Activity Recognition in Youth Using Single Accelerometer Placed at Wrist or Ankle

Andrea Mannini[1], **Mary Rosenberger**[2], **William L. Haskell**[3], **Angelo M. Sabatini**[1], and **Stephen S. Intille**[4]

[1]The BioRobotics Institute, Scuola Superiore Sant'Anna, Pisa, Italy

[2]Stanford Center on Longevity, Stanford University, Stanford, CA

[3]Stanford Prevention Research Center, Stanford University, Stanford, CA

[4]College of Computer and Information Science and Bouvé College of Health Sciences, Northeastern University, Boston, MA

Abstract

Purpose—State-of-the-art methods for recognizing human activity using raw data from body worn accelerometers have primarily been validated with data collected from adults. This study applies a previously available method for activity classification using wrist or ankle accelerometer to work on datasets collected from both adults and youth.

Methods—An algorithm for detecting activity from wrist-worn accelerometers, originally developed using data from 33 adults, is tested on a dataset of 20 youth (age 13±1.3). The algorithm is also extended by adding new features required to improve performance on the youth dataset. Subsequent tests on both the adult and youth data were performed using crossed tests (training on one group and testing on the other) and leave-one-subject-out cross-validation.

Results—The new feature set improved overall recognition using wrist data by 2.3% for adults and 5.1% for youth. Leave-one-subject-out cross-validation accuracy performance was 87.0% (wrist) and 94.8% (ankle) for adults, and 91.0% (wrist) and 92.4% (ankle) for youth. Merging the two datasets, overall accuracy was 88.5% (wrist) and 91.6% (ankle).

Conclusions—Previously available methodological approaches for activity classification in adults can be extended to youth data. Including youth data in the training phase and using features designed to capture information on the peculiar activity fragmentation of young participants allows a better fit of the methodological framework to the characteristics of activity in youth, improving its overall performance. The proposed algorithm differentiates ambulation from sedentary activities that involve gesturing in wrist data, such as that being collected in large surveillance studies.

Corresponding Author. Andrea Mannini, Polo Sant'Anna Valdera, viale Rinaldo Piaggio 34, 56025 Pontedera (Pisa), Italy Tel: +39-050-88-3410, fax: +39-050-88-3101, a.mannini@sssup.it.

Conflicts of Interest The authors have no conflicts of interest to disclose.

Introduction

Traditional methods of measurement of physical activity include using a wearable device on the hip over a seven-day period of activity to classify participants into activity categories. Accelerometer-based activity monitors that output activity "count" values aggregate the motion of the device over a short window of time ("epoch"). The count value does provide a useful summary of overall motion along with some indication of gross motion and ambulation when placed at the hip. A recent trend, however, is to move the activity monitor from the hip to the wrist location, to increase wear-time compliance and capture sleep-related behavior, as in the UK BioBank and U.S. National Health and Nutrition Examination Survey (NHANES) studies (26, 30). This trend recently started to involve physical activity estimation in youth using wrist worn accelerometers (7). Interpretation of wrist-based count values is challenging due to hand gesturing, which may confound the mapping between overall body ambulation and motion of the sensor (20). More detailed information about wrist motion captured in raw accelerometer data sampled at a high rate, however, may permit automatic identification of classes of activity types, such as ambulation vs. sedentary behavior. This information could be used directly to characterize activity, or perhaps to improve accuracy of accelerometer-based energy expenditure estimation (1, 8).

Prior work detecting activity type from raw accelerometer data from a variety of sensor locations on the body has primarily focused on detecting the activity of adults (e.g., (2, 16, 34)). Recent work on the detection of activity type from raw accelerometer data on the wrist has focused mainly on adults (16, 24, 34) with the only exception being the work by Trost *et al* (29). In this study, we tested the applicability of an activity recognition algorithm based on ankle or wrist raw accelerometer data, previously developed for and validated in adults with youth aged 11 to 15 (16). The type and amount and intensity of activity of children and youth may differ from those of adults (5, 27). Algorithms developed for adults may not work well on children for one of two reasons: activities that children perform may not be represented in the adult models, and features used by the models for adults may not adequately capture important distinctions between activities in children, if children perform those activities in dramatically different ways. Whereas physical activity assessment in children and youth using activity count values obtained using accelerometry is common (7, 21), researchers are still exploring whether the same models that worked for adults can be applied to children.

Activity recognition in youth: related work

Previous studies involving automatic activity recognition in youth from accelerometers have been surveyed (see Table 1). Four studies use activity count values. One such study using classification features extracted from activity counts (1 s epochs) gathered from 41 youth aged 10.8 ± 1.3 years classified activity into 10 categories (stationary, biking, crawling,

walking, scooter, horseback riding, jumping and *floor exercise*) with 67% accuracy using monitor data from the hip and wrist simultaneously (23). De Vries *et al.* tested single sensor solutions on 58 participants, using ankle and hip sensors worn by 9-12 yo youth that output 1 s count values (9). When classifying seven activity types (sitting, standing, walking, running, rope skipping, playing soccer, and *cycling*) from 20 minutes of data per person, they obtained an overall accuracy of 68% ankle data, and 77% using hip data. More recently Trost *et al.* attempted the classification of five activity types (sedentary, walking, running, light intensity household activities or games, moderate-to-vigorous intensity games or sports) from activity count data with 1 s epochs, evaluating the algorithm on data from 100 participants (5 to 15 yo), with two minutes of classified data for each activity (28). An overall accuracy of 88.4 % from a hip-worn sensor was reported. The most recent paper to include 1 s epoch counts processing for activity recognition was by Hagenbuchner *et al* who involved 11 pre-school children (3 to 6 yo, with a total of 264 minutes of classified data) (12). Four classes were recognized (*sedentary activities*, *light activities*, *moderate to vigorous activities*, *walking* and *running*) reporting 82.6% accuracy.

An alternative approach to using count values is to use the raw accelerometer data and compute a richer set of features that may help differentiate specific activities. Four recent studies have explored this approach for activity type detection in children. In Hikihara *et al.* (13), 32 Hz data from a waist-worn triaxial accelerometer were used to distinguish between two *non-locomotive* and *locomotive* activities in 68 6-12 yo children with approximately one hour of data from each child, classifying 99.1% of examples correctly. Nam and Park (19) proposed a method using a waist-worn accelerometer and a barometric pressure sensor to classify 11 classes (wiggling, rolling, standing still, standing up, sitting down, walking, toddling, crawling, climbing up, climbing down and stopping) of 10 1.3-2.4 yo toddlers. With a total of 50 h of acquired data, they obtained 88.3% classification accuracy using the accelerometer alone and 98.4% using both the accelerometer and barometric pressure sensors. Del Rosario *et al.* (10) performed activity classification using a smartphone embedded accelerometer, gyroscope and barometric pressure sensor, evaluating performance on 20 young adults (21.9 ± 1.7 yo) and 37 older adults persons (83.9 ± 3.4 yo). The same feature set was used in both age groups to classify 9 activities (stand, sit, lie, walk, walk upstairs, walk downstairs) from 10-30 minutes of data per person collected when the smartphone was kept in a person's trousers front pocket. Overall recognition accuracies of 79.9% and 82.0% for young and older adults, respectively were reported. Finally, Trost *et al.* (29) proposed a method based on wrist or hip accelerometer for recognizing 12 activities merged to 7 categories (lying down, sitting, standing, walking, running, basketball and *dancing*). A total of 52 (13.7 ± 3.1 yo) children were included in tests and 2 min of each activity were classified obtaining 91% and 88.4% accuracy on average at the hip and at wrist, respectively.

Previous work did not test activity classification methods designed for adult activity on youth data. In this work, we fill this gap by testing an activity recognition algorithm we previously developed for adults (16) on data from youth performing a similar set of activities. The objective behind this approach is to check if it is possible to apply existing methods to this different age group and if it is possible to improve its results in a general solution that could effectively process data from both age groups. While it is theoretically

possible to create algorithms tuned for all age ranges, that would require large age-specific data collection efforts, with somewhat arbitrary age cutoffs. Moreover, deployment of algorithms would be simpler in large surveillance studies if the same algorithms could be used for both adults and children. Even if the algorithms could have to be tuned differently based on age, using the same set of features for adults and children might ultimately allow algorithms to be developed that adapt smoothly to different age groups, rather than using a hard (and unrealistic) age threshold to process data using two entirely different algorithms and models. . Both the youth and adults data in our work were acquired simulating free living conditions: participants were asked to do things as naturally as possible while doing a set of 23 activities for youth and 26 activities for adults classified among 4 general classes. Sensors were again placed at the ankle and wrist (each sensor was tested independently). Modifications to the algorithm were proposed – specifically in the selection of features computed from the raw data and fed to classification algorithm – that allow better detection of differences in the activities being performed by the youth. The resulting method was tested on the adults' data as well as the combined dataset.

Materials and methods

1. Datasets

Participants performed a set of simulated daily activities in a lab environment while wearing a suite of synchronized sensors, following a similar protocol used when collecting data from 33 adults in prior work (16). Twenty youth (12 boys 8 girls, ages 13±1.3) were recruited from the Stanford, California community. The Stanford University's Institutional Review Board approved the data collection protocol, and written informed consent was obtained before participation. Triaxial Wocket accelerometers (14) were secured to the wrist and ankle positions on the body using custom Velcro bands. Wockets are small, thin, and lightweight devices ($43 \times 30 \times 7$ mm, 13 g) making them particularly suitable for long-term physical activity monitoring studies. Raw acceleration data (range ± 4 g) were acquired at 90 Hz and sent using the Bluetooth wireless protocol to a smartphone. The wrist sensor was placed on the dorsal aspect of the dominant wrist midway between the radial and the ulnar process. The ankle sensor was placed on the outside of the ankle, just above the lateral malleolus. The ankle placement site was chosen because it is an ideal site for ambulation detection (16). The wrist is a practical site for long term monitoring, because sensors can be attached using watch-like bands, can be worn during sleep comfortably, and do not need to be removed when changing clothes.

Participants were asked to perform a guided sequence of laboratory-based physical activities and common daily activities lasting 3-5 min each. Activities were annotated during the execution of tasks using a voice recorder, and then timings on the voice recording were used to annotate start/stop times for specific activities being observed. Data and annotation were synchronized using custom software (14). Data collected from the youth using this procedure will be identified as the Y-dataset (Youth). Similar data collected in prior work from adults will be identified as the A-dataset (Adult). Table 2 summarizes the list of available activities in each dataset. Activities were grouped into four classes: *sedentary, cycling, ambulation* and *other activities*. Multitasking behaviors were not included, except

for the activity "walking-carrying a load" in A. Activity categories were reduced to four categories used in previous work on wrist/ankle activity recognition in adults (16). Activities that were done in an upright posture but that can be physical demanding, such as cleaning and wall painting, were included in an *other activities* because they did not seem appropriate for the *sedentary*, *ambulation* and *cycling* classes. By collapsing activities into categories, the machine learning algorithms had more training data; future work with much larger datasets could explore detection of specific activities as well. To facilitate comparison with past work, new sport and leisure activities of the Y dataset involving movement in the upright position and not represented in A, were included in the *other activities* class.

2. Data preprocessing and features evaluation

Three-axis raw accelerometer data were preprocessed to extract the Signal Magnitude vector (SM):

$$SM = \sqrt[2]{acc_x{}^2 + acc_y{}^2 + acc_z{}^2}$$

where *acc* indicated the recorded raw data in *g*-units (1 g = 9.81 m/s^2). The resulting 90 Hz SM signal was independent of the orientation of the sensing node. SMs were low pass filtered using a 15 Hz cut-off 4th order Butterworth filter to limit the bandwidth of the signal to the frequencies common in human motion (3). To classify data within the four defined activity classes, the SMs data were divided into 12.8 s non-overlapping windows. This window size was proposed by Zhang *et al.* (34) and also applied in Mannini *et al.* (16). Although prior work has shown that other window sizes (e.g., 4s) can be used with only modest degradation of performance (16), here use of the same window length value as in two prior studies allows a direct comparison between these results and those from the previous one about activity classification in adults(16).

The datasets used in this work were collected in the lab but include semi-structured activities labeled in real time, introducing small errors in annotating activity transitions due to reaction time and the difficulty of labeling activities when transitions occur quickly. For this reason, one window (12.8 s) was discarded before and after each label transition. Another type of annotation error is that some short activity changes during semi-structured activities were not labeled at all. For example, the dataset contains examples where participants stop briefly during non-treadmill walking, such as at a door that had to be opened. In such cases, even though a participant is standing still briefly, the label for the data is still *ambulation*. Some errors can be detected using the ankle acceleration recordings because in data labeled as *ambulation*, the SM of the ankle sensor is expected to be significant. Therefore the ankle sensor SM was used both to identify these labeling errors and to correct labels indicating ambulation. In particular, 2 s windows labeled as *ambulation* with a standard deviation less than 0.1 g were marked as labeling errors and discarded. This value was set by empirical observation so that ambulation with a cadence of 1 impact per window (i.e., nearly any movement) would be detected with an ankle sensor. Data-loss due to Bluetooth wireless transmission errors was handled by discarding windows with less than the 80% of the number of expected samples at the nominal 90 Hz sampling rate. In such cases, a new

window was started at the end of the data gap. Some fluctuations in the sampling rate may occur in the remaining windows due to the wireless connection. Before extracting frequency domain features, SMs in each window were linearly interpolated to obtain the same number of samples in every window.

Initially, the feature set defined in previous work was tested (16). Those features, listed in Table 3, encoded both temporal and frequency domain information, computed from acceleration SM. As in our prior work (16), the feature vectors (computed on every 12.8 s window) were used as input for a Support Vector Machine (SVM) with radial basis function kernel (31). Also as in prior work (16), the results were evaluated using cross-validation with Leave-One-Subject-Out (LOSO). Finally, the parameters of the radial basis function used for SVM classification were retained from the previous study (upper complexity bound $C = 100$ and $\gamma = 0.1$) for evaluating the previous version of the algorithm and then optimized by running a grid search using classification accuracy as the optimization criterion. As described below, our initial testing suggested that the youth dataset recognition could be improved by adding some additional features to the algorithm. The adult dataset upon which the algorithm was developed did not include sports activities such as basketball, soccer, and tennis that were included in the youth data. To capture such activities that are more frequent in youth than in adults, the signal power at frequencies higher than 3.5 Hz, normalized by the total power, was included as a new frequency domain feature. The 3.5 Hz cutoff frequency was selected because previous studies pointed out that most of the energy of human movement during daily activities lies in the 0.3-3.5 Hz band (25). High frequency components, which are present especially in lower limb recordings, are mainly from high impacts (4); therefore, their presence suggests that the movement being performed involves high impact, as is common in ambulation or sport.

New features were introduced to capture information about starts and stops of short activity bouts that may be common for children (21, 22). Evaluating relevant acceleration bouts within the window allows to extract information about the activity fragmentation, while keeping the same window length used in previous work (16, 34). The fragmentation of the acceleration signal magnitude was evaluated as it is done in evaluating onset of electromyographic signals (18):

- SM were rectified by subtracting a constant value corresponding to gravitational acceleration (1 g) and removing the sign of the result.

- Windowed data were then low pass filtered (Butterworth filter, 5 Hz cut-off frequency).

A threshold ($Th = 0.2*$g, i.e. 1.96 m/s^2) to identify active samples was then applied. In previous studies, periods with unfiltered acceleration SM lower than 0.4 m/s^2 are considered static (32). Our threshold results from preliminary observation of rectified and filtered SM; it is significantly different, because our aim is to identify with high specificity periods of relevant activity, as opposed to periods with little motion.

Four different activity fragmentation features were evaluated as follows:

1. *Fragmentation, active samples:*

$$F_S = \frac{\#\text{samples} > Th}{W}$$

Where W is the window length in samples. This feature identifies the amount of activity within the window that is over the threshold, thereby providing a rough estimate of the amount of relevant activity being recorded in the window. This feature could distinguish between activities that result in relevant acceleration in most of the window and activities in which the relevant activity takes place for only a portion of the window.

2.   *Fragmentation, number of activations:*

$$F_{AN} = \frac{\#\text{threshold\_ exceedings}}{\#\text{samples} > Th}$$

This feature identifies the number of threshold crossings within the window (rising edges only), normalized to the number of active samples F_S, thereby capturing movement fragmentation within the window. This feature could discriminate impulsive events from longer-lasting acceleration events, since it quantifies how many times within the window the acceleration passed from the inactive to the active condition.

3.   *Fragmentation, mean activation interval duration*

$$F_{AM} = \frac{m}{W} \quad \text{where,}$$
$$m = \frac{1}{\#\text{threshold\_ exceedings}} \sum \#consecutive \quad fragmentation \quad samples$$

This captures the mean duration of activation intervals within the window, normalized by the window length. An activation interval is defined as the amount of samples between two consecutive threshold crossings. This feature provides information on movement bout fragmentation within the window that could help discriminate between activities that involve stable movements, such as those in natural walking, and those with short bouts, such as sport ones.

4.   Fragmentation, activation interval duration variability

$$F_{AV} = \frac{s}{W} \quad \text{where,}$$
$$s = \sqrt{\frac{\sum (\#consecutive \quad fragmentation \quad samples - m)^2}{\#\text{threshold\_ exceedings} - 1}}$$

This feature captures the standard deviation of the duration of activation intervals within the window, normalized by the window length, thereby providing information on uniformity of activation intervals within the window. This feature may help discriminate between activities with cyclic

movements with a very stable ratio between active and inactive phases, and more random activities. A stable cyclic movement would result in lower variability of activation intervals that would repeat themselves within the window. Fast and aperiodic movements, however, such as those in recreational activities would result in highly variable activation bouts.

If no samples over the threshold were observed, F_S, F_{AN}, F_{AM} and F_{AV} were set to 0. F_{AV} was set to zero also in the case of *#threshold_exceedings* less than 2.

In summary, six new features specifically developed for activities common in children were computed and available to the classification algorithm. All features are computationally efficient, permitting real-time recognition in future systems.

3 Feature selection, classification and validation strategies

In prior work, manual testing was used to assess the contribution of some group of features to overall algorithm performance (16). Here, new features were added to the feature space, and an automatic sequential forward search feature selection process (15) was adopted, where the LOSO validation output was used as the selection criterion. Sequential forward search is a suboptimal algorithm for feature selection that cannot guarantee the optimality of the selected set (15). However, it is a computationally acceptable (linear time complexity) dimensionality reduction strategy that may improve classification results by discarding redundant features prior to the classification step (15). SVM classifier training does not rank features (6). Therefore, feature set dimensionality reduction is effective for SVM classifiers. Reducing feature space dimensionality typically reduces the amount of training examples required to obtain reliable recognition results, because there is a lower risk of over-fitting examples and better generalization (15, 33). Moreover, by reducing the number of features, the complexity of the classifier is reduced (less parameters are needed) then the computational cost of both training and real-time recognition is reduced (33). In prior work the same feature sets were used for both the ankle and wrist algorithm training (16). Here, however, the algorithm could select different feature sets for the two sites; automatic feature selection was run on each site separately to obtain a location-specific feature set, given that signals may have different characteristics at different body sites (3, 17).

The LOSO cross-validation approach was preferred over standard n-fold cross-validation. In standard n-fold cross-validation, data are mixed from all subjects and held-out data are randomly selected; LOSO, alternatively, prevents similar data collected from the same participant at about the same time from ending up in both the training and test datasets (11, 16). Therefore, LOSO results are more likely to demonstrate how a method may work under realistic conditions where a new participant, not included in the training data, is tested. LOSO is particularly challenging if testing occurs across different populations of people, such as training on adults and testing on youth, as done here. Most previous studies on activity classification did not apply LOSO cross-validation (13, 19, 23, 28, 29). A few studies did use this type of cross-validation (9, 10, 12), but they limited their testing and evaluation to a homogeneous pool of healthy adult users, with the exception of Del Rosario *et al.* who involved two age groups: 37 elderly (average age 84 yo) and 20 young adults (average age 22 yo) (10).

The two datasets used in this work, adult activity (A) and youth activity (Y), were collected as part of two separate experiments. All activities in both datasets were chosen to represent common activities, and these are significantly different in an adult and youth population, as shown in Table 2. In LOSO validation using data from a single age group, all activities for that class listed in Table 2 were used. For LOSO validation using both age groups, classes were removed if no training data were available for that class. For example, when training on A and testing on Y, all the activities in A were trained, but those activities not available in A but in Y were removed from Y (i.e., the sport activities in the *other activities* class and the video gaming activities in the *sedentary* class were removed from Y for testing, because no training data were available for them). Alternatively, when training on Y and testing on A, activities in A that did not have training data were removed (i.e., *painting with roller* and *painting with brush*).

Testing proceeded as follows:

> – *Experiment 1:* Test the algorithm originally developed for A on Y. This first experiment was aimed at evaluating if the previous existing methodology could be extended without modification to Y by running a LOSO cross-validation on Y.

> – *Experiment 2*: Extend the feature set to incorporate additional features intended to capture more information about the youth activities, and train and the test using LOSO validation separately on the Y and A datasets. This was done to check if the proposed modification to the methodology improves the recognition accuracy respect to the original algorithm on both datasets.

> – *Experiment 3*: Cross-test the algorithm by training on A and testing on Y, and training on Y and testing on A. Both the original feature set and the newly proposed set are evaluated. As mentioned above, activities not represented in the training set (i.e., without training data) were removed from the test set. The goal of this experiment was to assess whether information learned from one of the groups could be used to recognize similar activities in the other.

> – *Experiment 4*: Test the algorithm with both datasets combined. LOSO validation is performed on A+Y, without removing any activities from either dataset. This test was performed to check if it is possible to obtain a general classifier that works equally-well on both groups.

Results

1. Experiment 1: Test the algorithm originally developed for A on Y

The first test consisted of running a LOSO validation on the youth dataset, using the previously existing feature set (16). Ankle and wrist classifications were correct in the 85.9% and 89.7% of cases, respectively. Results are reported in Table 4, part A.

2. Experiment 2: Extend the feature set

Table 3 shows the feature sets obtained for wrist and ankle activity classification after (1) adding the new features, (2) applying the feature selection approach and (3) optimizing SVM parameters.

The feature selection procedure was run independently for the wrist and ankle, selecting 9 wrist features and 7 ankle features. These location-optimized feature sets (including features designed to capture more information about the youth activities) led to improved results, as summarized in Table 4, part B, and in Table 5. By using the new feature set with the automatic feature selection, an improvement in overall classification accuracy of 2.7% was obtained for the ankle and 5.1% for the wrist. The detailed classification results for each type of activity are reported in Table 5.

SVM parameter optimization was used to find radial basis function parameters for the ankle ($C = 16$, $\gamma = 0.25$) and wrist ($C = 128$, $\gamma = 0.0625$) classifiers. However, comparing these parameter optimization results with the previously proposed configuration ($C = 100$, $\gamma = 0.1$) improved overall accuracy by less than 0.5% for both ankle-based and wrist-based classification.

Solution robustness respect to different window sizes (8 s , 6.4 s, 4 s and 3.2 s) was tested. Classification accuracy decreased when reducing the window length. However, even in the worst case, the accuracy remained higher than 80% (83.5% at the wrist and 86.9% at ankle for the smallest window).

3. Experiments 3 and 4: Tests on both A and Y datasets

Table 6 shows results obtained by training on one dataset and testing on the other (experiment 3). Both the new and old feature sets were tested. Classification accuracies varied in this case from 79.3% to 87.4% at the ankle and from 58.8% to 71.8% at wrist.

Table 4, part C shows the results of running LOSO cross validation with the combined datasets including all 53 participants (experiment 4). The classifier was trained without respect to age group. The overall accuracy for this validation test reached 88.5 % at wrist and 91.6% at ankle. The contribution of the youth and adult results to the overall accuracy are reported in Table 6. Table 6 also summarizes accuracy results for all the tested experiments, showing the overall recognition accuracy and the recognition accuracy of each activity class.

Discussion

Feature sets

The extension of the methodology presented in Mannini *et al*. (16) to the Y dataset (experiment 1), using the same feature set and classification approach used on the original A dataset, produced classification results comparable to those in most previous studies (see Table 4 part A and Table 1). However, a significant improvement on this new group was obtained in experiment 2 by adding 4 (wrist) or 3 (ankle) new features and pruning 8 of the old features, as identified by automatic selection (table 4 part B). In addition to features

related to basic signal structure (mean, standard deviation, acceleration range), activity fragmentation features appear to be an important source of information (see Table 3) for both ankle and wrist sensing sites. The ankle classifier exploited the ratio between the dominant frequency of the currently tested window and the dominant frequency of the previous window (i.e., the prior 12.8 s). This feature, used in the previous adult-only study as well (16), captures temporal information that is useful for identification of consistently periodic behaviors such as ambulation or cycling at the ankle site. This feature was discarded at the wrist site; the activities tested here may exhibit more constant, periodic motion at the ankle than at the wrist.

The activity fragmentation features that were included to capture information about some of the youth activity, such as sports, encode information about signal power, taking into account the amplitude of the signal, the time duration and frequency of significant acceleration episodes within the window. Accordingly, when these features were available, several power-related features were discarded by the feature selection strategy. Similarly, the introduction of the "range" feature, jointly with the "maximum value" feature, led to the algorithm discarding the minimum-value feature at both ankle and wrist. The minimum may be less informative than the maximum acceleration, given that a low SM depends upon slight variations of the SM around 1 g. Such values can result from noise or from downward accelerations that compensate for the gravitational acceleration measured by the sensor. Such direction of movement at the ankle site, is necessarily followed by an impact, that is already observed by the "maximum value" feature.

Not all the frequency domain features proposed previously (34) and confirmed in our previous work (16) were actually selected by the automatic selection strategy. In particular, the first dominant frequency was retained for both ankle and wrist classifier, but the feature selector discarded the second dominant frequency and the dominant frequency in the band 0.6-2.6 Hz. The number of activity fragments within the window may capture the information missed by discarding those features: a large number of activations within the window should be associated with fast movements that involve frequent accelerations and decelerations; a low number of activations may be associated to slow or sporadic movement episodes in which the amplitude of the acceleration is smoother. At the wrist site, a newly introduced frequency domain feature, the power at frequency components higher than 3.5 Hz, was selected.

Activity classification using youth data

When using the new feature set on the newly available youth dataset, both ankle and wrist overall accuracy exceeded 90% (see table 4A and 4B). Updating the feature set resulted in a significant improvement on youth activity classification accuracy, especially for the *other activities* class. This result was expected after the introduction of the new features that are capable of extracting information about movement fragmentation, typical of recreational activities. Moreover, misclassifications are consistent with intuition. With the ankle classifier, for example, 24.5% of *basketball-passing* windows were misclassified as *sedentary*, which is consistent with someone temporarily keeping feet still doing this task (for a short 12.8 s window). Similarly, *cleaning room* and *tennis-ball:-throwing-catching*

were misclassified as *sedentary* from ankle data in 16.1 % and 20.2 % of cases, respectively – likely as a result of short bouts of no leg motion. Wii games were misclassified as *other activities* using ankle data in approximately the 10% of cases. Despite the sedentary nature of video-gaming, some Wii-games can actually be similar to sport activities, such as those included in *other activities*, if players engage in full-body motion when playing. When assigning activities to classes, we clustered the Wii games with sedentary video gaming in the "sedentary" class instead of grouping it with sport activities in the "other" class because Wii games can be played without moving feet, and even while sitting, and some children were observed doing so. At the ankle, misclassifications also occur between *ambulation* and *other activity*, *basketball dribbling* and *ambulation*, and *walking natural* and other categories. In all instances, the relatively small percentage of errors can be explained by the variability in the behaviors being studied.

Table 5 illustrates challenges with wrist-only activity recognition. Approximately one fourth of *walking natural* data were misclassified as *other activities,* whereas slow speed treadmill walking (2 mph) had the highest number of misclassifications in the *ambulation* class. Wrist movement during slow treadmill walking may not be significant. *Exercise bike pedaling* was classified as *sedentary* in almost all cases (88.9%). During this activity, the wrist was placed on the exercise bike handlebar and its movement could be negligible. Conversely, *outdoor cycling* was characterized by more significant wrist movement due to more variable wrist movement and vibrations that may result from real bicycle riding. Three of the youth participants were not comfortable with riding a bike and so they only used the exercise bike. As a consequence their cycling data were misclassified (see Table 5). Sport activities may involve ambulation and active bouts followed by quasi-sedentary periods. This resulted in higher error rates for *basketball passing*, *soccer dribbling* and *tennis ball fielding*. As with the ankle, *Wii* activities at the wrist were sometimes classified as *other activities* instead of *sedentary* as expected.

Cross-tests (experiment 3)

When training the SVM classifier on all available adult data and testing on all available youth data (or vice versa, see Table 5), *ambulation* and *sedentary* activities were generally correctly recognized, whereas more errors are found in the *other activities*. *Cycling* detection using wrist data was problematic if the model was trained on youth data and tested on adults. Both the ankle and wrist classifiers obtained overall recognition accuracies higher than 75% for *sedentary* and *ambulation* classes, even if the tested population was completely different in term of age of participants.

Merged dataset LOSO tests (experiment 4)

Merging both datasets in a single, larger dataset allowed us to use LOSO validation to verify the classifier independently from the age group (Table 4 part C and Table 6). When training data reflects testing data well, in this case including both examples from adults and youth, better performance would be expected, and we confirmed that here. For what concern the wrist classifier it was confirmed that the correct classification rate for cycling in adults was low in the merged dataset as well. This was because most of the adults cycling data were

acquired on an exercise bike. As stressed before, the wrist movement in exercise bike pedaling may be negligible.

Overall, the results confirm that the activity recognition method proposed previously (16) can be applied to recognize activity on youth data as well on adults. The LOSO validation results on the complete dataset (with both young and adults data) show that training classifiers with data from different age groups does not significantly reduce activity classification performances (Table 4C). Merging data from the two groups into a single dataset results in an accurate classifier that is not specific to one age group.

Comparison to literature

Prior work on activity classification using accelerometers in youth uses different activity sets, age ranges, number of participants, amount of processed data and experimental setups, and validation approaches (see Table 1), making direct comparison of results challenging. This study demonstrates a solution with overall accuracy higher than 90% for both wrist and ankle in a four-class activity problem, tested on structured and semi-structured activities. Unlike prior work, adults and children are considered together. Del Rosario *et al*. included activities from elderly and young adults, classified with 82.0% and 79.9% accuracy respectively, but from a smartphone in the trouser pocket (10). Only two previous studies focus on wrist activity classification in youth. Accuracy results obtained by Ruch *et al*., (23) who used wrist and hip sensors *simultaneously*, 67.0% on 10 classes. Activity counts were used instead of raw data to extract features. Trost *et al.* recently achieved wrist activity classification in youth with 5-class accuracy results similar to those presented here (88.4 ± 3.0% at wrist). In that study, a smaller number of activities were merged into five broad categories. LOSO cross-validation was not conducted; however, the authors applied a modified version of n-fold cross-validation to prevent data of the subject being tested from being included in the training set (29). Two previous studies that did not use LOSO cross validation show recognition accuracies larger than ours. In Hikihara *et al.* the classification was a 2-class problem, discriminating data between *locomotion* and *non-locomotion* using raw accelerometer readings from the waist, obtaining a 99.1% accuracy (13). Nam and Park report 98.4% accuracy using sensors including the accelerometer and barometer, but in this case the target population was significantly younger (1.3 to 2.4 yo), resulting in a very different activity vocabulary (19).

The dataset and Matlab code used in this study will be available to interested researchers (https://mhealth.ccs.neu.edu/data/).

Conclusions

In this work, good overall classification results were obtained using previously defined methods and features. However, given the different nature of activities that are much more frequent in youth than in adults (such as video-gaming or sports), the accuracy of the classifier may take advantage of the introduction of a different set of dedicated features. When validating methods with a LOSO approach starting from a merged A+Y dataset, it was confirmed that the accuracy can be improved by using data from both groups, even if the tested subject was not included in the training set. At the same time, the algorithm

preserves all the advantages of the previously proposed method in term of real-time implementation suitability and high comfort for the user given the single sensor wrist or ankle-worn proposed configuration.

In conclusion, as large surveillance studies include young participants in their evaluations with wrist worn monitors, it will be possible to use previously available methods, provided that data from young participants are included in the definition of classification rules. Important accuracy improvement would be obtained if the feature set also includes quantities that can describe the impulsive and fragmented nature of many youth activities.

Acknowledgments

References

1. Albinali, F.; Intille, SS.; Haskell, W.; Rosenberger, M. Using wearable activity type detection to improve physical activity energy expenditure estimation.. Proceedings of the Int'l Conf. on Ubiquitous Computing (UbiComp); Copenhagen, Denmark. 2010; p. 311-20.

2. Bao L, Intille SS. Activity recognition from user-annotated acceleration data. Proc. of Pervasive. 2004; 301:1–17.

3. Bhattacharya A, McCutcheon EP, Shvartz E, Greenleaf JE. Body acceleration distribution and O2 uptake in humans during running and jumping. J Appl Physiol. 1980; 49(5):881–7. [PubMed: 7429911]

4. Bouten CV, Koekkoek KT, Verduin M, Kodde R, Janssen JD. A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. IEEE Transactions on Biomedical Engineering. 1997; 44(3):136–47. [PubMed: 9216127]

5. Caspersen CJ, Pereira MA, Curran KM. Changes in physical activity patterns in the United States, by sex and cross-sectional age. Medicine & Science in Sports & Exercise. 2000; 32(9):1601–9.

6. Chen Y, Lin C. Combining SVMs with various feature selection strategies. Feature Extraction. 2006; 207(1):1–10.

7. Crouter SE, Flynn JI, Bassett DR Jr. Estimating physical activity in youth using a wrist accelerometer. Medicine & Science in Sports & Exercise. 2015; 47(5):944–51.

8. Crouter SE, Horton M, Bassett DR Jr. Use of a 2-regression model for estimating energy expenditure in children. Medicine & Science in Sports & Exercise. 2012; 44(6):1177.

9. De Vries SI, Engels M, Garre FG. Identification of children's activity type with accelerometer-based neural networks. Medicine & Science in Sports & Exercise. 2011; 43(10):1994–9.

10. Del Rosario MB, Wang K, Wang J, et al. A comparison of activity classification in younger and older cohorts using a smartphone. Physiological measurement. 2014; 35(11):2269. [PubMed: 25340659]

11. Esterman M, Tamber-Rosenau BJ, Chiu Y-C, Yantis S. Avoiding non-independence in fMRI data analysis: Leave one subject out. NeuroImage. 2010; 50(2):572–6. [PubMed: 20006712]

12. Hagenbuchner M, Cliff DP, Trost SG, Van Tuc N, Peoples GE. Prediction of activity type in preschool children using machine learning techniques. Journal of Science and Medicine in Sport. 2015; 18(4):426–31. [PubMed: 25088983]

13. Hikihara Y, Tanaka C, Oshima Y, Ohkawara K, Ishikawa-Takata K, Tanaka S. Prediction models discriminating between nonlocomotive and locomotive activities in children using a triaxial

accelerometer with a gravity-removal physical activity classification algorithm. PloS one. 2014; 9(4):e94940. [PubMed: 24755646]

14. Intille, SS.; Albinali, F.; Mota, S.; Kuris, B.; Botana, P.; Haskell, WL. Design of a wearable physical activity monitoring system using mobile phones and accelerometers.. Proceedings of the IEEE Engineering in Medicine and Biology Society Meeting (EMBC); Boston, Massachusetts USA. 2011; p. 3633-9.

15. Jain AK, Duin RPW, Mao J. Statistical pattern recognition: a review. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2000; 22(1):4–37.

16. Mannini A, Intille SS, Rosenberger M, Sabatini AM, Haskell W. Activity recognition using a single accelerometer placed at the wrist or ankle. Medicine & Science in Sports & Exercise. 2013; 45(11):2193–203.

17. Mannini A, Sabatini AM, Intille SS. Accelerometry-based recognition of the placement sites of a wearable sensor. Pervasive and Mobile Computing. 2015; 21:62–74. [PubMed: 26213528]

18. Morey-Klapsing G, Arampatzis A, Brüggemann GP. Choosing EMG parameters: comparison of different onset determination algorithms and EMG integrals in a joint stability study. Clinical Biomechanics. 2004; 19(2):196–201. [PubMed: 14967584]

19. Nam Y, Park JW. Child activity recognition based on cooperative fusion model of a triaxial accelerometer and a barometric pressure sensor. IEEE Journal of Biomedical and Health Informatics. 2013; 17(2):420–6. [PubMed: 24235114]

20. Rosenberger ME, Haskell WL, Albinali F, Mota S, Nawyn J, Intille S. Estimating activity and sedentary behavior from an accelerometer on the hip or wrist. Medicine and science in sports and exercise. 2013; 45(5):964. [PubMed: 23247702]

21. Rowlands AV. Accelerometer assessment of physical activity in children: an update. Pediatric Exercise Science. 2007; 19(3):252. [PubMed: 18019585]

22. Ruch N, Melzer K, Mäder U. Duration, frequency, and types of children's activities: potential of a classification procedure. Journal of Exercise Science & Fitness. 2013; 11(2):85–94.

23. Ruch N, Rumo M, Mäder U. Recognition of activities in children by two uniaxial accelerometers in free-living conditions. European journal of applied physiology. 2011; 111(8):1917–27. [PubMed: 21249388]

24. Siirtola, P.; Laurinen, P.; Haapalainen, E.; Roning, J.; Kinnunen, H. Clustering-based Activity Classification with a Wrist-worn Accelerometer Using Basic Features.. Proceedings of the IEEE Symp on Computational Intelligence and Data Mining; Nashville, TN.. 2009; p. 95-100.

25. Sun M, Hill J. A method for measuring mechanical work and work efficiency during human activities. Journal of Biomechanics. 1993; 26(3):229–41. [PubMed: 8468336]

26. Troiano, R.; Mc Clain, J. Objective measures of physical activity, sleep, and strength in U.S. National Health and Nutrition Examination Survey (NHANES) 2011-2014.. Proceedings of the 8th Internat Conf on Diet and Activity Methods; Roma, Italy. 2012; p. 24

27. Trost SG, Pate RR, Sallis JF, et al. Age and gender differences in objectively measured physical activity in youth. Medicine & Science in Sports & Exercise. 2002; 34(2):350–5.

28. Trost SG, Wong W-K, Pfeiffer KA, Zheng Y. Artificial neural networks to predict activity type and energy expenditure in youth. Medicine & Science in Sports & Exercise. 2012; 44(9):1801–9.

29. Trost SG, Zheng Y, Wong W-K. Machine learning for activity recognition: hip versus wrist data. Physiological measurement. 2014; 35(11):2183. [PubMed: 25340887]

30. UBC. UK Biobank Coordinating Centre. Category 2 enhanced phenotyping at baseline assessment visit in last 100-150,000 participants. Stockport Cheshire. 2009:1–33.

31. Vapnik, V. The nature of statistical learning theory. Springer-Verlag; New York: 2000. p. 1-314.

32. Veltink PH, Bussmann H, de Vries W, Martens W, Van Lummel RC. Detection of static and dynamic activities using uniaxial accelerometers. Rehabilitation Engineering, IEEE Transactions on. 1996; 4(4):375–85.

33. Weston J, Mukherjee S, Chapelle O, Pontil M, Poggio T, Vapnik V. Feature selection for SVMs. Proceedings of the NIPS. 2000:668–74.

34. Zhang S, Rowlands AV, Murray P, Hurst TL. Physical activity classification using the GENEA wrist-worn accelerometer. Medicine & Science in Sports & Exercise. 2012; 44(4):742–8.

Table 1

Studies involving automatic activity recognition in youth from accelerometers

	De Vries (9)	Ruch et al. (23)	Trost et al. (28)	Hikihara et al. (13)	Del Rosario et al. (10)	Nam and Park (19)	Trost et al. (29)	Hagenbuchner et al (12)	This study
Year	2011	2011	2012	2014	2014	2013	2014	2015	
Number of participants	58	41	100	68	37 elderly, E 20 young adults, YA	10	52	11	33 adults. A 20 youth, Y
Age of participants	9 to 12	10.8 ± 1.3 (mean ± std dev.)	5 to 15 (11.0 ± 2.7)	6 to 12	E: 83.9 ± 3.4 YA: 21.9 ± 1.7	toddlers: 1.3 to 2.4	13.7 ± 3.1	Pre-school: 4.8 ± 0.87	A: 18 to 75 Y: 11 to 15
Classes	7 classes: sitting, standing, walking, running, rope skipping, playing soccer, and cycling	10 classes: biking, crawling, walking, scooter, horseback riding, jumping and floor exercise	12 activities merged to 5 classes: sedentary, walking, running, light intensity household activities or games, moderate-to-vigorous intensity games or sports	11 activities merged to 2 classes: non-locomotive, locomotive	9 classes: stand, sit, lie, walk, walk upstairs, walk downstairs	11 classes: wiggling, rolling, standing still, standing up, sitting down, walking, toddling, crawling, climbing up / down and stopping	12 activities merged to 7 classes: lying down, sitting, standing, walking, running, basketball, dancing	12 activities merged to 5 classes: sedentary activities, light activities, moderate to vigorous activities, walking, and running	23 activities for youth and 26 for adults merged to 4 classes: sedentary, ambulation, cycling and others
Sensor site	ankle or hip	hip and wrist	hip	waist	smartphone in trousers pocket	waist (diaper)	wrist or hip	hip	ankle or wrist
Raw data vs activity counts	AC	AC	AC	custom counts from acc. data (32 Hz)	Raw acc. and gyr. (100 Hz), barometer (10 Hz)	Raw acc. and barometer (95 Hz)	Raw acc. (30 Hz)	AC	Raw acc. (90 Hz)
AC epoch	1 s	1 s	1 s	10 s	n.a.	n.a.	n.a.	1 s	n.a.
Windowing approach	10 s, non overlapping	1 s	10 s to 60 s tested	10 s	2.5 s 50% overlapping	2.7 s, 50% overlapping	10 s	10, 30 or 60 s, non overlapping	12.8 s non overlapping
Acquired data (minutes)	20 per subject (tot. 1160)	438 on average per participant	5 per activity acquired (tot. 6000)	3-7 per activity (tot. 2244-5236)	10-15 per subject for the young adults and 30 for elderly (tot. 1310-1410)	3000	5 per activity acquired (tot. 3120)	4-5 per activity per participant (tot. 528-660)	3-5 per activity: Yankle 1848 Ywrist 1809 A ankle2976 A wrist 2917

	De Vries (9)	Ruch et al. (23)	Trost et al. (28)	Hikihara et al. (13)	Del Rosario et al. (10)	Nam and Park (19)	Trost et al. (29)	Hagenbuchner et al (12)	This study
Classified data (minutes)	not disclosed	108	2 per activity used (tot. 2400)	not disclosed	YA: 189, E: 683	not disclosed	2 per activity used (tot. 1248)	264	Yankle 1257 Ywrist 1248 A ankle1609 A wrist 1633
Features	time domain	time domain	time domain	time domain	time domain	time and frequency domain	time domain	time domain	time and frequency domain
Classifier	ANN	kNN + DT + NDTF + voting	ANN	threshold	decision tree	NB, SVM, kNN, DT, ANN	logistic regression	ANN or deep learning network	SVM
Validation approach	LOSO	split in two subset with half participants each	subjects were split into three groups: one for training one for validation and one for testing. 10 random splits tested.	split in two subset (48 participants for training, 20 participants for testing)	LOSO	10-fold cross validation	subjects were split in three groups: one for training one for validation and one for testing. 10 random splits tested.	LOSO	LOSO
Accuracy	ankle: 68.0% hip: 77.0%	67.0 %	88.4 %	99,10%	E: 82,0% YA: 79.9%	Acc. only: 88.3% (DT) acc.+ barometer: 98.4% (SVM)	hip: 91.0 ± 3.1% wrist: 88.4 ± 3.0%	ANN: 69.7% DLN: 82.6%	Yankle: 92.4% Ywrist: 90.7% A ankle: 94.5% A wrist: 86.9%

Table 2

Activities in the adult and youth datasets, grouped into four broad classes.

	Ambulation	Cycling	Other activities	Sedentary
Adult dataset (A)	• walking. carrying-load • stairs: inside and down • stairs: inside and up • treadmill: 3 mph 0% incline • treadmill: 3 mph 6% incline • treadmill: 3 mph 9% incline • treadmill: 2 mph 0% incline • treadmill: 4 mph 0% incline • walking, natural	• 70rpm. 50W. 0.7kg • cycling outdoor level • cycling outdoor uphill • cycling outdoor downhill	• painting: roller • painting: brush • sweeping with broom	• sitting, internet search • sitting, computer typing • sitting: writing • sitting: reading • sorting files / paperwork • lying: on back • lying on left side • lying on-right-side • sitting: legs straight • standing still
Youth dataset (Y)	• walking, natural • treadmill walking: 2 • treadmill walking: 3 - 4 mph • treadmill running:4.5 - 5 mph	• 70 rpm 50 watts • Outdoor cycling	• basketball:-dribbling • basketball:-passing • basketball:-shortshots • clean room • soccer:-dribbling • soccer:-passing • tennis-ball:-fielding • tennis-ball:-throwing-catching	• sitting: reading • play-computer-game • play-on-gameboy • watch TV • wii:-boxing • wii:-tennis • lying: on back • sitting: legs straight • standing still

Table 3

Feature sets considered in this work (shaded are new relative to prior work (16)). Features indicated with check marks are those retained by the automatic selection strategy.

Features	Selected for wrist classifier	Selected for ankle classifier
Mean value	✓	✓
Standard deviation	✓	✗
Maximum	✓	✓
Minimum	✗	✗
P1 = Power of the first dominant frequency	✗	✗
P2 = Power of the second dominant frequency	✗	✗
F1 = First dominant frequency	✓	✓
F2 = Second dominant frequency	✗	✗
PT = Total power	✗	✗
P1 / PT = Ratio of power at dominant frequency and total power	✓	✗
Power of the first dominant freq in the band 0.6-2.6 Hz	✗	✗
First dominant frequency in the band 0.6-2.6 Hz	✗	✗
Ratio between F1 of the current and previous windows	✗	✓
Range = maximum – minimum	✓	✗
(Power at frequencies higher than 3.5 Hz) / PT	✓	✗
fragmentation, samples	✓	✓
fragmentation, number of activations	✓	✓
fragmentation, mean activation interval duration	✗	✗
fragmentation, activation interval duration variability	✗	✓

Table 4

Wrist and ankle classification confusion matrices for the four target activity groups using the SVM classifier with LOSO cross-validation. Correct classifications are in bold.

Part A, experiment 1, original feature set, Youth dataset

Wrist			Ambulation	Cycling	Other activities	Sedentary
		Ambulation	**646 (75.6%)**	26 (3%)	52 (6.1%)	131 (15.3%)
		Cycling	26 (2.1%)	**1067 (87.5%)**	36 (3%)	90 (7.4%)
	Actual label	Other activities	38 (3%)	38 (3%)	**1038 (81.7%)**	157 (12.4%)
		Sedentary	16 (0.6%)	44 (1.8%)	170 (6.8%)	**2277 (90.8%)**
		Overall accuracy 85.9 %				

Ankle			Ambulation	Cycling	Other activities	Sedentary
		Ambulation	**752 (88%)**	41 (4.8%)	60 (7%)	2 (0.2%)
		Cycling	24 (2%)	**1074 (88%)**	89 (7.3%)	33 (2.7%)
	Actual label	Other activities	50 (3.9%)	58 (4.6%)	**1025 (80.5%)**	140 (11%)
		Sedentary	1 (0%)	16 (0.6%)	94 (3.7%)	**2433 (95.6%)**
		Overall accuracy 89.7 %				

Part B, experiment 2, new feature set, Youth dataset

Wrist			Ambulation	Cycling	Other activities	Sedentary
		Ambulation	**674 (78.8%)**	39 (4.6%)	60 (7%)	82 (9.6%)
		Cycling	14 (1.1%)	**1104 (90.6%)**	21 (1.7%)	80 (6.6%)
	Actual label	Other activities	36 (2.8%)	11 (0.9%)	**1149 (90.4%)**	75 (5.9%)
		Sedentary	8 (0.3%)	35 (1.4%)	63 (2.5%)	**2401 (95.8%)**
		Overall accuracy 91.0 %				

Ankle			Ambulation	Cycling	Other activities	Sedentary
		Ambulation	**818 (95.7%)**	10 (1.2%)	27 (3.2%)	0 (0%)
		Cycling	7 (0.6%)	**1101 (90.2%)**	79 (6.5%)	33 (2.7%)
	Actual label	Other activities	35 (2.7%)	19 (1.5%)	**1084 (85.2%)**	135 (10.6%)
		Sedentary	0 (0%)	7 (0.3%)	99 (3.9%)	**2438 (95.8%)**
		Overall accuracy 92.4 %				

Part C, experiment 4, new feature set, Youth +Adult dataset

Wrist			Ambulation	Cycling	Other activities	Sedentary
		Ambulation	**2842 (86.4%)**	149 (4.5%)	170 (5.2%)	127 (3.9%)
		Cycling	92 (4.1%)	**1741 (77.3%)**	36 (1.6%)	383 (17%)
	Actual label	Other activities	85 (3.8%)	37 (1.7%)	**1943 (87.4%)**	159 (7.1%)
		Sedentary	15 (0.3%)	153 (2.8%)	124 (2.3%)	**5197 (94.7%)**
		Overall accuracy 88.5 %				

Part A, experiment 1, original feature set, Youth dataset

Wrist			Ambulation	Cycling	Other activities	Sedentary
Ankle			Ambulation	Cycling	Other activities	Sedentary
		Ambulation	**3278 (97.9%)**	31 (0.9%)	37 (1.1%)	2 (0.1%)
		Cycling	24 (1.1%)	**2026 (89.7%)**	111 (4.9%)	97 (4.3%)
	Actual label	Other activities	54 (2.4%)	29 (1.3%)	**1766 (78.3%)**	407 (18%)
		Sedentary	0 (0%)	16 (0.3%)	315 (5.7%)	**5211 (94%)**
			Overall accuracy 91.6%			

Table 5

Wrist and ankle classification details showing category recognition for each specific activity type. LOSO validation results using the new feature set on the youth dataset (experiment 2). Correct classifications are in bold.

Wrist,	Ambulation	Cycling	Other Activities	Sedentary
Ambulation Activities				
walking, natural	**71 (71%)**	3 (3%)	26 (26%)	0 (0%)
treadmill walking: 2.0	**249 (66%)**	34 (9%)	21 (5.6%)	73 (19.4%)
treadmill walking: 3.0 – 4.0 mph	**9 (56.3%)**	0 (0%)	1 (6.3%)	6 (37.5%)
treadmill running: 4.5 – 5.0 mph	**344 (95.3%)**	2 (0.6%)	12 (3.3%)	3 (0.8%)
Cycling Activities				
Exercise bike 70 rpm 50 watts	2 (3.2%)	**0 (0%)**	5 (7.9%)	56 (88.9%)
Outdoor cycling	12 (1%)	**1104 (95.5%)**	16 (1.4%)	24 (2.1%)
Other activities				
basketball:-dribbling	0 (0%)	0 (0%)	**122 (97.6%)**	3 (2.4%)
basketball:-passing	0 (0%)	0 (0%)	**100 (90.9%)**	10 (9.1%)
basketball:-shortshots	0 (0%)	0 (0%)	**126 (97.7%)**	3 (2.3%)
clean room	13 (3.3%)	5 (1.3%)	**357 (90.4%)**	20 (5.1%)
soccer:-dribbling	21 (17.8%)	3 (2.5%)	**86 (72.9%)**	8 (6.8%)
soccer:-passing	1 (0.8%)	3 (2.3%)	**117 (90.7%)**	8 (6.2%)
tennis-ball:-fielding	1 (0.7%)	0 (0%)	**117 (86.7%)**	17 (12.6%)
tennis-ball:-throwing-catching	0 (0%)	0 (0%)	**124 (95.4%)**	6 (4.6%)
Sedentary activities				
sitting: reading	1 (0.4%)	3 (1.1%)	0 (0%)	**275 (98.6%)**
play-computer-game	0 (0%)	5 (1.5%)	0 (0%)	**321 (98.5%)**
play-on-gameboy	0 (0%)	0 (0%)	0 (0%)	**320 (100%)**
watch TV	1 (5.9%)	0 (0%)	0 (0%)	**16 (94.1%)**
wii:-boxing	3 (0.7%)	10 (2.5%)	41 (10%)	**354 (86.8%)**
wii:-tennis	1 (0.2%)	4 (0.9%)	21 (4.9%)	**403 (93.9%)**
lying: on back	0 (0%)	3 (1.2%)	0 (0%)	**242 (98.8%)**
sitting: legs straight	1 (0.4%)	7 (2.9%)	0 (0%)	**233 (96.7%)**
standing still	1 (0.4%)	3 (1.2%)	1 (0.4%)	**237 (97.9%)**

Ankle,	Ambulation	Cycling	Other Activities	Sedentary
Ambulation Activities				
walking, natural	**86 (86%)**	0 (0%)	14 (14%)	0 (0%)
treadmill walking: 2.0	**367 (97.3%)**	7 (1.9%)	3 (0.8%)	0 (0%)
treadmill walking: 3.0 – 4.0 mph	**14 (87.5%)**	2 (12.5%)	0 (0%)	0 (0%)
treadmill running: 4.5 – 5.0 mph	**350 (97%)**	1 (0.3%)	10 (2.8%)	0 (0%)
Cycling Activities				
Exercise bike 70 rpm 50 watts	0 (0%)	**63 (100%)**	0 (0%)	0 (0%)

Ankle,	Ambulation	Cycling	Other Activities	Sedentary
Outdoor cycling	7 (0.6%)	**1038 (89.7%)**	79 (6.8%)	33 (2.9%)
Other activities				
basketball:-dribbling	26 (20.8%)	2 (1.6%)	**97 (77.6%)**	0 (0%)
basketball:-passing	0 (0%)	2 (1.8%)	**81 (73.6%)**	27 (24.5%)
basketball:-shortshots	0 (0%)	2 (1.6%)	**119 (92.2%)**	8 (6.2%)
clean room	0 (0%)	11 (2.8%)	**323 (81.2%)**	64 (16.1%)
soccer:-dribbling	9 (7.6%)	0 (0%)	**106 (89.8%)**	3 (2.5%)
soccer:-passing	0 (0%)	0 (0%)	**129 (100%)**	0 (0%)
tennis-ball:-fielding	0 (0%)	1 (0.7%)	**127 (94.1%)**	7 (5.2%)
tennis-ball:-throwing-catching	0 (0%)	1 (0.8%)	**102 (79.1%)**	26 (20.2%)
Sedentary activities				
sitting: reading	0 (0%)	2 (0.7%)	1 (0.3%)	**291 (99%)**
play-computer-game	0 (0%)	1 (0.3%)	4 (1.2%)	**323 (98.5%)**
play-on-gameboy	0 (0%)	0 (0%)	3 (0.9%)	**334 (99.1%)**
watch TV	0 (0%)	0 (0%)	0 (0%)	**17 (100%)**
wii:-boxing	0 (0%)	2 (0.5%)	42 (10.2%)	**366 (89.3%)**
wii:-tennis	0 (0%)	2 (0.5%)	46 (10.7%)	**381 (88.8%)**
lying: on back	0 (0%)	0 (0%)	0 (0%)	**250 (100%)**
sitting: legs straight	0 (0%)	0 (0%)	2 (0.8%)	**238 (99.2%)**
standing still	0 (0%)	0 (0%)	1 (0.4%)	**238 (99.6%)**

Table 6

Accuracy results are summarized for all the experiments conducted (Single group LOSO cross validation tests (CV), crossed test and LOSO CV on the merged dataset)

| Exp ID | Wrist activity recognition | Classification accuracy (%) | | | | |
		Ambulation	Cycling	Other	Sedentary	Overall
Mannini et al. (16)	LOSO CV on A (original feature set)	87.2	62.9	81.6	91.2	**84.7**
Exp 1	LOSO CV on Y (original feature set)	75.6	87.5	81.7	90.8	**85.9**
Exp 2	LOSO CV on A (new feature set)	88.9	66.4	83.5	93.6	**87.0**
	LOSO CV on Y (new feature set)	78.8	90.6	90.4	95.8	**91.0**
Exp 3	Crossed tests:					
	Training on A testing on Y (original feature set)	75.1	29.0	44.1	92.2	**58.8**
	Training on A testing on Y (new feature set)	66.6	59.2	44.1	94.0	**69.8**
	Training on Y testing on A (original feature set)	76.3	8.5	47.9	93.7	**71.7**
	Training on Y testing on A (new feature set)	79.4	12.4	30.4	93.0	**71.8**
Exp 4	LOSO CV on A+Y	86.4	77.3	87.4	94.7	**88.5**
	(Youth data contribution only)	*77.9*	*92.5*	*90.0*	*94.6*	***90.7***
	(Adult data contribution only)	*89.4*	*59.4*	*83.8*	*94.8*	***86.7***

Exp ID	Ankle activity recognition	Ambulation	Cycling	Other	Sedentary	Overall
(16)	LOSO CV on A (original feature set)	99.5	93.9	81.6	96.0	**95.0**
Exp 1	LOSO CV on Y (original feature set)	88.0	88.0	80.5	95.6	**89.7**
Exp 2	LOSO CV on A (new feature set)	99.0	91.6	81.1	97.1	**94.8**
	LOSO CV on Y (new feature set)	95.7	90.2	85.2	95.8	**92.4**
Exp 3	Crossed tests:					
	Training on A testing on Y (original feature set)	92.7	76.2	37.4	92.7	**79.3**
	Training on A testing on Y (new feature set)	93.7	85.5	47.2	93.8	**84.7**
	Training on Y testing on A (original feature set)	89.4	91.3	26.0	94.9	**87.4**
	Training on Y testing on A (new feature set)	94.1	89.9	28.2	85.2	**84.9**
Exp 4	LOSO CV on A+Y	97.9	89.7	78.3	94.0	**91.6**
	(Youth data contribution)	*96.8*	*89.3*	*88.3*	*88..8*	***90.0***
	(Adult data contribution)	*98.3*	*90.3*	*65.3*	*98.5*	***92.9***