2020

Prediction of Sudden Cardiac Death Using Ensemble Classifiers

Ayman Momtaz El-Geneidy
Nova Southeastern University, aelgeneidy@compu-house.com

Prediction of Sudden Cardiac Death Using Ensemble Classifiers

by

Ayman M. El-Geneidy

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in
Computer Science

College of Computing and Engineering
Nova Southeastern University

2020

We hereby certify that this dissertation, submitted by Ayman M. El-Geneidy conforms to acceptable standards and is fully adequate in scope and quality to fulfill the dissertation requirements for the degree of Doctor of Philosophy.

Sumitra Mukherjee — July 21, 2020

Sumitra Mukherjee, Ph.D.
Chairperson of Dissertation Committee

Date

July 21, 2020

Michael J. Laszlo Ph.D.
Dissertation Committee Member

Date

July 21, 2020

Francisco J. Mitropoulos, Ph.D.
Dissertation Committee Member

Date

Approved:

Meline Kevorkian — July 21, 2020

Meline Kevorkian, Ed.D.
Dean, College of Computing and Engineering

Date

College of Computing and Engineering
Nova Southeastern University

2020

An Abstract of a Dissertation Submitted to Nova Southeastern University
in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Prediction of Sudden Cardiac Death Using Ensemble Classifiers

by
Ayman M. El-Geneidy
June 2020

Sudden Cardiac Death (SCD) is a medical problem that is responsible for over 300,000 deaths per year in the United States and millions worldwide. SCD is defined as death occurring from within one hour of the onset of acute symptoms, an unwitnessed death in the absence of pre-existing progressive circulatory failures or other causes of deaths, or death during attempted resuscitation. Sudden death due to cardiac reasons is a leading cause of death among Congestive Heart Failure (CHF) patients. The use of Electronic Medical Records (EMR) systems has made a wealth of medical data available for research and analysis. Supervised machine learning methods have been successfully used for medical diagnosis. Ensemble classifiers are known to achieve better prediction accuracy than its constituent base classifiers. In an effort to understand the factors contributing to SCD, data on 2,521 patients were collected for the Sudden Cardiac Death in Heart Failure Trial (SCD-HeFT). The data included 96 features that were gathered over a period of 5 years. The goal of this dissertation was to develop a model that could accurately predict SCD based on available features. The prediction model used the Cox proportional hazards model as a score and then used the ExtraTreesClassifier algorithm as a boosting mechanism to create the ensemble. We tested the system at prediction points of 180 days and 365 days. Our best results were at 180-days with accuracy of 0.9624, specificity of 0.9915, and F_1 score of 0.9607.

Keywords: Sudden Cardiac Death, SCD, Congestive Heart Failure, CHF, Ensemble Classifiers, Machine Learning, Classification, Cox, ExtraTreesClassifier.

Acknowledgements

My deepest thanks and sincerest gratitude are dedicated to my advisor, Dr. Sumitra Mukherjee, for all his help, guidance, and support. I most certainly appreciate his patience with me when I was going through a very difficult time. Thanks are also expressed to Dr. Michael Laszlo and Dr. Francisco Mitropoulos for their efforts in the dissertation committee. Their valuable comments and suggestions helped shape this work.

I would also like to thank Dr. Ling Wang for helping me obtain the research data. I would additionally like to thank my academic advisor, Laura Macias for her help with acquiring the data and for all her help throughout my academic study at Nova.

Very special thanks are dedicated to two very special people. First, I would like to thank my very dear friend, Dr. Aiman Darwiche, for all the moral support, brainstorming, technical discussions, and sometimes arguments, without which the current work would not be in the shape it is. I would also like to thank my best friend, Amr Abdel-Wahed, who helped me make the switch from electrical engineering to computer science, and mentored me in my early years in that field.

I would like to thank my parents for their endless and unconditional love and support. Thanks are also extended to my wife's family for all the encouragements and good wishes.

Additionally, I would like to thank my wife, Mai, for being there for me and for everything she does for us; and yes, she provided the medical expertise for this work. Finally, I would like to thank my lovely daughters, Maram and Hadil.

This work is dedicated to my father, Dr. Momtaz El-Geneidy. Words cannot express what he did for his family. I will be very happy if one day I am to my daughters just half what he is to me. I would also like to dedicate this work to the departed souls of my father-in-law, Dr. Mohamed Shehata, who inspired me to take on this research, being a heart patient himself, and who always encouraged me throughout this work, of my uncle, Mostapha El-Geneidy, and of my cousin, Wafaa El-Geneidy. Finally, I dedicate this work to my late grandparents. They are all missed dearly!

Table of Contents

List of Tables

Tables

List of Figures

Figures

Chapter 1

Introduction

Background

Sudden Cardiac Death (SCD) is one of the main clinical challenges in modern medicine (Shen, Shen, Lin, & Ou, 2007). It is responsible for more than 300,000 casualties in the United States and millions worldwide (Ebrahimzadeh, Pooyan, & Bijar, 2014; Murukesan, Murugappan, & Iqbal, 2013; Murukesan, Murugappan, Iqbal, & Saravanan, 2014).

SCD is defined as a witnessed death occurring within one hour of acute symptoms of cardiac events (Ayesta, Martínez-Sellés, de Luna, & Martínez-Sellés, 2018; Devi, Tyagi, & Kumar, 2016; Deyell, Krahn, & Goldberger, 2015; O'Mahony et al., 2014; Pascual-Figal et al., 2009; Ramírez, Orini, Mincholé, et al., 2017; Shiga et al., 2018), an unwitnessed death in the absence of pre-existing condition of circulatory failure (Pascual-Figal et al., 2009), or death during resuscitation (Ayesta et al., 2018; Pascual-Figal et al., 2009; Ramírez, Orini, Mincholé, et al., 2017; Shiga et al., 2018). It is the result of precipitous loss of heart function (Murukesan et al., 2014; Shen et al., 2007) and is related to an electrical problem in the heart (Murukesan et al., 2014; Vanitha, Suresh, & JenefarSheela, 2014). Most of the SCD incidents are related to arrhythmia or coronary heart diseases (Murukesan et al., 2013).

Ayesta et al. (2018) stated that SCD could occur in any of the following population groups: 1) patients who were never diagnosed with heart disease; 2) patients who were diagnosed with heart disease but had no or mild cardiac dysfunction; 3) patients who were diagnosed with heart disease and had severe cardiac dysfunction; and 4) patients who were diagnosed with cardiac arrhythmia due to a genetically-based cause. The data used in this dissertation satisfies the second group.

Several factors can lead to SCD. These included ventricular tachycardia (VT), ventricular flutter (VFL), ventricular fibrillation (VFib), left ventricular systolic dysfunction, arrhythmia, coronary heart diseases, previous heart attack, long QT-syndrome, myocardial infarction, transient ischemia, heart failure, prolonged QRS duration, asystole, and obesity (U. Rajendra Acharya et al., 2015; Deyell et al., 2015; Ebrahimzadeh & Pooyan, 2011; Ebrahimzadeh et al., 2014; Eranti et al., 2016; Fishman et al., 2010; Goldberger et al., 2014; Kurbanov, Mullabaeva, & Kilichev, 2015; Lee, Shin, Seo, Nam, & Joo, 2016; Liew, 2011; Murugappan, Murukesan, Omar, Khatun, & Murugappan, 2015; Murukesan et al., 2013; Pascual-Figal et al., 2009; Piccini et al., 2010; Shastri et al., 2012; Shen et al., 2007; Sotto, Coelho, & Melo, 2016; Vadakkumpadan, Trayanova, Younes, & Wu, 2012).

Patients with New York Heart Association (NYHA) class II (mild) and class III (moderate) heart failure constitute a high-risk population for SCD (Bardy et al., 2005; Ramírez, Orini, Mincholé, et al., 2017). Congestive Heart Failure (CHF) is defined as the inability of the heart to pump enough blood to meet the body's need (MedlinePlus, 2017c) or pump out venous blood returned to it by the veins.

The prediction of SCD is based on two main approaches. The first approach involves the study and analysis of electrocardiogram (ECG) and heart rate variability (HRV) signals in both normal and SCD-impacted subjects (U. Rajendra Acharya et al., 2015; U Rajendra Acharya et al., 2015; Devi et al., 2016; Ebrahimzadeh & Pooyan, 2011; Ebrahimzadeh et al., 2014; Murugappan et al., 2015; Murukesan et al., 2013; Murukesan et al., 2014; Sheela & Vanitha, 2014; Shen et al., 2007; Toshniwal, Goel, & Sharma, 2015; Vanitha et al., 2014). This approach predicts SCD with a small lead time (in the minutes range) and can only benefit patients that are actually in the hospial. Therefore, it was not considered for this dissertation. The second approach involves gathering data through a clinical trial or a long-term study, then using statistical analyses or machine learning techniques for prediction (Adabag et al., 2014; Bardy et al., 2005; Deyell et al., 2015; Fan et al., 2014; Kurbanov et al., 2015; O'Mahony et al., 2014; Pascual-Figal et al., 2009; Piccini et al., 2010; Shastri et al., 2012). This dissertation used data from the Sudden Cardiac Death in Heart Failure Trial (SCD-HeFT) (Bardy et al., 2005) to predict SCD in NYHA class II and class III patients.

SCD prediction is formulated as a binary classification problem where a patient is classified as an SCD or non-SCD, based on observed input features. No one algorithm is optimal for all problems (Johansson, Bostrom, & Karlsson, 2008). Each algorithm has its strengths and weaknesses. Ensemble classifiers are based on the idea of building a prediction model by combining the strengths of a collection of simpler base models (Hastie, Tibshirani, & Friedman, 2009). The individual decisions of the classifiers that make up the ensemble are combined in some way (typically by weighted or non-weighted

voting) to classify new examples (Dietterich, 2000). This dissertation used ensemble classifiers to predict SCD in CHF patients.

Problem Statement and Goal

SCD has a high mortality rate (Ebrahimzadeh et al., 2014; Murukesan et al., 2013; Murukesan et al., 2014) and its early prediction is a challenge for the medical community (Shen et al., 2007). Patients with congestive heart failure are more vulnerable to SCD (Bardy et al., 2005). The ability to accurately predict SCD is key to saving patients' lives and reducing the mortality rate (Ebrahimzadeh et al., 2014; Murugappan et al., 2015).

In this dissertation, SCD prediction was formulated as a binary classification problem. Given available data for a patient at a point in time, the goal was to predict SCD in NYHA class II and III heart failure patients within the next 180 days. Data from Sudden Cardiac Death in Heart Failure Trial (SCD-HeFT) study was used to train and test the classifiers (Bardy et al., 2005).

Relevance and Significance

The proposed research aimed to develop models to predict SCD within 180 days before its occurrence. The key contribution of this dissertation was a prediction model for SCD for patients, using ensemble classifiers.

SCD prediction has generated interest within the academic community in the last few years. Extensive studies have been conducted in the SCD prediction area starting from 2003. The Literature Review chapter highlights some of such studies.

This dissertation developed an improved SCD prediction model using ensemble classification. By combining time series prediction and ensemble classification, it was expected to achieve better results than the previously mentioned papers.

Issues

The data that was used in this dissertation came from a study that was not intended to be used for SCD prediction. It was meant to be employed to compare the effectiveness of a couple of SCD treatment methods (Bardy et al., 2005). While it offered a clear definition of SCD cases, it required a lot of pre-processing in order for it to be used.

The other issue was that the number of positive cases in the data was low with respect to the dataset. This mandated the use of standard sampling techniques (Kotsiantis, Kanellopoulos, & Pintelas, 2006).

Summary

The goal of this dissertation was to increase the prediction accuracy of SCD for patients with NYHA class II and class III heart failure. This was achieved using the Cox proportional hazard model as a score in the features list, then applying the ExtraTreesClassifier algorithm to generate the ensemble. Data from the SCD-HeFT was used to train and test the model.

Chapter 2

Literature Review

Introduction

The study of the cause and prediction of SCD has drawn the attention of many researchers. This review covers the on-going research in SCD prediction and risk stratification. Also, it provides an overview of ensemble classifiers and the use of the Cox model in the medical field.

Sudden Cardiac Death

Huikuri et al. (2003) analyzed several techniques to predict arrhythmic deaths and suggested several measures of risk of sudden arrhythmic death. Such measures include: Ejection Fraction (EF), ventricular arrhythmias, and ECG markers (Huikuri et al., 2003). These could be used as features in SCD prediction research.

Shen et al. (2007) developed a personal cardiac homecare system by sensing Lead-I ECG signals for detecting and predicting SCD events. This system achieved 87.5% accuracy and 75% sensitivity (Shen et al., 2007). The study presented a machine learning prediction model and suggested a good feature to utilize.

Pascual-Figal et al. (2009) studied whether the measurement of the soluble form of ST2 (sST2) could identify heart failure (HF). ST2 is a protein biomarker of cardiac stress. They demonstrated that elevated sST2 concentrations are predictive of SCD. sST2 could be used as a feature for future SCD prediction models.

Piccini et al. (2010) investigated whether SCD factors could vary with time after a patient suffers myocardial infraction (MI). They identified features, such as heart failure, stroke, atrial fibrillation, and hypertension, which could be used for SCD prediction.

Liew (2011) reviewed some of the evidence for ECG-based tests as predictors of SCD in patients with coronary artery disease (CAD). The reviewed tests consisted of 12-lead ECG, signal-averaged ECG, standard 24-hr Holter, heart rate variability, heart rate turbulence, and t-wave alternans. He concluded that the 24-hr Holter test, combined with other parameters such as heart rate variability and heart rate turbulence, are the most promising features to predict SCD (Liew, 2011).

Stecker and Chugh (2011) and Vadakkumpadan et al. (2012) emphasized that left ventricular ejection fraction (LVEF) cannot be used as the sole predictor of SCD. Any future research will need to include more features for SCD prediction.

Shastri et al. (2012) listed age, diabetes, peripheral vascular disease, ischemic heart disease, serum creatinine, and alkaline phosphatase as independent predictors of SCD. They used Cox proportional hazard model with the mentioned features to assess the risk of suffering SCD. The current dissertation investigated the use of Cox proportional hazard model in SCD prediction.

Sheela and Vanitha (2014) used a support vector machine classifier to predict SCD given HRV inputs. The model was able to predict sudden cardiac arrest (SCA) before 30 minutes of its occurrence based on time and frequency domains features (Sheela & Vanitha, 2014). SCD patients are those who suffered SCA and could not be revived (García Iglesias, Roqueñi Gutiérrez, De Cos, & Calvo, 2018). In a similar effort, Vanitha et al. (2014) used a hybrid classifier to classify SCD patients. Their classifier

consisted of probability neural networks (PNN), K-nearest neighbor and support vector machine (SVM). The decision of the three classifiers were combined using a simple voting system. The model could predict sudden cardiac arrest (SCA) before 30 minutes of its occurrence based on time and frequency domains features. The hybrid classifier achieved 90% prediction accuracy (Vanitha et al., 2014). The use of machine learning techniques was the core of this dissertation.

Ebrahimzadeh et al. (2014) used linear, time-frequency and non-linear features that were extracted from HRV to predict SCD using K-nearest neighbor (KNN) and multilayer perception neural network (MLP). Their results showed that the combination of features can predict SCD by the accuracy of 99.73%, 96.52%, 90.37% and 83.90% for the first, second, third and fourth one-minute intervals respectively, before SCD occurrence. This study was used as a basis for model comparison.

Fan et al. (2014) conducted an observational study to determine the incidence and predicators of SCD in patients with LVEF $\leq$ 30% and New York Heart Association class II/III heart failure that were myocardial infraction (MI) survivors. They concluded that SCD may be predicted by age, LVEF $\leq$ 25%, and non-revascularization.

Wellens et al. (2014) offered additional features for SCD prediction such as age, gender, ethnicity, blood pressure, heart rate, ischemic vs. non-ischemic cause, diabetes, kidney function, findings from cardiac imaging, electrical instability, ANS balance, biochemical markers, and the genetic profile. These offered features are important for future SCD prediction models.

O'Mahony et al. (2014) developed an SCD prediction model that provided individualized risk estimates. Their dataset was derived from a hypertrophic

cardiomyopathy (HCM) cohort study and the model was developed using Cox proportional hazards model. The key predictors were age, maximal left ventricular wall thickness, left atrial diameter, left ventricular overflow tract gradient, family history of SCD, non-sustained ventricular tachycardia, and unexplained syncope. The study utilized Cox hazard model and offered SCD predictors.

Adabag et al. (2014) derived a prediction model for SCD that consisted of age, gender, history of diabetes mellitus, history of myocardial infraction, Left Bundle Branch Block (LBBB) on ECG, and natural logarithm NT-proBNP (lnNT-proBNP). Their model was generated using Cox regression analysis. This is another study that demonstrated the use of Cox regression model.

Toshniwal et al. (2015) proposed a system to predict the risk of cardiovascular disease that can lead to SCD. The system was based on ECG signals. It consisted of a 2-stage classification model where the first stage differentiated normal from abnormal records, while the second stage aimed to improve the accuracy by reducing the false negatives. They used Random Forest with 120 trees for classification and achieved a 98.57% accuracy. The study showed the superiority of ensemble classification over single classification algorithms.

Deo et al. (2016) suggested 12 independent risk factors for SCD. These were age, gender, race, use of smoking products, systolic blood pressure, use of antihypertensive medication, diabetes mellitus, serum potassium, serum albumin, high-density lipoprotein, estimated glomerular filtration rate, and QT interval. Using such features, they developed a regression predictive model, which achieved a c-statistic of 0.808 and 0.743

in two different datasets. The researchers demonstrated a different way of utilizing predictive features.

Devi et al. (2016) used a KNN classifier to predict SCD one hour before its occurrence. They used time and frequency domain features and compared the values of such features between affected and normal subjects. Their model achieved 95% accuracy. The researchers formulated another use of single classification.

Lee et al. (2016) developed a prediction model for VT, which can lead to SCD. The model used artificial neural network (ANN) to predict the event one hour before its occurrence. The dataset contained parameters that were obtained from heart rate variability and respiratory rate variability (RRV) analysis (Lee et al., 2016). This study was used for model comparison.

Weeks, Sieg, Gass, and Rajapreyar (2016) provided a review on drug therapies that aimed at suppressing arrhythmias that can cause sudden cardiac death in patients with heart failure. They covered the SCD-HeFT trial and the effects of amiodarone. They also influenced the initial choice of some predictors.

Rai and Agrawal (2016) provided a review that focused on the etiology, risk factors, prognostic features, and importance of risk stratification of SCD. The key contribution of this research was the inclusion of multiple risk factors that were used as predictors.

Al-Gobari, Le, Fall, Gueyffier, and Burnand (2017) attempted to assess the available evidence that evaluated the effectiveness of statins in clinical outcomes for heart failure patients. The researchers concluded that while statins did not reduce SCD and other causes of mortality, it might decrease hospitalization for worsening heart

failure. The presence of statins in a patient is a feature that was initially considered in this dissertation and this study added information about this subject.

Ramírez, Orini, and Laguna (2017) used Cox hazard model with four different components of ECG to determine the component with the highest predictive value for SCD risk stratification. They achieved that by calculating the hazard ration of each component.

Desai et al. (2018) researched the effectiveness of the European Society of Cardiology's (ESC) SCD risk score on the prognosis of obstructive hypertrophic cardiomyopathy (HCM) disease and assessed whether additional factors could adjust SCD risk. The research was conducted on patients with obstructive HCM, a disease that causes the heart muscle to grow abnormally thick, which impacts the amount of blood the ventricular can hold and the amount pumped out with each heartbeat (Mayo Clinic, 2018a). This paper provided a 5-year lead time prediction model using the Cox proportional hazards analysis and helped verify the predictors' list.

Kubik, Dąbrowska-Kugacka, Lewicka, Daniłowicz-Szymanowicz, and Raczak (2018) provided a summary of the literature about left ventricular noncompaction (LVNC), which is a unique inherited cardiomyopathy characterized by an increased risk of heart failure, arrythmia, and SCD. This paper helped in the choice of SCD predictors and introduced LVNC as a possible cause.

Agoston-Coldea et al. (2018) utilized Cox proportional hazards regression to test the combined use of left ventricular global longitudinal strain (GLS) and Galectin-3 (protein used as a marker of fibrosis where the latter is defined as the formation of excess fibrous) to predict major cardiovascular adverse events (MACEs) in patients with severe

aortic stenosis (AS). This research effort benefited SCD because it is considered part of MACEs. They concluded that GLS and the NT-proBNP (a prohormone used in the prognosis of heart failure) were the most reliable predictors of MACEs in patients with severe AS and were superior to Galectin-3.

Shiga et al. (2018) used Cox model to study the effects of obesity on SCD in Japanese patients after myocardial infarction. They provided a list of predicators such as age, gender, and blood pressure, which were used in this dissertation.

Mohanty, Sahoo, Biswal, and Sabut (2018) provided a process to detect and classify ventricular tachycardia (VT) and ventricular fibrillation (VF) arrhythmias (two leading causes of SCD) using time-frequency domain features of processed ECG signals. They used support vector machine (SVM) and C4.5 for classification of selected features, where the latter had a better performance with 90.97% sensitivity, 97.86% specificity, and 97.02% accuracy. The results of this paper were used for comparison with the results of the current dissertation.

Su, Xia, Cao, and Gao (2018) provided a list of predictors. They stated that premature ventricular contractions (PVCs) can lead to ventricular tachycardia (VT), which in turn can lead to ventricular fibrillation (VF) and SCD. They assessed the cardiac characteristics and risk prediction in PVC patients with or without VT. They concluded that PVC couplets (two consecutive premature ventricular contractions) is the highest risk factor for the development of VT in patients with frequent PVC. Other risk factors included blood potassium, LVEF, extensive PVC burden (amounts of abnormal beats), and alcohol consumption.

Ayesta et al. (2018) suggested that a combination of clinical, biochemical, echocardiographic, and electrical parameters is better than a single risk marker to best predict SCD in elderly patients with heart failure. They provided a general survey on risk predictors, which influenced the choice of the ones used in this dissertation.

Ng, Mistry, Li, Schlindwein, and Nicolson (2018) developed two ECG markers as predictors of ventricular arrhythmias and SCD. They applied them as part of an under-development tool (LifeMap) for SCD risk stratification.

Kakimoto, Tanaka, Hayashi, Yokoyama, and Osawa (2018) studied the changes in miRNA expression from subjects who suffered SCD with cardiac hypertrophy (SCH). They compared cardiac tissues that were sampled at autopsy from SCH patients, compensated cardiac hypertrophy (CCH) subjects who died from causes other than heart failure, and control cases that were free from both cardiac hypertrophy and heart failure. They concluded that miR-221 had the most increase in SCH patients, hence established the relation between miR-221 levels and SCH patients.

Rosset et al. (2018) asserted that arrhythmic drugs have no impact on preventing SCD. Their findings are in line with the SCD-HeFT study. They also provided justification to use amiodarone patients in this dissertation.

Thomsen, Nielsen, Aarup, Bisgaard, and Pedersen (2018) attempted to establish links between chronic kidney disease (CKD) and sudden cardiac death. Their research highlighted the effect of kidney diseases and QRS duration as predictors of SCD.

Li et al. (2018) aimed to characterize the expression patterns of miRNAs in hypertrophic cardiomyopathy (HCM) and dilated cardiomyopathy. These two forms can

develop into heart failure and can cause SCD. They included a list of clinical data such as age, gender, and blood pressure, which were used in this dissertation.

Özyılmaz, Satılmışoğlu, Gül, Uyarel, and Serdar (2018) provided a list of features such as age, gender, NYHA class, creatinine, uric acid, and galectin-3 level. They obtained a risk score for SCD using Cox model at 5 years-time.

Ebrahimzadeh, Fayaz, Ahmadi, and Dolatabad (2018) produced two models to predict SCD within 4 minutes of its occurrence. They used MLP and KNN classifiers that delivered an accuracy of 83.96% for the first and 81.49% for the second. This research was used for comparison with the current dissertation.

Ensemble Classifiers

The idea behind ensemble classifiers stems from the human nature of seeking multiple expert advice when making important decisions, instead of trusting a single opinion (Harrington, 2012). Based on empirical observations and specific machine learning applications, a given learning algorithm can outperform all others for a specific problem or for a specific subset of the input data, but not on the overall problem domain (Valentini & Masulli, 2002).

Ensemble classifiers enhance the accuracy and the reliability of the overall inductive learning system because the overall system can recover if some base learners fail (Valentini & Masulli, 2002). Dietterich (2002) described the supervised machine learning setting as one where each data point consists of a vector of features x, a class label y, and an underlying function f, where $y = f(x)$ for each data point (x, y). The machine learning algorithm searches through a space of possible functions, or *hypotheses*, to find one function, h, with the best apporoximation to the unknown fuction

f (Dietterich, 2002). Ensemble classifiers help resolve three main problems that are experienced by learning algorithms that output only a single hypothesis (Dietterich, 2002). These are: 1) a statistical problem, when the learning algorithm is searching a space of hypotheses that is too large for the amount of training data; 2) a computational problem, when the learning algorithm cannot guarantee to find the best hypothesis within the hypothesis space; and 3) a representational problem, when the hypothesis space does not contain any hypothesis that is a good approximation to the true function *f* (Dietterich, 2002). A simple vote of all equally-good classifiers can reduce the risk of the statistical problem (Dietterich, 2002). A weighted combination of several different local minima can reduce the risk of choosing the wrong local minimum to output, and hence handle the computational problem (Dietterich, 2002). For the representational problem, a weighted sum of hypotheses can expand the space of functions that can be represented and hence forming a more accurate approximation to *f* (Dietterich, 2002).

Ensemble classifiers can be divided into two distinct groups: non-generative ensembles and generative ensembles (Abad, Zare-Mirakabad, & Rezaeian, 2014; Valentini & Masulli, 2002). The two types are briefly described below:

Non-generative ensembles

Non-generative ensemble methods attempt to combine a set of base learners (Valentini & Masulli, 2002). They do not actively generate new base learners but try to combine a set of existing base classifiers in a suitable way (Valentini & Masulli, 2002). They are divided into ensemble fusion and ensemble selection methods (Abad et al., 2014). Ensemble fusion methods combine the outputs of the base classifiers, while ensemble selection methods try to choose the best classifiers between the set of available

base learners (Abad et al., 2014). The most popular ensemble fusion method is expressed by majority voting ensembles (Abad et al., 2014). Other techniques include Naïve-Bayes decision rule, Behavior-Knowledge Space (BKS) method and multiple operators such as Minimum, Maximum, Average, Product and Ordered Weight (Abad et al., 2014). Ensemble selection methods identify the best base classifiers among base learners where the output of the ensemble is equal to the output of the best classifier (Abad et al., 2014).

Generative ensembles

Generative ensemble methods generate sets of base learners acting on the base learning algorithm or on the structure of the data set and try to actively improve diversity and accuracy of the base learners (Valentini & Masulli, 2002). They do so by either modifying the structure and the characteristics of the available input data, manipulating the aggregation of the classes, selecting base learners specialized for a specific input region, selecting a proper set of base learners, evaluating the performance and the characteristics of the component base learners or randomly modifying the base learning algorithm (Valentini & Masulli, 2002).

Combining methods of ensemble classifiers

Combining methods are used to improve a weak classifier (Skurichina & Duin, 2002). Weak classifiers refer to badly performing classifiers, unstable classifiers, classifiers of a low complexity, or classifiers that depend on certain assumed models that are not always true (Skurichina & Duin, 2002). The below methods work well on *homogenous ensembles* and attempt to generate diversity by sampling from or assigning weights to training sets (Whalen & Pandey, 2013). They generally use a single type of classifier to build the ensemble (Whalen & Pandey, 2013). Diversity refers to the case

where base classifiers produce different errors (Valentini & Masulli, 2002). Diversity is key to the overall performance of ensemble classifiers (Whalen & Pandey, 2013).

Bagging. Breiman (1996) introduced bagging, which is a technique where the data is taken from the original dataset *S* times, resulting in creating *S* new datasets (Harrington, 2012). Each of these new datasets are of the same size as the original (Harrington, 2012). Each dataset is built by randomly selecting an example from the original dataset (Harrington, 2012). The same example can be selected more than once (which is referred to as *selection with replacement*) resulting in repeated values in the new dataset and some values from the original dataset not being present in the new dataset (Harrington, 2012). The learning algorithm is then individually applied to the new set (Harrington, 2012). To classify a new piece of data, the *S* classifiers are applied to the new piece of data and classification is performed through a majority vote (Harrington, 2012).

Boosting. Freund and Schapire (1996) introduced boosting, in which the different classifiers are trained sequentially (Harrington, 2012). Each new classifier is trained based on the performance of those already trained (Harrington, 2012). The focus of boosting is on data that was previously misclassified by previous classifiers (Harrington, 2012). The output is calculated from a weighted sum of all classifiers (Harrington, 2012). The weights are based on how successful the classifier was in the previous iteration (Harrington, 2012). Initially, all objects have equal weights, and the first classifier is constructed on this data set (Skurichina & Duin, 2002). Then, weights are changed according to the performance of the classifier (Skurichina & Duin, 2002). Erroneously classified objects get larger weights, and the next classifier is boosted on the

reweighted training set (Skurichina & Duin, 2002). AdaBoost is a popular boosting algorithm (Bauer & Kohavi, 1998; Ghavidel, Yazdani, & Analoui, 2013; Harrington, 2012).

Random subspace. Ho (1998) introduced the random subspace method. In this technique, classifiers are constructed in random subspaces of the data feature space and are usually combined by simple majority voting in the final decision rule (Skurichina & Duin, 2002). Another approach is to build the ensemble from the predictions of a wide variety of *heterogeneous classifiers* such as decision trees, support vector machines, and neural networks as base classifiers (Whalen & Pandey, 2013).

Stacking. Stacking constructs a higher-level predictive model over the predictions of base classifiers (Whalen & Pandey, 2013).

Ensemble selection. Ensemble selection uses an incremental strategy to select base predictors for the ensemble while balancing diversity and performance (Whalen & Pandey, 2013).

Ensemble classifiers usage

Johansson et al. (2008) showed that by using a class-specific reliability measure instead of one based on the overall accuracy, the predictive performance of applying combination rules in an evidential framework can be improved. Ramos-Jimenez, del Campo-Avila, and Morales-Bueno (2009) presented a two-layers system where the first layer consisted of an ensemble classifier of 10 decision trees, and the second layer was a single classifier that was induced using the examples that were not unanimously voted by the ensemble. In addition, the examples that reached the second layer incorporated additional information added by the ensemble. Such information was the class estimated

by each base classifier in the ensemble and the class estimated by the ensemble itself. The idea was that second level can provide more informed classification, and hence improve the overall accuracy of the system. Bagheri and Gao (2012) proposed a classifier selection method based on the divide-and-conquer technique. The idea was to break down the classification problem into simpler binary classification problems. A first-guesser classifier was used to find the two most probable classes. Based on those two classes, one of the previously trained classes was called to resolve the binary classification problem. The proposed method slightly improved the overall classification accuracy but significantly lowered the execution time compared with existing ensemble classification methods. Verma and Rahman (2012) presented a cluster-oriented ensemble classifier (COEC) that was based on learning of cluster boundaries by the base classifiers. The proposed ensemble classifier clusters classified data into multiple clusters, learned the decision boundaries between clusters using a set of base classifiers, and combined the cluster decisions produced by the base classifiers into a class decision by a fusion classifier. The proposed system was evaluated on benchmark datasets from UCI machine learning repository. Results showed that: 1) Homogeneous clustering performs significantly better than heterogeneous clustering with COEC; 2) The proposed COEC performs significantly better than its base counterparts; 3) Fusion classifier performs significantly better than algebraic fusion with COEC; 4) COEC outperforms classical ensemble classifiers namely bagging, boosting, and random subspace method significantly on benchmark data sets. Gupta, Audhkhasi, and Narayanan (2014) addressed the limitations of training algorithms. They defined such limitations as existing algorithms either train an ensemble classifier on pre-defined feature sets or

independently perform classifier training and feature selection. They defined feature subset selection and training of diverse classification as an optimization problem and proposed a greedy algorithm to resolve this problem. The algorithm performed joint optimization to determine class boundaries and the feature set for each classifier in the ensemble. The authors introduced a loss function that introduced data-driven diversity. They sequentially optimized the loss function for each classifier in the ensemble. They used an oracle fusion function and an equal weighting function to obtain the final decision from the ensemble. They presented the results on several datasets and observed that not only the algorithm trained better classifier ensembles, but also filtered out features unrelated to class assignments. Yu, Li, Liu, and Han (2015) identified some of the limitations of traditional random subspace-based ensemble approaches (RSCE) as viewing the same importance for the base classifiers trained in different subspaces, and not considering how to find the optimal random subspace set. They designed a general hybrid adaptive ensemble learning framework (HAEL) that addressed the limitations of RSCE. HAEL consisted of two adaptive processes: base classifier competition and classifier ensemble interaction, to adjust the weights of the base classifiers in each ensemble and to explore the optimal random subspace set simultaneously. Their proposed framework was characterized with the following two properties: 1) the adoption of a base classifier competition adaptive process (BCCAP) to adjust the weights of the classifiers in the ensemble and 2) the adoption of a classifier ensemble interaction adaptive process (CEIAP) to search for the optimal random subspace set

Cox Model

The Cox hazard model has been used extensively in the medical field. The selected research papers below demonstrate the usage and importance of the Cox hazard model.

Van Dijk, Steyerberg, Stenning, Dusseldorp, and Habbema (2004) used the Cox regression model to examine the extent of any loss in discrimination within the classification provided by the International Germ Cell Consensus (IGCC). The latter identified a 3-level prognosis groups (good, intermediate and poor) among patients with metastatic nonseminomatous germ cell tumors (NSGCT), based on some risk factors.

The SCD-HeFT trial used the Cox model to test the interactions between the NYHA class and the type of treatment (ICD, amiodarone, or placebo), and also between the cause of CHF and the treatment (Bardy et al., 2005). This paper influenced the use of the SCD-HeFT data in this dissertation, because it demonstrated that the Cox model was previously applied to it.

Zhao (2005) extended the Cox model to enable it to consider familial correlation between family members. He applied his framework to data from the Collaborative Study on the Genetics of Alcoholism. He also used R for his data analysis.

Bellera et al. (2010) emphasized the importance of the assumption that the Cox model relies on the proportional hazards (PH). This meant that the factors investigated have a constant impact on the hazard over time. They argued that checking the proportionality of hazards should be an integral part of a survival analysis based on a Cox model. They analyzed prognostic factors of metastases in 979 women treated for breast

cancer with surgery. With a median follow-up of 14 years, they showed that conventional Cox model did not catch all the effective factors.

Royston and Altman (2013) described statistical approaches to external validation of a published Cox model and applied their methods to two datasets for breast cancer, where one was a derivation dataset, and the other was a validation dataset. They concluded that their methods are applicable to a wide range of prognostic studies.

Darwiche and Mukherjee (2018) used the Cox model to obtain a score 20 hours before septic shock occurrence, then added the score to features and applied random forest ensemble to classify patients. The paper demonstrated the use of Cox model as a score and the use of ensemble classification to improve the overall classification.

Chapter 3

Methodology

The aim of this dissertation was to build an ensemble classifier to predict SCD among NYHA class II and class III heart failure patients, within 180 days from admission point. The main tasks involved data collection, features selection, data cleanup and preparation, prediction model, validation process, and results evaluation.

Data Collection

Data was obtained from the Sudden Cardiac Death in Heart Failure Trial (SCD-HeFT) (Bardy et al., 2005). It was made available by the National Heart, Lung and Blood Institute (NHLBI). The trial involved a recruitment period from September, 1997 to July, 2001, and a follow-up period until October, 2003 (Bardy et al., 2005).

The aim of SCD-HeFT was to determine whether amiodarone, a medication that is used to treat and prevent certain types of serious, life-threatening ventricular arrhythmias (MedlinePlus, 2017a), or the implantable cardioverter-defibrillator (ICD) would decrease the overall mortality in patients with coronary artery disease or nonischemic cardiomyopathy who are in heart failure NYHA class II or III and have a left ventricular ejection fraction (LVEF) ≤35% (Klein, Auricchio, Reek, & Geller, 1999). Ejection fraction is a measurement of the percentage of blood leaving the heart each time it contracts. The left ventricle is the heart's main pumping chamber, so ejection fraction is

usually measured only in the left ventricle ("Ejection fraction: What does it measure?," 2018).

To be included in that trial, patients must have been 18 years of age or older, must have had heart failure NYHA class II or III for at least 3 months prior to enrollment, must have had LVEF ≤35%, must have had their CHF treated with a vasodilator, a medication that opens (dilates) blood vessels (Mayo Clinic Staff, 2016), and have been requested to have a coronary angiogram to detect possible coronary disease or to determine the nature of their disease (Klein et al., 1999).

Exclusion criteria from the trial consisted of: 1) patients that showed no symptoms of heart failure (asymptomatic patients); 2) patients who have survived a cardiac arrest or who have experienced sustained ventricular tachycardia (VT), except during the acute phase of myocardial infarction; 3) patients who were judged to have high probability of death from noncardiac causes within 12 months; 4) patients in heart failure due to causes other than dilated cardiomyopathy or coronary artery disease; 5) patients who were taking amiodarone or who had a contraindication for amiodarone, where a contraindication is a specific situation in which a drug, procedure, or surgery should not be used because it may be harmful to the person (MedlinePlus, 2017b); 6) patients who have had an unexplained syncope within the 5 years prior to the trial, where a syncope is a temporary loss of consciousness usually related to insufficient blood flow to the brain; (American Heart Association, 2017) and, 7) patients who were expected to have a heart transplant (Klein et al., 1999).

SCD-HeFT was a randomized trial where patients were randomly assigned to one of three groups of study. The first group was a control group that received conventional

heart failure treatment and placebo (Klein et al., 1999). A placebo is an inert or innocuous substance used especially in controlled experiments testing the efficacy of another substance ("Merriam-Webster Medical Dictionary," 2017). The second group combined conventional heart failure treatment and amiodarone (Klein et al., 1999). The third group employed conventional heart failure therapy and a single-lead pectoral ICD that was inserted through an outpatient procedure (Klein et al., 1999). Placebo and amiodarone were delivered in a double-blind fashion to avoid treatment bias in favor of amiodarone (Klein et al., 1999). A double-blind fashion is an experimental procedure in which neither the subjects nor the experimenters know which subjects are in the test and control groups during the actual course of the experiments ("Merriam-Webster Medical Dictionary," 2017).

There were 2,521 patients enrolled in the SCD-HeFT study, with 847 patients assigned to the placebo group, 845 patients assigned to the amiodarone group, and 829 patients assigned to the ICD group (Bardy et al., 2005). Table 1 demonstrates the patients' distribution over the three groups of the study, along with the death count in each group.

Table 1.

Patients Population

Group	Total Patients	Total Deaths	Total SCDs
Placebo	847	244	88
Amiodarone	845	240	72
ICD	829	182	34
Totals	2,521	666	194

An ICD is a device that is placed in the patient's chest to reduce the risk of SCD by detecting and stopping abnormal heartbeats through electrical pulses to regulate the patient's heartbeat (Mayo Clinic, 2018b). Therefore, it has a direct impact on the SCD outcome and patients using it were excluded from this dissertation. Given that amiodarone had no statistical significance on survival (Bardy et al., 2005), this group was included in the dissertation. Thus, the total initial population for this dissertation was 1,692 (Placebo and Amiodarone) subjects, with a total of 160 SCD cases (Placebo and Amiodarone).

Features Selection

The features shown in table 2 existed in the available SCD-HeFT data (Bardy et al., 2005), and were considered based on the literature (Adabag et al., 2014; Ayesta et al., 2018; Deo et al., 2016; Desai et al., 2018; Eranti et al., 2016; Fan et al., 2014; Goldstein, Chang, Mitani, Assimes, & Winkelmayer, 2014; Huikuri et al., 2003; Kubik et al., 2018; Kurbanov et al., 2015; Li et al., 2018; Liew, 2011; O'Mahony et al., 2014; Pascual-Figal et al., 2009; Piccini et al., 2010; Rai & Agrawal, 2016; Shastri et al., 2012; Shiga et al., 2018; Stecker & Chugh, 2011; Thomsen et al., 2018; Weeks et al., 2016; Wellens et al., 2014). There were 31 features listed in table 2 which were a subset of the 96 features used in the SCD-HeFT study. This was a preliminary list and further research resulted in updating the features list, as explained in later sections.

Table 2.

Potential Features to Be Used for Prediction

Feature Name	Type	Description	Data Type	Allowable Values
Age	Demographic	Patient's age	Numeric	18 – 100
Gender	Demographic	Patient's gender	Numeric	1 – Male 2 – Female

Ethnicity	Demographic	Patient's race	String	African American Asian Caucasian Latin American Other
Weight	Vital	Patient's weight	Numeric	90 – 350
Systolic blood pressure	Vital	Self-explanatory	Numeric	55 – 225
Diastolic blood pressure	Vital	Self-explanatory	Numeric	30 – 140
Heart rate	Vital	Self-explanatory	Numeric	30 – 156
History of SCD	History	Family history of SCD	Numeric	0 – No 1 – Yes
History of CHF	History	Family history of CHF	Numeric	0 – No 1 – Yes
LVEF	Diagnosis	Left ventricular ejection fraction	Numeric	1 – 100
Diabetes	Diagnosis	Patient has diabetes	Numeric	0 – No 1 – Yes
CrCl	Measurement	Creatinine clearance (related to kidney function)	Numeric	3 – 300
NYHA class	Measurement	New York Heart Association class	Numeric	1 – I 2 – II 3 – III 4 – IV
Pulmonary disease	Diagnosis	Patient is diagnosed with pulmonary disease	Numeric	0 – No 1 – Yes
Hypertension	Diagnosis	Patient is diagnosed with hypertension	Numeric	0 – No 1 – Yes
Atrial fibrillation/flutter	Diagnosis	Patient suffered atrial fibrillation or atrial flutter	Numeric	0 – No 1 – Yes

Syncope	Diagnosis	Patient suffered symptom of syncope	Numeric	0 – No 1 – Yes
CABG	History	Coronary Artery Bypass Graft	Numeric	0 – No 1 – Yes
Alchool abuse	History	Self-explanatory	Numeric	0 – No 1 – Yes
Smoking	History	Self-explanatory	Numeric	1 – Never 2 – Current 3 – Past
Cocaine abuse	History	Self-explanatory	Numeric	0 – No 1 – Yes
Stroke	History	Patient suffered a stroke	Numeric	0 – No 1 – Yes
Serum sodium	Measurement	Sodium measurement	Numeric	113 – 187
Serum potassium	Measurement	Potassium measurement	Numeric	2.0 – 7.0
Serum magnesium	Measurement	Magnesium measurement	Numeric	0.0 – 9.0
Serum creatinine	Measurement	Creatinine measurement	Numeric	0 – 566
Aspirin	Medication	Patient is on aspirin	Numeric	0 – No 1 – Yes
Beta-blocker	Medication	Patient is on beta-blocker	Numeric	0 – No 1 – Yes
Statin	Medication	Patient is on statin	Numeric	0 – No 1 – Yes
History of myocardial infarction	History	Patient suffered myocardial infarction	Numeric	0 – No 1 – Yes
QT Interval	Measurement	Self-explanatory	Numeric	308.00 – 604.00

SCD-HeFT was intended to evaluate the theory that amiodarone or a conservatively programmed shock-only, single-lead ICD would decrease the risk of death from any cause in patients with mild-to-moderate heart failure (Bardy et al., 2005). It was not aiming at predicting SCD itself. However, the data gathered by the SCD-HeFT

study was used for SCD prediction, which was the goal of this dissertation. Therefore, not all SCD-HeFT features were used by this dissertation. While the features in table 2 were part of the 96 starting features of the SCD-HeFT, the difference was due to the following reasons:

- Some features were specific to SCD-HeFT study and were not mentioned in other references.
- SCD-HeFT expanded some features, which was not useful for this dissertation. For example, SCD-HeFT contained *age*, *age65*, and *age75* features to describe the patient's age. Only the *age* feature was used in this dissertation.

Some SCD-HeFT features were ICD-related, and ICD patients were excluded from this dissertation.

Data Cleanup and Preparation

This section refers to placing the data in a format that can be used by the prediction model. Due to the Health Insurance Portability and Accountability Act (HIPAA) rules, patients in the SCD-HeFT database are de-identified. In the SCD-HeFT database, patients are uniquely identified by a patient id code (PID). The start time of any patient in the study is considered time $t = 0$, and all dates of vitals and lab works are calculated as the number of days/months from the start date.

The dataset consisted of multiple records per patient, one at each timestamp. Each single record consisted of the patient id, the start of the time bucket, the end of the time bucket, the input features, and the class column (event). Time buckets are one way

to encode time-dependents features (Therneau, Crowson, & Atkinson, 2013). This means that for example at day 7, SBP is 130 and DBP is 70 (Therneau et al., 2013).

Taking just the systolic blood pressure (SBP) and diastolic blood pressure (DBP) features as examples, table 3 shows the records for a patient who completed the trial without suffering from SCD.

Table 3.

A sample patient (PID SCD_HEFT0001) with 2 sample features who did not suffer SCD

PID	StartTime	EndTime	SBP	DBP	SCD
SCD_HEFT0001	0	7	130	70	0
SCD_HEFT0001	7	34	100	70	0
SCD_HEFT0001	34	129	128	64	0
SCD_HEFT0001	129	183	140	78	0
SCD_HEFT0001	183	323	142	78	0
SCD_HEFT0001	323	432	150	74	0
SCD_HEFT0001	432	569	150	82	0
SCD_HEFT0001	569	702	120	82	0
SCD_HEFT0001	702	919	140	80	0
SCD_HEFT0001	919	1063	128	80	0
SCD_HEFT0001	1063	1281	138	78	0
SCD_HEFT0001	1281	1541	108	64	0

Table 4 demonstrates a patient who suffered from SCD prior to completing the trial. The last row consists of the last known vital signs that were recorded before death.

Table 4.

A sample patient (PID SCD_HEFT0023) with 2 sample features who suffered SCD

PID	StartTime	EndTime	SBP	DBP	SCD
SCD_HEFT0023	0	7	130	80	0
SCD_HEFT0023	7	33	122	78	0
SCD_HEFT0023	33	68	120	62	0
SCD_HEFT0023	68	223	130	62	0
SCD_HEFT0023	223	292	140	75	0
SCD_HEFT0023	292	398	130	82	0
SCD_HEFT0023	398	525	118	68	0

SCD_HEFT0023	525	616	120	80	0
SCD_HEFT0023	616	732	120	80	1

The problem of missing data is common in medical trials (Dziura, Post, Zhao, Fu, & Peduzzi, 2013; Kenward, 2013). Data can be incorrectly recorded or not recorded at all. These were handled via simple imputation techniques, such as last observation carried forward (LOCF), and last observation carried backward (LOCB) (Dziura et al., 2013). A subject was eliminated from the dataset if any of its features could not be imputed (had no readings for that feature).

In addition, due to the low number of positive SCD cases in the dataset, random sampling techniques were used. Undersampling was applied to randomly eliminate majority class subjects (class = 0), while oversampling was used to randomly replicate minority class subjects (class = 1) (Kotsiantis et al., 2006).

For prediction purposes, the dataset was randomly partitioned into 80% training set and 20% testing set. The training set was equally divided into *k* partitions with $k = 5$ to use the k-fold cross-validation technique.

Performance Measures

To assess the performance of the prediction model, the following metrics were calculated:

- True Positive (TP): The patient suffers an SCD and the system correctly predicts SCD.
- False Positive (FP): The patient is normal, and the system incorrectly identifies him as an SCD case.
- True Negative (TN): The patient is normal, and the system correctly identifies him as such.

- False Negative (FN): The patient had an SCD and the system incorrectly identifies him as normal.
- Confusion Matrix: It demonstrates a comparison between predicted and actual outcome (Lantz, 2015). Table 5 shows the confusion matrix.

Table 5.

Confusion matrix

	Actual True	Actual False
Predicted True	TP	FP
Predicted False	FN	TN

- Accuracy: It is the ratio of the total number of correct assessment to the total number of evaluations (Sheela & Vanitha, 2014; Vanitha et al., 2014).

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

- Sensitivity: It refers to the ability of the system to correctly identify patients with cardiac arrest (Parikh, Mathai, Parikh, Sekhar, & Thomas, 2008; Sheela & Vanitha, 2014; Vanitha et al., 2014).

$$Sensitivity = \frac{TP}{(TP + FN)}$$

- Specificity: It refers to the ability of the system to correctly identify normal cases (Parikh et al., 2008; Sheela & Vanitha, 2014; Vanitha et al., 2014).

$$Specificity = \frac{TN}{(TN + FP)}$$

- Positive Predictive Value (PPV): It refers to the probability that the patient is having a cardiac arrest when the system identifies her as one having a cardiac arrest (Parikh et al., 2008; Sheela & Vanitha, 2014; Vanitha et al., 2014).

$$PPV = \frac{TP}{(TP + FP)}$$

- Negative Predictive Value (NPV): It refers to the probability that the patient is not having a cardiac arrest when the system identifies her as one not having a cardiac arrest (normal) (Parikh et al., 2008; Sheela & Vanitha, 2014; Vanitha et al., 2014).

$$NPV = \frac{TN}{(TN + FN)}$$

- Receiver Operating Characteristic (ROC) curve: It is used to examine the trade-off between the detection of true positives, while avoiding the false positives (Lantz, 2015). The Y-axis designates the true positive rate (sensitivity), while the X-axis designates false positive rate (1 – specificity). To summarize how well a clssifier performs, the area under the ROC curve (AUC) is used (Wang et al., 2015). Wang et al. (2015) define the AUC as the probability that the decision value assigned to a randomly-drawn positive sample is greater than the value assigned to a randomly-drawn negative sample. The AUC ranges from 0.5 to 1, with larger values representing higher system performance (Wang et al., 2015).
- F_1 Score: It is a measure of a test's accuracy through the use of the harmonic mean between sensitivity and PPV (Forte, 2015; Lantz, 2015).

$$F_1 = 2 \times \frac{PPV \times Sensitivity}{PPV + Sensitivity}$$

Prediction Model

The outcome of this dissertation was an SCD prediction model. This was done by using the Cox proportional hazard model to add a hazard score. The Cox proportional hazard model is the most widely used method for survival analysis (Fox, 2002). The term survival analysis refers to examining and modeling the elapsed time for events to occur (Fox, 2002). The Cox proportional hazard model relies on the assumption of the proportionality of the hazards, where the investigated factors have a constant impact on the hazard over time (Bellera et al., 2010). The Cox proportional hazard model is defined with the below equation:

$$h(t) = h_0(t) \times e^{\sum_{i=1}^{n} \beta_i \times X_i}$$

The value $h_0(t)$ is referred to as the baseline hazard rate (Bellera et al., 2010; Fox, 2002; Zhou, 2001). It is a non-negative function of time that constitutes the time-dependent part of the hazard (Bellera et al., 2010). It is the hazard rate when the values of all features are zeros (Bellera et al., 2010). β_i is the regression coefficient which gives the proportional change that can be expected, related to the change in the explanatory variables (Walters, 2009). X_i is the i^{th} feature value from the features' list and n is the total number of features. The regression coefficients are calculated using appropriate computer programs (Walters, 2009). This dissertation used the survival library in R. The model is constructed using the *coxph()* function where the coefficients were calculated.

The Hazard Ratio (HR) for each patient was calculated based on the coefficients obtained from fitting the Cox hazard model multiplied by the value of the features at the instant of time desired (180 days and 365 days before the occurrence of SCD in our case), using the formula below:

$$HR = \frac{h(t)}{h_0(t)} = e^{\sum_{i=1}^{n} \beta_i \times x_i}$$

The Extra-Trees algorithm (Geurts, Ernst, & Wehenkel, 2006) was used as an ensemble classifier to predict whether a patient will suffer an SCD at a prediction point. To enhance the accuracy of the model, we obtained the HR at prediction point, then added it to the dataset as a feature. The algorithm built an ensemble of unpruned trees (decision or regression) using the standard top-down procedure (Geurts et al., 2006). It differs from other tree-based ensemble methods in that it splits nodes by choosing cut-points fully at random, and that it uses the whole learning sample, not just a bootstrap replica, to grow the trees (Geurts et al., 2006). The explicit randomization of the cut-point and attribute, along with the ensemble averaging is aimed at reducing the variance in a stronger way than the randomization schemes used by other methods, while the usage of the full original learning sample is intended to minimize bias (Geurts et al., 2006). Using the Extra-Trees algorithm, the model with the dataset enhanced by Cox was trained, validated, and tested.

Model Training

With a Cox-enhanced training dataset supplied as input, the following steps were performed to train the model:

1. Import the necessary libraries. These include data handling, computations, and ensemble classifier libraries.
2. Read the training data file.
3. Pre-process training data, as necessary, to handle any unwanted characters in the file, or set categoral data as distinct numbers instead of strings.
4. Train the classifier by fitting the training dataset.

Model Validation

The k-fold cross validation technique was used to validate the proposed prediction model. A number of folds, k = 5 were chosen (Beleites, Neugebauer, Bocklitz, Krafft, & Popp, 2013). The Accuracy, Sensitivity, Specificity, F_1 Score, and Area under ROC metrics were used at each iteration. The mean and standard deviation (SD) were calculated to evaluate the overall performance of the system. The measurements were used to compare the results of all iterations to verify consistency across the whole validation process. Validation helped reduce bias and variance that might impact the parameters and performance of the predictive model (Beleites et al., 2013). Besides, it helped prevent overfitting the training data (Refaeilzadeh, Tang, & Liu, 2009). The k-fold validation process works as follows:

1. Divide the training set into *k* subsets, where the size of each subset is *n/k* (n = size of training set).
2. For *i = 1* to *k*
    a. Train each classifier of the ensemble using all folds but *i*.
    b. Test the ensemble using fold *i*.
    c. Compute performance metrics Accuracy, **S**ensitivity, **S**pecificity, **F_1 S**core, and **A**rea under ROC, for the current iteration.
3. Compute the mean and standard deviation of all iterations for each metric.

Model Testing

Using the Cox-enhanced test set, the following steps were performed to test the model at 6 months before SCD:

1. Read the test file, which included patient data at predict time.

2. Pre-process test data as necessary, to handle any unwanted characters in the file, or set categoral data as distinct numbers instead of strings.
3. Predict the outcome using the trained ensemble.
4. Calculate Accuracy, Sensitivity, Specificity, F_1 Score, and Area under ROC.

Model Comparison

The proposed prediction model in this dissertation was compared against three SCD prediction models published in peer-reviewed journals. The aim was to exceed the accuracy, the lead time, or both over the considered models. The first model was developed by Devi et al. (2016), which was a KNN classifier that used the Normal Sinus Rhythm dataset and the SCD Holter database from Physionet, and predicted SCD one hour before its occurrence at 95% accuracy. The second model was a C4.5 classification model developed by Mohanty et al. (2018), which utilized the CU Ventricular Tachyarrhythmia Database (CUDB) and MIT-BIH Malignant Ventricular Ectopy Database (VFDB) from Physionet, and achieved 90.97% sensitivity, 97.86% specificity, and 97.02% accuracy. The third model was an MLP model developed by Ebrahimzadeh et al. (2018), which was based on the MIT-BIH Sudden Cardiac Death Holter and Normal Sinus Rhythm databases from Physionet, and achieved 83.96% accuracy for SCD prediction within 4 minutes of its occurrence.

Chapter 4

Results

Overview

The goal of this dissertation was to develop a prediction model to predict SCD 180 days and 365 days before occurrence. We used a combination of Cox hazard model and the ExtraTreesClassifier ensemble technique. This chapter illustrates the results of the experiments we ran, along with the validation of the final model, and the comparison against the selected literature.

Model Results

We ran four experiments with the selected features. First, we generated the training and test data files with selected features and valid subjects. As previously stated, records containing features that had no measures could not be imputed, and had to be dropped. Also, patients that did not follow the study guidelines and had their first follow-up visit after the prediction point were dropped as well. Second, we reviewed our features selection process. Features that had no data in many records prevented the imputation and caused the majority of the records to be dropped. Therefore, these features were excluded in the first place. The used features were the ones that were statistically significant, were of medical importance (based on literature and on expert medical opinion), and were populated with data for the majority of the records. The statistical significance was determined by the R Boruta library. Boruta is a feature

selection wrapper algorithm that uses Random Forest and provides a variable importance measure (VIM) (Kursa & Rudnicki, 2010), which was used to guide the final feature selection. Based on the above, the final set of selected features is shown in table 6.

Table 6.

Final set of features

Feature
age
gender
diabetes
cabg
alcohol
systolic
diastolic
myocardialinfarction
nyha
weight

Finally, to compensate for the low number of positive cases in the dataset, we applied oversampling and undersampling techniques. We ran two variations of the test at 180 days and at 365 days as prediction points. These were as follows:

- We used the selected features at prediction point and added an additional field called *duration*, which is the time elapsed between the last test and the current one.
- We first obtained the Cox coefficients using the R survival library by fitting a Cox model on the entire training set, and then calculated the HR for both training and test sets at prediction point.
- We used the selected features at prediction point and added Cox model HR at prediction point.

180 Days with Cox

The resulting files from fitting the Cox model and adding the HR as a feature were used to train and test the ensemble classifier. Figure 1 shows the resulting coefficients.

Figure 1.

Cox coefficients when running at 180 days

```
Call:
coxph(formula = Surv(starttime, endtime, scd) ~ age + gendercode +
    diabetes + cabg + alcohol + systolic + diastolic + myocardialinfarction +
    nyha + weight, data = input.df)

  n= 4923, number of events= 112

                          coef  exp(coef)   se(coef)       z  Pr(>|z|)
age                  -0.0178141  0.9823437  0.0093429 -1.9067 0.0565611 .
gendercode            0.2355598  1.2656171  0.2613380  0.9014 0.3673966
diabetes              0.3677527  1.4444848  0.2051386  1.7927 0.0730204 .
cabg                  0.2179045  1.2434683  0.2224389  0.9796 0.3272763
alcohol               0.6196809  1.8583350  0.2206978  2.8078 0.0049877 **
systolic              0.0108148  1.0108735  0.0060934  1.7748 0.0759267 .
diastolic            -0.0206576  0.9795543  0.0103568 -1.9946 0.0460876 *
myocardialinfarction  0.9841024  2.6754094  0.7336124  1.3414 0.1797753
nyha                  0.5142974  1.6724630  0.1503349  3.4210 0.0006239 ***
weight               -0.0024480  0.9975550  0.0022821 -1.0727 0.2834194
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The training set size was at 534 records while the test set size was at 133 records. Table 7 lists the performance metrics of this test, while table 8 illustrates the confusion matrix.

Table 7.

Performance metrics – 180 days with Cox

Accuracy	Sensitivity	Specificity	F_1 Score	AUC
0.9624	0.7500	0.9915	0.9607	0.8707

Table 8.

Confusion matrix – 180 days with Cox

	0	1
0	116	1
1	4	12

Figure 2 shows the outcome of the folds' iterations, while figure 3 shows the overall outcome of the folds test.

Figure 2.

Outcome at each fold iteration – 180 days with Cox

	acc	sensitivity	specificity	f1_score	area_under_ROC
0	0.962617	0.833333	0.978947	0.962617	0.906140
1	0.990654	0.916667	1.000000	0.990476	0.958333
2	0.953271	0.750000	0.978947	0.952378	0.864474
3	0.962617	0.750000	0.989474	0.961112	0.869737
4	0.962264	0.750000	0.989362	0.960747	0.869681

Figure 3.

Overall Outcome of k-folds – 180 days with Cox

	measure	acc	sensitivity	specificity	f1_score	area_under_ROC
0	mean	0.966285	0.800000	0.987346	0.965466	0.893673
1	std	0.012699	0.066667	0.007871	0.013009	0.035602

180 Days without Cox

We first read the training and test files at prediction point. We then pre-processed the data then used it to train and test the ensemble classifier. The training set size was at 534 records while the test set size was at 133 records. Table 9 lists the performance metrics of this test, while table 10 illustrates the confusion matrix.

Table 9.

Performance metrics – 180 days without Cox

Accuracy	Sensitivity	Specificity	F_1 Score	AUC
0.9549	0.7500	0.9829	0.9536	0.8665

Table 10.

Confusion matrix – 180 days without Cox

	0	1
0	115	2
1	4	12

Figure 4 shows the outcome of the folds' iterations, while figure 5 shows the overall outcome of the folds test.

Figure 4.

Outcome at each fold iteration – 180 days without Cox

	acc	sensitivity	specificity	f1_score	area_under_ROC
0	0.981308	0.916667	0.989474	0.981308	0.953070
1	0.981308	0.916667	0.989474	0.981308	0.953070
2	0.981308	0.833333	1.000000	0.980556	0.916667
3	0.971963	0.750000	1.000000	0.970178	0.875000
4	0.971698	0.833333	0.989362	0.971158	0.911348

Figure 5.

Overall Outcome of k-folds – 180 days without Cox

	measure	acc	sensitivity	specificity	f1_score	area_under_ROC
0	mean	0.977517	0.850000	0.993662	0.976902	0.921831
1	std	0.004644	0.062361	0.005175	0.005107	0.029263

365 Days with Cox

The resulting files from fitting the Cox model and adding the HR as a feature were used to train and test the ensemble classifier. Figure 6 shows the resulting coefficients.

Figure 6.

Cox coefficients when running at 365 days

```
Call:
coxph(formula = Surv(starttime, endtime, scd) ~ age + gendercode +
    diabetes + cabg + alcohol + systolic + diastolic + myocardialinfarction +
    nyha + weight, data = input.df)

  n= 4548, number of events= 118

                          coef  exp(coef)   se(coef)       z  Pr(>|z|)
age                  -0.0022624  0.9977402  0.0095299 -0.2374   0.81235
gendercode            0.4854736  1.6249444  0.2763315  1.7569   0.07894 .
diabetes              0.4137317  1.5124512  0.1987803  2.0814   0.03740 *
cabg                  0.0853000  1.0890437  0.2154068  0.3960   0.69211
alcohol               0.4735166  1.6056306  0.2089742  2.2659   0.02346 *
systolic              0.0054649  1.0054798  0.0061680  0.8860   0.37561
diastolic            -0.0171534  0.9829929  0.0101203 -1.6949   0.09009 .
myocardialinfarction  1.4735938  4.3648934  0.6174876  2.3864   0.01701 *
nyha                  0.5925822  1.8086528  0.1500523  3.9492 7.842e-05 ***
weight               -0.0026418  0.9973617  0.0023129 -1.1422   0.25337
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The training set size was at 464 records while the test set size was at 116 records. Table 11 lists the performance metrics of this test, while table 12 illustrates the confusion matrix.

Table 11.

Performance metrics – 365 days with Cox

Accuracy	Sensitivity	Specificity	F_1 Score	AUC
0.9310	0.8333	0.9565	0.9310	0.8949

Table 12.

Confusion matrix – 365 days with Cox

	0	1
0	88	4
1	4	20

Figure 7 shows the outcome of the folds' iterations, while figure 8 shows the overall outcome of the folds test.

Figure 7.

Outcome at each fold iteration – 365 days with Cox

	acc	sensitivity	specificity	f1_score	area_under_ROC
0	0.956989	0.888889	0.973333	0.956989	0.931111
1	0.956989	0.944444	0.960000	0.957830	0.952222
2	0.935484	0.842105	0.959459	0.935484	0.900782
3	0.924731	0.684211	0.986486	0.920259	0.835349
4	0.923913	0.666667	0.986486	0.919020	0.826577

Figure 8.

Overall Outcome of k-folds – 365 days with Cox

	measure	acc	sensitivity	specificity	f1_score	area_under_ROC
0	mean	0.939621	0.805263	0.973153	0.937917	0.889208
1	std	0.014757	0.110982	0.011968	0.016942	0.050367

365 Days without Cox

We first read the training and test files at prediction point. We then pre-processed the data then used it to train and test the ensemble classifier. The training set size was at 464 records while the test set size was at 116 records. Table 13 lists the performance metrics of this test, while table 14 illustrates the confusion matrix.

Table 13.

Performance metrics – 365 days without Cox

Accuracy	Sensitivity	Specificity	F_1 Score	AUC
0.9483	0.8333	0.9783	0.9474	0.9058

Table 14.

Confusion matrix – 365 days without Cox

	0	1
0	90	2
1	4	20

Figure 9 shows the outcome of the folds' iterations, while figure 10 shows the overall outcome of the folds test.

Figure 9.

Outcome at each fold iteration – 365 days without Cox

	acc	sensitivity	specificity	f1_score	area_under_ROC
0	0.978495	0.888889	1.000000	0.978004	0.944444
1	0.989247	1.000000	0.986667	0.989357	0.993333
2	0.956989	0.894737	0.972973	0.956989	0.933855
3	0.924731	0.736842	0.972973	0.922253	0.854908
4	0.923913	0.722222	0.972973	0.921210	0.847598

Figure 10.

Overall Outcome of k-folds – 365 days without Cox

	measure	acc	sensitivity	specificity	f1_score	area_under_ROC
0	mean	0.954675	0.848538	0.981117	0.953563	0.914828
1	std	0.026873	0.105009	0.010829	0.027990	0.055700

Summary

This chapter provided the detailed results of test and validation of the SCD prediction model. Four tests were conducted for 180 days and 365 days, with and without Cox HR. The model was validated using a k-fold cross validation technique with $k = 5$. The assessment used the accuracy, sensitivity, specificity, F_1 score, and AUC to evaluate the model's performance.

The 180-days with Cox yielded the best accuracy at 0.9624, the best specificity at 0.9915, and the best F_1 score at 0.9607. The best AUC was obtained from the 365-days without Cox, at 0.9058. Cox had no effect on sensitivity for both 180-days and 365-days. It was at 0.7500 and 0.8333 respectively.

We compared the model with three established ones. The first model was developed by Devi et al. (2016), which is a KNN classifier that used the Normal Sinus Rhythm dataset and the SCD Holter database from Physionet, and predicted SCD one hour before its occurrence at 95% accuracy. The second model was a C4.5 classification model developed by Mohanty et al. (2018), which utilized the CU Ventricular Tachyarrhythmia Database (CUDB) and MIT-BIH Malignant Ventricular Ectopy Database (VFDB) from Physionet, and achieved 90.97% sensitivity, 97.86% specificity, and 97.02% accuracy. This paper used the Physionet normal sinus rhythm (NSR), the MIT-BIH Malignant Ventricular Ectopy, and the CU Ventricular Tachyarrhythmia Database. These are research databases and are characterized by the low number of subjects (18, 22, and 35 respectively). In addition, the research did not indicate any prediction time. Even after sampling, we used 676 subjects for 180-days prediction point, and 580 subjects for 365-days prediction point. The third model is an MLP model

developed by Ebrahimzadeh et al. (2018), which was based on the MIT-BIH Sudden Cardiac Death Holter and Normal Sinus Rhythm databases from Physionet, and achieved 83.96% accuracy for SCD prediction within 4 minutes of its occurrence.

Our system outperformed the other three systems in terms of lead time which compensates for the small difference in accuracy with respect to the work present by Mohanty et al. (2018).

Chapter 5

Conclusions, Implications, Recommendations, and Summary

Overview

This dissertation presented an ensemble model to predict SCD within 180 and 365 days. The model blended the strengths of the Cox Hazard Model with the power of the ExtraTreesClassifier ensemble. This chapter concludes the dissertation and highlights the implications of the current work in the medical field, particularly in the area of SCD prediction. It then provides recommendations for future work and ends with a summary of the work.

Conclusions

The focus of this study was to answer the following question:

RQ

Can one develop an ensemble model to predict SCD with acceptable accuracy and practical lead time?

The above question was answered through the development of an SCD prediction system that can predict the condition both at 6 months to one year before its occurrence. The system outperformed the work it is compared to in terms of lead time, and was very competitive in terms of accuracy, sensitivity, and specificity. The performance of the current work compared to other models is summarized in table 15.

Table 15.

Model Comparison

		Accuracy	Sensitivity	Specificity	F_1 Score	AUC	Lead Time
Current Model	Cox	0.9624	0.7500	0.9916	0.9607	0.8707	180 days
	No Cox	0.9549	0.7500	0.9829	0.9536	0.8665	180 days
	Cox	0.9310	0.8333	0.9565	0.9310	0.8949	365 days
	No Cox	0.9438	0.8333	0.9783	0.9474	0.9058	365 days
	(Devi et al., 2016)	0.9500	N/A	N/A	N/A	N/A	1 hour
	(Mohanty et al., 2018)	0.9702	0.9097	0.9786	N/A	N/A	N/A
	(Ebrahimzadeh et al., 2018)	0.8396	N/A	N/A	N/A	N/A	4 min.

Implications

Successful prediction of SCD can have huge impact on the medical system. SCD is generally treated with ICD, which is expensive and sometimes dangerous to implant, especially with patients with heart condition. Being able to predict it in advance can enable doctors to implant the device before it is too late and while the patient is in a good condition to sustain the procedure.

On the other hand, ICDs will only be used with patients that actually need them. This is particularly important in under-privileged areas where medical supplies are scarce. In addition, it will help drive the insurance costs down, since the ICD implants will be done on an as-needed basis not as a precautionary measure.

The work presented in this dissertation contributed to the current state of SCD prediction. The tool is able to predict the potential SCD with a comfortable lead time, allowing medical professionals the time to react and decide on treatment options.

Recommendations

As previously discussed, the SCD-HeFT data that was used in the current work was characterized by the low number of positive cases with respect to the entire dataset. For better validation, it is recommended to apply other datasets to the prediction model.

In addition, the current model can be applied in other medical areas, or even non-medical areas that incorporate time variant data, such as weather or agriculture. Finally, other ensemble mechanisms can be applied to see if that can outperform the Extra-Trees Classifier that was used in the current work.

Summary

In this dissertation, we developed an SCD prediction system for patients with NYHA class II and class III heart failure, using Cox hazard model and the ExtraTrees ensemble classifier. Data from the Sudden Cardiac Death in Heart Failure Trial (SCD-HeFT) (Bardy et al., 2005). We started with 1692 subjects, and due to sampling and exclusions, ended up using 676 subjects for 180-days prediction point, and 580 subjects for 365-days prediction point. The highest accuracy we achieved was 0.9624 at 180-days with Cox. The best specificity was 0.9915, and the best F1 score was 0.9607, at the same time. The best AUC was obtained from the 365-days without Cox, at 0.9058. Cox had no effect on sensitivity, which was at 0.7500 and 0.8333 for 180-days and 365-days respectively.

In conclusion, the current model was able to predict SCD at lead times up to a year, while maintaining a high-level of prediction accuracy, thus adding a contribution to the general body of knowledge.

References

Abad, S. K. J., Zare-Mirakabad, M. R., & Rezaeian, M. (2014, 29-30 Oct. 2014). *An approach for classifying large dataset using ensemble classifiers.* Paper presented at the Computer and Knowledge Engineering (ICCKE), 2014 4th International eConference on.

Acharya, U. R., Fujita, H., Sudarshan, V. K., Ghista, D. N., Lim, W. J. E., & Koh, J. E. W. (2015, 9-12 Oct. 2015). *Automated Prediction of Sudden Cardiac Death Risk Using Kolmogorov Complexity and Recurrence Quantification Analysis Features Extracted from HRV Signals.* Paper presented at the Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on.

Acharya, U. R., Fujita, H., Sudarshan, V. K., Sree, V. S., Eugene, L. W. J., Ghista, D. N., & San Tan, R. (2015). An integrated index for detection of sudden cardiac death using discrete wavelet transform and nonlinear features. *Knowledge-Based Systems, 83*, 149-158.

Adabag, S., Rector, T. S., Anand, I. S., McMurray, J. J., Zile, M., Komajda, M., . . . Carson, P. E. (2014). A prediction model for sudden cardiac death in patients with heart failure and preserved ejection fraction. *European journal of heart failure, 16*(11), 1175-1182.

Agoston-Coldea, L., Bheecarry, K., Petra, C., Strambu, L., Ober, C., Revnic, R., . . . Fodor, D. (2018). The value of global longitudinal strain and galectin-3 for predicting cardiovascular events in patients with severe aortic stenosis. *Medical ultrasonography, 20*(2), 205-212.

Al-Gobari, M., Le, H.-H., Fall, M., Gueyffier, F., & Burnand, B. (2017). No benefits of statins for sudden cardiac death prevention in patients with heart failure and reduced ejection fraction: A meta-analysis of randomized controlled trials. *PloS one, 12*(2), e0171168.

American Heart Association. (2017). Syncope (Fainting). Retrieved from http://www.heart.org/HEARTORG/Conditions/Arrhythmia/SymptomsDiagnosisMonitoringofArrhythmia/Syncope-Fainting_UCM_430006_Article.jsp#.WscVJJch2Uk

Ayesta, A., Martínez-Sellés, H., de Luna, A. B., & Martínez-Sellés, M. (2018). Prediction of sudden death in elderly patients with heart failure. *Journal of geriatric cardiology: JGC, 15*(2), 185.

Bagheri, M. A., & Gao, Q. (2012). *An efficient ensemble classification method based on novel classifier selection technique.* Paper presented at the Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics.

Bardy , G. H., Lee , K. L., Mark , D. B., Poole , J. E., Packer , D. L., Boineau , R., . . . Ip , J. H. (2005). Amiodarone or an Implantable Cardioverter–Defibrillator for Congestive Heart Failure. *New England Journal of Medicine, 352*(3), 225-237. doi:10.1056/NEJMoa043399

Bauer, E., & Kohavi, R. (1998). An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine learning, 36*(1), 2.

Beleites, C., Neugebauer, U., Bocklitz, T., Krafft, C., & Popp, J. (2013). Sample size planning for classification models. *Analytica chimica acta, 760*, 25-33.

Bellera, C. A., MacGrogan, G., Debled, M., de Lara, C. T., Brouste, V., & Mathoulin-Pélissier, S. (2010). Variables with time-varying effects and the Cox model: Some statistical concepts illustrated with a prognostic factor study in breast cancer. *BMC Medical Research Methodology, 10*(1), 20. doi:10.1186/1471-2288-10-20

Breiman, L. (1996). Bagging predictors. *Machine learning, 24*(2), 123-140.

Darwiche, A., & Mukherjee, S. (2018). *Machine Learning Methods for Septic Shock Prediction.* Paper presented at the Proceedings of the 2018 International Conference on Artificial Intelligence and Virtual Reality.

Deo, R., Norby, F. L., Katz, R., Sotoodehnia, N., Adabag, S., DeFilippi, C. R., . . . Folsom, A. R. (2016). Development and Validation of a Sudden Cardiac Death Prediction Model for the General PopulationClinical Perspective. *Circulation, 134*(11), 806-816.

Desai, M. Y., Smedira, N. G., Dhillon, A., Masri, A., Wazni, O., Kanj, M., . . . Lever, H. M. (2018). Prediction of sudden death risk in obstructive hypertrophic cardiomyopathy: potential for refinement of current criteria. *The Journal of thoracic and cardiovascular surgery*.

Devi, R., Tyagi, H. K., & Kumar, D. (2016). Heart Rate Variability Analysis for Early Stage Prediction of Sudden Cardiac Death. *World Academy of Science, Engineering and Technology, International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering, 10*(3), 401-404.

Deyell, M. W., Krahn, A. D., & Goldberger, J. J. (2015). Sudden cardiac death risk stratification. *Circulation research, 116*(12), 1907-1918.

Dietterich, T. G. (2000). Ensemble methods in machine learning *Multiple classifier systems* (pp. 1-15): Springer.

Dietterich, T. G. (2002). Ensemble learning. *The handbook of brain theory and neural networks, 2*, 110-125.

Dziura, J. D., Post, L. A., Zhao, Q., Fu, Z., & Peduzzi, P. (2013). Strategies for dealing with missing data in clinical trials: from design to analysis. *The Yale journal of biology and medicine, 86*(3), 343.

Ebrahimzadeh, E., Fayaz, F., Ahmadi, F., & Dolatabad, M. R. (2018). Linear and nonlinear analyses for detection of sudden cardiac death (SCD) using ECG and HRV signals. *Trends Med. Res., 1*(1), 1-8.

Ebrahimzadeh, E., & Pooyan, M. (2011). Early detection of sudden cardiac death by using classical linear techniques and time-frequency methods on electrocardiogram signals. *Journal of Biomedical Science and Engineering, 4*(11), 699.

Ebrahimzadeh, E., Pooyan, M., & Bijar, A. (2014). A novel approach to predict sudden cardiac death (SCD) using nonlinear and time-frequency analyses from HRV signals. *PloS one, 9*(2), e81896.

Ejection fraction: What does it measure? (2018). Retrieved from https://www.mayoclinic.org/ejection-fraction/expert-answers/FAQ-20058286

Eranti, A., Aro, A. L., Kerola, T., Tikkanen, J. T., Rissanen, H. A., Anttonen, O., . . . Huikuri, H. V. (2016). Body Mass Index as a Predictor of Sudden Cardiac Death and Usefulness of the Electrocardiogram for Risk Stratification. *The American journal of cardiology, 117*(3), 388-393.

Fan, X., Hua, W., Xu, Y., Ding, L., Niu, H., Chen, K., . . . Zhang, S. (2014). Incidence and predictors of sudden cardiac death in patients with reduced left ventricular ejection fraction after myocardial infarction in an era of revascularisation. *Heart*, heartjnl-2013-305144.

Fishman, G. I., Chugh, S. S., DiMarco, J. P., Albert, C. M., Anderson, M. E., Bonow, R. O., . . . Jouven, X. (2010). Sudden cardiac death prediction and prevention report from a National Heart, Lung, and Blood Institute and Heart Rhythm Society workshop. *Circulation, 122*(22), 2335-2348.

Forte, R. M. (2015). *Mastering predictive analytics with R*: Packt Publishing Ltd.

Fox, J. (2002). Cox proportional-hazards regression for survival data. *An R and S-PLUS companion to applied regression, 2002*.

Freund, Y., & Schapire, R. E. (1996). *Experiments with a new boosting algorithm.* Paper presented at the ICML.

García Iglesias, D., Roqueñi Gutiérrez, N., De Cos, F. J., & Calvo, D. (2018). Analysis of the High-Frequency Content in Human QRS Complexes by the Continuous

Wavelet Transform: An Automatized Analysis for the Prediction of Sudden Cardiac Death. *Sensors, 18*(2), 560.

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine learning, 63*(1), 3-42.

Ghavidel, J., Yazdani, S., & Analoui, M. (2013, 28-30 May 2013). *A new ensemble classifier creation method by creating new training set for each base classifier.* Paper presented at the Information and Knowledge Technology (IKT), 2013 5th Conference on.

Goldberger, J. J., Basu, A., Boineau, R., Buxton, A. E., Cain, M. E., Canty, J. M., . . . Exner, D. V. (2014). Risk stratification for sudden cardiac death a plan for the future. *Circulation, 129*(4), 516-526.

Goldstein, B. A., Chang, T. I., Mitani, A. A., Assimes, T. L., & Winkelmayer, W. C. (2014). Near-term prediction of sudden cardiac death in older hemodialysis patients using electronic health records. *Clinical Journal of the American Society of Nephrology, 9*(1), 82-91.

Gupta, R., Audhkhasi, K., & Narayanan, S. (2014, 4-9 May 2014). *Training ensemble of diverse classifiers on feature subsets.* Paper presented at the Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on.

Harrington, P. (2012). *Machine Learning in Action*: Manning Publications Co.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning* (Second ed.): Springer.

Ho, T. K. (1998). The random subspace method for constructing decision forests. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 20*(8), 832-844.

Huikuri, H. V., Mäkikallio, T. H., Raatikainen, M. P., Perkiömäki, J., Castellanos, A., & Myerburg, R. J. (2003). Prediction of sudden cardiac death appraisal of the studies and methods assessing the risk of sudden arrhythmic death. *Circulation, 108*(1), 110-115.

Johansson, R., Bostrom, H., & Karlsson, A. (2008, 20-22 Aug. 2008). *A study on class-specifically discounted belief for ensemble classifiers.* Paper presented at the Multisensor Fusion and Integration for Intelligent Systems, 2008. MFI 2008. IEEE International Conference on.

Kakimoto, Y., Tanaka, M., Hayashi, H., Yokoyama, K., & Osawa, M. (2018). Overexpression of miR-221 in sudden death with cardiac hypertrophy patients. *Heliyon, 4*(6), e00639.

Kenward, M. G. (2013). The handling of missing data in clinical trials. *Clinical Investigation, 3*(3), 241-250.

Klein, H., Auricchio, A., Reek, S., & Geller, C. (1999). New primary prevention trials of sudden cardiac death in patients with left ventricular dysfunction: SCD-HEFT and MADIT-II. *The American journal of cardiology, 83*(5, Supplement 2), 91-97. doi:https://doi.org/10.1016/S0002-9149(99)00040-5

Kotsiantis, S., Kanellopoulos, D., & Pintelas, P. (2006). Handling imbalanced datasets: A review. *GESTS International Transactions on Computer Science and Engineering, 30*(1), 25-36.

Kubik, M., Dąbrowska-Kugacka, A., Lewicka, E., Daniłowicz-Szymanowicz, L., & Raczak, G. (2018). Predictors of poor outcome in patients with left ventricular noncompaction: Review of the literature. *Advances in clinical and experimental medicine: official organ Wroclaw Medical University*.

Kurbanov, R. D., Mullabaeva, G. U., & Kilichev, A. A. (2015). Main Predictors of Sudden Cardiac Death in Patients with Q-Wave Myocardial Infarction. *INTERNATIONAL JOURNAL OF BIOMEDICINE, 5*(4), 195-197.

Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the Boruta package. *J Stat Softw, 36*(11), 1-13.

Lantz, B. (2015). *Machine Learning with R* (Second ed.): Packt Publishing.

Lee, H., Shin, S.-Y., Seo, M., Nam, G.-B., & Joo, S. (2016). Prediction of Ventricular Tachycardia One Hour before Occurrence Using Artificial Neural Networks. *Scientific Reports, 6*.

Li, M., Chen, X., Chen, L., Chen, K., Zhou, J., & Song, J. (2018). MiR-1-3p that correlates with left ventricular function of HCM can serve as a potential target and differentiate HCM from DCM. *Journal of translational medicine, 16*(1), 161.

Liew, R. (2011). Electrocardiogram-Based Predictors of Sudden Cardiac Death in Patients With Coronary Artery Disease. *Clinical cardiology, 34*(8), 466-473.

Mayo Clinic. (2018a). Hypertrophic cardiomyopathy. Retrieved from https://www.mayoclinic.org/diseases-conditions/hypertrophic-cardiomyopathy/symptoms-causes/syc-20350198

Mayo Clinic. (2018b). Implantable cardioverter-defibrillators (ICDs). Retrieved from https://www.mayoclinic.org/tests-procedures/implantable-cardioverter-defibrillators/about/pac-20384692

Mayo Clinic Staff. (2016). Vasodilator. Retrieved from https://www.mayoclinic.org/diseases-conditions/high-blood-pressure/in-depth/high-blood-pressure-medication/ART-20048154?p=1

MedlinePlus. (2017a). Amiodarone. Retrieved from https://medlineplus.gov/druginfo/meds/a687009.html

MedlinePlus. (2017b). Contraindication. Retrieved from https://medlineplus.gov/ency/article/002314.htm

MedlinePlus. (2017c). Heart Failure. Retrieved from https://medlineplus.gov/heartfailure.html

Merriam-Webster Medical Dictionary. (2017). Retrieved from https://www.merriam-webster.com/dictionary/congestive%20heart%20failure

Mohanty, M., Sahoo, S., Biswal, P., & Sabut, S. (2018). Efficient classification of ventricular arrhythmias using feature selection and C4. 5 classifier. *Biomedical Signal Processing and Control*, *44*, 200-208.

Murugappan, M., Murukesan, L., Omar, I., Khatun, S., & Murugappan, S. (2015). Time Domain Features Based Sudden Cardiac Arrest Prediction Using Machine Learning Algorithms. *Journal of Medical Imaging and Health Informatics*, *5*(6), 1267-1271.

Murukesan, L., Murugappan, M., & Iqbal, M. (2013, 8-10 March 2013). *Sudden cardiac death prediction using ECG signal derivative (Heart Rate Variability): A review.* Paper presented at the Signal Processing and its Applications (CSPA), 2013 IEEE 9th International Colloquium on.

Murukesan, L., Murugappan, M., Iqbal, M., & Saravanan, K. (2014). Machine learning approach for sudden cardiac arrest prediction based on optimal heart rate variability features. *Journal of Medical Imaging and Health Informatics*, *4*(4), 521-532.

Ng, G. A., Mistry, A., Li, X., Schlindwein, F. S., & Nicolson, W. B. (2018). LifeMap: towards the development of a new technology in sudden cardiac death risk stratification for clinical use. *EP Europace*.

O'Mahony, C., Jichi, F., Pavlou, M., Monserrat, L., Anastasakis, A., Rapezzi, C., . . . McKenna, W. J. (2014). A novel clinical risk prediction model for sudden cardiac death in hypertrophic cardiomyopathy (HCM Risk-SCD). *European heart journal*, *35*(30), 2010-2020.

Özyılmaz, S., Satılmışoğlu, M. H., Gül, M., Uyarel, H., & Serdar, O. A. (2018). Evaluation of the association between serum uric acid level and the predicted risk

score of sudden cardiac death in five years in patients with hypertrophic cardiomyopathy. *Turk Kardiyol Dern Ars, 46*(2), 111-120.

Parikh, R., Mathai, A., Parikh, S., Sekhar, G. C., & Thomas, R. (2008). Understanding and using sensitivity, specificity and predictive values. *Indian journal of ophthalmology, 56*(1), 45.

Pascual-Figal, D. A., Ordoñez-Llanos, J., Tornel, P. L., Vázquez, R., Puig, T., Valdés, M., . . . Bayes-Genis, A. (2009). Soluble ST2 for predicting sudden cardiac death in patients with chronic heart failure and left ventricular systolic dysfunction. *Journal of the American College of Cardiology, 54*(23), 2174-2179.

Piccini, J. P., Zhang, M., Pieper, K., Solomon, S. D., Al-Khatib, S. M., Van de Werf, F., . . . Velazquez, E. J. (2010). Predictors of sudden cardiac death change with time after myocardial infarction: results from the VALIANT trial. *European heart journal, 31*(2), 211-221.

Rai, V., & Agrawal, D. K. (2016). Role of risk stratification and genetics in sudden cardiac death. *Canadian journal of physiology and pharmacology, 95*(3), 225-238.

Ramírez, J., Orini, M., & Laguna, E. P. P. (2017). Comparison of ECG T-wave Duration and Morphology Restitution Markers for Sudden Cardiac Death Prediction in Chronic Heart Failure. *Computing, 44*, 1.

Ramírez, J., Orini, M., Mincholé, A., Monasterio, V., Cygankiewicz, I., Bayés de Luna, A., . . . Laguna, P. (2017). T-Wave Morphology Restitution Predicts Sudden Cardiac Death in Patients With Chronic Heart Failure. *Journal of the American Heart Association, 6*(5). doi:10.1161/jaha.116.005310

Ramos-Jimenez, G., del Campo-Avila, J., & Morales-Bueno, R. (2009, Nov. 30 2009-Dec. 2 2009). *Hybridizing Ensemble Classifiers with Individual Classifiers.* Paper presented at the Intelligent Systems Design and Applications, 2009. ISDA '09. Ninth International Conference on.

Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-Validation (pp. 532-538).

Rosset, S., Domingo, A. M., Asimaki, A., Graf, D., Metzger, J., Schwitter, J., . . . Pruvot, E. (2018). Reduced Desmoplakin immunofluorescence signal in arrhythmogenic cardiomyopathy with epicardial right ventricular outflow tract tachycardia. *HeartRhythm Case Reports*.

Royston, P., & Altman, D. G. (2013). External validation of a Cox prognostic model: principles and methods. *BMC Medical Research Methodology, 13*(1), 33.

Shastri, S., Tangri, N., Tighiouart, H., Beck, G. J., Vlagopoulos, P., Ornt, D., . . . Cheung, A. K. (2012). Predictors of sudden cardiac death: a competing risk approach in the hemodialysis study. *Clinical Journal of the American Society of Nephrology, 7*(1), 123-130.

Sheela, C. J., & Vanitha, L. (2014, 20-21 March 2014). *Prediction of Sudden Cardiac Death using support vector machine.* Paper presented at the Circuit, Power and Computing Technologies (ICCPCT), 2014 International Conference on.

Shen, T.-W., Shen, H.-P., Lin, C.-H., & Ou, Y.-L. (2007, 22-26 Aug. 2007). *Detection and Prediction of Sudden Cardiac Death (SCD) For Personal Healthcare.* Paper presented at the Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE.

Shiga, T., Kohro, T., Yamasaki, H., Aonuma, K., Suzuki, A., Ogawa, H., . . . Kasanuki, H. (2018). Body Mass Index and Sudden Cardiac Death in Japanese Patients After Acute Myocardial Infarction: Data From the JCAD Study and HIJAMI-II Registry. *Journal of the American Heart Association, 7*(14), e008633.

Skurichina, M., & Duin, R. P. (2002). Bagging, boosting and the random subspace method for linear classifiers. *Pattern Analysis & Applications, 5*(2), 121-135.

Sotto, L. F. D. P., Coelho, R. C., & Melo, V. V. d. (2016). *Classification of Cardiac Arrhythmia by Random Forests with Features Constructed by Kaizen Programming with Linear Genetic Programming*. Paper presented at the Proceedings of the Genetic and Evolutionary Computation Conference 2016, Denver, Colorado, USA.

Stecker, E. C., & Chugh, S. S. (2011). Prediction of sudden cardiac death: next steps in pursuit of effective methodology. *Journal of interventional cardiac electrophysiology, 31*(2), 101-107.

Su, Y., Xia, M., Cao, J., & Gao, Q. (2018). Cardiac characteristics in the premature ventricular contraction patients with or without ventricular tachycardia. *INTERNATIONAL JOURNAL OF CLINICAL AND EXPERIMENTAL MEDICINE, 11*(6), 6106-6112.

Therneau, T., Crowson, C., & Atkinson, E. (2013). Using time dependent covariates and time dependent coefficients in the cox model. *Red, 2*, 1.

Thomsen, M. B., Nielsen, M. S., Aarup, A., Bisgaard, L. S., & Pedersen, T. X. (2018). Uremia increases QRS duration after β-adrenergic stimulation in mice. *Physiological reports, 6*(13), e13720.

Toshniwal, D., Goel, B., & Sharma, H. (2015). *Multistage Classification for Cardiovascular Disease Risk Prediction.* Paper presented at the International Conference on Big Data Analytics.

Vadakkumpadan, F., Trayanova, N., Younes, L., & Wu, K. C. (2012, Aug. 28 2012-Sept. 1 2012). *Left-ventricular shape analysis for predicting sudden cardiac death risk.* Paper presented at the Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE.

Valentini, G., & Masulli, F. (2002). Ensembles of learning machines *Neural Nets* (pp. 3-20): Springer.

Van Dijk, M., Steyerberg, E., Stenning, S., Dusseldorp, E., & Habbema, J. (2004). Survival of patients with nonseminomatous germ cell cancer: a review of the IGCC classification by Cox regression and recursive partitioning. *British journal of cancer, 90*(6), 1176.

Vanitha, L., Suresh, G. R., & JenefarSheela, C. (2014, 13-14 Feb. 2014). *Sudden Cardiac Death prediction system using Hybrid classifier.* Paper presented at the Electronics and Communication Systems (ICECS), 2014 International Conference on.

Verma, B., & Rahman, A. (2012). Cluster-Oriented Ensemble Classifier: Impact of Multicluster Characterization on Ensemble Classifier Learning. *Knowledge and Data Engineering, IEEE Transactions on, 24*(4), 605-618. doi:10.1109/TKDE.2011.28

Walters, S. J. (2009). What is a Cox model? Retrieved from https://pdfs.semanticscholar.org/5142/55da06320e50add5b2e6f00287c28c7b2168.pdf

Wang, S., Li, D., Petrick, N., Sahiner, B., Linguraru, M. G., & Summers, R. M. (2015). Optimizing area under the ROC curve using semi-supervised learning. *Pattern Recognition, 48*(1), 276-287. doi:https://doi.org/10.1016/j.patcog.2014.07.025

Weeks, P. A., Sieg, A., Gass, J. A., & Rajapreyar, I. (2016). The role of pharmacotherapy in the prevention of sudden cardiac death in patients with heart failure. *Heart failure reviews, 21*(4), 415-431.

Wellens, H. J., Schwartz, P. J., Lindemans, F. W., Buxton, A. E., Goldberger, J. J., Hohnloser, S. H., . . . Malik, M. (2014). Risk stratification for sudden cardiac death: current status and challenges for the future. *European heart journal, 35*(25), 1642-1651.

Whalen, S., & Pandey, G. (2013, 7-10 Dec. 2013). *A Comparative Analysis of Ensemble Classifiers: Case Studies in Genomics.* Paper presented at the Data Mining (ICDM), 2013 IEEE 13th International Conference on.

Yu, Z., Li, L., Liu, J., & Han, G. (2015). Hybrid Adaptive Classifier Ensemble. *Cybernetics, IEEE Transactions on, 45*(2), 177-190.

Zhao, h. J. (2005). *Mixed-effects Cox models of alcohol dependence in extended families.* Paper presented at the BMC genetics.

Zhou, M. (2001). Understanding the Cox Regression Models With Time-Change Covariates. *The American Statistician, 55*(2), 153-155. doi:10.1198/000313001750358491