Predicting human movement type based on multiple accelerometers using movelets

Bing He[1], Jiawei Bai[1], Annemarie Koster[2], Caserotti Paolo[3], Glynn Nancy[4]
Tamara B. Harris[5], Ciprian M. Crainiceanu[1]
[1]Bloomberg School of Public Health, The Johns Hopkins University, Maryland, U.S.
[2]University of Maastricht, Maastricht, Netherlands
[3]University of Southern Denmark, Odense, Denmark
[4]University of Pittsburgh, Pittsburgh, PA, U.S.
[5]National Institute on Aging, Bethesda, Maryland, U.S.

We introduce statistical methods for prediction of types of human movement based on three tri-axial accelerometers worn simultaneously at the hip, left, and right wrist. We compare the individual performance of the three accelerometers using movelets and propose a new prediction algorithm that integrates the information from all three accelerometers. The development is motivated by a study of 20 older subjects who were instructed to perform 15 different types of activities during in-laboratory sessions. The differences in the prediction performance for different activity types among the three accelerometers reveal subtle yet important insights into how the intrinsic physical features of human movements could be effectively utilized in prediction. The proposed integrative movelet method takes into account those findings to augment the prediction accuracy and improve our understanding of human movement measurements.

Keywords: Accelerometer; physical activity; signal processing; pattern recognition; time series

1 Introduction

The objective and detailed characterization of daily physical activity is crucial for research studies where physical activity is involved as an exposure or outcome of interest. However, self-report based instruments for physical activity require people's cognitive and mental input, and are frequently subject to error due to recall bias, selection bias,

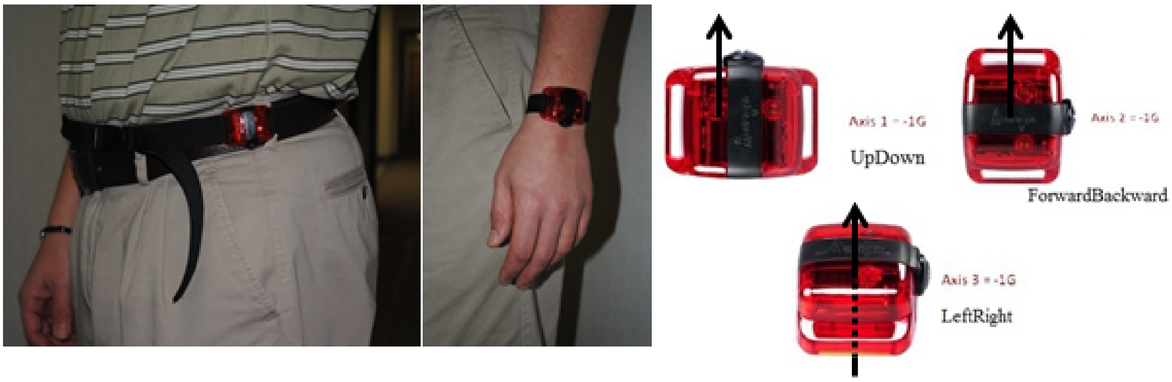

Figure 1: *Accelerometer placement on the body. The left and middle panels present how an accelerometer was attached to hip and wrist, respectively. Accelerometers have the same orientation. The right panel illustrates how the three axes of signals in the output are denoted.*

or potential cognitive decline and impairment in the target population. In the search for objective and refined measurements of physical activity, researchers have increasingly relied on accelerometers in observational studies and clinical trials (??????????). A tri-axial accelerometer is a wearable electromechanical sensor that records ultra-high density real-time dynamic accelerations in three mutually orthogonal directions. Accelerometers are relatively small in size and can be attached to different parts of the human body. A fundamental question is how to decipher and interpret the acceleration signals into meaningful information such as duration, intensity, and type of physical activities. Here we provide methods for predicting activity type based on single and multiple accelerometers worn at different parts of human body, including hip, right wrist, and left wrist.

Our methods are motivated by the Aging Research Evaluating Accelerometry (AREA) study, which is designed to investigate how well accelerometry data collected from the hip versus/and from the wrist reflect a given program of activities, including movements that emphasize either lower or upper body activities. The AREA study serves as a preceding study for evaluating the accelerometry data from NHANES assessment that is using a wrist accelerometer to increase compliance, as well as other epidemiological accelerometry studies to be conducted by the Laboratory of Epidemiology and Population Science group at the National Institute on Aging. In the study, 20 older participants were instructed

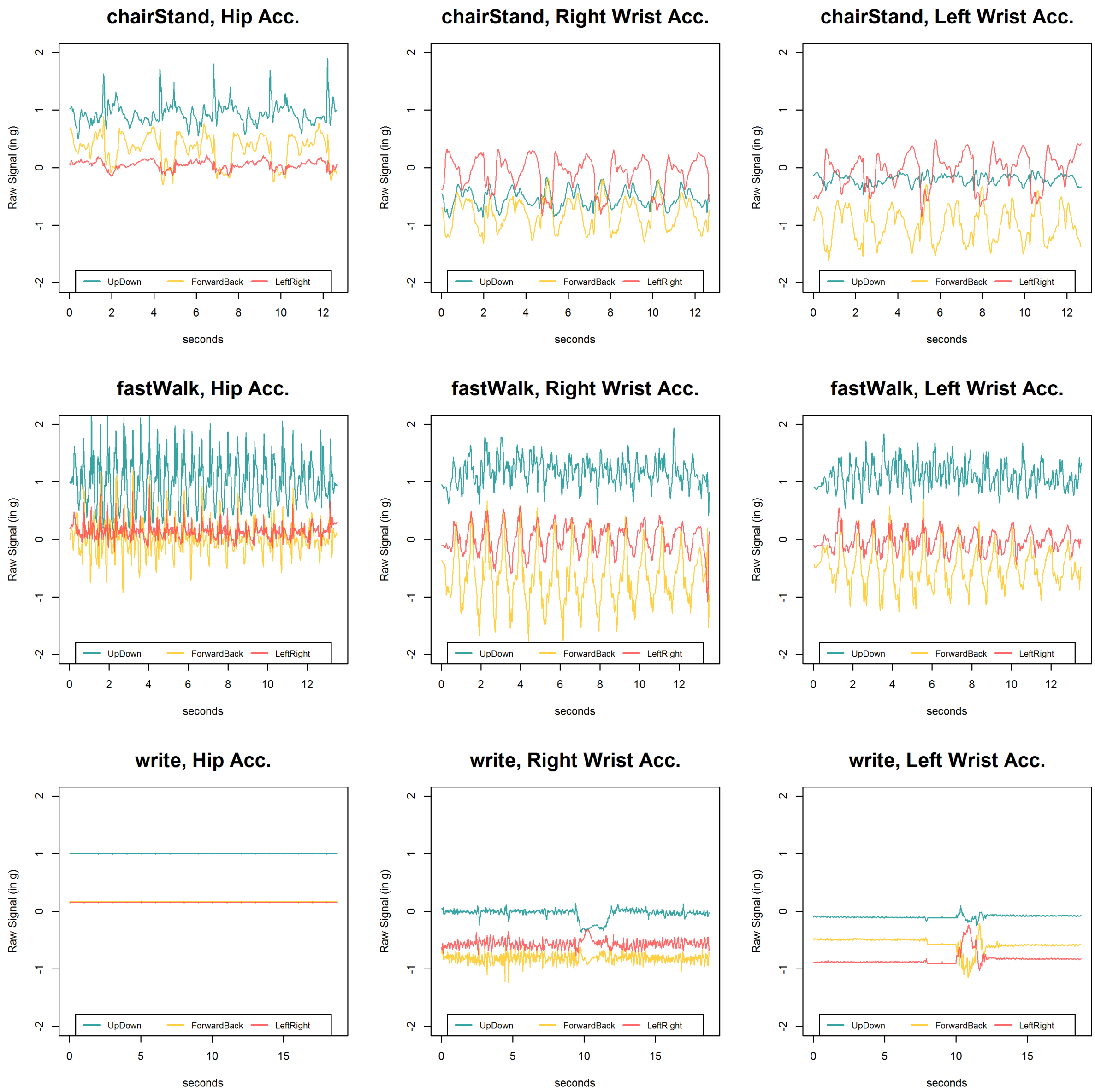

Figure 2: *The hip-worn, right wrist-worn, and left wrist-worn accelerometry data from Subject 1 for activities chairStand, fastWalk, and write*

Groups	Labels	Description	Duration
Resting	lying	lay still face-up on a flat surface with arms at sides and legs extended	10 mins
	stand	standing still with arms hanging at sides	3 mins
Upper body (while standing)	washDish	fetch wet plates from a drying rack, dry them using a trying towel, and stack adjacent to the drying rack one-by-one	3 mins
	knead	knead a ball of playdough as if for cooking/baking	3 mins
	dressing	unfold lab jacket, put jacket on (no buttoning), then remove, place the jacket on a hanger, and put the hanger on a nearby hook	3 mins
	foldTowel	fold towels and stack them nearby	3 mins
	vacuum	vacuum a specified area of the carpet	3 mins
	shop	walk along a long shelf, remove labeled items from the upper shelf about chest height, and place them on the lower shelf about waist height	3 mins
Upper body (while sitting)	write	write a specified sentence on one page of the notebook, then turn to the next page and repeat	3 mins
	dealCards	hold a full deck, and deal cards one-by-one to six positions around a table	3 mins
Lower body	chairStand	starting in a sitting position, rise to a normal standing position, then sit back down	5 cycles
	normalWalk Swing	starting from standing still, walk 20 meters at a comfortable pace	20 meters
	normalWalk NoSwing	starting from standing still, walk 20 meters at a comfortable pace with arms folded in front of chest	20 meters
	fastWalk Swing	starting from standing still, walk 20 meters at the fastest pace	20 meters
	fastWalk NoSwing	starting from standing still, walk 20 meters at the fastest pace with arms folded in front of chest	20 meters

Table 1: *The labels, detailed description, and durations of the 15 activity types.*

to perform 15 different types of activities sequentially according to a protocol during a 65-minute period, while wearing an Actigraph GT3X+ at the hip, right wrist, and left wrist, respectively. Table 1 provides the detailed description, duration, and labels for the 15 activities. The selection and design of these activities are intended to simulate a free-living context. In the rest of the paper, the activity types are referred to by their labels. For each accelerometer, the collected data contain a tri-axial time series of accelerations in the units of the standard gravity on earth, i.e., g. The sampling frequency of the Actigraph GT3X+ used is 80HZ. As presented in Figure 1, the accelerometers were attached to hip, right, and left wrist in a consistent orientation with respect to the human body in a standard standing up position, based on which, the three mutually orthogonal axes are denoted "UpDown", "ForwardBackward", and "LeftRight", respectively. Corresponding to the protocol and the start/end times for each activity, a time series of labels of activity types is constructed to annotate the accelerometry data for each subject.

Figure 2 displays the hip-worn, right wrist-worn, and left wrist-worn accelerometry data from Subject 1 for activities chairStand (top panels), fastWalk (middle panels), and write (bottom panels). In the top panels, Subject 1 repeats standing up and sitting back down 5 times, which is recognizable from the accelerometry signals. For example, it is easy to identify 5 spikes on the UpDown axis (shown in blue) from the hip-worn accelerometry data. The middle row displays fastWalk, indicating a strong periodic pattern in all three axes. For chairStand and fastWalk, the left and right wrist-worn accelerometers yield similar patterns, while the hip-worn accelerometer provides different patterns from the wrist-worn accelerometers for each activity. The bottom panels dedicated to writing indicate that the hip-worn signals remain largely unchanged around approximately 1 unit of gravity ($9.81m/s^2$) on the UpDown axis, as subjects were sitting while writing. The left wrist-worn accelerometer also indicates lack of movement, though the position of the hand is different from that of the hip, as illustrated by the different mean signals. The right wrist-worn accelerometer for write produces accelerations in a consistent pattern with substantially lower magnitude than those for fastWalk. We can expect that Subject

1 uses the dominant right hand to write, while placing the left hand on the table. The short segment of irregular signals from the 10th to 12th second likely corresponds to a 1.5 second interval when Subject 1 used the left hand to turn over one page (See description of write in Table 1). As it is shown in Figure 2, for accelerometers attached to the same parts of human body, the tri-axial signals differ in sign (i.e., direction), magnitude, and variability among different activities; whereas for the same activity, accelerometers attached to different parts of human body reveal different aspects of the movements.

Here we introduce methods for predicting activity type and provide answers to the following questions: 1) how well do accelerometry data reflect a given program of activities; and 2) given available accelerometry data from hip, right, and left wrist, how could we effectively integrate and take advantage of the combined information? The intrinsic features of accelerometry data present a range of challenges for answering these questions. First, the relatively high sampling frequency of accelerometers produces ultra-dense and massive amount of acceleration data ($80 \times 9 = 720$ observations per second with around $720 \times 65 \times 60 = 2,808,000$ observations per subject during a 65-minute period). Second, the output of accelerometers are tri-axial time series, which increases the dimension of the activity prediction problem. Third, when more accelerometry data from multiple parts of the human body are available, they also contain additional sources of variability that may not necessarily help activity prediction. Meanwhile, human activities consists of different movements of different parts of human body. Last but not least, novel graphic tools are needed to visualize the accelerometry data and the prediction process.

The intuition behind our methods for predicting activity type is that the movements with similar accelerometry patterns are likely to correspond to the same type of activity. The movelet approach proposed in (**?**) developed for one tri-axial accelerometer provides insight into what exactly leads to differences in prediction. A movelet is the entire time series collected in a window of given length, say 1 second. The sets of movelets constructed from the accelerometry data with annotated labels are organized by activity types, i.e., "chapters", which play the role of accelerometry dictionaries for different activity types.

Predictions of accelerometry data without annotated labels are provided through identifying the chapter that is most similar to the data in terms of mean squared error. This can be extended to multiple accelerometers in at least two ways: 1) by building separate movelet dictionaries and then combining predictions using voting or sequential decisions; or 2) by designing a joint dictionary, where a movelet is now a collection of nine time series (3 for each accelerometer).

In the fields of electronic engineering and computer science, researchers have developed various methods for recognition of physical activity type, but less intense methodological development has been seen in Biostatistics, especially in the context of epidemiological studies in public health. Many machine learning techniques have been developed for activity recognition, including linear/quadratic discriminant analysis (**??**), hidden Markov Chain (**??**), artificial neural networks (**?**), support vector machines (**?**) and combined methods (**??**). (**?**) and (**?**) reviewed and evaluated methods used in classification of normal activities and identifying falls from accelerometry data. However, these prediction approaches were usually developed and evaluated based on accelerometry data from subjects and activities that are of marginal interest in the settings of public health. Additionally, the prediction process tends to be very difficult computationally and hard to interpret, which reduces its appeal in realistic scenarios occurring in observational studies. Some studies did take advantage of multiple accelerometers to predict the activity type using machine learning techniques and to compare the performance of different accelerometers (**???**). However, lack of transparency of these prediction algorithms provides little insight into what exactly leads to differences in prediction. Meanwhile, subtle, intrinsic characteristics of human movements have not been incorporated in these prediction algorithms. Here we focus on accurate, fast methods that are easy to understand and mimic the natural human pattern recognition.

The remainder of the paper is organized as follows. In section 2, we describe the single movelets approach and compare prediction performance among hip, left wrist, and right wrist-worn accelerometers. Section 3 proposes movelets integration approaches and presents the prediction results. We conclude with Section 4.

2 Single-accelerometer movelets

We first review the movelets approach developed for predicting activity type using one tri-axial accelerometer in (?). The basic idea is to decompose the tri-axial time series into overlapping short-segment pieces of data, i.e., the movelets. The prediction of the activity type of an observation is based upon the similarity (distance) between the movelets that contain the observation and the ones that are labeled. The similarity between two movelets is measured by the mean square error. The entire process is similar to having a dictionary of words (movelets) with their associated meaning (labels). Given a new word (unlabeled movelet), the procedure simply requires looking up the word with closest meaning in the dictionary (labeled movelet) and assigning the corresponding label to the new word (unlabeled movelet). The idea is simple and easy to explain, though notations can be quite involved, albeit necessary. Below we provide these notations and the detailed description of the approach.

2.1 Single-accelerometer movelets definition

Denote the tri-axial accelerometry time series by $\mathcal{X}_i^p(t) = \{X_{i1}(t), X_{i2}(t), X_{i3}(t)\}$, where $t = 1, 2, \ldots, T_i$ and T_i is the length of time series, $i = 1, 2, \ldots, I$ and I is the total number of subjects, $p = 1, 2, 3$, representing hip, right wrist, and left wrist-worn accelerometers, respectively. Define the labeling function $L_i^p(t)$ that maps time $t \in \{1, 2, \ldots, T_i\}$ to $a \in \mathcal{A} = \{Act_1, Act_2, \ldots, Act_A\}$ based on data from the pth accelerometer, where Act_a designates the label of activity a. Let $\mathbb{U}_i$ and $\mathbb{W}_i$ be a partition of observation time into training data and testing data for the ith subject. $L_i^p(t)$ is known for labeled movelets in the training data and needs to be estimated for movelets in the testing data. We can now define a movelet as the basic unit of analysis. Let $M_i^p(t) = \{\mathcal{X}_i^p(t), \mathcal{X}_i^p(t+1), \ldots, \mathcal{X}_i^p(t+H-1)\}$ define the movelet starting at t with length H. Note that movelets overlap; for example, the observation $\mathcal{X}_i^p(t+H-1)$ belongs to $M_i^p(t)$, $M_i^p(t+1)$, ..., $M_i^p(t+H-2)$ and $M_i^p(t+H-1)$. We use overlapping movelets, as we are interested in predicting movements without knowing precisely what part of a movement is captured by a particular movelet. Movelets $M_i^p(t)$ in the training data are labeled with known activity types and grouped into activity-specific chapters $\mathcal{C}_i^p(a), a \in \mathcal{A}$. More specifically, $\mathcal{C}_i^p(a)$ is defined as $\mathcal{C}_i^p(a) = \{M_i^p(t), t \in \mathbb{U}_i : L_i^p(t) = Act_a\}$. The set of chapters forms a subject-specific dictionary of activities.

Given an unlabeled movelet, we find its closest match among labeled movelets in the dictionary. The chapter title to which the closest-matched movelet belongs is used to predict the unknown label. The intuition behind the approach is that movelets with similar visual appearance are likely to represent a similar activity. Formally, given an unlabeled movelet $M_i^p(t_0)$, we can identify a movelet $M_i^p(t^*), t^* \in \mathbb{U}_i$ as its closet match in the dictionary that maximizes the similarity, or minimizes the distance. More specifically, $t^* = \text{argmin}_{t \in \mathbb{U}_i} \{D[M_i^p(t), M_i^p(t_0)]\}$ and the distance function $D(.,.)$ is defined as $D[M_i^p(s), M_i^p(t)] = \frac{1}{3}\sum_{d=1}^{3}\sqrt{\sum_{h=1}^{H}[X_{id}(s-1+h) - X_{id}(t-1+h)]^2}$. Denote the activity label for $M_i^p(t^*)$ by a^* and thus $M_i^p(t^*) \in \mathcal{C}_i^p(a^*)$. For each observation in the movelet $M_i^p(t_0)$, i.e., $\{\mathcal{X}_i^p(t), \mathcal{X}_i^p(t+1), \ldots, \mathcal{X}_i^p(t+H-1)\}$, the activity type a^* gets a vote. Given an observation $\mathcal{X}_i^p(t_1), t_1 \in \mathbb{W}_i$, it is contained in H movelets, i.e., $M_i^p(t_1), M_i^p(t_1+1), \ldots, M_i^p(t_1+H-1)$. Then we say each of those movelets would generate a vote for a certain activity type, referred to as $\left\{a^*_{M_i^p(t_1)}, a^*_{M_i^p(t_1+1)}, \ldots, a^*_{M_i^p(t_1+H-1)}\right\}$, for this observation. Define the number of votes for a certain activity a for the observation $\mathcal{X}_i^p(t_1)$ by $V_i^p(a, t_1) = \sum_{h=t_1}^{t_1+H-1} \mathbb{1}\left(a^*_{M_i^p(h)} = a\right), a \in \mathcal{A}$. The activity type that gets the largest number of votes is adopted as the final prediction for $\mathcal{X}_i^p(t_1)$. Equivalently, define $L_i^p(t_1) = \text{argmax}_{a \in \mathcal{A}} V_i^p(a, t_1)$.

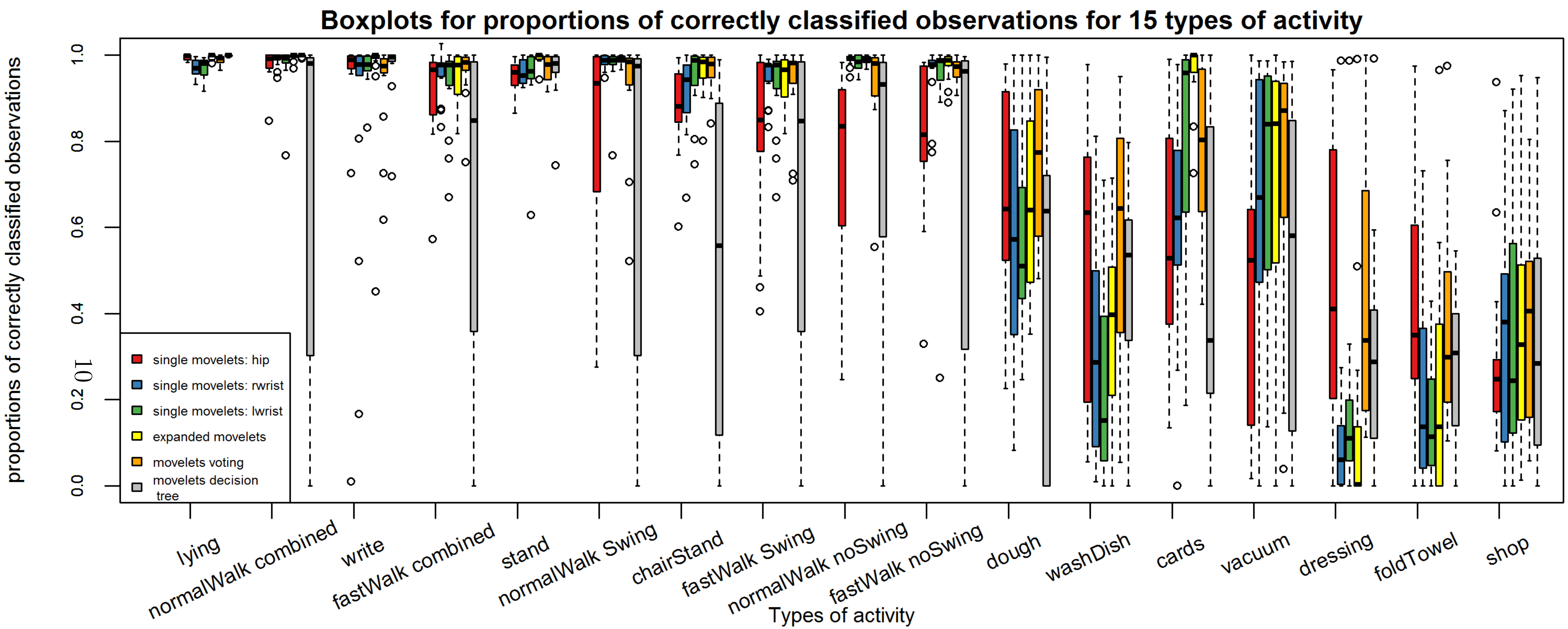

Figure 3: *Boxplot for prediction accuracy (i.e., proportions of correctly classified observations) for 15 types of activity using hip-worn accelerometer (red), right wrist-worn accelerometers (blue), and left wrist-worn accelerometers (green), and using integrative approaches expanded movelets (yellow), movelets voting (orange), and movelets decision tree (grey). normalWalk combined is the combined category of normalWalk Swing and normalWalk noSwing; and fastWalk combined is the combined category of fastWalk Swing and normalWalk noSwing. The activity labels on the X axis are ordered decreasingly by the median prediction accuracy of hip-worn accelerometers.*

2.2 Single-accelerometer movelets results

We now apply the single movelets approach to the AREA study. In this approach each accelerometer is treated independently of the other two accelerometers. As is described in Section 1, data from hip, right, left wrist-worn accelerometers were collected from 20 subjects during 65-minute in-laboratory sessions. The subjects were instructed to perform 15 activities according to a protocol (see Table 1). The protocol is used to construct a time series of activity labels accompanying the accelerometry time series. Four subjects are excluded because they did not perform all 15 activities. The breaks between two successive activities when subjects rested for around 3 minutes are removed from the data; the transitional periods at the beginning and end of each activity where subjects were transitioning from one activity to another are also dropped. In the following analysis, data for a half-a-minute period for each activity of each subject are used as the testing data. The length of the movelet H is taken to be 80, i.e., a 1-second window. A dictionary of 15 chapters of activities is created for each accelerometer of each subject. For activities with explicit beginning and end, such as chairStand, one replicate is used as training data. For movements with periodic features, such as normalWalk and fastWalk, a 2-second segment is utilized as training data. For other movements, a segment of length 2.5 seconds is adopted as training data.

The boxplot of the prediction accuracy, i.e., the proportions of correctly classified observations, for 15 activity types of 16 subjects using hip-worn accelerometer (red), right wrist-worn accelerometers (blue), and left wrist-worn accelerometers (green) is shown in Figure 3. On the x-axis we indicate the various activities and on the y-axis we display the proportion of correctly classified observations for that particular activity for all subjects. Because of the variability in the prediction accuracy across individuals, we present the boxplots of those proportions of correctly classified observations for all subjects. For each type of prediction approach, we plot a box in different color: 1) red for movelets based on single accelerometer at hip; 2) blue for movelets based on single accelerometer at right wrist; 3) green for movelets based on single accelerometer at left wrist; 4) yellow for movelets based on integrating the three accelerometers by expanding the movelets from 3 to 9 time series; and 5) orange for movelets based on integrating the three accelerometers by allowing the single movelet approaches to vote for the type of movement and then by accepting the majority vote. We have created also a new activity type, normalWalk combined, through combining normalWalk Swing and normalWalk noSwing into one category; fastWalk combined is created in a similar way, combining fastWalk Swing and fastWalk noSwing. The activity labels on the X axis are ordered decreasingly by the median prediction accuracy of hip-worn accelerometers (red). Table 2 presents the median prediction accuracy for each type of activities across subjects.

For resting activities, i.e., lying and standing, all accelerometers provide accurate predictions. This is reasonable, as visual data inspection reveals obvious differences between these signals, even though they both have very low variability. The main difference between the accelerometry signals for lying and standing is that the local average of individual time series are different in magnitude and rank. This happens because accelerometers have different angles with the vertical direction of gravity in the two different positions lying and standing. For example, while standing still, the gravity would appear as acceleration signals mainly on the UpDown axis (simply because 1 earth gravity will be added to any acceleration in the UpDown direction). Consequently, gravity affects differently each accelerometer axis and the size of the effect depends fundamentally on the angles the axes of the accelerometer form with the UpDown direction. Thus, the relative magnitude of local average accelerations is a proxy for the angles of hip, right and left wrist with the direction of gravity. While far from being a perfect proxy for position, this is enough to differentiate between standing and lying. This is a case where the variability of time series along their long term averages is of secondary importance, while the discrimination between the two resting positions is done by the shift in the relative magnitude of the mean functions.

Now we consider activities that mainly involving lower limbs. An arresting finding

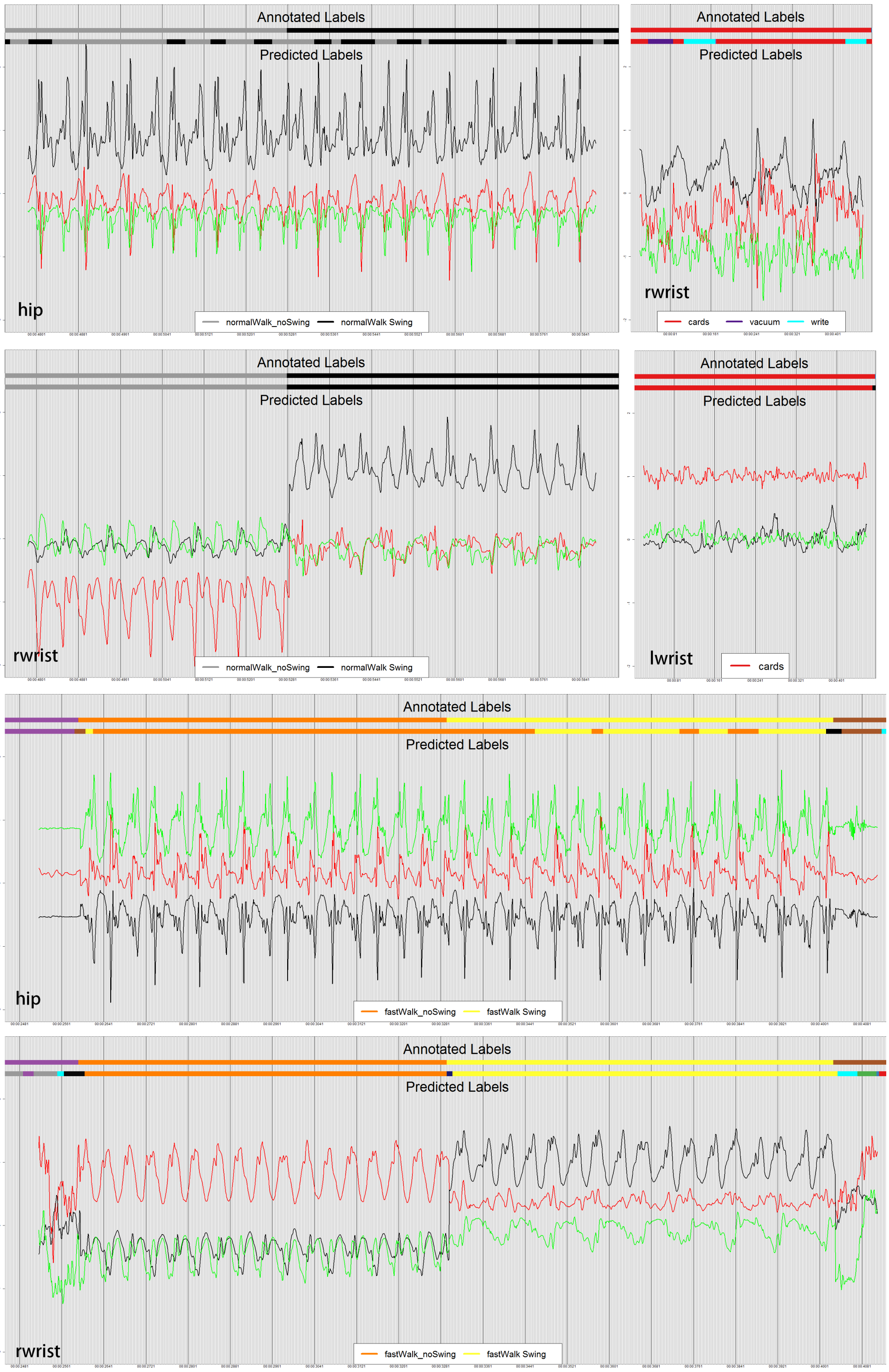

Figure 4: *Examples of prediction results using single movelets approach. Time series of raw signals on three axes are plotted, accompanied by annotated labels and predicted labels. Activities are color-coded. The examples are: 1) wrist-worn accelerometers can better distinguish between normalWalk Swing and normalWalk noSwing (the left panels in 1st and 2nd rows) and between fastWalk Swing and fastWalk noSwing (the panels in 3rd and 4th rows); and 2) left wrist-worn accelerometers can predict dealCards with higher accuracy than right wrist-worn accelerometers (the right panels in 1st and 2nd rows).*

	single movelets: hip %	single movelets: right wrist %	single movelets: left wrist %	expanded movelets %	movelets voting %	movelets decision tree %
lying	99.87	97.05	98.05	100.00	99.33	100.00
normalWalk combined	99.15	99.36	99.35	100.00	100.00	98.13
write	98.91	97.91	97.81	100.00	97.52	99.58
fastWalk combined	96.61	96.05	97.76	97.76	98.40	84.80
stand	96.09	95.31	96.25	100.00	97.81	98.13
normalWalk Swing	93.50	98.87	98.86	99.19	98.45	97.55
chairStand	88.09	94.31	98.81	98.42	98.04	55.76
fastWalk Swing	84.97	95.87	97.68	96.70	98.22	84.75
normalWalk noSwing	83.52	99.22	98.45	99.13	98.16	93.19
fastWalk noSwing	81.56	97.85	98.73	98.89	97.32	96.30
knead	64.28	57.22	51.09	64.06	77.44	63.75
washDish	63.44	28.71	15.20	39.70	64.45	53.61
cards	52.81	62.21	95.94	100.00	80.31	33.75
vacuum	52.32	67.02	84.05	84.08	87.19	58.03
dressing	41.07	6.00	11.00	0.27	33.78	28.77
foldTowel	34.93	13.66	11.35	13.66	29.92	30.83
shop	24.72	37.97	24.43	32.80	40.54	28.40

Table 2: *Median prediction accuracy for different types of activities using single movelets approach based on hip, right wrist, and left wrist-worn accelerometers, and using expanded movelets, movelets voting, and movelets decision tree. Activities are ordered decreasingly by the median prediction accuracy of hip-worn accelerometers.*

is that the right wrist-worn and left wrist-worn accelerometers outperform hip-worn accelerometers in predicting normalWalk Swing, normalWalk noSwing, fastWalk Swing, fastWalk noSwing, and chairStand. This is unexpected and counter intuitive: why would wrist worn accelerometers predict walking and chair standing better while these movement are fundamentally performed by lower body acceleration? This requires more in-depth analysis. Figure 4 shows three representative examples of prediction results for normalWalk, fastWalk, and dealCards. In each panel, labels of activities are coded in different colors; the annotated labels and predicted labels are plotted in parallel accompanying the original signals. Given a time point, if the annotated label and the predicted labels are of the same color, then the activity type is correctly predicted.

First consider accelerometry data for walking with and without arm swing as measured by the hip (first row, left panel) and right wrist (second row, left panel) in Figure 4. It is interesting that, as the person transitions from normal walk no arm swing to normal walk with arm swing, the time series associated with hip movement does not seem to display visually observable changes. In contrast, the wrist accelerometry indicates a strong change. Most interestingly, the black accelerometry curve shifts to a much higher level than before, probably because of the change in the angle of the accelerometer. Such strong angle changes can be easily observed and detected using movelets and should explain how information is being combined. The single movelets approach is confused between the two types of movements when using the hip data only (see the predicted labels alternating between gray and black, left panel in 1st row). In contrast, the predicted labels based on wrist data is quite accurate for detecting differences between these two types of movements. A very similar story holds for fast walking with or without arm swing (row 3 and 4 panels). These findings have potential implication on accelerometers' placement decisions in epidemiological studies. For example, consider a scenario when investigators decide to use hip-worn accelerometers and are interested in distinguishing between different types of walking. It is clear that it will be quite difficult to differentiate between walking normally and walking carrying a small object (no arm swing). Thus, in such situations it seems reasonable to simply define a label called normal walking that includes both arm swing and no arm swing. Alternatively, a wrist worn accelerometer could be used instead or in addition to the hip worn one.

Write and dealCards belong to the group of upper body activities while sitting. It is somewhat surprising that accelerometers worn at three different positions yield very accurate predictions for writing (see the blue red, blue and green boxplots in Figure 3 corresponding to write). Writing is a subtle movement that mainly involves hands compared to other upper limb activities, such as foldTowel. While this may be viewed as a potential problem it actually helps discriminate between writing and other, more intense, upper limb movements. The most counter intuitive finding was that hip accelerometry also distinguishes writing from any other activity. This happens because people sit down when writing and the posture of sitting helps distinguish writing from other upper limb activities. It would probably be very hard to distinguish between sitting and writing versus just sitting using hip accelerometry alone.

For dealCards, left wrist-worn accelerometers provide higher median prediction accuracy than right wrist-worn accelerometers and both outperform the hip-worn accelerometers. The reason is that Hip-worn accelerometers cannot acquire the subtle movements of hands, and often incorrectly classified dealCards as knead (5.4% averaged across subjects) and foldTowel (5.3% average across subjects) etc. Right wrist-worn accelerometers falsely predict on average 10.20% of observations of dealCards to be the activity write, which is not completely surprising. The right panels in 1st and 2nd rows of Figure 4 shed light on the difference of prediction performance between left and right wrist-worn accelerometers. The right wrist-worn accelerometer yields more variable signals with larger magnitude than the left-wrist accelerometer.

For activities knead, washDish, vacuum, dressing, foldTowel, and shop, all the three accelerometers show lower median prediction accuracy and larger variability across subjects, compared to the other activities. This probably happens because of the high level of overlap in movement and ambiguity of some sub-movements across labeled activities. The median prediction accuracy of each individual accelerometer for foldTowel, shop, and dressing is below 50%. These six activities belong to the group of upper limb activities while standing, are more complex, and require a series of distinct sub-movements (see Table 1). Thus, selecting a short segment of training data that can characterize the main features of the activities becomes a difficult task. Meanwhile, some of these activities contain similar sub-movements; for example, in foldTowel, subjects are required to stack the folded towels, and in washDish, subjects are also required to stack dried plates. This also adds to the difficulty of distinguishing between those activities. On average, 12.80%, 13.05%, and 9.40% of dressing is falsely classified as foldTowel based on hip-worn, right wrist-worn, and left wrist-worn accelerometers, respectively.

In summary, both accelerometers worn at the dominant hand and non-dominant hand can capture lying, standing, normal walking, and fast walking as well as the hip-worn accelerometers. The order of magnitudes of accelerations among the three axes is very useful in detecting and differentiating various postures for all accelerometers. Handedness may cause left and right wrist-worn accelerometers to record very different acceleration signals, which may effect prediction performance. For activities like dealCards where the dominant hand moves much more intensely than the non-dominant hand, the accelerometers worn at the non-dominant hand yield higher prediction accuracy. Truncal movements like walking are well recognized by all of the accelerometers worn at hip, right wrist, and left wrist, while fine movements that occur at the distal extremities of human body are predominantly captured by wrist-worn accelerometers. Thus, it is reasonable to expect that integrating information from accelerometers worn at different positions would help to better predict different activities, and thus leads to higher prediction accuracy. In the next section, we propose several integrative movelets approaches.

3 Movelets integration

So far, we have applied and analyzed only the single-accelerometer movelets approach. However, we have already shown that accelerometers worn at different positions contain nontrivial complementary information about movement type. Here we propose simple approaches for combining this information. Combining classifiers to improve prediction is a intensely-studied topic in statistics and machine learning. In statistics, Breiman and Wolpert discussed model stacking and averaging (**???**). Some good reviews on different methods of combining classifiers in machine learning include (**?????**). We emphasize that all these methods are developed for the case when a large space of predictors, X, is used to predict an outcome, Y. Typically, these methods combine black box algorithms that are known to work well in many applications. Our case is different, as we have three different predictor spaces, X_1, X_2, and X_3 (i.e., accelerometers worn at hip, right and left wrist, respectively), for each of which we have a good prediction approach: single-accelerometer movelets. Meanwhile, we are fundamentally interested in preserving the interpretation of the movelets approach, which allows fast exploration and visualization of how they work and provide important insights into when they do not. Using black box prediction algorithms when learning how to predict complex human activities would be akin to starting to learn a new language by covering ones' ears. Thus, we propose three easy to understand and scale up integrative approaches: *movelets voting*, *movelets decision tree* and *expanded movelets*. In movelets voting, among the three predictions proposed by three accelerometers using single movelets approach, the prediction with the largest number of votes is adopted as the combined prediction. Movelets decision tree builds up a simple hierarchy of decisions based on movelets. The hip-worn accelerometer first discriminate top-level groups of activities, followed by the low-level prediction using wrist-worn accelerometers for specific activity types. The expanded movelets expands the dimension of a movelet from three to nine, which incorporates the tri-axial time series from all the three accelerometers. Now we provide the formal definition of these three integrative movelet approaches.

3.1 Movelets integration definition

Following the notation from Section 2.1, for movelets voting, define $L_i^{MV}(t) = L_i^{p^*}(t)$, where $p^* = \text{argmax}_p V_i^p(a,t)$ and represents the accelerometer that delivers the prediction with the largest number of votes across activities. In movelets decision tree, the top-level groups of activities are defined as $\tilde{\mathcal{A}} = \{\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3\}$, where $\mathcal{G}_1 = \{\text{lying}\}$, $\mathcal{G}_2 = \{\text{dealCards, write}\}$, and $\mathcal{G}_3 = \{$stand, washDish, kneading, dressing, foldTowel, vacuum, shop, chairStand, normalWalk swing, normalWalk noSwing, fastWalk swing, fastWalk noSwing$\}$. The three top-level groups of activities correspond to lying, activity while sitting, and activity while standing (see Table 1). We use hip-worn accelerometers in the top level and define the top-level chapters as $\tilde{\mathcal{C}}_i^1(g) = \{M_i^1(t), t \in \tilde{\mathbb{U}}_i : L_i^1(t) \in \mathcal{G}_g\}$, where $g = 1, 2, 3$ and $\tilde{\mathbb{U}}_i$ denote the top-level training data. The prediction results from the top-level classifier reduces the number of candidate activity types for each observation; for example, the observation predicted as resting by the top-level classifier will not be assigned any activity types other than lying in the low level. Within each of the three top-level groups, we use the prediction of the single-accelerometer movelets approach with left wrist-worn accelerometers as our final prediction. For the expanded movelets, define $\mathcal{M}_i(t) = \{M_i^1(t), M_i^2(t), M_i^3(t)\}^T$. Now movelets in testing data and training data are all extended to nine dimensions. We define a variant of the distance function as $\tilde{D}\left[\mathcal{M}_i(s), \mathcal{M}_i(t)\right] = \sum_{p=1}^3 \{\omega_p \cdot D[M_i^p(s), M_i^p(t)]\}$, where $\omega_p \geqslant 0, p = 1, 2, 3$ and $\sum_{p=1}^3 \omega_p = 1$. ω_p represents the level of importance we assign to the information captured by the pth accelerometer. If all the weights are assigned to a single accelerometer, the expanded movelets method reduces to the single movelet approach.

3.2 Movelets integration results and discussion

Figure 3 presents the prediction accuracies for the expanded movelets approach (yellow), movelets voting (orange), and movelets decision tree (grey). Table 2 provides the median prediction accuracy for different activities. Equal weights across hip-worn, right wrist-worn, and left wrist-worn accelerometers are imposed in the expanded movelets approach. For the activities lying, normalWalk combined, write, stand, normalWalk Swing, fastWalk Swing, and dealCards, the expanded movelets approach yields the highest prediction accuracy with least variability across subjects among all the approaches. A substantial increase in prediction accuracy for dealCards is observed for the expanded movelets approach. In Figure 5, for both subjects, the expanded movelets approach provides better prediction for dealCards than all the single movelets approach on its own. The movelets voting approach is inferior to either one of the single movelets approaches or the expanded

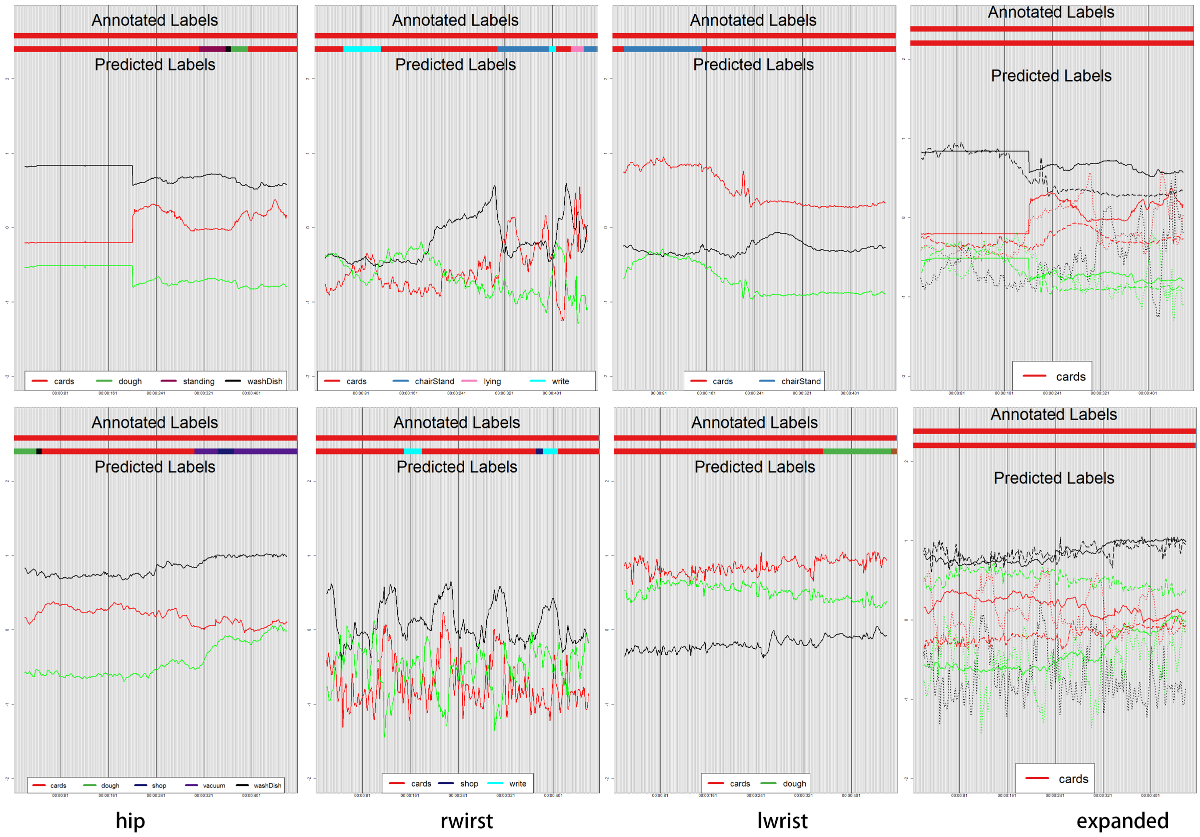

Figure 5: *Examples of prediction results using expanded movelets approach. Each row of panels corresponds to one subjects. The four columns display the prediction results for deal-Cards using the single hip-worn accelerometer, the single right wrist-worn accelerometer, the single left wrist-worn accelerometer, and the expanded movelets approach.*

movelets approaches for all the activities with exceptions for knead, washDish, vaccum, and shop. The movelets decision tree gives comparably good performance for lying, write, and stand; whereas for the rest of the activities, it tends to present lower median accuracy with larger variability across subjects than other approaches. The main reason is that, in movelets decision tree approach, the prediction error in the first level fully propagates into the second level. The left panel of Figure 6 displays the prediction accuracy for the top level of the movelets decision tree. It is not surprising that the top-level group lying presents very high prediction accuracy. However, there exists much variation in the prediction accuracy across subjects for standing and sitting. The median prediction accuracy for standing is higher than that for sitting. The right panel in Figure 6 shows an example of the top level prediction of movelets decision tree. The subject performed the activities standing still, vacuum while standing, and write while sitting in sequence. Standing still and vacuum while standing belong to the top-level standing group, and write while sitting belongs to the top-level sitting group. As it is shown in the right panel, standing still and vacuum while standing are mostly recognized by the movelets decision tree. In contract, the whole period of vacuum is misclassified in the sitting group at the top level. This is probably due to the leaning forward while vacuuming, which makes the orientation of accelerometers more similar to that of sitting. Such misclassifications fully propagated to the second level when using the movelets decision tree approach.

All integration approaches are flexible and can easily be generalized to more devices. The expanded movelets approach yields the best overall performance among the three integrated movelets approaches. However, each approach has strengthes and weaknesses. Movelets voting is conceptually straightforward. Since the integration of information occurs after using single-accelerometer movelets for each accelerometer, it can simultaneously process acceleration information from different sources and save computing time. At the same time, integration with majority votes of categorical activities labels loses some of the rich information embedded in the original signals. The expanded movelets approach merges all available information and provides the flexibility of weighting differ-

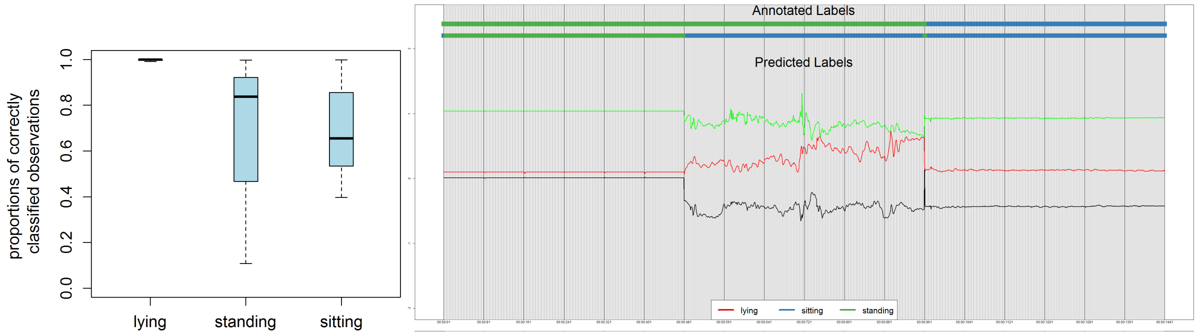

Figure 6: *The left panel presents the prediction accuracy for the top-level movelets decision tree. The right panel shows an example of the top-level prediction of movelets decision tree approach. The subject performed standing, vacuum (while standing), and write in sequence. The whole period of vacuum is misclassified as an activity while sitting, which is truly an activity while standing.*

ent information sources. This method yields exceptional good prediction in well controlled environments, though may be more prone to errors when one of the devices malfunctions or moves to a different position on the body. The movelets decision tree approach integrates information adaptively. An important assumption is that the top level of the decision tree is well designed to provide coarse discrimination between activity groups, while lower level decisions are well designed to make within-group predictions. Thus designing the tree hierarchy is rather delicate, will probably be application specific, and require refinement as more information becomes available.

4 Conclusion

In this paper we have addressed a tantalizing actual problem: can one predict the type of human activity from small sensors attached to the human body? The answer is yes, as otherwise this paper would never get published. However, here we want to qualify and nuance our response, clearly underline remaining roadblocks, and propose sampling designs that can dramatically improve data collection and analysis.

All predictions here are done on subjects in the lab wearing devices installed by trained technicians and subjects doing a given sequence of activities, which are intended to represent activities that happen in the subject's own environment. As much as one tries to standardize lab experiments, the data is likely to provide only a partial view of the heterogeneous activities individuals perform in their own houses. It remains unclear how in-lab data prediction algorithms perform in real life environments, especially in the absence of labeled in-home data on hundreds of individuals. Also, we have not yet investigated how well methods could be trained on one or multiple subjects and then applied to other subjects. Our experience seems to indicate that more work needs to be done and we remain mildly optimistic about the problem.

Probably the most important problem epidemiological studies will face in the future is the lack of in-home activity labeling as the ground truth. Moreover, increased quality control and standardization for in-home measurement will be necessary in future studies. Here we propose that observational studies involving accelerometers could dramatically improve data quality by incorporation of a set of standardized "life" activities that could be performed at the time the devices were initially placed. While the home setting is preferable, if not feasible, then even a clinic setting would still result in enhanced data quality. Our paper indicates that the activities can be far shorter with fewer repetitions and with far less physical and mental burden on the participants. Careful data annotation would be necessary, and careful checks of the location and orientation of the accelerometer devices should be performed. We suggest these to be done for several days, while instructing the individual to correctly install the device.

Needless to say, our proposal has met with strong resistance from scientists collecting the data and the main arguments merit in-depth discussion. First, it can be argued that this raises the burden to the individual by adding one or multiple short in-home visits. We actually think that, if done correctly and with sufficient planning, this will dramatically *decrease* the burden on the individuals. Indeed, a visit to the clinic or lab would require individuals, many of whom are older or impaired, to visit the clinic, spend hours traveling from home to the clinic, and possibly require assistance from a family member. Instead, the burden is shifted to trained professionals who travel to the home of individuals, help them with basic device installation in their familiar environment, thus reducing the physical and psychological burden on the individuals in the study. Second, it can be argued that asking individuals to go through a pre-determined set of activities may be physically prohibitive. We agree that performing strenuous physical activities should not be a requirement for individuals in the study. Moreover, we also agree that requiring any type of activity by phone or other means of communication without direct supervision should be avoided to prevent any possibly induced adverse health effects. However, the set of activities we propose is *actually much less intense* than the one required in *current studies.* For example, we would only require 2 repetitions of standing up from a chair, one normally, and one without using arms. For normal walking we can require as few as 10 seconds of combined walking in an area of the house where most walking is likely to occur. Third, it is argued that devices can simply be sent to individuals with instructions and activity can then be almost "magically" predicted using machine learning methods. As much as we enjoy "magic shows", the technology is not there yet, though important progress has being made. Here we argue in favor of fully transparent methods where the scientist understands the complex measurement, has access to the entire processing pipeline, and can access different levels of data compression via reproducible code and verifiable results.

5 Acknowledgement

This research was supported, in part, by the Intramural Research Program of the National Institute on Aging. He, Bai, and Crainiceanu were supported by Grant Number R01NS060910 from the National Institute of Neurological Disorders and Stroke. This work represents the opinions of the researchers and not necessarily that of the granting organizations.