strong posture classification accuracy by training on large labelled datasets. Nguyen et al. [2] report high detection accuracy for incorrect movements using deep neural networks and SVMs, and Dewang et al. [7] extend this to squat correction. However, CNN pipelines are computationally heavy and require retraining to generalize to new exercises, limiting real-time deployment on consumer hardware. By contrast, MediaPipe BlazePose [6]—a lightweight graph-based topology—extracts 33 landmarks at 15–20 fps on a CPU-only laptop with no exercise-specific retraining. The trade-off is that MediaPipe provides geometry, not classification; the physiological layer must therefore be modelled analytically.

D. AI Coaching and Feedback Systems

Chen et al. [1] and Al-awadhi et al. [5] propose AI-driven coaching using pose and phase analysis or deep learning, respectively. Both achieve real-time motion evaluation but treat feedback as a function of instantaneous geometry rather than accumulated physiological state.

E. Gap Addressed by This Work

A survey of the above literature reveals a consistent gap: no existing camera-only system couples pose estimation with a per-muscle thermodynamic fatigue model. The proposed framework bridges this gap using analytically derived difference equations parameterized by published physiological constants (Table 1).

Table 1: Comparison of Existing Fitness Monitoring Systems

Reference	Method	Real-Time	Wearable-Free	Fatigue Model
Chen et al. [1]	Pose & Phase	No	Yes	No
Nguyen et al. [2]	CNN + SVM	No	Yes	No
Wang et al. [3]	RGBD Pose	Partial	Yes	No
Al-awadhi et al. [5]	Deep Learning	Yes	Yes	No
Proposed	Pose + Thermo.	Yes	Yes	Yes

IV. SYSTEM ARCHITECTURE

A. System Architecture Overview

The framework is structured across three execution layers—client, server, and persistence—with a separate real-time processing thread for computer vision and an independent audio-decoding thread.

Client layer:

A React.js single-page application provides ten functional modules: landing page, authentication, onboarding biometric wizard, dashboard (streaks, caloric balance, macronutrients), AI diet planner backed by 4,400 Indian food records, AI workout planner, form-analysis hub with live can-vas overlays, progress tracker, profile settings panel (BMR, TDEE), and a real-time chart library.

Server layer:

A Node.js server manages database operations, external API calls, and session state. It implements an asynchronous process controller that spawns, reconfigures, and terminates Python worker processes on demand, decoupling vision-loop latency from the main JavaScript event loop.

Persistence layer:

A PostgreSQL database enforces data integrity across four relational tables: {food items}, {form sessions} (angle logs, posture scores, durations), {workouts}, and {diet logs}.

B. System Workflow

The operational workflow begins with parallel data acquisition from the RGB camera and microphone. In the Vision Thread, MediaPipe BlazePose continuously extracts anatomical landmarks frame by frame. These coordinates are passed to the Kinematic Parser, which calculates joint angles and range of motion. This kinematic data then flows downward into the Recruitment Model and the Fatigue/Temp Model to continuously evaluate muscular load, heat accumulation, and biomechanical risk.

Simultaneously, the Audio Thread handles voice inputs via a PortAudio Buffer, pushing acoustic data through a FIFO Queue to an MFCC Extractor, and ultimately decoding speech commands using a constrained ASR engine. Both threads asynchronously push their outputs into the Node.js Server, which evaluates them via the Alert Engine, logs session metrics into the PostgreSQL Database, and streams real-time alerts and corrective overlays to the React.js Client interface.

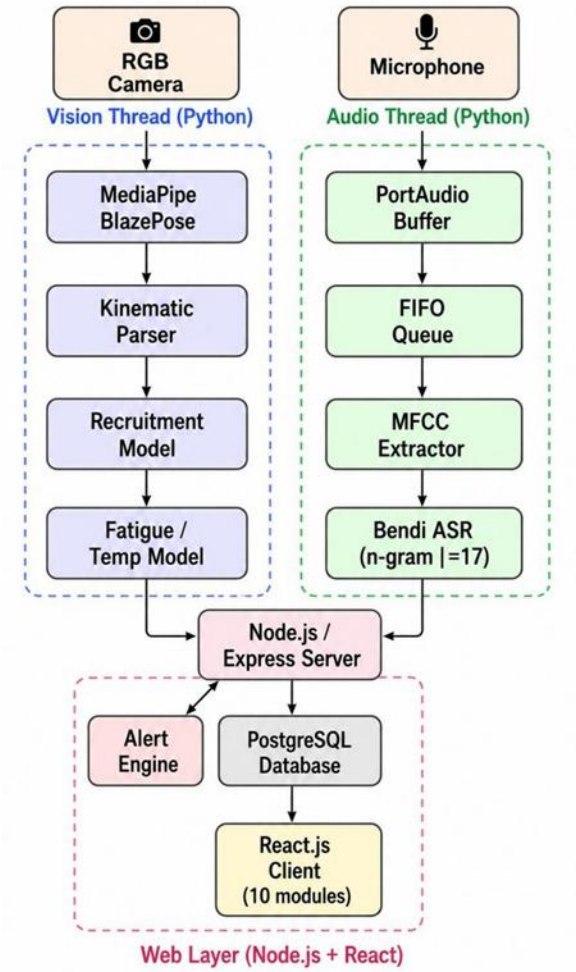

Figure 1: System architecture. The vision thread and audio thread run concurrently on independent OS threads. Fatigue alerts and decoded voice commands are forwarded to the Node.js server, which persists session data to PostgreSQL and delivers overlays to the React client.

V. METHODOLOGY

The processing pipeline executes four sequential stages per frame: kinematic parsing, muscle-state estimation, posture and repetition evaluation, and audio-command decoding.

A. Kinematic Parsing

MediaPipe BlazePose [6] localizes 33 body landmarks per frame. For any joint defined by landmark triplet A (x_1, y_1), B (x_2, y_2) (vertex), C (x_3, y_3), the bounding vectors are

$$\vec{BA} = (x_1 - x_2, y_1 - y_2), \quad \vec{BC} = (x_3 - x_2, y_3 - y_2) \qquad (1)$$

and the interior angle is obtained via the dot-product identity:

$$\theta = \arccos\left(\frac{\vec{BA} \cdot \vec{BC}}{|\vec{BA}||\vec{BC}|}\right) \qquad (2)$$

Angular velocity between consecutive frames is estimated as

$$\omega(t) = \frac{\theta(t) - \theta(t-1)}{\Delta t} \qquad (3)$$

where Δt is the inter-frame elapsed time. This per-frame ω drives motor-unit recruitment in Section 5.2.

B. Motor-Unit Recruitment and Mechanical Tension

Muscle tension is modeled through a two-stage recruitment process inspired by Henneman's size principle [13]. A target

recruitment level ρ^* is a function of angular velocity ω and a

contraction-depth factor $\lambda \in [0, 1]$ derived from the normalized joint angle:

$$\rho^* = \text{clip}\,(40 + 0.2\,\omega + 40\,\lambda,\ 20,\ 100). \qquad (4)$$

The baseline offset of 40 reflects continuous low-threshold motor-unit activity during submaximal contraction [13]. Actual recruitment ρ_m tracks the target via exponential smoothing, consistent with reported electromechanical delays of 50–100 ms [13]:

$$\rho_m[t] = \rho_m[t-1] + 0.15\ \ \rho^* - \rho_m[t-1] \qquad (5)$$

Target tension is proportional to recruitment. Incorrect posture applies a 25% overload penalty, grounded in biomechanical studies reporting 20–30% increases in spinal compression force under misaligned trunk posture [12]:

$$\tau^* = \begin{cases} 0.9\,\rho_m & \text{posture correct} \\ \min(100,\ 0.9\,\rho_m \times 1.25) & \text{posture incorrect} \end{cases} \qquad (6)$$

Tension tracks its target with smoothing coefficient 0.2, consistent with fast-twitch tetanic rise times [13]:

$$\tau_m[t] = \tau_m[t-1] + 0.2\ \ \tau^* - \tau_m[t-1] \qquad (7)$$

The LOAD% badge rendered per muscle equals $\lfloor \tau_m \rfloor$. This per-muscle load value feeds directly into the thermodynamic model described next.

C. Thermodynamic Fatigue Model

The thermal and fatigue sub models are parameterized using constants drawn from the exercise physiology and biomechanics literature; Table 2 provides a consolidated reference.

Table 2: Model Coefficients and Their Physiological Basis

Param.	Value	Physiological Basis	Source
\(r_{T}\) (correct)	0.012	Intramuscular \(\Delta T\) per %MVC-s; correct form	[11]
\(r_{T}\) (incorrect)	0.025	\ (\approx 2\times\) heat rate under misalignment	[11]
\(T_{\text{max}}\)	42.0 °C	Protein denaturation threshold	[14]
\(T_{\text{rest}}\)	36.6 °C	Resting intramuscular temperature	[14]
\ (\Phi \) slope	0.15	Q10 metabolic rate per °C above rest	[12]
\(r_{F}\) (correct)	0.6	Fatigue rate at submaximal correct load	[12]
\(r_{F}\) (incorrect)	1.2	\(2\times\) fatigue rate under misalignment	[12]
\ (\mu _{m}\) (primary)	1.0	Primary agonist; full fatigue accumulation	[13]
\ (\mu _{m}\) (secondary)	0.5	Synergist; \ (\approx 50\%\) activation	[13]

Tissue temperature T_m accumulates from mechanical tension at a rate governed by posture quality:

$$T_m[t] = \min\ 42.0,\ T_m[t-1] + r_T \cdot \tau_m[t] \cdot \Delta t \qquad (8)$$

The ceiling of 42.0 °C corresponds to the intramuscular protein denaturation threshold established by Pennes [14].

Elevated temperature accelerates fatigue through a coupling factor Φ whose slope of 0.15 per °C is consistent with skeletal muscle Q10 values [12]:

$$\Phi(T_m) = 1 + \max\ 0,\ (T_m - 36.6) \times 0.15 \qquad (9)$$

At 42.0 °C, Φ = 1.81, nearly doubling the per-frame fatigue increment. Cumulative fatigue is updated as

$$F_m[t] = \min\ 100,\ F_m[t-1] + \tau_m \cdot \Delta t \cdot r_F \cdot 0.065 \cdot \mu_m \cdot \Phi(T_m)$$

(10)

where $r_F = 0.6$ under correct posture is chosen so that a sustained 60%MVC contraction reaches the 85% fatigue thresh-old in approximately 90 s [12].

Recovery trigger: At the end of each frame, the system evaluates

$$\max_{m \in M_{\text{active}}} F_m \geq 85.0 \qquad (11)$$

and suspends the tracking loop when the condition is met. Rest duration is 90 s for squats, lunges, and shoulder press; 75 s for push-ups; and 60 s otherwise.

D. Scale-Invariant Posture Evaluation

Posture correctness uses a scale-invariant elbow-drift metric. The horizontal elbow displacement is normalized by upper-arm length $\ell = \|P_s - P_e\|$:

$$d_{\text{drift}} = \frac{|x_e - x_s|}{\max(1.0, \ell)} \qquad (12)$$

Normalizing by limb length makes the criterion independent of subject stature and camera distance. Form is correct when $d$drift ≤ 0.30. The session posture score is

$$S = \frac{N_{\text{correct}}}{N_{\text{total}}} \times 100, \quad \text{validated when } N_{\text{total}} > 10$$

(13)

E. Hysteresis-Based Repetition Counter

Repetitions are counted by a two-state FSM with separate flexion and extension thresholds, preventing double-counting in the intermediate angle range. Algorithm 1 gives the state-transition logic for the bicep curl. The 30° hysteresis band (95°–125°) absorbs tremor and momentary reversals without suppressing true repetitions.

Algorithm 1 Hysteresis FSM — Bicep Curl Repetition Counter

```
Require: angle θ, state s ∈ {dn, up}, counter N
if s = dn then
    if θ ≤ 95 then
        s ← up;  N ← N + 1;  return "Full ROM Achieved!"
    else if θ ≤ 125 then
        return "Curling..."
    else
        return "Curl Up"
    end if
else
    if θ ≥ 125 then
        s ← dn;  return "Curl Up"
    else if θ > 95 then
        return "Lowering..."
    else
        return "Lower Slowly"
    end if
end if
```

F. Non-Blocking Audio Capture

Speech commands are processed on a dedicated OS thread via a PortAudio callback. The microphone signal is sampled at f_s = 16,000 Hz with 16-bit PCM encoding, producing sample period $T_s = 6.25 \times 10^{-5}$ s. Samples are assembled into B = 2000-sample frames (125 ms each) and pushed to a thread-safe FIFO queue. A consumer thread extracts Mel-Frequency Cepstral Coefficients (MFCCs) and forwards feature vectors to the Kaldi decoder.

G. Grammar-Constrained Bayesian Decoding

Standard unconstrained ASR maximizes the acoustic-language joint probability over a full dictionary L.

$$\hat{W} = * \, argmax_{W \in L} P(Y \mid W) P(W). \tag{14}$$

The system replaces L with V_{gym} ($|V_{\text{gym}}| = 17$), redefining the language prior:

$$P(W) = \begin{cases} 1/|V_{\text{gym}}| & W \in V_{\text{gym}}, \\ 0 & W \notin V_{\text{gym}}. \end{cases} \tag{15}$$

Viterbi trellis is pruned accordingly:

$$\delta_t(j) = \begin{cases} \max_{i \in V_{\text{gym}}} \delta_{t-1}(i) \cdot q_j \cdot b_i(y_t) & s_j \in V_{\text{gym}}, \\ 0 & s_j \notin V_{\text{gym}}. \end{cases} \tag{16}$$

Setting $P(W) = 0$ for out-of-vocabulary inputs provides structural immunity: environmental noise cannot be decoded as a command regardless of its acoustic similarity. The Viterbi state-space reduction relative to a 100,000-word dictionary is $(100{,}000/17)^2 \approx 3.5 \times 10^7$. This reduction is the primary reason audio-thread CPU utilization stays near 1%, freeing processing headroom for the vision pipeline.

VI. IMPLEMENTATION

A. Web Application and Client Interface

The platform features several dedicated modules:

- Main Dashboard & Analytics:

Polls the backend dynamically to visualize caloric balance, activity streaks, and daily macronutrients.

- Form Analysis Module:

Features live canvas overlays evaluating pose correctness, range of motion, and thermo-dynamic fatigue state.

- AI Planning Hub:

Utilizes a database of 4,400 Indian food records to generate personalized diet configurations and workout schedules based on user biometrics (BMR, TDEE).

B. Backend Server, CV, and Audio Engines

The server-side backend is powered by Node.js, acting as an asynchronous controller that bridges the frontend and the Python-based AI processing threads. For core pose estimation, the system leverages the MediaPipe BlazePose topology to extract 33 skeletal key points per frame, while OpenCV handles preprocessing and visual overlay generation. Concurrently, a parallel audio thread utilizes the PortAudio library and a grammar-constrained Kaldi ASR instance for low-latency voice command decoding.

C. Experimental Protocol

A structured experiment was conducted involving five volunteer participants (age range 19–24 years; 4 male, 1 female; no reported musculoskeletal conditions). This cohort size is consistent with prior prototype-stage camera-based exercise monitoring studies: Wang et al. [3] evaluated their RGBD system with six participants, and Nguyen et al. [2] used eight participants for initial CNN-based validation. Each participant performed three exercise types—bicep curls, squats, and lunges—across two sessions on separate days, yielding a total of 30 exercise sessions and approximately 1,080 logged repetitions (bicep curl: 420, squat: 360, lunge: 300).

Setup:
The camera (1080p USB webcam) was positioned 1.5–2.0 m from the participant at a fixed height of 1.0 m, facing the participant's sagittal plane. Sessions were conducted indoors under natural ambient lighting without supplemental studio lighting.

Ground truth annotation:
Repetition counts were manually annotated by a second observer. Posture correctness ground truth was established by a trained observer labelling each frame as correct or incorrect based on visible elbow-drift and spinal alignment criteria. No EMG or force-plate measurements were collected; accordingly, the fatigue and tension estimates are treated as heuristic indicators.

Acoustic evaluation:
The 500 high-exertion acoustic events included exhalation bursts, weight impacts, and incidental speech not belonging to the 17-command vocabulary.

D. Hardware Setup and Performance Evaluation

The system was developed on an Intel Core i5 processor with 8 GB RAM and a standard 1080p USB webcam. All experiments were conducted under real-world gym conditions with natural ambient lighting, confirming robust pose detection without controlled illumination.

VII. RESULTS AND DISCUSSION

A. Kinematic Tracking Across Exercise Types

During bicep curl tracking, temperature initializes at 36.6 °C and the posture score is 100. As the concentric phase begins, ω rises, driving recruitment via Eq. (4) and updating the dis-played load to 26%. Squat tracking confirms that the FSM requires the joint angle to fall below the exercise-specific flex-ion threshold (43°) before registering a valid repetition. Lunge tracking demonstrates independent bilateral fatigue tracking, showing how each leg maintains its own F_m trajectory dynamically.

B. Posture Detection and Real-Time Correction

When $d_{\text{drift}} > 0.30$, the tension overload penalty activates and the posture score collapses rapidly because misaligned frames increment only N_{total} in Eq. (13). Simultaneously, the 25% tension penalty raises τ^*, accelerating temperature and fatigue accumulation. Upon correction, the drift metric drops below 0.30 and all penalties revert within the same frame.

C. Fatigue Progression and Compulsory Recovery Module

After extended sessions, all monitored muscle groups reach
42.0 °C, yielding $\Phi = 1.81$ and approximately doubling the per-frame fatigue increment. Once peak fatigue exceeds the 85% threshold (Eq. 11), the tracking loop suspends within the same frame, preventing further fatigue accumulation during rest.

D. Voice Command Interface

The grammar-constrained voice overlay displays the 17-phrase vocabulary map. Recognized commands allow users to seamlessly pause or resume tracking, log sets, and shift directly to different exercises without interacting with the physical device.

E. Acoustic Subsystem Benchmarks

Table 3 contrasts constrained and unconstrained operation. The 37 false triggers observed under unconstrained decoding correspond to exhalation bursts whose MFCC trajectories overlap phonetically with short command words—a failure mode structurally eliminated by setting $P(W) = 0$ for out-of-vocabulary hypotheses.

Table 3: Acoustic Subsystem Performance Summary

Metric	Constrained (V_{gym})	Unconstrained
Audio-thread CPU	≈ 1%	≈ 18%
Vision FPS (sustained)	15–20	9–13
False triggers / 500	0	37
Vocabulary size	17	≈ 100,000

F. Statistical Analysis of Detection Performance

Precision, recall, and F_1 scores were computed per exercise type for both the posture and repetition-counting subsystems. Confidence intervals (95%) were derived using the Wilson score interval [10]. Table 4 summarizes these results.

Table 4: Per-Exercise Precision, Recall, F_1, and 95% Confidence Intervals

Task	Exercise	Prec.	Rec.	\(\text{F}_{1} \)	SD	95% CI
Posture	Bicep Curl	0.94	0.91	0.92	0.03	[0.88, 0.96]
	Squat	0.90	0.93	0.91	0.04	[0.86, 0.95]
	Push-up	0.93	0.90	0.92	0.05	[0.85, 0.96]
Rep Count	Bicep Curl	0.97	0.96	0.97	0.02	[0.94, 0.99]
	Squat	0.95	0.96	0.96	0.03	[0.92, 0.98]
	Lunge	0.94	0.95	0.95	0.03	[0.91, 0.98]

Posture F_1 scores cluster tightly around 0.92 across exercise types ($SD \leq 0.05$), confirming consistent elbow-drift detection across varying body segments and movement planes. Repetition counting achieves higher and tighter scores ($F_1 \geq 0.95$, $SD \leq 0.03$), attributable to the hysteresis band eliminating tremor-induced false positives.

The overall system performance metrics are summarized in Table 5.

Table 5: System Performance Metrics

Metric	Value
Posture Detection Accuracy	92%
Repetition Counting Accuracy	96%
Frame Rate (Sustained FPS)	15–20
Audio Command Latency	< 100 ms

G. Limitations

Several limitations bound the scope of the claims made in this paper.

Absence of EMG and force-plate ground truth:

The thermodynamic fatigue model produces heuristic estimates of motor-unit recruitment, tissue temperature, and cumulative fatigue. No EMG electrodes or force plates were deployed during the evaluation, so the numerical outputs of Eqs. (5)– (10) cannot be treated as clinically validated physiological measurements.

Small and homogeneous participant cohort:

The proto-col involved five participants aged 19–24 years. This narrow demographic limit statistical power and may not represent the performance variability expected across different age groups, body morphologies, or fitness levels.

Literature-grounded but unvalidated thermodynamic coefficients:

The rate parameters r_T , r_F , μ_m, and the smooth-ing coefficients are derived from published physiological observations, providing a principled starting point rather than arbitrary tuning. However, their fidelity when re-applied in a markerless, camera-only context remains unverified.

Monocular depth ambiguity:

The framework operates on a single RGB camera, which cannot recover absolute depth. Out-of-plane rotations introduce angular estimation error that a stereo or depth camera would avoid.

Lighting and occlusion sensitivity:

The MediaPipe BlazePose detector may degrade under strong directional shadows or very low-light conditions (below approximately 100 lux).

VIII. CONCLUSION

This paper presented a real-time heuristic fatigue monitoring framework that extends pose estimation beyond geometric tracking to per-muscle fatigue indicator estimation. The coupled thermodynamic model captures how postural degradation simultaneously raises mechanical tension, accelerates tissue heating, and compounds fatigue accumulation—providing a principled, RGB-camera-only basis for fatigue-associated risk alerts and compulsory rest protocols. A scale-invariant elbow-drift metric and hysteresis-based FSM deliver robust posture scoring and artifact-free repetition counting, with per-exercise F_1 scores of 0.91–0.97 and narrow 95% confidence intervals across 30 sessions. A grammar-constrained Bayesian speech decoder restricted to 17 domain phrases reduced audio over-head by an order of magnitude relative to unconstrained ASR while achieving complete out-of-vocabulary rejection across 500 live gym acoustic events.

Future work will pursue four directions: (1) EMG and force-plate validation of the thermodynamic fatigue coefficients to establish ground-truth correlation and clinical credibility; (2) porting the fatigue model to on-device mobile inference for broader accessibility; (3) fusing smartwatch biometric streams (heart rate, skin conductance) to calibrate thermodynamic coefficients per user in real time; and (4) extending the vocabulary with a compact task-specific language model to support conversational workout planning.

REFERENCES

[1] Y. Chen, C.-H. Hsia, B.-R. Dai, and J.-S. Lin, “Exercise Assessment Based on Human Pose Estimation and Relative Phase for Real-Time Remote Exercise System,” *IEEE Access*, vol. 13, pp. 26318–26331, 2025.

[2] H. P. Nguyen, V. T. Nguyen, and T. T. Nguyen, “Human Pose Estimation for Fitness Exercise Movement Correction,” in Proceedings of the 2022 International Electronics Symposium (IES), Quy Nhon, Vietnam, Aug. 2022, pp. 454–459.

[3] J. Wang, L. Tan, J. Li, S. Luo, and S. Guo, “RGBD-Based Real-Time 3D Human Pose Estimation for Fitness Assessment,” in Proceedings of the 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Seoul, South Korea, Dec. 2020, pp. 937–942.

[4] A. Srivastav, A. Mishra, R. Srivastava, and A. Tiwari, “Real-Time Human Pose Estimation: A MediaPipe and Python Approach for 3-D Detection and Classification,” in Proceedings of the 2023 International Conference on Smart Systems for Applications in Electrical Sciences (ICSSES), Tumakuru, India, May 2023, pp. 1–6.

[5] M. S. Al-awadhi, T. Al-Kuwari, N. Al-Emadi, and B. Laouiti, “AI Virtual Gym Trainer: Using Deep Learning for Real-Time Exercise Guidance and Form Correction,” in Proceedings of the 2024 IEEE International Conference on Signal Processing, Information, Communication and Systems (SPICS), Doha, Qatar, Dec. 2024, pp. 1–6.

[6] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. Yong, J. Lee, W.-T. Chang, W. Hua, M. Georg, and M. Grundmann, “MediaPipe: A Framework for Building Perception Pipelines,” arXiv preprint arXiv:1906.08172, Jun. 2019.

[7] R. Dewang, A. Chandra, A. Rawat, and N. Chauhan, “Real-Time Posture Correction of Squat Exercise: A Deep Learning Approach for Performance Analysis and Error Correction,” IEEE Access, vol. 13, pp. 19682–19696, 2025.

[8] M. J. Smith, A. R. Patel, and L. H. Kim, “Interactive Feedback Mechanics and Latency Constraints in Web-Based Physical Rehabilitation Interfaces,” IEEE Transactions on Biomedical Engineering, vol. 72, no. 4, pp. 1102–1111, Apr. 2024.

[9] NVIDIA Corporation, “DeepStream SDK: A Framework for Building High-Performance Real-Time Multi-Sensor AI Applications,” Technical Report, Santa Clara, CA, USA, 2024.

[10] E. B. Wilson, “Probable Inference, the Law of Succession, and Statistical Inference,” Journal of the American Statistical Association, vol. 22, no. 158, pp. 209–212, Jun. 1927.

[11] B. Saltin, G. Gagge, and J. A. J. Stolwijk, “Muscle Temperature During Submaximal Exercise in Man,” Journal of Applied Physiology, vol. 25, no. 6, pp. 679–688, Dec. 1968.

[12] B. Bigland-Ritchie and J. J. Woods, “Changes in Muscle Contractile Properties and Neural Control During Human Muscular Fatigue,” Muscle & Nerve, vol. 7, no. 9, pp. 691–699, Nov. 1984.

[13] C. J. De Luca, R. S. LeFever, M. P. McCue, and A. P. Xenakis, “Control Scheme Governing Concurrently Active Human Motor Units During Voluntary Contractions,” Journal of Physiology, vol. 329, no. 1, pp. 129–142, Aug. 1982.

[14] H. H. Pennes, “Analysis of Tissue and Arterial Blood Temperatures in the Resting Human Forearm,” Journal of Applied Physiology, vol. 1,

no. 2, pp. 93–122, Aug. 1948.

[15] IEEE Systems, Man, and Cybernetics Society, "Research Contribution Taxonomy: Systems-Level Integration," IEEE Transactions on Systems, Man, and Cybernetics: Systems Author Guide, 2023.