8-2021

W8-Scope: Fine-grained, practical monitoring of weight stack-based exercises

Meera RADHAKRISHNAN
Singapore Management University, meeralakshm.2014@phdis.smu.edu.sg

Archan MISRA
Singapore Management University, archanm@smu.edu.sg

Rajesh K. BALAN
Singapore Management University, rajesh@smu.edu.sg

Pervasive and Mobile Computing 75 (2021) 101418

Contents lists available at ScienceDirect

Pervasive and Mobile Computing

journal homepage: www.elsevier.com/locate/pmc

W8-Scope: Fine-grained, practical monitoring of weight stack-based exercises

Meera Radhakrishnan *, Archan Misra, Rajesh K. Balan

School of Information Systems, Singapore Management University, 80 Stamford Rd, Singapore 178902, Singapore

ARTICLE INFO

Article history:
Received 21 July 2020
Received in revised form 20 March 2021
Accepted 21 May 2021
Available online 26 May 2021

ABSTRACT

Fine-grained, unobtrusive monitoring of gym exercises can help users track their own exercise routines and also provide corrective feedback. We propose *W8-Scope*, a system that uses a simple magnetic-cum-accelerometer sensor, mounted on the weight stack of gym exercise machines, to infer various attributes of gym exercise behavior. More specifically, using multiple machine learning models, *W8-Scope* helps identify who is exercising, what exercise she is doing, how much weight she is lifting, and whether she is committing any common mistakes. Real world studies, conducted with 50 subjects performing 14 different exercises over 103 distinct sessions in two gyms, show that *W8-Scope* can, at the granularity of individual exercise sets, achieve high accuracy—e.g., identify the weight used with an accuracy of 97.5%, detect commonplace mistakes with 96.7% accuracy and identify the user with 98.7% accuracy. By incorporating an additional, simple IR sensor on the weight stack, the exercise classification accuracy (across the 14 exercises) further increases from 96.93% to 97.51%. Moreover, by adopting incremental learning techniques, *W8-Scope* can also accurately track these various facets of exercise over longitudinal periods, in spite of the inherent natural changes in a user's exercising behavior. Our comprehensive analysis also reveals open challenges, such as adapting to the expertise level of individuals or providing in-situ, early feedback, that remain to be addressed.

1. Introduction

There is strong interest in developing pervasive sensing technologies to derive fine-grained & individualized insights into a person's gym exercise activities. By subsequently enabling personalized feedback, such monitoring can help support practically important objectives, such as preventing injuries [1] and curbing the propensity of early drop out among gym-goers [2]. Most approaches for gym exercise monitoring employ either body-worn, wearable devices (e.g., [3,4]), infrastructure-based video sensing [5] or the instrumentation of individual gym equipment [6,7]. Each approach has its own drawbacks: (a) *usability*: wearable devices may not be popular with the casual gym-going population (specifically, our survey with 107 users in a public gym revealed that over 59% were not in favor of using wearables), especially as a single wearable may not be sufficient (e.g., arm-worn sensors cannot help track leg or hip exercises); (b) *privacy*: video capture of workouts may be viewed as overly intrusive in public gym environments; and (c) *high deployment complexity*: approaches such as Jarvis [7] attach multiple sensors to different parts of an individual gym equipment, and additionally also instrument individuals with wearable sensors. Moreover, the efficacy of such approaches has typically been evaluated over relatively short observational periods (e.g., 1-2 gym sessions).

* Corresponding author.
E-mail addresses: radhakrism@smu.edu.sg (M. Radhakrishnan), archanm@smu.edu.sg (A. Misra), rajesh@smu.edu.sg (R.K. Balan).

https://doi.org/10.1016/j.pmcj.2021.101418

Fig. 1. Multi-purpose cable pulley machine and proposed sensor placement on the weight stack.

In this paper, we propose and evaluate a novel technique for wearable-free and *non-intrusive* monitoring of gym exercises performed using *weight stack*-based machines (which are widely used to perform activities for a variety of muscle groups). Our approach requires *no user instrumentation* and utilizes novel machine learning-based inferencing, over data from a single inexpensive accelerometer and magnetic sensor mounted on the weight stack (as illustrated in Fig. 1), to infer various individual-specific, exercise-related attributes.

Given our minimalist approach (a single sensor, mounted at a single point and capturing just the vertical motion of the weight stack), we explore two fundamental **research questions**: (1) Can data from only one simple weight-stack mounted sensor provide meaningful, fine-grained insights into the underlying exercise routine, such as 'amount of weight lifted' or 'which exercise is performed?' while accommodating exercise-and-user specific variations? And, how does our accuracy compare with a wearable-based alternative? (2) Can the inferencing logic, typically built through supervised learning based on labeled activity data collected over 1–2 sessions, be made robust enough to capture the medium-term *evolution* in an individual's gym activities?

Key contributions: We demonstrate the following key innovations and results:

- *Novel 'Weight-Stack Sensor'-based Inferencing for Exercise Monitoring:* We propose the use of a simple device, mounted rigidly to the top plate of a weight stack (illustrated in Fig. 1) to obtain fine-grained insights about the different exercises being performed. The device combines a 3-axis accelerometer and 3-axis magnetometer sensor, which capture distinct facets of the motion dynamics of the weight stack. Based on observed characteristics of these sensors, we develop a *multi-stage* pipeline (called *W8-Scope*[1]) that infers multiple novel facets of exercises, including (i) weight used; (ii) the type of exercise performed; (iii) the individual performing the exercise and (iv) common mistakes made.
- *Real-world Demonstration of W8-Scope:* We conduct real world (*in-the-wild*) studies with regular gym-goers at two separate gyms: (a) a *University* gym with a single multi-exercise cable pulley machine, and (b) a *Community* gym (open to the public) with 6 individual weight machines. Across these two gyms, using 1728 distinct *sets* of exercise data from 50 participants, we show that *W8-Scope* can (1) identify the *weight used* with 97.5% accuracy, (2) *distinguish among users* performing the same exercise with 98.7% accuracy, (3) distinguish among 14 distinct *exercises* with over 96.9% accuracy, and (4) identify commonplace *mistakes made* with 96.7% classification accuracy. Our results are also comparable to those achieved with a wrist-worn wearable (e.g., 84.3% for weight used, 96.4% for identifying mistakes). Finally, we also characterize the performance tradeoffs, in terms of accuracy vs. robustness vs. latency, between different variants of our core, set-based classification techniques.
- *Longitudinal Tracking & Incremental Learning:* While our approach provides high accuracy on unlabeled samples collected during the same or coterminous sessions (which is how most prior work has also been evaluated), we show that the inferencing accuracy degrades when applied to test data spaced weeks apart—e.g., exercise discrimination accuracy drops to 78%. To overcome this, we develop and validate an incremental learning strategy, which uses only highly confident samples to continually update the *W8-Scope* classifiers. This approach achieves an accuracy of 90.2% for classifying exercises and 87.4% in distinguishing users, even as an individual's exercise behavior evolves over a 12–15 week period.

Compared to other solutions that require more extensive instrumentation or wearable devices, we believe that *W8-Scope* demonstrates how low-cost instrumentation of commonplace gym equipment (specifically weight machines) can help obtain fine-grained, individual-specific insight in a privacy-sensitive manner. Such insight may be augmented with selective inputs from wearable devices or via the use of simple additional sensors (e.g., an IR sensor) in the future.

[1] Pronounced Weight-Scope.

2. Related work

We describe prior work on "exercise-monitoring" using mobile, wearable, infrastructural sensors and compare our approach against those.

Mobile, wearable & IoT sensor-based exercise monitoring: Chang et al. [8] were one of the first to propose a wearable solution (involving multiple accelerometers) for tracking the type and repetition count of free-weight exercises. Similarly, other works [9,10] also utilize multiple body-worn inertial sensors to detect different gym exercises. RecoFit [3] is also a wearable system based on an arm-worn inertial sensor to segment exercise and non-exercise periods and to detect different strength training exercises. Works such as [11,12] present smartwatch-based systems for recognizing and counting repetitions of various gym exercises. Zhou et al. [4] proposed a wearable fabric pressure sensor system that measures the muscle movement, action and repetitions of four leg machine exercises. Bian et al. [13] have demonstrated a wearable, body capacitance-based sensor for recognizing and counting seven different gym exercises. The recently proposed LiftRight system [14] utilizes an arm-worn initial sensor to quantify and analyze the performance of three different weight training exercises. The MuscleSense [15] system utilizes multiple sEMG sensors on the upper limb to assess the amount of workload while performing weight-based exercises. There are also other apps and wearables such as TrackMyFitness [16], Atlas Wristband [17], Samsung Watch Active2 [18] that detect exercises, record repetitions and track workout progress. Unlike *W8-Scope*, all these approaches require the user to have some body-worn devices. Among the various exercise attributes inferred, we believe that 'weight identification' and 'mistake identification' are harder to perform with wearable devices, while recognizing the exercise type (albeit limited to upper limb exercises) and user identification are easier to achieve using wearable sensors.

An alternate body of prior work assesses exercise characteristics using sensors attached to different parts of the exercise machine. Moller et al. [19] explored the use of a smartphone-based trainer for assessing the quality of exercises performed on a balance board. FEMO [20] is a platform for monitoring dumbbell exercises using passive RFID tags attached to individual dumbbells. Sundholm et al. [6] developed a pressure sensor mat that recognizes and counts repetitions of strength training exercises performed on a mat. Similarly, ExerTrack [21] is a recently proposed floor-based sensing system using a capacitive proximity sensor to recognize and count the repetitions of 8 different body weight exercises. The Jarvis system [7] utilizes multiple IoT sensors, attached to different moving parts of exercise machine to segment repetitions, recognize exercise type and provide feedback to the user through a VR headset. Closest in spirit to our work, Jarvis also uses wearable EMG sensors to incorporate muscle activation activity as part of the feedback. In contrast, our approach uses a single sensor device mounted on a novel location (the weight stack) to extract novel insights, such as the amount of weight lifted (besides exercise recognition) and commonplace mistakes made; we also consider the challenge of evolving the classifiers over medium time-scales.

Infrastructural sensor-based exercise monitoring: Prior work has explored the use of WiFi [22,23] and infrastructure-driven video sensing [5,24] for exercise activity recognition. SEARE [22] utilizes WiFi CSI waveform-based features to distinguish between 4 exercises. Similarly, Guo et al. [23] use CSI information to analyze workouts within a home/work environment. However, these WiFi-based systems may not work in a multi-user gym environment and in non line-of-sight scenarios. The GymCam [25] system leverages a single camera to track multiple people exercising simultaneously and recognize their exercise type and repetitions. However, this system does not track other aspects of exercising such as the weight lifted or mistakes made. Gonzalez-Ortega et al. [5] developed a 3D vision-based system to track the trajectories of human body parts during psychomotor exercises. Velloso et al. [24] presented a comparison of wearable sensor and Kinect model-based approaches for qualitative recognition of weight lifting exercises. All of these vision-based methods pose privacy concerns and are affected by external factors, such as lighting and line-of-sight. In contrast, *W8-Scope* is simpler to deploy, cost-effective and more privacy-friendly.

3. Overall goals and approach of W8-Scope

W8-Scope's broader goal is to quantify various attributes related to exercises performed using weight-based equipment in a gym or a fitness facility. To analyze their own progress, gym-goers are interested in tracking their exercises, weight lifted etc. [26]. A review of physical activity apps found that only 2% provided evidence-based guidelines for resistance training [27]. Automatically logging the exercise performed, as well as the amount of weight lifted, helps users (especially novice or intermediate users who lack knowledge about the proper exercise posture or use of gym equipment) to track their exercise performance and receive personalized feedback, such as: Am I committing more mistakes when performing *shoulder* exercises compared to exercises targeting other muscle groups? In this work, we focus on identifying the following facets: (a) the amount of **weight** used, (b) the **exercise** performed, (c) **incorrect patterns** of performing exercise and (d) which **user** is performing the exercise (the assumption being that each user has a unique signature while performing a specific exercise).

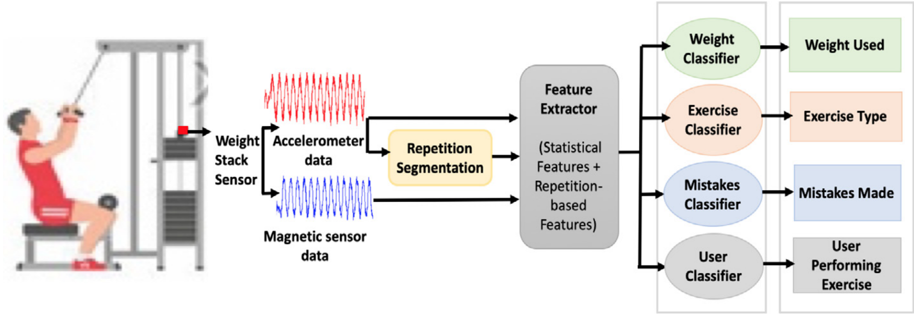


Fig. 2. Overview of *W8-Scope's* workflow.

3.1. Design goals and challenges

Design goals: One of our key goals is to devise a *wearable-free* and *non-intrusive* monitoring approach—i.e. infer the facets mentioned above without instrumenting the user's body with any wearable device, or without using privacy-violative infrastructural video sensing [5,25]. Our decision to avoid wearables is influenced not just by prior work [28] that suggests possible inconvenience from such devices, but also based on a survey conducted on 107 users of a *Community* gym: 59% of such users indicated an unwillingness to adopt wearable-based solutions (the antipathy to wearables was even higher (63%) among users in the 55+ age group). Our goal is to also provide a *simple* and *cost-effective* solution. As such, we propose to use a simple small form-factor sensor device mounted externally (i.e., after-market) on the top plate of a weight stack (unlike Jarvis [7], which uses multiple machine-attached sensors) to infer the exercise and related attributes. Such an approach does not interfere with the normal usage of the exercise machine.

Practical challenges: Our proposed approach, based on the attachment of a sensor to a single location, poses several practical challenges: (i) Distinguishing between different exercises becomes more challenging, given that the weight stack's motion is predominantly vertical and is likely to be similar across multiple exercises. This requires us to identify additional differentiating features; (ii) As the sensor is placed on the weight stack itself, it is thus exposed to noise, interference, and other confounding effects caused by nearby objects and users—e.g., the magnetic sensor is very sensitive to several environmental factors, including metallic equipment (e.g., dumbbells) carried by other gym-users; (iii) Different users perform the same exercise differently (with the motion dynamics potentially also varying with the choice of weight), implying the need to identify *robust* features; (iv) Over longer time periods, users exhibit natural "drift" in their exercising styles.

3.2. Overview of W8-Scope design

We utilize a combination of 3-axis accelerometer and 3-axis magnetometer sensor streams from a *weight-stack attached* sensor device (DA14583 IoT Sensor [29]), attached to the top-most slab, to uncover various attributes of a set of exercises performed on the weight machine. In our approach (illustrated in Fig. 2), we mainly leverage the magnetic sensor data to identify the amount of weight that is lifted, as the magnetic field strength is affected by this weight. We also combine features from accelerometer data to disambiguate magnetic sensor data which might look similar for different (weight, height) combinations. We then use a combination of features, extracted from both sensors, to identify the exercise performed and detect anomalous or incorrect exercise executions. The baseline *W8-Scope* pipeline operates at the granularity of individual sets—i.e., it infers the {weight, exercise type, user} attributes over an entire set, although we shall also study the implications of adapting it to support in-situ inferences, during an ongoing exercise set.

4. Dataset

We conduct extensive studies and experiments with 50 users performing a variety of exercises on weight stack-based machines under varying conditions. The data collection was performed in multiple phases at two different gym facilities (a *University* gym and a *Community* gym). The collected data included 2 distinct types of studies: (a) an initial *Validation Study* used to identify discriminative features and build the classification models, and (b) multiple *Real-World Studies*, conducted across 2 gyms, to evaluate *W8-Scope's* real-world accuracy.

For the studies, we focus on a class of 14 exercises that target different muscle groups and that the gym trainers indicated to be among the most popular exercise choices. At *University* gym, we monitored ten exercises performed using a weight stack-based "cable-pulley" multi-purpose equipment (shown in Fig. 1). This machine has a set of 20 free-weights (each weighing 2.5 kg, except the top-most slab (1.25 kg)), and permits at least 30 different weight training exercises [30]. Fig. 3 shows the position of the exerciser and the weight stack during the upward motion of these ten exercises. In the

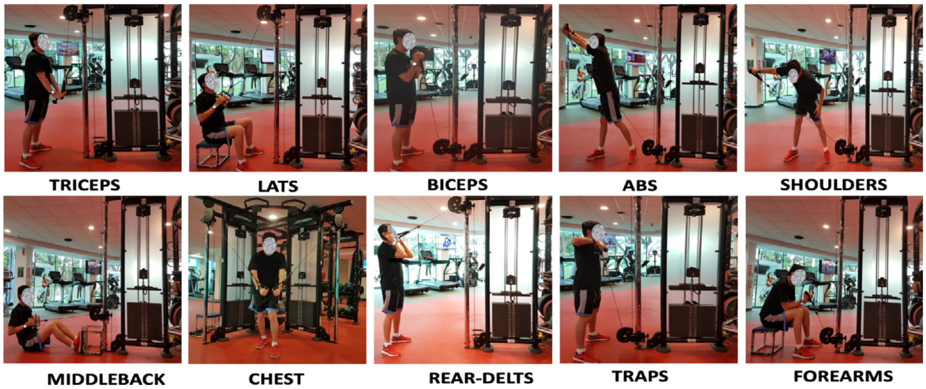


Fig. 3. Exercise positions for 10 exercises (on cable pulley machine).

Community gym, we utilize six dedicated single purpose weight machines for performing exercises such as *leg curls*, *leg press*, *triceps pushdown*, *biceps curls*, *chest press* and *shoulder press*. These machines have varying number of weight slabs, weighing 7.5 kg each.

4.1. Initial validation study

For the feasibility studies, we conducted several experiments using the cable-pulley machine in our *University* gym, over various controlled conditions across several days. The key parameters varied are: (i) the exercise performed (10 different exercises), (ii) amount of weight lifted (9 different weights), (iii) range of motion of the weight stack (4 different heights), (iv) different positions of placement of the sensor device (4 different positions), and (v) correctness of performing the exercise (2 incorrect executions). In total, we collected 252 sets of exercise data (where a *set* is the number of cycles of *reps* completed; an exercise set in our study consisted of 10 reps) for different combinations of these parameters across 8 subjects (5 males, 3 females).

4.2. Real world study

For the user study at *University* gym, we recruited 35 (23 males, 12 females) university students and staff. For the study at the *Community* gym, 15 (9 males, 6 females) participants were recruited. The studies were approved by our Institutional Review Board.

4.2.1. Overall study procedure

Prior to data collection, each weight stack exercise machine was instrumented with a sensor device, capturing both accelerometer and magnetometer sensors at 50 Hz. The participants who agreed to take part in the study were required to visit the gym and perform a set of specified exercises. At the *University* gym, the participants were also given a smartwatch (LG-Urbane), to be worn on their dominant hand, where a custom application captured accelerometer and magnetometer data (at 50 Hz). All the exercise sessions were video recorded for ground truth purpose. The number of sets and repetitions are as recommended by gym trainers. *Note:* For every exercise set, we collected data for 10 repetitions each. The participants were advised to take breaks (as required) in between exercise sets and were allowed to perform the exercises at a pace they are comfortable with. Except for the simulated incorrect executions, the subjects were not given any other special instructions and so, performed exercises *naturally*. An exercise session per subject ranged from about 35 to 55 min for $Study1\_univ$ and for 12 to 24 min for $Study2\_comm$. For participating in the study, we provided each participant a monetary compensation of $10.

4.2.2. Study in University gym ($Study1\_univ$)

At our University gym, we focused on collecting data for different exercises, different weights and simulated incorrect executions. Among the 35 participants, 30 performed: (i) 2 sets each of the ten exercises shown in Fig. 3, (ii) 3 sets of two exercises (*triceps* and *lats*) while simulating mistakes such as "pulling too fast", "releasing too fast" and "lifting only half through". For obtaining data for different set of weights, 18 out of the 35 participants performed three exercises (namely, *triceps*, *biceps* and *lats* exercise) using 6 different weights (from 3.75 kg to 16.25 kg). In total, we collected 1148 sets of exercise data. The details of this study are tabulated in column 2 of Table 1.

Table 1
Summary of real-world exercise dataset collected from *University* gym and *Community* gym.

	Study1_univ	Study2_comm
No. of participants	35 (23 males, 12 females)	15 (9 males, 6 females)
Age variation	21–35 years	18–65 years
Self-rated expertise	13 (Novice); 16 (Intermediate); 6 (Expert)	9 (Novice); 3 (Intermediate); 3 (Expert)
No. of exercises	10 (targeted muscles: forearms, biceps, triceps, chest, abs, shoulders, rear-delts, lats, traps, middleback)	6 (targeted muscles: biceps, hamstrings, chest, quadriceps, shoulders, triceps)
No. of sets of exercises	Total 1148 sets of 10 reps each 320 sets (6 weights for 3 exercises from 18 subjects) 588 sets (10 exercises with 2 weights from 30 subjects) 240 sets (4 incorrectness for 2 exercises from 30 subjects)	Total 180 sets of 10 reps—2 sets each of 6 exercises (with weights of subject's choice)
Variation of weights	6 weights (3.75 kg to 16.25 kg)	Weights used varied from 5 kg to 80 kg
Incorrect exercise variations	4 (pulling too fast, releasing too fast, pulling half way through, lifting heavier weight)	N/A
Average duration of exercise session across subjects	48 min	19 min
Aggregated duration across all sessions	36 h 50 min	5 h 46 min

Table 2
Summary of real-world *longitudinal* exercise dataset collected from *University* gym.

	Study3_long
No. of participants	10 (7 males, 3 females)
Age variation	21–35 years
Self-rated expertise	4 (Novice); 4 (Intermediate); 2 (Expert)
No. of exercises	5 (targeted muscles: triceps, biceps, abs, middleback, rear-delts)
No. of sets of exercises	Total 400 sets of 10 reps– 2 sets each of 5 exercises (with weights of subject's choice) on 4 different sessions
Variation of weights	Weights used varied from 3.75 kg to 43.75 kg
Average duration of exercise session across subjects	14 min
Aggregated duration across all sessions	8 h 20 min

4.2.3. Study in Community gym (Study2_comm)

At the publicly-accessible community gym, our focus was to obtain data from other demographic groups (e.g., working adults) and from different dedicated weight stack-based exercise machines (including leg exercises). The 15 subject in this study (referred to as *Study2_comm*) varied widely in their age, & expertise in weight training, and performed 2 sets each of 6 different exercises (with weights of their choice) on the dedicated weight stack machines. In total, 180 sets of exercise data were recorded (see column 3 of Table 1 for summary).

4.2.4. Longitudinal study in University gym (Study3_long)

In both *Study1_univ* and *Study2_comm*, the users performed exercises in a single session. We further conducted a *multi-session* study (*Study3_long*) with a subset of 10 users from the subject pool of *Study1_univ*. In addition to the original session, these users performed exercises on 4 additional days (separated by a week); furthermore, there was a gap of over 3 months between the original session and these 4 sessions (Fig. 4 illustrates the study period). In each of these session, the participant performed 5 exercises (namely, *triceps*, *biceps*, *abs*, *middleback* and *rear-delts*) with weights of their choice, resulting in a total of 400 sets of exercise data (details listed in Table 2).

5. Design and implementation of W8-Scope

To design *W8-Scope*, we first describe the sensor data patterns that occur during different exercises and detail the features extracted. We then explain how *W8-Scope* identifies different facets of such exercises.

5.1. Accelerometer sensor analysis

On inspecting the accelerometer sensor data across exercises, we observed that the accelerometer z-axis data clearly shows the variation with each repetition and also varies across different exercises, indicating the possibility of using an accelerometer to distinguish between exercises.

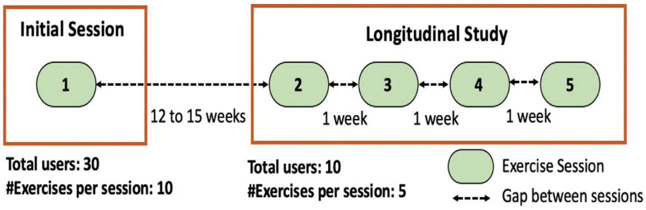


Fig. 4. Longitudinal study period.

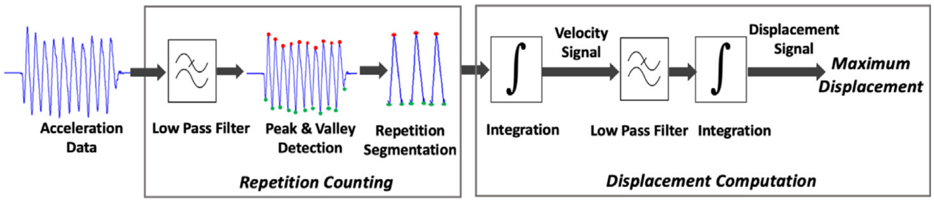


Fig. 5. Steps involved in counting repetitions and computing displacement.

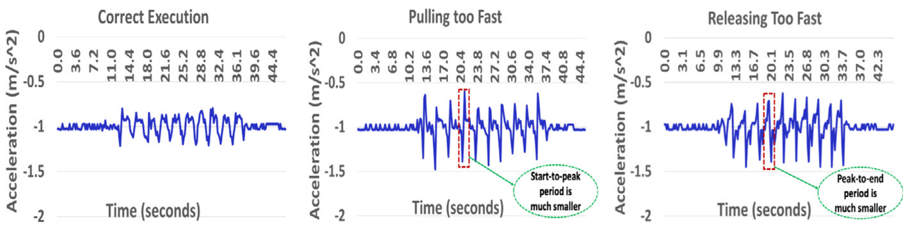


Fig. 6. Variation in accelerometer readings while performing Triceps Pushdown exercise (a) correctly, (b) by pulling weights too fast & (c) by releasing/slamming down the weights fast.

5.1.1. Identifying and counting repetitions

To segment and count individual repetitions in an exercise set from accelerometer data, the following approach is taken. The raw accelerometer data is initially filtered, then we obtain the local maxima and local minima (for z-axes)–i.e., points around which all other neighboring samples are lower/higher by δ (empirically set to 60% of the highest/lowest sample amplitude for our work). As certain repetitions were observed to have multiple peaks and valleys, an additional constraint on a minimum time threshold ΔT (empirically set to 2 s) between successive peaks is used to avoid over counting. The segment between two consecutive valleys is assumed to represent a repetition.

5.1.2. Computing the range of motion of weight stack

During our feasibility studies, we observed that one of the evident difference between exercises is in terms of the height to which the weight stack could be lifted (for the same amount of weights used). In addition, the inter-repetition time also vary for different exercises and different amounts of weight lifted (e.g., lifting heavier weights would take longer time). To compute the weight stack displacement (outlined in Fig. 5), we first extracted the z-axis acceleration signal, integrated it using cumulative trapezoidal integration [31] to obtain velocity, then low-pass filtered and then integrated again to obtain the displacement. As shown in Section 5.5, this approach results in a mean displacement error of $\pm$1.15 cm.

5.1.3. Understanding quality of exercise repetitions

To understand the common mistakes made while exercising, we first consulted the professional trainers in our campus gym. They reported that, (a) pulling or releasing the weights too fast, or (b) lifting the weight only half way through corresponded to some "common mistakes" made by novice users.

As a preliminary study, we collected data from 6 trainers at the gym for 3 sets of 10 *reps* of six exercises (out of the 10 exercises on cable pulley machine). Out of the 3 sets, they were instructed to perform one set correctly and two sets incorrectly—i.e., pull the weights too fast or release the weights too fast. We found (e.g., see Fig. 6) that the accelerometer data contains visible *signatures*, that can help distinguish between such correct and incorrect execution patterns (as shown later in Section 5.5).

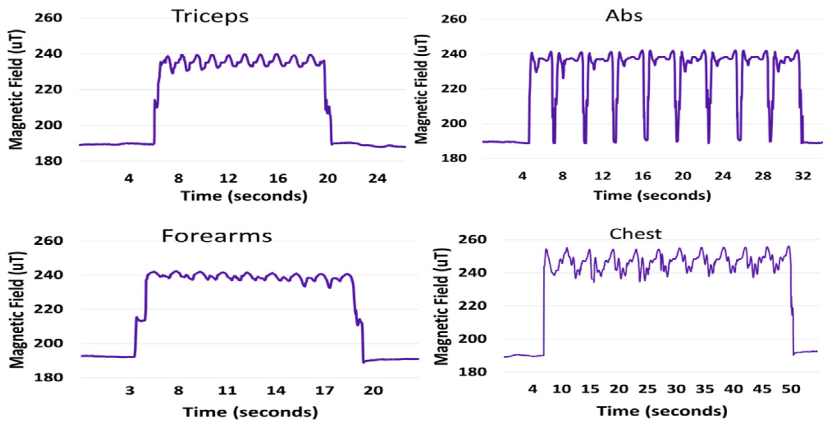
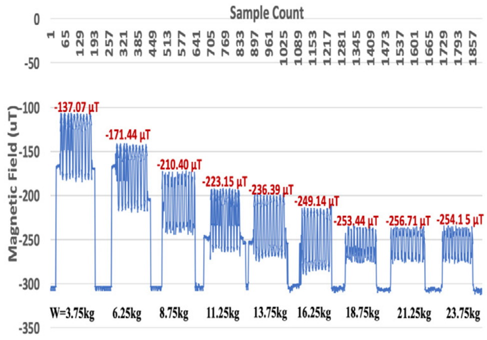
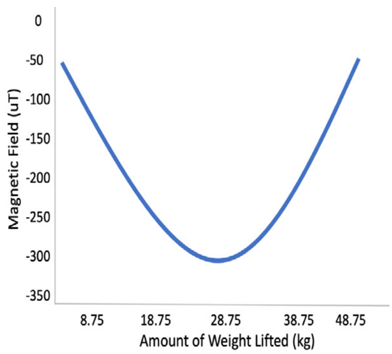


Fig. 7. Variation in magnetic field for four exercises performed with weight, $w = 6.25$ kg.



(a) Variation in magnetic field for 9 sets of 10 repetitions of lats exercise with weights per set varied between $w = 3.75kg$ and $w = 23.75kg$. (The number on top of each set shows the mean value of the magnetic field (μT) when lifting specific weight).



(b) Expected "U-shape" of the magnetic readings as the amount of weights is varied from minimum weight ($3.75kg$) to maximum weight ($48.75kg$).

Fig. 8. Variation in magnetic field for different weights.

5.2. Magnetic sensor analysis

We next studied how the magnetic field, sensed by a magnetometer, varies when performing different exercises using the cable pulley weight stack machine.

We found that the magnetic field indeed varies as the weight stack goes up and down, indicating individual repetitions of each exercise and also shows distinct pattern across exercises. Fig. 7 shows the distinct pattern of the magnetic field for a sample of four exercises performed with a weight, $w = 6.25$ kg.

5.2.1. Variation in magnetic field vs. weight lifted

We also observed that the magnetic field not only changes with the motion of the weight stack, but also as a function of weight lifted. Consider the weight stack has a set of m weight slabs, each slab with mass $= w$. Let d_i be the distance of the ith slab from the sensor, while at rest, and let D be the distance (height) moved by the set of K ($K \leq M$ weight slabs that are lifted). Eq. (1) represents magnetic field strength (which varies inversely with the square of the distance), B as a function of K. The first term represents the K slabs that move up (leaving the slab-sensor distance unchanged) and the second term represents the $M - K$ slabs) that do not move.

$$B = \sum_{i=1}^{K} \frac{w_i}{d_i^2} + \sum_{i=K}^{M} \frac{w_i}{(D + d_i)^2} \tag{1}$$

Table 3
Features extracted from each time window of accelerometer and magnetometer data. The *Count* represents the number of signal axes on which the feature is computed. For e.g., *count* = *4* means features are extracted on the *x-axis*, *y-axis*, *z-axis* and the *magnitude* of the signal.

Feature	Count	Description
Mean	4	Average of the values for the time window for each axes and the Euclidean norm (magnitude) of the signal
Max	4	Maximum value in a time window for each axis and signal's magnitude
Min	4	Minimum value in a time window for each axis and signal's magnitude
Range	4	Total change in values within the time window for each axis and signal's magnitude
Variance	4	Variance of the values in a time window for each axis and magnitude of signal
Spectral entropy	4	Normalized information entropy of the FFT components of each axis and magnitude of signal
Spectral energy	4	Mean value of the square of the FFT coefficients of the signal for each axis and magnitude value
Mean crossing rate	4	Number of times the values cross the mean of the time window
Covariance	3	Covariance between each pair of axes of the sensor
Correlation	3	Correlation between each pair of axes of the sensor
Repetition time	1	Average time taken to complete a repetition in a exercise set
Repetition height	1	Average height to which the weight stack was lifted within a set
Repetition velocity mean	1	Average of the speed with which the weight stack was lifted in a set
Repetition velocity Std.dev	1	Standard deviation of the speed with which the weight stack was lifted in a set

Accordingly, as illustrated in Fig. 8(b), the magnetic field at the zenith should exhibit a U-shape curve, initially decreasing (as K increases from a small value) but then eventually increasing (as the first term begins to dominate when K becomes larger).

Fig. 8(a) shows the variation in magnetic field while performing 10 repetitions each of *lats* exercise with 9 different set of weights ranging from 3.75 kg to 23.75 kg. The figure is annotated (in red color) with the mean value of the sensed magnetic field as experienced by the sensor when lifting varying amount of weights, and shows how the magnetic sensor values can help distinguish between different weights. Initially as the amount of weight is increased, the strength of the magnetic field keeps decreasing, thus making it easier to distinguish between the lighter weights. However, at higher weight values, the differentiation in the magnetic field is less pronounced (e.g., the mean magnetic field is -255 μT for $w = 21.25$ kg or $w = 23.75$ kg).

5.2.2. Magnetic field vs. (height, weight) variation

Given that the magnetic sensor is affected by both the height (D) and the weight lifted, we next study if there are cases where the magnetic sensor would be unable to distinguish between "weight $= w1$, height $= h1$" and "weight $= w2$, height $= h2$" combinations? We conducted an experiment in which *lats* exercise was performed with 3 different weights (3.75 kg, 8.75 kg, 13.75 kg) lifted to 4 different controlled heights (6 cm, 12 cm, 18 cm, 24 cm). We observed that the change in magnetic field for weight, $w = 8.75$ kg and height, $h = 6$ cm looked very similar to that of $w = 13.75$ kg and $h = 24$ cm (mean and total changes being approx. 45 μT and 32 μT respectively for both cases). A magnetic sensor alone is thus insufficient for resolving ambiguity: both magnetic and accelerometer sensor data are thus needed to accurately distinguish between different weights.

5.3. Sensor data analysis: Key takeaways and features

Based on our initial validation experiments and analysis, our major takeaways are: (i) the weight stack movement is clearly identifiable from the magnetometer data, (ii) the accelerometer sensor can provide an accurate estimate of the precise exercise-related $z - axis$ movements, as well as two useful motion-related features: the *time taken* to complete a repetition as well as the *height* to which the weight stack is lifted, (iii) the combination of accelerometer and magnetometer readings can help identify the amount of weight that is being lifted, and (iv) the accelerometer data also contains latent temporal features that are characteristic of selected mistakes in exercise motion dynamics.

Accordingly, in our approach, both the accelerometer and magnetic sensor streams are first pre-processed (for each individual set) to remove any outliers. The pre-processed sensor data is divided into frames of length w ($w = 2$ s, based on the observed duration of a single *rep*). On each frame, we first extract statistical features for each axis and the magnitude of both sensors. As described in Section 5.1, we also compute *repetition-based* features such as average *time taken* per repetition, average *height* to which the weight stack was lifted, and the average & standard deviation of *speed* with which the weight stack was lifted/brought down. See Table 3 for the complete set of features used in our classifier models.

5.4. The W8-Scope classification pipeline

Based on the insights gathered, we develop the *W8-Scope* classification pipeline. After evaluating different machine learning models, we use a Random Forest (RF) classifier (that gave best performance, similar to prior works (e.g., [24,32])) throughout our multi-stage pipeline. The key components in the classification pipeline (see Fig. 9) are as follows:

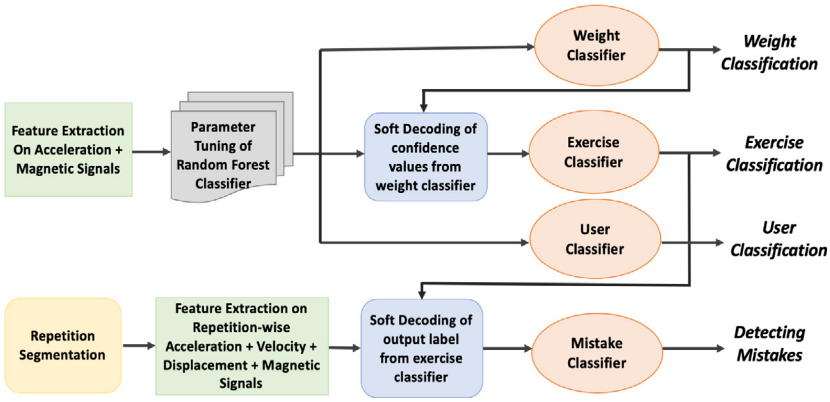


Fig. 9. The *W8-Scope* multi-stage classification pipeline.

Table 4
Average error (in cm) in displacement computation for varying heights to which weight stack is lifted.

Actual height	6 cm	12 cm	18 cm	24 cm
Average error	±0.67 cm	±0.87 cm	±1.1 cm	±1.96 cm

- *Amount of Weight Lifted Identification* – We train a *weight classifier* using the parameter tuned random forest classifier. The weight classifier provides the probability for each of the different weights (i.e., the probabilities that weight $= [w1, w2, \ldots, wn]$) for each distinct *set*. Note that the probability is computed first using each individual repetition, and then combined across the consecutive repetitions to determine the probability over an entire set.
- *Exercise Identification* – For the *exercise classifier*, we follow a *soft decoding* approach: we include an additional feature vector, consisting of the probability values for each of the candidate weight classes, instead of just using the 'most likely' weight value. The exercise classification is performed on the new feature set with the parameter tuned RF classifier. As before, the exercise probability values are computed for each individual repetition, and then combined to determine a set-level probability. In addition, in Section 6.3.2, we shall look at the possibility of performing such exercise classification *during* (i.e., after 3–6 repetitions) a set, possibly by including additional inexpensive IR (infrared) sensors.
- *Detecting Mistakes in Exercise Execution* – We next attempt to detect the mistakes made, at a *per-repetition* level. (This is necessary as users may incorrectly execute only a subset of the multiple repetitions in a set.) We first segment both sensor signals corresponding to the upward and downward motion of the weight stack during a repetition using techniques described earlier in Section 5.1.1. We then obtain the *velocity* and *displacement* corresponding to each upward & downward transition for each repetition. We also feed in the output of the *exercise classifier* (obtained by taking majority output labels over an entire set)–i.e., mistake identification is not performed real-time, but only at the end of an entire set (usually lasting 30–40 s). We use another RF classifier, and this new set of features, to classify the commonplace different mistakes such as {"pulling the weight stack too fast", "releasing fast or slamming down the weight stack", "lifting the weights only half-way"}.
- *User Identification* — To identify the specific user, we used the initial set of features used for weight classification to build multiple *per-exercise* classifiers, and use the specific classifier (corresponding to the identified exercise) to identify the user for an entire exercise *set*.

5.5. Initial validation results

We now present summarized results on the performance of different *W8-Scope* components, evaluated on validation studies (explained earlier in Section 4.1). The repetition counting mechanism (Section 5.1.1) achieves an accuracy of **98%** in counting the 10 repetitions in each set. For displacement computation, we observed an average estimation error of ±1.15 cm compared to the ground truth height. Table 4 shows the breakdown of the average error in displacement computed for each height. Additional results (summarized in Table 5) show that the combination of accelerometer and magnetic sensing features hold promise in achieving high accuracy (over **97%** using 10-fold cross validation) in inferring different exercise-related attributes.

Table 5
10-fold cross validation results of *W8-Scope* classifier models with initial validation study data.

	Weight classification	Exercise classification	Mistakes classification
Only accelerometer	77.49%	91.53%	90.43%
Only magnetometer	92.96%	79.37%	83.85%
Accelerometer and magnetometer	**99.41%**	**98.74%**	**97.34%**

Table 6
Performance of **amount of weight** identification.

Weight classification	Accuracy	Precision	Recall
10-fold CV ($Study1\_univ$)	97.5%	0.978	0.971
LOOCV ($Study1\_univ$)	93.75%	0.937	0.938

6. Real-world W8-Scope evaluation

We now present the performance evaluation of *W8-Scope*, along with insights gained, based on real world, naturalistic exercise data collected (described in Section 4.2) from two gyms. We focus on the primary attributes of interest {*Weight Used, Exercise Performed, Mistake Identification, User Identity*}. For the *University* gym, we also compare our proposed approach against that obtained via a wearable (smartwatch).

6.1. Counting repetitions

We first evaluate the performance of *repetition counting*. Using 908 sets of data collected from different weights and different exercises experiment in $Study1\_univ$, we obtained a performance of **97%** in accurately counting the 10 repetitions per set. Out of the 28 incorrectly counted sets (that caused 3% error in counting reps), 12 sets are off by ± 1, 9 sets are off by ± 2, 4 set are off by ± 3, 2 sets are over counted by 4 and 1 set is under-counted by 5. *W8-Scope* under-counted the repetitions primarily for the *forearms* exercise, because the range of motion of the weight stack was too short to show evident peaks in acceleration data. Over counting of repetitions happened due to human artifacts, when the subject moved the weight stack up and down while 'prepping' at the beginning of the set. For the 180 sets of additional data collected from $Study2\_comm$, the repetitions were accurately counted for 177 sets (98% accuracy), indicating that this estimation was accurate across gym environments.

6.2. Identify the amount of weight lifted

We evaluate the performance of weight classification on different weights' data obtained from $Study1\_univ$. Based on 10-fold cross validation with RF classifier (which outputs the dominant label observed across all the repetitions in a set), we achieved an accuracy of **97.5%** in distinguishing between six set of weights, $w = [3.75, 6.25, 8.75, 11.25, 13.75, 16.25]$ in the weight stack, with the classification error confined to the heavier weights – 13.75 kg and 16.25 kg.

We also performed a *leave-one-subject-out cross validation* (LOOCV) in which the *weight-classification* model was trained with data from all users, except the test user, and then tested on the data from test user. Using this approach, we obtained an average accuracy of **93.75%**, with a precision of 0.937 and recall of 0.938 in classifying the weights, i.e., the mean percentage error was 6.25%, with the maximum error (11%) in recalling weight, w = 16.25 kg. Table 6 presents the summary of results from weight classifier.

6.3. Identify the exercise performed

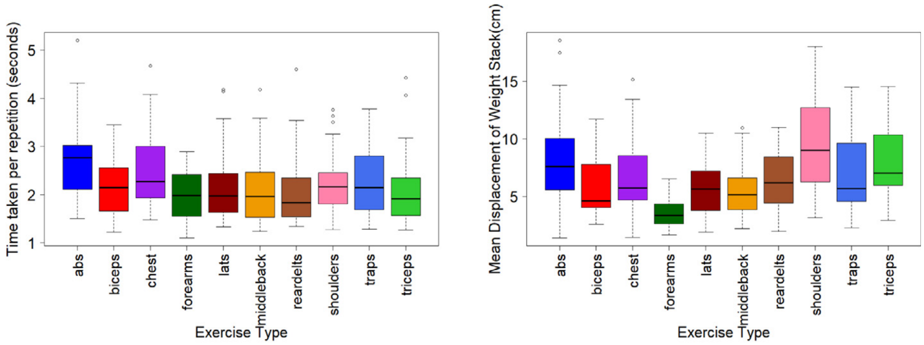
University gym: We first evaluate the accuracy of classifying the 10 exercises (performed on the multi-purpose cable pulley machine) from 588 sets of data collected from 30 subjects in $Study1\_univ$. We obtained a performance accuracy of **96.93%**, with a precision of 0.962 and recall of 0.969, in classifying the exercises. This is a mixed person model as it includes training data from all the users for all the exercises. From the confusion matrix(Fig. 10), we found that the classification errors occurred primarily during *middleback*, *rear-delts* and *biceps* exercises, due to the higher *within-exercise* variability across users.

Using *InfoGainAttributeEval* in Weka, we further evaluated the features with the highest information gain. We found that the *repetition-height* and *repetition-time* (both of which are derived from accelerometer data) were the most distinguishing features in exercise classification. To illustrate this, Fig. 11a plots the distribution of the average time *per repetition* of each exercise across all 30 subjects. For most users, *abs* exercise took the longest ($\geq$2.65 s) and *rear-delts* exercise took the least amount of time ($\leq$2 s). Similarly, Fig. 11b plots the boxplot of the variation of the height to which the weight stack was lifted for each of these 10 exercises.

Community gym: To further evaluate the exercise classification accuracy, we analyzed the $Study2\_comm$ data (where users performed exercises using exercise-specific weight machines) by withholding the machine label. We applied a

Predicted Label	Actual Label: abs	biceps	chest	forearms	lats	middleback	reardelts	shoulders	traps	triceps
abs	98.0%	0.2%	0.1%	0.1%	0.0%	0.2%	0.2%	0.7%	0.2%	0.0%
biceps	0.1%	96.1%	0.4%	0.3%	0.2%	0.5%	0.5%	0.3%	0.4%	0.2%
chest	0.2%	0.1%	96.5%	0.0%	0.3%	0.3%	0.0%	0.7%	0.2%	0.2%
forearms	0.0%	0.3%	0.1%	98.8%	0.1%	0.7%	0.4%	0.0%	0.1%	0.3%
lats	0.1%	0.9%	0.4%	0.3%	97.9%	1.5%	0.7%	0.1%	0.2%	0.5%
middleback	0.0%	0.2%	0.1%	0.2%	0.6%	94.9%	1.6%	0.1%	0.4%	0.2%
reardelts	0.0%	1.0%	0.5%	0.3%	0.4%	1.2%	95.2%	0.1%	0.8%	0.5%
shoulders	1.3%	0.1%	1.5%	0.0%	0.0%	0.1%	0.1%	97.4%	0.2%	0.1%
traps	0.2%	0.3%	0.2%	0.1%	0.1%	0.5%	0.6%	0.1%	96.8%	0.3%
triceps	0.1%	0.7%	0.3%	0.0%	0.3%	0.1%	0.8%	0.4%	0.7%	97.7%

Fig. 10. Confusion matrix of exercise classification (with *Study*1_*univ* data).



(a) Average **time taken** to complete repetition (b) Average **height** to which weight-stack was lifted

Fig. 11. Variation in repetition time & height per exercise (across subjects).

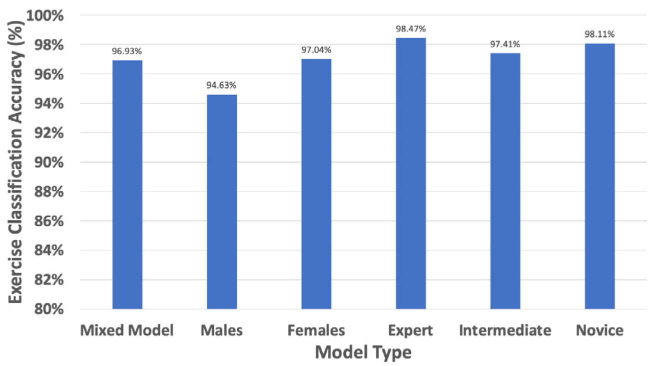


Fig. 12. Performance comparison of different exercise classification models.

Table 7
Performance of identifying **exercise performed**.

Exercise classification	Accuracy	Precision	Recall
10-fold CV (*Study*1_*univ*)	96.93%	0.962	0.969
10-fold CV (*Study*2_*comm*)	97.79%	0.978	0.982

10-fold CV approach, where the data consisted of exercises performed across *all* the 6 machines. *W8-Scope* achieved an accuracy of 97.79% (precision = 0.978, recall = 0.982) in classifying the 6 exercises performed by 15 subjects. With a *leave-one-exercise-set-out* cross validation approach, the accuracy drops slightly to 94.4%. Table 7 summarizes the performance of exercise classifier.

6.3.1. Performance of different segregated models

The results presented to date are based on cross-validation studies, where the training data involves labeled data combined across all users. Given an assumption that exercising styles are likely to be affected by certain demographic

(a) Sitting on the floor (b) Sitting on an exercise bench (c) Standing

Fig. 13. Thermal images captured by IR sensor for three different exercise positions of the individual.

attributes (e.g., gender or expertise level), we further studied the benefits of building such demographics-specific classification models. We utilize the data collected from *Study*1*_univ* for this purpose.

We first trained exercise classification models separately for *males* and *females*. Using a 10-fold CV approach for classifying exercises, we achieved an average accuracy of 94.63% with *males-only* model and 97.84% with *females-only* model (compared to our reported baseline of 96.93%). The *males-only* model exhibited comparatively poorer performance, with exercise classification accuracy being especially lower for *reardelts* and *middleback* exercises (precision and recall both less than 0.92). This degradation was due to the wider variance in exercise dynamics, observed among male participants. For the *middleback* exercise, some of the male individuals interchangeably used two distinct positions (''on the bench'' vs. ''on the floor'') to perform the exercise, leading to poorer classification accuracy (an aspect that we shall address further in Section 6.3.2). However, this variability was absent among any of the female subjects, resulting in a slight improvement in the overall exercise classification performance.

We further divide the data based on expertise levels of the individuals and train three different models for *expert*, *intermediate* and *novice* category of users. Based on 10-fold cross validation, each of the three models achieved an average classification accuracy of 98.47%, 98.01% and 97.41% respectively. Compared to the mixed-person model, there is a 1%–2% improvement in the performance of classifying the exercises. The highest accuracy was achieved for *expert* category of users, who are the most familiar with these specific weight machine exercises. On the other hand, the highest fraction of mis-classfications are for novice users with lowest performance in classifying *shoulders* exercise (which achieved a precision of 0.93 and recall of 0.91). The shoulders exercise was also reported as the toughest exercise, especially among novices.

Fig. 12 shows the performance comparison of different exercise classification models. Overall, we observed that data segregation, based on different categories such as the gender and expertise level, helps in slight improvement in the overall exercise classification accuracy compared to a mixed-person model.

6.3.2. Improving exercise classification with additional cheap sensors

In our prior analysis, we observed that the exercise classification accuracy drops as individuals maintain different postures while performing the same exercise. This was primarily evident for the *middleback* exercise where certain male subjects altered the position between ''on the bench'' and ''on the floor'' to perform the exercise. To overcome this variability introduced by varying exercise postures, we propose an approach where additional cheap IR sensors are attached to the top of the weight stack. These IR sensors help to obtain the contour of the exercising individual, in a privacy-preserving fashion (as they provide only a thermal silhouette), when the exercise starts and identify if the person is sitting on the floor or sitting on a bench or standing. To study the feasibility of this approach, we attached a thermal camera to the top of the weight stack and the individual who performed the exercise maintained different positions. We used the FLIR Lepton micro thermal camera module and attached it to a raspberry pi to capture the images. Fig. 13 shows the images captured by the thermal camera when the person performing the exercise (either middleback or reardelts exercise) was (a) sitting on the floor, (b) sitting on a bench and, (c) standing. We can observe that the contour of the exercising individual is clearly visible in the thermal images.

To validate if such additional hints can better classify among certain exercises that have higher variability across individuals, we included an additional feature which specifies the exercise position (i.e., stand, sit on bench, sit on floor) into the exercise classification model. For this, we assume that this extra information is available as ground truth when training the exercise classifier. Using a 10-fold CV approach, we achieved a slightly higher accuracy of 97.51% (compared to earlier 96.93%) in classifying the exercises. We also evaluated the performance separately for the gender-wise segregated models. We achieved an average accuracy of 96.45% and 97.89% for the *males-only* and *females-only* model respectively. We found that the performance of the *males-only* model improved by ≈2%. Upon inspecting the confusion matrix, we found that the accuracy of classifying *middleback* and reardelts exercises has improved. Middleback exercise was often confused with *lats* and *reardelts* exercises due to the similarity in the positions followed during these exercises. Therefore, including the additional postural information helped to more accurately disambiguate these exercises.

Table 8
Performance of identifying **mistakes made**.

Mistakes classification	Accuracy	Precision	Recall
10-fold CV (*Study*1_*univ*)	96.75%	0.968	0.967
LOOCV (*Study*1_*univ*)	79.2%	0.78	0.82

6.4. Identify exercise "Mistakes"

For evaluating the performance of this component, we utilized the data collected for three variations of incorrect executions (explained earlier in Section 5.4) of two exercises (triceps and lats) from 30 subjects in *Study*1_*univ*. We also included data from one *correct* execution set for each exercise.

Using 10-fold CV, we obtained an overall performance accuracy of **96.75%** in classifying the mistakes. Using LOOCV, we observed a sharp drop in accuracy to 79.2% (precision = 0.78; recall = 0.82). The performance drop in LOOCV is explained by the fact that *mistakes are often person-specific*, with mistakes for one person appearing very similar to the correct execution by another user–e.g., the weight stack motion dynamics for a tall user *lifting half way* are very similar to a short user performing *correct lifting*. The performance of classifying exercise mistakes is tabulated in Table 8.

6.4.1. Additional insights into 'Typical Mistakes'

Because our long-term goal is to provide individuals actionable feedback to correct mistakes, we also performed manual annotation of the exercise videos (which provide "ground truth") to understand a few additional characteristics of such mistakes. Table 9 details the various fine-grained insights that we gained from this analysis.

1. *Does lifting 'too heavy a weight' result in disproportionately higher mistakes (e.g., 'releasing too fast', 'lifting only halfway', 'making postural mistakes')?*
For this purpose, we manually annotated the 60 ground truth videos recorded, across 30 subjects, for the *triceps* and *lats* exercises performed with "heavy weight". The annotation was performed separately for the two individual transitions (upward and downward) of each repetition. We observed that, out of the 584 repetitions from 60 sets of lifting heavy weight, the subjects committed some kind of mistake (details listed in Table 9) during 93 repetitions across 21 sets (35% of heavy sets). The prominent mistakes were 'releasing too fast' and 'lifting only half-way through'. The other common mistake of 'pulling too fast' was not observed in this data. Also, compared to exercise sets performed with lighter/comfortable weight, on an average the time taken to complete one repetition for *triceps* and *lats* exercises also increased by 0.65 s. We used the previously trained *mistake-classifier* model and provided the data from the sets which had manually labeled 'mistake labels' as a test set. We obtained an overall accuracy of 81% (precision = 0.84; recall = 0.80) in classifying the two mistakes ('lift half way', 'release too fast'). In contrast, applying the same classifier to the 120 sets (of the same 30 users) which involved lighter weights resulted in the identification of mistakes in 57 repetitions across 14 sets (11.6% of non-heavy sets). This strongly suggests that mistakes in exercise motion dynamics are significantly more likely (almost double) when gym-goers attempt to exercise with heavier weights.
2. *Are mistakes isolated (singletons) in a set, or do they consistently manifest across an entire set?* To answer this question, we randomly selected 10 subjects and manually annotated 197 videos of their 10 exercises performed naturally with two different weights (3.75 kg, 6.25 kg). Out of the 197 exercise sets, 64 *reps* within 20 sets (10%) across 6 subjects had incorrect executions (i.e., had at least one rep with any of the 3 mistakes: 'pulling too fast', 'releasing too fast', 'lifting only halfway'). Moreover, *mistakes are often repeated*: 75% of the incorrect sets (15 out of 20) had 3–5 consecutive incorrect repetitions. The *W8-Scope* classifier was able to correctly identify 83% of the mistakes performed in these manually-curated sets.

Key takeaway: Our analyses suggests that *W8-Scope* can be used to reliably identify the majority of instances (repetitions) within an exercise set/session where a user makes commonplace "motion dynamics-related" errors. Such knowledge can then be used to tailor useful actionable feedback: e.g., observations of more frequent mistakes during *shoulders* exercise likely indicate weak shoulder muscles, and the gym-goer may be recommended additional shoulder exercises. However, our purely weight-stack based approach does not currently provide insights into other *postural* mistakes that may be committed by novice users.

6.4.2. Mistakes classification with 'weight' as an input

The *mistakes classifier* in *W8-Scope* pipeline does not use the 'weight used' as a feature in classifying the various mistakes that are made by individuals. Based on manual annotation and analyses of the "heavier weights" exercise data, we observed that there is a greater probability among individuals in making mistakes when lifting heavier weights compared to lighter weights. As such intuitively, incorporating the amount of weight used as a feature should help to improve the performance of mistakes classification.

Table 9
Insights into *Typical Mistakes* that people make — Observations from exercise videos.

Key observations	Supporting evidence
People tend to make more mistakes while lifting heavy weights	35% of heavier weight lifted sets had mistake — lift half way (62 reps), release too fast (31 reps)
Postural mistakes such as "hunching the back", "leaning forward", "moving elbow during triceps exercise", "swinging body during lats exercise" are commonly made while lifting heavy weights	41% of heavy weight sets had mistakes with body postures — hunch (33 reps), lean forward (16 reps), move elbow (54 reps), swing body (67 reps),
People tend to mistakes constantly in an exercise set	75% (15 sets) of the incorrect sets had 3–5 consecutive reps that were incorrect
Most mistakes are made towards the end of an exercise set and in the second set of the same exercise	90% of incorrect sets have mistakes made from rep 6 and onwards
Lifting the weight half way through followed by releasing the weight too fast were the prominent mistakes	Out of 64 incorrect reps — lift half way (48 reps), release too fast (10 reps), pull too fast (6 reps)
Most number of mistakes were made while performing shoulders exercise followed by chest and abs exercises	Incorrect reps: Shoulders (53%), Chest (17%), Abs (12%)

Table 10
Performance of **user identification**.

User classification	Accuracy	Precision	Recall
10-fold CV (*Study1_univ*)	98.97%	0.989	0.988
10-fold CV (*Study2_comm*)	98.74%	0.985	0.987

Table 11
Summary of performance accuracy – *W8-Scope* vs. smartwatch approach.

	W8-Scope *Study1_univ*	*W8-Scope* *Study2_comm*	Smartwatch *Study1_univ*
Weight classification	97.50%	N/A	84.37%
Exercise classification	96.93%	97.79%	98.75%
Mistakes classification	96.75%	N/A	96.46%
User classification	98.97%	98.74%	99.31%

To investigate this, we included 'weight' as a feature in the mistakes classifier. We utilized the 21 sets of data (earlier annotated to have mistakes) corresponding to heavier weights to evaluate the performance of the modified classifier. We obtained an average classification accuracy of 84.7% in identifying the mistakes ('release too fast' and 'lift half way'). This is about ≈4% improvement in performance, compared to the original classifier model which achieved 81% accuracy on the same data.

6.5. Identify users performing exercises

W8-Scope's final component helps to distinguish between the different users performing the same exercise. Table 10 summarizes our numerical results.

University gym: Applying the 'User Classifier' across the 30 university gym users results in a classification accuracy (using 10-fold cross validation) of **98.97%**. Out of the 10 exercises, the classification errors are primarily confined to the *shoulders*, *forearms*, *middleback* and *triceps* exercises. On more careful inspection, we found that the users who were typically misclassified had highly similar repetition-based features– i.e., having similar range of motion for the weight stack and taking the same amount of time to complete a repetition. By ranking the features based on its information gain, we found the most significant features to include: (a) *repetition time*, *displacement height* and *velocity* for the accelerometer sensor, and (b) *minimum*, *maximum* and *energy* of the 3-axes, for the magnetometer sensor.

Community gym: *W8-Scope's* 'User Classifier' achieves an accuracy of 98.74% (precision = recall = 0.98), when applied to the case of 15 users who performed 180 total sets of 6 different exercises. Note that the Community gym-goers were more diverse (in terms of various demographic factors and gym expertise). Our results thus demonstrate that *W8-Scope* can indeed be applied robustly to distinguish among users, across a wide variety of demographics.

6.6. Performance comparison: W8-Scope vs. smartwatch

Using the *Study1_univ* data, we compared (and summarize in Table 11) the performance of each component of *W8-Scope* with that of an alternative smartwatch-based approach. Key results include: (a) A weight-stack mounted sensor is able to identify the weight lifted more accurately than a hand-worn sensor (overall accuracy of 84.37%, precision =

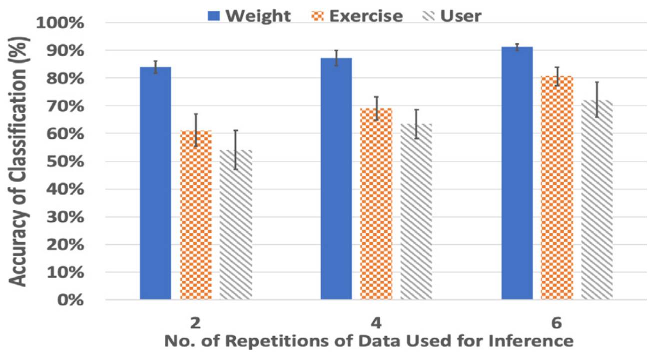


Fig. 14. Early classification performance accuracy for identifying weight used, exercise type and the user performing the exercise.

0.822 & recall = 0.845); (b) The smartwatch achieves slightly higher accuracy (98.75%) for exercise classification. (b) As expected, because of its ability to track the 3D arm motion precisely, the smartwatch has a slightly better accuracy of 99.31% (precision = recall = 0.99) in identifying the user. (c) For identifying the exercise performed or any mistakes made, the performance of *W8-Scope* and the smartwatch is roughly comparable.

6.7. In-situ/early classification

In *W8-Scope*, the identification of weight used, exercise type and the user are based on 'majority voting' across a set. While obtaining such insights and providing suggestions (in a retrospective manner at the end of a set) would help the individuals to improve their performance during the subsequent sets, it might be more useful (especially for novice users) to enable feedback earlier (i.e., during an exercise set). To study this potential 'accuracy' vs. 'real-time feedback' trade-off, we tested the *W8-Scope* classifiers for identifying weight, exercise and user with inferences obtained only from {2,4,6} repetitions and the corresponding majority voting (instead of majority voting across a set). We utilized the data collected during $Study1\_univ$ for this purpose. Fig. 14 shows the variation in performance accuracy for the different models (for classifying weight, exercise, user) tested based on the data from only {2,4,6} repetitions respectively. We observe the following:

- For identifying the amount of weight used, we obtained an average accuracy of 83.9%, 87.26%, 91.14% respectively with the three different models. While the accuracy is comparatively still high, there is about 6%–14% drop in performance compared to the approach of majority voting across a set (which achieved 97.5% accuracy).
- There was a more significant drop in the accuracy of classifying exercises and users with the early-classification approach. Using the models tested with data from only the first *two reps* of an exercise set, the average accuracy of classification of exercises and users dropped to 61.13% and 54.09% respectively. For exercise classification, the classifier was getting more confused between *lats* vs. *middleback* exercises and *triceps* vs. *reardelts* exercises. When data from first *six reps* of the exercise set was utilized in the test set, the classification accuracy for exercises and users significantly increased to 80% and 72.3% respectively.
- For user classification, 8 subjects (5 novices, 3 intermediates) out of the total 30 users had a classification accuracy lower than 50%. The performance improved to 81% and 72% respectively when data from the first six repetitions were used in training the models.

This analysis suggests that, while delaying the inference till a larger set of repetitions have been observed is likely to result in improved accuracy, it is possible to achieve reasonable accuracy after about 5–6 repetitions. The ERICA system [33] has recently studied the effectiveness of offering personalized, real-time corrective feedback to individuals during workouts. Results in ERICA [33] show that intra-set feedback given after 5 repetitions is favored by individuals and it also results in lesser repetition-level mistakes. This also suggests that *W8-Scope* classifiers can be enhanced to enable such in-situ feedback capabilities to the individuals.

6.8. Feedback accuracy

While the mistake detection module of *W8-Scope* detects mistakes at individual repetition-level, it is worth noting that the system only requires the mistakes to be detected once within a feedback delivery window. The feedback delivery window can vary depending on the application requirement and user preferences. For example, the feedback can be delivered after every repetition, once or twice within an exercise set or after few sets of the same exercise is over. As mentioned earlier, the ERICA system [33] shows that feedback delivered once in a sequence of 5 reps is found to be effective in minimizing the mistakes at repetition-level and is also preferred by individuals over a more frequent rep-level feedback. Given these insights, we study the accuracy of delivering feedback after first 5 repetitions of the exercise

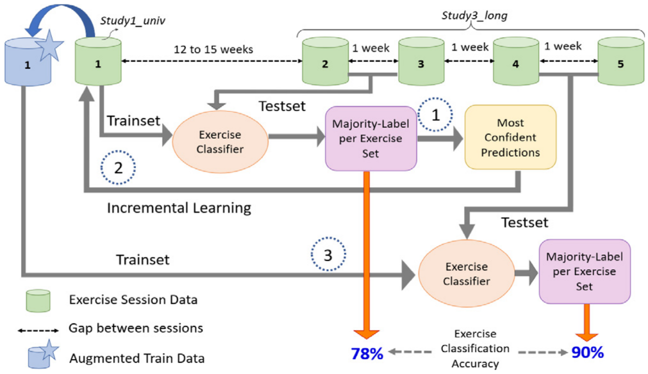


Fig. 15. Incremental learning with longitudinal exercise data.

set. For this, we analyze the result of aggregating the per-repetition output of mistake detector based on two different strategies–(i) provide 'mistake' feedback if at least one repetition has a mistake and, (ii) provide mistake feedback if at least 60% (i.e., majority) of the first 5 repetitions have mistakes and study how accurate such a feedback would be.

For the first strategy, when any of the first five repetitions had a mistake, we obtained a feedback accuracy of 94.2%. When the mistake feedback was delivered only if at least 60% of the first 5 reps (i.e., 3 reps) had mistakes, the feedback accuracy improved to 97.3%. Pulling the weight stack too fast and lifting it only half-way through were the two mistakes that consistently manifested across multiple repetitions.

7. Medium time-scale robustness: Adapting W8-Scope classifiers

Results in Section 6 demonstrate *W8-Scope's* impressive accuracy under real-world usage. However, these results utilize training and test data collected from coterminous (or closely spaced in time) sessions. We now further investigate whether *W8-Scope's* supervised models (especially those based on user-driven motion dynamics, such as exercise or user classification) are able to perform as well over medium-timescales (e.g., across weeks or months), as an individual's exercise pattern evolves over such time periods.

To validate the robustness of our approach across exercise activities that are spaced weeks apart, we initially use the data from first two sessions of $Study3\_long$ (i.e., 10 users performing 5 exercises, across 2 different weeks) as the test set, applying our previously trained models with $Study1\_univ$ data (i.e., from 30 users performing 10 exercises). (*Note:* As illustrated in Fig. 15, $Study1\_univ$ and $Study3\_long$ are separated by a gap of over 3 months, with each of the 4 sessions in $Study3\_long$ occurring in 4 consecutive weeks.) For these two sessions, we obtained an accuracy of 90.5% for weight classification, 78.3% for exercise classification and 75.2% for user classification, when the classifier outputs are ascertained *per-set* (using the "dominant-label" output across all the repetitions of an exercise set). This drop in accuracy, especially for exercise (previously 96.9%) and user classification (previously 98.9%), suggests that a single-shot training of *W8-Scope* classifiers may indeed prove incapable of accommodating the *evolutionary* changes in an individual's exercise patterns. This is further confirmed by training new classifiers using the first two sessions of $Study3\_long$ data, and testing them using the last two sessions. Such coterminous training is able to replicate the higher accuracy values (weight classification = 93.1%, exercise classification = 89.3%, user classification = 90.4%) observed previously, on single sessions, in Section 6.

7.1. Incremental learning

To better incorporate such temporal evolution in exercise dynamics, we propose an enhanced *Incremental Learning-based W8-Scope* framework. Under this approach (Fig. 15 illustrates the specifics, using "exercise classification" as an example, for our dataset), the labeled training data for the initially-trained *W8-Scope* classifier is continually augmented with those unlabeled exercise samples on which the classifier has *high confidence*. Very specifically, our *W8-Scope* instance starts off with the initial labeled training set (the $Study1\_univ$ data). As an individual sporadically visits the gym, *W8-Scope* classifies the observed exercise activities, and then chooses those activity instances whose classification probability exceeds a given threshold t. The modified training set (augmented with such "highly confident" samples) is then used to retrain the classifier (on a per-weekly basis)–this is illustrated in step 2 (indicated within dotted circle) of Fig. 15.

The performance of such incremental learning obviously depends on the right choice of the threshold t. Intuitively, very low values of t will add too many many noisy, likely misclassified, samples to the training set. Conversely, very high values of t will ensure the use of only 'clean' samples, but might suffer from data paucity. We empirically found $t = 0.6$ to provide an appropriate choice between these two extremes.

Table 12
Medium time-scale *W8-Scope* performance (with and without incremental learning).

	Weight	Exercise	User
Without incremental learning	90.5%	78.3%	75.2%
With incremental learning	**95.1%**	**90.2%**	**87.4%**

7.2. Performance results with incremental learning strategy

We present below the changes in the performance of *W8-Scope*, after adopting this incremental learning strategy—i.e., the classification accuracy of activities performed during weeks 3 & 4 of *Study*3*_long*, based on a classifier augmented using 'highly confident' activity samples from weeks 1 & 2.

Weight classification: The accuracy of weight classification was **95.1%**, with a precision and recall of 0.942 and 0.958 respectively. We observed that the classifier performance was poorer for certain heavier weights (e.g., 36.25 kg, 43.75 kg). This is due to both the inability of a *single* magnetic sensor to perform fine-grained differentiation of heavier weights (elaborated further in Section 8), as well as the lack of sufficient training data for heavier weights (most users exercise with lower weights).

Exercise classification: We achieved an average set-level accuracy of **90.2%** (an improvement of over 12%), with a precision of 0.881 and recall of 0.923, in classifying the 5 exercises in the test set. When we analyzed the confusion matrix, we found that *biceps* exercise and *middleback* exercise were the ones typically mis-classifed, as they exhibited the greatest variability in the way these exercises were performed (e.g., high variance in repetition time, repetition height etc.) across various sessions and individuals, with individual often also performing them incorrectly– e.g., not keeping the elbow fixed during *biceps curls* exercises. This dramatic improvement in exercise classification accuracy corroborates our belief that user exercise dynamics do indeed evolve over the span of several weeks and months.

User classification: We achieved an accuracy of **87.4%** (with a precision = 0.845 and recall = 0.893) for discriminating among the 10 users (from a training subject pool of 30 total users) participating in *Study*3*_long*. The somewhat lower values of user classification accuracy were often due to *significant* changes in an individual's exercise style observed from the video feeds—e.g., when performing the *middleback exercise*, a subject initially used a bench to sit and perform the exercise, while in latter sessions, the user performed the same exercise while sitting on the floor and thereby altering the weight stack's overall range of motion.

Table 12 shows the comparative performance of *W8-Scope* without and with incremental learning strategies. Overall, there was an increase of ∼12% in the accuracy of classifying exercises and users after reinforcing the existing training set with such highly confident samples from newly collected exercise data. These results suggest that as long as an individual visits the gym reasonably frequently (e.g., once every 1–2 weeks), *W8-Scope* can evolve its classifier models to capture the evolution of an individual's exercise motion dynamics. We also observe that, at medium time scales, user classification suffers higher loss in accuracy, compared to other metrics. Indeed, we anticipate that user classification accuracy might degrade further as the number of users scale to hundreds & thousands. However, we should note that user identification is the ''least interesting'' of our demonstrated capabilities, as alternative, relatively low user-effort mechanisms (e.g., tapping a smart card on a reader, or entering a user-specific passcode) can achieve this objective.

8. Discussion

While our results demonstrate the promise of our approach of instrumenting gym equipment with low-cost sensors, our work also raises additional questions and possibilities.

Additional sensor instrumentation: In several cases, additional sensors on the weight stack may enable finer-grained discrimination. For example, we experimented with a configuration where two sensors were attached to the weight stack (one at the top and another at the bottom). An expert gym staff member performed *lats* and *middleback* exercises (19 distinct sets of 8 *reps* each) with weights varying between {3.75,48.75} kg on the cable pulley equipment. Across the *entire range* of weight slabs, the use of both top and bottom sensors results in an improved weight classification accuracy of 98%, compared to 92% and 87% when one considers only the top or bottom sensor, respectively. The cost-accuracy trade-off involved in deploying multiple sensors thus needs further investigation.

Extension to additional gym equipment: To study the possible application of the *W8-Scope* approach to other gym equipment, we conducted a small study with 4 users (2 sets, 10 reps) performing 6 different exercises using a sensor-attached dumbbell. By utilizing only the accelerometer sensor data, we obtained an exercise classification accuracy of 85%; however, user identification using this data proved more challenging. In our preliminary work [34], we have recently explored the alternative approach of combining data sensed from an equipment-attached sensor and a more widely-used wearable device (an 'earable') to monitor weight-based exercises by multiple *concurrent* users. Our results show that the combined inertial signals from ear-worn and equipment-mounted sensors can identify the correct {user, equipment} pairings in 83% of the cases, and can help classify exercises (from among 8 distinct choices) with 92% accuracy. These results suggest the promise of exploring techniques that judiciously combine data from sensor-instrumented equipment & wearable devices.

Interleaved usage of equipment: From field observations, we noted that weight stack machines occasionally saw "interleaved usage"–e.g., two users would perform their sets alternately. Our decision to perform exercise classification and user identification on a per-set basis are driven by this observation. In particular, we do not perform any additional 'majority voting' across sets. Of course, different users might also alter the settings of the weight stack during their exercises—such additional features might help to further improve our ability to discriminate among distinct users.

Impact of alterations to gym equipment: In this work we show that individuals' exercise behavior may evolve over time and such changes could be captured by approaches such as incremental-learning. Another factor that may possibly confuse our classifiers would be due to certain artifacts on the gym equipment itself. For example, replacing the cables of the exercise machine with newer ones may make it much stiffer, and consequently, it may affect the way individuals perform the weight training exercises. Additional investigations are required to better understand the impact of such practical situations and explore ways to accordingly fine-tune our approaches.

Real-time, in-situ corrective feedback: While our analysis demonstrates *W8-Scope's* ability to identify a selected set of motion-related exercising mistakes, we do not consider the challenges of incorporating such mistake detection into a practical, real-time system that provides corrective feedback. Such challenges go beyond the drop in early classification studied in Section 6.7, and require a deeper study of how feedback efficacy is affected by the competing objectives of providing *early* and *accurate* feedback. Moreover, the frequency and timing of such feedback (e.g., multiple times within a single set vs. feedback at the end of a set) must also be designed to strike the right balance between minimizing user distraction and being actionable.

The recently proposed ERICA system [33] studies the effectiveness of providing personalized, corrective feedback while performing free-weights exercises. This work examines the impact that the feedback timeliness has on the interplay between the feedback accuracy and its efficacy. It shows that providing intra-set feedback, every 5–6 repetitions, provides a judicious balance of effectiveness, accuracy and user-acceptance. The ERICA system could identify more than 94% of mistakes during the first 5 repetitions of an exercise set and the resulting feedback helped in minimizing the repetition-level mistakes for remaining repetitions by 10%. The user studies conducted also revealed that more frequent feedback (i.e., per repetition) was too much and annoying. These results expand the possibilities for *W8-Scope* system to enable in-situ, corrective feedback by adopting similar strategies.

Identifying incorrect body forms/postures: Weight training requires the user to adhere to specific exercise techniques as well as body forms/postures. Although our proposed approach can track incorrect exercise executions, it is not possible to infer the postural mistakes using only the weight-stack based sensor. To overcome this, we could extend *W8-Scope* by combining it with video-based contour tracking of participants, using privacy-preserving thermal cameras. Such sensor fusion may allow us to track incorrect body postures and provide corrective feedback to prevent serious injuries. In Section 6.3.2, we showed example IR images which captured body contours of the individuals exercising. Prior work [35] also explores the use of thermal-imaging and optical flow techniques to estimate energy expenditure during treadmill exercises. Similar techniques could be extended to infer postural mistakes made by individuals during exercises performed with a weight stack machine.

9. Conclusions and future work

In this paper, we described the design and evaluation of *W8-Scope*, a system which can obtain quantified insights on various exercise-related attributes. We introduce a novel sensing mode (a combination of magnetometer & accelerometer) and sensor location (on top of a weight stack plate) for monitoring weight training exercises. Through extensive user studies conducted with 50 subjects in two real gyms, we consistently obtained an accuracy of 95%+ across all attributes, including the weight used, exercise performed, mistakes made and exercising user. We also show the need to adapt the classification model to accommodate real-world, longitudinal changes in user exercising behaviors, and show that an incremental learning-based approach provides sufficient robustness to our classifiers. As future work, we aim to utilize such low-cost sensing to capture free weights-based exercising behavior (especially in multi-user environments) and then integrate these insights into a mobile application offering gym-goers personalized, *real-time* feedback and recommendations.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This material is supported partially by the National Research Foundation, Prime Minister's Office, Singapore under its International Research Centres in Singapore Funding Initiative, and partially by the National Research Foundation, Prime Minister's Office, Singapore under its IDM Futures Funding Initiative. The authors are grateful for the support of several collaborators and partners at the SMU Gym, Singapore's Government Technology Office and Sport Singapore.

References

[1] Zachary Y Kerr, Christy L Collins, R Dawn Comstock, Epidemiology of weight training-related injuries presenting to united states emergency departments, 1990 to 2007, Amer. J. Sports Med. 38 (4) (2010) 765–771.
[2] Bonnie G Berger, David Pargman, Robert Stephen Weinberg, et al., Foundations of Exercise Psychology, Fitness Information Technology, Inc., 2002.
[3] Dan Morris, T Scott Saponas, Andrew Guillory, Ilya Kelner, RecoFit: using a wearable sensor to find, recognize, and count repetitive exercises, in: Proc. of ACM CHI, 2014.
[4] Bo Zhou, Mathias Sundholm, Jingyuan Cheng, Heber Cruz, Paul Lukowicz, Never skip leg day: A novel wearable approach to monitoring gym leg exercises, in: Proc. of IEEE PerCom, 2016.
[5] D González-Ortega, FJ Díaz-Pernas, Mario Martínez-Zarzuela, Miriam Antón-Rodríguez, A kinect-based system for cognitive rehabilitation exercises monitoring, Comput. Methods Programs Biomed. 113 (2) (2014) 620–631.
[6] Mathias Sundholm, Jingyuan Cheng, Bo Zhou, Akash Sethi, Paul Lukowicz, Smart-mat: Recognizing and counting gym exercises with low-cost resistive pressure sensing matrix, in: Proc. of ACM UbiComp'14, 2014.
[7] Md Fazlay Rabbi, Taiwoo Park, Biyi Fang, Mi Zhang, Youngki Lee, When virtual reality meets IoT in the gym: Enabling immersive and interactive machine exercise, in: Proc. of ACM IMWUT 2018, Vol. 2, ACM, 2018, pp. 78:1–78:21.
[8] Keng-Hao Chang, Mike Y Chen, John Canny, Tracking free-weight exercises, in: Proc. of UbiComp'07, Springer, 2007.
[9] Yousef Kowsar, Masud Moshtaghi, Eduardo Velloso, Lars Kulik, Christopher Leckie, Detecting unseen anomalies in weight training exercises, in: Proc. of ACM OzCHI, 2016.
[10] Christian Seeger, Kristof Van Laerhoven, Alejandro Buchmann, Myhealthassistant: An event-driven middleware for multiple medical applications on a smartphone-mediated body sensor network, IEEE journal of biomedical and health informatics 19 (2) (2015) 752–760.
[11] Chenguang Shen, Bo-Jhang Ho, Mani Srivastava, Milift: Efficient smartwatch-based workout tracking using automatic segmentation, IEEE Trans. Mob. Comput. 17 (7) (2018) 1609–1622.
[12] Bobak Jack Mortazavi, Mohammad Pourhomayoun, Gabriel Alsheikh, Nabil Alshurafa, Sunghoon Ivan Lee, Majid Sarrafzadeh, Determining the single best axis for exercise repetition recognition and counting on smartwatches, in: Proc. of BSN'14, IEEE, 2014, pp. 33–38.
[13] Sizhen Bian, Victor Rey, Peter Hevesi, Paul Lukowicz, Passive capacitive based approach for full body gym workout recognition and counting, in: Proc. of IEEE PerCom, 2019.
[14] Slobodan Milanko, Shubham Jain, Liftright: Quantifying strength training performance using a wearable sensor, Smart Health (2020) 100115.
[15] Chin Guan Lim, Chin Yi Tsai, Mike Y Chen, MuscleSense: Exploring weight sensing using wearable surface electromyography (sEMG), in: Proceedings of the Fourteenth International Conference on Tangible, Embedded, and Embodied Interaction, 2020, pp. 255–263.
[16] Vimo Labs Inc., Trackmyfitness, 2019, http://trackmy.fit (Last Accessed: September 2019).
[17] Atlas wearables, 2019, https://atlaswearables.com (Last Accessed: September 2019).
[18] Samsung galaxy watch active2, 2020, https://www.samsung.com/global/galaxy/galaxy-watch-active2/ (Last Accessed: June 2020).
[19] Andreas Möller, Luis Roalter, Stefan Diewald, Johannes Scherr, Matthias Kranz, Nils Hammerla, Patrick Olivier, Thomas Plötz, Gymskill: A personal trainer for physical exercises, in: Proc. of IEEE PerCom, 2012.
[20] Han Ding, Longfei Shangguan, Zheng Yang, Jinsong Han, Zimu Zhou, Panlong Yang, Wei Xi, Jizhong Zhao, Femo: A platform for free-weight exercise monitoring with rfids, in: Proc. of ACM SenSys'15, ACM, 2015, pp. 141–154.
[21] Biying Fu, Lennart Jarms, Florian Kirchbuchner, Arjan Kuijper, Exertrack—Towards smart surfaces to track exercises, Technologies 8 (1) (2020) 17.
[22] Fu Xiao, Jing Chen, Xiao Hui Xie, Linqing Gui, Juan Li Sun, Wang none Ruchuan, SEARE: A system for exercise activity recognition and quality evaluation based on green sensing, IEEE Trans. Emerg. Top. Comput. (2018).
[23] Xiaonan Guo, Jian Liu, Cong Shi, Hongbo Liu, Yingying Chen, Mooi Choo Chuah, Device-free personalized fitness assistant using wifi, Proc. ACM IMWUT 2 (4) (2018) 165.
[24] Eduardo Velloso, Andreas Bulling, Hans Gellersen, Wallace Ugulino, Hugo Fuks, Qualitative activity recognition of weight lifting exercises, in: Proc. of AH'13, 2013, pp. 116–123.
[25] Rushil Khurana, Karan Ahuja, Zac Yu, Jennifer Mankoff, Chris Harrison, Mayank Goel, Gymcam: Detecting, recognizing and tracking simultaneous exercises in unconstrained scenes, Proc. ACM IMWUT 2 (4) (2018) 185.
[26] Paul Krebs, Dustin T Duncan, Health app use among us mobile phone owners: a national survey, JMIR mHealth uHealth 3 (4) (2015).
[27] Emily Knight, Melanie I Stuckey, Harry Prapavessis, Robert J Petrella, Public health guidelines for physical activity: is there an app for that? A review of android and apple app stores, JMIR mHealth uHealth 3 (2) (2015).
[28] Misha Patel, Aisling Ann O'Kane, Contextual influences on the use and non-use of digital technology while exercising at the gym, in: Proc. of ACM CHI'15.
[29] Dialog Semiconductor, The DA14583 IoT multi sensor development kit, 2018, https://www.dialog-semiconductor.com/iotsensor; Last Accessed: September 2019.
[30] BodyBuilding.com, Cable exercises and exercise guides, 2018, https://www.bodybuilding.com/exercises/equipment/cable; Last Accessed: September 2019.
[31] Hans W. Borchers, Practical numerical math functions, 2019, https://cran.r-project.org/web/packages/pracma/pracma.pdf; Last Accessed: September 2019.
[32] Bo-Jhang Ho, Renju Liu, Hsiao-Yun Tseng, Mani Srivastava, Myobuddy: Detecting barbell weight using electromyogram sensors, in: Proceedings of the 1st Workshop on Digital Biomarkers, ACM, 2017, pp. 27–32.
[33] Meera Radhakrishnan, Darshana Rathnayake, Ong Koon Han, Inseok Hwang, Archan Misra, ERICA: enabling real-time mistake detection & corrective feedback for free-weights exercises, in: Proceedings of the 18th Conference on Embedded Networked Sensor Systems, 2020, pp. 558–571.
[34] Meera Radhakrishnan, Archan Misra, Can earables support effective user engagement during weight-based gym exercises?, in: Proc. of EarComp'19 (Colocated with UbiComp/ISWC), 2019.
[35] Martin Møller Jensen, Mathias Krogh Poulsen, Thiemo Alldieck, Ryan Godsk Larsen, Rikke Gade, Thomas Baltzer Moeslund, Jesper Franch, Estimation of energy expenditure during treadmill exercise via thermal imaging, Med. Sci. Sports Exerc. 48 (12) (2016) 2571–2579.