Human Kinetics

ORIGINAL RESEARCH

A Machine Learning Classifier for Detection of Physical Activity Types and Postures During Free-Living

Kerstin Bach,[1] Atle Kongsvold,[2] Hilde Bårdstu,[3,4] Ellen Marie Bardal,[2,3] Håkon S. Kjærnli,[1] Sverre Herland,[1] Aleksej Logacjov,[1] and Paul Jarle Mork[2]

[1]Department of Computer Science, Norwegian University of Science and Technology (NTNU), Trondheim, Norway; [2]Department of Public Health and Nursing, Norwegian University of Science and Technology (NTNU), Trondheim, Norway; [3]Department of Neuromedicine and Movement Science, Norwegian University of Science and Technology (NTNU), Trondheim, Norway; [4]Department of Sports, Food and Natural Sciences, Western Norway University of Applied Sciences, Sogndal, Norway

Introduction: Accelerometer-based measurements of physical activity types are commonly used to replace self-reports. To advance the field, it is desirable that such measurements allow accurate detection of key daily physical activity types. This study aimed to evaluate the performance of a machine learning classifier for detecting sitting, standing, lying, walking, running, and cycling based on a dual versus single accelerometer setups during free-living. *Methods*: Twenty-two adults (mean age [SD, range] 38.7 [14.4, 25–68] years) were wearing two Axivity AX3 accelerometers positioned on the low back and thigh along with a GoPro camera positioned on the chest to record lower body movements during free-living. The labeled videos were used as ground truth for training an eXtreme Gradient Boosting classifier using window lengths of 1, 3, and 5 s. Performance of the classifier was evaluated using leave-one-out cross-validation. *Results*: Total recording time was ~38 hr. Based on 5-s windowing, the overall accuracy was 96% for the dual accelerometer setup and 93% and 84% for the single thigh and back accelerometer setups, respectively. The decreased accuracy for the single accelerometer setup was due to a poor precision in detecting lying based on the thigh accelerometer recording (77%) and standing based on the back accelerometer recording (64%). *Conclusion*: Key daily physical activity types can be accurately detected during free-living based on dual accelerometer recording, using an eXtreme Gradient Boosting classifier. The overall accuracy decreases marginally when predictions are based on single thigh accelerometer recording, but detection of lying is poor.

Keywords: epidemiology, human activity recognition, sedentary behavior, sitting, validity, wearable technology

Accelerometer-based measurements of physical activity types (e.g., by hip- or wrist-worn sensors) are commonly implemented in large population-based studies and longitudinal studies to supplement or replace self-reports (Doherty et al., 2017; Fuzeki et al., 2017; Van Der Velde et al., 2017). A common method for processing the raw accelerometer signal is to map the accelerometer output to intensity level of physical activity by applying a set of predefined cut points (Heesch et al., 2018; Migueles et al., 2017; Watson et al., 2014). An alternative and complementary approach is to classify different postures and physical activity types by use of rule-based algorithms (Crowley et al., 2019; Skotte et al., 2014) or machine learning classifiers (Arvidsson et al., 2019; Narayanan et al., 2020; Stewart et al., 2018). Which postures and activity types that can be detected and with which accuracy depends on several factors such as the positioning and number of sensors and the

window length used for extracting information from the acceleration signal (Arvidsson et al., 2019; Twomey et al., 2018).

Regarding the use of machine learning, a recent study indicates that a random forest classifier can detect sitting, lying, standing, walking, running, and cycling with high overall accuracy based on training data obtained in a semi free-living environment, using two accelerometers positioned on the thigh and low back (Narayanan et al., 2020). Although a multiaccelerometer setup is likely to provide superior validity, a single-accelerometer setup will often be more desirable (e.g., lower cost, convenience for the participants). Thus, it is important to investigate to what extent a single sensor setup can provide valid estimates of key daily physical activity types in a free-living setting. Interestingly, a recent study showed that a random forest classifier can perform reasonably well in detecting most everyday physical activity types based on a single sensor on the thigh or low back (Stewart et al., 2018). However, the study was carried out in a controlled laboratory setting, and it is not clear if the results can be generalized to free-living (Farrahi et al., 2019). Moreover, while the abovementioned studies have used a random forest classifier, it has been suggested that extreme gradient boosting (XGBoost) is a more powerful approach for supervised learning problems (Chen & Guestrin, 2016).

The aim of this study was to evaluate the performance of an XGBoost classifier in detecting key daily physical activity types (i.e., sitting, standing, lying, walking, running, and cycling) during free-living based on a dual (thigh and back) versus single accelerometer setup (thigh or back). We also explored to what extent the window length for extracting information from the accelerometer signal influences the performance of the classifier. The latter is

Mork (paul.mork@ntnu.no) is corresponding author, https://orcid.org/0000-0003-3355-268.

1

important because it will provide information about the possible trade-off between performance of the classifier and the computational cost in analyzing data.

Methods

Participants

Twenty-two adults (eight females, 38.7 ± 14.4 years [range 25–68], weight 72.9 ± 10.9 kg [range 56–92], height 177 ± 8.5 cm [range 157–191], body mass index 23.1 ± 2.4 kg/m^2 [range 19.2–28.4]) were recruited via word of mouth among university/hospital staff to participate in the study. Potential participants were excluded if they had any physical or cognitive impairment that would prevent them from fully participating in the study protocol. All participants provided written informed consent and ethical approval was granted by the Regional Committee for Ethics in Medical Research, Mid-Norway (2015/1432).

Measurement of Physical Activity Types

Participants were equipped with two AX3 accelerometers (Axivity Ltd., Newcastle, United Kingdom) and a video camera (GoPro Hero 3+, San Mateo, CA). One accelerometer was positioned centrally on the lower back at the third lumbar segment (L3) and one on the front of the right thigh approximately 10 cm above the upper border of the patella. These placements were chosen based on user testing and piloting as well as findings in recent studies, showing promising results in detecting key physical activity types with similar sensor placements (Narayanan et al., 2020; Stewart et al., 2018). To attach the sensors, a 5×7 cm moisture permeable film (Opsite Flexifix; Smith & Nephew, Watford, United Kingdom) was attached to the skin. The sensor was then positioned on top of the film using double-side tape and covered with a second film layer of 10×8 cm. The video camera was placed on the chest with a chest harness (GoPro Chesty), pointing toward the feet to record leg and lower body movements. This positioning of the camera also allows identification of the orientation of the body relative to the surroundings.

The participants met with a research assistant at their workplace to have the sensors and camera mounted. After mounting the sensors and camera, the participants were free to carry out their activities as usual in their work or home environment for about 1.5–2.0 hr, that is, there were no restrictions on the participants' whereabouts or physical activities during the data collection. However, to ensure that data were recorded for the target physical activity types, the participants were instructed to accumulate at least 2–3 min of sitting, standing, lying down, walking, and jogging/running during the recording period. There was no enforced order or time period for carrying out these activities, and participants were encouraged to carry out their everyday free-living activities as normally as possible. Cycling was not an activity that was naturally included within the recording period, and a round of outdoor cycling was therefore added as an extended part of the protocol.

Instrumentation

The AX3 is a small and lightweight ($23 \times 32.5 \times 7.6$ mm, weight 11 g) triaxial accelerometer with a configurable sampling frequency between 12.5 and 3,200 Hz and range between ±2 and ±16 g. In the current study, the accelerometers were configured to record at 50 Hz and a range of ±8 g. The raw data was stored on a 512 MB internal memory and downloaded as binary file (Continuous Wave Accelerometer format) for visualization and further analysis. The AX3 OmGui software (version 1.0.0.37; Open Movement, Newcastle University, United Kingdom) was used to configure the sensors and download the accelerometer data. The GoPro Hero3+ (GoPro, Inc., San Mateo, CA) is small and lightweight video camera ($41 \times 59 \times 21$ mm, weight 135 g). Video files were recorded at 30 fps at $1,280 \times 720$ pixels in an MP4 format and stored in 20-min lengths on a 64 GB SanDisk Ultra XC I micro SD card.

Data Processing and Machine Learning

The videos were downloaded from the SD card and converted into an AVI file format with a resolution of 640 pixels × 360 pixels
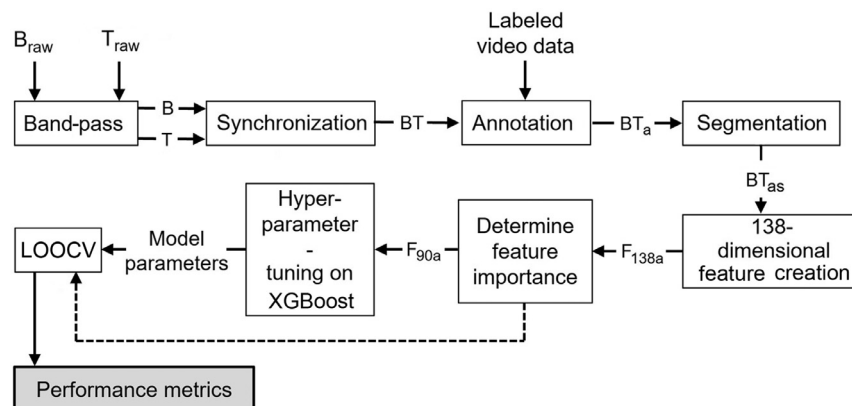

Figure 1 — Illustration of the main steps in developing the machine learning classifier. The raw acceleration data was first band-pass filtered before the data from back and thigh sensors were synchronized. The accelerometer data was thereafter synchronized with the labeled data from the annotated videos and segmented into 1-, 3-, and 5-s windows. Based on the six data streams (three from each sensor), we created 138 features and determined their importance using Gini importance. The final classifier includes 90 features for the dual accelerometer setup and 45 features for the single accelerometer setups. During the training of the classifier, hyperparameter tuning was carried out to determine the best possible configuration. The resulting configuration was then used to build individual classifiers for the dual and single accelerometer setups. LOOCV was thereafter performed to determine the performance of the classifier. See text for more details. B = back; T = thigh; a = annotated; F = features; LOOCV = leave-one-out cross-validation; XGBoost = extreme gradient boosting.

(Ahead of Print)

maintaining a frame rate of 25 fps. The videos were annotated frame by frame using the Anvil software package (Kipp, 2014) following a coding scheme for 10 physical activity types. The coding scheme was linked to a set of definitions, describing the onset and offset of the 10 target physical activity types (Supplementary Table S1 [available online]).

Figure 1 illustrates the main step in the development of the machine learning classifier. To prepare the accelerometer data for the training of the machine learning classifier, the data from each sensor were exported from the raw files, resampled, and filtered using a fourth-order Butterworth band-pass filter (Jackson, 2018). Finally, the back and thigh data were synchronized. The labeled data from the video annotations were synchronized with the accelerometer data, using three heel drops that were visible both in the accelerometer signal and video recording. As a result, six data streams of back and thigh accelerometer data were labeled with the respective activities.

The labeled accelerometer data were used as the ground truth to train, tune, and fit an XGBoost machine learning classifier. XGBoost was introduced by Chen and Guestrin (2016) as a novel implementation of the gradient boosting decision trees to increase efficiency, improve robustness, and prevent overfitting. An XGBoost classifier is an additive model that combines multiple sequentially aligned decision trees. Throughout the layers, each tree is built based on the results of trees in previous layers. Boosted ensembles are highly expressive. To prevent overfitting, the complexity of each tree and the total number of trees generated must be controlled.

The XGBoost works on features computed from the accelerometer data segments (e.g., 1-, 3-, or 5-s windows). We conducted preliminary investigations to determine relevant features by assessing all possible feature count combinations (i.e., from 1 to 138 features) and their contribution to overall accuracy using the Gini importance (Nembrini et al., 2018). This revealed that the performance of the model was saturated at about 40–45 features (i.e., no further improvement in performance with adding more features). Based on this, we selected 12 time-domain features (mean, standard deviation, skewness, kurtosis, average crossing rate, min/q25/q50/q75/max of the acceleration, total signal energy, and correlation between all axes) and three frequency domain features (mean amplitude of the signal, *SD*, and median frequency). Each window is therefore represented by 90 features for the dual accelerometer setup and 45 features for the single accelerometer setup, which are based on the 15 feature types computed for each axis with a 50% overlap of the windows. Furthermore, during the training of the classifier, we performed hyperparameter tuning using a grid search for the best F1 score over the following parameter ranges: learning rate set to 0.1; the maximum tree depth range between 15, 20, and 25; the maximum number of estimators between 20 and 70; subsample size is 0.6; and the lambda used for the L2 regularization of leaf node weights set to 1.0. The hyperparameter sweep was performed over two of the 22 leave-one-subject-out folds. Once the best configuration was identified, we built an individual model for the back and thigh, as well as for back and thigh individually. Notably, the probability of including multiple activity types increases with increasing window length and the XGBoost uses majority voting for each window to predict the activity type.

Figure 2 shows an example for one subject of accelerometer recordings from the thigh and back, the labeled physical activity types based on the video recordings (i.e., ground truth), the predictions produced by the XGBoost classifier, and the instances with true and false predictions.

Statistics and Performance Metrics

Two raters individually labeled the videos from two participants, and Cohen's kappa statistic was used to assess the interrater reliability. The predictive performance of the XGBoost classifier was assessed by using leave-one-out cross-validation. In this approach, the classifier is trained on the data from all participants except one, which is kept out and used as the test data set. Averaging the results across all participant data provide an estimate of the overall performance of classifier. This process was repeated for the dual and single accelerometer setups (back and thigh, thigh, and back) and the three window lengths (1, 3, and 5 s).

Performance metrics for the XGBoost classifier included precision, sensitivity (also termed recall), specificity, and overall accuracy. Precision for each of the target physical activity types was calculated as the sum of true positives divided by the sum of true and false positives. Sensitivity is the proportion of true positives that are correctly identified, calculated as the number of true positives divided by the sum of true positive and false negatives. Specificity is the proportion of true negatives that are correctly identified and was calculated as the number of true negatives divided by the sum of true negatives and false positives. Overall accuracy was calculated as the proportion of correctly classified instances divided by the total number of samples. The result of each of these performance metrics is a value between 0 and 1 with a higher number indicating better performance of the classifier. Due to the imbalance between activity classes (see Figure 3), we report the weighted performance metrics. Finally, Kappa statistics were calculated to provide a measure of agreement beyond that expected by chance.

Results

Figure 3 shows the distribution of the labeled video data and the merging of the 10 labeled subcategories into the six main target physical activity types. Total time with annotated video recordings was 37.9 hr, ranging from 1.6 hr for running to 17.5 hr for sitting. The Kappa value for the interrater agreement between the scoring by the two independent raters was .96.

For the dual accelerometer setup, overall accuracy in detecting the six target physical activity types was similar for the 1 s (95%), 3 s (96%), and 5 s (96%) window lengths. The corresponding overall accuracy was marginally lower when based on the single thigh accelerometer recordings, that is, 92%, 93%, and 93%, respectively. The overall accuracy was reduced to 80%, 84%, and 84% when based on the single back accelerometer recording. Furthermore, the Kappa values for the dual accelerometer setup and different window lengths were 0.91 (1 s), 0.92 (3 s), and 0.92 (5 s), respectively. The corresponding Kappa values for the thigh setup were .82 (1 s), .85 (3 s), and .85 (5 s), whereas for the back setup the values were .69 (1 s), .74 (3 s), and .74 (5 s), respectively.

Table 1 shows the performance metrics for detecting the six target physical activity types, based on the dual and single accelerometer setups and using 5-s windowing. Overall, precision, sensitivity, and specificity were high for detecting all six target physical activity types when based on the dual accelerometer recording. When based on a single accelerometer setup, the performance metrics were high (≥80%) for all target physical activity types except lying down when based on the single thigh accelerometer recording (i.e., precision 77%, sensitivity 64%) and standing when based on the single back accelerometer recording (i.e., precision 64%, sensitivity 54%).
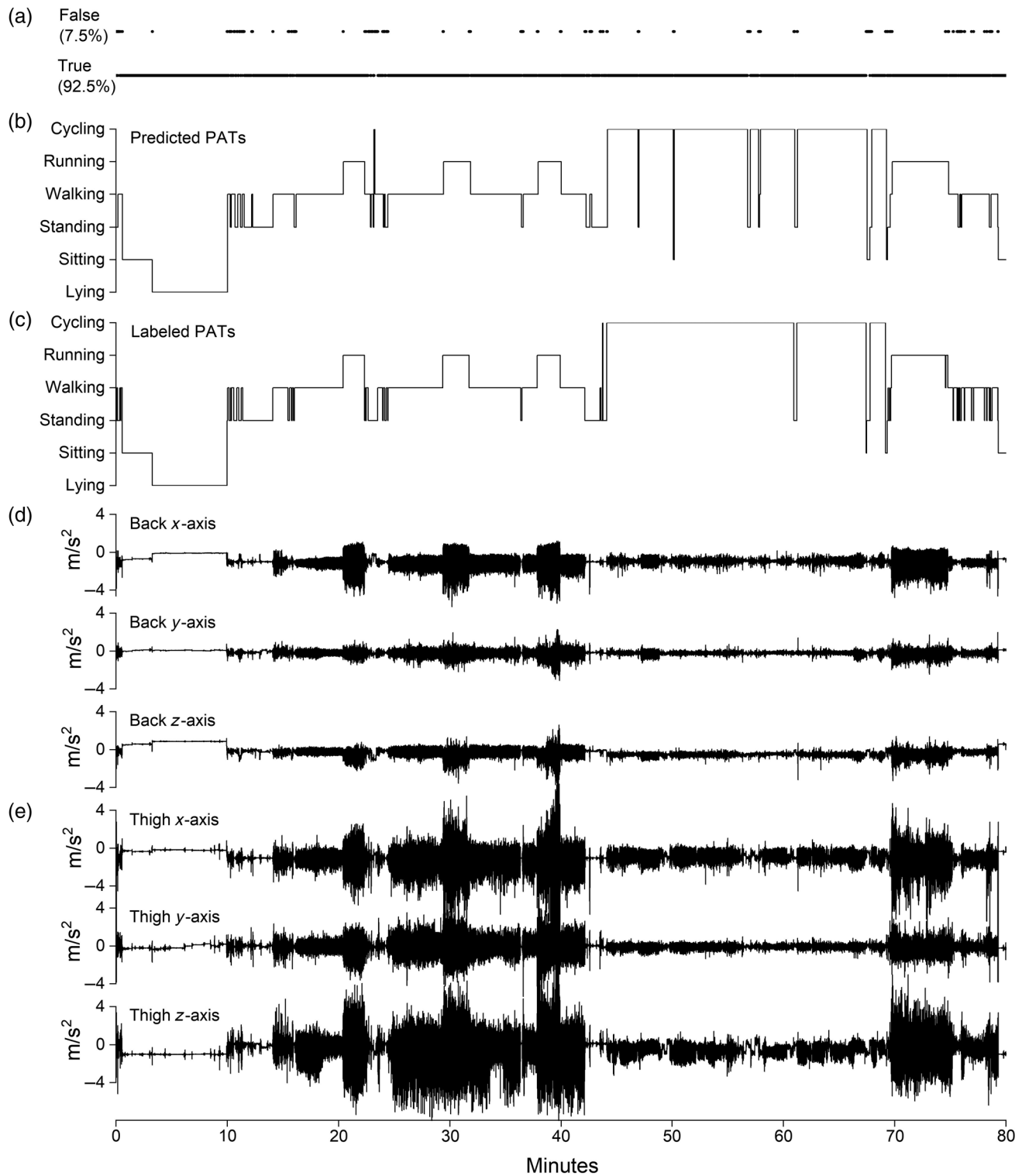
Figure 2 — Example of an 80-min recording period from one subject. The lower panels show the accelerometer recordings from thigh (e) and back (d). The middle panels show the labeled PATs based on the video recordings (c) and the predicted PATs (b) using a window size of 5 s. The upper panel shows the instances with true and false predictions (a). PATs = physical activity types.

Figure 4 shows the confusion matrixes for the dual (a) and single (b and c) accelerometer setups using 5-s windowing. For the back accelerometer recording, misclassifications were most pronounced for standing, that is, 37.7% of the time labeled as standing was misclassified as sitting (Figure 4b). For the thigh accelerometer recording, the poor detection of lying was mainly explained by 28.7% of the samples being misclassified as sitting (Figure 4c).
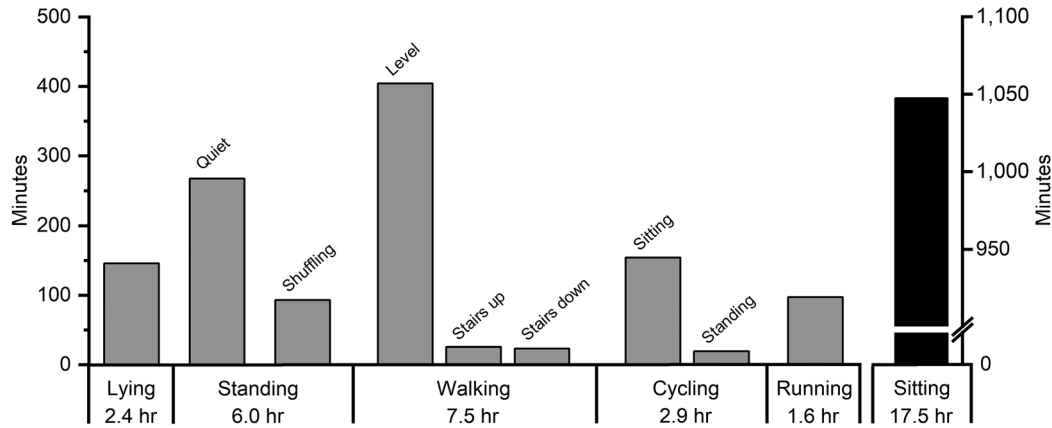
(Ahead of Print)

Figure 3 — Time distribution of the labeled video data for each subcategory and the six target physical activity types.

Table 1   Performance Metrics for Detecting the Six Target Physical Activity Types and Postures for the Dual and Single Accelerometer Setups, Using 5-s Windowing for Data Extraction

Sensor setup and activity type	Precision	Sensitivity	Specificity
Back and thigh			
Lying down	0.93 [0.83, 1.00]	0.99 [0.97, 1.00]	0.99 [0.99, 1.00]
Sitting	0.99 [0.98, 1.00]	0.99 [0.97, 1.00]	0.99 [0.98, 1.00]
Standing	0.88 [0.82, 0.93]	0.87 [0.84, 0.91]	0.98 [0.97, 0.99]
Walking	0.88 [0.83, 0.93]	0.90 [0.85, 0.96]	0.98 [0.97, 0.99]
Running	0.89 [0.81, 0.98]	0.91 [0.83, 1.00]	1.00 [0.99, 1.00]
Cycling	0.94 [0.92, 0.97]	0.92 [0.86, 0.98]	1.00 [0.99, 1.00]
Back			
Lying down	0.93 [0.83, 1.00]	0.99 [0.97, 1.00]	0.99 [0.97, 1.00]
Sitting	0.80 [0.71, 0.88]	0.86 [0.81, 0.92]	0.88 [0.85, 0.92]
Standing	0.64 [0.54, 0.74]	0.54 [0.47, 0.60]	0.95 [0.92, 0.97]
Walking	0.82 [0.74, 0.91]	0.88 [0.82, 0.95]	0.97 [0.95, 0.98]
Running	0.89 [0.79, 0.98]	0.87 [0.76, 0.97]	1.00 [0.99, 1.00]
Cycling	0.86 [0.74, 0.97]	0.83 [0.72, 0.94]	0.99 [0.99, 1.00]
Thigh			
Lying down	0.77 [0.61, 0.94]	0.64 [0.49, 0.79]	0.98 [0.97, 1.00]
Sitting	0.86 [0.78, 0.95]	0.88 [0.76, 1.00]	0.95 [0.93, 0.97]
Standing	0.90 [0.86, 0.94]	0.88 [0.85, 0.91]	0.99 [0.98, 1.00]
Walking	0.86 [0.79, 0.93]	0.89 [0.84, 0.95]	0.97 [0.96, 0.98]
Running	0.91 [0.83, 0.99]	0.91 [0.87, 0.95]	0.98 [0.96, 1.00]
Cycling	0.93 [0.89, 0.98]	0.90 [0.80, 1.00]	1.00 [0.99, 1.00]

Note. Values are mean calculated across participants with 95% confidence intervals in parentheses.

Discussion

The current results indicate that key daily physical activity types can be detected by a XGBoost classifier in a free-living setting with high overall accuracy. The overall accuracy was marginally better for the dual accelerometer setup compared to the single thigh accelerometer setup, while the overall accuracy for the single back accelerometer setup was markedly lower. The lower overall accuracy for the single accelerometer setups was mainly explained by poor detection of lying down (based on the thigh accelerometer recording) and standing (based on the back accelerometer recording). In other words, a single back accelerometer seems to provide limited possibilities for discriminating between sitting and standing, while a single thigh accelerometer does not discriminate well between lying and sitting. Thus, future studies that aim to assess physical activity types may consider these trade-offs when choosing sensor setup. The window length for processing the accelerometer data had minor influence on overall accuracy, except when using the single back accelerometer setup.

Numerous studies have been published in recent years, focusing on the validity of machine learning techniques to detect physical activity types based on accelerometer recording
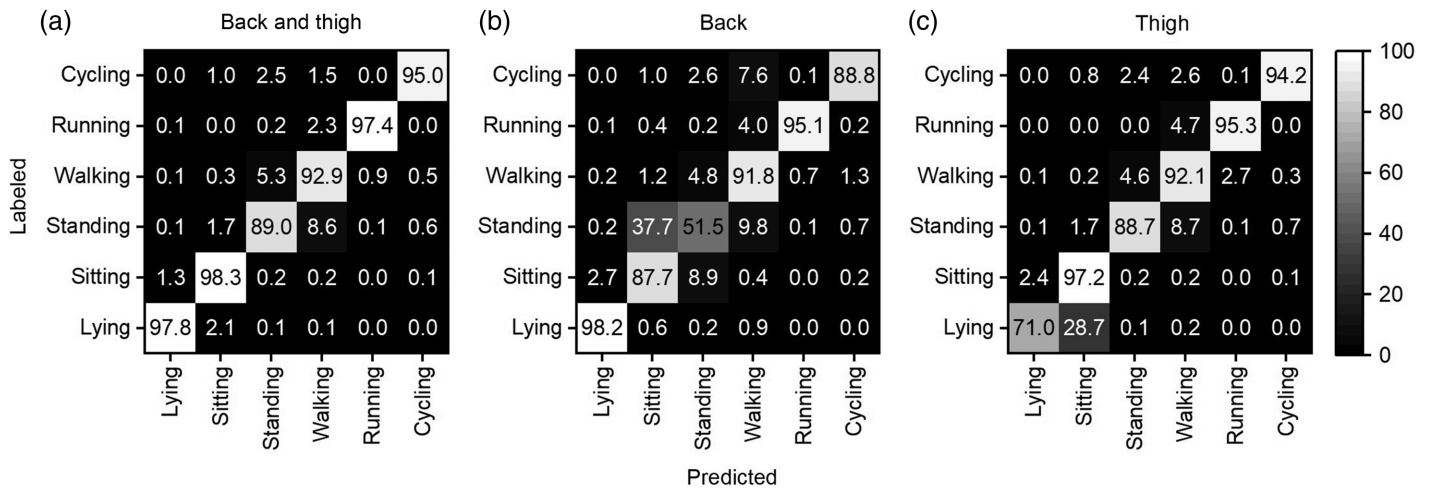
(Ahead of Print)

Figure 4 — Confusion matrixes for the dual (a) and single accelerometer setups (b and c) based on classification using 5-s windows. The labeled physical activity types are shown in the rows, while the predicted physical activity types are shown in the columns. Values are row percentages. The bar to the right indicates the shading according to percentages.

(Demrozi et al., 2020). Narayanan et al. (2020) used a dual accelerometer setup like what we used in the present study and showed that the same key physical activity types can be detected by a random forest classifier. However, they applied a semi free-living protocol for recording of training data, and they did not investigate if recordings from a single thigh or back accelerometer can be used to detect key daily physical activity types. Moreover, Giurgiu et al. (2020) found that thigh-worn accelerometers provide a valid detection of sedentary behavior, that is, merging sitting/lying postures. However, the training data were obtained using a semi-structured protocol in a laboratory setting, and they did not investigate to what extent the thigh accelerometer recording could be used to differentiate between a sitting and lying posture. The current study extends on the above studies, showing that an XGBoost classifier can be used to accurately detect key daily physical activity types during free-living based on a dual accelerometer setup. Moreover, a single accelerometer on the back or thigh can be used to accurately detect most key daily physical activity types. Although the overall accuracy was better for the thigh accelerometer recording compared to the back accelerometer recording (i.e., 93% vs. 84% using 5-s windowing), it should be noted that this difference was mainly explained by a poor precision in detecting standing based on the back accelerometer recording (i.e., standing misclassified as sitting; see Figure 4b). Moreover, it should be noted that the precision in detecting lying down was markedly attenuated when based on the single thigh accelerometer recording. A similar drop in accuracy of detecting standing and lying down was found in a recent study when based on a single thigh or lower back accelerometer recording compared to the dual accelerometer recording; however, the study was performed in a controlled lab setting (Stewart et al., 2018).

An apparent strength of machine learning classifiers is the possibility to integrate input from multiple sensors, thereby utilizing more data to achieve improved accuracy. Our study confirms this assumption, showing that a dual accelerometer setup outperforms a single accelerometer setup on all performance metrics. However, depending on the research questions, it may be enough to apply a single accelerometer on the thigh or lower back to record most key physical activity types. If the primary research question is

related to assessment of "time-on-feet" (standing) and physical activity (walking, running, and cycling), applying a single thigh-worn accelerometer can be enough (Crowley et al., 2019; Stevens et al., 2020). However, if the main purpose is to study sedentary behaviors and a necessity of differentiating between lying down and sitting, a back-worn accelerometer is recommended. It should also be noted that applying a dual-sensor setup in large-scale studies add both cost and some complexity related to the processing of data (e.g., the need of synchronizing sensor recordings). Moreover, placing the back sensor requires assistance while a single thigh sensor in principle can be placed by the participants himherself. These additional factors should be considered when choosing sensor setup in large-scale studies.

Window length has been shown to have impact on the performance of machine learning classifiers in detecting key daily physical activity types (Twomey et al., 2018). We used a sampling frequency of 50 Hz and found that window lengths of 3 or 5 s with 50% overlap for feature generation yielded similar precision, sensitivity, and specificity for both the dual and single accelerometer setups. Applying a 1-s window with 50% overlap resulted in a somewhat poorer performance of the XGBoost classifier, especially when based on the single back accelerometer recording. With large amounts of the data, it may therefore be advisable to use 5-s windowing to reduce the computational burden of the machine learning classifier. We did not investigate window lengths longer that 5 s, but the probability that a given window contains more than one physical activity type increases with increased window length. However, it should be noted that a longer window size can provide accurate measurement of physical activity with other sensor placements (Farrahi et al., 2019; Willetts et al., 2018).

There are apparent strengths of the current study, such as the recording of physical activity types during free-living, the exploration of the impact of window lengths on the performance of the XGBoost classifier, and a training data set of sufficient size to achieve high overall accuracy. However, there are some limitations that merit further discussion. First, we included a convenience sample of healthy adults. Thus, it is not clear if the performance of the current XGBoost classifier can be generalized to analyzing back/thigh accelerometer data from other population groups, such

as children, adolescents, and older adults; that is, this depends on whether the training data cover the variation in the target population. However, for the performance of the XGBoost classifier to differ significantly between these populations requires a substantial difference in the movement patterns during walking and running (static postures [i.e., standing, sitting, lying] and movements during cycling are not likely to differ between children, adolescents, and older adults). Although it is unlikely that such differences will have significant impact on the performance of the classifier, this needs to be confirmed in further studies. Second, to advance the field further, it has been advocated that it is necessary to consider the 24-hr compositional nature of key daily physical activity types to fully understand the physical activity–health associations (Biddle et al., 2018; Debache et al., 2019; Grgic et al., 2018; Rosenberger et al., 2019). In addition to the daily physical activity types detected in the current study, this requires the detection of intensity of dynamic physical activity types (i.e., walking, running, cycling) along with accurate detection of sedentary behaviors. For the latter, it is important to differentiate between sleep and awake time lying down. Thus, to advance the field further, it is essential that future work focus on developing measurement systems and analytic approaches that capture both activity type, intensity within activities, as well as the sleep–wake phase. Finally, it may be discussed to what extent the training data in the current study truly represent a free-living condition. Although there were no constraints on the physical activity types, we instructed the participants to accumulate at least a 2–3 min of sitting, standing, lying down, walking, and running during the recording period. Moreover, the participants were wearing the chest-mounted GoPro camera throughout the recording period. Thus, it is possible that the instructions given and the wearing of the camera influenced the behavior of the participants. However, the example included in Figure 2 illustrates the randomness and erratic pattern in shifts between different physical activity types. The exception was cycling that did not constitute a natural part of the physical activity types during the recording period.

Conclusion

Using an XGBoost classifier provides accurate detection of key daily physical activity types during free-living when based on a dual accelerometer recording. The overall accuracy in detecting key daily physical activity types decreased marginally when predictions were based on single thigh accelerometer recording but were somewhat poorer when based on the single back accelerometer recording. However, the lowered overall performance was mainly due to poor detection of lying down (when based on the thigh accelerometer recording) and standing (when based on the back accelerometer recording). Depending on the research questions, it may therefore be enough to record physical activity types with a single accelerometer on the thigh or lower back.

References

Arvidsson, D., Fridolfsson, J., & Borjesson, M. (2019). Measurement of physical activity in clinical practice using accelerometers. *Journal of Internal Medicine, 286*(2), 137–153. https://doi.org/10.1111/joim.12908

Biddle, G.J.H., Edwardson, C.L., Henson, J., Davies, M.J., Khunti, K., Rowlands, A.V., & Yates, T. (2018). Associations of physical behaviours and behavioural reallocations with markers of metabolic health: A compositional data analysis. *International Journal of Environmental Research and Public Health, 15*(10), 2280. https://doi.org/10.3390/ijerph15102280

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining.* San Francisco, CA. https://doi.org/10.1145/2939672.2939785

Crowley, P., Skotte, J., Stamatakis, E., Hamer, M., Aadahl, M., Stevens, M.L., Rangul, V., Mork, P.J., & Holtermann, A. (2019). Comparison of physical behavior estimates from three different thigh-worn accelerometers brands: A proof-of-concept for the Prospective Physical Activity, Sitting, and Sleep consortium (ProPASS). *International Journal of Behavioral Nutrition and Physical Activity, 16*(1), 65. https://doi.org/10.1186/s12966-019-0835-0

Debache, I., Bergouignan, A., Chaix, B., Sneekes, E.M., Thomas, F., & Sueur, C. (2019). Associations of sensor-derived physical behavior with metabolic health: A compositional analysis in the record multisensor study. *International Journal of Environmental Research and Public Health, 16*(5), 741. https://doi.org/10.3390/ijerph16050741

Demrozi, F., Pravadelli, G., Bihorac, A., & Rashidi, P. (2020). Human activity recognition using inertial, physiological and environmental sensors: A comprehensive survey. *IEEE Access, 8,* 210816–210836. https://doi.org/10.1109/ACCESS.2020.3037715

Doherty, A., Jackson, D., Hammerla, N., Plotz, T., Olivier, P., Granat, M.H., White, T., van Hees, V.T., Trenell, M.I., Owen, C.G., Preece, S.J., Gillions, R., Sheard, S., Peakman, T., Brage, S., & Wareham, N.J. (2017). Large scale population assessment of physical activity using wrist worn accelerometers: The UK Biobank Study. *PLoS One, 12*(2), e0169649. https://doi.org/10.1371/journal.pone.0169649

Farrahi, V., Niemelä, M., Kangas, M., Korpelainen, R., & Jämsä, T. (2019). Calibration and validation of accelerometer-based activity monitors: A systematic review of machine-learning approaches. *Gait & Posture, 68,* 285–299. https://doi.org/10.1016/j.gaitpost.2018.12.003

Fuzeki, E., Engeroff, T., & Banzer, W. (2017). Health benefits of light-intensity physical activity: A systematic review of accelerometer data of the national health and nutrition examination survey (NHANES). *Sports Medicine, 47*(9), 1769–1793. https://doi.org/10.1007/s40279-017-0724-0

Giurgiu, M., Bussmann, J.B.J., Hill, H., Anedda, B., Kronenwett, M., Koch, E.D., Ebner-Priemer, U.W., & Reichert, M. (2020). Validating accelerometers for the assessment of body position and sedentary behavior. *Journal for the Measurement of Physical Behaviour, 3*(3), 253–263. https://doi.org/10.1123/jmpb.2019-0068

Grgic, J., Dumuid, D., Bengoechea, E.G., Shrestha, N., Bauman, A., Olds, T., & Pedisic, Z. (2018). Health outcomes associated with reallocations of time between sleep, sedentary behaviour, and physical activity: A systematic scoping review of isotemporal substitution studies. *International Journal of Behavioral Nutrition and Physical Activity, 15*(1), 69. https://doi.org/10.1186/s12966-018-0691-3

Heesch, K.C., Hill, R.L., Aguilar-Farias, N., van Uffelen, J.G.Z., & Pavey, T. (2018). Validity of objective methods for measuring sedentary behaviour in older adults: A systematic review. *International Journal of Behavioral Nutrition and Physical Activity, 15*(1), 119. https://doi.org/10.1186/s12966-018-0749-2

Jackson, D.G. (2018). OpenMovement. *GitHub.* https://github.com/digitalinteraction/openmovement/tree/master/Software/AX3/omconvert

Kipp, M. (2014). ANVIL: The Video Annotation Research Tool. In J. Durand, U. Gut, & G. Kristoffersen (Eds.), *Handbook of corpus phonology* (pp. 420–436). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199571932.001.0001

Migueles, J.H., Cadenas-Sanchez, C., Ekelund, U., Delisle Nystrom, C., Mora-Gonzalez, J., Lof, M., Labayen, I., Ruiz, J.R., & Ortega, F.B. (2017). Accelerometer data collection and processing criteria to assess physical activity and other outcomes: A systematic review and practical considerations. *Sports Medicine, 47*(9), 1821–1845. https://doi.org/10.1007/s40279-017-0716-0

Narayanan, A., Stewart, T., & Mackay, L. (2020). A dual-accelerometer system for detecting human movement in a free-living environment. *Medicine & Science in Sports & Exercise, 52*(1), 252–258. https://doi.org/10.1249/MSS.0000000000002107

Nembrini, S., König, I.R., & Wright, M.N. (2018). The revival of the Gini importance? *Bioinformatics, 34*(21), 3711–3718. https://doi.org/10.1093/bioinformatics/bty373

Rosenberger, M.E., Fulton, J.E., Buman, M.P., Troiano, R.P., Grandner, M.A., Buchner, D.M., & Haskell, W.L. (2019). The 24-hour activity cycle: A new paradigm for physical activity. *Medicine & Science in Sports & Exercise, 51*(3), 454–464. https://doi.org/10.1249/MSS.0000000000001811

Skotte, J., Korshoj, M., Kristiansen, J., Hanisch, C., & Holtermann, A. (2014). Detection of physical activity types using triaxial accelerometers. *Journal of Physical Activity and Health, 11*(1), 76–84. https://doi.org/10.1123/jpah.2011-0347

Stevens, M.L., Gupta, N., Inan Eroglu, E., Crowley, P.J., Eroglu, B., Bauman, A., Granat, M., Straker, L., Palm, P., Stenholm, S., Aadahl, M., Mork, P.J., Chastin, S., Rangul, V., Hamer, M., Koster, A., Holtermann, A., & Stamatakis, E. (2020). Thigh-worn accelerometry for measuring movement and posture across the 24-hour cycle: A scoping review and expert statement. *BMJ Open Sport & Exercise Medicine, 6*(1), e000874. https://doi.org/10.1136/bmjsem-2020-000874

Stewart, T., Narayanan, A., Hedayatrad, L., Neville, J., Mackay, L., & Duncan, S. (2018). A dual-accelerometer system for classifying physical activity in children and adults. *Medicine & Science in Sports & Exercise, 50*(12), 2595–2602. https://doi.org/10.1249/MSS.0000000000001717

Twomey, N., Diethe, T., Fafoutis, X., Elsts, A., McConville, R., Flach, P., & Craddock, I. (2018). A comprehensive study of activity recognition using accelerometers. *Informatics, 5*(2), 27. https://doi.org/10.3390/informatics5020027

Van Der Velde, J.H.P.M., Koster, A., Van Der Berg, J.D., Sep, S.J.S., Van Der Kallen, C.J.H., Dagnelie, P.C., Schram, M.T., Henry, R.M.A., Eussen, S.J.P.M., Van Dongen, M.C.J.M., Stehouwer, C.D.A., Schaper, N.C., & Savelberg, H.H.C.M. (2017). Sedentary behavior, physical activity, and fitness—The Maastricht Study. *Medicine & Science in Sports & Exercise, 49*(8), 1583–1591. https://doi.org/10.1249/MSS.0000000000001262

Watson, K.B., Carlson, S.A., Carroll, D.D., & Fulton, J.E. (2014). Comparison of accelerometer cut points to estimate physical activity in US adults. *Journal of Sports Sciences, 32*(7), 660–669. https://doi.org/10.1080/02640414.2013.847278

Willetts, M., Hollowell, S., Aslett, L., Holmes, C., & Doherty, A. (2018). Statistical machine learning of sleep and physical activity phenotypes from sensor data in 96,220 UK Biobank participants. *Scientific Reports, 8*(1), 7961. https://doi.org/10.1038/s41598-018-26174-1