# Designing a Robust Activity Recognition Framework for Health and Exergaming Using Wearable Sensors

Nabil Alshurafa, *Student Member, IEEE*, Wenyao Xu, *Member, IEEE*, Jason J. Liu, *Student Member, IEEE*, Ming-Chun Huang, *Student Member, IEEE*, Bobak Mortazavi, *Student Member, IEEE*, Christian K. Roberts, *Member, IEEE*, and Majid Sarrafzadeh, *Fellow, IEEE*

*Abstract*—Detecting human activity independent of intensity is essential in many applications, primarily in calculating metabolic equivalent rates and extracting human context awareness. Many classifiers that train on an activity at a subset of intensity levels fail to recognize the same activity at other intensity levels. This demonstrates weakness in the underlying classification method. Training a classifier for an activity at every intensity level is also not practical. In this paper, we tackle a novel intensity-independent activity recognition problem where the class labels exhibit large variability, the data are of high dimensionality, and clustering algorithms are necessary. We propose a new robust stochastic approximation framework for enhanced classification of such data. Experiments are reported using two clustering techniques, K-Means and Gaussian Mixture Models. The stochastic approximation algorithm consistently outperforms other well-known classification schemes which validate the use of our proposed clustered data representation. We verify the motivation of our framework in two applications that benefit from intensity-independent activity recognition. The first application shows how our framework can be used to enhance energy expenditure calculations. The second application is a novel exergaming environment aimed at using games to reward physical activity performed throughout the day, to encourage a healthy lifestyle.

*Index Terms*—Classification, clustering, energy expenditure (EE), exergaming, intensity-varying activity, mixture models, stochastic approximation model.

## I. Introduction And Related Works

**W**EARABLE sensors, specifically inertial sensors, continue to be used in activity recognition systems and devices [1]. For example, previous literature has shown the value of using accelerometer-based activity monitors [2], [3]. These activity monitors are used in combination with regression models
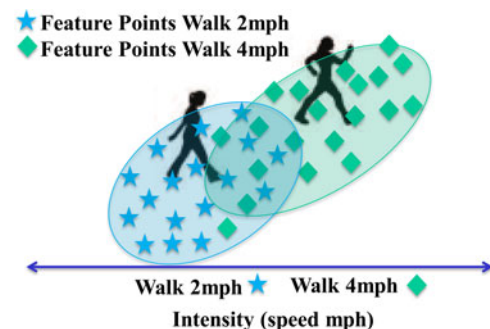
Fig. 1. Features collected from a wearable accelerometer at varying intensity levels can significantly overlap.

to enhance and automate the calculation of metabolic equivalent of task (MET), which is then used to approximate the calories burned [2]. The gaming industry continues to benefit from the accuracy of activity recognition systems using wearable kinematic sensors by developing interactive games that make gaming near-realistic [4], [5]. However, most research surrounding activity recognition either trains and tests activity at a single intensity level [6], or omits the intensity of the activity all together [7], [8]. Features collected on an activity categorized as running or walking will not only vary from one subject to the next, but will also vary with intensity. Yet when we physically see someone running or walking at any intensity level we can readily distinguish between the two, even though our eyes have not seen running and walking at every possible intensity level. The challenge remains to train a classifier on a subset of activity intensity levels, and to test it on different intensity levels. Only when a classifier performs well in classifying an activity at new intensity levels can it be claimed to be a robust provider of intensity-independent activity classification. We primarily focus our efforts on distinguishing activities of walking and running across multiple intensity levels.

Due to the large variance and overlap in the feature sets, it is quite challenging for a classification algorithm trained on one intensity level to distinguish walking and running at another intensity level [9]. Due to large intersubject variability, the nature of the data collected from an accelerometer at multiple intensities will be scattered with high variance, exhibiting overlapping categories across intensity levels [10], [11]. Fig. 1 provides an illustration of the large variance in the features collected across subjects and intensity levels, where someone walking at 2 mph may present similar characteristics to someone walking at 4 mph.

Class models are typically generated using raw sample points within each class label (activity category), resulting in complex and computationally inefficient systems. Some use clustering to enhance the computational efficiency [12]. We attempt to generate an activity model that uses clustering combined with uncertainty within the clusters to learn a robust activity model that can classify the same activity at a range of intensity levels, resulting in a more efficient formulation. Our framework is compared against other well-known classification schemes such as SVM, kNN, and C4.5 Decision Trees (DTs).

### A. EE Calculation

Typically, physiological variable such as METs are predicted using values from a tri-axis accelerometer [13]. The relationship between the accelerometer and MET is determined using multiple regression techniques [14]. However, there are more than 30 such techniques that produce very different results [15], [16]. Also, differences in energy expenditure (EE) values from accelerometer data result from inconsistencies in the calibration process, which makes comparing results among studies difficult [17]. Our ability to detect an activity type independent of its intensity empowers EE calculations to use separate regression models for each activity type, resulting in an improvement in EE calculations.

In many health-related fields, interest in EE has sparked the need to analyze the intensity of human activity. Ainsworth *et al.* attempted to compile a compendium using regression models to map common physical activities to METs [18]. However, Kozey *et al.* [14], showed that linear regression models are inappropriate for accurately predicting METs from accelerometer data. Albinali *et al.* [6] argued that automatic detection of physical activity type prior to using activity regressions enhances the estimate of EE. Moreover, activity monitors that rely on such equations have no physically interpretable meaning of the activity being performed. In this paper, we focus on identifying a model for a physical activity type that is independent of intensity, which we believe will aid in understanding the underlying physical meaning behind each activity type.

In other works, Vathsangam *et al.* [19] analyzed walking at only two intensity levels, predicting EE from treadmill walking using hip-worn inertial sensors; they applied regression techniques such as least-squares regression, Bayesian linear regression, and Gaussian process regression, without classifying the activities. Ravi *et al.* [7] and Kwapisz *et al.* [20] reached subject-dependent classification accuracies above 90%; however, they did not focus on distinguishing the intensity of the activity. Tapia *et al.* [9] attempted to recognize physical activities and their intensity levels using a C4.5 DT classifier, yet achieved poor subject-independent accuracies of 58.2%, and also did not train on a subset of activity intensity levels or test on another intensity.

### B. Exergaming

To encourage activity throughout the day, not only when playing the video game, we developed a video gaming system that maps specific activities performed throughout the day to the avatar in a video game. To test this, we modified an existing open-source role-playing game (RPG) video game named FreedroidRPG, to receive input from a wearable accelerometer that is embedded in a belt designed to be placed around the waist. The game uses our classification framework to detect activity type and time spent in each activity type to convert energy exerted in the real world to the potential energy of the avatar of the game. The goal of the video gaming system is to promote exercise throughout the day, thus encouraging a more active lifestyle. Our activity recognition framework will aid such video games in classifying a given activity independent of its intensity level.

### C. Methodology

Miller and Browning [21] have shown success using mixture models and expectation maximization (EM)-based algorithms for new class discovery with mixed labeled and unlabeled datasets. Sun *et al.* [22] attempted to classify EEG signals using stochastic approximation to update parameters of Bayesian classifiers on EEG signals. Compared to existing studies, this study differs accordingly:

1) Intensity-independent classification: Our model-based approach attempts to represent activities holistically, and is capable of inferring other intensity levels from a subset of intensity levels.
2) Stochastic approximation framework: We utilize clustering algorithms and stochastic approximation to create a model for each activity, and then use a stochastic approximation nearest-neighbor heuristic for classification.
3) MET approximation: We use our stochastic approximation framework to enhance the accuracy of detecting intensity-varying activities, and apply a different regression model based on each activity type.
4) Exergaming platform: We analyze our classification in real life on a new gaming platform that converts energy exerted in the real world to the potential energy of the avatar of a video game.

This paper is organized as follows. In Section II, we briefly review background on clustering algorithms and stochastic approximation. Then, we describe our intensity-independent activity recognition framework in Section III. We present our experimental results and analysis in Section IV, and related applications in EE and Exergaming in Section V. Finally, we conclude and discuss future work in Section VI.

## II. PRELIMINARY

### A. Data Modeling and Representation via Clustering

Clustering has been widely used in pattern recognition applications [23] as a means of organizing data into groups or clusters based on a similarity metric. It is mainly applied to discover classes and structures of data in an unsupervised-based learning fashion. It is also an efficient method for representing data, allowing for fast retrieval, and dimensionality reduction. In this paper, we use clustering algorithms as a means of efficient data representation. In this section, we mainly cover two types of clustering algorithms that are further examined in the

remainder of this paper: K-Means exclusive clustering, and a Gaussian mixture model (GMM)-based probabilistic clustering algorithm.

Let us assume we have $n$ sample points, denoted by $x_i \in R^m$. Each $x_i$ represents a vector of features extracted from a sample. The K-Means clustering technique is one of the simplest and most popular unsupervised learning algorithms [24], yet because our training samples are labeled, we deploy K-Means to further cluster each activity category $A_i$, providing a more efficient representation of each activity category. K-Means attempts to partition the sample points into $k$ clusters $C = C_1, C_2, \ldots C_k$, where $k < n$. With initial random placement of $k$ centroids $< \mu_1, \mu_2, \ldots, \mu_k >$, where $\mu_j \in R^m$, and each point $x_i$ is assigned to the closest cluster $C_j$ with centroid $\mu_j$. $k$ centroids are iteratively recalculated and points reassigned such that the within-cluster sum of squares is minimized

$$\arg\min_C \sum_{j=1}^{k} \sum_{x_i \in C_j} \|x_i - \mu_j\|^2. \qquad (1)$$

Training samples tend to be tightly clustered around the centroids, which can then serve as a compact representation for the training data. K-Means is simple to implement, but does not necessarily find a global optimal solution based on the objective function. K-Means is also known to be sensitive to the initial randomly selected cluster centers, and has problems when clusters differ in size, density, and shape [25].

The GMM [26] is another widely used, model-based clustering technique. GMMs optimize the fit between data and a parametric distribution (like a Gaussian or Poisson distribution), where the entire data are modeled by a mixture of such distributions. Each activity category $A_i$ will be modeled using a mixture of Gaussian distributions separately, with the parameters to be learned: $\omega_j$ is the prior probability of the $j$th distribution, mean $\mu_j$, and variance $\sigma_j^2$. Each Gaussian distribution is represented by $N(\mu_j, \sigma_j^2)$, and the parameters for each distribution are collectively represented by $\theta$. For each activity category $A_i$, the probability $p(x_i|\theta)$ that a data point $x_i$ belongs to a category $A_i$ can be represented by a weighted sum of $k$ component Gaussian densities of activity $A_i$, given by the following equation:

$$p(x_i|\theta) = \sum_{j=1}^{k} \omega_j N(x_i|\mu_j, \sigma_j^2). \qquad (2)$$

For each activity category $A_i$, there belongs a set of points $X = \langle x_1, x_2, \ldots, x_n \rangle$, drawn from an unknown distribution, we estimate the parameters of the GMM model $\theta$ that fits the data. The solution is to maximize the likelihood $p(X|\theta)$ of the data with regards to the model parameters

$$\arg\max_\theta p(X|\theta) = \arg\max_\theta \Pi_{i=1}^{n} p(x_i|\theta). \qquad (3)$$

This expression is a nonlinear function of the parameters $\theta$ and direct maximization is not possible. However, the maximum likelihood parameter estimates can be calculated using a special case of the EM algorithm [26], and is used in practice to find the mixture of Gaussians that can model the data.

### B. Stochastic Approximation

Stochastic approximation is a robust method used to solve objective functions while taking into account uncertainty or possible variation in the data [27]. It typically involves a nontractable convex optimization problem, where the objective or its derivatives are difficult to evaluate.

\tight We consider a matrix $A \in R^{m \times n}$ with the basic objective $\|Ax - b\|$, where $b \in R^m$ is an unknown test sample, but also wish to take into account uncertainty. In many cases, the number of unknowns is larger than the feature space, $n > m$, making it an underdetermined system. In our application, we use clustering algorithms to ensure that $n < m$, resulting in an overdetermined system. The mean of $A$ is $\bar{A}$, so it can be described as

$$A = \bar{A} + U \qquad (4)$$

where $U$ is a random matrix with zero mean. $U$ describes the statistical variation of $A$ and can be used to model uncertainty or large variation in the data matrix $A$. A natural objective is to minimize the expected value of $Ax - b$

$$\min_{x \in R^n} E\|Ax - b\|. \qquad (5)$$

This is a general form of the stochastic robust approximation problem. Some variations of the statistical robust approximation problem are tractable, for example the statistical robust least squares problem, where we minimize the Euclidean norm of $Ax - b$

$$\min_{x \in R^n} E\|Ax - b\|_2^2. \qquad (6)$$

This objective function has a closed-form solution that can be expressed as follows:

$$\begin{aligned} E\|Ax - b\|_2^2 &= E(\bar{A} - b + Ux)^T (\bar{A}x - b + Ux) \\ &= (\bar{A}x - b)^T (\bar{A}x - b) + Ex^T U^T Ux \\ &= \|\bar{A}x - b\|_2^2 + x^T Px \end{aligned} \qquad (7)$$

where $P = E(U^T U)$. This is exactly the form of a regularized least-squares problem [27]

$$\min_{x \in R^n} \|\bar{A}x - b\|_2^2 + \|P^T x\|_2^2. \qquad (8)$$

And the solution is

$$x = (\bar{A}^T \bar{A} + P^{-1} \bar{A}^T b). \qquad (9)$$

### III. OUR FRAMEWORK

In this section, we present our framework for intensity-independent classification of activity. Fig. 2 illustrates our framework comprising four important components: feature extraction, clustering-based representation, stochastic activity modeling (SAM), and stochastic approximation decision classification. The training phase begins by extracting features from the accelerometer time series data, which is then fed into a clustering-based algorithm. The data are then grouped by activity into clusters, and the clusters' mean and variance are combined to form a corresponding SAM. In the recognition phase,
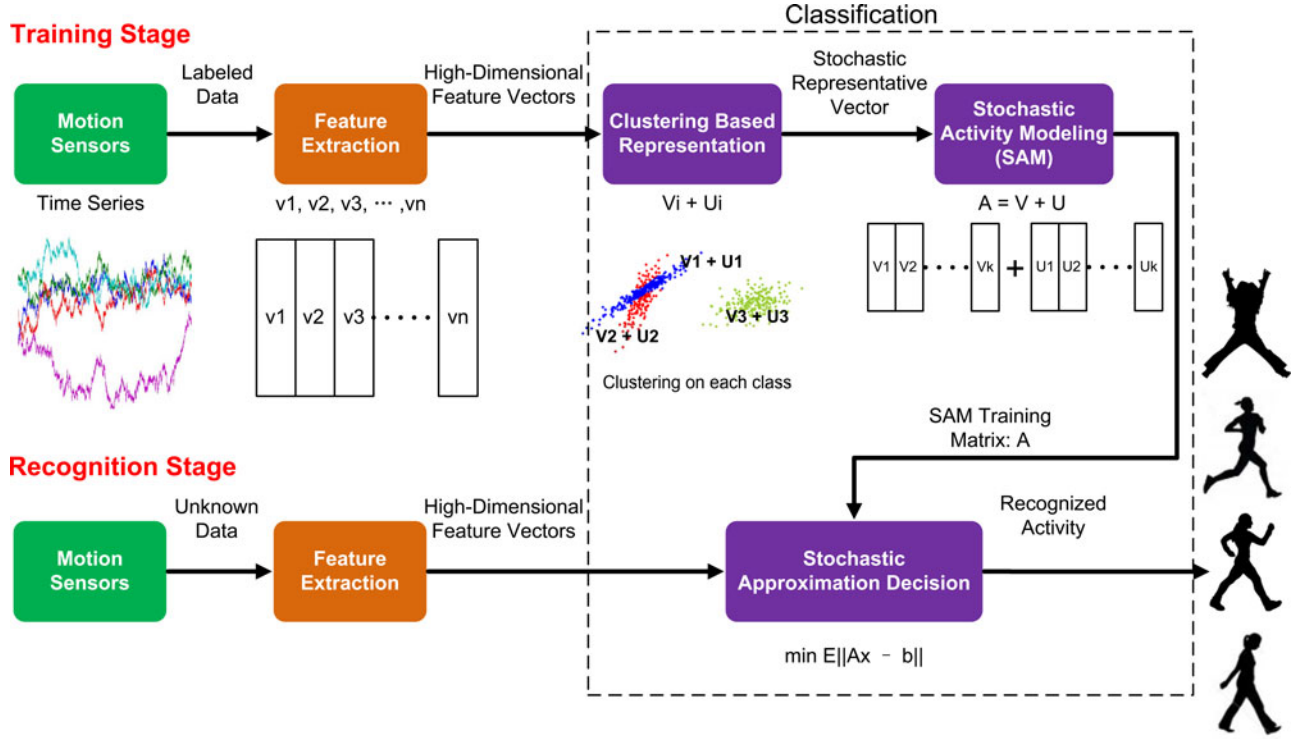
Fig. 2. Stochastic approximation framework comprises a training and a recognition stage. In the training stage, a high-dimensional set of features are extracted from motion sensors, labeled, and presented to the stochastic approximation classifier which deploys clustering algorithms to generate Stochastic Activity Models. In the recognition stage, features are extracted and compared against multiple SAMs in a stochastic approximation decision process to find the best label for the data.

TABLE I
FEATURE TABLE

| Mean | Standard Deviation | Mean Derivatives |
|------|--------------------|--------------------|
| Median | Pairwise Correlation | Interquartile Range |
| Skewness | Root Mean Square | Zero Crossing Rate |
| Variance | Mean Crossing Rate | Kurtosis |

the sample test points are compared against each activity category's SAM in a stochastic decision-making component. We describe the details of these components in this section.

### A. Feature Extraction

There are several studies that analyze varying features that are best utilized for human activity based on the accelerometer data. Table I lists the main features that have been shown to be useful in classifying activity [7], [28], [29]. Using the data from each $X, Y, Z$ acceleration axis generates a total of 45 features per segment, where a segment is a 2-s fixed time subdivision of the accelerometer data. Each segment results in a feature vector that belongs to one of $N$ categories: $A_1, A_2, \ldots, A_N$, where each activity category represents a specific activity type.

### B. Applying Clustering Algorithms

In order to solve the least squares problem defined in (8), we need to obtain the matrix $\bar{A}_i \in R^{m \times k}$ for each activity. We collect all the samples from class $A_i$ into $k$ clusters $C_1, C_2, \ldots, C_k$ using K-Means or the GMM clustering algorithm. We ensure that the generated number of clusters $k$ is less than the number

of features $m$ in order to make sure we have an overdetermined system. Each cluster $C_i$ can be represented by $V_i + U_i$, where $V_i$ is the cluster center, and $U_i$ is the random noise based on the distribution of cluster $C_i$.

We tested both the K-Means and GMM clustering methods. From each generated cluster, we extracted the mean and variance, and combined them for each activity category $A_i$ to form the $V_i$ and $U_i$ matrix, respectively.

### C. SAM and Classification

Once the data are clustered, we then combine all the clusters of each activity to form a SAM matrix $B_i$ for each activity $A_i$. Each category contains $k$ clusters which are generated from $n_i$ training samples, each having $m$ features

$$B_l = [V_1, V_2, \ldots, V_i, \ldots, V_k] + [U_1, U_2, \ldots, U_i, \ldots, U_k]$$

$$= \begin{bmatrix} v_{11} & \cdots & v_{1k} \\ v_{21} & \cdots & v_{2k} \\ \vdots & \ddots & \vdots \\ v_{m1} & \cdots & v_{mk} \end{bmatrix} + \begin{bmatrix} u_{11} & \cdots & u_{1k} \\ u_{21} & \cdots & u_{2k} \\ \vdots & \ddots & \vdots \\ u_{m1} & \cdots & u_{mk} \end{bmatrix} \quad (10)$$

where $v_{i,j}$ represents the center of the $i$th feature of the $j$th cluster, and $u_{i,j}$ represents the variance of the $i$th feature of the $j$th cluster. Given a test segment $b$, we want to make a decision as to which SAM $B_l$ belongs. To do this, we solve the following using (8)

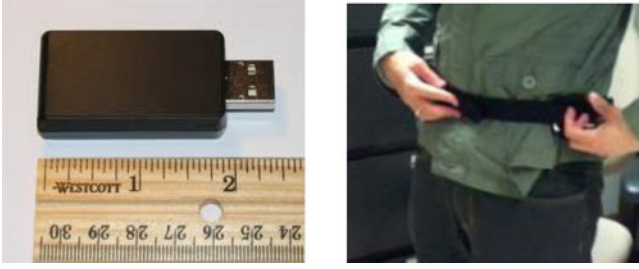$$\min_{i=1\ldots N} E\|B_l x - b\|_2^2. \quad (11)$$

Fig. 3. Image on the left is an illustration of the sensing platform. The image to the right is that of a participant wearing the motion sensing node.

TABLE II
ACTIVITY TYPES

| W1 | Walking 2.5mph | R1 | Running 4mph |
|----|----------------|----|--------------|
| W2 | Walking 3.5mph | R2 | Running 5mph |
| W3 | Walking 5mph   | R3 | Running 6mph |

In other words, we iterate through each of the $N$ SAMs and determine to which activity category $b$ belongs by choosing the one with the minimum reconstruction error, a process similar to searching for the nearest-neighbor SAM. We validate the classification performance on the collected data in a lab setting.

## IV. EXPERIMENTS AND EVALUATION

We performed three experiments to test the performance of our activity recognition framework. The first experiment was designed to collect data from a wearable accelerometer to create a classifier for each activity type, where participants performed a predefined set of activities in lab. The second and third experiments were designed to validate the usefulness of our framework in both enhancing caloric expenditure calculations and in a video game setting; this will be further discussed in Section V. During each experiment each participant wore a belt-like strap around the waist prior to performing any activity. A Gulf Coast Data Concepts X6-2 mini tri-axis accelerometer is embedded inside the belt, positioned on the participants front left hip as illustrated in Fig. 3. It is capable of measuring acceleration with 12-bit resolution, setting the detection range at $+/- 2g$. The accelerometer was set to record readings at a frequency of 80 Hz. Features in Table I were extracted from the accelerometer.

### A. Data Collection

We collected data from 12 subjects, ranging in age from 20 to 28 years. Each subject wore the accelerometer-embedded strap around their front left hip and performed several exercises on a Merit Fitness 715T Plus treadmill–walking and running at three intensity levels each–and performed jumping jacks to the side of the treadmill. Each activity was performed for 5 min, with 5 min of rest between each activity. During analysis of results from generic classifiers, we found the majority of misclassifications occurred between the running and walking activity types. For this reason, our intensity-independent model targets correct classification of walking and running. Table II provides the labels for each category, we selected typical walking and running speeds where MET values have been estimated before in order
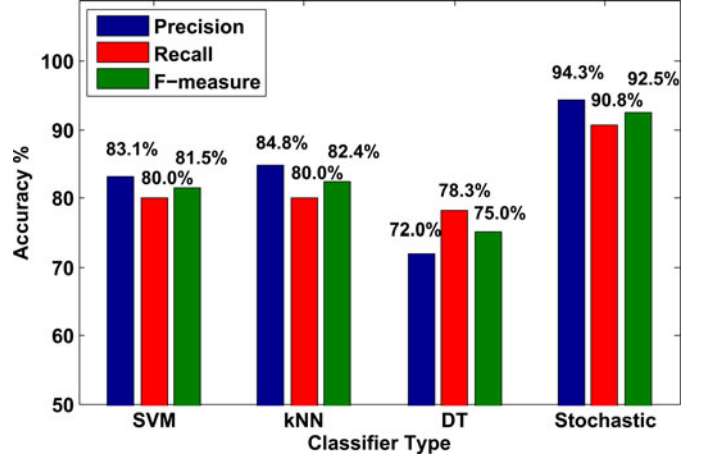


Fig. 4. Precision, recall, and F-measure across multiple classifiers applied to walking and running at multiple intensity levels. The stochastic approximation classifier with GMM clustering outperforms SVM, kNN, and C4.5 DT.
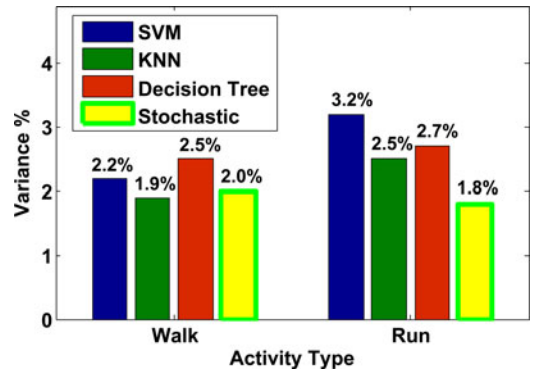


Fig. 5. Variance across classifiers during leave-one-out cross validation shows that the stochastic approximation classifier is a robust classifier with low variance.

to analyze the benefits of our classification scheme toward EE calculations [14], [18].

### B. Classification Performance and Evaluation

In order to discover the potential of our stochastic approximation framework, we compared our classifier against other well-known classifiers. Each classifier was trained with 11 subjects and tested on one, using six class labels (or activity categories), performing leave-one-out cross validation by subject. We tested each classifier after applying the data clustering algorithms. Fig. 4 shows the results for each classifier. Using the SVM, kNN, and C4.5 DT classification algorithms, we achieved F-measures (the harmonic mean of precision and recall) of 81.5%, 82.4%, and 75.0%, respectively. Our stochastic approximation framework with the GMM clustering algorithm yielded a 94.3% precision, 90.8% recall, and a 92.5% F-measure. Based on these results, we conclude that the stochastic approximation classification framework is better able to handle the large variation within each activity type.

We also evaluate the stability of our classification algorithm by analyzing the variation during cross validation. Stability is another important measure that describes how closely a classifier

TABLE III
CONFUSION MATRIX FOR THE STOCHASTIC APPROXIMATION CLASSIFIER TRAINING ON SIX CATEGORIES

| | | True Class | | | | | | | | Category Precision |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Walk | | | Run | | | | | |
| Category | Activity | W1 | W2 | W3 | R1 | R2 | R3 | Total | Precision | |
| Walk | W1 | 75 | 5 | 1 | 0 | 0 | 0 | 81 | 92.6% | 93.8% |
| | W2 | 6 | 71 | 1 | 2 | 2 | 0 | 82 | 86.6% | |
| | W3 | 3 | 6 | 76 | 5 | 5 | 2 | 97 | 78.4% | |
| Run | R1 | 1 | 2 | 2 | 73 | 2 | 2 | 82 | 89.0% | 95.6% |
| | R2 | 0 | 1 | 4 | 3 | 74 | 6 | 88 | 84.1% | |
| | R3 | 0 | 0 | 1 | 1 | 2 | 76 | 81 | 93.8% | |
| | Total | 85 | 85 | 85 | 85 | 85 | 85 | | | |
| | Recall | 88.2% | 83.5% | 89.4% | 85.9% | 87.1% | 89.4% | | | |
| Category Recall | | 95.7% | | | 93.7% | | | | | |

TABLE IV
CLASSIFICATION RESULTS FOR THE STOCHASTIC APPROXIMATION CLASSIFIER TRAINING ON A SUBSET OF CATEGORIES

| | Training (X means selected in training) | | | | | | Testing | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Average Category F-measure | | Intensity Independent F-measure | |
| Classifier Number | W1 | W2 | W3 | R1 | R2 | R3 | $Walk_A^*$ | $Run_A^*$ | $Walk_T^\dagger$ | $Run_T^\dagger$ |
| 1 | X | X | | X | X | | 84.5% | 85.5% | 82.1% (W3) | 83.2% (R3) |
| 2 | X | X | | X | | X | 81.5% | 83.5% | 79.4% (W3) | 80.5% (R2) |
| 3 | X | X | | | X | X | 86.5% | 84.5% | 82.8% (W3) | 80.9% (R1) |
| 4 | X | | X | X | X | | 89.0% | 88.5% | 83.6% (W2) | 83.2% (R3) |
| **5** | **X** | | **X** | **X** | | **X** | **94.0%** | **92.5%** | **84.1% (W2)** | **85.6% (R2)** |
| 6 | X | | X | | X | X | 86.5% | 85.5% | 83.4% (W2) | 82.1% (R1) |
| 7 | | X | X | X | X | | 89.5% | 89.0% | 85.1% (W1) | 84.3% (R3) |
| 8 | | X | X | X | | X | 90.5% | 91.5% | 84.2% (W1) | 83.2% (R2) |
| 9 | | X | X | | X | X | 86.5% | 85.5% | 79.7% (W1) | 80.1% (R1) |
| 10 | X | | | X | | | 53.4% | 55.6% | 45.6% (W2,W3) | 49.2% (R2,R3) |
| 11 | | X | | | X | | 65.3% | 67.8% | 57.8% (W1,W3) | 58.3% (R1,R3) |
| 12 | | | X | | | X | 56.2% | 58.2% | 49.3% (W1,W2) | 51.2% (R1,R2) |

\*$Walk_A$ and $Run_A$ represent the average F-measure across all intensity values.
†$Walk_T$ and $Run_T$ represent the F-measure of the tested intensity independent activity not used in the training set.

evaluates results if given different data. Fig. 5 shows the classification variation for each of the classifiers. We can conclude that the stochastic approximation framework not only outperforms other classifiers in precision, recall, and F-measure results, but also produces stable results across multiple runs, resulting in low variance when compared to other classifiers.

### C. Stochastic Nature of Human Activity

When we separated each intensity level and activity type into its own category, training the classifier on six categories, the data exhibited less variation. The stochastic approximation classifier was then trained on each intensity level separately, performing well on the test data. In order to determine the performance of the classifier in classifying the correct activity type, we must look at the category precision and recall. Table III shows the confusion matrix, where results show high average category precision and recall. While the classifier performs sufficiently well when trained on six categories, our primary goal has been to demonstrate the classifiers generalizability when trained on a subset of intensity levels. Table IV shows the results of 12 classifiers, the first nine classifiers trained on four activities and tested on two, while Classifiers 10–12 trained on two activities and tested on four. We show the average category F-measure, which is the average F-measure across intensity levels ($Walk_A$

column is an average of W1, W2, and W3, and the $Run_A$ column is an average of R1, R2, and R3). Table IV also shows the F-measure for the specific intensity-independent activity excluded from the training set, e.g., the first row trained on W1, W2 and R1, R2 and tested on W3 and R3, as shown in the $Walk_T$ and $Run_T$ columns in Table IV.

It is interesting to note that of the first nine classifiers the one that performs the best is the one that trains on the extremes of each category, W1, W3 and R1, R3. This result is illustrated in the fifth row of Table IV, where the stochastic approximation classifier is capable of extrapolating information about W2 from W1 and W3, and R2 from R1 and R3. Typically when a subject runs at 2 mph, they are not really running at 2 mph, but fluctuate between intensities slightly above and below 2 mph, and this noise is represented in the SAM of each the activity type, since it takes into account the variance in the clusters. The SAM is capable of extrapolating information about adjacent intensity levels from the variance in the data of each cluster. This supports our claim that in order to train a classifier for a given physical activity, we only need to train the stochastic approximation classifier on a subset of intensity levels.

Of the first nine classifiers, the classifier with the worst performance was the one trained on W1, W2 and R1, R3, as shown in row 2 of Table IV. This is due to the fact that there is much overlap between W3 and R2, since they represent walking and

TABLE V
COMPARING K-MEANS AND GMM WITH THE STOCHASTIC
APPROXIMATION CLASSIFIER

| K-Means (Average) | | | GMM (Average) | | |
|---|---|---|---|---|---|
| Precision | Recall | F-measure | Precision | Recall | F-measure |
| 89.2% | 86.3% | 87.7% | 94.4% | 93.7% | 94.0% |

running at 5 mph, respectively. Testing on the run category R2 yields a slightly higher F-measure, 80.5% compared to 79.4% for W3. This shows the classifier is able to perform a little better on R2 given it was trained on R1 and R3; however, the higher intensity activity, W3, is more challenging to classify given the training of lesser intensity activities W1 and W2.

Table IV also shows that of the last three classifiers that train on two categories, the best classifier is Classifier 12, which trains on W2 and R2 and tests on the remaining activities. This further supports the classifiers ability to extrapolate information about adjacent intensity levels from the variance of a single activity; however, it still does not outperform Classifier 5, which provides larger variation in the training set.

We wanted to investigate the choice of clustering algorithm on the performance of our stochastic approximation framework. We compared the outcomes of two clustering techniques, K-Means and GMM. It is interesting to note that in our application using GMM outperforms K-Means on average. This happens because K-Means essentially represents each category by a cluster of centroids, while GMM represents each category by a cluster of distributions, each with a mean, variance, and probability. As a result, GMMs can better represent the model when the data are sparse with high variability. Table V shows GMM outperforming K-Means in average precision, recall, and F-measure.

## V. APPLICATIONS

In this section, we attempt to define two potential applications that benefit from detecting activity independent of intensity. Each application was verified with an experimental setup. The first application shows how using our classification framework results can result in enhanced accuracy in calculating EE. The second application tests the generalizability of our classification framework in an exergaming environment.

### A. Calculating EE

Activity monitoring systems need to be able to improve their accuracy when calculating EE, as this accuracy is critical to the ability to determine relationships between EE and related variables [30]. Crouter et al. [31] showed improvement in EE estimates by using two regressions one for sedentary activities and another for nonsedentary activities. Albinali et al. [6] found it useful to perform separate regressions on each specific activity. We reaffirm Albinali's claim, and show that detecting the activity type first, and then using a separate regression model yields more accurate EE calculations.

*1) Experimental Setup:* In this experiment, we repeated the first experiment with 15 new participants performing the activities using a metabolic cart that estimates oxygen uptake, as illustrated in Fig. 6; however, due to the challenges of moving
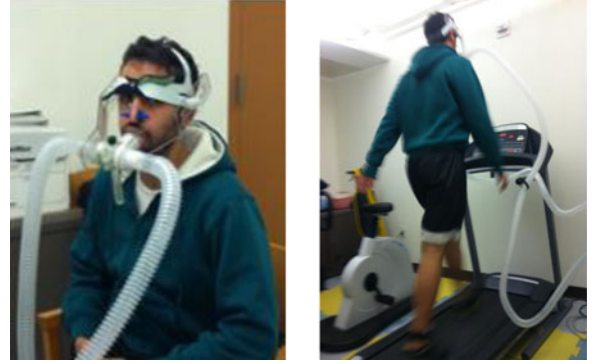


Fig. 6. Subject performing trials with a metabolic cart and accelerometer attached to a belt on the hip. The photo to the left shows the subject sitting wearing the face mask connected to the metabolic cart, with his nose clipped shut to ensure breathing occurs only through the mouth to enable the measuring of oxygen uptake. The photo to the right shows the user performing activity on the treadmill.

while using the metabolic cart, we decreased the duration of the activities performed and the rest between each activity to 3 min. Fifteen participants were selected, six females and nine males between the ages of 23 and 30. The metabolic cart estimates oxygen uptake, defined by $VO_2\left(\frac{ml}{min}\right)$, which is then used to calculate MET levels accordingly

$$MET = \frac{VO_2}{3.5 * m} \tag{12}$$

where $m$ is the mass of the participant in kilograms. METs are then used to approximate the calories burned and can be calculated by

$$Calories = \frac{MET * 3.5 * m * t}{200} \tag{13}$$

where $t$ is the time in minutes.

We separated the participants walking and running accelerometer data and MET values obtained from the metabolic cart. We then performed regressions on each activity type independently.

*2) Results:* We found that separating the data according to activity type greatly enhances the accuracy of the EE estimate. Fig. 7 shows the actual MET values obtained from the metabolic cart, and attempts to calculate the MET using two methods. It compares the estimates of EE using a single regression model trained without separation on all the activity types (labeled "All Regression" in Fig. 7), and using the separation method to apply a different regression model for each activity type (labeled "Separate Regression" in Fig. 7). In the "Separate Regression" method, the stochastic approximation classifier was used to detect the activity type and then apply the corresponding activity specific regression model.

Our first observation from Fig. 7 is that the regression equations when developed using all activity types, under the "All Regression" method, consistently underestimated the EE of activities in the walking category, while subsequently overestimating activities of running. This could be due to the position of the accelerometer on the hip, where it does not effectively capture the movement in the lower part of the body as well as the upper.
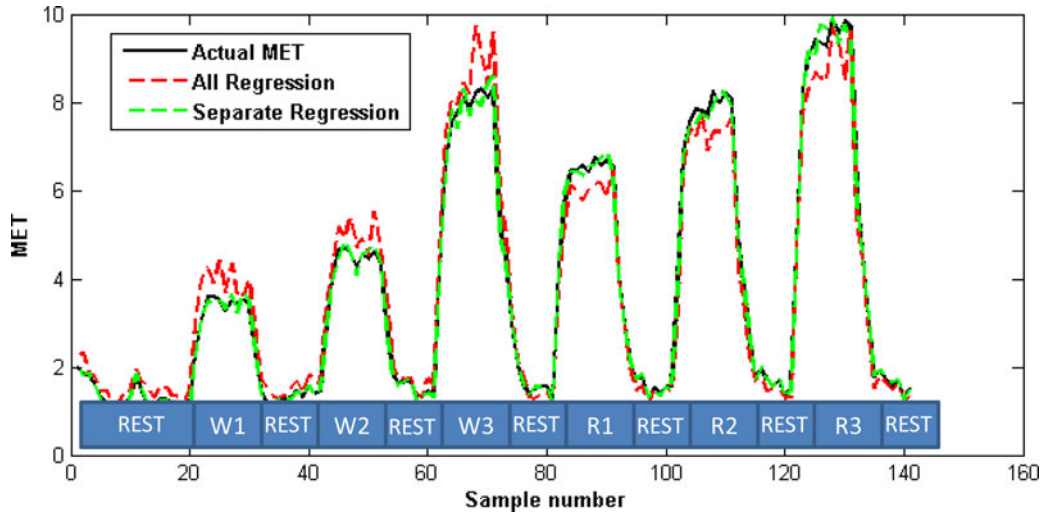
Fig. 7. Actual MET values compared against two regression methods for one subject performing predefined activities. Each sample represents 15 s of activity, which was the output of the MET value from the metabolic cart. Prior to performing each activity, the subject would rest for 5 min until reaching their resting metabolic rate of approximately 1 MET. It can be seen from the "Separate Regression" curve that using an activity-dependent regression yields closer results to the actual MET value when compared to the "All Regression" method.

TABLE VI
COMPARISON OF DIFFERENT METHODS IN ESTIMATING EE ACROSS ALL 15 SUBJECTS

| Activity | Actual MET | All Regress | Separate Regress |
|---|---|---|---|
| W1 | 12.8 | 15.1 | 12.6 |
| W2 | 16.7 | 18.5 | 16.3 |
| W3 | 29.5 | 32.1 | 29.9 |
| R1 | 24.1 | 22.4 | 24.2 |
| R2 | 28.9 | 26.6 | 28.5 |
| R3 | 34.7 | 28.5 | 34.3 |
| Mean Errors | | 11.74% | 1.38% |

As a result similar features extracted from the accelerometer on the hip from both walking and running activity types yield different MET values, and for this reason, a single regression model is unable to truly capture accurate MET values of both activity types.

Table VI shows that when we generate a separate regression model for each activity type independent of its intensity, we are able to achieve more accurate calculations of EE. Fig. 7 illustrates the accuracy of the "Separate Regression" method for a single subject. This bolsters the findings of Albinali *et al.* [6], however, we also observe that using a separate regression model for each activity type irrespective of their intensity levels also yields positive results. We expand on Crouter's approach to using a separate regression model for sedentary and nonsedentary activities, by showing that among nonsedentary activities, separate regression models should be used for different activity types, and only a single model needs to be used for an activity type with varying intensity levels. Abinali *et al.* [6] did not test similar activities across multiple intensity levels. Table VI compares the mean errors (averaged over all activities) of using different methods in estimating EE. Detecting the activity type and then applying the corresponding regression model outperforms using a single regression model, resulting in a mean error

rate of 1.38%, compared to 11.74% using a single regression model for all activities.

### B. Exergaming Platform

Our gaming platform consists of two main components: an accelerometer-embedded belt and an RPG video game called FreedroidRPG [32]. The video game was used as an incentive for the participant to perform physical activity throughout the day. FreedroidRPG tells a story of conflict between robots and their human masters. Tux is the avatar that is out to save the world from the evil robots. The best classifier resulting from the first experiment, which was capable of performing well on subject- and intensity-independent activity, was embedded into FreedroidRPG. The classifier is capable of detecting the correct activity type and duration performed by the participant after processing data from a wearable accelerometer. As a result, when Tux was low on energy, the participant would wear an accelerometer-embedded belt and perform a set of defined activities throughout the day including running, walking, and jumping-jacks. Fig. 8 shows Tux dying and the player performing activity in order to reenergize him.

The video game classifies each activity type and duration into the following categories: running, walking, and jumping-jacks. Running is mapped to avatar physique, where strengthening physique increases avatar health/lifetime. When enemy bots strike Tux, it decreases his health/lifetime. Walking is mapped to character dexterity, which affects the avatars success hit rate when attacking an enemy, and his success in defending himself from enemy bots. Jumping jacks are mapped to character strength, where increases in strength result in stronger blows to the enemy bots. The goal of this video game is to encourage physical activity, since the only way to increase avatar physique, dexterity, and strength is by walking, running, and performing jumping jacks throughout the day. Table VII
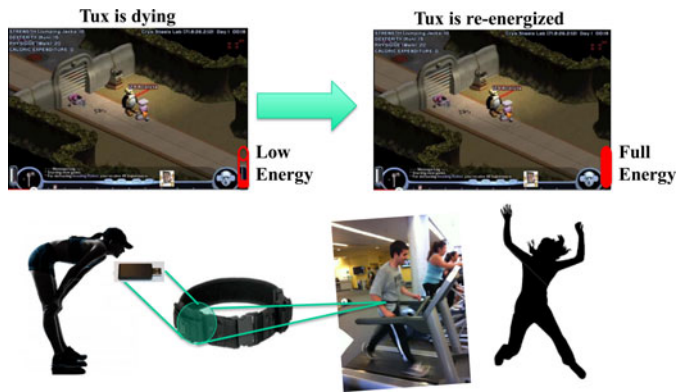
Fig. 8. FreedroidRPG exergaming version being played by a participant. When the avatar, Tux, has poor physique it effects his energy, and so the participant must pause the game, perform activity while wearing the accelerometer, and then reconnect the accelerometer with the game in order to reenergize Tux.

TABLE VII
MAPPING ACTIVITY TYPE TO AVATAR ATTRIBUTE

| Duration (min) | 1.0 | 1.0 | 1.0 |
|---|---|---|---|
| Activity Type | Walking | Running | Jumping Jacks |
| Mapped Attribute | Dexterity | Physique | Strength |
| Points | 5 | 5 | 10.0 |

describes the mapping of each activity to the points in the avatar of the video game.

*1) Experimental Setup:* This setup involved testing the system on the 15 participants in the second experiment. The goal of this experiment was to see the performance of the classifiers on running and walking at varying speeds in an uncontrolled fashion, as well as analyze the potential for such a game to succeed. We trained each participant for 1 h on the video game without the exergaming component. Once the player was comfortable with the rules, we activated the exergaming version of the game. The exergaming version provided rules for mapping physical activity performed while wearing the accelerometer to Tux's in-game attributes. This forced participants to perform the physical activity using the accelerometer in order to replenish Tux's energy. The video game would use the embedded classifier to detect the activity type and map treadmill walking, treadmill running, and jumping jacks to Tux's specific qualities. In this experiment, we did not instruct the participant at which intensity level to walk and run, and we left them to freely decide on the necessary activity to perform to continue playing the game. We did, however, use a stopwatch to manually record the amount of time they performed each activity type, and compared it with the results from the classifier embedded in the video game. We tested the outcomes of several classifiers including: SVM, kNN, C4.5 DTs, and the Stochastic Classifier.

When Tux would lose strength, physique or dexterity the participant would: pause the video game, perform physical activity while wearing the accelerometer, then select a transmit button to transmit the accelerometer data to the game. The video game would process the accelerometer data and apply a chosen classifier to detect the activity type and duration performed by the participant. Each subject played the video game for at-least 1 h,

TABLE VIII
COMPARING CLASSIFIERS ON DATA IN THE FREEDROIDRPG EXERGAME

| | Test by Category Average F-measure | |
|---|---|---|
| | Walk | Run |
| SVM | 72.6% | 70.4% |
| kNN | 74.1% | 73.2% |
| C4.5 DT | 70.0% | 68.3% |
| Stochastic* | 85.4% | 81.1% |



Fig. 9. Exergaming survey.

collecting a total of 51.4 min of walking, 40.3 min of running, and 15.6 min of jumping jack data for all subjects. Jumping jacks are less challenging to distinguish in accelerometer data, and as a result we focus our analysis of the classification results on walking and running. Each participant answered a six-question survey at the end of the experiment.

*2) RPG Video Game Classification Performance and Evaluation:* We then chose the optimal classifier identified by the stochastic approximation classifier in Table IV and used it in the video game to classify walking and running accurately. The classifier was tested on activities with intensity levels defined by the participant. Table VIII shows the Category F-measure of the data collected from the video game setting, showing the stochastic approximation framework's ability to outperform other well-known classification methods.

*3) Human–Computer Interface:* In order to evaluate the potential of such an exergame, a user experience survey is essential to see the feasibility of such an application. Several works focus on the appropriate methods in evaluating a user's experience in a gaming environment [33], [34]; however, we focus on the following three: encouragement of physical activity [35], [36], comparison to other games and general enjoyment [34], [37], and realism of exergames [37]. The survey questions are illustrated in Fig. 9.

It is very important that participants enjoy a video game, otherwise it will not succeed commercially. The responses from the survey given to each participant show that the majority of the participants enjoyed the FreedroidRPG exergame version even more than the normal game. On average, they rated a 4.2 out of 5 for the overall enjoyment of the regular FreedroidRPG game, compared to an average 4.6 for the exergame version of FreedroidRPG. The goal of most exergames is to use the gaming platform as an incentive for physical activity, and it is shown from the responses to Q.5 that the majority rated a 5 out of 5 in level of fatigue, with an average 4.8. It is also critical when designing a video game to ensure that the participants feel that it is accurate and reliable, for this we asked them to rate their perceived accuracy of the system in detecting their activity and mapping it to the correct attribute in the avatar, and the majority felt the system accurately mapped their activities to attributes of the avatar in the game, with an average rating of 4.6 out of 5.

## VI. CONCLUSION

Tri-axis accelerometers are increasingly prevalent in wearable sensors to derive behavioral patterns and context from human motion. Subject-independent activity classification using such accelerometers is critical in developing systems that are robust and scalable, and perform well in a the real world setting. However, for activities with varying intensity levels, one needs to also analyze the ability of a system to detect activity independent of intensity. We show promise in our stochastic approximation framework in its ability to extrapolate unknown intensity levels from a few known intensity levels. Our framework is capable of outperforming other well-known algorithms, and also performs better using a GMM-clustering algorithm as opposed to a K-Means clustering algorithm. We verify the motivation of our framework by showing its ability to be used to enhance EE calculations by detecting the activity type independent of its intensity, and using a separate regression model for each intensity varying activity. We also demonstrate the potential of our framework when used in an exergaming environment. Our system will be useful in many healthcare applications including calculating EE and providing context awareness.

In the future, we can attempt to test our system on other intensity varying activities, such as bicycling, as well as perform testing on more subjects. We would also like to test our classifier with other clustering algorithms: HMM and Fuzzy C-means along with hierarchical clustering. We also hope to test our framework on individuals running and walking at a greater range of intensity levels. We would like to test classification frameworks that can accurately detect cheating in exergames to prevent participants from gaming the system.

## REFERENCES

[1] S. Patel, H. Park, P. Bonato, L. Chan, and M. Rodgers, "A review of wearable sensors and systems with application in rehabilitation," *J. NeuroEng. Rehabil.*, vol. 9, no. 1, pp. 1–17, 2012.

[2] P. S. Freedson, E. Melanson, and J. Sirard, "Calibration of the Computer Science and Applications, Inc. accelerometer," *Med. Sci. Sports Exerc.*, vol. 30, no. 5, pp. 777–781, May 1998.

[3] P. S. Freedson, K. Lyden, S. Kozey-Keadle, and J. Staudenmayer, "Evaluation of artificial neural network algorithms for predicting METs and activity type from accelerometer data: Validation on an independent sample," *J. Appl. Physiol.*, vol. 111, no. 6, pp. 1804–1812, Dec. 2011.

[4] T. Shiratori and J. K. Hodgins, "Accelerometer-based user interfaces for the control of a physically simulated character," *ACM Trans. Graph.*, vol. 27, no. 5, pp. 123:1–123:9, Dec. 2008.

[5] D. Jung, D. Lee, and H. Cho, "Implementation of realistic game controllers using three axis accelerometer and wireless communication modules," in *Proc. 23rd Int. Conf. Inf. Netw.*, 2009, pp. 152–155.

[6] F. Albinali, S. Intille, W. Haskell, and M. Rosenberger, "Using wearable activity type detection to improve physical activity energy expenditure estimation," in *Proc. 12th ACM Int. Conf. Ubiquit. Comput.*, 2010, pp. 311–320.

[7] N. Ravi, N. D, P. Mysore, and M. L. Littman, "Activity recognition from accelerometer data," in *Proc. 17th Conf. Innovat. Appl. Artif. Intell.*, 2005, pp. 1541–1546.

[8] J. Yang, "Toward physical activity diary: motion recognition using simple acceleration features with mobile phones," in *Proc. 1st Int. Workshop Interact. Multimedia Consum. Electron.*, 2009, pp. 1–10.

[9] E. M. Tapia, S. S. Intille, W. Haskell, K. Larson, J. Wright, A. King, and R. Friedman, "Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart monitor," in *Proc. IEEE 11th Int. Symp. Wearable Comp.*, 2007, pp. 37–40.

[10] N. Amini, M. Sarrafzadeh, A. Vahdatpour, and W. Xu, "Accelerometer-based on-body sensor localization for health and medical monitoring applications," *Pervas. Mob. Comput.*, vol. 7, no. 6, pp. 746–760, Dec. 2011.

[11] N. Alshurafa, W. Xu, J. J. Liu, M.-C. Huang, B. Mortazavi, M. Sarrafzadeh, and C. Roberts, "Robust human intensity-varying activity recognition using stochastic approximation in wearable sensors," in *Proc. IEEE Int. Conf. Body sensor Netw.*, 2013, pp. 1–6.

[12] D. A. Bader and R. Pennington, "Cluster computing: Applications," *Int. J. High Perform. Comput.*, vol. 15, no. 2, pp. 181–185, May 2001.

[13] X. Long, B. Yin, and R. Aarts, "Single-accelerometer-based daily physical activity classification," in *Proc. IEEE Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2009, pp. 6107–6110.

[14] S. L. Kozey, K. Lyden, C. A. Howe, J. W. Staudenmayer, and P. S. Freedson, "Accelerometer output and MET values of common physical activities," *Med. Sci. Sports Exerc.*, vol. 42, no. 9, pp. 1776–1784, Sep. 2010.

[15] S. E. Crouter, J. R. Churilla, and D. R. Bassett, "Estimating energy expenditure using accelerometers," *Eur. J. Appl. Physiol.*, vol. 98, no. 6, pp. 601–612, Dec. 2006.

[16] M. P. Rothney, E. V. Schaefer, M. M. Neumann, L. Choi, and K. Y. Chen, "Validity of physical activity intensity predictions by ActiGraph, Actical, and RT3 accelerometers," *Obesity (Silver Spring)*, vol. 16, no. 8, pp. 1946–1952, Aug. 2008.

[17] D. Hendelman, K. Miller, C. Baggett, E. Debold, and P. Freedson, "Validity of accelerometry for the assessment of moderate intensity physical activity in the field," *Med. Sci. Sports Exerc.*, vol. 32, no. 9 (Suppl), pp. S442–449, Sep. 2000.

[18] B. E. Ainsworth, W. L. Haskell, S. D. Herrmann, N. Meckes, D. R. Bassett, C. Tudor-Locke, J. L. Greer, J. Vezina, M. C. Whitt-Glover, and A. S. Leon, "Compendium of Physical Activities: A second update of codes and MET values," *Med. Sci. Sports Exerc.*, vol. 43, no. 8, pp. 1575–1581, Aug. 2011.

[19] H. Vathsangam, A. Emken, E. Schroeder, D. Spruijt-Metz, and G. Sukhatme, "Determining energy expenditure from treadmill walking using hip-worn inertial sensors: An experimental study," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 10, pp. 2804–2815, Oct. 2011.

[20] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *SIGKDD Explor. Newslett.*, vol. 12, no. 2, pp. 74–82, Mar. 2011.

[21] D. J. Miller and J. Browning, "A mixture model and EM-based algorithm for class discovery, robust classification, and outlier rejection in mixed labeled/unlabeled data sets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 11, pp. 1468–1483, Nov. 2003.

[22] S. Sun, M. Lan, and Y. Lu, "Adaptive EEG signal classification using stochastic approximation methods," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 31–Apr. 4 2008, pp. 413–416.

[23] V. Estivill-Castro, "Why so many clustering algorithms: A position paper," *SIGKDD Explor. Newslett.*, vol. 4, no. 1, pp. 65–75, Jun. 2002.

[24] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probability*, 1967, pp. 281–297.

[25] A. McCallum, K. Nigam, and L. H. Ungar, "Efficient clustering of high-dimensional data sets with application to reference matching," in *Proc. 6th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2000, pp. 169–178.

[26] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J. Roy. Statist. Soc. Series B (Methodology)*, vol. 39, no. 1, pp. 1–38, 1977.

[27] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[28] M. Zhang and A. A. Sawchuk, "A feature selection-based framework for human activity recognition using wearable multimodal sensors," in *Proc. 6th Int. Conf. Body Area Netw.*, 2011, pp. 92–98.

[29] W. Xu, M. Zhang, A. Sawchuk, and M. Sarrafzadeh, "Robust human activity and sensor location corecognition via sparse signal representation," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 11, pp. 3169–3176, Nov. 2012.

[30] J. O. Hill, H. R. Wyatt, G. W. Reed, and J. C. Peters, "Obesity and the environment: where do we go from here?," *Science*, vol. 299, no. 5608, pp. 853–855, Feb. 2003.

[31] S. E. Crouter, D. M. Dellavalle, J. D. Haas, E. A. Frongillo, and D. R. Bassett, "Validity of actigraph 2-regression model and matthews and NHANES and cut-points for assessing free-living physical activity," *J. Phys. Act Health*, vol. 10, pp. 504–514, May 2013.

[32] (2010)."Freedroidrpg is an open source role playing game," [Sourceforge], [Online]. Available: http://www.freedroid.org/

[33] A. P. O. S. Vermeeren, E. L.-C. Law, V. Roto, M. Obrist, J. Hoonhout, and K. Väänänen-Vainio-Mattila, "User experience evaluation methods: Current state and development needs," in *Proc. 6th Nordic Conf. Human-Comput. Interact.: Extend. Bound.*, 2010, pp. 521–530.

[34] J. L. G. SÃnchez, F. L. G. Vela, F. M. Simarro, and N. P. Zea, "Playability: Analysing user experience in video games," *Behav. IT*, vol. 31, pp. 1033–1054, 2012.

[35] R. Bernhaupt, W. Ijsselsteijn, F. F. Mueller, M. Tscheligi, and D. Wixon, "Evaluating user experiences in games," in *Proc. Extended Abstracts Human Fact. Comput. Syst.*, 2008, pp. 3905–3908.

[36] R. Bernhaupt and K. Isbister, "Games and entertainment community SIG: Shaping the future," in *Proc. Extended Abstracts Human Factors Computing Systems*, 2012, pp. 1173–1176.

[37] B. Mortazavi, K. C. Chu, X. Li, J. Tai, S. Kotekar, and M. Sarrafzadeh, "Near-realistic motion video games with enforced activity," in *Proc. 9th Int. Conf. Wearable Implantable Body Sensor Netw.*, 2012, pp. 28–33.

Authors', photographs and biographies not available at the time of publication.