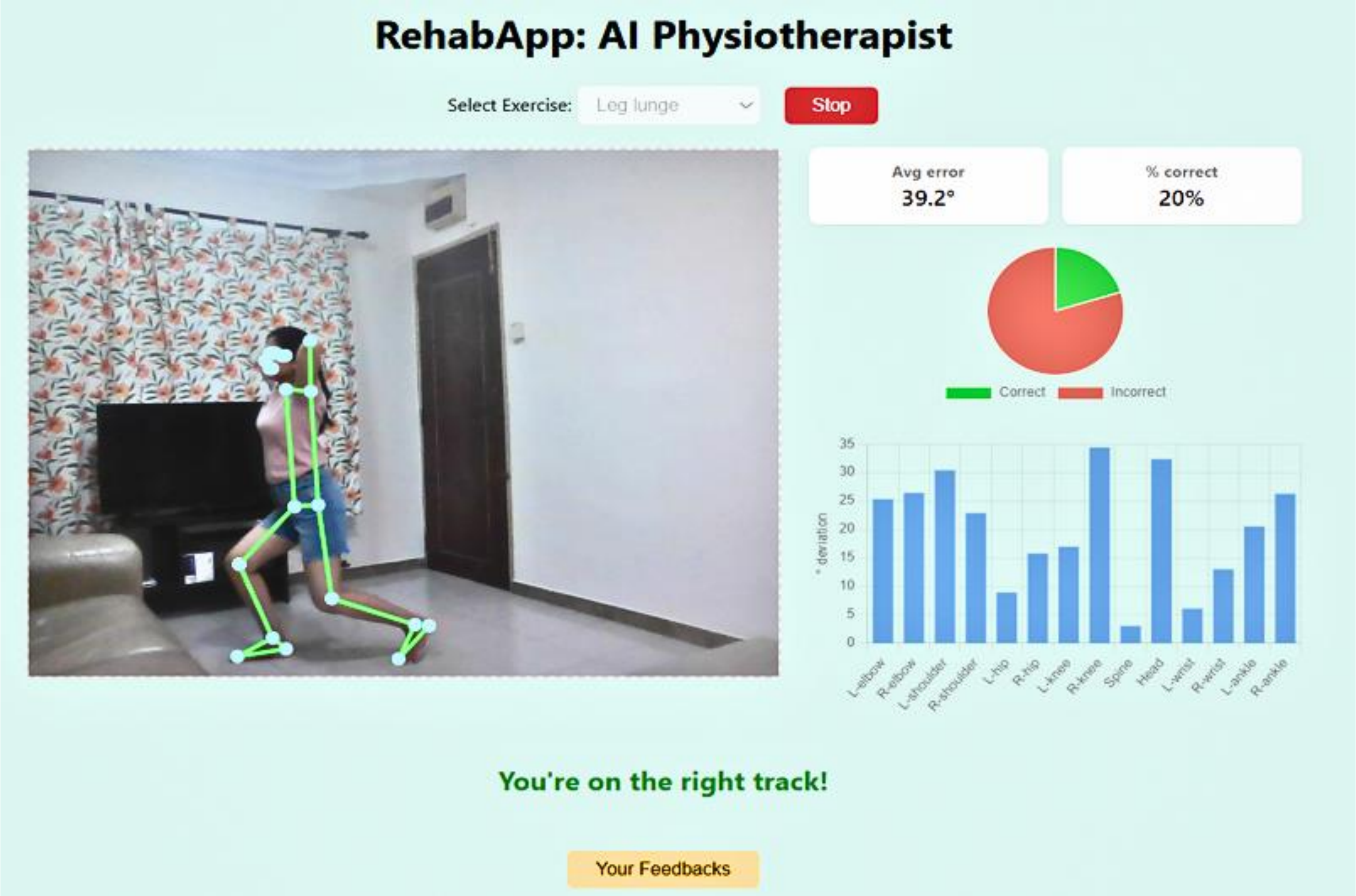
Pose Correction System for Physical Therapy and Rehabilitation Using Computer Vision



Jithin Krishnan (A0249481W)

Master of Technology in Artificial Intelligence Systems
NUS-ISS, National University of Singapore

Motivation: Why home Physical Therapy still needs a coach

Widespread form errors

- 55% of patients perform ≥ 1 drill incorrectly [1]
- Errors usually go unnoticed at home

Real clinical impact

- Mis-execution delays healing
- Re-injury risk increases when cues are missed

Most tools fall short

- Binary pass / fail feedback ⇒ no guidance
- Cloud / GPU needed ⇒ cost, latency, privacy

Lightweight way forward

- Real-time, joint-level, spoken advice on any laptop
- Summarizations and Analysis

[1] Faber et al., PeerJ 2015 [2] Almalty et al., Sport TK 2025.

Improving Rehabilitation :
AI-Powered Physical Therapy

Technical problem

• Laptop-only CV system, must spot 14 joint-angle errors and speak corrective cues in < 0.5s at ≥ 90% accuracy.

Business upside

• Cuts therapist load and follow-up visits → lower costs
• Scales telerehab to thousands without cloud / GPU fees

Impact for Singapore

• Shrinks wait-times for an ageing, physio-heavy population
• Backs Healthier SG & Smart-Nation tele-medicine goals

Key Contributions: RehabApp - AI Physiotherapist

Lightweight Architecture

3-head CNN + Bi-LSTM, only 3.4 M parameters

Granular Feedback

Joint-level advice on 14 angles – not just pass/fail

Built-in Safety

“Wrong-exercise” guard blocks unsafe cues automatically

Real-time & Open

handles 30 video fps on a mid-range CPU; Open-source React + FastAPI

Dataset: REHAB24-6[2]

Key facts	Details
Purpose	Benchmark pose-estimation & exercise-quality algorithms under real rehab conditions
Subjects & drills	10 adults (6♂ / 4♀, 25–50 y) perform 6 physiotherapy exercises Arm-abduction, Arm-VW, Push-ups, Leg-abduction, Lunge, Squats
Volume	65 recordings = 184,825 frames
Ground-truth	41-marker 3-D mocap Derived 26-joint skeleton in 3-D & 2-D Synced RGB videos
Why we chose it	Contains both visual & Motion-Capture Ground Truth (mocap GT) Balanced correct/wrong reps Multiple views → tests occlusion robustness

Exercise ID	Reps	Correct	Wrong	Frames
1	178	90	88	27 442
2	208	94	114	33 641
3	107	52	55	12 054
4	210	120	90	18 329
5	174	78	96	17 608
6	195	134	61	19 373
Total	**1 072**	**568**	**504**	**128 447**

Summary Table

video_id	repetition_nui	exercise_id	person_id	first_frame	last_frame	cam17_orient	mocap_erron	exercise_subt	lights_on	extra_person_	extra_person_	correctness
PM_001	1	1	1	130	324	front	0	right arm	0	1	0	1
PM_001	2	1	1	325	537	front	0	right arm	0	1	0	1
PM_001	3	1	1	538	731	front	0	right arm	0	1	0	1
PM_001	4	1	1	732	919	front	0	right arm	0	1	0	1
PM_001	5	1	1	920	1090	front	0	right arm	0	1	0	1
PM_001	6	1	1	1690	1940	front	0	right arm	0	1	0	0
PM_001	7	1	1	2060	2346	front	0	right arm	0	1	0	0
PM_001	8	1	1	2347	2554	front	0	right arm	0	1	0	0
PM_001	9	1	1	2555	2746	front	0	right arm	0	1	0	0
PM_001	10	1	1	2747	2961	front	0	right arm	0	1	0	0
PM_002	1	1	1	152	363	half-profile	0	right arm	0	1	0	1
PM_002	2	1	1	364	573	half-profile	0	right arm	0	1	0	1
PM_002	3	1	1	574	766	half-profile	0	right arm	0	1	0	1
PM_002	4	1	1	767	910	half-profile	0	right arm	0	1	0	1
PM_002	5	1	1	1000	1225	half-profile	0	right arm	0	2	0	0
PM_002	6	1	1	1226	1407	half-profile	0	right arm	0	2	0	0
PM_002	7	1	1	1408	1605	half-profile	0	right arm	0	1	0	0
PM_002	8	1	1	1606	1800	half-profile	0	right arm	0	1	0	0
PM_002	9	1	1	1801	1967	half-profile	0	right arm	0	2	0	0

Author-supplied frame-level segmentation

[2] Černek A. et al., “REHAB24-6: Dataset for Analyzing Pose Estimation Methods,” Springer LNCS, 2024.

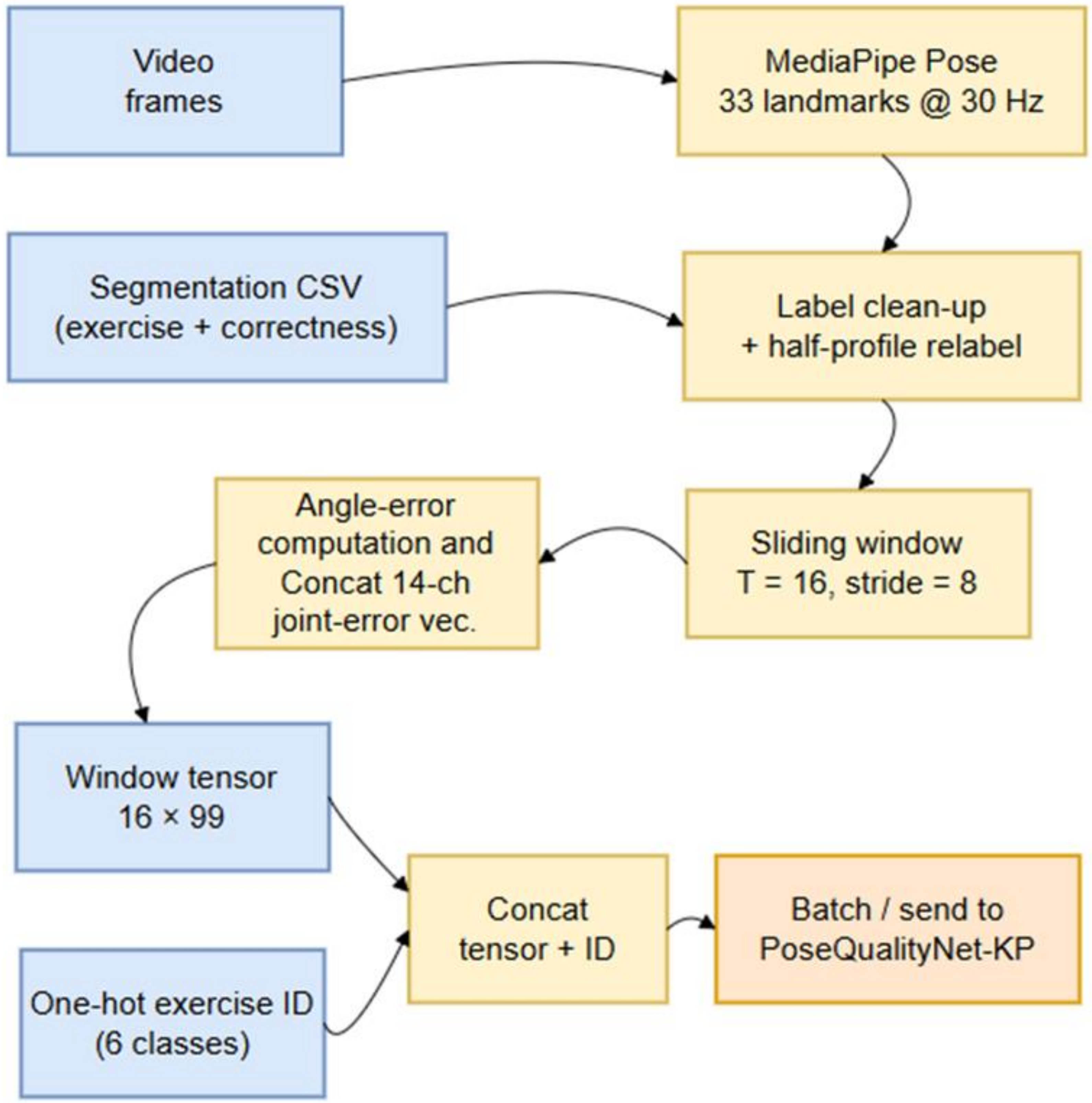
Data:
Augmentations and Pipeline

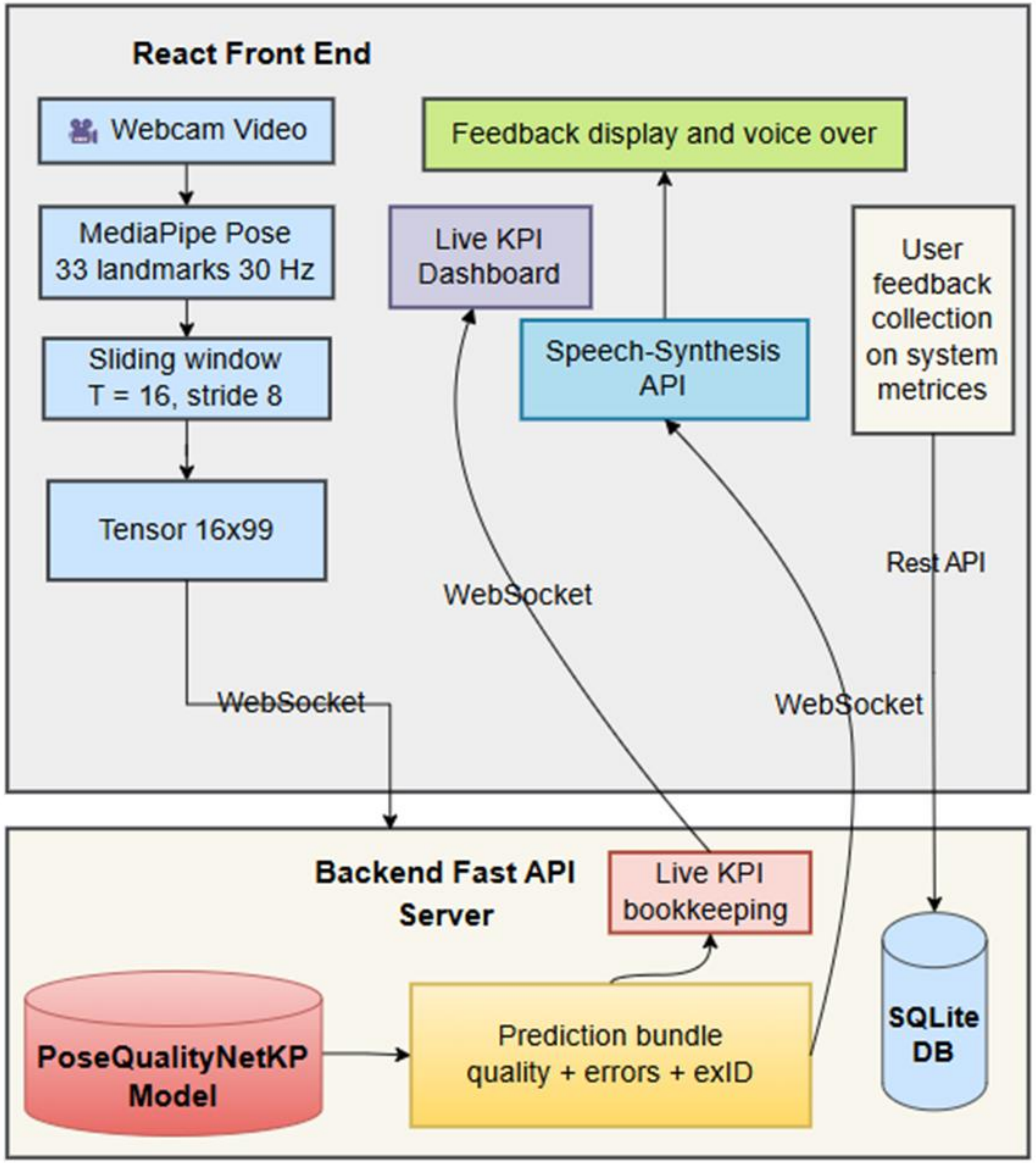
Temporal & Label Augmentations

- Sliding window T = 16 frames, stride = 8 → ×8 more samples
- Half-profile relabel
- Oblique views auto-marked incorrect (for 5 / 6 drills)

Feature Engineering

- 99-D vector / frame (33 key-points × 3 coords)
- One-hot exercise ID (6 classes) appended once per window
- 14 joint-angle errors kept as labels for the regression head





System Architecture:
End-to-End Overview

Front-end (React, runs in browser)

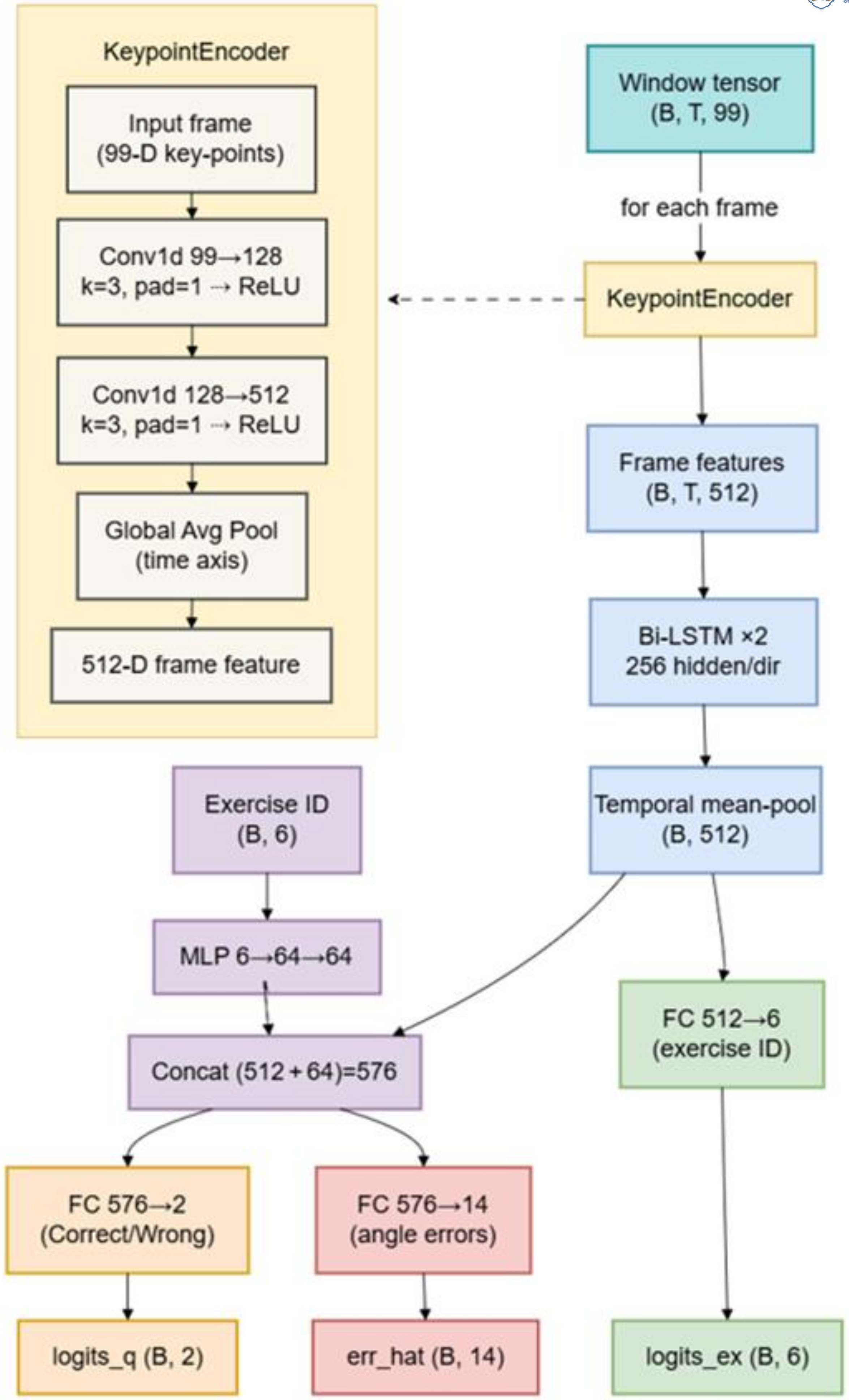
- Captures 640 × 480 webcam video and runs MediaPipe Pose @ 30 Hz
- Builds 16 × 99 tensors + one-hot exercise ID; streams them via WebSocket
- Renders colour-coded skeleton, speaks cues (Speech-Synthesis API) and shows live KPI dashboard

Back-end (FastAPI, on same laptop)

- Hosts PoseQualityNet-KP → predicts quality, 14 joint-angle errors & exercise ID in ≈ 2 ms
- Majority-vote buffer (5 windows) suppresses flicker; “wrong-exercise” guard blocks unsafe advice
- Sends compact JSON to front-end, logs KPIs & survey answers in SQLite (REST endpoint)

Model Architecture:
PoseQualityNet-KP

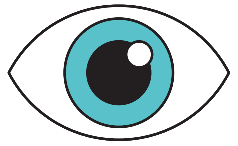
Block	Shape / Notes
Key-point encoder	1-D CNN 99 → 128 → 512
Temporal encoder	2-layer Bi-LSTM (256 h / dir)
Exercise embedding	6 → 64 MLP
Prediction heads	quality (2-way) • angle errors (14-D) • exercise (6-way)
Footprint	3.41 M params ≈ 13 MB



Features: Live Feedback & Analytics

Visual feedback

- Landmarks color change– sky-blue by default; top-3 joints with |Δθ| ≥ 8° turn amber.
- Textual feedback display – green = good, red = bad form, amber = wrong exercise.
- suggestion (angle to to adjust) display in amber color

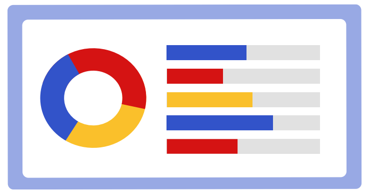


Spoken feedback

- Browser Speech-Synthesis speaks the same message; joint-level advice (e.g. “straighten left knee and right elbow”) has higher priority than generic banners.

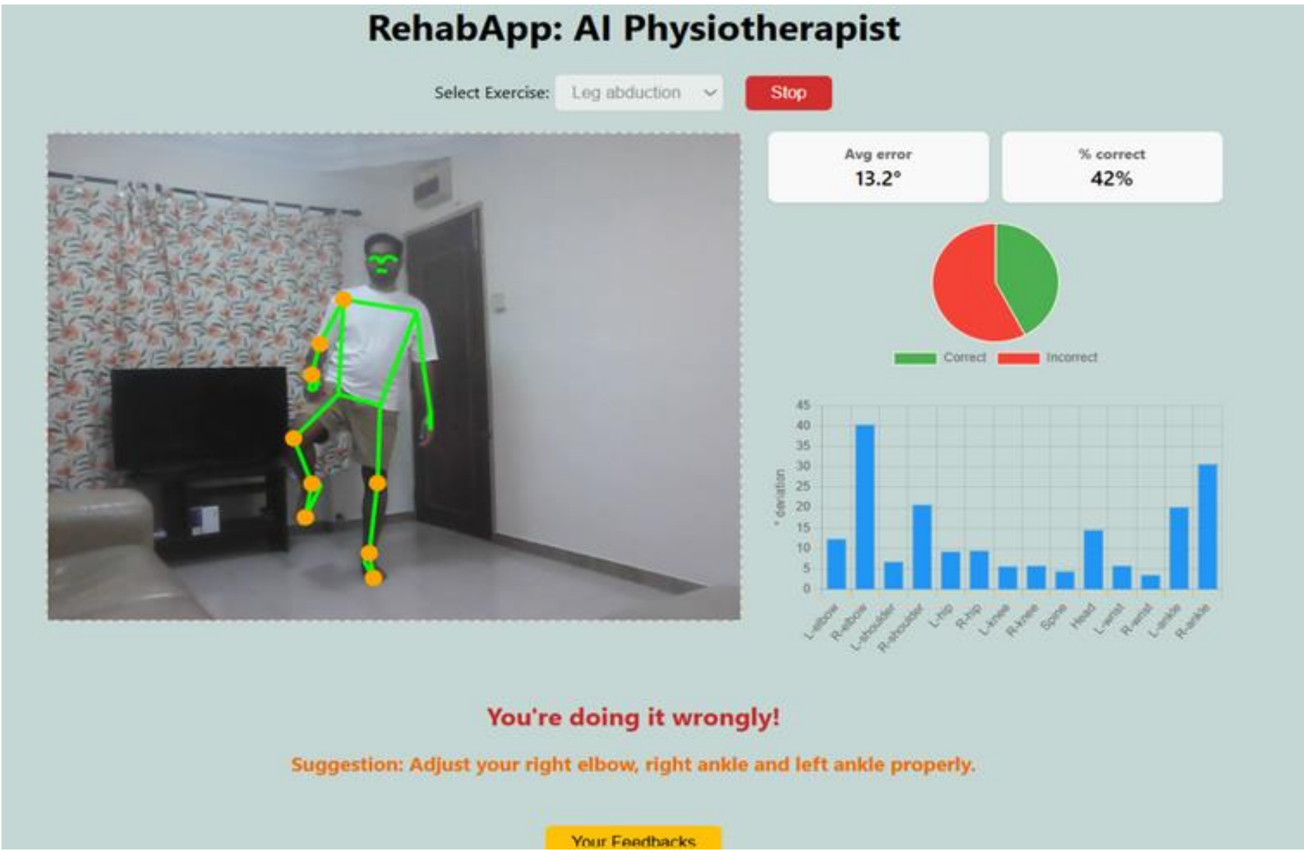
Real-time dashboard

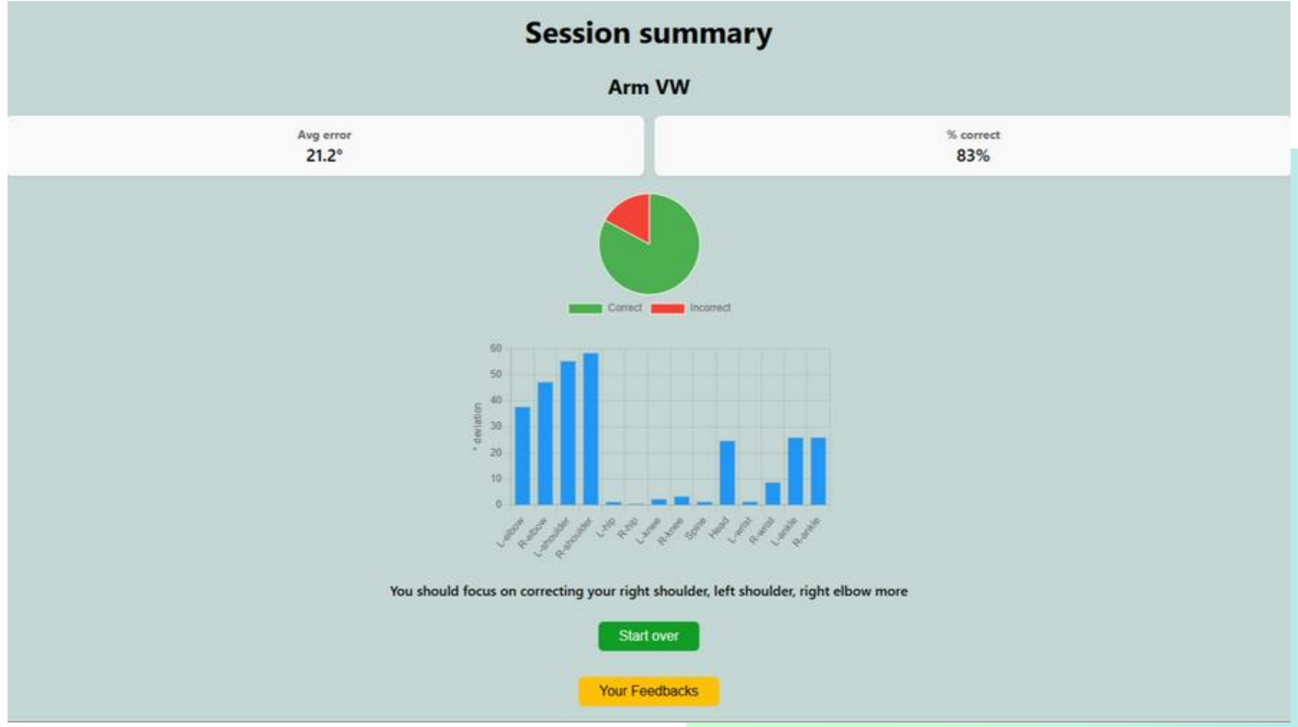
- Updates every window (≈ 0.5 s):
- Pie chart correct / wrong
- Numeric tiles mean angle error, % correct
- 14-bin histogram of joint errors with the three worst joints called out.



Session summary

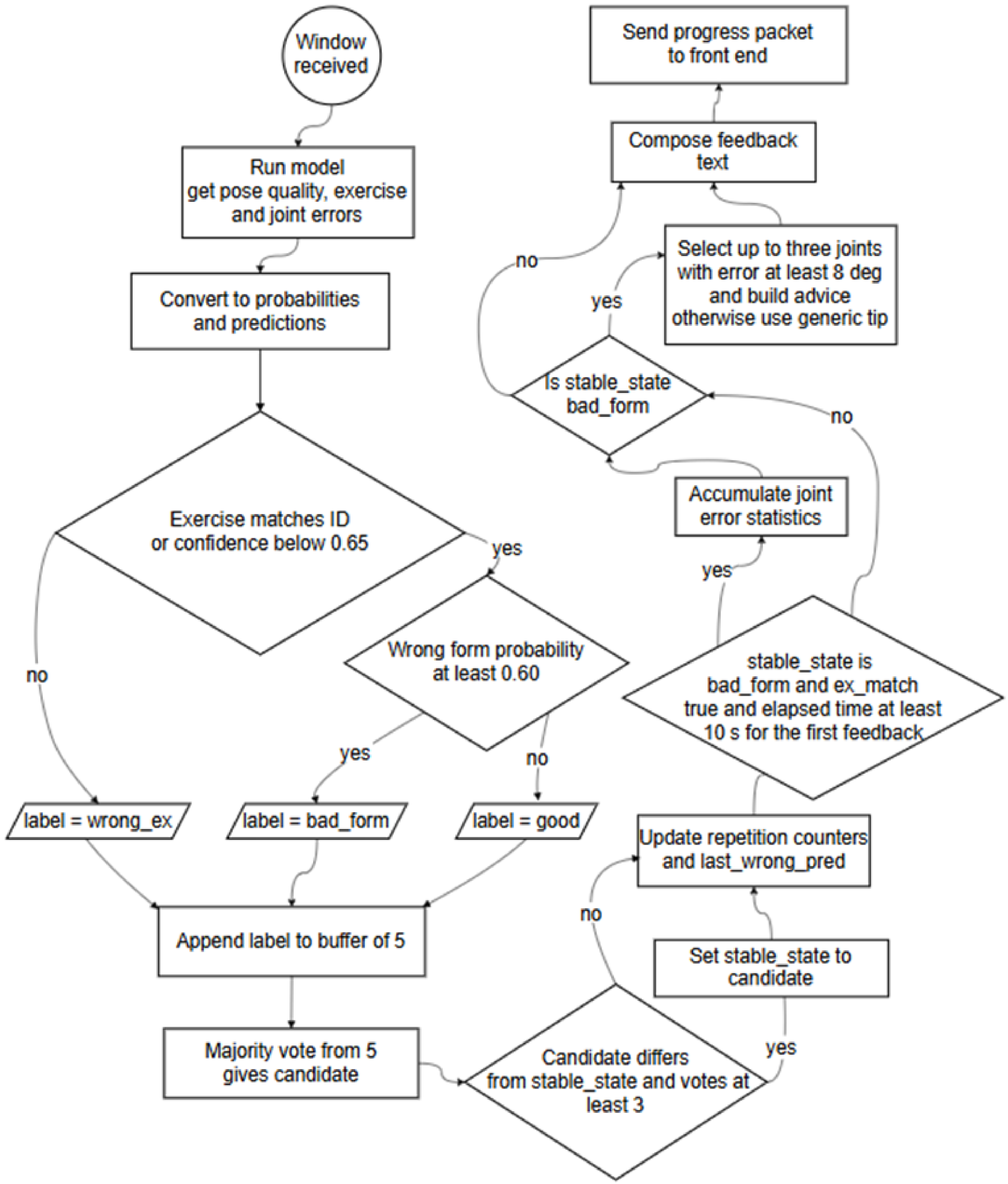
- Same charts from the real-time dashboard after the exercise end for the user to review and improve



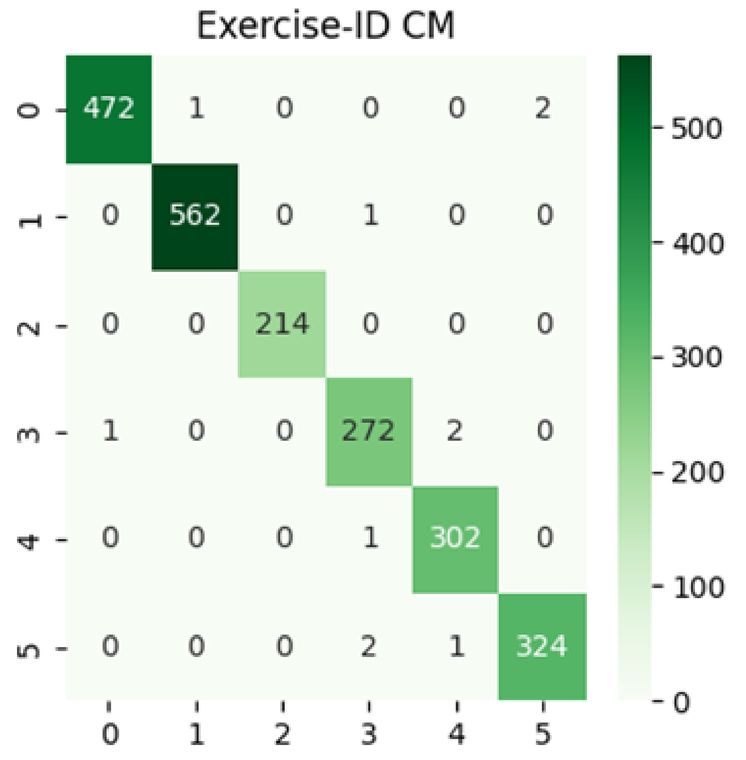


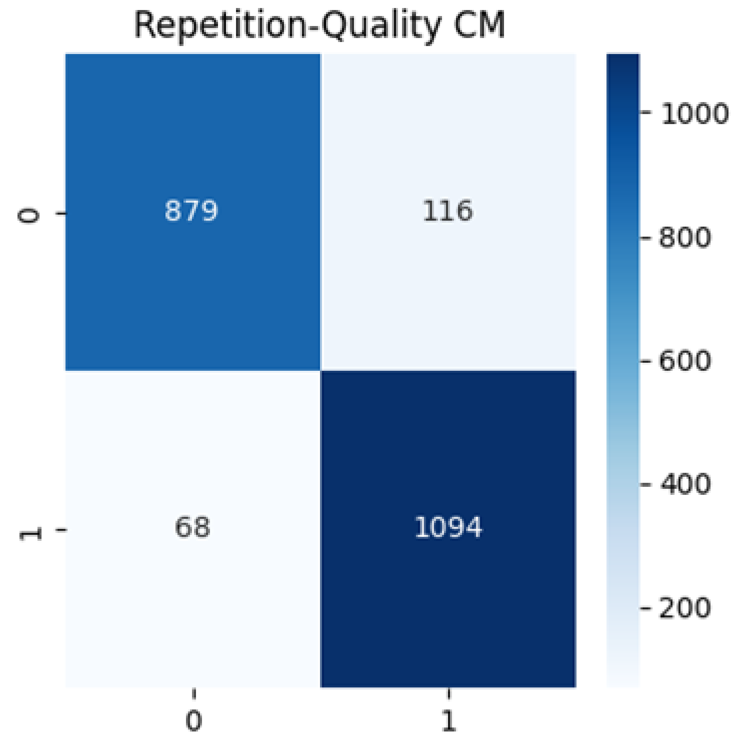
Inference: Feedback Logic

Scenario	Example sentences	How they are produced
Correct form	Feedback : “You’re on the right track!” Suggestion : —	1. Quality head ⇒ p(wrong) < 0.60 and exercise ID agrees with the drop-down → state = good. 2. All joint errors < 8° → corrections list is empty.
Bad form	Feedback : “You’re doing it wrongly!” Suggestion (example): “Adjust your left hip, spine and right ankle properly.”	1 p(wrong) ≥ 0.60 and exercise matches with selected → state = bad_form 2 Keep joints with |Δθ| ≥ 8°, sort by magnitude, take top 3 → e.g. {left-hip, spine, right-ankle}.
Wrong exercise	Feedback : “Wrong exercise! Looks like Push-ups.” Suggestion : —	1 Exercise head ≠ selected and conf ≥ 0.65 → state = wrong_ex 2 Skip joint-error logic entirely.



Results : Quantitative





Window-level performance (4112 test windows)

- Repetition quality: Acc 91.5 %, F1_w 0.915
- Exercise ID: Acc 99.5 %, F1_w 0.995
- Joint-angle error MAE: 4.73° across 14 DoF

Confusion–matrix take-aways

- Rep-quality: TP: 1094, TN: 879, FP: 116, FN: 68
- Exercise ID: only 5 mis–labels
- largest confusion Arm-VW ↔ Arm-Abduction
- 93 % of windows give clinically actionable (< 5°) angle errors

Efficiency & footprint

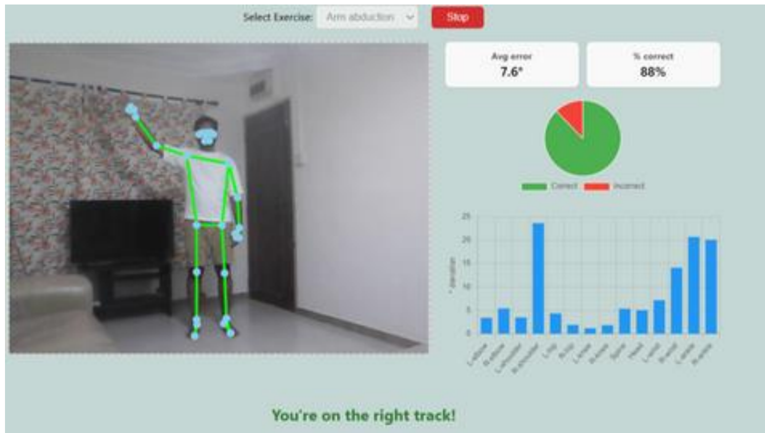
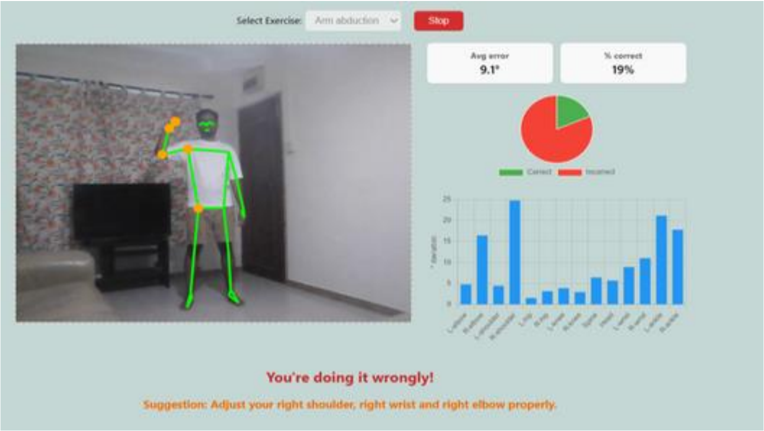
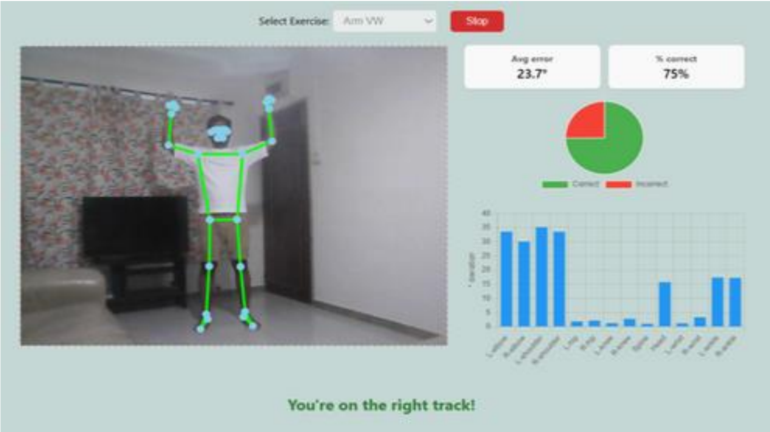
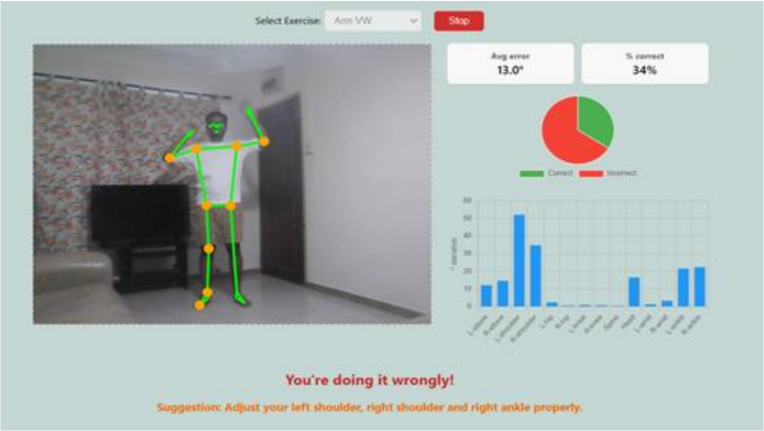
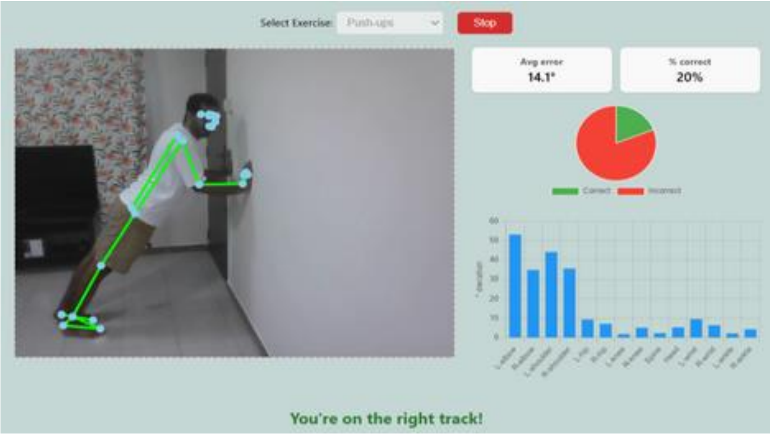
- 7500 windows s^{-1} ≈ 30 fps end-to-end on mid-range CPU
- 3.41M parameters (< 8MB), no GPU required
- Latency: ≈ 2.5 ms (browser → server → browser)

Efficiency & footprint

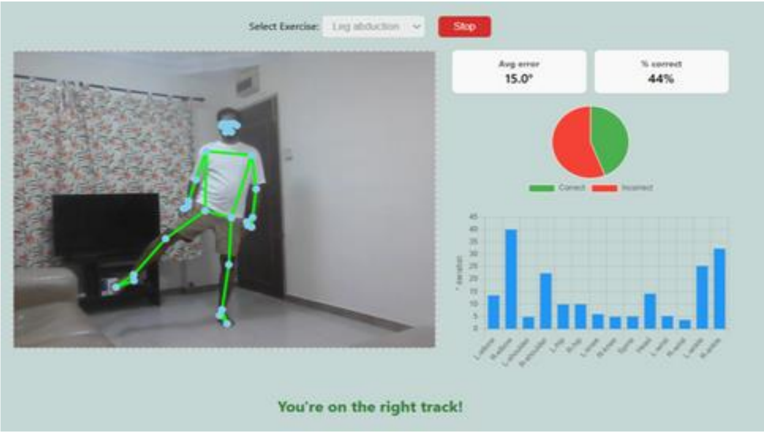
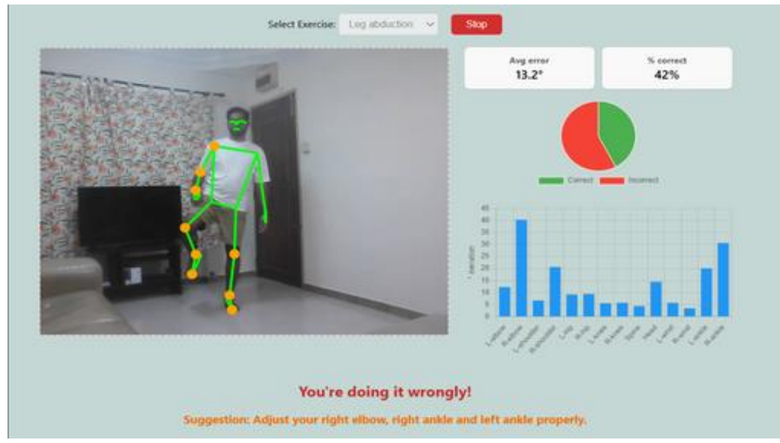
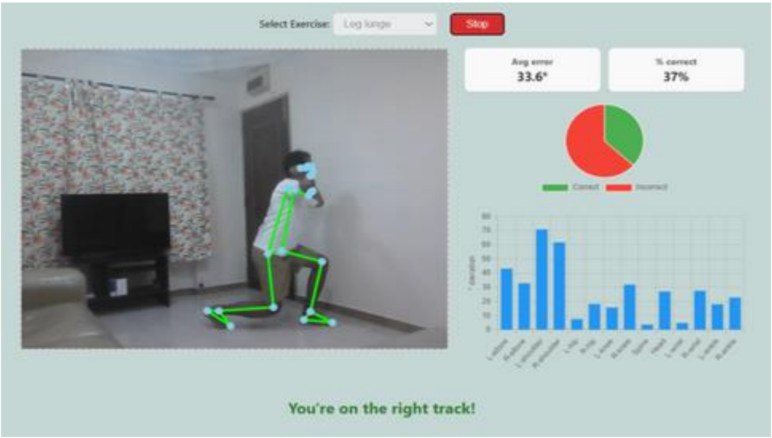
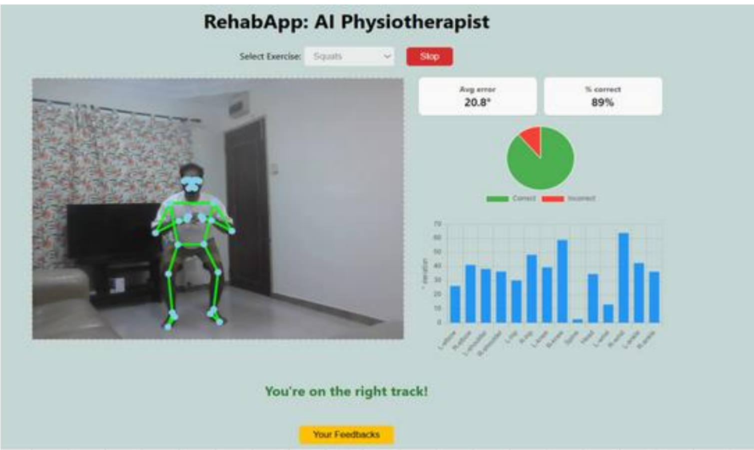
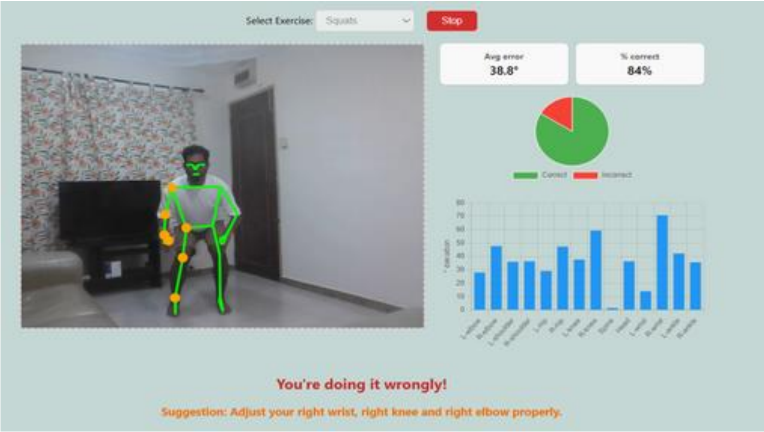
- Mobile-ready: real-time, < 10 MB, runs on-device
- Rich feedback: joint-level errors + wrong-exercise guard
- State-of-the-art accuracy on six diverse rehab drills

Head	Acc	F1_w	MAE (°)
Rep-quality	**0.915**	0.915	–
Exercise ID	**0.995**	0.995	–
Angle error	–	–	**4.73**

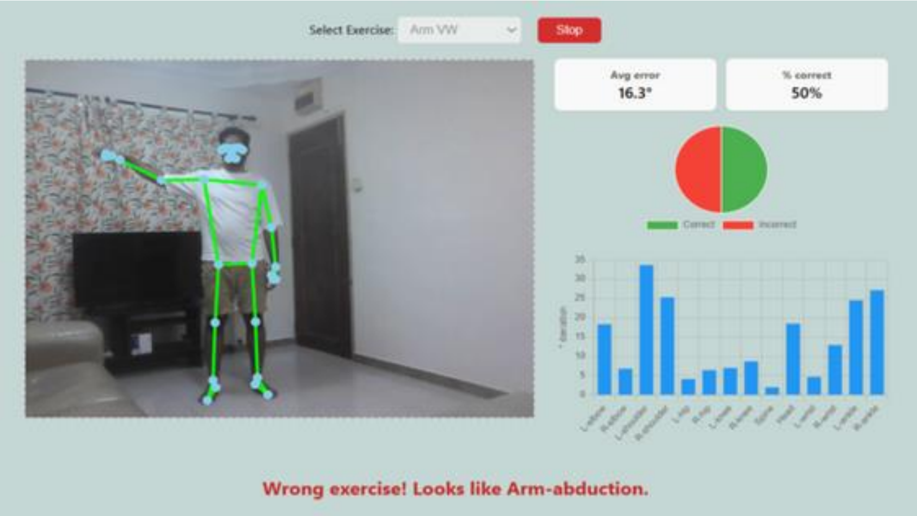
Results : Qualitative

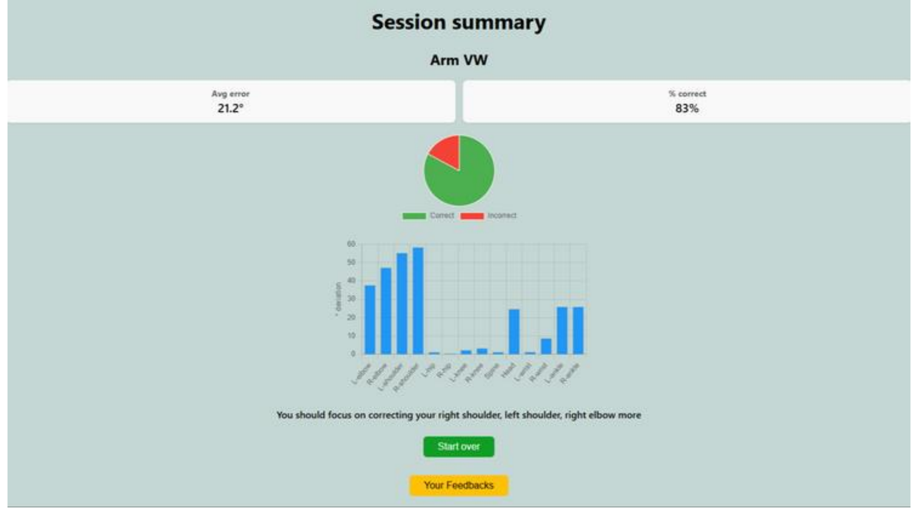
Excercise	Correct Form	Incorrect Form	Feedback/Suggestion
Ex1: Arm-abduction			**Correct form:** **"You're on the right track"** **Incorrect form :** **"You're doing it wrongly!"** **"Suggestion: Adjust your right shoulder, right wrist and right elbow properly"**
Ex2: Arm-VW			**Correct form:** **"You're on the right track"** **Incorrect form :** **"You're doing it wrongly!"** **"Suggestion: Adjust your left shoulder, right shoulder and right angle properly"**
Ex3: Push-ups			**Correct form:** **"You're on the right track"** **Incorrect form :** **"You're doing it wrongly!"** **"Suggestion: Adjust your left elbow, left shoulder and right shoulder properly"**

Results : Qualitative

Excercise	Correct Form	Incorrect Form	Feedback/Suggestion
Ex4: Leg-abduction			**Correct form:** “You’re on the right track” **Incorrect form :** “You’re doing it wrongly!” “Suggestion: Adjust your right elbow, right angle and left angle properly”
Ex5: Lunge			**Correct form:** “You’re on the right track” **Incorrect form :** “You’re doing it wrongly!” “Suggestion: Adjust your head, left shoulder and right angle properly”
Ex6: Squats			**Correct form:** “You’re on the right track” **Incorrect form :** “You’re doing it wrongly!” “Suggestion: Adjust your right wrist , right knee and right elbow properly”

Results : Qualitative

Safety demo – "Wrong exercise" guard		Feedback/Suggestion
Exercise Selected: Ex1: arm-abduction " Exercise Performed Ex2: arm-VW"		"Wrong exercise! Looks like arm-abduction " ✓

Session summary page		Advice
The pie chart, numeric tiles, and joint-error histogram from the session history, and an advice to guide the user		"You should focus on correcting your right shoulder, left shoulder, right elbow more"

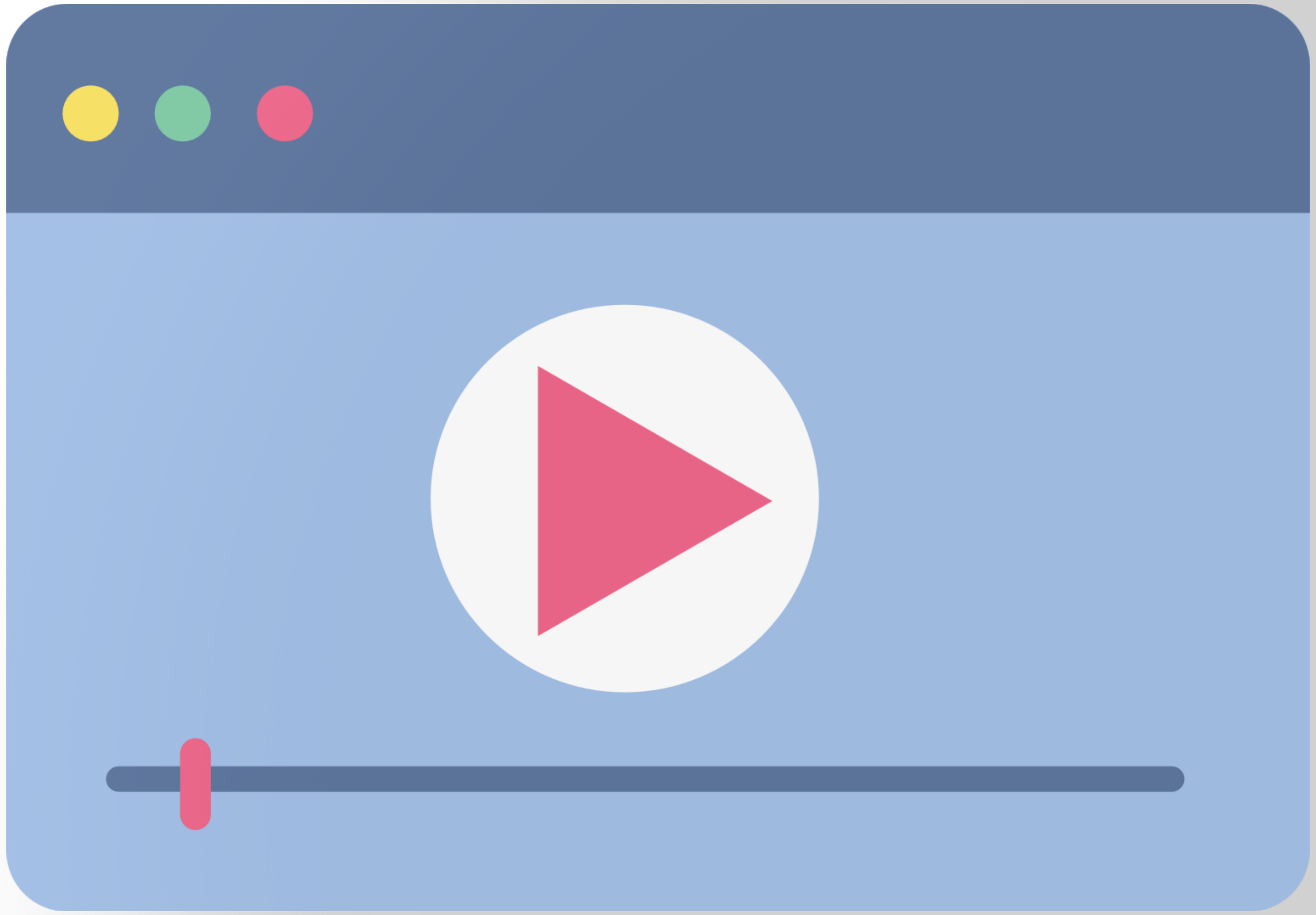
Ablation Study:
Architecture Variants

Variant	Added Block(s)	Rep-F1	Ex-F1	MAE (°)	Δ Rep-F1	Δ MAE	FPS	Params
A	**Baseline** 1-D CNN only	0.797	0.993	8.49	–	–	**9 160**	**0.25 M**
B	+ **Exercise embed** (6 → 64 MLP)	0.819	0.994	6.18	**↑ +2.2 pp**	**↓ -27 %**	9 126	0.25 M
C	+ **Bi-LSTM** (2 × 256)	0.853	0.996	6.12	**↑ +5.6 pp**	↓ -28 %	7 256	3.40 M
FULL	**B + C** (embed + Bi-LSTM)	**0.903**	**0.997**	**3.86**	**↑ +10.6 pp**	**↓ -55 %**	7 383	3.41 M

Which Blocks Really Pay Off?

- **Exercise context is cheap and valuable**
  - 0% extra params → +2pp Rep-F1 & –2.3° MAE.
- **Temporal reasoning matters**
  - Bi-LSTM alone adds +5.6 pp Rep-F1 despite a 13× size jump.
- **Synergy, not mere addition**
  - Combining both cuts angular error in half (8.49 → 3.86°) and yields the largest quality gain, yet model still fits in ~ 11 MB.
- **Real-time preserved**
  - Even the full model sustains ≈30 FPS on CPU (7 k+ on GPU), meeting the live-feedback requirement.

System: Demo



Limitations

Self-occlusion

When a limb is hidden (e.g. arms resting on thighs, crossed legs) MediaPipe drops landmarks, triggering spurious “bad-form” flags.

Camera-pose sensitivity

Accuracy degrades if the webcam is pitched or rolled by > ±20 ° or if lighting is uneven.

Scope constraints

Model was trained on 10 healthy young adults; it is untested on elderly, post-operative or high-BMI users (risk of domain shift).

Low-resolution extremities

Single-view, single-person, six drills, no on-device personalisation or multi-person support (yet).

Future Works

Robustness

- Add a lightweight self-supervised pre-text task to denoise landmarks under occlusion / harsh lighting.
- Fuse two camera views when available to recover hidden joints and improve 3-D angle accuracy.

Personalized Coaching

- 30-second calibration routine to learn each user's neutral joint baselines.
- Dynamic tolerance tightening as the patient improves, plus on-device few-shot fine-tuning.

Ultra-light Edge Deployment

- Quantise the model to 4-bit QAT ONNX and export as a cross-platform mobile SDK (Android / iOS / WebAssembly).
- Optimise CPU scheduling to keep real-time feedback while cutting battery drain.

Clinical Validation

- Run a 6-week field study with post-operative patients (n≈30) vs. standard care; measure adherence, recovery time, and user satisfaction.
- Collect qualitative therapist feedback to refine cue phrasing.

Feature Expansion

- Double the exercise library (e.g. balance drills, gait training) and introduce multi-person support.
- Add multi-language text-to-speech and haptic cues for broader accessibility.

References

[1] Faber, M., Andersen, M. H., Sevel, C., Thorborg, K., Bandholm, T., & Rathleff, M. The majority are not performing home-exercises correctly two weeks after their initial instruction—an assessor-blinded study. PeerJ, 3:e1102, 2015. DOI:10.7717/peerj.1102.

[2] A. Černek, J. Sedmidubsky, and P. Budíková, "REHAB24-6: Physical Therapy Dataset for Analyzing Pose Estimation Methods," in *Proc. 17th Int. Conf. on Similarity Search and Applications* (SISAP), LNCS 14512, pp. 18–33, Oct. 2024, doi: 10.1007/978-3-031-75823-2_2.


THANK
YOU