


Predicting lying, sitting, walking and running using Apple Watch and Fitbit data

Daniel Fuller ,^{1,2} Javad Rahimpour Anaraki,³ Bongai Simango,¹ Machel Rayner,¹ Faramarz Dorani,² Arastoo Bozorgi,² Hui Luan,⁴ Fabien A Basset¹

To cite: Fuller D, Anaraki JR, Simango B, *et al.* Predicting lying, sitting, walking and running using Apple Watch and Fitbit data. *BMJ Open Sport & Exercise Medicine* 2021;**7**:e001004. doi:10.1136/bmjsem-2020-001004

Accepted 22 February 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹School of Human Kinetics and Recreation, Memorial University of Newfoundland, St. John's, Newfoundland, Canada

²Department of Computer Science, Faculty of Science, Memorial University of Newfoundland, St. John's, Newfoundland, Canada

³Institute of Biomedical Engineering, University of Toronto, Toronto, Ontario, Canada

⁴Department of Geography, University of Oregon, Eugene, Oregon, USA

Correspondence to

Daniel Fuller;
fuller.daniel@gmail.com

ABSTRACT

Objectives This study's objective was to examine whether commercial wearable devices could accurately predict lying, sitting and varying intensities of walking and running.

Methods We recruited a convenience sample of 49 participants (23 men and 26 women) to wear three devices, an Apple Watch Series 2, a Fitbit Charge HR2 and iPhone 6S. Participants completed a 65 min protocol consisting of 40 min of total treadmill time and 25 min of sitting or lying time. The study's outcome variables were six movement types: lying, sitting, walking self-paced and walking/running at 3 metabolic equivalents of task (METs), 5 METs and 7 METs. All analyses were conducted at the minute level with heart rate, steps, distance and calories from Apple Watch and Fitbit. These included three different machine learning models: support vector machines, Random Forest and Rotation forest.

Results Our dataset included 3656 and 2608 min of Apple Watch and Fitbit data, respectively. Rotation Forest models had the highest classification accuracies for Apple Watch at 82.6%, and Random Forest models had the highest accuracy for Fitbit at 90.8%. Classification accuracies for Apple Watch data ranged from 72.6% for sitting to 89.0% for 7 METs. For Fitbit, accuracies varied between 86.2% for sitting to 92.6% for 7 METs.

Conclusion This preliminary study demonstrated that data from commercial wearable devices could predict movement types with reasonable accuracy. More research is needed, but these methods are a proof of concept for movement type classification at the population level using commercial wearable device data.

INTRODUCTION

The introduction of commercial wearable devices for physical activity monitoring has been an exciting development with the potential to increase physical activity at the population level.^{1,2} We define commercial wearable devices as those used primarily by individual consumers for physical activity monitoring rather than research purposes.¹

Research examining commercial wearable devices has primarily focused on two areas. First is examining the reliability and validity of the devices' measures, including step counts,

heart rate and energy expenditure.^{3–5} Our recently published systematic review shows that heart rate measures are valid for some brands (Apple Watch, Garmin), but no brand correctly measures energy expenditure.⁶ The second research area for commercial wearable devices is how available measures, particularly steps, from commercial devices, translate to current physical activity recommendations. For example, Tudor-Locke *et al* found that approximately 8000 steps/day is a good proxy for 30 min of daily moderate to vigorous physical activity (MVPA) and 7000 steps/day, 7 days a week is consistent with obtaining 150 min of weekly MVPA.^{7,8}

Despite the research examining commercial wearables to date, we believe that overall, these data are understudied. Specifically, the focus on concurrent validation studies⁹ to directly compare a commercial wearable device measure to a criterion measure is limiting. To improve our understanding of commercial wearable device data, researchers should use combinations of variables from these devices and machine learning methods to predict movement types, including physical activity and sedentary behaviour, using these data. These new combinations of variables (ie, features) and machine learning methods have commonly been applied to research-grade accelerometer data to predict movement types.^{10–13} For example, Staudenmayer *et al* used features including the distribution of counts (10th, 25th, 50th, 75th and 90th percentiles) and temporal dynamics of counts (lag, one autocorrelation) from wrist-worn accelerometers combined with Artificial Neural Networks to predict 18 different activities including lying down, running and raking leaves. To our knowledge, no research has developed new features and applied machine learning methods to predict movement types using commercial wearable device data. While specific measures from commercial wearable devices (ie, Heart rate,

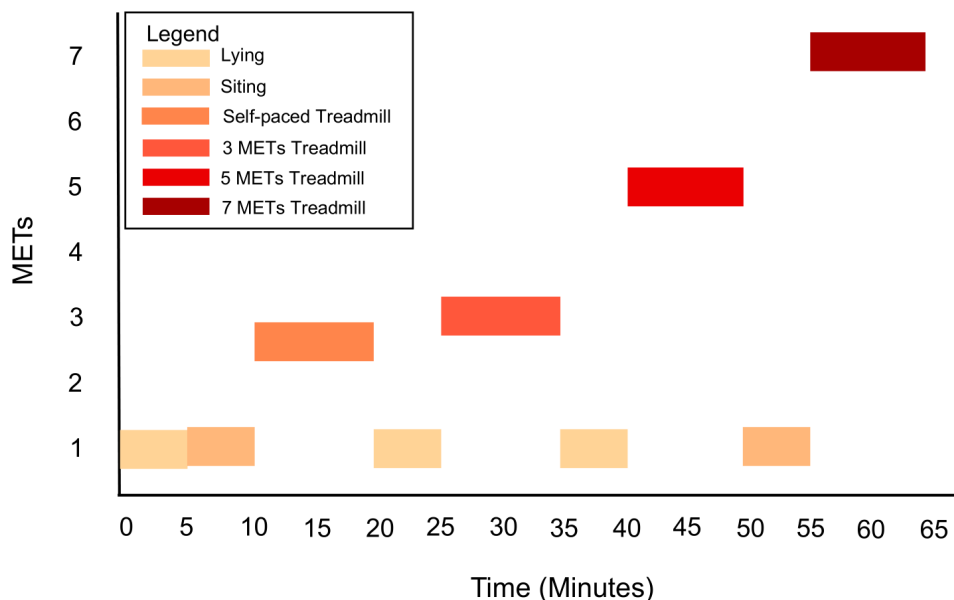


Figure 1 65 min lab-based activity protocol. METs, metabolic equivalents of task.

steps) are known to include measurement error, there is the potential to create new measures of physical activity and sedentary behaviour using devices that are already well adopted in the population. By examining multiple devices and including the device type as a feature in the collected data, we can investigate if the hardware and firmware play an essential role in deciding the outcome. If this is the case, machine learning methods may bypass hidden algorithms used in commercial devices and allow researchers to provide movement type estimates independent of device type (ie, standardise estimates across devices).

The purpose of this study is to examine whether using data from commercial wearable devices, Apple Watch and Fitbit and machine learning methods, we can predict movement types. We hypothesise that commercial wearable devices will accurately predict movement types associated with moderate to vigorous activity including running, but may not differentiate well between less intense movement types, including sitting. As a secondary objective, we examine whether accounting for the type of device could improve classification results. If the device type is an important feature for classification, this may allow for standardisation between devices.

METHODS

Study design

We used a lab-based protocol combined with a cross-sectional concurrent validation study design.^{9 14} Participants engaged in a 65 min protocol with 40 min of total treadmill time and 25 min of sitting or lying time. The protocol was similar to previous studies testing the reliability and validity of different commercial wearable devices.¹⁵ Figure 1 shows the lab-based protocol. Participant energy expenditure was measured for the entire study using the Oxycon Pro metabolic cart (Oxycon Pro, Jaeger, Hochberg, Germany). The Oxycon Pro is a

valid and reliable method for measuring energy expenditure.¹⁶ The metabolic cart was calibrated according to manufacturer specifications every morning of data collection.

The protocol's first two phases involve sedentary activity (ie, lying on a cot and sitting on a chair) for 5 min each. Following this, participants moved to the treadmill and selected a self-paced speed for 10 min. A 5 min lying period followed. Participants then moved to the treadmill and walked at a pace of 3 metabolic equivalents of task (METs) for 10 min. Following the 3 MET treadmill activity, the participants spent 5 min lying on a cot. Participants walked at an effort of 5 METs for 10 min, then had a 5 min sitting period. Finally, each participant completed 10 min at 7 METs. The 5 min rest periods were sufficient to lower participant heart rate and maintain a steady-state for these sedentary activities.¹⁷ The 10 min treadmill periods are sufficient to estimate O_2 uptake at steady state during each movement type involving activity. For each stage involving a specified MET value, a VO_2 to METs calculator was used to calculate the METs of each individual based on age, gender, height and weight.

Participants

We recruited a convenience sample of 49 participants (23 men and 26 women) from St. John's, Canada. Participants were recruited using social media posts and through word of mouth among lab members. Inclusion criteria included being over 18 years of age and completing the Physical Activity Readiness Questionnaire (PAR-Q).¹⁸ Participants were not provided with any compensation. All participants provided signed informed consent. Patients or the public were not involved in the design, conduct, or reporting, or dissemination plans of our research.

Table 1 Descriptions and descriptive statistics for each feature included in all models averaged over the entire study protocol

Variable	Description	Apple Watch Mean (SD)	Fitbit Mean (SD)
Steps/min	Number of steps/min	181.4 (270.4)	7.7 (21.8)
Heart rate (bpm)	Average heart rate/min	91.1 (26.8)	75.3 (38.7)
Energy expenditure	Amount of calories expended	5.8 (7.3)	40.8 (33)
Distance (m)	Distance ran in metres	0.1 (0.1)	23.8 (58)
Heart rate entropy	Measure of heart rate variability	6.1 (0.2)	5 (2.3)
Steps entropy	Measure of steps variability	6.1 (0.2)	3.9 (2.7)
Resting heart rate (bpm)	10th percentile of HR data	68.3 (11.9)	59.1 (38.9)
Correlation heart rate and steps	Correlation coefficient between heart rate and steps	0 (0.7)	0.7 (0.6)
Intensity (Karvonen formula)	Intensity zone during activity	0.2 (0.2)	0.1 (0.2)
Steps*distance	Product measure of total amount of steps and distance covered in metres	40.3 (113.5)	967.6 (5275.1)
SD normalised heart rate	SD of normalised heart rate	8.8 (10.1)	7.7 (15.7)

Wearable devices

Participants used three devices, an Apple Watch Series 2, a Fitbit Charge HR2 and an iPhone 6S. We chose Apple Watch and Fitbit because they have the highest market share among wearable devices.⁵ We randomly assigned the wrist for wearing the Fitbit and placed the Apple Watch on the opposite wrist. Participants were given an iPhone 6S with a custom iOS App called Physical Activity, Sleep, and Sedentary Behaviour Mobile (PASS Mobile) developed in our lab. PASS Mobile collects minute-by-minute data from Fitbit and Apple Watch. For Fitbit, the App connects to the Fitbit SDK.¹⁹ For Apple Watch, the App connects to Apple HealthKit.²⁰ PASS Mobile was installed through Test Flight, the Apple development platform, and is not available publicly in the Apple App Store.

Measures

The study's outcome variable was movement types based on the activities performed and the measures from the Oxycon Pro metabolic cart. For every minute of the protocol, the outcome variable includes a label for one of six movement types: lying, sitting, walking self-paced, 3 METs, 5 METs and 7 METs.

The variables collected through the PASS Mobile App were heart rate, steps, distance and calories. Each measure was collected at 1 Hz from Apple Watch and Fitbit, respectively. For heart rate, both devices collect the average heart rate for the minute. For steps, both

devices provide the total number of steps for the minute. For distance, both devices estimate the total distance travelled in metres. For calories, Apple Watch collects active calories, not including a constant to account for basal metabolic rate. Therefore, it was plausible that during sitting or lying participants had a true value of zero calories for Apple Watch. Fitbit provides total energy expenditure using the MD Mifflin-St Jeor equation,^{21,22} which means Fitbit reports energy expenditure every minute, even when the participant is sitting or lying. Additional variables included in the analysis are participants age in years, weight in kilograms, height in metres and sex (male or female).

Analyses

Statistical analyses were performed using R (V.3.6.1) and Weka (V.3.8.3). Data were downloaded from the metabolic cart. We used previously published methods to convert breath-by-breath data to minute by minute MET intensity estimates.²³ We have published the code for this analysis on GitHub (https://github.com/walkabillylab/jaeger_analysis).

Analyses were conducted separately for Apple Watch and Fitbit. We first cleaned the data and used linear interpolation on steps, heart rate, calories and distance to impute missing data. Following this, we developed a feature set that included intensity (Karvonen formula)^{24,25} which calculates individualised target heart rate parameters, steps entropy as a measure of predictability of step count and the correlation coefficient between heart rate and steps.²⁶ We developed the features to consider multiple physiological characteristics that could explain sitting, lying and different physical activities (see [table 1](#)).

We used three different classification methods, Random Forest,^{27,28} Rotation Forest,²⁹ and linear support vector machines (SVM),³⁰ in our analysis.³¹ Model accuracy was examined using k-fold cross-validation. Data were randomly split into 10 subsamples. For each subsample, classification algorithms were developed. Each algorithm

Table 2 Percentage accuracy per each classifier for Apple Watch and Fitbit for interpolated data

	Apple Watch	Fitbit
SVM	50.87	56.66
Random Forest	81.95	90.80
Rotation Forest	82.60	89.26

Bold value represents highest accuracy.
SVM, support vector machines.

Table 3 Confusion matrix for Apple Watch data from Rotation Forest model

True value*	Lying	Sitting	Self-paced walk	Running 3 METs	Running 5 METs	Running 7 METs
<i>Predicted as</i>						
Lying	613 (78.8%)	59 (10.7%)	24 (4.7%)	51 (8.9%)	30 (4.8%)	10 (1.6%)
Sitting	48 (6.2%)	399 (72.6%)	25 (4.9%)	7 (1.2%)	41 (6.5%)	32 (5.3%)
Self-paced walk	25 (3.22%)	28 (5.1%)	454 (88.2%)	10 (1.8%)	11 (1.8%)	1 (0.2%)
Running 3 METs	50 (6.4%)	7 (1.3%)	9 (1.8%)	493 (86.0%)	10 (1.6%)	3 (0.5%)
Running 5 METs	27 (3.5%)	25 (4.6%)	3 (0.6%)	9 (1.6%)	518 (82.2%)	21 (3.4%)
Running 7 METs	15 (1.9%)	32 (5.8%)	0 (0%)	3 (0.5%)	20 (3.2%)	543 (89.0%)

*Values in absolute number (representing minutes) and per cent.

was then used to predict the error associated with each one of the subsamples. A sum of prediction errors was calculated over all subsamples to produce a final accuracy.³² In each model, we included the features described in table 1 and age, gender, height and weight. We chose these models because linear SVM³³ and Random Forest models³⁴ are common in physical activity research using research-grade accelerometers and Rotation Forest are similar methods to Random Forest.

We evaluated model fit using accuracy, confusion matrices and feature ranking. Finally, to answer our second research question, we combined the Fitbit and Apple Watch data and added an additional feature, device type and reran a Rotation Forest model to see the difference between devices.

RESULTS

Participants included 26 women and 23 men. The average age was 29.3 (min 18–max 56). Table 1 shows mean and SD values for continuous variables or count and per cent for categorical predictors for Apple Watch and Fitbit, respectively. The average height and weight were 1.7 m and 70.6 kg, respectively. Average heart rate during the entire study protocol was 91.1 for Apple Watch and 75.3 for Fitbit. Average steps per minute were 181.4 and 7.7 for Apple Watch and Fitbit, respectively. Table 1 also shows the feature descriptions and descriptive statistics for each feature included in the models.

Table 2 shows the overall classification accuracies for the Random Forest, SVM and Rotation Forest models.

The Rotation Forest model had the highest accuracy for Apple Watch, and the Random Forest model had the highest accuracy for Fitbit. However, the difference between the Random Forest and Rotation Forest models was small. As a result, we present the Rotation Forest models. Tables 3 and 4 show the confusion matrices from the Rotation Forest model for Apple Watch and Fitbit data, respectively. Table 5 shows the top eight features from the χ^2 feature ranking for the Rotation Forest models.

Finally, we included the device type as a feature to the Rotation Forest model to examine the device's potential as important in predicting activity type. The accuracy, including device type, was 85.9% with the device type variable being ranked 13th overall in terms of feature importance.

DISCUSSION

This study used minute by minute data collected from two different commercial wearable devices combined with machine learning models to predict different movement types. We show that data from commercial wearable devices can correctly predict movement types in 82% and 90% of instances for Apple Watch and Fitbit, respectively. We also developed and used new features that combine existing data from commercial wearable devices, including heart rate, step count and calories.

Table 4 Confusion matrix for Fitbit data from Rotation Forest model

True value*	Lying	Sitting	Self-paced walk	Running 3 METs	Running 5 METs	Running 7 METs
<i>Predicted as</i>						
Lying	505 (88.1%)	15 (4.1%)	24 (6.4%)	26 (6.9%)	11 (2.8%)	11 (2.1%)
Sitting	16 (2.8%)	319 (86.2%)	8 (2.1%)	0 (0%)	20 (5.0%)	15 (2.9%)
Self-paced walk	6 (1.1%)	13 (3.5%)	333 (88.6%)	4 (1.1%)	1 (0.3%)	3 (0.6%)
Running 3 METs	25 (4.4%)	1 (0.3%)	5 (1.3%)	338 (89.9%)	7 (1.8%)	2 (0.4%)
Running 5 METs	16 (2.8%)	10 (2.7%)	3 (0.8%)	6 (1.6%)	357 (89.5%)	7 (1.4%)
Running 7 METs	5 (0.9%)	12 (3.2%)	3 (0.8%)	2 (0.5%)	3 (0.8%)	476 (92.6%)

*Values in absolute number (representing minutes) and per cent.

Table 5 χ^2 variable importance results for Apple Watch and Fitbit from the Rotation Forest model

Ranking	Apple Watch	Fitbit
1	Heart rate	Distance
2	Normalised heart rate	Steps
3	Steps	Heart rate
4	Intensity	Normalised heart rate
5	Calories	Steps*distance
6	Distance	Intensity_low energy
7	Steps*distance	Calories
8	SD normalised HR	SD normalised HR
9	Correlation heart rate and steps	Entropy steps per day
10	Weight	Resting heart rate

Comparison with past research

The Rotation Forest algorithm achieved the highest accuracy for classifying sedentary, light, moderate and vigorous activity from Apple Watch. The Random Forest model had the highest accuracy for Fitbit data. Overall percentage accuracy for all four movement types is slightly lower than previous research using research-grade wearable devices. The accuracy of models for predicting movement types in studies using research-grade wearables is typically greater than 90% in lab-based studies.^{28 35 36} The lower accuracy in our study is expected as the frequency of data collection is at the minute level, and each individual measure from the wearable devices has measurement error. We show that the overall accuracy is slightly higher for Fitbit compared with Apple Watch. This is surprising given that, previous studies have shown that the Apple Watch is more accurate for individual measures such as heart rate.^{37–40} The results are also consistent with previous work, showing that the models' accuracy was higher for activities greater than 3 METs and lower for low intensity and sedentary activities for both Fitbit and Apple Watch.¹ However, given the preliminary nature of this work, we believe these models are promising.

Important variables

To understand which variables were the most important in each model, we used feature ranking methods. The results of feature ranking showed that for Apple Watch, heart rate was the most important, while for Fitbit, steps were more important. Among the top six most important features for both Fitbit and Apple Watch were heart rate, steps, calories and distance. We developed new features based on the literature, including normalised heart rate and intensity using the Karvonen formula, which were important in model accuracy. Conversely, previous features thought to be important for classifying moderate activity (100 steps/min) were not important in our models.⁸ This may be because our study design is different, and our outcome variable includes multiple

movement types and features that differ from previous research.

Device type

This study found that the type of commercial wearable device does impact accuracy. However, we only examine the differences between the two devices. This may relate to different methods used by the devices to estimate heart rate, steps, calories, and distance. Including more devices and specifically including the software or firmware version of devices may be important.

Limitations

There are several limitations to this study. First, unlike research-grade devices, commercial wearables include much more missing data. There are many instances when the device cannot collect a reading, or there are errors in the transfer of data from the device to our app. For example, the total step count between Apple Watch and Fitbit is dramatically different (see [table 1](#)). We attempted to deal with these missing data by imputation. Future research should examine imputation methods and their impact on model accuracy. Second, each individual measure from commercial wearable devices has a measurement error. The impact of these errors on our models is unknown. Third, the devices we used for our research are now not the most current versions available. This is common with wearable device research. We cannot know if newer devices provide fewer missing data or more accurate measures. Fourth, our results show that device type was not important in predictions. However, we believe that given the unknown nature of the algorithms used to measure heart rate, steps, calories and distance, by commercial companies, researchers should continue to develop methods to attempt to account for algorithmic differences (ie, include device and/or firmware/software version) when these algorithms are unknown. Finally, we have not used neural network type methods because our dataset is small.

CONCLUSION

This preliminary study demonstrated that commercial wearable devices such as Apple Watch and Fitbit could predict six different movement types, including sitting, lying down and different intensities of walking/running with reasonable accuracy. The results support the use of raw data from Apple Watch and Fitbit combined with our machine learning approach for scalable movement type classification at the population level.

Twitter Daniel Fuller @walkabilly

Acknowledgements The authors would like to thank Machel Rayner for assistance with participant recruitment and data collection.

Contributors DF conceptualised the paper. All authors assisted with data collection. DF, JRA, BS, AB and HL conducted data analysis. All authors contributed to writing the manuscript and approved the submitted version.

Funding Funding for this research was provided by Dr. Fuller's Canada Research Chair (# 950–2 30 773).

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval Ethical approval was obtained by the Memorial University Interdisciplinary Committee on Ethics in Human Research (ICEHR #20180188-EX). All participants signed paper consent forms.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. Data are available at this link: <https://doi.org/10.7910/DVN/ZS2Z2J>.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Daniel Fuller <http://orcid.org/0000-0002-2015-2955>

REFERENCES

- 1 Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. *Int J Behav Nutr Phys Act* 2015;12:159.
- 2 Wright SP, Hall Brown TS, Collier SR, et al. How consumer physical activity monitors could transform human physiology research. *Am J Physiol Regul Integr Comp Physiol* 2017;312:R358–67.
- 3 Brooke SM, An H-S, Kang S-K, et al. Concurrent validity of wearable activity Trackers under free-living conditions. *J Strength Cond Res* 2017;31:1097–106.
- 4 Fokkema T, Kooiman TJM, Krijnen WP, et al. Reliability and validity of ten consumer activity trackers depend on walking speed. *Med Sci Sports Exerc* 2017;49:793–800.
- 5 Bunn JA, Navalta JW, Fountaine CJ, et al. Current state of commercial wearable technology in physical activity monitoring 2015-2017. *Int J Exerc Sci* 2018;11:503–15.
- 6 Fuller D, Colwell E, Low J, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. *JMIR Mhealth Uhealth* 2020;8:e18694.
- 7 Sisson SB, Camhi SM, Church TS, et al. Accelerometer-determined steps/day and metabolic syndrome. *Am J Prev Med* 2010;38:575–82.
- 8 Tudor-Locke C, Leonardi C, Johnson WD, et al. Accelerometer steps/day translation of moderate-to-vigorous activity. *Prev Med* 2011;53:31–3.
- 9 Fox MP, Lash TL, Bodnar LM. Common misconceptions about validation studies. *Int J Epidemiol* 2020;49:1392–6.
- 10 Ravi D, Wong C, Lo B, et al. A deep learning approach to on-Node sensor data analytics for mobile or wearable devices. *IEEE J Biomed Health Inform* 2017;21:56–64.
- 11 Zeng M, Nguyen LT, Yu B. Convolutional neural networks for human activity recognition using mobile sensors. *Proceedings of the 6th International Conference on Mobile Computing, Applications and Services. ICST*, 2014:197–205.
- 12 Staudenmayer J, Pober D, Crouter S, et al. An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. *J Appl Physiol* 2009;107:1300–7.
- 13 Liu S, Gao RX, Freedson PS. Computational methods for estimating energy expenditure in human physical activities. *Med Sci Sports Exerc* 2012;44:2138–46.
- 14 Marshall RJ. Validation study methods for estimating exposure proportions and odds ratios with misclassified data. *J Clin Epidemiol* 1990;43:941–7.
- 15 Saunders TJ, Gray CE, Borghese MM, et al. Validity of SC-StepRx pedometer-derived moderate and vigorous physical activity during treadmill walking and running in a heterogeneous sample of children and youth. *BMC Public Health* 2014;14:1–9.
- 16 Ismail M, Loucks-Atkinson A, Atkinson M. Multiple propane gas flow rates procedure to determine accuracy and linearity of indirect calorimetry systems : An experimental assessment of a method. *Peer J Prepr* 2019;709:1–30.
- 17 Phillipson EA, Bowes G, Townsend ER, et al. Role of metabolic CO₂ production in ventilatory response to steady-state exercise. *J Clin Invest* 1981;68:768–74.
- 18 Health Canada. *Par-q & you: physical activity readiness*, 2002.
- 19 Fitbit, Inc. Fitbit SDK. Available: <https://dev.fitbit.com/> [Accessed 26 Nov 2019].
- 20 Apple Inc. Apple healthKit. Available: <https://developer.apple.com/healthkit/> [Accessed 26 Nov 2019].
- 21 Fitbit, Inc. Fitbit activity logging API. Available: <https://dev.fitbit.com/build/reference/web-api/activity/#activity-logging> [Accessed 7 Feb 2019].
- 22 Gerrerio S, Juan W, Basiotis P. An easy approach to calculating estimated energy requirements. *Prev Chronic Dis* 2006;3:A129.
- 23 Robergs R, Burnett A. Methods used to process data from indirect calorimetry and their application to VO₂max. *J Exerc Physiol Online* 2003;6:44–57.
- 24 Karvonen J, Chwalbinska-Moneta J, Saynajakangas S. Comparison of heart rates measured by ECG and microcomputer. *Phys Sportsmed* 1984;12:65–9.
- 25 She J, Nakamura H, Makino K, et al. Selection of suitable maximum-heart-rate formulas for use with Karvonen formula to calculate exercise intensity. *Int J Autom Comput* 2015;12:62–9.
- 26 Freedson PS, Miller K. Objective monitoring of physical activity using motion sensors and heart rate. *Res Q Exerc Sport* 2000;71:21–9.
- 27 Breiman L. Random forests. *Mach Learn* 2001;45:5–32.
- 28 Ellis K, Kerr J, Godbole S, et al. A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers. *Physiol Meas* 2014;35:2191–203.
- 29 Rodríguez JJ, Kuncheva LI, Alonso CJ. Rotation forest: a new classifier ensemble method. *IEEE Trans Pattern Anal Mach Intell* 2006;28:1619–30.
- 30 Platt JC. *Support vector learning*. MIT Press, 1999.
- 31 Quinlan JR. *C4.5: programs for machine learning*. Morgan Kaufmann, 2014.
- 32 Crown WH. Potential application of machine learning in health outcomes research and some statistical cautions. *Value Health* 2015;18:137–40.
- 33 SW S, Wang L, Celler BG. Estimation of walking energy expenditure by using support vector regression. *IEEE Engineering in Medicine and Biology 27th Annual Conference*, 2005:3526–9.
- 34 Bonomi AG, Goris AHC, Yin B, et al. Detection of type, duration, and intensity of physical activity using an accelerometer. *Med Sci Sports Exerc* 2009;41:1770–7.
- 35 Montoye AHK, Mudd LM, Biswas S, et al. Energy expenditure prediction using RAW accelerometer data in simulated free living. *Med Sci Sports Exerc* 2015;47:1735–46.
- 36 Bunney PE, Zink AN, Holm AA, et al. Orexin activation counteracts decreases in nonexercise activity thermogenesis (NEAT) caused by high-fat diet. *Physiol Behav* 2017;176:139–48.
- 37 Dooley EE, Golaszewski NM, Bartholomew JB. Estimating accuracy at exercise intensities: a comparative study of self-monitoring heart rate and physical activity wearable devices. *JMIR Mhealth Uhealth* 2017;5:e34.
- 38 Wang R, Blackburn G, Desai M, et al. Accuracy of Wrist-Worn heart rate monitors. *JAMA Cardiol* 2017;2:104–6.
- 39 Wallen MP, Gomersall SR, Keating SE, et al. Accuracy of heart rate watches: implications for weight management. *PLoS One* 2016;11:e0154420–9.
- 40 Gillinov S, Etiwy M, Wang R, et al. Variable accuracy of wearable heart rate monitors during aerobic exercise. *Med Sci Sports Exerc* 2017;49:1697–703.