AI Personal Trainer System using Machine Learning based on Pose Landmark Detection and Natural Language Reasoning

Milzam Wafi Azhar
EECS International Graduate Program
National Yang Ming Chiao Tung University
Hsinchu, Taiwan
azharmilzam.ee11@nycu.edu.tw

Azam Khan
EECS International Graduate Program
National Yang Ming Chiao Tung University
Hsinchu, Taiwan
azamkhan5556.ee13@nycu.edu.tw

Sirapop Nuannimnoi
EECS International Graduate Program
National Yang Ming Chiao Tung University
Hsinchu, Taiwan
sirapop.film@outlook.com

Himmatur Rijal
EECS International Graduate Program
National Yang Ming Chiao Tung University
Hsinchu, Taiwan
rijal6jari@gmail.com

Chia-Hsin Lai
Electronics Engineering
National Yang Ming Chiao Tung University
Hsinchu, Taiwan
chiahsinlai9180@gmail.com

Ching-Yao Huang
Electronics Engineering
National Yang Ming Chiao Tung University
Hsinchu, Taiwan
cyhuang@nycu.edu.tw

***Abstract*—Holo-Wellness is a smartphone fitness coach that delivers frame-level form correction and bilingual coaching *entirely on-device*. MediaPipe first extracts 33 landmarks; joint angles and keypoints feed a classification models that labels each repetition as *correct* or *fault*. The performance achieves a macro-F_1 score more than $97\,\%$ on four exercises—biceps curl, plank, squat–stage, and lunge—annotated over 45 k frames.**

A 32-billion-parameter Qwen-QwQ model, paired with a 1 B translation bridge trained on 1.2 M parallel fitness sentences, lets users switch fluidly between Traditional Chinese and English. The chatbot answers health queries, performs few-shot pain triage, and composes workout plans that adapt to logged performance.

In a three-day trial with ten users, instant voice cues and code-switching were rated most helpful; participants asked for deeper long-term personalisation. We outline privacy safeguards, a cloud portal for model updates, and future work on multi-view depth fusion and on-device continual learning.

***Index Terms*—Fitness assistant system, pose landmark detection, natural language reasoning, chatbot**

I. Introduction

Evolving consumer habits and the rapid growth of on-demand, at-home workout programs have created unprecedented demand for real-time, personalized coaching. However, most commercial apps still fall back on scripted videos or simple repetition counters that neither detect subtle posture faults nor offer advice tailored to users' language, culture, or pre-existing musculoskeletal conditions.

Monocular pose-estimation frameworks such as MediaPipe MediaPipe now make it feasible to extract body landmarks with a single smartphone camera, and research prototypes have begun to classify exercise quality in controlled settings. Yet two critical gaps remain: a) responsiveness on commodity hardware: many systems rely on cloud inference or heavyweight models that cannot sustain real-time feedback on mid-range phones; and b) Conversational depth and cultural fit. Even when chatbots are added, they often struggle with domain-specific terminology or fail to reason about pain, schedules, and goals in the user's preferred language.

We present Holo-Wellness, a mobile fitness-assistant that integrates:

- Real-time movement analysis. MediaPipe delivers 33 landmarks within 30 FPS; derived joint-angle and key-points features feed a machine learning models that grades each repetition as correct or fault.
- Bilingual natural-language reasoning. A translation bridge fine-tuned on 1.2 M parallel fitness sentences lets the 32-billion-parameter Qwen QwQ model switch seamlessly between Traditional Chinese and English while preserving technical accuracy.
- Personalized feedback loop. Instant voice or text corrections appear on-device, while an encrypted cloud portal lets coaches update models, curate chatbot datasets, and review anonymized user reports.

The rest of this paper is organized as follows: Section II reviews the recent advances and background knowledge on several software components including pose landmark detection, RAG and natural language reasoning technologies; Section IV briefly introduces the datasets used for developing Holo-Wellness prototype, the key components of our application, and initial user study with target users in Taiwan; Section V discusses the results from the user study; and finally, Section VI concludes our paper and explores potential future work directions.

II. Related Work

A. Mobile Health & Fitness

The global m-health market reached USD 37.5 B in 2024 and is forecast to hit USD 86 B by 2030 (15 % CAGR) thanks to ¿ 6 B active smartphones and 1.1 B wearable devices [1], [4]. Contemporary apps integrate AI for medication reminders, fall-risk triage, and adaptive coaching; FitBot, for instance, tailors elderly strength regimens via ChatGPT [5]. Wearable–cloud loops also improve chronic-disease dashboards but expose attack surfaces: 75 % of audited Android m-health apps lacked TLS certificate pinning [10]. Metaverse pilots add immersive PT sessions and remote ward rounds, yet raise fresh identity-linkage concerns [2].

B. On-device Pose Estimation

Lightweight CNN/transformers have displaced GPU-heavy roots. MediaPipe Pose (region proposal + landmark head) attains 33 keypoints at 30 FPS on a 2017 Pixel 2 [11]; MoveNet Lightning ($< 3MB$) and Thunder (6 MB) reach 300 and 120 fps, respectively, on a single A55 core [12]. For broader anatomy, AlphaPose adds hands/face (133 pts) while preserving 15 FPS on RTX-2070 [17]; ST-LineNet layers temporal kernels to capture martial-arts kicks with 3-D error $< 35mm$ [21]. OpenCap triangulates full-body kinematics from two consumer iPhones and yields joint-moment estimates within 8 % of lab-grade Vicon [19]. Clinical safety use-cases already run fully on-device for fall detection on low-end phones [23].

C. Retrieval-Augmented Generation (RAG)

RAG prepends retrieved passages to an LLM prompt, thereby reducing hallucination while keeping model size constant [24]. Standard pipeline: chunk corpus $\rightarrow$ vector-embed (FAISS) $\rightarrow$ top-k retrieve $\rightarrow$ optional COLBERT rerank $\rightarrow$ answer synthesis. Doshi *et al.* reported a 12-point EM jump on Natural-Questions after tuning only the retriever [25]. In medicine, MIRAGE demonstrates 18 % higher strict-accuracy over vanilla GPT-3 via PubMed retrieval [27]; GastroBot combines Chinese guidelines with llama-index to halve wrong-dose recommendations [28]. Cross-lingual work adds self-verification (CLARA) or query-by-example embeddings to level performance across 7 languages [30], [31].

D. Large-Language-Model Reasoning

Early chatbots were FAQ lists; modern LLM's ($> 7Bparams$) can plan progressive workouts or flag red-flag symptoms. HealthQ couples chain-of-thought prompts with RAG, lifting "clinically useful" follow-up questions from 54 % to 79 % per GPT-4 adjudication [32]. GPTCoach streams Apple-Watch metrics into a mobile LLM, improving weekly activity minutes by 18 % in a four-week crossover trial [33]. Multilingual fine-tuning scales reach: MMed-Llama3 (8 B) digests a 25.5 B-token med-corpus and matches GPT-4 on six-language HealthQA (75 vs 76 F_1) [35]. Prompt engineering—e.g. few-shot+CoT—can raise diagnostic F_1 a further 5-10 points while surfacing reasoning traces, crucial for clinician trust [37].

III. Realtime Feedback System design

Fig. 1 illustrates the life-cycle of an exercise classifier in three stages. First, raw *video frames* are streamed from the phone's camera and processed by a pose-estimation module, which yields 33 landmark coordinates per frame. These landmarks are converted into joint angles, visibility scores, and first-order velocities (x, y, z, v, θ), then normalised to build a compact feature tensor.

Second, the dashed green path shows how the same normalised features feed an *offline data-experiment loop*: labelled clips are partitioned into training and test sets, multiple models are explored, and the best network is exported as a lightweight on-device classifier. Each repetition is ultimately assigned to one of n possible fault categories (or *correct*).

Finally, the right-hand lane depicts the *deployment* stage. During a workout the classifier runs on the device; its decisions trigger immediate visual and auditory cues ("Feet too narrow", "Loose upper arm"). These low-latency prompts travel directly from the classifier to the text-to-speech (TTS) engine, while a richer language-model–driven coach delivers longer-form feedback and session summaries from the cloud.

IV. Datasets and Methods

A. Exercise-Pattern Data

a) Exercise roster and labeled faults.: The dataset covers *biceps curl, basic plank, basic squat* and *lunge*. Each exercise is annotated with the fault taxonomy listed in Table I.

TABLE I: Exercises and fault classes

Exercise	Fault classes
Biceps curl	*Loose upper arm*, *Weak peak contraction*, *Lean–back torso*
Plank	*Lower-back sag*, *High back (pike)*
Squat	*Feet too narrow/wide*, *Knees too narrow /wide*
Lunge	*Wrong knee angle*, *Knee over toe*

b) Collection protocol.: One contributors recorded every **proper** demonstration and each **error** variant from two camera angles. Static moves (plank) were recorded for at least 30 s; dynamic moves contained ≥ 15 repetitions per clip. All videos used uniform lighting, and the full body remained inside the frame.

c) Frame-level annotation.: MediaPipe extracted 33 landmarks per frame. For each exercise, we retained only the relevant subset (e.g., 17 landmarks for plank, 13 for curl) and exported the data as {`landmark_x, y, z, v`} together with a categorical `fault_label`.

d) Feature Extraction.: For every three-point combination P, Q, R that forms an articulated segment (e.g. hip–knee–ankle or shoulder–elbow–wrist) we calculate the interior angle at the middle joint Q using Algorithm 1. The procedure first constructs vectors $\mathbf{u} = QP$ and $\mathbf{v} = QR$, derives their Euclidean norms and dot product, and then

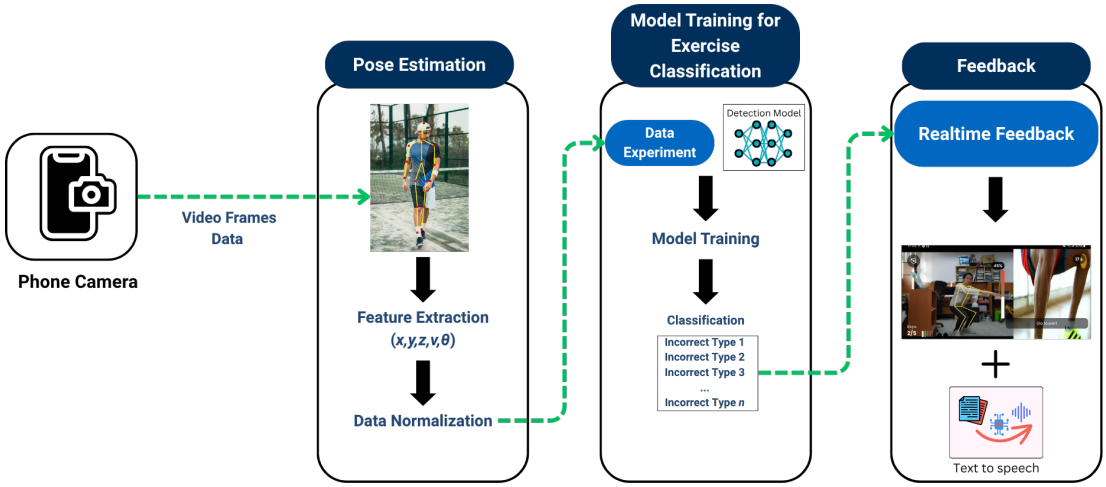

Fig. 1: Realtime Feedback System design.

TABLE II: Size of the final training splits

Exercise	Proper frames	Error frames	Total frames	Distinct fault labels	Important landmarks
Biceps curl	8 238	7 134	15 372	3	9 (shoulders–hips–elbows–wrists–nose)
Plank	9 630	18 993	28 623	2	17 (full body)
Squat	188 (up)	186 (down)	374†	2 (stage)	8 (shoulders–hips–knees–ankles)
Lunge	545	531	1 076	2	15 (lower-body focus)

†Squat frames are further split into *up* and *down* stages; fault detection is performed stage-conditionally.

recovers the angle φ via the inverse–cosine rule. The result is expressed in degrees and normalised to the $[0°, 180°]$ range by mapping any reflex value ($\varphi > 180°$) to its complementary angle ($360° - \varphi$). This scalar angle—together with landmark joint angle scores and first-order joint velocities—constitutes the feature vector that feeds the downstream fault-classification models.

Algorithm 1 Joint–angle computation at joint Q

Require: $P(p_x, p_y, p_z)$, $Q(q_x, q_y, q_z)$, $R(r_x, r_y, r_z)$
Ensure: φ_{deg}

$\mathbf{u} \leftarrow (p_x - q_x,\ p_y - q_y,\ p_z - q_z)$
$\mathbf{v} \leftarrow (r_x - q_x,\ r_y - q_y,\ r_z - q_z)$
$\|\mathbf{u}\| \leftarrow \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2 + (p_z - q_z)^2}$
$\|\mathbf{v}\| \leftarrow \sqrt{(r_x - q_x)^2 + (r_y - q_y)^2 + (r_z - q_z)^2}$
$\cos\varphi \leftarrow \dfrac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|\,\|\mathbf{v}\|}$
$\varphi_{\text{rad}} \leftarrow \arccos(\cos\varphi)$
$\varphi_{\text{deg}} \leftarrow |\varphi_{\text{rad}} \times 180/\pi|$
if $\varphi_{\text{deg}} > 180$ **then**
  $\varphi_{\text{deg}} \leftarrow 360 - \varphi_{\text{deg}}$
end if
return φ_{deg}

e) Public Dataset.: Because plank shows little intra-video motion, we merged 30 carefully screened images from Kaggle's "Yoga Poses" set into the proper-plank class to increase visual diversity.

f) Train/Validation Split.: Each exercise CSV was shuffled and divided 80 % / 20 % (stratified by label) before model training.

B. Evaluation Metrics

We assess classification quality with the standard, class-balanced metrics below.

- **Confusion matrix**. A 2×2 table of *true positives* (TP), *false positives* (FP), *true negatives* (TN) and *false negatives* (FN). A well-behaved classifier has high TP and TN, low FP and FN.
- **Precision** (Prec). Fraction of predicted positives that are correct:
$$\text{Prec} = \frac{\text{TP}}{\text{TP} + \text{FP}}.$$
- **Recall / Sensitivity** (Rec). Fraction of actual positives that are recovered:
$$\text{Rec} = \frac{\text{TP}}{\text{TP} + \text{FN}}.$$
- **F_1 score**. Harmonic mean of precision and recall, maximized when both are high:
$$\text{F}_1 = \frac{2\,\text{Prec}\,\text{Rec}}{\text{Prec} + \text{Rec}}.$$

C. RAG Data

Our Retrieval-Augmented Generation (RAG) pipeline (Fig. 2) has six main stages:

1) **Multimodal ingestion & preprocessing.**

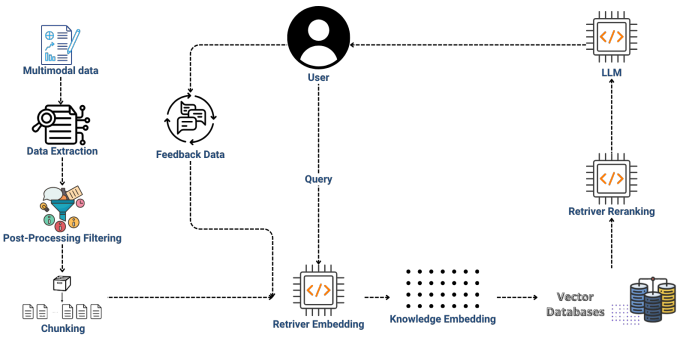

Fig. 2: RAG pipeline.

- Source documents (e.g. bilingual health guides, exercise tutorials, pain-diagnosis references) are loaded.
- Text is *chunked* into passages and run through *post-processing filters* (e.g. deduplication, length control).

2) **Vectorization.**
   - A retriever-embedding model encodes each chunk into a dense vector.
   - Vectors (with metadata) are stored in a *vector database*.
3) **Query handling.**
   - User input passes through an *input guardrail* and, if needed, a query-rewriter.
   - The sanitized query is encoded into a vector.
4) **Retrieval & reranking.**
   - The vector database returns the top-K nearest chunks.
   - A reranker refines their order based on relevance and policy.
5) **Generation & output filtering.**
   - The LLM consumes the top passages to generate a response.
   - An *output guardrail* vets the draft for safety and factuality.
6) **Feedback & observability.**
   - User feedback and system logs are captured.
   - All documents, embeddings, histories, and metrics are persisted for continuous improvement.

To enable seamless Chinese–English code-switching we prepend or post-process through a small (1 B-param) translation model.

D. Application UI Design

The Holo-Wellness application is designed to deliver a smooth and responsive experience on mid-range smartphones, emphasizing clarity, cultural accessibility, and real-time interactivity. The UI architecture is modular and layered, separating core functions such as camera tracking, chatbot dialog, and feedback visualization to reduce cognitive load during active workouts.

E. Application User Interface Overview

Fig. 3 illustrates the three main screens of the Holo-Wellness mobile app:

- **Real-time detection (Fig. 3.a).** A full-screen live camera feed is overlaid with pose landmarks, repetition counters, and visual error highlights. Simultaneous audio prompts are synthesized on-device via a VITS-Tiny TTS model and delivered in the user's preferred language to guide form corrections in real time.
- **AI-Generated Personalized Workout Plan (Fig. 3.b).** After logging performance, the system proposes a tailored regimen. Each card shows the focus area (e.g., "Build strength"), difficulty level, recommended frequency, and duration. Users may swipe through or tap "View all" for more options.
- **Chatbot (Fig. 3.c).** The HoloDiagnose panel lets users upload an image of their posture or pain area. The conversational interface then gathers symptom details, reasons about possible causes via few-shot prompting, and synthesizes recommendations—while reminding users to seek professional care as needed.

F. User Study

To evaluate the usability and perceived value of the Holo-Wellness prototype, we conducted a qualitative user study involving 10 participants representative of our target demographic. These individuals, recruited through local fitness clubs, included both native Taiwanese and international residents (5 Taiwanese and 5 international participants), ensuring a balanced mix of linguistic and cultural backgrounds. We adopted a semi-structured interview format to explore user experiences in depth while allowing flexibility to probe emerging themes. Each participant interacted with the Holo-Wellness system over a 3-day period, completing at least two different exercise routines guided by the AI coach.

Following the trial sessions, we conducted individual interviews lasting 10–15 minutes. The interview protocol covered themes such as perceived accuracy of motion feedback, naturalness and helpfulness of chatbot responses, language-switching fluency, and trust in health-related suggestions. Interviews were audio-recorded and transcribed, then analyzed using thematic coding to identify recurring sentiments and usability issues.

The questions for the semi-structured interviews are listed as follows:

- How did you feel when first interacting with the Holo-Wellness system?
- Was the system easy to set up and start using? Why or why not?
- How would you describe your overall experience with the exercise guidance?
- Did you find the real-time movement corrections helpful or distracting?
- How accurate did the system feel in identifying your posture or errors?
- What kinds of questions did you ask the chatbot, and how well did it respond?
- How natural or artificial did the language feel during conversations?

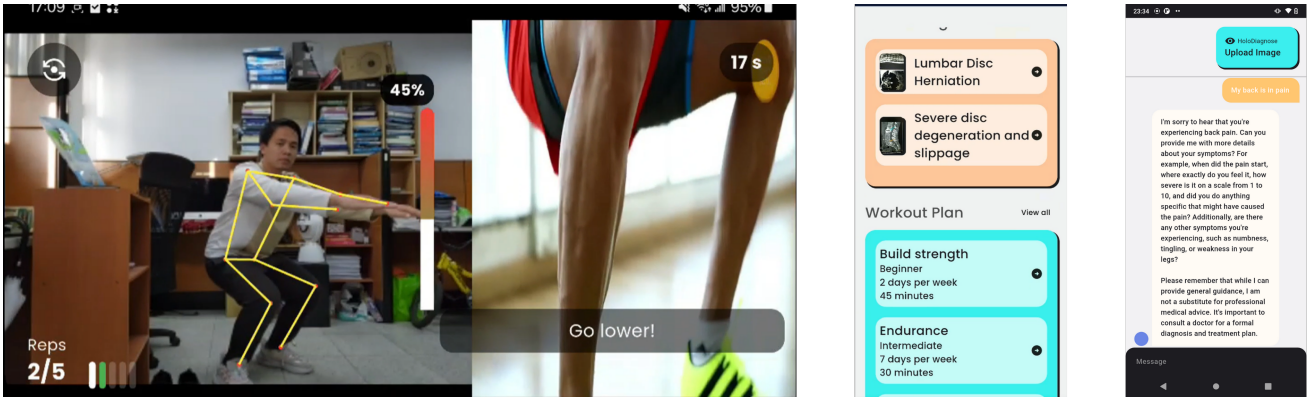

Fig. 3: Application User Interfaces: (a) real-time detection; (b) AI-Generated Personalized Workout Plan; (c) Multimodal chatbot.

- Did you feel the guidance and suggestions were relevant to your fitness habits or goals?
- How much do you trust the chatbot's advice, especially around pain or injury?
- Is there anything that would help build your trust in its health knowledge?
- What features did you like the most?
- What aspects of the system need improvement?
- Is there anything else you'd like to share about your experience?

V. Results and Discussion

In this section, we discuss the accuracies of motion capturing and analysis AI models and the results of overall user experience with the Holo-Wellness prototype.

A. Biceps Curl

1) Error Detection Methods:

a) Loose upper arm.: Compute the angle between the elbow, shoulder, and the projection of the shoulder onto the ground. If this angle exceeds $40°$, the repetition is classified as a "loose upper arm" error.

b) Weak peak contraction.: Compute the angle between the wrist, elbow, and shoulder at the top of the curl. If this angle exceeds $60°$ before the arm descends, the repetition is classified as a "weak peak contraction" error.

c) Lean too far back.: Because of its complexity, a lightweight machine learning classifier was trained on video frames labeled "correct" vs. "lean back."

Table III summarizes the precision, recall and accuracy of all models evaluated on this task.

TABLE III: Model training experiments for Biceps Curl error detection

Model	NN	Precision	Recall	Accuracy
KNN	No	0.975	0.968	0.972
7-layer MLP	Yes	0.972	0.962	0.967
5-layer MLP	Yes	0.963	0.949	0.955
SVC	No	0.930	0.934	0.932
Random Forest	No	0.947	0.925	0.931
7-layer + dropout	Yes	0.936	0.924	0.930
3-layer MLP	Yes	0.939	0.920	0.929
Logistic Regression	No	0.792	0.738	0.762
SGD Classifier	No	0.712	0.715	0.715

B. Basic Plank

1) Technique and Errors: The plank engages core muscles by holding the body straight on forearms and toes. Two common faults are targeted:

- *High lower back* (pike): torso raised too high.
- *Low lower back* (sag): hips dropped too low.

Important landmarks: nose, shoulders, elbows, wrists, hips, knees, ankles, heels, foot indices.

2) Error Detection Methods: A classifier was trained on 28,623 frames: 9,630 (33.6%) correct (C), 8,982 (31.4%) high back (H), 10,011 (35.0%) low back (L).

TABLE IV: Model training experiments for Plank error detection

Model	NN	Precision	Recall	Accuracy
Logistic Regression	No	0.996	0.996	0.996
7-layer + dropout	Yes	0.994	0.994	0.994
SVC	No	0.987	0.987	0.987
SGD Classifier	No	0.981	0.981	0.981
KNN	No	0.955	0.949	0.949
5-layer MLP	Yes	0.934	0.930	0.930
7-layer MLP	Yes	0.935	0.924	0.924
Random Forest	No	0.922	0.899	0.899
3-layer MLP	Yes	0.869	0.848	0.848

C. Basic Squat

1) Technique and Errors: A squat is a compound, lower-body movement in which the trainee flexes the hip and knee

joints to descend and then extends them to return to standing. The two most frequent faults we target are

- **Feet placement** — a stance that is either too narrow or too wide relative to shoulder width;
- **Knee placement** — knees that collapse inward or flare excessively outward with respect to the feet.

The detection pipeline relies on the shoulder, hip, knee, and ankle landmarks.

2) Stage Detection: Before fault analysis, we identify the movement phase. A lightweight classifier was trained to distinguish the *eccentric* (*down*) and *concentric* (*up*) stages on a balanced set of 374 annotated frames (188 *up*, 186 *down*).

TABLE V: Model training experiments for Squat stage detection

Model	NN	Precision	Recall	Accuracy
Logistic Regression	No	0.994	0.994	0.994
5-layer MLP	Yes	0.994	0.993	0.993
SGD Classifier	No	0.993	0.993	0.993
3-layer MLP	Yes	0.990	0.992	0.992
KNN	No	0.985	0.985	0.985
7-layer MLP	Yes	0.985	0.982	0.982
7-layer+dropout	Yes	0.981	0.985	0.985
SVC	No	0.978	0.977	0.977
Random Forest	No	0.254	0.504	0.504

3) Error-Detection Methods:

a) Feet placement.: Compute the ratio

$$x = \frac{\text{feet width}}{\text{shoulder width}}.$$

Using 851 reference samples (**Fig. 4**) we assign

$$\text{label}(x) = \begin{cases} \text{too narrow}, & x < 1.2, \\ \text{correct}, & 1.2 \le x \le 2.8, \\ \text{too wide}, & x > 2.8. \end{cases}$$

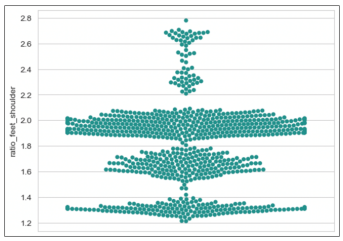

Fig. 4: Distribution of feet-to-shoulder width ratio in correct squat.

b) Knee placement.: Compute $x = \frac{\text{knee width}}{\text{feet width}}$ and evaluate per stage (down, middle, up) against thresholds shown in **Fig. 5**.

D. Lunge

1) Technique and Errors: A lunge places one leg forward and bends the knee. Targeted errors:

- *Knee angle*: should be $\approx 90^\circ$ at bottom.

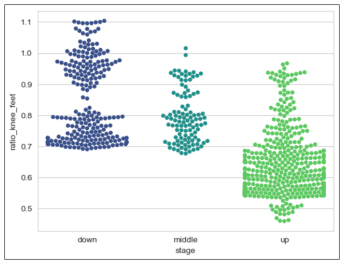

Fig. 5: Knee-to-feet width ratio across squat stages.

- *Knee over toe*: forward knee passes the toes.

Key landmarks: nose, shoulders, hips, knees, ankles, heels, foot indices.

2) Error Detection Methods:

a) Knee angle.: From 9,906 correct-form datapoints, knee angles during the low position lie in $[60^\circ, 135^\circ]$. See **Fig. 6**.

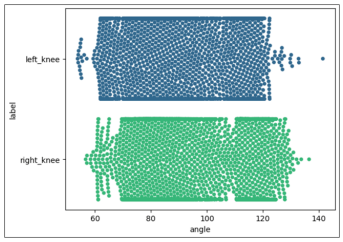

Fig. 6: Distribution of left/right knee angles in correct lunge.

b) Knee over toe.: A classifier was trained on 17,907 frames: 8,793 correct (49.1%), 9,114 knee-over-toe (50.9%).

TABLE VI: Model training experiments for Lunge error detection

Model	NN	Precision	Recall	Accuracy
Logistic Regression	No	0.973	0.972	0.972
SGD Classifier	No	0.961	0.956	0.956
3-layer MLP	Yes	0.937	0.928	0.928
7-layer+dropout	Yes	0.892	0.865	0.865
Random Forest	No	0.855	0.842	0.842
5-layer MLP	Yes	0.861	0.831	0.831
KNN	No	0.768	0.765	0.765
7-layer MLP	Yes	0.838	0.773	0.773
SVC	No	0.752	0.752	0.720

E. Model Summary

Table VII shows A simple logistic-regression baseline are the top three of the four tasks each above 99, while only the biceps-curl "lean-back" fault benefits from a nearest neighbour model.

Table VIII Interviews reveal that real-time feedback and seamless CN/EN switching drive engagement, yet users still seek some fault-detection, clearer clinical boundaries, and longer-term personalization.

TABLE VII: Top model and metrics per exercise

Exercise	Model	Prec.	Rec.	F_1	Acc.
Biceps curl (lean-back)	KNN	0.976	0.968	0.971	0.972
Plank (high/low back)	Log. reg.	0.996	0.996	0.996	0.996
Squat stage (up/down)	Log. reg.	0.994	0.994	0.995	0.994
Lunge (knee-over-toe)	Log. reg.	0.973	0.972	0.972	0.972

TABLE VIII: Key themes from semi-structured interviews

Theme	Codes	Illustrative quote
Feedback accuracy	Helpful, occasional mismatch	"It felt like having a coach beside me… but once or twice it said my form was off when I was careful."
Bilingual engagement	CN/EN switching, slang	"I typed in Chinese and English, even slang, and it still understood me."
Cautious trust	Safety boundary, seek pro advice	"I liked the tips, but I wasn't sure about pain advice without a real trainer."
Need for personalisation	Remember history, adapt goals	"It would be great if it remembered what I did yesterday or asked about my goals."

VI. Conclusion and Future Work

This paper introduced Holo-Wellness, a smartphone-based fitness assistant that fuses (1) on-device pose-landmark detection for real-time movement grading, (2) a bilingual large-language-model dialog layer with retrieval-augmented grounding, and (3) a cloud portal for continuous model and content management.

In the future, there are several research and development directions for this Holo Wellness project. First direction is the incorporation of stereo depth or multi-view fusion, which could potentially extend feedback to partner workouts and complex sports drills. We also plan to explore federated or on-device continual learning so classifiers personalize to each user's biomechanics without exporting raw video. Moreover, extending the RAG corpus and translation layer to additional East-Asian and South East Asian languages (e.g., Japanese, Korean, Thai, Bahasa Indonesia) and dialectal Chinese can broaden accessibility. Last but not least, as functionality edges toward clinical assessment, aligning with Taiwan TFDA and international MDR guidelines will be essential.

Acknowledgment

We would like to extend our sincere gratitude to a Taiwanese startup "Holo Wellness" and their dedicated team for their invaluable support and collaboration in our research. The insights and expertise provided by the team (Website: https://www.holowellness.net/) have significantly contributed to the development and success of this work. This work would not have been possible without the collective effort and commitment of all parties involved.

References

[1] M. Arbaoui, M.-E. Brahmia, and A. Rahmoun, "A review of IoT architectures in smart healthcare applications," Dec. 2022.

[2] A. Athar, Shah Mahsoom Ali, A. Islam, S. Ali, and H.-C. Kim, "Applications and Possible Challenges of Healthcare Metaverse," Feb. 2023.

[3] S. Oniani, G. Marques, Ivan Miguel Pires, S. Muhkashavria, and N. M. Garcia, "E-health and M-health applications in Georgia: A review on the free available applications for Android Devices," Dec. 2020.

[4] R. Talmale and S. Sonwane, "Mobile Healthcare Applications and Platforms," pp. 293–295, Jan. 2024.

[5] Y. M. Yee, T.-N. Li, Y.-H. Fu, H. L. Olinger, and W.-S. Chiu, "FitBot: A ChatGPT Mobile Application-Based Fitness Tracker for Elderly Users," pp. 301–302, Jul. 2024.

[6] S. N. Thakur, A. Sinha, M. K. Singh, M. K. Bagaria, R. Grover, and K. Shrivastava, "Optimizing Wellness: A Comprehensive Examination of a Conversational AI-Driven Healthcare BOT for Personalized Fitness Guidance," Dec. 2023.

[7] S. Goswami and A. Dubey, "Fitness Industry Propelling on IoT," Dec. 2022.

[8] V. Bhatt and S. Chakraborty, "Real-time healthcare monitoring using smart systems: A step towards healthcare service orchestration Smart systems for futuristic healthcare," IEEE Xplore, Mar. 01, 2021.

[9] R. Bokde and P. Dongare, "Evolution of Wearable Healthcare Technology: Opportunities, Challenges and Applications," pp. 1752–1755, Feb. 2025.

[10] S. Anwar, D. Anwar, and S. Abdulla, "Security Evaluation of Android Mobile Healthcare and Fitness Applications," 2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), pp. 1–6, Jun. 2020.

[11] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. L. Zhu, F. Zhang, and M. Grundmann, "MediaPipe: On-device Real-time Body Pose tracking", ArXiv, vol. abs/2006.10204, 2020.

[12] R. Bajpai and D. Joshi, "MoveNet: A Deep Neural Network for Joint Profile Prediction Across Variable Walking Speeds and Slopes," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-11, 2021, Art no. 2508511, doi: 10.1109/TIM.2021.3073720.

[13] Z. Cao, G. Hidalgo, T. Simon, S. -E. Wei and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 1, pp. 172-186, 1 Jan. 2021, doi: 10.1109/TPAMI.2019.2929257.

[14] K. Duan, S. Bai, L. Xie, H. Qi, Q. Huang, and Q. Tian, "CenterNet: Keypoint Triplets for Object Detection", in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019.

[15] M. Sandler, A. G. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks", 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4510–4520, 2018.

[16] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection", arXiv [cs.CV]. 2017.

[17] H.-S. Fang et al., "AlphaPose: Whole-Body Regional Multi-Person Pose Estimation and Tracking in Real-Time", IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 6, pp. 7157–7173, Jun. 2023.

[18] S. Choi, S. Choi and C. Kim, "MobileHumanPose: Toward real-time 3D human pose estimation in mobile devices," 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville, TN, USA, 2021, pp. 2328-2338, doi: 10.1109/CVPRW53098.2021.00265.

[19] S. D. Uhlrich et al., "OpenCap: Human movement dynamics from smartphone videos", PLOS Computational Biology, vol. 19, no. 10, pp. 1–26, 10 2023.

[20] J. R. Terven and D. M. Córdova-Esparza, "KinZ an Azure Kinect toolkit for Python and Matlab", Science of Computer Programming, vol. 211, p. 102702, 2021.

[21] Y. Sun, R. Deng, and D. Wei, "ST-LineNet: A spatiotemporal network for real-time 3D Pose estimation in martial arts training", Alexandria Engineering Journal, vol. 117, pp. 136–147, 2025.

[22] T. Jiang, X. Xie, and Y. Li, "RTMW: Real-Time Multi-Person 2D and 3D Whole-body Pose Estimation", ArXiv, vol. abs/2407.08634, 2024.

[23] E. Alam, A. Sufian, P. Dutta, and M. Leo, "Real-Time Human Fall Detection Using a Lightweight Pose Estimation Technique", in Computational Intelligence in Communications and Business Analytics, 2024, pp. 30–40.

[24] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks", Advances in Neural Information Processing Systems, vol. 33, pp. 9459–9474, 2020.

[25] N. Doshi and R. Patel, "Optimizing Retrieval-Augmented Generation for Open-Domain Tasks", Neural Networks, vol. 141, pp. 67–78, 2022.
[26] J. Park, H. Lee, and M. Kim, "Enhancing Retrieval-Augmented Generation with Adaptive Prompt Engineering", Journal of Information Retrieval and NLP, vol. 15, pp. 45–56, 2023.
[27] G. Xiong, Q. Jin, Z. Lu, and A. Zhang, "Benchmarking Retrieval-Augmented Generation for Medicine", in Findings of the Association for Computational Linguistics: ACL 2024, 2024, pp. 6233–6251.
[28] Q. Zhou et al., "GastroBot: a Chinese gastrointestinal disease chatbot based on the retrieval-augmented generation", Frontiers in Medicine, vol. 11–2024, 2024.
[29] E. Zuo, C. Pan, J. Chen, and Z. Yi, "Chinese Medicine Question Answering Robot Based on RAG and Self-Built Dataset", in Proceedings of the 2nd International Conference on Machine Learning and Automation, CONF-MLA 2024, November 21, 2024, Adana, Turkey, 2025.
[30] D. Restrepo et al., "Multi-OphthaLingua: A Multilingual Benchmark for Assessing and Debiasing LLM Ophthalmological QA in LMICs", ArXiv, vol. abs/2412.14304, 2024.
[31] B. Kang, J. Kim, T.-R. Yun, and C.-E. Kim, "Prompt-RAG: Pioneering Vector Embedding-Free Retrieval-Augmented Generation in Niche Domains, Exemplified by Korean Medicine", ArXiv, vol. abs/2401.11246, 2024.
[32] Z. Wang, H. Li, D. Huang, and A. M. Rahmani, "HealthQ: Unveiling Questioning Capabilities of LLM Chains in Healthcare Conversations", ArXiv, vol. abs/2409.19487, 2024.
[33] M. Jörke et al., "GPTCoach: Towards LLM-Based Physical Activity Coaching", in Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, 2025.
[34] A. Yuan, E. Colato, B. Pescosolido, H. Song, and S. Samtani, "Improving Workplace Well-being in Modern Organizations: A Review of Large Language Model-based Mental Health Chatbots", ACM Transactions on Management Information Systems, vol. 16, 10 2024.
[35] P. Qiu et al., "Towards building multilingual language model for medicine", Nature Communications, vol. 15, no. 1, p. 8384, Sep. 2024.
[36] C. Liu et al., "CPMI-ChatGLM: parameter-efficient fine-tuning ChatGLM with Chinese patent medicine instructions", Scientific Reports, vol. 14, no. 1, p. 6403, Mar. 2024.
[37] J. Miao, C. Thongprayoon, S. Suppadungsuk, P. Krisanapan, Y. Radhakrishnan, and W. Cheungpasitporn, "Chain of Thought Utilization in Large Language Models and Application in Nephrology", Medicina (Kaunas), vol. 60, no. 1, Jan. 2024.