Enhancing Strength Training for Fitness Trackers

Isha Saikumar MS in Computer Science isaikuma@iu.edu Nitin Chowdary Koduru MS in Information Science nkoduru@iu.edu

Rahul Palivela MS in Data Science rpalive@iu.edu

April 29, 2024

GitHub Repository: https://github.iu.edu/nkoduru/Data-Mining-Final-Project

Abstract

In the realm of strength training, automated tracking mechanisms for free weight exercises remain underexplored. This study delves into context-aware applications by analyzing data from wristband accelerometers and gyroscopes during strength training sessions. Our dataset comprises observations from 5 participants engaged in diverse barbell exercises. The objective is to develop and assess models capable of imitating human personal trainers—identifying exercises, counting repetitions, and detecting improper form. Employing supervised learning techniques, we evaluate various machine learning algorithms to identify effective models.

1 Introduction

Over the past decade, significant progress has been made in overcoming practical challenges related to wearable sensors like accelerometers, gyroscopes, and GPS receivers. These advancements have paved the way for monitoring and classifying human activity using smartwatches and other wearables. The field of pattern recognition and machine learning has witnessed substantial growth due to the commercial potential of context-aware applications and user interfaces. Furthermore,

activity recognition holds promise in addressing critical societal issues such as rehabilitation, sustainability, elderly care, and overall health.

Historically, efforts to promote healthier lifestyles have centered around tracking movements and gathering user feedback through exercise management systems. These systems partially replace the functions traditionally performed by personal trainers. For instance, in aerobic exercises like bicycling, swimming, and running, we have seen the use of accelerometers and GPS-based pedometers to monitor pace and distance, ECG monitors to track exertion, and electronic exercise machines (such as treadmills and elliptical trainers). However, when it comes to free weight exercises, comprehensive tracking mechanisms are still lacking. Currently, only one fitness wearable brand claims to automatically identify exercises and record repetitions, but the literature on context-aware applications within strength training remains limited, likely due to untapped commercial potential.

As technology continues to evolve, the vision of fully digital personal trainers becomes more attainable. To define the requirements for such digital trainers, we must examine the roles of existing personal trainers. These roles encompass not only knowledge of human anatomy, exercise science, and nutrition but also the ability to design tailored exercise programs that motivate individuals effectively. Additionally, personal trainers play a crucial role in tracking workouts to ensure proper form and progressive overload, ultimately helping clients achieve their fitness goals safely.

While recent years have witnessed significant progress in digitizing the first two roles of personal trainers, the third role—ensuring safety and progress—remains in-adequately addressed in today's fitness wearables and applications. Consequently, this paper aims to explore the untapped potential of context-aware applications within the strength training domain. Specifically, we analyze accelerometer and gyroscope data collected from wristbands during free weight workouts. Our dataset includes information from five participants performing various barbell exercises with medium to heavy weights. Our goal is to develop and evaluate models that mimic human personal trainers by tracking exercises, counting repetitions, and identifying improper forms. We base our approach on the work of Hoogendoorn and Funkuse, employing supervised learning techniques for classification. By comparing accuracies across various machine learning algorithms, we aim to identify the most effective models for this context-aware application.

2 Related Work

The initial work on activity recognition using wearable sensors was conducted by Van Laerhoven in 2000 [7]. In their study, they connected accelerometers to a pair of pants and a laptop to interpret raw sensor data collected during daily activities. By employing machine learning techniques such as Kohonen maps and probabilistic models, they built a system capable of recognizing various activities. Even today, recognizing daily activities remains a heavily researched area, as evidenced by subsequent studies. These findings have been widely incorporated into commercial products like Fitbit [13], Apple Watch [15], and Samsung Gear [14]. Additionally, modern smartphones come equipped with features for tracking fitness activities such as walking, running, and cycling, provided by Google [16] and Apple [17]. Some newer devices and applications even support the recognition of specific sports activities like aerobic workouts, elliptical training, and swimming. Fitbit's SmartTrack feature automatically identifies and records workouts, capturing details like activity duration, calories burned, and heart rate zones [18]. While these products already introduce the concept of a digital personal trainer, there remains a lack of support for free weight exercises.

Several studies have delved into the realm of recognizing gym exercises using wearable sensor data. Notably, Chang et al. equipped a workout glove with a three-axis accelerometer to monitor hand movements, while another accelerometer was placed on users' waists to track body posture. Their data collection involved ten different subjects performing three sets of fifteen repetitions for nine distinct exercises, each with varying weights. To identify these exercises, they evaluated two methods: Naive Bayes Classifier and Hidden Markov Models. Remarkably, both approaches achieved an overall recognition accuracy of approximately 90% based on data from individual users. When cross-validating training results with new user data, the accuracy remained around 85%. Additionally, they developed a peak counting algorithm to accurately tally repetitions, achieving a mere 5% miscount rate.

Koskimki et al. [21] highlighted a limitation in prior research—the lack of exercise variety. In their study, they employed two accelerometers—one on the left wrist and another on the torso. Their goal was to create automatic activity diaries that reliably indicated the number of sets performed for a given exercise. Data collection involved a single participant performing thirty different exercises, each comprising three sets of ten repetitions. The resulting personalized model achieved an impressive average true positive rate of 96%. However, when tested against a separate validation set, accuracies significantly declined for certain exercise classes.

Li et al. [19] introduced a novel recognition method based on a single accelerometer attached to a glove. Instead of fixed-length windows, they divided the filtered acceleration data stream into unequal time series for peak analysis. Dynamic Time Warping was then employed to recognize exercises. Their experiments involved three male and one female subject, each performing exercises across three sets with varying weights. The proposed approach demonstrated feasibility and achieved commendable performance.

The research outcomes indicate significant promise in recognizing exercises using accelerometer data. However, when considering strength training and body-building, certain critical aspects related to weight training have been overlooked. Specifically, the principles of progressive overload and proper form deserve closer examination.

Progressive overload, a fundamental concept in strength training, involves gradually increasing the stress on the body during exercise. This can be achieved by adjusting either the weights or the repetitions over time. Previous studies mentioned various weight levels, but the relationship between weights and repetitions was not proportional. Regardless of the weight used, subjects were consistently instructed to perform the same number of repetitions. Notably, the study participants completed sets with either 10 or 15 repetitions, which significantly exceeds the typical practice of limiting sets to 5 repetitions or fewer [24].

Additionally, the studies did not address the perceived intensity of the exercises. For context-aware applications intended for use during workouts, training data should better reflect realistic scenarios. Furthermore, the quality of exercise execution especially proper form remains unexplored. Ensuring correct form is crucial in free weight training and is an essential feature for any digital personal trainer, as improper form can lead to serious injuries, particularly when using heavier weights.

3 Methodology

3.1 Data Collection

In contrast to previous studies that primarily focused on accelerometer data, contemporary smart devices incorporate an array of additional sensors, including gyroscopes. These sensors, however, have been placed in unconventional locations, such as wristbands or workout gloves. To transition these findings into a possible commercial product, it is prudent to integrate them into a practical device like a smartwatch, which already houses the necessary sensors.

For our research, we employed the MbientLabs wristband sensor research kit [12]. This wristband emulates the placement and orientation of a watch, facilitating controlled experiments. Data collection occurred using the default sensor settings: accelerometer at 12.500Hz and gyroscope at 25.000Hz. To create a comprehensive dataset, we enlisted five participants to perform barbell exercises in three sets of five repetitions. Additionally, we conducted another training session with the same participants, this time using three sets of ten repetitions. These higher-rep sets allow us to explore how models generalize across varying weights.

In total, we gathered 187 sets of data. During intervals between sets, we also collected resting data. Importantly, test subjects experienced no restrictions during these resting periods, resulting in a mix of standing, walking, and sitting. This resting data will be instrumental in analyzing the transition from rest to exercise.

Participant	Gender	Age	Weight	(Kg) Height (cm)	Experience (years)
A	Male	23	95	194	5+
В	Male	24	76	183	5+
$^{\mathrm{C}}$	Male	16	65	181	<1
D	Male	21	85	197	3
E	${\bf Female}$	20	58	165	1

Figure 1: Participants Demographics

3.2 Weight

In the context of weight training, determining the appropriate load for exercises hinges on the concept of the one-rep max (1RM). The 1RM represents the maximum weight an individual can lift for a single repetition [27]. Calculating the 1RM can be achieved directly through maximal testing or indirectly via submaximal estimation. The latter approach is favored due to its safety and efficiency.

Various formulas exist for estimating 1RM using the submaximal method, with the Epley and Brzycki equations being the most prevalent [26]. For our experiment, we employed Epley's formula:

$$1RM = w\left(1 + \frac{r}{30}\right)$$

Here, r denotes the number of repetitions performed, and w represents the weight lifted. Once the 1RM is determined for each exercise, Epley's formula can further guide us in calculating the appropriate weight for 5 or 10 repetitions. Specifically, this corresponds to approximately 85% (for 5 reps) and 75% (for 10 reps) of the 1RM. By adhering to this method, we ensure that participants engage in exercises tailored to their individual strength levels.

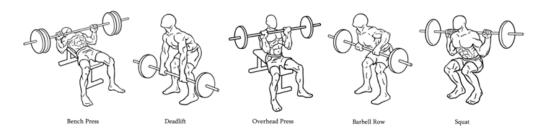


Figure 2: Exercise Types

3.3 Outlier Detection

During strength training sessions, participants often engage in auxiliary movements such as walking or resting between sets. These incidental movements introduce noise into the accelerometer and gyroscope data, potentially leading to outliers. To ensure the integrity of our dataset, we employed Chauvenet's outlier detection method.

Chauvenet's criterion is a statistical technique that identifies data points significantly deviating from the expected distribution. By applying this method, we systematically identified and removed values considered outliers. Our rigorous approach ensured that the dataset remained uniformly distributed, allowing us to confidently utilize Chauvenet's outlier detection for data cleaning.

This meticulous process contributes to the robustness of our analysis, enhancing the accuracy of exercise classification, repetition counting, and form detection models. As aspiring researchers, we recognize the importance of data quality in advancing the field of wearable technology for strength training.

3.4 Handling Missing Values

In our study, preserving data integrity was paramount. To achieve this, we employed an imputation method specifically, interpolation to fill in missing values

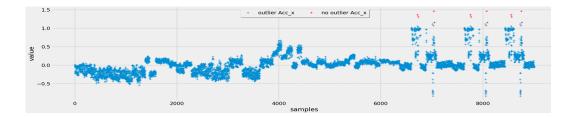


Figure 3: Chauvnet's Outlier Detection

within our dataset.

The initial dataset consisted of 69,677 entries, each capturing epoch timestamps and triaxial accelerometer (x, y, z) measurements. These individual sensor readings were organized into separate files, each associated with a unique timestamp. Aggregating this granular data was essential to minimize information loss.

For continuous numerical attributes, such as accelerometer readings, we aggregated the data using the mean. This approach ensured that the imputed values aligned with the overall trend of the exercise. Labels and categorical attributes were aggregated using the mode, preserving the dominant characteristics of each exercise type.

The resulting aggregated dataset served as a robust foundation for subsequent visualizations, model training, and analysis. Our meticulous approach to data imputation contributes to the reliability of our findings and enhances the accuracy of our machine learning models.

3.5 Feature Engineering

3.5.1 Low-pass Filter:

The low-pass filter is a valuable tool for processing temporal data, assuming the presence of periodicity. Specifically, the Butterworth low-pass filter effectively mitigates high-frequency noise within the dataset, thereby enhancing the learning process. In our case, we applied this filter to all attributes except the target features.

Upon visual examination, we observed that the movements exhibit a frequency of approximately 2 seconds per repetition. To fine-tune the filter, we conducted further visual inspections, guided by the trial-and-error approach outlined by van den Bogert. Ultimately, we set the cutoff frequency at 1.3 Hz.

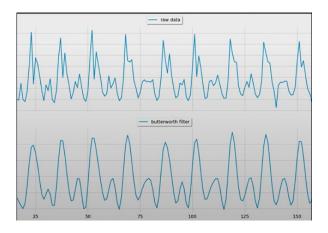


Figure 4: Low-pass Filter to Mitigate Noise

3.5.2 Principal Component Analysis:

In our study, we embarked on a comprehensive exploration of feature reduction techniques, with a specific focus on Principal Component Analysis (PCA). Prior to delving into PCA, we employed the elbow technique to ascertain the optimal number of principal components. Remarkably, the visual analysis of our results unequivocally pointed to three components as the ideal choice.

The subsequent PCA transformation yielded three distinct components: pca_1, pca_2, and pca_3. These components encapsulate the essential variance within our feature space, allowing us to succinctly represent the underlying data.

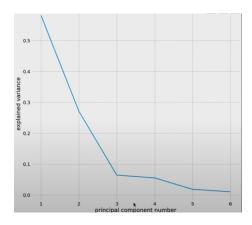


Figure 5: Elbow Curve for PCA

gyr_x	gyr_y	gyr_z	participant	label	category	set	duration	pca_1	pca_2	pca_3
96395	2.439080	0.939616								
29685										
		-0.929804								0.065980
	-1.219897								-0.095463	
					medium					

Figure 6: Principal Component Analysis

3.6 Clustering:

we investigate the clustering of movement patterns using K-means, leveraging the elbow method to determine the optimal number of clusters. Our analysis reveals that five clusters provide an effective representation of the underlying data.

We visualize the results through a 3D scatter plot, contrasting predicted values with actual observations. Notably, the proximity of gray and green points in the plot corresponds to the similarity between deadlift and barbell row exercises, as their sensor readings exhibit comparable patterns. Similarly, bench press and overhead press movements align due to their shared characteristics.

Our modeling efforts encompass a diverse set of algorithms, including neural networks, random forests, decision trees, naïve Bayes, and k-nearest neighbors (KNN). To enhance feature representation, we construct four distinct feature sets:

- 1. Raw Sensor Data (1st Feature Set): This includes accelerometer and gyroscope measurements along the x, y, and z axes.
- 2. Enriched Features (2nd Feature Set): Building upon the raw data, we incorporate square features, capturing statistical moments (mean and standard deviation), and apply principal component analysis (PCA).
- 3. **Temporal Abstraction (3rd Feature Set):** We extend the second feature set by introducing time-related features, achieved through temporal abstraction techniques.

4. Frequency-Domain Features (4th Feature Set): Finally, we augment the third feature set with Fourier transformation-based features and cluster-based representations.

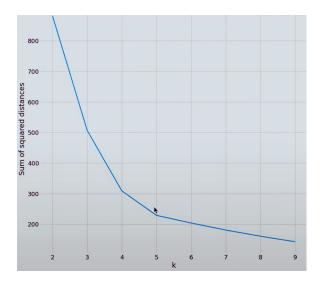


Figure 7: Elbow Curve for K-Means

Figure 8: K-Means Cluster

3.7 Counting Repetitions

In the context of monitoring exercise movements for repetition counting, the decision between accelerometers and gyroscopes hinges on the clarity and simplicity of the data they yield. Accelerometers, which measure linear acceleration, provide a direct understanding of the movement of the monitored device or body part. This straightforwardness results in clearer waveforms, particularly advantageous for repetitive exercises. In contrast, gyroscopes capture angular velocity, leading to potentially noisy readings due to the intricate nature of human motion and changes in device orientation. The challenge lies in interpreting these intricate signals to extract meaningful patterns

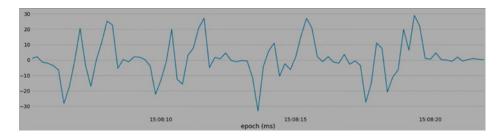


Figure 9: Gyroscope Readings

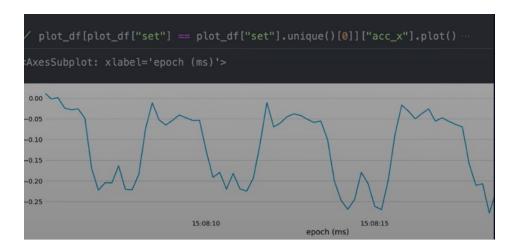


Figure 10: Accelerometer Readings

To enhance the accuracy of repetition counting, a low-pass filter is applied to the accelerometer data. This filter effectively reduces high-frequency noise and fluctuations, resulting in a clearer waveform. By analyzing these smoothed waveforms often characterized by peaks or extremas, we can accurately count repetitions based on the distinct motion signatures exhibited during each repetition. This approach offers a practical solution for automated repetition counting, capitalizing on the directness of accelerometers and the refinement achieved through signal processing techniques.

4 Results

4.1 Classification

Given the temporal nature of our dataset, we partitioned it into training and test sets based on exercise sets. Specifically, the training data comprises the first two sets for each exercise, weight, and participant combination, while the test set contains the remaining sets. This approach ensures that we have valid test data that the models have not encountered previously.

To optimize model performance, we employed forward feature selection. This technique allowed us to identify the most influential features, as irrelevant ones could adversely affect algorithm performance. By iteratively enhancing a simple decision tree with the best features, we observed that performance ceased to significantly improve after considering 15 features. Among these, the five most predictive features are: pca_1 , acc_y , pca_3 , gyr_x , $temp_{std}$, and $acc_{r_{pse}}$.

To mitigate model complexity, we introduced a regularizer into the objective functions. Notably, the accuracy of the test set slightly improves with higher regularization, but only up to a certain point. Beyond that threshold, accuracy decreases for both the training and test sets.

Our initial test run evaluated the performance of several models and features. The models considered included Neural Networks, Random Forests, Support Vector Machines, K-nearest Neighbors, Decision Trees, and Naive Bayes. We conducted a grid search across all models to fine-tune their parameters.

4.2 Classification Results:

The model performance results reveal distinct patterns for different feature sets, as outlined in [2]. Notably, the feature set encompasses all available features. Based

on these findings, we further optimized a Random Forest model using the 15 best-performing features.

Employing a 5-fold cross-validation approach, we fine-tuned the model parameters through grid search. The resulting optimal parameters for the Random Forest were as follows: minimum samples per leaf: 2, n-estimators: 100, and criterion: gini. Subsequently, we evaluated the model's performance using a confusion matrix, which demonstrated an impressive overall accuracy of 99.58%.

/	score_df	.sort_values(by=	"accuracy"	, ascending=False)
	model	feature_set	accuracy	
1		Feature Set 4	0.995863	
0			0.994829	
3	DT	Feature Set 4	0.991727	
0			0.986556	
1		Selected Features	0.982420	
1			0.982420	
0		Selected Features	0.982420	
2	KNN		0.972079	
3	DT	Selected Features	0.967942	
4		Feature Set 4	0.963806	
1			0.955533	
1			0.950362	
3	DT		0.946225	
0			0.937952	
4		Feature Set 3	0.935884	
3	DT		0.930714	
3	DT		0.929679	
0			0.928645	
4			0.927611	
2	KNN		0.922441	
2	KNN		0.870734	
4				
4			0.854188	
2	KNN			
2	KNN	Feature Set 2	0.789038	

Figure 11: Results After Modeling

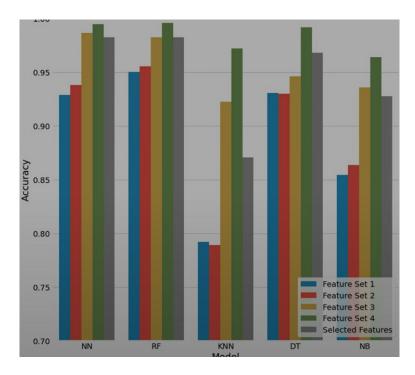


Figure 12: Graphical Representation of Results

4.3 Counting Repetitions:

To quantify repetitions, we employed a straightforward peak counting algorithm on the scalar magnitude acceleration data. To mitigate the impact of small local peaks, we initially applied a robust low-pass filter with a cutoff frequency of 0.4 Hz. Notably, the effectiveness of this repetition counting method necessitates individual adjustments for specific exercises to achieve optimal performance. Overall, the error rate for counting repetitions across the collected dataset stood at approximately 5%.

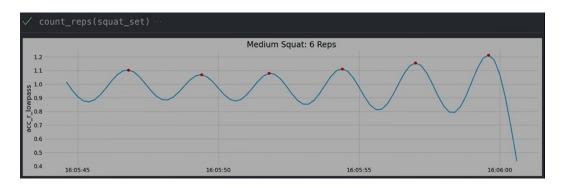


Figure 13: Extrimas to Count Repetetions

4.4 Detecting Improper Form:

During the experiments, additional data was collected from a participant performing the bench press with improper form. Specifically, the participant intentionally executed sets where the bar was either placed too high on the chest or did not make contact with the chest at all. This data served as the basis for training a Random Forest model, comparable to the one described in this section, to classify form quality. The dataset included three labels: correct form, too high, and no touch.

Applying the same training and testing methodology as previously explained, we evaluated the model's performance on a dataset comprising 1,098 instances. Remarkably, the model achieved an impressive accuracy of 98.53% on the test set.

5 Conclusion

This study aimed to explore context-aware applications within the strength training domain, addressing a gap in the existing literature that often overlooks strength programs. Leveraging wristband accelerometer and gyroscope sensor data, we collected real-world strength training sessions involving five participants performing the fundamental barbell lifts with varying weights. To ensure participants adhered to their strength levels, we utilized the one-rep max (1RM) metric.

Applying the machine learning framework for quantified self [2] to this data, we found that a Random Forest emerged as the optimal exercise classification model. Remarkably, the model achieved an overall accuracy of 99.58% when classifying unseen instances. To mitigate potential accuracy paradoxes, we also evaluated precision, recall, and F1-score, confirming the robustness of our results. However, the model occasionally misclassified bench press instances as overhead press and vice versa, as well as deadlifts and rows. A plausible explanation lies in the similarity between wrist orientations during these vertical pressing and pulling movements.

To count repetitions, we employed a straightforward peak counting algorithm combined with a strong low-pass filter. Although basic, this method yielded decent results, with a miscount rate of 5%. Additionally, we trained another Random Forest to classify bench press form, using data intentionally collected from a participant exhibiting improper form. Remarkably, the model accurately distinguished correct instances from improper ones, achieving an accuracy of 98.53%. While the form detection dataset remains limited, this approach serves as a proof of concept.

References

- 1. The Presidents Council on Physical Fitness and Sports. Fitness Fundamentals Guidelines for Personal Exercise Programs. Online Council Publications (2003). Available at: https://www.hhs.gov/fitness/index.html
- 2. Hoogendoorn, M., & Funk, B. Machine Learning for the Quantified Self.
- 3. Van Den Bogert, M. (1996). Practical Guide to Data Smoothing and Filtering.
- 4. Atlas Wristband 2. Available at: https://atlaswearables.com/
- 5. American College of Sports Medicine. (2013). ACSM's Resources for the Personal Trainer. Lippincott Williams & Wilkins.
- 6. Read, R., Kasparian, M., Li, P. (2015). USD725512S1.
- 7. Van Laerhoven, K., & Cakmakci, O. (2000, October). What Shall We Teach Our Pants? In Wearable Computers, The Fourth International Symposium on (pp. 77-83). IEEE.
- 8. Ermes, M., Prkk, J., Mntyjrvi, J., & Korhonen, I. (2008). Detection of Daily Activities and Sports with Wearable Sensors in Controlled and Uncontrolled Conditions. IEEE Transactions on Information Technology in Biomedicine, 12(1), 20-26.
- 9. Banos, O., Damas, M., Pomares, H., Prieto, A., & Rojas, I. (2012). Daily Living Activity Recognition Based on Statistical Feature Quality Group Selection. Expert Systems with Applications, 39(9), 8013-8021.
- 10. Zhang, M., & Sawchuk, A. A. (2013). Human Daily Activity Recognition with Sparse Representation Using Wearable Sensors. IEEE Journal of Biomedical and Health Informatics, 17(3), 553-560.
- 11. Leutheuser, H., Schuldhaus, D., & Eskofier, B. M. (2013). Hierarchical, Multi-Sensor-Based Classification of Daily Life Activities: Comparison with State-of-the-Art Algorithms Using a Benchmark Dataset. PloS One, 8(10), e75196.
- 12. MbientLab. Available at: https://mbientlab.com/
- 13. Fitbit. Available at: https://www.fitbit.com/
- 14. Samsung Gear. Available at: https://www.samsung.com/us/mobile/wearables/

- 15. Apple Watch. Available at: https://www.apple.com/lae/apple-watch-series-4/
- 16. Location API. Available at: https://developers.google.com/location-context/
- 17. MotionActivity. Available at: https://developer.apple.com/documentation/coremotion/cmmotionactivity
- 18. Fitbit Smarttrack. Available at: https://www.fitbit.com/nl/smarttrack
- 19. Li, C., Fei, M., Hu, H., & Qi, Z. (2012, September). Free Weight Exercises Recognition Based on Dynamic Time Warping of Acceleration Data. In International Conference on Intelligent Computing for Sustainable Energy and Environment (pp. 178-185). Springer, Berlin, Heidelberg.
- 20. Chang, K. H., Chen, M. Y., & Canny, J. (2007, September). Tracking Free Weight Exercises. In International Conference on Ubiquitous Computing (pp. 19-37). Springer, Berlin, Heidelberg.
- 21. Koskimki, H., & Siirtola, P. (2014, December). Recognizing Gym Exercises Using Acceleration Data from Wearable Sensors. In CIDM (pp. 321-328).
- 22. Ward, J. A., Lukowicz, P., & Gellersen, H. W. (2011). *Performance Metrics for Activity Recognition*. ACM Transactions on Intelligent Systems and Technology (TIST), 2(1), 6.
- 23. Kraemer, W. J., Ratamess, N. A., & French, D. N. (2002). Resistance Training for Health and Performance. Current Sports Medicine Reports, 1(3), 165-171.
- 24. Rippetoe, M., Kilgore, L. (2007). Starting Strength: Basic Barbell Training. Wichita Falls, Texas, USA: Aasgaard Company.
- 25. One Rep Max Calculator. Available at: https://strengthlevel.com/one-rep-max-calculator
- 26. Wood, T. M., Maddalozzo, G. F., Harter, R. A. (2002). Accuracy of Seven Equations for Predicting 1-RM Performance of Apparently Healthy, Sedentary Older Adults. Measurement in Physical Education and Exercise Science, 6(2), 67-94.
- 27. Marchese, R., Hill, A. (2011). The Essential Guide to Fitness: For the Fitness Instructor. Pearson Australia.