RECOGNIZING CARDIOVASCULAR DISEASE PATTERNS WITH MACHINE LEARNING USING NHANES ACCELEROMETER DETERMINED PHYSICAL ACTIVITY DATA

BY

ELENA BOIARSKAIA

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Kinesiology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Doctoral Committee:

Professor Weimo Zhu, Chair
Professor David Buchner
Professor Feng Liang
Professor Kenneth Wilund

ABSTRACT

The relationship between physical activity (PA) and cardiovascular disease (CVD) is well established; however, questions about the appropriate dose of PA to reduce CVD risk still remain (Blair, LaMonte, & Nichaman, 2004; Pate et al., 1995). The optimal dose and the effects of intensity, duration, and frequency of PA are not fully understood (Haskell et al., 2007). This study connects objectively measured PA with a cross-sectional measure of CVD risk for an in-depth analysis of PA patterns that contribute to higher risk of CVD. Specifically, this study applied machine learning algorithms to NHANES accelerometer data from the 2003-2006 cohorts with the Reynolds cardiovascular risk score as the outcome.

Using accelerometer data as a proxy for the Reynold's risk score to study cardiovascular disease risk allows the use of cross-sectional data when the longitudinal outcome is not known. A major benefit of using accelerometers to objectively measure of PA is that the data is easy and inexpensive to obtain. Furthermore, most locomotive activities are measured with a high degree of accuracy. Accelerometers can gather highly detailed information about an individual's PA pattern over extended periods of time. This produces a large amount of data that requires specialized techniques to analyze. The analysis for this study was conducted using a variety of machine learning techniques to identify individual patterns in the data and evaluate what contributes most to high CVD risk.

Comparison of machine learning algorithms shows that all classifiers perform well when given appropriate features. Using predefined intensity thresholds to compute average time spent in a PA category yielded good classification results in identifying study participants at high and low risk for CVD (Troiano et al., 2008). Adding PA pattern-related features to the model did not appear to improve classification. Features derived using k-means and the Hidden Markov Model

(HMM) performed on the level of using predefined intensity thresholds, indicating that data driven methods may be used for feature extraction without relying on prior knowledge of the data.

In general, the lasso regression, support vector machines (SVM) and random forest (RF) classifiers all performed well on large sets of data-driven features, achieving greater than 82% classification accuracy when time spent in PA intensity categories was combined with k-means and HMM-derived inputs. Neural networks performed well on smaller uncorrelated feature sets, and decision trees produced consistent results with the most transparency and interpretability.

With respect to physical activity recommendations, the findings indicate that gender and time spent in lifestyle minutes (760-2019 intensity counts) play a key role in classifying CVD risk. Thus, a greater emphasis on gender specific recommendations focusing on lifestyle minutes in addition to moderate and vigorous activity may be necessary. Furthermore, time spent in the activity categories, not how PA is spread throughout the day and week appear to be most important for classification of CVD risk.

Dedicated to the memory of my grandfather, Eldgor Giterman

AKNOWLEDGEMENTS

I would like to thank my committee members: Dr. David Buchner, Dr. Feng Liang and Dr. Kenneth Wilund. Thank you for broadening my horizons, sharing your expertise, and providing guidance and advice on this project. A special thank you to my adviser, Dr. Weimo Zhu. I have learned so much from you over the years. Thank you for inspiring me to do my best work, for shaping me as a researcher, and for mentoring me as a professional in the field.

Thank you to my friends and colleagues: John Kalohn, David Scrams, Ray Yan, Farai Munyari Madison, and my editor, Elizabeth Deanna Morris Lakes. Your words of encouragement and support helped me see this project through to the end.

My family: my parents Alla Boiarskaia and Alexandre Boiarski and my grandmother, Albina Boiarskaia. Without you, this journey would have been impossible. Thank you for your enthusiasm and support. Finally, I am deeply grateful to my husband, Bradley Paul Uban. Thank you for your patience and for always being there for me.

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION

The connection between physical activity (PA) and health outcomes is a central theme in kinesiology and public health research. This relationship is well documented with studies showing that physical inactivity leads to increased risk of all-cause mortality, is associated with coronary heart disease, osteoporosis, diabetes and some cancers (Pate et al., 1995; *NIH Consensus Development Panel on Physical Activity and Cardiovascular Health.,* 1996). PA offers a wide range of health benefits that include weight control, longevity, reduced risk of various diseases, and improved mental health (Hebebrand & Hinney, 2015). Furthermore, PA has been shown to improve cardiovascular function and reduce the risk of cardiovascular disease, the number one killer of Americans (Go et al., 2014). Given the numerous preventive benefits of engaging in PA, physicians are recommending PA to their patients (Blair et al., 2004).

However, in order to recommend PA as a way of improving cardiovascular health and preventing risk of chronic diseases, appropriate dosage for different populations must be established. Currently, to enjoy the benefits of PA, the US Federal Physical Activity Guidelines recommend at least 150 minutes of moderate intensity or 75 minutes of vigorous intensity PA, along with biweekly strength training for adults (United States Department of Health and Human Services., 2008). Nevertheless, unanswered questions remain about the appropriate length of continuous PA bouts, optimal intensity, and the effects of light PA (Haskell et al., 2007). The dose-response relationship between PA and various health outcomes is less understood and requires further investigation (Kesaniemi et al., 2001).

Measuring Cardiovascular Health Status

A particular interest in studying the benefits of PA is disease prevention. Examining the impact of PA on cardiovascular health status as an outcome is a challenge since there is no single

criterion measure of cardiovascular health. Longitudinal studies using mortality rates due to cardiovascular disease as outcomes have established a number of modifiable and non-modifiable risk factors that may lead to cardiovascular disease (Blair et al., 1989; Pate et al., 1995). While age, gender, and family history cannot be changed, regular exercise has been shown to have the potential to reduce a number of modifiable risk factors that include high blood pressure, poor glucose tolerance, high cholesterol, and obesity (D'Agostino et al., 2008).

Therefore, cross-sectional and intervention studies measuring cardiovascular health status as an outcome may focus on specific risk factors such as blood pressure and cholesterol levels. However, to study a more general representation of cardiovascular health status, it may be of interest to account for several known risk factors simultaneously by using a composite risk score. Several cardiovascular risk scores have been developed beginning with the Framingham Heart Study 10 year cardiovascular risk score (D'Agostino et al., 2008). These risk scores aim to represent the percent risk that an individual will have a cardiovascular event in the next 10 years and offer a possible solution for representing cardiovascular health status in cross-sectional data and intervention studies.

Measuring Physical Activity

Accurate and objective measures of PA are also imperative for establishing recommendations for cardiovascular health improvement (Strath et al., 2013). However, these measures may be difficult to obtain. PA takes on various forms and can include exercise with the intention of improving fitness, playing sports, engaging in leisure activity, and performing household chores. The choice of PA varies greatly amongst individuals and may be difficult to monitor. Multiple methods including questionnaires and direct observation exist to estimate PA;

however there is a tradeoff between accuracy and cost effectiveness (Ainsworth & Coleman, 2006).

Recent technological advances may offer a solution to some of these challenges. Electronic devices such as accelerometers, heart rate monitors, and GPS trackers provide insight into when and at what intensity, frequency, and duration the activity was performed at every preset interval. These devices provide a very large amount of detailed data with frequent readings for each individual. Furthermore, accelerometers, heart rate monitors, and GPS trackers have become fairly inexpensive and readily available, making these useful for a wide range of studies (Chen, Janz, Zhu, & Brychta, 2012).

For example, accelerometers have become increasingly popular in PA and health research (Troiano, 2006) . Studies have used accelerometers to establish the connection between PA and breast cancer, cardiometabolic syndrome, kidney disease, insomnia, and depression (Chasens & Yang, 2012; Hawkins et al., 2011; Healy, Matthews, Dunstan, Winkler, & Owen, 2011; Lynch et al., 2011; Vallance et al., 2011). These findings indicate that the use of these electronic devices is appropriate in a variety of settings, offering objective and fairly accurate ways of assessing PA with possible insights into optimal dosage on an individual basis.

At the same time, data collection using electronic devices creates a new set of problems. While a large volume of detailed data is inexpensive and relatively easy to obtain, special analytic techniques are required to derive useful information from the data while retaining its richness. Each individual has a pattern to his or her daily PA based on daily habits, occupation, and health status. Identifying the complex individual's PA patterns derived by electronic devices in conjunction with health outcomes may be very helpful for establishing PA guidelines and

recognizing at risk populations. This type of research requires a way of accurately classifying health outcomes based on the individual's PA patterns.

Machine Learning

Machine learning techniques are useful for analyzing large volumes of data, recognizing patterns, and classifying outcomes with existing applications in a broad range of topics. Machine learning is a subset of artificial intelligence that utilizes a collection of algorithms that help computers learn from the data, predicting a set of outcomes or recognizing patterns, where the prediction gets better with experience. These techniques allow for individualized analysis, and while it takes more time, it also yields more accurate results. Furthermore, advances in computational technology help make this process fast.

Applications of machine learning algorithms include speech and handwriting recognition, DNA sequencing, stock market analysis and robotics (Hastie, Tibshirani, & Friedman, 2009). Machine learning and data mining techniques have been used in various disciplines to find patterns in accelerometer data but have yet to be used to their full potential in the field of PA and health research. So far, the main objective of accelerometer data analyzed using machine learning has been to recognize the individual's mode of PA (Ermes, Pärkka, Mantyjarvi, & Korhonen, 2008; Freedson, Lyden, Kozey-Keadle, & Staudenmayer, 2011; Mannini & Sabatini, 2010). These studies are typically conducted in a controlled environment, such as a laboratory, where the study participants are asked to perform a choreographed routine while wearing an accelerometer. Prediction of PA type is used to estimate energy expenditure. The goal is to accurately predict energy expenditure from accelerometer data for applications in heath related research or in engineering to provide feedback to users tracking their PA using electronic devices (Bouten, Koekkoek, Verduin, Kodde, & Janssen, 1997; Long, Yin, & Aarts,

2009; Rothney, Neumann, Béziat, & Chen, 2007; Rothney, Schaefer, Neumann, Choi, & Chen, 2008). Few studies have attempted to classify health outcomes directly from accelerometer data.

NHANES Accelerometer Data

Health monitoring and research into chronic disease prevention that will further inform public health policy, prevention strategies, and medical treatment options is of national importance according to the Center of Disease Control's (CDC) mission. The National Health and Nutrition Examination Survey (NHANES) collected for the years 2003–2004 and 2005–2006 by the CDC offers objective measures of PA of a representative sample of the U.S. population using ActiGraph accelerometers. Step counts are available for each minute of the week the participants wore the PA monitors. The NHANES data also provides an extensive number of health related outcome variables and is ideal for establishing a connection between health status and objectively measured PA.

The accelerometers used in the NHANES are an example of how electronic devices are able to provide highly detailed data for each individual and thus have tremendous potential in numerous research settings. Though the NHANES data has many advantages, the large volume of readings provided by the accelerometers and individual variation of the participants is also problematic as discussed above. So far, studies using the NHANES accelerometer data have shown connections between objectively derived PA and various health outcomes measured by the NHANES (Healy et al., 2011; Luke, Dugas, Durazo-Arvizu, Cao, & Cooper, 2011). However, these studies compress the detailed accelerometer data into a few data points using standardized thresholds (Atienza et al., 2011; Camhi, Sisson, Johnson, Katzmarzyk, & Tudor-Locke, 2011b; Evenson & Wen, 2011). Machine learning techniques have yet to be used to classify individual health status via PA patterns in this data.

Statement of Problem

Electronic devices such as accelerometers provide large volumes of readings that offer rich information about PA patterns. However, this data requires special pattern recognition, data reduction and classification techniques to arrive at meaningful conclusions about the relationship between PA and health outcomes and consequently provide PA recommendations.

Solutions

Utilize machine learning algorithms to analyze PA data collected using electronic devices. These techniques may prove to be useful not only for categorizing the type of PA but also for classifying health outcomes based on PA patterns.

Purpose

This study is dedicated to selecting features and applying machine learning algorithms to find associations and patterns in accelerometer-derived free-living daily PA that can help predict the participant's cardiovascular health status based on cardiovascular risk scores using several machine learning techniques using the 2003–2004 and 2005–2006 NHANES data. More specifically, this study will attempt to:

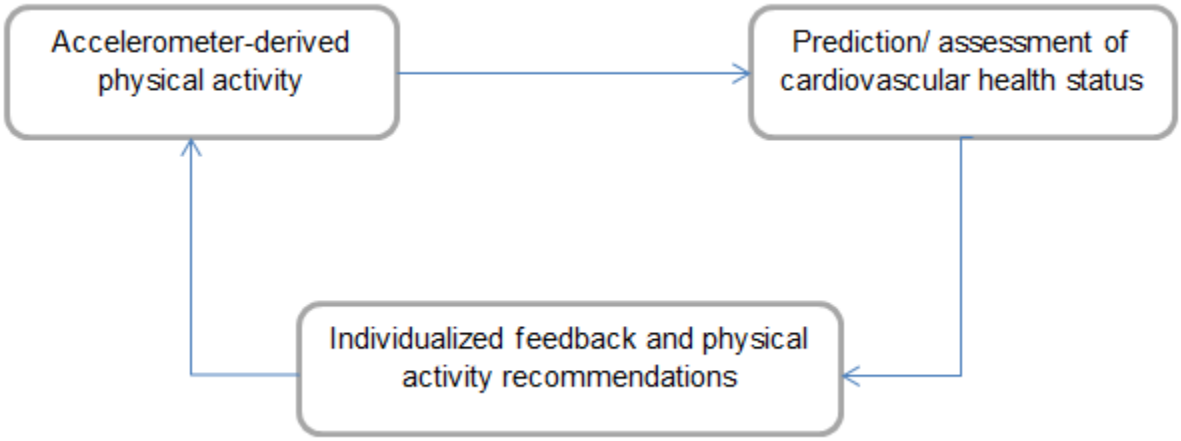
- connect intensity readings from accelerometer data recoded over the course of a week to a cardiovascular risk score;
- extract, compare and select appropriate features from accelerometer data to use with machine learning algorithms;
- predict cardiovascular health status of an individual based on accelerometer-derived PA using machine learning approaches;
- compare machine learning algorithms for classification of accelerometer-derived PA by cardiovascular health status.

Significance of Study

So far, machine learning techniques have been applied to accelerometer-derived PA data that were collected in controlled, clinical conditions with a main focus on classifying activity type, not predicting health status. The NHANES data offers accelerometer-derived, free-living PA collected over the course of one week. Previous analysis of this data was either descriptive in nature or established connections between various health markers and objectively measured PA by condensing a week's worth of step counts recorded every minute into a single data point.

The analysis of the rich information offered by the NAHNES PA monitor data has tremendous potential in providing valuable feedback on an individual basis. As Figure 1 shows, machine learning algorithms could help classify cardiovascular health status based on accelerometer-derived PA. Subsequently, these findings will help develop individualized interventions and prevention programs. Additionally, these methods may be successful in detecting unknown and future health problems, allowing the development of PA recommendations based on health status. Furthermore, the NHANES 2011–2012 study will have PA data using waterproof accelerometers for seven consecutive days without breaks for swimming, bathing or sleeping. Innovative methods to analyze this data are needed and will be proposed in this study.

Figure 1. The potential of using objectively measured PA for prediction of cardiovascular health status, which will in turn lead to individualized recommendation for PA.



CHAPTER 2 LITERATURE REVIEW

This chapter reviews potential measures of PA and cardiovascular health status, and how accelerometer data has been used to further research in PA and health. The challenges and possible solutions associated with analyzing large volumes of accelerometer data are discussed. Specifically, studies utilizing the NHANES accelerometer data are examined in detail, and machine learning techniques for recognizing PA patterns are introduced.

Prevalence of Cardiovascular Disease

Although death rates due to heart disease have declined from 2000 to 2010 by 31%, cardiovascular disease is still the leading cause of death in the U.S., killing nearly one in three Americans (Go et al., 2014). Increased public awareness, better preventive measures, improved treatment options, and quicker response times have all contributed to the decrease in cardiovascular disease death rate; however, known risk factors remain poorly controlled. Although fewer American adults smoke, few adults meet the recommendations for body composition, PA, cholesterol, and glucose levels (Go et al., 2014). At the same time, the financial burden of cardiovascular disease is extremely high, estimated to be $445 billion in 2010 in the U.S. (Heidenreich et al., 2011).

Forms of Cardiovascular Disease

Cardiovascular disease refers to any disease that affects the cardiovascular system. Although heart disease manifests itself in many forms, coronary heart disease kills the greatest number of Americans by far, causing 1 in 6 deaths (Go et al., 2014). The next leading killer is stroke, causing 1 in 19 deaths in the U.S., though death from stroke decreased by 35.8% from 2000 to 2010 (Go et al., 2014). Hypertension affects 29.6% of Americans, while 48.0% are successfully controlling the disease (Gillespie & Hurvitz, 2013). Other types of heart disease

affecting the U.S. population are heart failure, arterial disease, and congenital heart disease (Kenney, Wilmore, & Costill, 2012). The prevalent types of heart disease with controllable risk factors are discussed below.

Coronary heart disease. As described by Kenney et al. (2012), coronary heart disease is characterized by narrowing of the arteries, a process known as atherosclerosis. Damage to the innermost lining of the arterial wall leads to an inflammatory response, with platelets, smooth muscle cells and connective tissue attaching to the affected site and beginning the formation of plaque. Low-density lipoprotein (LDL) cholesterol attaches itself to the area, contributing to the size of the plaque. Overtime, the plaque grows and leads to restricted blood flow, or ischemia, to the heart. The plaque may become unstable and rupture, forming a blood clot or thrombus and eventually blocking the artery. Myocardial infarction, or a heart attack, occurs when the heart muscle is deprived of oxygen due to the impeded blood flow (p. 524).

Stroke. Kenney et al. (2012) characterize stroke is the disease of the cerebral arteries. Ischemic **s**troke occurs when blood flow to the brain is restricted. This type of stroke may be the result of thrombosis, where an artery in the brain is obstructed due to atherosclerosis, or an embolism, where a thrombus or fat globule have become dislodged and traveled to a cerebral artery, causing blockage. Stroke may also result from a hemorrhage in the brain due to a ruptured artery. An artery may rupture due to atherosclerotic damage to its walls. Both types of stroke lead to brain tissue damage (p. 525).

Hypertension. As explained by Kenney et al. (2012), blood pressure is the pressure exerted on the arterial walls by the circulation of blood. High blood pressure, or hypertension, means the blood is circulating with greater pressure on the arterial walls that what is considered normal for the individual's size and age. Hypertension signals that the heart has to pump harder

to circulate the same amount of blood due to the increased resistance in the arteries. Overtime, the heart muscle becomes strained and enlarged, and the arteries become less elastic and damaged. This mechanism eventually leads to atherosclerosis and other cardiovascular diseases (p.525).

Heart failure. According to Kenney et al. (2012), when the heart muscle is too weak to adequately circulate the blood due to hypertension and atherosclerosis, the condition is referred to as heart failure. The disease is characterized by reduced force of contraction, an enlarged heart muscle, and increased heart rate. Heart failure results in poor circulation and fluid accumulation, or edema. When the lungs are affected, heart failure causes shortness of breath and exercise intolerance (p.526).

Risk Factors of Cardiovascular Disease

Although cardiovascular disease usually presents itself in late adulthood, changes that signal the beginning of atherosclerosis have been shown to occur in the early stages of life with some infants exhibiting early formation of plaque in the aorta (Kannel & Dawber, 1972). Of the 300 autopsied American soldiers with average age of 22.1 killed in the Korean war, 77.3% exhibited some signs of atherosclerosis in the aorta (Enos, Holmes, & Beyer, 1953). The progress of the decease is determined by genetics, lifestyle, and environmental factors and is largely asymptomatic until later in life. While some factors like genetics, age, and gender could not be changed, risk of developing heart disease may be significantly reduced by modifying lifestyle factors (Go et al., 2014).

Obesity and overweight. Obesity and overweight describe the degree of excess accumulation of body fat that poses a risk to one's health. Due to the relative difficulty and cost of measuring body fat directly, to classify individuals as overweight or obese on a population

level, the body mass index (BMI), popularized by Ancel Keys in 1972, is used. BMI is an index of the relationship between the weight and height of an individual, calculated by dividing the weight in kilograms by the height in meters squared (Keys, Fidanza, Karvonen, Kimura, & Taylor, 1972). Overweight is defined by a BMI greater than 25 but less than 30, and obesity is defined by a BMI of 30 and greater (Hebebrand & Hinney, 2015).

Obesity affects more than a third of the U.S. population with levels rising from 30.3% in 2000 to 35.9% in 2010 according to the CDC (Ogden, Lamb, Carroll, & Flegal, 2010). The adverse health effects associated with excess weight are well documented and mortality rates increase exponentially with increases in BMI (Bray, 1985). It has been shown that excess body fat leads to impaired bodily functions and has been linked to various forms of cardiovascular disease, diabetes, cancer, and respiratory problems. Specifically, obesity leads to decreased oxygenation of the blood, causing hypertension, enlargement of the heart and thrombosis (Kenney et al., 2012). To lower the risk of cardiovascular disease it is recommended that BMI be kept between 18 and 25 through weight control (Hebebrand & Hinney, 2015).

Inflammation. Inflammation is a natural immune response of the body to injury. Characterized by increase blood flow that causes redness, heat, swelling and pain at the site of the injury, inflammation prevents the spread of infection, removes damaged tissue, and assists healing (Kenney et al., 2012). Infections or tissue injuries such as a cut on the hand or a sprained ankle are all types of acute inflammation that is elicited by an external stimulus and is necessary to promote the healing process. However, inflammation may also be chronic in response to stress, smoking, poor diet, and environmental factors. Chronic inflammation contributes to heart disease by hardening the arterial wall and increasing its permeability, leading to the formation of plaque (Danesh, 2000; Kaptoge et al., 2010; Medzhitov, 2008).

C-reactive protein (CRP) is released by the liver in response to inflammation and is considered to be a relatively new indicator of cardiovascular disease risk (Danesh et al., 2007). CRP levels are measured by a blood test where < 1.0 mg/L is considered normal. CRP levels between 1.0 and 3.0 mg/L are indicators of chronic inflammation, while 3.0 mg/L and greater indicates a high risk of heart attack (Go et al., 2014; G. L. Myers et al., 2004).

Abnormal lipoprotein levels. As described by Kenney et al. (2012), cholesterol is a type of lipid molecule, a sterol that carries out a number of vital functions in the body. Cholesterol is produced by the liver but may also be obtained from foods from animal sources. Cholesterol is essential for building cell membranes and determining cell membrane permeability. Furthermore, it helps in the production of various hormones including sex hormones and those released by the adrenal gland. Additionally, cholesterol is used for the producing bile, insulating nerve fibers, converting sunlight exposure to vitamin D, and metabolizing fat soluble vitamins (p. 530).

While cholesterol is essential for numerous mechanisms in the body, it also plays a role in atherogenesis, or plaque formation (D'Agostino et al., 2008; National Cholesterol Education Program (NCEP) Expert Panel, 2002). Because cholesterol is a fat-like substance, it is transported through the water-based bloodstream by lipoproteins. While low-density lipoproteins (LDL) help transport cholesterol molecules to the cells, high-density lipoproteins (HDL) transport cholesterol back to the liver. When the blood vessels are inflamed, LDL containing cholesterol molecules may permeate and get deposited in the arterial wall thus contributing to plaque formation and further inflammation (Kenney et al., 2012). The optimal ranges for total cholesterol, LDL, and HDL are described in Table 1below.

Table 1. Optimal levels of cholesterol from ATP III (National Cholesterol Education Program (NCEP) Expert Panel, 2002)

Total cholesterol	**Category**
Below 200 mg/dL	Desirable
200-239 mg/dL	Borderline high
240 mg/dL and above	High
LDL cholesterol	**Category**
Below 70 mg/dL	Ideal for people at very high risk of heart disease
Below 100 mg/dL	Ideal for people at risk of heart disease
100-129 mg/dL	Near ideal
130-159 mg/dL	Borderline high
160-189 mg/dL	High
190 mg/dL and above	Very high
HDL cholesterol	**Category**
Below 40 mg/dL (men)	Poor
Below 50 mg/dL (women)	
40-49 mg/dL (men)	Better
50-59 mg/dL (women)	
60 mg/dL and above	Best

Hypertension. Hypertension is a form of cardiovascular disease when the arterial blood pressure is chronically high, placing additional stress on the arterial walls and the heart (Kenney et al., 2012). Therefore, not only is high blood pressure a form of cardiovascular disease, but it is also an important risk factor for coronary heart disease, stroke, and heart failure.

According to Kenney et al. (2012), blood pressure is characterized by systolic and diastolic blood pressure. Systolic blood pressure refers to the maximum force that occurs when blood is driven out of the heart due to the contraction of the left ventricle. Diastolic blood pressure occurs when the heart is filling with blood, and the ventricle is relaxing. Although blood pressure depends on body size, for the average adult, the normal ranges of systolic blood pressure are shown in Table 2 below. Individuals with blood pressure above these ranges are considered at risk and treatments for lowering and controlling blood pressure are recommended (p. 525).

Table 2. Optimal blood pressure levels (Chobanian et al., 2003)

Blood Pressure Category	Systolic mmHg		Diastolic mmHg
Normal	less than 120	and	less than 80
Prehypertension	120 – 139	or	80 – 89
High Blood Pressure Stage 1	140 – 159	or	90 – 99
High Blood Pressure Stage 2	160 or higher	or	100 or higher
Hypertensive Crisis	higher than 180	or	higher than 110

Diabetes and insulin resistance. Diabetes mellitus (DM) is a health condition characterized by elevated blood sugar due to either a lack of insulin production (Type I DM) or insulin resistance (Type II DM). Insulin is a hormone released by the pancreas that helps regulate carbohydrate metabolism. While a disease in itself, Type II diabetes is a major cardiovascular disease risk and is also strongly associated with obesity (Bray, 1985). High levels of sugar or glucose in the blood lead to damage of the arterial walls contributing to atherogenesis (Kenney et al., 2012).

Blood glucose and glycated hemoglobin (HbA1c) levels are used to diagnose diabetes. The glucose tolerance test measures the baseline blood sugar level after an overnight fast, and the blood sugar level 2 hours after drinking a glucose solution. Glycated hemoglobin (HbA1c) measures the plasma glucose levels over the past three months (the half-life of red blood cells) (Q. Yang et al., 2012). The optimal ranges of these indicators of diabetes are presented in the Table 3 below.

Table 3. Optimal Blood glucose and glycated hemoglobin (HbA1c) levels (American Diabetes Association, 2010)

Condition	2 hour glucose	Fasting glucose	HbA_{1c}	
Unit	mmol/l(mg/dl)	mmol/l(mg/dl)	mmol/mol	DCCT %
Normal	<7.8 (<140)	<6.1 (<110)	<42	<6.0
Impaired fasting glycaemia	<7.8 (<140)	≥6.1 (≥110) & <7.0 (<126)	42-46	6.0–6.4
Impaired glucose tolerance	≥7.8 (≥140)	<7.0 (<126)	42-46	6.0–6.4
Diabetes mellitus	≥11.1 (≥200)	≥7.0 (≥126)	≥48	≥6.5

Smoking. Smoking has been shown to increase the risk of heart disease nearly twofold (Lloyd-Jones, Adams, & Brown, 2010). Smoking increases the inflammatory response of the

body, contributing to the calcification of the arterial walls and atherosclerosis (Ambrose & Barua, 2004). While smoking history may be assessed via a questionnaire, recent exposure to nicotine may be measured by blood cotinine levels (Caraballo, R. S., Giovino, G. A., Pechacek, T. F., & Mowery, 2001). Cotinine is a byproduct of metabolizing nicotine and may be detected days after nicotine exposure. Thus, even exposure to second hand smoke may be detected by a cotinine test. Levels of ≥ 10 ng/mL are indicative of some smoking or exposure, while active smokers will have cotinine levels of 100 ng/mL or greater (Wall & Johnson, 1988).

Non-modifiable risk factors. Certain indicators of cardiovascular disease risk cannot be changed. Age and gender are strongly associated with cardiovascular disease risk where the majority of heart disease related deaths occur in people over 65 and males are at a higher risk of a cardiovascular event (Wilson et al., 1998). Having a close relative who had cardiovascular disease before the age of 60 increases one's risk of heart disease as well (D'Agostino et al., 2008). Race has also been shown to be a factor in cardiovascular disease risk with African Americans at a higher risk than white Americans (Hozawa, Folsom, Sharrett, & Chambless, 2007).

Thus, recommendations for cardiovascular disease prevention focus on modifiable risk factors that may be influenced by drugs, diet, and exercise (Lloyd-Jones, Hong, et al., 2010). Figure 2 shows the factors that influence cardiovascular disease risk and which of these may be modified by PA.

Reducing Cardiovascular Disease Risk with Physical Activity

Research shows that PA effects cardiovascular health through a variety of pathways. Numerous studies have demonstrated that engaging in regular exercise has to the potential to reduce cholesterol and inflammation, lower blood pressure, improve glucose tolerance, and

change body composition by increasing bone density and muscle mass while reducing body fat. These benefits, in turn, contribute to the reduction of cardiovascular disease risk (Lloyd-Jones, Adams, et al., 2010).

Epidemiological evidence. Following World War II, the development of penicillin as a mass-produced drug contributed to a reduction in the prevalence of infectious diseases in the U.S. and Europe. Non-communicable, chronic diseases, specifically cardiovascular disease, emerged as the leading cause of death (Mahmood, Levy, Vasan, & Wang, 2014). Several landmark epidemiological studies around this time led to the identification of risk factors of cardiovascular disease, establishing a connection between PA and various health outcomes.

To promote better understanding and prevention of heart disease, President Harry Truman signed the National Heart Act of 1948, thus establishing the National Heart Institute, now known as the National Heart, Lung, and Blood Institute. The Framingham heart study was initiated in 1948 by the Institute as an ongoing longitudinal study of risk factors of cardiovascular disease (Chapman, 1958). This landmark study initially recruited 5,209 participants from Framingham, MA aged 30 to 62 with a follow up every two years. In 1971, children of the original participants and spouses were recruited by the study. The third generation consisting of grandchildren of the original participants was recruited in 2002. To reflect a more diverse population, an Omni cohort was recruited in 1994, and a second cohort in 2003 (Mahmood et al., 2014).

After six years of follow up, an inverse relationship between levels of PA and cardiovascular disease mortality was noted (Kannel, Dawber, Kagan, Revotskie, & Stokes, 1961). Since PA data in the study was collected by a 24-hour history and the study population was generally sedentary, effects of PA were difficult to establish (Kannel & Sorlie, 1979). The

investigators noted that after 24 years of follow up, the participants increased their PA levels and a stronger relationship between PA and reduced cardiovascular disease risk was observed (Kannel et al., 1986).

Research into heart disease in Europe was initiated by Jeremy Morris who conducted one of the first large scale epidemiological studies linking PA and cardiovascular health (Paffenbarger, Blair, & Lee, 2001). Morris led the London Transport workers study comparing the incidence of cardiovascular disease in 31,000 men employed by the London Transport Executive from 1949 to 1952. The study found that a higher proportion of drivers experienced coronary heart disease than conductors. The investigators attributed this difference to the amount of PA (total energy expenditure) performed by the two jobs, with conductors being more active than the sedentary drivers (Morris, Heady, Raffle, Roberts, & Parks, 1953).

To further study the effects of PA on cardiovascular disease, Morris et al. controlled for PA demands in a professional setting by studying British civil servants with sedentary office jobs only. The study compared reported weekend leisure time activity in 16,882 male participants between the ages of 40 and 64. The study revealed that the intensity of the PA, not the total energy expenditure had the strongest effect on cardiovascular disease (Chave, Morris, Moss, & Semmence, 1978; Morris et al., 1973).

Numerous large-scale longitudinal studies followed, confirming the positive health benefits of PA. The Harvard Alumni study that was started in 1962 focused specifically on the effects of PA on cardiovascular disease risk. The data was collected via a simple questionnaire from 16,936 male Harvard graduates, and the researchers found that PA lowered the risk of heart disease even when controlling for the effects of age, BMI, family history and smoking (Paffenbarger, Wing, & Hyde, 1978; Sesso, Paffenbarger, & Lee, 2000). Another ongoing

longitudinal study, the Nurses' Health Study, started in 1976, recruited 121,700 female nurses with the aim of identifying chronic disease risk factors specific to women (Belanger, Hennekens, Rosner, & Speizer, 1978). The study found that PA not only reduces risk of coronary heart disease and stroke but also various types of cancer (Colditz, Manson, & Hankinson, 1997; Holmes, Chen, Feskanich, Kroenke, & Colditz, 2005).

The Cooper Center Longitudinal study is an ongoing and comprehensive assessment of the effects of PA on various health outcomes. Founded in 1970 by Kenneth Cooper, the "father of aerobics" and a strong believer in the importance of physical fitness, the study recruited 10,224 male and 3,120 female participants with an average eight-year follow up (Blair et al., 1989). Similar to the Harvard study and the Nurses' Health Study findings, the researchers found that participants with higher baseline PA levels had lower all-cause mortality, cardiovascular disease and cancer rates even after adjusting for age, smoking status and various heart disease risk factors (Blair et al., 1989).

Exercise training adaptations. While epidemiological studies suggest a relationship between PA and cardiovascular disease risk, intervention studies shed light on the pathways through which PA affects cardiovascular fitness. Studies have shown that a dose-response relationship exists between PA and cardiovascular fitness in children, adults, and the elderly (Haskell et al., 2007; Lloyd-Jones, Hong, et al., 2010). Furthermore, evidence suggests that PA offers benefits to fit and unfit populations and has potential of reversing the effects of aging on the cardiovascular system (Blair et al., 1995). It appears that beginning an exercise program has potential of reducing cardiovascular disease risk even in high-risk populations (Chamnan, Simmons, Sharp, Griffin, & Wareham, 2009; Chobanian et al., 2003; Hawkins et al., 2011; Holmes et al., 2005; Smith, Nolan, Robison, Hudson, & Ness, 2011).

Engaging in PA has been shown to lead to various cardiovascular adaptations such as increased maximal oxygen uptake, maximal cardiac output and stroke volume and a decreased heart rate (Blomqvist & Saltin, 1983). While improving indicators of cardiorespiratory fitness, PA has also been shown to directly improve specific risk factors of cardiovascular disease. Regular PA has been shown to reduce cholesterol, cRP levels, risk of type II diabetes, and obesity (Dunn et al., 1999).

The numerous benefits of PA have been noted in various training protocols. While steady endurance aerobic activity has often been recommended for optimal health, in recent years, high intensity interval training has been shown to produce similar cardiovascular adaptations and health benefits (*NIH Consensus Development Panel on Physical Activity and Cardiovascular Health.*, 1996). Engaging in short bursts of vigorous PA has been shown to improve cardiorespiratory fitness measured by VO2max as much as regular endurance aerobic activity (Tabata et al., 1996). Furthermore, high intensity interval training has been shown to improve cardiovascular disease risk factors to a greater degree than moderate intensity steady state exercise (Wisløff et al., 2007). These findings are in line with the British Civil Servant Study conclusion that the intensity of leisure time PA, not total energy expenditure contributes to a greater reduction in mortality risk (Morris et al., 1973).

Physical activity recommendations. It has been well established that PA carries positive health benefits (Blair et al., 1989), while physical inactivity may lead to numerous health problems that include diabetes and cardiovascular disease (D'Agostino et al., 2008; Healy et al., 2011). Medical professionals are beginning to recognize the importance of recommending PA for overall health improvement and prevention of chronic disease to patients (Blair et al., 2004). Consequently, a main focus of kinesiology and public health research includes establishing a

dose-response relationship between PA and various health outcomes to eventually establish PA recommendations for different populations and health conditions.

Although numerous studies successfully showed positive health effects of PA in various populations, questions in PA and health-related research remain. Establishing recommendations for health improvement and disease risk reduction via PA is a challenge. The current guidelines are based on existing findings from longitudinal data and intervention studies establishing a dose-response relationship between PA and health outcomes. However, the optimal dose and the effects of intensity, duration, and frequency of PA remain unclear (Haskell et al., 2007).

Attempts to establish a set of recommendations for PA were initiated by the American College of Sports Medicine (ACSM) in 1978. Three to five days a week of aerobic exercise at 15–60 minutes per session and an intensity between 50% and 85% of the maximum heart rate were advised. The guidelines were geared toward improving and maintaining cardiorespiratory fitness and were therefore perceived as highly structured. Revised versions of the guidelines were released in 1990 and again in 1998 to emphasize health rather than performance-oriented fitness. The 1998 guidelines were less rigid, suggesting at least 20 minutes of aerobic activity three to five times a week (Blair et al., 2004).

The shift toward prescribing exercise as a way to improve health related outcomes led to more simplified guidelines. The CDC and ACSM released a joint recommendation for the general public to accumulate at least 30 minutes of moderate-intensity PA on most days of the week (Pate et al., 1995). This suggestion was based on epidemiologic studies showing an inverse dose-response relationship between PA and disease risk. To make the guidelines accessible to the general population, the researchers focused on the minimum amount of activity necessary to achieve the greatest health benefit (Blair et al., 2004).

Next, a set of PA guidelines were issued by the U.S. Department of Health and Human Services in 2008. These guidelines target several populations that include children, older adults, pregnant and post-partum women and provide more individualized recommendations. For the general adult population, at least 150 minutes of moderate or 75 minutes of vigorous-intensity aerobic activity spread throughout the week are recommended for important health benefits, however the idea that more activity leads to greater benefits is emphasized. Additionally, the guidelines stress that accumulating 10 minute bouts of sustained moderate to vigorous activity are sufficient to benefit from the activity (United States Department of Health and Human Services., 2008).

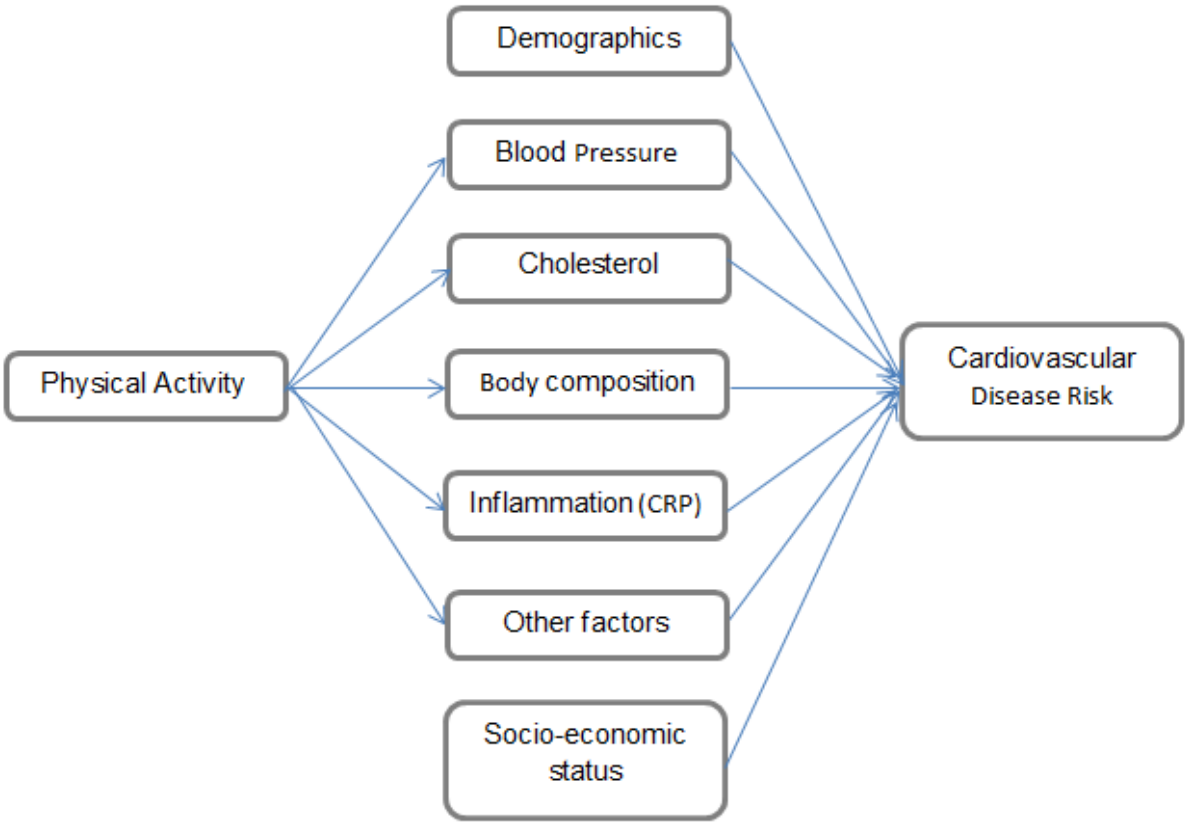
Since PA recommendations are geared toward the general population, focusing on the least amount of exercise for the greatest benefit for most people, individual activity patterns are still overlooked. Further studies to examine different patterns of PA throughout the day and week, the effects of light and highly vigorous intensity activity, and continuous bout length of exercise are needed.

Measuring Cardiovascular Disease Risk

Since numerous mechanisms through which PA impacts cardiovascular health have been identified, summary outcome measures that combine several risk factors have become popular in PA and health research. Cardiometabolic risk scores that combine risk factors for both cardiovascular disease and type II diabetes have been used in cross-sectional and intervention studies as measures of general health outcomes (Camhi et al., 2011b; Holman, Carson, & Janssen, 2011; Sisson et al., 2010). However, for explicit study of cardiovascular health outcomes, a summary score that combines several know risk factors for cardiovascular disease

may provide an interesting solution for representing cardiovascular health outcomes in cross-sectional data.

Figure 2. The relationship between PA and cardiovascular risk as mediated by other variables.



Cardiovascular risk scores. Several cardiovascular risk scores that estimate the risk of having a cardiovascular event in the next 10 years have been developed and are compared below. The goal for this study is to choose the most appropriate risk score to represent cardiovascular health of the U.S. adult population using NHANES data.

Framingham risk scores (FRS) consist of several models with slightly different outcomes. The Coronary Heart Disease (CHD) Risk Score controls for gender, age, smoking status, total cholesterol (TC), HDL cholesterol, systolic blood pressure (SBP), whether the patient is receiving treatment for high blood pressure (BP) and diabetes (Wilson et al., 1998). This model

is applicable for patients 30 to 74 years old and without overt CHD at the baseline examination. The hard CHD (myocardial infarction or coronary death) risk score controls for gender, age, smoking status, TC, HDL, SBP and BP treatment and is appropriate for individuals free of CHD, intermittent claudication and diabetes, 30–79 years of age (National Cholesterol Education Program (NCEP) Expert Panel, 2002). The General Heart Disease Risk Score controls for gender, age, diabetes, smoking, BP, TC, LDL cholesterol and is also applicable to patients 30 to 74 years old and without overt CHD at the baseline examination (D'Agostino et al., 2008). Comparison of the FRS with metabolic syndrome (MetSyn) classification showed that MetSyn is better at predicting diabetes while FRS is more successful at predicting cardiovascular events (Wannamethee, Shaper, Lennon, & Morris, 2005).

The Reynolds Risk Score was developed to improve prediction of CVD risk in women and is based on an initially healthy cohort of 24,558 U.S. women age 45 and over followed for a median of 10.2 years (Ridker, Buring, Rifai, & Cook, 2007). A model for men was later developed based on 10,724 men with a median follow up of 10.8 years (Ridker, Paynter, Rifai, Gaziano, & Cook, 2008). The Reynolds risk score controls for the same variables as FRS with the addition of C-reactive protein (CRP) levels and family history of mother or father having a heart attack under the age of 60. Reynolds performed better than FRS when applied to data form NIH's multiethnic Women's Health Initiative Observational Cohort with the latter model tending to overestimate CVD risk (Cook et al., 2012).

There are many other 10-year cardiovascular risk scores that were developed following the FRS to better fit a specific population. The SCORE (Systematic COronary Risk Evaluation) project focused on development of a risk score based on 12 European cohort studies for Europeans. Calibrated by country, it controls for gender, age, SBP, smoking status and

cholesterol ratio (TC/HDL) (Conroy et al., 2003). QRisk2 was developed specifically for England and Wales, controlling for gender, age, ethnicity, SBP, BMI, smoke status, BP treatment, type 2 diabetes, rheumatoid arthritis, renal disease, arterial fibrillation, TC/HDL, and Townsend deprivation score that describes material deprivation based on census data and family history of CHD (Hippisley-Cox et al., 2008). The SCORE risk score utilizes similar variables as FRS, but is specifically calibrated for the European population, while QRisk2 includes many more variables including a measure of socioeconomic status that is particular to the region the model was calibrated on.

Gaziano et al. (2008) developed laboratory and non-laboratory based CVD risk assessment models using NAHNES I follow up study cohort. The laboratory model uses age, gender, SBP, smoking status, TC, reported diabetes status and treatment for high BP for prediction while the non-laboratory model replaces TC with BMI yielding similar results. This model is very interesting in practice as the non-laboratory version does not require a blood draw; however, the goal of this study is to use the most accurate representation of cardiovascular risk, not the most practical.

Mora et al. (2007) found that the relationship between PA (assessed via questionnaire) and cardiovascular risk is mediated in large part by inflammatory/hemostatic factors, BP, cholesterol, and BMI and less so diabetes based on the Women's Health Study data. The mediation effect of homocysteine and creatinine levels was almost negligible. The only inflammatory/hemostatic biomarker available in the NHANES is CRP, a variable used in the Reynolds risk score. Additionally, studies show that risk scores such as the FRS and Reynolds developed on the general population do well in predicting CVD risk in diabetics (Chamnan et al.,

2009). Thus, it appears that the Reynolds risk score is a good candidate for the outcome variable for this study without the need for a separate model for diabetics.

Measuring Physical Activity

PA is bodily movement produced by contracting skeletal muscle that in turn increases energy expenditure. This broad definition leads to much confusion and consequent difficulty in measuring PA. Many forms of PA exist and may be classified by purpose, mode, or intensity. Methods of measuring PA strive to assign a number to the PA an individual has performed for comparative purposes.

To better study the dose-response relationship between PA and cardiovascular health, accurate and objective measures of PA are needed. Frequency, intensity and duration of the PA are often studied to identify the optimal dose of PA needed to elicit a positive health benefit and reduce disease risk. Currently, several methods of measuring PA are used in practice. Since there is no gold standard measure of PA, indirect and direct measures are used as estimates. All methods have advantages and disadvantages and are chosen based on the target population, study costs and administrative burden (Ainsworth & Coleman, 2006; Strath et al., 2013).

Indirect measures. These measures are retrospective and are based on PA that has already occurred. Some examples of indirect measures include surveys and questionnaires that ask study participants to recall the amount and type of PA performed over a period of time in the past. These methods are by far the easiest and least expensive to administer to a large number of participants and are therefore most often used in large-scale longitudinal studies.

The main disadvantages of indirect measures are their low accuracy and potential bias on the part of both researchers and participants. Because questionnaires rely on memory, study participants tend to either overestimate or inadequately recall the amount of PA performed.

Numerous studies have documented the low correlation of self-report PA questionnaires and actual activity performed (Atienza et al., 2011; Clark et al., 2011; Lynch et al., 2011; Troiano et al., 2008). At the same time, questionnaires may be constructed such that a major part of someone's PA is overlooked. An example may be a compendium of activities that does not include household chores and gardening. A study participant may be physically active, but not according to the categories provided by the researchers.

Additionally, the intent of PA questionnaires is to obtain a general picture of an individual's PA in a large-scale study. Thus, these approaches focus on aggregate measures, likely ignoring individual PA patterns and specifics regarding frequency, duration and time. An avid exerciser who has a sedentary job versus someone who does no exercise at all, but walks all day due to his occupation have very different PA patterns. These patterns of PA are highly individual and lifestyle dependent and may not be adequately captured via am indirect measure such as a questionnaire.

Direct measures. These measures are collected as the PA is happening. Methods of direct measures include diary entries, where the participant records all PA preformed, and direct observation, where all PA performed by the participant is observed and recorded by the researcher. Direct measures yield objective and fairly accurate assessments of PA performed. Additionally, the measures provide information about the pattern of PA that includes the frequency, duration, and intensity. Data obtained with direct measures allows to further study bout length of the activity. However, the numerous advantages of direct measures carry a much greater burden than indirect measures.

First, these methods are only accurate for a specified time frame of the study; however, participants may alter their behavior in response to having to record their actions or to being

observed. Thus, findings may not be representative of the typical PA of an individual. Next, to obtain detailed direct measures of PA, careful planning, monitoring and training are often required. Additionally, the recorded qualitative data is time consuming to code and difficult to score and analyze, making these approaches inefficient for large scale studies. Thus, likely applications of direct measures are for smaller scale interventions and exercise training studies.

Recent technological advances offer a solution to some of the challenges associated with diary and observation methods. Electronic devices such as accelerometers measure when and at what intensity, frequency and duration the PA was performed and are continuously becoming more accurate and less expensive. Furthermore, wearing an accelerometer for a study solves issues with recall or overestimation while being unobtrusive for the participants. Thus, these devices offer objective and fairly accurate measures that are appropriate for large-scale studies and carry little administrative burden.

Accelerometers contain a small mass on a spring that is displaced by the movement of the case and presses on either a crystal (piezoelectric) or a capacitor (capacitative) to convert the motion into an electric signal. This configuration is only able to measure acceleration in one plane of motion, so the basic accelerometers are uniaxial. Triaxial accelerometers consist of three uniaxial accelerometers measuring acceleration in three different planes of motion. These types of accelerometers are included in many cell phones and fitness devices.

However, these devices are somewhat limited in detection of certain activities. Due to their construction, accelerometers are most appropriate for detecting ambulatory activities such as walking and running. Activities that do not involve undulating motion such as swimming or rowing may not be detected by certain types of accelerometers. Also, accelerometer accuracy varies depending on placement of the device on the body. It has been shown that hip placement

is more accurate for detection of most ambulatory activities, while wrist or ankle placement works better for finer movements (Cleland et al., 2013). It also appears that using two accelerometers on different locations on the body works better for detection of finer grained activities (Cleland et al., 2013).

Finally, a major issue in research using accelerometer-derived PA data is usability and interpretability of the findings. Raw accelerometer results are provided as step counts for a preset interval. Thus, the number of raw readings is usually very large. To render the data useful, the counts must be summarized in a meaningful way. Often, accelerometer data is reduced using predetermined intensity thresholds for light, moderate, and vigorous activity or converted to energy expenditure using existing models (Alhassan & Robinson, 2010; Bouten et al., 1997; Crouter, Clowers, & Bassett, 2006; Rothney et al., 2008; Troiano, 2006; Tudor-Locke, Johnson, & Katzmarzyk, 2009; Vanhelst, Béghin, Turck, & Gottrand, 2011; Wong, Colley, Connor Gorber, & Tremblay, 2011). The use of cutoffs, however, is a "crude categorization of activity status" and overlooks the unique pattern of PA on the individual level (Ainsworth & Coleman, 2006). Approaches to extracting useful findings from accelerometer data are still in development with noteworthy methods discussed below.

Accelerometers in health research. The use of accelerometers has become popular in health research (Troiano, 2006). Accelerometer data has been used to assess the adherence to PA recommendations in population-based studies and to examine the relationship between PA and a variety of health outcomes such as cancer, diabetes, kidney disease, and cardiometabolic risk factors (Camhi et al., 2011b; Gerber et al., 2012; Hawkins et al., 2011; Lynch et al., 2010, 2011; Smith et al., 2011; Tudor-Locke, Brashear, Johnson, & Katzmarzyk, 2010; Tudor-Locke et al., 2009; Wong et al., 2011). Additionally, the popularity of accelerometers in various handheld

devices such as smart phones and wristbands has sparked interest in providing real-time feedback to users tracking their own PA for fitness purposes (Cleland et al., 2013; Pärkkä, Cluitmans, & Ermes, 2010; Tapia et al., 2007). Finally, accelerometers have been used to assess PA in nonhuman subjects, notably cattle, with growing interest in assessment of pets (Guo et al., 2009; Michel & Brown, 2011; O'Connell, Tøgersen, Friggens, Løvendahl, & Højsgaard, 2010; Robert, White, Renter, & Larson, 2009).

NHANES Accelerometer Data

The National Health Survey Act of 1956, signed by President Eisenhower, commenced an on-going, cross-sectional survey representative of the American people, focusing on disease, injury, impairment, and disability ("NHIS - About the National Health Interview Survey," 2015). Conducted by the National Center for Health Statistics, initial surveys in the 1960's concentrated on chronic diseases in the adult population. In 1970, a new component focusing on nutrition and its relationship to health outcomes was introduced. In the 1980's the survey expanded to include a major focus on representing minority ethnic groups in the U.S. The current NHANES survey began in 1999, collecting numerous health related variables using questionnaires, examinations and laboratory based tests from 7,000 randomly chosen American residents every year ("NHANES - History," 2011).

In 2003, the NHANES introduced accelerometers to the examination. As described earlier, gathering free-living PA data is a challenge and a survey may not capture true activity patterns accurately. Including accelerometers in a large-scale study aimed at gathering an objective view of free-living PA patterns of a representative sample of US residents ("NHANES 2003–2004: Physical Activity Monitor Data Documentation, Codebook, and Frequencies," 2007). The NHANES 2003–2004 and 2005–2006 cycles used uniaxial accelerometers to gather a

week of PA readings. Thus, the data offers a unique opportunity to objectively study PA patterns of a representative sample of the U.S. population in conjunction with a variety of health outcomes.

Studies using NHANES accelerometer data. One of the goals of studies using the NHANES accelerometer data is to compare self-reported PA measures to those derived by the ActiGraph PA monitor. Atienza et al. (2011) show a large discrepancy between self-reported and objectively measured PA and the relationship of the two measures to various biomarkers. Studies also use the NHANES accelerometer data to look at PA patterns of the general U.S. adult population, as well as special populations including children, pregnant women, and cancer survivors (Belcher et al., 2010; Camhi, Sisson, Johnson, Katzmarzyk, & Tudor-Locke, 2011a; Clark et al., 2011; Evenson & Wen, 2011; Smith et al., 2011; Troiano et al., 2008; Tudor-Locke, Brashear, et al., 2010; Tudor-Locke et al., 2009; Tudor-Locke, Johnson, & Katzmarzyk, 2010, 2011).

Since the NHANES data also offers a variety of health-related outcomes derived both by questionnaire and laboratory testing, numerous studies strive to establish a relationship between accelerometer-derived PA and various health outcomes. These include breast cancer, kidney disease, non-alcoholic fatty liver disease, depression, and insomnia (Chasens & Yang, 2012; Gerber et al., 2012; Hawkins et al., 2011; Lynch et al., 2010, 2011; Vallance et al., 2011).

Several approaches were used to study cardiovascular health outcomes. Some studies focused on metabolic syndrome defined as having three or more cardiometabolic risk factors that include large waist circumference, high levels of triglycerides, low HDL, high blood pressure and elevated fasting glucose while others concentrated on the risk factors individually adding high BMI as an outcome (Camhi et al., 2011a; Holman et al., 2011; Luke et al., 2011; Sisson et

al., 2010). Healy et al. (2011) considered the effects of sedentary time on cardiometabolic and inflammatory markers using CRP in addition to the outcomes listed above. Atienza et al. (2011) also used skinfold measures, glycohemoglobin, C-peptide and homocysteine levels as outcomes to compare self-reported and accelerometer derived PA. No studies using NHANES accelerometer data used 10-year cardiovascular risk scores as outcomes.

Reduction of NHANES accelerometer data. The NHANES PA monitor data provides readings for every consecutive minute of the week the subject wore the accelerometer. Thus, 10,080 intensity readings are available for every individual included in the study, provided that the data is complete. Techniques are needed to extract useful information from such a large volume of data for further analysis. The majority of the studies using the NHANES accelerometer data rely on a SAS macro provided by the National Cancer Institute (NCI) to summarize the data and render it usable.

The approach suggested by NCI involves dividing the week's data into days and classifying the data into intensity categories defined by predetermined thresholds of intensity counts (Troiano et al., 2008; Trost, Pate, Freedson, Sallis, & Taylor, 2000). Once categorized, the amount of time spent in each category during the day is averaged for each person over the course of the week. This way, the average time per day spent in low, moderate, and vigorous intensity activity is calculated and the continuous nature of the data is discretized (Tudor-Locke et al., 2009). For example, every minute of an adult's intensity reading above 2,020 would be categorized as vigorous activity. Total or average time spent in vigorous activity over the week would be used for the analysis.

This approach successfully reduces the large volume of data to just one reading for the three intensity categories per person rendering it possible to use regression models to establish

the connection between PA and health status. Some studies used slightly different approaches to categorization such as analyzing cadence instead of intensity or using different thresholds for a closer look at lifestyle activity instead of the low, moderate and vigorous intensity categories (Camhi et al., 2011a; Tudor-Locke, Camhi, et al., 2011). Nevertheless, the use of thresholds to separate accelerometer data into PA categories may be masking some individual differences that may be of interest with respect to predicting health outcomes of the participants.

The rich details the data may offer are lost due to the data reduction via averaging the intensity readings using thresholds. Each individual has a unique pattern to his daily PA based on daily habits, occupation, and health status. Important clues to an individual's health status may be averaged out with this technique and remain unnoticed. Additionally, Tudor-Locke et al. (2011) suggest that the population estimates are distorted due to non-wear time at the end of the day. Furthermore, the use of thresholds may distort the results since these cutoff points are the same for everyone while moderate intensity for one person may be vigorous intensity for another who is less fit. Loprinzi et al. (2012) compare different thresholds for PA and show that changes in cut points for PA in both children and adults influenced the resulting adherence to PA guidelines and relationship to health outcomes. Analysis of the NHANES accelerometer data may be improved by using a more individualized approach to data reduction rather than the use of cutoff scores.

Analysis of NHANES accelerometer data. Following the data reduction described above, studies analyzing the NHANES accelerometer data utilized linear or logistic regression models to study the relationship between PA and health outcomes. Logistic regression is a type of regression model where the dependent variable is categorical. This approach was used when the outcome was coded as a binary variable dependent on average accelerometer-derived PA

(Belcher et al., 2010; Camhi et al., 2011b; Sisson et al., 2010; Smith et al., 2011). Linear regression models were used when continuous variables such as triglycerides or blood glucose levels were studied as outcomes (Atienza et al., 2011; Chasens & Yang, 2012; Clark et al., 2011; Gerber et al., 2012; Hawkins et al., 2011; Healy et al., 2011; Lynch et al., 2010, 2011; Tudor-Locke, Johnson, et al., 2011). Correlations, coefficients, and analysis of variance (ANOVA) were also used to study differences in PA levels in various groups (Chasens & Yang, 2012; Hawkins et al., 2011; Luke et al., 2011; Tudor-Locke, Camhi, et al., 2011; Van Domelen et al., 2011).

Studies analyzing NHANES accelerometer data focused on either the adherence to PA recommendations or establishing the relationship between PA and various health outcomes. These approaches to NHANES accelerometer data used standard statistical procedures for analysis and did not take advantage of machine learning methodology that would allow for a more tailored approach to recognizing patterns in the data. Although prior studies of the accelerometer-derived PA data show a clear relationship with health-related outcomes measured by the NHANES, analysis of the data with the help of machine learning techniques may potentially augment the findings. Machine learning techniques may be helpful for extracting useful information while retaining the individualized and highly detailed nature of the NHANES accelerometer data.

Analysis of Accelerometer Data in Other Settings

Studies that use machine learning algorithms for analysis of accelerometer-derived PA largely focus on predicting the mode of PA performed in clinical settings (Baek, Lee, Park, & Yun, 2004; Bao & Intille, 2004; Gaura, Rider, Steele, & Naguib, 2001; Mannini & Sabatini, 2010; Pärkkä et al., 2006; Pober, Staudenmayer, Raphael, & Freedson, 2006; Sprager & Zazula,

2009; Staudenmayer, Pober, Crouter, Bassett, & Freedson, 2009). Some studies aim to connect PA patterns to posture recognition and fall detection but are also conducted in a controlled environment with known activities (Giansanti, 2006; Gjoreski, Lustrek, & Gams, 2011; S.-H. Liu & Chang, 2009).

Some accelerometer studies also focusing on activity recognition have been conducted in realistic conditions outside of a clinical environment (Bao & Intille, 2004; Ermes et al., 2008; Pärkkä et al., 2006). Accelerometer-derived PA patterns in cattle, data that was collected in a free-living environment, have also been studied using machine learning algorithms with the main focus of classifying cattle movements into lying, standing, grazing etc. (Guo et al., 2009; Martiskainen et al., 2009; Robert et al., 2009).

One study did aim to connect cattle behavior monitored by accelerometers with reproductive status based on progesterone levels (O'Connell et al., 2010). That application suggests that machine learning methods may successfully be applied not only for classifying accelerometer-derived PA into activity types but also for recognizing patterns in movement that help predict health status.

Machine learning algorithms have also been applied to accelerometers data for diagnosis of tremor related disease such as Parkinson's, the classification and assessment of severity of Levodopa-Induced Dyskinesia, and recognition of involuntary gestures in babies with cerebral palsy (Gaura et al., 2001; Keijsers, Horstink, & Gielen, 2003; Keijsers, Horstink, van Hilten, Hoff, & Gielen, 2000; Singh & Patterson, 2010; Tsipouras et al., 2010). Thus, machine learning methods show promise in recognizing abnormal movement patterns for classification of disease status.

The NHANES accelerometer data offers a look at a week of PA in a natural setting connected with a variety of health-related biomarkers of the participants. No known studies have attempted to connect accelerometer-derived PA in a free-living environment and health status in human participants using machine learning.

Classification of Health Status Using Machine Learning

Machine learning algorithms have been successfully used for classification of health outcomes using other types of data in the medical field. Machine learning applications include detection of breast cancer from biopsied tissue, prediction of functional health status of HIV/AIDS patients, genetic research using mass spectrometry for detection of various types of cancers, and detection of pulmonary disease from breath data (Hauschild & Baumbach, 2012; Kwak & Lee, 1997; Shipp et al., 2002; Wolberg, Street, & Mangasarian, 1995; Wu et al., 2003).

Studies have shown that machine learning algorithms show greater accuracy for classification of at-risk populations in public health settings when compared with more traditional methods such as the logistic regression (Lemon, Roy, Clark, Friedmann, & Rakowski, 2003). Furthermore, Song, Mitnitski, Cox, and Rockwood (2004) show that machine learning algorithms are superior to traditional methods for predicting health outcomes, particularly when nonlinear relationships between the inputs and outcome variables are present. It appears that health status may successfully be classified by machine learning, but the method has yet to be applied to the NHANES accelerometer data.

Machine Learning Methods

There are two major types of problems that machine learning algorithms help solve: supervised and unsupervised. Supervised learning occurs when the outcomes are known and the machine learns to predict outcomes given new cases. A set of training data, where both inputs

and outcome variables are known, is used to build a model. The model is then applied to a set of new test data where the input variables are classified and compared to actual outcome variables. Supervised learning algorithms include regression (for continuous variables) and classification (for discrete variables) problems. Unsupervised learning problems do not assume a set of specific outcome variables and the algorithms used are aimed at finding patterns and clusters in the input variables. In this scenario the machine learns by itself (Hastie et al., 2009). Given a set of accelerometer-derived PA data from the NHANES, the purpose of this study is to classify the cardiovascular health status of the participant, a known outcome, thus supervised learning algorithms are of particular interest.

As described by Hastie et al. (2009), there are two main approaches to classification in machine learning: discriminative and generative. The first relies on the conditional probability of the outcome given a set of inputs or features. The goal is to build a decision or boundary to separate the data, for example positive and negative cardiovascular health status. Discriminative algorithms include logistic regression, decision trees, support vector machines, and neural networks. Generative algorithms, on the other hand, focus on specifying the distribution from which the input data is generated. Separate models are estimated for each outcome and each new case is classified according to which model fits best. Algorithms using this approach include hidden Markov models and naïve Bayes. To improve the predictive power of all of these algorithms, ensemble classifiers consisting of multiple models may also be used.

All of the machine learning algorithms mentioned above will be discussed in the context of accelerometer data and predicting disease outcomes or health status. A discussion of ensemble learning methods will follow. Finally, because the choice of algorithm is secondary to the choice

of features to be used as inputs in the model, the selection of features in prior work related to accelerometers will also be discussed.

Naïve Bayes classifier. Naïve Bayes is another type of generative algorithm that is popular with text classifications i.e., spam filtering (Hastie et al., 2009). The classifier "naively" assumes that different features that contribute to classification are independent of each other. Bayes' formula is used to estimate the probability that a test case belongs to a particular category. The Naïve Bayes classifier has been used for classifying activities in a gym setting using accelerometers and HR monitors (Tapia, 2008).

The simplicity of this algorithm yields quick and easy implementation and is therefore a good starting point for analysis. However, studies comparing the performance of several machine learning classifiers on accelerometer data found that other machine learning algorithms described below performed better than Naïve Bayes (Bao & Intille, 2004; Bao, 2003; Gjoreski et al., 2011; Hauschild & Baumbach, 2012; Patel, Mancinelli, Healey, Moy, & Bonato, 2009; Tapia, 2008).

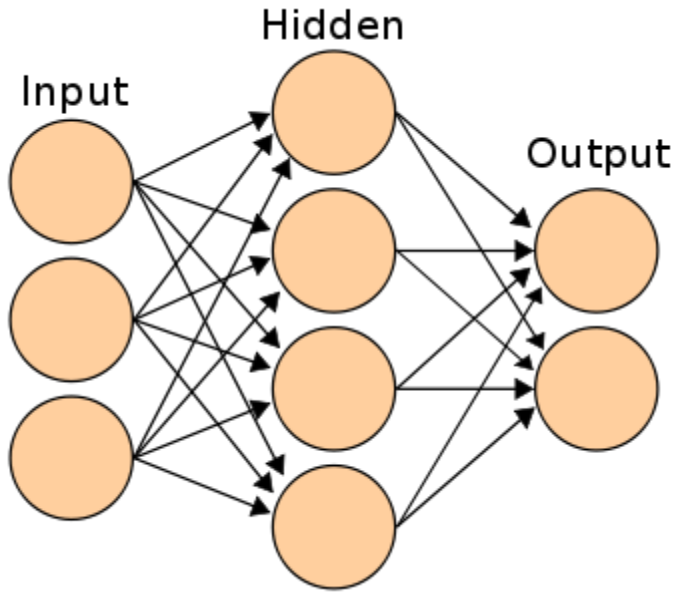
Penalized logistic regression. Logistic regression is a popular technique that has been applied to the NHANES accelerometer data to classify binary health outcomes like cardiometabolic syndrome and cancer status (Camhi et al., 2011b; Sisson et al., 2010; Smith et al., 2011). However, Song et al. (2004) showed that since the logistic regression assumes a linear separation between the observations, it does not perform as well as NN and SVM for classifying binary health outcomes. Furthermore, multicollinearity between the input variables poses a problem (Hastie et al., 2009).

To prevent overfitting and achieve more accurate classification results in the presence of highly correlated features, a regularization penalty has been proposed (Friedman, Hastie, & Tibshirani, 2010). The penalty term, called the least squares absolute shrinkage operator (lasso)

or the L1 regularization, shrinks coefficients not contributing to improved classification to zero. This method improves classification accuracy of a logistic regression and works well when there are correlated predictors in the feature set, essentially performing variable selection (Friedman et al., 2010). The lasso regression has been used on accelerometer data for activity recognition in multiple studies (Bai et al., 2014; Trost, Zheng, & Wong, 2014; Zheng, 2012).

Neural networks. Artificial neural networks (NN) are another type of discriminative classifier, as described by Hastie et al. (2009). Considering the brain is an advanced learning machine, neural networks were originally explored in an attempt to mimic the neurons in the brain for applications in artificial intelligence. The model consists of inputs that are evaluated by the neuron and sent out as outputs. The hidden layer computes a set of new features using some function of the combination of inputs to help classify them. In other words, NN is a multistage regression model that consists of input, output, and at least one hidden layer (see Figure 3 below). Various NN architectures may contain multiple hidden layers with a second layer building on the first to compute a more complex function (Hastie et al., 2009).

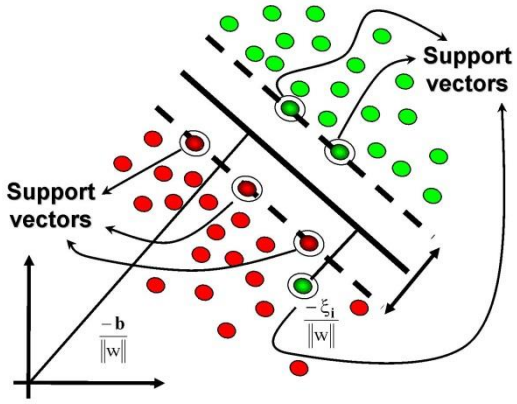
Figure 3. Neural networks chart.



Popular applications of NN include handwriting and handwritten zip code recognition (Hastie et al., 2009). The algorithm has also been popular for estimating energy expenditure and PA recognition using accelerometer data (Baek et al., 2004; Ermes et al., 2008; Freedson et al., 2011; Rothney et al., 2007; Staudenmayer et al., 2009; Trost, Wong, Pfeiffer, & Zheng, 2012). Other accelerometer-based applications include fall risk assessment in the elderly using self-constructing, fuzzy-logic NN and identification of abnormal gait patterns in patients with Complex regional pain syndrome using MLP (S. H. Liu & Chang, 2009; M. Yang et al., 2012). Comparison studies show that MLP outperformed logistic regression, single layer NN, and Support vector machines (SVM) for predicting health outcomes of at-risk Canadian subjects 65 and older (Song et al., 2004). NN work well with nonlinear hypotheses and are flexible when learning features; however they are slower to train than SVM and do not work well with many features.

Support vector machines. SVM is a discriminative classifier that separates data into outcomes by maximizing the margin or distance between the data belonging to different categories (Cortes & Vapnik, 1995). SVM utilizes optimization rather than search and is sometimes referred to as a large margin classifier. If it is not possible to separate the data by a straight line as shown in Figure 4 below, a more complex nonlinear classifier is fitted with the help of a kernel function.

Figure 4. Support vector machines chart.



SVM has been shown to be effective for studies using accelerometer data (Sprager & Zazula, 2009). Cow behavior accelerometer derived movement patterns were classified using SVM with excellent results (Martiskainen et al., 2009). In comparison studies, SVM outperformed nearest neighbor, decision tree, multilayer perceptron (MLP)—a type of neural network discussed below—Naïve Bayes, and Random Forrest classifiers for activity recognition of accelerometer data (Patel et al., 2009; S. W. S. Wang, Yang, Chen, Chen, & Zhang, 2005). Additionally, SVM was shown to perform better than decision trees, Naïve Bayes, nearest neighbor, and as good as ensemble classifiers such as Random Forests, bagging and boosting for fall detection (Luštrek & Kaluža, 2008).

Decision tree models. Decision trees are a popular type of discriminative algorithm that originated in psychology to model human decision-making. This algorithm partitions data recursively and develops rules for categorizing the data. The classifier begins at a node that tests the value of one feature deemed to be most informative. The node then branches the data and the resulting partitions are tested and classified separately based on new features.

Several studies comparing machine learning classifiers found that decision trees performed better than decision tables, nearest neighbor instance based learning (IBL), and Naïve Bayes classifiers for activity recognition using accelerometers (Bao & Intille, 2004; Bao, 2003).

Pärkkä et al., (2006) found that custom and automatically generated decision trees outperformed artificial neural networks discussed below for activity recognition. Decision trees have also been successfully used to classify movements based on accelerometer data in cattle (Robert et al., 2009). The decision tree is an efficient constructive search algorithm that builds on itself and may prove to be an excellent starting point for health status classification.

Random forests. Numerous machine learning algorithms exist for the task of pattern recognition and many appear to be promising for classifying accelerometer data by cardiovascular health outcomes. Instead of focusing on one classifier, a current trend in analysis is to combine several classifiers into a single, stronger classifier for improved classification performance. These techniques are called ensemble learning methods and include techniques such as boosting, stacking, plurality voting, and bootstrap aggregation, often referred to as bagging. Comparison of several ensemble classifiers for activity recognition using accelerometer data revealed that plurality voting performed better than boosting, bagging with stacking coming in a close second (Ravi, Dandekar, Mysore, & Littman, 2005).

An ensemble classifier that shows particular promise in classifying accelerometer data is the random forest (RF) framework that consists of bagged decision trees. RF have been used to classify PA and assess functional impairment of fine motor tasks in recovering stroke patients (Casale, Pujol, & Radeva, 2011; Kozina, Lustrek, & Gams, 2011; Patel et al., 2010). Additionally, the RF framework has been shown to perform better that than Naïve Bayes, NN, SVM, and decision trees for classification of pulmonary disease using breath data, for detecting ovarian cancer using mass spectrometer data, and for posture recognition and fall detection using accelerometer data (Gjoreski et al., 2011; Hauschild & Baumbach, 2012; Wu et al., 2003).

Various machine learning algorithms have been applied to accelerometer data and for classification of disease status and health outcomes. These methods have yet to be applied to accelerometer derived PA data with the intention of classifying cardiovascular health status. Based on the success of classification of accelerometer data using machine learning algorithms in previous studies, the methods described above appear to be promising for the purposes of this study.

Feature Extraction

All of the machine learning algorithms described above require a set of features to be derived from the raw data as inputs. Feature extraction and selection is the key to successful classification and will lead to shorter training times and reduce overfitting. Studies analyzing accelerometer data for activity recognition also use mean acceleration value, standard deviation for the range of acceleration values and correlation of pairs of axes for triaxial accelerometers (Bao & Intille, 2004; Ravi et al., 2005; S. W. S. Wang et al., 2005). Others also use root mean square and maximum value, median, peak frequency and sum of variances, and a jerk metric computed as the root mean square value of the derivative of the acceleration data (Ermes et al., 2008; Gjoreski et al., 2011; Pärkkä et al., 2006; Patel et al., 2010).

A study comparing various features for optimal activity classification with NN found that mean and standard deviation are best for distinguishing static from dynamic states and between static states for activity recognition while skewness and kurtosis best distinguish amongst dynamic states (Baek et al., 2004). Martiskainen et al. (2009) used mean, standard deviation, skewness, kurtosis, maximum value, and minimum value for activity recognition in cows using SVM. Smoothing techniques such as fast Fourier transforms (FFT), Radon (R) transforms, and Holt-Winters exponential smoothing have also been applied to accelerometer data for improved

activity recognition (Elle et al., 2005; J. Liu, Pan, & Xiangcheng, 2010; O'Connell et al., 2010; Ravi et al., 2005; Tapia, 2008; Y. Wang, Huang, & Tan, 2007).

Selection of accelerometer data features for the purposes of classifying cardiovascular health status have yet to be conducted. The NHANES offers accelerometer readings over the course of the week, aggregated by one minute. As mentioned earlier, the data has been processed with methods proposed by NCI relying on predetermined thresholds to categorize the observations into PA intensity categories (Evenson, Wen, Metzger, & Herring, 2015; Troiano et al., 2008; Tudor-Locke et al., 2009). Unsupervised learning algorithms may also be used to cluster the data into groups to derive useful features. These methods are entirely data driven, focusing on describing individual PA patterns, and are discussed below.

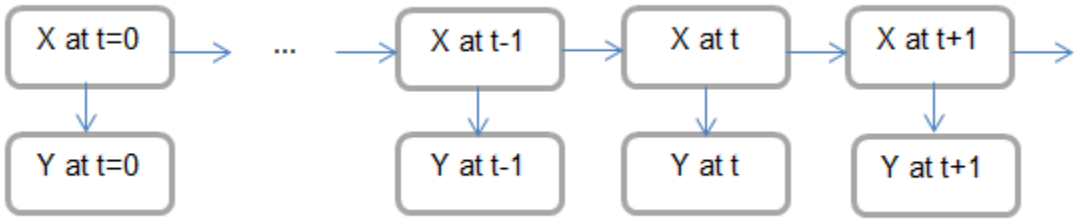
K-means. K-means is a type of clustering algorithm that aims to separate the data into k groups by minimizing the distance of the observations to the group means. This is a popular unsupervised learning technique that yields a non-overlapping partition of the data (Hastie et al., 2009). Multiple studies have used k-means clustering as an intermediate step to process accelerometer data (Choe, Min, & Cho, 2010; Krause, Siewiorek, Smailagic, & Farringdon, 2003; Laerhoven, 2001; J. Liu et al., 2010; Tapia, 2008; Zhang, Chen, & Li, 2011). This approach reduces dimensionality of the large volume of accelerometer readings, while retaining relevant data attributes.

Hidden Markov model. Hidden Markov model (HMM) is a type of generative algorithm that predicts the unobserved or hidden state of a system using a Markov chain. When the probability of an outcome or state depends only on the previous state, i.e,. the system has no memory of the events that preceded the state, the system has a Markov property (Rabiner, 1989). HMM has been used to learn from accelerometer data to recognize type of activity based on

intensity readings (Guo et al., 2009; Mannini & Sabatini, 2010; O'Connell et al., 2010; Pober et al., 2006; Reddy, Burke, Estrin, Hansen, & Srivastava, 2008). Applications of HMM to other fields include speech and gesture recognition (J. Liu et al., 2010; Pylvänäinen, 2005; Rabiner, 1989).

One advantage of the HMM is that it uses the sequential temporal information embedded in the accelerometer data and is therefore particularity useful in recognizing a sequence of physical activities performed. However, HMM may be useful as an intermediate step for feature extraction as well. Figure 5 shows a graphical representation of the HMM where X is the observed accelerometer data and Y is unobserved state to be predicted.

Figure 5. HMM with X as the observed state and Y as the hidden state.



Principal components analysis. Principal components analysis (PCA) is a popular dimensionality reduction technique that produces uncorrelated linear combinations of the data (Hastie et al., 2009). This is accomplished by a single value decomposition of the feature matrix that may be interpreted as revealing the underlying structure of the data. Because the possible features of accelerometer data described above are likely to be correlated with each other, this technique is very useful in ensuring that the feature set contains all of the relevant information while maintaining full rank. Furthermore, the number of features needed is reduced since the first principal component explains the largest amount of variance in the data. Many studies rely on PCA to process accelerometer data (Long et al., 2009; Sprager & Zazula, 2009; Trost et al., 2000).

Summary of Needs for Study

No known study has attempted to connect accelerometer-derived PA in a free-living environment and cardiovascular health status in human participants using machine learning. Machine learning algorithms have been successfully used to predict health outcomes in medical fields and have been used in various disciplines to find patterns in accelerometer data but have yet to be used to their full potential in the field of PA research. PA patterns have been studied to predict the activity type of an individual using machine learning, but not health status. Moreover, many of these studies were conducted in clinical settings, not free-living conditions.

The NAHNES offers accelerometer derived PA data that was obtained in free-living conditions in addition to an extensive number of health-related outcomes. The NHANES accelerometer data have been processed using a data reduction method that uses uniform thresholds to summarize the data that may miss more subtle individual patterns. The relationship between NHANES accelerometer data and various biomarkers has been established using and logistic regression and not machine learning techniques.

This study attempts to establish a set of features to be used with several machine learning algorithms and ensemble classifiers to best predict cardiovascular health status via a cardiovascular risk score using objective measures of PA in a nationally representative sample. Using a 10-year cardiovascular risk score as the outcome controls for age, gender, SBP, TC, HDL, and CRP by default, thus eliminating the need to control for these variables in the model. This approach allows researchers to focus the analysis on features of accelerometer data as inputs. Thus, classification algorithms would only use features obtained from the accelerometer data for prediction yielding a very versatile model.

Specific Aims

1. Explore the possibility of using the Reynolds 10-year cardiovascular risk score as an outcome measure of cardiovascular health status.
2. Connect intensity readings from accelerometer data recoded over the course of a week to a Reynolds cardiovascular risk score.
3. Extract, compare and select appropriate features from accelerometer data to use with machine learning algorithms for classification by Reynolds cardiovascular risk score.
4. Classify individuals by Reynolds cardiovascular risk score based on accelerometer-derived PA using machine learning methods.
5. Compare machine learning algorithms for classification of accelerometer-derived PA by Reynolds cardiovascular risk score.

CHAPTER 3 METHODS

The goal of this study was to assess methods for recognizing patterns in PA data that will help predict individual health outcomes. The methodology for extracting and selecting features from free-living accelerometer-derived PA NHANES data to accurately classify cardiovascular health status characterized by a cardiovascular risk score is discussed in this chapter.

Data

Free-living PA data obtained using ActiGraph accelerometers from both the NHANES 2003–2004 and 2005–2006 was used. The participants wore the activity monitors for seven consecutive days and were instructed to remove the device for sleep and water activities such as showering and swimming. 10,080 readings of intensity counts, representing each minute of the seven day period for each individual wore the PA monitor during the study were recorded (Troiano et al., 2008; Tudor-Locke, Johnson, et al., 2011).

Exclusion Criteria

Although data for ages 6 to 85 was available, only individuals 30 and older were included in the analysis since none of the risk models apply to a younger population. Furthermore, it has been shown that children's data is less consistent (Lloyd-Jones, Hong, et al., 2010; Trost et al., 2000). It must be noted that the ages of participants were censored above 85 years old. Pregnant women were also excluded from analysis as gait and PA levels are affected by pregnancy. Women who were pregnant were identified by urine sample results.

Participants returned the activity monitors by mail following the seven-day study period. Monitors that were not calibrated when returned were flagged as uncalibrated and were excluded from analysis as the resulting data may be unreliable. Additionally, the National center for Health Statistics (NCHS) and the National Cancer Institute (NCI) reviewed the data and flagged

intensity values that were outside a reasonable range (intensity ≥ 32,767). Thus, the data flagged as unreliable was also excluded from analysis.

Finally, additional screening for incomplete data was performed to identify consecutive wear of the ActiGrah monitor for the full week (Masse et al., 2005). Participants with insufficient valid data were excluded from analysis. As a starting point, the methodology proposed by the NCI was used to define non-wear time (Tudor-Locke, Camhi, & Troiano, 2012). Non-wear time was determined by consecutive intensity counts of 0, uninterrupted by either 1 minute of an intensity count > 100 or 3 consecutive intensity counts between 1 and 100. Valid wear time was summarized for each participant and those not meeting the criteria, 4 days of at least 10 hours of wear time were excluded (Luke et al., 2011; Masse et al., 2005). Participants with missing values for the outcome variables used in the analysis were excluded as well.

Measures

Due to the large number of variables collected for the participants in the NHANES in addition to accelerometer derived PA, the choice of variables to be used as outcomes was not straightforward. Studies that assessed the relationship between health outcomes and accelerometer data in the NHANES focused on cardiometabolic and inflammation biomarkers available in the data. Classification of cardiovascular health status was difficult since the NHANES is not a longitudinal data set. A potential solution explored in this study is the Reynolds 10-year cardiovascular risk score (Ridker et al., 2007).

Reynolds risk score women*:

10-year CVD risk [%] = [1 - 0.98634^(exp[B - 22.325])] x 100 [1]

where *B = 0.0799*age + 3.137*ln(systolic BP (mmHg)) + 0.180*ln(CRP (mg/L)) + 1.382*ln(total cholesterol (mg/dL)) - 1.172*ln(HDL-C (mg/dL)) + 0.134*HbA1c (if diabetic) + 0.818 (if smoker) + 0.438 (if family history).*

Reynolds risk score men*:

10-year CVD risk [%] = P = [1 - 0.8990^(exp[B - 33.097])] x 100 [2]

where *B = 4.385*ln(age) + 2.607*ln(systolic BP (mmHg)) + 0.102*ln(CRP (mg/L)) + 0.963*ln(total cholesterol (mg/dL)) - 0.772*ln(HDL-C (mg/dL)) + 0.405 (if smoker) + 0.541 (if family history).*

*CVD = cardiovascular disease; BP = blood pressure; CRP = c-reactive protein; HDL-C = high density lipoprotein cholesterol

Reynolds risk scores were calculated using gender, age, smoking status, SBP, TC, HDL CRP and family history of the mother or father having a heart attack before the age of 60 (Ridker et al., 2007). Smoking status was categorized according to serum-cotinine levels and questionnaire (Wall & Johnson, 1988). 10 year cardiovascular disease risk of 10% and above was considered high risk and coded as "1" while risk lower than 10% was considered low risk and coded as "0" (Cook et al., 2012). The resulting binary variable of cardiovascular risk was used for classification.

Additionally, the major cardiovascular disease risk factors that make up the Reynolds risk score were assessed individually. Systolic blood pressure, total cholesterol, HDL cholesterol, cRP levels were coded as binary variables according to Table 4 below, with "1" being considered

high risk and “0” low risk, and were used as outcome measures. Self-reported questionnaire data was combined with blood serum levels to determine diabetes and smoking status.

Table 4. Cardiovascular disease risk factors as individual outcomes.

Cardiovascular Disease Risk Factors	Low Risk ("0")	High Risk ("1")
Systolic blood pressure (mm Hg)	< 140	≥ 140
Total cholesterol (mg/dL)	< 240	≥ 240
HDL cholesterol (mg/dL) men	≥ 40	< 40
HDL cholesterol (mg/dL) women	≥ 50	< 50
cRP levels (mg/L)	< 1	≥ 1
Diabetes measured by HbA1c (%)	< 6.5	≥ 6.5
Diabetes (self-reported)	No	Yes
Smoker status measured by cotinine (ng/mL)	< 10	≥ 10
Smoker status (self-reported)	No	Yes
Family history of myocardial infarction before age 60?	No	Yes

Features

Prior to applying any of the machine learning algorithms, features from the accelerometer data were identified to be used as inputs. The NHANES accelerometer data contains intensity readings for every minute of the week the participants wore the accelerometer and thus contains redundant and uninformative data, for example, non-wear periods. As discussed earlier, the NHANES accelerometer data was originally processed using thresholds and the features used for classification included mean and total minutes at a low, moderate, or vigorous intensity PA. Many studies using accelerometer data focus on activity recognition and therefore utilize by the second intensity readings for precision (Kozina et al., 2011). The NHANES data on the other hand, focuses on health outcomes and only by the minute intensity readings are provided. Thus, the NHANES accelerometer data poses a unique problem of feature extraction that was explored in this study.

Certain features used for activity recognition from accelerometer data were not applicable to the data available from the NHANES. For example, Actigraph monitors used in NHANES

were uniaxial and therefore correlations of accelerations of axis pairs of triaxial accelerometers such as by Ravi et al. (2005) were not applicable. Also, the mean, root mean square, standard deviation, and other features were derived over sliding or overlapping windows of around six seconds by Gjoreski et al. (2011) were again not applicable for the by minute readings in the NHANES data. Feature selection for this data focused on representing a general pattern of PA of an individual over the course of the week, not concentrating on the type of activity performed.

Feature extraction. Multiple features sets were extracted for this study. First, NCI based features that focus on time spent in each intensity category outlined in Table 5 were computed (Freedson, Bowles, Troiano, & Haskell, 2012; Tudor-Locke et al., 2012). Features that further describe the pattern of PA throughout the week, including the number of 10 minute bouts of moderate and vigorous activity (MVPA), the number of sedentary bouts greater than 60 minutes, and the difference in lifestyle and MVPA minutes and bouts between week and weekend days were also considered.

Table 5. Thresholds for PA intensity categories.

Activity type	Intensity counts
Sedentary	<100
Light	100 - 759
Lifestyle	760 - 2019
Moderate	2020 - 5998
Vigorous	≥5999

Feature extraction using clustering algorithms was also explored. K-means and HMM were applied to the data, grouping the accelerometer readings of each individual study participant into five categories. Five categories were chosen to correspond to the five PA intensity categories used for the NCI features. Means, variances, skewness, and kurtosis were computed for the distribution of each cluster group for both k-means and HMM. Additionally,

the probability of each individual transitioning between the five states was estimated by the HMM and used as features for classification.

Feature selection. First, each set of features was checked for full rank and linear combinations were eliminated. Next each feature set was reduced using a high correlation filter with a threshold of 0.7. Thus, subsets of the NCI, k-means and HMM derived features sets were created to only contain features whose absolute correlation coefficient with any other feature in the set was less than 0.7. A subset of the NCI features was also manually chosen to accomplish the task of creating a low correlation subset. Next, each feature set was broken down into principal components. The principal components that explain at least 90% of the variance were chosen and used as features. Finally, the three features sets were combined and used for model training all together. Because the Reynold's risk score was developed separately for men and women, gender was included as a control variable in all feature sets.

Classification

Lasso penalized logistic regression, neural networks, SVM, decision trees, and random forest classifiers were trained on the accelerometer-derived feature sets described above. The binary outcomes used for classification were the Reynolds CVD risk score as well as its individual components, systolic BP, cRP, total and HDL cholesterol, as outlined in Table 4. Due to the difference in the Reynold's risk score between men and women, the classifiers were also trained separately by gender.

Comparison and Validation

The classifiers were trained using randomly selected 80% of the data and tested on the rest of the data. The models were compared using several performance indicators. First, classification accuracy, the model prediction agreement with the true outcome, was computed.

Because the sample contained more low risk than high risk cases, the outcome measure was imbalanced. Therefore, relying solely on classification accuracy for model assessment would be misleading. To account for bias, the kappa coefficient was also computed, thus adjusting the observed agreement rate by the expected agreement due to chance. If either the false positive or the false negative rate was heavily favored, the kappa coefficient would be lower than if the misclassification was more balanced.

Another measure of model performance used was the area under the receiver operating characteristic (ROC) curve, or AUC. The ROC curve plots the false positive rate (1-specificity) against the true positive rate (sensitivity). Therefore, the AUC represents the probability of correctly identifying a randomly selected high risk participant over a randomly selected low risk participant. The AUC ranges from 0.5 to 1. Unlike classification accuracy, this measure is not biased by an unbalanced sample. Additionally, performance of the machine learning algorithms was assessed using sensitivity (proportion of correctly identified positives), specificity (proportion of correctly identified negatives), false negative and false positive rates (Bao, 2003; Loprinzi et al., 2012; Patel et al., 2009; Song et al., 2004).

Software

All analysis was performed using R statistical software (R Core Team, 2014). The penalized logistic regression was fit using the 'glmnet' package (Friedman et al., 2010). The neural network with the 'nnet' package (Venables & Ripley, 2002); random forest classifier with the 'randomForest' package (Liaw & Wiener, 2002); decision trees with the 'rpart' package (Therneau, Atkinson, & Ripley, 2015); and SVM with the 'e1071' package (Meyer, Dimitriadou, & Hornik, 2015). Model tuning was conducted with the 'caret' (Kuhn, 2008) and 'e1071'

packages depending on the model; visualizations were created with the ‘ggplot2’ (Wickham, 2009), ‘corrplot’ (Wei & Simko, 2016), and ‘tabplot’ packages (Tennekes & Jonge, 2012).

CHAPTER 4 RESULTS

This chapter describes the results of the analysis described in Chapter 3 to meet the specific aims of this study. First, the characteristics of the study participants, including the Reynolds 10-year cardiovascular risk score are summarized. Next appropriate features from the accelerometer data are extracted, compared and selected. Then, individuals are classified by their binary Reynolds cardiovascular risk score using the various feature sets. Finally, the machine learning algorithms used for classification are compared.

Summary Statistics

As a result of filtering the 2003-2006 NHANES data by the exclusion criteria outlined in Chapter 3, data for 4,236 individuals remained and was used in this study. Table 6 shows the characteristics of the remaining participants, including demographics and CVD risk factors. The sample data consists of 2029 women and 2207 men with an average age of 56.24 (SD = 15.67). The average Reynold's risk score was 9% with a standard deviation of 13%. However, the distribution of this measure was positively skewed, with a maximum computed Reynold's risk score of 96% in the sample; the median Reynold's risk score was 4%.

Table 6. Summary statistics of study participants.

Cardiovascular Disease Risk Factors	Mean	SD	Median	Min	Max
Age	56.24	15.67	55	30	85
BMI	28.65	5.82	27.85	13.36	63.87
Systolic blood pressure (mm Hg)	129.39	21.4	126	80	270
Total cholesterol (mg/dL)	203.12	40.93	200	84	458
HDL cholesterol (mg/dL)	54.3	15.96	52	17	154
cRP levels (mg/L)	0.43	0.78	0.21	0.01	18.5
HbA1c (%)	5.7	1.02	5.5	3.9	14
Reynold's risk score	0.09	0.13	0.04	0	0.96

Next, the summary statistics are presented by gender in Table 7. There did not appear to be any notable differences between men and women in this study except for HDL, HbA1c and Reynold's risk score. Women appeared to have a higher HDL (0.49 mg/dL) and HbA1c

(60.05%), and a lower Reynold's risk score (6%) than men (0.38 mg/dL, 49.01% and 11%, respectively). Figure 6 illustrates the CVD risk factors partitioned by gender with the Reynold's risk score presented separately in Figure 7.

Table 7. Summary statistics of study participants by gender.

	Female (N = 2029)		Male (N = 2207)	
Cardiovascular Disease Risk Factors	**Mean**	**SD**	**Mean**	**SD**
Age	56.41	15.57	56.09	15.77
BMI	28.94	6.53	28.38	5.08
Systolic blood pressure (mm Hg)	129.92	23.82	128.91	18.91
Total cholesterol (mg/dL)	206.73	40.54	199.79	41
HDL cholesterol (mg/dL)	60.05	16.29	49.01	13.68
cRP levels (mg/L)	0.49	0.71	0.38	0.83
HbA1c (%)	5.67	0.95	5.73	1.07
Reynold's risk score	0.06	0.11	0.11	0.14

Figure 6. Individual CVD risk factors grouped by male (green) and female (coral) participants.

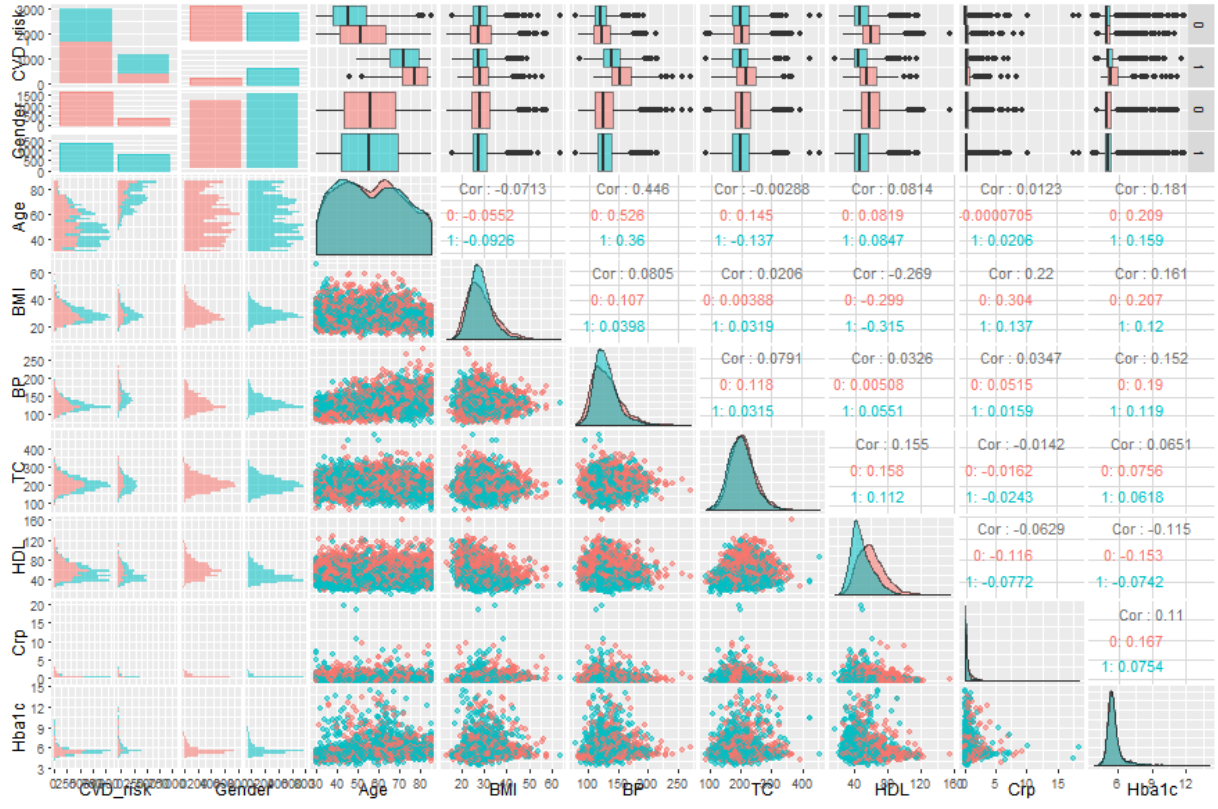


Figure 7. Reynold's risk score by male (green) and female (coral) participants.

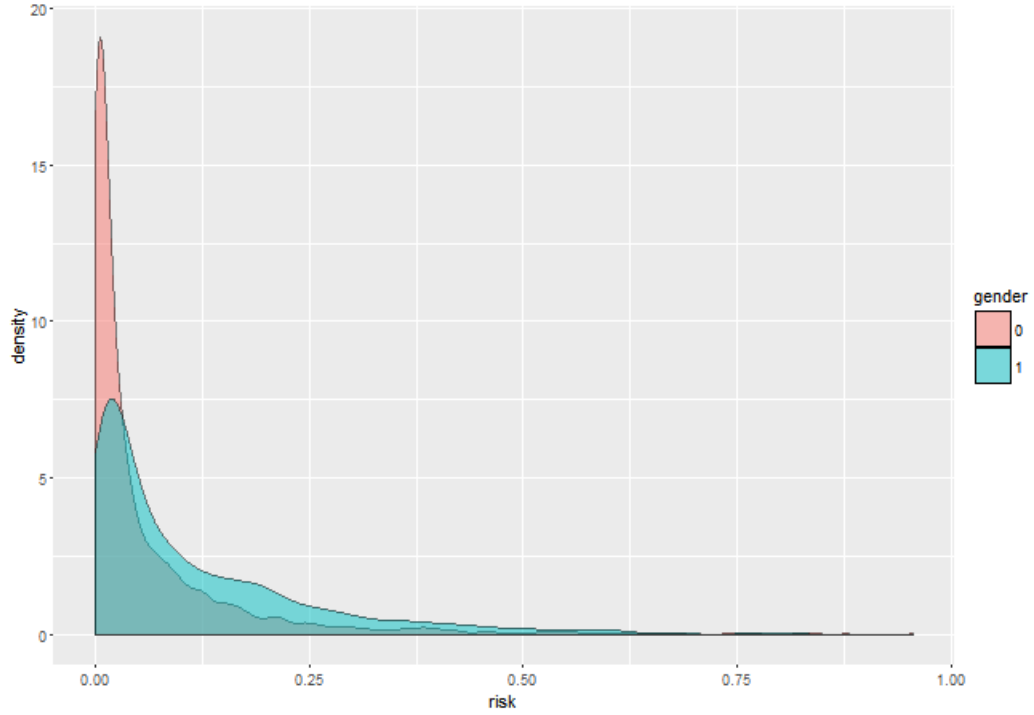


Table 8 summarizes the CVD risk factors in terms of high and low risk (with the thresholds for high risk described in Table 4 in Chapter 3). 28.38% of the participants had a Reynold's risk score of 10% or greater, indicating a high risk of a cardiovascular event in the next 10 years. More than half of the participants had a family history of myocardial infarction, and less than 10% were considered at a high risk based on their cRP levels. Interestingly, the self-reported smoking status was slightly lower than that determined by serum cotinine levels, while the self-reported diabetes status was slightly higher than the one based on HbA1c levels.

Figure 8 shows the CVD risk factors by high and low CVD risk score. The most apparent difference between high and low risk categories was in age. Unsurprisingly, participants who were at a high risk of CVD, appeared to be older. Another risk factor that appeared to be different for the high and low risk categories was BP. Again, since high BP was in itself a form of CVD, those at high risk of CVD appeared to have high BP. The rest of the risk factors,

including TC, HDL, cRP and Hba1c did not appear to be different between the high and low risk categories.

Table 8. Binary CVD risk factors.

Cardiovascular Disease Risk Factors	Low Risk	High Risk	% High Risk
Systolic blood pressure (mm Hg)	3088	1148	27.10%
Total cholesterol (mg/dL)	3519	717	16.93%
HDL cholesterol (mg/dL)	3137	1099	25.94%
cRP levels (mg/L)	3829	407	9.61%
Diabetes measured by HbA1c (%)	3777	451	10.67%
Diabetes (self-reported)	3719	517	12.20%
Smoker status measured by cotinine (ng/mL)	3276	955	22.57%
Smoker status (self-reported)	3412	824	19.45%
Family history of myocardial infarction before age 60?	1835	2401	56.68%
Reynold's Risk score >= 10%	3034	1202	28.38%

Figure 8. Individual CVD risk factors grouped by high (green) and low (coral) CVD risk score.

Figure 9 shows a visual representation of the binary risk factors for CVD. The first bar represents the outcome measure based on the Reynold's risk score. The rest of the data was ordered according to high and low risk CVD category of the outcome measure. Again, there did

not appear to be a difference between the amount of smokers, people with diabetes, high cRP, total and HDL cholesterol in high and low risk categories. However, there did appear to be more participants with high blood pressure in the high CVD risk category.

Figure 9. Binary CVD risk factors ordered by high and low risk categories.

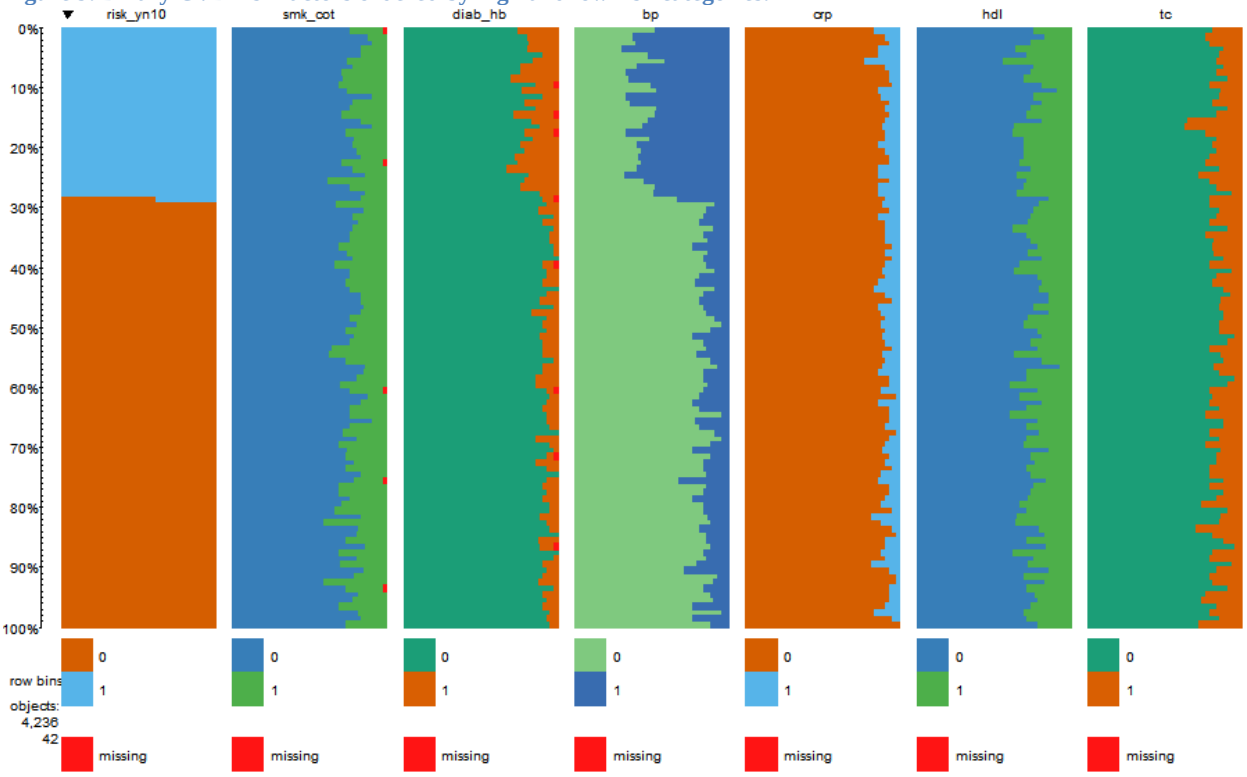


Table 9 shows the CVD risk factors separated by gender. There were fewer female smokers, and the discrepancy between self–reported and cotinine–derived smoking status appeared only in men. Slightly more women seemed to have high systolic BP and cRP levels than men. Notably, the percentage of participants at high CVD risk based on their Reynold's risk score differed by almost 20% between men and women, with more men at high risk.

Counts of participants in high and low risk categories based on CVD factors that make up the Reynold's risk score are shown in Table 10. While no one in the sample had all six risk factors flagged, three participants, one male and two female had five of the six risk factors. There did not appear to be a gender difference in terms of the number of CVD risk factors.

Table 9. Binary CVD risk factors by gender.

Gender	Female (N = 2029)			Male (N = 2207)		
Cardiovascular Disease Risk Factors	**Low Risk**	**High Risk**	**% High Risk**	**Low Risk**	**High Risk**	**% High Risk**
Systolic blood pressure (mm Hg)	1419	610	30.06%	1669	538	24.38%
Total cholesterol (mg/dL)	1639	390	19.22%	1880	327	14.82%
HDL cholesterol (mg/dL)	1463	566	27.90%	1674	533	24.15%
cRP levels (mg/L)	1763	266	13.11%	2066	141	6.39%
Diabetes measured by HbA1c (%)	1820	209	10.30%	1957	242	11.01%
Diabetes (self-reported)	1783	246	12.12%	1936	271	12.28%
Smoker status measured by cotinine (ng/mL)	1702	324	15.99%	1574	631	28.62%
Smoker status (self-reported)	1709	320	15.77%	1703	504	22.84%
Family history of myocardial infarction before age 60?	844	1185	58.40%	991	1216	55.10%
Reynold's Risk score ≥10%	1660	369	18.19%	1374	833	37.74%

Table 10. Count and percentage of participants having a combination of CVD risk factors.

Number of CVD risk factors	0		1		2		3		4		5	
	N	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
Female	334	16.46	718	35.39	636	31.35	285	14.05	54	2.66	2	0.10
Male	412	18.67	812	36.79	683	30.95	261	11.83	38	1.72	1	0.05
All	746	17.61	1530	36.12	1319	31.14	546	12.89	92	2.17	3	0.07

Overall, there did not appear to be a gender effect for the individual CVD risk factors. While the average Reynold's risk score did not appear to be substantially different between men and women, there were almost 20% more men with a Reynold's risk score of ≥ 10% in the sample. These findings are a strong indicator that gender should be controlled for when using the Reynold's risk score as the outcome measure.

Features

The features extracted from the accelerometer data are discussed in this section. Three categories of features were chosen for this study, those based on the established PA intensity thresholds (Troiano et al., 2008), those derived by k-means clustering and those based on the HMM.

NCI methods. The methods proposed by NCI to process the NHANES accelerometer data and the thresholds proposed by Troiano et al. (2008) were used to establish the first set of features in this study. The average number of minutes spent in each PA intensity category is presented in Table 11. The participants spent the most time in sedentary and light categories. At the same time, it appears that individuals spent less than a minute in vigorous activity, on average.

Table 11. NCI PA categories.

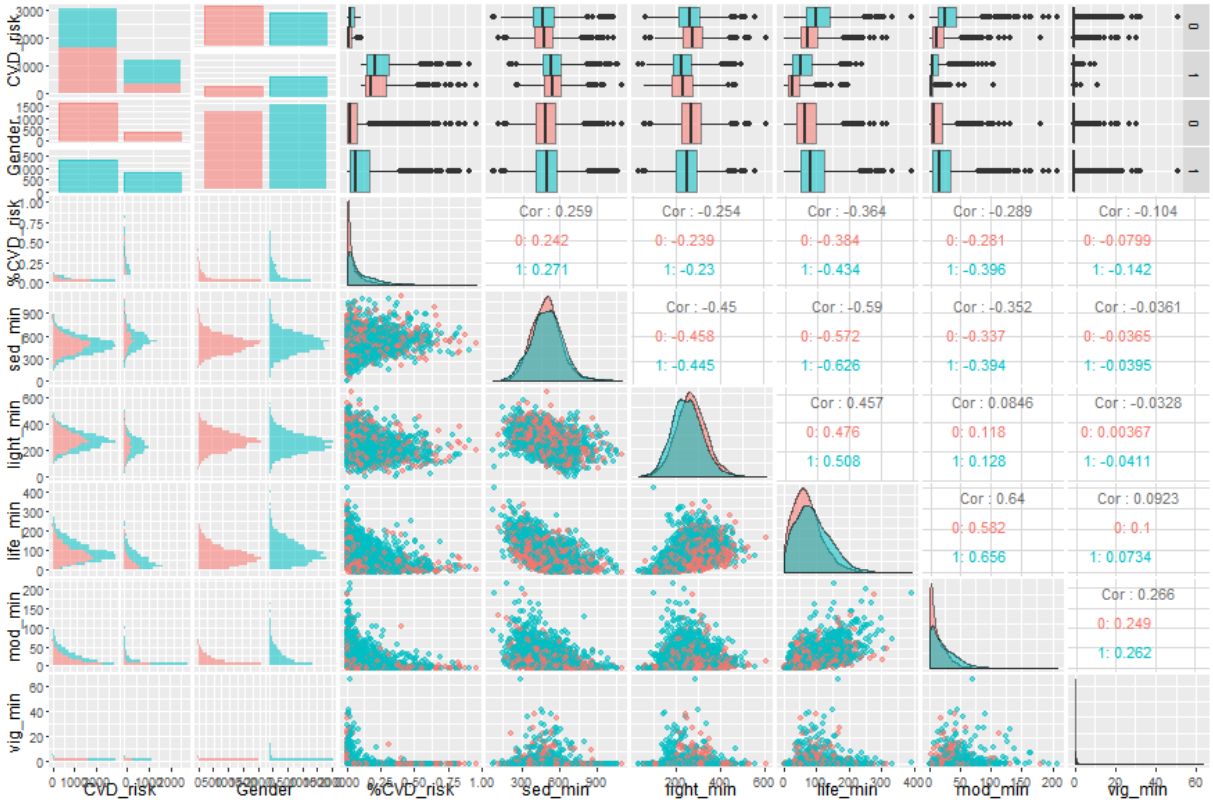
PA type minutes	Mean	SD	Median	Min	Max
Sedentary	494.89	124.81	492.15	67.5	1088.33
Light	258.68	69.71	257.67	31	608
Lifestyle	81.62	50.73	73.86	0.67	393.2
Moderate	19.26	21.41	12.29	0	208.5
Vigorous	0.63	2.93	0	0	63.4

Table 12 shows the breakdown of minutes spent in various PA categories by gender. While sedentary and light minutes appeared to be similar, there did appear to be a difference in the amount of lifestyle and moderate activity performed by men and women. The activity categories separated by gender are illustrated in Figure 10.

Table 12. NCI PA categories by gender.

	Female (N = 2029)		Male (N = 2207)	
PA type minutes	Mean	SD	Mean	SD
Sedentary	491.31	117.13	498.19	131.42
Light	268.47	68.38	249.68	69.73
Lifestyle	72.46	45.11	90.05	54.05
Moderate	13.88	15.62	24.2	24.59
Vigorous	0.42	2.26	0.82	3.42

Figure 10. Time spent in NCI PA categories by male (green) and female (coral) participants.

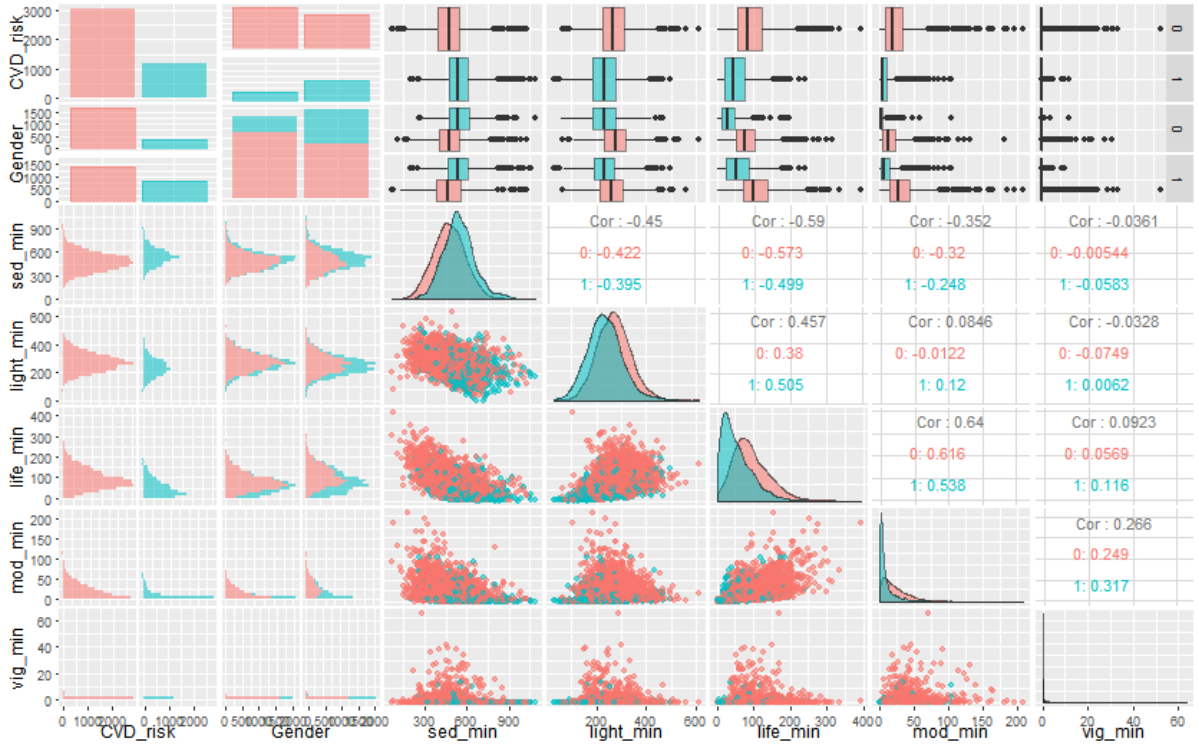


The PA intensity features separated by CVD risk based on the Reynold's risk score are presented in Table 13 and illustrated in Figure 11. Those at high risk of CVD appeared to have spent more time in the sedentary category and less time in lifestyle and moderate activity.

Table 13. NCI PA categories by CVD risk.

	Low CVD Risk		**High CVD Risk**	
PA type minutes	**Mean**	**SD**	**Mean**	**SD**
Sedentary	475.12	121.2	544.8	119.81
Light	269.36	66.98	231.71	69.21
Lifestyle	92.99	49.88	52.94	40.5
Moderate	23.25	22.61	9.19	13.57
Vigorous	0.83	3.41	0.12	0.76

Figure 11. Time spent in NCI PA categories separated by high (green) and low (coral) CVD risk.



Finally, Table 14 shows the PA intensity categories by gender and CVD risk. There did not appear to be a gender effect for sedentary and light time, with men and women spending roughly the same amount of time for the high and low risk categories. For both high and low CVD risk categories, women spent less time in lifestyle and moderate PA than men.

Table 14. NCI PA categories by gender and CVD risk.

	Low CVD Risk				**High CVD Risk**			
	Female		**Male**		**Female**		**Male**	
PA type minutes	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Sedentary	478.11	112.62	471.52	130.76	550.7	118.75	542.19	120.25
Light	275.95	64.28	261.4	69.31	234.83	75.84	230.33	66.06
Lifestyle	80.61	43.74	107.95	52.69	35.83	30.74	60.52	41.98
Moderate	15.98	16.06	32.02	26.02	4.42	8.5	11.3	14.81
Vigorous	0.5	2.47	1.23	4.25	0.07	0.81	0.14	0.74

In addition to the NCI PA categories, features that account for how PA was spread throughout the day and week were calculated. The features still relied on the NCI methods

(Tudor-Locke et al., 2009) and thresholds outlined in Table 5. The features included the amount of time spent in sedentary bouts of one hour or longer, the number of times per day the participant engaged in at least 10 minute bouts of MVPA, the number of days per week with 2 or more bouts if MVPA, and the difference between the number of MVPA bouts accumulated during the week and the weekend. Also, the total minutes spent in lifestyle and MVPA categories per day were included for the top four days for each participant. The summary of all NCI features, along with the descriptions of each variable is presented in Table 15.

K-means clustering. To further explore various PA activity patterns, the data was clustered into categories for each participant using k-means. With this method, thresholds for each participant would vary given the activity levels recorded throughout the week. The data was divided into five clusters to correspond to the five PA intensity levels established by Troiano (2008) and used as part of the NCI feature set. Table 16 summarizes the results of applying a k-means cluster to the data.

HMM clustering. The accelerometer data was also clustered using the HMM. Similar to k-means, the data was clustered into 5 categories; however, in addition to computing the distributions for the categories for each participant, the probabilities of transitioning between the categories were also estimated. In order to fit HMM, the zero readings were removed from the data. Table 17 shows a summary of the HMM features.

Table 15. NCI PA pattern features.

Variable Name	Variable Description	Mean	SD	Median	Min	Max
sed_min	Sedentary minutes per day	494.89	124.81	492.15	67.5	1088.33
light_min	Light minutes per day	258.68	69.71	257.67	31	608
life_min	Lifestyle minutes per day	81.62	50.73	73.86	0.67	393.2
mod_min	Moderate minutes per day	19.26	21.41	12.29	0	208.5
vig_min	Vigorous minutes per day	0.63	2.93	0	0	63.4
sed_bouted_60min	Sedentary minutes accumulated in bouts of length ≥ 60 min	21.66	35.66	10.71	0	441.57
num_mvpa_bouts	Number of MVPA bouts per day ≥ 10	0.29	0.6	0	0	7.67
num_vig_bouts	Number of vigorous bouts per day ≥10	0.02	0.15	0	0	3.29
mvpa_bouted	MVPA minutes accumulated in bouts of length ≥ 10 min	5.72	12.46	0	0	141.83
vig_bouted	Vigorous intensity minutes accumulated in bouts of length ≥ 10 min	0.36	2.49	0	0	56.4
tot_mv_bouts	Total week MVPA bouts	1.79	3.64	0	0	46
tot_mv_min	Total week MVPA Minutes	120.27	136.41	74	0	1251
tot_li_min	Total week Lifestyle Minutes	494.85	317.29	442	4	2332
wk_mv_dif_bout	Week - weekend MVPA bouts	0.07	0.64	0	-7.5	6.4
avg_wk_mv_dif_min	Average week - weekend MVPA minutes	5.29	21.04	1.8	-186.5	194.8
avg_wk_li_dif_min	Average week - weekend lifestyle minutes	15.54	49.54	7.5	-178.67	393.2
tot_wk_mv_dif_min	Total week - weekend MVPA minutes	68.12	95.18	35	-266	974
tot_wk_li_dif_min	Total week - weekend lifestyle minutes	264.06	222.13	213	-178	1966
perc_wk_mv	Percent week MVPA minutes	0.77	0.19	0.8	0	1
perc_we_mv	Percent weekend MVPA minutes	0.23	0.18	0.2	0	1
perc_wk_li	Percent week lifestyle minutes	0.76	0.13	0.76	0.16	1
perc_we_li	Percent weekend lifestyle minutes	0.24	0.13	0.24	0	0.84
mv_wk2	Number of days with at least two MVPA bouts >= 10	0.4	0.99	0	0	7
top_mv.1	Highest MVPA minutes in one day	37.46	38.95	26	0	365
top_mv.2	Second highest MVPA minutes in one day	26.57	30.11	16	0	278
top_mv.3	Third highest MVPA minutes in one day	20.56	24.92	11	0	257
top_mv.4	Fourth highest MVPA minutes in one day	15.85	20.59	8	0	194
top_li.1	Highest lifestyle minutes in one day	123.32	72.65	112	2	470
top_li.2	Second highest lifestyle minutes in one day	99.71	62.4	90	0	445
top_li.3	Third highest lifestyle minutes in one day	85.6	56.35	75	0	441
top_li.4	Fourth highest lifestyle minutes in one day	72.81	50.06	64	0	340

Table 16. K-means clustering features.

Variable Name	Variable Description	Mean	SD	Median	Min	Max
min.1	Minutes in category 1	8223.14	560.5	8254.5	5858	9805
min.2	Minutes in category 2	1042.16	290.55	1027	168	2657
min.3	Minutes in category 3	517.03	196.84	502.5	36	1394
min.4	Minutes in category 4	223.24	115.93	208	7	889
min.5	Minutes in category 5	74.43	55.91	63	1	550
threshold.1	Threshold for category 1	198.5	82.81	185	40	806
threshold.2	Threshold for category 2	649.44	278.35	598	143	3979
threshold.3	Threshold for category 3	1346.05	648.01	1218	315	11364
threshold.4	Threshold for category 4	2745.86	2096.29	2273	564	32767
threshold.5	Maximum for category 5	5856.37	3526.98	4981.5	1192	32767
center.1	Centers of category 1	16.11	8.88	14.26	1.25	77.98
center.2	Centers of category 2	379.9	158.25	354.72	76.24	1552.91
center.3	Centers of category 3	916.57	401.67	844.21	203.86	6579.47
center.4	Centers of category 4	1763.67	891.05	1580.99	417.43	15664.24
center.5	Centers of category 5	3517.13	2565.77	2898.34	707.2	32767
variance.1	Variance of category 1	1770.91	1695.23	1278.12	28.47	20808.17
variance.2	Variance of category 2	18711.43	18699.16	13575.72	803.5	351468.19
variance.3	Variance of category 3	44559.42	105371.38	28547.3	2226.8	5795191.63
variance.4	Variance of category 4	137420.87	409281.55	71276.34	4885.76	13800445.2
variance.5	Variance of category 5	869534.31	3715371.22	313846.95	0	105947456
skewness.1	Skewness of category 1	2.95	0.62	2.88	1.13	7.18
skewness.2	Skewness of category 2	0.39	0.16	0.38	-0.51	1.72
skewness.3	Skewness of category 3	0.44	0.21	0.41	-0.56	2.73
skewness.4	Skewness of category 4	0.62	0.4	0.56	-2.32	5.35
skewness.5	Skewness of category 5	1.41	1.13	1.29	-4.83	11.65
kurtosis.1	Kurtosis of category 1	8.62	4.46	7.76	-0.08	57.04
kurtosis.2	Kurtosis of category 2	-0.98	0.22	-1.03	-1.31	2.67
kurtosis.3	Kurtosis of category 3	-0.88	0.43	-0.97	-1.41	9.78
kurtosis.4	Kurtosis of category 4	-0.43	1.48	-0.75	-1.91	35.63
kurtosis.5	Kurtosis of category 5	3.55	7.66	1.36	-2.75	146.24

Table 17. HMM clustering features.

Variable Name	Variable Description	Mean	SD	Median	Min	Max
min.1	Minutes in category 1	735	245.76	702.5	126	4628
min.2	Minutes in category 2	920.06	263.75	893.5	0	2175
min.3	Minutes in category 3	1083.65	312.39	1057	0	2501
min.4	Minutes in category 4	975.99	348.99	951.5	0	2800
min.5	Minutes in category 5	347.21	230.47	295	0	1661
trans_prob.1	Transition probability from category 1 to 1	0.49	0.09	0.48	0	0.84
trans_prob.2	Transition probability from category 1 to 2	0.27	0.05	0.27	0	0.47
trans_prob.3	Transition probability from category 1 to 3	0.14	0.05	0.14	0	0.42
trans_prob.4	Transition probability from category 1 to 4	0.08	0.04	0.08	0	0.31

Table 17 (cont.)

trans_prob.5	Transition probability from category 1 to 5	0.01	0.01	0.01	0	0.14
trans_prob.6	Transition probability from category 2 to 1	0.2	0.05	0.2	0	0.42
trans_prob.7	Transition probability from category 2 to 2	0.48	0.08	0.48	0	0.92
trans_prob.8	Transition probability from category 2 to 3	0.21	0.05	0.21	0	0.44
trans_prob.9	Transition probability from category 2 to 4	0.08	0.04	0.08	0	0.28
trans_prob.10	Transition probability from category 2 to 5	0.02	0.02	0.01	0	0.22
trans_prob.11	Transition probability from category 3 to 1	0.08	0.03	0.08	0	0.31
trans_prob.12	Transition probability from category 3 to 2	0.18	0.05	0.18	0	0.38
trans_prob.13	Transition probability from category 3 to 3	0.52	0.09	0.52	0	0.94
trans_prob.14	Transition probability from category 3 to 4	0.19	0.05	0.19	0	0.4
trans_prob.15	Transition probability from category 3 to 5	0.03	0.02	0.02	0	0.14
trans_prob.16	Transition probability from category 4 to 1	0.05	0.02	0.05	0	0.21
trans_prob.17	Transition probability from category 4 to 2	0.08	0.03	0.07	0	0.29
trans_prob.18	Transition probability from category 4 to 3	0.21	0.06	0.2	0	0.46
trans_prob.19	Transition probability from category 4 to 4	0.59	0.09	0.58	0	0.98
trans_prob.20	Transition probability from category 4 to 5	0.08	0.05	0.08	0	0.32
trans_prob.21	Transition probability from category 5 to 1	0.02	0.02	0.01	0	0.26
trans_prob.22	Transition probability from category 5 to 2	0.03	0.03	0.03	0	0.28
trans_prob.23	Transition probability from category 5 to 3	0.08	0.05	0.07	0	0.41
trans_prob.24	Transition probability from category 5 to 4	0.22	0.1	0.22	0	0.78
trans_prob.25	Transition probability from category 5 to 5	0.65	0.14	0.64	0	0.99
center.1	Ceneter of category 1	7.72	5.26	6.47	0	119.86
center.2	Ceneter of category 2	52.26	33.46	44.95	0	798.98
center.3	Ceneter of category 3	210.08	117.18	182.63	0	1342.5
center.4	Ceneter of category 4	651.38	330.67	585.54	0	3617.77
center.5	Ceneter of category 5	1918.42	1353.81	1534.42	0	24706.97
variance.1	Variance of category 1	52.13	201.77	22.14	0	9913.77
variance.2	Variance of category 2	1287.84	3983.14	670.27	0	218373.75
variance.3	Variance of category 3	15182.79	21054.2	8577.43	0	334599.18
variance.4	Variance of category 4	124407.8	147885	78043.9	0	2023938.2
variance.5	Variance of category 5	1255096	5073612	555482	0	123735091
skewness.1	Skewness of category 1	0.64	0.3	0.63	0.14	13.89
skewness.2	Skewness of category 2	0.58	0.12	0.59	0	1.1
skewness.3	Skewness of category 3	0.5	0.14	0.51	-0.18	1.1
skewness.4	Skewness of category 4	0.57	0.22	0.56	-0.51	2.21
skewness.5	Skewness of category 5	1.17	0.97	1.05	-2.43	10.38
kurtosis.1	Kurtosis of category 1	-0.45	6.93	-0.64	-1.22	381.05
kurtosis.2	Kurtosis of category 2	-0.54	0.21	-0.56	-1.16	1.04
kurtosis.3	Kurtosis of category 3	-0.48	0.22	-0.5	-1.11	0.9
kurtosis.4	Kurtosis of category 4	-0.02	0.48	-0.11	-0.89	11.35
kurtosis.5	Kurtosis of category 5	4.17	7.71	2.15	-2.33	181.66

Feature Selection

The results of applying several feature selection methods are presented here. The derived feature subsets are summarized for each of the feature sets, NCI, k-means and HMM.

NCI features. For this feature set, the first subset was selected manually to be a basic set of PA activity descriptors, the average time spent in each of the five intensity categories. This subset is summarized in Tables 6-9 and represents the classic approach to aggregating accelerometer data using PA intensity thresholds.

The matrix of all features was checked for full rank, and no linear combinations were found. Next, highly correlated features were identified and removed, creating a new subset of NCI features with low correlation. The correlation matrices of the full feature set and the low correlation subset are visually summarized in Figure 12 and Figure 13.

Because the features describing the patterns of PA throughout the week were chosen to provide a more detailed description of time spent in PA categories, the basic features, particularly lifestyle and moderate minutes were highly correlated with the PA pattern features. The low correlation subset did not include lifestyle, moderate, and vigorous PA minutes, but instead included the average number of vigorous activity bouts, the difference in MVPA during the week and weekend, and the highest number of minutes spent in lifestyle activity during one day.

Since the choice of low correlation set variables was automated, another subset of NCI features was chosen manually. The aim was also to reduce the correlation between variables. Figure 14 shows the resulting correlation plot. Like in Figure 13, the correlations of the features in the subset were under |0.7|. Table 18 summarizes the sets of NCI based features to be used for classification.

Figure 12. Correlation plot of all NCI features, with dark blue indicating strong positive correlation and dark red indicating strong negative correlation.

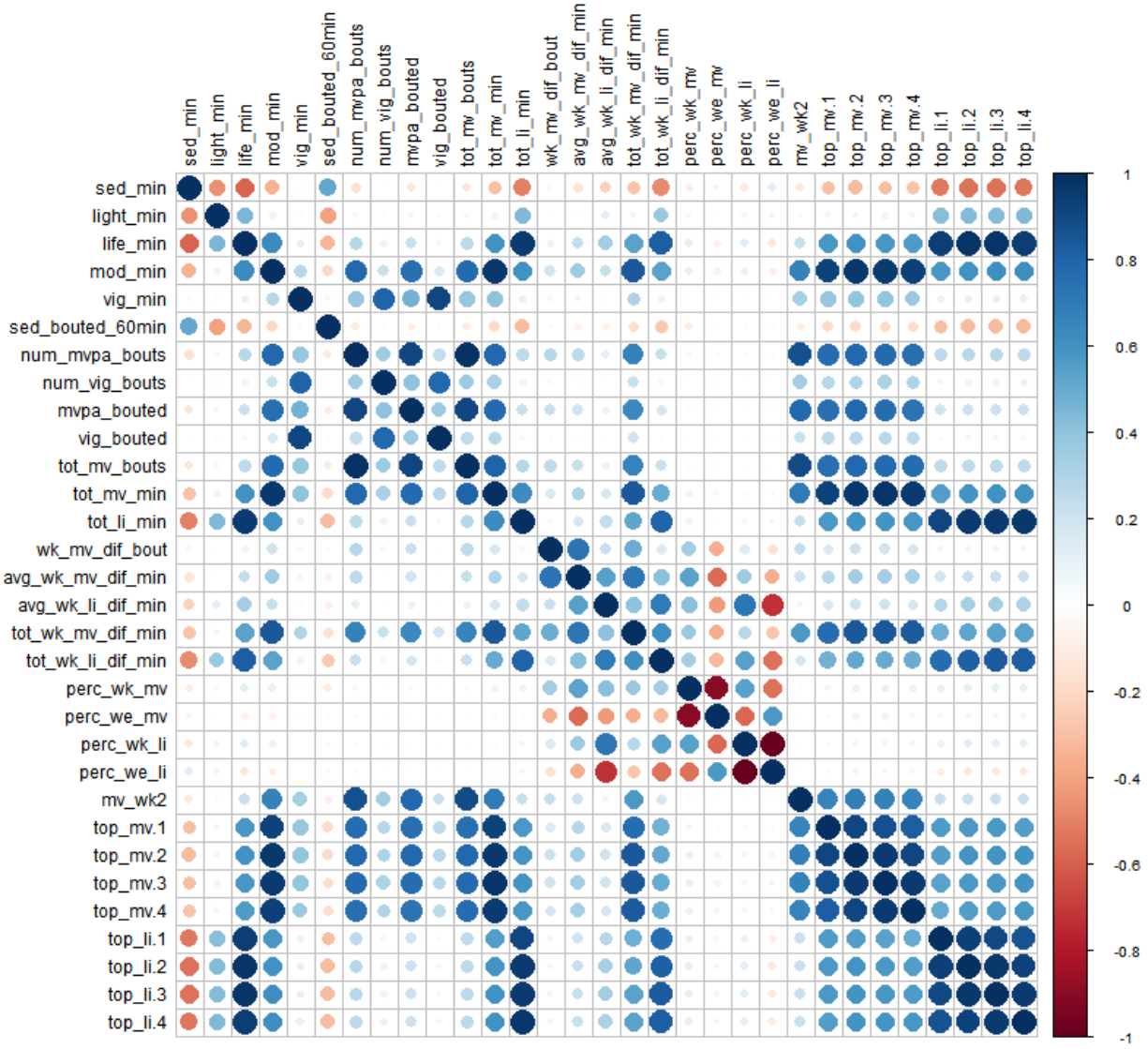


Figure 13. Correlation plot of NCI low correlation subset, with dark blue indicating strong positive correlation and dark red indicating strong negative correlation.

Figure 14. Correlation plot of NCI manually chosen variables subset, with dark blue indicating strong positive correlation and dark red indicating strong negative correlation.

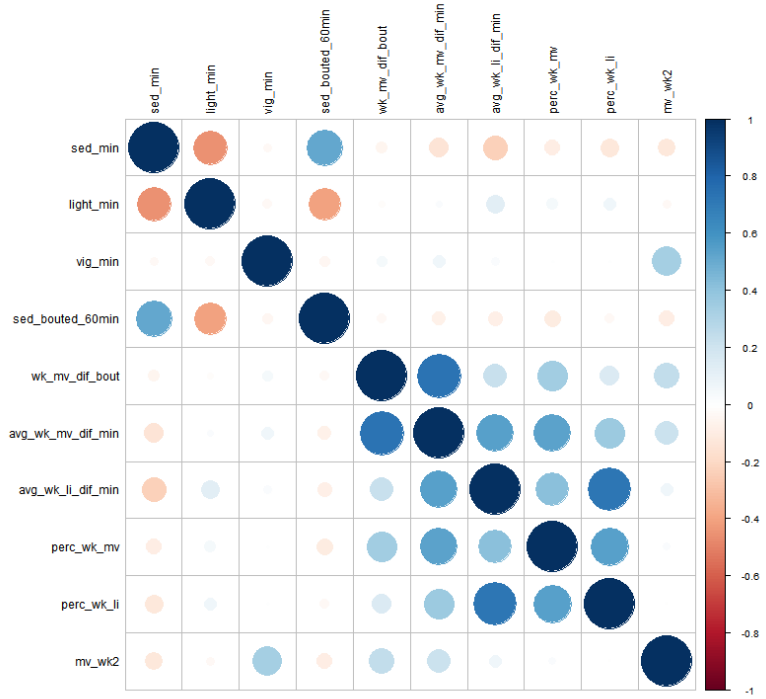


Table 18. Subsets of NCI features.

NCI features	Basic Set (N=7)	Full Set (N=32)	Low Corr (N=10)	Chosen (N=11)
gender	x	x	x	x
sed_min	x	x	x	x
light_min	x	x	x	x
life_min	x	x		
mod_min	x	x		
vig_min	x	x		x
sed_bouted_60min		x	x	x
num_mvpa_bouts		x		
num_vig_bouts		x	x	
mvpa_bouted		x		
vig_bouted		x		
tot_mv_bouts		x		
tot_mv_min		x		
tot_li_min		x		
wk_mv_dif_bout		x	x	x
avg_wk_mv_dif_min		x		x
avg_wk_li_dif_min		x		x
tot_wk_mv_dif_min		x		
tot_wk_li_dif_min		x		
perc_wk_mv		x		x
perc_we_mv		x	x	
perc_wk_li		x		x
perc_we_li		x	x	
mv_wk2		x	x	x
top_mv.1		x		
top_mv.2		x		
top_mv.3		x		
top_mv.4		x		
top_li.1		x	x	
top_li.2		x		
top_li.3		x		
top_li.4		x		

K-means features. The features derived by clustering the accelerometer readings into five categories were produced using a purely data driven technique. The feature selection process also relied on a data driven method by applying an automated high correlation filter. Unlike the NCI features, no manual feature sets were chosen. Figure 15 and Figure 16 illustrate the

correlation plots of the all features and those chosen by the filter, respectively. The full set of 30 features was reduced to 11. Table 19 summarizes the two feature sets.

Figure 15. Correlation plot of k-means full set of features, with dark blue indicating strong positive correlation and dark red indicating strong negative correlation.

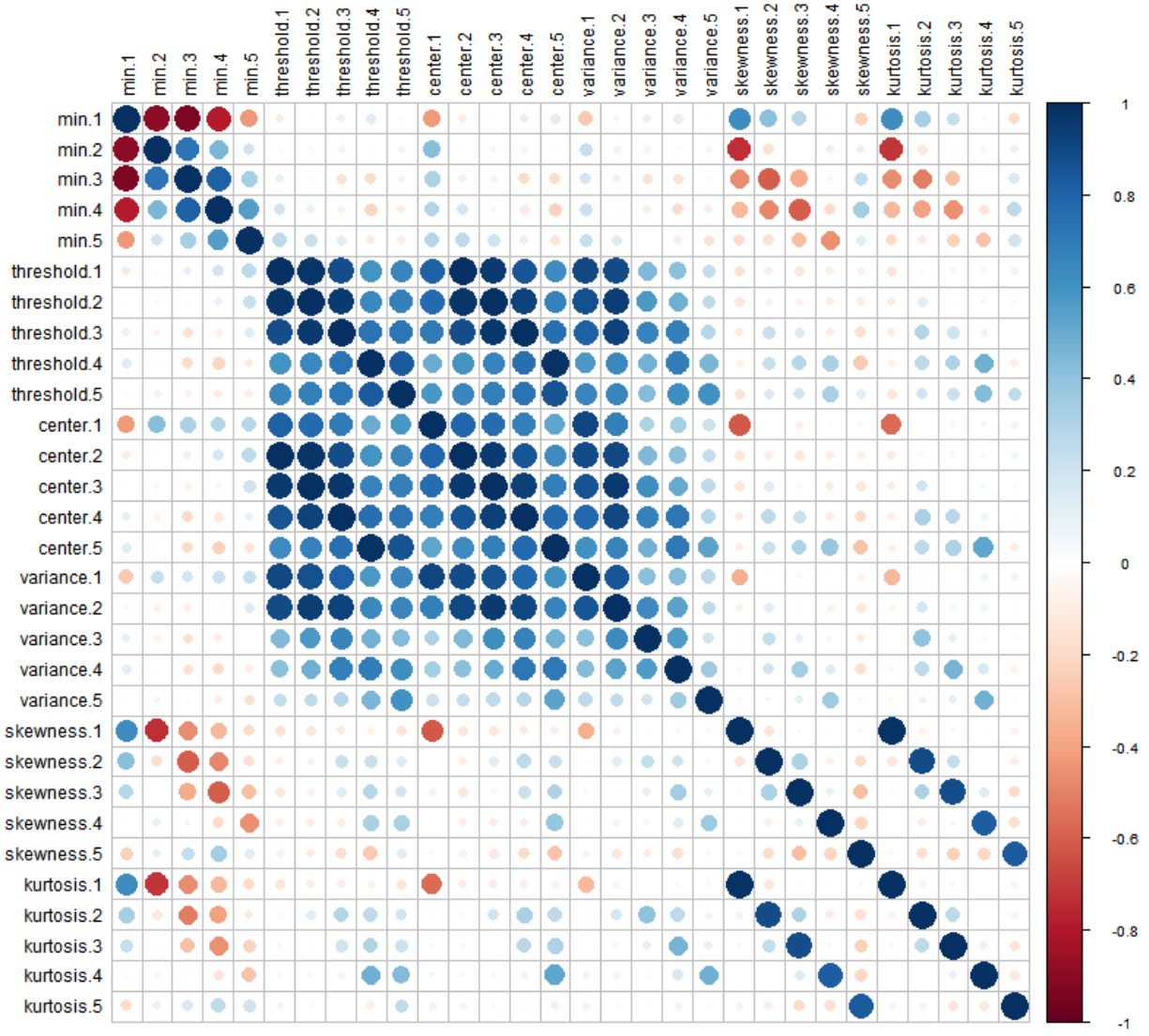


Figure 16. Correlation plot of k-means low correlation subset, with dark blue indicating strong positive correlation and dark red indicating strong negative correlation.

Table 19. Subsets of k-means features.

K-means Features	Full Set (N=31)	Low Corr (N=12)
gender	x	x
min.1	x	
min.2	x	
min.3	x	
min.4	x	
min.5	x	x
threshold.1	x	
threshold.2	x	
threshold.3	x	
threshold.4	x	
threshold.5	x	x
center.1	x	
center.2	x	x
center.3	x	
center.4	x	
center.5	x	
variance.1	x	

Table 19 (cont.)

variance.2	x	
variance.3	x	x
variance.4	x	x
variance.5	x	x
skewness.1	x	x
skewness.2	x	x
skewness.3	x	
skewness.4	x	
skewness.5	x	x
kurtosis.1	x	
kurtosis.2	x	
kurtosis.3	x	x
kurtosis.4	x	x
kurtosis.5	x	

HMM features. The matrix of all HMM features was not full rank and three variable were identified to be linear combinations of each other and were removed. Similar to the k-means feature set, the HMM set was reduced using an automated high correlation filter. The correlation plots for the full set and the low correlation subsets are presented in Figure 17 and Figure 18, respectively. Unlike the NCI and k-means feature sets, the HMM set did not have as many highly correlated variables, hence many remained in the low correlation subset. The full set of 46 variables was reduced to 36. The summary of the two HMM feature sets is presented in Table 20.

Figure 17. Correlation plot of HMM full set of features, with dark blue indicating strong positive correlation and dark red indicating strong negative correlation.

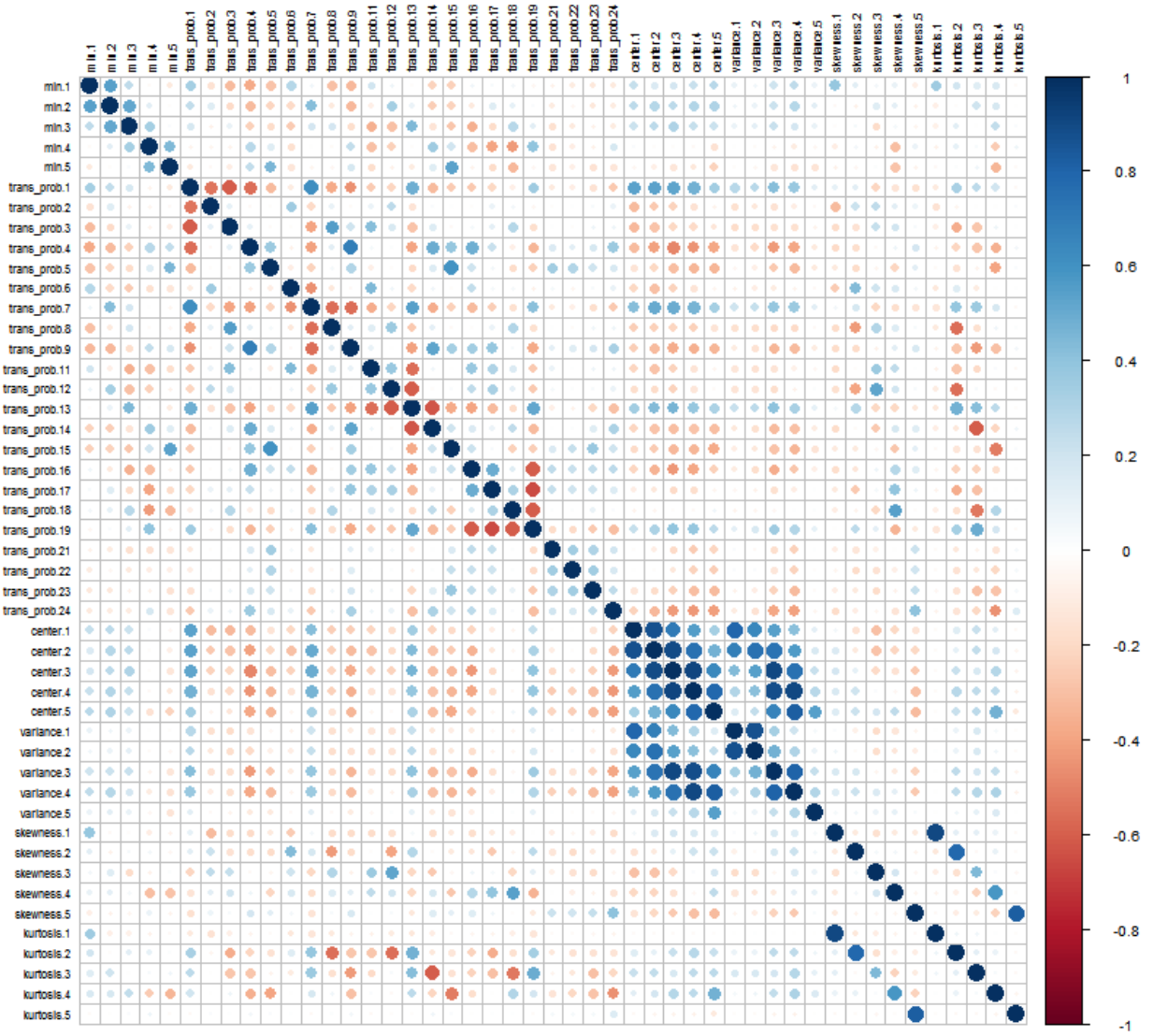


Figure 18. Correlation plot of HMM low correlation subset, with dark blue indicating strong positive correlation and dark red indicating strong negative correlation.

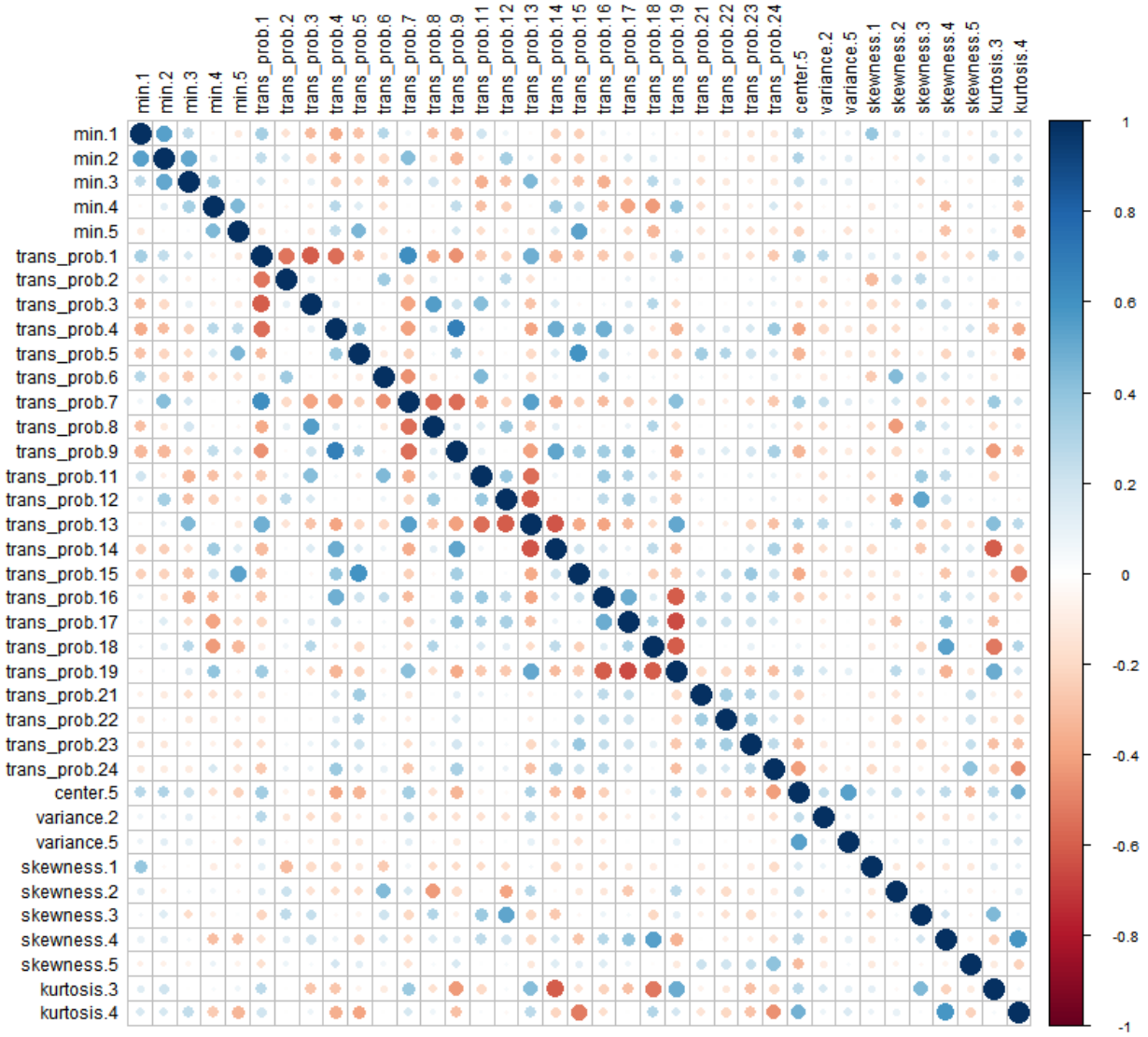


Table 20. Subsets of HMM features.

HMM features	Full Set (N=47)	Low Corr (N=37)
gender	x	x
min.1	x	x
min.2	x	x
min.3	x	x
min.4	x	x
min.5	x	x
trans_prob.1	x	x
trans_prob.2	x	x
trans_prob.3	x	x
trans_prob.4	x	x
trans_prob.5	x	x
trans_prob.6	x	x
trans_prob.7	x	x
trans_prob.8	x	x
trans_prob.9	x	x
trans_prob.10		
trans_prob.11	x	x
trans_prob.12	x	x
trans_prob.13	x	x
trans_prob.14	x	x
trans_prob.15	x	x
trans_prob.16	x	x
trans_prob.17	x	x
trans_prob.18	x	x
trans_prob.19	x	x
trans_prob.20		
trans_prob.21	x	x
trans_prob.22	x	x
trans_prob.23	x	x
trans_prob.24	x	x
trans_prob.25		
center.1	x	
center.2	x	
center.3	x	
center.4	x	
center.5	x	x
variance.1	x	
variance.2	x	x
variance.3	x	
variance.4	x	

Table 20 (cont.)

variance.5	x	x
skewness.1	x	x
skewness.2	x	x
skewness.3	x	x
skewness.4	x	x
skewness.5	x	x
kurtosis.1	x	
kurtosis.2	x	
kurtosis.3	x	x
kurtosis.4	x	x
kurtosis.5	x	

PCA. Finally, each full set of features was transformed using PCA. Because principal components are linear combinations of the variables in the original matrix that are uncorrelated with each other, the high correlation between the variables is resolved without the need to apply a filter or chose the variables manually. Figure 19, Figure 20, and Figure 21 show the scree plots for the NCI, k-means and HMM feature sets, and Table 21, Table 22, and Table 23 show the amount of variance explained by each principal component for the respective feature sets. Based on this information, 8 principal components were chosen for the NCI set, 8 for the k-means set and 19 for the HMM set to explain at least 90% of the variance in each feature set. The PCA loadings for the NCI, k-means and HMM feature sets are shown in the Appendix Table 25, Table 26, and Table 27, respectively.

Figure 19. Scree plot of NCI principal components.

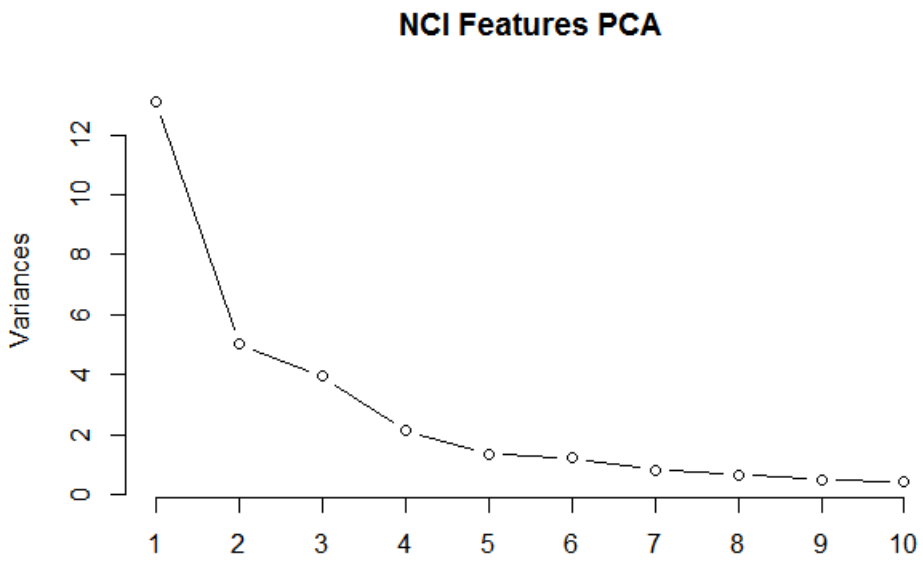


Table 21. Variance explained by NCI principal components.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Standard deviation	3.62	2.24	1.99	1.46	1.17	1.10	0.91	0.81
Proportion of Variance	0.42	0.16	0.13	0.07	0.04	0.04	0.03	0.02
Cumulative Proportion	0.42	0.59	0.71	0.78	0.83	0.86	0.89	0.91

Figure 20. Scree plot of k-means principal components.

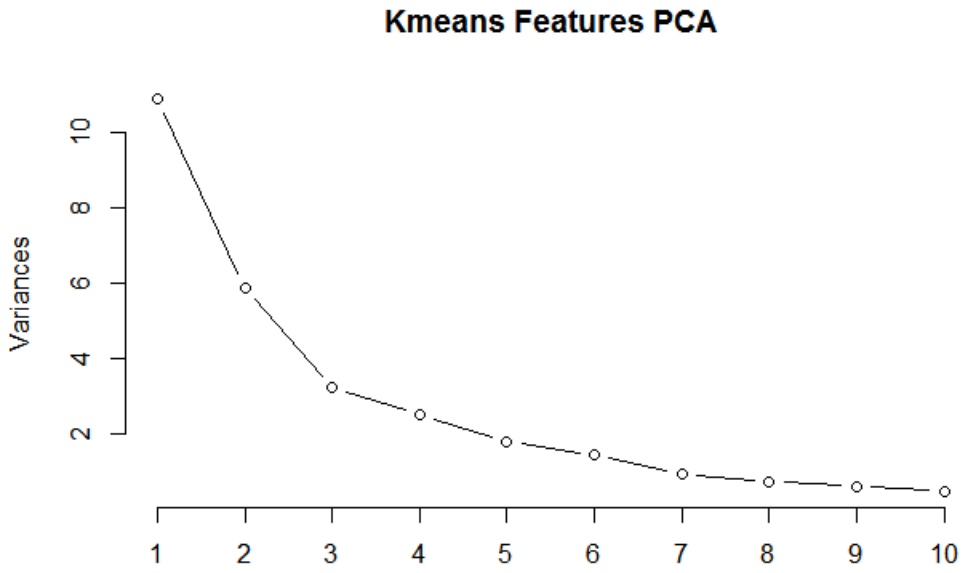


Table 22. Variance explained by k-means principal components.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Standard deviation	3.30	2.42	1.79	1.58	1.34	1.19	0.95	0.85
Proportion of Variance	0.36	0.20	0.11	0.08	0.06	0.05	0.03	0.02
Cumulative Proportion	0.36	0.56	0.67	0.75	0.81	0.86	0.89	0.91

Figure 21. Scree plot of HMM principal components.

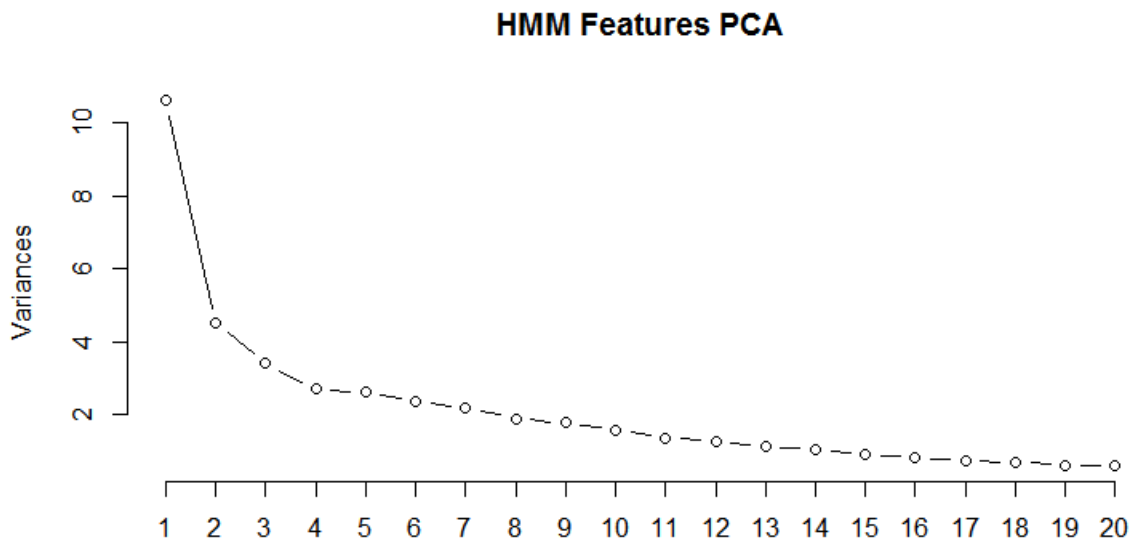


Table 23. Variance explained by HMM principal components.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Standard deviation	3.26	2.12	1.85	1.65	1.62	1.54	1.49	1.38	1.34	1.26
Proportion of Variance	0.23	0.10	0.07	0.06	0.06	0.05	0.05	0.04	0.04	0.03
Cumulative Proportion	0.23	0.32	0.39	0.45	0.51	0.56	0.60	0.65	0.68	0.72
	PC11	**PC12**	**PC13**	**PC14**	**PC15**	**PC16**	**PC17**	**PC18**	**PC19**	**PC20**
Standard deviation	1.17	1.13	1.07	1.04	0.96	0.92	0.87	0.84	0.79	0.77
Proportion of Variance	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01
Cumulative Proportion	0.75	0.77	0.80	0.82	0.84	0.86	0.87	0.89	0.90	0.92

Classification

The results of classifying participants by CVD risk based on the feature sets described above are presented in this section. The participants were classified altogether and separately by gender. The algorithms yielding high classification accuracies and interpretable results are discussed in more detail below.

Model training. The models were trained using 10-fold cross-validation where appropriate. The lambda parameter for the lasso penalty (alpha = 1) was chosen to be 0.00238332. The number of units in the hidden layer of the neural network was chosen to be 5 with a decay of 0.1. SVM with a radial basis kernel, decision tree and random forest classifiers were trained with the default parameters in the respective R packages, as these showed the best performance. Model performance using the training data is presented in APPENDIX B: CLASSIFICATION RESULTS Table 28 in the Appendix.

Overview of results. Table 24 summarizes the classification accuracy of training the classification models on the various feature sets for the data overall. For the combined data set, the classification accuracy ranged from 70.99% to 83.25%. Kappa coefficients ranged from 0 to 56.30% indicating that some classifiers were heavily biased, while others showed adequate performance. Similarly, AUC ranged from 50% to 76.19%. The feature set with the worst overall performance was the NCI features chosen by hand, followed by the NCI features derived with a high correlation filter. Surprisingly, the basic NCI features showed consistently high performance across the classifiers.

Table 24. Classification accuracy by feature set for combined data.

	Classifier	Accuracy (%)	Kappa (%)	AUC (%)	Sensitivity (%)	Specificity (%)	False Negative (%)	False Positive (%)
Basic NCI	Lasso	79.60	44.68	70.01	92.86	47.15	52.85	7.14
	Neural Network	83.25	56.30	76.19	93.02	59.35	40.65	6.98
	SVM	81.72	51.48	73.54	93.02	54.07	45.93	6.98
	Decision Tree	81.60	53.08	75.26	90.37	60.16	39.84	9.63
	Random Forest	81.13	51.88	74.28	90.03	59.35	40.65	9.97
NCI complete	Lasso	80.78	49.38	72.76	91.86	53.66	46.34	8.14
	Neural Network	78.42	40.85	68.09	92.69	43.50	56.50	7.31
	SVM	82.43	53.72	74.76	93.02	56.50	43.50	6.98
	Decision Tree	81.60	52.84	75.02	90.70	59.35	40.65	9.30
	Random Forest	81.37	51.12	75.04	92.03	55.28	44.72	7.97
NCI Low Corr	Lasso	78.30	41.25	68.49	91.86	45.12	54.88	8.14
	Neural Network	78.89	41.51	68.19	93.69	42.68	57.32	6.31
	SVM	78.77	40.94	67.86	93.85	41.87	58.13	6.15
	Decision Tree	78.66	39.02	66.58	95.35	37.80	62.20	4.65
	Random Forest	79.13	44.30	69.79	91.53	48.78	51.22	8.47
NCI chosen	Lasso	76.77	34.55	64.89	93.19	36.59	63.41	6.81
	Neural Network	72.88	14.39	55.54	96.84	14.23	85.77	3.16
	SVM	77.83	39.48	67.56	92.03	43.09	56.91	7.97
	Decision Tree	77.48	38.93	67.43	91.36	43.50	56.50	8.64
	Random Forest	79.36	44.79	70.49	91.86	48.78	51.22	8.14
NCI Prcomp	Lasso	79.13	44.30	70.15	91.53	48.78	51.22	8.47
	Neural Network	81.60	53.32	75.50	90.03	60.98	39.02	9.97
	SVM	80.90	48.82	72.12	93.02	51.22	48.78	6.98
	Decision Tree	80.54	46.81	70.79	94.02	47.56	52.44	5.98
	Random Forest	78.30	42.48	69.86	90.53	48.37	51.63	9.47
Km complete	Lasso	82.31	52.99	74.20	93.52	54.88	45.12	6.48
	Neural Network	76.53	41.40	70.13	85.38	54.88	45.12	14.62
	SVM	81.60	49.92	72.26	94.52	50.00	50.00	5.48
	Decision Tree	80.78	50.54	73.84	90.37	57.32	42.68	9.63
	Random Forest	81.84	52.85	75.27	91.69	57.72	42.28	8.31
Km low corr	Lasso	81.37	50.22	72.81	93.19	52.44	47.56	6.81
	Neural Network	71.11	0.58	50.00	100.00	0.41	99.59	0.00

Table 24 (cont.)

	SVM	82.19	51.19	72.67	95.35	50.00	50.00	4.65
	Decision Tree	80.42	50.32	74.07	89.20	58.94	41.06	10.80
	Random Forest	80.54	49.80	74.38	90.37	56.50	43.50	9.63
Km prcomp	Lasso	80.66	48.19	71.84	92.86	50.81	49.19	7.14
	Neural Network	82.90	55.66	73.36	92.36	59.76	40.24	7.64
	SVM	82.19	50.93	72.43	95.68	49.19	50.81	4.32
	Decision Tree	81.01	50.89	73.89	90.86	56.91	43.09	9.14
	Random Forest	81.49	52.11	73.40	91.20	57.72	42.28	8.80
HMM all	Lasso	81.96	51.73	73.47	93.69	53.25	46.75	6.31
	Neural Network	72.64	28.78	50.00	85.38	41.46	58.54	14.62
	SVM	81.01	49.20	72.32	93.02	51.63	48.37	6.98
	Decision Tree	78.66	40.03	67.30	94.35	40.24	59.76	5.65
	Random Forest	80.07	46.11	71.44	93.02	48.37	51.63	6.98
HMM low corr	Lasso	80.19	45.92	70.42	93.69	47.15	52.85	6.31
	Neural Network	70.99	0.00	59.87	100.00	0.00	100.00	0.00
	SVM	80.78	46.88	70.60	94.85	46.34	53.66	5.15
	Decision Tree	78.42	39.37	67.01	94.19	39.84	60.16	5.81
	Random Forest	80.78	48.02	71.03	93.52	49.59	50.41	6.48
HMM prcomp	Lasso	79.83	44.88	69.93	93.52	46.34	53.66	6.48
	Neural Network	80.19	48.56	72.76	90.53	54.88	45.12	9.47
	SVM	81.13	49.05	72.05	93.69	50.41	49.59	6.31
	Decision Tree	73.58	26.10	61.32	90.53	32.11	67.89	9.47
	Random Forest	77.71	35.22	65.85	95.68	33.74	66.26	4.32
All	Lasso	82.78	54.48	75.01	93.52	56.50	43.50	6.48
	Neural Network	74.17	38.55	50.73	80.40	58.94	41.06	19.60
	SVM	82.19	52.23	73.64	94.02	53.25	46.75	5.98
	Decision Tree	80.42	50.07	73.83	89.53	58.13	41.87	10.47
	Random Forest	82.19	53.10	74.03	92.86	56.10	43.90	7.14
All Low corr	Lasso	82.43	52.86	73.92	94.19	53.66	46.34	5.81
	Neural Network	70.99	0.34	54.31	99.83	0.41	99.59	0.17
	SVM	81.49	50.86	73.26	92.86	53.66	46.34	7.14
	Decision Tree	80.66	50.30	73.76	90.20	57.32	42.68	9.80
	Random Forest	81.25	50.10	73.35	92.86	52.85	47.15	7.14

Models by gender. Tables 20 and 21 in the appendix summarize the classification results for women and men separately. At first glance it appeared that models based on women's data had higher classification accuracy (80.30%-85.35%) than men's (67.04%-81.42%), however the kappa coefficients told a different story. Men's kappa coefficients ranged from 17.56% to 59.75% while women's kappa ranged from 1.48% to 45.64%. As discussed earlier, there was a difference in the proportion of men and women with high CVD risk. This is where classification accuracy is misleading while the kappa coefficient and AUC account for bias.

It appears that the combined data was able to achieve more consistent results when gender was added as a control variable to the models. No model for women was able to achieve greater than 46% kappa coefficient or greater than 70% AUC. Thus, it appears that modeling CVD risk separately by gender does not lead to a final model that is as accurate as one that may be achieved with the combined data set.

Alternate outcomes. The results of individual CVD risk factors as outcomes are presented in the appendix Table 31 through Table 34. None of the classifiers were able to identify any patterns for the individual risk factors. Total cholesterol, HDL and cRP had false positive rates of 100% for all combinations of classifiers and feature sets. Blood pressure did have a few classifiers with nonzero kappa coefficients, however the highest kappa coefficient was 19.24% indicating very poor performance overall.

Lasso regression. Overall the lasso regression performed well with classification accuracy as high as 82.78% and kappa coefficient of 54.48% for the full feature set combining NCI, k-means, and HMM derived variables (N = 108). The performance of the lasso regression seems to improve as number of features increases. The lasso regression provides additional

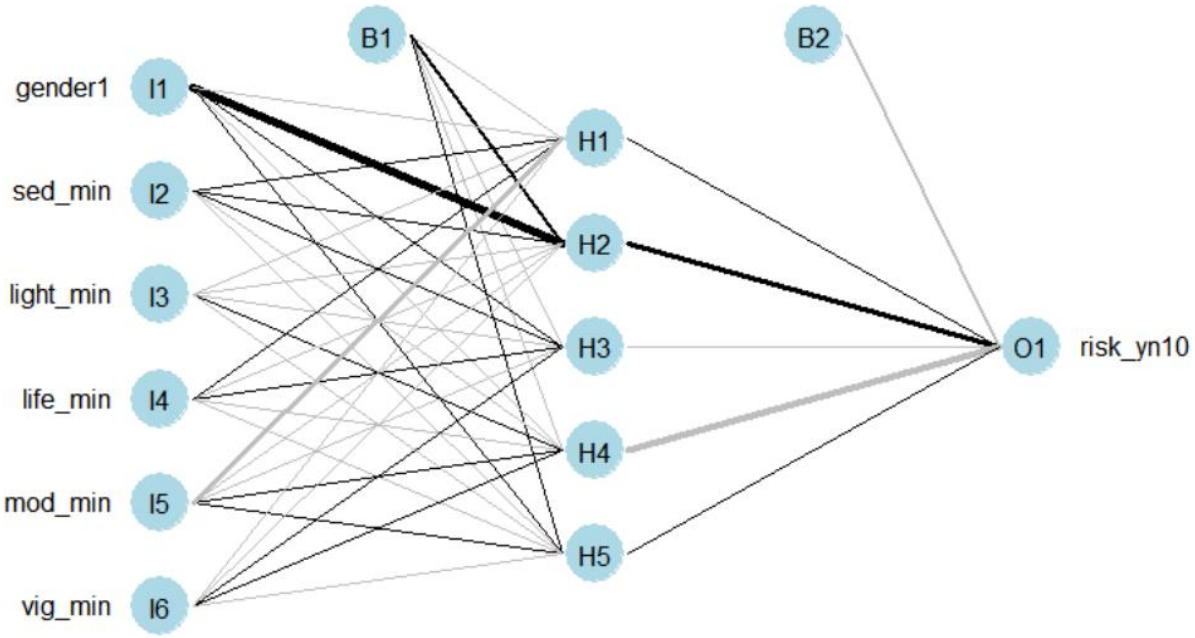
information for variable selection. Lasso coefficients are presented in the appendix APPENDIX C: FEATURE IMPORTANCE MEASURES

Table 35 through Table 43 for each feature set the model was trained on. The variables whose coefficients are equal to zero are not used for classification, and thus may be interpreted as unimportant to the model. However, in the presence of correlation between variables, the lasso choses only one feature that is correlated with the others. Therefore, a zero coefficient may be misleading in the larger features sets where several variables are all important, but are highly correlated with each other.

Neural Networks. In contrast to the lasso regression, neural networks performed worse on the full features sets and better on smaller subsets. In fact, the highest overall classification accuracy of 83.25% was achieved by a neural network using only the basic NCI features (N = 6). Interestingly, the lowest classification accuracy with a kappa coefficient of 0 was also achieved by a neural network using HMM low correlation subset (N = 38). That is, all cases were classified as low risk, resulting in a 100% false negative rate. This indicates that neural networks don't perform well on highly dimensional data sets and tend to bias the classification against the underrepresented class.

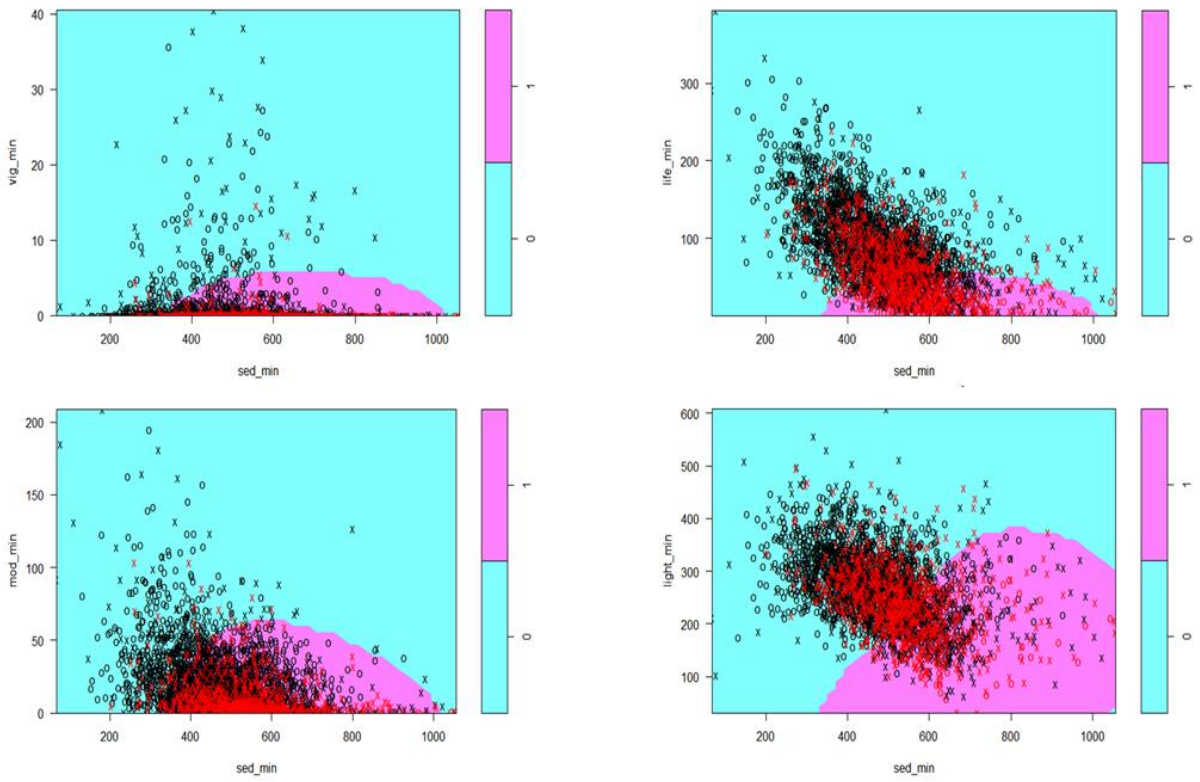
Figure 22 shows a visual representation of the neural network trained on the basic NCI features. Five hidden layers were created as combinations of the six input features. The hidden layers are labeled as H1 to H5 and the bias factors as B1 and B2. While the classifier arrived at a high classification accuracy, kappa coefficient and AUC, the model yielded itself to little further interpretation.

Figure 22. Neural network using NCI basic features with the inputs indicated on the left, the hidden layers shown as H1 – H5, the bias terms B1 and B2, and the outcome measure on the right. The line thickness represents the weight of each feature contributing to each hidden layer.



SVM. SVM performed best for the full NCI feature set with classification accuracy of 82.43% and kappa coefficient of 53.72%, though the classifier showed consistently good performance for all feature sets. However, similar to neural networks, the results from SVM were not straightforward to interpret. Figure 23 illustrates the SVM plot comparing two NCI features at a time. The support vectors are labeled "x" and the decision boundary for low and high CVD risk is shown in contrasting colors. It appears that the contrast between sedentary and vigorous, and sedentary and moderate minutes yielded to the clearest distinction between high and low risk participants.

Figure 23. SVM classification plots for pairs of the NCI basic features. The support vectors are labeled "x" and high risk observations are in red. The decision boundary for high CVD risk is shown in purple.



Decision Trees. Decision trees performed consistently well for all feature sets, but the classification accuracy of 81.60% with a kappa coefficient of 53.10% and AUC of 75.42% was highest for the basic NCI feature set. The decision tree for the combined data using the basic NCI features is presented in Figure 24. Unlike the previously discussed models, in addition to yielding a classification by high and low CVD risk, decision trees provide additional insight into how the classifier arrived at the decision.

The first partition was chosen at greater than or equal to 37 minutes spent in lifestyle PA category. The CVD risk scores resulting from this partition are illustrated in Figure 25 where the separation between the high and low risk participants becomes evident.

Next, both branches were partitioned by gender. Two terminal nodes were reached at this point. If a participant got at least 37 minutes of lifestyle PA on average per day and was female,

she was classified as low risk of CVD. However, if the participant did not get the specified number of minutes and was a man, he was considered high risk.

Women who did not get at least 37 minutes of lifestyle PA were partitioned by lifestyle minutes again, this time greater than or equal to 26. Thus, if a woman got at least 26 lifestyle minutes per day on average, she was classified as low risk. If not, the next partition used the threshold of less than 547 sedentary minutes per day. Thus, finally, women who got less than about 9 hours of sedentary time were considered low risk.

Men who did get at least 37 minutes of lifestyle PA did not arrive at a terminal node. This group was partitioned by 12 minutes or greater of moderate PA per day on average. Those who met these criteria were classified as low risk. Those who did not have at least 12 minutes of moderate activity were partitioned further by at least 7.4 moderate minutes. Those who did not meet either requirement were considered high risk. In short, it appears that lifestyle minutes were sufficient for women to be considered at a low risk of CVD while a combination of lifestyle and moderate intensity minutes is required for men.

Figure 24. Decision tree with basic NCI features.

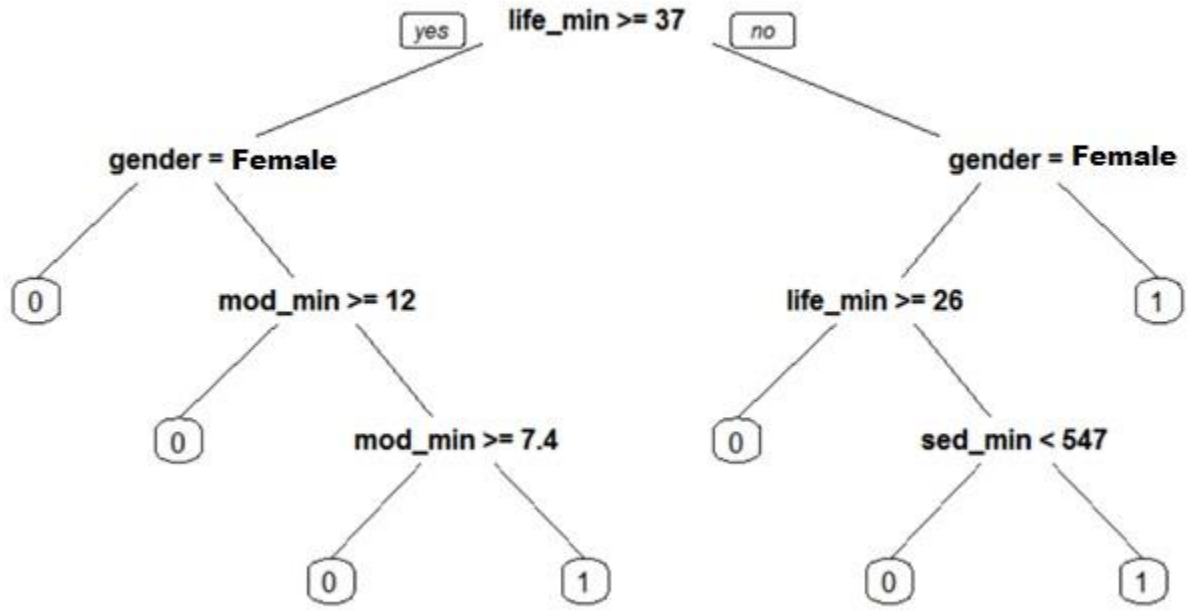
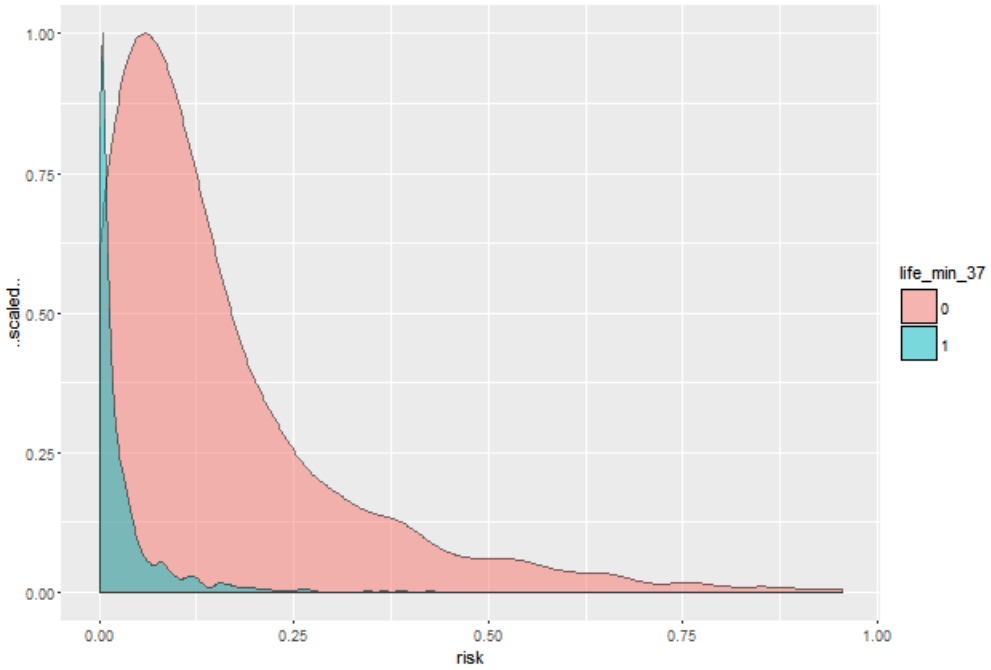


Figure 25. Separation of CVD risk scores by first decision tree split.

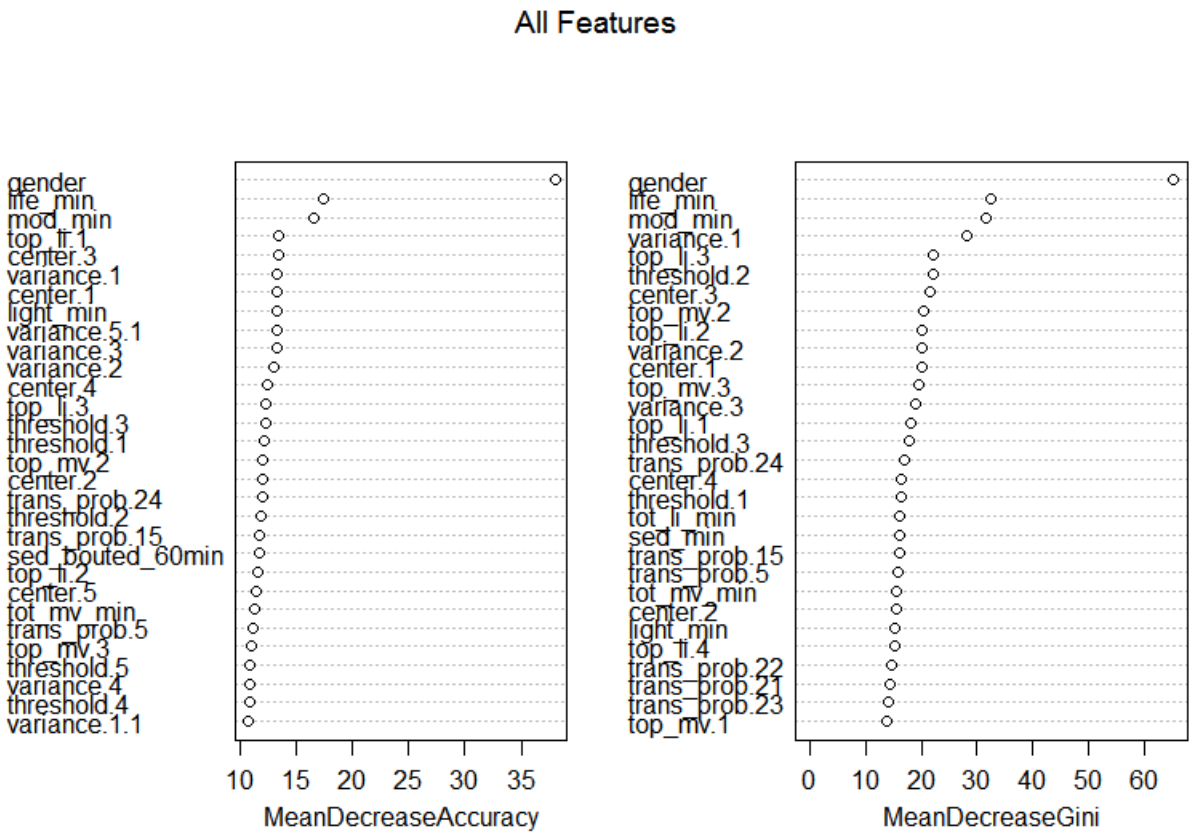


Random Forests. Random forest classifiers performed well for the full feature sets, achieving the highest classification accuracy of 82.19% and kappa coefficient of 53.10% when trained on the set combining NCI, k-means, and HMM derived features. It appears that similar to the lasso regression and SVM, RF performs best on highly dimensional data.

In addition to classification, random forest classifiers provide information for feature selection. Figure 26 shows a visual representation of mean decrease in accuracy and Gini index for all features combined. APPENDIX C: FEATURE IMPORTANCE MEASURES

Table 35 through Table 43 and Figure 27 through Figure 33 in the appendix show the mean decrease in accuracy and Gini index for the rest of the feature sets used to train the RF classifier. While gender appears to be the most important predictor of CVD risk, lifestyle minutes, followed by moderate PA minutes are still considered the most informative.

Figure 26. Mean decrease in accuracy and Gini index for all features based on the random forest classifier.



Summary of results. All classifiers achieved at least 80% classification accuracy, at least a 50% kappa coefficient, and at least 70% AUC on different feature sets. However, some classifiers were more consistent than others. Neural networks did not perform well on larger feature sets and yielded a very high false negative rate when trained on k-means, HMM and combined low correlation feature subsets. The other classifiers provided consistently good results across the feature sets with only slight variations in accuracy.

CHAPTER 5 DISCUSSION

The results of the study are discussed in this chapter. First, the major findings are summarized in the context of the specific aims of this study. The appropriateness of the Reynolds risk score as and outcome measure and the interpretability of various machine learning algorithms are discussed. Next, findings are evaluated in terms of implications for PA recommendations. Finally, limitations and suggestions for further study are described.

Major Findings

The main goal of this study was to explore the possibility of using accelerometer data as a proxy for the Reynolds risk score to predict CVD health status. Utilizing a variety of machine learning techniques to obtain relevant features from the data and train several classification algorithms, this study showed that accelerometer data alone may be used to classify individuals into high and low CVD risk status. Applying these techniques to accelerometer data can yield a better understanding of the relationship between PA and CVD patterns.

Accelerometer data as an input. This study attempted to connect intensity readings from accelerometer data recoded over the course of a week to a Reynolds risk score. Using accelerometer data in health research offers several benefits. First, the data is easy and inexpensive to collect using a variety of devices, including cell phones, hip and wrist monitors. Next, the data is a direct measure of PA that offers insights into the rich patterns of individual activity. This study attempted to assess several methods of processing accelerometer data to retaining maximum information necessary to predict CVD risk.

Several methods to extract and select appropriate features from accelerometer data were used. The NCI features focused on using already established PA thresholds to calculate various PA pattern measures. Additionally, two clustering approaches, k-means and HMM, attempted to

use data-driven techniques to identify relevant features. PCA was also applied to the feature sets for comparison.

The basic set of NCI features consisting of time spent in five established PA categories proved surprisingly successful in terms of classification accuracy. This feature set consisted of the smallest number of variables, and all five classifiers trained on this feature set performed very well. At the same time, a manually chosen subset of NCI features that contained only a few more variables showed the worst performance across the classification algorithms. It appears that the basic NCI set of features avoided overfitting, yet contained the right amount of information to achieve good classification results. Furthermore, using the feature set yielded interpretable results that carry clinical significance as discussed below.

Data driven approaches also yielded good results for most classifiers. Both k-means and HMM clustered the accelerometer data into five categories individually, with features describing the distributions of each cluster for each participant. These techniques reinterpreted the PA activity thresholds, recognizing individual patterns in accelerometer data. These features provided a closer look at relative, rather than absolute PA levels and still led to good classification results.

K-means partitioned the data into clusters by minimizing the distance of the observations to the center (mean) of each cluster. This approach effectively established the PA thresholds for each individual. HMM relies on sequential temporal information to fit a distribution for each specified state. Thus, transition probabilities between each category in addition to the means and variances of each category were computed. This provided further information into the PA patterns of the individual.

Reducing the feature sets using a high correlation filter did not appear to improve classification. Furthermore, deleting variables that are highly correlated with each other in an automated way may have lost some meaningful features along the way. Likewise, features obtained using PCA did not appear to yield better results than the original feature sets. Principal components are combinations of variables that are orthogonal (uncorrelated) to each other. These linear combinations of variables form a black box and the carefully chosen features lose meaning and interpretability. PCA is a good option for dimensionality reduction, however the interpretability of the model is lost and classification accuracy is not improved.

Reynolds risk score as an outcome. This study examined the Reynolds risk score as an outcome measure of cardiovascular health status. The Reynolds risk score was developed based on a large longitudinal study to predict 10-year CVD risk. Therefore, in the likely event that a longitudinal study is not feasible, the Reynolds risk score may be used to study the effects of PA on CVD risk. The measure already accounts for various CVD risk factors and demographic variables. Using the Reynolds risk score as an outcome of a study that uses accelerometers to measure PA allows researchers to focus solely on identifying appropriate features from the accelerometer data.

It must be noted that since the Reynolds risk score was developed separately and is calculated differently for men and women, gender must be used as a control variable, or models should be trained separately by gender. Nevertheless, in this study, some models achieved classification accuracy of nearly 83% and a kappa coefficient of 56% using only accelerometer data as features. This finding is a good indicator that accelerometer data may be connected to the Reynolds risk score and may be interpreted as a proxy for CVD risk.

Using the individual components of the Reynolds risk score as outcomes would significantly increase the number of participants in the study since not as many laboratory measures would be required. However, this study showed that individual CVD risk factors, including BP, TC, HDL, and cRP, are not sufficient to act as outcomes for classifying accelerometer determined PA. While none of the outcomes produced good classification results, BP showed the strongest relationship with the Reynolds risk score. However, the Reynolds risk score controls for a variety of factors including age and gender in addition to several CVD risk factors. Also, since some of the risk factors may be controlled by medication, the full picture of an individual's cardiovascular health may not be captured by a single risk factor.

Machine learning algorithms. Several machine learning classifiers were trained and compared for classification of accelerometer-derived PA by Reynolds risk score. While all classifiers achieved good classification accuracy on different combinations of features, some appeared to have more consistent and interpretable results.

The penalized logistic regression performed very well, particularly on the larger feature sets, achieving 82% classification accuracy with a 53% kappa coefficient for all features combined. During model training, the lasso regression also performed feature selection due to the lasso penalty. The coefficients of some features were shrunken to zero and not used by the model for classification. While this may be useful for feature selection, because the features are correlated, the choice of variables is not unique and should therefore be used with caution. Overall, this approach is useful for large feature sets to prevent overfitting and yields interpretable results.

Neural networks provided inconsistent and at times biased results across the feature sets. The highest classification accuracy of 83% and kappa coefficient of 56% was achieved on the

smallest feature set consisting of time spent in each PA category. The classifier performed similarly well using principal components of the feature sets. For larger, correlated feature sets, neural networks defaulted to classifying everyone as low risk and yielded a kappa coefficient equal or near zero. Furthermore, aside for classification, the algorithm did not provide an interpretable result. These findings indicate that neural networks may not be an appropriate choice when studying accelerometer PA data and its effects on CVD risk.

SVM performed well for most feature sets, yielding consistent results. However, similar to neural networks, much of the meaning and interpretability was lost in the model. Features could only be visualized in pairs to understand the derived decision boundaries. While this may work with small feature sets, this is not desirable when accelerometer data is being studied with the intension to identify meaningful variables.

The decision tree models, on the other hand, yielded the most interpretable results. When the decision tree was applied to just the basic NCI features, only one split was needed to yield high classification accuracy of 78%. The split indicated that those who spent at least 37 minutes in lifestyle activity on average throughout the week are likely at a low risk for CVD. Overall, the decision trees achieved high classification accuracy on par with the other algorithms while providing the exact path of how the classification was achieved along the way. This transparency may be very useful for better understanding PA patterns and how they affect CVD risk.

It must be noted that decision trees are weak learners and are easily influenced by changes in the sample data. Furthermore, the variable for each split is chosen based on maximum information and therefore may overlook some equally meaningful but slightly less informative features. Therefore, decision tree partitions are highly useful for understanding the path of classification, but should be interpreted with caution.

Random forest classifiers, did not outperform decision trees when trained on the NCI feature sets, and only did slightly better for the data-driven methods. In addition to classifying the data, the algorithm does provide insight into the importance of features. However, as with the lasso regression, the ranking is influenced by highly correlated variables and should be interpreted with caution. Overall, the random forest classifier added a layer of unnecessary complexity without providing improved classification accuracy, at least for this sample.

Classification. The individuals in this study were classified as high or low risk based on their Reynolds risk score. All classifiers achieved good classification accuracy for some combination of features. However, the results show that very high classification accuracy is achieved when identifying low risk individuals, and much lower accuracy is achieved for those at high risk. In other words, low risk is identified with less than 10% error by most algorithms, while high risk is misclassifies at least 40% of the cases. This finding indicates that the PA patterns of high risk individuals are rather varied while those who are low risk have more in common.

PA Recommendations

A major benefit of connecting accelerometer intensity readings to a Reynolds risk score is the opportunity to study the PA patterns that lead to high or low CVD risk. Better understanding of these patterns has direct implications for establishing appropriate PA recommendations to reduce CVD risk. This study attempted to assess a few different methods of processing and analyzing accelerometer data to achieve an interpretable result.

Interpretability of results. Neural networks and SVM trained on certain feature sets provided good results, however did not provide additional meaning that could be used to establish PA recommendations. Other models, the lasso regression, random forests, and decision

trees, provided further interpretation of the input features and their effect on classification. For example, random forests that were trained on feature sets that included lifestyle minutes as an input showed that it was one of the most important features for classifying individuals by CVD risk aside from gender. Further, the decision tree provided a completely transparent result that could eventually be used for establishing recommendations for lifestyle and moderate intensity PA meeting.

Feature selected as inputs for the various models also play a major role for interpretability. NCI features provide the most relatable results as time spent in PA intensity categories is often used in the field. For example, the decision tree suggests an average of 37 daily minutes of lifestyle PA as a recommendation to reduce PA risk. The current PA guidelines also focus on time spent in the intensity categories and may be directly compared to the finding based on the NCI feature set (United States Department of Health and Human Services., 2008).

Data driven methods provide the benefit of capturing individual patterns and identifying novel representations of PA. The k-means and HMM based feature sets used in this study provided classification results on par with the NCI features, though did not outperform them. The data-driven features, however, did provide new insights into the relationship between PA and CVD risk. For example, measures of variance in the established intensity category clusters were ranked as highly important by the random forest classifier. This feature may be interpreted as the spread within the category, with more variability leading to lower CVD risk. Using unsupervised learning techniques to study accelerometer data may be useful to gain novel insights into setting future PA recommendations.

It must be noted that attempts to reduce the multicollinearity of the features using high correlation filters and PCA did not improve classification accuracy and instead lead to reduced

interpretability of the models. Principal components applied to large feature sets work well for dimensionality reduction, but lose the meaning of the features along the way. The three classifiers deemed most interpretable for this research have built in mechanisms that deal with highly correlated features. The lasso penalty shrinks the coefficients of unimportant features to zero, without being swayed by multicollinerity like the logistic regression. Decision trees and random forests are nonparametric and focus on maximum information of the feature being used.

The results of these classifiers must be interpreted with caution in the presence of highly correlated features; however, they do allow for a large number of features to be studied and evaluated without reduction in classification accuracy, as is the case with neural networks. For the purposes of establishing PA recommendations, unsupervised learning techniques show promise for identifying novel PA patterns, especially on an individual basis.

Gender differences. In recent years, studies have shown that there is a gender difference in the way CVD manifests itself (Galiuto & Locorotondo, 2015). Men's risk of CVD increases linearly with age, while women's CVD risk increases with menopause. These differences indicate that estrogen may have cardio-protective benefits (Barrett-Connor & Bush, 1991). Thus, the mechanism that affects men and women appears to be different and is reflected by the Reynold's risk score, which was developed separately and contains different sets of predictors by gender.

In addition to physiological gender differences, studies show that women tend to engage in less PA than men and that the amount of PA decreases with age (Cooper et al., 2000; Evenson et al., 2002; Trost et al., 2002). The results of this study show that gender effects are indeed present in the accelerometer data with the Reynolds risk score as the outcome. Men and women

engage in different amounts of PA and have different levels of CVD risk. Specifically, women engage in less PA and exhibit less CVD risk than men.

As the analysis shows, gender appears to be the most important input for classification for all feature sets. The decision tree trained on NCI features shows that there is a different mechanism for classifying men and women into high and low CVD risk categories. Specifically, an average of at least 37 minutes of lifestyle activity a day appears to be sufficient to identify women who are at low risk of CVD and men who are at high risk. For men to be considered low risk in the sample, at least 12 minutes of moderate PA is also required. These findings suggest that the mechanism of CVD affects men and women differently, and it may be necessary to have gender specific recommendations (Cooper et al., 2000; Zisko et al., 2015).

PA thresholds. Several studies have described the importance of relative PA thresholds as individuals of different fitness levels exhibit very different PA patterns that effect accelerometer readings (Alhassan & Robinson, 2010; Ozemek, Cochran, Strath, Byun, & Kaminsky, 2013). Unlike the NCI features, data driven methods can help identify relative PA patterns and thresholds for the activity categories. K-means clustering established PA category thresholds separately for each participant. The thresholds for the first cluster, ones that correspond to the lowest intensity category ranged from 40 to 806. This finding alone indicates that the participants in the study ranged widely in terms of activity. The average thresholds for the five categories across the sample were as follows: <198; 198- 649; 650-1349; 1350-2745; ≥2745.

Using a data driven approach to establish the PA intensity thresholds would significantly lower the moderate and vigorous cut offs, while increasing the one for light PA. While this finding is sample specific, it is in line with previous research suggesting the need to redefine the

PA intensity cut points (Matthews, 2005; Ozemek et al., 2013). Applying these clustering techniques may be used to obtain clinically significant results that will carry over into establishing new thresholds of PA intensity categories.

PA patterns. To evaluate the importance of how PA is spread throughout the day and week, several features measuring these patterns were included in the NCI feature set. When the basic NCI features were included as inputs, the other representations of PA patterns weren't as important according to the random forest classifiers' mean decrease in accuracy and Gini index. For data driven feature extraction methods, PA patterns are described by skewness and kurtosis, as well as transition probabilities between the states identified by HMM. These features did not appear to be more important for determining CVD risk when combined with lifestyle minutes.

These findings indicate that the amount of PA, not how and when it is performed is important for CVD risk classification. Other studies looking into PA patterns show similar results. Evenson et al. (2015) applied latent component analysis (LCA) to the NHANES accelerometer data from 2003-2006 identified some common PA patterns that include individuals who are generally sedentary or moderately active, those who are primarily active on the weekends, and those who are primarily active on the weekdays. However, Lee, Sesso, Oguma, and Paffenbarger (2004) showed that some amount of PA, whether performed on weekends or spread throughout the week carries cardio protective benefits that help reduce CVD mortality risk. Additionally, J. Myers et al. (2004) showed that fitness level, not PA patterns were most important for predicting all-cause mortality in men.

Limitations

Some limitations that should be considered when interpreting the results of this study are outlined in this section. First, the NHANES accelerometer data used in this study was collected using a uniaxial ActiGraph AM-7164 monitor. This device is thought to have the most accurate representation of walking, running, and other activities with an up and down motion. Therefore, activities such as biking or rowing may not be accurately reflected in the data. It may be inferred that such activities would still be recorded by the accelerometer and would likely end up in the lifestyle intensity category.

Next, only aggregated rather than raw data was available and used in this study. The intensity was averaged over each minute and therefore some intense bouts of activity may have been averaged out. An example is climbing the stairs when an intense reading lasts only a few seconds followed by a moment of rest. Since the activity does not last an entire minute at a vigorous intensity, stair climbing may present as a lower intensity activity in the data. Therefore, this effort may not be accurately captured by the data used in this study. The readings for this type of activity likely end up in lifestyle minutes as well. Unsurprisingly, lifestyle minutes appear to be the best indicators of CVD risk.

Overall, the sample size used for this study was relatively small. The protocol allowed for nonwear time, so participants simply took the device of for sleep and forgot to put it back on, thus lowering the number of available data points. Using the Reynolds risk score further diminished the sample by excluding participants under the age of 30 and those with incomplete data. Machine learning techniques are data driven and therefore work better on larger data sets. To better study the nuances of daily PA patterns, a larger sample of participants and raw accelerometer data should be used.

Future Studies

The main benefit of machine learning algorithms is that the models are constantly updated and improved as more data becomes available. In other words, higher accuracy will be achieved with more data. Accelerometer data perfectly lends itself to be analyzed using these data driven techniques. Large volumes of data are easy to collect and provide detailed information about individual PA patterns.

Future studies should focus on large scale data collection, leveraging personal devices of the participants. An app can record the accelerometer data of participating individuals using their personal devices and provide large amounts of raw, objectively measured PA data for research. This data can be used to study the effects of PA patterns on CVD risk using the Reynolds risk score, eventually providing real time feedback to the participants regarding CVD risk status and PA recommendations. Additional devices that include a heart rate monitor and other relevant sensors may be included for more accuracy.

With the availability of detailed accelerometer PA measures and machine learning techniques for analysis, the approach to PA recommendations may shift toward individualized feedback. Research shows that activity thresholds for different fitness levels should be established separately, and difference in PA levels between men and women have long been observed (Alhassan & Robinson, 2010; Cooper et al., 2000; Ozemek et al., 2013). By using personal devices that are collecting data and training classification algorithms in real-time, each person may be provide a personal set of PA recommendations.

Conclusion

Accelerometer data can be analyzed with machine learning techniques to act as a proxy for the Reynolds risk score to predict CVD health status. High classification accuracy was

achieved by all classifiers with some performing slightly better than others when using different features sets. In general, lasso regression, SVM and RF all performed well on large feature sets that included data-driven features, achieving greater than 82% classification accuracy when the NCI, k-means and HMM features were combined. Neural networks performed well on smaller uncorrelated feature sets, and decision trees produced consistent results with the most transparency.

Overall, PA recommendations that may be derived from this study indicate that at least 37 minutes of daily lifestyle activity are key for reducing CVD risk. This finding is supported by the decision tree classifier whenever this input feature is included for classification. Furthermore, the random forest classifier indicates that lifestyle minutes are second only to the indicator of gender as the most important variable for determining CVD risk.

The approaches discussed in this study may be used to better understand the effects of PA patterns on CVD risk and can eventually lead to individualized PA recommendations. Ultimately, training machine learning algorithms on large volumes of accelerometer data collected by the study participants themselves will change the way PA is measured and analyzed in the field of Kinesiology and Public Health, with a greater focus on individualized feedback provided in real-time.

REFERENCES

Ainsworth, B. E., & Coleman, K. (2006). Physical activity measurement. In A. McTiernan (Ed.), *Cancer Prevention and Management Through Exercise and Weight Control* (pp. 13–24). Boca Raton, FL: Taylor & Francis.

Alhassan, S., & Robinson, T. N. (2010). Defining accelerometer thresholds for physical activity in girls using ROC analysis. *Journal of Physical Activity & Health*, *7*(1), 45–53. doi:10.1016/j.biotechadv.2011.08.021.Secreted

Ambrose, J. A., & Barua, R. S. (2004). The pathophysiology of cigarette smoking and cardiovascular disease: an update. *Journal of the American College of Cardiology*, *43*(10), 1731–7. doi:10.1016/j.jacc.2003.12.047

American Diabetes Association. (2010). *Standards of medical care in diabetes—2010*. *Diabetes care* (Vol. 33). doi:10.2337/dc10-S004

Atienza, A. a, Moser, R. P., Perna, F., Dodd, K., Ballard-Barbash, R., Troiano, R. P., & Berrigan, D. (2011). Self-reported and objectively measured activity related to biomarkers using NHANES. *Medicine and Science in Sports and Exercise*, *43*(5), 815–21. doi:10.1249/MSS.0b013e3181fdfc32

Baek, J., Lee, G., Park, W., & Yun, B. (2004). Accelerometer signal processing for user activity detection. In M. G. Negoita, R. J. Howlett, & L. C. Jain (Eds.), *Knowledge-Based Intelligent Information and Engineering Systems* (Vol. 3215, pp. 610–617). Springer Berlin Heidelberg. doi:10.1007/978-3-540-30134-9_82

Bai, J., He, B., Shou, H., Zipunnikov, V., Glass, T. A., & Crainiceanu, C. M. (2014). Normalization and extraction of interpretable metrics from raw accelerometry data. *Biostatistics (Oxford, England)*, *15*(1), 102–16. doi:10.1093/biostatistics/kxt029

Bao, L. (2003). *Physical activity recognition from acceleration data under semi-naturalistic conditions*. Massachusetts Institute of Technology. Retrieved from http://18.181.0.31/afs/athena/dept/cron/group/house_n/documents/Bao03.pdf

Bao, L., & Intille, S. S. (2004). Activity recognition from user-annotated acceleration data. In A. Ferscha & F. Mattern (Eds.), *Pervasive Computing: Second International Conference* (pp. 1–17). Springer Berlin Heidelberg. doi:10.1007/b96922

Barrett-Connor, E., & Bush, T. L. (1991). Estrogen and coronary heart disease in women. *JAMA : The Journal of the American Medical Association*, *265*(14), 1861–7. doi:10.1001/jama.1991.03460140089033

Belanger, C. F., Hennekens, C. H., Rosner, B., & Speizer, F. E. (1978). The nurses' health study. *The American Journal of Nursing*, *78*(6), 1039–40. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/248266

Belcher, B. R., Berrigan, D., Dodd, K. W., Emken, B. A., Chou, C.-P., & Spruijt-Metz, D. (2010). Physical activity in US youth: effect of race/ethnicity, age, gender, and weight status. *Medicine and Science in Sports and Exercise*, *42*(12), 2211–21. doi:10.1249/MSS.0b013e3181e1fba9

Blair, S. N., Kohl, H. W., Barlow, C. E., Paffenbarger, R. S., Gibbons, L. W., & Macera, C. A. (1995). Changes in physical fitness and all-cause mortality. A prospective study of healthy and unhealthy men. *The Journal of the American Medical Association*, *273*(14), 1093–98. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7707596

Blair, S. N., Kohl, H. W., Paffenbarger, R. S., Clark, D. G., Cooper, K. H., & Gibbons, L. W. (1989). Physical Fitness and All-Cause Mortality. *The Journal of the American Medical Association*, *262*(17), 2395. doi:10.1001/jama.1989.03430170057028

Blair, S. N., LaMonte, M. J., & Nichaman, M. Z. (2004). The evolution of physical activity recommendations: how much is enough? *The American Journal of Clinical Nutrition*, *79*(5), 913S–920S. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15113739

Blomqvist, C. G., & Saltin, B. (1983). Cardiovascular adaptations to physical training. *Annual Review of Physiology*, *45*, 169–89. doi:10.1146/annurev.ph.45.030183.001125

Bouten, C. V, Koekkoek, K. T., Verduin, M., Kodde, R., & Janssen, J. D. (1997). A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. *IEEE Transactions on Bio-Medical Engineering*, *44*(3), 136–47. doi:10.1109/10.554760

Bray, G. A. (1985). Complications of obesity. *Annals of Internal Medicine*, *103*(6 (Pt 2)), 1052–62. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4062125

Camhi, S. M., Sisson, S. B., Johnson, W. D., Katzmarzyk, P. T., & Tudor-Locke, C. (2011a). Accelerometer-determined lifestyle activities in US adults. *Journal of Physical Activity & Health*, *8*(3), 382–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21487137

Camhi, S. M., Sisson, S. B., Johnson, W. D., Katzmarzyk, P. T., & Tudor-Locke, C. (2011b). Accelerometer-determined moderate intensity lifestyle activity and cardiometabolic health. *Preventive Medicine*, *52*(5), 358–60. doi:10.1016/j.ypmed.2011.01.030

Caraballo, R. S., Giovino, G. A., Pechacek, T. F., & Mowery, P. D. (2001). Factors associated with discrepancies between self- reports on cigarette smoking and measured serum cotinine levels among persons aged 17 years or older. *American Journal of Epidemiology*, *153*(8), 807–814. Retrieved from http://aje.oxfordjournals.org/content/153/8/807.short

Casale, P., Pujol, O., & Radeva, P. (2011). Human activity recognition from accelerometer data using a wearable device. In J. Vitrià, J. M. Sanches, & M. Hernández (Eds.), *Pattern Recognition and Image Analysis: 5th Iberian Conference* (pp. 289–296). Springer Berlin Heidelberg. doi:10.1007/978-3-642-21257-4

Chamnan, P., Simmons, R. K., Sharp, S. J., Griffin, S. J., & Wareham, N. J. (2009). Cardiovascular risk assessment scores for people with diabetes: a systematic review. *Diabetologia*, *52*(10), 2001–14. doi:10.1007/s00125-009-1454-0

Chapman, A. L. (1958). Program of the United States Public Health Service in heart disease in industry. *The American Journal of Cardiology*, *1*(3), 361–364. doi:10.1016/0002-9149(58)90304-7

Chasens, E. R., & Yang, K. (2012). Insomnia and physical activity in adults with prediabetes. *Clinical Nursing Research*, *21*(3), 294–308. doi:10.1177/1054773811411295

Chave, S. P., Morris, J. N., Moss, S., & Semmence, A. M. (1978). Vigorous exercise in leisure time and the death rate: a study of male civil servants. *Journal of Epidemiology & Community Health*, *32*(4), 239–243. doi:10.1136/jech.32.4.239

Chen, K. Y., Janz, K. F., Zhu, W., & Brychta, R. J. (2012). Redefining the roles of sensors in objective physical activity monitoring. *Medicine and Science in Sports and Exercise*, *44*(SUPPL. 1), 13–23. doi:10.1249/MSS.0b013e3182399bc8

Chobanian, A. V, Bakris, G. L., Black, H. R., Cushman, W. C., Green, L. A., Izzo, J. L., … Roccella, E. J. (2003). Seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. *Hypertension*, *42*(6), 1206–52. doi:10.1161/01.HYP.0000107251.49515.c2

Choe, B., Min, J., & Cho, S. (2010). Online gesture recognition for user interface on accelerometer built-in mobile phones. *Neural Information Processing. Models and Applications*, *6444*, 650–657. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-17534-3_80

Clark, B. K., Healy, G. N., Winkler, E. a H., Gardiner, P. a, Sugiyama, T., Dunstan, D. W., … Owen, N. (2011). Relationship of television time with accelerometer-derived sedentary time: NHANES. *Medicine and Science in Sports and Exercise*, *43*(5), 822–8. doi:10.1249/MSS.0b013e3182019510

Cleland, I., Kikhia, B., Nugent, C., Boytsov, A., Hallberg, J., Synnes, K., … Finlay, D. (2013). Optimal placement of accelerometers for the detection of everyday activities. *Sensors (Basel, Switzerland)*, *13*(7), 9183–200. doi:10.3390/s130709183

Colditz, G. A., Manson, J. E., & Hankinson, S. E. (1997). The Nurses' Health Study: 20-year contribution to the understanding of health among women. *Journal of Women's Health*, *6*(1), 49–62. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9065374

Conroy, R. M., Pyörälä, K., Fitzgerald, A. P., Sans, S., Menotti, A., Backer, G. De, … Graham, I. M. (2003). Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *European Heart Journal*, *24*(11), 987–1003. doi:10.1016/S0195-668X(03)00114-3

Cook, N. R., Paynter, N. P., Eaton, C. B., Manson, J. E., Martin, L. W., Robinson, J. G., … Ridker, P. M. (2012). Comparison of the Framingham and Reynolds Risk scores for global cardiovascular risk prediction in the multiethnic Women's Health Initiative. *Circulation*, *125*(14), 1748–56, S1–11. doi:10.1161/CIRCULATIONAHA.111.075929

Cooper, R., Cutler, J., Desvigne-Nickens, P., Fortmann, S. P., Friedman, L., Havlik, R., … Thom, T. (2000). Trends and Disparities in Coronary Heart Disease, Stroke, and Other Cardiovascular Diseases in the United States : Findings of the National Conference on Cardiovascular Disease Prevention. *Circulation*, *102*(25), 3137–3147. doi:10.1161/01.CIR.102.25.3137

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. doi:10.1007/BF00994018

Crouter, S. E., Clowers, K. G., & Bassett, D. R. (2006). A novel method for using accelerometer data to predict energy expenditure. *Journal of Applied Physiology*, *100*(4), 1324–31.

doi:10.1152/japplphysiol.00818.2005

D'Agostino, R. B., Vasan, R. S., Pencina, M. J., Wolf, P. a, Cobain, M., Massaro, J. M., & Kannel, W. B. (2008). General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation*, *117*(6), 743–53. doi:10.1161/CIRCULATIONAHA.107.699579

Danesh, J. (2000). Low grade inflammation and coronary heart disease: prospective study and updated meta-analyses. *BMJ*, *321*(7255), 199–204. doi:10.1136/bmj.321.7255.199

Danesh, J., Erqou, S., Walker, M., Thompson, S. G., Tipping, R., Ford, C., … Wood, A. M. (2007). The Emerging Risk Factors Collaboration: analysis of individual data on lipid, inflammatory and other markers in over 1.1 million participants in 104 prospective studies of cardiovascular diseases. *European Journal of Epidemiology*, *22*(12), 839–69. doi:10.1007/s10654-007-9165-7

Dunn, A. L., Marcus, B. H., Kampert, J. B., Garcia, M. E., Kohl, H. W., & Blair, S. N. (1999). Comparison of lifestyle and structured interventions to increase physical activity and cardiorespiratory fitness: a randomized trial. *The Journal of the American Medical Association*. Retrieved from http://jama.jamanetwork.com/article.aspx?articleid=188405

Elle, O. J., Halvorsen, S., Gulbrandsen, M. G., Aurdal, L., Bakken, A., Samset, E., … Fosse, E. (2005). Early recognition of regional cardiac ischemia using a 3-axis accelerometer sensor. *Physiological Measurement*, *26*(4), 429–40. doi:10.1088/0967-3334/26/4/009

Enos, W. F., Holmes, R. H., & Beyer, J. (1953). Coronary disease among United States soldiers killed in action in Korea; preliminary report. *The Journal of the American Medical Association*, *152*(12), 1090–1093. doi:10.1001/jama.1986.03380200097028

Ermes, M., Pärkka, J., Mantyjarvi, J., & Korhonen, I. (2008). Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. *IEEE Transactions on Information Technology in Biomedicine*, *12*(1), 20–6. doi:10.1109/TITB.2007.899496

Evenson, K. R., & Wen, F. (2011). Prevalence and correlates of objectively measured physical activity and sedentary behavior among US pregnant women. *Preventive Medicine*, *53*(1-2), 39–43. doi:10.1016/j.ypmed.2011.04.014

Evenson, K. R., Wen, F., Metzger, J. S. J., & Herring, A. A. H. (2015). Physical activity and sedentary behavior patterns using accelerometry from a national sample of United States adults. *International Journal of Behavioral Nutrition and Physical Activity*, *12*(20), 20. doi:10.1186/s12966-015-0183-7

Evenson, K. R., Wilcox, S., Pettinger, M., Brunner, R., King, A. C., & McTiernan, A. (2002). Vigorous Leisure Activity through Women's Adult Life: The Women's Health Initiative Observational Cohort Study. *American Journal of Epidemiology*, *156*(10), 945–953. doi:10.1093/aje/kwf132

Freedson, P. S., Bowles, H. R., Troiano, R. P., & Haskell, W. L. (2012). Assessment of physical activity using wearable monitors: recommendations for monitor calibration and use in the field. *Medicine and Science in Sports and Exercise*, *44*(S1), S1–4. doi:10.1249/MSS.0b013e3182399b7e

Freedson, P. S., Lyden, K., Kozey-Keadle, S., & Staudenmayer, J. (2011). Evaluation of artificial neural network algorithms for predicting METs and activity type from accelerometer data: validation on an independent sample. *Journal of Applied Physiology*, *111*(6), 1804–12. doi:10.1152/japplphysiol.00309.2011

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, *33*(1), 1–22. doi:10.18637/jss.v033.i01

Galiuto, L., & Locorotondo, G. (2015). Gender differences in cardiovascular disease prevention. *Journal of Integrative Cardiology*, *14*(1), 20–22. doi:10.15761/JIC.1000107

Gaura, E. I., Rider, R. J., Steele, N., & Naguib, R. N. (2001). Neural-network compensation methods for capacitive micromachined accelerometers for use in telecare medicine. *IEEE Transactions on Information Technology in Biomedicine*, *5*(3), 248–52. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11550847

Gaziano, T. A., Young, C. R., Fitzmaurice, G., Atwood, S., & Gaziano, J. M. (2008). Laboratory-based versus non-laboratory-based method for assessment of cardiovascular disease risk: the NHANES I Follow-up Study cohort. *Lancet*, *371*(9616), 923–31. doi:10.1016/S0140-6736(08)60418-3

Gerber, L., Otgonsuren, M., Mishra, a, Escheik, C., Birerdinc, a, Stepanova, M., & Younossi, Z. M. (2012). Non-alcoholic fatty liver disease (NAFLD) is associated with low level of physical activity: a population-based study. *Alimentary Pharmacology & Therapeutics*, *36*(8), 772–81. doi:10.1111/apt.12038

Giansanti, D. (2006). Investigation of fall-risk using a wearable device with accelerometers and rate gyroscopes. *Physiological Measurement*, *27*(11), 1081–90. doi:10.1088/0967-3334/27/11/003

Gillespie, C. D., & Hurvitz, K. A. (2013). Prevalence of hypertension and controlled hypertension - United States, 2007-2010. *Morbidity and Mortality Weekly Report*, *62*(3), 144–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24264505

Gjoreski, H., Lustrek, M., & Gams, M. (2011). Accelerometer Placement for Posture Recognition and Fall Detection. In *2011 Seventh International Conference on Intelligent Environments* (pp. 47–54). Ieee. doi:10.1109/IE.2011.11

Go, A. S., Mozaffarian, D., Roger, V. L., Benjamin, E. J., Berry, J. D., Blaha, M. J., … Turner, M. B. (2014). Heart disease and stroke statistics - 2014 update: a report from the American Heart Association. *Circulation*, *129*(3), e28–e292. doi:10.1161/01.cir.0000441139.02102.80

Guo, Y., Poulton, G., Corke, P., Bishop-Hurley, G. J., Wark, T., & Swain, D. L. (2009). Using accelerometer, high sample rate GPS and magnetometer data to develop a cattle movement and behaviour model. *Ecological Modelling*, *220*(17), 2068–2075. doi:10.1016/j.ecolmodel.2009.04.047

Haskell, W. L., Lee, I.-M., Pate, R. R., Powell, K. E., Blair, S. N., Franklin, B. a, … Bauman, A. (2007). Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Circulation*, *116*(9), 1081–93. doi:10.1161/CIRCULATIONAHA.107.185649

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. *Springer*. doi:10.1007/b94608

Hauschild, a-C., & Baumbach, J. I. (2012). Integrated statistical learning of metabolic ion mobility spectrometry profiles for pulmonary disease identification. *Genetics and Molecular Research*, *11*(3), 2733–44. doi:10.4238/2012.July.10.17

Hawkins, M. S., Sevick, M. A., Richardson, C. R., Fried, L. F., Arena, V. C., & Kriska, A. M. (2011). Association between physical activity and kidney function: National Health and Nutrition Examination Survey. *Medicine and Science in Sports and Exercise*, *43*(8), 1457–64. doi:10.1249/MSS.0b013e31820c0130

Healy, G. N., Matthews, C. E., Dunstan, D. W., Winkler, E. a H., & Owen, N. (2011). Sedentary time and cardio-metabolic biomarkers in US adults: NHANES 2003-06. *European Heart Journal*, *32*(5), 590–7. doi:10.1093/eurheartj/ehq451

Hebebrand, J., & Hinney, A. (2015). Obesity and overweight. *World Health Organization*. Retrieved from http://www.who.int/mediacentre/factsheets/fs311/en/

Heidenreich, P. a, Trogdon, J. G., Khavjou, O. a, Butler, J., Dracup, K., Ezekowitz, M. D., … Woo, Y. J. (2011). Forecasting the future of cardiovascular disease in the United States: a policy statement from the American Heart Association. *Circulation*, *123*(8), 933–44. doi:10.1161/CIR.0b013e31820a55f5

Hippisley-Cox, J., Coupland, C., Vinogradova, Y., Robson, J., Minhas, R., Sheikh, A., & Brindle, P. (2008). Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. *BMJ (Clinical Research Ed.)*, *336*(7659), 1475–82. doi:10.1136/bmj.39609.449676.25

Holman, R. M., Carson, V., & Janssen, I. (2011). Does the fractionalization of daily physical activity (sporadic vs. bouts) impact cardiometabolic risk factors in children and youth? *PloS One*, *6*(10), e25733. doi:10.1371/journal.pone.0025733

Holmes, M. D., Chen, W. Y., Feskanich, D., Kroenke, C. H., & Colditz, G. A. (2005). Physical activity and survival after breast cancer diagnosis. *The Journal of the American Medical Association*, *293*(20), 2479–86. doi:10.1001/jama.293.20.2479

Hozawa, A., Folsom, A. R., Sharrett, A. R., & Chambless, L. E. (2007). Absolute and attributable risks of cardiovascular disease incidence in relation to optimal and borderline risk factors: comparison of African American with white subjects--Atherosclerosis Risk in Communities Study. *Archives of Internal Medicine*, *167*(6), 573–9. doi:10.1001/archinte.167.6.573

Kannel, W. B., & Dawber, T. R. (1972). Atherosclerosis as a pediatric problem. *The Journal of Pediatrics*, *80*(4), 544–554.

Kannel, W. B., Dawber, T. R., Kagan, A., Revotskie, J., & Stokes, J. (1961). Factors of Risk in the Development of Coronary Heart Disease—Six-Year Follow-up Experience. *Annals of Internal Medicine*, *55*(1), 33. doi:10.7326/0003-4819-55-1-33

Kannel, W. B., Neaton, J. D., Wentworth, D., Thomas, H. E., Stamler, J., Hulley, S. B., & Kjelsberg, M. O. (1986). Overall and coronary heart disease mortality rates in relation to

major risk factors in 325,348 men screened for the MRFIT. Multiple Risk Factor Intervention Trial. *American Heart Journal*, *112*(4), 825–36. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3532744

Kannel, W. B., & Sorlie, P. (1979). Some Health Benefits of Physical Activity. *Archives of Internal Medicine*, *139*(8), 857. doi:10.1001/archinte.1979.03630450011006

Kaptoge, S., Di Angelantonio, E., Lowe, G., Pepys, M. B., Thompson, S. G., Collins, R., & Danesh, J. (2010). C-reactive protein concentration and risk of coronary heart disease, stroke, and mortality: an individual participant meta-analysis. *Lancet*, *375*(9709), 132–40. doi:10.1016/S0140-6736(09)61717-7

Keijsers, N. L. W., Horstink, M. W. I. M., & Gielen, S. C. (2003). Automatic assessment of levodopa-induced dyskinesias in daily life by neural networks. *Movement Disorders*, *18*(1), 70–80. doi:10.1002/mds.10310

Keijsers, N. L. W., Horstink, M. W. I. M., van Hilten, J. J., Hoff, J. I., & Gielen, S. C. (2000). Detection and assessment of the severity of Levodopa-induced dyskinesia in patients with Parkinson's disease by neural networks. *Movement Disorders*, *15*(6), 1104–1111. doi:10.1002/1531-8257(200011)15:6<1104::AID-MDS1007>3.0.CO;2-E

Kenney, W., Wilmore, J., & Costill, D. (2012). Cardiovascular disease and physical activity. In *Physiology of sport and exercise* (5th ed., pp. 521–543). Champaign, IL: Human Kinetics.

Kesaniemi, Y. K., Danforth, E., Jensen, M. D., Kopelman, P. G., Lefèbvre, P., & Reeder, B. A. (2001). Dose-response issues concerning physical activity and health: an evidence-based symposium. *Medicine and Science in Sports and Exercise*, *33*(6 Suppl), S351–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11427759

Keys, A., Fidanza, F., Karvonen, M. J., Kimura, N., & Taylor, H. L. (1972). Indices of relative weight and obesity. *Journal of Chronic Diseases*, *25*(6), 329–343. Retrieved from http://www.sciencedirect.com/science/article/pii/0021968172900276

Kozina, S., Lustrek, M., & Gams, M. (2011). Dynamic signal segmentation for activity recognition. In *Proceedings of International Joint Conference on Artificial Intelligence* (pp. 1–12). Barcelona, Spain.

Krause, A., Siewiorek, D., Smailagic, A., & Farringdon, J. (2003). Unsupervised, dynamic identification of physiological and activity context in wearable computing. In *Seventh IEEE International Symposium on Wearable Computers* (p. 88). Retrieved from http://www.computer.org/csdl/proceedings/iswc/2003/2034/00/20340088.pdf

Kuhn, M. (2008). Caret package. *Journal of Statistical Software*. Retrieved from http://download.nextag.com/cran/web/packages/caret/caret.pdf

Kwak, N. K., & Lee, C. (1997). A neural network application to classification of health status of HIV/AIDS patients. *Journal of Medical Systems*, *21*(2), 87–97. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9297617

Laerhoven, K. Van. (2001). Combining the self-organizing map and k-means clustering for on-line classification of sensor data. *Artificial Neural Networks—ICANN*. Retrieved from http://link.springer.com/chapter/10.1007/3-540-44668-0_65

Lee, I.-M., Sesso, H. D., Oguma, Y., & Paffenbarger, R. S. (2004). The "weekend warrior" and risk of mortality. *American Journal of Epidemiology*, *160*(7), 636–41. doi:10.1093/aje/kwh274

Lemon, S. C., Roy, J., Clark, M. A., Friedmann, P. D., & Rakowski, W. (2003). Classification and regression tree analysis in public health: Methodological review and comparison with logistic regression. *Annals of Behavioral Medicine*, *26*(3), 172–181. doi:10.1207/S15324796ABM2603_02

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*. Retrieved from ftp://131.252.97.79/Transfer/Treg/WFRE_Articles/Liaw_02_Classification and regression by randomForest.pdf

Liu, J., Pan, Z., & Xiangcheng, L. (2010). An accelerometer-based gesture recognition algorithm and its application for 3D interaction. *Computer Science and Information Systems*, *7*(1), 177–188. doi:10.2298/CSIS1001177L

Liu, S.-H., & Chang, Y.-J. (2009). Using accelerometers for physical actions recognition by a neural fuzzy network. *Telemedicine Journal and E-Health : The Official Journal of the American Telemedicine Association*, *15*(9), 867–76. doi:10.1089/tmj.2009.0032

Lloyd-Jones, D. M., Adams, R., & Brown, T. (2010). Heart disease and stroke statistics—2010 update A report from the American Heart Association. *Circulation*. Retrieved from http://circ.ahajournals.org/content/121/7/e46.short

Lloyd-Jones, D. M., Hong, Y., Labarthe, D., Mozaffarian, D., Appel, L. J., Van Horn, L., … Rosamond, W. D. (2010). Defining and setting national goals for cardiovascular health promotion and disease reduction: the American Heart Association's strategic Impact Goal through 2020 and beyond. *Circulation*, *121*(4), 586–613. doi:10.1161/CIRCULATIONAHA.109.192703

Long, X., Yin, B., & Aarts, R. M. (2009). Single-accelerometer-based daily physical activity classification. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Vol. 2009, pp. 6107–10). doi:10.1109/IEMBS.2009.5334925

Loprinzi, P. D., Lee, H., Cardinal, B. J., Crespo, C. J., Andersen, R. E., & Smit, E. (2012, September). The relationship of actigraph accelerometer cut-points for estimating physical activity with selected health outcomes: results from NHANES 2003-06. *Research Quarterly for Exercise and Sport*. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22978192

Luke, A., Dugas, L. R., Durazo-Arvizu, R. a, Cao, G., & Cooper, R. S. (2011). Assessing physical activity and its relationship to cardiovascular risk factors: NHANES 2003-2006. *BMC Public Health*, *11*(1), 387. doi:10.1186/1471-2458-11-387

Luštrek, M., & Kaluža, B. (2008). Fall Detection and Activity Recognition with Machine Learning. *Informatica*, *33*, 205–212. doi:10.1371/journal.pone.0036556

Lynch, B. M., Dunstan, D. W., Healy, G. N., Winkler, E., Eakin, E., & Owen, N. (2010). Objectively measured physical activity and sedentary time of breast cancer survivors, and associations with adiposity: findings from NHANES (2003-2006). *Cancer Causes & Control : CCC*, *21*(2), 283–8. doi:10.1007/s10552-009-9460-6

Lynch, B. M., Friedenreich, C. M., Winkler, E. a H., Healy, G. N., Vallance, J. K., Eakin, E. G., & Owen, N. (2011). Associations of objectively assessed physical activity and sedentary time with biomarkers of breast cancer risk in postmenopausal women: findings from NHANES (2003-2006). *Breast Cancer Research and Treatment*, *130*(1), 183–94. doi:10.1007/s10549-011-1559-2

Mahmood, S. S., Levy, D., Vasan, R. S., & Wang, T. J. (2014). The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. *Lancet*, *383*(9921), 999–1008. doi:10.1016/S0140-6736(13)61752-3

Mannini, A., & Sabatini, A. M. (2010). Machine learning methods for classifying human physical activity from on-body accelerometers. *Sensors (Basel, Switzerland)*, *10*(2), 1154–75. doi:10.3390/s100201154

Martiskainen, P., Järvinen, M., Skön, J.-P., Tiirikainen, J., Kolehmainen, M., & Mononen, J. (2009). Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines. *Applied Animal Behaviour Science*, *119*(1-2), 32–38. doi:10.1016/j.applanim.2009.03.005

Masse, L. C., Fuemmeler, B. F., Anderson, C. B., Matthews, C. E., Trost, S. G., Catellier, D. J., & Treuth, M. (2005). Accelerometer Data Reduction: A Comparison of Four Reduction Algorithms on Select Outcome Variables. *Medicine & Science in Sports & Exercise*, *37*(Supplement), S544–S554. doi:10.1249/01.mss.0000185674.09066.8a

Matthews, C. E. (2005). Calibration for Accelerometer Output for Adults. *Medicine and Science in Sport and Exercise*, *S512*(Supplement), S512–S522. doi:10.1249/01.mss.0000185659.11982.3d

Medzhitov, R. (2008). Origin and physiological roles of inflammation. *Nature*, *454*(7203), 428–35. doi:10.1038/nature07201

Meyer, D., Dimitriadou, E., & Hornik, K. (2015). Package "e1071." Retrieved from ftp://200.236.31.10/CRAN/web/packages/e1071/e1071.pdf

Michel, K. E., & Brown, D. C. (2011). Determination and application of cut points for accelerometer-based activity counts of activities with differing intensity in pet dogs. *American Journal of Veterinary Research*, *72*(7), 866–70. doi:10.2460/ajvr.72.7.866

Mora, S., Cook, N., Buring, J. E., Ridker, P. M., & Lee, I.-M. (2007). Physical activity and reduced risk of cardiovascular events: potential mediating mechanisms. *Circulation*, *116*(19), 2110–8. doi:10.1161/CIRCULATIONAHA.107.729939

Morris, J. N., Chave, S. P., Adam, C., Sirey, C., Epstein, L., & Sheehan, D. J. (1973). Vigorous exercise in leisure-time and the incidence of coronary heart-disease. *The Lancet*, *301*(7799), 333–339. doi:10.1016/S0140-6736(73)90128-1

Morris, J. N., Heady, J. A., Raffle, P. A., Roberts, C. G., & Parks, J. W. (1953). Coronary heart disease and physical activity of work. *Lancet*, *265*, 1053–1057.

Myers, G. L., Rifai, N., Tracy, R. P., Roberts, W. L., Alexander, R. W., Biasucci, L. M., … Waymack, P. P. (2004). CDC/AHA Workshop on Markers of Inflammation and Cardiovascular Disease: Application to Clinical and Public Health Practice: report from the

laboratory science discussion group. *Circulation*, *110*(25), e545–9. doi:10.1161/01.CIR.0000148980.87579.5E

Myers, J., Kaykha, A., George, S., Abella, J., Zaheer, N., Lear, S., … Froelicher, V. (2004). Fitness versus physical activity patterns in predicting mortality in men. *The American Journal of Medicine*, *117*(12), 912–8. doi:10.1016/j.amjmed.2004.06.047

National Cholesterol Education Program (NCEP) Expert Panel. (2002). Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). *Archives of Internal Medicine*, (6), 284. doi:10.1001/archinte.1991.00400060019005

NHANES - History. (2011). *Centers for Disease Control and Prevention (CDC)*. Retrieved November 13, 2015, from http://www.cdc.gov/nchs/nhanes/history.htm

NHANES 2003 - 2004: Physical Activity Monitor Data Documentation, Codebook, and Frequencies. (2007). *Centers for Disease Control and Prevention (CDC)*. Retrieved November 28, 2015, from http://www.cdc.gov/nchs/nhanes/nhanes2003-2004/PAXRAW_C.htm

NHIS - About the National Health Interview Survey. (2015). *Centers for Disease Control and Prevention (CDC)*. Retrieved November 28, 2015, from http://www.cdc.gov/nchs/nhis/about_nhis.htm

O'Connell, J., Tøgersen, F. A., Friggens, N. C., Løvendahl, P., & Højsgaard, S. (2010). Combining Cattle Activity and Progesterone Measurements Using Hidden Semi-Markov Models. *Journal of Agricultural, Biological, and Environmental Statistics*, *16*(1), 1–16. doi:10.1007/s13253-010-0033-7

Ogden, C. L., Lamb, M. M., Carroll, M. D., & Flegal, K. M. (2010). Obesity and socioeconomic status in adults: United States, 2005-2008. *NCHS Data Brief*, (50), 1–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21211165

Ozemek, C., Cochran, H. L., Strath, S. J., Byun, W., & Kaminsky, L. A. (2013). Estimating relative intensity using individualized accelerometer cutpoints: the importance of fitness level. *BMC Medical Research Methodology*, *13*, 53. doi:10.1186/1471-2288-13-53

Paffenbarger, R. S., Blair, S. N., & Lee, I.-M. (2001). A history of physical activity, cardiovascular health and longevity: the scientific contributions of Jeremy N Morris, DSc, DPH, FRCP. *International Journal of Epidemiology*, *30*(5), 1184–1192. doi:10.1093/ije/30.5.1184

Paffenbarger, R. S., Wing, A. L., & Hyde, R. T. (1978). Physical activity as an index of heart attack risk in college alumni. *American Journal of Epidemiology*, *108*(3), 161–75. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/707484

Pärkkä, J., Cluitmans, L., & Ermes, M. (2010). Personalization algorithm for real-time activity recognition using PDA, wireless motion bands, and binary decision tree. *IEEE Transactions on Information Technology in Biomedicine*, *14*(5), 1211–5. doi:10.1109/TITB.2010.2055060

Pärkkä, J., Ermes, M., Korpipää, P., Mäntyjärvi, J., Peltola, J., & Korhonen, I. (2006). Activity

classification using realistic data from wearable sensors. *IEEE Transactions on Information Technology in Biomedicine*, *10*(1), 119–128. doi:10.1109/TITB.2005.856863

Pate, R. R., Pratt, M., Blair, S. N., Haskell, W. L., Macera, C. A., Bouchard, C., … Wilmore, J. H. (1995). Physical activity and public health. A recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. *The Journal of the American Medical Association*, *273*(5), 402–407.

Patel, S., Hughes, R., Hester, T., Stein, J., Akay, M., Dy, J. G., & Bonato, P. (2010). A Novel Approach to Monitor Rehabilitation Outcomes in Stroke Survivors Using Wearable Technology. In *Proceedings of the IEEE* (Vol. 98, pp. 450–461). doi:10.1109/JPROC.2009.2038727

Patel, S., Mancinelli, C., Healey, J. A., Moy, M., & Bonato, P. (2009). Using Wearable Sensors to Monitor Physical Activities of Patients with COPD: A Comparison of Classifier Performance. In *2009 Sixth International Workshop on Wearable and Implantable Body Sensor Networks* (pp. 234–239). doi:10.1109/BSN.2009.53

Physical activity and cardiovascular health. NIH Consensus Development Panel on Physical Activity and Cardiovascular Health. (1996). *Journal of the American Medical Association* (Vol. 276). Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8667571

Pober, D. M., Staudenmayer, J., Raphael, C., & Freedson, P. S. (2006). Development of novel techniques to classify physical activity mode using accelerometers. *Medicine and Science in Sports and Exercise*, *38*(9), 1626–34. doi:10.1249/01.mss.0000227542.43669.45

Pylvänäinen, T. (2005). Accelerometer Based Gesture Recognition Using Continuous HMMs. *Pattern Recognition and Image Analysis*, 639–646. doi:10.1007/11492429_77

R Core Team. (2014). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. In *Proceedings of the IEEE* (Vol. 77, pp. 257–286). doi:10.1109/5.18626

Ravi, N., Dandekar, N., Mysore, P., & Littman, M. (2005). Activity recognition from accelerometer data. In *Proceedings of the National Conference on Artificial Intelligence* (pp. 1541–1546). doi:10.1007/978-3-642-02481-8_120

Reddy, S., Burke, J., Estrin, D., Hansen, M., & Srivastava, M. (2008). Determining transportation mode on mobile phones. In *2008 12th IEEE International Symposium on Wearable Computers* (pp. 25–28). Ieee. doi:10.1109/ISWC.2008.4911579

Ridker, P. M., Buring, J. E., Rifai, N., & Cook, N. R. (2007). Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score. *The Journal of the American Medical Association*, *297*(6), 611–9. doi:10.1001/jama.297.6.611

Ridker, P. M., Paynter, N. P., Rifai, N., Gaziano, J. M., & Cook, N. R. (2008). C-reactive protein and parental history improve global cardiovascular risk prediction: the Reynolds Risk Score for men. *Circulation*, *118*(22), 2243–51, 4p following 2251. doi:10.1161/CIRCULATIONAHA.108.814251

Robert, B., White, B. J., Renter, D. G., & Larson, R. L. (2009). Evaluation of three-dimensional accelerometers to monitor and classify behavior patterns in cattle. *Computers and Electronics in Agriculture*, *67*(1-2), 80–84. doi:10.1016/j.compag.2009.03.002

Rothney, M. P., Neumann, M., Béziat, A., & Chen, K. Y. (2007). An artificial neural network model of energy expenditure using nonintegrated acceleration signals. *Journal of Applied Physiology*, *103*(4), 1419–27. doi:10.1152/japplphysiol.00429.2007

Rothney, M. P., Schaefer, E. V., Neumann, M. M., Choi, L., & Chen, K. Y. (2008). Validity of Physical Activity Intensity Predictions by ActiGraph, Actical, and RT3 Accelerometers. *Obesity*, *16*(8), 1946–1952. doi:10.1038/oby.2008.279.Validity

Sesso, H. D., Paffenbarger, R. S., & Lee, I.-M. (2000). Physical Activity and Coronary Heart Disease in Men : The Harvard Alumni Health Study. *Circulation*, *102*(9), 975–980. doi:10.1161/01.CIR.102.9.975

Shipp, M. A., Ross, K. N., Tamayo, P., Weng, A. P., Kutok, J. L., Aguiar, R. C. T., … Golub, T. R. (2002). Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nature Medicine*, *8*(1), 68–74. doi:10.1038/nm0102-68

Singh, M., & Patterson, D. J. (2010). Involuntary gesture recognition for predicting cerebral palsy in high-risk infants. In *International Symposium on Wearable Computers* (pp. 1–8). IEEE. doi:10.1109/ISWC.2010.5665874

Sisson, S. B., Camhi, S. M., Church, T. S., Tudor-Locke, C., Johnson, W. D., & Katzmarzyk, P. T. (2010). Accelerometer-determined steps/day and metabolic syndrome. *American Journal of Preventive Medicine*, *38*(6), 575–82. doi:10.1016/j.amepre.2010.02.015

Smith, W. a, Nolan, V. G., Robison, L. L., Hudson, M. M., & Ness, K. K. (2011). Physical activity among cancer survivors and those with no history of cancer- a report from the National Health and Nutrition Examination Survey 2003-2006. *American Journal of Translational Research*, *3*(4), 342–50.

Song, X., Mitnitski, A., Cox, J., & Rockwood, K. (2004). Comparison of machine learning techniques with classical statistical models in predicting health outcomes. *Studies in Health Technology and Informatics*, *107*(Pt 1), 736–40. doi:10.3233/978-1-60750-949-3-736

Sprager, S., & Zazula, D. (2009). A Cumulant-Based Method for Gait Identification Using Accelerometer Data with Principal Component Analysis and Support Vector Machine, *5*(11), 369–378.

Staudenmayer, J., Pober, D. R., Crouter, S., Bassett, D., & Freedson, P. S. (2009). An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. *Journal of Applied Physiology*, *107*(4), 1300–7. doi:10.1152/japplphysiol.00465.2009

Strath, S. J., Kaminsky, L. A., Ainsworth, B. E., Ekelund, U., Freedson, P. S., Gary, R. A., … Swartz, A. M. (2013). Guide to the assessment of physical activity: Clinical and research applications: a scientific statement from the American Heart Association. *Circulation*, *128*(20), 2259–79. doi:10.1161/01.cir.0000435708.67487.da

Tabata, I., Nishimura, K., Kouzaki, M., Hirai, Y., Ogita, F., Miyachi, M., & Yamamoto, K.

(1996). Effects of moderate-intensity endurance and high-intensity intermittent training on anaerobic capacity and VO2max. *Medicine and Science in Sports and Exercise*, *28*(10), 1327–30. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8897392

Tapia, E. M. (2008). *Using machine learning for real-time activity recognition and estimation of energy expenditure*. Massachusetts Institute of Technology. Retrieved from http://dspace.mit.edu/handle/1721.1/44913

Tapia, E. M., Intille, S. S., Haskell, W. L., Larson, K., Wright, J., King, A., & Friedman, R. (2007). 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* (pp. 1–4). Ieee. doi:10.1109/ISWC.2007.4373774

Tennekes, M., & Jonge, E. de. (2012). tabplot: Tableplot, a visualization of large datasets. *R Package Version*.

Therneau, T., Atkinson, B., & Ripley, B. (2015). rpart: Recursive Partitioning and Regression Trees. R package version 4.1-10.

Troiano, R. P. (2006). Translating accelerometer counts into energy expenditure: advancing the quest. *Journal of Applied Physiology*, *100*(4), 1107–8. doi:10.1152/japplphysiol.01577.2005

Troiano, R. P., Berrigan, D., Dodd, K. W., Mâsse, L. C., Tilert, T., & McDowell, M. (2008). Physical activity in the United States measured by accelerometer. *Medicine and Science in Sports and Exercise*, *40*(1), 181–8. doi:10.1249/mss.0b013e31815a51b3

Trost, S. G., Pate, R. R., Freedson, P. S., Sallis, J. F., & Taylor, W. C. (2000). Using objective physical activity measures with youth: how many days of monitoring are needed? *Medicine and Science in Sports and Exercise*, *32*(2), 426–31. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10694127

Trost, S. G., Pate, R. R., Sallis, J. F., Freedson, P. S., Taylor, W. C., Dowda, M., & Sirard, J. (2002). Age and gender differences in objectively measured physical activity in youth. *Medicine and Science in Sports and Exercise*, *34*(2), 350–5. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11828247

Trost, S. G., Wong, W.-K., Pfeiffer, K. a, & Zheng, Y. (2012). Artificial neural networks to predict activity type and energy expenditure in youth. *Medicine and Science in Sports and Exercise*, *44*(9), 1801–9. doi:10.1249/MSS.0b013e318258ac11

Trost, S. G., Zheng, Y., & Wong, W.-K. (2014). Machine learning for activity recognition: hip versus wrist data. *Physiological Measurement*, *35*(11), 2183–9. doi:10.1088/0967-3334/35/11/2183

Tsipouras, M. G., Tzallas, A. T., Rigas, G., Bougia, P., Fotiadis, D. I., & Konitsiotis, S. (2010). Automated Levodopa-induced dyskinesia assessment. In *Conference of the IEEE Engineering in Medicine and Biology Society*. (pp. 2411–4). doi:10.1109/IEMBS.2010.5626130

Tudor-Locke, C., Brashear, M. M., Johnson, W. D., & Katzmarzyk, P. T. (2010). Accelerometer profiles of physical activity and inactivity in normal weight, overweight, and obese U.S. men and women. *The International Journal of Behavioral Nutrition and Physical Activity*,

7, 60. doi:10.1186/1479-5868-7-60

Tudor-Locke, C., Camhi, S. M., Leonardi, C., Johnson, W. D., Katzmarzyk, P. T., Earnest, C. P., & Church, T. S. (2011). Patterns of adult stepping cadence in the 2005-2006 NHANES. *Preventive Medicine*, *53*(3), 178–81. doi:10.1016/j.ypmed.2011.06.004

Tudor-Locke, C., Camhi, S. M., & Troiano, R. P. (2012). A catalog of rules, variables, and definitions applied to accelerometer data in the National Health and Nutrition Examination Survey, 2003-2006. *Preventing Chronic Disease*, *9*, E113. doi:10.5888/pcd9.110332

Tudor-Locke, C., Johnson, W. D., & Katzmarzyk, P. T. (2009). Accelerometer-determined steps per day in US adults. *Medicine and Science in Sports and Exercise*, *41*(7), 1384–91. doi:10.1249/MSS.0b013e318199885c

Tudor-Locke, C., Johnson, W. D., & Katzmarzyk, P. T. (2010). Accelerometer-determined steps per day in US children and youth. *Medicine and Science in Sports and Exercise*, *42*(12), 2244–50. doi:10.1249/MSS.0b013e3181e32d7f

Tudor-Locke, C., Johnson, W. D., & Katzmarzyk, P. T. (2011). Relationship between accelerometer-determined steps/day and other accelerometer outputs in US adults. *Journal of Physical Activity & Health*, *8*(3), 410–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21487141

United States Department of Health and Human Services. (2008). *2008 Physical Activity Guidelines for Americans. Washington DC: Department of Health and Human Services USA*. Retrieved from http://www.health.gov/paguidelines/guidelines/

Vallance, J. K., Winkler, E. a H., Gardiner, P. a, Healy, G. N., Lynch, B. M., & Owen, N. (2011). Associations of objectively-assessed physical activity and sedentary time with depression: NHANES (2005-2006). *Preventive Medicine*, *53*(4-5), 284–8. doi:10.1016/j.ypmed.2011.07.013

Van Domelen, D. R., Koster, A., Caserotti, P., Brychta, R. J., Chen, K. Y., McClain, J. J., … Harris, T. B. (2011). Employment and physical activity in the U.S. *American Journal of Preventive Medicine*, *41*(2), 136–45. doi:10.1016/j.amepre.2011.03.019

Vanhelst, J., Béghin, L., Turck, D., & Gottrand, F. (2011). New validated thresholds for various intensities of physical activity in adolescents using the Actigraph accelerometer. *International Journal of Rehabilitation Research. Internationale Zeitschrift Für Rehabilitationsforschung. Revue Internationale de Recherches de Réadaptation*, *34*(2), 175–7. doi:10.1097/MRR.0b013e328340129e

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer. Retrieved from https://books.google.com/books?id=GeX9CYd_JTkC&pgis=1

Wall, M., & Johnson, J. (1988). Cotinine in the serum, saliva, and urine of nonsmokers, passive smokers, and active smokers. *American Journal of Public Health*. Retrieved from http://ajph.aphapublications.org/doi/abs/10.2105/AJPH.78.6.699

Wang, S. W. S., Yang, J. Y. J., Chen, N. C. N., Chen, X. C. X., & Zhang, Q. Z. Q. (2005). Human Activity Recognition with User-Free Accelerometers in the Sensor Networks. In *2005 International Conference on Neural Networks and Brain* (Vol. 2, pp. 1212–1217).

IEEE. doi:10.1109/ICNNB.2005.1614831

Wang, Y., Huang, K., & Tan, T. (2007). Human Activity Recognition Based on R Transform. *2007 IEEE Conference on Computer Vision and Pattern Recognition*, 1–8. doi:10.1109/CVPR.2007.383505

Wannamethee, S. G., Shaper, a G., Lennon, L., & Morris, R. W. (2005). Metabolic syndrome vs Framingham Risk Score for prediction of coronary heart disease, stroke, and type 2 diabetes mellitus. *Archives of Internal Medicine*, *165*(22), 2644–50. doi:10.1001/archinte.165.22.2644

Wei, T., & Simko, V. (2016). *corrplot: Visualization of a Correlation Matrix*. *R package version 0.77*. Retrieved from http://cran.r-project.org/package=corrplot

Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. Springer-Verlag New York.

Wilson, P. W. F., D'Agostino, R. B., Levy, D., Belanger, a. M., Silbershatz, H., & Kannel, W. B. (1998). Prediction of Coronary Heart Disease Using Risk Factor Categories. *Circulation*, *97*(18), 1837–1847. doi:10.1161/01.CIR.97.18.1837

Wisløff, U., Støylen, A., Loennechen, J. P., Bruvold, M., Rognmo, Ø., Haram, P. M., … Skjærpe, T. (2007). Superior cardiovascular effect of aerobic interval training versus moderate continuous training in heart failure patients: A randomized study. *Circulation*, *115*(24), 3086–3094. doi:10.1161/CIRCULATIONAHA.106.675041

Wolberg, W. H., Street, W. N., & Mangasarian, O. L. (1995). Image analysis and machine learning applied to breast cancer diagnosis and prognosis. *Analytical and Quantitative Cytology and Histology / the International Academy of Cytology [and] American Society of Cytology*, *17*(2), 77–87. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7612134

Wong, S. L., Colley, R., Connor Gorber, S., & Tremblay, M. (2011). Actical accelerometer sedentary activity thresholds for adults. *Journal of Physical Activity & Health*, *8*(4), 587–91. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21597132

Wu, B., Abbott, T., Fishman, D., McMurray, W., Mor, G., Stone, K., … Zhao, H. (2003). Comparison of statistical methods for classification of ovarian cancer using mass spectrometry data. *Bioinformatics*, *19*(13), 1636–1643. doi:10.1093/bioinformatics/btg210

Yang, M., Zheng, H., Wang, H., McClean, S., Hall, J., & Harris, N. (2012). A machine learning approach to assessing gait patterns for Complex Regional Pain Syndrome. *Medical Engineering & Physics*, *34*(6), 740–6. doi:10.1016/j.medengphy.2011.09.018

Yang, Q., Cogswell, M. E., Flanders, W. D., Hong, Y., Zhang, Z., Loustalot, F., … Hu, F. B. (2012). Trends in cardiovascular health metrics and associations with all-cause and CVD mortality among US adults. *The Journal of the American Medical Association*, *307*(12), 1273–83. doi:10.1001/jama.2012.339

Zhang, X., Chen, X., & Li, Y. (2011). A framework for hand gesture recognition based on accelerometer and EMG sensors. *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, *41*(6), 1064 – 1076. doi:10.1109/TSMCA.2011.2116004

Zheng, Y. (2012, August 17). *Predicting activity type from accelerometer data*. Retrieved from http://ir.library.oregonstate.edu/xmlui/handle/1957/33727

Zisko, N., Carlsen, T., Salvesen, Ø., Aspvik, N. P., Ingebrigtsen, J. E., Wisløff, U., & Stensvold, D. (2015). New relative intensity ambulatory accelerometer thresholds for elderly men and women: the Generation 100 study. *BMC Geriatrics, 15*(1), 97. doi:10.1186/s12877-015-0093-1

APPENDIX A: PRINCIPAL COMPONENT LOADINGS

Table 25. NCI principal components loadings

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
sed_min	0.13	-0.17	0.11	-0.05	0.17	-0.42	0.06	-0.23
light_min	-0.07	0.20	-0.15	0.11	-0.21	0.36	-0.10	-0.47
life_min	-0.22	0.21	-0.16	0.04	-0.01	-0.09	0.00	-0.05
mod_min	-0.26	-0.09	-0.01	-0.13	0.08	-0.01	0.07	0.16
vig_min	-0.10	-0.21	0.04	0.51	-0.09	-0.07	-0.01	0.06
sed_bouted_60min	0.08	-0.10	0.09	-0.06	0.30	-0.59	-0.12	-0.39
num_mvpa_bouts	-0.20	-0.24	0.07	-0.12	0.05	0.17	-0.07	-0.23
num_vig_bouts	-0.08	-0.20	0.04	0.49	-0.09	-0.05	-0.04	-0.12
mvpa_bouted	-0.19	-0.26	0.07	-0.04	0.06	0.16	-0.01	-0.12
vig_bouted	-0.07	-0.18	0.03	0.55	-0.11	-0.10	-0.02	0.08
tot_mv_bouts	-0.20	-0.25	0.06	-0.12	0.05	0.16	-0.04	-0.25
tot_mv_min	-0.26	-0.13	-0.02	-0.06	0.07	-0.04	0.10	0.12
tot_li_min	-0.22	0.18	-0.19	0.03	-0.01	-0.14	0.04	-0.11
wk_mv_dif_bout	-0.07	-0.01	0.25	-0.16	-0.52	-0.14	-0.41	-0.10
avg_wk_mv_dif_min	-0.12	0.08	0.31	-0.12	-0.36	-0.18	-0.25	0.16
avg_wk_li_dif_min	-0.11	0.21	0.26	0.08	0.22	0.04	-0.37	0.11
tot_wk_mv_dif_min	-0.24	-0.03	0.16	-0.10	-0.12	-0.11	-0.03	0.17
tot_wk_li_dif_min	-0.20	0.26	0.05	0.08	0.15	-0.05	-0.12	-0.02
perc_wk_mv	-0.06	0.16	0.35	0.02	-0.19	0.00	0.53	-0.15
perc_we_mv	0.05	-0.15	-0.37	-0.01	0.16	0.05	-0.49	0.15
perc_wk_li	-0.06	0.20	0.35	0.12	0.34	0.17	-0.08	0.00
perc_we_li	0.06	-0.20	-0.35	-0.12	-0.34	-0.17	0.08	0.00
mv_wk2	-0.18	-0.23	0.06	-0.11	0.04	0.19	-0.08	-0.36
top_mv.1	-0.24	-0.12	-0.03	-0.06	0.08	-0.01	0.08	0.10
top_mv.2	-0.25	-0.12	0.00	-0.06	0.05	-0.03	0.07	0.14
top_mv.3	-0.25	-0.12	0.00	-0.06	0.05	-0.04	0.08	0.17
top_mv.4	-0.25	-0.12	-0.01	-0.07	0.05	-0.04	0.08	0.16
top_li.1	-0.21	0.19	-0.17	0.03	-0.01	-0.11	0.02	-0.12
top_li.2	-0.22	0.20	-0.17	0.04	0.00	-0.12	0.00	-0.08
top_li.3	-0.22	0.20	-0.16	0.04	0.00	-0.12	0.00	-0.07
top_li.4	-0.22	0.20	-0.17	0.04	0.00	-0.12	0.01	-0.06

Table 26. K-means principal components loadings

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
min.1	0.00	0.37	0.19	-0.05	0.03	-0.10	0.19	-0.01
min.2	-0.01	-0.28	-0.32	0.16	-0.01	0.08	-0.07	-0.09
min.3	0.02	-0.36	-0.15	-0.09	-0.06	0.14	-0.20	-0.12
min.4	0.01	-0.36	0.05	-0.07	0.00	0.00	-0.30	0.11
min.5	-0.03	-0.23	0.19	0.12	-0.05	0.08	-0.26	0.69
threshold.1	-0.27	-0.10	0.14	-0.09	-0.08	0.02	0.16	0.03
threshold.2	-0.29	-0.05	0.15	-0.03	-0.05	-0.04	0.07	-0.01
threshold.3	-0.29	0.03	0.09	0.05	-0.02	0.04	-0.03	-0.03
threshold.4	-0.25	0.11	-0.16	-0.10	0.03	0.00	-0.07	0.10
threshold.5	-0.25	0.04	-0.14	-0.14	0.29	0.01	0.07	0.05
center.1	-0.24	-0.19	-0.06	0.09	-0.06	-0.05	0.25	0.00
center.2	-0.27	-0.09	0.15	-0.10	-0.08	0.02	0.15	0.03
center.3	-0.29	-0.03	0.15	0.00	-0.04	-0.06	0.04	-0.02
center.4	-0.29	0.05	0.07	0.08	-0.01	0.08	-0.04	-0.03
center.5	-0.26	0.11	-0.18	-0.11	0.03	-0.01	-0.05	0.10
variance.1	-0.27	-0.13	0.02	0.00	-0.06	0.01	0.19	-0.02
variance.2	-0.28	-0.01	0.14	-0.02	-0.04	-0.03	0.00	-0.10
variance.3	-0.19	0.07	0.07	0.10	0.05	-0.04	-0.54	-0.43
variance.4	-0.20	0.11	-0.10	0.03	0.06	0.24	-0.34	-0.17
variance.5	-0.12	0.05	-0.22	-0.21	0.16	-0.08	-0.14	0.33
skewness.1	0.07	0.23	0.28	-0.32	0.00	0.15	-0.19	0.07
skewness.2	-0.06	0.21	-0.04	0.40	0.15	-0.36	-0.02	0.15
skewness.3	-0.04	0.23	-0.22	0.21	-0.01	0.46	0.14	0.08
skewness.4	-0.02	0.08	-0.38	-0.34	-0.01	-0.18	0.02	-0.14
skewness.5	0.05	-0.16	0.11	-0.05	0.62	0.08	0.08	-0.11
kurtosis.1	0.06	0.23	0.28	-0.31	0.00	0.15	-0.18	0.07
kurtosis.2	-0.07	0.20	-0.02	0.38	0.16	-0.30	-0.24	0.09
kurtosis.3	-0.05	0.20	-0.19	0.19	0.01	0.55	0.05	0.11
kurtosis.4	-0.06	0.09	-0.35	-0.33	0.03	-0.20	-0.06	0.19
kurtosis.5	0.00	-0.13	0.11	-0.03	0.64	0.10	0.12	0.01

Table 27. HMM principal components loadings.

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
min.1	0.10	0.13	-0.01	0.19	-0.23	0.15	-0.02	0.40	-0.05	0.20
min.2	0.11	0.12	0.03	-0.10	-0.08	0.02	0.22	0.49	0.21	0.04
min.3	0.11	0.01	0.01	-0.06	0.32	0.14	0.24	0.24	0.08	0.08
min.4	0.00	-0.27	-0.11	-0.17	0.21	0.02	-0.06	0.35	-0.08	0.06
min.5	-0.04	-0.24	-0.05	-0.19	0.00	0.03	-0.22	0.22	0.19	0.02
trans_prob.1	0.21	-0.05	0.05	0.14	-0.08	-0.02	0.11	0.05	0.16	0.06
trans_prob.2	-0.07	0.10	-0.19	-0.07	0.00	-0.23	0.07	0.04	-0.04	0.07
trans_prob.3	-0.12	0.17	-0.02	-0.21	0.22	0.04	0.00	-0.20	-0.16	0.05
trans_prob.4	-0.20	-0.12	0.08	-0.03	0.09	-0.04	-0.18	0.07	0.02	-0.19
trans_prob.5	-0.13	-0.21	0.10	-0.09	-0.09	-0.05	-0.08	-0.03	0.13	-0.18
trans_prob.6	-0.06	0.13	-0.18	0.23	-0.11	-0.29	-0.14	0.02	-0.16	0.28
trans_prob.7	0.21	-0.09	-0.01	0.05	-0.04	-0.04	0.18	0.04	0.25	-0.11
trans_prob.8	-0.11	0.15	0.03	-0.34	0.18	0.11	0.11	-0.13	-0.10	0.06
trans_prob.9	-0.18	-0.07	0.14	-0.02	0.17	-0.04	-0.21	0.08	-0.04	-0.15
trans_prob.11	-0.11	0.25	-0.02	0.01	-0.16	-0.14	-0.15	-0.02	-0.13	0.16
trans_prob.12	-0.10	0.24	0.04	-0.27	-0.25	-0.05	0.05	0.16	0.03	0.05
trans_prob.13	0.21	-0.12	-0.09	0.07	0.09	-0.03	0.23	-0.21	0.11	-0.08
trans_prob.14	-0.15	-0.05	0.09	0.08	0.30	-0.02	-0.21	0.20	-0.06	0.05
trans_prob.15	-0.14	-0.22	0.07	-0.06	-0.11	-0.08	-0.16	0.07	0.11	-0.11
trans_prob.16	-0.16	0.10	0.14	0.16	-0.18	-0.12	-0.08	0.05	0.13	-0.08
trans_prob.17	-0.11	0.21	0.24	0.07	-0.09	-0.06	0.00	0.07	0.15	-0.15
trans_prob.18	-0.04	0.28	0.17	0.16	0.27	0.07	0.10	0.01	0.10	0.07
trans_prob.19	0.18	-0.15	-0.25	-0.16	0.02	-0.09	0.06	-0.12	-0.13	0.03
trans_prob.21	-0.08	-0.03	0.16	0.10	-0.15	-0.11	0.16	-0.15	0.03	-0.13
trans_prob.22	-0.07	-0.07	0.20	0.07	-0.12	-0.08	0.16	-0.11	0.05	-0.12
trans_prob.23	-0.12	-0.05	0.16	0.12	-0.04	-0.07	0.14	-0.07	0.01	-0.07
trans_prob.24	-0.15	-0.09	0.06	0.08	0.00	-0.10	0.14	0.15	-0.16	0.04
center.1	0.21	-0.07	0.30	-0.04	-0.03	-0.09	-0.02	0.04	-0.03	0.18
center.2	0.24	-0.04	0.28	-0.08	0.01	-0.07	-0.04	-0.01	-0.05	0.07
center.3	0.26	0.02	0.16	-0.06	0.08	-0.06	-0.10	-0.02	-0.08	-0.06
center.4	0.26	0.08	0.06	-0.07	0.04	-0.09	-0.18	0.06	-0.08	-0.17
center.5	0.21	0.20	-0.02	-0.04	0.03	-0.06	-0.15	0.06	-0.02	-0.30
variance.1	0.13	-0.07	0.30	-0.03	-0.04	-0.15	-0.06	-0.05	-0.13	0.31
variance.2	0.16	-0.05	0.27	-0.05	-0.02	-0.15	-0.09	-0.07	-0.15	0.23
variance.3	0.24	0.05	0.10	-0.05	0.04	-0.10	-0.15	-0.01	-0.13	-0.13
variance.4	0.23	0.14	0.01	-0.05	0.01	-0.09	-0.19	0.07	-0.07	-0.25
variance.5	0.08	0.09	0.01	-0.01	0.00	-0.07	0.00	0.08	-0.10	-0.39
skewness.1	0.08	-0.03	0.04	0.13	-0.18	0.51	-0.13	-0.03	-0.20	-0.08
skewness.2	0.07	-0.02	-0.30	0.31	0.11	-0.24	-0.13	0.03	-0.02	0.00
skewness.3	-0.03	0.18	-0.24	-0.29	-0.23	-0.12	-0.05	0.05	-0.07	-0.07
skewness.4	-0.04	0.33	0.00	0.12	0.15	0.00	-0.03	0.05	0.11	-0.11

Table 27 (cont.)

skewness.5	-0.10	-0.08	0.05	0.09	-0.01	-0.12	0.34	0.16	-0.43	-0.18
kurtosis.1	0.01	-0.02	0.02	0.13	-0.20	0.49	-0.14	0.00	-0.25	-0.02
kurtosis.2	0.14	-0.13	-0.21	0.34	0.07	-0.09	-0.09	0.02	0.02	-0.03
kurtosis.3	0.14	-0.07	-0.17	-0.21	-0.32	-0.02	0.11	-0.04	0.00	-0.05
kurtosis.4	0.12	0.26	-0.06	0.05	0.15	0.10	0.07	-0.12	0.06	-0.01
kurtosis.5	-0.03	-0.05	0.03	0.06	0.03	-0.08	0.29	0.18	-0.45	-0.18
Variable	**PC11**	**PC12**	**PC13**	**PC14**	**PC15**	**PC16**	**PC17**	**PC18**	**PC19**	**PC20**
min.1	-0.11	0.04	-0.11	0.26	0.00	-0.08	0.06	0.00	-0.03	-0.03
min.2	-0.13	0.07	-0.06	-0.10	0.03	-0.16	-0.11	-0.02	0.01	-0.04
min.3	-0.23	0.21	-0.14	0.25	0.14	0.03	0.00	0.04	0.03	0.00
min.4	0.01	0.00	-0.25	0.08	-0.10	-0.24	0.02	0.02	0.03	0.00
min.5	-0.05	0.16	0.35	0.18	-0.06	0.11	-0.07	0.01	-0.15	-0.11
trans_prob.1	0.20	-0.32	0.04	0.22	-0.10	0.15	0.18	0.05	0.03	0.09
trans_prob.2	-0.30	0.29	0.08	-0.53	-0.06	-0.06	0.15	0.03	-0.03	-0.17
trans_prob.3	-0.15	0.00	-0.02	0.17	0.19	-0.14	-0.46	-0.05	-0.12	-0.08
trans_prob.4	0.24	0.15	-0.17	-0.03	0.13	-0.08	-0.02	0.03	0.10	0.08
trans_prob.5	-0.19	0.20	0.16	0.18	-0.14	0.04	0.02	-0.30	0.00	0.07
trans_prob.6	-0.11	0.05	-0.01	0.17	-0.08	0.07	0.32	0.04	-0.08	-0.09
trans_prob.7	0.12	-0.15	0.09	-0.22	0.02	-0.09	-0.40	-0.08	0.00	-0.02
trans_prob.8	-0.09	0.00	-0.04	0.18	-0.02	0.21	0.19	-0.03	0.12	0.19
trans_prob.9	0.16	0.10	-0.15	0.02	0.15	-0.14	0.18	0.15	-0.01	-0.14
trans_prob.11	-0.01	-0.20	0.00	0.26	0.06	-0.20	-0.32	-0.01	-0.17	-0.16
trans_prob.12	0.02	-0.16	0.06	-0.20	-0.14	0.03	0.02	-0.05	0.10	0.10
trans_prob.13	0.03	0.22	-0.06	0.11	0.24	0.08	0.13	0.08	-0.06	-0.09
trans_prob.14	0.10	-0.20	-0.03	-0.17	-0.27	-0.07	0.01	-0.16	0.03	0.14
trans_prob.15	-0.24	-0.03	0.21	0.12	-0.10	0.10	-0.09	0.19	0.16	0.05
trans_prob.16	0.19	0.09	-0.06	0.14	0.20	0.02	-0.06	0.07	0.02	0.10
trans_prob.17	0.07	0.11	0.00	-0.08	0.25	-0.08	0.10	0.11	-0.07	-0.31
trans_prob.18	-0.13	-0.01	0.16	-0.02	-0.02	0.23	-0.03	-0.15	0.01	0.18
trans_prob.19	0.08	-0.12	-0.17	0.02	-0.18	-0.27	0.05	0.09	0.05	-0.04
trans_prob.21	-0.13	0.11	-0.18	0.18	-0.20	-0.24	0.05	-0.60	0.26	-0.23
trans_prob.22	-0.24	0.03	-0.10	0.02	-0.12	-0.35	0.09	0.18	-0.54	0.53
trans_prob.23	-0.35	-0.18	-0.13	0.04	-0.06	-0.04	-0.10	0.48	0.54	-0.04
trans_prob.24	0.01	-0.11	-0.37	-0.16	0.24	0.26	-0.03	-0.26	-0.02	0.17
center.1	0.09	0.03	-0.03	0.04	-0.06	0.00	0.00	-0.04	0.00	-0.07
center.2	-0.01	0.00	0.02	-0.06	0.01	-0.04	0.04	0.00	-0.02	-0.05
center.3	-0.13	-0.11	0.06	-0.01	0.08	-0.10	0.07	-0.03	-0.01	-0.04
center.4	-0.10	-0.12	0.06	-0.02	0.13	-0.06	0.07	-0.06	0.05	0.05
center.5	-0.03	0.03	-0.09	0.03	-0.09	0.10	0.00	0.00	0.01	0.03
variance.1	0.12	0.29	-0.06	-0.02	-0.11	0.11	-0.19	0.05	0.07	0.06
variance.2	0.03	0.27	-0.02	-0.09	-0.06	0.09	-0.17	0.09	0.05	0.12
variance.3	-0.18	-0.09	0.06	0.02	0.15	-0.08	0.11	-0.07	0.03	0.03

Table 27 (cont.)

variance.4	-0.07	-0.03	0.04	0.01	0.14	-0.03	0.10	-0.08	0.10	0.14
variance.5	-0.06	0.01	-0.34	0.04	-0.42	0.43	-0.18	0.12	-0.31	-0.27
skewness.1	-0.07	0.05	-0.01	-0.09	0.00	-0.03	-0.02	0.01	0.03	0.04
skewness.2	-0.14	0.08	0.02	0.00	0.06	0.01	-0.16	-0.03	0.06	0.12
skewness.3	0.05	0.11	-0.07	0.11	0.08	0.04	-0.08	0.00	0.15	0.28
skewness.4	0.22	0.24	0.04	0.04	-0.21	-0.13	-0.04	0.06	0.18	0.20
skewness.5	0.10	0.02	0.21	0.05	-0.01	0.00	-0.03	0.02	0.00	0.00
kurtosis.1	-0.11	0.17	-0.04	-0.12	-0.02	-0.04	-0.05	-0.01	0.04	0.01
kurtosis.2	-0.05	0.12	-0.02	0.00	0.05	0.03	-0.20	-0.08	0.08	0.18
kurtosis.3	0.17	0.20	-0.11	0.02	0.06	0.06	0.01	0.00	0.08	0.12
kurtosis.4	0.19	0.18	0.07	0.08	-0.32	-0.24	0.02	0.09	0.06	-0.04
kurtosis.5	0.16	0.04	0.41	0.09	0.01	-0.05	0.00	0.04	0.02	-0.02

APPENDIX B: CLASSIFICATION RESULTS

Table 28. Classification results with training data.

	Classifier	Accuracy (%)	Kappa (%)	AUC (%)	Sensitivity (%)	Specificity (%)	False Negative (%)	False Positive (%)
Basic NCI	Lasso	77.98	38.72	67.17	91.98	42.36	57.64	8.02
	Neural Network	79.99	46.45	71.43	91.08	51.78	48.22	8.92
	SVM	80.11	46.25	71.13	91.74	50.52	49.48	8.26
	Decision Tree	81.58	51.71	74.35	90.95	57.74	42.26	9.05
	Random Forest	80.25	48.42	72.88	89.80	55.96	44.04	10.20
NCI complete	Lasso	80.05	46.28	71.21	91.49	50.94	49.06	8.51
	Neural Network	76.03	32.33	64.13	91.45	36.82	63.18	8.55
	SVM	81.70	50.76	73.29	92.60	53.97	46.03	7.40
	Decision Tree	82.02	52.80	74.81	91.37	58.26	41.74	8.63
	Random Forest	80.84	49.35	73.04	90.95	55.13	44.87	9.05
NCI Low Corr	Lasso	77.10	35.43	65.51	92.11	38.91	61.09	7.89
	Neural Network	77.89	37.15	66.13	93.13	39.12	60.88	6.87
	SVM	78.93	39.35	66.88	94.53	39.23	60.77	5.47
	Decision Tree	77.57	33.38	63.87	95.31	32.43	67.57	4.69
	Random Forest	76.95	37.13	66.80	90.09	43.51	56.49	9.91
NCI chosen	Lasso	74.68	25.80	60.87	92.56	29.18	70.82	7.44
	Neural Network	73.44	13.29	55.02	97.29	12.76	87.24	2.71
	SVM	78.19	39.79	67.79	91.65	43.93	56.07	8.35
	Decision Tree	79.07	41.95	68.66	92.56	44.77	55.23	7.44
	Random Forest	78.51	42.46	69.57	90.09	49.06	50.94	9.91
NCI Prcomp	Lasso	78.07	39.55	67.71	91.49	43.93	56.07	8.51
	Neural Network	80.73	49.76	73.56	90.01	57.11	42.89	9.99
	SVM	79.16	42.21	68.79	92.60	44.98	55.02	7.40
	Decision Tree	79.69	43.16	69.03	93.50	44.56	55.44	6.50
	Random Forest	78.75	43.49	70.18	89.84	50.52	49.48	10.16
Km complete	Lasso	79.66	45.00	70.54	91.49	49.58	50.42	8.51
	Neural Network	74.09	34.95	67.19	83.02	51.36	48.64	16.98
	SVM	82.11	51.34	73.29	93.54	53.03	46.97	6.46
	Decision Tree	80.31	48.20	72.61	90.30	54.92	45.08	9.70
	Random Forest	79.75	47.34	72.47	89.19	55.75	44.25	10.81
Km low corr	Lasso	79.43	43.61	69.64	92.11	47.18	52.82	7.89
	Neural Network	71.78	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	81.64	49.31	72.04	94.08	50.00	50.00	5.92
	Decision Tree	80.58	50.36	74.31	88.69	59.94	40.06	11.31
	Random Forest	79.55	46.75	72.17	89.10	55.23	44.77	10.90
Km prcomp	Lasso	79.63	44.04	69.78	92.39	47.18	52.82	7.61
	Neural Network	80.96	49.84	73.34	90.83	55.86	44.14	9.17
	SVM	81.58	49.54	72.32	93.59	51.05	48.95	6.41

Table 28 (cont.)

	Decision Tree	79.81	46.31	71.49	90.58	52.41	47.59	9.42
	Random Forest	79.40	45.73	71.43	89.72	53.14	46.86	10.28
HMM all	Lasso	80.25	46.30	71.01	92.23	49.79	50.21	7.77
	Neural Network	71.81	0.24	50.08	99.96	0.21	99.79	0.04
	SVM	85.48	61.09	78.14	94.98	61.30	38.70	5.02
	Decision Tree	79.43	41.72	68.18	94.00	42.36	57.64	6.00
	Random Forest	79.87	45.31	70.58	91.90	49.27	50.73	8.10
HMM low corr	Lasso	79.40	43.11	69.27	92.52	46.03	53.97	7.48
	Neural Network	71.66	21.37	59.50	87.42	31.59	68.41	12.58
	SVM	85.42	59.82	76.80	96.59	57.01	42.99	3.41
	Decision Tree	78.93	39.98	67.32	93.96	40.69	59.31	6.04
	Random Forest	79.84	44.28	69.77	92.89	46.65	53.35	7.11
HMM prcomp	Lasso	79.34	42.62	68.94	92.80	45.08	54.92	7.20
	Neural Network	81.29	50.71	73.76	91.04	56.49	43.51	8.96
	SVM	84.77	58.68	76.69	95.23	58.16	41.84	4.77
	Decision Tree	77.30	35.80	65.62	92.43	38.81	61.19	7.57
	Random Forest	77.80	34.23	64.26	95.35	33.16	66.84	4.65
All	Lasso	81.97	52.06	74.17	92.06	56.28	43.72	7.94
	Neural Network	71.84	0.84	50.30	99.75	0.84	99.16	0.25
	SVM	86.13	63.33	79.58	94.61	64.54	35.46	5.39
	Decision Tree	82.02	53.14	75.16	90.91	59.41	40.59	9.09
	Random Forest	80.31	47.98	72.42	90.54	54.29	45.71	9.46
All Low corr	Lasso	80.61	47.33	71.51	92.39	50.63	49.37	7.61
	Neural Network	72.34	12.33	54.80	95.07	14.54	85.46	4.93
	SVM	86.42	63.97	79.75	95.07	64.44	35.56	4.93
	Decision Tree	82.29	54.40	76.08	90.34	61.82	38.18	9.66
	Random Forest	80.79	48.78	72.59	91.41	53.77	46.23	8.59

Table 29. Classification results by feature set for women's data.

	Classifier	Accuracy (%)	Kappa (%)	AUC (%)	Sensitivity (%)	Specificity (%)	False Negative (%)	False Positive (%)
Basic NCI	Lasso	85.35	40.81	66.44	96.88	36.00	64.00	3.12
	Neural Network	85.35	45.63	70.02	94.70	45.33	54.67	5.30
	SVM	83.84	32.10	62.44	96.88	28.00	72.00	3.12
	Decision Tree	84.85	43.76	69.20	94.39	44.00	56.00	5.61
	Random Forest	83.08	36.08	65.55	93.77	37.33	62.67	6.23
NCI complete	Lasso	84.60	37.35	64.95	96.57	33.33	66.67	3.43
	Neural Network	85.10	45.64	70.37	94.08	46.67	53.33	5.92
	SVM	84.34	34.22	63.26	97.20	29.33	70.67	2.80
	Decision Tree	84.34	40.50	67.35	94.70	40.00	60.00	5.30
	Random Forest	85.10	43.71	68.84	95.02	42.67	57.33	4.98
NCI Low Corr	Lasso	84.09	30.88	61.58	97.82	25.33	74.67	2.18
	Neural Network	83.33	21.73	57.53	99.07	16.00	84.00	0.93
	SVM	83.59	26.69	59.73	98.13	21.33	78.67	1.87
	Decision Tree	84.09	39.17	66.69	94.70	38.67	61.33	5.30
	Random Forest	82.83	31.47	62.84	95.02	30.67	69.33	4.98
NCI chosen	Lasso	83.08	19.95	56.87	99.07	14.67	85.33	0.93
	Neural Network	82.58	21.07	57.58	97.82	17.33	82.67	2.18
	SVM	83.33	19.35	56.51	99.69	13.33	86.67	0.31
	Decision Tree	82.58	33.38	64.22	93.77	34.67	65.33	6.23
	Random Forest	83.33	32.64	63.15	95.64	30.67	69.33	4.36
NCI Prcomp	Lasso	83.59	30.58	61.78	96.88	26.67	73.33	3.12
	Neural Network	82.83	33.95	64.37	94.08	34.67	65.33	5.92
	SVM	83.33	22.86	58.04	98.75	17.33	82.67	1.25
	Decision Tree	82.07	35.33	65.95	91.90	40.00	60.00	8.10
	Random Forest	84.34	36.72	64.80	96.26	33.33	66.67	3.74
Km complete	Lasso	84.34	36.72	63.62	96.26	33.33	66.67	3.74
	Neural Network	81.06	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	83.59	26.69	59.73	98.13	21.33	78.67	1.87
	Decision Tree	84.60	39.66	66.49	95.64	37.33	62.67	4.36
	Random Forest	84.09	38.44	66.17	95.02	37.33	62.67	4.98
Km low corr	Lasso	84.09	34.46	63.62	96.57	30.67	69.33	3.43
	Neural Network	81.57	30.36	63.08	92.83	33.33	66.67	7.17
	SVM	83.33	23.97	58.55	98.44	18.67	81.33	1.56
	Decision Tree	82.83	23.84	58.75	97.51	20.00	80.00	2.49
	Random Forest	83.59	34.08	63.82	95.64	32.00	68.00	4.36
Km prcomp	Lasso	83.59	31.49	62.29	96.57	28.00	72.00	3.43
	Neural Network	84.85	39.53	66.13	96.26	36.00	64.00	3.74
	SVM	83.84	26.27	59.38	98.75	20.00	80.00	1.25
	Decision Tree	82.07	28.00	61.35	94.70	28.00	72.00	5.30
	Random Forest	84.09	37.68	65.66	95.33	36.00	64.00	4.67

Table 29 (cont.)

HMM all	Lasso	84.60	38.14	65.46	96.26	34.67	65.33	3.74
	Neural Network	82.32	41.84	70.70	89.41	52.00	48.00	10.59
	SVM	83.33	23.97	58.55	98.44	18.67	81.33	1.56
	Decision Tree	82.58	29.13	61.66	95.33	28.00	72.00	4.67
	Random Forest	83.84	33.84	63.46	96.26	30.67	69.33	3.74
HMM low corr	Lasso	84.34	35.08	63.78	96.88	30.67	69.33	3.12
	Neural Network	81.82	19.36	57.11	96.88	17.33	82.67	3.12
	SVM	82.07	12.57	54.20	99.07	9.33	90.67	0.93
	Decision Tree	83.08	32.05	63.00	95.33	30.67	69.33	4.67
	Random Forest	83.84	33.84	63.46	96.26	30.67	69.33	3.74
HMM prcomp	Lasso	85.10	38.62	65.26	97.20	33.33	66.67	2.80
	Neural Network	82.83	34.74	64.88	93.77	36.00	64.00	6.23
	SVM	81.82	14.61	55.07	98.13	12.00	88.00	1.87
	Decision Tree	81.06	20.95	58.17	95.02	21.33	78.67	4.98
	Random Forest	82.83	19.35	56.71	98.75	14.67	85.33	1.25
All	Lasso	84.85	40.28	66.64	95.95	37.33	62.67	4.05
	Neural Network	80.30	-1.48	49.53	99.07	0.00	100.00	0.93
	SVM	84.60	34.85	63.42	97.51	29.33	70.67	2.49
	Decision Tree	82.58	36.43	66.26	92.52	40.00	60.00	7.48
	Random Forest	84.85	39.53	66.13	96.26	36.00	64.00	3.74
All Low corr	Lasso	83.84	34.68	63.98	95.95	32.00	68.00	4.05
	Neural Network	81.06	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	82.58	21.07	57.58	97.82	17.33	82.67	2.18
	Decision Tree	81.31	28.98	62.42	92.83	32.00	68.00	7.17
	Random Forest	82.83	26.89	60.29	96.57	24.00	76.00	3.43

Table 30. Classification results by feature set for men's data.

	Classifier	Accuracy (%)	Kappa (%)	AUC (%)	Sensitivity (%)	Specificity (%)	False Negative (%)	False Positive (%)
Basic NCI	Lasso	79.20	55.07	77.09	85.77	68.42	31.58	14.23
	Neural Network	80.75	58.27	78.57	87.54	69.59	30.41	12.46
	SVM	80.09	56.58	77.58	87.90	67.25	32.75	12.10
	Decision Tree	80.53	58.99	79.76	82.92	76.61	23.39	17.08
	Random Forest	76.99	50.52	74.97	83.27	66.67	33.33	16.73
NCI complete	Lasso	79.20	55.99	78.01	82.56	73.68	26.32	17.44
	Neural Network	75.22	49.74	76.07	72.60	79.53	20.47	27.40
	SVM	80.31	57.61	78.44	86.12	70.76	29.24	13.88
	Decision Tree	80.53	58.99	79.76	82.92	76.61	23.39	17.08
	Random Forest	80.53	57.24	77.70	89.32	66.08	33.92	10.68
NCI Low Corr	Lasso	75.44	46.13	72.35	85.05	59.65	40.35	14.95
	Neural Network	74.56	41.85	69.46	90.39	48.54	51.46	9.61
	SVM	74.56	42.97	70.38	87.54	53.22	46.78	12.46
	Decision Tree	74.78	44.74	71.70	84.34	59.06	40.94	15.66
	Random Forest	74.78	44.21	71.25	85.77	56.73	43.27	14.23
NCI chosen	Lasso	71.68	36.14	67.04	86.12	47.95	52.05	13.88
	Neural Network	73.23	41.96	70.57	81.49	59.65	40.35	18.51
	SVM	73.67	42.11	70.36	83.99	56.73	43.27	16.01
	Decision Tree	73.01	42.36	71.08	79.00	63.16	36.84	21.00
	Random Forest	75.88	46.97	72.71	85.77	59.65	40.35	14.23
NCI Prcomp	Lasso	76.55	49.45	74.39	83.27	65.50	34.50	16.73
	Neural Network	78.10	52.73	75.97	84.70	67.25	32.75	15.30
	SVM	77.21	49.89	74.12	86.83	61.40	38.60	13.17
	Decision Tree	76.77	48.18	72.96	88.61	57.31	42.69	11.39
	Random Forest	74.12	42.67	70.48	85.41	55.56	44.44	14.59
Km complete	Lasso	79.42	55.08	76.75	87.54	66.08	33.92	12.46
	Neural Network	79.20	56.58	78.81	80.43	77.19	22.81	19.57
	SVM	80.09	56.06	77.00	89.68	64.33	35.67	10.32
	Decision Tree	77.65	51.44	75.16	85.41	64.91	35.09	14.59
	Random Forest	80.09	57.08	78.15	86.12	70.18	29.82	13.88
Km low corr	Lasso	80.31	56.81	77.52	88.97	66.08	33.92	11.03
	Neural Network	74.78	45.76	72.62	81.49	63.74	36.26	18.51
	SVM	79.20	53.56	75.49	90.75	60.23	39.77	9.25
	Decision Tree	78.98	54.22	76.46	86.83	66.08	33.92	13.17
	Random Forest	79.87	56.65	77.97	85.77	70.18	29.82	14.23
Km prcomp	Lasso	79.42	55.08	76.81	87.54	66.08	33.92	12.46
	Neural Network	79.42	54.54	76.24	89.32	63.16	36.84	10.68
	SVM	80.97	57.41	77.26	92.53	61.99	38.01	7.47
	Decision Tree	77.21	52.80	77.09	77.58	76.61	23.39	22.42
	Random Forest	80.97	58.51	78.52	88.61	68.42	31.58	11.39

Table 30 (cont.)

HMM all	Lasso	79.42	54.65	76.36	88.97	63.74	36.26	11.03
	Neural Network	69.69	39.38	70.93	65.84	76.02	23.98	34.16
	SVM	78.98	54.00	76.23	87.54	64.91	35.09	12.46
	Decision Tree	75.88	48.32	73.97	81.85	66.08	33.92	18.15
	Random Forest	78.32	52.27	75.24	87.90	62.57	37.43	12.10
HMM low corr	Lasso	76.11	46.25	71.85	89.32	54.39	45.61	10.68
	Neural Network	67.48	30.46	65.14	74.73	55.56	44.44	25.27
	SVM	77.88	50.59	74.08	89.68	58.48	41.52	10.32
	Decision Tree	75.88	48.20	73.85	82.21	65.50	34.50	17.79
	Random Forest	78.54	53.36	76.10	86.12	66.08	33.92	13.88
HMM prcomp	Lasso	75.88	46.47	72.25	87.19	57.31	42.69	12.81
	Neural Network	77.43	50.55	74.52	86.48	62.57	37.43	13.52
	SVM	76.99	49.47	73.94	86.48	61.40	38.60	13.52
	Decision Tree	69.69	31.57	64.86	84.70	45.03	54.97	15.30
	Random Forest	76.33	45.76	71.23	92.17	50.29	49.71	7.83
All	Lasso	81.42	59.75	79.33	87.90	70.76	29.24	12.10
	Neural Network	76.11	49.55	74.95	79.72	70.18	29.82	20.28
	SVM	80.75	58.17	78.45	87.90	69.01	30.99	12.10
	Decision Tree	78.10	53.38	76.66	82.56	70.76	29.24	17.44
	Random Forest	81.19	59.42	79.27	87.19	71.35	28.65	12.81
All Low corr	Lasso	79.20	53.67	75.61	90.39	60.82	39.18	9.61
	Neural Network	67.92	26.68	62.41	85.05	39.77	60.23	14.95
	SVM	77.65	50.28	74.02	88.97	59.06	40.94	11.03
	Decision Tree	77.88	52.64	76.14	83.27	69.01	30.99	16.73
	Random Forest	80.31	57.11	77.87	87.90	67.84	32.16	12.10

Table 31. Classification results of alternative outcome – Blood Pressure.

	Classifier	Accuracy (%)	Kappa (%)	AUC (%)	Sensitivity (%)	Specificity (%)	False Negative (%)	False Positive (%)
Basic NCI	Logistic Regression	72.52	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	73.47	10.37	54.98	97.40	10.30	89.70	2.60
	SVM	72.64	1.38	50.73	99.67	1.29	98.71	0.33
	Decision Tree	72.52	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	72.41	14.31	56.64	92.68	18.88	81.12	7.32
NCI complete	Lasso	72.52	0.77	50.53	99.67	0.86	99.14	0.33
	Neural Network	71.34	2.14	52.90	96.42	5.15	94.85	3.58
	SVM	72.52	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	72.52	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	73.23	12.40	55.78	95.77	13.73	86.27	4.23
NCI Low Corr	Logistic Regression	72.52	0.00	50.61	100.00	0.00	100.00	0.00
	Neural Network	72.17	13.30	60.53	92.68	18.03	81.97	7.32
	SVM	72.76	2.00	50.93	99.67	1.72	98.28	0.33
	Decision Tree	72.52	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	73.23	13.30	54.99	95.28	15.02	84.98	4.72
NCI chosen	Logistic Regression	72.64	0.62	50.53	100.00	0.43	99.57	0.00
	Neural Network	72.41	-0.24	49.92	99.84	0.00	100.00	0.16
	SVM	72.41	-0.24	50.20	99.84	0.00	100.00	0.16
	Decision Tree	72.52	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	72.17	7.80	52.64	95.61	10.30	89.70	4.39
NCI Prcomp	Logistic Regression	72.52	0.00	50.73	100.00	0.00	100.00	0.00
	Neural Network	73.11	12.17	56.27	95.61	13.73	86.27	4.39
	SVM	72.52	0.39	50.41	99.84	0.43	99.57	0.16
	Decision Tree	72.52	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	72.41	10.78	55.48	94.63	13.73	86.27	5.37
Km complete	Lasso	72.64	8.40	56.28	96.42	9.87	90.13	3.58
	Neural Network	72.52	0.39	50.41	99.84	0.43	99.57	0.16
	SVM	72.76	2.74	51.17	99.35	2.58	97.42	0.65
	Decision Tree	73.23	7.92	52.09	98.05	7.73	92.27	1.95
	Random Forest	71.46	10.50	57.42	92.52	15.88	84.12	7.48
Km low corr	Logistic Regression	72.52	1.15	54.75	99.51	1.29	98.71	0.49
	Neural Network	72.52	0.39	50.12	99.84	0.43	99.57	0.16
	SVM	73.47	6.33	51.82	99.35	5.15	94.85	0.65
	Decision Tree	72.64	10.62	55.69	95.28	12.88	87.12	4.72
	Random Forest	70.64	6.51	56.45	92.68	12.45	87.55	7.32
Km prcomp	Logistic Regression	72.52	1.52	54.34	99.35	1.72	98.28	0.65
	Neural Network	71.93	6.36	54.11	95.77	9.01	90.99	4.23
	SVM	72.41	0.53	50.44	99.51	0.86	99.14	0.49
	Decision Tree	72.52	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	71.82	9.34	54.95	93.98	13.30	86.70	6.02

Table 31 (cont.)

HMM all	Lasso	73.23	8.91	55.60	97.56	9.01	90.99	2.44
	Neural Network	72.52	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	72.52	0.77	50.81	99.67	0.86	99.14	0.33
	Decision Tree	73.82	19.24	60.03	93.33	22.32	77.68	6.67
	Random Forest	72.76	11.47	58.31	95.12	13.73	86.27	4.88
HMM low corr	Logistic Regression	72.29	-0.08	53.69	99.51	0.43	99.57	0.49
	Neural Network	72.52	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	72.64	1.38	51.02	99.67	1.29	98.71	0.33
	Decision Tree	72.88	7.88	53.23	97.24	8.58	91.42	2.76
	Random Forest	72.17	7.80	56.64	95.61	10.30	89.70	4.39
HMM prcomp	Logistic Regression	72.17	-0.70	53.97	99.51	0.00	100.00	0.49
	Neural Network	71.70	10.04	56.65	93.33	14.59	85.41	6.67
	SVM	72.41	0.15	50.61	99.67	0.43	99.57	0.33
	Decision Tree	72.52	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	72.76	5.95	53.97	97.89	6.44	93.56	2.11
All	Lasso	72.64	9.68	57.34	95.77	11.59	88.41	4.23
	Neural Network	72.52	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	72.29	1.43	51.87	98.86	2.15	97.85	1.14
	Decision Tree	71.93	16.70	62.61	90.08	24.03	75.97	9.92
	Random Forest	72.17	10.33	58.18	94.31	13.73	86.27	5.69
All Low corr	Logistic Regression	71.82	-0.63	55.31	98.70	0.86	99.14	1.30
	Neural Network	72.52	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	72.05	0.21	51.46	98.86	1.29	98.71	1.14
	Decision Tree	70.99	11.95	61.20	90.57	19.31	80.69	9.43
	Random Forest	72.05	8.86	57.90	94.80	12.02	87.98	5.20

Table 32. Classification results of alternative outcome – Total Cholesterol.

	Classifier	Accuracy (%)	Kappa (%)	AUC (%)	Sensitivity (%)	Specificity (%)	False Negative (%)	False Positive (%)
Basic NCI	Logistic Regression	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	85.02	-0.70	50.04	99.59	0.00	100.00	0.41
NCI complete	Lasso	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	85.14	-0.47	50.12	99.72	0.00	100.00	0.28
NCI Low Corr	Logistic Regression	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	85.38	0.00	50.00	100.00	0.00	100.00	0.00
NCI chosen	Logistic Regression	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	85.26	0.89	50.04	99.72	0.81	99.19	0.28
NCI Prcomp	Logistic Regression	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	85.26	-0.23	49.92	99.86	0.00	100.00	0.14
	SVM	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	85.14	-0.47	50.41	99.72	0.00	100.00	0.28
Km complete	Lasso	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	85.38	0.00	50.00	100.00	0.00	100.00	0.00
Km low corr	Logistic Regression	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	85.26	-0.23	49.92	99.86	0.00	100.00	0.14
Km prcomp	Logistic Regression	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	85.14	-0.47	50.12	99.72	0.00	100.00	0.28

Table 32 (cont.)

HMM all	Lasso	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	85.38	0.00	50.00	100.00	0.00	100.00	0.00
HMM low corr	Logistic Regression	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	85.38	0.00	50.00	100.00	0.00	100.00	0.00
HMM prcomp	Logistic Regression	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	84.91	-0.92	49.95	99.45	0.00	100.00	0.55
	SVM	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	85.38	0.00	50.00	100.00	0.00	100.00	0.00
All	Lasso	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	85.38	0.00	50.00	100.00	0.00	100.00	0.00
All Low corr	Logistic Regression	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	85.38	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	85.38	0.00	50.00	100.00	0.00	100.00	0.00

Table 33. Classification results of alternative outcome – HDL.

	Classifier	Accuracy (%)	Kappa (%)	AUC (%)	Sensitivity (%)	Specificity (%)	False Negative (%)	False Positive (%)
Basic NCI	Logistic Regression	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	72.64	2.97	50.58	97.74	4.39	95.61	2.26
NCI complete	Lasso	72.88	-0.47	49.83	99.68	0.00	100.00	0.32
	Neural Network	72.76	-0.30	49.58	99.35	0.44	99.56	0.65
	SVM	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	72.52	-0.76	49.99	99.03	0.44	99.56	0.97
NCI Low Corr	Logistic Regression	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	72.52	1.21	49.73	98.23	2.63	97.37	1.77
NCI chosen	Logistic Regression	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	73.11	0.80	49.67	99.68	0.88	99.12	0.32
	SVM	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	72.76	0.50	49.54	99.03	1.32	98.68	0.97
NCI Prcomp	Logistic Regression	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	73.00	-0.24	49.92	99.84	0.00	100.00	0.16
	SVM	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	71.93	0.46	50.30	97.26	3.07	96.93	2.74
Km complete	Lasso	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	72.64	-0.13	49.62	99.03	0.88	99.12	0.97
Km low corr	Logistic Regression	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	73.11	0.80	49.67	99.68	0.88	99.12	0.32
Km prcomp	Logistic Regression	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	72.88	0.73	50.77	99.19	1.32	98.68	0.81
	SVM	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	72.05	-1.68	49.95	98.39	0.44	99.56	1.61

Table 33 (cont.)

HMM all	Lasso	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	73.00	-0.24	49.92	99.84	0.00	100.00	0.16
HMM low corr	Logistic Regression	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	73.00	-0.24	49.92	99.84	0.00	100.00	0.16
HMM prcomp	Logistic Regression	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	71.23	2.11	50.19	95.00	6.58	93.42	5.00
	SVM	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	72.88	-0.47	50.12	99.68	0.00	100.00	0.32
All	Lasso	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	72.88	-0.47	49.83	99.68	0.00	100.00	0.32
All Low corr	Logistic Regression	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	73.11	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	73.11	0.00	50.00	100.00	0.00	100.00	0.00

Table 34. Classification results of alternative outcome – CRP.

	Classifier	Accuracy (%)	Kappa (%)	AUC (%)	Sensitivity (%)	Specificity (%)	False Negative (%)	False Positive (%)
Basic NCI	Logistic Regression	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	89.15	-0.23	50.20	99.87	0.00	100.00	0.13
	SVM	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	89.15	-0.23	50.20	99.87	0.00	100.00	0.13
NCI complete	Lasso	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	89.27	0.00	50.00	100.00	0.00	100.00	0.00
NCI Low Corr	Logistic Regression	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	89.15	1.45	50.61	99.74	1.10	98.90	0.26
	SVM	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	89.27	0.00	50.00	100.00	0.00	100.00	0.00
NCI chosen	Logistic Regression	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	89.15	-0.23	50.20	99.87	0.00	100.00	0.13
NCI Prcomp	Logistic Regression	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	89.15	-0.23	49.92	99.87	0.00	100.00	0.13
	SVM	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	89.27	0.00	50.00	100.00	0.00	100.00	0.00
Km complete	Lasso	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	89.03	-0.46	50.41	99.74	0.00	100.00	0.26
Km low corr	Logistic Regression	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	89.27	0.00	50.00	100.00	0.00	100.00	0.00
Km prcomp	Logistic Regression	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	89.15	-0.23	49.92	99.87	0.00	100.00	0.13

Table 34 (cont.)

HMM all	Lasso	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	89.27	0.00	50.00	100.00	0.00	100.00	0.00
HMM low corr	Logistic Regression	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	89.27	0.00	50.00	100.00	0.00	100.00	0.00
HMM prcomp	Logistic Regression	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	88.92	0.98	49.58	99.47	1.10	98.90	0.53
	SVM	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	89.27	0.00	50.00	100.00	0.00	100.00	0.00
All	Lasso	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	89.03	-0.46	50.41	99.74	0.00	100.00	0.26
All Low corr	Logistic Regression	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Neural Network	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	SVM	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Decision Tree	89.27	0.00	50.00	100.00	0.00	100.00	0.00
	Random Forest	89.27	0.00	50.00	100.00	0.00	100.00	0.00

APPENDIX C: FEATURE IMPORTANCE MEASURES

Table 35. Lasso coefficients and RF importance measures of NCI features.

Features	Lasso Coefficients	Mean Decrease in Accuracy	Mean Decrease in Gini
gender	1.86	64.41	98.36
sed_min	0	18.83	58.18
light_min	0	16.11	58.27
life_min	-0.04	28.77	85.52
mod_min	-0.12	28.48	80.18
vig_min	-0.17	17.16	14.62
sed_bouted_60min	0.01	17.74	49.28
num_mvpa_bouts	-0.12	13.96	12.04
num_vig_bouts	-1.32	7.45	1.17
mvpa_bouted	0.11	12.37	17.04
vig_bouted	-0.08	5.24	1.12
tot_mv_bouts	-0.01	10.56	9.19
tot_mv_min	0	22.51	60.17
tot_li_min	0	18.16	59.15
wk_mv_dif_bout	-0.1	11.18	12.8
avg_wk_mv_dif_min	0.01	16.39	40.68
avg_wk_li_dif_min	0	18.06	47.85
tot_wk_mv_dif_min	0	17.19	42.32
tot_wk_li_dif_min	0	16.91	41.79
perc_wk_mv	0.11	17.3	35.03
perc_we_mv	0.8	17.24	34.45
perc_wk_li	0.08	18.42	39.45
perc_we_li	-0.02	15.37	40.2
mv_wk2	0.04	4.9	3.59
top_mv.1	0	20.08	42.59
top_mv.2	0	22.86	52.46
top_mv.3	0	23.21	53.39
top_mv.4	0.02	21.28	42.66
top_li.1	0.01	20.14	53.16
top_li.2	0.01	18.18	62.02
top_li.3	0.01	23.01	67.92
top_li.4	0.01	20.19	54.26

Table 36. Lasso coefficients and RF importance measures of NCI basic features.

Features	Lasso Coefficients	Mean Decrease in Accuracy	Mean Decrease in Gini
gender	1.7	84.42	112.86
sed_min	0	17.98	223.69
light_min	0	18.97	219.2
life_min	-0.02	56.71	331.23
mod_min	-0.04	72.02	329.42
vig_min	-0.13	20.17	60.57

Table 37. Lasso coefficients and RF importance measures of NCI low correlation features.

Features	Lasso Coefficients	Mean Decrease in Accuracy	Mean Decrease in Gini
gender	1.49	59.11	78.64
sed_min	0	24.99	199.5
light_min	0	10.69	185.44
sed_bouted_60min	0.01	27.28	166.85
num_vig_bouts	-3.28	16.87	10.22
wk_mv_dif_bout	0.03	18.97	68.58
perc_we_mv	0.58	24.97	149.72
perc_we_li	0	19.38	161.51
mv_wk2	-0.18	10.58	25.63
top_li.1	-0.01	66.3	297.09

Table 38. Lasso coefficients and RF importance measures of NCI chosen features.

Features	Lasso Coefficients	Mean Decrease in Accuracy	Mean Decrease in Gini
gender	1.12	68.28	82.07
sed_min	0	26.34	178.08
light_min	0	19.85	167.82
vig_min	-0.29	30.91	66.1
sed_bouted_60min	0.01	30.4	150.63
wk_mv_dif_bout	0.28	18.59	56.56
avg_wk_mv_dif_min	-0.01	54.16	218.13
avg_wk_li_dif_min	0	31.21	155.75
perc_wk_mv	-0.28	28.63	131.79
perc_wk_li	0.1	25.91	134.86
mv_wk2	-0.25	9.41	19.53

Table 39. Lasso coefficients and RF importance measures of K-means features.

Features	Lasso Coefficients	Mean Decrease in Accuracy	Mean Decrease in Gini
gender	2.04	67.68	108.38
min.1	0	15.76	34.57
min.2	0	14.62	33.86
min.3	0	18.75	36.68
min.4	0	19.74	39.17
min.5	0	18.44	39.06
threshold.1	0	18.57	42.83
threshold.2	-0.01	19.36	53.14
threshold.3	0	19.47	45.18
threshold.4	0	18.16	37.21
threshold.5	0	17.84	37.05
center.1	-0.04	20.97	61.47
center.2	0	19.52	46.77
center.3	0	18.83	52.65
center.4	0	19.91	45.5
center.5	0	16.93	34.51
variance.1	0	21.38	75.85
variance.2	0	20.36	58.34
variance.3	0	18.05	47.84
variance.4	0	16.48	35.31
variance.5	0	20.8	38.74
skewness.1	-0.2	15.46	32.2
skewness.2	-3.66	10.98	35.77
skewness.3	-0.87	13.83	38.3
skewness.4	0.27	18.96	44.16
skewness.5	-0.05	11.08	33.68
kurtosis.1	0	15.49	32.57
kurtosis.2	2.16	11.45	36.43
kurtosis.3	0.29	13.86	39.12
kurtosis.4	0.08	20.01	42.64
kurtosis.5	0.02	8.05	33.52

Table 40. Lasso coefficients and RF importance measures of K-means low correlation features.

Features	Lasso Coefficients	Mean Decrease in Accuracy	Mean Decrease in Gini
gender	1.9	84.8	125.91
min.5	0	23.5	99.03
threshold.5	0	28.65	111.81
center.2	-0.01	47.95	184.1
variance.3	0	43.09	165.36
variance.4	0	33.86	109.8
variance.5	0	30.77	97.16
skewness.1	0.3	6.25	95.15
skewness.2	-0.67	13.19	90.88
skewness.5	0	14.02	87.41
kurtosis.3	0.44	19.93	100.46
kurtosis.4	0.08	26.5	103.51

Table 41. Lasso coefficients and RF importance measures of HMM features.

Features	Lasso Coefficients	Mean Decrease in Accuracy	Mean Decrease in Gini
gender	1.82	42.19	64.49
min.1	0	4.87	19.52
min.2	0	7.76	23.33
min.3	0	6.38	22.2
min.4	0	12.65	33.81
min.5	0	21.07	39.5
trans_prob.1	0	9.52	21.28
trans_prob.2	-2.33	4.5	21.19
trans_prob.3	1.18	7.3	21.39
trans_prob.4	-0.8	11.39	26.97
trans_prob.5	-12.83	21.42	37.67
trans_prob.6	0.43	5.01	20.97
trans_prob.7	0.05	8.93	21.04
trans_prob.8	0	5.7	20.25
trans_prob.9	0	8.52	24.09
trans_prob.11	1.16	7.72	20.49
trans_prob.12	-0.29	6.16	19.85
trans_prob.13	0	9.53	21.41
trans_prob.14	0	6.76	23.44
trans_prob.15	-8.99	22.63	42.24
trans_prob.16	-5.33	11.04	23.04
trans_prob.17	3.11	6.84	21.75
trans_prob.18	2.18	5.47	20.76
trans_prob.19	3.4	11.89	23.56
trans_prob.21	-0.83	11.71	25.17
trans_prob.22	-3.62	9.23	25.37
trans_prob.23	-3.24	10.05	23.56

Table 41 (cont.)

trans_prob.24	-0.49	14.38	25.97
center.1	-0.09	18.7	28.89
center.2	0	16.81	28.34
center.3	0	15.14	29.31
center.4	0	24.51	54.28
center.5	0	26.7	61.7
variance.1	0	18.6	30.04
variance.2	0	15.5	29.25
variance.3	0	16.18	28.8
variance.4	0	29.21	65.42
variance.5	0	29.84	59.53
skewness.1	0	7.51	21.94
skewness.2	0	6.73	21.24
skewness.3	-0.42	6.66	20.61
skewness.4	-0.63	10.96	22.3
skewness.5	-0.05	9.06	23.13
kurtosis.1	-0.01	8	20.83
kurtosis.2	0	9.74	21.46
kurtosis.3	0	7.13	20.02
kurtosis.4	0.21	15.78	29.42
kurtosis.5	0	5.44	20.34

Table 42. Lasso coefficients and RF importance measures of HMM low correlation features.

Features	Lasso Coefficients	Mean Decrease in Accuracy	Mean Decrease in Gini
gender	1.68	46.19	71.01
min.1	0	7.87	25.71
min.2	0	11.69	35.09
min.3	0	7.09	28.83
min.4	0	15.22	40.25
min.5	0	25.45	51.66
trans_prob.1	0	10.79	25.95
trans_prob.2	-0.56	3.52	26.95
trans_prob.3	2.6	12.29	29.7
trans_prob.4	0	10.32	29.04
trans_prob.5	-8.88	19.78	40.72
trans_prob.6	1.91	8.18	28.09
trans_prob.7	0	9.41	25.33
trans_prob.8	0	9.63	26.4
trans_prob.9	0	10.8	27.66
trans_prob.11	0.7	7.45	26.77
trans_prob.12	0	4.36	23.69
trans_prob.13	0.46	8.03	25.04
trans_prob.14	0	7.9	26.89
trans_prob.15	-5.73	24.41	44.51
trans_prob.16	-3.68	10.09	26.32
trans_prob.17	1.03	6.4	27.95
trans_prob.18	0	6.98	24.34
trans_prob.19	1.91	11.46	26.7
trans_prob.21	-0.77	11.45	32.03
trans_prob.22	-3.32	10.96	31.13
trans_prob.23	-2.89	10.62	28.46
trans_prob.24	0	12.68	30.44
center.5	0	43.7	114.88
variance.2	0	31.42	66.17
variance.5	0	38.61	96.84
skewness.1	-0.02	9	30.45
skewness.2	0.4	7.63	27.71
skewness.3	0.35	8.69	26.7
skewness.4	0	10.32	28.26
skewness.5	-0.09	10.56	33.78
kurtosis.3	0	7.08	24.12
kurtosis.4	0.74	19.51	36.1

Table 43. Lasso coefficients and RF importance measures of all features.

	Features	Lasso Coefficients	Mean Decrease in Accuracy	Mean Decrease in Gini
	gender	2.02	37.99	65.24
	sed_min	0	10.07	16.15
	light_min	-0.01	13.29	15.33
	life_min	-0.01	17.46	32.43
	mod_min	-0.03	16.66	31.58
	vig_min	0	5.97	2.12
	sed_bouted_60min	0	11.72	13.44
	num_mvpa_bouts	0	6.43	2.38
	num_vig_bouts	-0.47	2.58	0.16
	mvpa_bouted	0.04	6.8	3.48
	vig_bouted	-0.05	1.92	0.2
	tot_mv_bouts	0	3.97	1.99
	tot_mv_min	0	11.33	15.59
	tot_li_min	0	10.62	16.17
	wk_mv_dif_bout	0	2.22	3.87
	avg_wk_mv_dif_min	0	6.6	9.43
	avg_wk_li_dif_min	0	8.45	11.35
	tot_wk_mv_dif_min	0	7.79	9.9
	tot_wk_li_dif_min	0	8.57	10.06
	perc_wk_mv	0	5.29	8.25
	perc_we_mv	0.24	5.82	8.46
	perc_wk_li	0.05	3.46	9.06
	perc_we_li	0	2.54	9
	mv_wk2	0	2.53	0.86
	top_mv.1	0	10.41	13.86
	top_mv.2	0	12.12	20.31
	top_mv.3	0	11.04	19.49
	top_mv.4	0	9.54	12.93
	top_li.1	0	13.52	18.16
	top_li.2	0	11.66	20.13
	top_li.3	0	12.4	22.12
NCI	top_li.4	0	9.68	15.21
	min.1	0	7.46	9.19
	min.2	0	5.97	9.57
	min.3	0	8.08	9.71
	min.4	0	8.4	10.58
	threshold.1	0	12.2	16.43
	threshold.2	0	11.85	22.02
	threshold.3	0	12.29	17.92
	threshold.4	0	10.89	11.99
K-means	threshold.5	0	10.94	11.8

Table 43 (cont.)

	center.1	0	13.31	20.01
	center.2	0	12.09	15.57
	center.3	0	13.5	21.66
	center.4	0	12.46	16.47
	center.5	0	11.48	10.94
	variance.1	0	13.4	28.16
	variance.2	0	13	20.02
	variance.3	0	13.27	19.04
	variance.4	0	10.89	10.75
	variance.5	0	8.51	11.26
	skewness.1	0	6.33	9.4
	skewness.2	-0.83	6.84	10.33
	skewness.3	0	7.5	10.6
	skewness.4	0.02	10.39	13.59
	skewness.5	0	5.35	10.36
	kurtosis.1	0	6.44	9.23
	kurtosis.2	0.33	5.05	10.46
	kurtosis.3	0	6.67	11.09
	kurtosis.4	0.02	9.59	13.03
K-means	kurtosis.5	0	3.77	10.42
	min.1.1	0	5.15	9.68
	min.2.1	0	6.86	10.37
	min.3.1	0	5.1	9.34
	min.4.1	0	7.44	10.2
	min.5.1	0	6.17	9.93
	trans_prob.1	0	6.41	10.19
	trans_prob.2	-1.28	1.09	11.35
	trans_prob.3	0.85	6.92	10.75
	trans_prob.4	-1.31	7.89	12.2
	trans_prob.5	-11.54	11.23	15.88
	trans_prob.6	0	4.55	11.26
	trans_prob.7	1.07	5.69	10.87
	trans_prob.8	-0.19	3.32	10.85
	trans_prob.9	0	5.06	12.19
	trans_prob.11	0	3.83	10.25
	trans_prob.12	0	3.4	10.8
	trans_prob.13	0	5.1	10.58
	trans_prob.14	0.1	4.53	10.96
	trans_prob.15	-5.64	11.75	15.97
	trans_prob.16	-7.2	7.7	12.77
	trans_prob.17	0	5.08	11.1
HMM	trans_prob.18	0	3.6	11.28

Table 43 (cont.)

	trans_prob.19	1.75	8.33	13.31
	trans_prob.21	-0.52	8.33	14.45
	trans_prob.22	-3.11	9.27	14.52
	trans_prob.23	-2.63	7.46	14.08
	trans_prob.24	-0.3	12.08	17.07
	center.1.1	-0.01	9.05	10.89
	center.2.1	0	8.9	8.99
	center.3.1	0	6.65	8.42
	center.4.1	0	9.05	10
	center.5.1	0	10.42	11.49
	variance.1.1	0	10.71	11.2
	variance.2.1	0	7.85	9.05
	variance.3.1	0	8.31	8.5
	variance.4.1	0	8.98	11.15
	variance.5.1	0	13.28	12.94
	skewness.1.1	0	5.11	11.03
	skewness.2.1	0	1.93	11.63
	skewness.3.1	-0.13	4.67	9.77
	skewness.4.1	-0.02	7.59	10.95
	skewness.5.1	0.1	5.34	10.74
	kurtosis.1.1	0	4.61	10.82
	kurtosis.2.1	0	7.19	12.07
	kurtosis.3.1	0	6.48	10.69
	kurtosis.4.1	0	8.62	11
HMM	kurtosis.5.1	0	4.17	9.5

Figure 27. Mean decrease in accuracy and Gini coefficient for all NCI features based on the random forest classifier.

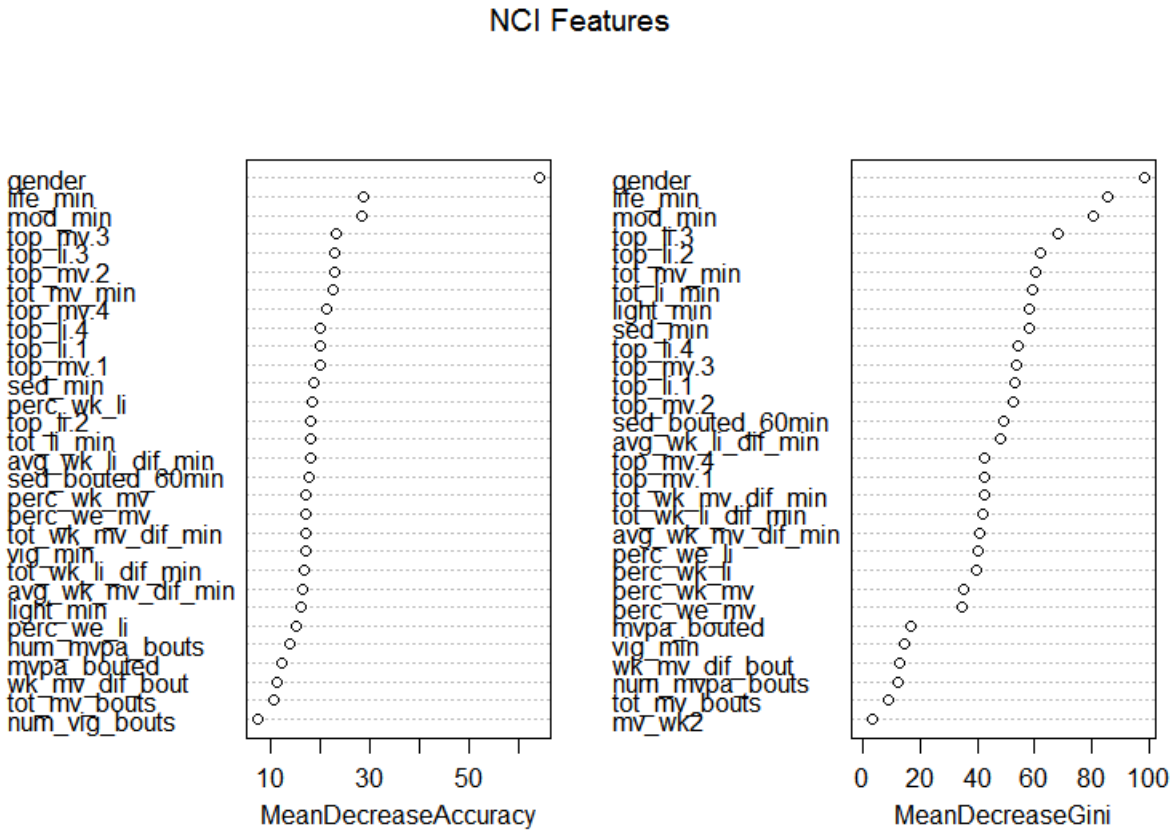


Figure 28. Mean decrease in accuracy and Gini coefficient for NCI low correlation features based on the random forest classifier.

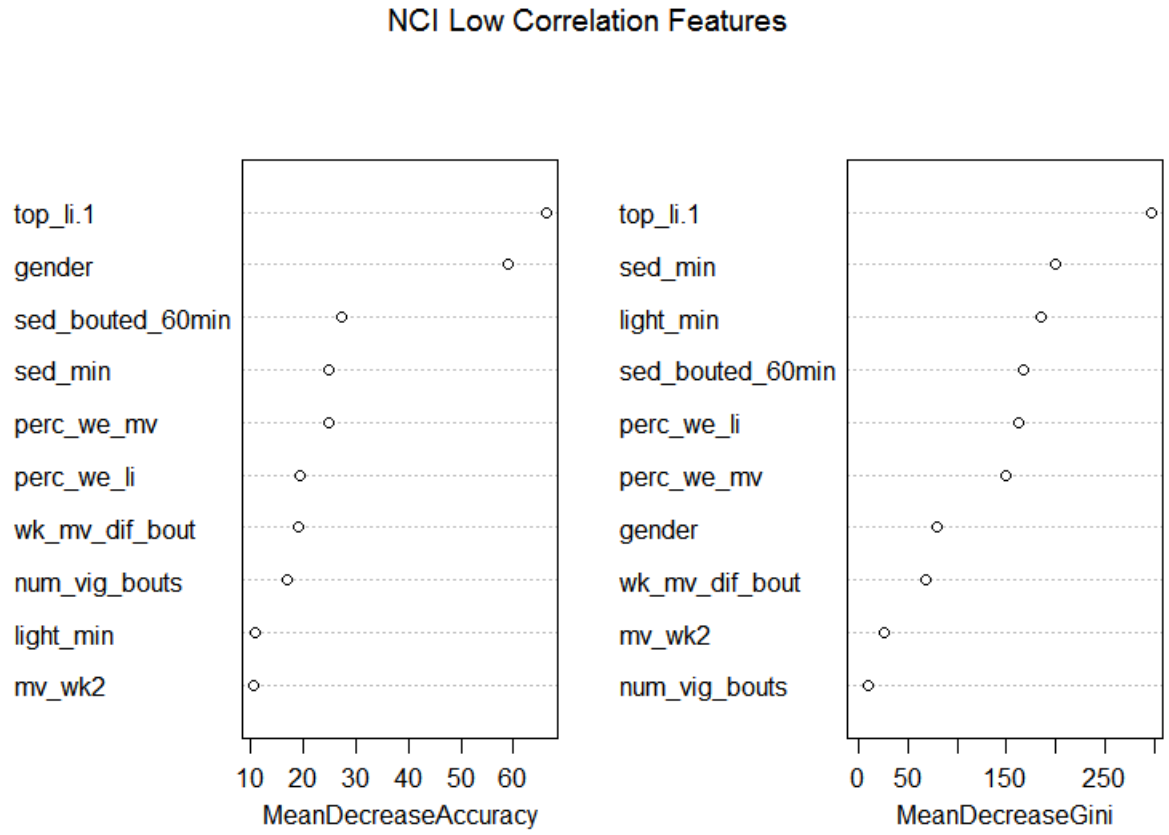


Figure 29. Mean decrease in accuracy and Gini coefficient for all k-means features based on the random forest classifier.

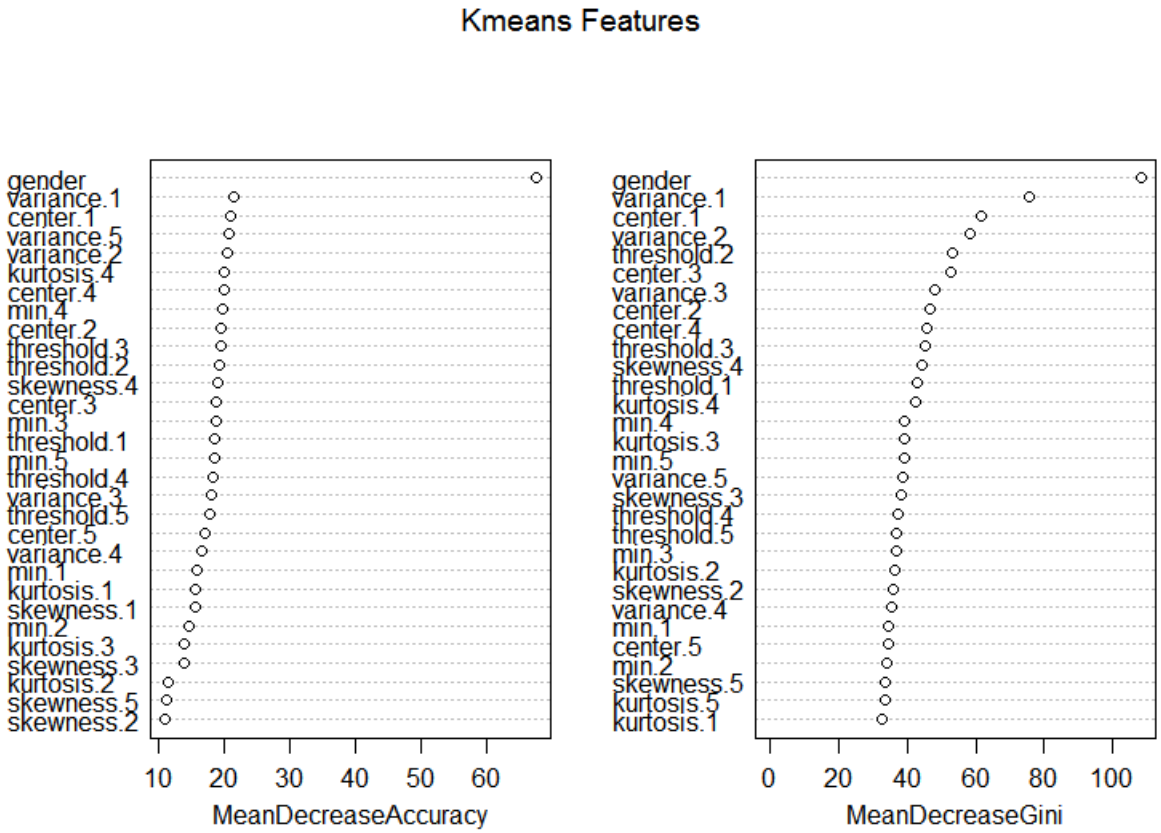


Figure 30. Mean decrease in accuracy and Gini coefficient for k-means low correlation features based on the random forest classifier.

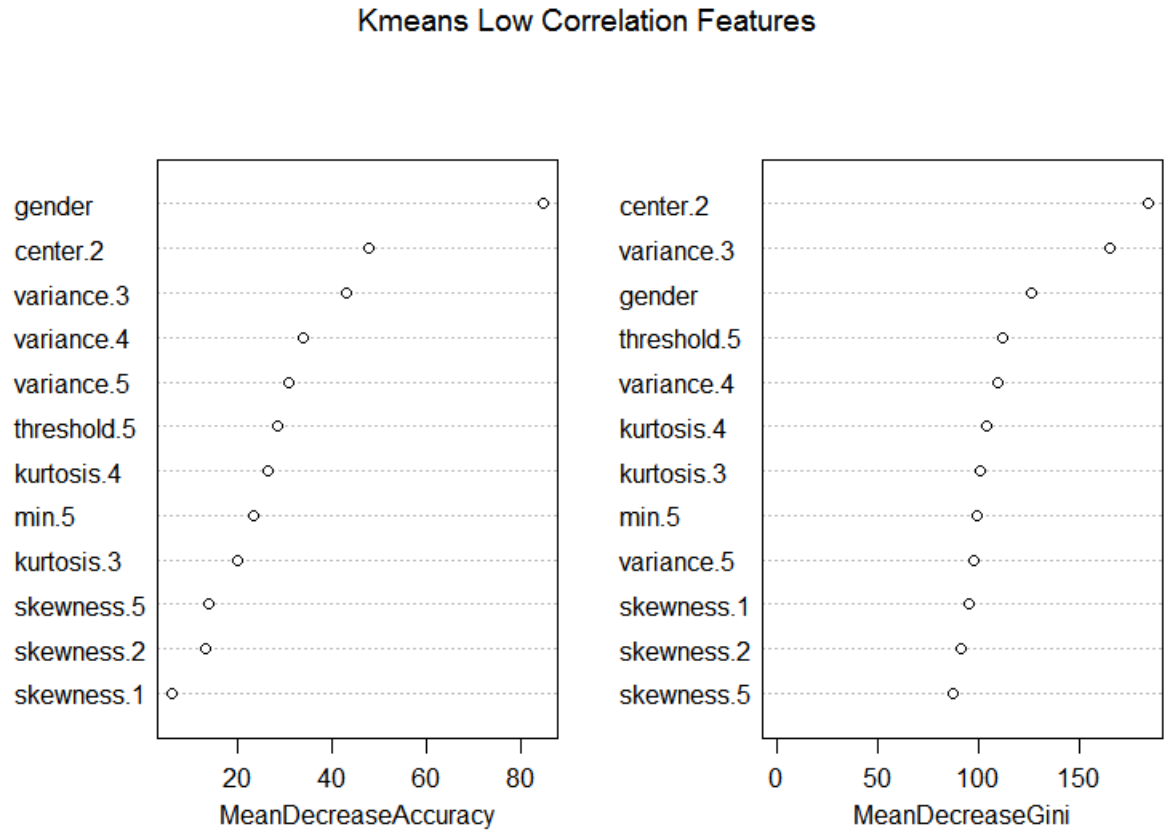


Figure 31. Mean decrease in accuracy and Gini coefficient for HMM features based on the random forest classifier.

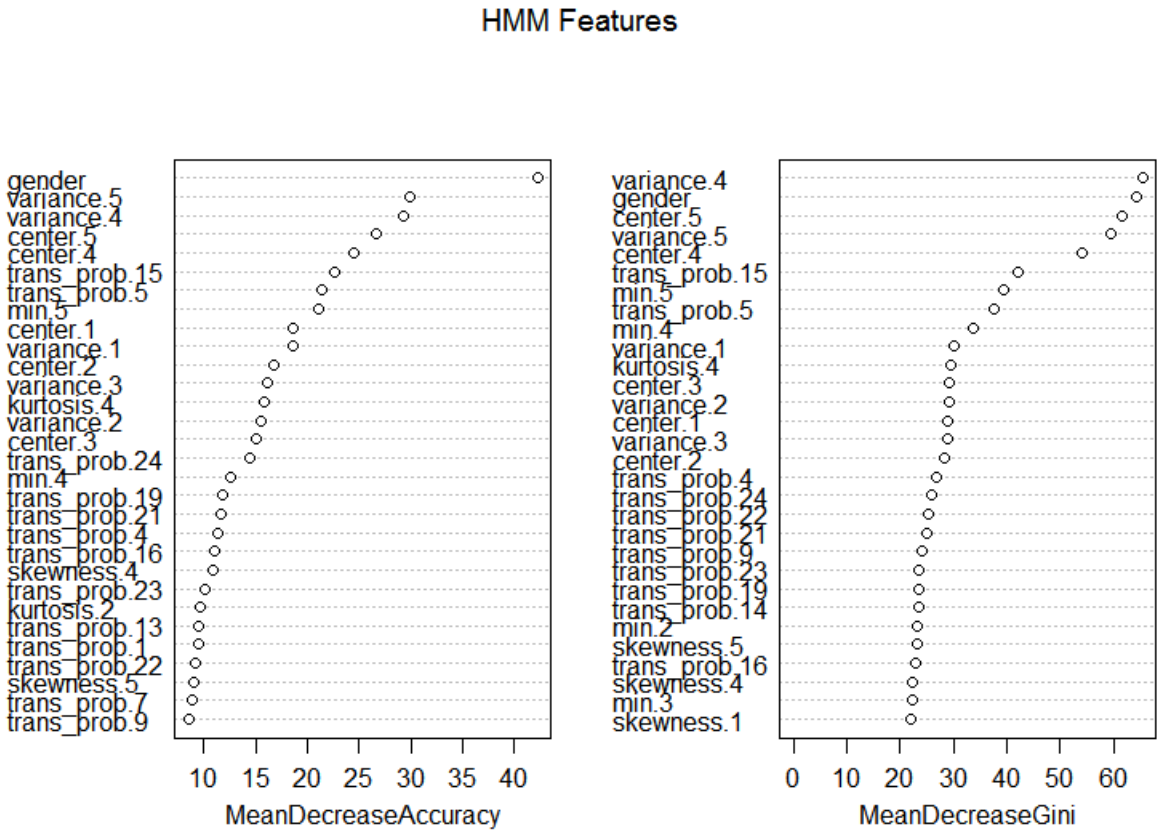


Figure 32. Mean decrease in accuracy and Gini coefficient for HMM low correlation features based on the random forest classifier.

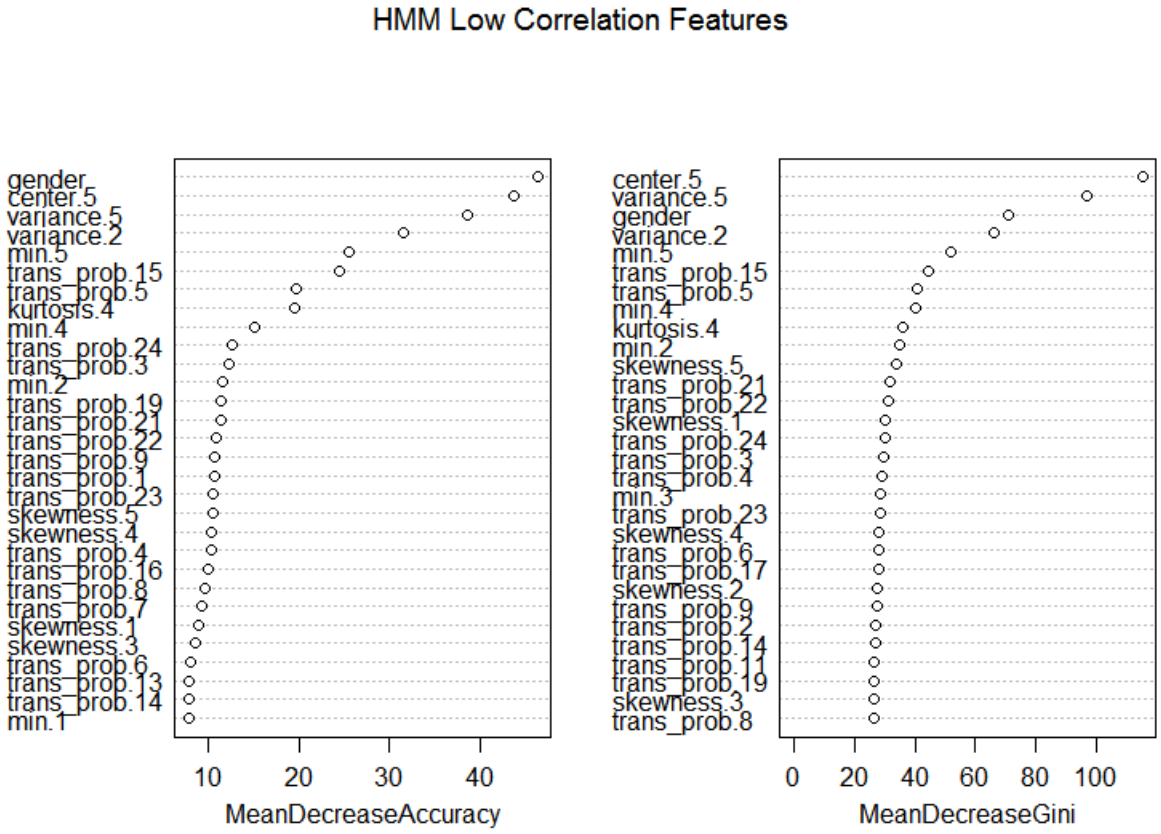


Figure 33. Mean decrease in accuracy and Gini coefficient for all low correlation features based on the random forest classifier.