**Title: Predicting sleep, sedentary behaviour, and physical activity with Apple Watch and Fitbit using Rotation Forest models**

Daniel Fuller1,2, Javad Rahimipour Anaraki2, Henry Luan3, Bo Simango1, Faramarz Dorani2, Arastoo Bozorgi2

1. School of Human Kinetics and Recreation  
   Physical Education Building  
   Memorial University  
   St. John’s, NL  
   A1C 5S7
2. Department of Computer Science, Faculty of Science, Memorial University of Newfoundland, St. John’s, Canada
3. Department of Geography, University of Oregon, Eugene, Oregon, USA

Daniel Fuller, [dfuller@mun.ca](mailto:dfuller@mun.ca), corresponding author

Javad Rahimipour Anarak, [jra066@mun.ca](mailto:jra066@mun.ca)

Henry Luan, [hluan@uoregon.edu](mailto:hluan@uoregon.edu)

Bongai Simango, [b.s.simango@mun.ca](mailto:b.s.simango@mun.ca)

Faramarz Dorani, [fd6713@mun.ca](mailto:fd6713@mun.ca)

Arastoo Bozorgi, [ab1502@mun.ca](mailto:ab1502@mun.ca)

Keywords: commercial wearable devices; physical activity; sleep; sedentary; machine learning

**Abstract**

Objectives

There is considerable promise for using commercial wearable devices for measuring physical activity at the population level. The objective of this study was to examine whether commercial wearable devices could accurately predict sleep, sedentary behaviour, and physical activity in free living conditions.

Methods

We recruited a convenience sample of 19 participants (7 female) to wear four devices, a GENEActiv, and Apple Watch Series 2, a Fitbit Charge HR2, and an iPhone 6S, for seven days. Data were collected between November 12, 2017 and August 15, 2018. Commercial wearable device data were collected using a custom built iOS App. The outcome variable was minute-by-minute sleep, sedentary, light, moderate, and vigorous activities calculated using the Sedentary Sphere method for GENEActiv devices. Minute-by-minute heart rate, steps, distance, and calories from Apple Watch and Fitbit, and self-report height, and self-report weight were used as features in Random Forest models.

Results

Our analysis dataset included 57 097 and 21 489 minutes of Apple Watch and Fitbit data, respectively. Classification accuracies for Apple Watch data ranged from 87.67% for sedentary behaviour to 0.32% for light activities. For Fitbit, accuracies varied between 72.35% for sedentary to 0% for vigorous activity. Weight, age, and height were the most important features in the models.

Conclusion

Our results suggest that commercial wearable devices are not able to correctly predict sleep, sedentary behaviour, and physical activity when using a research grade device and open source algorithms as the outcome.

**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.1–5 We define commercial wearable devices as those used primarily by individual consumers for physical activity monitoring rather than for research purposes.6–14 The market share for commercial wearable devices is dominated by Fitbit, Apple Watch, and Garmin.12,15,16 Commercial wearable devices are often contrasted with research grade wearable devices include Actigraph, GENEActiv, and Actical. For example, using open-source algorithms, GENEActiv is considered to have good reliability and validity in predicting sleep, sedentary behaviour, and physical activity and has been studied in various populations.17,18

Research examining commercial wearable devices has primarily focused on two areas. First, examining the reliability and validity of the measures that the devices provide, including step counts, heart rate, and energy expenditure.6–14 A systematic review found that many smartphone applications and wearable devices were accurate for tracking step count.19

The second primary 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, Sissons, Camhi, Church, and Katzmarzyk, 2011, found that approximately 8000 steps/day is a good proxy for 30 minutes of daily moderate to vigorous physical activity (MVPA) and 7000 steps/day, seven days a week is consistent with obtaining 150 minutes of weekly MVPA.20,21 Evidence also shows that 100 steps per minute is consistent with moderate to vigorous activity.22–26

Despite the promising research examining commercial wearables, two important areas remain unexplored. First, commercial wearable devices tend to focus on step counts as a user goal rather than sleep, sedentary behaviour, and physical activity intensity (i.e., light, moderate, and vigorous physical activity). Yet, physical activity guidelines are based on minutes of moderate to vigorous physical activity.27 Second, commercial wearable devices use proprietary methods for estimating steps, heart rate, and calories, sleep, sedentary behaviour, and physical activity. Proprietary methods are unknown, make standardization between different commercial wearable devices difficult or impossible.

The purpose of this exploratory study is to examine whether commercial wearable devices (Apple Watch and Fitbit) could accurately predict sleep, sedentary behaviour, and physical activity in free living conditions. Data collected from research grade wearable device, GENEActiv was used as the ground truth. We hypothesize that commercial wearable devices will accurately predict moderate and vigorous physical activity, but may not differentiate well between sleep and sedentary behaviour. As a secondary objective, we examined whether account for the type of device (Apple Watch or Fitbit) could improve classification results. If device type is an important feature for classification, this may be an important first step in standardization between devices.

**Method**

*Design*

We recruited 19 participants to use four devices, a GENEActiv, and Apple Watch Series 2, a Fitbit Charge HR2, and an iPhone 6S, for seven days. Data were collected between November 12, 2017 and August 15, 2018. We chose Apple Watch and Fitbit for this study because they have the highest market share among wearable devices.13 Participants were instructed to go about their daily life and not change their behaviours. They wore the devices at all times except during water-based activities and device charging, when all devices were removed to ensure that missing data would be consistent for all devices. GENEActiv was worn on the non-dominant wrist, Fitbit wear location was randomly assigned and Apple Watch was worn on opposite wrist to Fitbit.

Participants were given an iPhone 6S with a custom iOS App called Physical Activity, Sleep, and Sedentary Behaviour Mobile (PASS Mobile). PASS Mobile collects minute-by-minute data from Fitbit and Apple Watch. For Fitbit, the App connects to the Fitbit SDK.28 For Apple Watch, the App connects to Apple HealthKit.29 PASS Mobile was installed through Test Flight, the Apple development platform, and is not available publicly in the App Store.

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

*Measures*

We collected raw accelerometer data at 100Hz using GENEActiv devices. We aggregated data to the 1Hz using standard approaches.30 We applied a method known as Sedentary Sphere to the GENEActiv data to identify minute-by-minute sleep, sedentary, light, moderate, and vigorous activities.31,32 Thus, our outcome variable was minute-by-minute sleep, sedentary, light, moderate, and vigorous activities. Through the PASS Mobile App we collected 1Hz heart rate, steps, distance, and calories from Apple Watch and Fitbit. We also collected participants age, self-report height, and self-report weight.

*Data Analysis*

All data cleaning and preparation were conducted using R (Version 3.5.1). The analysis code is available on GitHub.33 We cleaned datasets individually before merging all data on time and at the minute level. Standardized protocols were used to clean and create class labels for GENEActiv data.17,18 Measures of heart rate and energy expenditure differ between Apple Watch and Fitbit. It is challenging to differentiate between true zeros and missing data. Apple Watch records heart rate 1 minute every 5 minutes when the Heart App is in the background or display is off. It also collects active calories, that do not include a constant to account for basal metabolic rate. For Apple Watch, a minute of valid data could have a value of 0 for calories if the participant was at rest or the device is not being worn. For Apple Watch, a participant minute could have a zero value for heart rate (if between 5-minute measurements), steps, distance, and calories when the device is being worn during sedentary activity or sleep. We assumed that if all variables were zero for greater than 10 minutes the device was not being worn during that period. However, if the heart rate returned a biologically plausible value within 10 minutes, we assumed the device was being worn.

In contrast, Fitbit collects heart rate every minute and provides a total energy expenditure using the MD Mifflin-St Jeor equation,34,35 which means Fitbit always reports some energy expenditure, if the participant is at rest or the device is not being worn. For Fitbit, we assumed that if all variables, except calories, were zero for greater than 10 minutes the device was not being worn during that 10-minute period. We assume that energy expenditure values are based on participants’ height, weight, and age and existing functions for estimating energy expenditure based on height, weight, and gender.35

We used WEKA (Version 3.6.15)36, a machine learning software. Two datasets were created for each device for all participants (four datasets total). We used 70% of the data for training and 30% for validation and testing. We included heart rate, steps, calories, distance, age, height, and weight as features in our model. Because the number of features was small and not suitable for feature selection methods, we used a Rotation Forest37 classifier which is a combination of Principal Component Analysis38 and a Decision Tree39 classifier. We assessed the performance of the Rotation Forest classifier model for both Apple Watch and Fitbit, compared with four other classifiers, Support Vector Machine (SVM)40, Naïve Bayes41, Radial Basis Functions Network (RBFNetwork)42, and PARtial decision Tree (PART).43 We chose these models because SVM44, Naïve Bayes45, and Random Forest models46 are common in physical activity research using research grade accelerometers.

We evaluated model fit using accuracy, sensitivity, specificity, confusion matrices, and feature ranking. Accuracy is the percentage of correctly classified samples in a dataset (i.e., summation of all values in the diagonal of the resulting confusion matrix). Sensitivity is the number of samples of each class which are correctly classified, and specificity is the number of samples of a different class which are classified correctly. A Confusion matrix is a table, in which each row represents the number of samples of a class and each column represents the number of samples in a predicted class. Chi Square feature ranking methods to evaluate the importance of each feature. The Chi-Squared method47, calculates statistic for each feature and returns it as the rank of that feature.

Finally, to answer our second research question we combined the Fitbit and Apple Watch data and added an additional feature, device type. We reran a Rotation Forest model and feature importance measures including the device type feature.

**Results**

Participants included 7 females and 12 males. The average age was 31.1 (min 21 – max 44). The average height and weight were 1.71m and 80.6Kg, respectively. Our analysis dataset included 57097 and 21489 minutes of data for Apple Watch and Fitbit, respectively. Table 1 shows the sum of minutes of data in each activity category for Apple Watch and Fitbit. Differences in the number of minutes of data between Apple Watch and Fitbit are due to our wear time definitions for each device and upload challenges with the PASS Mobile App.

Table 1. Participant specific and average minutes

|  |  |  |
| --- | --- | --- |
|  | **Apple Watch** | **Fitbit** |
| Sleep | 18985 | 9503 |
| Sedentary | 34656 | 11061 |
| Light | 2118 | 558 |
| Moderate | 1215 | 310 |
| Vigorous | 123 | 57 |
| Total | 57097 | 21489 |

We applied five classifiers to the Apple Watch and Fitbit data (See Table 2). Rotation Forest had the best accuracy, sensitivity and specificity for Apple Watch data and accuracy and sensitivity for Fitbit. The best specificity for Fitbit was achieved with the Naïve Bayes classifier.

Table 2. Resulting classification accuracies using Naïve Bayes, PART, RBFNetwork, Rotation Forest and SVM classifiers for Apple Watch and Fitbit datasets

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Device** | **Classifier** | **Classification Accuracy (%)** | **Sensitivity** | **Specificity** |
| Apple Watch | Rotation Forest | **70.79** | **0.708** | **0.639** |
| Naïve Bayes | 58.93 | 0.589 | 0.419 |
| PART | 67.07 | 0.671 | 0.607 |
| RBFNetwork | 60.38 | 0.604 | 0.397 |
| SVM | 60.42 | 0.604 | 0.396 |
| Fitbit | Rotation Forest | **64.37** | **0.644** | 0.667 |
| Naïve Bayes | 46.04 | 0.460 | **0.695** |
| PART | 63.15 | 0.631 | 0.657 |
| RBFNetwork | 59.76 | 0.598 | 0.624 |
| SVM | 57.90 | 0.579 | 0.618 |

Table 3 and 4 show the confusion matrices from the Rotation Forest model for Apple Watch and Fitbit data, respectively. Confusion matrices were generated based on our models trained with 70% of the data and tested against 30% of the data, indicating that the values presented in Tables 3 and 4 sum to 30% of the total.

Table 3. Confusion matrix of AppleWatch for the test data

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | Predicted as | | | | |
|  |  | Sleep | Sedentary | Light | Moderate | Vigorous |
| Actual outcome | Sleep | 3036 (52.99%) | 2690 | 1 | 2 | 0 |
| Sedentary | 1263 | 9074 (87.67%) | 4 | 8 | 1 |
| Light | 73 | 554 | 2 (0.32%) | 4 | 0 |
| Moderate | 47 | 322 | 2 | 10 (2.62%) | 0 |
| Vigorous | 1 | 31 | 0 | 0 | 4 (11.11%) |

Table 4. Confusion matrix of Fitbit for the test data

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | Predicted as | | | | |
|  |  | Sleep | Sedentary | Light | Moderate | Vigorous |
| Actual outcome | Sleep | 1730 (60.89%) | 1106 | 5 | 0 | 0 |
| Sedentary | 910 | 2402 (72.35%) | 8 | 0 | 0 |
| Light | 35 | 124 | 15 (8.52%) | 2 | 0 |
| Moderate | 23 | 56 | 8 | 3 (3.33%) | 0 |
| Vigorous | 5 | 15 | 0 | 0 | 0 (0.00%) |

Classification accuracies for Apple Watch data ranged from 87.67% for sedentary behaviour to 0.32% for light activities. For Fitbit, accuracies varied between 72.35% for sedentary to 0% for vigorous activity. For both Apple Watch and Fitbit, the classes with the least data had lowest classification accuracy. However, in absolute terms, more sleep and sedentary minutes were misclassifited because these categories are more common.

To benefit from the data generated using Apple Watch and Fitbit, we merged both datasets and added a feature that defined the device type. We then applied Rotation Forest models to the dataset including the previous features and device type. The classification accuracy was 69.28%, sensitivity was 0.693, and specificity was 0.649. The results were similar to the results with separate Apple Watch and Fitbit data. The confusion matrix for this analysis is available in the online supplement.

We evaluated the importance of each feature for both devices using Chi-Squared47 and ReleifF48 feature ranking method and the results are shown in Table 5.

Table 5. Chi-Squared feature ranking results for Apple Watch and Fitbit

|  |  |  |
| --- | --- | --- |
| **Ranking** | **Apple Watch** | **Fitbit** |
| 1 | Weight | Weight |
| 2 | Age | Age |
| 3 | Height | Height |
| 4 | Calories | Distance |
| 5 | Distance | Calories |
| 6 | Steps | Gender |
| 7 | Heart | Heart |
| 8 | Gender | Steps |

The most important features using Chi-Squared feature ranking were weight, age, and height. To examine the discrepancy between the two feature ranking methods we removed the highest ranking features and reran our analysis. Removing weight, age, and height classification accuracies decreased from 70.79% to 63.16% and from 64.37% to 61.59% for Apple Watch and Fitbit datasets, respectively. The experiments show the importance of weight, age, and height in providing valuable information to the classifier.

**Discussion**

The purpose of this exploratory study was to examine whether commercially available devices could accurately predict sleep, sedentary behaviour, and physical activity. Our results suggest that commercial wearable devices are not able to correctly predict sleep, sedentary behaviour, and physical activity when using a research grade device and open source algorithms as the ground truth. Our results are consistent with past research showing that compared to indirect calorimetry, commercial wearable devices do not appear to be able to correctly predict activity intensity when proprietary algorithms are used.6,10,13,14

The modest classification accuracy for the data is evident when examining the confusion matrices. There were very few minutes of light, moderate, and vigorous activities, meaning it is difficult for the model to accurately predict physical activities for both types of devices. The confusion matrices also show differences in performance between devices. Apple Watch performed better in predicting sedentary and vigorous activities, whereas, Fitbit outperformed Apple Watch in predicting sleep, light, and moderate activities, with less data.

Both devices performed reasonably well in predicting sleep and sedentary behaviour. This was somewhat surprising as predicting sleep and sedentary behaviour are challenging when using research grade devices.49–58 However, misclassification between sleep and sedentary behaviour remains an important challenge. The amount of data played a role in the classification accuracy for sleep and sedentary behaviour.

Our results suggest that the use of existing features and development of new features will be important in order to improve classification accuracy. Our feature ranking analysis showed that heart rate, steps, calories, height, weight, and age are critical features, while gender and distance travelled may provide limited additional information to the model. As has previously been done with raw accelerometer data59–65, we believe that the inclusion of additional features, including heart rate entropy or quintiles of steps per minute could provide additional useful information to the model. It is also important to note that we did not have access to any accelerometer data for the commercial devices. Having access to accelerometer data from commercial devices could allow us to combine our approach with well-established methods for predicting activity using raw accelerometer data.66

Our results show that the device type did not improve model performance. We conducted this analysis with the hypothesis that device type could play an important role in standardizing across commercial wearable devices. Two types of devices, however, were included in the analysis, which limited the generalizability of our conclusions.

Future research in this area could test different methods for the ground truth. We used the Sedentary Sphere method but other methods exist and are available in the GGIR package.67 Future research should also test whether additional features improve model performance. Finally, additional devices should be included with the goal to standardize predictions across device types.

**Limitations**

Our study has numerous limitations including stability of the PASS Mobile App, error propagation, and firmware and operating system changes. The PASS Mobile App was developed by our research team. Throughout the course of data collection bugs in the App slowed data transfer and resulted in some data loss. During the data collection, there were firmware and operating system updates for all of the devices. Participants were instructed not to update any devices as updates could change the underlying proprietary algorithms and would require updates to the PASS Mobile App. While the GENEActiv is shown to be accurate, there are still classification errors with the GENEActiv for sleep, sedentary behavior, and physical activity. Our modeling approach did not propagate error from the GENEActiv into the models for the commercial wearable devices.

**Conclusion**

This preliminary study showed that Apple Watch and Fitbit were not able to correctly predict sleep, sedentary behaviour, and physical activity using a feature set including data available from the device (heart rate, steps, calories, and distance) and demographic characteristics (age, height, weight, gender). Additional research could improve classification accuracies and potentially help standardize between commercial wearable devices.

**Ethics Approval and Consent to Participate**

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

**Consent for Publication**

Not applicable

**Availability of Data and Material**

Data are publicly available on the BeapLab Dataverse : <https://doi.org/10.7910/DVN/TO3YNT> Code for analysis are available : <https://github.com/walkabillylab/wearable_device_classification>

**Competing Interest**

The authors do not have any competing interests.

**Funding**

Funding for this research was provided by MaRS via the Health Behaviour Data Challenge and by Dr. Fuller’s Canada Research Chair.

**Authors’ Contributions**

DF, JRA, and FD conceptualized the paper. JRA, FD, and AB developed the PASS Mobile App. JRA, HL, BS, and DF conducted data analysis. DF, JRA, and BS drafted the manuscript. All authors contributed to writing the manuscript and approved the submitted version.

**Acknowledgements**

The authors thank Machel Rayner for assisting with data collection. The authors thank Richard Buote, Melissa Tobin, and Kassia Orychock for comments on early drafts of this manuscript.

**References**

1. Bagot KS, Matthews SA, Mason M, et al. Current, future and potential use of mobile and wearable technologies and social media data in the ABCD study to increase understanding of contributors to child health. *Dev Cogn Neurosci*. 2018;32(March):121-129. doi:10.1016/j.dcn.2018.03.008.

2. Ridgers ND, Timperio A, Brown H, et al. Wearable activity tracker use among Australian adolescents: Usability and acceptability study. *J Med Internet Res*. 2018;20(4). doi:10.2196/mhealth.9199.

3. Coughlin SS, Stewart J. Use of Consumer Wearable Devices To Promote Physical Activity: a Review of Health Intervention Studies. *J Environ Heal Sci*. 2016;2(6):1-6. doi:10.15436/2378-6841.16.1123.

4. Sullivan AN, Lachman ME. Behavior Change with Fitness Technology in Sedentary Adults: A Review of the Evidence for Increasing Physical Activity. *Front Public Heal*. 2017;4(January). doi:10.3389/fpubh.2016.00289.

5. Kerner C, Goodyear VA. The Motivational Impact of Wearable Healthy Lifestyle Technologies: A Self-determination Perspective on Fitbits With Adolescents. *Am J Heal Educ*. 2017;48(5):287-297. doi:10.1080/19325037.2017.1343161.

6. Simunek A, Dygryn J, Gaba A, Jakubec L, Stelzer J, Chmelik F. Validity of Garmin Vivofit and Polar Loop for measuring daily step counts in free-living conditions in adults. *Acta Gymnica*. 2016. doi:10.5507/ag.2016.014.

7. Andrés Calabró M, Lee J-M, Saint-Maurice PF, Yoo H, Welk GJ. *Validity of Physical Activity Monitors for Assessing Lower Intensity Activity in Adults*.; 2014. doi:10.1186/s12966-014-0119-7.

8. Brooke SM, An HS, Kang SK, Noble JM, Berg KE, Lee JM. Concurrent Validity of Wearable Activity Trackers under Free-Living Conditions. *J Strength Cond Res*. 2017;31(4):1097-1106. doi:10.1519/JSC.0000000000001571.

9. Alberto FP, Nathanael M, Mathew B, Ainsworth BE. Wearable monitors criterion validity for energy expenditure in sedentary and light activities. *J Sport Heal Sci*. 2017;6(1):103-110. doi:10.1016/j.jshs.2016.10.005.

10. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. 2015. doi:10.1186/s12966-015-0314-1.

11. Driscoll O. How well do activity monitors estimate energy expenditure? A systematic review and meta-analysis of the validity of current technologies. *Br J Sport Med*. 2018;0:1-11. doi:10.1136/bjsports-2018-099643.

12. Xie J, Wen D, Liang L, Jia Y, Gao L, Lei J. Evaluating the validity of current mainstream wearable devices in fitness tracking under various physical activities: Comparative study. *J Med Internet Res*. 2018;20(4). doi:10.2196/mhealth.9754.

13. Bunn JA, Navalta JW, Fountaine CJ, Reece JD. Current State of Commercial Wearable Technology in Physical Activity Monitoring 2015-2017. *Int J Exerc Sci*. 2018;11(7):503-515.

14. Alsubheen SA, George AM, Baker A, Rohr LE, Basset FA. Accuracy of the vivofit activity tracker. *J Med Eng Technol*. 2016;40(6):298-306. doi:10.1080/03091902.2016.1193238.

15. Ramkumar PN, Muschler GF, Spindler KP, Harris JD, McCulloch PC, Mont MA. Open mHealth Architecture: A Primer for Tomorrow’s Orthopedic Surgeon and Introduction to Its Use in Lower Extremity Arthroplasty. *J Arthroplasty*. 2017;32(4):1058-1062. doi:10.1016/j.arth.2016.11.019.

16. Schneider M, Chau L. Validation of the Fitbit Zip for monitoring physical activity among free-living adolescents. *BMC Res Notes*. 2016;9(1):1-9. doi:10.1186/s13104-016-2253-6.

17. Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W, Pollak CP. The role of actigraphy in the study of sleep and circadian rhythms. *Am Acad Sleep Med Rev Pap*. 2003;26(3):342-392. doi:10.1093/sleep/26.3.342.

18. Zdravevski E, Stojkoska BR, Standl M, Schulz H. Automatic machine-learning based identification of jogging periods from accelerometer measurements of adolescents under field conditions. *PLoS One*. 2017;12(9):1-29. doi:10.1371/journal.pone.0184216.

19. Case MA, Burwick HA, Volpp KG, Patel MS. Accuracy of smartphone applications and wearable devices for tracking physical activity data. *JAMA - J Am Med Assoc*. 2015;313(6):625-626. doi:10.1001/jama.2014.17841.

20. Sisson SB, Camhi SM, Church TS, Tudor-Locke C, Johnson WD, Katzmarzyk PT. Accelerometer-Determined Steps/Day and Metabolic Syndrome. *Am J Prev Med*. 2010;38(6):575-582. doi:10.1016/j.amepre.2010.02.015.

21. Tudor-Locke C, Leonardi C, Johnson WD, Katzmarzyk PT, Church TS. Accelerometer steps/day translation of moderate-to-vigorous activity. *Prev Med (Baltim)*. 2011;53(1-2):31-33. doi:10.1016/j.ypmed.2011.01.014.

22. Tudor-Locke C, Schuna JM, Han H, et al. Cadence (steps/min) and intensity during ambulation in 6-20 year olds: The CADENCE-kids study. *Int J Behav Nutr Phys Act*. 2018;15(1):1-11. doi:10.1186/s12966-018-0651-y.

23. Slaght J, Sénéchal M, Hrubeniuk TJ, Mayo A, Bouchard DR. Walking Cadence to Exercise at Moderate Intensity for Adults: A Systematic Review. *J Sports Med*. 2017;2017:1-12. doi:10.1155/2017/4641203.

24. Wang H, Zhang YF, Xu LL, Jiang CM. Step rate-determined walking intensity and walking recommendation in Chinese young adults: A cross-sectional study. *BMJ Open*. 2013;3(1):1-6. doi:10.1136/bmjopen-2012-001801.

25. O’brien MW, Kivell MJ, Wojcik WR, D’entremont G, Kimmerly DS, Fowles JR. Step rate thresholds associated with moderate and vigorous physical activity in adults. *Int J Environ Res Public Health*. 2018;15(11). doi:10.3390/ijerph15112454.

26. Abel M, Hannon J, Mullineaux D, Beighle A. Determination of Step Rate Thresholds Corresponding to Physical Activity Intensity Classifications in Adults. *J Phys Act Heal*. 2011;8(1):45-51. doi:10.1123/jpah.8.1.45.

27. Organization WH, others. Information Sheet: Global Recommendations on Physical Activity for Health: 5--17 years old. *World Heal Organ*. 2011.

28. Fitbit SDK. https://dev.fitbit.com/. Accessed February 8, 2019.

29. Apple HealthKit. https://developer.apple.com/healthkit/. Accessed February 8, 2019.

30. Wickham H, Francois R, Henry L, Müller K. dplyr: A grammar of data manipulation. *R Packag version 04*. 2015;3.

31. Schaeffer CA, Nigg CR, Hill JO, Brink LA, Raymond BC. Establishing and Evaluating Wrist Cutpoints for the GENEActiv Accelerometer in Youth. *Med Sci Sport Exerc*. 2014;46(4):826-833. doi:10.1249/MSS.0000000000000150.

32. van Hees VT, Sabia S, Anderson KN, et al. A Novel, Open Access Method to Assess Sleep Duration Using a Wrist-Worn Accelerometer. *PLoS One*. 2015;10(11):e0142533. doi:10.1371/journal.pone.0142533.

33. Wearable Device Classification. https://github.com/walkabillylab/wearable\_device\_classification. Accessed February 8, 2019.

34. Fitbit Activity Logging API. https://dev.fitbit.com/build/reference/web-api/activity/#activity-logging. Accessed February 7, 2019.

35. Gerrior S, Juan W, Basiotis P. An easy approach to calculating estimated energy requirements. *Prev Chronic Dis*. 2006;3(4):A129-A129.

36. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. *ACM SIGKDD Explor Newsl*. 2009;11(1):10-18.

37. Rodriguez JJ, Kuncheva LI, Alonso CJ. Rotation forest: A new classifier ensemble method. *IEEE Trans Pattern Anal Mach Intell*. 2006;28(10):1619-1630.

38. Jolliffe I. Principal component analysis. In: *International Encyclopedia of Statistical Science*. Springer; 2011:1094-1096.

39. Quinlan JR. *C4. 5: Programs for Machine Learning*. Elsevier; 2014.

40. Platt JC. 12 fast training of support vector machines using sequential minimal optimization. *Adv kernel methods*. 1999:185-208.

41. John GH, Langley P. Estimating continuous distributions in Bayesian classifiers. In: *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*. ; 1995:338-345.

42. Broomhead DS, Lowe D. *Radial Basis Functions, Multi-Variable Functional Interpolation and Adaptive Networks*.; 1988.

43. Frank E, Witten IH. Generating accurate rule sets without global optimization. 1998.

44. Su SW, Wang L, Celler BG, Ambikairajah E, Savkin AV. Estimation of Walking Energy Expenditure by Using Support Vector Regression. In: *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*. Vol 4. IEEE; 2005:3526-3529. doi:10.1109/IEMBS.2005.1617240.

45. Staudenmayer J, Pober D, Crouter S, Bassett D, Freedson P. An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. *J Appl Physiol*. 2009;107(4):1300-1307. doi:10.1152/japplphysiol.00465.2009.

46. BONOMI AG, GORIS AHC, YIN B, WESTERTERP KR. Detection of Type, Duration, and Intensity of Physical Activity Using an Accelerometer. *Med Sci Sport Exerc*. 2009;41(9):1770-1777. doi:10.1249/MSS.0b013e3181a24536.

47. Forman G. An extensive empirical study of feature selection metrics for text classification. *J Mach Learn Res*. 2003;3(Mar):1289-1305.

48. Kira K, Rendell LA. A practical approach to feature selection. In: *Machine Learning Proceedings 1992*. Elsevier; 1992:249-256.

49. Van Loo CMT, Okely AD, Batterham MJ, et al. Wrist accelerometer cut points for classifying sedentary behavior in children. *Med Sci Sports Exerc*. 2017;49(4):813-822. doi:10.1249/MSS.0000000000001158.

50. Janssen X, Cliff DP, Reilly JJ, et al. Predictive validity and classification accuracy of actigraph energy expenditure equations and cut-points in young children. *PLoS One*. 2013;8(11). doi:10.1371/journal.pone.0079124.

51. Kerr J, Patterson RE, Ellis K, et al. Objective Assessment of Physical Activity: Classifiers for Public Health HHS Public Access. doi:10.1249/MSS.0000000000000841.

52. Kim Y, Barry VW, Kang M. Validation of the ActiGraph GT3X and activPAL Accelerometers for the Assessment of Sedentary Behavior. *Meas Phys Educ Exerc Sci*. 2015;19(3):125-137. doi:10.1080/1091367X.2015.1054390.

53. Zhang S, Rowlands A V., Murray P, Hurst TL. Physical activity classification using the GENEA wrist-worn accelerometer. *Med Sci Sports Exerc*. 2012;44(4):742-748. doi:10.1249/MSS.0b013e31823bf95c.

54. S.R. G, N. N, N.W. B, T.G. P, N.D. G, Brown W.J. AO - Gomersall Norman; ORCID: http://orcid.org/0000-0002-7800-7654 AO - Burton, Nicola W.; ORCID: http://orcid.org/0000-0002-3221-2265 AO - Pavey, Toby G.; ORCID: http://orcid.org/0000-0002-4946-6683 AO - Gilson, Nicholas D.; ORCID: http: SR. O http://orcid. org/000.-0001-6808-0180 AO-N. Estimating Physical Activity and Sedentary Behavior in a Free-Living Context: A Pragmatic Comparison of Consumer-Based Activity Trackers and ActiGraph Accelerometry. *J Med Internet Res*. 2016;18(9):e239. doi:http://dx.doi.org/10.2196/jmir.5531.

55. Chastin SFM, Granat MH. Methods for objective measure, quantification and analysis of sedentary behaviour and inactivity. *Gait Posture*. 2010;31(1):82-86. doi:10.1016/j.gaitpost.2009.09.002.

56. Atallah L, Lo B, Ali R, King R, Yang G. Real-Time Activity Classification Using Ambient and Wearable Sensors. *IEEE Trans Inf Technol Biomed*. 2009;13(6):1031-1039. doi:10.1109/TITB.2009.2028575.

57. van Wouwe NC, Valk PJL, Veenstra BJ. Sleep Monitoring: A Comparison Between Three Wearable Instruments. *Mil Med*. 2011;176(7):811-816. doi:10.7205/MILMED-D-10-00389.

58. Aguilar-Farías N, Brown WJ, Peeters GMEEG. ActiGraph GT3X+ cut-points for identifying sedentary behaviour in older adults in free-living environments. *J Sci Med Sport*. 2014;17(3):293-299. doi:10.1016/j.jsams.2013.07.002.

59. He Z, Jin L. Activity recognition from acceleration data based on discrete consine transform and SVM. *Conf Proc - IEEE Int Conf Syst Man Cybern*. 2009;(October):5041-5044. doi:10.1109/ICSMC.2009.5346042.

60. Garcia-Ceja E, Brena R. Long-Term Activity Recognition from Accelerometer Data. *Procedia Technol*. 2013;7:248-256. doi:10.1016/j.protcy.2013.04.031.

61. Twomey N, Diethe T, Fafoutis X, et al. A Comprehensive Study of Activity Recognition Using Accelerometers. *Informatics*. 2018;5(2):27. doi:10.3390/informatics5020027.

62. Erdaş B, Atasoy I, Açici K, Oǧul H. Integrating Features for Accelerometer-based Activity Recognition. *Procedia Comput Sci*. 2016;58(Eicn):522-527. doi:10.1016/j.procs.2016.09.070.

63. Arenas Eduardo, Najera Francisco, Hernandez Ricardo. Tejido Epitelial I Tejido Epitelial De Revestimiento. 2010.

64. Pires IM, Garcia NM, Pombo N, Flórez-Revuelta F, Spinsante S. Pattern Recognition Techniques for the Identification of Activities of Daily Living using Mobile Device Accelerometer. 2017.

65. Olukoya F. Addendum to National Offender Management Service ( NOMS ): Strategic and Business Plans 2009 / 10 to. *Business*. 2011;45(11):2010. doi:10.1249/MSS.0b013e31829736d6.Activity.

66. ROWLANDS A V., YATES T, DAVIES M, KHUNTI K, EDWARDSON CL. Raw Accelerometer Data Analysis with GGIR R-package. *Med Sci Sport Exerc*. 2016;48(10):1935-1941. doi:10.1249/MSS.0000000000000978.

67. van Hees VT, Fang Z, Langford J, et al. Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents. *J Appl Physiol*. 2014;117(7):738-744. doi:10.1152/japplphysiol.00421.2014.

Online supplement

Table 1. Confusion matrix from rotation forest models predicting sleep, sedentary behaviour, light, moderate, and vigorous physical activity using Apple Watch and Fitbit data.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | Predicted as | | | | |
|  |  | Sleep | Sedentary | Light | Moderate | Vigorous |
| Actual outcome | Sleep | 4585 (53.87%) | 3922 | 3 | 1 | 0 |
| Sedentary | 2003 | 11717 (85.27%) | 13 | 8 | 0 |
| Light | 112 | 671 | 23 (2.83%) | 5 | 1 |
| Moderate | 51 | 396 | 4 | 8 (1.74%) | 0 |
| Vigorous | 5 | 47 | 0 | 0 | 1 (1.89%) |

Figure 1. Decision tree map from rotation forest models predicting sleep, sedentary behaviour, light, moderate, and vigorous physical activity using Apple Watch.



Figure 2. Decision tree map from rotation forest models predicting sleep, sedentary behaviour, light, moderate, and vigorous physical activity using Fitbit.



STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

|  |  |  |  |
| --- | --- | --- | --- |
|  | Item No | Recommendation | Page No |
| **Title and abstract** | 1 | (*a*) Indicate the study’s design with a commonly used term in the title or the abstract | X |
| (*b*) Provide in the abstract an informative and balanced summary of what was done and what was found | X |
| Introduction | | | |
| Background/rationale | 2 | Explain the scientific background and rationale for the investigation being reported | X |
| Objectives | 3 | State specific objectives, including any prespecified hypotheses | X |
| Methods | | | |
| Study design | 4 | Present key elements of study design early in the paper | X |
| Setting | 5 | Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection | X |
| Participants | 6 | (*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up | X |
| (*b*)For matched studies, give matching criteria and number of exposed and unexposed | NA |
| Variables | 7 | Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable | X |
| Data sources/ measurement | 8\* | For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group | X |
| Bias | 9 | Describe any efforts to address potential sources of bias | X |
| Study size | 10 | Explain how the study size was arrived at | X |
| Quantitative variables | 11 | Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why | X |
| Statistical methods | 12 | (*a*) Describe all statistical methods, including those used to control for confounding | X |
| (*b*) Describe any methods used to examine subgroups and interactions |  |
| (*c*) Explain how missing data were addressed | X |
| (*d*) If applicable, explain how loss to follow-up was addressed |  |
| (*e*) Describe any sensitivity analyses | X |
| Results | | |  |
| Participants | 13\* | (a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed | X |
| (b) Give reasons for non-participation at each stage |  |
| (c) Consider use of a flow diagram |  |
| Descriptive data | 14\* | (a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders | X |
| (b) Indicate number of participants with missing data for each variable of interest |  |
| (c) Summarise follow-up time (eg, average and total amount) |  |
| Outcome data | 15\* | Report numbers of outcome events or summary measures over time | X |

|  |  |  |  |
| --- | --- | --- | --- |
| Main results | 16 | (*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included |  |
| (*b*) Report category boundaries when continuous variables were categorized |  |
| (*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period |  |
| Other analyses | 17 | Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses |  |
| Discussion | | | |
| Key results | 18 | Summarise key results with reference to study objectives | X |
| Limitations | 19 | Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias | X |
| Interpretation | 20 | Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence | X |
| Generalisability | 21 | Discuss the generalisability (external validity) of the study results | X |
| Other information | | | |
| Funding | 22 | Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based | X |

\*Give information separately for exposed and unexposed groups.

**Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.