5-15-2017

ActivityAware: Wearable System for Real-Time Physical Activity Monitoring among the Elderly

George G. Boateng
Dartmouth College

ActivityAware: Wearable System for Real-Time Physical Activity Monitoring among the Elderly

A Thesis

Submitted to the Faculty

in partial fulfillment of the requirements for the

degree of

Master of Science

by

George G. Boateng

Thayer School of Engineering

Dartmouth College

Hanover, New Hampshire

May, 15 2017

Dartmouth Computer Science Technical Report TR2017-824

Examining Committee:

Chairman________________________

(Ryan Halter, PhD)

Member________________________

(Kofi Odame, PhD)

Member________________________

(David Kotz, PhD)

____

F. Jon Kull, Ph.D.

Dean of Graduate and Advanced Studies

Abstract

Physical activity helps reduce the risk of cardiovascular disease, hypertension and obesity. The ability to monitor a person's daily activity level can inform self-management of physical activity and related interventions. For older adults with obesity, the importance of regular, physical activity is critical to reduce the risk of long-term disability. In this work, we present *ActivityAware*, an application on the Amulet wrist-worn device that monitors the daily activity levels (low, moderate and vigorous) of older adults in real-time. The app continuously collects acceleration data on the Amulet, classifies the current activity level, updates the day's accumulated time spent at that activity level, displays the results on the screen and logs summary data for later analysis.

The app implements an activity-level detection model we developed using a Linear Support Vector Machine (SVM). We trained our model using data from a user study, where subjects performed common physical activities (sit, stand, lay down, walk and run). We obtained accuracies up to 99.2% and 98.5% with 10-fold cross validation and leave-one-subject-out (LOSO) cross-validation respectively. We ran a week-long field study to evaluate the utility, usability and battery life of the *ActivityAware* system where 5 older adults wore the Amulet as it monitored their activity level. The utility evaluation showed that the app was somewhat useful in achieving the daily physical activity goal. The usability feedback showed that the *ActivityAware* system has the potential to be used by people for monitoring their activity levels. Our energy-efficiency evaluation revealed a battery life of at least 1 week before needing to recharge. The results

are promising, indicating that the app may be used for activity-level monitoring by individuals or researchers for epidemiological studies, and eventually for the development of interventions that could improve the health of older adults.

are promising, indicating that the app may be used for activity-level monitoring by individuals or researchers for epidemiological studies, and eventually for the development of interventions that could improve the health of older adults.

Acknowledgements

First, I'm grateful to my advisors, Professor Ryan Halter, Professor David Kotz and Dr. John Batsis for the opportunity to work on this thesis under their guidance and the assistance throughout the process. I'm grateful to Professor Kofi Odame for his mentorship throughout my academic journey at Dartmouth.

I'm grateful to Professor Douglas Van-Citters for the insightful discussions on accelerometry, Dr. Diane Gilbert-Diamond for the assistance with setting up and using the Actigraph device, Dr. Summer Cook for her input on the activity-level categorization, Alexandra Zagaria for assisting with setting up the user studies we run for this work and Justice Amoh for his advice on the machine learning components of this work.

I thank the members of the Amulet team, including Ron Peterson, Taylor Hardin, Patrick Proctor, Varun Mishra, Josiah Hester, Joseph Skinner, and Steven Hearndon. I'm grateful for the assistance with resolving issues with the Amulet device as I worked on this thesis.

I'm grateful to the volunteers that participated in the user studies without whom I would not have gotten data on which to perform my analysis.

Finally, I'm grateful to family and friends for their support throughout the journey of working on this thesis.

This research is supported by the National Science Foundation under award numbers CNS-1314281, CNS-1314342, CNS- 1619970, and CNS-1619950 and the National Institute On Aging under NIH Award Number K23AG051681.

Table of Contents

List of Tables

List of Figures

1 Introduction

Physical inactivity increases the risk for cardiovascular disease and chronic diseases such as diabetes, hypertension and obesity [1]. The prevalence of obesity continues to increase in Western societies, and with the aging of the population, an increasing number of older adults are classified as obese. Older adults with obesity who are sedentary are at higher risk of long-term disability, and physical activity in this population is critical to reducing their risk of functional impairment. The American College of Sports Medicine (ACSM) and the Centers for Disease Control (CDC) recommend 30 minutes of moderate intensity activity or 15 minutes of vigorous activity daily for adults, including older adults [2]. Hence, there is a need for a system that tracks the amount of time spent doing moderate or vigorous activities to encourage positive changes in behavior, which we believe will enable this population to achieve this important health goal and ultimately allow them to remain living independently in the community.

In this work, we developed *ActivityAware,* a wrist-worn, energy-efficient system that uses a lightweight machine-learning algorithm to monitor and encourage physical activity among older adults. Our *ActivityAware* app monitors the activity level of individuals in real time using acceleration data recorded from an Amulet, a low-power wrist-worn device [3]. The app continuously collects acceleration data, classifies the activity level of an individual, updates the day's accumulated time spent at that activity level, displays the results on the screen as feedback to the wearer, and logs the data for later analysis.

The app uses an implementation of a Support Vector Machine (SVM)-based machine-learning model to detect the activity level of a person. We developed this activity-level detection model using data from a study approved by the Dartmouth College Institutional Review Board (CPHS#28905). We collected acceleration data from younger and older volunteers who wore the Amulet as they performed various activities.

Our primary contribution is the development, implementation and evaluation of an open-source wearable system for real-time monitoring and encouragement of physical activity among older adults. Our secondary contribution is the development and implementation of an SVM-based activity-level model validated on older adults. Our tertiary contribution is a review of the current methods for physical activity monitoring using accelerometry and wearables.

In the remainder of this thesis, we describe the Amulet platform on which *ActivityAware* runs, our approach to physical activity-level categorization, and an overview of accelerometry in Section 2. We describe the components of *ActivityAware* and how we characterized the system in Sections 3 and 4 respectively. We describe our approach to developing the *ActivityAware* machine-learning model and the evaluation of the system in Sections 5 and 6 respectively. We describe limitations and future work in Section 7. We describe related work in Section 8 and conclude in Section 9.

2 Background

In this section, we describe the Amulet platform on which the *ActivityAware* app runs and why it is suitable for running the app. Then, we describe the categorization of the physical activity levels we use in this work. We also give an overview of accelerometry and its relation to activity monitoring.

2.1 Amulet Wearable Platform

The Amulet is an open-source hardware and software platform for writing energy- and memory-efficient sensing applications, which achieve long battery life [3]. The Amulet is a wrist-worn device that has two microcontrollers: an MSP430 running applications, and an nRF51822 for communicating with peripheral Bluetooth Low Energy (BLE) devices such as a heart-rate monitor and a galvanic skin response sensor (Figure 1,2)

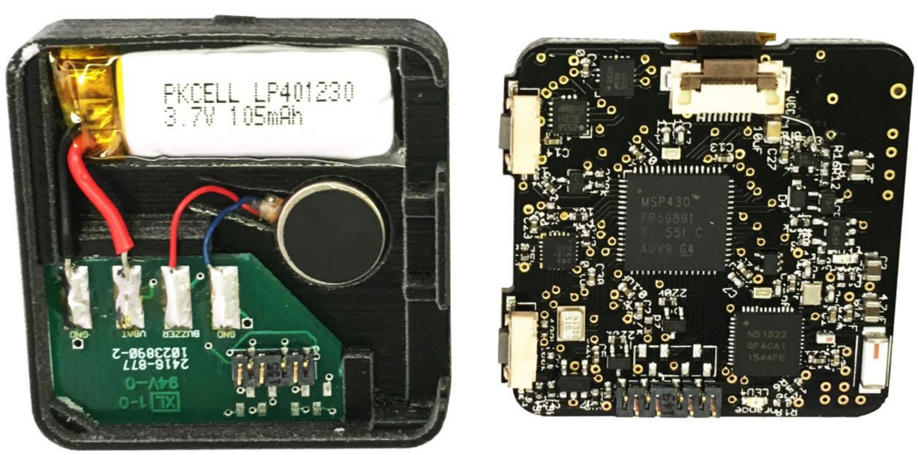

Figure 1: Internal Amulet peripherals (left), custom Amulet circuit board (right)

Figure 2: Amulet prototypes running various apps: heart rate app(left), EMA app (middle) and clock app (right)

It has built-in sensors to measure acceleration, rotation, ambient sound, ambient light, and ambient temperature. It has two buttons, a capacitive touch slider, a battery, a haptic buzzer, two LEDs, a micro-SD card reader, and a low-power display. The energy-efficient Amulet platform is useful for creating and running mHealth applications that monitor the physiological and behavioral health of its wearer, often lasting weeks before needing to recharge.

2.2 Physical Activity Level Categorization

Physical activity levels are defined using the Compendium of Physical Activities, which capture the intensity of activities expressed in metabolic equivalents

(METs): 1 MET corresponds the metabolic rate obtained during quiet sitting [4]. According to the CDC guidelines, activities can be categorized into *low, moderate* and *vigorous* based on METs [5]. *Low* corresponds to activities with METs less than 3 (e.g., sit, stand, lay down), *moderate* corresponds to activities with METs between 3 and 6 (e.g., walking at a moderate pace, walking fast), and *vigorous* corresponds to activities with METs greater than 6 (e.g., running) [5]. In this work, we use these example activities to categorize our activity levels.

2.3 Accelerometry for Physical Activity Monitoring

Accelerometers have been used as an objective measure of physical activity because of their ability to capture the intensity, duration and frequency of human movement [6]. An accelerometer captures the acceleration of objects along each of its axes (Figure 3).

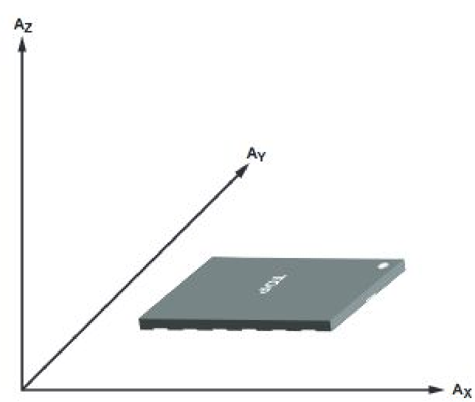

Figure 3: An accelerometer shown along its axes. Source: http://www.analog.com/media/en/technical-documentation/data-sheets/ADXL362.pdf

Acceleration values are measured in *g*s or milli*gs* (m*g*), where 1 *g* corresponds to the acceleration due to gravity (9.8m/s^2). Various features can be derived from the raw acceleration values to describe the physical activity of a person. Accelerometers are worn on various parts of the body such as the waist, wrist and ankle when used for physical-activity monitoring.

3 Overview of System: *ActivityAware*

ActivityAware is an Amulet application that measures the daily activity levels of individuals (low, moderate and vigorous). The app continuously collects acceleration data, classifies the activity level, updates the day's accumulated time spent at that activity level, logs the data for later analysis, and displays the results on the screen as feedback to the wearer. The app consists of four components: data collector, activity-level detector, activity-level monitor, and activity-level display (Figure 4).

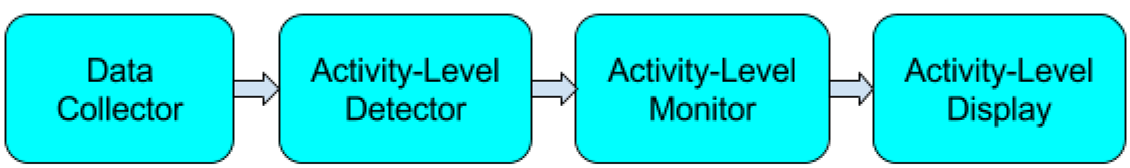

Figure 4: Components of *ActivityAware* App

3.1 Data Collector

The data collector samples data from a 3-axis accelerometer (Analog Devices ADXL362, range: ±2g) at a frequency of 20Hz, and parses the data stream into

5-second windows. Previous studies have shown that a frequency of 20Hz is sufficient for capturing the frequency range of physical human activities for classifying activities [7].

3.2 Activity-Level Detector

The activity-level detector determines the activity level of the user. It computes a vector of features from each 5-second window of accelerometer data. This feature vector is then fed to the activity-level classifier that determines the activity level as low, moderate or vigorous. We describe the selection and implementation of this classifier in a later section. Before performing a classification, the app checks whether the Amulet is being worn, in which case the app skips the classification operation to conserve energy. Also, this check ensures that the system does not accumulate minutes of low activity (which is the mostly likely level that will be classified) when the Amulet is instead not being worn. To infer whether the Amulet is being worn, we assume that the Amulet is unworn when it is still, which we infer when there is low variability in the acceleration data. This approach is an approximation to assess whether the Amulet is being worn, but was the best option since the Amulet does not have a dedicated sensor for detecting skin contact. We use a threshold of the variance of the magnitude of the acceleration values. If the variance for that 5-second time window is below the threshold, we set the non-wear state to be true and then skip the classification operation. To develop this threshold, we first recorded acceleration values with the Amulet placed flat on a table (non-wear acceleration values). We then computed the variance of 5-sec time windows of the non-wear acceleration values and selected a threshold corresponding to the 75th percentile value. We picked this value rather than the maximum since there were some variance values of low activities between the 75th percentile and the maximum. Hence selecting the 75th percentile would reduce the likelihood of low activities being determined as non-wear states.

3.3 Activity-Level Monitor

The activity-level monitor is responsible for keeping track of the number of minutes spent per day, for each of the three activity-level categories. This component tracks two data points for each activity level and non-wear state: t*otal minutes today* and *total minutes over all days* (*all days* refers to the set of days since the app was started). The value for each of these data points is updated after each classification result, and the *total minutes today* is reset at midnight each day.

This component logs summary information every hour to a microSD card inserted into the Amulet. Specifically, it logs date, time (hour, minute and second), battery level (ADC value and percentage), and total minutes spent at each of the activity levels and non-wear state. This logged data can be used to analyze the activity patterns of individuals during epidemiological studies.

This component also sets a daily activity goal and tracks progress towards this goal. The current implementation uses CDC's recommendation of 30 minutes of moderate activity or 15 minutes of vigorous activity as the daily goal. We also implement an equation that counts 1 minute of vigorous activity as 2 minutes of moderate activity towards the goal based the CDC's recommendation of the minutes for either moderate or vigorous activity:

$$y = mod + 2*vig$$

where y is the result that is compared against the 30 minutes, *mod* is the amount of moderate minutes today and *vig* is the amount of vigorous minutes today.

The user receives three encouragement alerts daily at 12pm, 3pm and 6pm based on the progress made (Figure 5). The mode of this alert is via buzzing of the Amulet and displaying a red LED, which stays on for 5 seconds. When the user has achieved less than 33% of the goal, the alert message says "You can do it". When the user has achieved between 33% and 66% of the goal, the encouragement alert says "Keep at it". When the user has achieved between 66% and 99%, the encouragement alert says "Almost there." Once the goal is achieved, the Amulet buzzes, turns on a green LED for a few seconds, and displays an alert message "Goal Achieved!". No alert is given if the goal has already been surpassed.

Figure 5: Encouragement alerts of the *ActivityAware* app

3.4 Activity-Level Display

The activity-level display component displays information about the progress made towards the daily activity goal tracked by the activity-level monitor. The display presents the progress pictorially and numerically in 3 ways: percentage, progress bar and number of minutes left for either moderate or vigorous activity (Figure 6).

Figure 6: Modes of the *ActivityAware* app

4 Characterization of System

We performed various experiments to characterize the noise and power draw of the *ActivityAware* system. We describe our characterization in this section.

4.1 Noise Characterization

We sought to characterize the noise of the system, and estimate the signal-to-noise ratio (SNR). We recorded acceleration data while the Amulet lay flat on a table. We then computed the variance of 5-second non-overlapping windows of the acceleration magnitude, which corresponds to the noise power of the signal. We repeated this process for the three activity levels of all the older adult dataset. We then created a boxplot to compare the power of the noise and the power of the activity levels signal (Figure 7). The box plot shows noticeable difference between the noise power and the moderate and vigorous activity power. There difference between the noise and low activity is not obvious from the scale of the boxplot. We assessed the difference quantitatively by estimating the SNR of each of the three activity levels. We did this by computing the ratio of the average of the activity power values and the average of the noise power values. We got SNR values of 14 dB, 32dB and 40dB for low, moderate and vigorous activity respectively. Our minimum SNR of 14 dB is not high and hence low-level activities might be difficult to distinguish from noise. On the other hand, the moderate and vigorous activities have SNR values 32dB and 40dB, and hence those signals can be adequately distinguished from the noise of the system.

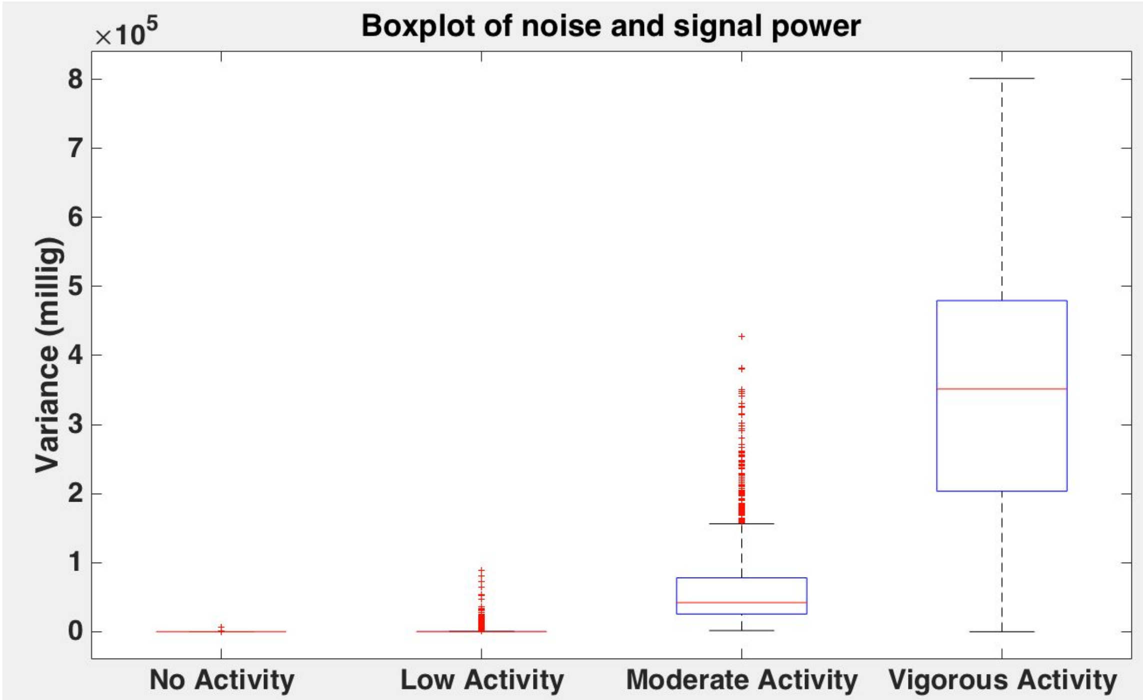

Figure 7: Boxplot of noise and activity power

4.2 Power Characterization

We estimated the power draw of various computational modes of the *ActivityAware* app. We stepped through the various modes of the app and used an oscilloscope to measure the voltage across a 50 ohm resistor connected in series with the Amulet's circuitry (Figure 8).

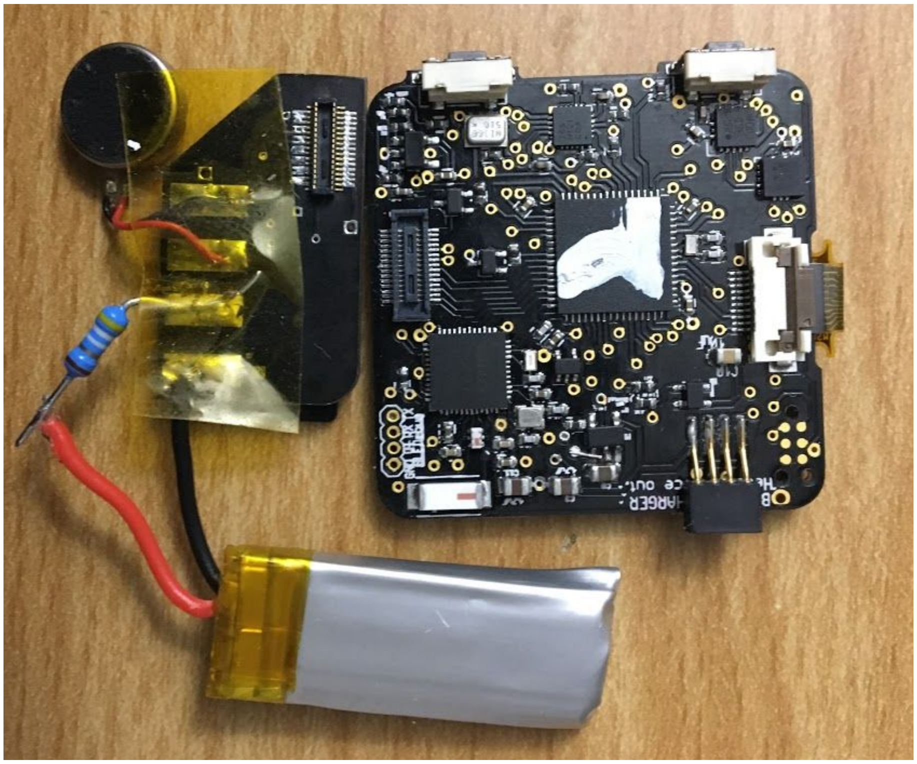

Figure 8: Circuit for power draw measurement

We summarize the power-draw measurements in Table 1. The table shows that the most power hungry operations were log and double buzz, which draw 22.5 *mA* and 19.6 *mA* respectively. The least power hungry operations are the button tap and operating system, which draw 0.6 *mA* and 0.42 *mA* respectively. We calculated the average current in the system with the equation:

$$mAavg = \frac{\sum curr(i)dur(i)}{time}$$

where *mAavg* is the average current in *mA, curr(i)* is the current of each computational mode in *mA*, *dur(i)* is the duration of each computational mode per day in *ms,* and *time* is the number of milliseconds in a day. Our estimate shows that the system draws an average of 0.67 *mA*. We then estimated the battery life for the 110mAh battery in the Amulet. We estimate that the system can run for 6.9 days (165.4 hours) before needing to be recharged. This estimate is based on the various components of the app being completely used and hence corresponds to a lower bound on the battery life. As a result, the battery life could be longer depending on how the app gets used. For example, the less the Amulet gets worn, the more the app skips the classification operation as mentioned in section 3.3, which results in a longer battery life. Also, if the user achieves the daily activity goal before 12pm, the user does not receive the three encouragement alerts consisting of double buzz, red LED and alert display, which will result in a longer battery life.

Table 1: Summary of power draw measurements

Mode	Current (mA)	Duration (ms)	No of times / day	Description
Getting Acceleration	1.24	140	86400	Every second in a 24 hour period
Feature Extraction	1.33	310	17280	Every 5 sec in a 24 hour period
Classification	1.20	16	17280	Every 5 sec in a 24 hour period - duration of nonwear (8 hours)
Display	1.53	170	32	Maximum of 30 times for 30 minutes of moderate activity + start display + midnight update
Log	22.49	47	24	Once per hour
Button tap	0.60	225	1	Once when the app is started
Alert display	0.70	176	3	3 times a day
Alert Double Buzz	19.58	500	4	4 times a day
Alert Red LED	3.01	5,000	3	3 times a day
Alert Green LED	3.01	5,000	1	Once when the goal is achieved
Operating System Interrupts	0.76	200	86400	Once every second
Operating System	0.42			Total time in day - sum of time in other modes

5 Activity Level Detection Model - Machine Learning

We developed an activity-level detection model using a common machine learning algorithm, Support Vector Machine (SVM). SVM is a classifier that constructs a high-dimensional hyperplane and uses it to perform classification [8]. SVM chooses a hyperplane that maximizes distance to the nearest points on the either side of the plane for the binary classification case (Figure 9).

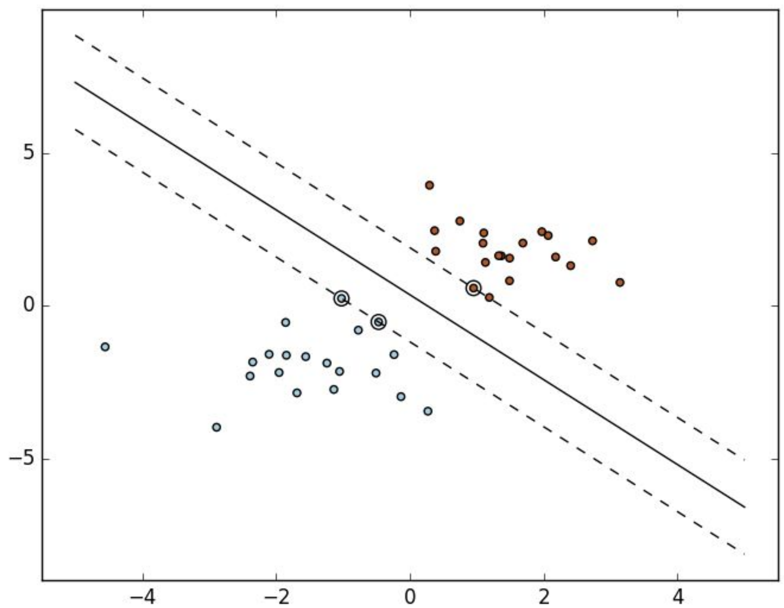

Figure 9: Hyperplane separating two classes in SVM [9]

We use SVM because it uses a subset of the training set – “the support vectors” – for its prediction function. Models like k-nearest neighbor (kNN), on the other hand, need to store all the data points in memory for prediction. SVM is more memory efficient and thus well suited for low-memory platforms like the Amulet.

We trained a linear SVM model to distinguish low, moderate, and vigorous activity levels using the scikit-learn library [9]. We use scikit-learn's default parameters for the linear SVM model.

5.1 Data Collection

We collected data from volunteer subjects under a study protocol approved by Dartmouth's Institutional Review Board. All individuals completed a basic baseline demographic questionnaire that assessed age, gender, race, height, weight and handedness (left or right). All data was collected online via Research Electronic Data Capture software (REDCap) into a centralized, HIPAA compliant repository. REDCap is a secure, web-based application designed to support data capture for research studies, providing 1) an intuitive interface for validated data entry; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for importing data from external sources.

5.1.1 Activity Data Collection App

We developed an app similar to *ActivityAware* for the purpose of collecting data from the study. The app has three states: Ecological Momentary Assessment (EMA), Data Collection, and Data Logging (Figure 10).

Figure 10: States of Activity Data Collection App

The app begins in the EMA state. Within this state, the user selects which activity they are about to perform from a list of activities using the capacitive-touch slider on the Amulet (Figure 11). After the user selects the specific activity and presses the button on the Amulet, the app switches to the data collection state.

Figure 11: EMA state (left), Data collection state (middle), Logging state (right)

In the data collection state, the app collects and stores acceleration data from a 3-axis accelerometer with range ±2g at a frequency of 20 Hz. We discard the first 5 seconds of data. After a specified time duration (either 1 or 2 minutes), the app switches to the data-logging state in which it logs the collected acceleration data along with the activity level onto a micro-SD card on the Amulet.

The app then switches back to the EMA mode to allow the user to select the next activity to perform. We accompanied the subjects when they performed the activities so we could ensure they completed all activities correctly and the appropriate number of times.

5.1.2 Study Protocol

We collected acceleration data from 29 subjects (n=29) as they performed various physical activities. We had 2 cohorts: younger adults (n=14) and older adults (n=15). The younger adults were college students 18–23 years old and the older adults were all above 65 years old. For the younger adults, we collected data from them at Dartmouth College's Alumni Gymnasium. For the older adults, we collected data from them at the Dartmouth-Hitchcock Aging Resource Center. Subjects wore the Amulet on their left wrist, irrespective of their hand dominance, and performed each of the following activities for a duration from the range 1 to 10 minutes as the Amulet ran the Activity Data Collection App: sit, stand, lay down, walk at a regular pace, walk fast and run (Figures 12, 13). The plots show that the run activity has the most variability, followed by walk fast and walk moderate, and sit, stand, and lay down. We collected data using the Amulet placed on the same wrist to ensure the data is is consistent since the orientation of the accelerometer with reference to the wrist changes when switched between wrists. Four older adults were unable to perform the run activity and as a result we had no running data from them.

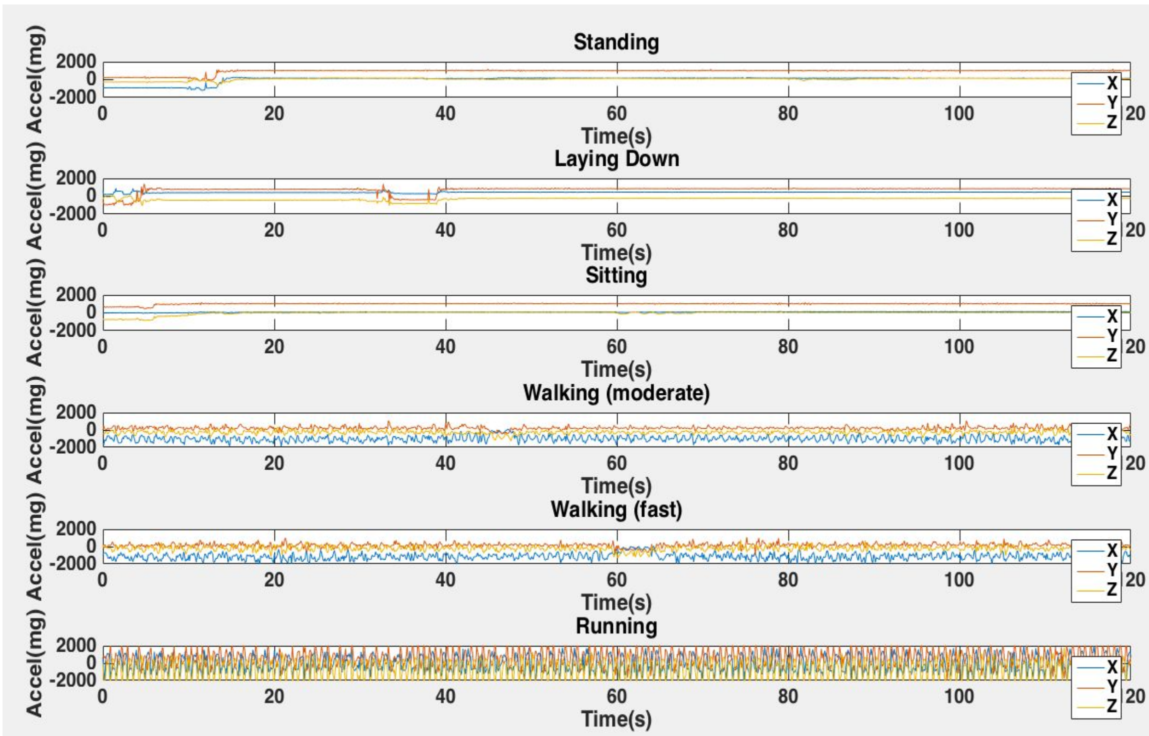

Figure 12 : Plots of acceleration data from one younger subject for each of the 6 activities over a 1-minute period

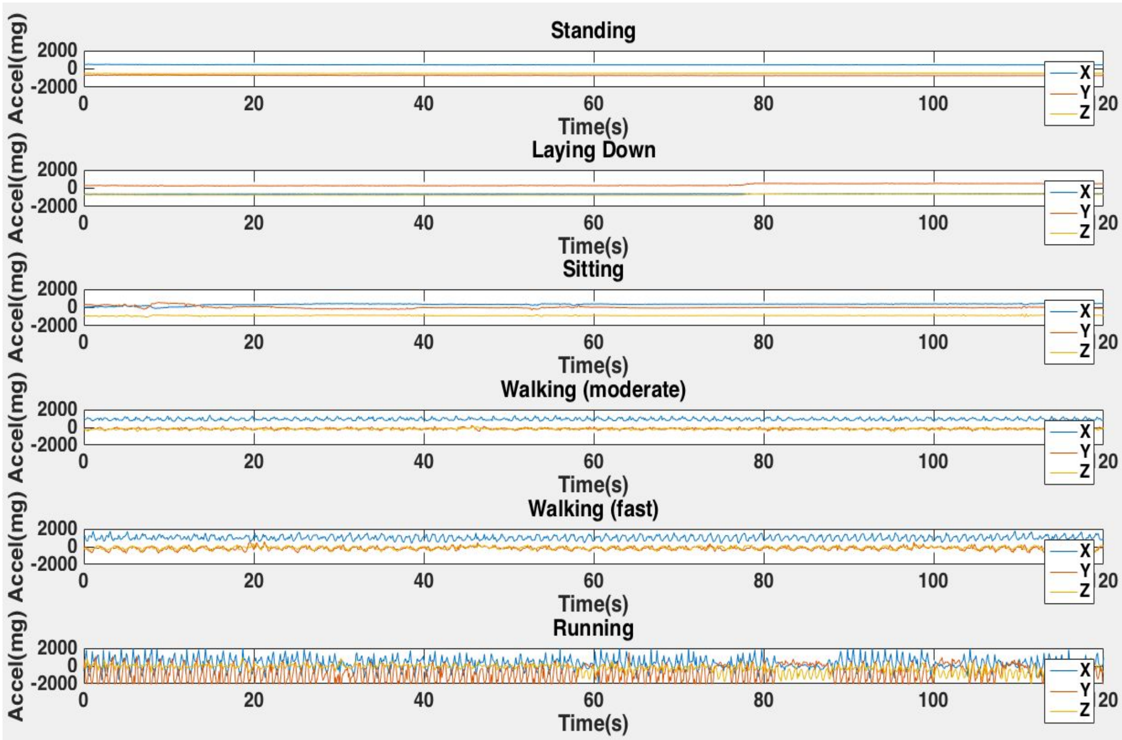

Figure 13 : Plots of acceleration data from one older subject for each of the 6 activities over a 1-minute period

We had 1282 minutes of data in total (younger - 447 minutes, older - 835 minutes). We categorized the data from these 6 activities into the following classes: low (sit, stand and lay down); moderate (walk at a regular pace and walk fast); and vigorous (run). We then split the data into 5-second non-overlapping time windows that previous studies have shown to be suitable for activity classification [10].

5.2 Feature Extraction

From each 5-second window of each subject's data, we extracted 6 temporal and 6 spectral features from the (*x, y, z axes*) and magnitude of the acceleration vector that previous studies have shown to be relevant for activity detection [7][10][11][12]. We had a total of 2x6x4=48 different features (Table 2,3). To compute the frequency-based features, we first computed the discrete Fourier transform (DFT) of the signals using the Fast Fourier Transform (FFT) algorithm. The result of the feature extraction was a training dataset containing 10,018 and 5,364 feature vectors for the younger and older-adult datasets respectively.

Table 2: Description of temporal features

Features	**Description**
Mean	Sum of values divided by total number of values
Median	Middle value of sorted values
Range	Difference between maximum and minimum of values
Interquartile range	Difference between 75th and 25th percentiles of values
Standard deviation	Square root of average square difference of values from mean
Root mean square	Square root of sum of square of values

Table 3: Description of spectral features

Features	**Description**
Energy	Sum of the squared DFT component magnitudes of the signal normalized by window length
Dominant frequency	Frequency value corresponding to the maximal spectral coefficient between 0.6 and 2.5 Hz
Dominant power	Maximal spectral coefficient between 0.6 and 2.5 Hz
Power ratio	Dominant power divided by total energy
Coefficients sum	Sum of coefficients from 0.5 Hz to 3 Hz
DC value	First coefficient in DFT

5.3 Training and Evaluation of Models

We used all 48 features in our experiments. We trained different models and ran various experiments to evaluate the models. We used the following metrics: accuracy, confusion matrix, precision, recall and F1-score, which have been used in previous studies [7][10]. TP refers to true positives, TN refers to true negatives, FP refers to false positives, and FN refers to false negatives. *Accuracy* is the percentage of correctly classified data, computed as follows: $\frac{TP+TN}{TP+TN+FP+FN}$. Precision tells what percentage of the positively predicted class was correctly classified, computed as follows: $\frac{TP}{TP+FP}$. Recall tells what percentage of the positively labeled class is classified correctly, computed as follows: $\frac{TP}{TP+FN}$. F1-score is the harmonic mean of the precision and recall, computed as follows: 2 $\frac{precision * recall}{precision + recall}$. We performed our evaluation using 10-fold cross-validation (stratified), hold-out validation, and leave-one-subject-out (LOSO) cross-validation.

With 10-fold cross-validation, the dataset is divided into 10 equal parts with 9 parts used for training and the remaining 1 used for testing. This process is repeated 10 times with each part being used for testing once. The metrics described above are computed at each iteration and then averaged at the end. We say the process is stratified because each part contains the same ratio for all the classes as there are in the whole dataset.

With hold-out cross validation, the dataset is divided into 2 parts: training and testing datasets. This division is based on a specified criteria and the parts do not have to have the same number of samples.

With LOSO, training is done on the dataset from all subjects but 1 whose dataset is used for testing. This processes is repeated as many times as there are subjects ensuring that each subject's dataset is used as the test dataset once.

5.3.1 Testing and Results I: Various Datasets

We trained and evaluated three models: one using the younger adult dataset only, another using the older adult dataset only, and one using both datasets. We evaluated each model using 10-fold and LOSO cross validation with the corresponding dataset from which the model was developed. The results for the 10-fold cross validation are better than LOSO (Table 4). This result is expected since for 10-fold cross validation, a subject's data might be in both the train and test dataset resulting in a better performance. Thus, we consider LOSO a more rigorous evaluation metric. The results show that the younger adult model performed better than the older adult model and both model (Table 4).

Table 4: Classification results of various datasets

Data	**Accuracy**		**Precision**		**Recall**		**F1-score**	
	LOSO	**10-fold**	**LOSO**	**10-fold**	**LOSO**	**10-fold**	**LOSO**	**10-fold**
Younger	98.5%	99.2%	98.6%	99.2%	98.5%	99.2%	98.4%	99.2%
Older	94.1%	96.4%	94.4%	96.4%	94.2%	96.4%	93.4%	96.4%
Both	94.3%	96.5%	95.6%	96.7%	94.3%	96.5%	93.3%	96.5%

A further analysis of the result using the confusion matrices in Table 5 and 6 (*A* corresponds to actual classes and *P* corresponds to predicted classes) show that the older adult model misclassified 25% of vigorous activities as moderate compared to the 5.4% misclassification of the corresponding case in the younger adult model. This result may be due to the fact that unlike the younger adults, the older adults did not perform the vigorous activity with intensities that were much different from the moderate activities. In fact, some older adults struggled to run and as result their running activity looked like walking fast. Also, as mentioned earlier, 4 older adults could not run, which is an example of older adults' struggle running. Their running data was not collected and is thus not part of either the training or testing datasets. These points could explain the misclassification.

Table 5: Confusion matrix of younger model using LOSO

	Low (P)	**Mod (P)**	**Vig (P)**
Low (A)	99.6%	0.4%	0.0%
Mod (A)	1.1%	98.7%	0.2%
Vig (A)	0.2%	5.4%	94.4%

Table 6: Confusion matrix of older model using LOSO

	Low (P)	Mod (P)	Vig (P)
Low (A)	98.6%	1.4%	0.0%
Mod (A)	1.9%	95.8%	2.3%
Vig (A)	2.0%	25.0%	73.0%

The result using both datasets performed slightly better than the older adult model (Table 7). This result might suggest that we should use an activity-level detection model trained on data from both older and younger adults rather than on data from only older adults since it is easier to get data from younger adults. However, the results might be inflated due to the larger amount of data from younger adults (almost twice that from older adults).

Table 7: Confusion matrix of both model using LOSO

	Low (P)	Mod (P)	Vig (P)
Low (A)	99.0%	0.9%	0.1%
Mod (A)	2.8%	95.3%	1.9%
Vig (A)	0.9%	18.5%	80.6%

5.3.2 Testing and Results II: Train on One Dataset & Test on the Other

We performed an experiment to find out how well a model trained on data from only younger adults would perform when tested on a dataset from older adults and vice versa. We used hold-out cross validation in which the younger adult dataset was used as the training dataset and the older adult dataset was used as the testing dataset, and vice versa. Our results (Table 8) show that the model trained on the older adult dataset and tested on the younger adult dataset performed better.

Table 8: Classification results from training on one dataset and testing on the other

Dataset		**Metrics**			
Train	**Test**	**Accuracy**	**Precision**	**Recall**	**F1-score**
Younger	Older	74.9%	77.3%	74.9%	69.7%
Older	Younger	88.0%	90.7%	88.0%	88.3%

A further analysis using the confusion matrix shows that the younger adult model misclassified 95% of older adults' vigorous activities as moderate and 35% of older adults' moderate activities as low (Table 9). This result is expected since the older adults performed vigorous and moderate activities with much less intensity than younger adults, explaining the poor results.

Table 9: Confusion matrix of younger adult model tested on older adult dataset

	Low (P)	Mod (P)	Vig (P)
Low (A)	99.7%	0.3%	0.0%
Mod (A)	35.1%	64.9%	0.0%
Vig (A)	2.3%	94.9%	2.8%

The older adult model misclassified 29% of younger adults' moderate activities as vigorous (Table 10). Again, this result is not unexpected and corroborates the intuition that activities that are moderate intensity for younger adults might in fact be vigorous for older adults.

Table 10: Confusion matrix of older adult model tested on younger adult dataset

	Low (P)	Mod (P)	Vig (P)
Low (A)	98.4%	1.1%	0.5%
Mod (A)	1.0%	70.5%	28.5%
Vig (A)	0.0%	7.9%	92.1%

5.3.3 Testing and Results III: Different Feature Sets (Older Adults)

We performed an experiment to compare different feature subsets and evaluate their performance. We performed this evaluation using only the older adult datasets to aid in picking a small number of features that work best for older adults – who are, after all, the target population for the *ActivityAware* system. We used only LOSO cross validation for this evaluation since it is a better reflection of how well the model will perform on a new subject. The feature subsets along with the total number of features and results are shown in Table 11.

Table 11: Classification results using various feature sets with LOSO

Feature Sets	**No of Features**	**Accuracy**	**Precision**	**Recall**	**F1-score**
All	48	94.1%	94.4%	94.2%	93.4%
Magnitude All	12	91.2%	91.4%	91.2%	90.5%
Temporal	24	94.3%	94.7%	94.3%	93.6%
Magnitude Temporal	6	93.9%	95.8%	93.9%	93.7%
Spectral	24	92.1%	93.8%	92.0%	91.3%
Magnitude Spectral	6	92.4%	93.2%	92.5%	91.5%

The first key observation is that all 48 features are not necessary to have good performance. In fact, the temporal features consisting of 24 features outperform the 'All' feature set in all the metrics and the 'Magnitude Temporal' feature set with only 6 features has comparable results. A surprising result is that the spectral features did not perform better than the temporal features, despite their computational complexity. This result suggests that it is not necessary to use spectral features, especially considering their complexity if implemented on a low-power device like the Amulet. Nevertheless, various subsets of spectral features – or others not included in this evaluation –might have better performance.

We used the recursive feature elimination (RFE) algorithm to select features within each of the features subsets mentioned above. RFE is a feature selection algorithm that recursively eliminates features based on the coefficients of a linear model that is initially trained on all the features [9]. This elimination process continues until the desired number of features to be selected has been reached. Our implementation uses the coefficients of the linear SVM for eliminating features. We selected features ranging from 1 to the maximum number of features within the feature subset. We then evaluated the performance of the selected features using LOSO and accuracy as the metric and plotted the results separately in Figure 14 and then all together in Figure 15. Overall, as features are removed, the accuracy decreases with some fluctuations.

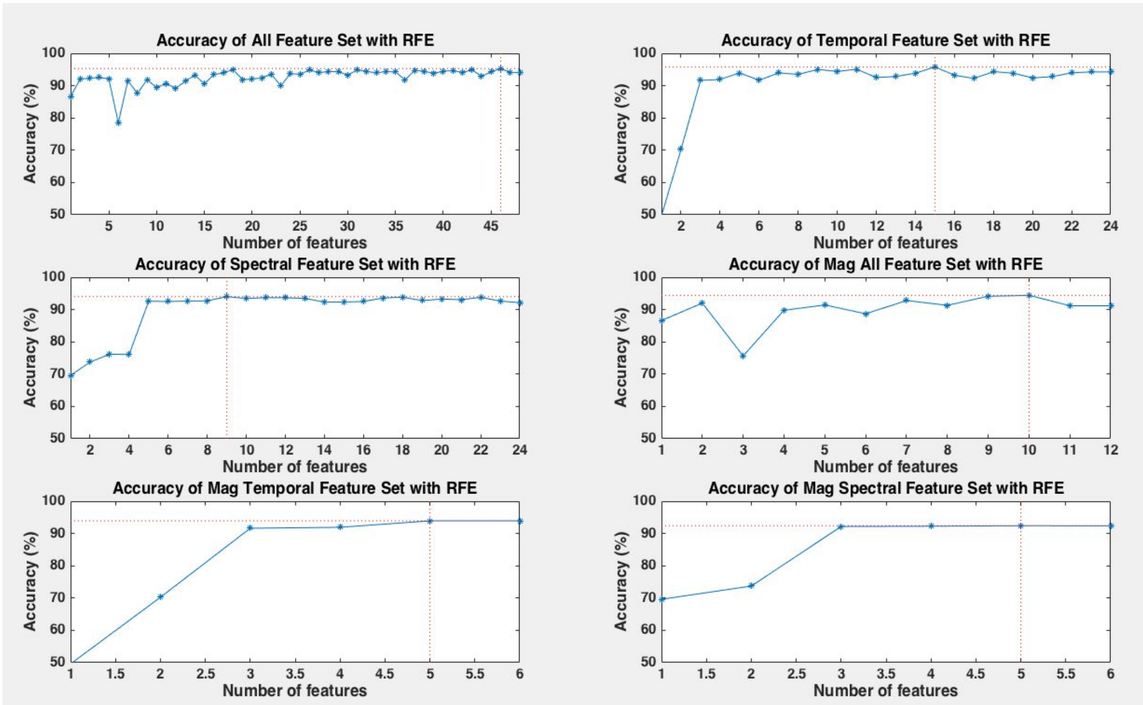

Figure 14: Accuracy of selected features in feature subsets (plotted separately)

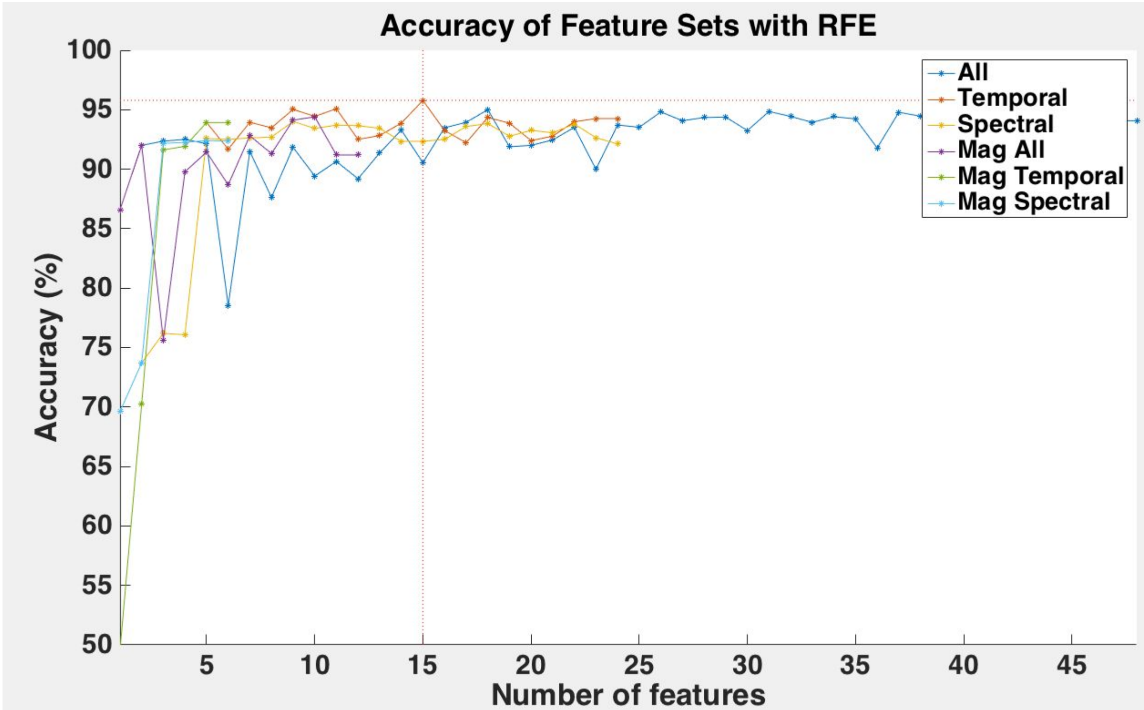

Figure 15: Accuracy of selected features in feature subsets (plotted together)

We then found the optimal number of features in each subset (features with maximum accuracy) and summarize the results in Table 12. The best performing feature set overall was the temporal feature subset having 15 features and accuracy of 95.8%. The features are as follows: standard deviation (*x, y, z, magnitude*), mean (*x, y, z, magnitude*), median (*x, y, z, magnitude*), interquartile range (*x, magnitude*), and root mean square (*magnitude*). We also found the smallest number of features that give an accuracy within 95th percentile of the maximum accuracy. There are 3 features within the temporal features subset that give an accuracy of 91.6%: root mean square (*magnitude*), mean (*magnitude*) and standard deviation (*magnitude*). This result shows that a small percentage drop in accuracy (4.2%) can be traded for significant decrease in features used (80%).

Table 12: Accuracy of optimal features selected using RFE

Feature Sets	Maximum Number	Selected Number	Accuracy
All	48	46	95.3%
Magnitude All	12	10	94.4%
Temporal	24	15	95.8%
Magnitude Temporal	6	5	93.9%
Spectral	24	9	94.0%
Magnitude Spectral	6	5	92.4%

5.4 Selection and Implementation of Activity Level Detection Model

We selected a model using only a subset of features and implemented the model on the Amulet. We chose a model that works best using the older adult dataset before we performed feature selection with RFE. We describe the model selection and implementation in this section.

5.4.1 Model Selection

We sought to pick a model with low computational complexity and good performance. The first choice we made was to use temporal features since they performed best and are less computationally intensive than the spectral features – all of which run in O(NlogN) time for computing the DFT components with the FFT algorithm.

The next choice was to eliminate temporal features that are computationally complex. Specifically, we eliminated features that run in time more complex than O(N). The two features that fit this criteria are median and interquartile range since they need the data to be first sorted before they are computed and sorting runs in O(NlogN) time. We then picked 2 of the remaining 4 temporal features (mean and standard deviation) and extracted the features from the *x, y, z* accelerations, and magnitude of the acceleration) resulting in an 8-feature vector.

We trained our linear SVM model with the older adults dataset and tuned the hyperparameters to improve the performance. We did this by trying various combinations of scikit-learn's linear SVM parameter options. Using LOSO, our best model had an accuracy of 91.7%, precision of 93.2%, recall of 91.6%, and F1 score of 91.5% with the following parameters: C=100, penalty = 'l1' and dual=false. An analysis of the confusion matrix (Table 12) shows that this model would misclassify about 10% of moderate activity as vigorous and 21% of vigorous activity as moderate.

Table 13: Confusion matrix of chosen activity-detection model

	Low (P)	Mod (P)	Vig (P)
Low (A)	98.3%	1.7%	0.0%
Mod (A)	3.6%	86.8%	9.6%
Vig (A)	2.1%	21.3%	76.6%

These are significant misclassification percentages, which could lead to an overestimation or underestimation of the activity minutes of older adults. Further work is needed to obtain a model with fewer misclassifications and yet has minimal computational complexity.

5.4.2 Model Implementation

We implemented the model in the activity-level detector component of the *ActivityAware* app. The component computes the 8 features that were selected using each 5-second window of accelerometer data. This 8-feature vector is fed to the activity-level classifier, which is an implementation of the decision function of a Linear SVM:

$$y = wx + b$$

Here, y is the vector that holds the result of the evaluation for the three activity levels, x is the computed feature vector (number of features), w is the coefficient matrix (number of classes × number of features) and b is the intercept vector (number of classes). The values for w and b are obtained from the linear model that we train offline using the scikit-learn library. Because this is a multi-class classification, we implemented the "one-vs-the-rest" approach for multi-class classification since the scikit-learn Linear SVM function uses this method [9]. In this approach, one classifier is trained for each of the classes that correspond to each row in the matrix w. The result of solving the equation is a vector y that contains a value for each of the three classes. The class with the maximum value is the predicted class.

6 Evaluation of System

We evaluated the *ActivityAware* system by running a week-long field study and analyzing whether the system was useful in achieving the CDC's recommended daily activity goal (utility), whether the system was easy to use (usability), and how long the battery might last before needing to be recharged (energy efficiency). We describe our evaluation below.

6.1 Field Study

We ran a five-day field study in which five older adults (ages: 73, 73, 83, 86 and 87 years) each wore an Amulet as it monitored their activity level. The app tracked how much time they spent doing low, moderate, or vigorous activity, and the duration the Amulet spent in a non-wear state. The app also tracked battery life and logged a summary of this information hourly for later analysis. The app displayed to subjects how close they were to achieving the daily activity goal and gave encouragement alerts 3 times a day as described earlier.

6.2 Utility Evaluation

We sought to determine whether the *ActivityAware* system was useful in helping older adults achieve the CDC's recommended daily activity goal. Specifically, we were interested in knowing whether the three displays of progress (percentage, progress bar, and number of minutes left for either moderate or vigorous activity) as well as the encouragement alerts helped to achieve the activity goal. We summarize the number of minutes per activity level for all five subjects (S1, S2, S3, S4, & S5) and for all 5 days (Table 14). We also include the time each subject achieved the activity goal for each day. An analysis of the activity data showed that all five subjects achieved the activity goal for all the five days (Table 14).

Table 14: Summary of activity log data

	Activity Data	**S1**	**S2**	**S3**	**S4**	**S5**
Day 1	Low	443 mins	403 mins	32 mins	154 mins	412 mins
	Mod	41	63 mins	4 mins	108 mins	110 mins
	Vig	10 mins	9 mins	24 mins	17 mins	3 mins
	Goal Reach Time	2PM	1PM	3PM	3PM	2PM
Day 2	Low	608 mins	497 mins	357 mins	618 mins	247 mins
	Mod	85 mins	95 mins	13 mins	68 mins	23 mins
	Vig	17 mins	10 mins	16 mins	7 mins	4 mins
	Goal Reach Time	12PM	12PM	10AM	2PM	5PM
Day 3	Low	650 mins	505 mins	118 mins	396 mins	327 mins
	Mod	94 mins	93 mins	4 mins	79 mins	38 mins
	Vig	25 mins	11 mins	13 mins	7 mins	6 mins
	Goal Reach Time	10AM	12PM	11AM	11AM	8PM
Day 4	Low	685 mins	470 mins	270 mins	340 mins	226 mins
	Mod	92 mins	62 mins	8 mins	125 mins	38 mins
	Vig	41 mins	8 mins	21 mins	7 mins	6 mins
	Goal Reach Time	11AM	3PM	11AM	10AM	11AM
Day 5	Low	465 mins	539 mins	505 mins	570 mins	340 mins
	Mod	59 mins	117 mins	39 mins	93 mins	55 mins
	Vig	36 mins	13 mins	31 mins	19 mins	5 mins
	Goal Reach Time	11AM	1PM	9AM	10AM	4PM

To find out whether the display and alerts were helpful, we looked at the corresponding usability questionnaire questions (rated on 'Strongly Disagree - 1' to 'Strongly Agree - 5'). Three people selected "Agree" to the statement "The display of progress (progress bar, percentage and minutes left) was useful in achieving the activity goals" whereas the remaining two selected "Neutral" (Figure 16). The results provide preliminary evidence that suggests that the display about progress was somewhat helpful in achieving the daily goal. Additionally, the subjects in their written and verbal feedback mentioned that they liked seeing the values and progress bar change on the screen as they performed various activities. The current implementation of the app stopped updating the display of the progress once the goal was achieved, to conserve battery life. Subjects suggested that the display would have encouraged them to perform more activity if it kept updating even after the goal was achieved.

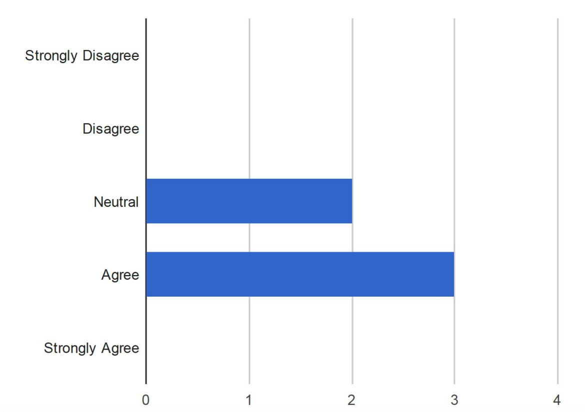

Figure 16: Questionnaire response about usefulness of progress display

In response to the statement “The daily encouragement alerts were useful in achieving the activity goals”, only one person selected “Agree” with two being neutral, and the remaining two split between disagree and strongly disagree (Figure 17).

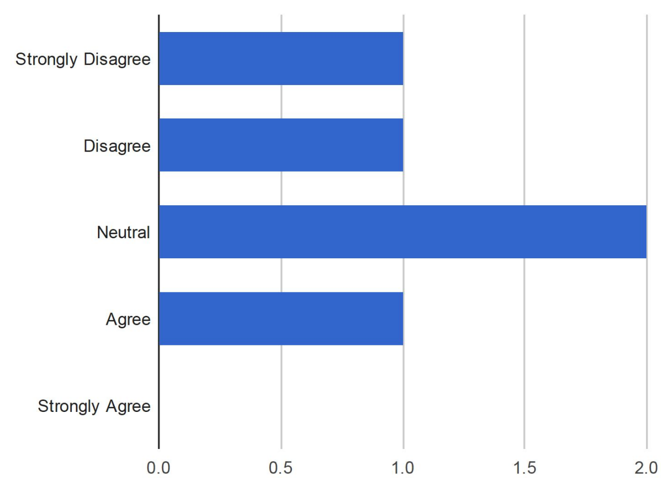

Figure 17: Questionnaire response about usefulness of encouragement alerts

This result seems to suggest that the encouragement alerts were not particularly useful. The subjects mentioned that the goal was very easy to achieve and hence not challenging. The data shows that most subjects achieved the goal by morning or early afternoon. As a result, they did not get the encouragement alerts, which explains why they did not find them useful. There are three possible reasons for the easiness with achieving this goal.

First, conversations with the five subjects revealed that they are generally active. In fact, two of the subjects were recruited from a bi-weekly exercise class. The activeness of this subject population could have contributed to the easiness in achieving this goal.

Second, the app in tracking progress towards the daily activity goal does not take into consideration the CDC's additional recommendation that activities should be done for at least 10 continuous minutes. The current implementation of the app just accumulates time intervals, which may just be sporadic activities, which could have contributed to the ease with which subjects achieved the daily activity goal.

Lastly, misclassifications of activity levels could have contributed to the easiness of achieving this goal. As was noted in section 5.4.1, the activity-level detection model misclassifies about 10% of moderate activity as vigorous. Because the app counts 1 minute of vigorous activity as equivalent to 2 minutes of moderate, such misclassifications could have contributed to the easiness of achieving the goal.

Further experiments will need to be conducted using a more challenging goal or a less active subject population to adequately evaluate the usefulness of the encouragement alerts. Also, the feedback obtained suggests that subjects should be given the chance to adjust their activity goal to make it more challenging if needed. Additionally, the app could automatically adjust the goal based on the user's activity pattern.

Finally, we note that our field study involved only five subjects, from which we can only draw preliminary observations and no significant conclusions – further field studies are planned.

6.3 Usability Evaluation

We sought to determine whether the *ActivityAware* system is easy to use and whether older adults might be willing to use it for monitoring their activity or during epidemiological studies. We asked subjects to react to various statements pertaining to the usability of the system and summarize the mean responses in Table 15 and Table 16.

Table 15: Summary of Usability Questionnaire for Positive Statements

	Survey Statements (Positive)	Mean (1-5)	SD
1	My overall experience using Amulet was satisfactory	4.6	0.55
2	Wearing Amulet was enjoyable and interesting	4.2	0.45
3	The Amulet is comfortable to wear	3.2	0.45
4	I could easily feel the buzzer when it buzzed me	3.4	1.3
5	The display was easy to read, even in varying light conditions	4	0
6	The buttons were easy to use	3.6	0.55
7	I would consider wearing Amulet for a longer period of time	4.2	0.84
8	I think that Amulet can be used to help with activity monitoring in older adults	4.6	0.55

Table 16: Summary of Usability Questionnaire for Negative Statements

	Survey Statements (Negative)	**Mean (1-5)**	**SD**
1	Wearing Amulet interfered with my daily activities	1.4	0.55
2	Wearing Amulet interfered with my social interactions	1.4	0.55
3	Wearing Amulet made me feel self-conscious in public	1.4	0.55
4	I felt that wearing Amulet was a nuisance	2.4	0.55

Overall, there were high scores for the positive statements and low scores for the negative statements. These results suggest that the *ActivityAware* system has the potential to be used by older adults for activity monitoring.

6.4 Energy Efficiency Evaluation

We evaluated the battery life of the *ActivityAware* system by analyzing the hourly log over the 5-day period. All 5 Amulets were still running the *ActivityAware* app upon return of the devices and none of them had been charged. This suggests that the system has a battery life of at least 5 days. To predict exactly how long the system would run before needing to be recharged, we plotted the battery life over the 5-day period for each device (Figure 18).

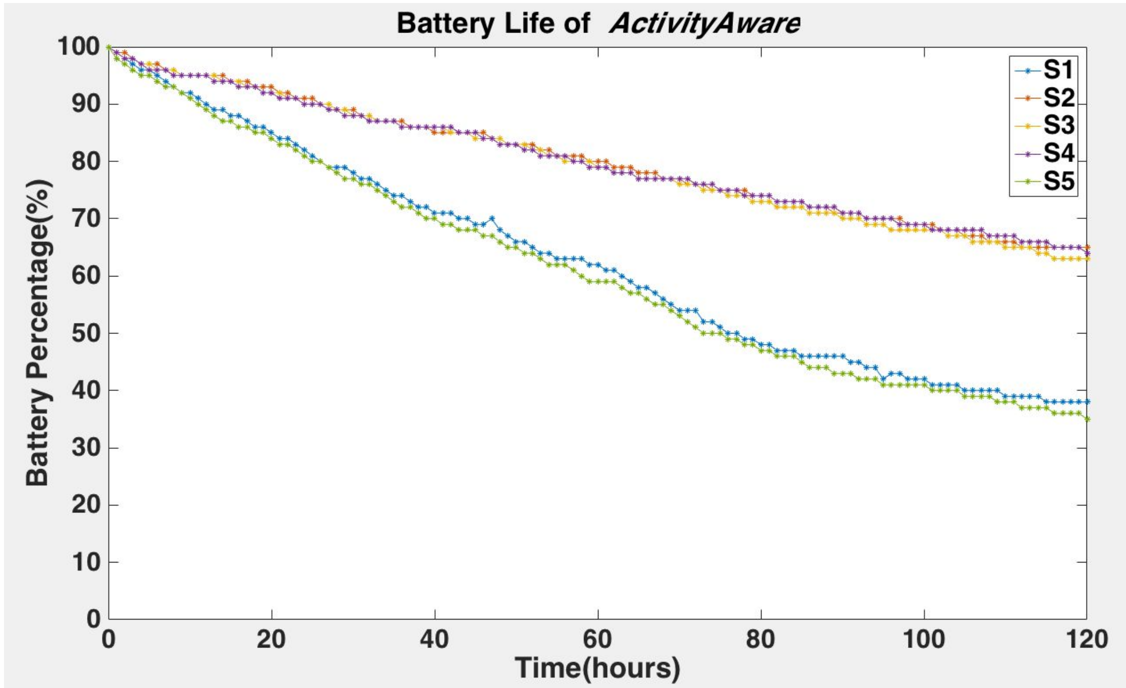

Figure 18: Plots of battery charge of 5 devices for 5 days

We computed a linear extrapolation of the battery data from Figure 18 to estimate battery life. The results summarized in Table 17 show that the system could run for at least 7 days (178 hours) before needing to be recharged, and in the best case the system might last 14 days. This result matches the battery life prediction of at least 7 days from the power-draw measurements in Section 4.2. We observed there was a difference in battery life of all the 5 devices. A further investigation is necessary to pinpoint whether the difference in battery life is due to the specific batteries in the devices, the difference in how the system was used (there is a slight increase in battery life when the device is not worn for longer periods), a combination of the two, or some other factors. This result demonstrates that the *ActivityAware* system is sufficiently energy efficient.

Table 17: Projected Battery Life of 5 devices

	D1	**D2**	**D3**	**D4**	**D5**
Battery Life (Days)	7.4	13.9	13.5	14.2	7.2

7 Related Work

Several methods have employed accelerometers for monitoring physical activity. Some of these methods perform activity classification in real time whereas others do it offline. These works range from systems that have been developed by researchers for activity classification to commercial devices used for personal physical-activity monitoring. There are three main categories of approaches: systems that use linear regression, systems that use machine learning, and systems that use proprietary algorithms. This section describes these three approaches to physical activity monitoring.

7.1 Linear-Regression Algorithm

Several researchers have developed cut points of activity counts per minute for activity levels such as light, moderate and vigorous. These cut points are estimated from a linear-regression equation fitted to data corresponding to acceleration and metabolic costs of subjects. These approaches use commercial accelerometers (such as the Actigraph) whose outputs are 'activity counts'. Activity counts are derived using proprietary algorithms and computed over various epochs such as a minute. We describe two such approaches.

Freedson et al. were one of the first to develop cut points of acceleration counts/minute for four physical activity levels, light, moderate, hard, and very hard [13]. They simultaneously collected accelerometry and oxygen consumption data from 50 adults (25 males, 25 females) as they walked and ran on a treadmill at various speeds. They collected the accelerometry data using the Computer Science and Applications, Inc. (CSA) activity monitor (currently called the Actigraph) placed at the subject's hip. They collected the oxygen consumption data using an open-circuit spirometer. They used the oxygen-consumption data to estimate metabolic equivalents (METs), which are a standard metric to express the intensity of activities. They used linear regression to establish the relationship between METs and counts/min. They found a linear relationship ($r = 0.88$) between counts/min and METs. They used the regression equation for estimating METs from counts/min to find the count ranges for MET categories for the defined activity levels: light (<= 2.99 METs), moderate (3.0 - 5.99 METs), hard (6.0 - 8.99 METs), and very hard activity (>=9.0 METs). They then ran a field study where a subject wore a CSA device on the hip during non-sleep time over a three-day period as it logged activity counts. The activity counts were used to estimate how much time the subject spent in each of the activity levels daily. The subject also used a diary to record an hourly summary of time spent in all of the activity levels to the nearest 15 minutes, which was used for offline analysis. The authors calculated the amount of time that was spent in each of these three activity levels per day. Their estimate showed that 84-96% of each day was spent in light activity 4-16%: (45-135 min) moderate and above in comparison with the dairy recordings, which showed that 83-97% of each day spent in light activity and 9-17% (30-150 min) was spent in activity level moderate and above. They did not perform a correlation of the hourly estimate with the diary and also did not estimate the error rate or accuracy of their cut points in their analysis.

Miller et al. primarily sought to examine the estimation of activity intensity across different age groups since most previous studies had focused on younger adults [14]. They developed cut points of activity counts/minute for three physical activity levels (light, moderate, and vigorous) for each of three age groups (20-29, 40-49, and 60-69 years). They used a study methodology similar to those by Freedson et al. [13]. They simultaneously collected accelerometry and oxygen-consumption data from 90 healthy adults (30 per age group) as they walked and ran on a treadmill at various speeds. They collected the accelerometry data using the Actigraph 7164 accelerometer placed at the hip (Figure 19)[15]. They collected the oxygen-consumption data using open-circuit indirect calorimetry techniques. They used the oxygen-consumption data to estimate METs. They developed linear regression equations for counts and oxygen consumption for each age group. They used the equations to find the count ranges for MET categories for the three defined activity levels (light, moderate and vigorous). They found a strong linear relationship for each of the age groups (r = 0.94 for the 20-29 age group, r = 0.89 for the 40-49 age group, and r = 0.79 for the 60-69 age group) and overall (r = 0.90). As in the previous study, the authors did not estimate the error of their regression equations.

Figure 19: Actigraph GT9X (left) and Actigraph wGT3X-BT (right)[15]

7.2 Machine-Learning Algorithms

Several studies have used machine-learning algorithms to classify different activities and activity groups. These approaches use raw acceleration readings from accelerometers. Researchers collect acceleration data corresponding to specific activities, extract features from the data, and then train a machine-learning algorithm on that data so that given a new piece of data, the algorithm would be able to correctly assign the activity category. We describe two such approaches.

In their work, Maurer et al. developed a real-time activity recognition system using a custom built multisensor system called the eWatch, which they placed on various parts of the body including the wrist, belt and pocket [7]. The eWatch contains the following sensors: 2-axis accelerometer, temperature sensor, light sensor and microphone (Figure 20). Their system classified six activities: sitting, standing, walking, ascending stairs, descending stairs, and running. They collected acceleration and light data from six subjects as they performed these activities. They extracted various temporal features from the data such as mean, standard deviation, variance, root mean square, and zero crossing rate. They used a decision tree as their classifier and ran 5-fold cross validation to evaluate the performance of their model. They had a classification accuracy of up to 87% for both the wrist and the belt positions with the 20 Hz down-sampled data. They had a subject wear the eWatch on the wrist as the subject performed the following sequence of activities: walked to a restaurant, sat down, ate lunch, returned to office and sat down to continue working. Their plot of the classified activities against the actual activities showed that their predictions qualitatively matched the actual activities except for eating lunch, which was partially interpreted as walking or running (possibly due to arm movements).

Figure 20: The eWatch device [7]

Manini et al. developed a computationally efficient algorithm to classify four activity categories: ambulation, cycling, sedentary, and other [10]. They used data from triaxial accelerometers (called Wockets) placed at the ankle and wrist of 33 subjects. The subjects performed 26 activities, which were categorized as follows: ambulation (natural walking, treadmill walking, carrying a box, and stairs up/down), cycling (indoor and outdoor), sedentary (lying, sitting, Internet search, reading, typing, writing, sorting files on paperwork, and standing still) and other (sweeping with broom and painting with roller or brush). They computed the signal magnitude vector of the data from which they extracted temporal features (mean, standard deviation, minimum and maximum), Fourier transform features and wavelet transform features. They trained a support vector machine (SVM) as their classifier. They had an accuracy of up to 84.7% for wrist and 95% for ankle data using leave-one-subject-out cross validation.

7.3 Proprietary Algorithms

There are several commercial devices that people use for physical-activity monitoring, such as Fitbit, Apple Watch, Jawbone and Garmin (Figure 21). We describe two such devices.

Figure 21: Fitbit (left) and Apple Watch (right)

Fitbit is a wrist-worn device that monitors several fitness parameters such as sleep, steps taken and activity level using data from an accelerometer, a gyroscope, and a heart-rate monitor (for some models). Fitbit calculates 'active minutes' when a person performs activities with METs above 3: moderate-tointense activities such as brisk walking, cardio workout, and running for 10 continuous minutes [16]. Fitbit uses a proprietary algorithm for computing active minutes. As a result, there is no way for external researchers to validate the specific algorithms being used. It is likely the developers of Fitbit used a linear-regression based model to develop cut points for active minutes since Fitbit outputs activity counts and also estimates METs to calculate active minutes. It is also not clear what experimental conditions they used to develop and validate their algorithms.

The Apple Watch is a smartwatch that tracks various fitness parameters of users [17]. The watch has accelerometer, gyroscope and heart-rate sensors. It runs an Activity app that tracks how much a user moves, exercises and stands daily. The app tracks how active a user is and displays the information to the user using three rings: Move, Exercise and Stand. The Move ring shows how many calories that a user burns daily. The Exercise ring shows the number of minutes of brisk activity (such as brisk walking) that a user does daily. The app sets a 30-minute daily exercise goal. The Stand ring shows how many hours a user has stood or moved for at least 1 minute. Like the Fitbit, Apple Watch uses proprietary algorithms to track these fitness parameters. The watch generally lasts a day on a single charge.

7.4 Strengths and Weaknesses of Related Work

The linear-regression based systems have been widely used for physical-activity monitoring studies among the elderly. For example, Davis et al. ran a user study where they compared activity levels of younger and older adults over a 7-day period [18]. They used the Actigraph 7164 and the cut points developed by Freedson et al. for their analysis. These systems seem to be common in such studies because the cut points are easy to use to estimate how much time adults spend doing various activity levels. These systems, however, do not perform any real-time analysis, which is a crucial feature if some intelligence needs to be built into the system to encourage behaviors that will increase physical activity of older adults. Additionally, most of these cut points are derived using accelerometers placed at the hip. Placements such as the wrist, however, are more likely to improve wear-time compliance, which is crucial for a system that needs to encourage physical activity for the elderly [10]. These linear regression based studies do not give an estimate of the accuracy or error of their systems. The assumption is that because there is strong correlation, the linear regression model works well. Some research has shown that these cut points tend to have high classification error rates [19].

The machine-learning based systems can capture the intensity of activities, as in the linear-regression based systems, but could also be trained to identify specific activities performed. They can also be implemented on low-power devices and run in real-time. Also, unlike the linear-regression studies, these machine-learning based studies explore various locations of the body such as the ankle or wrist, which might lead to better wear-time compliance. Additionally, these systems have well-defined validation metrics, which provide an assessment of their accuracy, unlike the linear-regression studies. Several of these studies, however, like the two studies described above, use data from younger adults for training their models. The systems developed in these studies have not yet been validated for activity classification of the elderly. Also, most of these machine-learning studies focus on offline analysis of physical activity just like linear-regression studies [10]. For those like the system developed by Maurer et al., they do not track how much time is spent in specific activity groups and also do not focus on providing feedback to users to improve their physical activity habits [11].

Some activity trackers like Fitbit and Apple Watch have the advantage that they track in real-time the activity levels of users, and can provide their wearer with immediate feedback. As a result, users can make changes to their physical activity patterns when necessary. The purpose-built trackers are able to last for days or weeks on a single charge whereas smartwatches like Apple Watch tend to last only a day on a single charge. All these systems, however, are closed systems that use proprietary algorithms. As a result, it is not clear how activity values such as active minutes or exercise minutes are calculated, and the accuracy of the algorithms is unknown. Specifically, it is not clear whether the algorithms were validated on older adults. Also, these devices only track active or exercise minutes, which is a coarse assessment of activity levels as compared to the three levels (light, moderate and vigorous) used in many research studies. Additionally, their algorithms cannot be modified to track additional information such as the amount of sporadic activity versus longer bouts of activity, which might be needed to get a much better understanding of activity patterns of older adults in epidemiological studies [20].

7.5 Comparison to *ActivityAware* system

Our *ActivityAware* system addresses the weaknesses of these three main approaches to activity monitoring and combines the strengths into a comprehensive physical activity monitoring system to encourage physical activity among the elderly.

First, our system tracks three activity levels (low, moderate and vigorous). Using three activity levels provides a more granular assessment of physical activity patterns. This tracking can be optimized for older adults, and this thesis presents a preliminary validation of this system's algorithm on older adults.

Second, our system performs analysis of the activity levels of older adults in real-time (unlike the ActiGraph). This real-time analysis makes it possible for our system to provide feedback to the wearer concerning progress towards the CDC's recommended daily goal. It also has a long battery life (like the Fitbit) that enables activity tracking without the interruptions associated with charging mobile systems. We achieve this goal by implementing a lightweight algorithm on a low-power device and duty cycling a lot of the computational components of the system.

Third, our system is wrist-worn and hence is more likely to be worn than one placed on the hips. To this effect, our algorithm has been developed and works well using wrist data only. As a wrist-worn device, it has the potential for longer wear time.

Fourth, our system uses an algorithm (machine learning) that could be extended to detect specific activities such as sitting, standing, laying down, walking and running (although the current implementation does not focus on monitoring specific activities). With an understanding of an individual's specific activities, researchers and clinicians could devise better interventions.

Finally, our system is open-source and could be modified to compute important statistics such as sporadic minutes versus longer bouts of minutes, unlike devices like Fitbit. Additional intelligence could be built into the system based on these data to encourage users to improve physical activity when they are falling short of the recommended daily activity goal.

8 Limitations and Future Work

This work has several limitations, some of which suggest opportunities for future work.

Our field study had only 5 subjects. As a result, the conclusions made from the study are preliminary. We have planned future studies that will use a larger number of subjects.

We estimated the whether the Amulet was being worn by using a threshold of the variance of acceleration values in a time window. Our method assumes that if the Amulet is still, then it is not being worn, which is not necessarily true. Hence, better approaches need be explored such as adding a capacitive touch sensor onto the Amulet that will infer contact with the skin; motion and skin contact could be used to determine wear state.

The *AcivityAware* app in tracking progress towards the daily activity goal does not take into consideration the CDC's additional recommendation that activities should be done for at least 10 continuous minutes. The current implementation of the app just accumulates time intervals, which may just be sporadic activities, which could have contributed to the ease with which subjects achieved the daily activity goal. In the future, we should only count minutes towards the goal if the activity has lasted for at least 10 continuous minutes. Also, the app should track the percentage of the total minutes corresponding to sporadic and long bouts of activity to help better understand the activity patterns of older adults.

The activity-level display currently does not show trends over the current week or previous weeks. This information could be added onto the current display, which might be challenging because of the small size of the Amulet screen. The information could be made accessible via a button press or scrolling up or down on the capacitive touch sensor. This approach however, might add to the complexity of using the *ActivityAware* system. Currently, once the app is started, the subject does not need to interact with it, which simplifies its usage. Adding interactivity to the app might prove challenging for older adults. Further experiments are needed to find the right balance between interactivity and information to add.

Our activity-level detection model had high misclassification results in certain circumstances, which could have contributed to the ease of reaching the daily activity goal. Further experiments need to be conducted to develop a model with better performance and low computational complexity.

We used a population of mostly healthy older adults for developing our activity-level detection model. This model might not generalize to older adults classified as having obesity or physical limitations. Further experiments need to be conducted to collect data from such older adult population groups to develop a model adapted to these populations. These experiments might entail the inclusion of subject-specific information such as weight or body mass index to make the algorithm more accurate.

9 Conclusion

In this thesis, we developed a wrist-worn, energy-efficient system that uses a lightweight machine-learning algorithm to monitor and encourage physical activity among older adults. Our *ActivityAware* app runs on the Amulet wearable platform and measures the activity levels of individuals continuously and in real time. The app continuously collects acceleration data on the Amulet, classifies the activity level of an individual, updates the day's accumulated time spent at that activity level, displays the results on the screen as feedback to the wearer, and logs the data for later analysis.

We developed an activity-level detection model using a Linear Support Vector Machine (SVM). We obtained classification accuracies of up to 99.2% and 98.5% with 10-fold cross validation and leave-one-subject-out (LOSO) cross validation respectively.

We ran a week-long field study to evaluate the utility, usability and battery life of the *ActivityAware* system where five older adults wore the Amulet as it monitored their activity level. The utility evaluation showed that the app was somewhat useful in achieving the daily physical activity goal. The usability feedback showed that the *ActivityAware* system has the potential to be used by people for monitoring their activity levels. Our energy-efficiency evaluation revealed a battery life of at least 1 week before needing to recharge.

The results are promising, indicating that the system may be useful for activity-level monitoring by individuals or researchers for epidemiological studies, and eventually for the development of interventions that could improve the health of older adults.

References

[1] Warburton, Darren ER, Crystal Whitney Nicol, and Shannon SD Bredin. "Health benefits of physical activity: the evidence." *Canadian Medical Association Journal* 174.6 (2006): 801-809.

[2] American College of Sports Medicine. ACSM issues new recommendations on quantity and quality of exercise, February 2017. http://www.acsm.org/about-acsm/media-room/news-releases/2011/08/01/acsm-issues-new-recommendations-on-quantity-and-quality-of-exercise

[3] Josiah Hester, Travis Peters, Tianlong Yun, Ronald Peterson, Joseph Skinner, Bhargav Golla, Kevin Storer, Steven Hearndon, Kevin Freeman, Sarah Lord, Ryan Halter, David Kotz, and Jacob Sorber. "Amulet: An energy-efficient, multi-application wearable platform". *Proceedings of the ACM Conference on Embedded Networked Sensor Systems (SenSys)*. ACM Press, November 2016.

[4] Barbara E Ainsworth, William L Haskell, Melicia C Whitt, Melinda L Irwin, Ann M Swartz,Scott J Strath, William L O Brien, David R Bassett, Kathryn H Schmitz, Patricia O Emplaincourt,et al. Compendium of physical activities: an update of activity codes and met intensities. *Medicine and science in sports and exercise*, 32(9; SUPP/1):S498–S504, 2000.

[5] Centers for Disease Control and others. *Promoting Physical Activity: A Guide for Community Action*. Human Kinetics, 1999.

[6] Che-Chang Yang and Yeh-Liang Hsu. A review of accelerometry-based wearable motion detectors for physical activity monitoring. *Sensors*, 10(8):7772–7788, 2010.9

[7] Maurer, Uwe, Asim Smailagic, Daniel P. Siewiorek, and Michael Deisher. "Activity recognition and monitoring using multiple sensors on different body positions." *International Workshop on Wearable and Implantable Body Sensor Networks (BSN)*. IEEE, 2006.

[8] Hsu, Chih-Wei, Chih-Chung Chang, and Chih-Jen Lin. "A practical guide to support vector classification." (2003): 1-16.

[9] Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." *Journal of Machine Learning Research*, volume 12. Oct (2011): 2825-2830.

[10] Mannini, Andrea, Stephen S. Intille, Mary Rosenberger, Angelo M. Sabatini, and William Haskell. "Activity recognition using a single accelerometer placed at the wrist or ankle." *Medicine and Science in Sports and Exercise* 45.11 (2013): 2193.

[11] Figo, D., Diniz, P. C., Ferreira, D. R., & Cardoso, J. M. (2010). Preprocessing techniques for context recognition from accelerometer data. *Personal and Ubiquitous Computing*, 14(7), 645-662.

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

[13] Patty S Freedson, Edward Melanson, and John Sirard. Calibration of the computer science and applications, inc. accelerometer. *Medicine and science in sports and exercise*, 30(5):777–781,1998

[14] Nora E Miller, Scott J Strath, Ann M Swartz, and Susan E Cashin. Estimating absolute and relative physical activity intensity across age via accelerometry in adults. *Journal of aging and physical activity*, 18(2):158–170, 2010.8

[15] ActiGraph LLC. Actigraph, February 2017. http://actigraphcorp.com/

[16] Fitbit. Fitbit official site for activity trackers and more, February 2017. https://www.fitbit.com/

[17] Apple Inc. Apple watch, February 2017. https://www.apple.com/watch/

[18] Mark G Davis and Kenneth R Fox. Physical activity patterns assessed by accelerometry in older people.European journal of applied physiology, 100(5):581–589, 2007.

[19] Kerr J, Marshall SJ, Godbole S, et al. Using the SenseCam to improve classifications of sedentary behavior in free-living settings. Am J Prev Med. 2013;44(3):290–6.

[20] Jennifer L Copeland and Dale W Esliger. Accelerometer assessment of physical activity inactive, healthy older adults. Journal of aging and physical activity, 17(1):17–30, 2009.