W8-Scope: Fine-Grained, Practical Monitoring of Weight Stack-based Exercises

Meera Radhakrishnan, Archan Misra and Rajesh Krishna Balan
School of Information Systems, Singapore Management University

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 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. 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.

I. Introduction

There is strong interest in using pervasive & IoT devices to derive fine-grained 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 [11] and reducing the likelihood of early drop out among gym-goers [2]. Most approaches for gym exercise monitoring employ either body-worn, wearable devices (e.g., [17], [29]) or infrastructure-based video sensing ([8]). 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. Moreover, the efficacy of such approaches has typically been evaluated over relatively short observational periods (e.g., 1-2 gym sessions).

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 inexpensive accelerometer and magnetic sensors mounted on the weight stack (as illustrated in Figure 1), to infer various individual-specific, exercise-related attributes.

Fig. 1: Multi-Purpose Cable Pulley Machine and Proposed Sensor Placement on the Weight Stack

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 Figure 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

[1] pronounced Weight-Scope

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).
- *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 in the future.

II. 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. [6] 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 [14], [22] also utilize multiple body-worn inertial sensors to detect different gym exercises. RecoFit [17] 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 [24], [18] present smartwatch-based systems for recognizing and counting repetitions of various gym exercises. Zhou et al. [29] proposed a wearable fabric pressure sensor system that measures the muscle movement, action and repetitions of four leg machine exercises. Recently, Bian et al. [3] introduced a wearable, body capacitance-based sensor for recognizing and counting seven different gym exercises. There are also other emerging apps and wearables such as TrackMyFitness [27] and Atlas Wristband [1] 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. [16] explored the use of a smartphone-based trainer for assessing the quality of exercises performed on a balance board. FEMO [7] is a platform for monitoring dumbbell exercises using passive RFID tags attached to individual dumbbells. Sundholm et al. [25] developed a pressure sensor mat that recognizes and counts repetitions of strength training exercises performed on a mat. The Jarvis system [20] 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 [28], [9] and infrastructure-driven video sensing [8], [26] for exercise activity recognition. SEARE [28] utilizes WiFi CSI waveform-based features to distinguish between 4 exercises. Similarly, Guo et al. [9] 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 [12] 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. [8] developed a 3D vision-based system to track the trajectories of human body parts during psychomotor exercises. Velloso et al. [26] 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.

III. Overall Goals and Approach of W8-Scope

W8-Scope's broader goal is to quantify various attributes related to exercises performed in a gym or a fitness facility. To analyze their own progress, gym-goers are interested in tracking their exercises, weight lifted etc. [15]. A review of physical activity apps found that only 2% provided evidence-based guidelines for resistance training [13]. 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).

A. 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([8], [12]). Our decision to avoid wearables is influenced not just by prior work [19] 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[20], 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, implying the need to identify *robust* features; (iv) Over longer time periods, users exhibit natural "drift" in their exercising styles.

B. 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 [23]), 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 Figure 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

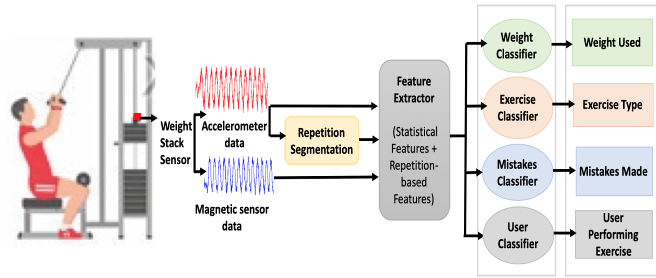


Fig. 2: Overview of *W8-Scope's* Workflow.

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.

IV. 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 Figure 1). This machine has a set of 20 free-weights (each weighing $2.5kg$, except the top-most slab ($1.25kg$)), and permits at least 30 different weight training exercises [4]. Figure 3 shows the position of the exerciser and the weight stack during the upward motion of these ten exercises. In the *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 $7.5kg$ each.

A. 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

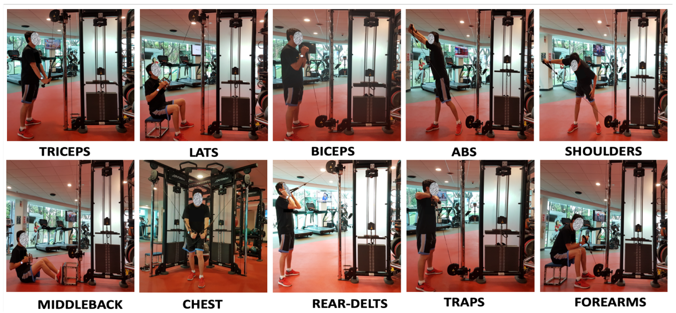


Fig. 3: Exercise positions for 10 exercises (on cable pulley machine)

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).

B. 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.

*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 50Hz. 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 50Hz). 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 minutes for $Study1\_univ$ and for 12 to 24 minutes for $Study2\_comm$. For participating in the study, we provided each participant a monetary compensation of $10.

*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 Figure 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 $3.75kg$ to $16.25kg$). In total, we collected 1148 sets of exercise data. The details of this study are tabulated in column 2 of Table I.

*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 I for summary).

*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. 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 column 4 of Table I).

V. 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.

A. 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.

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 secs) between successive peaks is used to avoid over counting. The segment between two consecutive valleys is assumed to represent a repetition.

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).

TABLE I: Summary of real-world exercise dataset collected from *University* gym and *Community* gym.

	Study1_univ	Study2_comm	Study3_long
No. of participants	35 (23 males, 12 females)	15 (9 males, 6 females)	10 (7 males, 3 females)
Age Variation	21–35 years	18–65 years	21–35 years
Self-rated expertise	13 (Novice); 16 (Intermediate); 6 (Expert)	9 (Novice); 3 (Intermediate); 3 (Expert)	4 (Novice); 4 (Intermediate); 2 (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)	5 (targeted muscles: triceps, biceps, abs, middleback, rear-delts)
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)	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	6 weights ($3.75kg$ to $16.25kg$)	Weights used varied from $5kg$ to $80kg$	Weights used varied from $3.75kg$ to $43.75kg$
Incorrect exercise variations	4 (pulling too fast, releasing too fast, pulling half way through, lifting heavier weight)	N/A	N/A
Average duration of exercise session across subjects	48 minutes	19 minutes	14 minutes
Aggregated duration across all sessions	36 hours 50 minutes	5 hours 46 minutes	8 hours 20 minutes

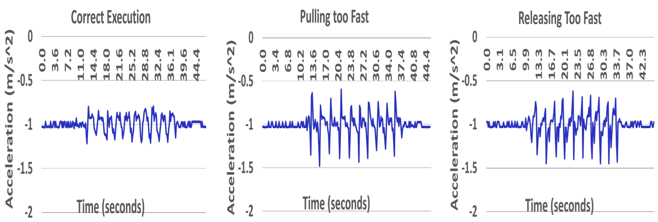


Fig. 4: Variation in accelerometer readings while performing Triceps Pushdown exercise (a) correctly, (b) by pulling weights too fast & (c) by releasing/slamming down the weights fast

To compute the weight stack displacement, we first extracted the z-axis acceleration signal, integrated it using cumulative trapezoidal integration [5] to obtain velocity, then low-pass filtered and then integrated again to obtain the displacement. As shown in Section V-E, this approach results in a mean displacement error of ±1.15 cm.

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 Figure 4) that the accelerometer data contains visible *signatures*, that can help distinguish between such correct and incorrect execution patterns (as shown later in Section V-E).

B. 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.

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 i^{th} 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). Equation 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.

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

Accordingly, 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).

Figure 5 shows the variation in magnetic field while performing 10 repetitions each of *lats* exercise with 9 different set of weights ranging from $3.75kg$ to $23.75kg$. 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 -255μT for $w = 21.25kg$ or $w = 23.75kg$).

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 ($3.75kg$, $8.75kg$, $13.75kg$) lifted to 4 different controlled heights ($6cm$, $12cm$, $18cm$, $24cm$). We observed that the change in magnetic field for weight, $w = 8.75kg$ and height, $h = 6cm$ looked very similar to that of $w = 13.75kg$ and $h = 24cm$ (mean and total changes being approx. 45μT and 32μT 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.

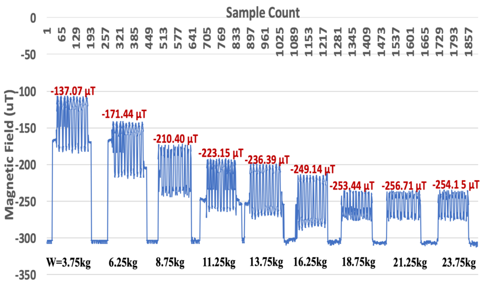


Fig. 5: Magnetic field variation 9 sets of 10 reps of **lats** exercise; weight varied between $w = \{3.75, 23.75\}kg$. (The number on top shows the mean value of the magnetic field (μT) for each set).

TABLE II: Features extracted from each time window of accelerometer and magnetometer data

Feature	Description
Mean	Average of the values for the time window for each axes and the Euclidean norm (magnitude) of the signal
Max	Maximum value in a time window for each axis and signal's magnitude
Min	Minimum value in a time window for each axis and signal's magnitude
Range	Total change in values within the time window for each axis and signal's magnitude
Variance	Variance of the values in a time window for each axis and magnitude of signal
Spectral Entropy	Normalized information entropy of the FFT components of each axis and magnitude of signal
Spectral Energy	Mean value of the square of the FFT coefficients of the signal for each axis and magnitude value
Covariance	Covariance between each pair of axes of the sensor
Correlation	Correlation between each pair of axes of the sensor
Repetition Time	Average time taken to complete a repetition in a exercise set
Repetition Height	Average height to which the weight stack was lifted within a set
Repetition Velocity Mean	Average of the speed with which the weight stack was lifted in a set
Repetition Velocity Std.dev	Standard deviation of the speed with which the weight stack was lifted in a set

C. 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, and (iii) the combination of accelerometer and magnetometer readings can help identify the amount of weight that is being lifted.

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 = 2secs, 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 V-A, 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 II for the complete set of features used in our classifier models.

D. 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., [10], [26])) throughout our multi-stage pipeline. The key components in the classification pipeline (see Figure 6) are as follows:

- *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 = $[w1, w2, ..., wn]$) for each distinct *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.

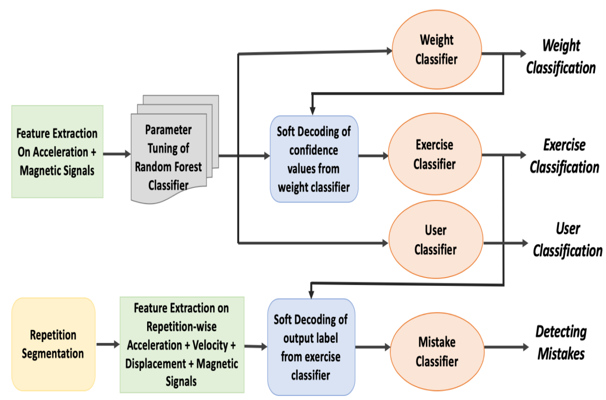


Fig. 6: The *W8-Scope* Multi-Stage Classification Pipeline

- *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 V-A1. 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 30-40secs). 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*.

E. Initial Validation Results

We now present summarized results on the performance of different *W8-Scope* components, evaluated on validation studies (explained earlier in Section IV-A). The repetition

TABLE III: 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	$\pm\mathbf{0.67}$ cm	$\pm\mathbf{0.87}$ cm	$\pm\mathbf{1.1}$ cm	$\pm\mathbf{1.96}$ cm

TABLE IV: 10-Fold Cross Validation Results of *W8-Scope* Classifier Models

	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%**

counting mechanism (Section V-A1) achieves an accuracy of $\mathbf{98\%}$ in counting the 10 repetitions in each set. For displacement computation, we observed an average estimation error of $\pm\mathbf{1.15}\boldsymbol{cm}$ compared to the ground truth height. Table III shows the breakdown of the average error in displacement computed for each height. Additional results (summarized in Table IV) show that the combination of accelerometer and magnetic sensing features hold promise in achieving high accuracy (over $\mathbf{97\%}$ using 10-fold cross validation) in inferring different exercise-related attributes.

VI. 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 IV-B) 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).

A. Counting Repetitions

We first evaluate the performance of *repetition counting*. Using 908 sets of data collected from different weights and different exercises experiment in $\boldsymbol{Study1\_univ}$, we obtained a performance of $\mathbf{97\%}$ 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 $\pm\mathbf{1}$, 9 sets are off by $\pm\mathbf{2}$, 4 set are off by $\pm\mathbf{3}$, 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 artefacts, 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 $\boldsymbol{Study2\_comm}$, the repetitions were accurately counted for 177 sets (98% accuracy), indicating that this estimation was accurate across gym environments.

B. Identify the Amount of Weight Lifted

We evaluate the performance of weight classification on different weights' data obtained from $\boldsymbol{Study1\_univ}$. 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 $\mathbf{97.5\%}$ in distinguishing between six

TABLE V: Performance of **amount of weight** identification

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

TABLE VI: Performance of identifying **exercise performed**

Exercise Classification	Accuracy	Precision	Recall
10-fold CV ($\boldsymbol{Study1\_univ}$)	96.93%	0.962	0.969
10-fold CV ($\boldsymbol{Study2\_comm}$)	97.79%	0.978	0.982

set of weights, w=[$\mathbf{3.75, 6.25, 8.75, 11.25, 13.75, 16.25}$] in the weight stack, with the classification error confined to the heavier weights – $\mathbf{13.75}\boldsymbol{kg}$ and $\mathbf{16.25}\boldsymbol{kg}$.

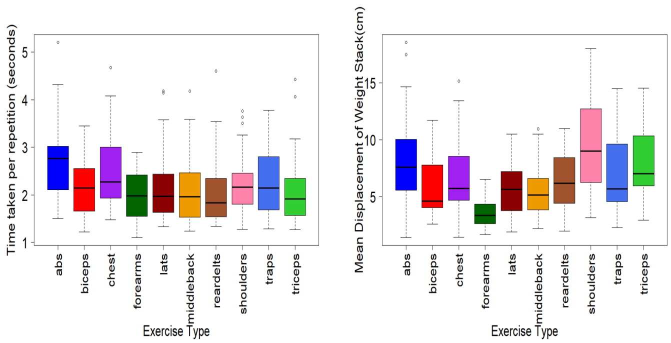
We also performed a *leave-one-subject-out cross validation* (LOOCV) in which the *weight-classification* model was trained with data from all users, except the test user, and then tested on the data from test user. Using this approach, we obtained an average accuracy of $\mathbf{93.75\%}$, with a precision 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=$\mathbf{16.25}\boldsymbol{kg}$. Table V presents the summary of results from weight classifier.

C. 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 $\boldsymbol{Study1\_univ}$. We obtained a performance accuracy of $\mathbf{96.93\%}$, 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, 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, Figure 7a 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$2.65 secs) and *rear-delts* exercise took the least amount of time ($\leq$2 secs). Similarly, Figure 7b 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 $\boldsymbol{Study2\_comm}$ data (where users performed exercises using exercise-specific weight machines) by withholding the machine label. We applied a 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 VI summarizes the performance of exercise classifier.



(a) Average **time taken** to complete repetition

(b) Average **height** to which weight-stack was lifted

Fig. 7: Variation in repetition time & height per exercise (across subjects)

TABLE VII: Performance of identifying **mistakes made**

Mistakes Classification	Accuracy	Precision	Recall
10-fold CV (***Study1_univ***)	96.75%	0.968	0.967
LOOCV (***Study1_univ***)	79.2%	0.78	0.82

D. 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 V-D) of two exercises (triceps and lats) from 30 subjects in ***Study1_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 VII.

E. Identify Users Performing Exercises

W8-Scope's final component helps to distinguish between the different users performing the same exercise. Table VIII 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.

TABLE VIII: 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

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.

F. Performance Comparison: W8-Scope vs Smartwatch

Using the ***Study1_univ*** data, we compared (and summarize in **Table IX**) 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=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.

TABLE IX: 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%

VII. Medium Time-Scale Robustness: Adapting W8-Scope Classifiers

Results in Section VI demonstrate *W8-Scope's* impressive accuracy under real-world usage. However, these results were based on the use of training and test data from coterminous (or closely spaced in time) sessions. It is natural to ask whether *W8-Scope's* supervised models (especially those based on user-driven motion dynamics, such as exercise or user classification) will remain valid 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 Figure 8, ***Study1_univ*** and ***Study3_long*** are separated by a gap of over 3 months, with each of the 4 sessions in $\boldsymbol{Study3\_long}$ 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 $\boldsymbol{Study3\_long}$ 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 VI.

A. Incremental Learning

To better incorporate such temporal evolution in exercise dynamics, we propose an enhanced *Incremental Learning-based W8-Scope* framework. Under this approach (Figure 8 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 $\boldsymbol{Study1\_univ}$ 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 $\boldsymbol{t}$. 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 Figure 8.

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

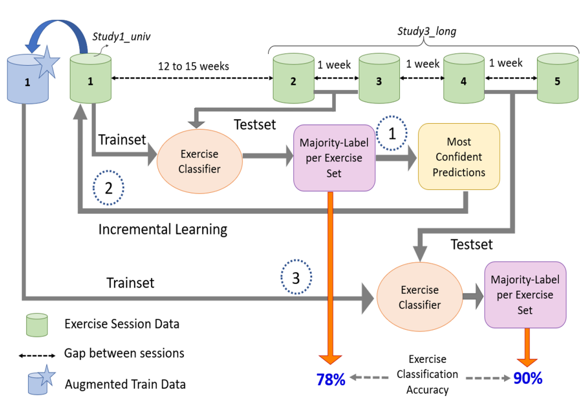


Fig. 8: Incremental Learning with Longitudinal Exercise Data

TABLE X: 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%**

B. Performance Results with Incremental Learning Strategy

Table X shows the comparative performance of *W8-Scope* without and with incremental learning strategies. Overall, there was an increase of $\sim$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.

VIII. 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 [21], 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.

IX. 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.

Acknowledgements

This material is supported partially by the National Research Foundation, Prime Ministers Office, Singapore under its International Research Centres in Singapore Funding Initiative, and partially by the National Research Foundation, Prime Ministers Office, Singapore under its IDM Futures Funding Initiative. The authors are grateful for the support of several collaborators and partners, including Roy Chua, Gary Chua at the SMU Gym, Tony Varghese from Singapore's Government Technology Office and Serene Yang from Sport Singapore.

References

[1] Atlas wearables. https://atlaswearables.com; Last Accessed: September 2019.

[2] Bonnie G Berger, David Pargman, Robert Stephen Weinberg, et al. *Foundations of exercise psychology.* Fitness Information Technology, Inc., 2002.

[3] Sizhen Bian, Victor Rey, Peter Hevesi, and Paul Lukowicz. Passive capacitive based approach for full body gym workout recognition and counting. In *Proc. of IEEE PerCom*, 2019.

[4] BodyBuilding.com. Cable exercises and exercise guides, 2018. https://www.bodybuilding.com/exercises/equipment/cable; Last Accessed: September 2019.

[5] Hans W. Borchers. Practical numerical math functions, 2019. https://cran.r-project.org/web/packages/pracma/pracma.pdf; Last Accessed: September 2019.

[6] Keng-Hao Chang, Mike Y Chen, and John Canny. Tracking free-weight exercises. In *Proc. of UbiComp'07.* Springer, 2007.

[7] Han Ding, Longfei Shangguan, Zheng Yang, Jinsong Han, Zimu Zhou, Panlong Yang, Wei Xi, and Jizhong Zhao. Femo: A platform for free-weight exercise monitoring with rfids. In *Proc. of ACM SenSys'15*, pages 141–154. ACM, 2015.

[8] D González-Ortega, FJ Díaz-Pernas, Mario Martínez-Zarzuela, and Miriam Antón-Rodríguez. A kinect-based system for cognitive rehabilitation exercises monitoring. *Computer methods and programs in biomedicine*, 113(2):620–631, 2014.

[9] Xiaonan Guo, Jian Liu, Cong Shi, Hongbo Liu, Yingying Chen, and Mooi Choo Chuah. Device-free personalized fitness assistant using wifi. *Proc. of ACM IMWUT*, 2(4):165, 2018.

[10] Bo-Jhang Ho, Renju Liu, Hsiao-Yun Tseng, and Mani Srivastava. Myobuddy: Detecting barbell weight using electromyogram sensors. In *Proceedings of the 1st Workshop on Digital Biomarkers*, pages 27–32. ACM, 2017.

[11] Zachary Y Kerr, Christy L Collins, and R Dawn Comstock. Epidemiology of weight training-related injuries presenting to united states emergency departments, 1990 to 2007. *The American Journal of Sports Medicine*, 38(4):765–771, 2010.

[12] Rushil Khurana, Karan Ahuja, Zac Yu, Jennifer Mankoff, Chris Harrison, and Mayank Goel. Gymcam: Detecting, recognizing and tracking simultaneous exercises in unconstrained scenes. *Proc. of ACM IMWUT*, 2(4):185, 2018.

[13] Emily Knight, Melanie I Stuckey, Harry Prapavessis, and 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 and uHealth*, 3(2), 2015.

[14] Yousef Kowsar, Masud Moshtaghi, Eduardo Velloso, Lars Kulik, and Christopher Leckie. Detecting unseen anomalies in weight training exercises. In *Proc. of ACM OzCHI*, 2016.

[15] Paul Krebs and Dustin T Duncan. Health app use among us mobile phone owners: a national survey. *JMIR mHealth and uHealth*, 3(4), 2015.

[16] Andreas Möller, Luis Roalter, Stefan Diewald, Johannes Scherr, Matthias Kranz, Nils Hammerla, Patrick Olivier, and Thomas Plötz. Gymskill: A personal trainer for physical exercises. In *Proc. of IEEE PerCom*, 2012.

[17] Dan Morris, T Scott Saponas, Andrew Guillory, and Ilya Kelner. Recofit: using a wearable sensor to find, recognize, and count repetitive exercises. In *Proc. of ACM CHI*, 2014.

[18] Bobak Jack Mortazavi, Mohammad Pourhomayoun, Gabriel Alsheikh, Nabil Alshurafa, Sunghoon Ivan Lee, and Majid Sarrafzadeh. Determining the single best axis for exercise repetition recognition and counting on smartwatches. In *Proc. of BSN'14*, pages 33–38. IEEE, 2014.

[19] Misha Patel and 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*.

[20] Md Fazlay Rabbi, Taiwoo Park, Biyi Fang, Mi Zhang, and Youngki Lee. *When Virtual Reality Meets IoT in the Gym: Enabling Immersive and Interactive Machine Exercise*, volume 2. ACM, 2018.

[21] Meera Radhakrishnan and Archan Misra. Can earables support effective user engagement during weight-based gym exercises? In *Proc. of EarComp'19 (colocated with UbiComp/ISWC)*, 2019.

[22] Christian Seeger, Kristof Van Laerhoven, and 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):752–760, 2015.

[23] Dialog Semiconductor. The da14583 iot multi sensor development kit, 2018. https://www.dialog-semiconductor.com/iotsensor; Last Accessed: September 2019.

[24] Chenguang Shen, Bo-Jhang Ho, and Mani Srivastava. Milift: Efficient smartwatch-based workout tracking using automatic segmentation. *IEEE Transactions on Mobile Computing*, 17(7):1609–1622, 2018.

[25] Mathias Sundholm, Jingyuan Cheng, Bo Zhou, Akash Sethi, and Paul Lukowicz. Smart-mat: Recognizing and counting gym exercises with low-cost resistive pressure sensing matrix. In *Proc. of ACM UbiComp'14*, 2014.

[26] Eduardo Velloso, Andreas Bulling, Hans Gellersen, Wallace Ugulino, and Hugo Fuks. Qualitative activity recognition of weight lifting exercises. In *Proc. of AH'13*, pages 116–123, 2013.

[27] Vimo Labs Inc. Trackmyfitness. http://trackmy.fit; Last Accessed: September 2019.

[28] Fu Xiao, Jing Chen, Xiao Hui Xie, Linqing Gui, Juan Li Sun, and Wang none Ruchuan. Seare: A system for exercise activity recognition and quality evaluation based on green sensing. *IEEE Trans. on Emerging Topics in Computing*, 2018.

[29] Bo Zhou, Mathias Sundholm, Jingyuan Cheng, Heber Cruz, and Paul Lukowicz. Never skip leg day: A novel wearable approach to monitoring gym leg exercises. In *Proc. of IEEE PerCom*, 2016.