RECOGNITION OF EVERYDAY ACTIVITIES THROUGH WEARABLE SENSORS AND MACHINE LEARNING

A Thesis

by

JOSH A. CHERIAN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Tracy Hammond
Co-Chair of Committee,	Theresa Maldonado
Committee Members,	Daniel Goldberg
	Costas N. Georghiades
Head of Department,	Miroslav M. Begovic

May 2017

Major Subject: Electrical Engineering

ABSTRACT

Over the past several years, the use of wearable devices has increased dramatically, primarily for fitness monitoring, largely due to their greater sensor reliability, increased functionality, smaller size, increased ease of use, and greater affordability.These devices have helped many people of all ages live healthier lives and achieve their personal fitness goals, as they are able to see quantifiable and graphical results of their efforts every step of the way (i.e. in real-time). Yet, while these device systems work well within the fitness domain, they have yet to achieve a convincing level of functionality in the larger domain of healthcare.

As an example, according to the Alzheimer's Association, there are currently approximately 5.5 million Americans with Alzheimer's Disease and approximately 5.3 million of them are over the age of 65, comprising 10% of this age group in the U.S. The economic toll of this disease is estimated to be around $259 billion. By 2050 the number of Americans with Alzheimer's disease is predicted to reach around 16 million with an economic toll of over $1 trillion. There are other prevalent and chronic health conditions that are critically important to monitor, such as diabetes, complications from obesity, congestive heart failure, and chronic obstructive pulmonary disease (COPD) among others.

The goal of this research is to explore and develop accurate and quantifiable sensing and machine learning techniques for eventual real-time health monitoring by wearable device systems. To that end, a two-tier recognition system is presented that is designed to identify health activities in a naturalistic setting based on accelerometer data of common activities. In Tier I a traditional activity recognition approach is employed to classify short windows of data, while in Tier II these classified windows are grouped to identify instances of a specific activity. Everyday activities that were explored in this research include

brushing one's teeth, combing one's hair, scratching one's chin, washing one's hands, taking medication, and drinking. Results show that an F-measure of 0.83 is achievable when identifying these activities from each other and an F-measure of of 0.82 is achievable when identifying instances of brushing teeth over the course of a day.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to a number of people without whom I would never have been able to put together this thesis. First I would like to thank my adviser, Dr. Tracy Hammond for her constant and unwavering enthusiasm and support. This thesis is a testament to how much I have grown and learned under her guidance.

I would like to thank my co-chair, Dr. Theresa Maldonado for her positivity and encouragement throughout this whole process. From the beginning, she was instrumental in helping me navigate the process of doing my Master's work in two different departments and her insights and perspectives throughout have been invaluable. I would also like to thank Dr. Daniel Goldberg and Aaron Harmon for their perspectives on my work. In addition they were instrumental in helping me throughout the data collection process, and for that I am immensely grateful. I would like to thank Dr. Costas Georghiades for his encouragement and support throughout.

Of course, none of this work would not have been possible without the support, feedback, and encouragement of the Sketch Recognition Lab. In particular I would like to thank Vijay Rajanna for his constant support, guidance, and willingness to help in any way and Paul Taele for his guidance and insight throughout this process.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a thesis committee consisting of Dr. Tracy Hammond (advisor) of the Department of Computer Science & Engineering, Dr. Theresa Maldonado (co-advisor) and Dr. Costas N. Georghiades of the Department of Electrical and Computer Engineering and Dr. Daniel Goldberg of the Department of Geography.

All work conducted for the thesis was completed by the student independently.

Funding Sources

This work is funded in part by TEES AggieChallenge and Future Position X. Furthermore the data collected in Stage III of data collected was made possible by the grant from NSF (#SMA-1560106)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Recent advances in functionality, design, and affordability have led to the widespread adoption of wearable devices such as fitness trackers and smartwatches. As a result, people are now able to play a much larger role in managing their day-to-day fitness, as many of these devices can accurately track basic metrics such as steps walked, calories burned, and even several simple exercises. To allow for this functionality, these devices often utilize sensors such as accelerometers, gyroscopes, GPS, and magnetometers. Despite these advances in functionality and ubiquity, current wearable devices are predominantly used by those who are healthy and affluent, as devices are largely limited to fitness-related applications.

One area where wearable devices could make a significant impact is healthcare [1–4]. The global population is expected to age dramatically over the coming decades [5] (shown in Fig. 1.1) which is expected to coincide with a sharp increase in the number of cases of chronic age-related diseases such as Alzheimer's, diabetes, arthritis, congestive heart failure, and chronic obstructive pulmonary disease (COPD) [6, 7], posing significant societal, financial, and systemic problems to an already strained system [8, 9].

There are several areas within healthcare where wearable devices could make an immediate and dramatic impact. One study found that of the nearly 70% of U.S. adults who track at least one health metric, half of this population track this metric "in their heads" while another third keep track of this metric in some type of notebook or journal [10]. Both of these methods are unreliable as the former relies on the person's memory while the latter relies both on the user's diligent and accurate recording and their recollection of where those records are kept. The alternative to patients keeping track of their own health metrics requires trained observers to do so for them; however this is expensive, obtrusive,

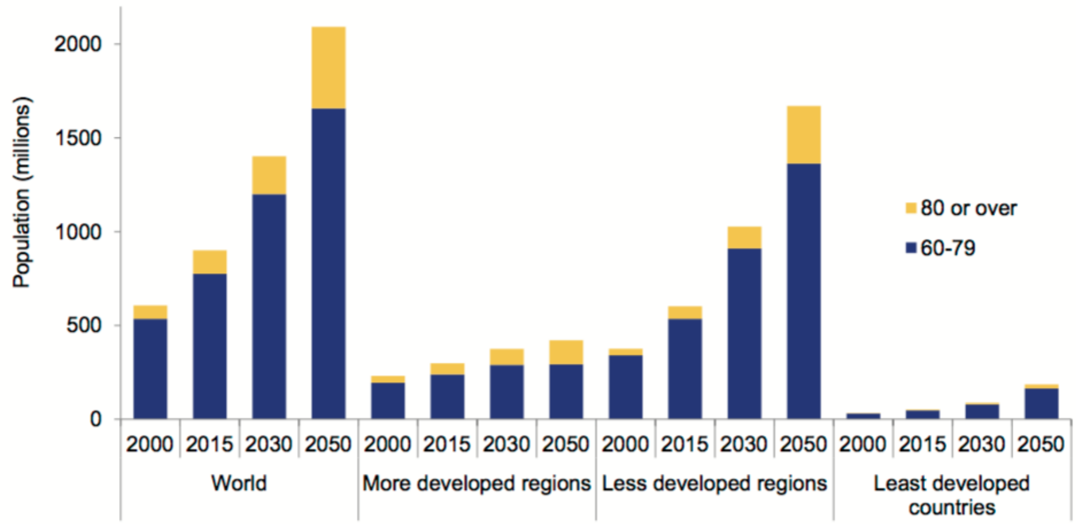

Figure 1.1: United Nations world population projections. Reprinted from [5]

and tiring for the observers. By leveraging the assortment of sensors within wearable devices today, it is entirely possible to not only accurately, securely, and unobtrusively track these metrics but also to develop more sophisticated and personalized applications and interventions based on these metrics [11–15].

Sophisticated and accurate monitoring of chronic health conditions is especially important. By 2030 it is estimated that one half of American households will include someone who suffers from one or more chronic conditions. Furthermore, about one-fifth of patients who suffer from chronic conditions experience some degree of inability to perform normal everyday activities [16], commonly referred to as Activities of Daily Living (ADLs). Such activities can be divided into two categories: Basic and Instrumental. Basic ADLs, characterized by care and movement of the body, consist of bathing/showering, toilet hygiene, dressing, eating & feeding, functional mobility, personal device care, personal hygiene and grooming, and sexual activity. Instrumental ADLs, characterized by more complex daily interactions, consist of activities such as health management, home man-

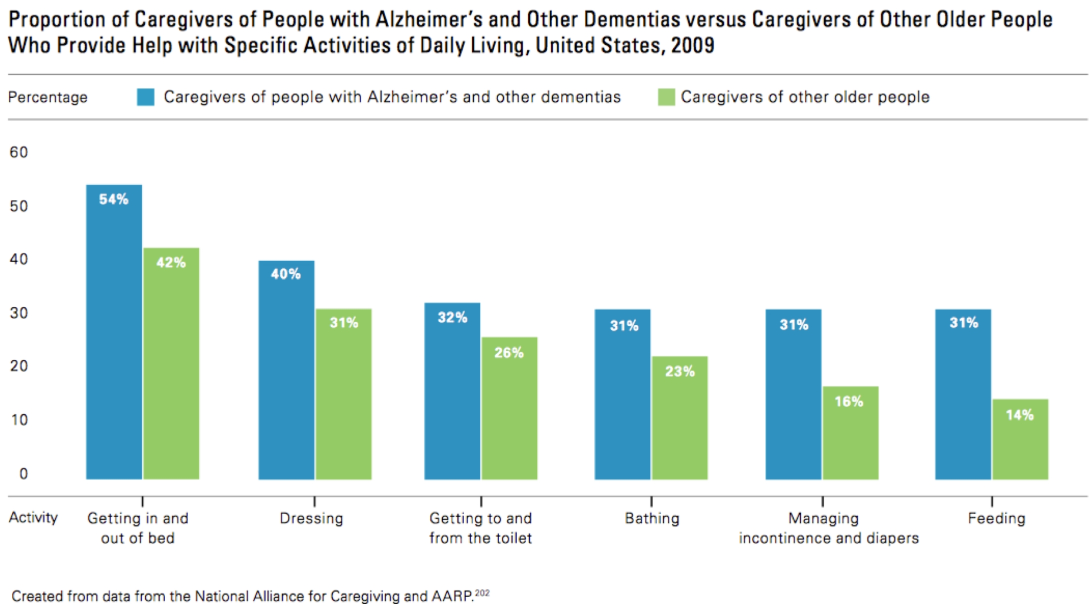

Figure 1.2: A comparison of the amount of care caregivers give to people with Alzheimer's and the elderly without Alzheimer's. Reprinted from [22]

agement, driving/community mobility, child rearing, meal preparation and cleanup, and shopping [17]. Often patients can live relatively independently if they are unable to perform one or two instrumental ADLs; however, if they are unable to perform three or more, they typically require some form of nursing care [18]. This is especially problematic for people suffering from cognitive decline who lose the ability to perform these activities over time. Common activities that such patients need additional help with are shown in Fig. 1.2, thus, not only is monitoring an ability to perform such an actions vital, but also equally critical is being able to determine when they are unable to perform such an action. This type of monitoring has been shown to be possible through the use of wearable sensors [19]. By using devices and algorithms that can determine when a person is performing these activities, physicians and/or family members can be alerted if activities are performed incorrectly or not at all, a reassurance which can allow individuals with chronic conditions to live more independently [20, 21].

In this study we target six everyday activities: brushing one's teeth, combing one's hair, scratching one's chin, washing one's hands, taking medication, and drinking. We focus on identifying brushing teeth in a naturalistic setting as it can lead to severe health consequences if not performed adequately or at all. Failing to brush one's teeth regularly and maintaining poor oral hygiene, in general, can lead to a variety of systemic diseases including atherosclerosis, COPD, diabetes, and bacteremia as well as conditions such as brain abscess [23–25]. Furthermore, periodontal disease, one of the most common results of poor oral hygiene is estimated to affect 10-15% of the world's population [26], and individuals with this disease are 3.2 times more likely to die from Type 2 diabetes [27], 4.3 times more likely to have a cerebral ischemic stroke [28], susceptible to a 25% increased risk of heart disease [29], and disposed to an increased risk of developing Alzheimer's and pancreatic cancer [30, 31]. Periodontal disease and poor oral hygiene have also been shown to have a stronger association with total mortality than with coronary heart disease [29].

Given these risks it is especially important to ensure that patients with dementia are maintaining proper oral hygiene. In the early stages of dementia, patients may need to be reminded that they need to brush their teeth and/or be supervised while doing so. In later stages, patients may be unable to brush their teeth, stop understanding that they need to clean their teeth, or lose interest in doing so [32].

To this end, the goal of this research is to explore the following two questions:

1. Can we accurately distinguish between six everyday activities: brushing one's teeth, combing one's hair, scratching one's chin, washing one's hands, taking medication, and drinking using a commercially available smartwatch and smartphone?

2. Is our two-tier recognition system able to identify brushing one's teeth in a naturalistic setting?

To address these questions, we employ activity recognition, which is the use of machine learning to identifying human activities. In Chapter 2 we describe previous work in this field. In Chapter 3 we describe the recognition system we developed to identify activities in a naturalistic setting. In Chapter 4 we describe the data collection process, from the system we used to collect data to the three user studies we conducted. In Chapters 5 and 6 we discuss our efforts to apply our recognition system to our dataset and the implications of these results. In Chapter 7 we describe the next steps in this research and then offer conclusions in Chapter 8.

2. RELATED WORK

Research into activity recognition has seen considerable interest since the late 1990s given its potential to facilitate more personalized interventions and interactions in fields ranging from health care and assisted living to sports, learning environments, and security [33–36]. Early on, before the widespread adoption of mobile phones, much of the research centered around sensor placement; however since smartphones and wearable devices have become popular, research has shifted to take advantage of their ubiquity and the easy access to sensor data they provide. In the following sections, we focus our discussion on works that recognize health and/or exercise activities and feature systems worn on the body as opposed to in the surrounding environment.

2.1 Sensor Type and Placement

Initially, much of the research in the field was focused on identifying both the optimal collection of sensors as well as their optimal location on the body. Researchers demonstrated varying levels of accuracy with sensors such as accelerometers, gyroscopes, and microphones placed on users' wrists [37–46], hands [41, 47–50], upper arms [38, 42, 46, 51], hands [52, 53], waist [38–40, 42, 44–46, 54–58], upper leg [37, 38, 42, 45, 59], lower leg [37, 38, 42, 43], shoulder [39, 41, 60], head [41], foot [57], eyes [61–63], and chest [37, 40, 41, 43, 44, 64–66]. These studies almost uniformly found that the best locations for sensors depended on the activity being measured. For example, Bao & Intille [38] placed bi-axial accelerometers on the upper arm, wrist, hip, thigh, and ankle and found that accelerometers placed on the thigh, hip, and ankle were the best indicators for activities that had some form of ambulation or posture, while accelerometers placed on the wrist and arm were better indicators for activities that involved mostly the upper body. Berchtold et al. [67] took a unique approach in developing ActiServ, which was designed to detect activities no matter where the sensors were placed or how they were oriented; however it required several days of calibration to achieve a high accuracy.

By and large, in terms of types of sensors, the accelerometer is the most commonly used sensor for activity recognition. In most cases it is used individually; however, in several cases it is used in combination with other sensors [68]. Notably, Pärkkä et al. [41] conducted a comprehensive study using 19 different types of sensors to identify several common scenarios such as visiting the library, a restaurant, or a shop as well as common physical activities such as walking, running, and cycling. They found that accelerometers proved to be the most accurate indicator of what activity was being performed.

2.2 Smartphone Mediated Recognition

Over the past several years, as smartphones have achieved widespread popularity, activity recognition research has shifted to utilizing smartphone sensors, commonly consisting of accelerometers, gyroscopes, GPS, magnetometers, proximity sensors, and light sensors, for activity recognition. In addition to their available sensors and ubiquity, smartphones are also advantageous for this purpose due to their unobtrusiveness, their high computational power and storage capacity, and the ease with which they can be programmed [69–71]. Brezmes et al. [72] had users carry a smartphone in their hand without regard for orientation, and they were able to identify walking, climbing-up stairs, climbing-down stairs, standing-up, sitting-down, and falling using the phone's triaxial accelerometer. Kwapisz et al. [73] were able to identify similar activities (walking, jogging, ascending stairs, descending stairs, sitting, and standing) with an overall accuracy of 91.7%; however, they had users keep a smartphone in their front pant pocket. Anguita et al. [74] looked at similar activities (walking, standing, ascending stairs, descending stairs, sitting, and laying), and they were able to distinguish between them with an overall accuracy of 89% with the smartphone attached to the user's waist.

Several studies have taken advantage of this combination of sensors present in smartphones to inform their recognition. Reddy et al. [75] used both an accelerometer and GPS to differentiate between modes of transportation (staying still, walking, running, biking, motor transport) using six phones attached to the arm, waist, chest, hands, pocket, and in a bag. World of Workout, a mobile role-playing game (RPG), allowed the user to improve their game character through a variety of exercises such as jumping jacks, crunches, and cycling, which were automatically recognized using the phone's accelerometer and GPS data [76].

2.3 Wearable Device Mediated Recognition

In more recent years several studies have utilized wrist-mounted devices and commercially available wearable devices to improve the accuracy and feasibility of recognition systems. Saponas et al. [77] used the Nike+iPod Sport kit to detect when the user's were using their feet, and by combining this output with features extracted from data gathered from the accelerometer of an iPhone placed in the user's pocke, they were able to recognize running, walking, bicycling, and sitting with an accuracy of 97%. Thomaz et al. [78] recognized eating movements with an overall F-score of 76% using the 3-axis accelerometer in a Pebble smartwatch for both controlled and in-the-wild studies.

2.4 Activities Being Recognized

The majority of the work in this field has looked at various forms of ambulation (walking, running, standing, sitting, lying down, etc.); however, many studies do include other activities in addition to these. In particular, numerous studies have looked at brushing one's teeth; however, it has commonly been investigated simultaneously with other ambulation and posture activities [12, 38, 39, 78–82]. In most of these cases, this activity was not compared to other similar motions: brushing teeth is a highly repetitive wrist motion, while most forms of ambulation are full body movements with much longer delays between repetitions. Amft et al. [83] looked at eating and drinking through a combination of inertial sensors to detect the gesture, a microphone to detect chewing, and electromyography electrodes to detect swallowing, and achieved an average detection rate of about 80%. Dernbach et al. [84] made a distinction between simple activities and complex activities, and they sought to recognize both categories using an Android smartphone. They were able to reliably differentiate between simple activities such as driving, lying, running, sitting, standing, and walking using an Android smartphone, but they were unable to do the same with complex activities, which included cleaning, cooking, retrieving and sorting medication, sweeping, washing hands, and watering plants. Varkey et al. [85] made a distinction between static and dynamic activities, and they were able to differentiate between walking, standing, writing, smoking, jumping jacks, and jogging using an accelerometer worn on the wrist and a gyroscope worn on the ankle. Hong et al. [86] used a combination of accelerometers and radio-frequency identification (RFID) sensors to differentiate between many different forms of ambulation and posture as well as activities such as brushing hair, phone calling, reading, brushing teeth, and taking pictures.

Our system builds on this body of work by seeking to identify ADL activities that all solely consist of some form of wrist and upper arm movement. By including activities that, in several cases, have only slight differences in orientation and vigor, we aim to show that we can reliably and relatively unobtrusively differentiate between many of the more physically-nuanced activities that one performs on a daily basis using solely a commercially available accelerometer and smartphone. Furthermore, we aim to show that recognition of these activities is not limited to a controlled setting but can be realistically used in real-time systems.

3. TWO-TIER RECOGNITION

To identify activities in a real time naturalistic setting, we propose a two tier recognition system. Tier I consists of traditional activity recognition, i.e. data are segmented into static sliding windows which are classified as a particular activity. Tier II features a dynamic window constructed from the smaller windows of Tier I to identify specific instances of the specific activity actually occurring. In this chapter we describe this two-tier system in detail; in Chapters 5 & 6 we describe the implementation of this system to identify brushing teeth in a naturalistic setting.

3.1 Tier I Recognition: Static Sliding Windows

For Tier I recognition, we take the traditional approach to activity recognition [87]. Raw accelerometer data is first segmented into static sliding windows. From each of these windows we then extract a set of features, which are then collectively inputted into various classification algorithms to test their efficacy in differentiating between the activities being observed.

For this particular study, collected data were segmented into four second sliding windows with a a one second overlap. We discuss the implications of window size in Sec. 3.3, as the characteristics of Tier II recognition influence this parameter. Samples of a single window of recorded data for each activity can be seen in Fig. 3.1. For each window we extracted fifty-one features. Thirty of these features were features that have previously been shown to be effective in activity recognition (described in Sec. 5.1) [88], while the remaining twenty-one features were novel features discerned from visualizing the data for each activity and quantifying their differences (described in Sec. 6.1). Every feature we extracted was either calculated for each axis or from a combination of axes.

3.2 Tier II Recognition: Dynamic Activity Windows

While the recognition of activities segmented into windows is important, these results are not directly applicable to a real-life and real-time situation, which is the ultimate goal of this work. To that end, we developed an additional layer of recognition to integrate the shorter sliding classified windows from Tier I into a longer dynamic window representing an actual instance of the user performing the activity. To do this step we established three thresholds a *minimum activity duration*, a *maximum inactivity duration*, and an *activity percentage*. These thresholds are defined below. The values of these thresholds are dependent on the activity being classified. Examples of the functioning of this algorithm are shown in Fig. 3.2.

3.2.1 Minimum Activity Duration Threshold

The minimum activity duration threshold is the minimum amount of time that the user needed to have spent doing the activity for it to be considered an instance of that activity. In the case of brushing teeth, we initially considered setting this threshold to 60 seconds given that the recommended duration for brushing teeth is 2-3 minutes [89]; however we found that this produced several false negatives, as people on average spend less than this recommended time brushing their teeth [90]. Thus, for brushing teeth we set this threshold to be 45 seconds.

3.2.2 Maximum Inactivity Duration Threshold

The maximum inactivity duration threshold accounts for instances when clusters of windows classified as the particular activity being identified are classified as another activity or as Inactive. If the number of sequential windows classified as something other than the desired activity does not exceed this threshold, these windows are added to the total number of windows, thus counting towards the overall activity window. In other words, the maximum inactivity duration is the maximum amount of time that the algorithm can predict that a user is not doing the activity being identified. In the case of brushing teeth this threshold was set to 15 seconds; setting this metric to be much less than this value led to sequences that should be considered to be one instance of brushing teeth being divided into separate instances.

3.2.3 Activity Percentage Threshold

With just the first two thresholds a situation could arise where most of the windows in a sequence are Inactive, as so long as the number of sequential windows of inactivity is less than the maximum inactivity duration threshold, the inactivity sections will be counted towards an instance of the activity. To ensure that most of the sequence does consist of windows classified as the activity we established a third threshold: the activity percentage threshold. This threshold is the minimum percentage of windows within the sequence that need to be classified as the activity. We found that setting this threshold to 75% worked well for brushing teeth, as any value significantly lower introduced a number of false positives.

3.3 Window Size

As previous studies have noted, the size of the window can have a significant impact on the performance of the classification methodology [91]. Purely in terms of recognition accuracy the optimal window size varies by activity; however, other metrics such as computational load and recognition time, both larger with smaller window sizes, need to be considered depending on the application. In our study, the dynamic window recognition system we employ as part of Tier II recognition has further implications on the window and overlap size. The larger the window and overlap size are, the smaller the number of windows required to meet the minimum activity duration becomes. This view of Tier II recognition in turn increases the likelihood of false positives occurring. However, since we envision this system being implemented into a real time system, we cannot set these values to be too small in the interest of minimizing computational load and real-time delay. Both of these considerations were taken into account when finally settling on a window size of four seconds with a one second overlap.

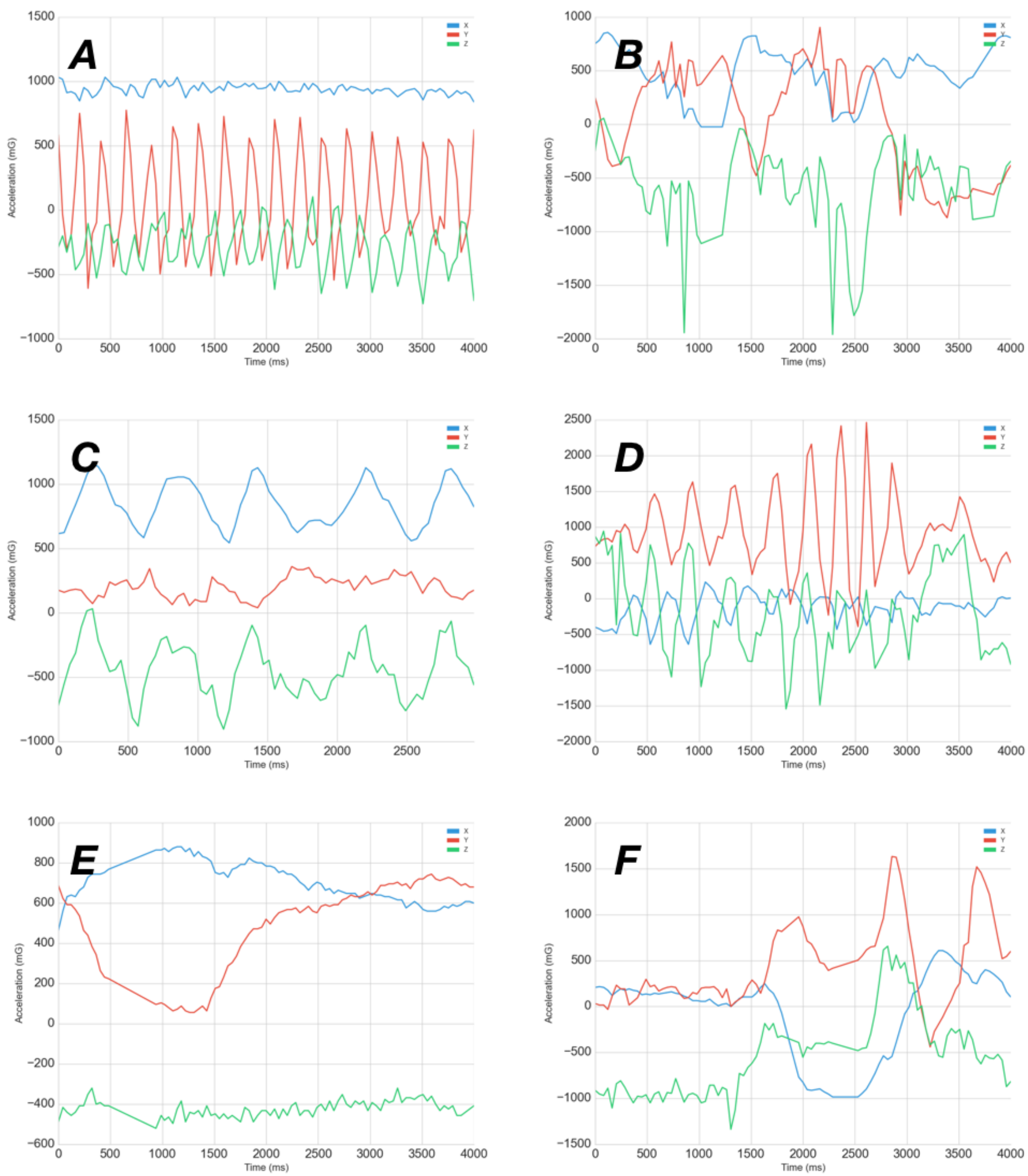

Figure 3.1: Sample graph of each activity where A = Brushing Teeth, B = Combing Hair, C = Scratching Chin, D = Washing Hands, E = Drinking, & F =Taking Medication

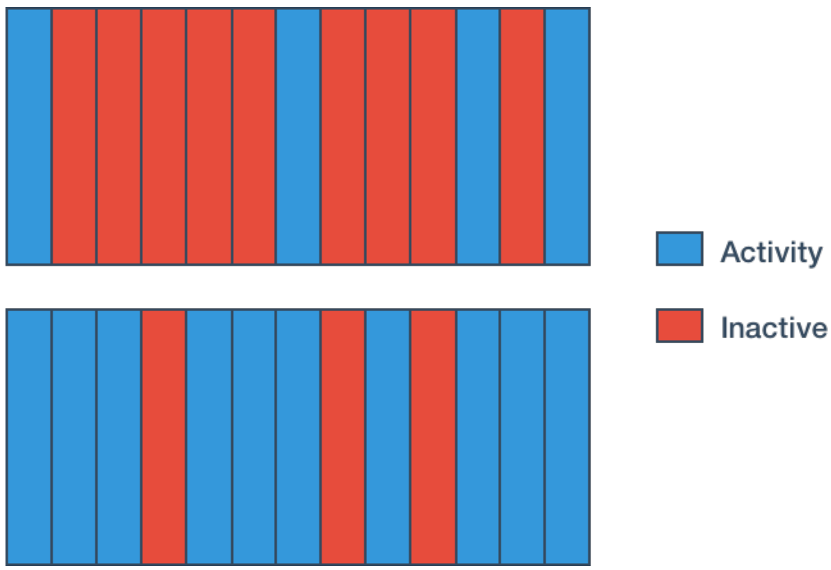

Figure 3.2: To illustrate how Tier II recognition works in practice, let us assume that for this particular activity we'd like to identify (shown in blue) the minimum activity duration is 13 windows, the maximum inactivity duration is 6 windows, and the activity percentage is 60%. In both cases the maximum number of sequential windows classified as Inactive (shown in red) is less than 5, so every single one of these will be treated as having been misclassified as Inactive and will count towards the number of windows signifying the activity. In both cases the minimum activity duration threshold is now met since every window is now treated as being classified as the activity; however, only the bottom sequence passes the activity percentage threshold, as 10 out of the 13 windows were initially classified as the activity (77%). In the top sequence, the activity percentage is only 31%, which does not meet the threshold of 60%. Thus only the bottom sequence is considered an instance of the activity.

4. DATA COLLECTION

To test our proposed two-tier recognition system we conducted a series of user studies. In this chapter we describe the activities studied, the system used to facilitate data collection, and the user studies we ran.

4.1 Activities

To evaluate our recognition system's ability to identify activities in a natural setting, we tested our system on the data collected for six activities: brushing one's teeth, combing one's hair, scratching one's chin, washing one's hands, taking medication, and drinking. These activities were chosen for at least one of the two following reasons:

1. The activity plays a role in the average person's everyday health, and
2. The activity requires physical movements similar to those required by another activity being studied.

In the case of the former, activities such as brushing one's teeth, washing one's hands, and for some people taking medication, are, or at least should be, preformed regularly, as they impact our day-to-day health. The first two of these are Basic ADLs, as they are important forms of personal hygiene/grooming, while taking medication is an instrumental ADL indicative of the person's ability to manage their personal health. For the purposes of data collection, taking medication was separated into two separate activities, taking a pill (simulated with an M&M and a push and turn pill bottle) and drinking a glass of water.

In the case of the latter, four of the activities (brushing one's teeth, combing one's hair, scratching one's chin, washing one's hands) consist of back and forth movements of the wrist. In this sense it is important to note that the definition of several of these activities was limited to their most literal meaning. For example, the colloquial definition of brushing teeth often includes the actions of putting toothpaste on the toothbrush, rinsing out one's mouth, and cleaning one's toothbrush, in addition to the actual physical action of brushing one's teeth. Similarly washing one's hands can consist of activities such as turning on the faucet, applying soap to one's hands, and drying one's hands off on a towel or under/in a hand dryer in addition to actually washing one's hands under running water. In these cases, we aimed to recognize the actual activity, as not only did those sections solely display the back and forth movements we sought to classify, but they also represent the clearest indication that the activity has actually been performed.

4.2 System Implementation

To collect data, a system consisting of a Pebble smartwatch [92] worn on the user's dominant hand and an Android application [93] was developed. The Pebble contains a 4G 3-axis accelerometer (shown in Fig. 4.1); these data, combined with corresponding timestamps, was transmitted via Bluetooth using the Pebble SDK to the Android application. Data was recorded at a sampling rate of 25 Hz, and any collected data that did not fall within $\pm 10\%$ of the sampling rate (22.5 Hz - 27.5 Hz) were not used in data analysis. For Stages I and II of the user study, data were stored on the phone, while in Stage III data was stored on a server. In addition to facilitating the storage of this data, the Android application also allowed for precise labeling of the data (via a drop-down menu) and mediated the starting and stopping of data collection. Any data that was not labeled as one of the six specified activities were labeled as "Inactive".

4.3 Controlled vs. Naturalistic Data Collection

Historically, research in the area of activity recognition has relied predominantly on training systems based on activities completed in a laboratory setting. This characteristic severely limits the efficacy of these systems for several reasons. Several studies have shown that people tend to perform activities differently in a laboratory setting than they

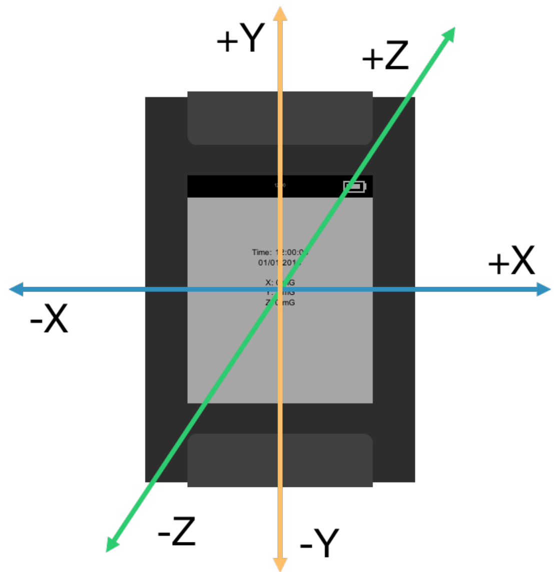

Figure 4.1: Pebble accelerometer axes

would naturally as part of their normal routine [37, 38, 70, 94]. Furthermore, training models purely on data obtained from a controlled setting removes much of the context surrounding the activities being recognized in a natural setting. Therefore these models would not account for the other activities actually present nor the motions a user would make in between performing the measured activities in real life. Thus, although many of these systems have achieved a high accuracy, they remain flawed candidates for deployment into real-world settings.

4.4 User Studies

To address this pitfall, we conducted a user study with three distinct stages. Each subsequent stage was characterized by a shift towards a more naturalistic setting.

4.4.1 Stage I: Controlled Study

To establish a baseline understanding of what these activities looked like as quantified by an accelerometer, we first collected data in a completely controlled environment, where participants were asked to complete each of the six activities consecutively but in no particular order. For this stage we had 20 participants.

4.4.2 Stage II: Hour-Long Semi-Naturalistic Study

For Stage II, data were collected for an hour, during which participants went about their day as they normally would except that they performed each of the activities at some point during the hour. Six participants participated in this study, with data being collected for four of them for a second day. It is worth highlighting that this stage introduced "Inactive" data, or data that are representative of none of the six activities being measured. Additionally participants were instructed in how to label the data themselves (described in more detail in Sec. 4.2), and they were not supervised during data collection.

4.4.3 Stage III: Full-Day Naturalistic Study

For the Stage III of our user study, data collection was conducted over a 31-day period, during which the users were only asked to perform each of the activities at least once during each day of the study. As in Stage II, participants were taught how to label the data themselves, and they were completely unsupervised during data collection. Over this time period, data were collected on average for 6.25 days per user, with data collection being run for 4.9 hours on average per day. While these averages still represent a significant amount of data, the difference between these numbers and the actual duration of the study can largely be explained by the fact that the onus of data collection was placed on the participants, which leads to varying compliance [95]. This stage had 12 participants, none of whom took part in the first two stages of this study.

5. RECOGNITION BASED ON TRADITIONAL FEATURES

Having collected data for each of the activities being studied, we implemented our two-tier recognition system based on features commonly used in activity recognition. In this section we describe these features, our implementation of the proposed two-tier recognition system, and the results of distinguishing between the six activities being studied and identifying brushing teeth in a naturalistic setting.

5.1 Recognition Methodology

As stated in Sec. 3.1, collected accelerometer data were segmented into four-second sliding windows with a one second overlap. From these windows, we initially extracted thirty features. These features were taken from prior work, in which they have been shown to be effective in activity recognition [96]. Subset selection was then performed on these features to identify the optimal set of features for recognizing the activities being observed. This subset of features was then inputted into several classification algorithms with 10-fold cross-validation. To further identify brushing teeth in a naturalistic setting, the best model produced by these algorithms was evaluated on a test data set. We describe each of these steps in detail below.

5.1.1 Features

The thirty extracted features can be divided into time-domain features and frequency-domain features. Time-domain features are advantageous because of their lack of computational complexity, and as a result, these features can provide a basic yet practical quantitative sense of the data. Frequency-domain features, on the other hand, are advantageous as they can highlight the repetitiveness of signals.

5.1.1.1 Time-Domain Features

For our purposes we calculated the average (X, Y, Z), standard deviation (X, Y, Z) (Equation 5.1), average jerk (X, Y, Z), average distance between axes (XY, XZ, YZ) (Equation 5.2), correlation (XY, XZ, YZ) (Equation 5.3), number of peaks (X, Y, Z), number of valleys (X, Y, Z), and root mean square (RMS) (X, Y, Z) (Equation 5.4). In these equations n represents the total number of data points in the window, while x_i and y_i represent the ith data point in the window.

$$\sigma_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i - \overline{x}} \tag{5.1}$$

$$avg_{xy} = \frac{\sum_{i=1}^{n} |x_i - y_i|}{n} \tag{5.2}$$

$$Corr(x_i, y_i) = \frac{Cov(x_i, y_i)}{\sigma_x \sigma_y} \tag{5.3}$$

$$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} \tag{5.4}$$

5.1.1.2 Frequency-Domain Features

The frequency-domain features included energy (X, Y, Z) and entropy (X, Y, Z). Energy has been shown to be effective in differentiating between activities by intensity, while entropy has been shown to be capable of further distinguishing between activities that have similar energy levels. The formulas for energy and entropy can be seen in Equations 5.5 and 5.6 respectively, where n represents the total number of data points in the window and a and b are the real and imaginary components respectively of each data point in the window after they have been converted to the frequency domain using a discrete Fast Fourier Transform (FFT).

$$energy = \sum_{j=1}^{n} \frac{a_j^2 + b_j^2}{n} \tag{5.5}$$

$$entropy = \sum_{j=1}^{n} c_j * \log(c_j), c_j = \frac{\sqrt{a_j^2 + b_j^2}}{\sum_{k=1}^{n} \sqrt{a_k^2 + b_k^2}} \tag{5.6}$$

These two features have commonly been successfully used together in activity recognition [88, 97].

5.1.2 Subset Selection

To determine the optimal subset of these features, we performed subset selection on the thirty extracted features. This step is critical, as it can make the difference between an algorithm being able to successfully build a model of the activity being studied or not. Often times, several of the features initially selected can be redundant or misleading. While the redundant features only lead to additional computation time and an increased load on the system, misleading features can sometimes dramatically reduce the efficacy of the algorithm. Even if the metrics of the model are high, irrelevant features can lead to overfitting, rendering the produced model sub-optimal or even useless when applied to a test data set.

For our purposes we used the CFsSubset Evaluator with the BestFirst search method which is available in the Weka Data Mining Toolkit [98]. The CFsSubset Evaluator rewards features within a subset that correlate highly with recognition accuracy but correlate poorly with other features in the subset. More specifically, it evaluates the "merit" of adding a new feature using Equation 5.7, where $\overline{r_{zc}}$ is the "merit" of the subset of features, $\overline{r_{zi}}$ is the average of the correlations between the features and the classification results, k is the number of features, and r_{ii} is the average of the correlations between the features.

This formula is identical to Pearson's correlation coefficient with the exception that the variables are standardized [99].

$$r_{zc} = \frac{k\overline{r_{zi}}}{\sqrt{k + k(k-1)\overline{r_{ii}}}} \tag{5.7}$$

The BestFirst method of searching explores the subset space using a greedy hill climbing approach that can backtrack if a certain subset is producing sub-optimal results [99].

5.1.3 Recognition Algorithms

In this work we tested our features on five distinct classifiers: C4.5 (The Java implementation known as J48 was used here.), Random Tree, K-Nearest Neighbor (kNN), Multilayer Perceptron, and Random Forest. As our data were labeled, all the chosen classifiers were supervised learning techniques. With each of these algorithms, 10-fold cross-validation was done to ensure that the algorithm did not overfit to the training data, and that it could be counted on to produce similar results when applied to a test data set.

5.1.3.1 Decision Trees - C4.5 & Random Tree

Decision Trees are one of the most commonly used machine learning algorithms for several reasons. For one thing, they are amongst the simplest algorithms to understand, often elucidating the underlying relationships in the data. Although performing subset analysis separately usually yields better results, decision trees also implicitly determine the optimal subset features when building their model.

In this work we tested two Decision Tree algorithms, Random Tree and C4.5. Random Tree is a relatively simple decision tree algorithm that considers a random subset of attributes at each node in the tree and does not do any pruning. C4.5 is one of the most widely used decision trees, as it can handle missing values and both categorical and numeric values. It addresses overfitting through the use of pruning; however, it is still susceptible to overfitting when supplied with noisy data.

5.1.3.2 Random Forest

Random Forest is an ensemble method that constructs a model based on the average of many randomized decision trees. It has shown to be widely applicable to a diverse range of problems, consistently achieving high degrees of accuracy [100]. Moreover it is relatively fast and easy to implement, and it can avoid overfitting when handling a large number of input features.

5.1.3.3 kNN

kNN is a lazy learner that classifies an instance based on the majority vote of its k-nearest neighbors. As its mechanics suggest, kNN works well on problems that focus on finding the similarity between observations; however, it struggles with large datasets and high dimensionality. Another challenge associated with kNNs is picking the optimal value of k: choosing a small k value allows noise to more severely impact the results, while choosing a large k value is very computationally expensive. In this work the initial value of k was set to be the square root of the number of features, with nearby values being tested to see if the resulting model improved.

5.1.3.4 Multilayer Perceptron

Multilayer Perceptron is a feed-forward artificial neural network that utilizes back-propagation to train its model.This algorithm works well at finding non-linear patterns in data that are not linearly-separable, but that can be extremely slow. It has been used in other studies to recognize hand gestures, notably for recognizing sign language [101].

5.2 Results

In evaluating the efficacy of these algorithms in both distinguishing between the six activities and identifying when a user brushed their teeth over the course of a day, we used accuracy and F-measure as indicators of the algorithm's performance. While accuracy by itself can provide an initial indicator of performance, it can be misleading when the dataset being analyzed is imbalanced in favor of a particular class. In our case, this scenario held true when evaluating the data from Stage II, and especially from Stage III of our user study which featured an enormous amount of data not representative of any of the activities being analyzed. In this situation one could simply use an algorithm such as ZeroR, which simply classifies every instance as an instance of the majority class and achieves an accuracy of nearly 100%. This algorithm, however, would obviously fail completely at identifying the minority classes, which is one of the primary goals of this work.

For this reason we report F-measure in addition to accuracy, as it is a good indicator of an algorithm's performance when given an imbalanced dataset. Looking at the formula (shown in Equation 5.8) we can see that the F-measure takes true negatives into account and relies solely on the true positives that occur.

$$F\text{-}measure = \frac{2TP}{2TP + FP + FN} \tag{5.8}$$

5.2.1 Distinguishing between Activities

To first test the effectiveness of our features, we tested their ability to discriminate between the six activities. As described in Sec. 4.1, distinguishing between these activities is in and of itself a challenging problem, given that several of these activities are quite similar gesturally. Performing subset selection on the thirty extracted features yielded a subset of seventeen features. This subset is shown in Table 5.1.

These features were then run through each of the classifiers with 10-fold cross-validation. As with the subset selection methods, these algorithms are available in the Weka Data Mining Toolkit [98]. The results of running these algorithms can be seen in Table 5.2. Table 5.3 shows the confusion matrix of the best classifier (Random Forest).

Table 5.1: Optimal subset of traditional features for discerning between the six activities

Features
Average Jerk Z
Energy Z
Entropy Z
Average (X, Y, Z)
Average Distance Between Axes (XY, YZ)
Standard Deviation (X, Y, Z)
Correlation YZ
Number of Peaks (X, Y, Z)
Number of Valleys (X, Z)

Table 5.2: Performance of classifiers for distinguishing between the six activities using the subset of features shown in Table 5.1

Classifier	Overall Accuracy (%)	F-measure
C4.5	71.2	0.71
kNN (k=4)	76.6	0.762
Multilayer Perceptron	68.0	0.672
Random Tree	69.8	0.
Random Forest	82.1	0.818

Table 5.3: Confusion matrix for discerning between the six activities using Random Forest and the feature subset shown in Table 5.1

Activity	Classified As					
	Brush Teeth	Comb Hair	Drinking	Scratch Chin	Take Meds	Wash Hands
Brush Teeth	91.7	1.07	0.98	0.27	1.87	4.09
Comb Hair	24.5	64.0	2.37	2.37	3.16	3.56
Drinking	7.93	0.51	86.7	2.05	1.53	1.28
Scratch Chin	12.05	5.42	3.61	71.1	3.01	4.82
Take Meds	17.1	1.25	4.17	1.67	72.1	3.75
Wash Hands	20.6	0.9	2.02	0.22	2.91	73.3

Table 5.4: Subset of thirty features for identifying brushing teeth

Features
Entropy (Y, Z)
Average (X, Z)
Standard Deviation Z
Number of Peaks (X, Y, Z)
Number of Valleys (X, Y)

5.2.2 Recognizing Brushing Teeth in a Naturalistic Setting

Having shown that these activities could be differentiated from each other, we shifted our focus to testing the ability of our two-tier recognition system to recognize if and when a user brushed their teeth over the course of a day. For this analysis, any activity that was not brushing teeth was relabeled as Inactive, such that our model was simply required to differentiate between brushing teeth and everything else. For Tier I of our recognition system we followed the same steps as we did for distinguishing between the six activities, the only difference being the data that were used for training and test data. Since we were interested in identifying when a person brushed their teeth over the course of a day our training dataset needed to contain data representative of activities other than the six specified. For this reason training data consisted of data from both the first and second stages of data collection. Since the dataset changed, we again conducted subset selection (using the CFsSubset Evaluator with the BestFirst search method as before); the optimal feature subset is shown in Table 5.4. These features were extracted from each window of data and were run through the same classifiers with 10-fold cross-validation. These results can be seen in Table 5.5. Table 5.6 shows the results of the confusion matrix of the best classifier, Random Forest.

To test Tier II recognition, we evaluated these models on data from Stage III of data collection. The results of this can be seen in Table 5.7. Random Forest, the classifier

Table 5.5: Performance of classifiers for identifying brushing teeth using features shown in Table 5.4

Classifier	Overall Accuracy (%)	Overall F-measure
C4.5	96.1	0.957
kNN (k=3)	96.3	0.96
Multilayer Perceptron	96.3	0.958
Random Tree	94.7	0.947
Random Forest	96.7	0.963

Table 5.6: Confusion matrix for identifying 4 second windows of brushing teeth using C4.5 and the feature subset shown in Table 5.4

Activity	Classified As	
	Brush Teeth	Inactive
Brush Teeth	**53.1**	46.9
Inactive	0.29	**99.7**

that had produced the most accurate model, produced six false negatives and seven false positives leading to F-measure of 0.67. As these results are not really sufficient for real world applications, we introduce new features to reduce the number of errors in the next chapter. As such we leave the discussion of these errors to the next chapter.

Table 5.7: Performance of two-tier recognition on Stage III data

Classifier	F-measure
J48	0.65
kNN	0.59
Multilayer Perceptron	0.57
Random Forest	0.67
Random Tree	0.37

6. IMPROVING RECOGNITION WITH ADDITIONAL FEATURES

The 30 features we selected were able to discriminate between the six activities fairly accurately and achieve a high accuracy on Tier I recognition. However, the results of Tier II recognition make the models impractical for real-world use. Thus, to improve the recognition of brushing teeth in a naturalistic setting, we added an additional 21 features by observing and quantifying differences in the accelerometer data between activities. We discuss each feature as it relates to brushing teeth below and then present the results of adding these features.

6.1 Additional Features

As can be seen in Fig. 3.1 many of the activities displayed the see-saw pattern in the accelerometer data that one would expect from a back and forth physical gesture; however, the frequency and amplitude of this pattern tended to vary between activities. To capture these qualities we calculated 21 additional features: average and standard deviation side height (X, Y, Z), average and standard deviation of the peaks (X, Y, Z), average and standard deviation of the valleys (X, Y, Z), and axis order (XY, XZ, YZ). The side height represents the height of the segments connecting the peaks and valleys. The axis order represents the number of times the axes crossed over each other over the course of the window. In total 51 features were now calculated for each window of data.

6.2 Results

As we did with solely the 30 original features, we started our analysis from Stage I data to ensure that the addition of these features had not negatively impacted our ability to differentiate between the six activities being studied. Subset selection was performed on these 51 features, again using the CFsSubset Evaluator with the BestFirst search method,

Table 6.1: Subset of original and new features for discerning between the six activities

Features
Average Jerk Z
Average Side Height (X, Z)
Standard Deviation Height X
Energy Z
Entropy Z
Average (X, Y, Z)
Average Distance Between Axes (XZ, YZ)
Standard Deviation (Y, Z)
Axis Order YZ
Number of Peaks (X, Z)
Average of Peak (X, Y, Z)
Standard Deviation of Peaks Y
Average of Valleys (X, Z)
Standard Deviation of Valleys (Y, Z)

producing an optimal subset of 24 features shown in Table 6.1. The results of running these features through the same classifiers as before are shown in Table 6.2. Table 6.3 shows the confusion matrix of the best classifier (Random Forest).

To test our two tier recognition system, we again trained on data from Stages I and II and tested on data from Stage III. As the dataset changed we again performed subset selection on the training data using the CFsSubset Evaluator with the BestFirst search method. As we did when performing this analysis with just the original features, any data that were not characteristic of brushing teeth were relabeled as Inactive. The selected subset of features are shown in Table 6.4. The classification performance of algorithms trained on these features with 10-fold cross validation is shown in Table 6.5. The confusion matrix for C4.5, the best-performing algorithm, can be seen in Table 6.6.

As can be seen in Table 6.7, when applying these models to the data from Stage III and testing Tier II Recognition, a marked improvement in recognition performance (with the exception of the Random Tree-based model) occurs. Using this model produced four false

Table 6.2: Performance of classifiers for distinguishing between the six activities using features shown in Table 6.1

Classifier	Overall Accuracy (%)	Overall F-measure
C4.5	71.0	0.71
KNN (K=5)	75.0	0.75
Multilayer Perceptron	66.5	0.66
Random Tree	70.5	0.704
Random Forest	83.4	0.83

Table 6.3: Confusion matrix for discerning between the six activities using Random Forest and the feature subset shown in Table 6.1

Activity	Classified As					
	Brush Teeth	Comb Hair	Drinking	Scratch Chin	Take Meds	Wash Hands
Brush Teeth	93.5	0.36	0.71	0.27	2.14	3.03
Comb Hair	22.9	64.0	3.56	2.77	3.95	2.77
Drinking	6.14	1.28	88.5	1.54	1.28	1.28
Scratch Chin	13.9	3.01	2.41	73.5	4.22	3.01
Take Meds	19.6	0.42	5.42	1.67	70.0	2.92
Wash Hands	19.51	0.45	1.79	0.22	2.92	75.1

Table 6.4: Feature subset for brushing teeth

Features
Entropy Z
Average (X, Z)
Axis Order (YZ)
Number of Peaks Y
Standard Deviation Peaks (X, Y, Z)
Number of Valleys (X, Y)
Standard Deviation Valleys (X, Y, Z)

Table 6.5: Performance of classifiers for identifying four-second windows of brushing teeth using the feature subset shown in Table 6.4

Classifier	Overall Accuracy (%)	F-measure
C4.5	96.1	0.620
KNN (K=5)	95.3	0.636
Multilayer Perceptron	96.0	0.594
Random Tree	94.4	0.560
Random Forest	96.5	0.657

Table 6.6: Confusion matrix for identifying 4 second windows of brushing teeth using C4.5 and the feature subset shown in Table 6.4

Activity	Classified As	
	Brush Teeth	Inactive
Brush Teeth	**49.7**	50.3
Inactive	0.73	**99.3**

negatives and one false positive.

6.3 Evaluating the Effect of Additional Features

When attempting to discriminate between the six activities, Table 5.2 clearly shows that the original features were more than sufficient to reliably achieve this goal. As can be seen in Table 6.2, the addition of the new features did not negatively impact the performance of the algorithms; instead their addition led to a slight improvement in recognition in the case of the best-performing algorithm.

Evaluating the effect of the new features on recognizing brushing teeth is not as clear cut. Comparing the results in Table 5.5 to those in Table 6.4, we can see that the performance of the algorithms did decrease with the addition of the new features. However, with the introduction of the new features the results of applying Tier II recognition did

Table 6.7: Performance of our two-tier recognition system with the feature subset shown in Table 6.4

Classifier	F-measure
C4.5	0.82
KNN	0.65
Multilayer Perceptron	0.71
Random Forest	0.76
Random Tree	0.33

improve (as can be seen by comparing Table 5.7 to Table 6.7). More specifically, the number of False Positives produced by applying Tier II Recognition on windows classified by C4.5 decreased by six and the number of False Negatives decreased by one. Thus the F-measure improved from 0.67 with the original features to 0.82 with the new features. In the following sections we discuss the feature subsets chosen in each case and the errors that were removed by the introduction of the new features. The errors that still occurred despite their introduction are also addressed

6.3.1 Algorithm Accuracy

It is important to note that when designing our algorithm (as discussion in Sec. 4.1), we limited the definition of brushing one's teeth to the back and forth movements typically associated with this activity. However, controlling how the data were labeled by users was a challenge (recall that in Stage III the users were tasked to self-annotate the data using the Android application). In other words, given that much of this study was not conducted in a controlled laboratory setting, it was impossible to limit the sections of data labeled as brushing teeth to just those consisting of the actual physical motions of brushing teeth. Therefore, actions such as using the application and putting the phone down were most likely mislabeled as brushing one's teeth.

When looking at the labeled data, this phenomenon manifested itself most often in the mis-classification of the first and/or last several windows labeled by the user as brushing teeth. In most cases the period during which the user literally brushed their teeth still meets the Tier II thresholds (the exceptions are discussed in Sec. 6.3.3.2). Furthermore, in a real-world application, which would be promoting brushing teeth for at least 2-3 minutes, the actions separating the literal and colloquial definition of brushing teeth would not factor into this time period.

6.3.2 Comparing Feature Subsets

Interestingly, most of the features selected by performing subset selection on just the original features were present in the subset of both the original and new features. Table 6.8 shows the subset of features selected for differentiating between the six activities. When recognition was based on just the original features, subset selection chose 16 out of the 30 features; when the new features were added subset selection chose 23 out of the 51 features. Of these selected features, 11 features were chosen in both instances.

Table 6.9 shows the subset of features selected for identifying brushing teeth. When subset selection was applied to just the original features, it chose 10 out of the 30 features; when it was applied to the combination of original and new features, it chose 13 out of the 51 features. These subset selections had 6 features in common.

Three features showed up in each case of subset selection: Entropy Z, Average X, and Average Z, suggesting that these features are fundamental in differentiating between activity and inactivity. When subset selection was performed on just the original features, there were five features that remained in common: Standard Deviation Z, Number of Peaks X, Number of Peaks Y, Number of Peaks Z, and Number of Valleys X, suggesting that of the original 30 features, these particular features were the most critical to distinguishing brushing teeth from other activities. When looking at the subsets of the combination of original and new features, there were four features they held in common: Standard

Table 6.8: Comparison of feature subsets for differentiating between the six activities. The common features are bold-faced.

Subset of Original Features	Subset of Original + New Features
Average Jerk Z	**Average Jerk Z**
Energy Z	Average Side Height (X, Z)
Entropy Z	Standard Deviation Side Height X
Average(X, Y, Z)	**Energy Z**
Average Distance Between Axes (XY, YZ)	**Entropy Z**
Standard Deviation (X, **Y**, **Z**)	**Average (X, Y, Z)**
Correlation YZ	**Average Distance Between Axes (XY, YZ)**
Number of Peaks (**X**, Y, **Z**)	**Standard Deviation of (Y, Z)**
Number of Valleys (X, Z)	Axis Order YZ
	Number of Peaks (X, Z)
	Average of Peaks (X, Y, Z)
	Standard Deviation of Peaks Y
	Average of Valleys (X, Z)
	Standard Deviation of Valleys (Y, Z)

Deviation of Valleys Y, Standard Deviation of Valleys Z, Standard Deviation of Peaks Y, and Axis Order YZ. The presence of these particular features in both subsets suggests that they were the most important features out of the 51 total features in identifying when a user was brushing teeth, as the performance of our system improved with the addition of the new features.

6.3.3 False Negatives and False Positives

With the addition of the new features, the best-performing algorithm (C4.5) produced five errors: four false negatives and one false positive. None of these errors were new, i.e. the addition of the new features did not lead to any errors that the original features did not produce. In the following sections we first discuss the errors that were fixed by adding these new features and then discuss the remaining errors.

Table 6.9: Comparison of feature subsets for identifying brushing teeth. The common features are bold-faced.

Subset of Original Features	Subset of Original + New Features
Entropy (Y, **Z**)	**Entropy Z**
Average (X, Z)	**Average (X, Z)**
Standard Deviation Z	Axis Order (YZ)
Number of Peaks (X, **Y**, Z)	**Number of Peaks Y**
Number of Valleys (X, Y)	Standard Deviation of Peaks (X, Y, Z)
	Number of Valleys (X, Y)
	Standard Deviation of Valleys (X, Y, Z)

6.3.3.1 Errors Removed by Introducing New Features

The biggest source of improvement by introducing the new features was in the resulting false positives: the original features produced seven false positives, while the original and new features produced only one false positive. Below we discuss each of these seven false positives and how the classification of their comprising windows differed when the features used changed.

One of the false positives produced is shown in Fig. 6.1. This figure shows a one minute period which the algorithm trained on the original 30 features thought consisted almost entirely of windows of brushing teeth. With the addition of the new features, the model correctly identified this entire period as Inactive.

Another false positive fixed by the introduction of the new features is shown in Fig. 6.2. Clearly in this example the introduction of Axis Order YZ feature had a significant impact on the recognition accuracy, as the sections of this false positive where the Y and Z axes intersected frequently were recognized with much higher accuracy. In both cases the introduction of this feature led to the amount of time the algorithm thought the user spent brushing their teeth to fall below the minimum activity threshold of 45 seconds.

In Fig. 6.3 there are two periods which are almost completely comprised of windows incorrectly classified as brushing teeth with the original features. With the introduction of the new features, the percentage of these windows within these sections dropped well below the required 75% activity percentage threshold. Similarly in Fig.s 6.4 & 6.5 the introduction of the new features decreased the number of windows classified as brushing teeth within the sequences, enough to cause the percentage of windows within this sequence classified as brushing teeth to fall below the activity percentage threshold. In Fig. 6.6 the introduction of the new features leads to the correct classification of a twenty second period towards the beginning of the sequence. As a result the sequence is no longer long enough to be considered an instance of brushing teeth.

The addition of the new features also removed one of the false negatives that had occurred when the system was built solely on the original features. As shown in Fig. 6.7, with the original features, the tail end of the sequence consisted entirely of misclassified windows, thereby causing the instance to fall short of the minimum activity duration. However, with the new features several windows at the very end were correctly identified. Looking at the data, it makes sense that the break in the repetitive motion that occurs from about 43 seconds to 52 seconds would be misclassified, as it does seem like the user performed a different gesture during that time and then resumed brushing their teeth for a few seconds afterwards.

6.3.3.2 Remaining Errors

In the case of the best-performing model (C4.5), our recognition system still produced six errors: 5 false negatives and 1 false positives. As each case occurred for different reasons, we address each separately.

The false negative shown in Fig. 6.8 depicts an approximately 30 second period during which the user indicated they had brushed their teeth. Our system did not classify any of the windows present in this period as brushing teeth. Even if it had, this sequence

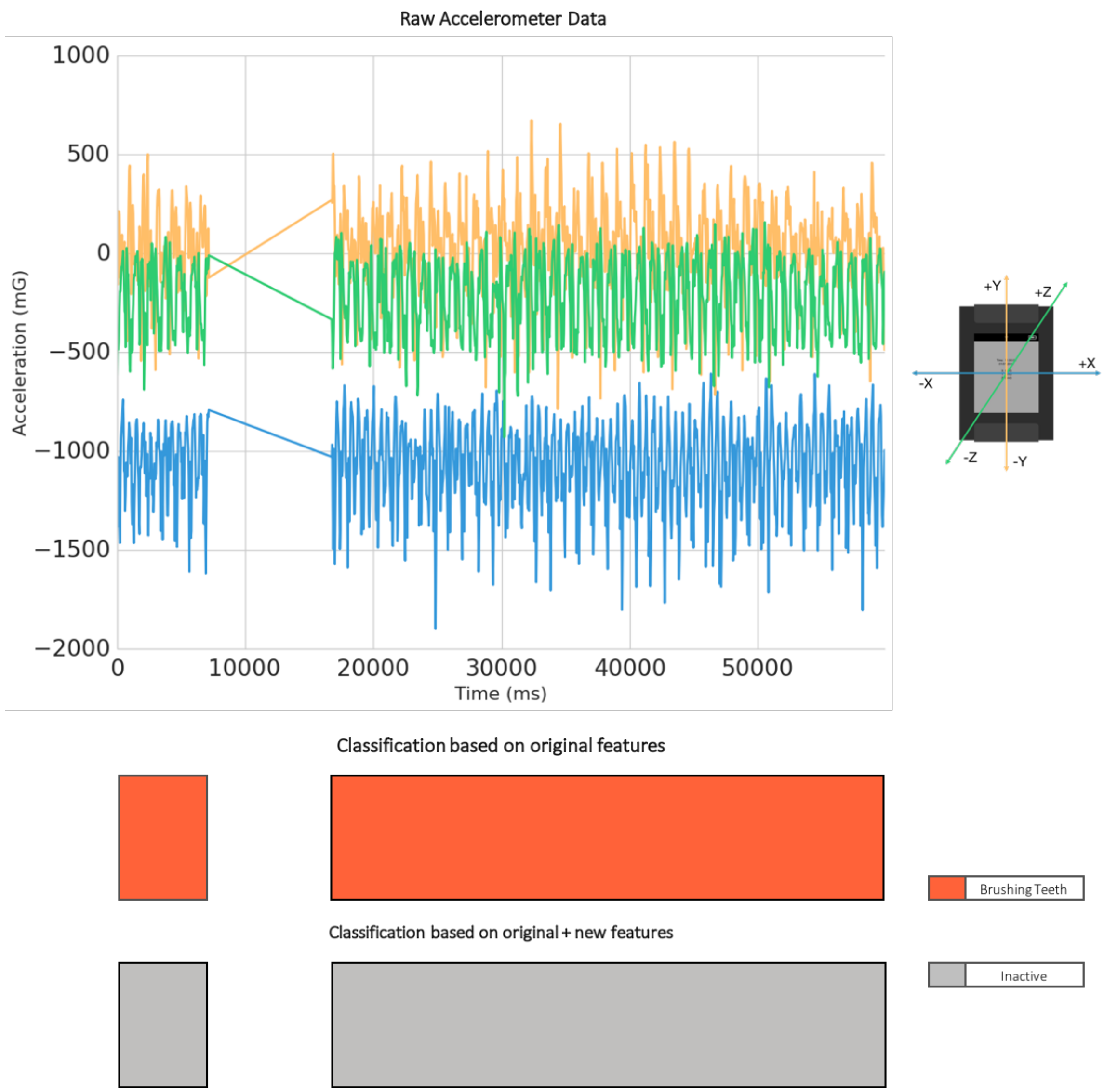

Figure 6.1: The graphed period shows an approximately one minute long false positive, which occurred when windows were classified using the original features. The brief gap in data (which lasted about 7 seconds) was not included in any window; it can be treated as an Inactive period. In other words, if that period had lasted longer than the 15 second maximum inactivity duration threshold, the two areas of data that it spans would not be considered part of the same possible activity. With just the original features every window of data in this period were classified as brushing teeth, while the addition of the new features led to this entire period being classified as Inactive.

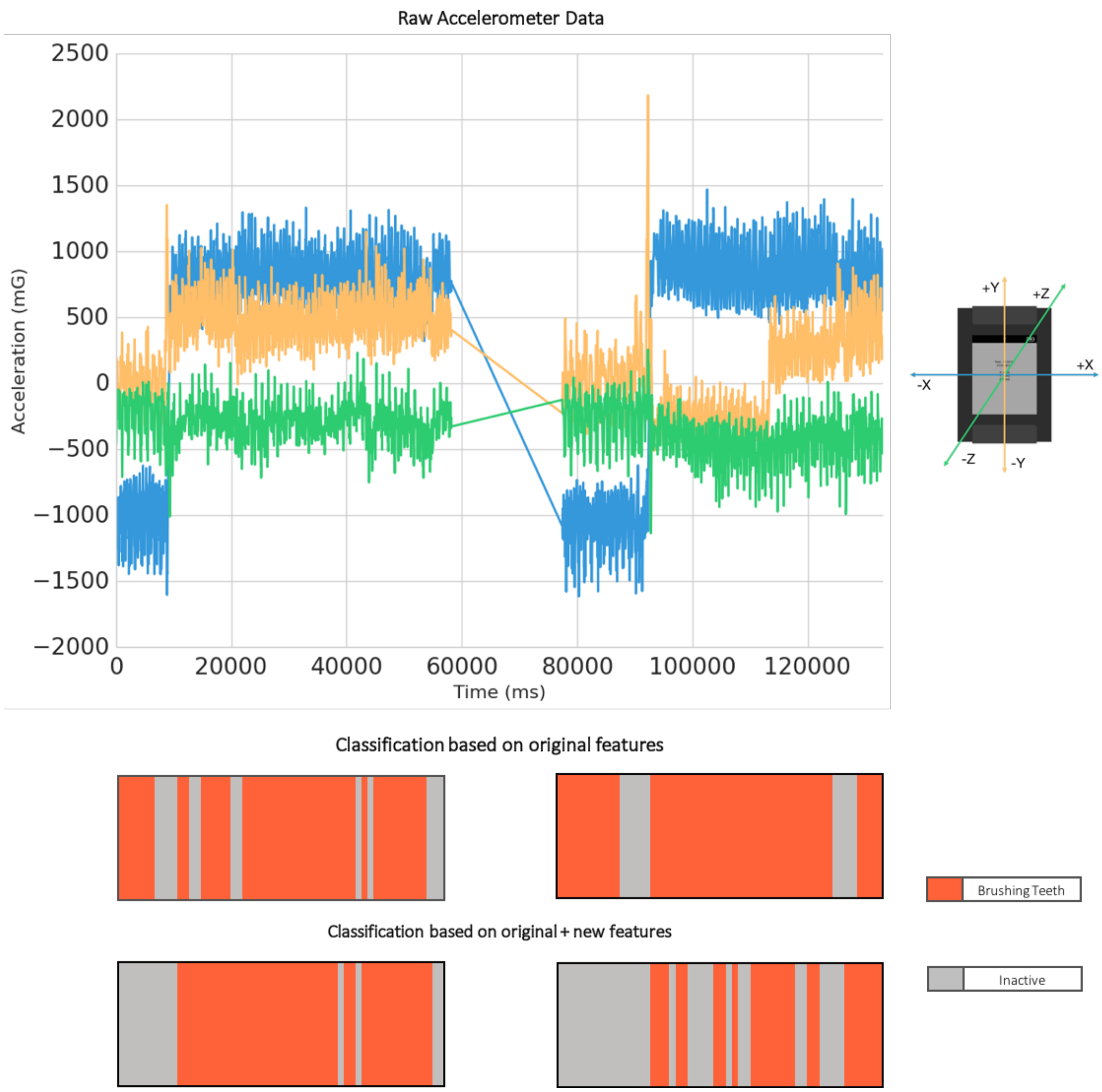

Figure 6.2: This graph shows a period during which two false positives occurred in quick succession when windows were classified using the original features. In this example, the gap in the data lasted for 19 seconds, which is longer than the minimum inactivity threshold of 15 seconds. However, since this gap is still relatively small and surrounded by two false positives, this example was treated as a single instance of a false positive. Looking at the two sections surrounding the gap, we can see that most of the improvement in classification with the introduction of the new features came from correctly recognizing the initial few seconds of each section (where the X values are below the Y and Z values).

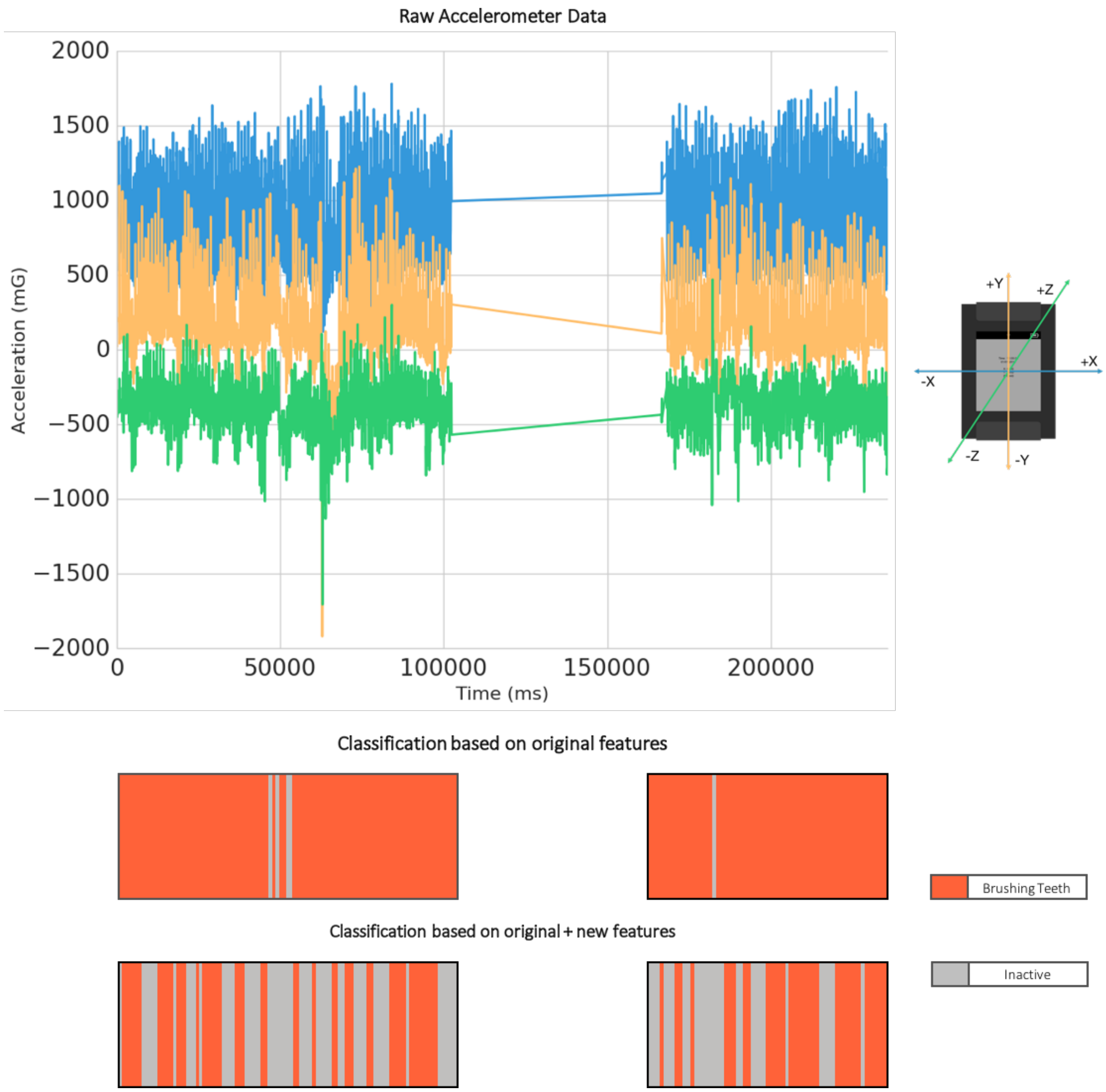

Figure 6.3: This graph shows a period during which two false positives occurred one minute apart from each other when the windows were classified using the original features alone. With the introduction of the new features, enough of the windows in each sequence are correctly identified as Inactive for Tier II recognition to determine the section is not an instance of brushing teeth.

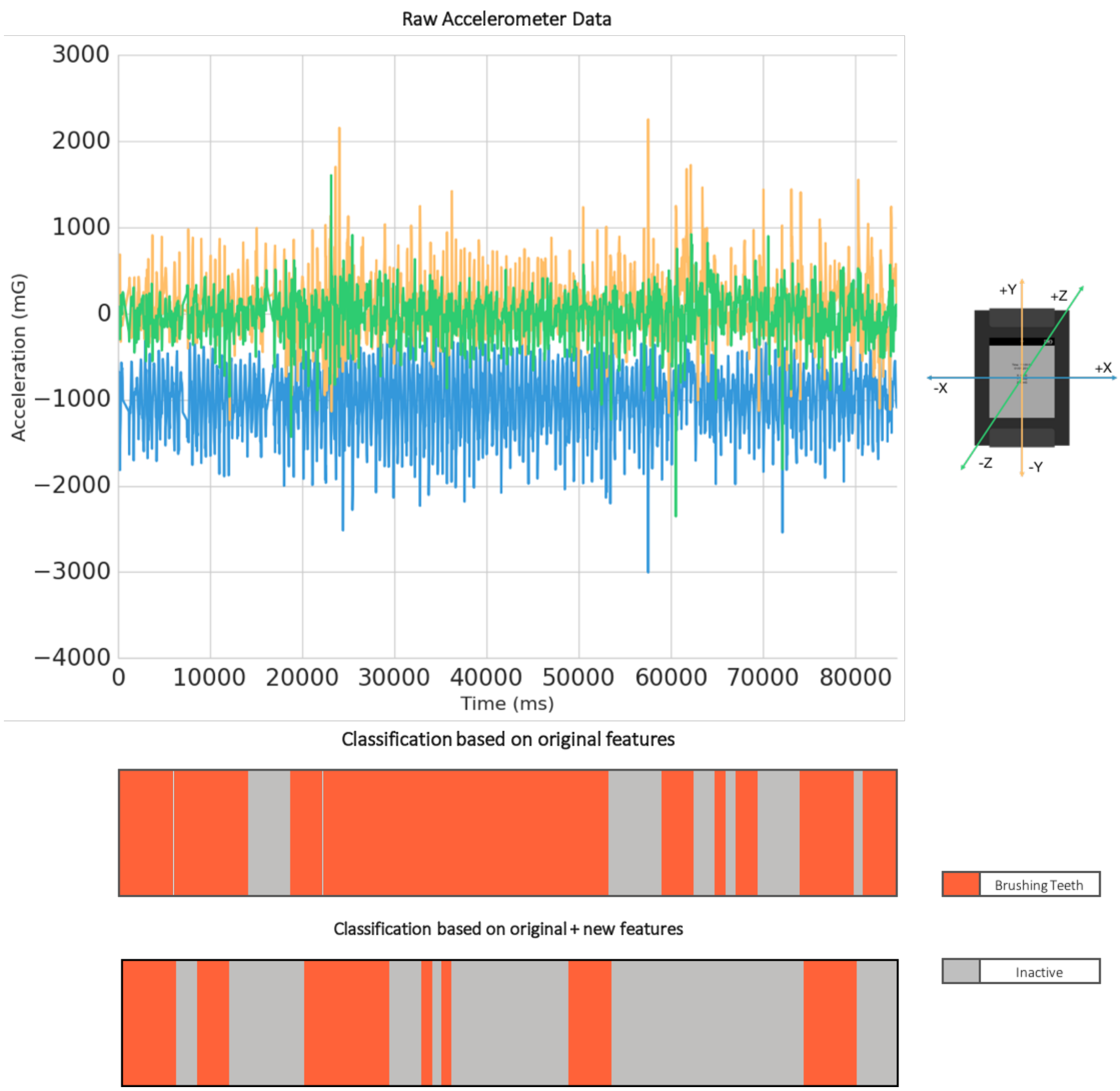

Figure 6.4: This graph shows a period during which a false positive occurred when its comprising windows were classified using the original features. As can be seen in the bottom sequence, when these windows were classified using the subset of the original and new features, the number of windows correctly identified as Inactive increased dramatically, allowing Tier II recognition to identify this section as a period of inactivity.

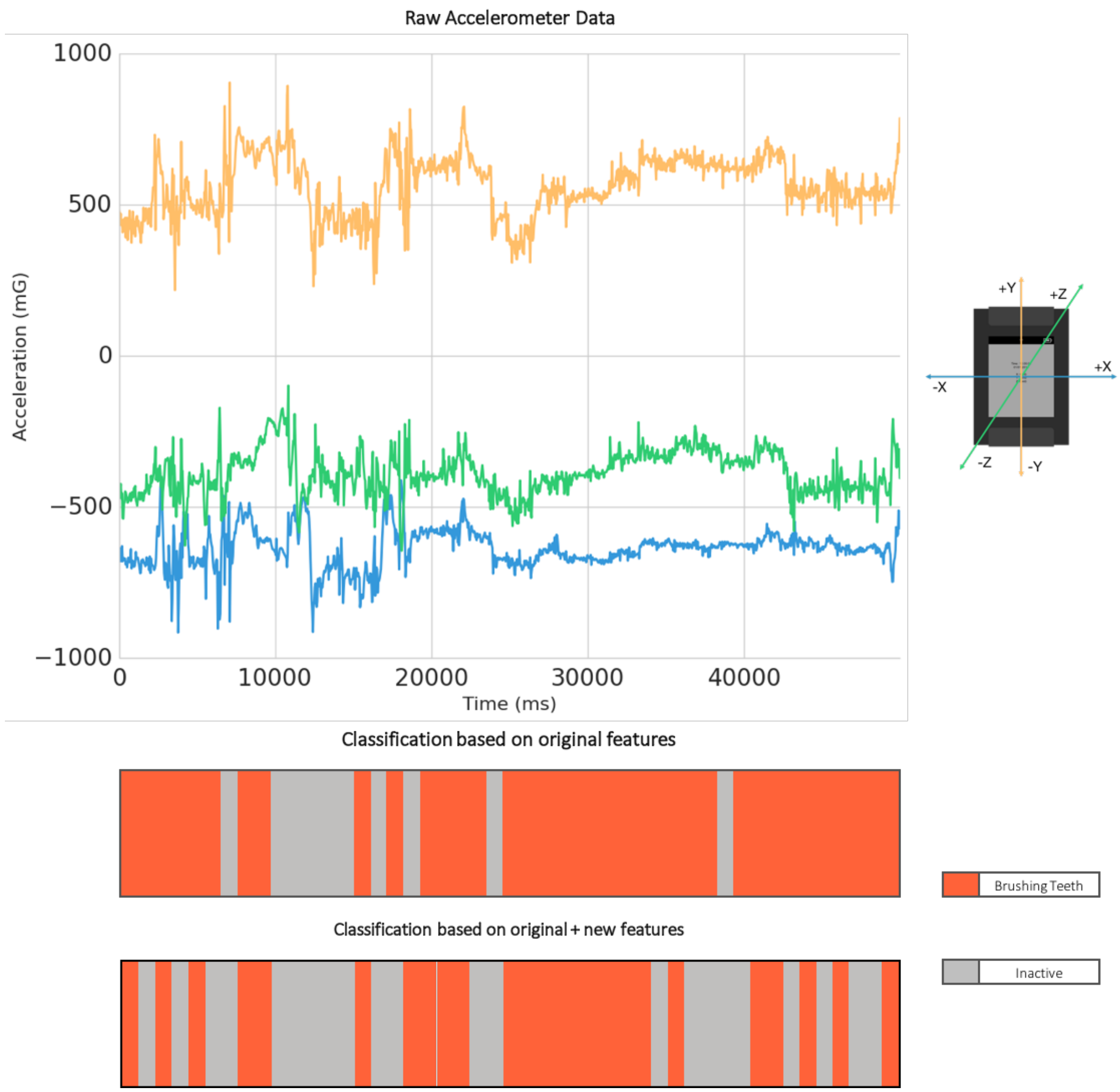

Figure 6.5: In this instance, the introduction of the new features increases the number of windows correctly classified as Inactive; as a result, the percentage of windows within the sequence classified as Brushing Teeth does not meet the activity percentage threshold of 75%.

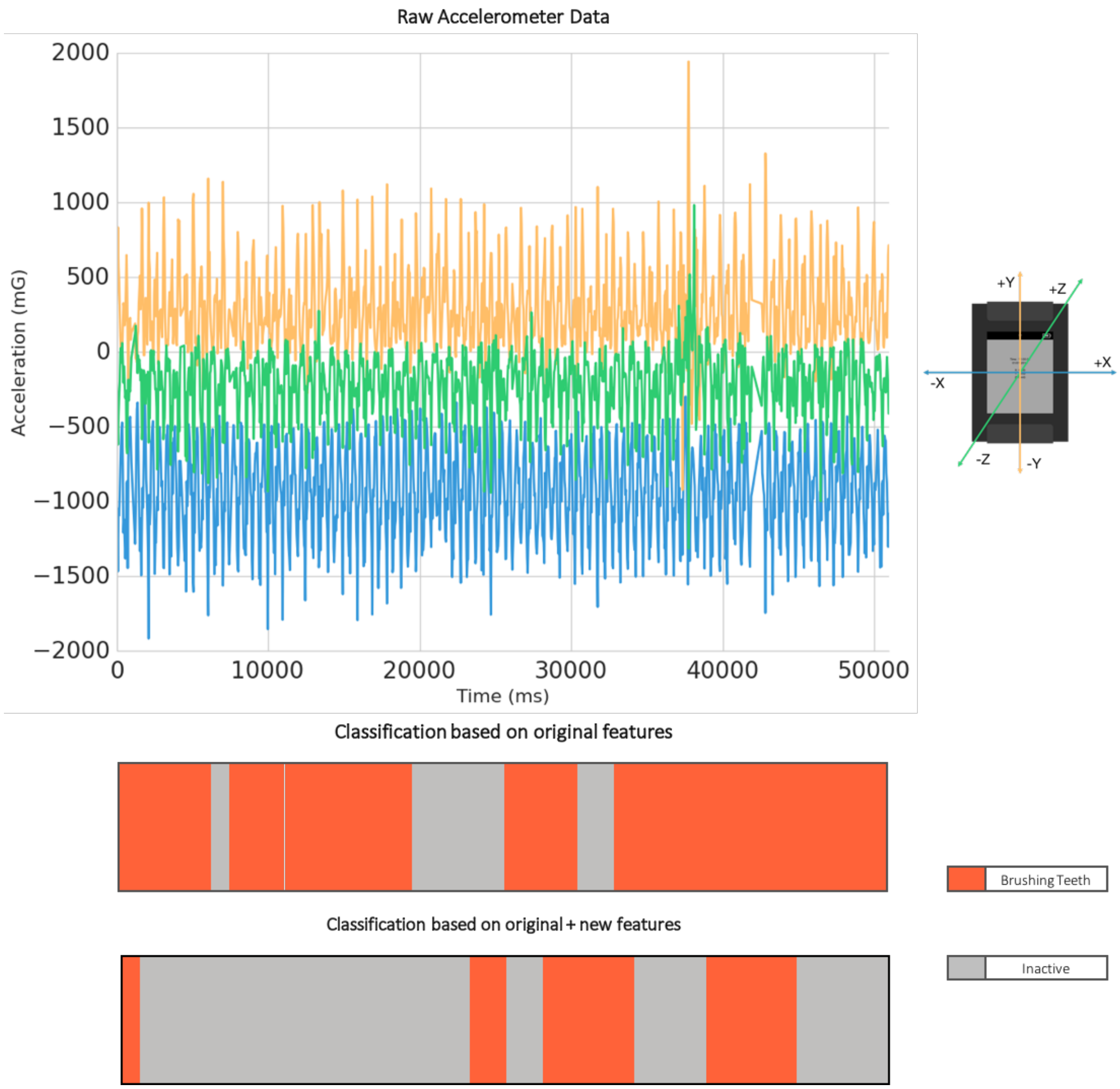

Figure 6.6: This graph shows an approximately 50 second period the two-tier recognition system mistakenly identified as an instance of brushing teeth when it relied solely on the original features. With the addition of the new features, this sequence is no longer recognized as an instance of brushing teeth.

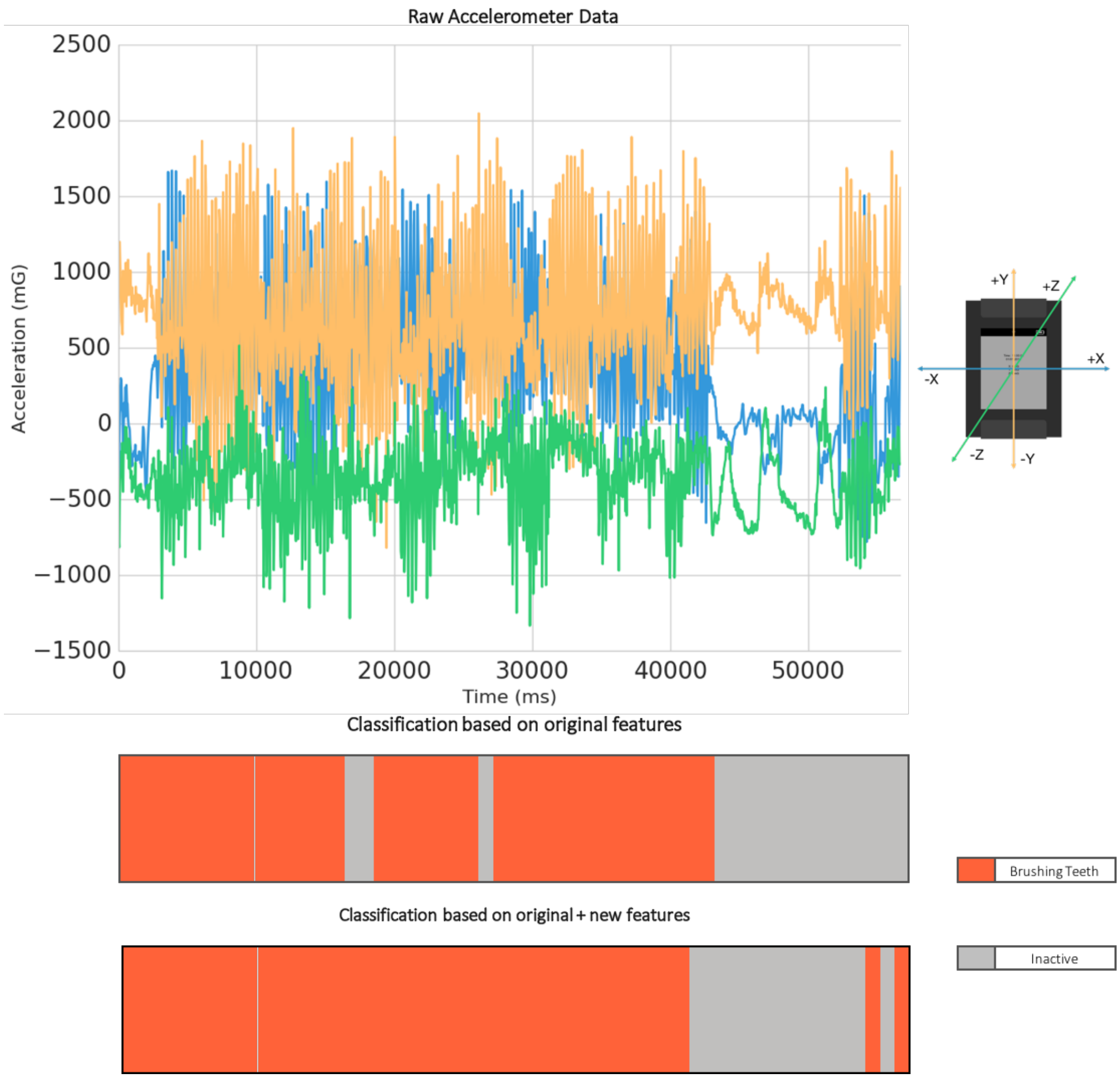

Figure 6.7: The graphed period shows an approximately one minute long false negative that occurred when the algorithm was run on the original features. With the introduction of the new features, more of the windows were correctly recognized as brushing teeth. Notably with the addition of the new features, the system was able to correctly recognize several windows at the end of the sequence.

still would have been considered a period of inactivity, since it occurred for less than the minimum activity duration threshold. These data appear to indicate that the user did not actually brush their teeth during this period, suggesting that the user may have accidentally indicated they brushed their teeth.

In the case of the false negative shown in Fig. 6.9 only 67% of the windows comprising the sequence were classified as brushing teeth. Looking at the sections that were classified as Inactive, the user appears to have paused multiple times while brushing their teeth, leading to the low activity percentage. This false negative could be averted by reducing the activity percentage threshold or by reducing the minimum activity duration. However, reducing either of these thresholds to accommodate this particular instance introduces a number of false positives. It is worth noting that if the number of false positives could be kept to a minimum or avoided altogether, recognition of smaller sections of brushing teeth could serve as the basis for motivational interventions encouraging participants to brush their teeth for longer periods of time.

In two cases the system recognized the user as having brushed their teeth for less than the time they indicated. Fig. 6.10 depicts one of these cases where the user indicated that they had brushed their teeth for 60 seconds; however, the system only identified the user as brushing their teeth for 39 seconds. Notably, the 21 seconds that were misclassified occurred at the beginning of the one minute period, and the data do seem to indicate that the user was doing something other than literally brushing their teeth. As discussed in Sec. 6.3.1 the user could have been doing other activities that comprise the colloquial definition of brushing teeth, e.g., opening the toothpaste, applying toothpaste to their toothbrush, and/or running their toothbrush under the faucet. In the other case, shown in Fig. 6.11, the user brushed their teeth for 49 seconds; however, the system predicted they had brushed their teeth for only 34 seconds. The data indicate that there does seem to be a five-second gap from about 40 to 45 seconds into the gesture where the user seems to have paused while brushing their teeth. Given that analysis was done using four-second windows with a one second overlap it is possible that this five second window altered the feature values for the windows around it enough for the model to not recognize those surrounding windows as brushing teeth.

Our algorithm also produced one false positive, which occurred right after a false negative, as can be seen in Fig. 6.12. Although the user only annotated the first 43 seconds as brushing their teeth (producing the false negative), in all likelihood the user actually brushed their teeth for 75 seconds, took a one minute break, and then continued brushing their teeth for another 67 seconds. Clearly, this approximately three minute period can be treated as one instance of the user brushing their teeth over the course of a day.

6.3.4 Noteworthy Cases

There is one unique case worth mentioning that we classified as a true positive. The user indicated that they brushed their teeth for 26 seconds, which is clearly below the minimum activity duration of 45 seconds that we set. However, our algorithm classified the user as having brushed their teeth for 80 seconds. As can be seen in Fig. 6.13, the data suggest that the user did in fact brush their teeth for longer than they indicated. We believe the brief interlude after the end of the labeled brushing teeth data is when the user switched the label from ”Brushing Teeth” to ”Nothing” on the Android application.

As we stated in Chapter 1, many users rely on memory or on written records to track health metrics. However, these approaches to tracking activity are susceptible to inaccuracy due to memory lapses and/or failing to diligently keep records. In that light, this particular example is encouraging, as it suggests that our system is more accurate than the traditional methods of activity tracking.

By discussing these errors that persisted even with the addition of new features,we aim to show that most of these errors would not change the experience of using such a

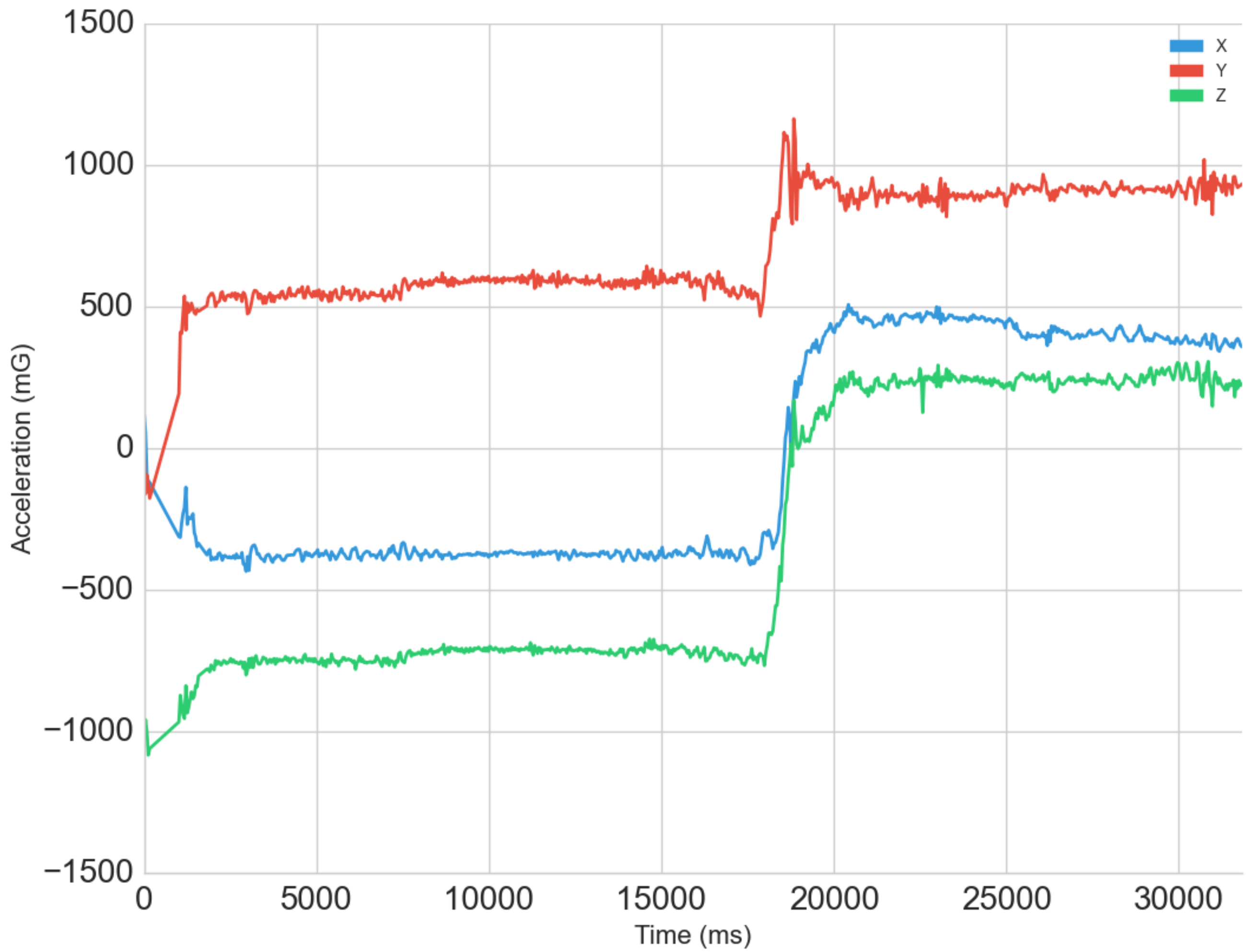

Figure 6.8: This graph shows a short period consisting of one false negative. Although the user indicated that he or she brushed their teeth during this time, the data seem to indicate that they were doing something else. None of the windows in this period were classified as brushing teeth.

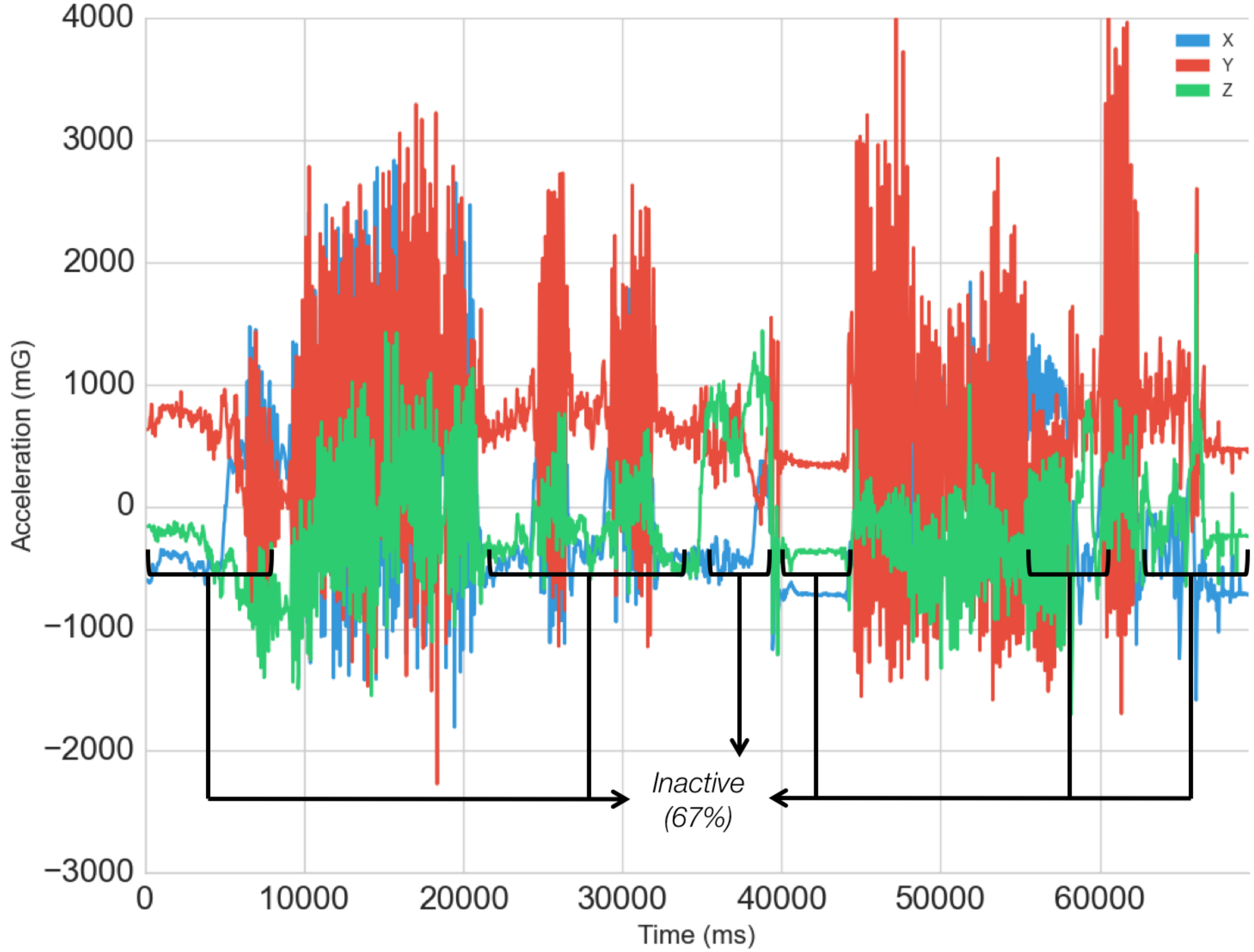

Figure 6.9: During this period the user appears to have been brushing their teeth; however, it looks like they took several pauses in between. Our system classified these pauses as Inactive, and as a result the activity percentage was only 67%; which did not meet the 75% threshold.

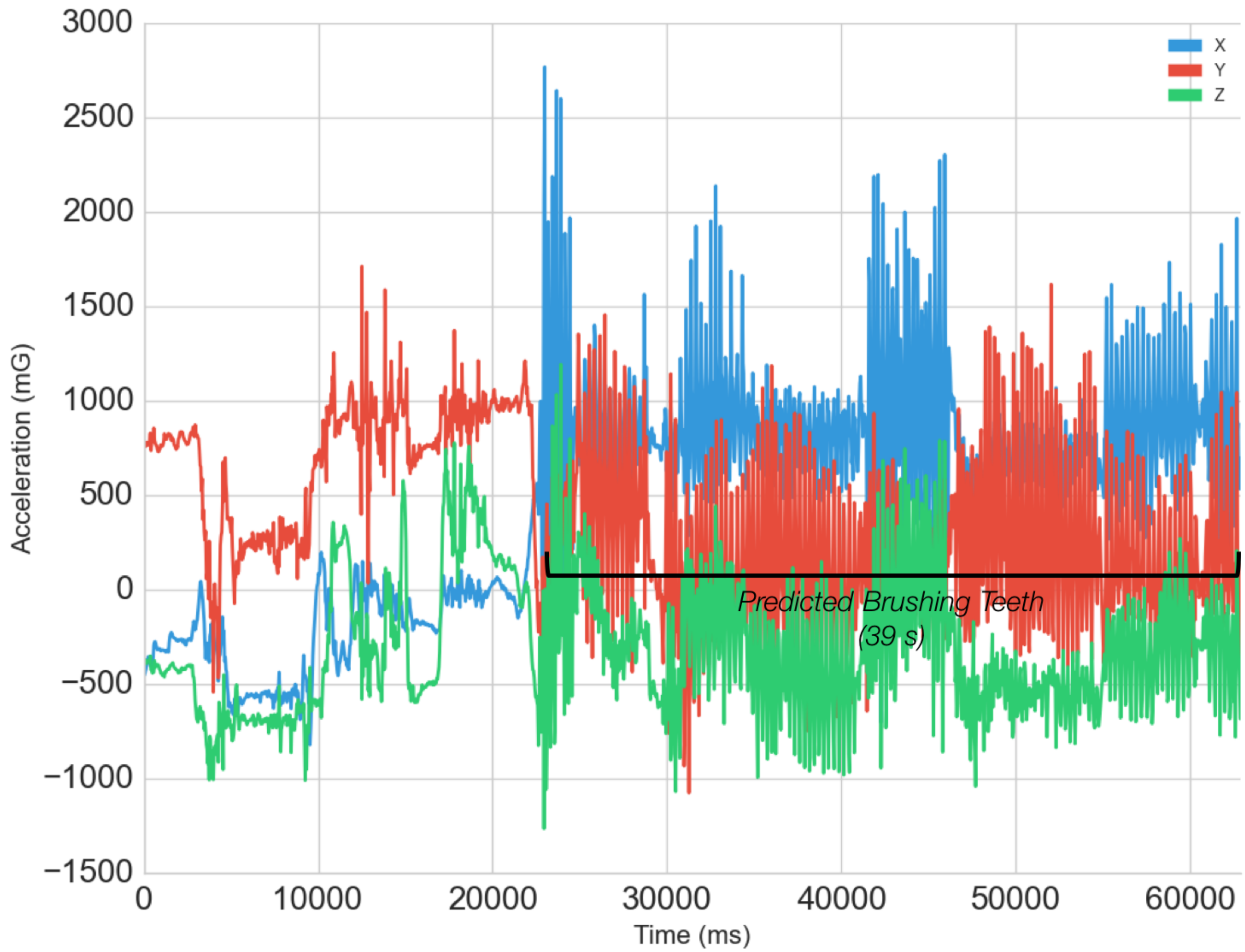

Figure 6.10: This approximately 60-second period resulted in a false negative. As can be seen the first 21 seconds of this period were not recognized as brushing teeth; most likely the users were not literally brushing their teeth during this period. The remaining 39 seconds of the period were correctly classified as brushing teeth.

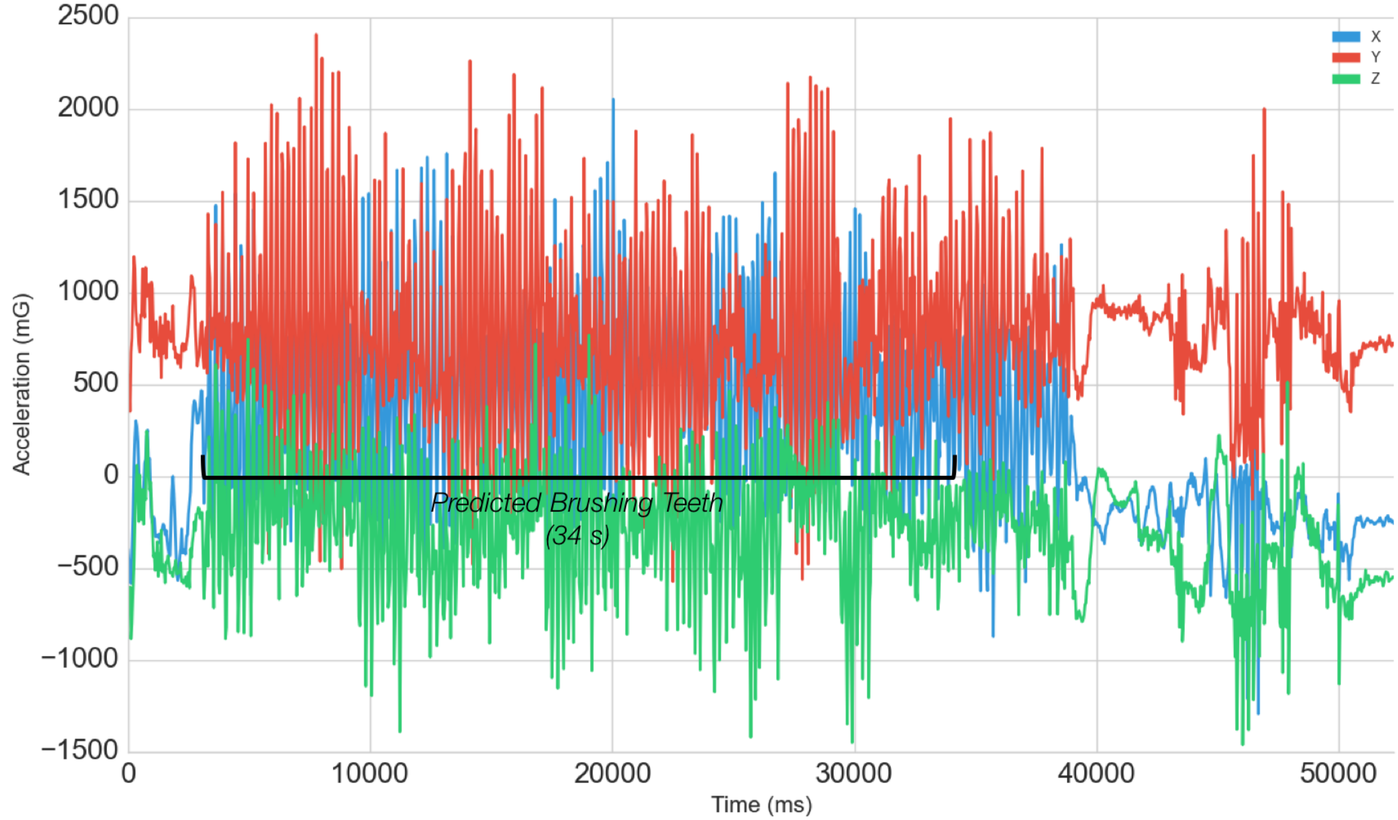

Figure 6.11: This image shows a three-minute period consisting of one false negative. The section labeled “Predicted Brushing Teeth” contained continuous instances of four-second windows classified as brushing teeth. However, because the entire length was less than one minute, the entire time was not ultimately classified as an instance of brushing teeth during the second pass.

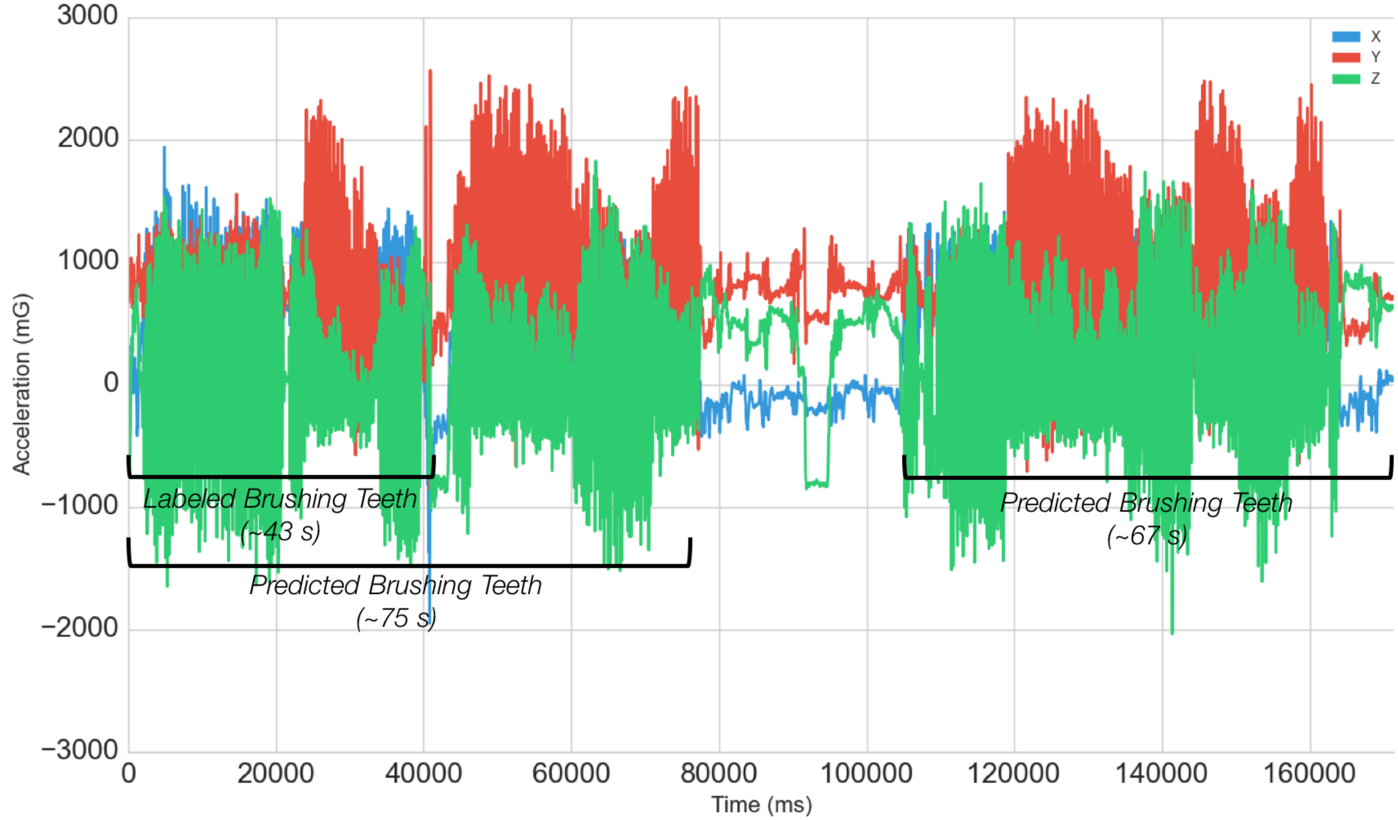

Figure 6.12: The image above shows both a three minute period consisting of both a true positive and the only false positive. The two sections labeled "Predicted Brushing Teeth" consisted of continuous stretches of four-second windows classified as brushing teeth. However, the user only labeled the first 43 seconds of the first stretch as actually brushing teeth. Looking at the data immediately following the 43 second period, a small break in the data can be seen during which the user most likely pressed the button on the Android app to change the activity from Brushing Teeth to Inactivity. We think there is a good chance that this person did continue brushing his/her teeth even after changing the label.

system in a real-time application. Several of the errors were likely due to human error, i.e., the data was mislabeled. As the user would not be tasked with labeling the data in a real-world application, these errors would not occur. In a system designed to ensure that users were properly brushing their teeth, failing to recognize that the user brushed their teeth when they did not brush their teeth for the recommended 1-2 minutes would be an understandable, if not optimal, response. Encouragingly, Tier I of our recognition system does seem to recognize the comprising windows of smaller periods of brushing teeth; this could serve as the basis of a more sophisticated Tier II. As stated earlier, such a system that recognized smaller periods of time spent brushing teeth could be used to drive motivational interventions. Additionally, efforts could be made to recognize the other actions comprising the colloquial definition of brushing teeth (putting toothpaste on the toothbrush, running the toothbrush under the faucet, cleaning the toothbrush, etc.) to improve the recognition of this activity.

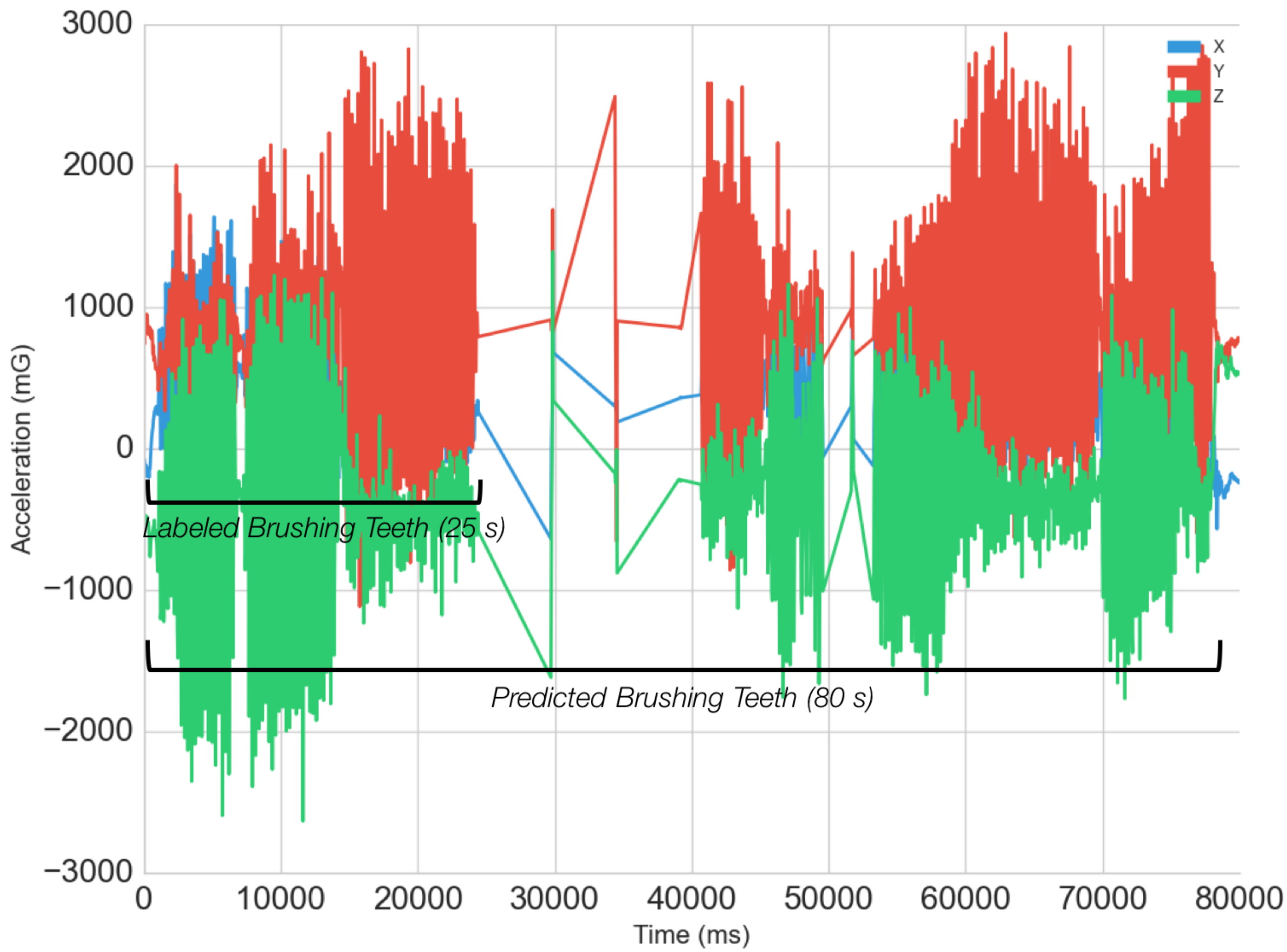

Figure 6.13: The image above shows a 80 second period consisting of a false negative. The user labeled the first 25 seconds as brushing teeth, which is below the minimum activity duration threshold for brushing teeth. As can be seen in the graph, there is a period immediately following this section where the user most likely used the Android application, and then continued to brush their teeth afterwards. In such a real life situation our system would correctly identify this occurence as an instance of brushing teeth.

7. FUTURE WORK

Our work thus far suggests several avenues for further research. Based on our results, we believe that our system can be extended to other activities and to a real-time application. Furthermore, we believe that this recognition algorithm could serve as the underlying intelligence of a system designed to provide users with personalized feedback and interventions.

7.1 Other Activities

In this work we outline a two-tier recognition system for activity recognition but have yet to test it on activities other than brushing one's teeth. Possibly the most promising candidate for recognition in our dataset is washing hands, as it also features a back and forth hand motion - which our features were able to identify in brushing teeth - and poses significant health implications [102, 103]. Furthermore, its Tier II thresholds can be easily established, as the CDC recommends that people wash their hands for at least 20 seconds [104]. Taking medication is also worth studying, as medication nonadherence is a widespread and expensive problem worldwide [105–107]; however given that this activity is not a back and forth hand motion, recognition is potentially much more challenging.

7.2 Real-Time Recognition

While our work thus far has focused on recognizing activities in a naturalistic setting, we have yet to recognize activities in real-time. We established the framework of the two-tier recognition with the end goal of real time recognition in mind, and we believe that it is entirely possible to implement this system into a real-time system. We envision our recognition running in the background on a smartphone, classifying accelerometer data from a smartwatch in real-time.

7.3 Personalized Feedback & Interventions

The eventual goal of this work is to develop intelligent personalized feedback and interventions based on this recognition. There are many potential use cases; however, we focus our discussion here on a few.

One use case is developing a system for people suffering from dementia. As mentioned in Chapter 1, in the early stages of dementia, patients may need to be reminded that they must brush their teeth and/or be supervised while doing so. In later stages, patients may be unable to brush their teeth, stop understanding that they need to clean their teeth, or lose interest in doing so [32].

Another use case would be using detection of washing hands as an indirect indicator of bathroom use, as incontinence is a leading factor in nursing home placement [108]. By recognizing these activities, we can provide users with interventions such as reminders, navigation to the bathroom with vibration cues (as in [109–112]), encouragement, and suggestions, or even alert the user's physician and/or family members if something might be wrong.

8. CONCLUSION

In this work we were able to differentiate between six everyday activities: brushing one's teeth, combing one's hair, scratching one's chin, washing one's hands, taking medication, and drinking using accelerometer data collected from a Pebble smartwatch with an overall accuracy of 83.4% and an F-measure of 0.83. Furthermore, we were able to determine when a person brushed their teeth over the course of a day with an F-measure of 0.82.

To achieve these results we presented two main contributions. First, we developed a two-tier recognition system, where the first tier employed traditional activity recognition (segmenting data into dynamic windows, extracting features, running those features through classification algorithms), and the second tier grouped classified windows into sequences based on three thresholds to find instances of brushing teeth. To test the efficacy of this system, we collected data from 38 participants over the course of three user studies. In the first study, users were asked to perform each activity in a controlled environment; however, in the second and third study users were asked to perform the activities at any time they wished over the course of one hour and one day, respectively.

Second, to improve the performance of this system, we developed 21 novel features specific to the action of brushing teeth. These 21 additional features improved the system's ability to distinguish between the six activities by 1.3% and improved the system's ability to identify brushing teeth by 15%. Although our eventual system did lead to several errors, we discussed each error and reached the conclusion that these errors would not necessarily occur or matter if such a system was implemented into a real-world application. Thus, through this work we show that it is possible to identify specific health activities through the use of features specifically designed to highlight the unique characteristics of those activities. Through recognition of these activities, specific personalized health interventions can be developed to improve the lives of the average person.

REFERENCES

[1] Richard Smith. The future of healthcare systems. *BMJ: British Medical Journal*, 314(7093):1495, 1997.

[2] Paul Rubel, Jocelyne Fayn, Lucas Simon-Chautemps, Hussein Atoui, Mattias Ohlsson, David Telisson, Stefano Adami, Sébastien Arod, Marie Claire Forlini, Cesare Malossi, et al. New paradigms in telemedicine: ambient intelligence, wearable, pervasive and personalized. *Stud Health Technol Inform*, 108:123–132, 2004.

[3] A Lymberis. Smart wearables for remote health monitoring, from prevention to rehabilitation: current r&d, future challenges. In *Information Technology Applications in Biomedicine, 2003. 4th International IEEE EMBS Special Topic Conference on*, pages 272–275. IEEE, 2003.

[4] Maulik D Majmudar, Lina Avancini Colucci, and Adam B Landman. The quantified patient of the future: Opportunities and challenges. In *Healthcare*, volume 3, pages 153–156. Elsevier, 2015.

[5] United Nations, Department of Economic and Social Affairs, Population Division (2015). World Population Ageing 2015 (ST/ESA/SER.A/390).

[6] Kaare Christensen, Gabriele Doblhammer, Roland Rau, and James W Vaupel. Ageing populations: the challenges ahead. *The Lancet*, 374(9696):1196–1208, 2009.

[7] Institute for Health Metrics and Evaluation (IHME). Rethinking Development and Health: Findings from the Global Burden of Disease Study. Seattle, WA: IHME, 2016.

[8] Robin Saxby. How silicon will transform healthcare. In *Medical Electrical Devices*

and Technology, 2007. MEDTECH 2007. 2007 3rd Institution of Engineering and technology International Conference on, pages 135–150. IET, 2007.

[9] Namita Jayaprakash, Ronan O'Sullivan, Tareg Bey, Suleman S Ahmed, and Shahram Lotfipour. Crowding and delivery of healthcare in emergency departments: the european perspective. *Western Journal of Emergency Medicine*, 10(4), 2009.

[10] Susannah Fox and Maeve Duggan. *Tracking for Health*. Pew Research Center's Internet & American Life Project, 2013.

[11] Tanzeem Choudhury, Matthai Philipose, Danny Wyatt, and Jonathan Lester. Towards activity databases: Using sensors and statistical models to summarize people's lives. *IEEE Data Eng. Bull.*, 29(1):49–58, 2006.

[12] Gary M Weiss, Jessica L Timko, Catherine M Gallagher, Kenichi Yoneda, and Andrew J Schreiber. Smartwatch-based activity recognition: A machine learning approach. In *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pages 426–429. IEEE, 2016.

[13] Natalie Wilde, Katrin Hänsel, Hamed Haddadi, and Akram Alomainy. Wearable computing for health and fitness: Exploring the relationship between data and human behaviour. *arXiv preprint arXiv:1509.05238*, 2015.

[14] Daniel W Goldberg, Myles G Cockburn, Tracy A Hammond, Geoffrey M Jacquez, Daniel Janies, Craig Knoblock, Werner Kuhn, Edward Pultar, and Martin Raubal. Envisioning a future for a spatial-health cybergis marketplace. In *Proceedings of the Second ACM SIGSPATIAL International Workshop on the Use of GIS in Public Health*, pages 27–30. ACM, 2013.

[15] Yan Hong, Deborah Vollmer Dahlke, Marcia Ory, Daniel Goldberg, Jessica Cargill, Debra Kellstedt, Jairus Pulczinski, Tracy Hammond, and Edgar Hernandez. Development of icanfit: A mobile device application to promote physical activity and access to health information among older cancer survivors. In *142nd APHA*

Annual Meeting and Exposition (November 15-November 19, 2014). APHA, 2014. 1 page.

[16] Gerard F Anderson. *Chronic Care: Making the Case for Ongoing Care*. Robert Wood Johnson Foundation, Feb 2010.

[17] Deborah Ann Amini, Kathy Kannenberg, Stefanie Bodison, Pei-Fen Chang, Donna Colaianni, Beth Goodrich, et al. Occupational therapy practice framework: Domain and process 3rd edition. *The American Journal of Occupational Therapy*, 68:S19–S20, 2014.

[18] Diane Rowland. Measuring the elderly's need for home care. *Health Affairs*, 8(4):39–51, 1989.

[19] Jie Yin, Qiang Yang, and Jeffrey Junfeng Pan. Sensor-based abnormal human-activity detection. *IEEE Transactions on Knowledge and Data Engineering*, 20(8):1082–1090, 2008.

[20] Kenton O'Hara, Cecily Morrison, Abigail Sellen, Nadia Bianchi-Berthouze, and Cathy Craig. Body tracking in healthcare. *Synthesis Lectures on Assistive, Rehabilitative, and Health-Preserving*, 5(1):1–151, 2016.

[21] Gregory D Abowd and Elizabeth D Mynatt. Charting past, present, and future research in ubiquitous computing. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7(1):29–58, 2000.

[22] Association Alzheimers. 2015 alzheimer's disease facts and figures. *Alzheimer's & Dementia: the Journal of the Alzheimer's Association*, 11(3):332, 2015.

[23] Cindy L Munro and Mary Jo Grap. Oral health and care in the intensive care unit: state of the science. *American Journal of Critical Care*, 13(1):25–34, 2004.

[24] Poul Erik Petersen and Tatsuo Yamamoto. Improving the oral health of older people: the approach of the who global oral health programme. *Community Dentistry and Oral Epidemiology*, 33(2):81–92, 2005.

[25] Janette C Rahamat-Langendoen, Marit GA van Vonderen, Lex J Engström, Willem L Manson, Arie Jan Van Winkelhoff, and Esther ANM Mooi-Kokenberg. Brain abscess associated with aggregatibacter actinomycetemcomitans: case report and review of literature. *Journal of Clinical Periodontology*, 38(8):702–706, 2011.

[26] Mara S Meyer, Kaumudi Joshipura, Edward Giovannucci, and Dominique S Michaud. A review of the relationship between tooth loss, periodontal disease, and cancer. *Cancer Causes & Control*, 19(9):895–907, 2008.

[27] Aramesh Saremi, Robert G Nelson, Marshall Tulloch-Reid, Robert L Hanson, Maurice L Sievers, George W Taylor, Marc Shlossman, Peter H Bennett, Robert Genco, and William C Knowler. Periodontal disease and mortality in type 2 diabetes. *Diabetes Care*, 28(1):27–32, 2005.

[28] Armin J Grau, Heiko Becher, Christoph M Ziegler, Christoph Lichy, Florian Buggle, Claudia Kaiser, Rainer Lutz, Stefan Bültmann, Michael Preusch, and Christof E Dörfer. Periodontal disease as a risk factor for ischemic stroke. *Stroke*, 35(2):496–501, 2004.

[29] Frank DeStefano, Robert F Anda, Henry S Kahn, David F Williamson, and Carl M Russell. Dental disease and risk of coronary heart disease and mortality. *BMJ*, 306(6879):688–691, 1993.

[30] Pamela Sparks Stein, Michelle J Steffen, Charles Smith, Gregory Jicha, Jeffrey L Ebersole, Erin Abner, and Dolph Dawson. Serum antibodies to periodontal pathogens are a risk factor for alzheimers disease. *Alzheimer's & Dementia*, 8(3):196–203, 2012.

[31] Dominique S Michaud, Kaumudi Joshipura, Edward Giovannucci, and Charles S Fuchs. A prospective study of periodontal disease and pancreatic cancer in us male health professionals. *Journal of the National Cancer Institute*, 99(2):171–175, 2007.

[32] Janet Griffiths and Vicki Jones. *Dental care and oral health*. Alzheimer's Society, 2015. https://goo.gl/wZfgCt.

[33] Akin Avci, Stephan Bosch, Mihai Marin-Perianu, Raluca Marin-Perianu, and Paul Havinga. Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: A survey. In *Architecture of Computing Systems (ARCS), 2010 23rd International Conference on*, pages 1–10. VDE, 2010.

[34] Emmanuel Munguia Tapia, Stephen S Intille, and Kent Larson. *Activity recognition in the home using simple and ubiquitous sensors*. Springer, 2004.

[35] Liming Chen, Jesse Hoey, Chris D Nugent, Diane J Cook, and Zhiwen Yu. Sensor-based activity recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(6):790–808, 2012.

[36] Graeme S Chambers, Svetha Venkatesh, Geoff AW West, and Hung Hai Bui. Hierarchical recognition of intentional human gestures for sports video annotation. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, volume 2, pages 1082–1085. IEEE, 2002.

[37] F Foerster, M Smeja, and J Fahrenberg. Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring. *Computers in Human Behavior*, 15(5):571–583, 1999.

[38] Ling Bao and Stephen S Intille. Activity recognition from user-annotated acceleration data. In *Pervasive Computing*, pages 1–17. Springer, 2004.

[39] Jonathan Lester, Tanzeem Choudhury, and Gaetano Borriello. A practical approach to recognizing physical activities. In *Pervasive Computing*, pages 1–16. Springer, 2006.

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

Workshop on, pages 4–pp. IEEE, 2006.

[41] Juha Pärkkä, Miikka Ermes, Panu Korpipää, Jani Mäntyjärvi, Johannes Peltola, and Ilkka Korhonen. Activity classification using realistic data from wearable sensors. *Information Technology in Biomedicine, IEEE Transactions on*, 10(1):119–128, 2006.

[42] Emmanuel Munguia Tapia, Stephen S Intille, William Haskell, Kent Larson, Julie Wright, Abby King, and Robert Friedman. Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor. In *2007 11th IEEE international symposium on wearable computers*, pages 37–40. IEEE, 2007.

[43] Miikka Ermes, Juha Parkka, and Luc Cluitmans. Advancing from offline to online activity recognition with wearable sensors. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4451–4454. IEEE, 2008.

[44] Daniel Olguın Olguın and Alex Sandy Pentland. Human activity recognition: Accuracy across common locations for wearable sensors. In *Proceedings of 2006 10th IEEE International Symposium on Wearable Computers, Montreux, Switzerland*, pages 11–14. Citeseer, 2006.

[45] David McGlynn and Michael G Madden. An ensemble dynamic time warping classifier with application to activity recognition. In *Research and Development in Intelligent Systems XXVII*, pages 339–352. Springer, 2011.

[46] Muhammad Shoaib, Stephan Bosch, Ozlem Durmaz Incel, Hans Scholten, and Paul JM Havinga. Fusion of smartphone motion sensors for physical activity recognition. *Sensors*, 14(6):10146–10176, 2014.

[47] Perng Kangchunn, Brian Fisher, Seth Hollar, and Kristofer S. J. Pister. Acceleration sensing glove (asg). In *The Third International Symposium on Wearable Computers*

(ISWC'99), pages 178–180. IEEE, 1999.

[48] Joshua M. Peschel, Brandon Paulson, and Tracy Hammond. A surfaceless pen-based interface. In *Proceedings of the Seventh ACM Conference on Creativity and Cognition*, C&C '09, pages 433–434, New York, NY, USA, 2009. ACM.

[49] Paul Taele and Tracy Hammond. Invisishapes: A recognition system for sketched 3d primitives in continuous interaction spaces. In *Proceedings of the 2015 International Symposium on Smart Graphics, Chengdu, China*, SG, page 12, 2015.

[50] Jace Miller and Tracy Hammond. Wiiolin: A virtual instrument using the wii remote. In *Proceedings of the 2010 Conference on New Interfaces for Musical Expression (NIME)*, page 497500, Sydney, Australia, June 15-18 2010.

[51] Vijay Rajanna, Patrick Vo, Jerry Barth, Matthew Mjelde, Trevor Grey, Cassandra Oduola, and Tracy Hammond. Kinohaptics: An automated, wearable, haptic assisted, physio-therapeutic system for post-surgery rehabilitation and self-care. *Journal of Medical Systems*, 40(3):1–12, 2015.

[52] Brandon Paulson, Danielle Cummings, and Tracy Hammond. Object interaction detection using hand posture cues in an office setting. *Int. J. Hum.-Comput. Stud.*, 69(1-2):19–29, January 2011.

[53] Brandon Paulson and Tracy Hammond. Office activity recognition using hand posture cues. In *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction - Volume 2*, BCS-HCI '08, pages 75–78, Swinton, UK, UK, 2008. British Computer Society.

[54] Jani Mantyjarvi, Johan Himberg, and Tapio Seppanen. Recognizing human motion with multiple acceleration sensors. In *Systems, Man, and Cybernetics, 2001 IEEE International Conference on*, volume 2, pages 747–752. IEEE, 2001.

[55] Korbinian Frank, Vera Nadales, María Josefa, Patrick Robertson, and Michael Angermann. Reliable real-time recognition of motion related human activities using

mems inertial sensors. 2010.

[56] Zhenyu He and Lianwen Jin. Activity recognition from acceleration data based on discrete consine transform and svm. In *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on*, pages 5041–5044. IEEE, 2009.

[57] Chun Zhu and Weihua Sheng. Human daily activity recognition in robot-assisted living using multi-sensor fusion. In *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*, pages 2154–2159. IEEE, 2009.

[58] Vijay Rajanna, Folami Alamudun, Daniel Goldberg, and Tracy Hammond. Let me relax: Toward automated sedentary state recognition and ubiquitous mental wellness solutions. In *Proceedings of the 5th EAI International Conference on Wireless Mobile Communication and Healthcare*, MOBIHEALTH'15, pages 28–33, ICST, Brussels, Belgium, Belgium, 2015. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

[59] Kristof Van Laerhoven and Ozan Cakmakci. What shall we teach our pants? In *Wearable Computers, The Fourth International Symposium on*, pages 77–83. IEEE, 2000.

[60] Jonathan Lester, Tanzeem Choudhury, Nicky Kern, Gaetano Borriello, and Blake Hannaford. A hybrid discriminative/generative approach for modeling human activities. In *International Joint Conference on Artificial Intelligence*, volume 5, pages 766–772. Citeseer, 2005.

[61] Folami Alamudun, Hong-Jun Yoon, Kathleen B Hudson, Garnetta Morin-Ducote, Tracy Hammond, and Georgia D Tourassi. Fractal analysis of visual search activity for mass detection during mammographic screening. *Medical Physics*, 2017.

[62] Folami Alamudun, Hong-Jun Yoon, Tracy Hammond, Kathy Hudson, Garnetta Morin-Ducote, and Georgia Tourassi. Shapelet analysis of pupil dilation for modeling visuo-cognitive behavior in screening mammography. In *Proc. SPIE*,

volume 9787 of *SPIE*, pages 97870M–97870M–13, 2016.

[63] Purnendu Kaul, Vijay Rajanna, and Tracy Hammond. Exploring users' perceived activities in a sketch-based intelligent tutoring system through eye movement data. In *ACM Symposium on Applied Perception (SAP '16)*, SAP, page 1, 2016.

[64] Óscar D Lara, Alfredo J Pérez, Miguel A Labrador, and José D Posada. Centinela: A human activity recognition system based on acceleration and vital sign data. *Pervasive and Mobile Computing*, 8(5):717–729, 2012.

[65] Adil Mehmood Khan, Young-Koo Lee, Sungyoung Y Lee, and Tae-Seong Kim. A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer. *IEEE Transactions on Information Technology in Biomedicine*, 14(5):1166–1172, 2010.

[66] Hristijan Gjoreski, Simon Kozina, Mitja Luštrek, and Matjaž Gams. Using multiple contexts to distinguish standing from sitting with a single accelerometer. In *Proceedings of the Twenty-first European Conference on Artificial Intelligence*, pages 1015–1016. IOS Press, 2014.

[67] Martin Berchtold, Matthias Budde, Dawud Gordon, Hedda R Schmidtke, and Michael Beigl. Actiserv: Activity recognition service for mobile phones. In *International Symposium on Wearable Computers (ISWC) 2010*, pages 1–8. IEEE, 2010.

[68] Muhammad Shoaib, Stephan Bosch, Ozlem Durmaz Incel, Hans Scholten, and Paul JM Havinga. A survey of online activity recognition using mobile phones. *Sensors*, 15(1):2059–2085, 2015.

[69] Alfredo J Perez, Miguel A Labrador, and Sean J Barbeau. G-sense: a scalable architecture for global sensing and monitoring. *IEEE Network*, 24(4):57–64, 2010.

[70] Stephen S Intille, Jonathan Lester, James F Sallis, and Glen Duncan. New horizons in sensor development. *Medicine and Science in Sports and Exercise*, 44(1 Suppl

1):S24, 2012.

[71] Muhammad Fahim, Iram Fatima, Sungyoung Lee, and Young-Tack Park. Efm: evolutionary fuzzy model for dynamic activities recognition using a smartphone accelerometer. *Applied Intelligence*, 39(3):475–488, 2013.

[72] Tomas Brezmes, Juan-Luis Gorricho, and Josep Cotrina. Activity recognition from accelerometer data on a mobile phone. In *International Work-Conference on Artificial Neural Networks*, pages 796–799. Springer, 2009.

[73] Jennifer R Kwapisz, Gary M Weiss, and Samuel A Moore. Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter*, 12(2):74–82, 2011.

[74] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge L Reyes-Ortiz. Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine. In *Ambient Assisted Living and Home Care*, pages 216–223. Springer, 2012.

[75] Sasank Reddy, Min Mun, Jeff Burke, Deborah Estrin, Mark Hansen, and Mani Srivastava. Using mobile phones to determine transportation modes. *ACM Transactions on Sensor Networks (TOSN)*, 6(2):13, 2010.

[76] Joey Bartley, Jonathon Forsyth, Prachi Pendse, Da Xin, Garrett Brown, Paul Hagseth, Ashish Agrawal, Daniel W Goldberg, and Tracy Hammond. World of workout: a contextual mobile rpg to encourage long term fitness. In *Proceedings of the Second ACM SIGSPATIAL International Workshop on the Use of GIS in Public Health*, pages 60–67. ACM, 2013.

[77] T Saponas, Jonathan Lester, Jon Froehlich, James Fogarty, and James Landay. ilearn on the iphone: Real-time human activity classification on commodity mobile phones. *University of Washington CSE Tech Report UW-CSE-08-04-02*, 2008, 2008.

[78] Edison Thomaz, Irfan Essa, and Gregory D Abowd. A practical approach for recognizing eating moments with wrist-mounted inertial sensing. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 1029–1040. ACM, 2015.

[79] Nishkam Ravi, Nikhil Dandekar, Preetham Mysore, and Michael L Littman. Activity recognition from accelerometer data. In *AAAI*, volume 5, pages 1541–1546, 2005.

[80] Tzu-Ping Kao, Che-Wei Lin, and Jeen-Shing Wang. Development of a portable activity detector for daily activity recognition. In *Industrial Electronics, 2009. ISIE 2009. IEEE International Symposium on*, pages 115–120. IEEE, 2009.

[81] Daniele Riboni and Claudio Bettini. Cosar: hybrid reasoning for context-aware activity recognition. *Personal and Ubiquitous Computing*, 15(3):271–289, 2011.

[82] Yen-Ping Chen, Jhun-Ying Yang, Shun-Nan Liou, Gwo-Yun Lee, and Jeen-Shing Wang. Online classifier construction algorithm for human activity detection using a tri-axial accelerometer. *Applied Mathematics and Computation*, 205(2):849–860, 2008.

[83] Oliver Amft, Martin Kusserow, and Gerhard Tröster. Probabilistic parsing of dietary activity events. In *4th International Workshop on Wearable and Implantable Body Sensor Networks (BSN 2007)*, pages 242–247. Springer, 2007.

[84] Stefan Dernbach, Barnan Das, Narayanan C Krishnan, Brian L Thomas, and Diane J Cook. Simple and complex activity recognition through smart phones. In *Intelligent Environments (IE), 2012 8th International Conference on*, pages 214–221. IEEE, 2012.

[85] John Paul Varkey, Dario Pompili, and Theodore A Walls. Human motion recognition using a wireless sensor-based wearable system. *Personal and Ubiquitous Computing*, 16(7):897–910, 2012.

[86] Yu-Jin Hong, Ig-Jae Kim, Sang Chul Ahn, and Hyoung-Gon Kim. Mobile health monitoring system based on activity recognition using accelerometer. *Simulation Modelling Practice and Theory*, 18(4):446–455, 2010.

[87] Andreas Bulling, Ulf Blanke, and Bernt Schiele. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR)*, 46(3):33, 2014.

[88] Davide Figo, Pedro C Diniz, Diogo R Ferreira, and João MP Cardoso. Preprocessing techniques for context recognition from accelerometer data. *Personal and Ubiquitous Computing*, 14(7):645–662, 2010.

[89] American Dental Association. Brushing your teeth. Dec 2016. http://www.mouthhealthy.org/en/az-topics/b/brushing-your-teeth.

[90] T Attin and E Hornecker. Tooth brushing and oral health: how frequently and when should tooth brushing be performed? *Oral Health & Preventive Dentistry*, 3(3), 2005.

[91] Tâm Huynh and Bernt Schiele. Analyzing features for activity recognition. In *Proceedings of the 2005 Joint Conference on Smart Objects and Ambient Intelligence: Innovative Context-Aware Services: Usages and Technologies*, pages 159–163. ACM, 2005.

[92] Pebble Inc. Pebble. May 2016. https://www.pebble.com.

[93] Google. Android. May 2016. https://www.android.com.

[94] Ling Bao. *Physical activity recognition from acceleration data under semi-naturalistic conditions*. PhD thesis, Massachusetts Institute of Technology, 2003.

[95] Stephen S Intille, Ling Bao, Emmanuel Munguia Tapia, and John Rondoni. Acquiring in situ training data for context-aware ubiquitous computing applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1–8. ACM, 2004.

[96] Oscar D Lara and Miguel A Labrador. A survey on human activity recognition using wearable sensors. *Communications Surveys & Tutorials, IEEE*, 15(3):1192–1209, 2013.

[97] Joyce Ho. *Interruptions: using activity transitions to trigger proactive messages*. PhD thesis, Massachusetts Institute of Technology, 2004.

[98] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The weka data mining software: an update. *ACM SIGKDD Explorations Newsletter*, 11(1):10–18, 2009.

[99] New Zealand Hamilton. Correlation-based feature subset selection for machine learning. *New Zealand*, 1998.

[100] GÃŠrard Biau. Analysis of a random forests model. *Journal of Machine Learning Research*, 13(Apr):1063–1095, 2012.

[101] Mu-Chun Su, Woung-Fei Jean, and Hsiao-Te Chang. A static hand gesture recognition system using a composite neural network. In *Fuzzy Systems, 1996., Proceedings of the Fifth IEEE International Conference on*, volume 2, pages 786–792. IEEE, 1996.

[102] Tamer Rabie and Valerie Curtis. Handwashing and risk of respiratory infections: a quantitative systematic review. *Tropical Medicine & International Health*, 11(3):258–267, 2006.

[103] Regina I Ejemot-Nwadiaro, John E Ehiri, Martin M Meremikwu, and Julia A Critchley. Hand washing for preventing diarrhoea. *The Cochrane Library*, 2008.

[104] Centers for Disease Control and Prevention. Handwashing: Clean hands save lives, when & how to wash your hands. *http://www.cdc.gov/handwashing/when-how-handwashing.html*, 2015.

[105] Marie T Brown and Jennifer K Bussell. Medication adherence: Who cares? In *Mayo Clinic Proceedings*, volume 86, pages 304–314. Elsevier, 2011.

[106] Eduardo Sabaté. *Adherence to Long-Term Therapies: Evidence for Action*. World Health Organization, 2003.

[107] Lars Osterberg and Terrence Blaschke. Adherence to medication. *New England Journal of Medicine*, 353(5):487–497, 2005.

[108] David H Thom, Mary N Haan, and Stephen K Van den Eeden. Medically recognized urinary incontinence and risks of hospitalization, nursing home admission and mortality. *Age and Ageing*, 26(5):367–374, 1997.

[109] Manoj Prasad, Murat I. Russell, and Tracy Anne Hammond. A user centric model to design tactile codes with shapes and waveforms. In *Haptics Symposium (HAPTICS), 2014 IEEE*, pages 597–602, Feb 2014.

[110] Manoj Prasad, Murat Russell, and Tracy A. Hammond. Designing vibrotactile codes to communicate verb phrases. *ACM Trans. Multimedia Comput. Commun. Appl.*, 11(1s):11:1–11:21, October 2014.

[111] Manoj Prasad, Paul Taele, Daniel Goldberg, and Tracy A. Hammond. Haptimoto: Turn-by-turn haptic route guidance interface for motorcyclists. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, pages 3597–3606, New York, NY, USA, 2014. ACM.

[112] Manoj Prasad, Paul Taele, Ayobami Olubeko, and Tracy Hammond. Haptigo: A navigational tap on the shoulder. In *Haptics Symposium (HAPTICS), 2014 IEEE*, pages 339–345, Feb 2014.