Improving a Smartphone Wearable Mobility Monitoring System with Feature Selection and Transition Recognition

Nicole Alexandra Capela

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements for the degree of

MASTER OF APPLIED SCIENCE

in Biomedical Engineering

Ottawa-Carleton Institute for Biomedical Engineering
University of Ottawa
Ottawa, Ontario

April 2015

Abstract

Modern smartphones contain multiple sensors and long lasting batteries, making them ideal platforms for mobility monitoring. Mobility monitoring can provide rehabilitation professionals with an objective portrait of a patient's daily mobility habits outside of a clinical setting.

The objective of this thesis was to improve the performance of the human activity recognition within a custom Wearable Mobility Measurement System (WMMS). Performance of a current WMMS was evaluated on able-bodied and stroke participants to identify areas in need of improvement and differences between populations. Signal features for the waist-worn smartphone WMMS were selected using classifier-independent methods to identify features that were useful across populations. The newly selected features and a transition state recognition method were then implemented before evaluating the improved WMMS system's activity recognition performance.

This thesis demonstrated: 1) diverse population data is important for WMMS system design; 2) certain signal features are useful for human activity recognition across diverse populations; 3) the use of carefully selected features and transition state identification can provide accurate human activity recognition results without computationally complex methods.

Table of Contents

List of Figures

List of Tables

Abbreviations and Definitions

ADL	Activities of daily living
C1	Classifier 1
C2	Classifier 2
C3	Classifier 3
CFS	Correlation-based feature selection
Covxz	Covariance of X and Z elements of gravity vector
DifftoY	X and Z components subtracted from Y component of gravity signal
ECG	Electrocardiogram
FCBF	Fast correlation based filter
FIM	Functional independence measure
FN	False negative
FP	False positive
G-SMA$_{var}$	Maximum slope of simple moving average of sum of variances of gravity
G-XZ	Sum of range of X and Z gravity components
HAR	Human activity recognition
HLA	Horizontal linear acceleration
IEEE	Institute of Electrical and Electronics Engineers
L-SMA	Simple moving average of sum of range of linear acceleration
Matlab	High-level technical computing language and interactive environment
MCE	Multiple context based ensemble
MeMeA	Medical measurements and applications
mRmR	Minimum-Redundancy-Maximum-Relevance
RGB	Red, green, blue
SD	Standard deviation
SMA	Simple moving average
SMAvar	Simple moving average of sum of variances of gravity
SoR	Sum of range of linear acceleration
SoSD	Sum of standard deviation of linear acceleration

SVM	Support vector machine
TN	True negative
TOHRC	The Ottawa Hospital Rehabilitation Centre
TP	True positive
WMMS	Wearable mobility monitoring system
WOT	Classifier without transition identification and previous-state awareness
WT	Classifier with transition identification and previous-state awareness
Xgrav	X-component of acceleration due to gravity signal
Xlin	X-component of linear acceleration signal
Ygrav	Y-component of acceleration due to gravity signal
Ylin	Y-component of linear acceleration signal
YMAskew	Moving average of the skewness of the rotated *y* linear acceleration
Zgrav	Z-component of acceleration due to gravity signal
Zlin	Z-component of linear acceleration signal

Acknowledgments

I would like to thank my supervisors Edward Lemaire and Natalie Baddour for their continuing guidance and direction throughout the completion of this thesis. I would also like to thank BlackBerry and the Natural Sciences and Engineering Research Council of Canada for funding the research in this thesis.

I would also like to thank the following individuals:

Staff of the University Research Institute
Helena Burger
Marko Rudolf
Nika Goljar

Staff of the Ottawa Hospital Rehabilitation Centre
Nancy Turcotte
Suzanne Larivière

Students at the Ottawa Hospital Rehabilitation Centre
Andrew Herbert-Copley
Andrew Smith
Emily Sinitski
Evan Beisheim
Jiayao Huang
Marco Tundo

Finally, I would like to sincerely thank the volunteers who donated their time to participate in this research.

1. Introduction

The use of smartphones for health related applications has gained significant interest in recent years due to the mobile device's integrated sensors and computing power. Human activity recognition has been a topic of interest for entertainment and security, as well as healthcare applications [1], [2]. Research that previously used multiple dedicated wearable sensors has shifted toward smartphones, which are compact, ubiquitous, and carried close to the body for most of a person's day. As the world's population ages, it becomes increasingly important for healthcare professionals to have an accurate idea of the mobility habits and daily activities of their patients [3]. Smartphone-based activity monitoring can provide insight into patient mobility outside a hospital setting. For rehabilitation and health monitoring, this tool has the potential to revolutionize the way decision-making data is obtained, opening doors for the involvement of patients and their family in their own wellbeing.

1.1 Rationale

In the broader field of Human Activity Recognition (HAR), Wearable Mobility Monitoring Systems (WMMS) use wearable technology to track a person's movements in their chosen environment. Many WMMS have been developed for able-bodied participants; however, few systems have been tested on the elderly or people with disabilities. For example, stroke is a leading cause of disability among adults and can lead to limited activities of daily living, poor balance, and walking problems, yet WMMS have not been sufficiently tested on this population. Research on WMMS performance for able-bodied and stroke populations is needed to guide system development and ensure that WMMS work for various mobility impaired populations with differing requirements and characteristics.

The selection of relevant signal features is a crucial step in the WMMS design process in order to reduce potentially large data dimensionality and to provide viable parameters for activity classification. Most WMMS are customized to an individual research group, including unique data sets, classes, algorithms, and signal features. These data sets are obtained predominantly from able-bodied participants. Research is needed to identify classifier independent features that are useful for developing and improving WMMS across and within populations.

WMMS using a single sensor location on the body typically rely on phone orientation to identify immobile states (stand, sit and lie). This is useful for identifying a lying down state, however prior research has shown that sitting and standing are often misclassified due to similar pelvis orientation. It is also typically difficult for WMMS to distinguish between walking and climbing stairs, since the accelerometer and gyroscope signals are very similar for these activities. Many classifiers in the literature use computationally intensive training and algorithms, which can be costly on a smartphone battery. Research is needed to design reliable sit-stand differentiation and stair recognition without the use of computationally intensive feature spaces or classifiers.

1.2 Objectives and hypotheses

This thesis examines the use of smartphone hardware and sensors for gait and mobility monitoring. The four objectives of this thesis are:

1) Evaluate performance of a current WMMS for young adult and stroke populations.

   Hypothesis:

   a. Performance of a WMMS developed for able-bodied participants will be worse for a stroke population.

2) Determine which smartphone signal features are best suited to activity recognition using waist-worn smartphones for able-bodied, elderly, and stroke populations. Select features independently of a classifier to identify feature subsets that improve activity classification.

   Hypotheses:

   a. Features selected using able-bodied group data will differ from features selected from a stroke group.
   b. Generic classifier accuracy does not degrade when using a selected subset of features instead of the entire set of features.

3) Assess WMMS performance after implementing selected features from objective 2 and a proposed transition identification method.

   Hypotheses:

   a. Transition identification and previous state awareness improves F-score, sensitivity and specificity for all populations when differentiating between immobile states.
   b. Implementation of new features reduces false positives for stair activities, improves walk recognition sensitivity, and improves stair recognition specificity.

1.3 Contributions

This thesis produced several contributions in the mobility monitoring area. WMMS performance evaluations on able-bodied and stroke participants identified areas of poor performance and performance differences between groups, since these groups had varying mobility characteristics. Human activity recognition using a smartphone based system can be accomplished for both able-bodied and stroke populations; however, an increase in activity classification complexity leads to a decrease in WMMS performance with a stroke population. The study results can be used to guide WMMS development for populations with differing movement characteristics.

Useful signal features that are independent of a classifier and applicable across able-bodied, elderly, and stroke populations were identified using filter-based feature selection methods. The resulting smartphone signal feature subsets can guide future smartphone-based WMMS development among the targeted users, regardless of the classifier.

Incorporation of context awareness and transition identification into a simple hierarchal classifier significantly improved sit and stand detection, thereby avoiding the use of computationally intensive classifiers. The "covariance of the x and z axes of the gravity vector" feature was identified and implemented and reduced the number of false positives in stair recognition. These results can help guide the development of computationally inexpensive WMMS that can function on realistic data from people with differing mobility characteristics.

1.4 Thesis outline

The thesis is presented in manuscript format and is divided into 7 chapters. Chapter 2 reviews the motivation, equipment, methods, and challenges of human activity recognition. Chapters 3 and 4 contain journal manuscripts that have been accepted or submitted for publication. Part of Chapter 5 was presented as a full paper at the IEEE International Symposium on Medical Measurements and Applications (Torino, May, 2015).

Chapter 3 contains a submitted journal manuscript that addresses objective 1 and compares the performance of a custom WMMS on able-bodied and stroke participants. This chapter identifies areas in need of improvement and differences between populations.

Chapter 4 contains an accepted journal manuscript that addresses objective 2 and uses filter feature selection methods to identify useful smartphone signal features for use in WMMS development. This chapter identifies features that were selected across multiple populations.

Chapters 5 and 6 address objective 3 and evaluate the effects of implementing new features and transition recognition on the performance of the custom WMMS. In particular, immobile state recognition and stair recognition are evaluated.

Chapter 7 presents a thesis summary, overview of the thesis contributions, and suggestions for future work.

2. Literature Review

This chapter introduces the concept of human activity recognition (HAR) and ubiquitous mobility monitoring. The motivation, equipment, and methods used to recognize human activities are explored, as well as the challenges and need for additional work that exist in this unique area of research. A review of studies that use smartphones for activity and gait recognition is presented.

2.1 Mobility monitoring

HAR and mobility monitoring has the potential to contribute to many areas of our lives, with applications ranging from entertainment and video games to military, security, and healthcare [1], [2]. Extensive knowledge of a person's movements and physical activities can be indicative of mobility level, chronic disease, and the aging process [3], which can lead to improved diagnoses [4]. Pervasive sensors provide family members with an avenue to monitor the activities of loved ones who are elderly or mentally unstable, or for healthcare practitioners to obtain an objective record of their patient's daily activities.

HAR systems can use either external or wearable sensors. External sensors include cameras, force plates, light sensors, and "smart homes" (living environments equipped with various sensors) [2]. These rely on the person coming into contact with or being near the external sensors. Cameras and similar sensors (such as the Kinect sensor; a combination of infrared laser, infrared camera, and RGB camera [5]) are useful for entertainment and security; however, privacy concerns can occur when used for activity recognition and complex and computationally expensive analysis may be required [2].

Wearable sensors are fixed to the person's body, ensuring that they are always in contact with the person, and are typically less expensive than external sensors and require fewer sensor locations. In addition, wearable sensors allow the person to move freely, without restricting them to a certain laboratory or home environment. This makes wearable sensors an affordable, convenient, and practical approach to human activity monitoring.

2.2 Wearable sensors

Many types of wearable sensors have been studied, depending on the application. Rising healthcare costs and extensive technological advancements in miniature electronics and biosensors have led to a large push in recent years towards the development of wearable sensor systems and ubiquitous health monitoring [6].

Various commercial products have been developed to monitor activity levels and count steps that incorporate accelerometers or other sensors. The Sensewear armband (Figure 2.1) incorporates accelerometers and measures energy expenditure as well as skin temperature and heat flux [3]. Other accelerometer-based devices can measure energy expenditure, count steps, and offer sleep monitoring (ex. waist worn RT3 [7], wrist worn Fitbit [8], Actigraph [9], and StepWatch 3 Activity Monitor [10]; Figure 2.1). The ankle-worn AMP 331 offers detailed gait parameters such as stride length, speed, and distance travelled [11].

a) b) c)

Figure 2.1: Commercial devices for activity monitoring. a) Sensewear, b) Fitbit, c) Actigraph

Many biological sensors have been implemented in wearable applications; such as, electroencephalogram (EEG), electromyogram (EMG), electrocardiogram (ECG), and others [6]. As the world’s population ages, particular attention has been directed towards activity monitoring and fall detection [12], which can give a broad portrait of the person’s mobility and their health risk in a home environment. Many researchers have gone further in seeking a detailed portrait of a person’s activities of daily living through recognition of distinct activities such as walking, sitting, climbing stairs etc. [2]. These wearable mobility monitoring systems can be labelled WMMS.

Accelerometers, gyroscopes, magnetometers, and GPS are typically used in WMMS since these sensors are affordable, small, and unobtrusive [3]. Typical accelerometer and gyroscope signals from a waist-worn smartphone are shown in Figure 2.2. In the figure, the right-left axis corresponds to the medial-lateral axis of the wearer, the vertical axis corresponds to the superior-inferior axis of the wearer, and the forward axis corresponds to the anterior-posterior axis of the person. Since these are raw acceleration signals, the vertical acceleration baseline is offset by 9.81 m/s^2, which corresponds to the acceleration due to gravity. Gyroscope signals describe rotation about each of these axes.

Sensor placement depends on the sensor and the application. Common locations include arm, wrist, chest, waist, thigh, and ankle [13]–[15]. Some studies use multiple sensor locations, such as sternum and thigh [16] or thigh and chest[17]. While multiple locations can provide more information about a person's movements, multiple sensors can be cumbersome and inconvenient for reliable implementation in everyday life. A single sensor location is a practical approach for activity classification that leads to increased user compliance and reduced device cost [18], [19].

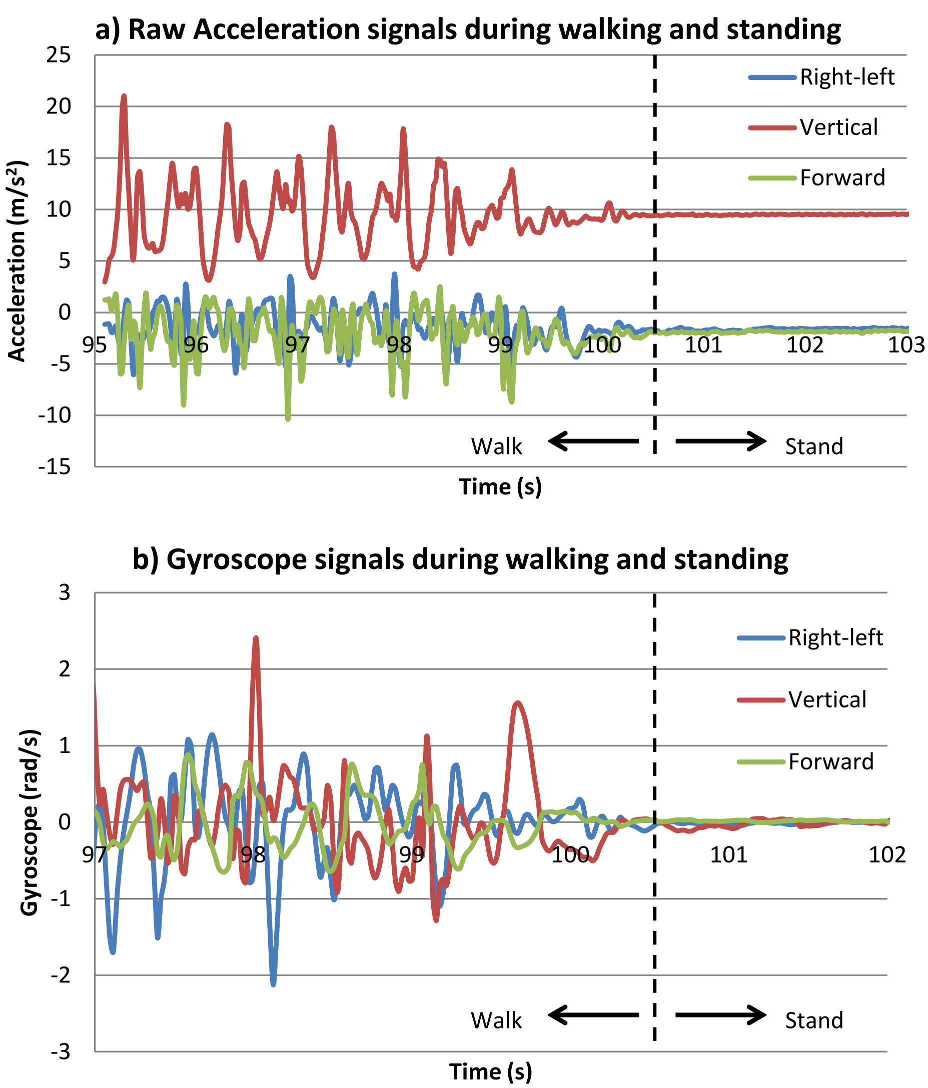

Figure 2.2 Typical signals during walking and standing from a waist-worn smartphone: a) accelerometer signal b) gyroscope signal

Smartphones are ubiquitous and are equipped with accelerometers, gyroscopes, magnetometers, and GPS. They have become computationally powerful and may be worn on or carried close to the body on a daily basis. This has made them an ideal sensor and processing platform for WMMS purposes.

Activity recognition has been accomplished using different smartphone platforms, such as Android, iOS, Symbian [20], and BlackBerry [21]. Google has created a real-time activity recognition API that identifies if the user is in a vehicle, on a bicycle, on foot (walk or run), running, walking, still, tilting, or unknown [22].

2.3 Human Activities

Human activities are diverse and complex, and in some cases may be performed simultaneously. In order to construct a WMMS that can classify human activities, a set of classes must be created. Activities are typically divided into groups, which can be classified at varying levels of detail. For example [1], [2]:

- **Ambulation**: sedentary (standing still, sitting, lying down), mobile (walking, running, ascending/descending stairs)
- **Transportation**: driving, cycling, riding a bus
- **Exercise**: bicycling, playing soccer, lifting weights
- **Phone**: making a phone call, texting, playing a game
- **Medical**: falls, rehabilitation tests, activities of daily living
- **Other daily activities**: using a computer, eating, using an ATM, etc.

The types of activities that are recognized by the WMMS affect its design. For physical rehabilitation, interest has focused on activities of daily living, which can be considered variations on basic ambulation activities (ex., eating dinner can be considered sitting down). Monitoring these activities can support rehabilitation and medical diagnosis, leading to improved treatment or therapy [4].

It is important to realistically represent human activities when evaluating the performance of a WMMS, otherwise the sensor signals acquired during evaluation may correspond to an artificial representation of an activity and not correspond to the signal that would be acquired when the WMMS is implemented. For example, if a participant sits straight and still on a chair for a discrete collection of sensor data, the signals will differ from sitting down at a table and using their hands to have dinner or play a board game. Realistic settings and continuous data collection should improve human activity recognition development.

2.4 Human Activity Recognition

The main HAR steps are data collection, pre-processing, data segmentation, feature extraction, dimensionality reduction (feature selection), and classification [23]. Raw signals generally contain noise and can be pre-processed using filtering, Fourier transforms, wavelet transforms, or other methods. When dealing with continuous data, such as a time series, the data can be divided into smaller sections using segmentation. The most widely used method in medical applications is sliding windows, which may or may not overlap. Some studies require the participant to record sensor data for each activity separately and then manually trim the excess sensor data collected [24].

Signal features are raw data abstractions, usually calculated over each window. Many features can be derived from raw sensor signals; however, simply increasing the number of features used by a classifier does not necessarily increase its accuracy since features may be redundant or not indicative of class [25]-[28]. In addition, adding more features may cause the classifier to over-fit to the training data, making it less adaptable to new data [29]. The goal of feature selection is to reduce data dimensionality (i.e. reduce the number of variables in the feature space) and only pass relevant and useful features to the classifier. This is particularly important when working with a battery-powered mobile device with limited computing resources, such as a smartphone. The classifier uses these features to identify the activity, or “class” to which the data corresponds.

2.4.1 Features and feature selection methods

The feature extraction step transforms a large set of sensor data into a reduced representation that captures the original data’s main characteristics [23]. Features may be in the time domain, which include signal characteristics and statistics (mean, variance etc.), or in the frequency domain, which look at the signal’s periodic structure. Typical HAR features include [2], [23]:

- **Time domain**: mean, variance, standard deviation, mean absolute deviation, root mean square, cumulative histogram, zero/mean crossing rate, derivative, peak count and amplitude, sign, interquartile range, kurtosis
- **Frequency domain**: discrete fast Fourier coefficient, spectral centroid, spectral energy/entropy, frequency range power
- **Time-frequency**: wavelet coefficient
- **Heuristic**: signal magnitude area, signal vector magnitude, inter-axis correlation
- **Other**: principle component analysis, linear discriminant analysis, autoregressive models

Certain features combine a signal's axes, such as using the acceleration vector's magnitude. These features are considered orientation independent, since they are the same regardless of phone orientation. These features can be useful for broad categorization of dynamic movement, but less useful for differentiating between immobile states.

Many features can be generated from a single sensor signal, therefore it is essential to identify features that are relevant or redundant to the classification problem. Feature selection methods can be grouped into filter methods, embedded methods, and wrapper methods [25].

Filter feature selection methods evaluate the relevance of a feature independently of the classifier, by observing how the features are related to the classes and to each other. These methods are typically faster than wrapper methods and are good for data with very high dimensionality [25]. Embedded methods incorporate feature selection as part of a classifier's training process. Wrapper methods use the classifier as part of the feature selection process, selecting features based on results obtained from the classifier. Wrapper methods can lead to better results than filter methods for a specific classifier, but are much slower to implement.

Examples of filter feature selection methods implemented in HAR studies include ReliefF, Correlation-based Feature Selection (CFS), Simba, and mRmR (Minimum-Redundancy-Maximum-Relevance) [30]–[34]. Weighted Naïve Bayes and feature selection using the weight vector of SVM are examples of embedded methods [35]. Random forest classifiers are a commonly used embedded method for HAR that incorporates feature selection into classifier construction [36], [37]. Many wrapper methods involving multiple classifiers have been implemented in HAR systems, including single feature classification, sequential forward selection, and greedy forward stepwise methods [30], [31], [37]. The danger of using wrapper or embedded methods is that they are classifier dependent and may tune feature selection specifically to the classifier being built [35]. This can be problematic for HAR, which has high intra-class and inter-class variability [4]. In addition, most HAR studies use unique data sets, classes, and algorithms [38] and are predominantly trained on able-bodied participants or small samples of mobility impaired participants [19]. Therefore, features selected in the HAR literature may not transfer well to more diverse populations and may be over-fit to specific classifiers. Research is needed to identify features that are useful for HAR across populations and classifiers.

2.4.2 Training and testing classifiers

For HAR, selected feature sets are used as inputs into classification and recognition algorithms. Pattern recognition techniques (classifiers) in machine learning have been a growing topic of interest in recent years. Algorithms can be constructed using unsupervised or supervised training. Unsupervised methods train the classifier using input data, but no output data (i.e. unlabeled data). The classifier searches for patterns in the data and forms natural groupings of input data. HAR systems tend to use supervised learning, which provides the classifier with both input and output data as a training set, allowing the classifier to observe the relationship between the inputs and outputs in order to make future predictions [29]. A few common classifiers used in HAR studies are listed in Table 2.1. Decision tables and decision trees have lower complexity than other classifiers, such as hidden Markov models and support vector machines that can be computationally intensive.

Machine learning classifiers are trained on one set of data and tested on new data. Strategies to maximize the use of limited data samples have emerged in machine learning applications. The simplest approach is hold-out testing, which divides the data set into training and validation sets. The classifier is trained on the training set repeatedly to find the best model parameters then tested on the validation set. m-fold cross validation consists of dividing the data set into m separate parts (m=10 is often used). The classifier is trained m times, each time withholding a different subset as the validation set. This gives an error estimate that may be different to the final classifier, since it was not trained on the full data set. Leave-one-out cross validation is an example of m-fold cross validation where $m = n$ (n = total number of samples) [29].

In the case of HAR systems, leave-one-out testing can leave one subject out, which forces the classifier to be trained and tested on different individuals [39]. Training and testing on the same individual is referred to as subject-dependent, while training and testing on different individuals is called subject-independent. Subject-dependent classifiers tend to have much better results than subject-independent, since the subject-dependent classifier has been designed for the individual's particular mobility habits [20], [40]. Unfortunately, many studies use randomly split 10-fold cross validation to evaluate HAR system performance [24], [41]–[44], where data from the same participants may be used for both training and validation, potentially resulting in excessively optimistic performance outcomes.

Table 2.1 Types of classifiers used in HAR systems [23]

Classifier	Description	Studies using the classifier
Decision tables	A set of decision rules	[13], [45]
Decision trees	A tree model of decision, outcomes and cost	[33], [42], [46], [47]
Nearest neighbor (NN)	Trained on closest samples in training feature space (K neighbors, K-NN)	[30], [33], [46]–[49]
Naïve Bayes	Simple probabilistic classifier based on Bayes theorem	[30], [33], [46]–[48]
Support vector machines	Supervised learning method	[45], [50]–[52]
Hidden Markov Models	Statistical model	[53]
Gaussian mixture models	Parametric representation of probability density functions using a weighted sum of multi-variate Gaussian distributions	[38]

Smartphone based classifiers may be trained and tested “offline” (on a computer) or “online” (locally on the phone) [20]. In many studies, the smartphone is used as a sensor and the signals are then imported to a computer or sent to a server to design, train, and test a classifier. Alternatively, a computationally expensive classifier can be trained offline and then implemented online, on the phone. Very few studies perform training online, since training tends to be a computationally expensive process [20].

There are some concerns with machine trained algorithms for HAR. Complex models may offer only marginal improvements over simple classifiers despite significant increases in computational cost, and these improvements may be due to the classifier tailoring itself to the specific idiosyncrasies of the training set [54]. Thus, improvements found in complex models over simpler models may disappear when the classifier is used with future samples from a real-world environment, particularly since human activities tend to have high variability, even in the same individual. On the other hand, a simple classifier uses the data’s most important aspects, capturing the underlying phenomena without over-fitting to the training data, making the classifier scalable and computationally efficient [54], [55]. Thus, a simple classifier may be more adaptable to different populations, as well as less costly on smartphone battery life.

Classifiers trained on ‘pure’ states or in laboratory settings may not transfer well to a real world setting. Some recent studies have acknowledged the importance of collecting realistic, continuous data sets, including complex and naturalistic activities in real-life settings [39], [56].

2.4.3 Classifier evaluation

Classifier performance is evaluated by examining how many instances were correctly and incorrectly classified. This can be descriptively represented in a confusion table, which identifies instances that were correctly classified or misclassified.

Each instance can be defined as a true positive (TP), true negative (TN), false positive (FP), or false negative (FN). A correctly classified activity is a true positive. An instance where the absence of the activity is correctly classified is a true negative. A false positive occurs when the classified activity did not really occur and false negative occurs when the activity did occur but another activity was classified. TP, FP, FN and FP can be combined to generate representative measures that describe HAR classifier performance: sensitivity, specificity, precision, F-score and accuracy (Table 2.2).

Table 2.2 Classifier outcomes

Outcome	Formula
Sensitivity	$\frac{TP}{TP+FN}$
Specificity	$\frac{TN}{TN+FP}$
Precision	$\frac{TP}{TP+FP}$
F-score	$\frac{2TP}{2TP+FN+FP}$
Accuracy	$\frac{(TP+TN)}{TP+FP+TN+FN}$

Sensitivity, also known as the true positive rate, describes how often the result will be positive if the true class is positive. A highly sensitive classifier will likely give a negative result if the true class is negative. Specificity, also called the correct rejection rate, describes how often the results will be negative, if the true class is negative. A highly specific classifier will likely give a positive result if the true class is positive. Precision, also called the positive predictive value, describes the fraction of retrieved instances that are relevant. In other words, precision reports on the probability that the true class is positive, if the classification was positive. F score is a measure of classifier accuracy, using a weighted average of the classifier's precision and sensitivity. Accuracy calculates the fraction of instances that were correctly classified. Accuracy values can be misleading when data is skewed or imbalanced [57] and provides little information on individual class performance when averaged over all classes. Ward et al. presented details on performance metrics and their shortcomings in continuous HAR [58].

In order to identify true classes and label input data for classifiers, a gold standard must be established. Some HAR systems require the user, or alternatively a research assistant, to annotate the data as it is collected, using an interface to input the activity before it is performed [24]. Many HAR studies use video recording as the gold standard comparator, which gives detailed contextual information, however this requires laborious manual labelling of activities by research staff. This is time consuming and one of the limiting factors in data availability for HAR research, which explains why many studies have small sample sets.

2.5 Challenges in HAR research

Despite the vast amount of research on HAR using smartphones, many challenges exist in this area of work. When using a mobile phone, energy consumption is a concern since a HAR program runs continuously on the device. Thus, every step of the HAR process must consider the trade-off between battery life and effectiveness. For example, employing many sensors and continuously calculating complicated frequency-domain features may improve classification accuracy but at the cost of battery life, decreasing the amount of time the HAR system can be used. Similarly, training complex classifiers needs to be performed offline, which makes it unlikely that users will take the time to personalize the training process to their specific mobility characteristics.

This leads to the challenge of classifier personalization. Classifiers trained for a specific person perform better for that person (i.e. subject-dependent training) [40], [59]. However, human activities exhibit both inter-subject and intra-subject variability. For example, a person might walk more slowly one day if they are tired or climb differently constructed stairs with varied styles. An overly specific classifier would require the user to re-train the classifier each time it is used, leading to reduced user compliance, particularly if the training phase is accomplished offline. For these reasons, HAR research is moving towards subject-independent systems [1]. A simple classifier that is subject-independent is more adaptable to future users and is also less costly on a smartphone battery.

Challenges exist when comparing HAR studies. Most studies report recognition accuracy (often an average over all activities) as the primary measure to evaluate classifier performance and use this to compare results to previous work in the literature. However, accuracy is an inappropriate measure if the classes are imbalanced [57], which is often the case in HAR datasets. In cases where activities are sampled in the same quantity, classifiers may be trained to predict certain classes too generously, since they appear more often in the training set than they do in real life. This could lead to a large number of false positives when the classifier is applied to a realistic distribution of data. Research focused on improving classifier performance for a realistic data distribution can give a more accurate portrait of a person's mobility habits.

As Allen et al. remarked [38], many HAR approaches exist in the literature and each research group presents a particular data sample, defined classes, algorithm, and feature set. Therefore, extracting meaningful information to guide HAR algorithm development is difficult. In addition, data sets are often small and consist primarily of able-bodied participants. With publicly available databases such as the UCI HAR database [60], classifiers can be compared on identical data sets and evaluated in a standardized fashion. Future studies will benefit from the use of common data and baselines for comparison. Identifying useful features for HAR across classifiers and populations will also benefit future research.

This thesis addresses these challenges by using a continuous data set with realistic class distributions and diverse populations to select features independently of a classifier. New features and transition recognition were implemented into a simple, custom, user-independent hierarchal WMMS system to improve immobile state detection and reduce stair false positives.

3. Data Collection and Processing

3.1 Protocol

Data collection took place at the Ottawa Hospital Rehabilitation Centre and the University Rehabilitation Institute in Slovenia. The collection process was designed to be controlled, but provide realistic data for activity recognition.

Before the trial, participant characteristics were recorded (i.e., age, sex, height, weight). The smartphone was placed in a holster and attached to the participant's right front pant waist with the camera facing forward, as shown in figure 3.1.

Figure 3.1: Phone placement on participant

Participants followed a predefined path, including living spaces within the rehab centres, and performed a consecutive series of mobility tasks including standing, walking, sitting, riding an elevator, brushing teeth, combing hair, washing hands, drying hands, setting dishes, filling the kettle with water, toasting bread, a simulated meal at a dining table, washing dishes, walking on stairs, lying on a bed, and walking outdoors [21]. A detailed list of the sequence of activities performed in each location is provided in Appendix A.

At the beginning of the trial, a project assistant started video recording with a separate smartphone, capturing the participant from the neck down. The participant was asked to give the phone a firm shake, providing a recognizable video and accelerometer signal event that was used to synchronize the video time to the smartphone sensor output. During the trial, the project assistant followed a few steps behind the participant and provided instructions on where to walk and what activities to perform. No guidance was given on how to perform the activities, allowing the participant to interpret each task in a realistic way. For example, there were cases when participants encountered other people in the hallway or on the elevator and had to manoeuver around them; and cases when participants chose to lie on their side rather than on their back.

After the trial, a project assistant watched the video and manually recorded the event time of each activity. The time of the shake was used to synchronize the gold-standard activity event times with the sensor signal output. All videos were processed before running the WMMS algorithm on the sensor data.

The collected sensor signal data were used to calculate features, which are abstractions of raw data (e.g., mean, standard deviation etc.). These features were calculated over 1 second sliding windows with no overlap. Each one second of the activity was considered an occurrence (i.e. 5 seconds of sitting is equal to 5 instances).

3.2 Population data sets

Data collected from three distinct populations were used in this thesis: a young adult population, a population of people who had suffered a stroke, and a population of elderly people. Population characteristics and the thesis chapters that included each dataset are shown in table 3.1.

Table 3.1: Population Data Characteristics

Group	Number of Participants	Age (years)	Height (cm)	Weight (kg)	Thesis chapter
Able-body	15	26 ± 8.9	173.9 ± 11.4	68.9 ± 11.1	4, 5, 6, 7
Elderly	17	74 ± 6.3	166.7 ± 9.3	69.6 ± 14.6	5
Elderly subset	15	73 ± 5.9	166.3 ± 9.4	68.8 ± 14.7	6, 7
Stroke	15	55 ± 10.8	171.6 ± 5.8	80.7 ± 9.7	4, 6, 7
Stroke subset	12	54 ± 8.9	170.8 ± 5.5	82.7 ± 12.6	5

The young population was a convenience sample of 15 students and staff at the Ottawa Hospital Rehabilitation Centre. None of the participants had any health issues that would affect their ability to perform the activities included in the trial.

Stroke participants were identified as capable of completing the mobility tasks by a physical medicine specialist. Thirteen stroke patients had ischemic stroke. In addition to having had a stroke, one participant had subarachnoid hemorrhage and one had impairment because of a benign cerebral tumor. Six stroke patients used one crutch, two had one arm in a sling, and one used an ankle-foot orthosis. The stroke event averaged 9.6 months before the study and the average FIM score was 107 points. At the beginning of the feature selection process described in Chapter 5, data from 12 stroke participants had been collected, which is why a subset of 12 participants was used in Chapter 5. Data from additional stroke participants were subsequently collected and included in Chapter 3 in order to have an equal number of able bodied and stroke participants for comparison.

The elderly population consisted of volunteers over the age of 65 who were capable of completing the mobility tasks in the study. One senior participant walked with a limp and was awaiting surgery on their left leg. Another participant had arthritis in his hip, which adversely affected walking gait. Another participant was cautious on his left foot due to a bunion and wore foot orthoses for the previous 2 years. He had also had his patella replaced in 2013. A total of 17 elderly participants were recruited. Data from all participants were included in Chapter 5 to maximize the amount of data included in the feature selection process. Data from a subset of 15 elderly participants were included in Chapters 6 and 7 to have the same number of participants from each population.

The study was approved by the Ottawa Health Science Network Research Ethics Board and the Ethics Board of University Rehabilitation Institute (Ljubljana, Slovenia). All participants provided informed consent.

3.3 Original WMMS algorithm

The original WMMS algorithm, evaluated and improved in this thesis, is a custom decision tree classifier developed at the Ottawa Hospital Rehabilitation Centre [61]. The general structure of the classifier is shown in figure 3.2.

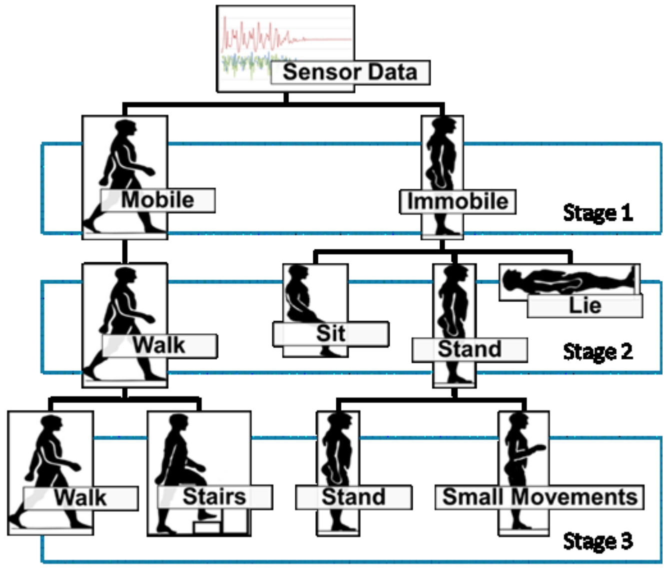

Figure 3.2: Structure of original WMMS classifier

The classifier has three activity stages. The first stage identifies if a person is mobile (walking, climbing stairs) or immobile (sitting, standing, lying down, or small movements) using a combination of the three orientation-independent features. At stage 2, pelvis orientation is examined to determine if the person is sitting, standing or lying down. At stage 3, if the person is standing, a weighting factor is calculated based on how much the person was moving. If the person was standing for more than 3 seconds and moving slightly for more than 2 seconds, the person is classified as performing a small movement (ex., standing and washing dishes at a sink).

4. Evaluation of a Smartphone Wearable Mobility Monitoring System (WMMS) with Able-Bodied and Stroke Participants

This chapter addresses objective 1 by evaluating the performance of a current custom WMMS classifier on a diverse dataset including able bodied participants and participants who have suffered a stroke [62]. Areas in need of improvement were identified, as well as differences in performance between population samples. Testing took place in two separate settings; the Ottawa Hospital Rehabilitation Centre (Ottawa, Canada) and the University Rehabilitation Institute (Ljubljana, Slovenia).

Appendix A: Data collection activity circuit presents the activity circuits from the Ottawa and Ljubljana locations.

A portion of this chapter including only the stroke participant results was presented at the 2015 European Forum for Research in Rehabilitation in Helsinki, Finland, May 2015.

Rudolf M, Goljar N, Burger H, Capela, NA, Lemaire, ED, Baddour, N, "Assessment of Stroke Patients Mobility with Wearable Mobility Measurement System (WMMS)", *European Forum for Research in Rehabilitation*, Helsinki, Finland, 2015.

This journal manuscript was submitted for publication in 2015

Capela, NA, Lemaire, ED, Baddour, N, Rudolf M, Goljar N, Burger H, "Evaluation of a Smartphone Human Activity Recognition Application with Able-Bodied and Stroke Participants", 2015.

Nicole Capela helped develop the algorithm, performed data analysis, evaluated the WMMS application, and was primary author of the manuscript. Edward Lemaire coordinated the study and contributed to software design, analysis, and the manuscript. Natalie Baddour contributed to the analysis and interpretation of data and the manuscript. Nika Goljar, Marko Rudolf, and Helena Burger assisted with data collection and contributed to the manuscript. All authors read and approved the final manuscript. The study was approved by the Ottawa Health Science Network Research Ethics Board and the Ethics Board of University Rehabilitation Institute (Ljubljana, Slovenia).

4.1 Abstract

Mobile health monitoring using wearable sensors is a growing area of interest. As the world's population ages and locomotor capabilities decrease, the ability to report on a person's mobility activities outside a hospital setting becomes a valuable tool for clinical decision-making and evaluating healthcare interventions. Smartphones are omnipresent in society and offer convenient and suitable sensors for mobility monitoring applications. To enhance our understanding of human activity recognition (HAR) system performance for able-bodied and populations with gait deviations, this research evaluated a custom smartphone-based HAR classifier on fifteen able-bodied participants and fifteen participants who suffered a stroke.

Participants performed a consecutive series of mobility tasks and daily living activities while wearing a BlackBerry Z10 smartphone on their waist to collect accelerometer and gyroscope data. Five features were derived from the sensor data and used to classify participant activities (decision tree). Sensitivity, specificity and F-scores were calculated to evaluate WMMS classifier performance.

The classifier performed well for both populations when differentiating mobile from immobile states (F-score > 94%). As activity recognition complexity increased, HAR system sensitivity and specificity decreased for the stroke population, particularly when using information derived from participant posture to make classification decisions.

Human activity recognition using a smartphone based system can be accomplished for both able-bodied and stroke populations; however, an increase in activity classification complexity leads to a decrease in HAR performance with a stroke population. The study results can be used to guide smartphone HAR system development for populations with differing movement characteristics.

4.2 Background

Mobile health monitoring using wearable sensors is a growing area of interest. As the world's population ages and locomotor capabilities decrease, the ability to monitor a person's mobility activities outside a hospital setting becomes valuable for clinical decision-making. Human Activity Recognition (HAR) systems combine wearable sensor and computing technologies to monitor human movement in the person's chosen environment.

HAR systems typically use accelerometer and gyroscope sensors since these are small, affordable, and generally unobtrusive [63]. Other HAR systems combine sensor types, such as accelerometer and ECG [64], or use multiple sensor locations, such as sternum and thigh [16], or thigh and chest [17]. However, multiple sensors can be cumbersome and inconvenient for reliable implementation in everyday life. Smartphones are ubiquitous, carried by most individuals on a daily basis, and many devices contain integrated accelerometer and gyroscope sensors, which are commonly used to measure posture and movement [6].

HAR systems typically follow a machine learning structure [2]. Raw sensor signals are collected, pre-processed, and segmented into time windows. Feature extraction is then performed to retrieve relevant information from sensor signals over each window. Features are abstractions of raw data; such as statistical calculations (mean, variance etc.) or frequency domain features that describe the signal's periodic structure. Since many features could be used in a model, a selection process is typically used to reduce the data's dimensionality. Feature selection methods may be filter-based, which evaluate features characteristics without a classifier, or wrapper based, which use classifier accuracy to evaluate features [65]. Finally, a classifier is constructed using training data and evaluated on testing data. The literature has previously focused on offline human activity recognition, although recent work is moving towards algorithms that can be implemented in real time using the onboard sensors and computational power of a smartphone [20].

Many HAR systems have been developed for able-bodied participants; however, few systems have been tested on the elderly or people with disabilities [19]. A recent study showed that an activity classification model trained on an older cohort and tested on a younger sample performed better than model training with the younger cohort and testing on the older sample. This suggested that a model trained on elderly participants may be more generalizable and result in more a robust classifier [39], since younger people may perform activities of daily living with more intensity than older or disabled people. Stroke is a leading cause of disability among adults and can lead to limited activities of daily living, balance and walking problems, and a need for constant care [66]. For a clinician, reliable data about a patient's activity is important, particularly information about the type, duration and frequency of daily activities (i.e., standing, sitting, lying, walking, climbing stairs).

The current research compared the performance of a smartphone-based wearable mobility monitoring system (WMMS) between able-bodied participants and people who had suffered a stroke. By studying differences in classifier performance between populations, we addressed the hypothesis that a WMMS developed using sensor data from able-bodied participants would perform worse on a population of stroke participants due to differences in walking biomechanics. This research also identified where the classifier performed poorly, thereby providing guidance for future research on HAR for populations with mobility problems.

4.3 Methods

4.3.1 Population

A convenience sample of 15 able-bodied participants (age 26 ± 8.9 yrs, height 173.9 ± 11.4 cm, weight 68.9 ± 11.1 kg) and 15 stroke participants (age 55± 10.8 yrs, height 171.6± 5.79 cm, weight 80.7± 9.65 kg) participated in this study. Stroke participants were identified by a physical and rehabilitation medicine specialist as capable of safely completing the mobility tasks and able to commit to the time required to complete the evaluation session (approximately 30 minutes). Six stroke patients had left hemiparesis and nine had right hemiparesis. Thirteen stroke patients had ischemic stroke. In addition to having had a stroke one participant had subarachnoid hemorrhage and one had impairment because of a benign cerebral tumor. Six stroke patients used one crutch, two had one arm in a sling, and one used an ankle-foot orthosis. The stroke event averaged 9.6 months before the study and the average FIM score was 107 points. The study was approved by the Ottawa Health Science Network Research Ethics Board and the Ethics Board of University Rehabilitation Institute (Ljubljana, Slovenia). All participants provided informed consent.

4.3.2 Equipment

Accelerometer, magnetometer, and gyroscope data were collected with a Blackberry Z10 smartphone using the TOHRC Data Logger [67]. Smartphone sampling rates can vary [68], therefore the Z10 sensors were sampled at approximately 50Hz, with a mean standard deviation of 15.37Hz across all trials. The WMMS used the Blackberry's gravity and linear acceleration output to calculate features. Linear acceleration is the Z10 acceleration minus the acceleration due to gravity. The gravity signal (acceleration due to gravity) is output from the phone and uses sensor fusion to subtract linear acceleration from the device acceleration..

Since the phone's orientation on the pelvis can differ between individuals due to a larger mid-section or different clothing, a rotation matrix method was used to correct for phone orientation [69]. Ten seconds of accelerometer data were collected while the participant was standing still and a 1-second data segment with the smallest standard deviation was used to calculate the rotation matrix constants. The orientation correction matrix was applied to all sensor data.

While the WMMS application can run entirely on the smartphone, for the purposes of this research, the raw sensor output was exported as a text file and run in a custom Matlab program to observe WMMS algorithm performance in detail and calculate outcome measures.

4.3.3 WMMS Algorithm

Raw sensor data from the smartphone were converted into features, over 1 second data windows. The features were used to classify movement activities. The features derived from acceleration due to gravity, linear acceleration, and gyroscope signals are displayed in Table 4.1. Features were selected based on the literature and observing feature behaviour from pilot data with the target activities.

Table 4.1: Features derived from smartphone sensor signals. Acceleration due to gravity = (Xgrav, Ygrav, Zgrav), linear acceleration = (Xlin, Ylin, Zlin), SD = standard deviaiton.

Signal Feature	Formula	Abbreviation
Simple moving average of sum of range of linear acceleration (4 windows)	$\frac{\Sigma^{4}_{i=1}[(range(Xlin_i)) + (range(Ylin_i)) + (range(Zlin_i))]}{4}$	L-SMA
Difference to Y	$Ygrav - Zgrav - Xgrav$	DifftoY
Sum of range of linear acceleration	$(range(Xlin_i)) + (range(Ylin_i)) + (range(Zlin_i))$	SOR
Sum of standard deviation of linear acceleration	$(SD(Xlin_i)) + (SD(Ylin_i)) + (SD(Zlin_i))$	SoSD
Maximum slope of simple moving average of sum of variances of gravity	$SMA_{var} = \frac{\Sigma^{4}_{i=1}(Var(Xgrav)_i + Var(Ygrav)_i + Var(Zgrav)_i)}{4}$ $max(SMA_{var}(2) - SMA_{var}(1), SMA_{var}(3) - SMA_{var}(2), SMA_{var}(4) - SMA_{var}(3))$	G-SMA_{var}

A custom decision tree (Figure 4.1) used these features to classify six activity states: mobile (walk, stairs) and immobile (sit, stand, lie, and small movements).

The WMMS has three activity stages. The first stage used a combination of three features (L-SMA, SOR, SoSD: Table 4.1) to identify if the person was mobile (walking, climbing stairs) or immobile (sitting, standing, lying down, or small movements). If all three features passed their respective thresholds, the state was classified as mobile. Otherwise, the state was classified as immobile. The thresholds for SMA, SOR and SoSD were 5 m/s^2, 1 m/s^2 and 1 m/s^2, respectively.

In stage 2, if the person was in an immobile state, trunk orientation was examined using the “difference to Y” signal feature (Table 4.1). Based on thresholds, the classifier determined if the person was upright (standing, DifftoY > 8 m/s^2), leaning back (sitting, -6 m/s^2 < DifftoY < 8 m/s^2), or horizontal (lying down, DifftoY < -6 m/s^2). If the person was standing, a weighting factor was calculated based on how many of the stage 1 features passed thresholds. If the weighting factor exceeded 1 (i.e., one of the three features passed their specific threshold) for two consecutive data windows and the person was standing for more than 3 seconds, the person was considered to be performing a small movement (i.e., standing and washing dishes at a sink, etc.).

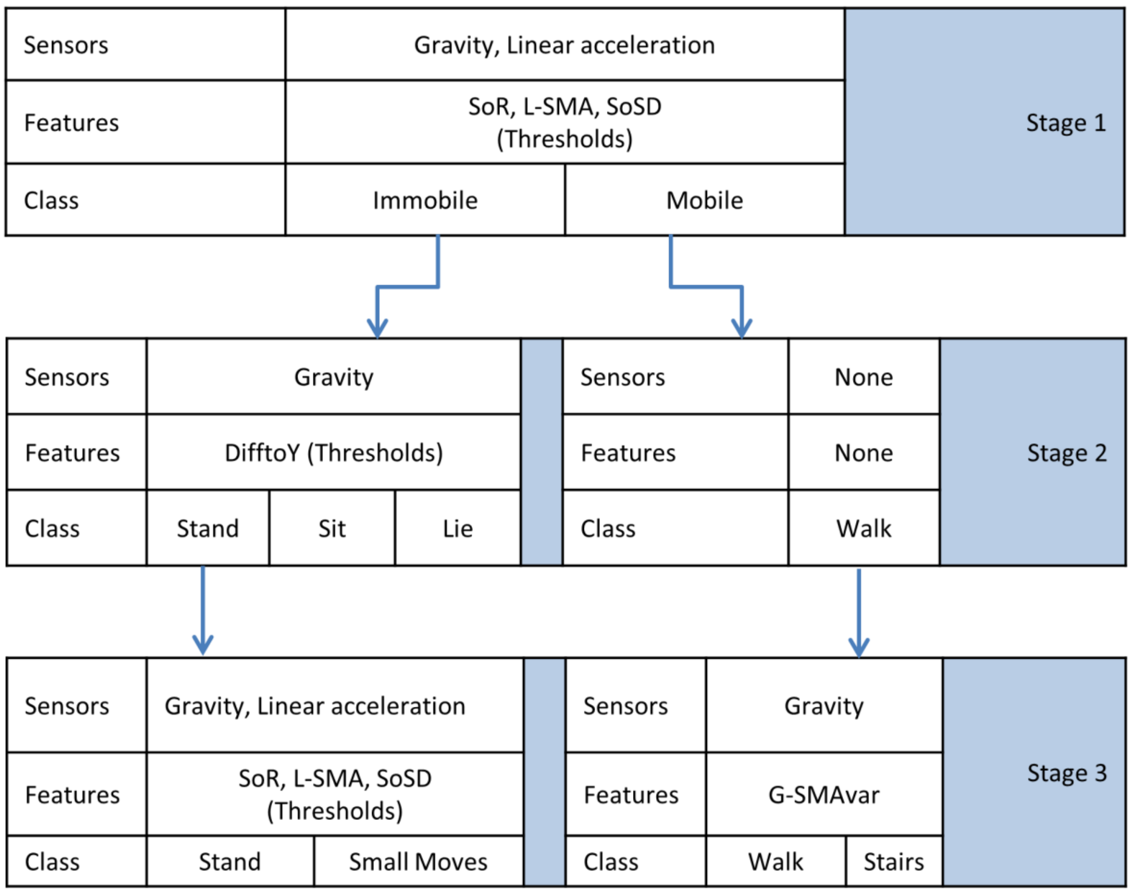

Figure 4.1: WMMS Decision Tree Structure

In stage three, the default classification was walking. If the participant walked for more than 5 seconds and the slope of G-SMA$_{var}$ feature passed a threshold of 8, then the activity was classified as climbing stairs.

4.3.4 Protocol

Data collection took place under realistic but controlled conditions. Participants follow a predefined path in The Ottawa Hospital Rehabilitation Centre or University Rehabilitation Institute , including living spaces within the rehab centres, and perform a consecutive series of mobility tasks: standing, walking, sitting, riding an elevator, brushing teeth, combing hair, washing hands, drying hands, setting dishes, filling the kettle with water, toasting bread, a simulated meal at a dining table, washing dishes, walking on stairs, lying on a bed, and walking outdoors [21].

Before the trial, participant characteristics were recorded (i.e., age, gender, height, weight). Participants wore the smartphone in a holster attached to their right-front belt or pant waist, with the camera pointed forward. Trials were video recorded using a separate smartphone for activity timing comparison and contextual information. Video time was synchronized with the smartphone sensor output by shaking the phone at the beginning and end of the trial, providing a recognizable accelerometer signal and video event.

Gold-standard activity event times were manually identified from the video recordings. Each 1 second window was considered an occurrence. For example, sitting for 5 seconds was considered 5 occurrences. When segmenting the data, a 1 second window on either side of a change of state was considered part of the transition; to reduce error from inter-rater variability in identifying the start of an activity. Transitions were not considered when calculating outcomes. The number of 1 second instances (class distribution) of each activity is shown in Table 4.2. Since this is a realistic data sample representing activities of daily living, class imbalances occur. For example, there were more instances of walking or sitting than climbing stairs or lying down.

Table 4.2: Class distributions at each level (average of all participants and standard deviation in brackets)

Activity	Able bodied	Stroke	Both
Stand	114 (27.1)	131 (37.0)	122 (32.9)
Sit	45 (6.6)	93 (26.9)	68.9 (30.8)
Lie	32 (7.0)	36 (4.3)	34 (6.0)
Walk	361 (32.9)	768 (239.4)	565 (266.5)
Upstairs	17 (2.5)	49 (25.9)	33 (24.4)
Small moves	95 (14.1)	135 (31.8)	115 (31.8)

Data analysis involved calculating the number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) in Matlab. Sensitivity, specificity, and F-scores were calculated for each individual, and the average and standard deviation of all participants were calculated for each activity. Results for each data window were compared to the gold-standard results from the video recording using descriptive statistics. Descriptive statistics and t-tests ($p<0.05$) were used to compare sensitivity, specificity, and f-scores between able-bodied and stroke groups.

4.4 Results:

The WMMS performed similarly with able-bodied and stroke populations when detecting immobile and mobile states (stage 1), with all sensitivity and specificity results greater than 0.92 and F-scores greater than 0.94 (Table 4.3). No significant differences were found between groups for stage 1, although sensitivity and F-score for the stroke population were lower for immobile states and specificity was higher for mobile states.

Table 4.3: Average, standard deviation (in brackets), and p-values for differences between able-bodied and stroke groups for sensitivity, specificity, and F-score at stage 1

Activity		Sensitivity	Specificity	F-score
Stand	Stroke	0.920 (0.076)	0.997 (0.006)	0.944 (0.053)
	Able-bodied	0.963 (0.048)	0.997 (0.005)	0.975 (0.028)
	p-value	0.08	0.95	0.06
Large Moves	Stroke	0.997 (0.005)	0.920 (0.076)	0.994 (0.006)
	Able-bodied	0.997 (0.006)	0.963 (0.048)	0.993 (0.008)
	p-value	0.95	0.08	0.68

In stage 2, specificity and F-scores for stroke participants were significantly lower for stand detection, but specificity was greater than 0.94 for both groups (Table 4.4). Specificity for lie detection was significantly greater for stroke participants, but results for both groups were greater than 0.97. Sitting sensitivity and F-Score were lower than the other activities, with results for both groups less than 0.68.

In stage 3, stand F-scores for stroke participants were significantly lower than the able-bodied group (Table 4.5). Lie specificity was significantly greater for stroke participants, but outcomes for both groups were greater than 0.98. For the stroke group, walk sensitivity and F-score were lower. Specificity was significantly lower for stair recognition and sensitivity and small movement recognition was poor for both groups.

Table 4.4: Average, standard deviation (in brackets), and p-values for differences between able-bodied and stroke groups for sensitivity, specificity, and F-score at stage 2

Activity		Sensitivity	Specificity	F-score
Stand	Stroke	0.826 (0.133)	0.940 (0.053)	0.701 (0.176)
	Able-bodied	0.903 (0.168)	0.987 (0.024)	0.917 (0.137)
	p-value	0.17	0.01	0.00
Sit	Stroke	0.533 (0.361)	0.987 (0.021)	0.568 (0.265)
	Able-bodied	0.646 (0.408)	0.983 (0.036)	0.673 (0.405)
	p-value	0.43	0.71	0.41
Lie	Stroke	0.794 (0.347)	1.000 (0.000)	0.824 (0.337)
	Able-bodied	0.943 (0.086)	0.979 (0.038)	0.871 (0.165)
	p-value	0.13	0.05	0.64
Walk	Stroke	0.997 (0.006)	0.961 (0.035)	0.993 (0.006)
	Able-bodied	0.997 (0.005)	0.966 (0.041)	0.989 (0.011)
	p-value	0.95	0.70	0.27

Table 4.5: Average, standard deviation (in brackets), and p-values for differences between able-bodied and stroke groups for sensitivity, specificity, and F-score at stage 3

Activity		Sensitivity	Specificity	F-score
Stand	Stroke	0.759 (0.163)	0.883 (0.051)	0.512 (0.145)
	Able-bodied	0.878 (0.169)	0.886 (0.044)	0.728 (0.128)
	p-value	0.06	0.89	0.00
Sit	Stroke	0.533 (0.360)	0.978 (0.037)	0.529 (0.262)
	Able-bodied	0.646 (0.408)	0.975 (0.049)	0.660 (0.400)
	p-value	0.43	0.88	0.30
Lie	Stroke	0.794 (0.347)	1.000 (0.000)	0.824 (0.337)
	Able-bodied	0.943 (0.086)	0.982 (0.033)	0.871 (0.165)
	p-value	0.13	0.05	0.10
Walk	Stroke	0.514 (0.161)	0.903 (0.074)	0.646 (0.123)
	Able-bodied	0.643 (0.226)	0.932 (0.052)	0.734 (0.162)
	p-value	0.08	0.22	0.10
Stairs	Stroke	0.622 (0.260)	0.672 (0.107)	0.101 (0.085)
	Able-bodied	0.711 (0.384)	0.805 (0.123)	0.168 (0.142)
	p-value	0.46	0.00	0.13
Small Movements	Stroke	0.154 (0.156)	0.987 (0.016)	0.209 (0.179)
	Able-bodied	0.091 (0.102)	0.994 (0.01)	0.149 (0.163)
	p-value	0.20	0.23	0.35

4.5 Discussion

This research demonstrated that a smartphone-based WMMS approach can provide relevant information on human movement activities for both able-bodied and stroke populations, at a broad level of detail; however, sensitivity and specificity decrease as the classification tasks become more complex. Thus, our hypothesis that the WMMS would perform worse for stroke participants was valid at higher detail levels, but invalid at a broad classification level.

For stage 1, mobile and immobile activity states were well classified for both able-bodied and stroke populations. From the accelerometer-based HAR literature, activity classification accuracy ranged from 71-97% [1], [2], with studies in the past two years typically reporting results from 92-96% for able bodied [44], [70] and 82-95% for older people [71]. Since this stage has only 2 classes, and the feature differences are large, thresholds can be set such that variability between people and populations has less of an effect on classification accuracy. Classification errors at stage1 may not be purely due to WMMS issues. For example, annotating gold-standard video can be difficult for small movements, such as washing dishes, since the person may move their body enough to be classified in a mobile state but human interpretation of the video could indicate an immobile state. The WMMS may provide a more consistent method of assessing an appropriate movement threshold for daily activity assessment since human raters could differ in their interpretation of movement-type and movement-onset during activities of daily living.

In stage 2, classification algorithm performance decreased when identifying if an immobile person was standing, sitting, or lying down. Specificity and F-score were significantly lower for stand detection and the algorithm performed poorly for sit identification, for both populations. Classification was based on static thresholds from a single feature (DifftoY). Since stroke can cause posture asymmetry during standing [72] and the stroke population was much older than the able-bodied sample, with posture changing with age [73], the DifftoY feature and threshold may not be sufficient to identify standing across populations, and could benefit from a combination of multiple features. In addition, inaccurate results could occur if the phone shifted or changed orientation during the trial. The therapist manually repositioned the phone during the trial for two stroke participants, one stroke participant unintentionally moved the smartphone with her paretic hand, and another participant intentionally re-adjusted his phone. The changed position may have affected application performance for activities that require a consistent phone orientation (i.e., standing, sitting, lying).

Inclination angle is typically used to classify posture when using a single accelerometer location [74]. In this case, sit identification relies on the pelvis tilting slightly back while sitting, which was not always the case in this study. For example, when a person sits at the dinner table they often lean forward to reach for objects or when eating. If the person did not sit back enough to pass the threshold before leaning forward, sitting was not identified. In many cases, stroke participants were detected as standing during the dinner table sequence. This reduced sitting sensitivity and standing specificity. Improvements in sit detection from one pelvis-worn sensor location could be achieved by using additional features or expanding the duration of sit analysis beyond the 1-second data window to compensate for forward-back transitions when sitting and performing daily activities (eating, office work, etc.). The DifftoY threshold setting also attributed to classification problems for three of the able-bodied participants, for whom some sit periods were classified as lie. This outcome also demonstrated the importance of assessing HAR systems across a range of daily activities since the results would have been much better if only "pure" sit, stand, and lie tasks were included.

In stage 3, lower walk detection sensitivity and F-score were observed for the stroke group. The smartphone was worn on the right side of the pelvis and nine of the participants had right hemiparesis, thereby reducing pelvis movement on the right side and affecting sensor and feature output. In most cases, the people with right hemiparesis had slightly lower outcomes than those with left hemiparesis (<0.18% difference in sensitivity and specificity), however the differences were not significant ($p<0.05$). Many stroke participants wore the phone with cotton pants that had an elastic waist strap, which may have provided an inferior anchor point for the phone's holster (i.e., as compared with a leather belt or fitted pants). This may have increased sensor signal variability for stroke participants. All able-bodied participants had a belt or more rigid pant waist. When used in practice, a viable HAR system must deal with mounting inconsistencies.

Stair specificity for the stroke group was significantly lower than the able-bodied group, and the algorithm performed poorly for stair recognition for all participants. F-score was low for both populations due to the high number of false positives detected, lowering the precision of classification. For five able-bodied people, the WMMS briefly detected "stairs" when lying down, then correctly re-identified the state as lie. This occurred because the feature used to detect stair climbing (covariance) increased during the stand-to-lie movement. Interestingly, this did not occur for stroke participants, perhaps due to a difference in bed height or a difference in mobility techniques when transitioning into a supine position. As with sitting, error correction over a longer duration would eliminate incorrect stair classification during the stand-to-lie transition. Stroke participants tended to rely more on the railing while climbing stairs. Multiple threshold settings for differing the stair ascent methods, or user-specific thresholds for stair identification, could be explored as a means of improving classification results. For example, one stroke participant ascended and descended the stairs in a step-by-step fashion that placed both feet on a single stair, thereby changing the sensor signals and hence affecting stair recognition. This is a common stair climbing strategy for the stroke population and persons with other mobility and walking limitations.

Small movements were not well classified for either population, resulting in a sensitivity of 0.09 for able-bodied participants and 0.15 for stroke participants. The small movements included in the trial (making toast, washing dishes, eating a meal etc.) did not always cause pelvis accelerations. Thus, accelerometer and gyroscope sensors located on the hip were not appropriate for detecting all activities. Other small movements, such as washing dishes or brushing teeth, caused the person to move their hips enough for the WMMS to classify a mobile state. The poor performance related to the difficulty in categorizing daily living human movements and difficulty setting small movement onsets when labeling gold-standard video. In future work, better methods are needed for gold file annotation, taking into account individual differences in how small movements are performed.

These results show that, while mobile and immobile classifications can be achieved with a relatively similar degree of accuracy for able-bodied and stroke participants, the WMMS had more difficulty with classification as the activity detail level increased, especially for the mobility affected stroke population. More research with pathological movement populations are required to understand how HAR algorithms need to be modified to accommodate for group and individual differences when performing activities of daily living.

Limitations in the current work include a moderate sample size from each population (15 people). The stroke group was not age matched to the able bodied group; therefore, age-related differences may have accounted for some differences in WMMS performance. However, the average ages for both groups were less than 60 years, so both groups would be considered adults, thereby minimizing potential age effects. Stroke participants were in the sub-chronic phase and were capable of completing 30 minutes of walking. In the community, post-stroke populations may have lower mobility levels that could introduce greater movement variability, thereby decreasing WMMS performance. Since this study only used one smartphone model for testing, future work could evaluate algorithm performance with other smartphone based systems.

4.6 Conclusions

In this paper, it was demonstrated that human activity recognition using a smartphone based system can be accomplished for both able-bodied and stroke populations. However, an increase in activity classification complexity leads to a decrease in WMMS performance with a stroke population. This validates the hypothesis that a WMMS developed using only able-bodied sensor data would perform worse when used to classify activities in a stroke population.

Sensor data and features produced by the different populations affected WMMS performance. The algorithm performed reasonably well for both stroke and able-bodied participants when differentiating between sit, stand, lie, and walk and between mobile and immobile states. When stair climbing and small movements were added to the classification, algorithm performance decreased. Additional features are recommended to more accurately identify sitting, standing, and lying, as well as stair identification, since stair signals are similar to level walking for many individuals. These features should be selected using data from people with differing mobility levels, so as not to over-fit the classifier to a young population with potentially more intense movements. The study results can be used to guide HAR system development for populations with differing movement characteristics.

5. Feature Selection for Wearable Smartphone-based Human Activity Recognition with Able bodied, Elderly, and Stroke patients

This chapter addresses objective 2 (select useful features for HAR) by using filter feature selection methods to select features useful for HAR with waist-worn smartphones. A diverse data set including young, elderly, and stroke participants was used to select features that improve activity classification while reducing computational burden by reducing data dimensionality. Appendix B presents the transition classes included in the study.

The content of this chapter was accepted for publication in PLoS ONE:

> Capela NA, Lemaire ED, Baddour N. Feature Selection for Wearable Smartphone-Based Human Activity Recognition with Able-bodied, Elderly, and Stroke Patients. PLoS ONE. 2015 Apr 17;10(4):e0124414.

Common features were identified across the able bodied, elderly and stroke participants. These features are shown in tables 5.7, 5.10, and 5.12. Filter feature selection methods produced classifier-independent feature subsets that should be useful for developing and improving HAR systems across and within populations.

Nicole Capela, Edward Lemaire, and Natalie Baddour conceived and designed the experiments and analyzed the data. Nicole Capela performed the experiments. Edward Lemaire contributed materials and analysis tools. Nicole Capela, Edward Lemaire and Natalie Baddour contributed to writing the manuscript.

5.1 Abstract

Human activity recognition (HAR), using wearable sensors, is a growing area with the potential to provide valuable information on patient mobility to rehabilitation specialists. Smartphones with accelerometer and gyroscope sensors are a convenient, minimally invasive, and low cost approach for mobility monitoring.

HAR systems typically pre-process raw signals, segment the signals, and then extract features to be used in a classifier. Feature selection is a crucial step in the process to reduce potentially large data dimensionality and provide viable parameters to enable activity classification. Most HAR systems are customized to an individual research group, including a unique data set, classes, algorithms, and signal features. These data sets are obtained predominantly from able-bodied participants.

In this paper, smartphone accelerometer and gyroscope sensor data were collected from populations that can benefit from human activity recognition: able-bodied, elderly, and stroke patients. Data from a consecutive sequence of 41 mobility tasks (18 different tasks) were collected for a total of 44 participants. Seventy-six signal features were calculated and subsets of these features were selected using three filter-based, classifier-independent, feature selection methods (Relief-F, Correlation-based Feature Selection, Fast Correlation Based Filter). The feature subsets were then evaluated using three generic classifiers (Naïve Bayes, Support Vector Machine, j48 Decision Tree).

Common features were identified for all three populations, although the stroke population subset had some differences from both able-bodied and elderly sets. Evaluation with the three classifiers showed that the feature subsets produced similar or better accuracies than classification with the entire feature set. Therefore, since these feature subsets are classifier-independent, they should be useful for developing and improving HAR systems across and within populations.

5.2 Introduction

Human activity monitoring and classification from wearable sensors can provide valuable information on patient mobility outside a hospital setting. While research in this area has received substantial attention in recent years, most research has involved able-bodied populations and proprietary hardware. An activity monitoring approach that works with ubiquitous technologies and is applicable across clinical populations would greatly benefit evidence-based decision making for people with mobility deficits.

Smartphones provide an ideal wearable computing environment that is convenient, easy to use, and rich with sensors, computing power, and storage. Many human activity recognition (HAR) systems have been developed for smartphone use [1], some using internal sensors and others interfacing with external biological sensors [6]. When measuring posture or movement, accelerometers and gyroscopes are popular choices since they are small, affordable, and easily worn on the body. Most commercial smartphones include accelerometers and gyroscopes, making them an ideal candidate for activity monitoring in real-world or rehabilitation settings.

Wearable sensors have been used to assess movement quality after stroke, such as upper extremity motion [75] or gait characteristics [76]. Activity levels, measured as the number of times total acceleration passes a threshold or minutes per day of activity, are typically collected using accelerometers or other body-worn sensors [77], [78]. However, activity level analysis lacks contextual information. A system that provides contextual information on a person's mobility activities would be of particular interest to healthcare professionals and researchers.

The typical signal processing steps for activity recognition are pre-processing, segmentation, feature extraction, dimensionality reduction (feature selection), and classification [23]. Features are raw data abstractions, usually calculated over a data segment or window (ex. signal magnitude area[31], correlations [31], [32], [34], interquartile range [33]). While numerous features can be extracted from a signal, increasing the number of features does not necessarily increase classifier accuracy since features may be redundant or not indicative of class (i.e., the activity being classified). Thus, feature selection is used to reduce data dimensionality and pass relevant and useful features to the classifier.

As Allen et al. remarked [38], many HAR approaches exist in the literature and each research group presents a particular data sample, defined classes, algorithm, and feature set. Therefore, extracting meaningful information to guide HAR algorithm development is difficult. Cheung et al. concluded that the most promising and practical activity classification solution would use a single, waist mounted, triaxial accelerometer, and future classifiers would be trained with larger samples from mobility impaired or older participants [19]. Smartphones with triaxial accelerometers meet the technical and practical hardware requirements but determining the best signal processing approach is still an open question. There is also uncertainty regarding the question of whether signal processing approaches need to be modified for different target populations since the majority of studies have involved able-bodied participants.

Cheung et al. [19] produced an extensive review of studies between 1980 and 2010 that used accelerometers to classify human movement. The majority of the 54 analyzed studies involved able-bodied participants. Nine studies involved patients who had various conditions, including Parkinson's, back pain, and hemiparesis [19], and six studies involved elderly participants. One study involved cardiac rehabilitation patients [79]. These studies were limited by small sample sizes and used multiple sensors placed in various locations on the body, which can be obtrusive and inconvenient in a real life setting. Multiple sensors are unlikely to be consistently used in the community for long term monitoring.

A study with a larger data set (20 older adults and 32 Parkinson's patients) evaluated a commercial activity monitor, the DynaPort MoveMonitor. This device achieved an accuracy of 65.1 to 98.9% for older adults and 57.5 to 96.9% for the Parkinson's population, who had large variability [80]. Allen et al used a single triaxial accelerometer, mounted at the waist of six older healthy participants. A Gaussian mixture-model based system achieved mean accuracies of 77.3 to 98.9% [38]. Again, this study had a small sample size. Bidargaddi et al. used wavelet decomposition based measures to identify walking from other high intensity activities by using a triaxial accelerometer worn on the waist of cardiac outpatients. Sensitivity of 89.14% and specificity of 89.97% were achieved [79]. This study only differentiated walking from other activity states. A recent smartphone study with 20 younger people and 37 older people achieved total class sensitivity of 80.5% when trained on the older cohort and tested on the young, compared to a sensitivity of 69.2% when training on the young and testing on the older cohort [39]. This demonstrated the importance of considering differing populations while developing HAR systems.

From smartphone sensor data, many features have been identified for HAR [23]. Signal feature selection is necessary to identify the most important features and eliminate redundant features. A feature is considered statistically relevant if removing it decreases the prediction power, and a feature is considered redundant if another relevant feature exists with similar predicting power [65]. Feature selection methods can be categorized as filter methods, wrapper methods, or embedded methods [25].

Filter methods look at the data's general characteristics to evaluate features without involving a classifier [65]. A wrapper method uses accuracy from a specific classifier to select features. Embedded methods incorporate feature selection as part of a classifier's training process. Thus, both wrapper and embedded methods produce results that are specific to the classifier used for the task. Therefore, features weights, or feature subset selection, may only be useful to researchers using that particular classifier. In addition, the classifier depends on the training data set, which is relatively small for most HAR systems.

The purpose of this study was to determine signal features that are best suited for activity recognition using waist-worn smartphones with various populations, independent of the chosen classifier. This was achieved by examining a diverse dataset, using three different populations, and using various filter methods to select signal features independent of a classifier. Identifying feature subsets that improve activity classification will improve mobility monitoring models for use in future classifiers. Feature subsets with similar classifier performance to the full feature set should reduce computational burden, thus facilitating real-time implementations. This research is an important step in the larger aim of developing an accurate and robust HAR system for diverse populations.

5.3 Materials and Methods

A convenience sample of 15 able-bodied participants, 17 participants over the age of 65, and 12 stroke patients were involved in this study (Table 4.1). The able bodied group were healthy students and staff at the Ottawa Hospital Rehabilitation Centre. The older participants were volunteers who were capable of completing the mobility tasks in the study. One senior participant walked with a limp and was awaiting surgery on their left leg. Another participant wore foot orthoses for the previous 2 years and had his patella replaced in 2013. He was also cautious on his left foot due to a bunion. Another participant had arthritis in their hip, which adversely affected walking gait. Seven stroke patients had left hemiparesis and three had right hemiparesis. Nine stroke patients had ischemic stroke and one had impairment because of a benign cerebral tumor. Two stroke patients used one crutch and one used an ankle-foot orthosis. All participants provided written informed consent and the study was approved by the Ottawa Health Science Network Research Ethics Board. Participant characteristics were recorded on a data sheet (i.e., age, sex, height, weight).

Table 5.1: Participant characteristics (mean and standard deviation)

Group	Number of Participants	Age (years)	Sex (% male)	Height (cm)	Weight (kg)
Able-body	15	26 ± 8.9	67	173.9 ± 11.4	68.9 ± 11.1
Elderly	17	74 ± 6.3	70	166.7 ± 9.3	69.6 ± 14.6
Stroke	12	54 ± 8.9	55	170.8 ± 5.5	82.7 ± 12.6

Participants performed a pre-determined set of daily living actions by moving through a continuous test circuit that included mobility activities (walking, standing, sitting, lying, ascending/descending stairs, ascending/descending ramps), daily living tasks (combing hair, brushing teeth, preparing food, eating, washing dishes), and environment changes (opening doors, using an elevator, traversing staircase landings, walking outdoors). Digital video was recorded while participants performed the activities, to establish a gold standard against which the sensor data could be compared. Activity timing was determined from the video and the video-based time was synchronized with smartphone sensor output by shaking the phone, thereby providing an easily recognizable accelerometer signal and video event.

Accelerometer and gyroscope data were collected using a Blackberry Z10 smartphone worn on the right-front hip and sampled at approximately 50Hz (smartphone sample rates vary, the Blackberry Z10 had a 3.84 Hz standard deviation [81]). Most modern smartphones contain comparable accelerometers and are capable of 50Hz or higher sampling rates [37], [39]. The phone's x and *y* axes were parallel to the phone`s face, with the *z*-axis pointed outward. Blackberry sensor data included raw acceleration (*x, y, z*), acceleration due to gravity (*x, y, z*), linear acceleration (raw acceleration minus gravity), and gyroscope data (*x, y, z*). The gravity signal (acceleration due to gravity) is calculated by the BlackBerry 10 operating system, using sensor fusion and a proprietary algorithm, and provides the device's true (gravity-free) linear acceleration. All data were collected using the TOHRC Data Logger application [67] and imported into Matlab to calculate all features.

For HAR in this study, six activity classes were defined and labelled from the video recordings: sit, stand, lie, large movements (including walking, small steps, opening doors), stairs, and small movements (common activities of daily living that are often used in HAR studies [19], [38], [80]). The sensor data was continuous and thus contained sections between each identifiable class, which were labelled as transition states. For feature selection, the data were labelled by level of detail, such that each level contained only sensor data from a subset of classes. In this way, the selected features can be used in HAR systems that have varying detail levels. The levels were:

- Level 1: (2 classes) - Mobile, immobile (large movements and stairs labeled as mobile; sit, stand, lie, and small movements labeled as immobile)
- Level 2 (2 classes) - Sit, stand (not including small movements)
- Level 3 (3 classes) - Sit, stand, lie
- Level 4 (2 classes) - Large movements, stairs
- Level 5 (5 classes) - Ramp up, ramp down, large movements, stairs up, stairs down
- Level 6 (2 classes) - Small movements (Yes or No, only while sitting, standing or lying)
- Level 7 (21 classes) - Transition states (transition between activities, Table B.1, Appendix B)

Level 1 represented the lowest level of detail, and can be used for monitoring a person's activity level without identifying individual activities. Level 2 differentiated between sit and stand, since these states are typically difficult to distinguish using a single waist worn accelerometer, causing them to be mutually misclassified [82], [83]. Sit and stand activities with small movements, such as standing and working in the kitchen, were not included in level 2. Level 3 represented the three common immobile states (sit, stand, lie), including small movements in these states. Level 4 separated mobile states into large movements (predominantly walking) and stairs. Ramps were not included in level 4 since the sensor signals are typically similar to level walking. Level 5 represented the highest level of detail, including level walking, ramps, and stairs, to investigate features that can be considered when differentiating between activities that can have similar signals. Level 6 represented small movements, since these are difficult to detect using a waist worn sensor. Level 7 represented transitions between states.

5.3.1 Features

Time domain features are typically used in HAR systems because they help preserve battery life by virtue of being less computationally intensive [52]. Seventy six features (Table 5.2) were selected from the literature and from observation of accelerometer and gyroscope pilot test data. These features were calculated over short sliding windows (1 second, no overlap) to allow a fast response in real time and to improve detection of short duration movements [38].

Feature files were generated from the sensor data and class files were created using the activities identified in the video recordings, and synchronized with the feature files. Each 1 second window was considered an occurrence. For example, sitting for 5 seconds was considered 5 occurrences. When segmenting the data, a 2 second window was selected on either side of a change of state to encompass the transition features; therefore, transition features were not included in the feature selection process for the surrounding states. Class distributions at each level are shown in Table 5.3. Since this is a realistic data sample representing activities of daily living, class imbalances occur. For example, there were more instances of walking or sitting than climbing stairs or lying down.

With a phone positioned on the front of the pelvis, the phone's orientation when the person is standing upright differs depending on the individual's body type or clothing. To address this, a quaternion based rotation matrix method was used to correct for these differences [69]. A ten second sample of accelerometer data was collected while the participant was standing still. One second of this sample with the smallest standard deviation was used to calculate the rotation matrix constants. The remaining raw linear accelerometer data were multiplied by this matrix to create a consistent linear acceleration signal that was corrected for initial phone orientation.

Table 5.2: Features derived from raw sensor data

Feature #	Feature Description
1	Sum of range of linear acceleration
2	Sum of standard deviation of linear acceleration
3	Simple moving average of sum of range of linear acceleration
4	Difference to y ($y_{\text{gravity}} - z_{\text{gravity}} - x_{\text{gravity}}$)
5, 6, 7	Range of gravity vector (x, y, z)
8,9,10	Mean of gravity vector (x, y, z)
11,12,13	Range of linear acceleration (x, y, z)
14,15,16	Mean of linear acceleration (x, y, z)
17,18,19	Kurtosis of gravity vector (x, y, z)
20	Sum of Kurtosis of gravity vector
21	Sum of standard deviation of linear acceleration
22	Sum of variances gravity (summed diagonal of covariance matrix)
23	Simple moving average of sum of variances
24	Maximum slope of simple moving average of sum of variances
25-30	Covariance matrix elements of gravity vector
31-36	Covariance matrix elements of linear acceleration vector
37-42	Covariance matrix elements of rotated linear acceleration vector [69]
43	Velocity from integral of rotated linear acceleration[1] (excluding y-axis)
44	Velocity for y-axis
45,46,47	Mean Euclidean norm (Linear, rotated linear[1], raw acceleration)
48,49,50	Skewness of rotated linear acceleration (x, y, z)[1]
51,52,53	Moving average of skewness of rotated linear acceleration (x, y, z)[1]
54	Sum of moving average of skewness (x, y)
55	Range of rotated linear acceleration (x)
56	Moving average of distance from rotated linear acceleration[1]
57,58,59	Mean absolute linear acceleration (x, y, z)
60,61,62	Harmonic mean linear acceleration (x, y, z)
63,64,65	Cumulative sum linear acceleration (x, y, z)
66	Correlation between acceleration along gravity and heading
67	Average velocity in gravity direction
68	Average velocity in heading direction
69,70,71	Gyroscope mean (x, y, z)
72,73,74	Interquartile range linear acceleration (x, y, z)
75	Zero cross rate
76	Mean cross rate

Table 5.3: Class distributions at each level

Level	Class	Number of instances
1	Immobile	12841
1	Mobile	17607
2	Stand	4072
2	Sit	2756
3	Stand	8649
3	Sit	2808
3	Lie	1353
4	Large Movements	16720
4	Stairs	887
5	Large Movements	15262
5	Stairs up	582
5	Stairs down	458
5	Ramp up	438
5	Ramp down	346
6	Small moves	4577
6	None	8233

5.3.2 Feature Selection

The current study focused on filter methods for feature selection, since these methods are independent of the selected classifier. Three filter methods were chosen: Relief-F, Correlation-based Feature Selection (CFS), and Fast Correlation Based Filter (FCBF). These feature selection methods are not biased towards features based on activities that occur more often in the data set.

Relief F is a commonly used filter method that ranks features by weighting them based on quality (relevance). For each instance, the algorithm finds the nearest hit (data point from same class) and nearest misses (data points from different classes). Feature relevance is based on how well instances from different classes and instances from the same class are distinguished [26], [65]. Rather than providing a subset of features, Relief-F weights all features according to relevance. The formula used to update the weight of each feature is equation 4.1.

$$w_i = \sum_{j=1}^{N} \left(x_i^j - nearmiss(x^j)_i\right)^2 - \left(x_i^j - nearhit(x^j)_i\right)^2 \tag{4.1}$$

where w is the weight of the i^{th} feature, x_i^j is the value of the i^{th} feature for point x^j, and N is the total number of data points. Nearhit x^j and nearmiss x^j are the nearest data point to x^j in the same and different classes, respectively [65].

One concern with Relief-F is that it does not evaluate redundancy in comparison to other features [65]. However, Relief-F has been reported to be useful in cases with strong interdependencies between fields [26]. Since Relief-F does not select a subset of features, an appropriate number of features to include in each subset was determined by processing the ranked feature list with three common classifiers (Naïve Bayes, Support Vector Machine (SVM), j48 Decision tree (j48)) using every possible number of features, added in order of rank. For all but 3 cases, the accuracy achieved using the ten highest ranked features was within 5% of the maximum accuracy achieved. Thus, subsets of the top 10 ranked features were used to compare populations.

Correlation based Feature Selection (CFS) evaluates the relevance of features from a correlation based heuristic that examines inter-correlation among features along with their ability to predict classes [84]. Thus, CFS selects features that are highly correlated with the class and uncorrelated with each other. Feature relevance is quantified using equation 4.2.

$$merit_s = \frac{k\overline{r_{cf}}}{\sqrt{k + k(k-1)\overline{r_{ff}}}} \tag{4.2}$$

The subset S contains k features, $\overline{r_{ff}}$ is the average feature correlation, and $\overline{r_{cf}}$ is the mean feature-class correlation. This equation is a version of Pearson's correlation with standardized variables. CFS uses a "forward best first search with a stopping criterion of five consecutive fully expanded non-improving subsets" [84].

The Fast Correlation Based Filter (FCBF) method evaluates feature merit by examining the predominant correlation between features and classes and selecting the predominant features from redundant peers. By using subsets of features based on symmetrical uncertainty, the algorithm can more efficiently analyze feature redundancy to perform a faster selection and achieve a high level of dimensionality reduction (selecting a small number of features) [27].

These three filter methods were chosen because they deal with potential issues when selecting multiple features from a common data set. Specifically, Relief F is useful in cases with strong interdependencies between fields and, since the features were derived primarily from the same accelerometer sensor data, interdependencies could occur. CFS selects features that are highly correlated with the class and uncorrelated with each other, which is desirable. Since the features were expected to correlate with each other, it is necessary to identify features that can be used together to increase performance, without being redundant. FCBF also compares correlations between features, yet tends to select smaller subsets than CFS. This is ideal because reducing computational cost by using fewer features is beneficial in a wearable system.

CFS, FCBF, and Relief-F filter methods were run in Matlab using the Arizona State University Feature Selection repository [65]. The algorithms were executed for each level and each population.

5.3.3 Evaluation of Feature Selection

To evaluate whether the feature subsets were more effective for classification than the entire feature set , three common classifiers were run using all features and then using the feature subsets: Naïve Bayes, SVM, and j48 from the Arizona State University Feature Selection repository [65]. For each population, leave-one out cross validation was performed (as in [39]) at each level. Each level contained a subset of data from all participants. Data from all but one participant were used to train the classifier, which was then tested on data from the one "left out" participant. This was repeated for each participant (i.e., cross validation) to generate an accuracy data set for statistical analysis. A paired samples sign test was used to identify significant differences in classifier accuracy between the full feature set and feature subsets ($p<0.05$), since the data were neither symmetrical nor normal. The Benjamini–Hochberg procedure was used to correct for multiple comparisons. This classification procedure evaluated the feature selection results and provided outcome measures to determine if the subsets should be implemented in a HAR system.

5.4 Results

The selected features for each population are shown in Table 5.4, Table 4.5, and Table 4.6. In general, the CFS method selected larger subsets that contained between 2 and 22 features, while FCBF selected subsets with 1 to 11 features.

5.4.1 Features Selected by Population

For the able-bodied group (Table 5.4), features selected by CFS and FCBF methods were similar. Often, the FCBF features were a subset of the features selected by CFS (levels 1, 6). This was expected since these algorithms are similar. Features 18 and 23 (y gravity kurtosis and simple moving average of sum of gravity variances, respectively) were selected by all three algorithms to separate sitting and standing states. Only feature 42 (element (2, 3) of the rotated linear acceleration covariance matrix) was selected by all three algorithms to distinguish sitting, standing, and lying states, and feature 52 (moving average skewness of rotated Y linear acceleration) was selected by all three algorithms to determine if the person was climbing stairs or walking. No feature was selected by all algorithms for differentiating between mobile and immobile states, although feature 12 (range of *y* linear acceleration) was selected by both Relief-F and CFS.

Table 5.4: Features selected for able bodied participants

Level	CFS	FCBF	Relief-F
1	3, 6, 12, 23, 27, 39, 43, 64, 66, 70, 71, 73	3, 6, 27, 39, 64, 66	1, 2, 10, 11, 12, 15, 21, 33, 54, 58
2	10, 18, 23, 47	10, 18, 23, 47	7, 12, 18, 20,23, 52, 58, 65, 66, 73,
3	4, 9, 10, 42, 68	4, 9, 10, 42, 68	17, 34, 42, 44, 52, 57, 60, 61, 712, 76
4	3, 4, 7, 8, 14, 15, 23, 24, 25, 29, 34, 42, 48, 52, 57, 68	1, 8, 14, 15, 17, 29, 52	35, 48, 51, 52, 57, 60, 65, 66, 70, 75
5	4, 5, 7, 10, 23, 28, 29, 39, 42, 56	4, 20, 28, 39	11, 16, 29,36, 42, 59, 64, 76
6	3, 4, 9, 10, 11, 35, 37, 47, 48, 52, 56, 61, 712, 64, 70	10, 37	13, 44, 49, 46, 48, 52, 58, 65, 72, 76
7	3, 23, 24, 47, 68	8	7, 11, 17, 18, 20, 48, 49, 50, 55, 75

Table 5.5: Features selected for senior participants

Level	CFS	FCBF	Relief-F
1	3, 7, 12, 22, 23, 25, 33, 35, 38, 44, 44, 57, 55, 56, 58, 58, 64, 66, 70, 71, 73, 74	25, 56, 58	15, 16,19, 44, 56,57, 59, 65, 66, 75
2	4, 10, 15, 24, 44	10, 15, 24, 28	1, 2, 3, 7, 11, 27, 30, 66, 67, 71
3	4, 10, 42, 45	4, 10, 42	16, 32, 47, 57, 56, 58,712, 64, 65, 76
4	4, 9, 10, 15, 20, 23, 24, 25, 27, 29, 38, 41, 44, 47, 48, 49, 52, 54, 66, 68	4, 29, 39, 44, 52, 61 63, 66	1, 12, 14, 18, 49, 50, 52, 63, 66, 76
5	4, 22, 24, 25, 29, 43, 44, 49, 52, 54, 66	4, 29, 42, 43, 44, 52	12, 14, 25, 40, 42, 63, 64, 67, 71, 72
6	3, 4, 8, 10, 11, 51, 56, 60, 61, 70, 72, 76	4, 11	22, 27, 43, 57, 56, 66, 68, 69, 73, 75
7	9, 44, 45, 48, 51, 72	4	13, 14, 15, 19, 40, 44, 50, 64, 67, 72

For elderly participants (Table 5.5), the features selected by FCBF were subsets of the features selected by CFS for levels 1, 3, and 6, with only one extra feature selected when separating sitting and standing (feature 28: element (1,2) of the gravity covariance matrix) and when identifying the mobile state at level 5 (feature 42). All three algorithms selected feature 56 (moving average of distance from rotated linear acceleration) to differentiate mobile and immobile states, and feature 52 to discriminate other large movements from stairs.

For stroke participants (Table 5.6), all algorithms selected feature 61 (harmonic mean y linear acceleration) to distinguish mobile and immobile states, and feature 44 (y velocity) to differentiate sitting, standing, and lying down. The stroke group had the least agreement between algorithms for selected features.

Table 5.6: Features selected for stroke participants

Level	CFS	FCBF	Relief-F
1	3, 5, 6, 23, 25, 32, 38, 43, 49, 57, 56, 61, 68, 69, 70, 71, 74, 76	3, 25, 27, 32, 57, 61, 63, 65, 66, 76	7, 45, 48, 52, 54, 55, 56, 61, 66, 68
2	9, 10, 38, 47, 66, 68	26, 66, 70	1, 2, 5, 11, 12, 13, 14, 15, 28
3	4, 10, 44, 67	10, 44	17, 18, 20, 44, 50, 51, 57, 64, 65
4	4, 18, 24, 29, 31, 47, 48, 52, 56	18, 24, 52, 56	14, 17, 19, 35, 37, 39, 41, 48, 63
5	3, 4, 7, 10, 16, 17, 18, 20, 24, 25, 29, 39, 47, 48, 47, 54, 56, 58, 59	10, 20, 29, 39, 63	2, 5, 21, 33, 34, 49, 46, 59, 74
6	10, 30, 47, 48, 67, 70, 72	70, 76	15, 17, 18, 20, 44, 49, 51, 57, 65
7	47, 48	30	11, 15, 43, 44, 49, 55, 58, 72, 76

Table 5.7 compares common features selected across populations by CFS. The CFS algorithm selected common features for all populations at every detail level, except for transitions.

Table **5.8** shows the classifier accuracy for all features and the feature subset from CFS. Classifier performance was unchanged or significantly improved when using the selected feature subsets, meaning that redundant features were eliminated without sacrificing classification accuracy. As an example, Table 5.9 shows confusion tables for each classifier, run on the entire Level 3 dataset (all populations) using all features and using the CFS feature subset. For level 3, it can be considered that the cost of misclassifying a sitting state as standing is higher than the cost of misclassifying a sitting state as lying down, since lying down and sitting are both sedentary states with little energy expenditure.

Table 5.7: Common features selected across populations using CFS

Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7
Able bodied and Senior						
3, 12, 23, 64, 66, 70, 71, 73	10	4, 10, 42	4, 15, 23, 24, 25, 29, 52, 68	4, 29	3, 4, 10, 11, 56, 61, 70	n/a
Able bodied and stroke patients						
3, 6, 23, 43, 70, 71	10	4, 10	4, 24, 29, 52	4, 7, 10, 29, 39, 56	10, 70	n/a
Stroke patients and senior						
3, 23, 25, 38, 57, 56, 70, 71, 74	10	4, 10	4, 24, 29, 52	4, 24, 25, 29, 54	10, 70, 72	n/a
All three populations						
3, 23, 70, 71	10	4, 10	4, 24, 29, 52	4, 29	10, 70	n/a

Table 5.8: Classifier accuracy with all features and with selected feature subset using CFS. Bold cells show significant differences after correction for multiple tests

	Features	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7
Able bodied								
Bayes	All	97.31	94.93	77.93	72.32	85.58	78.16	20.11
	Selected	97.52	95.62	94.42	81.94	94.32	82.72	21.87
	Sig	0.146	0.754	**0.001**	**0.001**	**0.013**	0.118	0.180
SVM	All	63.62	74.05	73.47	84.97	94.90	72.06	22.50
	Selected	86.79	96.70	88.12	84.97	94.90	78.72	20.25
	Sig	**< 0.001**	**< 0.001**	**0.001**	1.000	1.000	0.180	0.549
j48	All	94.69	96.85	95.02	71.38	90.52	78.16	20.36
	Selected	97.27	96.40	95.07	75.24	92.80	81.97	22.04
	Sig	0.581	1.000	0.109	0.035	1.000	0.302	0.791
Senior								
Bayes	All	94.45	86.75	79.02	76.57	91.57	82.61	17.85
	Selected	95.08	85.54	88.88	89.56	94.73	87.44	24.50
	Sig	**0.001**	1.000	**0.002**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**
SVM	All	65.57	67.81	70.14	91.07	94.49	67.14	18.38
	Selected	88.42	77.82	70.45	91.07	94.51	63.35	21.48
	Sig	**< 0.001**	0.143	0.210	1.000	1.000	0.143	0.332
j48	All	94.37	81.12	87.31	85.39	93.92	84.05	22.84
	Selected	94.78	80.87	85.87	87.68	94.55	83.51	23.19
	Sig	0.143	0.454	0.049	0.210	0.629	0.629	0.804
Stroke								
Bayes	All	96.18	76.35	77.88	74.31	83.05	75.42	26.98
	Selected	96.49	82.49	86.61	85.66	90.69	83.18	29.52
	Sig	0.039	0.180	0.012	0.012	0.012	0.227	0.344
SVM	All	67.14	61.02	66.38	90.59	95.97	64.09	26.58
	Selected	92.07	55.57	65.35	90.59	95.97	66.85	29.73
	Sig	**0.001**	0.125	1.000	1.000	1.000	1.000	0.065
j48	All	94.56	81.63	82.90	81.87	93.00	79.52	22.09
	Selected	95.20	84.07	80.23	85.89	94.76	77.37	23.40
	Sig	0.549	1.000	1.000	0.549	1.000	1.000	0.549
All Populations								
Bayes	All	95.62	84.81	78.36	76.74	87.18	78.26	11.97
	Selected	96.02	85.99	89.00	84.19	92.17	84.01	22.28
	Sig	**< 0.001**	0.617	0.165	1.000	1.000	**< 0.001**	0.522
SVM	All	66.90	68.38	70.76	88.82	95.01	68.59	21.35
	Selected	90.59	67.59	72.82	88.82	95.01	77.99	21.51
	Sig	0.029	1.000	0.118	0.874	0.871	0.877	0.127
j48	All	94.97	86.20	89.53	83.02	92.65	81.83	20.98
	Selected	95.84	85.30	88.52	84.01	93.16	82.44	20.12
	Sig	0.549	1.000	0.754	0.065	0.549	1.000	0.227

Table 5.9: Confusion Tables for all populations at Level 3 using all features and feature subsets selected by CFS (each instance represents 1 second)

		Bayes			SVM			j48		
Features	Class	Stand	Sit	Lie	Stand	Sit	Lie	Stand	Sit	Lie
All	Stand	2135	219	7	2576	468	221	2522	52	1
	Sit	441	635	0	0	388	0	54	804	2
	Lie	0	2	404	0	0	190	0	0	408
CFS subset	Stand	2526	240	0	2546	47	8	2544	72	1
	Sit	50	612	3	30	809	0	32	784	1
	Lie	0	4	408	0	0	403	0	0	409

Table 5.10 shows the common features selected between populations by the FCBF algorithm and Table 5.11 shows the classifier accuracy for all features and the FCBF feature subset. Since FCBF subsets were smaller, there were fewer common features between populations than CFS. Four of 168 cases showed decreased classifier performance, though three of the accuracy differences were 1.6% or less, which is not clinically significant. The other case was a change from 22.50 to 17.41% accuracy; however, both results were low and likely unacceptable for making decisions on a person's mobility status. These results were from the transition data set (level 7), which was not well identified using any of the classifiers. All other cases showed unchanged or significantly improved classifier performance, demonstrating that redundant features were eliminated without sacrificing classification accuracy.

Table 5.10: Common features between populations using FCBF

Level 1	**Level 2**	**Level 3**	**Level 4**	**Level 5**	**Level 6**	**Level 7**
Able bodied and Senior						
n/a	10	10, 4, 42	29, 52	4	n/a	n/a
Able bodied and stroke patients						
3, 27, 66	n/a	10	52	39, 20	n/a	n/a
Stroke patients and senior						
25	n/a	10	52	29	n/a	n/a
All three populations						
n/a	n/a	10	52	n/a	n/a	n/a

Table 5.11: Classifier accuracy with all features and with selected feature subset using FCBF. Bold cells show significant differences after correction for multiple tests

	Features	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7
Able bodied								
Bayes	All	97.31	94.93	77.93	72.32	85.58	78.16	20.11
	Selected	97.43	95.62	94.42	82.90	94.83	78.01	21.26
	Sig	0.549	0.754	**0.001**	**0.001**	0.035	0.607	0.791
SVM	All	63.62	74.05	73.47	84.97	94.90	72.06	22.50
	Selected	82.84	96.70	88.12	84.56	94.70	72.05	17.41
	Sig	**< 0.001**	**< 0.001**	**0.001**	0.031	0.031	1.000	**< 0.001**
j48	All	94.69	96.85	95.02	71.38	90.52	78.16	20.36
	Selected	97.12	96.40	95.07	77.80	93.73	77.84	16.24
	Sig	0.581	1.000	0.109	0.180	0.302	0.791	0.607
Senior								
Bayes	All	94.45	86.75	79.02	76.57	91.57	82.61	17.85
	Selected	95.51	84.24	88.88	92.09	95.82	85.75	24.49
	Sig	**0.013**	1.000	**0.002**	**< 0.001**	**< 0.001**	0.049	0.210
SVM	All	65.57	67.81	70.14	91.07	94.49	67.14	18.38
	Selected	88.51	72.37	70.45	91.07	94.30	62.79	21.13
	Sig	**< 0.001**	0.454	0.210	1.000	0.039	0.332	0.332
j48	All	94.37	81.12	87.31	85.39	93.92	84.05	22.84
	Selected	94.93	86.52	85.87	87.56	93.96	84.23	19.57
	Sig	0.077	0.332	0.049	0.332	1.000	1.000	0.210
Stroke								
Bayes	All	96.18	76.35	77.88	74.31	83.05	75.42	26.98
	Selected	96.72	76.34	81.34	88.17	95.35	75.80	31.48
	Sig	0.021	0.227	1.000	**0.001**	**0.001**	1.000	0.109
SVM	All	67.14	61.02	66.38	90.59	95.97	64.09	26.58
	Selected	77.31	70.81	64.75	88.97	95.95	73.40	24.57
	Sig	**0.001**	0.344	0.549	**0.001**	0.109	0.065	0.549
j48	All	94.56	81.63	82.90	81.87	93.00	79.52	22.09
	Selected	95.66	88.13	79.61	87.66	94.32	76.96	24.76
	Sig	1.000	0.754	0.065	0.549	1.000	1.000	0.227
All populations								
Bayes	All	95.62	84.81	78.36	76.74	87.18	78.26	11.97
	Selected	96.27	80.75	88.09	88.88	94.29	79.34	19.70
	Sig	**< 0.001**	0.061	**< 0.001**	**< 0.001**	**< 0.001**	0.360	**< 0.001**
SVM	All	66.90	68.38	70.76	88.82	95.01	68.59	21.35
	Selected	90.44	73.76	88.02	87.58	94.67	77.28	19.96
	Sig	**< 0.001**	0.118	**< 0.001**	**< 0.001**	**< 0.001**	**0.003**	0.877
j48	All	94.97	86.20	89.53	83.02	92.65	81.83	20.98
	Selected	96.04	86.20	88.80	86.73	94.62	80.20	17.63
	Sig	0.021	0.417	0.082	**0.003**	0.268	0.200	0.090

Table 5.12 shows the common features selected across populations by Relief-F. Despite comparing populations with a larger subset of ten features, less than three features were common between populations.

Table 5.12: Common features ranked in the top ten across populations using Relief-F

Level 1	**Level 2**	**Level 3**	**Level 4**	**Level 5**	**Level 6**	**Level 7**
Able bodied and Senior						
15	66,	76, 62	66, 52	64,42	n/a	50
Able bodied and stroke patients						
54	12	17	35, 48, 75	59	44, 48, 65	11, 49, 55, 75
Stroke patients and senior						
56, 66	1,2,11	57, 64	14, 63	n/a	57	15, 44, 72
All three populations						
n/a	n/a	n/a	n/a	n/a	n/a	n/a

5.4.2 Analysis of results

5.4.2.1 Level 1: Mobile, immobile states

When distinguishing between mobile and immobile states, the "simple moving average of sum of range of linear acceleration" (feature 3), "simple moving average of the sum of variances of the gravity vector" (feature 23), mean gyroscope output on the *y* and *z* axes (features 70, 71), and correlation between acceleration in gravity and heading direction (feature 66) were selected by multiple algorithms across populations.

The CFS method selected features 3 and 23 for all populations and FCBF selected feature 3 for both able bodied and stroke populations. Both of these features are moving averages, which filter the raw signal to provide a more consistent, smoothed feature. Filtering makes the features more effective in differentiating between states at a broad level of detail; however, these features were not selected for the higher detail levels.

Mean gyroscope output on the *y* and *z* axes (features 70, 71) were selected by CFS for all populations. Since many HAR systems employ only accelerometers, informative gyroscope data are unavailable for activity classification. Most commercial smartphones have gyroscopes built in, making gyroscope data accessible and feasible.

Feature 66 was selected by two algorithms for each population, but without consistent results, and examines the relationship between vertical and horizontal accelerations. This feature has been used to differentiate between activities that translate along a single axis, such as walking, from activities that translate along multiple axes, such as stair climbing [25]. Since acceleration in an immobile state would exist predominantly in one direction, between axis correlations would be small, which explains why feature 66 was selected at level 1. Interestingly, feature 66 was not selected for all groups. Some people may perform mobile activities, such as walking with pathological gait, with smaller correlations between acceleration axes (i.e., mobile and immobile states both having smaller correlations), which would result in this feature not being consistently selected across populations.

5.4.2.2 Level 2: Sit, stand and Level 3: Sit, stand, lie

When distinguishing between sitting, standing, and lying down, feature 10 (mean *z* gravity vector) and feature 4 (difference to *y* gravity) were repeatedly selected across populations.

For differentiating sitting and standing, the CFS method selected feature 10 for all populations. This feature is related to the phone's orientation, thus feature 10 is a reasonable choice for differentiating between sitting and standing since the pelvis angle changes during these activities. This feature was selected by FCBF for able bodied and senior participants, but not stroke. The mean *y*-gravity vector also changes with phone orientation; however, these changes are small in comparison to the initial value in an upright position (roughly 9.81 m/s^2). This small change would not have been identified as significant, which could be why the mean *y* gravity vector was not selected repeatedly, as opposed to mean *z* gravity that was near zero when upright.

Interestingly, feature 4 (difference to *y* gravity) was selected for all populations by CFS when including "Lie" as a class (level 3), even though feature 4 was only selected for the senior population when differentiating solely between sit and stand (level 2). This suggests that, if the pelvis remains relatively upright for sit and stand, the *z*-axis gravity vector (feature 10) is better at differentiating between these smaller changes in phone orientation than feature 4. For level 3, FCBF selected feature 4 for able bodied and senior populations, but not stroke patients.

For the Relief-F subset, no features were selected for all populations when comparing sitting, standing and lying (level 3). The features that ranked well were those that examined a single acceleration axis (cumulative sum of *y* linear acceleration, skewness of *z* acceleration, kurtosis of *x*-gravity etc.). Since these activities affect pelvis orientation, a combination of features examining behaviour along different axes can indicate the person's state. Interestingly, the features selected by Relief-F were not similar to the ones selected by CFS and FCBF at level 3. When implementing these features in a HAR system, the CFS and FCBF subsets could be considered first since these selection algorithms take feature redundancy into consideration.

5.4.2.3 Level 4: Large movements, stairs

When distinguishing stair climbing from other large movements, four features were selected for all populations by the CFS method: feature 4 (difference to *y* gravity), feature 24 (maximum slope of simple moving average of sum of variances), feature 29 (element (1, 3) of the gravity vector's covariance matrix) and Feature 52 (moving average of the skewness of the rotated *y* linear acceleration). Feature 52 was also selected for all populations by FCBF.

Feature 4 relates to pelvis orientation. A person who walks upright, but leans forward when navigating stairs could exhibit a change in pelvis orientation. Feature 24 (maximum slope of simple moving average of sum of variances) describes how the acceleration variance increases or decreases, which would change if the person slows down or speeds up while climbing stairs. Feature 29 describes how the variance along different axes change together, and feature 52 describes the asymmetry of a person's vertical acceleration, making them viable features to identify the difference in the direction of motion between stair climbing and walking.

5.4.2.4 Level 5: Ramp up, ramp down, large movements, stairs up, stairs down

Similarly to Level 4, feature 4 (difference to *y* gravity) and feature 29 (element (1, 3) of the gravity vector's covariance matrix) were selected by CFS for all populations. These features describe how the acceleration axes relate to each another, which is directly affected by changes in pelvis orientation and the direction of movement (i.e., if a person moved up a staircase or ramp). FCBF did not select any features that were common to all populations, although features 4 and 29 were selected for two populations each, agreeing with CFS selections. Interestingly, level 5 accuracy tended to be better than level 4. Therefore, different features should be used to classify stair ascent and descent, rather than combining ascent and descent into one class.

5.4.2.5 Level 6: Small movements

For all populations, feature 10 (mean *z* gravity vector) and feature 70 (mean gyroscope on *y*-axis) were selected by CFS. Since these classes describe when a person is slightly moving while seated, standing, or lying down, these motions can be characterized by pelvis rotation, as measured by the gyroscope. The mean *z* gravity vector is along the phone's forward axis, which changes when a person leans in to accomplish a small movement task, such as making toast or eating dinner.

5.4.2.6 Level 7: Transitions

No features were common to all populations when classifying transitions. Due to the short duration of transition states and movement variability between individuals, it is difficult to identify consistent features across multiple people or groups. Since transitions are defined by the two activities performed at the start and end of the transition period, other methods could be considered to classify a transition (i.e., without using specific transition features).

5.5 Discussion

Smartphone signal features that were consistent across able-bodied, elderly, and stroke groups were successfully identified. This established viable HAR feature subsets that can be used with waist-worn smartphones. Evaluation of these subsets with generic classifiers showed improvements in activity recognition accuracy. This indicates that the features eliminated in the feature selection process were redundant and did not significantly contribute to classifier accuracy. Thus, with appropriate feature subset selection, equivalent classifier performance can be obtained with a reduced feature set, effectively reducing computation burden on the HAR.

When differentiating between sitting, standing, and lying down, the mean gravity signal along the phone's *z*-axis was repeatedly selected across populations. This is similar to results obtained by Cruz-Silva et al. [37], who ranked mean horizontal acceleration third and mean *z* acceleration 15th out of 159 features for HAR (gravity vector was not included). For our study, the feature selection algorithms found the gravity signal to be more relevant, and the redundant mean acceleration signal was therefore excluded (i.e., the gravity signal along the *z*-axis and mean horizontal acceleration may be similar for the target activities). Maurer et al. [33] also selected mean acceleration in the *z*-axis for HAR, although neither Cruz-Silva nor Maurer separated their data into different levels. The mean *z*-axis gravity feature is related to the phone's orientation, which changes as the pelvis rotates when transitioning between immobile states, such as sitting or lying down. The mean *z* gravity feature was not selected by FCBF for the stroke group, who may have a different posture when standing that lead to incorrect classification as sitting [72].

Skewness (asymmetry) of the rotated linear acceleration along the phone's *y*-axis was selected by CFS and FCBF for differentiating between large movements and stairs. This result is supported by Hache et al. [74] who identified vertical skewness as a viable feature for identifying stair ascent or descent. In future work, other activities that produce vertical acceleration (i.e., jumping, hopping, jogging, etc.) could be included to verify if skewness would remain a viable feature. Acceleration covariance was frequently selected when detecting ramp and stair ascent or descent. This feature measures how acceleration axes change together, which is affected by the person's orientation as they move on an incline.

Differentiating between stair navigation and walking is difficult for HAR systems that only use one sensor location. Pelvis movements are similar for both activities, making it difficult to derive useful information from motion sensors such as accelerometers and gyroscopes. Classification accuracy is typically lower for stair recognition than other activities when using HAR systems with a single sensor location, such as a smartphone [34], [45], [46]. This is supported by the lower classification accuracies found at level 4.

While this research identified features that were commonly selected across populations, diversity between populations did occur for many features. For example, features selected for the stroke group tended to differ from the able-bodied and elderly groups. This may be due to the inhomogeneity of the stroke participants, whose mobility levels varied and some people used crutches and arm slings. Nine stroke patients had right hemiparesis, thereby reducing pelvis movement on the right side where the phone was attached, affecting sensor and feature output. Also, many stroke participants attached the holster to cotton pants that had an elastic waist strap, which may have provided an inferior anchor point compared to the leather belts and fitted pants that the able bodied and elderly populations typically wore. This may increase sensor signal variability for stroke participants and demonstrates the importance of including the target group when training and evaluating HAR systems, so that a classifier is not tailored to an inappropriate sample set.

The highest accuracies were achieved using feature subsets selected by the CFS algorithm. Some limitations exist in the current study. While the feature selection methods were designed to compensate for class imbalances in the feature set, selective sampling before performing feature selection could improve results in future work. The classifiers were used to evaluate the quality of the selected feature subsets; however, they were not customized to the specific HAR application. When implementing the features selected in this study in HAR classifiers, it is recommended that the classifier be tailored to the specific needs of the situation. Three separate populations were included in this study, but the total sample set was small at 44 participants. Larger data sets could contribute to future work.

5.6 Conclusion

This research selected smartphone signal feature subsets for human activity recognition that were applicable across able bodied, elderly, and stroke populations. Three filter-based feature selection methods were used so that identification of useful features could be performed independent of the classifier. This information can guide future smartphone-based HAR system development among the targeted users, regardless of the classifier. In particular, the following signal features were effective across multiples populations: (i) acceleration features using simple moving averages and correlations of acceleration along different axes, as well as mean gyroscope output on the y and z axes, to distinguish between immobile and mobile states; (ii) gravity signal mean in the phone's forward direction and the difference between the phone's y gravity to the x and z gravity signals when distinguishing between sit, stand, and lie; (iii) the gravity signal range in the forward direction for differentiating between sitting and standing; (iv) skewness of the rotated linear acceleration along the y-axis for classifying stair climbing from other large movements; (v) acceleration covariance when detecting ramp and stair ascent or descent; (vi) mean gyroscope signal on the y-axis for detecting small movements.

Future research could expand on the study results through feature selection with different pathological populations, such as amputees or people with neurological disorders, since different gait patterns may identify additional features to be included in a generalized feature set.

6. Using Previous-State Awareness and Transition Identification to Improve Sit, Stand, and Lie Classification in a Smartphone WMMS

This chapter addresses objective 3 by implementing previous-state awareness and transition recognition into the WMMS to improve sit, stand and lie classification. The performance was evaluated on a diverse dataset including participants form three population groups (young, older and stroke) in two settings.

A portion of this chapter including results from only the able bodied and elderly populations was presented at the IEEE International symposium on Medical Measurements and Applications (MeMeA).

> Capela, NA, Lemaire, ED, Baddour, N, "Improving Classification of Sit, Stand, and Lie in a Smartphone Human Activity Recognition System", *IEEE International symposium on Medical Measurements and Applications (MeMeA),* Turin, Italy, 2015.

A novel immobile state identification algorithm is described in section 6.3.3. Two new features were introduced: gravity range difference (equation 6.7) was used to identify stand-to-sit and stand-to-lie transitions, and sum of X and Z gravity components (equation 6.8) was introduced to recognize sit-to-stand or lie-to-stand transitions.

Nicole Capela developed the algorithm, performed data analysis, and was primary author of the manuscript. Edward Lemaire coordinated the study and contributed to data analysis and writing the manuscript. Natalie Baddour contributed to the analysis and interpretation of data and the manuscript. All authors read and approved the final manuscript. The study was approved by the Ottawa Health Science Network Research Ethics Board and the Ethics Board of University Rehabilitation Institute (Ljubljana, Slovenia).

6.1 Abstract

Human Activity Recognition (HAR) allows healthcare specialists to obtain clinically useful information about a person's mobility. When characterizing immobile states with a smartphone, HAR typically relies on phone orientation to differentiate between sit, stand, and lie. While phone orientation is effective for identifying when a person is lying down, sitting and standing can be misclassified since pelvis orientation is often similar. Therefore, training a classifier from this data is difficult. In this paper, a hierarchical classifier that includes the transition phases into and out of a sitting state is proposed to improve sit-stand classification. For evaluation, young (26 ± 8.9 years), senior (73 ± 5.9 years), and stroke (55± 10.8 years) participants wore a Blackberry Z10 smartphone on their right front waist and performed a continuous series of 16 activities of daily living. Z10 accelerometer and gyroscope data were processed with a custom HAR classifier that used previous state awareness and transition identification to classify immobile states. Immobile state classification results were compared with (WT) and without (WOT) transition identification and previous state awareness. The WT classifier had significantly greater sit sensitivity ($p<0.05$) than WOT for young and older populations. WT sit sensitivity was greater than WOT for the stroke population, though not significantly. Stand F-scores for WT were significantly greater than WOT for seniors. Sensitivity when detecting lying down states in the able bodied population decreased, though not significantly, and all other outcomes improved across all populations. These results indicated that examining the transition period before an immobile state can improve immobile state recognition. Sit-stand classification on a continuous daily activity data set was comparable to the current literature and was achieved without the use of computationally intensive feature spaces or classifiers.

6.2 Introduction

Human activity recognition (HAR) using wearable sensors can offer valuable information to healthcare specialists about a person's daily activities, thus providing insight into their mobility status and the frequency and duration of activities of daily living (ADL).

High mobility classification accuracy has been achieved with systems that have multiple sensor locations; however, specialized wearable sensors are inconvenient for long term use outside a hospital setting. Single sensor systems result in higher user compliance and lower device costs, making them more practical for HAR [19], [39]. Smartphones are ubiquitous, contain useful sensors, and many people carry them in their daily lives, making them an ideal HAR platform.

While most HAR devices can accurately distinguish when a person is active or inactive, knowing when a person is standing, sitting, or lying down can give a more representative portrait of a person's health state. Del Rosario et al. [39] proposed that the cost of misclassifying a sitting state as standing is higher than the cost of misclassifying a sitting state as lying down, since sitting and lying are both sedentary states with lower energy expenditure. When classifying immobile states, HAR devices typically rely on phone orientation to differentiate between sit, stand, and lie [74]. This is generally effective for identifying when a person is lying down, but often provides poor results for differentiating between sitting and standing, since pelvis and trunk orientation can be similar when the person is upright. For this reason, sitting and standing states are often mutually misclassified when using a single sensor location [82].

Gjoreski et al. noted that the sit-stand classification issue is largely sidestepped in the literature [83], with some researchers opting not to include both classes [85] and some studies merging the two classes [86]. Gjoreski et al. used a multiple context-based ensemble (MCE) of classifiers and additional signal features, specifically for standing and sitting, to improve sit-stand recognition accuracy from 62% to 86%. Their random forest classifier only used MCE when sit or stand was identified. The MCE consisted of eleven contexts, including current activity, previous activity, and last transition. A model was trained for each potential context value (ex., sitting or standing), creating a data subset. The selected models were aggregated and the decision on the recognized activity was made using majority voting. This added layers of complexity and additional variables and features to the classification system.

Newly added features should be mutually uncorrelated with existing features to provide beneficial gains in classifier accuracy [54], which is difficult when features are derived from a single sensor or sensor location. High dimensionality feature vectors, multiple contexts, and complex classifiers can quickly become computationally intensive. Algorithms intended for long term activity recognition must consider the trade-off between load requirements and battery life [71].

Complex models may offer only marginal improvements over simple classifiers despite significant increases in computational cost, and these improvements may disappear when the classifier is used in a real-world environment [54]. The improvements found in sophisticated models over simpler models can degrade on future samples [54], since future distributions are rarely identical to the training data distribution. A simple classifier should use the data's most important aspects, capturing the underlying phenomena without over-fitting to the training data, making the classifier scalable and computationally efficient [54], [55].

Laboratory classification accuracy is optimistic, and real world measurements may have unexpected variability [87]. Classifiers trained on ‘pure’ states in laboratory settings may not transfer well to real world settings. Recent studies have acknowledged the importance of collecting realistic, continuous data sets that include complex and naturalistic activities in real-life settings [39], [56]. Intra-class variability (same activity performed differently by different individuals or the same individual) exists not only for different states, but also for transitions between states [4], [83], creating a danger of classifier over-fitting.

The goal of this research was to determine if context awareness and single feature transition state identification can be used in a simple hierarchal HAR system to significantly improve sit and stand recognition, achieving results comparable to more computationally intensive classifiers trained on continuous datasets.

6.3 Algorithm

This work employed a custom hierarchal WMMS classifier that used gravity and linear acceleration signals from a Blackberry Z10 smartphone to calculate features. Linear acceleration is the Z10 acceleration minus acceleration due to gravity. The gravity signal (acceleration due to gravity) uses a BlackBerry proprietary sensor fusion algorithm to subtract linear acceleration from the device acceleration. When looking at the phone’s screen in its natural orientation, the positive X axis is to the right, the Y axis is up, and the Z axis projects out of the screen. Raw sensor output was exported to a custom Matlab program to evaluate HAR classifier performance. The following sections discuss the algorithm without (WOT) and with (WT) transition identification and context awareness.

6.3.1 Signal Features

Signal features were calculated over 1 second windows with no overlap to allow a fast response in real time and to detect short duration movements [38]. The features used to classify movement activities were:

1. Sum of Ranges (SoR)

$$range(Xlin_i) + range(Ylin_i) + range(Zlin_i) \tag{5.1}$$

2. Simple moving average of sum of range of linear acceleration (L-SMA)

$$\frac{\Sigma^{4}_{i=1}[range(Xlin_i) + range(Ylin_i) + range(Zlin_i)]}{4} \tag{5.2}$$

3. Sum of standard deviation of linear acceleration (SoSD)

$$SD(Xlin_i) + SD(Ylin_i) + SD(Zlin_i) \tag{5.3}$$

4. Difference to Y-axis (DifftoY)

$$mean(Ygrav_i) - mean(Xgrav_i) - mean(Zgrav_i) \tag{5.4}$$

5. Maximum slope of simple moving average of sum of variances of gravity (G-SMA_{var})

$$SMA_{var} = \frac{\Sigma^{4}_{i=1}(Var(Xgrav)_i + Var(Ygrav)_i + Var(Zgrav)_i)}{4} \tag{5.5}$$

$$\text{G-SMA}_{var} = max(SMA_{var}(2) - SMA_{var}(1)\ ,\ SMA_{var}(3) - SMA_{var}(2),\ SMA_{var}(4) - SMA_{var}(3)) \tag{5.6}$$

The components of "acceleration due to gravity" are referred to as *Xgrav*, *Ygrav*, *Zgrav* and the components of linear acceleration are *Xlin*, *Ylin*, *Zlin*. SD is standard deviation.

Two features were introduced to identify transitions into or out of immobile states. These were implemented in the WT classifier, along with the original five features.

A. Gravity range difference (GRD)

$$range(Xgrav_i) + range(Zgrav_i) - range(Ygrav_i) \tag{5.7}$$

B. Sum of range of X and Z gravity components (G-XZ)

$$range(Xgrav) + range(Zgrav) \tag{5.8}$$

6.3.2 Original Classifier (WOT)

A custom decision tree (Figure 5.1) used the signal features to classify activity states: mobile (walk, stairs) and immobile (sit, stand, lie, and small movements).

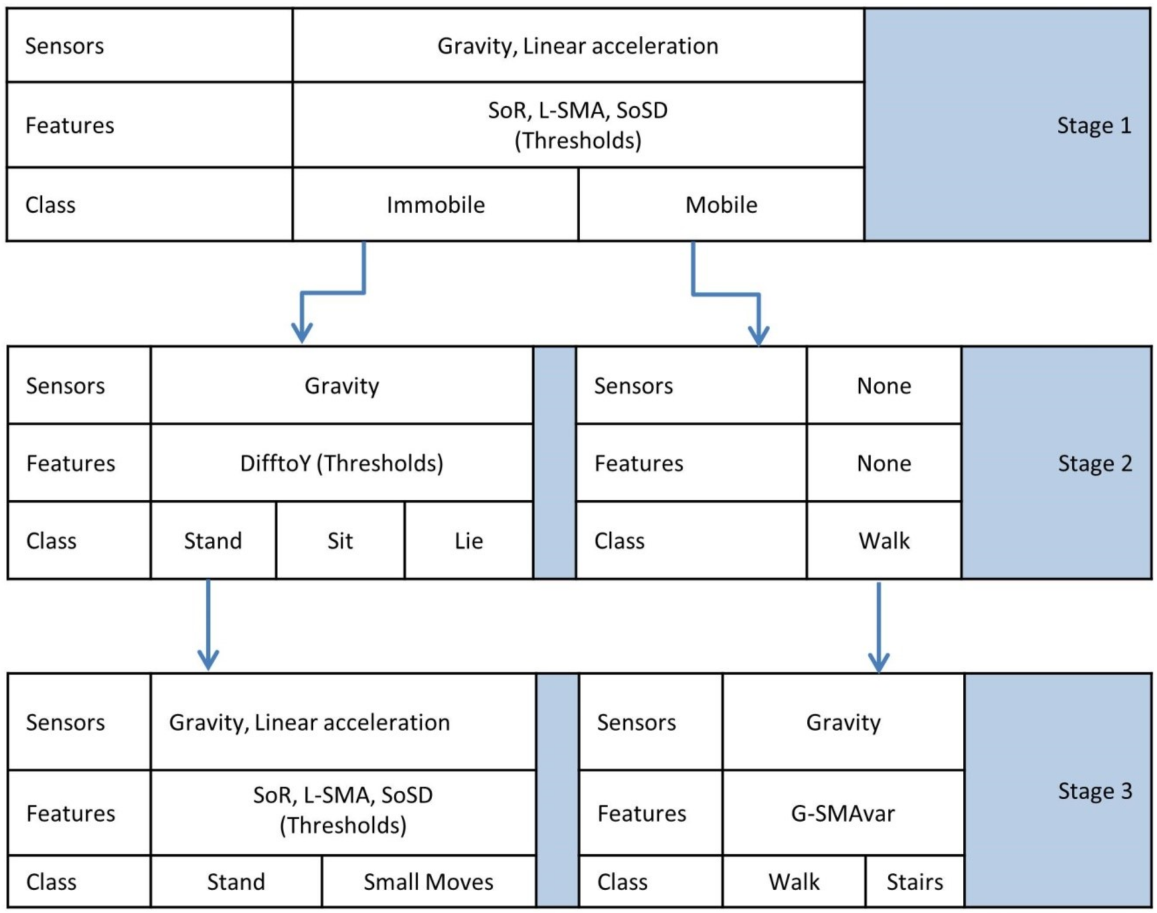

Figure 6.1: Hierarchal Structure of WMMS

The classifier has three activity stages. The first stage used a combination of L-SMA, SoR, and SoSD to identify if a person was mobile (walking, climbing stairs) or immobile (sitting, standing, lying down, or small movements).

At stage 2, pelvis orientation was examined to determine the person's immobile state. If DifftoY was above an 8 m/s^2 threshold, the person was classified as standing (upright). If this feature was below -6 m/s^2, the person was classified as lying down (horizontal). Values between these thresholds were classified as sitting (pelvis angled slightly back).

At stage 3, if the person was standing, a weighting factor was calculated based on how many of L-SMA, SoR, and SoSD passed thresholds. If the person was standing for more than 3 seconds and the weighting factor exceeded 1 for two consecutive data windows, the person was classified as performing a small movement (ex., standing and washing dishes at a sink).

If the person was mobile at stage 3, the default classification was walking. If the person walked for more than 5 seconds and the slope of G-SMA_{var} passed a threshold, then the activity was classified as climbing stairs. In all cases, thresholds were determined from observing signal feature behaviour for a separate young participant data set.

Since the phone's orientation on the pelvis can differ between individuals due to a larger mid-section or different clothing, a rotation matrix method was used to correct for phone orientation [69]. Ten seconds of accelerometer data were collected while the participant was standing still and a 1 second data segment with the smallest standard deviation was used to calculate the rotation matrix constants. The orientation correction matrix was applied to all sensor data. This ten second sample was also used to establish the person's initial standing state, for context awareness.

6.3.3 Classifier with transition state identification and context awareness

A combination of context awareness and transition state recognition (WT) was added to the classifier to improve sit-stand differentiation. A simple single feature threshold was used to identify a transition state and, in combination with knowledge of the previous state and phone orientation, was used to determine the immobile state (Figure 6.2).

The original algorithm (WOT) used a threshold of -6 m/s^2 to differentiate between sitting and lying down (i.e., if difftoY was less than -6, the person was identified as lying down). In the current research, the younger group tended to wear their pants lower on their waist than elderly or stroke groups, lowering the phone's position. This caused the phone to be more horizontal when sitting than in older populations, making it difficult to distinguish between sitting and lying states. Since the proposed algorithm (WT) does not rely solely on phone orientation to identify immobile states, a difftoY threshold representing a more horizontal orientation (less than -6 m/s^2) was calculated.

The gravity vector's X, Y and Z components were expected to sum to 9.81 m/s^2, however this was not the case in practice. The average value of the sum of gravity components was 9.57 m/s^2, ranging from -10.93 to 16.83 m/s^2. The smartphone's gravity vector output was generated by sensor fusion using the linear acceleration output. Linear acceleration is calculated by combining inputs from multiple device sensors [88]. The sequence of calculations may have introduced error, causing a wider range than expected. As a result, the difftoY range was -16.32 to 16.46 m/s^2.

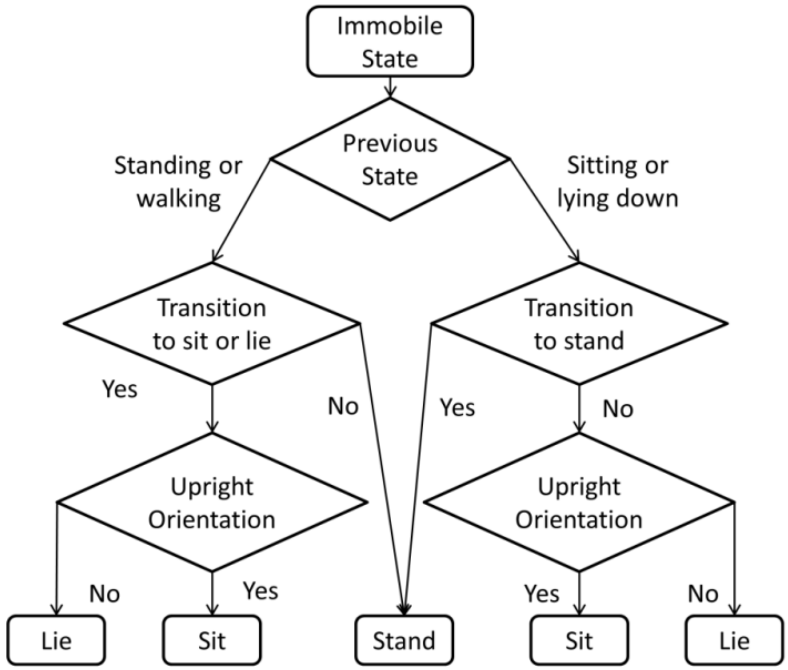

Figure 6.2: Immobile state identification

Data from five younger people were used to select an orientation threshold to identify lying versus sitting. The data were classified while varying the threshold from the minimum to the maximum, in increments of 0.01, and achieved the highest accuracy using a threshold value of -9.96 m/s^2. A final threshold value of -9.81 m/s^2 was chosen for the WT algorithm to allow for some variation and to match the theoretically horizontal position.

The gravity vector along the X and Z axes spiked when transitioning into or out of a sitting or lying state. These spikes were more prominent when transitioning into the sedentary states than transitioning out of the sedentary states since getting up from a chair or bed was often performed more slowly. As a person sits, they bend their trunk forward. At this time, the Y axis gravity component is very low while the gravity components along the X and Z axes increase. Similar results were found when a person lies down. Thus, the gravity range difference feature was created to identify a stand-sit or stand-lie transition.

If the classifier identified an immobile state and the previous state was standing or walking, the maximum gravity range difference over the previous 10 windows was calculated. This time frame accounted for a person shifting in their seat or pulling up their chair to a table before being classified as immobile, while still capturing the transition from stand-sit or stand-lie. If the maximum GRD passed a threshold of 5.5, a stand-sit or stand-lie transition was identified and the phone orientation was used to determine the state. Upright phone orientation was classified as sitting and horizontal was classified as lying. The 5.5 threshold provided the most accurate stand-sit transition, based on data from five senior participants. The senior participants were chosen since Del Rosario et al. [39] demonstrated that classifiers designed with data from elderly participants performed better on young participants than classifiers trained with data from young people and tested on elderly.

Since sit-stand and lie-stand transitions were more variable than stand-sit or stand-lie and the gravity Y-axis component varied, gravity range difference could not be reliably used for a transition threshold. Instead, the "sum of the range of the X and Z gravity vectors" (G-XZ) was used to identify sit-stand and lie-stand transitions. A 1.5 m/s^2 threshold, based on data analysis from the five senior participants, correctly identified transition states without being falsely triggered by small movements (i.e. shifting in their chair, reaching while seated, etc.).

The algorithm used context awareness of the previous state and single feature thresholds to identify transitions. This approach, combined with the phone orientation, was used to determine whether the person was sitting, standing or lying.

6.4 Evaluation

6.4.1 Population

A convenience sample of 15 young (age 26 ± 8.9 years, height 173.9 ± 11.4 cm, weight 68.9 ± 11.1 kg) and 15 seniors (age 73 ± 5.9 years, height 166.3 ± 9.4 cm, weight 68.8 ± 14.7 kg) were recruited from staff and volunteers at The Ottawa Hospital. Fifteen stroke patients (age 55± 10.8 years, height 171.6± 5.79cm, weight 80.7± 9.65 kg) were recruited from the University Rehabilitation Institute in Ljubljana, Slovenia. Stroke participants who were capable of safely completing the mobility tasks were identified by a physical therapist and rehabilitation medicine specialist. Nine stroke patients had right hemiparesis and six had left hemiparesis. Thirteen stroke patients had ischemic stroke, one subarachnoid hemorrhage, and one had impairment because of a benign cerebral tumor. One stroke patient used an ankle-foot orthosis, six used one crutch, and two had one arm in a sling. The stroke event averaged 9.6 months before the study and the average functional independence measure (FIM) [89] score was 107 points. All older participants were capable of completing the mobility tasks. One senior participant had a surgery scheduled for their left leg and walked with a limp and another participant had his kneecap replaced in 2013,wore foot orthoses, and had a bunion on his left foot, leading to caution when using his left foot. Another participant had hip arthritis, which resulted in slower movements. The study was approved by the Ottawa Health Science Network Research Ethics Board and the Ethics Board of University Rehabilitation Institute (Ljubljana, Slovenia). All participants provided informed consent.

6.4.2 Protocol

Participant age, sex, height, and weight were recorded before the trial. A Blackberry Z10 smartphone was placed in a holster attached to the participant's right-front belt or pant waist, with the camera pointed forward. Accelerometer, magnetometer, and gyroscope data were collected at approximately 50Hz (smartphones can have a variable sampling rate [68]). Trials were video recorded using a separate smartphone to establish a gold standard for activity timing comparison. This gold standard time was synchronized with the smartphone sensor output by shaking the phone at the beginning and end of the trial, providing a recognizable accelerometer signal and video event. Activity states were manually labelled from the video recordings. Each 1 second feature window was considered an occurrence. For example, if a person was sitting for 5 seconds, this was considered 5 occurrences of sitting.

Data collection took place under realistic but controlled conditions, providing a continuous data set of activities of daily living. Participants followed a predefined path and performed a predetermined sequence of activities; including, standing, walking, sitting, riding an elevator, brushing teeth, combing hair, washing hands, drying hands, setting dishes, filling a kettle with water, toasting bread, simulated meal at a dining table, washing dishes, walking on stairs, lying on a bed, and walking outdoors [38].

6.5 Results

Sensitivity, specificity and F-score for sit, stand, and lie classification, with and without transition recognition and context awareness, are shown in Table 6.1 – 6.3. Table 6.4 shows confusion tables for each population. P-values indicate the confidence that WT classifier differs from WOT. Sit recognition sensitivity significantly increased by 26% for the young population (from 64.6 to 90.6%).

No other outcomes changed significantly; however, all stand and sit outcomes improved. Lie sensitivity and F-score decreased, though not significantly.

Sensitivity and F-score for sit recognition significantly increased for the elderly participants (30.4% and 24.6% increase, respectively). Specificity and F-score increased for standing (5.8% and 11.4% increase respectively). All outcomes improved for the stroke population, though no improvements were statistically significant.

Table 6.1: Sensitivity, specificity, F-score for young participants with (WT) and without (WOT) transition recognition (standard deviation in brackets and significant differences in bold) and p-value comparing WOT and WT

Activity		Sensitivity (%)	Specificity (%)	F-score
Stand	WOT	90.3 (16.8)	98.7 (2.4)	91.7 (13.7)
	WT	90.5 (13.7)	99.4 (1.5)	93.2 (9.5)
	p	0.97	0.36	0.73
Sit	WOT	**64.6 (40.8)**	98.3 (3.6)	67.3 (40.5)
	WT	**90.6 (20.8)**	97.4 (4.2)	84.7 (20.5)
	p	**0.04**	0.54	0.15
Lie	WOT	94.3 (8.6)	97.9 (3.8)	87.1 (16.5)
	WT	83.1(34.0)	99.7 (0.9)	83.0 (34.1)
	p	0.24	0.09	0.68

Table 6.2: Sensitivity, specificity, F-score for elderly participants with (WT) and without (WOT) transition recognition (standard deviation in brackets and significant changes in bold) and p-value comparing WOT and WT

Activity		Sensitivity (%)	Specificity (%)	F-score
Stand	WOT	94.7 (11.8)	90.2 (8.2)	**79.1 (14.6)**
	WT	97.2 (4.7)	96.0 (6.0)	**90.5 (11.2)**
	p	0.22	0.07	**0.02**
Sit	WOT	**53.2 (38.0)**	99.3 (1.9)	**60.4 (32.4)**
	WT	**83.6 (33.0)**	99.7 (0.7)	**85.0 (29.6)**
	p	**0.02**	0.3	**0.02**
Lie	WOT	90.5 (25.4)	100.0 (0.0)	91.8 (25.5)
	WT	97.03 (4.3)	100.0 (0.0)	98.4 (2.3)
	p	0.17	0.33	0.17

Table 6.3: Sensitivity, specificity, F-score for stroke participants with (WT) and without (WOT) transition recognition (standard deviation in brackets) and p-value comparing WOT and WT.

Activity		Sensitivity (%)	Specificity (%)	F-score
Stand	WOT	82.5 (24.3)	94.0 (24.1)	70.1 (24.4)
	WT	85.9 (26.4)	96.0 (24.3)	77 (25.4)
	p	0.73	0.83	0.47
Sit	WOT	53.3 (37.3)	98.7 (24.8)	56.8 (29.3)
	WT	63.0 (38.9)	99.1 (24.8)	66.7 (33.0)
	p	0.51	0.96	0.41
Lie	WOT	79.4 (38.7)	100 (25.0)	82.4 (38.6)
	WT	94.8 (24.1)	100 (25.0)	97.27 (24.4)
	p	0.22	1.00	0.23

Table 6.4: Confusion tables for young, older, and stroke populations without (WOT) and with (WT) context awareness, transition states and new difftoY threshold

	WOT			WT		
	Stand	Sit	Lie	Stand	Sit	Lie
Young Population						
Stand	1237	53	0	1252	21	0
Sit	92	364	8	77	510	57
Lie	0	129	396	0	15	347
Senior Population						
Stand	1412	569	42	1444	200	5
Sit	48	693	3	16	1062	8
Lie	0	0	491	0	0	523
Stroke Population						
Stand	1072	673	84	1139	483	0
Sit	136	568	4	69	758	11
Lie	0	0	370	0	0	447
All Populations						
Stand	3721	1295	126	3835	704	5
Sit	276	1625	15	162	2330	76
Lie	0	129	1257	0	15	1317

6.6 Discussion

The WT custom hierarchal classifier improved sit and stand recognition compared to the WOT classifier. Specifically, sit sensitivity was significantly higher for the young and senior populations and higher for the stroke population when using the WT classifier. The WOT classifier identified immobile states using only the phone's tilt angle, which remained primarily upright during sitting and standing. This resulted in many sit states being identified as standing, hence low sit sensitivity.

The WT classifier achieved higher stand specificity than WOT for all populations, with fewer sitting states falsely classified as standing. WT stand sensitivity was also higher for all populations, implying that the increase in sit sensitivity did not cause an increase in standing states falsely classified as sitting (sit false positives). This can be seen in the confusion table (Table 5.4). The dinner table sitting sequence was not detected by the WT classifier for three senior participants, compared to eight with the WOT classifier. This stand-sit transition is especially difficult to detect since it involves the participant pulling out a chair, sitting, and adjusting the chair at the table. These movements can be performed in different ways, thereby introducing variability. One of these three senior participants did not have any sitting states detected by WT and WOT, since the phone remained upright the entire time and transitions were not detected. The GRD feature was within 1 unit of the 5.5 threshold, but did not cross the threshold, for the three participants where sitting was not detected by the WT classifier. People who have minimal pelvis rotation during sit-stand activities can be considered outliers for HAR systems with a single waist mounted sensor. A pre-test screening may be necessary to identify these outliers before long-term WMMS monitoring.

In the stroke population, the WT classifier classified some sitting states as standing for eight participants. However, all but three people had improved or unchanged sit recognition compared to the WOT classifier. The stand-sit transition at the dinner table was difficult to detect in the stroke population and caused most of the sit false negatives. One person was classified as sitting at the dinner table, but not when sitting on a chair that did not have a table nearby (i.e., sitting in a chair that was in the middle of a room). When a table or chair arm was unavailable for support when sitting, the participant leaned heavily on his crutch when transitioning onto the chair, which resulted in lower acceleration and thus lower GRD. Therefore, the transition was not recognized. He did not use his crutch while sitting down at the dinner table, therefore this transition was correctly identified. Future work can investigate differences in stand-sit transitions between people who use crutches and those who do not. The stroke patients had individually diverse physical impairments, which introduced different levels of variability in their movements and the resulting sensor signals. This may be why the classifier, which was developed with healthy participant data, did not perform as well with the stroke group. The WT classifier still performed better than WOT with this population. While not statistically significant, a 9.7% improvement in sit sensitivity was observed, bringing sensitivity for the stroke group into the range reported by current classifiers that tested on able-bodied participants. A 15.4% improvement in lie sensitivity was also observed.

The WT sit and stand results were comparable to recent studies involving single, waist worn sensors. Del Rosario et al. [1] reported sit and stand sensitivity ranges from 60.2%-89.0% and 83.2%-91.6% respectively, depending on the data used to train the classifier. Reported sit and stand specificity ranges for young participants were 85.1%-93.7% and 94.0%-97.9%, respectively. Sit and stand sensitivity for elderly participants were 73.6%-78.0% and 54.8%-70.8%, and sit and stand specificity ranged from 85.1%-98.0% and 98.0%-98.9%, respectively. Del Rosario's study [1] was conducted using continuous, realistic data, similar to the current study. The study presented here had superior results for the elderly population and results within the range of those reported by Del Rosario et al. for the young participants.

Sit accuracy in the current study was 92.5% for young participants, 93.1% for senior participants, and 80.6% for stroke participants. Gjoreski et al. reported sit accuracy of 86% [5]. Chavarriaga et al. reported F-measures from 0.41-0.90 when identifying four classes (sit, stand, walk, lie), using four well known pattern classification algorithms (K-nearest neighbor, nearest centroid classifier, linear discriminant analysis, quadratic discriminant analysis) on a continuous sample of realistic activities of daily living [12]. F-scores in the current study were comparable to these values.

Comparing outcomes between studies is difficult since test data and methods are typically unique to each study. The current dataset was continuous and let the individual choose how to perform each activity task; therefore, many activities were performed differently by different people. For example, some people chose to lie on their side and others chose to lie on their back, some people struggled or took longer to pull out a chair, sit down at the dinner table, and pull the chair forward, people of different heights had to lean further forward to reach objects on the dinner table, etc. By allowing individuals to perform activities in their own fashion and with minimal guidance, the dataset contains more variability and is representative of real-life scenarios.

While the inclusion of transition identification and context awareness was designed to improve sit and stand differentiation, sensitivity for lying down also improved in the WT classifier for the senior and stroke populations. For participants who sat down and then transitioned slowly into a lying state, the WOT algorithm identified an immobile state while the person was sitting and classified the state as standing, due to the phone's upright orientation. Since the second transition was not identified, the lying down state was never identified. However, the WT algorithm identified both transitions and correctly classified sitting and lying states. Since there were no false positives for the senior population for lying down, specificity for both WT and WOT was 100. Some participants shifted in the bed, which caused the phone's orientation to change to an upright position. With the WT algorithm, unless a transition to standing was identified, the state remained classified as lying down. The inclusion of transition identification and context awareness also allowed the implementation of a more stringent threshold to determine whether a person was sitting or lying, resulting in improved lie sensitivity in the WT classifier.

Some limitations in the WT classifier can be improved in future work. For example, lie identification was not 100% accurate using only the phone tilt angle. This was noted with the young population, who tended to wear their pants lower on their waist, lowering the phone's position. This caused the phone to sometimes tilt into a primarily horizontal position while sitting, causing sitting to be falsely classified as lying. In addition, the stand-sit transition at the dinner table, when pulling out a chair, was not well classified for the stroke population (sensitivity of 63.0%). Future work can investigate transitions in persons with crutches and arm casts.

The current WT classifier is a custom, threshold-based hierarchal classifier. The results of the current work show that outcomes comparable to current literature can be achieved using a simple classifier that works across age and mobility categories.

6.7 Conclusion

The current research has shown that sit and stand detection can be significantly improved in a simple hierarchal classifier by incorporating context awareness and transition identification, thereby avoiding the use of computationally intensive classifiers. Results obtained were comparable or better than recent work in the literature on able bodied populations. Continuous data collected in a realistic living environment was used to test the classifier, thereby providing a good approximation of real-world use of the WMMS. Future work will evaluate whether additional or other features are necessary to improve classifier performance, without adding computational complexity to the algorithm.

7. Reducing Stair False Positives in a Custom Wearable Mobility Monitoring System

This chapter addresses objective 3 (Improve WMMS performance) by implementing a "covariance of *x* and *z* gravity components" feature to reduce stair false positives, improving stair specificity and walk sensitivity.

This journal manuscript will be submitted for publication.

Capela, NA, Lemaire, ED, Baddour, N. "Reducing Stair False Positives in a Custom Wearable Mobility Monitoring System", 2015.

A new feature was introduced, covariance of *x* and *z* gravity components (equation 7.9). The classifier structure is shown in figure 7.1. This feature was implemented in the WMMS and reduced stair false positives, thereby improving stair specificity and walk sensitivity.

Nicole Capela developed the algorithm, performed data analysis, and was primary author of the manuscript. Edward Lemaire coordinated the study and contributed to data analysis, and writing the manuscript. Natalie Baddour contributed to the analysis and interpretation of data and the manuscript. All authors read and approved the final manuscript. The study was approved by the Ottawa Health Science Network Research Ethics Board and the Ethics Board of University Rehabilitation Institute (Ljubljana, Slovenia).

7.1 Abstract

Human activity recognition (HAR) is important for understanding how a person moves in their daily life. Therapists and rehabilitation specialists would benefit from accurate knowledge of patient activity levels outside of a hospital setting. In 2013, it was found that 91% of adult Americans own cell phones, 61% of whom own smartphones, and that number has increased significantly every year [90]. Smartphones have integrated sensors, such as accelerometers and gyroscopes, which are commonly used for activity recognition.

HAR systems with sensors at one body location have difficulty distinguishing between walking and climbing stairs, since the accelerometer and gyroscope signals produced by these activities are very similar. In this study, a new feature (covariance of *x* and z gravity components) is introduced into a custom waist-worn smartphone wearable mobility monitoring system (WMMS) to improve walk and stair recognition by reducing stair false positives, thereby improving walk sensitivity and stair specificity. The original classifier and the classifier with improved stair recognition are evaluated on a continuous, realistic data set including three distinct populations (young, elderly, stroke patients) in two separate settings (Ottawa, Canada; Ljubljana, Slovenia).

With the new feature, stair specificity and walk sensitivity significantly increased for all populations. Stair sensitivity decreased for all populations; however, stair F-score did not significantly change. Walk F-score significantly increased for senior, stroke, and all populations. Stair false positives were reduced by a total of 3500 instances for all groups, reduced by more than half. Thus, the newly-introduced "covariance of *x* and z gravity components" feature reduces stair false positives while maintaining overall classifier performance, improving stair specificity and walk sensitivity in a smartphone WMMS.

7.2 Introduction

Human activity recognition (HAR) is important for understanding how a person moves in their daily life. Therapists and rehabilitation specialists will benefit from accurate knowledge of a patient activity levels outside the clinic, informing clinical decision-making.

Smartphones are prevalent in society and have integrated sensors, such as accelerometers and gyroscopes, which are commonly used for activity recognition. Therefore, smartphones are an ideal tool for wearable HAR applications since they are regularly worn on the body and are computationally powerful. Some common activities of daily living that are recognized by smartphone HAR systems include standing, sitting, lying down, walking, and climbing stairs. However, smartphone HAR systems with sensors at one body location have difficulty distinguishing between walking and climbing stairs, since the accelerometer and gyroscope signals produced by these activities are similar. Results are better in cases where smartphones have a barometer [39], [52], however not all smartphones possess this sensor, creating a need for a method that can accurately distinguish between walking and stair signals using only accelerometer and gyroscope sensors. Lockhart et al. achieved stair recognition accuracies of 96.4-98.0% when trained and tested on the same subjects [59]. However, that same classifier experienced a drop in stair accuracy to 61.5-68.0% when trained and tested on different individuals. Similarly, Schindhelm et al. found that stair classification accuracy from the same individual who provided the training data (96.4%), was only 28.6% on another participant and 0% on three others [91]. Stair classification outcomes reported in the literature may be overly optimistic since many studies use randomly split 10-fold cross validation on data with many users, meaning that data from the same participants may be used for both training and testing [24], [41]–[44]. Kwapisz et al. reported upstairs recognition accuracy of 27.5 – 61.5% [42], and Catal et al. reported accuracies of 50.16 – 85.13% [41] for various classifiers on the same dataset. However, both used randomly split cross validation. Mandal also used randomly split data and a small sample set of five participants when reporting stair classification accuracy of 71.53% [43]. Other studies reported stair recognition from 60-95%, without providing confusion tables [50], [52], making it difficult to evaluate the number of false positives in these studies.

A classifier can have high overall accuracy and high stair sensitivity while still classifying walking states as stair climbing (i.e., stair false positives). Stair climbing is a more rigorous activity than walking, and an overestimation of stair climbing results in an inaccurate representation of mobility habits. A reduction in stair false positives allows the inclusion of stair activity recognition in a HAR system without adversely affecting walking recognition or giving an exaggerated prediction of daily mobility habits.

Recent studies have noted the importance of using realistic, continuous data sets that include transition states and complex movements when testing HAR systems [39], [56]. This helps to ensure that the performance observed in the laboratory can be expected in a real world. The current study used continuous data from three participant groups in two realistic settings, and used separate groups for threshold setting and evaluation.

The goal of this research is to reduce the number of false positives when detecting stairs by modifying a classifier to have high specificity, rather than high sensitivity. To achieve this, a new feature (covariance of *x* and z gravity components) is introduced to improve walk and stair recognition by reducing stair false positives, thereby improving walk sensitivity and stair specificity.

7.3 Methods

7.3.1 Algorithm

A custom hierarchal classifier was used to identify six activity states: standing, sitting, lying down, walking, stairs, and small movements. The hierarchal structure is shown in Figure 7.1.

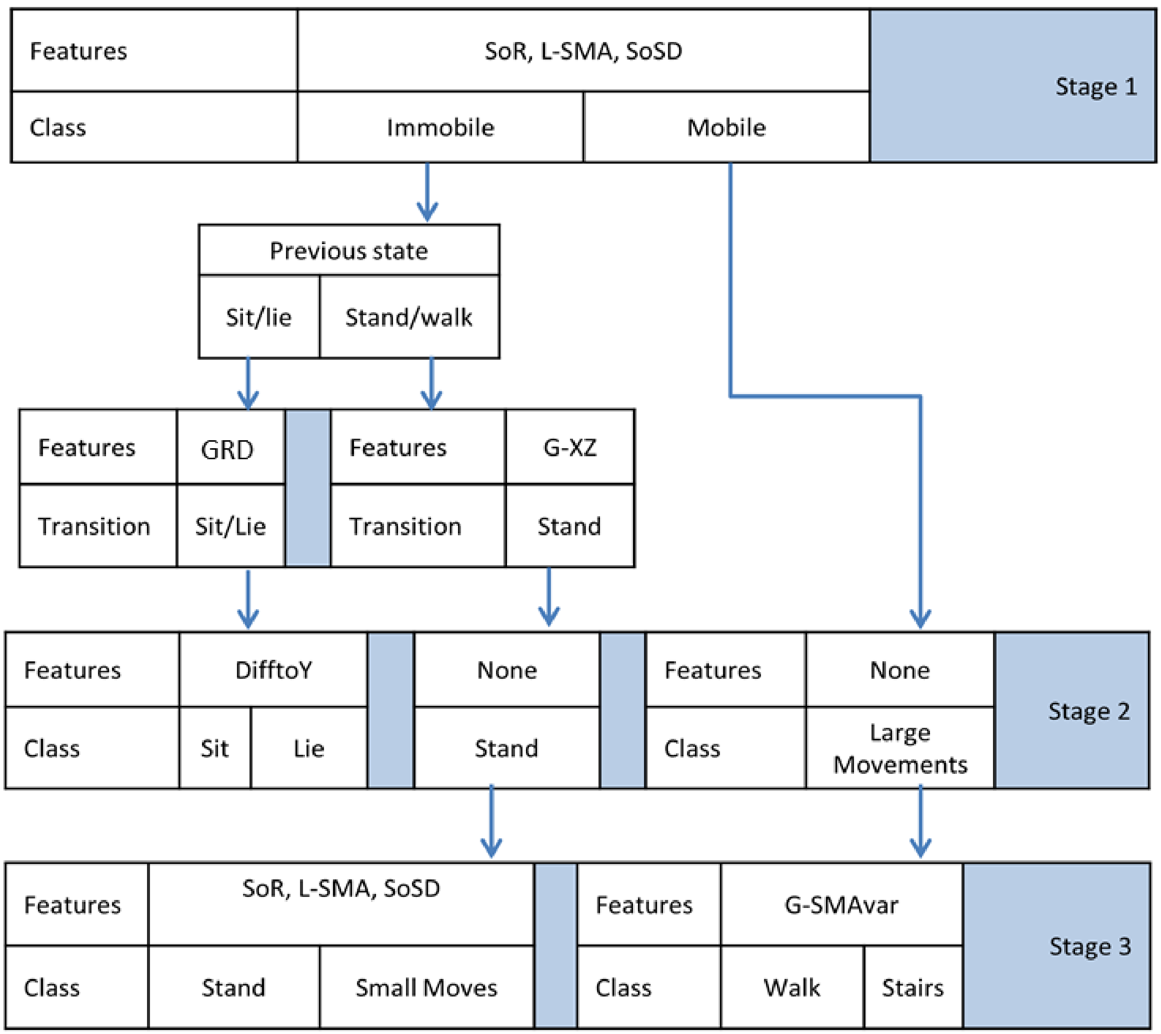

Figure 7.1: Hierarchal Structure of the WMMS

Stage 1 establishes whether the person is mobile or immobile, based on three features derived from the linear acceleration signal (*Xlin*, *Ylin, Zlin*).

Stage 1 Features (SD = standard deviation):

- Sum of Ranges (SoR)

$$range(Xlin_i) + range(Ylin_i) + range(Zlin_i) \tag{6.1}$$

- Simple moving average of sum of range of linear acceleration (L-SMA)

$$\frac{\Sigma^4_{i=1}[range(Xlin_i) + range(Ylin_i) + range(Zlin_i)]}{4} \tag{6.2}$$

- Sum of standard deviation of linear acceleration (SoSD)

$$SD(Xlin_i) + SD(Ylin_i) + SD(Zlin_i) \tag{6.3}$$

If all three of these features pass thresholds, then the window is considered a mobile state. The threshold for SoR is 1, L-SMA is 5, and SoSD is 1. If any of SoR, L-SMA or SoSD do not pass their threshold, the state is classified as immobile.

Stage 2 classifies the person as sitting, standing, lying down, or performing large movements. Large movements include walking, stairs, and other large movements such as shuffling or opening doors. If the state in stage 1 is identified as immobile, the classifier considers the previous window state in the algorithm. From a sitting or lying state, the WMMS checks if the person has transitioned to standing using a "sum of range of the x and z gravity components" feature threshold. From a standing or walking state, the WMMS uses a gravity range difference feature threshold to check for a transition into a sitting or lying state. The DifftoY feature is then used to determine trunk orientation and classify the state as sitting or lying down.

Stage 2 Features:

- Gravity Range Difference (GRD)

$$range(Xgrav) + range(Zgrav) - range(Ygrav) \tag{6.4}$$

- Sum of range of X and Z gravity components (G-XZ)

$$range(Xgrav) + range(Zgrav) \tag{6.5}$$

- Difference to Y-axis (DifftoY)

$$mean(Ygrav_i) - mean(Xgrav_i) - mean(Zgrav_i) \tag{6.6}$$

Stage 3 determines if the person is standing, performing small movements, walking, or climbing stairs. Small movements are classified using the same features as stage 1.

Stage 3 Features:

- Maximum slope of simple moving average of sum of variances of gravity ($G\text{-}SMA_{var}$)

$$SMA_{var} = \frac{\sum_{i=1}^{4}(Var(Xgrav)_i + Var(Ygrav)_i + Var(Zgrav)_i)}{4} \tag{6.7}$$

$$G\text{-}SMA_{var} = max(SMA_{var}(2) - SMA_{var}(1)\ , SMA_{var}(3) - SMA_{var}(2), SMA_{var}(4) - SMA_{var}(3)) \tag{6.8}$$

If two of the features pass their threshold and the person is in an immobile state, then they are classified as performing small movements. Small movements include activities that can be accomplished while standing, sitting, or lying down, such as preparing a meal at a counter or eating a meal at the dinner table. Stairs are classified at stage 3 if the person is walking and the $G\text{-}SMA_{var}$ feature passes a threshold of 8.

Since the phone's orientation on the pelvis can differ between individuals due to differently shaped bodies or clothing choices, a rotation matrix method was used to correct for phone orientation [69]. Ten seconds of accelerometer data were collected while the participant was standing still and a 1-second data segment with the smallest standard deviation was used to calculate the rotation matrix constants. The orientation correction matrix was applied to all sensor data. Previous research by the authors that evaluated the classifier on three groups (young, senior, stroke patients) found that participants were often classified as climbing stairs when they were walking (chapter 3). Thus, the aim of this research was to reduce the number of false positives for stairs, improving the specificity of the classifier for stair recognition. The original classifier (C1) [92] was compared to the improved classifier (C2), which has improved stair recognition.

7.3.2 Stair recognition

In order to improve stair and walk differentiation in C2, it is desirable to identify features that are specific to stairs, meaning that they correspond to stair climbing only, and will therefore not cause false positives. In previous research [93]], filter feature selection methods were used to identify four features that were useful for differentiating between walking and stairs. Two of these features were DifftoY and G-SMA$_{var}$, which were used in C1. The other two features are "covariance of X and Z elements of gravity vector (Cov$_{XZ}$)" and "moving average of the skewness of the rotated *y* linear acceleration (YMA$_{skew}$)".

- Covariance of X and Z elements of gravity vector (Cov$_{XZ}$)

$$Cov_{XZ} = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \mu_X) * (Z_i - \mu_Z) \tag{6.9}$$

- Moving average of the skewness of the rotated *y* linear acceleration (YMA$_{skew}$)

$$s_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)^3} \tag{6.10}$$

$$YMA_{skew} = \frac{\sum_{i=i-6}^{i} s_i}{7} \tag{6.11}$$

Sample data from 15 of the participants (5 young, 5 senior, 5 stroke) were used to calculate thresholds and evaluate the best number of features to use in combination in C2. Each of the four selected features was evaluated on its own and in combination with the other features, varying the thresholds from the feature minimum to the feature maximum. The accuracy (true positives/total stair instances) and the number of false positives were calculated. Only data containing walking and stair activities were used to select thresholds (i.e. immobile state data was not included). The data sample contained 353 instances of stair activity. Since the aim of this research was to reduce false positives, it was decided that a cutoff value of 176 would be used, which corresponded to having less false positives than half of all stair cases.

The maximum accuracy obtained with less than 176 false positives was 21%, obtained using two features in combination (YMA_{skew} , Cov_{XZ}). However, when this combination of features and thresholds was implemented on the data from the remaining subjects, no stair instance was identified. Thus, the feature combination that obtained the second best accuracy with less than 176 false positives was used to design C2. This accuracy (14.7%) occurred when using a single feature (Cov_{XZ}) and a threshold of -10.6. Cov_{XZ} describes how the gravity signal variance along the *x* and *z* axes change together, which would decrease when the person moves on an incline, such as a ramp or staircase. Since only one feature had the best results, the other features selected may be sensitive to stair climbing activities, but not specific to stairs, and including them did not increase classifier accuracy without increasing false positives. C1 and C2 were evaluated on a different sample of 10 people from each participant group (30 people) to compare stair recognition methods.

7.3.3 Evaluation

A BlackBerry Z10 smartphone was used to collect accelerometer, gyroscope, and magnetometer data at approximately 50Hz (smartphones can have a variable sampling rate [68]). Participants wore the smartphone in a holster on the hip while performing a designated series of activities of daily living. A video of the trial was recorded using a separate smartphone and used to identify activity times for gold standard comparison. The data was collected in one continuous trial, including transition data. Participants were not instructed on how to perform the different tasks, which lead to inter-subject variability. Data was collected in two locations: The Ottawa Hospital Rehabilitation Centre and the University Rehabilitation Institute (Ljubljana, Slovenia). The study was approved by the Ottawa Health Science Network Research Ethics Board and the Ethics Board of University Rehabilitation Institute. All participants provided informed consent.

The collected data were used to calculate features, which are abstractions of raw data (e.g., mean, standard deviation etc.). These features were calculated over short sliding windows (1 second, no overlap) to allow a fast response in real time and to improve detection of short duration movements [38]. A one second window results in a good trade-off between speed and accuracy [94]. The activities were identified in the video and time synchronized to the sensor data. Each one second of the activity was considered an occurrence (i.e. 5 seconds of sitting is equal to 5 instances).

A convenience sample of 15 young (age 26 ± 8.9 yrs, height 173.9 ± 11.4 cm, weight 68.9 ± 11.1 kg), 17 senior volunteers (age 75 ± 6.7 yrs, height 165.9 ± 9.2 cm, weight 66.2 ± 14.1 kg), and 15 stroke patients (age 55± 10.8 yrs, height 171.6± 5.79 cm, weight 80.7± 9.65 kg) participated in this study.

Sensitivity, Specificity and F-score were calculated for C1 and C2 for each population and for all populations together. The difference and p-value of sensitivity, specificity and F-score between C1 and C2 were calculated to compare classifier performance. The number of stair false positives and false negatives were also counted for C1 and C2 on all populations.

7.4 Results

As shown in Table 7.1, stair specificity and walk sensitivity significantly increased for all populations. However, stair sensitivity significantly decreased for all populations. F-score did not significantly change for stairs but significantly increased for senior, stroke, and all populations for walking.

Table 7.1: Sensitivity, specificity, and F-score for walk and stair using C1 and C2

		Stair			**Walk**		
Population		**Sensitivity**	**Specificity**	**F-score**	**Sensitivity**	**Specificity**	**F-score**
Able	C1	0.719	0.804	0.151	0.643	0.928	0.732
	C2	0.278	0.950	0.153	0.900	0.897	0.877
	Difference	-0.441	0.146	0.002	0.257	-0.031	0.145
	p-value	**0.031**	**0.023**	0.984	**0.035**	0.286	0.115
Senior	C1	0.808	0.744	0.162	0.503	0.900	0.614
	C2	0.261	0.929	0.132	0.859	0.866	0.856
	Difference	-0.547	0.186	-0.030	0.357	-0.034	0.241
	p-value	**0.001**	**<0.0005**	0.660	**<0.0005**	**0.043**	**0.001**
Stroke	C1	0.622	0.681	0.087	0.517	0.900	0.654
	C2	0.306	0.841	0.055	0.758	0.868	0.825
	Difference	-0.316	0.160	-0.033	0.241	-0.032	0.172
	p-value	**0.042**	**<0.0005**	0.236	**<0.0005**	0.470	**<0.0005**
All	C1	0.716	0.743	0.134	0.554	0.909	0.667
	C2	0.282	0.907	0.113	0.839	0.877	0.853
	Difference	-0.435	0.164	-0.020	0.285	-0.032	0.186
	p-value	**<0.0005**	**<0.0005**	0.628	**<0.0005**	0.073	**<0.0005**

As shown in Table 6.2, the number of false positives was reduced by a total of 3500 instances for all groups. False negatives increased by 273.

Table 7.2: False positives and false negatives for stairs using C1 and C2

		False positives	False Negatives
Able	C1	943	37
	C2	282	88
	Difference	-661	51
Senior	C1	1759	57
	C2	494	168
	Difference	-1265	111
Stroke	C1	2982	149
	C2	1408	260
	Difference	-1574	111
All	C1	5684	243
	C2	2184	516
	Difference	-3500	273

Table 7.3: Confusion tables using C1 and C2 for all populations

	Classified as	Activity					
		Stand	Sit	Lie	Walk	Stair	Small movements
C1	Stand	2384	404	0	53	0	1629
	Sit	84	1493	72	9	0	91
	Lie	0	3	829	0	0	0
	Walk	100	27	20	6580	235	479
	Stair	6	3	12	5522	389	50
	Small movements	122	65	0	3	0	416
C2	Stand	2384	404	0	53	0	1629
	Sit	84	1493	72	9	0	91
	Lie	0	3	829	0	0	0
	Walk	104	30	29	9867	494	502
	Stair	2	0	3	2235	130	27
	Small movements	122	65	0	3	0	416

Table 7.3 shows the confusion table using C1 and C2. The columns represent the true activity and the row indicates the classified activity. As the confusion table shows, the percentage of walk instances that were falsely classified as stair was reduced from 45% to 18%. This occurs at the cost of stair sensitivity; however, the reduction in stair false positives corresponds with the aims of this study, since we defined the cost of stair false positives as higher than false negatives.

7.5 Discussion

In this study, we considered the cost of misclassifying a walk as a stair (stair FP, walk FN) as more costly than misclassifying a stair as a walk (walk FP stair FN) since stair climbing is less frequent than regular walking in a typical person's day and stair ascent requires more energy expenditure. The average rate of oxygen consumption (VO_2) during stair ascension is 33.5 mL/kg/min [95] compared to 18.4 mL/kg/min for fast walking on level ground and 12.1 mL/kg/min for customary walking speeds on level ground [96]. This is a ratio of stair-walk energy expenditure of 1.82 for fast walking and 2.77 for customary walking. Thus, over-counting stair climbing instances in a day would lead to a drastically different portrait of a person's daily activities than an underestimation. In other words, it is more costly for a therapist to believe the person is climbing a lot of stairs than to believe that they are not climbing as many. This is especially true since a person typically spends much more time walking than they do climbing stairs.

Stair specificity and walk sensitivity were significantly improved for all populations using the covariance of X and Z elements of the gravity vector to identify stair climbing. The number of false positives was reduced by a total of 3500 instances; however, this resulted in stair sensitivity significantly decreasing for all populations. The F-score did not significantly change for stairs, indicating that the overall performance was not necessarily better or worse, while still increasing stair specificity and walk sensitivity, as was the objective. F-score significantly increased for senior, stroke and all populations for walking. It is worth noting that the specificity increase in stairs had higher levels of significance (lower p-values) than the decreases in sensitivity.

Considering accuracy, calculated as (TP+TN)/(TP+FP+FN+TN) and including all classes, the overall walk classification accuracy improved from 65.7-77.7% for C1 to 80.7-90.1% for C2 and the accuracy for stairs improved from 67.7-80.2% for C1 to 82.4-93.2% for C2. While the improvement is worth noting, the accuracy value itself is misleading. Due to the class imbalance, the number of true negatives is invariably much higher than the other values, thereby influencing the accuracy value. The confusion tables are more telling, displaying how the activities are classified. While false negatives increased by 273 instances, the cost of these false negatives was defined in this study as less than the cost of the stair false positives. As shown in Table 6.3, the percentage of walk instances that were falsely classified as stair was reduced from 45% to 18%. This reduction was most prominent in the senior population (51.8 to 15.4% reduction). The young group had a reduction of 36.4 to 11.0% and the stroke group had a reduction of 45.8 to 22.8%.

Wu et al. reported walk accuracy from 55.7-94.1% , which is a wider range than the results of the current study, and stair ascent accuracy from 19.7-70.4% [24], which is lower than the results presented here. In addition, Wu et al. segmented the data into activities during pre-processing and trimmed off excess data. This simplifies the classification task. Liang et al. reported a walk accuracy of 80% and stair accuracy of 82%, which matches the lower range of the current study. The percent of walk instances falsely classified as stairs was 10%, which is very close to the value obtained for young participants in this study. Liang used 5-fold cross validation and offline data training. As mentioned earlier, Lockhart et al. achieved accuracies of 56.8-73.0% for walking and 61.5-68.0% for stairs when training and testing on separate individuals [59]. The percent of walk instances falsely classified as stairs was 24.1%., which is higher than the results obtained in the current study. Mandal et al. only had 9.4% of walk classified as stairs, however they had a very small sample set of five subjects [43].

The number of walk instances falsely classified as stairs was reduced by 27% over all populations to 18%. The percent of walk instances classified as stairs was the highest for the stroke population at 23%. Following a stroke, gait is typically more variable; therefore, gravity values along the x and z axes may differ more in level walking than they would for a young or older participant (i.e., their pelvis movement may change orientation differently). In addition, two stroke patients used one crutch and one used an ankle-foot orthosis, which affects the direction of movement of the pelvis during walking.

There are some limitations to the current study. The phone was worn in a holster on the person's waist, which may have made stair recognition more difficult. Wearing the phone in a pocket could provide better insight into thigh angle during daily activities, improving stair recognition in future studies. Gold standard annotation can be difficult, with multiple video frames possible as the start or end of an activity. A tolerance of 2 windows (2 seconds) around a change of state was used in the comparison of classification results to gold standard classes; however, some individuals may take longer to transition between activities. This explains how some immobile states were classified as mobile and vice versa. Further research with more participants could give a more accurate evaluation of the classifier performance.

7.6 Conclusion

This study improved walk and stair recognition in a custom smartphone WMMS using "covariance of the x and z gravity vector axes" as a feature to identify stair activities. Stair and walk accuracy achieved was similar to the better results in the literature and was evaluated on a large set of continuous data from able bodied, senior, and stroke participant groups, in two settings. The number of stair false positives was reduced by 27%, improving walk sensitivity and stair specificity.

Future work will examine other potential features to increase stair sensitivity without increasing false positives and expand the data set to test the classification method on more populations with differing mobility characteristics.

8. Thesis Conclusions and Future Work

This thesis improved the performance of a waist-worn smartphone WMMS across populations with varying mobility characteristics. Smartphone human activity recognition was compared between stroke and able-bodied participants. Signal features that were useful for waist worn smartphone activity recognition were selected independently of a classifier. Lastly, the newly selected features and previous state awareness were implemented into a custom classifier to improve immobile state recognition and reduce stair false positives. Each objective and its associated hypotheses are discussed below.

8.1 Objective 1: Evaluate WMMS on stroke and able-bodied participants

8.1.1 Hypothesis: Performance of the WMMS system developed for able-bodied participants will be worse for the stroke population.

Consistent with the hypothesis, WMMS system performance was worse for stroke participants at higher detail levels; however, performance was approximately equal for both populations at a broad classification level. This research demonstrated that a smartphone-based WMMS approach can provide relevant information on human movement activities for both able-bodied and stroke populations, although sensitivity and specificity decreased as the classification tasks became more complex.

The hierarchical classifier performed well for both able-bodied and stroke participants when differentiating between mobile or immobile states (sensitivity, specificity, and F score >92%). However, classifier performance decreased when classifying activities at a higher detail level (sit, stand, lie, large movements). Specificity and F-score were significantly lower for the stroke participants than the able-bodied group. Sit, stand and lie were classified based on inclination angle and stroke can cause posture asymmetry during standing that affects pelvis inclination. The stroke population was also much older than the able-bodied sample, and age can affect posture. Additionally, walk detection had lower sensitivity and F-score for the stroke group, possibly due to hemiparesis in certain participants and the use of walking aids. These results demonstrated the need to develop HAR systems using data from the intended user population, since systems developed with able-bodied participants exhibited poor performance on participants with differing mobility characteristics.

8.2 Objective 2: Select signal features for WMMS independent of a classifier

8.2.1 Hypothesis 1: Features selected with data from populations with differing mobility levels differ.

Consistent with the hypothesis, different features were selected with data from different populations. For example, features selected for the stroke group differed from the able-bodied and elderly groups, possibly due to the stroke group inhomogeneity, whose mobility levels and use of crutches and arm slings varied. However, some features were identified that may be applicable across able-bodied, elderly, and stroke populations. In particular, the following signal features were effective across multiple populations:

- acceleration features using simple moving averages and correlations of acceleration along different axes, as well as mean gyroscope output on the *y* and *z* axes, were useful to distinguish between immobile and mobile states;
- gravity signal mean in the phone's forward direction and the difference between the phone's *y* gravity to the *x* and *z* gravity signals were useful to distinguish between sit, stand, and lie;
- gravity signal range in the forward direction was useful to differentiate between sitting and standing;
- skewness of the rotated linear acceleration along the *y*-axis was useful for classifying stair climbing from other large movements;
- acceleration covariance was useful to detect ramp and stair ascent or descent;
- mean gyroscope signal on the *y*-axis was useful for detecting small movements.

These results can guide smartphone WMMS system development for targeted users, regardless of the type of classifier.

8.2.2 Hypothesis 2: Classifier accuracy does not degrade when using a selected subset of features instead of the entire set of features.

Consistent with the hypothesis, evaluation of selected subsets with generic classifiers showed improvements in activity recognition accuracy. Therefore, the features eliminated in the feature selection process were redundant and did not significantly contribute to classifier accuracy. With appropriate feature subset selection, equivalent classifier performance was obtained with a reduced feature set, effectively reducing the computational burden on the WMMS system.

8.3 Objective 3: Assess WMMS improvement after implementing selected features and transition recognition

8.3.1 Hypothesis 1: Transition identification and previous state awareness improves WMMS outcomes on all populations when differentiating between immobile states.

Consistent with the hypothesis, sit, stand and lie classification improved significantly by incorporating context awareness and transition identification, without the use of computationally expensive classifiers. Sit and stand results were comparable to recent studies involving single, waist worn sensors and continuous data collection. These results showed that results consistent with the literature can be obtained without the use of computationally intensive and machine learned algorithms.

8.3.2 Hypothesis 2: New features reduce false positives for stair activities, improving walk recognition sensitivity and stair recognition specificity.

Consistent with the hypothesis, implementation of the 'covariance of x and z elements of the gravity vector' feature reduced the number of stair false positives, thereby improving stair specificity and walk sensitivity. This provides a more accurate representation of a person's daily mobility habits since a person spends a much smaller portion of their day climbing stairs than they do walking.

8.4 Future work

The research contributed to the literature on smartphone-based human activity recognition. The work also identified areas in need of improvement and exploration.

1. Better methods are needed for gold file annotation, taking into account individual differences in small movement performance.
2. Since this study only used one smartphone model for testing, algorithm performance could be evaluated with other smartphone based systems.
3. For signal feature selection, other activities that produce vertical acceleration (i.e., jumping, hopping, jogging, etc.) could be included to verify if the selected features remain viable. While the feature selection methods were designed to compensate for class imbalances in the feature set, selective sampling before performing feature selection could improve results in future work. Study results could be expanded upon through feature selection with larger data sets and different pathological populations, such as amputees or people with neurological disorders, since different gait patterns may identify additional features to be included in a generalized feature set.
4. Immobile state identification can investigate the difference in stand-sit transitions between people who use crutches and those who do not. Wearing the phone in a pocket could provide better insight into thigh angle during daily activities, improving stair recognition in future studies. A dataset that includes more instances of large movements that are substantially different from walking should be collected for further evaluation of the classifier.
5. Functionality could be added to recognize when the phone is in use for calling, texting, or games, and suspend the activity recognition algorithm for that time. Different algorithm design could be used for different populations. This could allow someone to enter a custom input, such as using a crutch on the right side, and allow the phone to select an algorithm appropriate for their specific mobility level without requiring them to go through a calibration phase.

In general, HAR system research would benefit from open source, large databases of sensor data that would allow researchers to compare their classifier performance on common signals and against a common baseline. Reporting detailed outcomes (i.e., sensitivity, specificity, confusion tables) is essential for an accurate assessment of classifier performance, as opposed to accuracy. Investigation into online implementation and battery consumption could help to move HAR classifier design toward more efficient and effective design. More research involving elderly and mobility challenged patients will improve HAR system performance for the populations who stand to benefit from this technology.

References

[1] O. D. Incel, M. Kose, and C. Ersoy, “A Review and Taxonomy of Activity Recognition on Mobile Phones,” *BioNanoScience*, vol. 3, no. 2, pp. 145–171, Jun. 2013.

[2] O. D. Lara and M. . Labrador, “A Survey on Human Activity Recognition using Wearable Sensors,” *IEEE Commun. Surv. Tutor.*, vol. 15, no. 3, pp. 1192–1209, Third 2013.

[3] C.-C. Yang and Y.-L. Hsu, “A review of accelerometry-based wearable motion detectors for physical activity monitoring,” *Sensors*, vol. 10, no. 8, pp. 7772–7788, 2010.

[4] A. Bulling, U. Blanke, and B. Schiele, “A Tutorial on Human Activity Recognition Using Body-worn Inertial Sensors,” *ACM Comput Surv*, vol. 46, no. 3, pp. 33:1–33:33, Jan. 2014.

[5] K. Khoshelham and S. O. Elberink, “Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications,” *Sensors*, vol. 12, no. 2, pp. 1437–1454, Feb. 2012.

[6] A. Pantelopoulos and N. G. Bourbakis, “A Survey on Wearable Sensor-Based Systems for Health Monitoring and Prognosis,” *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 40, no. 1, pp. 1–12, Jan. 2010.

[7] A. Hecht, S. Ma, J. Porszasz, and R. Casaburi, “Methodology for Using Long-Term Accelerometry Monitoring to Describe Daily Activity Patterns in COPD,” *COPD*, vol. 6, no. 2, pp. 121–129, Apr. 2009.

[8] J. Takacs, C. L. Pollock, J. R. Guenther, M. Bahar, C. Napier, and M. A. Hunt, “Validation of the Fitbit One activity monitor device during treadmill walking,” *J. Sci. Med. Sport*, vol. 17, no. 5, pp. 496–500, Sep. 2014.

[9] D. W. Esliger, A. Probert, S. Connor Gorber, S. Bryan, M. Laviolette, and M. S. Tremblay, “Validity of the Actical accelerometer step-count function,” *Med. Sci. Sports Exerc.*, vol. 39, no. 7, pp. 1200–1204, Jul. 2007.

[10] J. T. Cavanaugh, K. L. Coleman, J. M. Gaines, L. Laing, and M. C. Morey, “Using step activity monitoring to characterize ambulatory activity in community-dwelling older adults,” *J. Am. Geriatr. Soc.*, vol. 55, no. 1, pp. 120–124, Jan. 2007.

[11] Y.-L. Kuo, K. M. Culhane, P. Thomason, O. Tirosh, and R. Baker, “Measuring distance walked and step count in children with cerebral palsy: An evaluation of two portable activity monitors,” *Gait Posture*, vol. 29, no. 2, pp. 304–310, Feb. 2009.

[12] F. Bagalà, C. Becker, A. Cappello, L. Chiari, K. Aminian, J. M. Hausdorff, W. Zijlstra, and J. Klenk, “Evaluation of Accelerometer-Based Fall Detection Algorithms on Real-World Falls,” *PLoS ONE*, vol. 7, no. 5, p. e37062, May 2012.

[13] L. Bao and S. S. Intille, “Activity Recognition from User-Annotated Acceleration Data,” in *Pervasive Computing*, A. Ferscha and F. Mattern, Eds. Springer Berlin Heidelberg, 2004, pp. 1–17.

[14] M.-W. Lee, A. M. Khan, and T.-S. Kim, “A single tri-axial accelerometer-based real-time personal life log system capable of human activity recognition and exercise information generation,” *Pers. Ubiquitous Comput.*, vol. 15, no. 8, pp. 887–898, Jun. 2011.

[15] S.-Y. Chiang, Y.-C. Kan, Y.-C. Tu, and H.-C. Lin, “Activity Recognition by Fuzzy Logic System in Wireless Sensor Network for Physical Therapy,” in *Intelligent Decision Technologies*, J. Watada, T. Watanabe, G. Phillips-Wren, R. J. Howlett, and L. C. Jain, Eds. Springer Berlin Heidelberg, 2012, pp. 191–200.

[16] A. F. Dalton, C. Ní Scanaill, S. Carew, D. Lyons, and G. OLaighin, “A clinical evaluation of a remote mobility monitoring system based on SMS messaging,” *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2007, pp. 2327–2330, 2007.

[17] J. J. Guiry, P. van de Ven, J. Nelson, L. Warmerdam, and H. Riper, “Activity recognition with smartphone support,” *Med. Eng. Phys.*, vol. 36, no. 6, pp. 670–675, Jun. 2014.

[18] A. Mannini, S. S. Intille, M. Rosenberger, A. M. Sabatini, and W. Haskell, “Activity recognition using a single accelerometer placed at the wrist or ankle,” *Med. Sci. Sports Exerc.*, vol. 45, no. 11, pp. 2193–2203, Nov. 2013.

[19] V. H. Cheung, L. Gray, and M. Karunanithi, “Review of accelerometry for determining daily activity among elderly patients,” *Arch. Phys. Med. Rehabil.*, vol. 92, no. 6, pp. 998–1014, Jun. 2011.

[20] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. M. Havinga, “A Survey of Online Activity Recognition Using Mobile Phones,” *Sensors*, vol. 15, no. 1, pp. 2059–2085, Jan. 2015.

[21] H. H. Wu, E. D. Lemaire, and N. Baddour, “Change-of-state determination to recognize mobility activities using a BlackBerry smartphone,” in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society,EMBC*, 2011, pp. 5252–5255.

[22] "ActivityRecognitionApi | Android Developers." [Online]. Available: http://developer.android.com/reference/com/google/android/gms/location/ActivityRecognitionApi.html. [Accessed: 02-Apr-2015].

[23] A. Avci, S. Bosch, M. Marin-Perianu, R. Marin-Perianu, and P. Havinga, "Activity Recognition Using Inertial Sensing for Healthcare, Wellbeing and Sports Applications: A Survey," in *2010 23rd International Conference on Architecture of Computing Systems (ARCS)*, 2010, pp. 1–10.

[24] W. Wu, S. Dasgupta, E. E. Ramirez, C. Peterson, and G. J. Norman, "Classification accuracies of physical activities using smartphone motion sensors," *J. Med. Internet Res.*, vol. 14, no. 5, p. e130, 2012.

[25] M. Zhang and A. A. Sawchuk, "A Feature Selection-based Framework for Human Activity Recognition Using Wearable Multimodal Sensors," in *Proceedings of the 6th International Conference on Body Area Networks*, ICST, Brussels, Belgium, Belgium, 2011, pp. 92–98.

[26] H. Liu and H. Motoda, *Computational Methods of Feature Selection*. CRC Press, 2007.

[27] L. Yu and H. Liu, "Feature selection for high-dimensional data: A fast correlation-based filter solution," 2003, pp. 856–863.

[28] M. A. Hall and L. A. Smith, *Feature Selection for Machine Learning: Comparing a Correlation-based Filter Approach to the Wrapper*. 1999.

[29] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2 edition. New York: Wiley-Interscience, 2000.

[30] P. Gupta and T. Dallas, "Feature Selection and Activity Recognition System Using a Single Triaxial Accelerometer," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 6, pp. 1780–1786, Jun. 2014.

[31] M. Zhang and A. A. Sawchuk, "USC-HAD: A Daily Activity Dataset for Ubiquitous Activity Recognition Using Wearable Sensors," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, New York, NY, USA, 2012, pp. 1036–1043.

[32] L. Atallah, B. Lo, R. King, and G.-Z. Yang, "Sensor Placement for Activity Detection Using Wearable Accelerometers," in *2010 International Conference on Body Sensor Networks (BSN)*, 2010, pp. 24–29.

[33] U. Maurer, A. Smailagic, D. P. Siewiorek, and M. Deisher, "Activity recognition and monitoring using multiple sensors on different body positions," in *International Workshop on Wearable and Implantable Body Sensor Networks, 2006. BSN 2006*, 2006, p. 4 pp.–116.

[34] B. Fish, A. Khan, N. H. Chehade, C. Chien, and G. Pottie, “Feature selection based on mutual information for human activity recognition,” in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 1729–1732.

[35] Y. Saeys, I. Inza, and P. Larrañaga, “A review of feature selection techniques in,” *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, Oct. 2007.

[36] P. Casale, O. Pujol, and P. Radeva, “Human Activity Recognition from Accelerometer Data Using a Wearable Device,” in *Pattern Recognition and Image Analysis*, J. Vitrià, J. M. Sanches, and M. Hernández, Eds. Springer Berlin Heidelberg, 2011, pp. 289–296.

[37] J. M. M. Nuno Cruz Silva, “Features Selection for Human Activity Recognition with iPhone Inertial Sensors,” 2013.

[38] F. R. Allen, E. Ambikairajah, N. H. Lovell, and B. G. Celler, “Classification of a known sequence of motions and postures from accelerometry data using adapted Gaussian mixture models,” *Physiol. Meas.*, vol. 27, no. 10, pp. 935–951, Oct. 2006.

[39] M. B. Del Rosario, K. Wang, J. Wang, Y. Liu, M. Brodie, K. Delbaere, N. H. Lovell, S. R. Lord, and S. J. Redmond, “A comparison of activity classification in younger and older cohorts using a smartphone,” *Physiol. Meas.*, vol. 35, no. 11, pp. 2269–2286, Oct. 2014.

[40] Q. V. Vo, M. T. Hoang, and D. Choi, “Personalization in Mobile Activity Recognition System Using Medoids Clustering Algorithm,” *Int. J. Distrib. Sens. Netw.*, vol. 2013, p. e315841, Jul. 2013.

[41] C. Catal, S. Tufekci, E. Pirmit, and G. Kocabag, “On the use of ensemble of classifiers for accelerometer-based activity recognition,” *Appl. Soft Comput.*

[42] G. M. W. Jennifer R. Kwapisz, “Activity recognition using cell phone accelerometers.,” *SIGKDD Explor.*, vol. 12, pp. 74–82, 2010.

[43] I. Mandal, S. L. Happy, D. P. Behera, and A. Routray, “A framework for human activity recognition based on accelerometer data,” in *Confluence The Next Generation Information Technology Summit (Confluence), 2014 5th International Conference* -, 2014, pp. 600–603.

[44] M. Fahim, I. Fatima, S. Lee, and Y.-T. Park, “EFM: evolutionary fuzzy model for dynamic activities recognition using a smartphone accelerometer,” *Appl. Intell.*, vol. 39, no. 3, pp. 475–488, Oct. 2013.

[45] N. Ravi, N. D, P. Mysore, and M. L. Littman, “Activity recognition from accelerometer data,” in *In Proceedings of the Seventeenth Conference on Innovative Applications of Artificial Intelligence(IAAI*, 2005, pp. 1541–1546.

[46] S. L. Lau and K. David, “Movement recognition using the accelerometer in smartphones,” in *Future Network and Mobile Summit, 2010*, 2010, pp. 1–9.

[47] H. L. Jun Yang, “Physical Activity Recognition with Mobile Phones: Challenges, Methods, and Applications,” pp. 185–213, 2010.

[48] P. Siirtola and J. Röning, “User-Independent Human Activity Recognition Using a Mobile Phone: Offline Recognition vs. Real-Time on Device Recognition,” in *Distributed Computing and Artificial Intelligence*, S. Omatu, J. F. D. P. Santana, S. R. González, J. M. Molina, A. M. Bernardos, and J. M. C. Rodríguez, Eds. Springer Berlin Heidelberg, 2012, pp. 617–627.

[49] A. Rasekh, C.-A. Chen, and Y. Lu, “Human Activity Recognition using Smartphone,” *ArXiv14018212 Cs*, Jan. 2014.

[50] M. T. H. Quang Viet Vo, “Personalization in Mobile Activity Recognition System Using K-Medoids Clustering Algorithm,” *Int. J. Distrib. Sens. Netw.*, vol. 2013, 2013.

[51] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, “Human Activity Recognition on Smartphones Using a Multiclass Hardware-Friendly Support Vector Machine,” in *Ambient Assisted Living and Home Care*, J. Bravo, R. Hervás, and M. Rodríguez, Eds. Springer Berlin Heidelberg, 2012, pp. 216–223.

[52] A. M. Khan, A. Tufail, A. M. Khattak, and T. H. Laine, “Activity Recognition on Smartphones via Sensor-Fusion and KDA-Based SVMs,” *Int. J. Distrib. Sens. Netw.*, vol. 2014, p. e503291, May 2014.

[53] J. Lester, T. Choudhury, and G. Borriello, “A Practical Approach to Recognizing Physical Activities,” in *Pervasive Computing*, K. P. Fishkin, B. Schiele, P. Nixon, and A. Quigley, Eds. Springer Berlin Heidelberg, 2006, pp. 1–16.

[54] D. J. Hand, “Classifier Technology and the Illusion of Progress,” *Stat. Sci.*, vol. 21, no. 1, pp. 1–14, Feb. 2006.

[55] A. Sae-Tang, L. K. McDowell, M. Catasta, and K. Aberer, “Semantic Place Prediction using Mobile Data.” Nokia Mobile data Challenge, Jun-2012.

[56] R. Chavarriaga, H. Sagha, A. Calatroni, S. T. Digumarti, G. Tröster, J. D. R. Millán, and D. Roggen, “The Opportunity Challenge: A Benchmark Database for On-body Sensor-based Activity Recognition,” *Pattern Recogn Lett*, vol. 34, no. 15, pp. 2033–2042, Nov. 2013.

[57] P. Langley, “Crafting Papers on Machine Learning,” in *In Proceedings of the Seventeenth International Conference on Machine Learning (ICML*, 2000, pp. 1207–1212.

[58] J. A. Ward, P. Lukowicz, and H. W. Gellersen, “Performance Metrics for Activity Recognition,” *ACM Trans Intell Syst Technol*, vol. 2, no. 1, pp. 6:1–6:23, Jan. 2011.

[59] G. M. W. Jeffrey W Lockhart, “The Benefits of Personalized Smartphone-Based Activity Recognition Models,” *Proc SIAM Int. Conf. Data Min.*, 2014.

[60] J. L. Reyes-Ortiz, D. Anguita, A. Ghio, and X. Parra, “Human Activity Recognition Using Smartphones Data Set,” *UCI Machine Learning Repository*. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones.

[61] M. Tundo, “Development of a human activity recognition system using inertial measurement unit sensors on a smartphone,” University of Ottawa, Ottawa, 2014.

[62] H.-H. Wu, E. Lemaire, and N. Baddour, “Combining low sampling frequency smartphone sensors and video for a Wearable Mobility Monitoring System,” *F1000Research*, Jul. 2014.

[63] C.-C. Yang and Y.-L. Hsu, “A Review of Accelerometry-Based Wearable Motion Detectors for Physical Activity Monitoring,” *Sensors*, vol. 10, no. 8, pp. 7772–7788, Aug. 2010.

[64] W.-Y. Chung, S. Bhardwaj, A. Punvar, D.-S. Lee, and R. Myllylae, “A fusion health monitoring using ECG and accelerometer sensors for elderly persons at home,” *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2007, pp. 3818–3821, 2007.

[65] Z. Zhao, F. Morstatter, S. Sharma, S. Alelyani, A. Anand, and H. Liu, “Advancing Feature Selection Research-ASU Feature Selection Repository,” Arizona State University, technical.

[66] J. P. Bettger, X. Zhao, C. Bushnell, L. Zimmer, W. Pan, L. S. Williams, and E. D. Peterson, “The association between socioeconomic status and disability after stroke: findings from the Adherence eValuation After Ischemic stroke Longitudinal (AVAIL) registry,” *BMC Public Health*, vol. 14, p. 281, 2014.

[67] “TOHRC Data Logger - BlackBerry World.” [Online]. Available: http://appworld.blackberry.com/webstore/content/60743/?countrycode=CA&lang=en. [Accessed: 08-Oct-2014].

[68] S. Mellone, C. Tacconi, and L. Chiari, “Validity of a Smartphone-based instrumented Timed Up and Go,” *Gait Posture*, vol. 36, no. 1, pp. 163–165, May 2012.

[69] M. D. Tundo, E. Lemaire, and N. Baddour, “Correcting Smartphone orientation for accelerometer-based analysis,” in *2013 IEEE International Symposium on Medical Measurements and Applications Proceedings (MeMeA)*, 2013, pp. 58–62.

[70] Y. He and Y. Li, “Physical Activity Recognition Utilizing the Built-In Kinematic Sensors of a Smartphone,” *Int. J. Distrib. Sens. Netw.*, vol. 2013, p. e481580, Apr. 2013.

[71] L. Gao, A. K. Bourke, and J. Nelson, “Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems,” *Med. Eng. Phys.*, vol. 36, no. 6, pp. 779–785, Jun. 2014.

[72] J. Barra, L. Oujamaa, V. Chauvineau, P. Rougier, and D. Pérennou, “Asymmetric standing posture after stroke is related to a biased egocentric coordinate system,” *Neurology*, vol. 72, no. 18, pp. 1582–1587, May 2009.

[73] B. Ostrowska, K. Rożek-Mróz, and C. Giemza, “Body posture in elderly, physically active males,” *Aging Male*, vol. 6, no. 4, pp. 222–229, Jan. 2003.

[74] G. Hache, E. D. Lemaire, and N. Baddour, “Wearable Mobility Monitoring Using a Multimedia Smartphone Platform,” *IEEE Trans. Instrum. Meas.*, vol. 60, no. 9, pp. 3153–3161, Sep. 2011.

[75] S. Patel, R. Hughes, T. Hester, J. Stein, M. Akay, J. G. Dy, and P. Bonato, “A Novel Approach to Monitor Rehabilitation Outcomes in Stroke Survivors Using Wearable Technology,” *Proc. IEEE*, vol. 98, no. 3, pp. 450–461, Mar. 2010.

[76] C. Mizuike, S. Ohgi, and S. Morita, “Analysis of stroke patient walking dynamics using a tri-axial accelerometer,” *Gait Posture*, vol. 30, no. 1, pp. 60–64, Jul. 2009.

[77] G. D. Fulk and E. Sazonov, “Using sensors to measure activity in people with stroke,” *Top. Stroke Rehabil.*, vol. 18, no. 6, pp. 746–757, Dec. 2011.

[78] B. G. Steele, B. Belza, K. Cain, C. Warms, J. Coppersmith, and J. Howard, “Bodies in motion: monitoring daily activity and exercise with motion sensors in people with chronic pulmonary disease,” *J. Rehabil. Res. Dev.*, vol. 40, no. 5 Suppl 2, pp. 45–58, Oct. 2003.

[79] N. Bidargaddi, A. Sarela, L. Klingbeil, and M. Karunanithi, “Detecting walking activity in cardiac rehabilitation by using accelerometer,” in *3rd International Conference on Intelligent Sensors, Sensor Networks and Information, 2007. ISSNIP 2007*, 2007, pp. 555–560.

[80] B. Dijkstra, Y. Kamsma, and W. Zijlstra, “Detection of gait and postures using a miniaturised triaxial accelerometer-based system: accuracy in community-dwelling older adults,” *Age Ageing*, vol. 39, no. 2, pp. 259–262, Mar. 2010.

[81] E. D. L. Nicole A. Capela, “A Smartphone Approach for the 2 and 6 Minute Walk Tests,” *36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBC’14 Chic. IL*, 2014.

[82] H. G. Simon Kozina, “Three-layer Activity Recognition Combining Domain Knowledge and Meta- classification Author list,” *J. Med. Biol. Eng.*, vol. 33, no. 4, pp. 406–414, 2013.

[83] S. K. Hristijan Gjoreski, “Using multiple contexts to distinguish standing from sitting with a single accelerometer,” 2014.

[84] L. A. S. Mark A. Hall, “Feature Selection for Machine Learning: Comparing a Correlation-Based Filter Approach to the Wrapper.,” pp. 235–239, 1999.

[85] A. J. P. Óscar D. Lara, “Centinela: A human activity recognition system based on acceleration and vital sign data,” *Pervasive Mob. Comput.*, vol. 8, no. 5, p. 717, 2012.

[86] D. Curone, G. M. Bertolotti, A. Cristiani, E. L. Secco, and G. Magenes, “A real-time and self-calibrating algorithm based on triaxial accelerometer signals for the detection of human posture and activity,” *IEEE Trans. Inf. Technol. Biomed. Publ. IEEE Eng. Med. Biol. Soc.*, vol. 14, no. 4, pp. 1098–1105, Jul. 2010.

[87] K. R. W. A G Bonomi, “Advances in physical activity monitoring and lifestyle interventions in obesity: a review.,” *Int. J. Obes. 2005*, vol. 36, no. 2, pp. 167–77, 2011.

[88] “Sensors - BlackBerry Native,” 21-Dec-2013. [Online]. Available: http://developer.blackberry.com/native/documentation/core/sensor_intro.html. [Accessed: 10-Dec-2014].

[89] “Rehab Measures - Functional Independence Measure,” *The Rehabilitation Measures Database*. [Online]. Available: http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=889. [Accessed: 12-Feb-2015].

[90] A. Smith, “Smartphone Ownership 2013,” *Pew Research Center’s Internet & American Life Project*. .

[91] C. K. Schindhelm, “Activity recognition and step detection with smartphones: Towards terminal based indoor positioning system,” in *2012 IEEE 23rd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, 2012, pp. 2454–2459.

[92] N. A. Capela, E. D. Lemaire, and N. Baddour, “Improving Classification of sit, Stand, and Lie in a Smartphone Human Activity Recognition System,” in *IEEE international symposium on Medical Measurements and Applications*, Turin, Italy.

[93] N. A. Capela, E. D. Lemaire, and N. Baddour, “Feature Selection for Wearable Smartphone-Based Human Activity Recognition with Able bodied, Elderly, and Stroke Patients,” *PLoS ONE*, vol. 10, no. 4, p. e0124414, Apr. 2015.

[94] J.-M. G. Oresti Banos, “Window Size Impact in Human Activity Recognition,” *Sensors*, vol. 14, no. 4, pp. 6474–99, 2014.

[95] K. C. Teh and A. R. Aziz, “Heart rate, oxygen uptake, and energy cost of ascending and descending the stairs,” *Med. Sci. Sports Exerc.*, vol. 34, no. 4, pp. 695–699, Apr. 2002.

[96] R. L. Waters and S. Mulroy, “The energy expenditure of normal and pathologic gait,” *Gait Posture*, vol. 9, no. 3, pp. 207–231, Jul. 1999.

Appendix A: Data collection activity circuit

A.1 The Ottawa Hospital Rehabilitation Centre, Ottawa, Ontario

Follow the participant and video their actions, on a second smartphone, while they perform the following actions, spoken by the investigator.

- From a standing position, shake the smartphone to indicate the start of the trial.
- Continue standing for at least 10 seconds. This standing phase can be used for phone orientation calibration.
- Walk to a nearby chair and sit down.
- Stand up and walk 60 meters to an elevator.
- Stand and wait for the elevator and then walk into the elevator.
- Take the elevator to the second floor.
- Turn and walk into the home environment
- Walk into the bathroom and simulate brushing teeth.
- Simulate combing hair.
- Simulate washing hands.
- Dry hands using a towel.
- Walk to the kitchen.
- Take dishes from a rack and place them on the counter.
- Fill a kettle with water from the kitchen sink.
- Place the kettle on the stove element.
- Simulate placing bread in a toaster.
- Walk to the dining room.
- Sit at a dining room table.
- Simulate eating a meal at the table.
- Stand and walk back to the kitchen sink.
- Rinse off the dishes and place them in a rack.
- Walk from the kitchen back to the elevator.
- Stand and wait for the elevator and then walk into the elevator.
- Take the elevator to the first floor.

- Walk 50 meters to a stairwell.
- Open the door and enter the stairwell.
- Walk up stairs (13 steps, around landing, 13 steps).
- Open the stairwell door into the hallway.
- Turn right and walk down the hall for 15 meters.
- Turn around and walk 15 meters back to the stairwell.
- Open the door and enter the stairwell.
- Walk down stairs (13 steps, around landing, 13 steps).
- Exit the stairwell and walk into a room.
- Lie on a bed.
- Get up and walk 10 meters to a ramp.
- Walk up the ramp, turn around, then down the ramp (20 meters).
- Continue walking into the hall and open the door to outside.
- Walk 100 meters on the paved pathway.
- Turn around and walk back to the room.
- Walk into the room and stand at the starting point.
- Continue standing, and then shake the smartphone to indicate the end of trial.

A.2 University Rehabilitation Institute, Ljubljana Slovenia

Follow the participant and video their actions, on a second smartphone, while they perform the following actions, spoken by the investigator.

- From a standing position, shake the smartphone to indicate the start of the trial.
- Walk down the hall to a chair in another room and sit down.
- Stand up and walk into the hall.
- Walk around the lobby and into the home environment
- Walk up to the sink and simulate brushing teeth.
- Simulate combing hair.
- Simulate washing hands.
- Dry hands using a towel.

- Walk to the kitchen.
- Take dishes from a table and place them on the counter.
- Fill a kettle with water from the kitchen sink.
- Place the kettle on the stove element.
- Simulate placing bread in a toaster.
- Walk to the dining room.
- Sit at a dining room table.
- Simulate eating a meal at the table.
- Stand and walk back to the kitchen sink.
- Rinse off the dishes and place them in a rack.
- Walk from the kitchen to the elevator.
- Stand and wait for the elevator and then walk into the elevator.
- Take the elevator to the third floor.
- Walk down the hall to a stairwell.
- Open the door and enter the stairwell.
- Walk up stairs (11 steps, around landing, 11 steps).
- Open the stairwell door into the hallway.
- Walk down the hall.
- Open the door and enter a room.
- Lie on a bed.
- Get up and walk to the door.
- Open the door and walk back down the hallway.
- Open the door and enter the stairwell.
- Walk down stairs (11 steps, around landing, 11 steps).
- Exit the stairwell and walk down the hallway back to the elevator.
- Stand and wait for the elevator and then walk into the elevator.
- Take the elevator to the first floor.
- Walk to the exit and walk outside.
- Walk through the parking lot to the underground parking entrance.
- Stand at the top of the ramp.
- Walk down the ramp, turn around, then walk up the ramp.

- Walk through the parking lot back to the entrance.
- Walk inside and back to the elevator.
- Stand and wait for the elevator and then walk into the elevator.
- Take the elevator to the second floor.
- Walk down the hall and stand at the starting point.
- Continue standing, and then shake the smartphone to indicate the end of trial.

Appendix B: Transition classes

Table B.1: Transition classes

1	stand-sit (1-2)
2	stand-lie (1-3)
3	Stand-walk (1-4)
4	Stand-small move (1-6)
5	Sit-stand (2-1)
6	Sit-lie (2-3)
7	Sit-walk (2-4)
8	Lie-stand (3-1)
9	Lie-sit (3-2)
10	Lie-walk (3-4)
11	Walk-stand (4-1)
12	Walk-sit (4-2)
13	Walk-lie (4-3)
14	Walk-stairs (4-5)
15	Walk-small move (4-6)
16	Stairs-walk (5-1)
17	Stairs-walk (5-4)
18	Small move-stand (6-1)
19	Small move-sit (6-2)
20	Small move-lie (6-3)
21	Small move-walk (6-4)

Note that only transitions that follow the logic of the classifier structure shown in figure 7.1 were considered. Therefore, certain transitions were not included. For example, a mobile state was classified as walk (i.e. large movements) before being classified as stairs, so a transition from lie-to-stairs was not possible. Small movements may only be transitioned into from a standing state.