Cardiac Arrhythmia Monitoring and Severe Event Prediction System

by

Zhi Li

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Bioinformatics)
in The University of Michigan
2021

Doctoral Committee:

Professor Kayvan Najarian, Chair
Professor Harm Derksen
Assistant Professor Hamid Ghanbari
Associate Professor Alla Karnovsky
Assistant Professor Michael R. Mathis
Professor Gilbert S. Omenn

Zhi Li

zcli@umich.edu

ORCID iD: 0000-0002-9866-8396

For Abigail, Isaac and Shuyang

ACKNOWLEDGEMENTS

This project would not have been possible without the support of many people. First I would like to express my deepest gratitude to my advisor Professor Kayvan Najarian for his marvelous supervision, guidance and encouragement. I am always grateful for his encouragement to transfer into the field of bioinformatics, pursue this degree and for the opportunities to join BCIL. He is the true definition of a leader and the ultimate role model. His insight, breadth of knowledge and passion for research steered me through my Ph.D. research.

I would also like to thank Toyota Motor North America for funding the study and providing the financial means to complete the project.

I am fortunate to have my dissertation committee. I would like to thank Dr. Harm Derksen for his expertise in mathematics and guidance in algorithm development has greatly helped this thesis. The project would not be possible without his guidance. Dr. Hamid Ghanbari for all his clinical advices, help with paper reviews and all those rhythm annotations. This research would not be possible without his expertise. Dr. Gilbert Omenn for his valuable feedbacks in committee meetings and paper reviews. I am inspired from his dedication to science and attention to details. Dr. Alla Karnovsky for all the support in committee meetings and wonderful feedbacks in research. Dr. Michael Mathis for all his medical expertise and insightful comments in committee meetings and paper reviews.

I would like to thank the support from all the faculties and staff in Department of Computational Medicine and Bioinformatics. I would like to express special thanks

to Dr. Jonathan Gryak for his knowledge and expertise in the field, his countless support to all the projects and all the advice which has encouraged me through the years.

I would like to thank all my labmates and colleagues for being great friends throughout the years. I am very blessed to have these friends and make my time in the lab so enjoyable.

I am grateful for my parents whose constant love and support keep me motivated and confident. Abigail and Isaac, you have brought endless joy and new meanings to my life. Last but not least, Shuyang, thank you for always being there for me, whenever and wherever.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF ABBREVIATIONS

AFib atrial fibrillation

AHRE atrial high rate episodes

AUC Area under the ROC Curve

BPM beats-per-minute

CNN convolutional neural network

DB1 Database 1

DB2 Database 2

DB3 Database 3

DB4 Database 4

DB5 Database 5

DB6 Database 6

DNN deep neural network

DPFA deterministic probabilistic finite-state automata

DWT discrete wavelet transform

ECG electrocardiogram

GDA generalized discriminant analysis

HMM hidden Markov model

HR heart rate

HRV heart rate variability

ICA independent component analysis

IQR inter-quantile range

LSTM long short-term memory

MAD mean amplitude deviation

PCA principal component analysis

PFA probabilistic finite-state automata

RBF radial basis function

RNN recurrent neural network

RVR rapid ventricular rates

SCD sudden cardiac death

SVM support vector machine

SVT supraventricular tachycardia

VA ventricular arrhythmia

VF ventricular fibrillation

VFlutter ventricular flutter

VT ventricular tachycardia

ABSTRACT

Abnormalities of cardiac rhythms are correlated with significant morbidity. For example atrial fibrillation, affecting at least 2.3 million people in the United States, is associated with an increased risk of both stroke and mortality; supra-ventricular tachycardia, detected in approximately 90,000 cases annually in the United States, results in hospitalization in about 25% of all emergency department visits for supra-ventricular tachycardia; ventricular arrhythmias cause 75% to 80% of the cases of sudden cardiac death; bradyarrhythmias and other forms of conduction disease may cause syncope, fatigue from chronotropic incompetence, or sudden death from asystole or ventricular tachycardia. Due to the time-sensitive nature of cardiac events, it is of utmost importance to ensure that medical intervention is provided in a timely manner, which could benefit greatly from a cardiac arrhythmia monitoring system that can detect and preferably also predict the abnormal cardiac events.

In recent years, with the development of cardiac and other types of medical monitoring devices, vast amounts of physiological signal data have been collected and become available for analysis. Physiological signals such as electrocardiogram (ECG) have many clinical applications in cardiac arrhythmia, including diagnosis confirmation, drug effect monitoring and rate control. However, the extraction of the relevant information from physiological signals—despite its great value—is hindered by the complexity and variability within signal morphology, which leads to vague definitions and ambiguous guidelines, causing difficulties even for a well-trained medical expert.

Such variability-related issues are ubiquitous and manifested in different ways: via the ECG signals themselves, the measurements derived from such signals, and the diagnostic interpretations based on such measurements.

In order to address the variability-related issues, most traditional methods for physiological signal analysis depend heavily on pre-processing to identify specific morphology types (such as R peaks in ECG) and extract the related features. Despite many successes, one of the drawbacks of these methods is that they often require signal data of high quality and tend to be less effective in the presence of noise which could significantly distort the signal morphology. Although not an issue in almost noiseless situations such as bedside ECG monitoring, such pre-processing–based methods have become insufficient with the advent of portable arrhythmia monitoring devices in recent years capable of collecting physiological signals in real time, albeit with more noise. Therefore, in order to enable automated clinical decision making using such ECG sources, it is desirable to introduce new methods that require minimal pre-processing prior to analysis and are robust to noise.

This thesis aims to develop a cardiac arrhythmia monitoring and prediction system applicable to portable arrhythmia monitoring devices. The analysis is based on a novel algorithm which does not rely on the detailed morphological information contained within each heartbeat, thus minimizing the impact of noise. Instead, the method works by analyzing the similarity and variability within strings of consecutive heartbeats, relying only on the broad morphology type of each heartbeat and utilizing the computer's ability to store and process a large number of heartbeats beyond humanly possible. The novel algorithm is based on deterministic probabilistic finite-state automata which have found great success in the field of natural language processing by studying the relation among different words in a sentence rather than the detailed structure of the individual words.

The proposed algorithm has been employed in experiments on both detection

and prediction of various cardiac arrhythmia types, and has achieved an Area under the ROC Curve (AUC) in the range of 0.70 to 0.95 for detection and prediction of different types of cardiac arrhythmias and cardiac events including atrial fibrillation, supraventricular tachycardia, ventricular arrhythmia, ventricular tachycardia, bradycardia, and rapid ventricular rates with publicly available benchmark databases, hospital bedside database and data collected from portable devices. Comparing with other well-established methods, the proposed algorithm has achieved equal or better classification results, even though in some cases the advantage might not be statistically significant. In addition, the performance of the proposed algorithm is almost identical with or without any pre-processing on the data.

The work in the thesis could be deployed as a cardiac arrhythmia monitoring and severe event prediction system which could alert patients and clinicians of an impending event, thereby enabling timely medical interventions.

CHAPTER I

Introduction

1.1 Motivation and Background

The long-term goal of this thesis is to develop a cardiac arrhythmia monitoring system that not only detects but also predicts abnormal cardiac events.

Abnormalities of cardiac rhythms are correlated with significant morbidity. Atrial fibrillation, supra-ventricular tachycardia, ventricular arrhythmia and bradycardia are all common types of cardiac arrhythmia. Atrial fibrillation is the most common sustained cardiac arrhythmia occurring in 1-2 % of the general population *Camm et al.* (2010), the percentage will likely keep growing between 2012 and 2050, when the United States will experience considerable growth in its elder population *January et al.* (2014); *Ortman et al.* (2014). Atrial fibrillation (AFib) affects at least 2.3 million people in United States and is associated with an increased risk of stroke and mortality. Supraventricular tachycardia (SVT) is a general term describing a subgroup of arrhythmias whose mechanism involves or originates above the atrioventricular node. About 90,000 cases of SVT are detected annually in the United States and about 25% of all emergency department visits for SVT result in hospitalization. Ventricular arrhythmia (VA) encompasses a spectrum of abnormal heart rhythms originating from the ventricles, the heart's lower chambers. VA causes 75% to 80% of cases of sudden cardiac death. Bradyarrhythmias and other forms of conduction

disease may cause syncope, fatigue from chronotropic incompetence, or sudden death from asystole or ventricular tachycardia. Therefore it is crucial to develop a cardiac arrhythmia monitoring system that can not only detect the abnormal cardiac events but also predict these events.

Physiological signals such as ECG have many clinical applications in cardiac arrhythmia, including diagnosis confirmation, monitoring drug effects, and rate control *Lankveld et al.* (2014). In recent years, with development of cardiac monitoring devices, more physiological signals are obtained and digitalized and available for analysis. Despite its great value, it is often challenging to extract relevant information from physiological signals, even for a well-trained medical expert, due to variability in signal morphology. Such variability is ubiquitous and manifests itself in different ways: via the ECG signals themselves, the measurements derived from such signals, and the diagnostic interpretations based on such measurements *Schijvenaars et al.* (2008).

1.2 Current Approaches and Challenges

Most traditional methods for physiological signal analysis depend heavily on preprocessing to identify specific morphology (such as R peaks in ECG). As a result, these methods tend to be less effective on noisy data. In recent years, with the advent of portable arrhythmia monitoring devices, it has become possible to collect ECG data in real time, albeit with more noise. Therefore, in order to enable automated clinical decision making using such ECG sources, it is desirable to introduce new methods that require minimal pre-processing prior to analysis.

In this thesis, we aim to develop a cardiac arrhythmia monitoring and prediction system with a novel deterministic probabilistic finite-state automata (DPFA) based approach. The method transforms physiological signals into probabilistic strings over an alphabet which contains a list of symbols. Each symbol of the alphabet corre-

sponds to a broad morphology type which is present in the physiological signal in question. Using these probabilistic strings obtained from the physiological signals via symbolization, the algorithm seeks to extract the information related to the patterns and sequences of the symbols contained in the probabilistic string. This is achieved by constructing a DPFA from the probabilistic string, after which the DPFA serves as the primary means of classification. This approach frees the resultant classification algorithm from relying on the identification of specific morphological information such as the location of R peaks in ECG. The proposed approach does not rely on extensive clinical knowledge regarding the specific morphology of the signals, and is robust to noise in the signals.

1.3 Outline of Thesis Study

This thesis aims to develop a cardiac arrhythmia monitoring and prediction system. The central component is a novel DPFA-based algorithm which is developed and tested on different types of data. The rest of this thesis is organized as follows.

Chapter II provides an extensive literature review of the background and current development and applications of arrhythmia classification and prediction algorithms, together with various types of arrhythmia that are considered in our detection and prediction experiments, using physiological signals such as ECG and accelerometer data.

Chapter III contains the detailed information on the various algorithms used in this thesis. These include the automated annotation algorithms for different types of arrhythmia and cardiac events, which are used in experiments where there is no annotations done by clinicians, the pre-processing, noise detection and event extraction algorithms involved in various auxiliary steps of the experiments, and most importantly the DPFA algorithm.

Chapter IV describes data sources that have been used for this thesis. There is

a total of six databases, each has been experimented on for different purposes of the study. Database 1 (DB1) is used in the earlier stages of the thesis for algorithm development and events detection. Database 2 (DB2) is collected in-hospital from bed-sides devices and used for arrhythmia prediction. Database 3 (DB3) has not been collected as planned and eventually combined with Database 4 (DB4). DB4 and Database 5 (DB5) are prospective databases with portable devices for hospital discharged patients and healthy controls. These two databases have been used for arrhythmia predictions. Database 6 (DB6) is a database with signals collected from portable devices on AFib patients. This dataset is also used for cardiac event prediction.

Chapter V presents the results of all the experiments conducted during the thesis study. These include detection and prediction of AFib, SVT, VA, bradycardia and rapid ventricular rates (RVR) with low activities. Discussions and limitations for the monitoring and prediction events are also presented in this chapter.

1.4 Conclusion

This thesis presents a new method for classifying and predicting cardiac arrhythmia events based on a novel DPFA algorithm. The proposed method takes a probabilistic string extracted from physiological signals as input, and constructs the underlying state space and transition probabilities of the DPFA model, directly from the input data via frequency analysis. Experiments have been performed on physiological signals collected from different sources such as public databases, hospital bed-side devices and portable devices to detect and predict different types of cardiac arrhythmia events. The proposed algorithm requires minimal pre-processing and is robust to noise. The work in the thesis could be deployed as a cardiac event prediction system which could patients and clinicians of an impending cardiac event, thereby enabling timely medical intervention.

CHAPTER II

Review of Literature

2.1 Introduction

This chapter contains a background literature review conducted in three aspects. Firstly, a review in different types of arrhythmia, including atrial fibrillation, supraventricular tachycardia, ventricular arrhythmia and bradycardia. Secondly, a review in physiological signals and ECG in particular as ECG data is the main data used in the study for experiments. Finally a review in arrhythmia classification algorithms.

2.2 Types of Arrhythmia

One of the project objectives is to construct a reliable prediction system for cardiac arrhythmia, specifically AFib, SVT, VA and bradycardia as they are the most prevalent types of arrhythmia. In the sections below we will introduce different types of arrhythmias and their significance.

2.2.1 Atrial Fibrillation

AFib is the most common sustained cardiac arrhythmia occurring in 1-2 % of the general population *Camm et al.* (2010), the percentage will likely keep growing between 2012 and 2050, when the United States will experience considerable growth

in its elder population *January et al.* (2014); *Ortman et al.* (2014). AFib is the most common sustained cardiac arrhythmia. As of 2014, between 2.7 million and 6.1 million American adults were affected by AFib *January et al.* (2014). AFib is a major cause of many life-threatening diseases and adversely impacts quality of life. It is associated with a 5-fold increased risk of stroke *Wolf et al.* (1991), a 3-fold increased risk of heart failure, and a 2-fold increased risk of both dementia and mortality *January et al.* (2014); *Wang et al.* (2003); *Kannel et al.* (1998); *Camm et al.* (2010).

2.2.2 Supraventricular Tachycardia

SVT is a general term describing a subgroup of arrhythmias whose mechanism involves or originates above the atrioventricular node. The incidence of SVT is approximately 35 cases per 100,000 patients with a prevalence of 2.25 cases per 1,000 in the general population *Sohinki and Obel* (2014). SVT has different arrhythmia related symptoms ranging from nonexistent to severe. Symptoms include palpitation, fatigue, light-headedness, chest discomfort and dyspnea, while on the other extreme certain paroxysmal SVT could also be asymptomatic.

2.2.3 Ventricular Arrhythmia

VA encompasses a spectrum of abnormal heart rhythms originating from the ventricles, the heart's lower chambers. These types of arrhythmia have rates of over 100 beats per minute *Al-Khatib et al.* (2018). Types of VA include ventricular tachycardia (VT), ventricular flutter (VFlutter), and ventricular fibrillation (VF). Serious ventricular arrhythmia is associated with ischemic heart disease and can contribute to sudden cardiac death (SCD) events. These events constitute approximately 230,000 to 350,000 deaths annually in the United States and 50% of all cardiovascular deaths *Myerburg* (2001); *Myerburg and Junttila* (2012). Approximately half of SCD events can be attributed to VT or VF *John et al.* (2012). Therefore, monitoring and detect-

ing VT and VF is critical for the prevention of SCD events.

2.2.4 Bradycardia

Among various types of arrhythmia, tachycardia refers to a fast heart beat generally with a resting heart rate greater than 100 beats-per-minute (BPM), whereas bradycardia, on the other hand, refers to a slow heart beat commonly defined as having a heart rate of less than 60 BPM *Kusumoto et al.* (2019). Despite the general definition, a slow heart rate does not always indicate an underlying disease, since there is considerable variation in the resting heart rate among the healthy, asymptomatic population *Mangrum and DiMarco* (2000) such as the example of people with a family history of slow heart beats, or athletes and physically active individuals. However, bradycardia can cause inadequate blood flow to tissues and organs leading to symptoms such as fatigue, dizziness or heart failure *Mangrum and DiMarco* (2000).

2.3 Physiological Signals

Physiological signals reflect the functionality of various physiological systems. A trained medical expert can retrieve valuable information about the structure of the organ by examining the physiological signals generated by its electrical conduction system. This knowledge has far reaching impact in decision making in diagnosis, treatment monitoring, drug efficacy tests and quality of life *Faust et al.* (2018).

The ECG is a type of physiological signal with many clinical applications in cardiac arrhythmia, including diagnosis confirmation, monitoring drug effects, and rate control *Lankveld et al.* (2014). Despite its great value, it is often challenging to extract relevant information from ECG signals, even for a well-trained medical expert, due to variability in signal morphology. Such variability is ubiquitous and manifests itself in different ways: via the ECG signals themselves, the measurements derived

from the signals, and the diagnostic interpretations based on such measurements.

2.4 Arrhythmia Classification Algorithms

Arrhythmia classification using ECG signals typically involves three major tasks: signal pre-processing in order to remove noise and baseline wandering, feature extraction in time and/or frequency domains, and classification of the signals into different arrhythmia types according to the features extracted, often through the training of classifiers via machine learning algorithms.

In the literature, numerous feature extraction techniques have been proposed to analyze and classify ECG signals. These techniques include: heart rate variability (HRV) *Asl et al.* (2008); *Mei et al.* (2018), independent component analysis (ICA) *Yu and Chou* (2009); *Jiang et al.* (2006); *Naseri et al.* (2019), principal component analysis (PCA) *Ceylan and Özbay* (2007); *Martis et al.* (2012), wavelet transform *Mahmoodabadi et al.* (2005); *Zhao and Zhang* (2005); *Ashtiyani et al.* (2018), convolutional neural network (CNN) *Acharya et al.* (2017a); *Kiranyaz et al.* (2016). These feature extraction techniques are used together with learning algorithms for arrhythmia detection and classification. Each feature extraction technique focuses on different aspects of the signal, e.g., HRV features put emphasis on the features of heart beats, while ICA, PCA, and wavelet transform focus on the morphology of the signal. All of these methods have their unique strengths and drawbacks.

The machine learning algorithms used for arrhythmia classification include: support vector machine (SVM) *Ye et al.* (2010); *Nasiri et al.* (2009); *Song et al.* (2005); *Ozcan and Gurgen* (2010); *Bazi et al.* (2013), auto-regressive modeling *Ge et al.* (2002), hidden Markov model (HMM) *Coast et al.* (1990); *Andreao et al.* (2006); *Gomes et al.* (2010), a set of rules as determined by cardiologists *Tsipouras et al.* (2002); *Exarchos et al.* (2005), optimal path forest *Luz et al.* (2013), and artificial neural networks *Ceylan and Özbay* (2007); *Jadhav et al.* (2011); *Özbay et al.* (2006);

Elhaj et al. (2016). In the past two years, many studies have adopted recurrent neural network (RNN) or long short-term memory (LSTM) as the learning algorithms of choice *Oh et al.* (2018); *Chauhan and Vig* (2015); *Yildirim* (2018); *Zhang et al.* (2017); *Warrick and Homsi* (2017). A number of studies have also applied deep learning methods for the classification of ECG signals *Acharya et al.* (2017a); *Rajpurkar et al.* (2017); *Acharya et al.* (2017b). In general these algorithms do not require specific signal pre-processing, QRS detection, or feature extraction steps. However, this advantage comes with the cost of requiring larger training datasets. This can be challenging as arrhythmia cases are relatively rare in comparison to healthy control cases.

Contrary to the above, to the best of our knowledge few studies have applied DPFA-based methods in the analysis of physiological signals, and especially not in arrhythmia detection and prediction using ECG signals. DPFA are finite-state automata in which every state is assigned a probability to each distinct letter in a fixed finite alphabet, which then determines a unique new state. More generally, in probabilistic finite-state automata (PFA), given a letter and current state, multiple transitions are allowed to other states with potentially different probabilities *Vidal et al.* (2005). PFA have the same expressive power as HMM *Vidal et al.* (2005), despite the fact that only the latter has found wide application in ECG signal analysis.

Due to its deterministic assignment of transitions to the alphabet for a given state, DPFA have less expressive power than PFA. On the other hand, it is much easier to estimate the parameter values of DPFA than PFA *Vidal et al.* (2005). As a result, DPFA will likely perform better when running the algorithm in real-time. Some well-developed special cases of DPFA include: n-grams with smoothing *Ney et al.* (1997), Markov chains built by aggregating or mixing n-grams *Saul and Pereira* (1997), and probabilistic suffix trees (PST) *Esposito et al.* (2002). These algorithms have found applications in natural language processing *Brown et al.* (1992); *Hakki-*

nen and Tian (2001) as well as in biomedical research such as protein structural analysis and sequence analysis *Dorohonceanu and Nevill-Manning* (2000); *Oğul and Mumcuoğlu* (2006).

2.4.1 Arrhythmia Prediction

As detailed above, many algorithms and methods exist for the detection or classification of cardiac arrhythmias. A more challenging application of machine learning in this context is the prediction of cardiac events or arrhythmias before they occur. A reliable cardiac event prediction system could alert patients and clinicians alike of an impending event, thereby enabling timely intervention. However, prediction of arrhythmic events well before their onset is still an open research problem *Rizwan et al.* (2020).

CHAPTER III

Methodology

3.1 Introduction

This chapter contains a detailed description of the methodologies. There are four major sections in this chapter. In the first section, the DPFA algorithm is explained in detail including training method, symbolization, DPFA generation and classification method. In the second section, the annotation algorithms for different types of arrhythmia and cardiac events are described. In the third section, the pre-processing steps and prediction event extractions steps are illustrated. In the last section, several existing methods are listed and explained for comparison.

3.2 Deterministic Probabilistic Finite-State Automata

3.2.1 Training Method

A schematic diagram of the proposed method is illustrated in Figure 3.1. The training dataset consists of annotated ECG signals. From these the algorithm extracts windows that are indicative of future arrhythmia events and others that are not. The positive ECG signals (i.e., pre-arrhythmia) are symbolized into probabilistic strings, which are then fed into the DPFA Generation module that constructs the positive DPFA M_+. The same is done for the negative ECG windows (i.e., regions that are

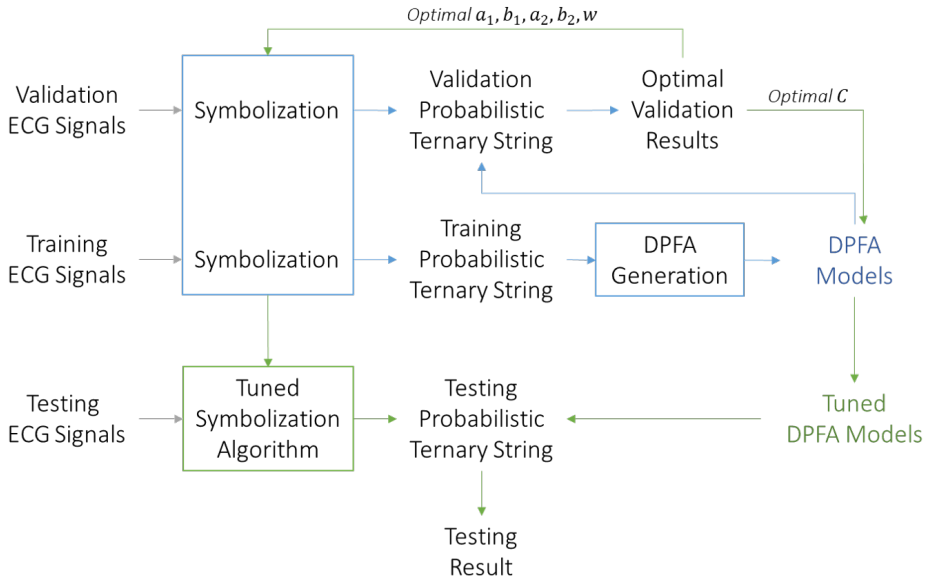


Figure 3.1: DPFA training, A schematic diagram depicting the training and testing steps for the proposed method.

neither arrhythmia nor pre-arrhythmia), producing the negative DPFA M_{-}. The Symbolization module is detailed in Section 3.2.2 and the DPFA Generation module is described in Section 3.2.3.

3.2.2 Symbolization

In this study, ECG signals are symbolized into strings of ternary words, but the method can be extended to larger alphabets. Let $g(t)$ be the input ECG signal, where t is time in seconds, $0 \leq t \leq L$, and L is the length of the ECG recording. $g(t)$ can be thought of as a continuous signal, but in implementation it is generally discrete with a high frequency (typically $\geq$ 200 Hz) depending on the ECG source. The symbolization process for a sample of ECG signal with normal rhythm is illustrated in Figure 3.2 and a sample of ECG signal with VA is illustrated in Figure 3.3.

A brief description of the main steps of the symbolization process is below, with pseudocode provided in Algorithm 1:

(a) The first step is to capture morphological features of width less than the dura-

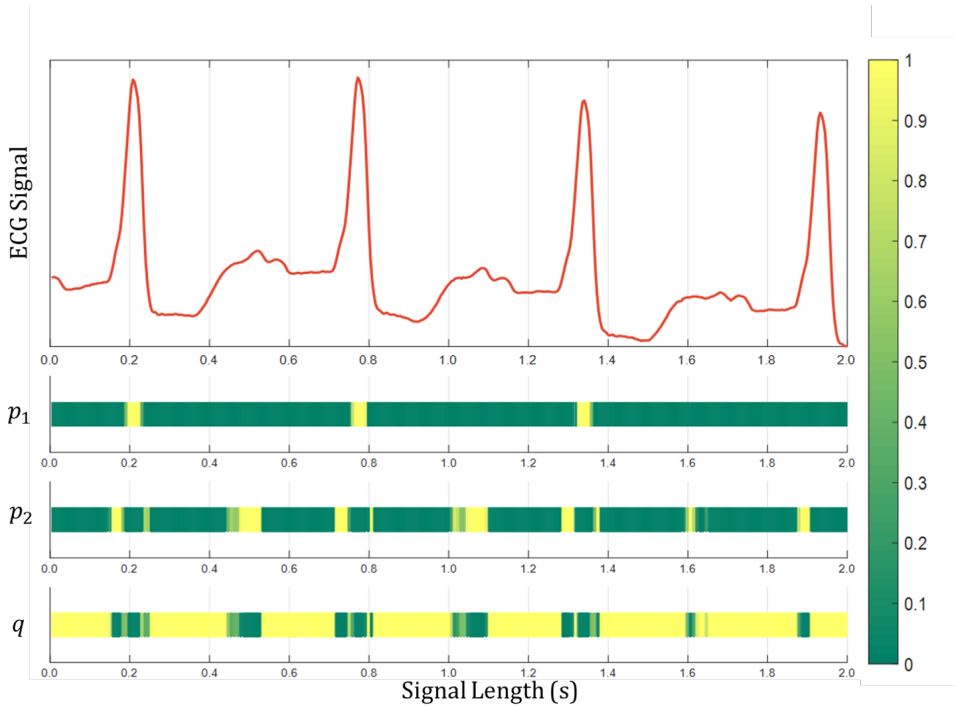


Figure 3.2: DPFA symbolization, an example of ECG signal symbolization into a ternary alphabet: $\{p_1,p_2,q\}$.

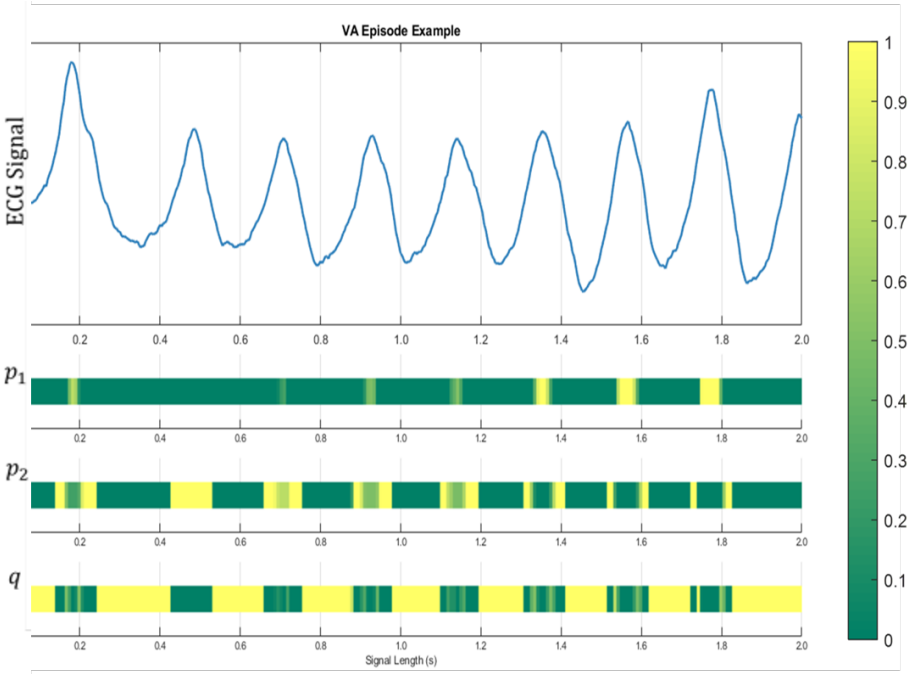


Figure 3.3: DPFA symbolization, an example of VA ECG signal symbolization into a ternary alphabet: $\{p_1,p_2,q\}$.

tion of a typical QRS-complex: this is achieved by first subtracting the moving average $\frac{1}{2h_0}\int_{t-h_0}^{t+h_0} g(r)dr$ over intervals of width $2h_0$. The signal is then smoothed

by convolving with the triangle function $\Lambda_{h_0}(t) = \max\{h_0 - |t|, 0\}$ and normalized by dividing by the local L^2 norm

$$\left| \int_{t-h_1}^{t+h_1} g^2 * \Lambda_{h_1}(r) dr \right|^{\frac{1}{2}}$$

over intervals of width $2h_1$.

The parameter h_0 controls the window size for baseline removal, while h_1 controls the length of the filter whose purpose is to capture the variability of magnitudes among adjacent heartbeats, while ensuring comparable average magnitudes over longer intervals. Therefore h_0 should be approximately the duration of a heart beat. Since a normal heart rate is around 60-100 beats per minute, so h_0 was chosen to be 2 seconds. h_1 needs to be longer than the duration of two or three heartbeats, but not too large as to reduce the local nature of normalization, so h_1 was chosen to be 5 seconds. Both parameters can also be tuned during the training steps for the purpose of facilitating the capture of local information for use by the DPFA algorithm. For different applications these parameters may need to be modified.

Finally a piecewise linear filter is applied by forming the convolution

$$g * \phi(t) = g^{(-2)} * \phi^{(2)}(t),$$

where $g^{(-2)}$ denotes the second antiderivative of the signal $g(t)$ and $\phi^{(2)}(t)$ denotes the second derivative of the piecewise linear function $\phi(t)$. The expression on the right hand side has the advantage that $\phi^{(2)}$ is a sum of delta functions, which is better behaved when working with discrete signals.

(b) The second step is to apply downsampling to discretize the signal into $x_1 x_2 \dots x_n$ by

$$x_i = \max\Big\{ g(t) \;\Big|\; (i-1)w \le t \le iw \Big\},$$

where w is the downsampling window size tuned during the training step and $n = \lfloor L/w \rfloor$.

(c) The next step is to locally normalize the discrete signal by dividing x_i by the local maximum $\max\{x_{i-\lfloor h_2/w \rfloor}, \dots, x_{i+\lfloor h_2/w \rfloor}\}$ over intervals of width $2h_2$, where $h_2 = 1$ second. h_2 was chosen to be long enough to cover several QRS complexes.

(d) In the final step, the probabilities p_{i1}, p_{i2} and q_i are obtained from the normalized signal $\tilde{x}_1 \tilde{x}_2 \dots \tilde{x}_n$ via soft-thresholding

$$\begin{cases} p_{i1} &= \psi_1(\tilde{x}_i) \\ p_{i2} &= \big(1 - \psi_1(\tilde{x}_i)\big) \cdot \psi_2(\tilde{x}_i) \\ q_i &= 1 - p_{i1} - p_{i2}, \end{cases}$$

where the soft-thresholding functions are chosen to be piecewise linear functions $\psi_1, \psi_2 : [0,1] \to [0,1]$ of the form

$$\psi_j(x) = \begin{cases} 1 & \text{if } x > b_j \\ \frac{x - a_j}{b_j - a_j} & \text{if } a_j \le x \le b_j \\ 0 & \text{if } x < a_j \end{cases}$$

The parameters a_1, b_1, a_2, b_2 are all different and tuned within the training step for different studies. A grid search was performed for a_1 and b_1 with $a_1 < b_1$ from 0.1 through 0.9 with 0.1 increment. A grid search for a_2 and b_2 was performed in a similar manner. The parameter values used in this study were selected tuned with training and validation data.

In summary, p_{i1}, p_{i2}, and q_i represent the probability of dominant patterns and local shapes depending on values of a_1, b_1, a_2, and b_2.

Algorithm 1 ECG Symbolization

Require: $g(t)$, $0 \le t \le L$ ▷ ECG
$h_0, h_1, h_2 > 0$ ▷ time parameters
$\phi(t)$, supported locally near $t = 0$ ▷ piecewise linear filter
$w > 0$ ▷ downsampling window size
$\psi_1, \psi_2 : [0,1] \to [0,1]$ ▷ soft-thresholding functions

Ensure: $\mathbf{P} = \mathbf{p}_1\mathbf{p}_2 \dots \mathbf{p}_n = \begin{bmatrix} p_{11} \\ p_{12} \\ q_1 \end{bmatrix}\begin{bmatrix} p_{21} \\ p_{22} \\ q_2 \end{bmatrix} \cdots \begin{bmatrix} p_{n1} \\ p_{n2} \\ q_n \end{bmatrix}$ ▷ probabilistic string

for all $t \in [0, L]$ **do**
$g(t) \leftarrow g(t) - \frac{1}{2h_0}\int_{t-h_0}^{t+h_0} g(r)dr$ ▷ subtract moving average
$g(t) \leftarrow g * \Lambda_{h_0}(t)$ ▷ triangular smoothing
$g(t) \leftarrow |\int_{t-h_1}^{t+h_1} g^2 * \Lambda_{h_1}(r)dr|^{-\frac{1}{2}} g(t)$ ▷ normalization
$g(t) \leftarrow g^{(-2)} * \phi^{(2)}(t)$ ▷ piecewise linear filter
end for
$n \leftarrow \lfloor L/w \rfloor$ ▷ length of $\mathbf{P}$
for $i \leftarrow 1, n$ **do**
$x_i \leftarrow \max\{g(t) \mid (i-1)w \le t \le iw\}$ ▷ downsampling
end for
for $i \leftarrow 1, n$ **do**
$\tilde{x}_i \leftarrow x_i / \max\{x_j \mid i - \lfloor h_2/w \rfloor \le j \le i + \lfloor h_2/w \rfloor\} \in [0,1]$
▷ local normalization
$p_{i1} \leftarrow \psi_1(\tilde{x}_i)$
$p_{i2} \leftarrow (1 - \psi_1(\tilde{x}_i))\psi_2(\tilde{x}_i)$
$q_i \leftarrow 1 - p_{i1} - p_{i2}$ ▷ soft-thresholding
end for

3.2.3 DPFA Generation

The output of the symbolization module is a probabilistic string over the alphabet Σ. Such a probabilistic string defines a word distribution, for which the DPFA generation module produces a DPFA that best approximates this distribution. The DPFA is constructed in two steps: first by building a frequency prefix tree (FPT) and then performing state merging within the FPT.

3.2.3.1 The Word Distribution Defined by a Probabilistic String

Given the alphabet $\Sigma = \{a_1, \ldots, a_d\}$ consisting of d symbols, let Σ^* denote the set of all words $a_{j_1}a_{j_2}\ldots a_{j_m}$ of finite length, with ε representing the empty word. A word distribution is defined to be a function $f : \Sigma^* \to \mathbb{R}$ such that

(a) $f(w) \geq 0$, and

(b) $f(w) = \sum_{j=1}^{d} f(wa_j)$

for all $w \in \Sigma^*$.

Let $\mathbf{P} = \mathbf{p}_1\mathbf{p}_2\ldots\mathbf{p}_n$ be a probabilistic string of length n consisting of d-dimensional probability vectors

$$\mathbf{p}_i = \begin{bmatrix} p_{i1} \\ \vdots \\ p_{id} \end{bmatrix}, \quad p_{ij} \geq 0 \text{ and } \sum_{j=1}^{d} p_{ij} = 1;$$

where p_{ij} denotes the probability that the ith letter of the probabilistic string is equal to $a_j \in \Sigma$. Then the probabilistic string defines a word distribution $f_{\mathbf{P}}$ by

$$f_{\mathbf{P}}(a_{j_1}a_{j_2}\ldots a_{j_m}) = \sum_{i=1}^{n-m+1} \prod_{k=i}^{i+m-1} p_{ij_k} \tag{3.1}$$

for all non-empty $a_{j_1}a_{j_2}\ldots a_{j_m} \in \Sigma^*$ and $f_{\mathbf{P}}(\varepsilon) = n$ (see Algorithm 2).

3.2.3.2 Constructing the FPT

A frequency prefix tree (FPT) is a tree-like automaton $T = \langle Q_0, \Sigma, \varepsilon, \text{Freq}\rangle$ with alphabet Σ, state space $Q_0 \subset \Sigma^*$, and initial state $\varepsilon \in Q_0$, which is also equipped with a frequency function $\text{Freq} : Q_0 \to \mathbb{R}$. Here the transition function is given by concatenation of words, which assigns $qa_j \in \Sigma^*$ to $(q, a_j) \in Q_0 \times \Sigma$ whenever $qa_j \in Q_0$.

Algorithm 2 Word Distribution of a Probabilistic String

Require: $\mathbf{P} = \mathbf{p}_1\mathbf{p}_2\ldots\mathbf{p}_n$ ▷ probabilistic string
Ensure: $f_{\mathbf{P}}(w) \in \mathbb{R}$ ▷ $w = a_{j_1}a_{j_2}\ldots a_{j_m} \in \Sigma^*$

```
w ← ε
p_{n+1}(w) ← 0
for i ← 1 to n do
    p_i(w) ← 1
end for
for k ← 0 to m − 1 do
    w ← a_{j_{m−k}} w
    p_i(w) ← p_{i j_{m−k}} p_{i+1}(w)
end for                          ▷ p_i(w) = ℙ(w occurs at the ith position of P)
f_P(w) ← Σ_{i=1}^{n} p_i(w)
```

In particular, a word distribution $f : \Sigma^* \to \mathbb{R}$ on the alphabet Σ defines an FPT with full state space $Q_0 = \Sigma^*$ and frequency function $\text{Freq} = f$. In general, such FPTs have infinite state spaces. However, in the special case when $\text{Freq}(w) = 0$ for all but finitely many $w \in \Sigma^*$, the FPT can be restricted to the finite state space

$$Q_0 = \{w \in \Sigma^* \mid \text{Freq}(w) \neq 0\}$$

without losing any information.

One such special case is when the word distribution f arises from a probabilistic string $\mathbf{P}$ over the alphabet Σ with finite length. This is because $f_{\mathbf{P}}(w) = 0$ for all words $w \in \Sigma^*$ whose length exceeds the length of $\mathbf{P}$.

3.2.3.3 Cutoff and State Merging

A deterministic probabilistic finite-state automaton (DPFA) is an automaton $M = \langle Q, \Sigma, \varepsilon, P, T\rangle$ with state space Q, alphabet Σ, initial state $\varepsilon \in Q$ and transition function $T : Q \times \Sigma \to Q$, which is also equipped with a probability function $P : Q \times \Sigma \to \mathbb{R}$ such that

(a) $P(q, a_j) \geq 0$, and

(b) $\sum_{j=1}^{d} P(q, a_j) = 1$

for all $q \in Q$.

Given an FPT $T = \langle Q_0, \Sigma, \varepsilon, \text{Freq} \rangle$, one can construct a DPFA $M = \langle Q, \Sigma, \varepsilon, P, T \rangle$ by the largest suffix merging (LSM) algorithm (Algorithm 3); a brief description of which is provided below:

(a) the algorithm requires a cutoff parameter C such that $0 < C < \text{Freq}(\varepsilon)$;

(b) the state space Q of the DPFA consists of all $q \in Q_0$ which have frequency $\text{Freq}(q) > C$;

(c) the alphabet Σ and initial state $\varepsilon \in Q$ are the same as the FPT;

(d) the probability function is defined to be

$$P(q, a_j) = \frac{\text{Freq}(qa_j)}{\text{Freq}(q)}$$

for all $q \in Q$, which is well-defined since $\text{Freq}(q) > C > 0$;

(e) if $\text{Freq}(qa_j) > C$, then the transition function is defined to be $T(q, a_j) = qa_j \in Q$;

(f) otherwise $T(q, a_j) = s \in Q$ is defined to be the largest suffix s of qa_j with $\text{Freq}(s) > C$.

In summary, the LSM algorithm selects those states $q \in Q_0$ of T with sufficiently high frequency to be in the state space M, and then defines the transition state $T(q, a_j)$ to be the largest suffix of qa_j that is itself contained in the state space of M.

Figure 3.4 provides an example of an FPT, while Figure 3.5 depicts the DPFA constructed by the LSM algorithm. For example, in Figure 3.5 the transition state $T(a, aba)$ is the largest suffix of $abaa$ contained in the state space. Since $abaa$ and aba are not in the state space, $T(a, aba) = aa$.

The mechanism for state merging in the LSM algorithm reflects the time-dependent nature of the underlying data (i.e., ECG signals). In particular, the transition state $T(q, a_j)$ is found by repeatedly deleting the left-most letter from the string qa_j, which represents the process of forgetting information from the most distant past, in order to update the system to the current state.

Algorithm 3 Largest Suffix Merging

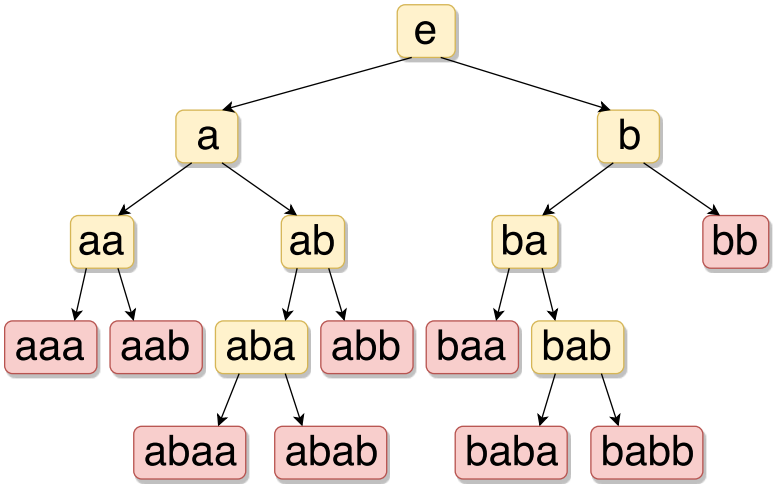
Require: $T = \langle Q_0, \Sigma, \varepsilon, \text{Freq} \rangle$ ▷ FPT
$0 < C < \text{Freq}(\varepsilon)$ ▷ cutoff parameter
Ensure: $M = \langle Q, \Sigma, \varepsilon, P, T \rangle$ ▷ DPFA
$Q \leftarrow \{\varepsilon\}$ ▷ initial state
for all $q \in Q_0$ **do**
while $\text{Freq}(q) > C$ **do**
for $j \leftarrow 1, d$ **do**
$P(q, a_j) \leftarrow \dfrac{\text{Freq}(qa_j)}{\text{Freq}(q)}$ ▷ probability function
$T(q, a_j) \leftarrow qa_j$
while $\text{Freq}\big(T(q, a_j)\big) \leq C$ **do**
$T(q, a_j) \leftarrow$ largest suffix of $T(q, a_j)$
end while ▷ transition function
$Q \leftarrow Q \cup \{T(q, a_j)\}$ ▷ state space
end for
end while
end for

3.2.3.4 Implementation

When training the DPFA $M_{\mathbf{P}}$ from the input probabilistic string $\mathbf{P}$, there is no need to construct the states of the FPT $T_{\mathbf{P}}$ with frequency less than C, and the LSM algorithm is executed to merge the states of $T_{\mathbf{P}}$ immediately after they are constructed.

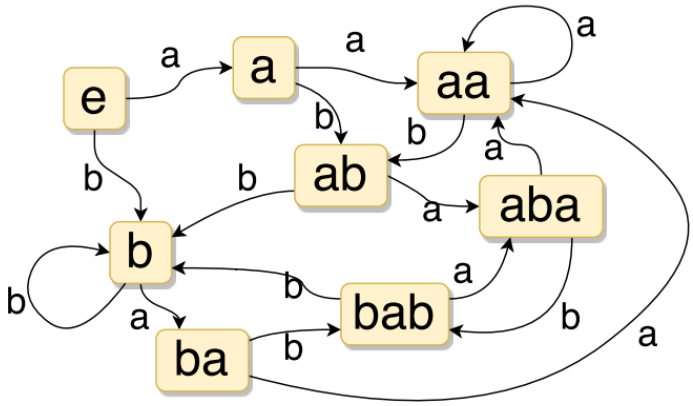
The combined algorithm has computational complexity $O(dns)$ and needs total memory $O(nl)$, where d is the size of the alphabet Σ, n is the length of the probabilistic string $\mathbf{P}$, s is the size of the state space Q, and l is the length of the longest word in the state space Q. Both s and l depend on the cutoff parameter C.

Figure 3.4: DPFA generation, an example of FPT



FPT - The DPFA generation module searches through the tree. The yellow words have frequency at least C, while the red words have frequency less than C. This example will generate a DPFA with 8 states.

Figure 3.5: DPFA generation, an example of transition states



The DPFA generated from the previous figure. The transition state $T(a, aba)$ is the largest suffix of $abaa$ contained in the state space. Since $abaa$ and aba are not in the state space, $T(a, aba) = aa$.

3.2.4 Classification Method

The classification scheme contains the training phase and the testing phase. By the end of the training phase, the algorithm has already learned the DPFA M_+ and M_- for the "positive ECG signals" and "negative ECG signals" classes respectively. Then to classify a given ECG signal $g(t)$ from the testing dataset, we first symbolize

the signal into a probabilistic string $\mathbf{P}_g$, and then compare the goodness-of-fit between $\mathbf{P}_g$ and M_+ versus $\mathbf{P}_g$ and M_-. The goodness-of-fit is measured using the expected likelihood between word distributions.

3.2.4.1 The Word Distribution Defined by a DPFA

Let $M = \langle Q, \Sigma, \varepsilon, P, T \rangle$ be a DPFA, then M defines a word distribution $f_M : \Sigma^* \to \mathbb{R}$ on the alphabet Σ by

$$f_M(a_{j_1} a_{j_2} \dots a_{j_m}) = \prod_{k=1}^{m} P(q_{k-1}, a_{j_k}) \tag{3.2}$$

for all non-empty $a_{j_1} a_{j_2} \dots a_{j_m} \in \Sigma^*$ and $f_M(\varepsilon) = 1$, where the state $q_k \in Q$ is defined recursively by $q_0 = \varepsilon$ and $q_k = T(q_{k-1}, a_{j_k})$ (see Algorithm 4).

Algorithm 4 Word Distribution of a DPFA

Require: $M = \langle Q, \Sigma, \varepsilon, P, T \rangle$ ▷ DPFA
Ensure: $f_M(w) \in \mathbb{R}$ ▷ $w = a_{j_1} a_{j_2} \dots a_{j_m} \in \Sigma^*$

```
w ← ε
q ← ε
f_M(w) ← 1
for k ← 1, m do
    f_M(w) ← f_M(w) P(q, a_{j_k})
    w ← w a_{j_k}
    q ← T(q, a_{j_k})
end for
```

3.2.4.2 Compute Expected Likelihood

The expected likelihood $\mathrm{EL}(f_{\mathbf{P}}, f_M)$ is defined as the expected value $\mathbb{E}_{f_{\mathbf{P}}}[\mathcal{L}(f_M)]$ of the likelihood function $\mathcal{L}(f_M|w)$ over all words $w \in \Sigma^*$:

$$\mathrm{EL}(f_{\mathbf{P}}, f_M) = \sum_{w \in \Sigma^*, f_{\mathbf{P}}(w) \neq 0} \mathcal{L}(f_M \mid w) f_{\mathbf{P}}(w) = \sum_{w \in \Sigma^*, f_{\mathbf{P}}(w) \neq 0} f_M(w) f_{\mathbf{P}}(w), \tag{3.3}$$

which is computed with respect to the word distribution $f_{\mathbf{P}}(w)$.

The summation in (3.3) is well-defined as there are only finitely many words $w \in \Sigma^*$ with $f_{\mathbf{P}}(w) \neq 0$ since $\mathbf{P}$ is a probabilistic string of finite length. As a measure of goodness-of-fit, a greater value of expected likelihood implies a better fit between the word distributions.

3.2.4.3 Classification

If $\mathrm{EL}(f_{\mathbf{P}_g}, f_{M_+}) > \mathrm{EL}(f_{\mathbf{P}_g}, f_{M_-})$, then the ECG signal $g(t)$ is classified as "positive ECG signal", otherwise it is classified as "negative ECG signal".

3.2.4.4 Implementation

The likelihood values $\mathcal{L}(f_{M_\pm}|w)$ can be calculated simultaneously with the word distribution $f_{\mathbf{P}_g}(w)$ as the algorithm searches through all the words $w \in \Sigma^*$ until $f_{\mathbf{P}_g}(w) = 0$. Once the values $\mathcal{L}(f_{M_\pm}|w)$ and $f_{\mathbf{P}_g}(w)$ are found for all relevant $w \in \Sigma^*$, one then applies formula (3.3) to compute the expected likelihoods $\mathrm{EL}(f_{\mathbf{P}_g}, f_{M_\pm})$.

The values $\mathrm{EL}(f_{\mathbf{P}_g}, f_{M_\pm})$ can be extremely small, which can lead to rounding errors. Thus, $\log \mathrm{EL}(f_{\mathbf{P}_g}, f_{M_\pm})$ is used instead of $\mathrm{EL}(f_{\mathbf{P}_g}, f_{M_\pm})$.

3.3 Arrhythmia and Event Annotation

Annotations of different types of arrhythmia are needed for using as labels in detection and prediction experiments. Databases in DB1 has their own annotations, most of them are annotated by clinicians. DB2 which is the retrospective database collected in UMHS, does not have any annotations. We have designed an automated annotation for different types of arrhythmia to encompass this difficulty. Continuous ECG signals recorded with Bodyguardian for DB4 and DB5 are annotated by Preventice's own BeatLogic platform which is a commercially available FDA-cleared system. The BeatLogic platform consists of 2 deep-learning models, BeatNet and RhythmNet, the results of which are consolidated using rules-based logic to produce

a single contiguous annotation file. We used Preventice's annotation in DB4 and DB5 labels. The Preventice annotation included three types of AFib, AFib normal (heart rate (HR) 60-100 BPM), AFib slow (HR $<$ 60 BPM) and AFib rapid (HR $>$ 100 BPM). All three types of AFibs for our analysis in DB4 and DB5 have been annotated with Preventice with $>$ 0.75 confidence. Please note that the AFib slow and AFib normal might be less severe forms of AFib which were not included according to our own annotations in DB2.

In the sections below, automated annotations are used and described below in sections below for different types of arrhythmias used for labeling data in DB2.

3.3.1 AFib Annotation

To label the unannotated signals, annotation algorithms have been developed based on the combination of heart rate and duration. Signals with extremely high heart rate that last for a short period of time and signals with relatively lower heart rate that last for a longer period of time are both considered to be indication for symptomatic AFib. Using a single heart rate and duration might not capture all of the symptomatic AFib events. Here we propose a heart rate-duration criteria region in terms of a linear inequality that involves the values of heart rate recorded during different duration periods.

3.3.1.1 AFib Annotation Algorithm

The annotation algorithm consists of 6 major parts: pre-processing of signal, R peak detection, noise detection, heart rate and duration calculation, grid search on combinations of heart rate and duration, and labelling signals according to the heart rate-duration criteria:

1. During the signal pre-process step, we used a double median filter on the raw ECG signals to remove baseline wandering and applied a Butterworth band

pass filter to remove noises.

2. For R peak detection, we applied the Pan Tompkins algorithm for peak detection *Pan and Tompkins* (1985).

3. Noise detection is performed by using 3 criteria. First, the percentage of missing signal in a defined time window (set to 300s) is calculated, with the window being classified as noise if the missing signal percentage is above a certain threshold (set to 15%). Secondly, if the signal passes the first criteria then we will check for missing peaks by total variation analysis. Thirdly, if the signal passes both the first and the second criteria, the third criteria will check if the percentage of missing R peaks is above a certain threshold (set to 15%) in the current window. The window of signal will be classified as noisy if it fails any of the three criteria. Noisy signals will not be used for subsequent annotation.

4. Heart rate is calculated based on R peaks in a certain time interval. Duration is calculated by counting the number of consecutive intervals that a particular heart rate spans. The time interval is set to be 30 seconds and count up to 6 intervals, which gives us a range from 30 to 180 seconds.

5. A grid search on different combinations of heart rate and duration is performed, for example using heart rates from 100 to 200 with an increment of 10 beats at each step and duration from 30 to 180 with an increment of 30 seconds will give us 66 different combinations.

6. Portions of signals with extreme high heart rate y_2 that lasts for a short time period x_1 and a relative lower heart rate y_1 that lasts for a longer time period x_2 are both considered to be indication for symptomatic AFib. Using the straight line passing through the points (x_1,y_2), (x_2,y_1), we check the heart rates between durations x_1,x_2, if any of the heart rates on the duration grid is above the heart

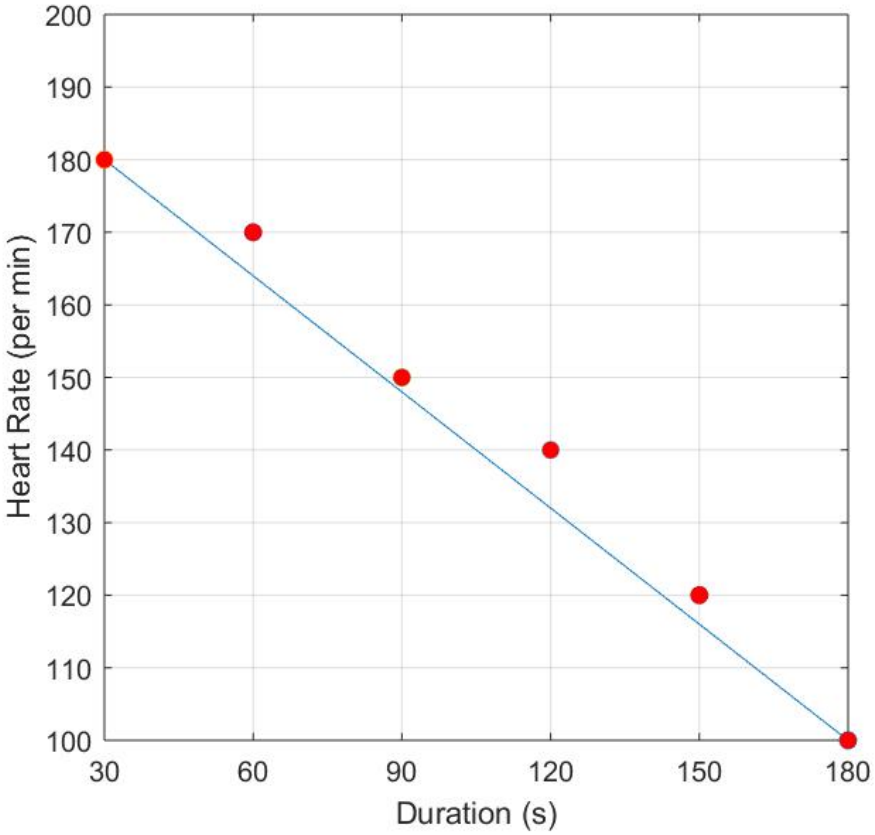


Figure 3.6: AFib annotation, an example of Heart Rate/Duration Criteria line goes through (30,180) and (180,100).

rate-duration criteria line, that point of the signal is classified as symptomatic. Figure 3.6 shows an example of Heart Rate/Duration Criteria line goes through (30, 180) and (180, 100). If the heart rate/duration is satisfied by any of the six red points in the grid search, it is classified as AFib.

This annotation method aggregates criteria found within the clinical literature for severe cardiac events, and enables us to determine the cardiac events wherein a patient will be considered symptomatic. However, it may not capture other cardiac events that, though less severe, may still be predictive of severe events to come. As such, for each type of cardiac event we wish to predict, clinician in our research team has helped in both verifying the symptomatic criteria and identifying additional portions of the ECGs that are indicative of the underlying event, but do not fall within the

above criteria. This will enable the research team to annotate the relevant portions of the collected waveforms, as will ensure that the predictions made by our algorithm are correlated with true events.

3.3.1.2 Self-Reported Symptom Analysis

To better understand the relationship between symptoms and it's effect on heart rates, we have performed an analysis with patient's self reported symptom data. Patients were monitored at home and asked to report the severity of multiple symptoms. Five measurements are included: shortness of breath, fatigue level, palpitations, chest pain and dizziness. We calculated a summary score z as the sum of these five measurements and normalized the score for each individual: $z_i = \frac{x_i - min(x)}{max(x) - min(x)}$. We then extracted the maximum heart rate that occurred within 5 minutes before the time that the symptom was reported and calculated the duration of that maximum heart rate period. The objective of this experiment is to use maximum heart rate and its duration to classify signals with high versus low symptom score z. The high symptom score group consists of signals with z above the median of the normalized summary score and low symptom score is defined as below the median of the normalized summary score. However, we did not find a clear correlation between symptomatic scores and heart rate and duration. This may be due the unreliability of the self-reported data, as different participants grade levels of discomfort and pain differently.

3.3.1.3 Literature Review

An extensive literature review has been conducted to identify severe AFib episodes. AFib is defined by *Calkins et al.* (2012) to be a cardiac arrhythmia with three characteristics: 'absolutely' irregular RR intervals, no distinct P waves, and variable atrial cycles of length $< 200ms$ ($> 300bpm$).

An episode is considered significant when it lasts for at least 30 seconds *January*

Table 3.1: The relationship between durations of AFib episodes and cardiovascular health.

Duration of an AFib episodes	**Notes**
30 seconds	Definition of AFib lasts for at least 30 seconds ECG at least 30 seconds for accepted estimation of HR.
3 minutes	Considered as severe AFib episodes by cardiologist in expert interview.
5 to 6 minutes	Associated with increased risk of stroke.
5 minutes to >24 hours	Associated with increased risk of stroke.
prolonged episodes of AFib	Higher AFib density given same AFib burden

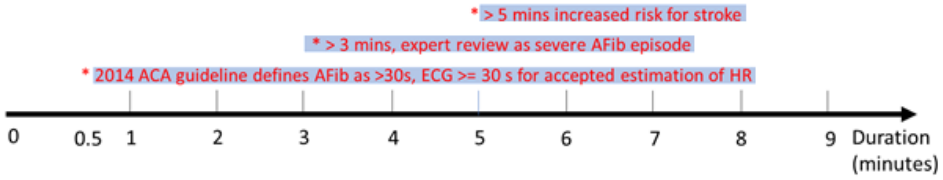


Figure 3.7: AFib annotation duration, a graphical depiction of the relationship between AFib duration and cardiovascular health.

et al. (2014), thus 30 seconds can be used as the lower limit for severe AFib in terms of duration. AFib duration anywhere from 5 minutes to > 24 hours is associated with increased stroke risk *Passman and Bernstein* (2016). Data from implantable electronic cardiac devices also suggest that even relatively short AFib episodes (5-6 minutes) are associated with an increased risk of stroke *Passman and Bernstein* (2016). Given the same AFib burden, a patient with a small number of prolonged episodes of AFib has a higher AFib density than a patient with many brief episodes of AFib *Passman and Bernstein* (2016). Table 3.1 and Figure 3.7 demonstrate the relationship between durations of AFib episodes and cardiovascular health.

In a nationwide community cohort of patients with permanent AFib, nearly all patients had resting heart rate < 110 bpm (99%) and the majority (70%) were < 80 BPM. Atrial rates above a certain cut-off (typically 170–220 BPM) can be categorized as atrial high rate episodes (AHRE). AHRE can be triggered by any of the atrial tachyarrhythmias, including AFib, atrial flutter, or atrial tachycardia, with AFib

Table 3.2: The relationship between heart rates of AFib episodes and cardiovascular health

Resting Heart Rate	**Notes**
80 bpm	Majority of participants with permanent AFib 70% were <80 bpm *Steinberg et al.* (2013).
100 bpm	CCS guidelines define the resting heart rate target as <100 bpm, >100 bpm were associated with adverse outcomes in AFFIRM and RACE trials *Macle et al.* (2016).
110 bpm	Nearly all participants had resting heart rates <110 bpm (99%), in a nationwide cohort with permanent AFib *Macle et al.* (2016). Effective treating target for AFib *Van Gelder et al.* (2010).
170-220 bpm	Categorized as atrial high rate episodes *Shuai et al.* (2016).
>180 bpm	Severe AFib episode by cardiologist interview

and atrial flutter representing the most common arrhythmias *Shuai et al.* (2016). In the Atrial Fibrillation Follow-up Investigation of Rhythm Management (AFFIRM) study, adequate rate control was defined as < 80 bpm at rest, < 110 bpm on moderate exertion (6-minute walk test), and average and maximal heart rates of < 100 bpm and $< 100\%$ of age-predicted maximum, respectively, on 24-hour Holter monitor *Olshansky et al.* (2004). The CCS guidelines define the resting heart rate target as $<$ 100 bpm *Macle et al.* (2016); *Andrade et al.* (2017), the rationale being the observation that resting heart rates > 100 bpm were associated with adverse outcomes in AFFIRM and RACE. Table 3.2 and Figure 3.8 demonstrate the relationship between heart rates of AFib episodes and cardiovascular health.

A criteria region is constructed for severe AFib episodes based on literature reviews for HR and duration. Figure 3.9 below shows the suggested severe AFib region by combining Figure 3.7 and Figure 3.2. Episodes less than 30 seconds of duration are considered 'No AFib', since 2014 ACA guideline defines AFib episodes to be 30 seconds and greater, also study suggests that ECG as least 30 seconds is needed for accepted estimation of HR. HR below 80 is considered as no AFib, since it is well below the treating targets for AFib (100 BPM). A linear line was selected using duration-HR points (0.5 minutes, 180 bpm) and (5 minutes, 80 bpm). 180 bpm is

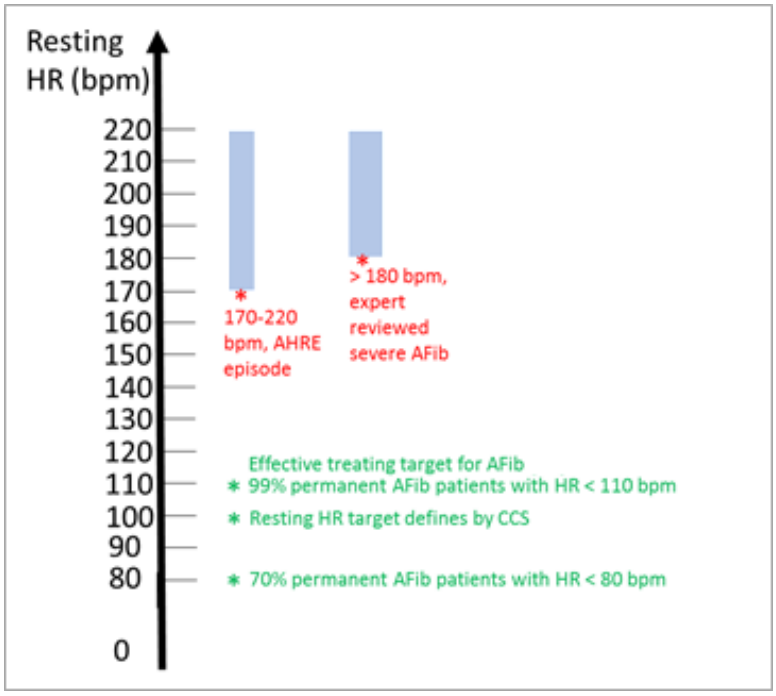


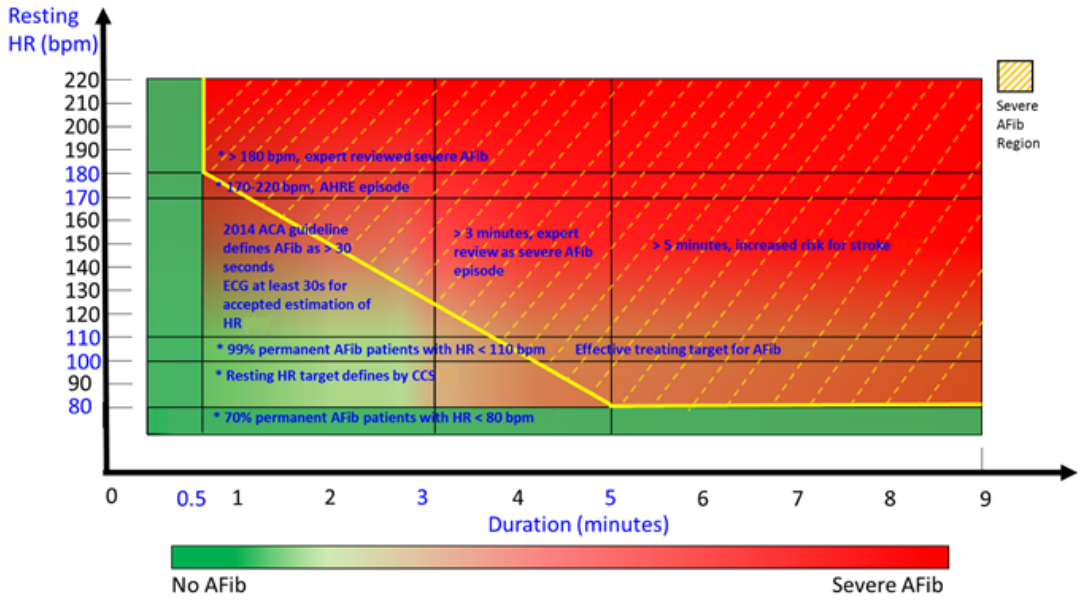
Figure 3.8: AFib annotation heart rates, a graphical depiction of the relationship between heart rate and severity of AFib.

chosen because it is our expert reviewed HR for severe AFib and 5 minutes is chosen because increased risk for stroke is associated for duration $>$ 5 minutes. The block bounded by HR 80 to 180 bpm and duration 30 seconds to 5 minutes is separated by the line in half. The upper triangle is considered more likely to be severe AFib episodes, since higher HR and longer duration is associated with higher severity.

3.3.2 SVT Annotation

To make predictions of SVT episodes waveforms from patients diagnosed with SVT required annotation, as some of the databases we have used are unannotated databases. An SVT annotation algorithm with input from cardiologists has been developed. The portions of signal annotated as SVT by our algorithm were then used predict SVT events. For the pre-processing of the data we applied a fourth-order Butterworth bandpass filter with cutoff at 0.5 and 40 Hz to the raw ECG for noise removal, and a double median filter with orders 0.2 and 0.6 times the sampling

Figure 3.9: AFib annotation, combining duration and heart rates



Assessing the severity of AFib at rest using a combination of heart rate and duration.

frequency to remove baseline wandering. We have also excluded the regions of the signal that are too noisy for analysis. An automated annotation algorithm was applied to determine SVT episodes with the ECG signal. Time windows consisting of three R-R intervals were annotated as SVT if they satisfied all of the following four SVT annotation criteria. First, an automated annotation algorithm based on the heart rate-duration criteria line was used to annotate high heart rate events. Portions of signals with extremely high heart rate y_2 that last for a short time period x_1, and a relatively lower heart rate y_1 that last for a longer time period x_2, are both considered to be indicative of high heart rate events. Using a straight line passing through the points (x_1,y_2) and (x_2,y_1), the lower bound for heart rates in the criteria region between duration x_1 and x_2 are determined. If any of the heart rates on the duration grid are above the heart rate-duration criteria line, then that portion of the signal lies in the criteria region and will be classified as high heart rate. In the case of SVT, the episodes lie above the heart rate-duration criterion line passing through the

points $(30, 150)$ and $(180, 100)$. Secondly, the difference between three consecutive R-R intervals should be less than 50 ms. Thirdly, there should be fewer than 4 P/T like waves within the window. Finally, the cross-correlations with the 500 P-wave templates extracted from the MIT-BIH P-wave database should be less than 0.85. A portion of the annotated SVT episodes were randomly selected and reviewed by a cardiologist to confirm the sensitivity of the annotation algorithm.

3.3.3 VT Annotation

To make prediction on VT episodes in our databases, waveforms diagnosed with VT required annotation. Two criteria were first applied to filter out the intervals that might contain VT episodes. The first criterion is the heart rate should be above 100 bpm and the second criterion is average width corresponding to half length of the peaks should be wider than 40 ms. The second criterion is defined because VT episodes have wide QRS complex (> 120 ms), with half length of the peaks > 40 ms, the width QRS complex will likely to be at least > 80 ms. This criterion is set to be lower than the 120 ms, to make sure that we capture all the possible VT episodes for manual review. The last step of VT annotation requires manual review by clinician of the episodes which satisfied the 2 criteria.

3.3.4 Bradycardia Annotation

In order to extract bradycardia events that have more clinical importance, several criteria were employed in terms of heart rate, medical diagnosis and medication.

1. Only the events with a heart rate of less than 60 BPM for 10 consecutive heart beats were considered. 60 BPM was chosen because the National Institutes of Health defines bradycardia as a heart rate < 60 BPM in adults other than well trained athletes *Kusumoto et al.* (2019). Some studies and guidelines have suggested to use < 50 bpm for the definition. We have chose < 60 bpm over

$<$ 50 bpm for because of first other additional criteria on co-morbidity and medication have already been used filter the population, secondly with $<$ 50 bpm there will be very limited bradycardia events for analysis (only 38 events from 23 participants if $<$ 50 bpm criteria was used). The choice of 10 consecutive beats was made to eliminate short intervals of accidental low heart rate.

2. Patients without CHF were excluded. CHF makes the heart pump less efficiently and effectively, thus causing symptomatic bradycardia.

3. The remaining population that satisfies both criteria is then separated according to whether the patient is taking beta blockers or not. Beta blockers are a type of medication used to manage abnormal heart rhythms including treatments of tachycardia, myocardial infarction, CHF and other conditions. The effect of beta blocker can result in a slower heart rate. The participants taking beta blockers were excluded to eliminate low heart rates which were potentially due to medication.

3.3.5 Activity Level

The raw accelerometer data are collected along three orthogonal axes in the device-specific frame of reference. The continuous accelerometer data are collected using a wireless monitoring device that adheres to patient's chest (Preventice Solutions, Inc) and sampled at 10 Hz. The accelerometer data were up-sampled to match the ECG sampling rate. For the analysis, the accelerometer magnitude $am(t)$ consisting of the vector magnitude of the accelerometer data at each time point is used in analysis together with the synchronized ECG signal.

$$\begin{aligned} am(t) &= \sqrt{x(t)^2 + y(t)^2 + z(t)^2} \\ &= \text{ accelerometer magnitude at the time point } t. \end{aligned} \tag{3.4}$$

The data is aggregated using mean amplitude deviation (MAD) which computes the deviation of $am(t)$ from its mean over the corresponding epoch, averaged over the length of the annotated AF signal *Bakrania et al.* (2016).

$$MAD = \frac{1}{n} \times \sum_{i=1}^{n} |am(t_i) - \overline{am}| \tag{3.5}$$

where:

$am(t_i)$ = accelerometer magnitude at the ith time point

$\overline{am}$ = mean accelerometer magnitude within the time period of interest

n = length of the time period

The $MAD \leq 0.75$ quantile relative to the measurements from the entire group id designated as low activity and $MAD > 0.75$ quantile as high activity in order to account for inter-individual differences in activity levels within the study group. Threshold of 0.75 quantile for distinguishing activity levels is an arbitrary choice since there is no well-established guideline for low vs. high activity level.

3.4 Pre-processing and Event Extraction

The aim of pre-processing is to remove noise from ECG signals. Methods focusing on heartbeat segmentation within the ECG signal tend to require pre-processing *Luz et al.* (2016). Automated annotation algorithms were used to annotate the arrhythmia event and activity events as described in section 3.3.1. Pre-processing, noise removal, and arrhythmia event extraction were performed in the same fashion for different types of arrhythmia. The same pre-processing was applied to all methods in order to make the results commensurable.

The following section describes the pre-processing and noise removal steps.

3.4.1 Pre-processing

During the signal pre-processing step, a fourth-order Butterworth band-pass filter with cutoff frequencies of 0.5 and 40 Hz is first applied to the raw ECG signal to remove noise, after which a double median filter with orders equal to 0.2 and 0.6 times the sampling frequency is applied to remove baseline wandering. To calculate heart rate and beat duration, the Pan–Tompkins algorithm for QRS detection is used *Pan and Tompkins* (1985) *Sedghamiz* (2014).

3.4.2 Noise Removal

3.4.2.1 Noise Detection Algorithm 1

After R peak detection, a stepwise noise detection method is performed. There are 3 criteria for noise detection. First, the method calculates the percentage of missing signal in a defined time window (300 seconds) and checks if the missing signal percentage is above a certain threshold (15%). If the percentage of the missing signal is above the threshold, the signal is classified as being noisy by the first criteria. Secondly, if the signal passes the first criterion then non-linear filter analysis is used to determine missing peaks. If the signal passes both the first and the second criteria, the third step determines if the percentage of missing R peaks is above a certain threshold (15%) in the current window. The window of signal will be classified as noisy if it fails any of these three criteria. Noisy signals will not be used for subsequent annotation. Signals that occur too close to the noisy signal will also be excluded from the analysis.

3.4.2.2 Noise Detection Algorithm 2

The ECG is then divided into time windows of 30 seconds each. R peaks are found using the Pan-Tompkins algorithm *Pan and Tompkins* (1985). Three criteria

are evaluated for each window based on cross-validated thresholds: high amplitude artifacts, abnormal R peak distributions, and high amounts of distortion artifacts. Distortion artifacts are characterized by clusters of critical points on the ECG curve. Critical points are points on the ECG curve with a derivative of zero and an absolute amplitude greater than 0.25 times the average R peak height within the window. A high number of critical points indicates that the underlying waveform is hidden by high frequency and high amplitude distortions that cannot be removed by simple filtering.

3.4.3 Pre-event Signal Extraction

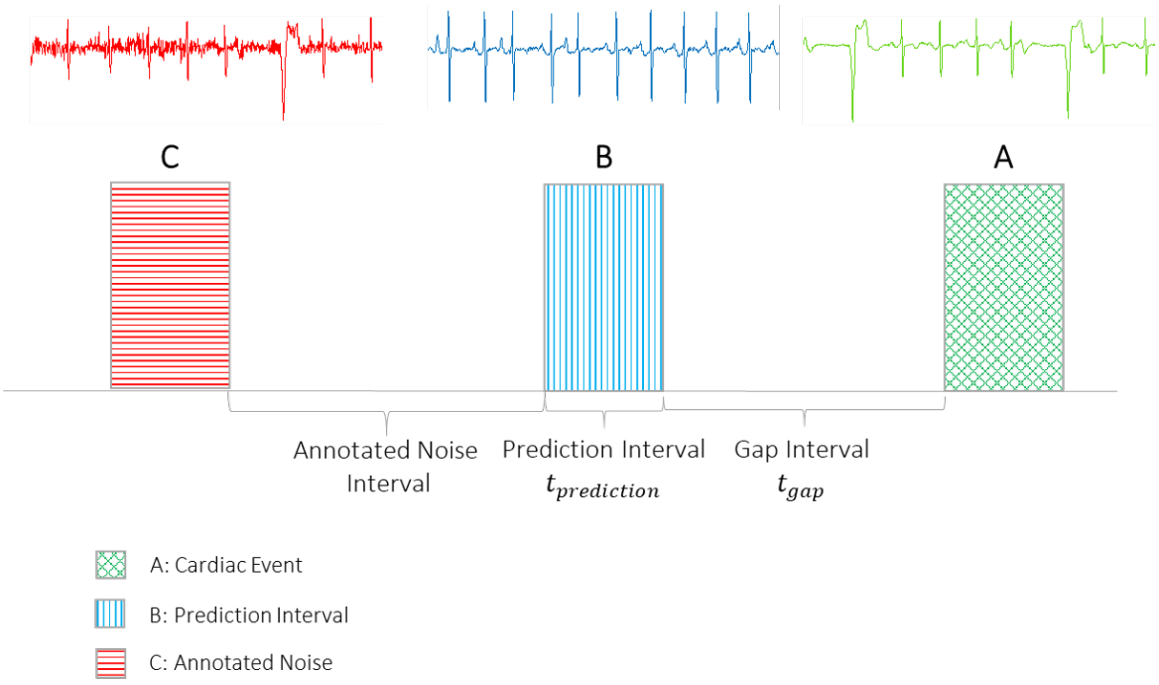
One of the aims of this study is to predict the onset of an cardiac event using pre-event signals. In order to assess the predictive power of the method, events that occurred too close to previous events were excluded, as the prediction interval should not overlap with arrhythmia events. Events that occurred within defined length of minutes of a noisy signal were also excluded, with the aim to ensure that the prediction interval is out of the noisy signal range. Figure 3.10 shows an example of an annotated ECG signal. After annotation, there is an annotated event A and attendant prediction interval B, which is located t_{gap} minutes before the annotated event A with a length of t_{signal} minutes. The interval C classified as noise will be excluded from the analysis, along with the neighboring intervals of defined length of minutes.

In the prediction experiments, different combinations of t_{gap} and t_{signal} for different types of arrhythmia and cardiac events.

3.5 Comparison Method

In order to validate DPFA performance, several comparison algorithms have been applied to the datasets. In most cases the algorithm outperforms the other more

Figure 3.10: Prediction interval extraction.



Intervals of signal are extracted for use in prediction that are t_{gap} minutes prior from the arrhythmia event and several minutes away from detected noise.

traditional and well-established approaches, including a heart rate variability HRV based method using SVM a method combining discrete wavelet transform (DWT), PCA, a deep neural network (DNN), a CNN, and a CNN with LSTM.

Automatic detection of types of arrhythmia or cardiac conditions encompasses several basic steps, including pre-processing/segmentation, feature extraction, followed by a classifier *Apandi et al.* (2018). Machine learning (ML) approaches for arrhythmia detection can be grouped into two main categories based on feature extraction strategies *Rizwan et al.* (2020). The first group uses features extracted from the raw ECG as input followed by classical ML algorithms such as SVM, KNN, and decision trees as classifiers. Such ML algorithms require dimensionality reduction after feature extraction from the ECG signals. The second group uses raw ECG data as input and does not require feature extraction. Information from the original physiological signals can be directly processed by algorithms in this group *Faust et al.* (2020). ML

algorithms like neural networks, including their basic and advanced versions, use the raw data for model training and detection of arrhythmia types. Compared to the first group, algorithms in this group generally have higher computational cost and require larger datasets to achieve equal or greater performance. Our method belongs to the second group, even though it is not based on neural networks.

In medical applications, datasets that are both sufficiently large and well-annotated are not always available due to nature of the various diseases. Thus both groups have their advantages and drawbacks. In this study we selected five methods from both groups for comparison.

The first method is an SVM based arrhythmia classification algorithm with features extracted from heart rate variability (HRV with SVM) by Asl et al. *Asl et al.* (2008). Following *Asl et al.* (2008), time and frequency domain features were extracted from the signals. However, generalized discriminant analysis (GDA) for feature reduction was not utilized in this study since GDA would reduce the number of features, and there are only two classes to begin with. The SVM model with radial basis function (RBF) kernel was used for classification.

The second method for comparison uses the principal components of the DWT coefficients as features and SVM as the classifier *Martis et al.* (2012).

The third comparison method is a DNN for arrhythmia prediction by *Hannun et al.* (2019). The authors used a CNN to detect types of arrhythmia. The network architecture has 34 layers, takes ECG signals as input and outputs arrhythmia classes. The same experimental setup was used as in *Hannun et al.* (2019). The network was trained de novo with random initialization and the Adam optimizer with default parameters.

The fourth and the fifth comparison methods are both based on CNN techniques. The fourth method, due to *Acharya et al.* (2017a), utilized a CNN algorithm to automatically detect different ECG segments. The algorithm consists of an eleven-

layer deep CNN. The fifth method, due to *Tan et al.* (2018), employed a LSTM (CNN with LSTM) to classify ECG signals.

To compare the performance with alternative algorithms, the same training, cross validation, and testing data sets were used.

CHAPTER IV

Data Sources and Description

4.1 Introduction

A total of six databases have been constructed and collected for the project. Database 1 (DB1) is comprised of several publicly available databases and has been used for initial algorithm development. These are also benchmark databases that also allow us to compare our algorithm with other existing methods. Database 2 (DB2) is a retrospective database collected by UMHS, it containing previous UMHS cardiac patients and non-cardiac control group. Clinicians have also reviewed DB2 and annotated the waveforms, these annotations are the gold standard for classifying arrhythmia types and can be used for further model training and prediction. Database 3 (DB3) is planned to be a prospective database that contains in-hospital data collected from inpatients at UMHS. However, DB3 has not been collected due to patient participation issues. Details about the collection difficulties can be found in section 4.4. Database 4 (DB4) consists of data collected from the patients after discharge, at-home, in-vehicle and in other daily activities. DB4 contains data collected from portable device and each participant has approximately 14 days of continuous ECG data. Database 5 (DB5) contains data collected from healthy subjects, it serves a control group. Each participant has approximately 7 days of continuous ECG data. Database 6 (DB6) consists of a total of 45 patients with history of AF who presented

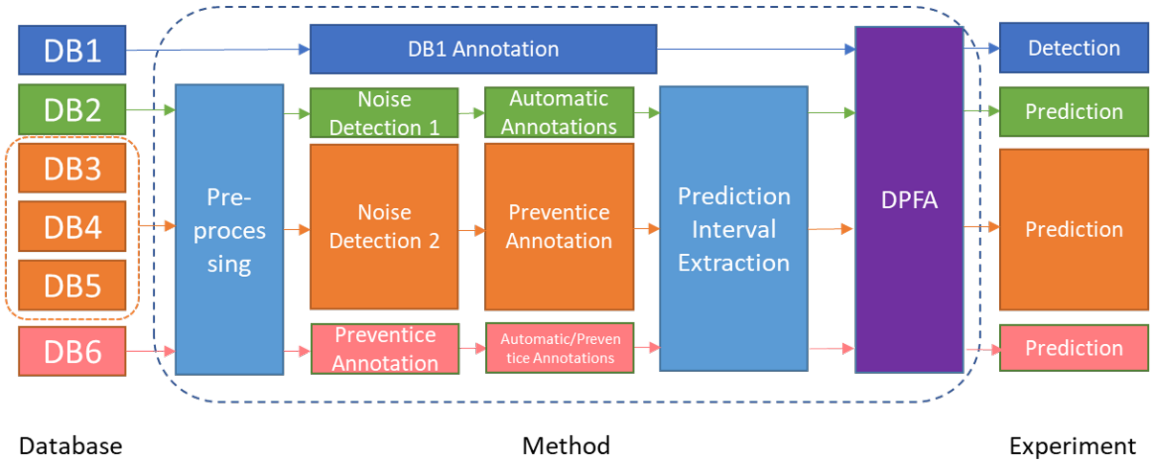


Figure 4.1: Overview of databases, methods and prediction experiments.

to University of Michigan and are recruited in the study. The ECG and accelerometer data were recorded continuously for up to 3 weeks. See Table 4.1 for a summary of the characteristics of the databases and Figure 4.1 for an overview of databases, methods and prediction experiments. Details of each database are described in sections below.

4.2 Database 1

DB1 is comprised of five external databases: PhysioNet/CinC 2017 *Goldberger et al.* (2000), MIT-BIH Arrhythmia database *Moody* (1983), European ST-T Database *Taddei et al.* (1992), PTB Diagnostic Database *Bousseljot et al.* (1995) , MIMIC-III database *Johnson et al.* (2016). Among the five databases, PhysioNet/CinC2017, MIT-BIH and European ST-T databases contain annotated signals. As the MIT-BIH and European ST-T databases contain very few AFib patients and PhysioNet/CinC2017 consists only of isolated single episodes, all three databases are unsuitable for developing and testing for predictions. To have more data for supraventricular arrhythmia and ventricular arrhythmia we have included three additional databases: MIT-BIH supraventricular arrhythmia database *Greenwald et al.* (1990), MIT-BIH malignant ventricular arrhythmia database *Greenwald* (1986) and Creighton university ventric-

Table 4.1: Summary of the characteristics of the databases.

Database	Number of Subjects	Source	Data	Data Collection Environment	Experiment
DB1	Various	Publicly available databases	ECG signals	Various	Arrhythmia detection
DB2	1270	UM Hospital Retrospective	ECG signals	In-hospital	Arrhythmia prediction
DB4	66	UM Hospital	Continuous signals from Bodyguardian	At home, in-vehicle, others	Arrhythmia prediction
DB5	80	Healthy Subjects	Continuous signals from Bodyguardian	At home, in-vehicle, others	Arrhythmia prediction
DB6	45	AFib Subjects	Continuous signals from Bodyguardian	At home, in-vehicle, others	Arrhythmia cardiac event prediction

ular tachyarrhythmia *Nolle et al.* (1986). These three databases have annotations for SVT and VT and we used them for testing our algorithm's performance on SVT and VT detections. See Table 4.2 for a summary of the databases we have used in papers and manuscripts.

In performing AFib detection on the PhysioNet/CinC2017 database, we correctly classified 84.2% AFib cases and 93.3% non-AFb cases in the testing set, with an F1 score of 0.74 and AUC of 0.95. For SVT detection, we correctly identified 41/56 SVT events (0.73 sensitivity) and 97/100 (0.97 specificity) non-SVT events. The AUC is 0.951 and F1 of 0.914. Details about the SVT detection are in section 5.3.1.

For VA detection, we correctly identified 243/261 (0.93 sensitivity) VA episodes and 227/261 (0.87 specificity) non-VA episodes for 5-seconds long signals, with an AUC of 0.96 and F1-score of 0.91. For 2-seconds long signals, the algorithm correctly identified 625/662 (0.94 sensitivity) VA episodes and 600/662 (0.91 specificity) non-VA episodes with an AUC of 0.97 and an F1-score of 0.93. Table 4.3 below summarizes

Table 4.2: Database 1 Overview

Source	No. of Recordings / Patients	Data	Length of Signal	Freq -uency (Hz)	Anno -tation	Experi -ment
PhysioNet /CinC 2017	8582 / NA	Single lead	30 - 60 seconds	300	Yes	AFib detection
Long-term AF Database	84 / NA	Holter Recordings	various	128	Yes	SVT detection
MIT-BIH Arrhythmia	48 / 47	ML II, modified lead V1	30 seconds	360	Yes	SVT detection
MIT-BIH Ventricular Arrhythmia Database	22 / NA	Single lead	30 seconds	250	Yes	SVT detection
Creighton University Ventricular Tachy-arrhythmia Database	35 / NA	Single lead	8 minutes	250	Yes	SVT detection

results based on DB1.

Arrhythmia prediction experiments were not tested in DB1, since most of the recordings in the publicly available databases were of insufficient duration. DB1 was mainly used for preliminary algorithm development, for prediction analysis we used other databases. Details about the prediction analysis can be found in DB2, DB4, DB5 and DB6.

4.3 Database 2

DB2 is a retrospective database containing previous UMHS cardiac patients and a non-cardiac control group. It increases size of data for algorithm development and

Table 4.3: DB1 Result summary for arrhythmia detection

Arrhythmia	**Sensitivity**	**Specificity**	**AUC**	**F1**
AFib Detection	0.84	0.93	0.95	0.74
SVT Detection	0.73	0.97	0.95	0.91
VA Detection	0.93	0.87	0.96	0.91

Table 4.4: Summary of DB2 by types of arrhythmia

Arrhythmia	**Number of Waveforms**	**Number of Subjects**
AFib	3540	514
Other	2464	355
VT	2156	309
VF	264	45
Block	1000	141
MI	1556	223
Brady	396	62
SVT	976	139
Total	8838	1270

training. One of our main objectives is to predict various types of arrhythmia. Length of ECG is too short for predicting these events, DB2 on the hand has continuous hospital bedside ECG recorded from cardiac patients. We used data for training the prediction models.

The ECG waveforms in DB2 were collected from leads I, II, III and IV. There is a total of 8838 waveforms from four leads from 1270 participants in DB2. Among these 8838 waveforms, there are 514 subjects with AFib episodes across 3540 waveforms. The data quality is monitored by UM experts according to UM standards. A summary of types arrhythmia contained in DB2 is shown in Table 4.4 below.

We have completed AFib prediction analysis, SVT prediction analysis and VT prediction using DB2 with the DPFA algorithms. Our own annotation algorithms for AFib, SVT and VT were applied to the waveforms in DB2 and used for classifying events. For AFib predictions, the AUC is above 0.85 for all 2-minute long prediction intervals. Prediction performances with 2-minute long signals are slightly better than

Table 4.5: DB2 result summary for arrhythmia prediction

Arrhythmia	Gap Interval (min)	Prediction Interval (min)	AUC
AFib Prediction	0.5	2	0.85
	1	2	0.85
	1.5	2	0.86
	2	2	0.86
	2.5	2	0.86
	3	2	0.86
SVT Prediction	0.5	2	0.85
	1	2	0.84
	1.5	2	0.83
	2	2	0.83
	2.5	2	0.82
	3	2	0.83
VT Prediction	0.5	2	0.78
	1	2	0.77
	1.5	2	0.76
	2	2	0.75
	2.5	2	0.75
	3	2	0.76

1-minute long and 0.5-minute long signals. For the SVT prediction analysis, we examined the predictive power of the DPFA algorithm with different pre-SVT lengths. The AUC is above 0.80 for all 2 minute-long prediction intervals. As the time to the SVT event increases, the mean prediction performance gradually declines while variance increases. The AUC for VT prediction is around 0.75, which are not as good as other types of arrhythmias like AFib or SVT. The lack of performance might be due to the low number of participants having VT compared to other arrhythmia types. Details about the prediction results can be found in sections in Chapter 6.

4.4 Database 3 and Database 4

4.4.1 Database Construction Challenges

Database 3 (DB3) is planned to be a prospective database that will contain in-hospital data collected from current inpatients at UMHS. It will contain continuous ECG, blood pressure and patient clinical data. Database 4 (DB4) consists of a subset of patients that will be included in DB3 with data collected after discharge, at-home and during daily activities including driving. Prior to the start of patient recruitment, several changes were made to the study protocol including the removal of the required sensor, the addition of the blood pressure cuff, and recruiting from the Cardiovascular Center (CVC) ICU instead of from the Emergency Department.

The purchase of additional devices required for the study was delayed. These were later resolved and the devices were approved for purchase after which patient recruitment started. However, the aforementioned recruitment strategy resulted in a number of challenges for both databases. For potential patients planned to include in DB3, due to their illness many of these patients were intubated and/or sedated and were thus unable to consent to in-hospital monitoring. Additionally, the severity of their illnesses curtailed patient interest in study participation. Due to these patient participation concerns, DB3 has not been collected as planned.

There were numerous challenges for recruiting patients for the at-home database DB4 as well. Many patients who were discharged from the CVC ICU had recently closed wounds on their chests as a result of recent surgery. These wounds prevented them from wearing the BodyGuardian monitor. Other potential patients were prescribed a separate at-home monitor for clinical use, which prevented them from wearing the BodyGuardian monitor. Finally, as UMHS serves as a regional referral hospital, many patients lived two or more hours away, and were unwilling to travel back to Ann Arbor to return equipment. These challenges resulted in only one study

participant being successfully recruited from the CVC ICU.

In order to resolve these aforementioned recruitment issues, we revised our recruitment strategy in consultation with our study coordinator and clinician. Rather than recruiting from the CVC ICU, where patients meet our inclusion criteria with respect to relevant arrhythmias but are often too sick to be recruited, we have been recruiting patients who use the services provided by the CVC Device Clinic.

The CVC Device Clinic services patients with pacemakers, ICD, and other cardiac equipment. These patients are at risk of developing life-threatening arrhythmias such as severe AFib or SVT, VFib, and VT, making them ideal subjects for our study. For Databases 3 and 4, 66 patients have been enrolled, 11 patients have exited the study early or have no data. A total 55 patients have completed the study and have data from BodyGuardian devices. Although we have enrolled less participants than originally planned in the, each participant has a much longer monitoring time. For each subject that completed the study we have approximately 14 days of continuous ECG data compared to 3 - 4 days as planned. Table 4.6 below shows number of participants with different diagnosis in DB4.

We have completed AFib prediction analysis, SVT prediction analysis using DB4 with the DPFA algorithms. Preventice annotation algorithm For AFib predictions, the AUC is around 0.71 for various prediction intervals. For the SVT prediction analysis, we examined the predictive power of the DPFA algorithm with different pre-SVT lengths. The AUC is above 0.75 - 0.88 for various prediction intervals. As the time to the SVT event increases, the mean prediction performance gradually declines while variance increases.

4.5 Database 5

Database 5 consists with healthy control participants for the study. Similar to DB4, participants in DB5 have data collected with Bodyguardian devices provided

Table 4.6: Patients in Database 3 and Database 4 by diagnosis

Diagnosis	**Number of Participants**
AFIB	9
AFIB, Non-sustained ventricular tachycardia	2
AFIB, VFIB	1
AT	1
AV block	2
Bradycardia	2
Hypertrophic cardiomyopathy	1
Heart Block	1
Implantable cardiac monitor/ complete heart block	1
Non-ST elevation myocardial infarction	1
Non-sustained ventricular tachycardia	5
Sinus brady	1
Tachy-Brady syndrome	1
VFIB	1
VFIB arrest	1
VT	23
VT, SVT	1
Wolff Parkinson white	1

by Preventice. A total of 80 subjects have been enrolled as planned, with 74 having completed data collection. For each subject that completed the study we have approximately 7 days of continuous ECG data which is much longer than what we have planned (2-3 hours each).

Table 4.7 below shows a summary of recorded time in DB4 and DB5. There is a total of 721 ECG recorded over 13963 hours in DB4 and a total of 555 ECG recorded over 10307 hours in DB5. 26328 AFib events for total of 364 hours are recorded in DB4 and 348 AFib events over a total of 7 hours are recorded in DB5 with Preventice annotations. Among the 55 participants in DB4, 16 of them had AFib events during driving. There is a total of 13963.5 hours ECG recording time in DB4, with 873.7 (6.26%) hours of them recorded in vehicle. There is a total of 363.5 hours of AFib recorded in DB4, with 16.4 (4.51%) hours of them recorded in vehicle. So there is no evidence suggesting more AFib were recorded in vehicle. A total of 532 hours of

Table 4.7: Summary of recorded ECG time

	DB4	**DB5**
Total number of ECG	721	555
Total ECG time (hrs)	13963.53	10306.73
Total number of AFib events	26328	348
Total AFib time (hrs)	363.52	7.56
Total Normal Sinus time (hrs)	4646.21	6201.41
Total in-vehicle AFib time (hrs)	16.41	0.01
Total in-vehicle time (hrs)	873.73	532.13

in-vehicle time is recorded in DB5, only 0.01 hours of AFib time is observed which is expected since DB5 is the control group.

4.6 Database 6

Database 6 (DB6) consists a total of 45 patients with history of AF who presented to University of Michigan are recruited in this database. All patients wore an event recorder (Preventice solutions Inc) for up to three weeks. IRB approved the protocol and written informed consent was obtained. The ECG and accelerometer data were recorded continuously for up to 3 weeks. DB6 is used for prediction analysis of RVR with low levels of activity, 18 participants are included in the prediction analysis. Table 4.8 describes the patient characteristics of DB6.

Table 4.8: Characteristics of patients

Variable	**All Participants (n=45)**	**Participants in Prediction Analysis (n=18)**
Female	14 (31.82%)	5 (27.78%)
Age	66.36 (11.67%)	69.13 (7.31%)
BMI	31.30 (6.08%)	30.93 (5.66%)
Hypertension	26 (59.09%)	11 (61.11%)
History of Stroke	0 (0.00%)	0 (0.00%)
Diabetes	12 (27.27%)	5 (27.78%)
Coronary artery disease	11 (25.00%)	5 (27.78%)
Peripheral vascular disease	2 (4.54%)	2 (11.11%)
Beta Blockers	31 (70.45%)	13 (72.22%)
Calcium channel blockers	14 (31.82%)	4 (22.22%)
Antiarrhythmic drugs	9 (20.45%)	2 (11.11%)

CHAPTER V

Presentation of Research

5.1 Introduction

This chapter presents the results for the experiments on the detection and prediction of different types of arrhythmia and cardiac events followed by discussion and limitation sections. The experiments include AFib detection based on DB1 with publicly available databases, AFib prediction based on DB2 with hospital bed-side data and DB4 and DB5 for data collected from portable devices and during driving; SVT detection based on DB1 with publicly available databases, SVT prediction based on DB2 with hospital bed-side data and DB4 and DB5 for data collected from portable devices; VA detection based on DB1 with publicly available databases, VT prediction based on DB2 with hospital bed-side data; Bradycardia prediction based on DB2 with hospital bed-side data and RVR with low activity prediction based on data collected on DB6 from portable devices.

5.2 Atrial Fibrillation

5.2.1 AFib Detection on Benchmark Datasets

5.2.1.1 Data

PhysioNet/Cinc2017*Goldberger et al.* (2000) described in Section 4.2 was used to help develop the algorithms. The dataset contains 738 AFib episodes and 7790 non-AFib episodes consists with normal, noisy or other episodes.

5.2.1.2 Method

Encoding and peak detection algorithms are applied to produce a training peak probability vector for AFib cases and a training peak probability vector for non-AFib cases. The DPFA algorithm is then applied to the training probability vectors to obtain the transition matrices.

Encoding and peak detection algorithms are applied to the testing data to produce a testing probability vector. By applying the results from AFib DPFA output and non-AFib DPFA output to the testing probability vector, we can get a probability matrix for AFib and non-AFib.

We take the logarithmic average of the two probability vectors and compare the two results. The class with the higher average of the probability will be the predicted class. The predicted class is then tested against the 'true' label. The method is described in Section 3.2.

5.2.1.3 Result

Two DPFA models are computed from the dataset:M_A for class 'AFib' with 5892 states and M_N for class 'Other' with 117058 states. We have correctly classified 84.2% of AFib cases and 93.3% of non-AFib cases in the testing set with an F1 score of 0.74 and AUC of 0.95. The results are shown in Table 5.1 and Figure 5.1.

Table 5.1: AFib detection results on publicly available benchmark datasets

Class	Classfied as "AFib"	Classified as "Non-AFib"	Accuracy
AFib	80	15	84.2%
Non-AFib	61	844	93.3%

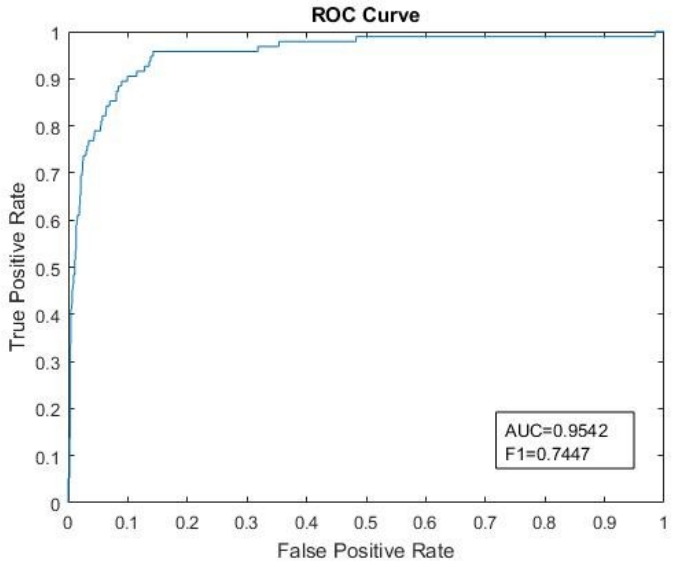


Figure 5.1: AFib detection, ROC curve for AFib detection on publicly available benchmark datasets (DB1).

Publicly available benchmark databases were used for algorithm development for earlier stage of the study. The result has validated the possibility of using DPFA algorithm to predict AFib events.

5.2.2 AFib Prediction on Hospital Bedside Dataset

Experiments for AFib predictions are performed on DB2 which consists of hospital bedside data, Methods, result and discussion for AFib prediction with hospital bedside dataset are described in details in sections below .

5.2.2.1 Data

A retrospective database with ECG signals from Michigan Medicine cardiac patients with AFib was used in this study. After excluding recordings with pacemakers, implantable cardioverter defibrillators (ICD), or ventricular assist devices (VAD), there were 1210 lead II ECG recordings for patients with AFib. The recordings were sampled at 240 Hz. Details about this dataset is described in Section 4.3.

5.2.2.2 Method

DB2 consists of data collected from hospital bedside and is used for prediction experiment. DB2 does not have annotation, thus an automated annotation algorithm has been developed to annotate the events with a combination of heart rate and duration. Heart rate is calculated based on the R peaks within a time interval, which were detected previously during test interval extraction. Duration is calculated by counting the number of consecutive intervals that a particular heart rate spans. For this study, the time interval was set to 30 seconds and counted up to 6 intervals, which allows for the duration to range from 30 to 180 seconds.

The automated annotation algorithm based on the heart rate-duration criteria region is described in detail in Section 3.3.1 to annotate AHRE events with different heart rate and duration limits. The lower duration limit of 30 seconds is chosen in accordance with the definition of an AFib event provided by the 2014 AHA guideline *January et al.* (2014). The lower heart rate limit is set to 110 bpm, as this rate has proven to be an effective treatment target for AFib *Van Gelder et al.* (2010). The higher heart rate limit is set at 160 bpm since a rapid heart rate is more likely to cause symptoms.

Rapid heart rate in AFib may also have an untoward effect on cardiac function, resulting in tachycardia-induced cardiomyopathy *Fuster et al.* (2006). Using this annotation method, it becomes possible to capture the portions of the signal that

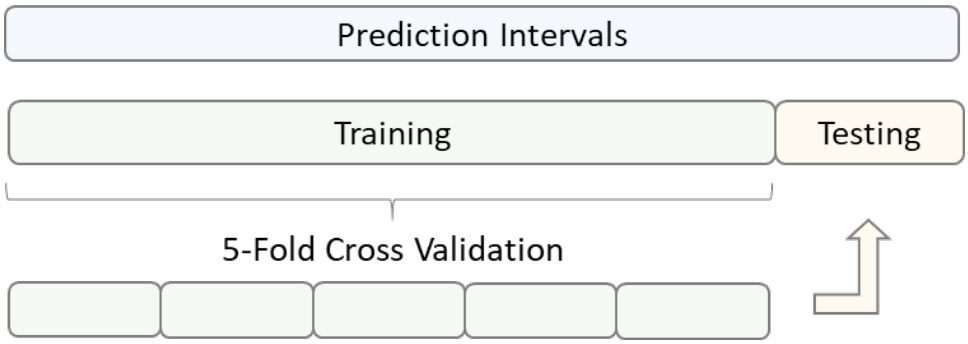


Figure 5.2: AFib prediction, data partitioning scheme used for AFib prediction in hospital bedside database (DB2).

correspond to AHRE with sufficiently high severity, either in the form of extremely high heart rate over a span of 30 seconds, or moderately high heart rate stretching over 180 seconds. A heart rate-duration criteria region that lies above the line passing through the points $(30, 160)$ and $(180, 110)$ is used. These AHRE events were annotated as surrogate for AFib.

A total of 417 AFib events were labeled by these criteria and a total of 7319 non-AFib intervals were randomly selected from non-AFib regions of the signals. A total of 353 of the AFib intervals were used in training with the remaining 64 intervals being held out for testing. Within the training data set, 5-fold cross validation was performed for parameter tuning (Figure 5.2). Training, cross validation, and testing sets/folds were partitioned on a participant level, meaning that signals from the same participant were only included in one set/fold so as to prevent over-fitting.

The five models (each with four folds of training data) with the best parameters tuned during cross validation were then applied to the testing set.

5.2.2.3 Result and Discussion

Different combinations of signal intervals (i.e., $t_{signal} = 0.5, 1.0, 2.0$ minutes) and gap intervals (i.e., $t_{gap} = 0.5, 1.0, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5$ minutes) up to 5 minutes prior to the event were used for prediction. These prediction intervals were tested

using the DPFA algorithm. Figure 5.3 depicts a summary of the performance. For the AHRE DPFA, around 758×3, 683×3, and 582×3 transition states were generated for 0.5, 1.0 and 2.0-minute long AHRE prediction intervals based on the training data set. For the non-AHRE, an average of 626×3 transition states were generated for the models using 0.5-minute long signals, 642×3 transitions states for the 1.0-minute long signal models, and 459×3 transition states for the 2.0-minute long signal models.

The AUC is above 0.80 for all prediction intervals. Performance is nearly steady when gap length increases and the prediction interval moves further away from the AHRE events. Prediction performances with 2-minute long signals are slightly better than 1-minute long and 0.5-minute long signals (Figure 5.3). Longer signals contain more local and global patterns and these information may have contributed to the better performances.

The performance of DPFA on raw ECG data was compared with its performance on data with pre-processing (Figure 5.3). The AUCs for prediction intervals are almost the same with or without pre-processing. Nor is there any statistically significant difference between these results.

For AHRE prediction, the DPFA algorithm has a comparable performance to deep learning with on average a 2% higher AUC. For the other comparison algorithms, DPFA has a 17.9% higher AUC than HRV, 26.1% higher AUC than DWT, 15.4% higher AUC than CNN and 10.3% higher AUC than CNN with LSTM. The results for these experiments are shown in Figure 5.4.

The McNemar test was also performed to evaluate the performance of the alternative methods against the proposed DPFA algorithm. The McNemar test used a 1:11 ratio for the cost for the imbalance of data. Tables 5.2, 5.3, and 5.4 depict the results for half-minute, 1-minute, and 2-minute long signals with various gap sizes, respectively. p-values representing the statistical significance of the different performance in terms of AUC ($< .05$) are shown in bold font in the referenced tables. The

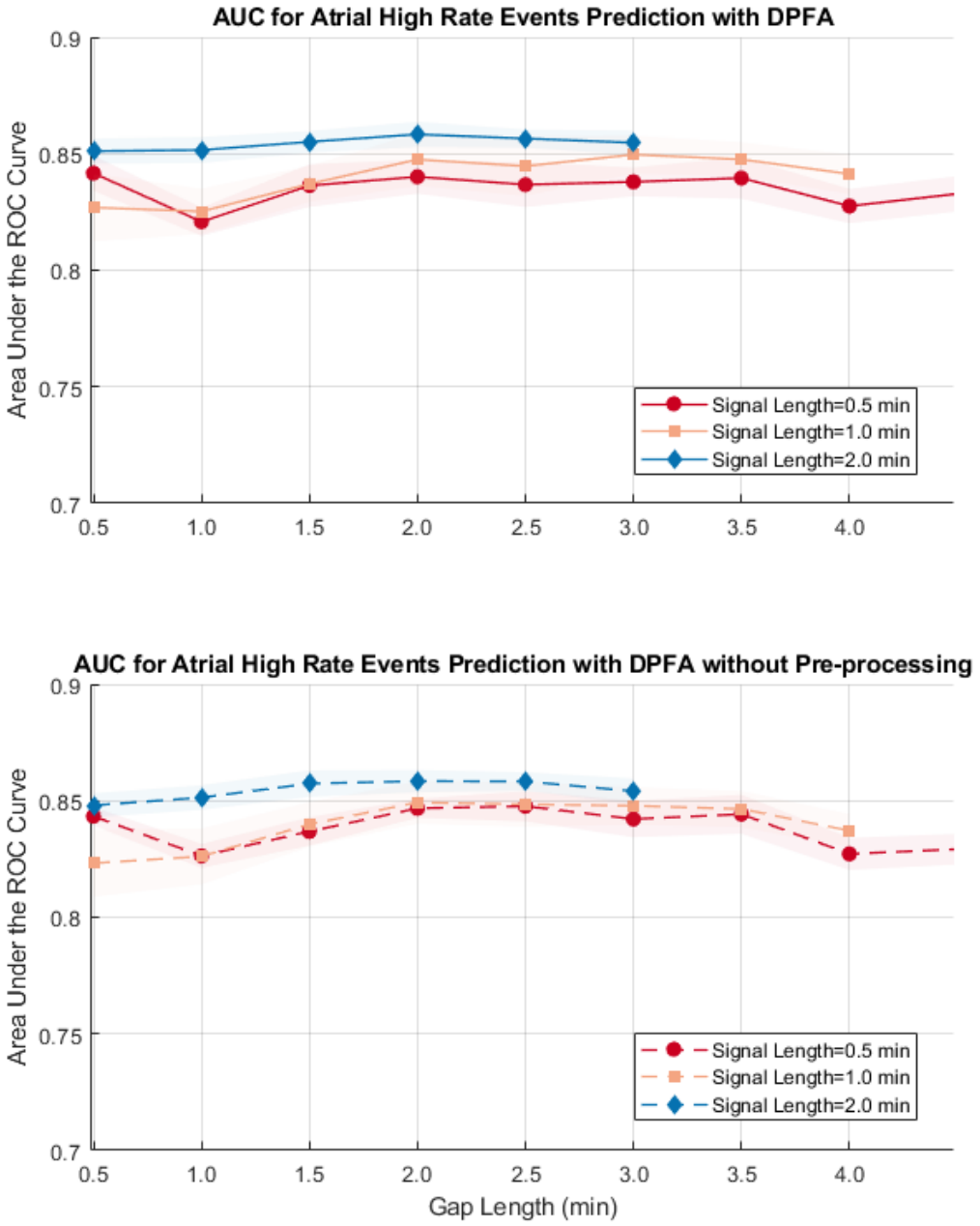


Figure 5.3: AFib prediction on hospital bedside database (DB2), A comparison of AUC for AHRE prediction using the DPFA with and without pre-processing method for various signal lengths and gap intervals.

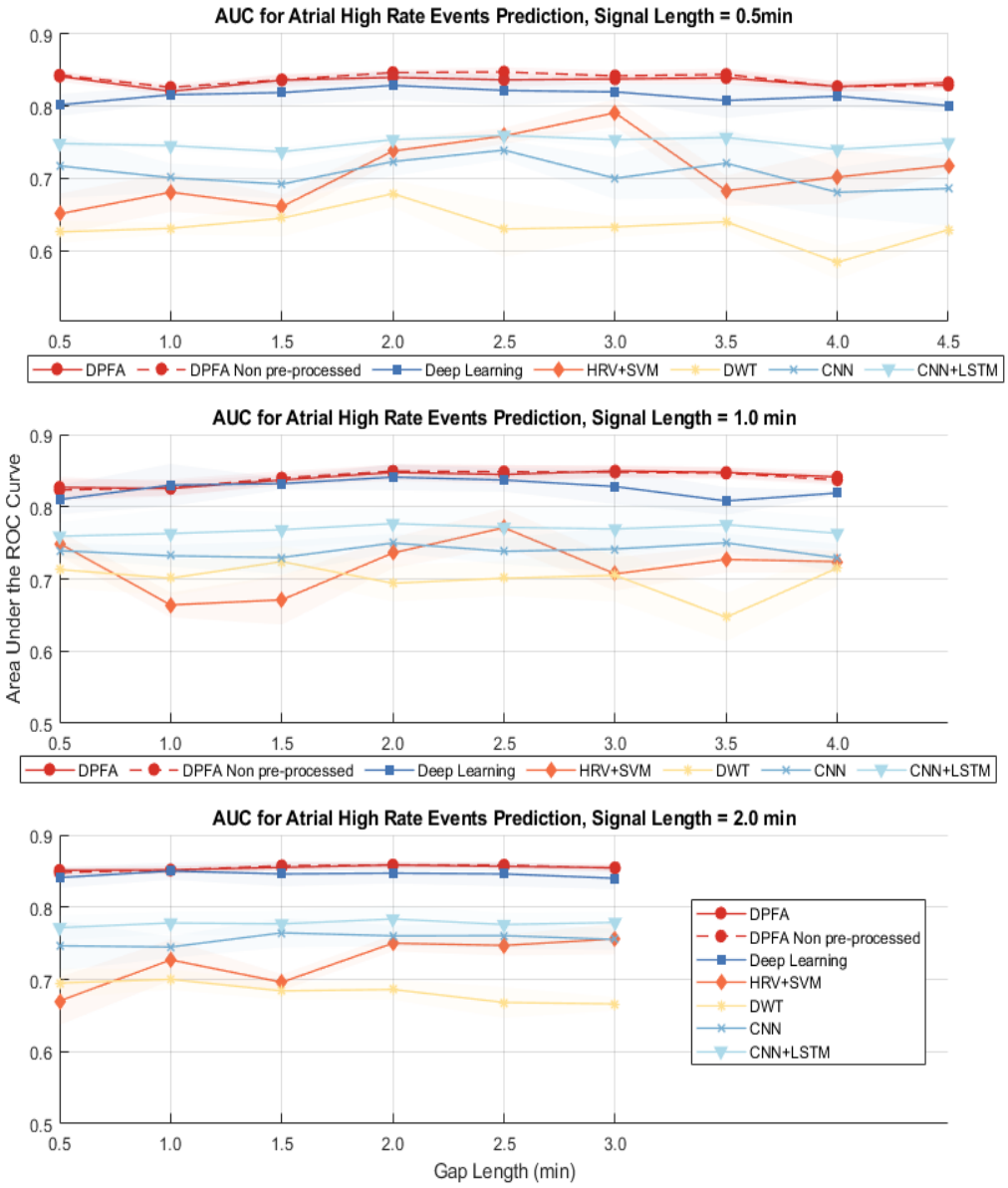


Figure 5.4: AFib prediction on hospital bedside database (DB2), A comparison of AUC for AHRE prediction for all models using various signal lengths and gap intervals.

Table 5.2: AFib prediction on hospital bedside database (DB2), a comparison of AUC for AHRE Prediction with different gap lengths for half-minute long signals.

Gap Length (min)	**DPFA**	**Deep Learning**		**HRV + SVM**		**DWT + PCA**		**CNN**		**CNN + LSTM**	
	Mean (STD)	Mean (STD)	p value	Mean (STD)	p value	Mean (STD)	p value	Mean (STD)	p value	Mean (STD)	p value
0.5	0.842 (0.007)	0.802 (0.015)	0.852	0.651 (0.028)	**0.004**	0.626 (0.015)	**0.001**	0.718 (0.044)	0.335	0.749 (0.004)	**0.000**
1.0	0.821 (0.006)	0.816 (0.012)	0.202	0.681 (0.027)	0.240	0.631 (0.008)	0.236	0.701 (0.020)	**0.001**	0.746 (0.005)	**0.000**
1.5	0.836 (0.009)	0.819 (0.017)	0.419	0.661 (0.016)	**0.002**	0.645 (0.025)	**0.023**	0.692 (0.020)	**0.000**	0.737 (0.006)	**0.000**
2.0	0.840 (0.007)	0.829 (0.020)	0.701	0.738 (0.015)	**0.007**	0.679 (0.018)	**0.000**	0.723 (0.019)	**0.015**	0.754 (0.007)	**0.000**
2.5	0.837 (0.010)	0.822 (0.020)	0.469	0.759 (0.011)	**0.001**	0.630 (0.038)	0.156	0.739 (0.013)	**0.027**	0.760 (0.004)	**0.000**
3.0	0.838 (0.006)	0.820 (0.012)	0.719	0.791 (0.019)	**0.003**	0.633 (0.014)	0.353	0.700 (0.029)	**0.020**	0.754 (0.011)	**0.000**
3.5	0.840 (0.009)	0.808 (0.024)	0.151	0.683 (0.021)	**0.000**	0.640 (0.011)	**0.000**	0.721 (0.048)	**0.028**	0.757 (0.008)	**0.000**
4.0	0.828 (0.007)	0.814 (0.015)	0.855	0.702 (0.037)	**0.010**	0.584 (0.023)	0.757	0.681 (0.034)	0.133	0.740 (0.007)	**0.003**
4.5	0.833 (0.008)	0.801 (0.009)	0.461	0.718 (0.009)	**0.006**	0.629 (0.009)	**0.061**	0.686 (0.051)	**0.005**	0.750 (0.007)	**0.000**

deep learning algorithm has no statistically significant difference in performance with respect to the McNemar test. However, the DPFA algorithm has 40.1% lower variance than deep learning. On the other hand, HRV, CNN and DWT have much lower AUC. The results for HRV with SVM were obtained after further excluding certain cases. Two AHRE cases and 20 of control cases were excluded from the test dataset because peak annotation could not be correctly applied to the signals. Although CNN with LSTM has lower AUC than the DPFA model, there is no significant difference in the McNemar test results for 1-minute and 2-minute long signals. CNN with LSTM performs better than the CNN model.

5.2.3 AFib Prediction with Data Collected from Portable Devices

5.2.3.1 Data

In this experiment, analysis for AFib prediction was performed with data collected from portable devices, (DB4 and DB5). Details about DB4 and DB5 are described in details in Sections 4.4 and 4.5. For DB4 and DB5, the AF events were anno-

Table 5.3: AFib prediction on hospital bedside database (DB2), a comparison of AUC for AHRE Prediction with different gap intervals for 1 minute long signals.

Gap Length (min)	**DPFA**	**Deep Learning**		**HRV + SVM**		**DWT + PCA**		**CNN**		**CNN + LSTM**	
	Mean (STD)	Mean (STD)	p value	Mean (STD)	p value	Mean (STD)	p value	Mean (STD)	p value	Mean (STD)	p value
0.5	0.827 (0.014)	0.810 (0.020)	0.751	0.749 (0.017)	**0.000**	0.713 (0.025)	**0.005**	0.739 (0.009)	0.056	0.759 (0.019)	0.780
1.0	0.825 (0.010)	0.830 (0.029)	0.104	0.664 (0.017)	**0.004**	0.701 (0.024)	0.232	0.732 (0.015)	0.345	0.763 (0.034)	0.803
1.5	0.837 (0.008)	0.832 (0.008)	0.105	0.671 (0.034)	0.083	0.724 (0.021)	0.080	0.730 (0.023)	0.081	0.768 (0.024)	0.830
2.0	0.847 (0.011)	0.841 (0.018)	0.265	0.736 (0.021)	**0.000**	0.694 (0.025)	**0.000**	0.750 (0.013)	0.101	0.777 (0.022)	0.822
2.5	0.845 (0.013)	0.837 (0.016)	0.297	0.771 (0.026)	0.190	0.701 (0.024)	0.080	0.738 (0.018)	0.562	0.771 (0.026)	0.902
3.0	0.850 (0.008)	0.828 (0.024)	0.385	0.707 (0.023)	**0.062**	0.705 (0.036)	0.176	0.742 (0.031)	**0.023**	0.769 (0.035)	0.377
3.5	0.848 (0.007)	0.808 (0.019)	0.720	0.727 (0.018)	**0.007**	0.647 (0.033)	**0.000**	0.750 (0.013)	0.217	0.775 (0.022)	0.620
4.0	0.841 (0.008)	0.819 (0.003)	0.294	0.724 (0.015)	**0.000**	0.715 (0.025)	**0.005**	0.729 (0.019)	**0.007**	0.763 (0.021)	0.366

Table 5.4: AFib prediction on hospital bedside database (DB2), a comparison of AUC for AHRE Prediction with different gap intervals for 2 minute long signals.

Gap Length (min)	**DPFA**	**Deep Learning**		**HRV + SVM**		**DWT + PCA**		**CNN**		**CNN + LSTM**	
	Mean (STD)	Mean (STD)	p value	Mean (STD)	p value	Mean (STD)	p value	Mean (STD)	p value	Mean (STD)	p value
0.5	0.851 (0.005)	0.841 (0.014)	0.219	0.670 (0.033)	0.174	0.695 (0.020)	**0.001**	0.746 (0.034)	0.071	0.772 (0.017)	0.658
1.0	0.852 (0.006)	0.850 (0.012)	0.102	0.727 (0.028)	**0.000**	0.700 (0.014)	**0.000**	0.745 (0.015)	0.096	0.778 (0.015)	0.812
1.5	0.855 (0.005)	0.846 (0.017)	0.961	0.696 (0.011)	**0.001**	0.684 (0.012)	**0.000**	0.764 (0.020)	0.602	0.777 (0.017)	0.949
2.0	0.858 (0.005)	0.847 (0.014)	0.208	0.750 (0.011)	**0.006**	0.686 (0.014)	**0.001**	0.760 (0.014)	0.175	0.784 (0.019)	0.828
2.5	0.856 (0.004)	0.846 (0.017)	0.449	0.747 (0.014)	**0.005**	0.668 (0.021)	**0.005**	0.761 (0.018)	0.211	0.776 (0.015)	0.715
3.0	0.855 (0.005)	0.840 (0.015)	0.434	0.756 (0.021)	**0.006**	0.666 (0.009)	**0.000**	0.755 (0.014)	0.190	0.779 (0.014)	0.957

tated using the FDA-cleared BeatLogic algorithm. Preventice BeatLogic platform, a comprehensive ECG annotation platform that leverages deep learning for beat and rhythm detection/classification *Teplitzky et al.* (2020).

5.2.3.2 Result and Discussion

Four sets of analyses were performed on the patient population:

- AFib only (participant number=9, event number=70)
- AFib with or without other arrhythmia (participant number=12, event number=303)
- all types of arrhythmia (participant number=55, event number=589)
- DB4 and DB5 (participant number=129, event number=654)

Details about the patient diagnosis are described in Table 4.6. The AUC for all arrhythmia group is around 0.71-0.73. The AUC for the AFib only group ranges from 0.65-0.69 which is lower than the all arrhythmia group. This might be due to the lower number of patients and episodes in the AFib only group. The AUC for the AFib with/without other arrhythmia group ranges from 0.61-0.65. The lower AUC in this group is due to these patients having AFib with other morbidities like VT or Vfib. The AUC for DB4 and DB5 combined has lower performance than DB4 only. We might need to further examine the events annotated in DB5, since it is the control group with healthy patients. The experiment results of these analysis are depicted in Table 5.5, Figures 5.5, 5.6, 5.7 and 5.8, which show the achieved AUC for different arrhythmia populations.

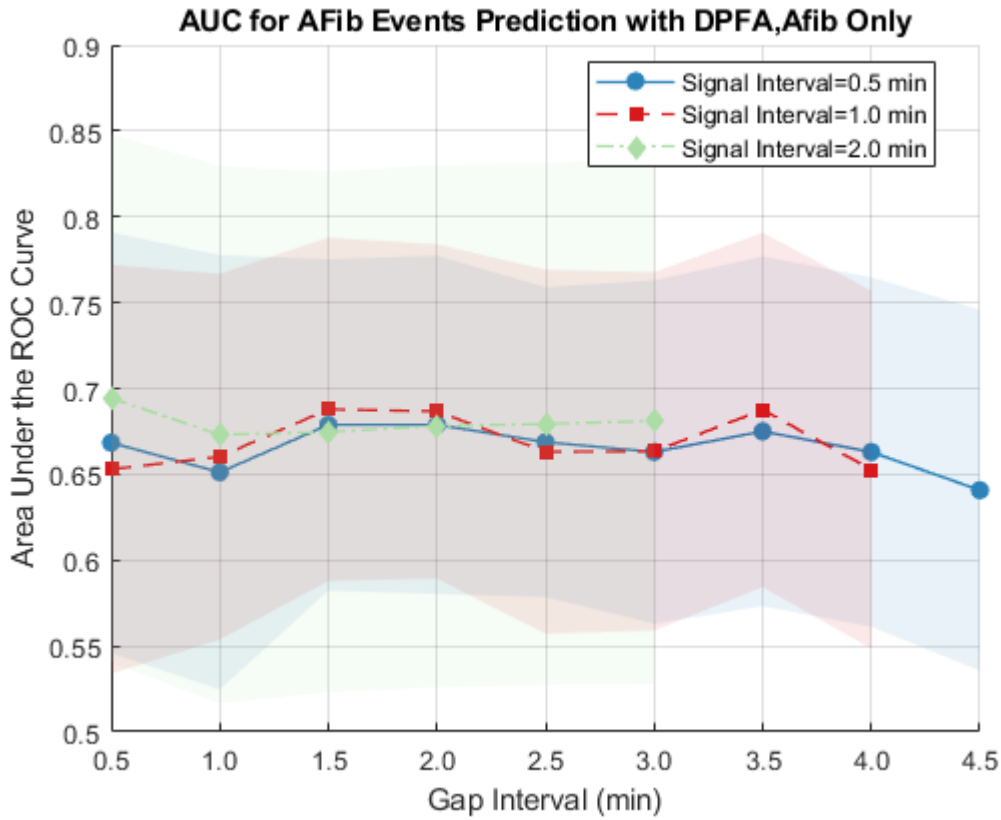


Figure 5.5: AFib predictions on data collected from portable devices (DB4, DB5), AUC for AFib events prediction, AFib only (n=9).

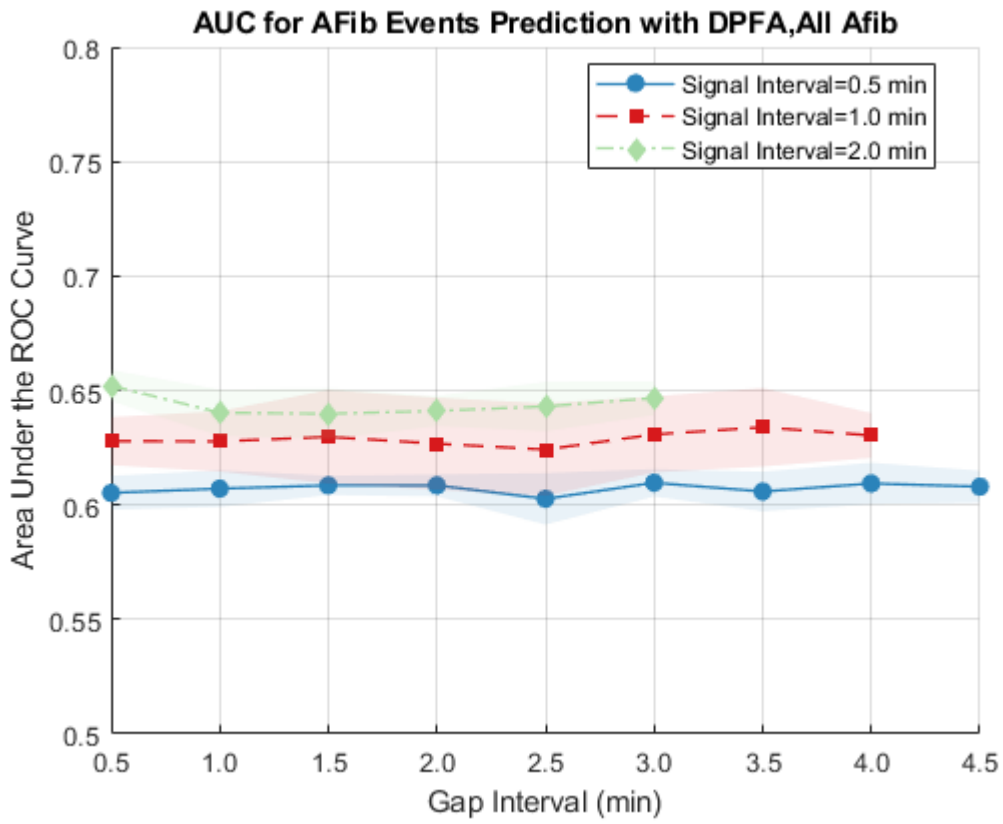


Figure 5.6: AFib prediction on data collected from portable devices (DB4, DB5), AUC for AFib events prediction, All AFib (n=12).

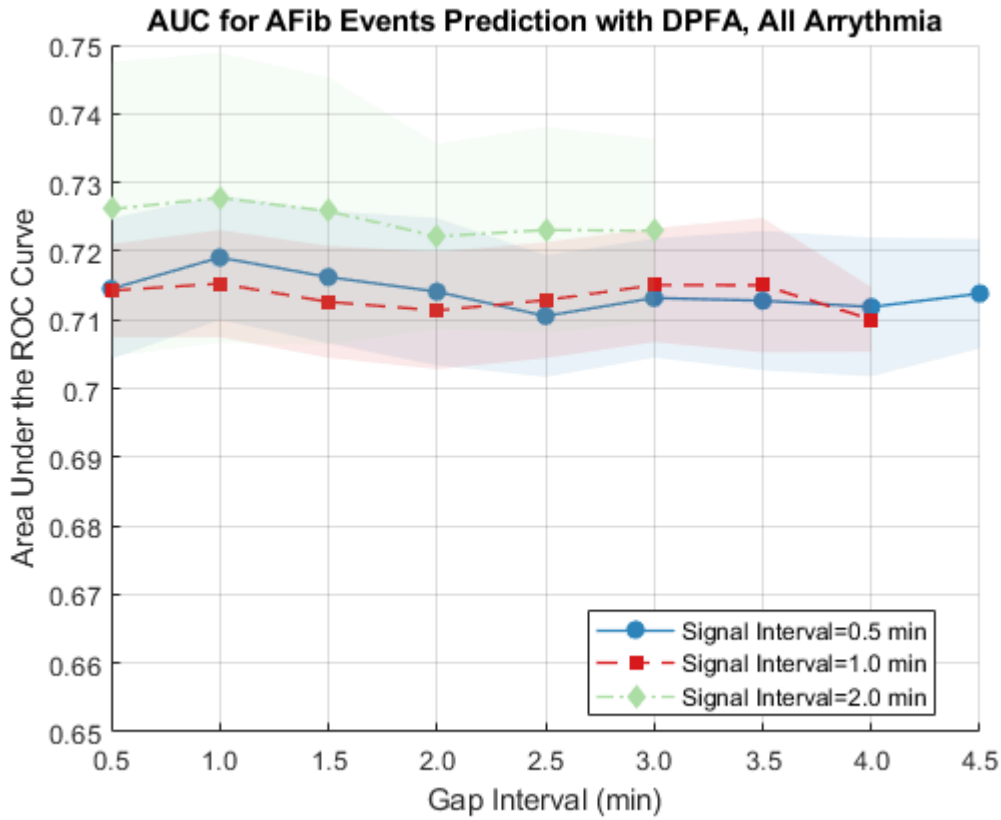


Figure 5.7: AFib predictions on data collected from portable devices (DB4, DB5), AUC for AFib events prediction, All Arrhythmia (n=55)

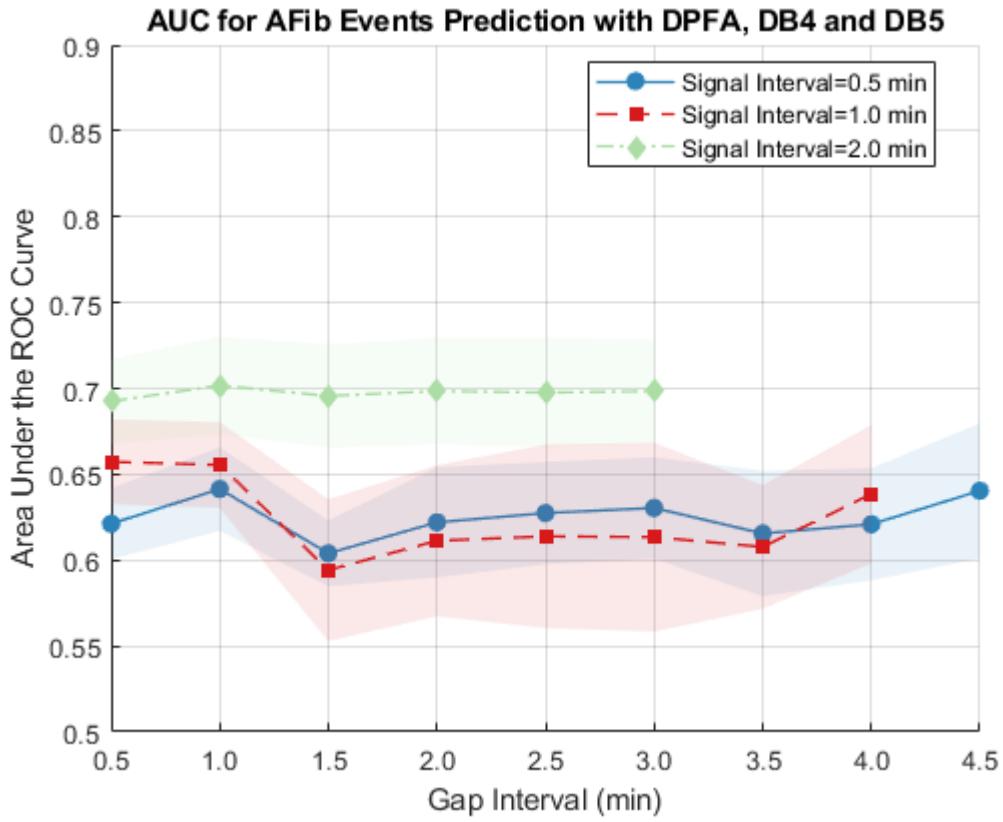


Figure 5.8: AFib predictions on data collected from portable devices (DB4, DB5), AUC for AFib events prediction, DB4 and DB5 (n=129).

Gap Interval (min)	Prediction Interval (min)	AFib only (n=9)		All AFib (n=12)		All Arrhythmia (n=55)		DB4 DB5 (n=129)	
		AUC Mean (STD)	ACC Mean (STD)	AUC Mean (STD)	ACC Mean (STD)	AUC Mean (STD)	ACC Mean (STD)	AUC Mean (STD)	ACC Mean (STD)
0.5	0.5	0.668 (0.122)	0.765 (0.035)	0.605 (0.007)	0.763 (0.013)	0.715 (0.010)	0.668 (0.026)	0.621 (0.021)	0.597 (0.010)
1	0.5	0.651 (0.127)	0.759 (0.016)	0.607 (0.008)	0.760 (0.005)	0.719 (0.009)	0.673 (0.022)	0.641 (0.024)	0.603 (0.030)
1.5	0.5	0.679 (0.096)	0.752 (0.012)	0.608 (0.004)	0.756 (0.006)	0.716 (0.010)	0.659 (0.031)	0.604 (0.019)	0.554 (0.025)
2	0.5	0.688 (0.098)	0.755 (0.010)	0.608 (0.005)	0.756 (0.002)	0.714 (0.011)	0.659 (0.029)	0.622 (0.032)	0.575 (0.025)
2.5	0.5	0.669 (0.090)	0.748 (0.006)	0.603 (0.011)	0.751 (0.004)	0.711 (0.009)	0.671 (0.004)	0.627 (0.030)	0.584 (0.013)
3	0.5	0.663 (0.099)	0.752 (0.006)	0.609 (0.006)	0.757 (0.005)	0.713 (0.009)	0.658 (0.010)	0.630 (0.029)	0.602 (0.055)
3.5	0.5	0.675 (0.102)	0.755 (0.010)	0.605 (0.009)	0.756 (0.004)	0.713 (0.010)	0.660 (0.013)	0.616 (0.036)	0.577 (0.020)
4	0.5	0.663 (0.102)	0.745 (0.000)	0.609 (0.009)	0.757 (0.003)	0.712 (0.010)	0.659 (0.014)	0.621 (0.033)	0.575 (0.053)
4.5	0.5	0.641 (0.105)	0.748 (0.006)	0.608 (0.007)	0.754 (0.001)	0.714 (0.008)	0.662 (0.008)	0.640 (0.039)	0.596 (0.044)
0.5	1	0.653 (0.119)	0.796 (0.031)	0.628 (0.011)	0.781 (0.010)	0.714 (0.008)	0.657 (0.030)	0.657 (0.025)	0.615 (0.006)
1	1	0.660 (0.106)	0.779 (0.036)	0.628 (0.013)	0.779 (0.009)	0.715 (0.007)	0.657 (0.031)	0.655 (0.025)	0.590 (0.025)
1.5	1	0.688 (0.099)	0.789 (0.026)	0.630 (0.020)	0.779 (0.009)	0.713 (0.008)	0.653 (0.027)	0.594 (0.041)	0.577 (0.026)
2	1	0.687 (0.100)	0.782 (0.016)	0.627 (0.020)	0.774 (0.005)	0.711 (0.008)	0.650 (0.024)	0.611 (0.044)	0.566 (0.045)

Gap Interval (min)	Prediction Interval (min)	AFib only (n=9)		All AFib (n=12)		All Arrhythmia (n=55)		DB4 DB5 (n=129)	
		AUC Mean (STD)	ACC Mean (STD)	AUC Mean (STD)	ACC Mean (STD)	AUC Mean (STD)	ACC Mean (STD)	AUC Mean (STD)	ACC Mean (STD)
2.5	1	0.663 (0.106)	0.772 (0.006)	0.624 (0.020)	0.774 (0.001)	0.713 (0.008)	0.642 (0.019)	0.614 (0.053)	0.581 (0.063)
3	1	0.663 (0.104)	0.776 (0.010)	0.631 (0.016)	0.775 (0.003)	0.715 (0.008)	0.650 (0.028)	0.613 (0.055)	0.568 (0.064)
3.5	1	0.687 (0.103)	0.782 (0.006)	0.634 (0.017)	0.778 (0.001)	0.715 (0.010)	0.646 (0.028)	0.608 (0.036)	0.556 (0.041)
4	1	0.652 (0.104)	0.772 (0.016)	0.630 (0.010)	0.774 (0.005)	0.710 (0.005)	0.645 (0.015)	0.638 (0.040)	0.578 (0.025)
0.5	2	0.694 (0.153)	0.786 (0.035)	0.652 (0.007)	0.793 (0.011)	0.726 (0.021)	0.682 (0.041)	0.693 (0.024)	0.624 (0.043)
1	2	0.673 (0.156)	0.779 (0.024)	0.641 (0.010)	0.790 (0.011)	0.728 (0.021)	0.672 (0.040)	0.702 (0.028)	0.648 (0.032)
1.5	2	0.675 (0.151)	0.765 (0.027)	0.640 (0.011)	0.786 (0.010)	0.726 (0.019)	0.674 (0.030)	0.696 (0.030)	0.626 (0.034)
2	2	0.678 (0.152)	0.765 (0.027)	0.641 (0.006)	0.783 (0.010)	0.722 (0.013)	0.672 (0.037)	0.699 (0.031)	0.634 (0.050)
2.5	2	0.679 (0.152)	0.765 (0.027)	0.643 (0.011)	0.784 (0.012)	0.723 (0.015)	0.678 (0.032)	0.698 (0.031)	0.645 (0.034)
3	2	0.682 (0.153)	0.759 (0.016)	0.647 (0.007)	0.784 (0.012)	0.723 (0.013)	0.671 (0.038)	0.698 (0.030)	0.628 (0.035)

Table 5.5: AFib predictions on portable devices dataset (DB4, DB5), results summary for AFib events prediction on DB4 and DB5.

5.2.4 AFib Prediction on In-vehicle Data

5.2.4.1 Data

Driving logs were collected from each study participant and were used to determine the portion of collected ECG signals that were recorded in-vehicle. In the analysis presented in this section, predictions were made on the annotated AFib events that occurred during the in-vehicle time periods only. There is a total of 35 annotated AFib episodes in all types of arrhythmia during driving from 7 patients. A total of 874 hours of in-vehicle time is recorded in DB4 with observing AFib evetns over 16 hours. A total of 532 hours of in-vehicle time is recorded in DB5, only 0.01 hours of AFib time in DB5 as described in Table 4.7.

5.2.4.2 Result

Figure 5.9 and Table 5.9 below shows the performance in AUC for all arrhythmia participants during driving. The performance of the algorithm on the in-vehicle only dataset in AUC is very similar to AUC for the entire group. The 2-minute prediction is lower than before; however the standard deviation for the 2-minute predictions are also very large. This could be due to the low number of patients and events in this analysis. We expect the performance to improve with more driving events.

5.3 Supraventricular Tachycardia

SVT detection and prediction experiments were performed on various databases. DB1 is constructed from publicly available databases and used for SVT detection experiment. In DB1, we have correctly identified 41/56 SVT events (0.73 sensitivity) and 97/100 (0.97 specificity). The AUC is 0.951 and F1 of 0.914. DB2 is a retrospective database from UMHS and used for SVT prediction experiment. There is a total of 139 subjects with SVT diagnosis, 38 of them have SVT episode met with

Table 5.6: AFib predictions on in-vehicle dataset (DB4, DB5), summary of results for AFib events prediction, in-vehicle only

Gap Interval (min)	**Prediction Interval (min)**	**AFib only (n=7)**	
		AUC Mean (STD)	ACC Mean (STD)
0.5	0.5	0.704 (0.011)	0.668 (0.026)
1	0.5	0.709 (0.014)	0.673 (0.022)
1.5	0.5	0.720 (0.014)	0.659 (0.031)
2	0.5	0.726 (0.013)	0.659 (0.029)
2.5	0.5	0.715 (0.020)	0.671 (0.004)
3	0.5	0.726 (0.024)	0.658 (0.010)
3.5	0.5	0.717 (0.028)	0.660 (0.013)
4	0.5	0.713 (0.004)	0.659 (0.014)
4.5	0.5	0.690 (0.016)	0.662 (0.008)
0.5	1	0.668 (0.040)	0.657 (0.030)
1	1	0.666 (0.039)	0.657 (0.031)
1.5	1	0.680 (0.005)	0.653 (0.027)
2	1	0.684 (0.006)	0.650 (0.024)
2.5	1	0.704 (0.007)	0.642 (0.019)
3	1	0.702 (0.012)	0.650 (0.028)
3.5	1	0.683 (0.005)	0.646 (0.028)
4	1	0.663 (0.033)	0.645 (0.015)
0.5	2	0.611 (0.126)	0.682 (0.041)
1	2	0.642 (0.128)	0.672 (0.040)
1.5	2	0.645 (0.129)	0.674 (0.030)
2	2	0.645 (0.122)	0.672 (0.037)
2.5	2	0.648 (0.112)	0.678 (0.032)
3	2	0.637 (0.115)	0.671 (0.038)

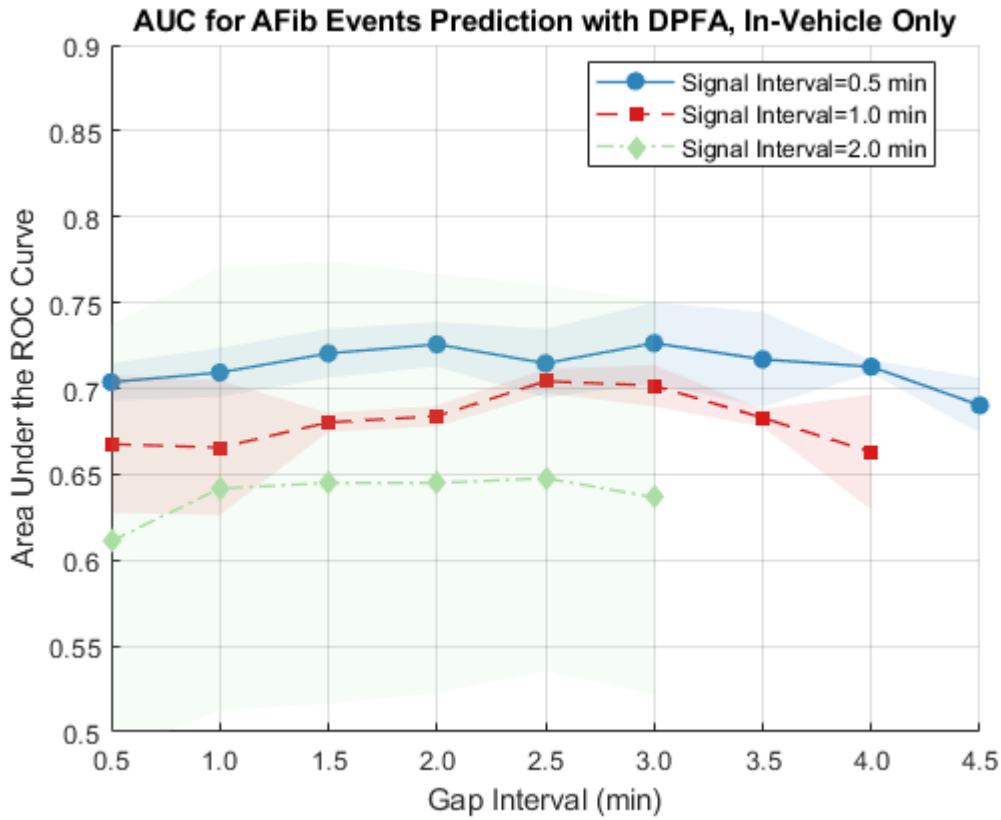


Figure 5.9: AFib predictions on in-vehicle dataset (DB4, DB5), AUC for AFib events prediction, in-vehicle only (n=7).

annotation criteria as in Section 3.3.2. DPFA algorithm is applied for SVT prediction with various prediction intervals and achieved an AUC greater than 0.75 for all prediction intervals in DB2. DB4 contains participants diagnosed with various types of arrhythmia while DB5 is the control group. There is a total of 27 SVT events that met annotation criteria from 12 patients from DB4 included in the analysis. Non-SVT events were randomly selected from all DB4 and DB5 episodes. The prediction AUC values are in the range of 0.79-0.87 for DB4 and 0.78-0.92 for DB4 and DB5 combined. SVT predictions while in-vehicle experiment was not performed due to the limited data size.

5.3.1 SVT Detection on Benchmark Datasets

5.3.1.1 Data

Three publicly available benchmark datasets in DB1 with SVT rhythm annotations were used in this study. The first dataset was the MIT-BIH Arrhythmia Database (mitdb) *Goldberger et al.* (2000), which contains 48 half-hour excerpts of two-channel ambulatory ECG recordings, obtained from 47 subjects enrolled in the study. The recordings were sampled at 360 per second. Two or more cardiologists have independently annotated each recording. The second database was the long-term AFib database (ltafdb) from physionet *Goldberger et al.* (2000); *Petrutiu et al.* (2007). This database includes 84 long-term ECG recordings of subjects with paroxysmal or sustained atrial fibrillation (AF). Each recording contains two simultaneously recorded ECG signals sampled at 128 Hz with duration around 24 to 25 hours. The third database was the MIT-BIH Malignant Ventricular Arrhythmia Database (vfdb) *Goldberger et al.* (2000) sampled at 250 Hz. All three databases contain some annotated SVT episodes other than AFib or atrial flutter.

Episodes of SVT were extracted from the three databases based on the rhythm annotations. These SVT episodes were then separated into 10-second-long segments.

We aim to detect these SVT events based only on the data contained within 10-second-long intervals.

There was a total of 667 10-second intervals. Among these episodes 611 were included in the training set and the remaining 56 episodes from a different patient cohort were reserved for the testing set. Additionally, 800 of the non-SVT episodes were randomly selected for the training set and 100 of non-SVT episodes were selected for the testing set.

5.3.1.2 Method

Heart Rate Variability with Machine Learning Techniques is described in the section below. Apart from the DPFA algorithm as described in Section 3.2, an algorithm based on heart rate variability (HRV) was used as comparison. This section will describe the algorithm.

During the signal pre-processing step, a fourth-order Butterworth bandpass filter with cutoff frequencies at 0.5 and 40 Hz was first applied to the raw ECG signal to remove noise, after which a double median filter with orders equal to 0.2 and 0.6 times the sampling frequency was applied to remove baseline wandering.

The signals from mitdb or vfdb were then down-sampled to 128 Hz to be consistent with ltafdb.

Episodes of SVT were extracted from the three databases based on the rhythm annotations. These SVT episodes were then separated into 10-second-long segments.

The R-peaks and QRS complex were detected by the Pan–Tompkins algorithm *Pan and Tompkins* (1985) *Sedghamiz* (2014). After identifying the R-peaks and QRS complex, we extracted a total of 16 features from the 10-second intervals as shown in Table 5.7.

Several machine learning algorithms including random forest (RF), support vector machine (SVM), and k-nearest neighbors (KNN) on the training dataset with 5-fold

Table 5.7: SVT detection, HRV features extracted for SVT detection on publicly available benchmark datasets (DB1)

HRV features	**Description**
ApEn	Approximate entropy, which measures the regularity and complexity of a time series
HR	Mean heart rate
SDHR	Standard Deviation of heart rate
RR	Mean RR interval
SDRR	Standard deviation of RR intervals
CV_RR	SDRR/meanRR intervals unitless scaled with factor 100
QS	Mean Q-S peak interval
SDQS	Standard deviation of Q-S peak interval
NN50	pairs of adjacent RR intervals differing that differ by more than 50 ms
pNN50	Percentage of successive RR intervals that differ by more than 50 ms
RMSSD	Root mean square of successive RR interval differences
SDSD	Standard deviation of differences between adjacent RR intervals
Poincaré SD1	Poincaré plot standard deviation perpendicular the line of identity
Poincaré SD2	Poincaré plot standard deviation along the line of identity
Lorentz OriginCount	Number of points (dRR(i-1),dRR(i)) within the radius normal sinus rhythm mask
Lorentz OriginCountRatio	Ratio between number of points (dRR(i-1),dRR(i)) within the radius normal sinus rhythm mask

cross validation. The cross validation procedure was performed on the training dataset only, enabling parameter optimization with respect to area under ROC curve (AUC) performance.

Among the three methods, RF seemed to have achieved the best result on the training dataset (Figure 5.10).

Among the 16 HRV features, mean QS interval length, Poincaré SD2, mean RR interval and mean HR had the highest importance based on the trained random forest algorithm on the training dataset. Importance was calculated as the weighted average

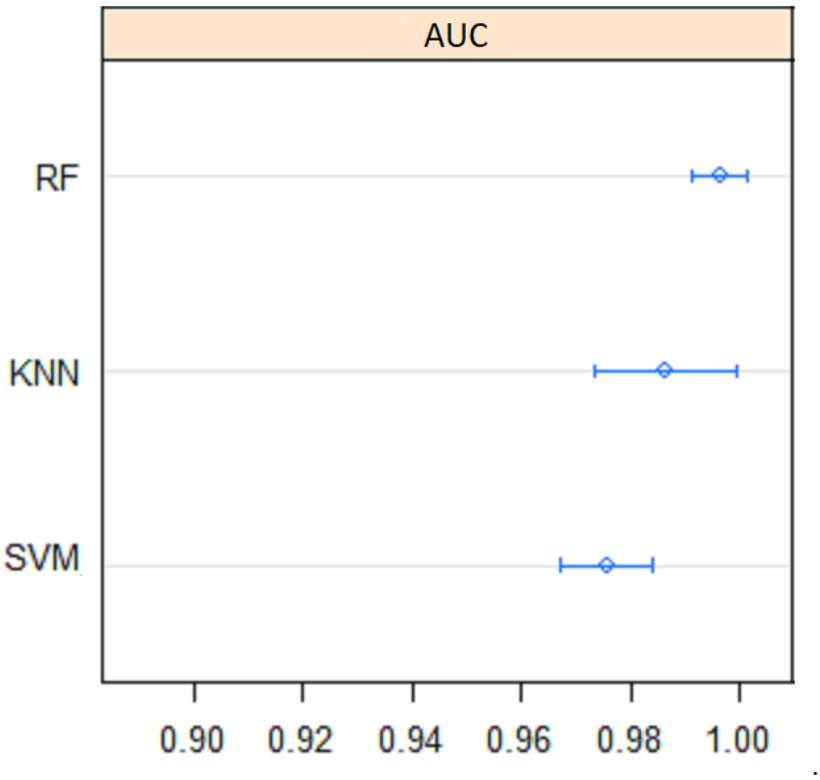


Figure 5.10: SVT detection on publicly available benchmark datasets, comparison of machine learning algorithms on training data

of the differences between the prediction accuracies on the out-of-bag portion or the entirety of the data. Regression was performed based on the weighted average of the differences of the respective MSEs instead.

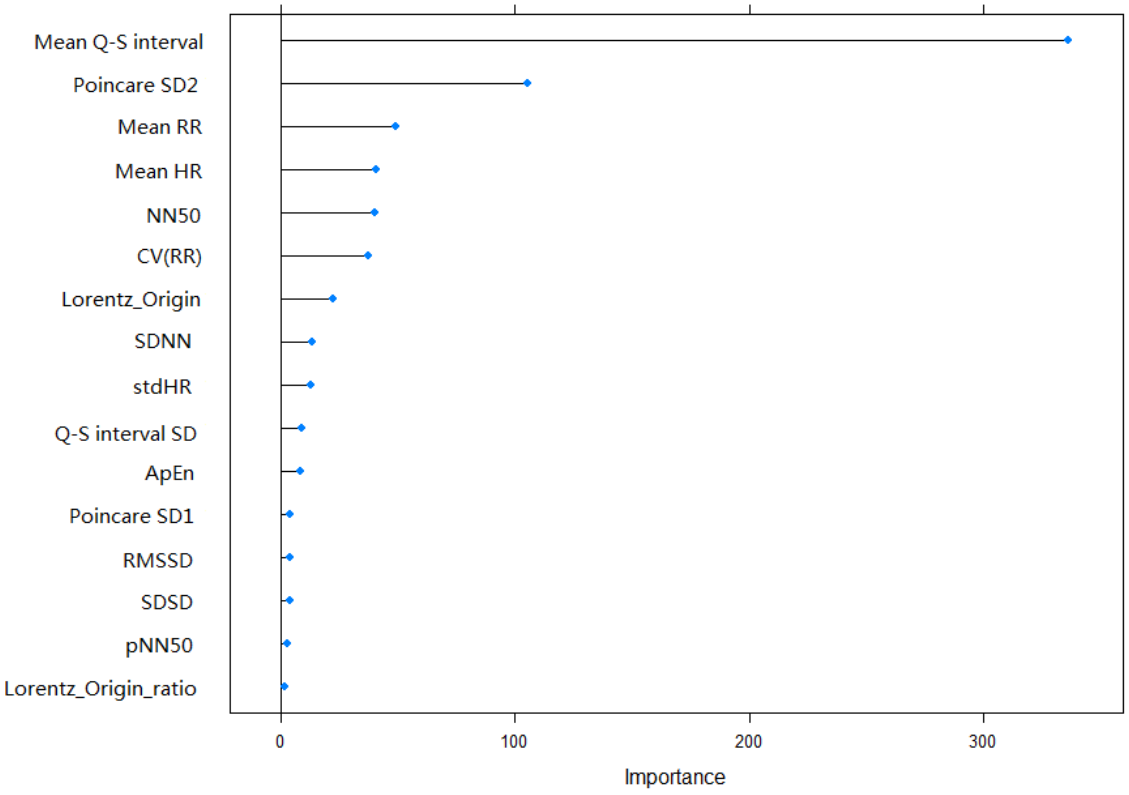
The RF model with the best AUC based on the training data was then applied on the testing data for evaluation.

DPFA model was applied for SVT detection with a binary encoding. Details of the algorithm can be found in Section 3.2.

5.3.1.3 Results

The result section begins by showing the algorithms' ability to detect annotated SVT episodes using HRV features and RF. Then we will also show the performance of the DPFA for SVT detection.

Figure 5.11: SVT detection on publicly available benchmark datasets, variable importance on training data



There was a total of 667 10-second intervals. 611 of these episodes were included in the training set and the remaining 56 episodes from a different patient cohort were reserved for the testing set. The algorithm correctly classified 51 out of 56 SVT episodes with sensitivity of 91.1% and 100 out of 100 non-SVT episodes with specificity of 100% in the testing set. The F1 score was 0.95 and the AUC was 0.995. The results are described in Table 5.8 and Figure 5.12.

With the DPFA algorithm, we have correctly identified 41/56 SVT events (0.73 sensitivity) and 97/100 (0.97 specificity). The AUC was 0.951 and F1 score was 0.914. Results are summarized in Table 5.9.

Table 5.8: SVT detection on publicly available benchmark datasets, confusion matrix for SVT Detection with HRV and RF

		Annotated Label		Total
		SVT	Non-SVT	
Prediction	SVT	51	0	51
	Non-SVT	5	100	105
	Total	56	100	156

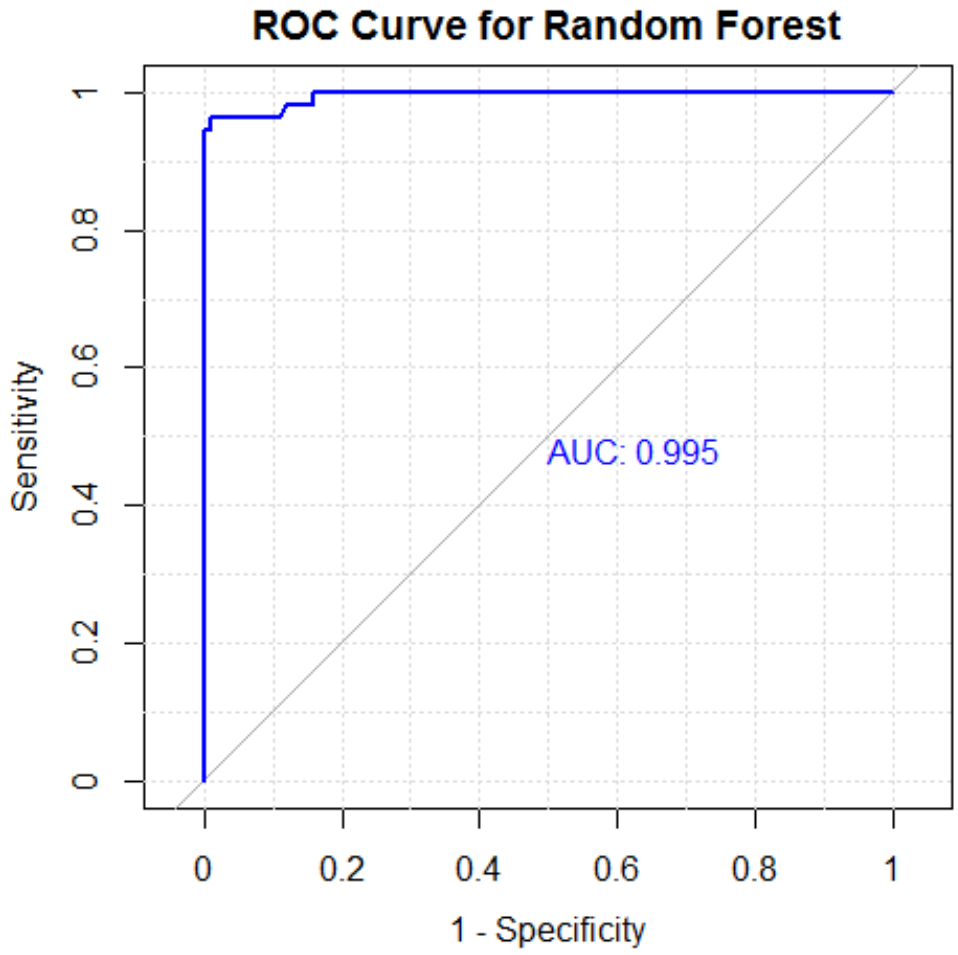


Figure 5.12: SVT detection on publicly available benchmark datasets, ROC curve for SVT Detection with HRV features and RF.

5.3.1.4 Discussion

In this part of the study, HRV features with RF successfully detected the SVT episodes with 0.995 AUC and 0.95 F1 score. The DPFA algorithm was able to detect

Table 5.9: SVT detection on publicly available benchmark datasets,,confusion Matrix for SVT Detection with DPFA

		Annotated Label		Total
		SVT	Non-SVT	
Prediction	SVT	41	3	44
	Non-SVT	15	97	112
	Total	56	100	156

the SVT episodes with high AUC of 0.95 and F1 of 0.91. The performance of the HRV features with RF is better than the DPFA algorithm. However, the DPFA algorithm does not require calculation of hand-crafted features and can be applied easily on real-time data collected from portable devices the DPFA algorithm is expected to perform better with a larger dataset with more complex morphologies in signals.

One limitation of the study is the lack of annotated samples with good quality labels. There are more databases with beat to beat annotations but not many of them has rhythm labels. Without these rhythm labels it is hard to classify different types of arrhythmia other than the beat types. We used the three publicly available databases with well annotated rhythms, however the number of SVT episodes we obtained from these records were still rather low. With limited amount of samples, the trained DPFA and RF model are more prone to problems such as generalizability and whether they can be directly applied to other datasets. Although in the testing set, we tested the algorithms on a different patient population, more ECG data with quality SVT labels are required to further validate the programs.

5.3.2 SVT Prediction on Hospital Bedside Dataset

5.3.2.1 Data

DB2 is used for SVT prediction. It is a retrospective database with ECG signals from Michigan Medicine cardiac patients with SVT. After excluding recordings with pacemakers, implantable cardioverter defibrillators (ICD), or ventricular assist devices

(VAD), there were 181 lead II ECG recordings for patients with SVT. Details about DB2 can be found in Section 4.3.

5.3.2.2 Method

After pre-processing and noise removal as described in Section 3.3.2, an automated annotation algorithm was applied to determine SVT episodes with the ECG signal. Time windows consisting of three R-R intervals were annotated as SVT if they satisfied all of the following four SVT annotation criteria. First, an automated annotation algorithm based on the heart rate-duration criteria line was used to annotate high heart rate events. The annotation criteria region is described in detail in Section 3.3.1. In the case of SVT, the episodes lie above the heart rate-duration criterion line passing through the points $(30, 150)$ and $(180, 100)$. Secondly, the difference between three consecutive R-R intervals should be less than 50 ms. Thirdly, there should be fewer than 4 P/T like waves within the window. Finally, the cross-correlations with the 500 P-wave templates extracted from the MIT-BIH P-wave database should be less than 0.85.

A total of 149 SVT events were annotated by the algorithm and a total of 755 non-SVT intervals were randomly selected from the non-SVT regions of the signals. A total of 119 (80%) of the SVT intervals were used in training with the remaining 30 intervals being held out for testing. Within the training data set, 5-fold cross validation was performed for parameter tuning (Figure 5.2). Training, cross validation, and testing sets/folds were partitioned on a participant level, meaning that signals from the same participant were only included in one set/fold so as to prevent overfitting.

The five models (each with four folds of training data) with the best parameters tuned during cross validation were then applied to the testing set.

Parameters are tuned with training and validation data as described in Section 3.2.2, the optimal parameters are listed in 5.1.

The parameters a_1, b_1, a_2, b_2 were all tuned in the training step and set at

$$\begin{cases} a_1 &= 0.4 \\ b_1 &= 0.8 \\ a_2 &= 0.04 \\ b_2 &= 0.06 \end{cases} \tag{5.1}$$

The performance of DPFA model were compared with five comparison methods described in Section 3.5. The McNemar test used a 1:5 ratio for the cost for the imbalance of data.

5.3.2.3 Results

Different combinations of signal intervals (i.e., $t_{signal} = 0.5, 1.0, 2.0$ minutes) and gap intervals (i.e., $t_{gap} = 0.5, 1.0, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5$ minutes) up to 5 minutes before the event were used for prediction. These prediction intervals were tested using the DPFA algorithm. Figure 5.13 shows a summary of the performance. The AUC is above 0.75 for all prediction intervals. As the time to the SVT event increases, the mean prediction performance gradually declines while variance increases. For the SVT DPFA, an average of 114×3 transition states were generated for the models using 0.5-minute long signals, 97×3 transitions states for the 1.0-minute long signal models, and 60×3 transition states for the 2.0-minute long signal models. For the non-SVT DPFA, an average of 114×3 transition states were generated for the models using 0.5-minute long signals, 95×3 transitions states for the 1.0-minute long signal models, and 56×3 transition states for the 2.0-minute long signal models.

The performance of the DPFA on raw ECG data was compared with its performance on pre-processed data (Figure 5.13). The AUC for all prediction intervals are almost the same with or without pre-processing and there is no statistically significant difference between these results. These results indicate that pre-processing does

not impact the performance of the DPFA algorithm.

The proposed DPFA algorithm has on average a 10.2% higher AUC than the deep learning method, 32.9% higher than DWT with PCA, 33.3% higher than CNN and 26.4% higher than CNN with LSTM for various prediction intervals. The HRV with SVM algorithm has comparable performance on SVT prediction, with a slightly lower average AUC compared to DPFA. Figure 5.14 summarizes the performance of DPFA together with the 5 alternative methods. The McNemar test was also performed to evaluate the performance of the alternative methods against the proposed DPFA algorithm. The McNemar test used a 1:5 ratio for the cost for the imbalance of data. The results are shown in Tables 5.10, 5.11, and 5.12. p-values representing the statistical significance of the different performances in terms of AUC ($< .05$) are shown in bold font in the referenced tables. There is no statistically significant difference between DPFA and the HRV method. DPFA is significantly better than the deep learning algorithm for 1-minute long prediction signals. DPFA is significantly better than the DWT, CNN and CNN with LSTM method for most of the prediction intervals. It is also worth noting that the performance of the DPFA algorithm with raw ECG as input is also higher than the deep learning, DWT, CNN, and CNN with LSTM methods.

5.3.2.4 Discussion

Many studies have focused on arrhythmia classification, but very few studies have attempted to predict the onset of an SVT event. As compared to other methods such as HRV, DWT, and deep learning, the method based on the DPFA algorithm has achieved around 0.8 AUC for all prediction intervals.

Remarkably, the proposed novel DPFA algorithm utilizes minimal pre-processing and does not require specific peak identification. The prediction results with raw ECG signals are almost the same as the results with pre-processing. This method represents

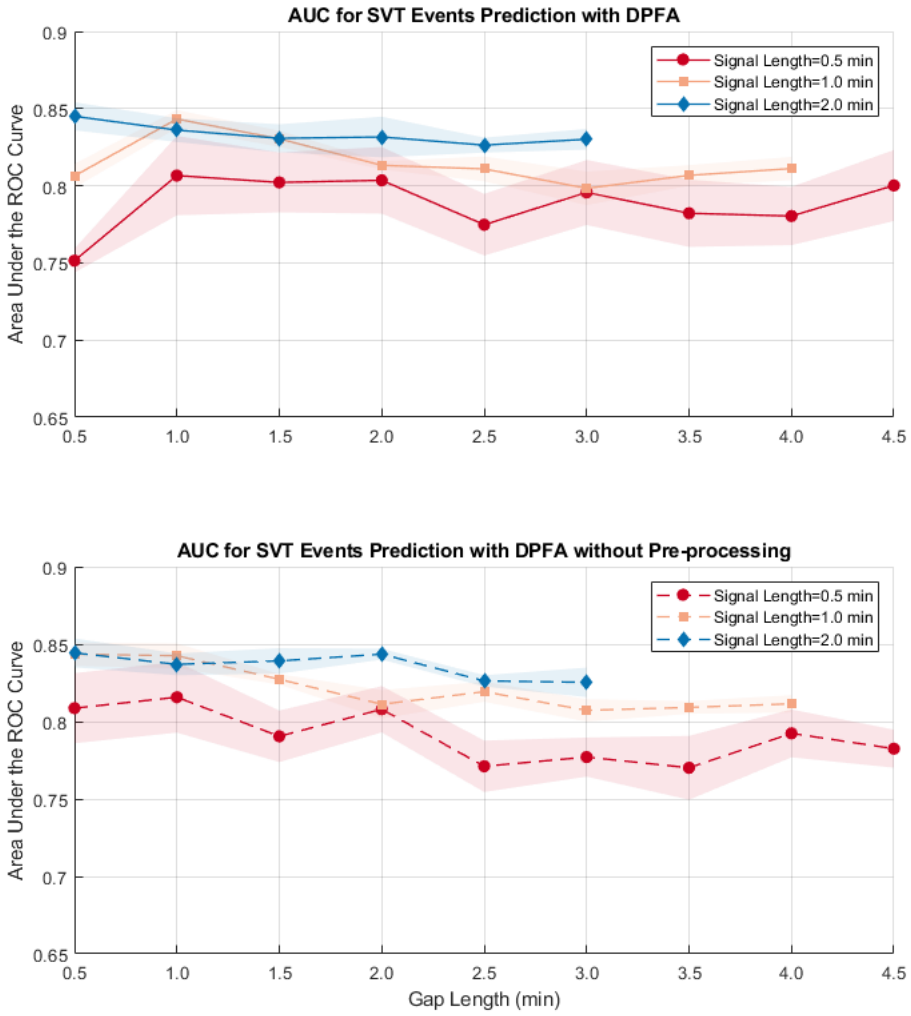


Figure 5.13: SVT predictions with DPFA, AUC for SVT prediction using the DPFA method with and without pre-processing for various signal lengths and gap intervals.

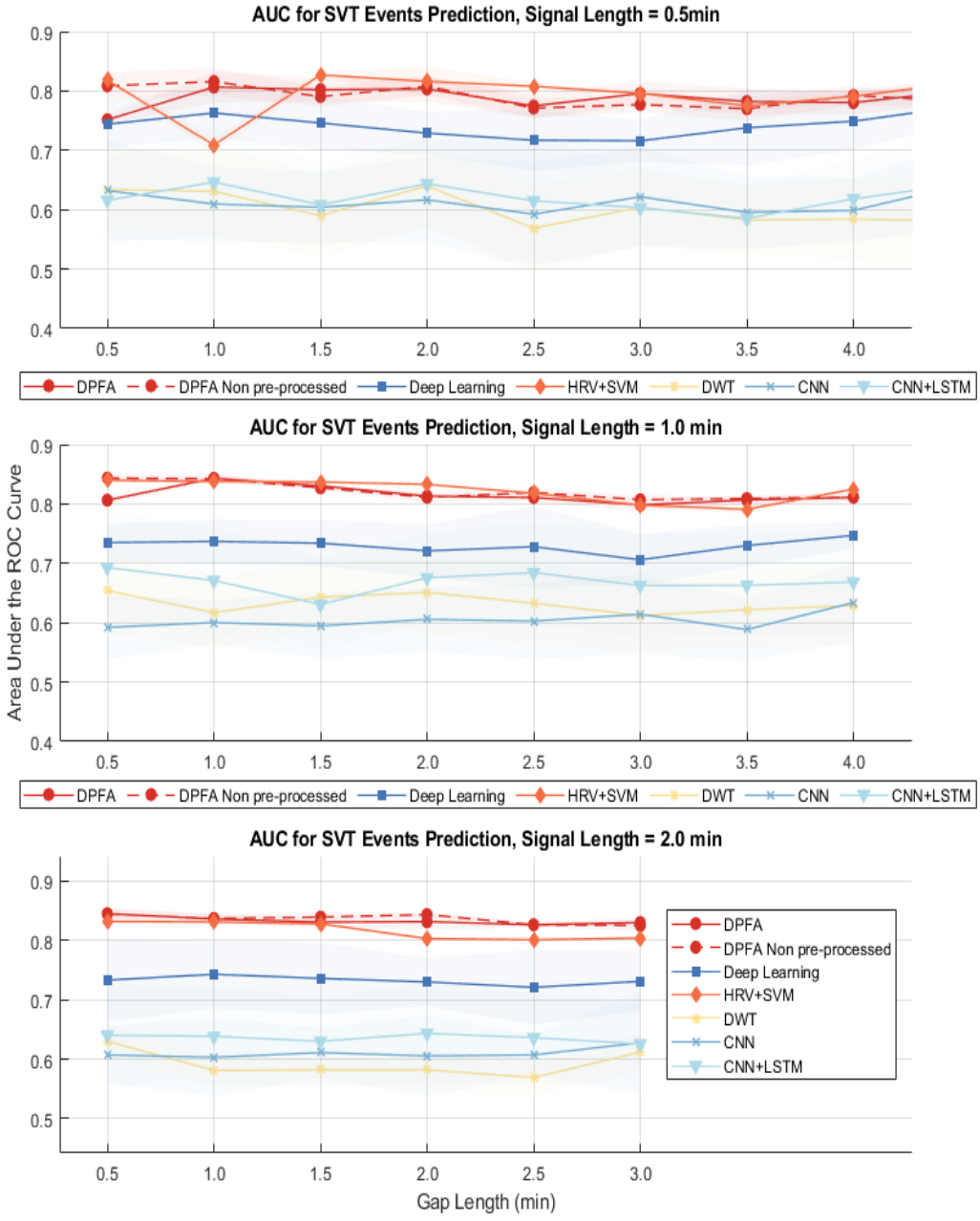


Figure 5.14: SVT prediction, AUC for SVT prediction for all models using various signal lengths and gap intervals.

Table 5.10: A comparison of AUC for SVT prediction with different gap intervals for half-minute long signals.

Gap Length (min)	**DPFA**	**Deep Learning**		**HRV + SVM**		**DWT + PCA**		**CNN**		**CNN + LSTM**	
	Mean (STD)	Mean (STD)	p value	Mean (STD)	p value	Mean (STD)	p value	Mean (STD)	p value	Mean (STD)	p value
0.5	0.751 (0.008)	0.744 (0.040)	0.090	0.818 (0.004)	0.101	0.634 (0.068)	0.812	0.634 (0.059)	0.897	0.616 (0.069)	0.084
1.0	0.807 (0.026)	0.763 (0.041)	0.542	0.709 (0.012)	0.126	0.631 (0.075)	0.122	0.664 (0.069)	**0.019**	0.646 (0.093)	**0.024**
1.5	0.802 (0.019)	0.746 (0.045)	0.679	0.827 (0.005)	0.869	0.590 (0.068)	0.117	0.664 (0.069)	**0.035**	0.608 (0.070)	**0.027**
2.0	0.804 (0.022)	0.729 (0.036)	0.404	0.816 (0.025)	0.858	0.640 (0.070)	0.407	0.653 (0.050)	0.172	0.644 (0.076)	**0.013**
2.5	0.775 (0.020)	0.717 (0.052)	0.130	0.808 (0.006)	0.805	0.569 (0.075)	0.119	0.630 (0.079)	**0.022**	0.615 (0.106)	**0.022**
3.0	0.796 (0.021)	0.716 (0.035)	0.227	0.796 (0.004)	0.686	0.604 (0.064)	**0.038**	0.646 (0.087)	0.082	0.603 (0.067)	**0.017**
3.5	0.782 (0.022)	0.738 (0.063)	0.364	0.775 (0.009)	0.789	0.583 (0.058)	**0.041**	0.650 (0.066)	0.113	0.585 (0.055)	**0.014**
4.0	0.780 (0.019)	0.749 (0.048)	0.718	0.791 (0.006)	0.591	0.584 (0.068)	**0.005**	0.653 (0.072)	0.126	0.618 (0.064)	**0.014**
4.5	0.800 (0.023)	0.774 (0.029)	0.716	0.814 (0.005)	0.353	0.581 (0.076)	**0.039**	0.684 (0.053)	0.124	0.643 (0.094)	**0.026**

a valuable alternative to traditional methods that require heavy pre-processing and feature extraction.

We have compared the DPFA algorithms with five other algorithms. The HRV with SVM method has good performance on SVT prediction when the dataset is relatively small. The performance drops when applied to AHRE prediction. Moreover, the automated annotation algorithm for both methods are directly related to the HRV features, which might have unknowingly affected the results. Several cases in AHRE prediction were excluded as a result of difficulties in peak annotation for the HRV with SVM method. The proposed DPFA method, on the other hand, does not require a peak annotation algorithm. As a consequence, the DPFA algorithm is more likely to perform better on other classification problems, where peak locations are not among the primary features.

Neural network based algorithms such as deep learning, CNN, and CNN with LSTM are similar to the DPFA algorithm in that no specific pre-processing is re-

Table 5.11: A comparison of AUC for SVT prediction with different gap intervals for 1 minute long signals.

Gap Length (min)	**DPFA**	**Deep Learning**		**HRV + SVM**		**DWT + PCA**		**CNN**		**CNN + LSTM**	
	Mean (STD)	Mean (STD)	p value	Mean (STD)	p value	Mean (STD)	p value	Mean (STD)	p value	Mean (STD)	p value
0.5	0.806 (0.008)	0.735 (0.031)	0.066	0.840 (0.004)	0.342	0.654 (0.072)	1.000	0.680 (0.068)	0.098	0.693 (0.094)	0.566
1.0	0.843 (0.006)	0.737 (0.037)	0.232	0.838 (0.004)	0.519	0.617 (0.061)	**0.010**	0.680 (0.069)	**0.005**	0.671 (0.090)	**0.004**
1.5	0.830 (0.005)	0.734 (0.039)	**0.030**	0.837 (0.002)	0.629	0.643 (0.054)	**0.020**	0.644 (0.076)	**0.007**	0.631 (0.077)	**0.002**
2.0	0.813 (0.003)	0.721 (0.044)	**0.038**	0.833 (0.003)	1.000	0.651 (0.070)	0.079	0.639 (0.075)	**0.005**	0.676 (0.074)	0.074
2.5	0.811 (0.008)	0.728 (0.069)	**0.015**	0.818 (0.002)	0.925	0.633 (0.072)	0.065	0.677 (0.098)	**0.002**	0.684 (0.084)	**0.053**
3.0	0.798 (0.011)	0.706 (0.043)	**0.041**	0.798 (0.003)	0.858	0.612 (0.060)	0.060	0.663 (0.069)	**0.024**	0.663 (0.088)	**0.017**
3.5	0.807 (0.007)	0.730 (0.035)	0.286	0.791 (0.007)	0.831	0.622 (0.066)	0.093	0.644 (0.064)	**0.009**	0.663 (0.084)	**0.022**
4.0	0.811 (0.008)	0.747 (0.022)	0.334	0.825 (0.002)	0.276	0.629 (0.065)	0.334	0.666 (0.086)	**0.049**	0.668 (0.090)	**0.016**

Table 5.12: A comparison of AUC for SVT prediction with different gap intervals for 2 minute long signals.

Gap Length (min)	**DPFA**	**Deep Learning**		**HRV + SVM**		**DWT + PCA**		**CNN**		**CNN + LSTM**	
	Mean (STD)	Mean (STD)	p value	Mean (STD)	p value	Mean (STD)	p value	Mean (STD)	p value	Mean (STD)	p value
0.5	0.845 (0.009)	0.733 (0.069)	0.924	0.832 (0.004)	0.917	0.630 (0.029)	**0.012**	0.663 (0.076)	**0.003**	0.640 (0.083)	**0.002**
1.0	0.836 (0.008)	0.743 (0.057)	0.916	0.831 (0.004)	0.575	0.581 (0.024)	**0.007**	0.664 (0.073)	**0.019**	0.639 (0.083)	**0.007**
1.5	0.831 (0.009)	0.736 (0.060)	0.925	0.828 (0.003)	0.812	0.582 (0.024)	**0.004**	0.659 (0.070)	**0.004**	0.630 (0.076)	**0.001**
2.0	0.832 (0.013)	0.730 (0.040)	0.276	0.803 (0.004)	0.644	0.582 (0.033)	**0.029**	0.673 (0.086)	**0.016**	0.643 (0.083)	**0.006**
2.5	0.826 (0.005)	0.721 (0.062)	0.753	0.801 (0.005)	0.680	0.569 (0.029)	**0.001**	0.654 (0.102)	**0.004**	0.636 (0.077)	**0.004**
3.0	0.830 (0.007)	0.731 (0.050)	0.488	0.804 (0.003)	0.741	0.612 (0.031)	**0.024**	0.668 (0.092)	**0.034**	0.626 (0.069)	**0.013**

quired. However, as shown in the experiments, they require a larger dataset for better performance. The performances of these three algorithms are worse than DPFA and HRV in SVT prediction due to the limited sample size.

There are several limitations of the study. The first limitation is the use of the automated annotation algorithm. To validate the performance of the algorithm, the positive cases were verified by a cardiologist, but not the negative cases. In a future study, we aim to have the dataset annotated by clinicians, which is the gold standard for annotations. The proposed DPFA model can be easily trained on other labeled datasets. Secondly, extremely noisy sections of the ECG signals have been removed from the beginning. These sections have been randomly checked by the clinicians and most of them are not suitable for classifications. For the purpose of annotations that rely on peak detection the signals were pre-processed and noisy sections were excluded. In the future, using events manually annotated by clinicians, these noisy sections can be included as an additional class for prediction instead of removing them. Thirdly, the database used in this study is imbalanced, with many more negative cases than arrhythmia cases. As a result, the costs and weights in the deep learning and SVM models were adjusted accordingly to make the prediction more balanced. In the future it is possible to employ additional data augmentation algorithms to increase the positive cases, which may benefit both the deep learning and DPFA methods. Lastly, the performance of the DPFA algorithm is still limited by the number of cases. Some participants had substantially more cases than others, thus inter-patient ECG signal variability may limit the performance of our algorithm as well as the other alternative algorithms.

5.3.3 SVT Prediction with Data Collected from Portable Devices

DB4 and DB5 contains data with portable devices nad were used for SVT prediction experiment. Details of DB4 and DB5 can be found in Section 4.4 and Section

4.5. In DB4 and DB5, the SVT events were annotated using the FDA-cleared BeatLogic algorithm. Preventice BeatLogic platform, a comprehensive ECG annotation platform that leverages deep learning for beat and rhythm detection/classification *Teplitzky et al.* (2020).

Two sets of analyses were performed on this patient population:

- DB4 (participant number=55, SVT event number=27)
- DB4 and DB5 (participant number=129, SVT event number=27)

Details about the patient diagnosis are described in Table 4.6. The AUC for all arrhythmia group in DB4 is around 0.79 - 0.87. The AUC for DB4 and DB5 combined has is in the range of 0.78 - 0.92. The experiment results of these analysis are depicted in Table 5.13 and Figure 5.15.

5.4 Ventricular Arrhythmia

Detection and prediction of ventricular arrhythmia events were performed on various databases. DB1 is constructed from publicly available databases and was used VA detection experiment. Details of DB1 can be found in Section 4.2. For VA detection in DB1, the algorithm correctly identified 625/662 (0.94 sensitivity) VA episodes and 600 /662 (0.91 specificity) non-VA episodes with an AUC of 0.97 and an F1-score of 0.93 for the 2 seconds long signals. VT prediction experiment was performed with DB2 using our own annotations. There are 77 reviewed episodes are obtained from only 23 patients. The AUC for VT prediction is around 0.75. We did not perform VT prediction in DB4/DB5 because we did not have any annotated events in these 2 databases.

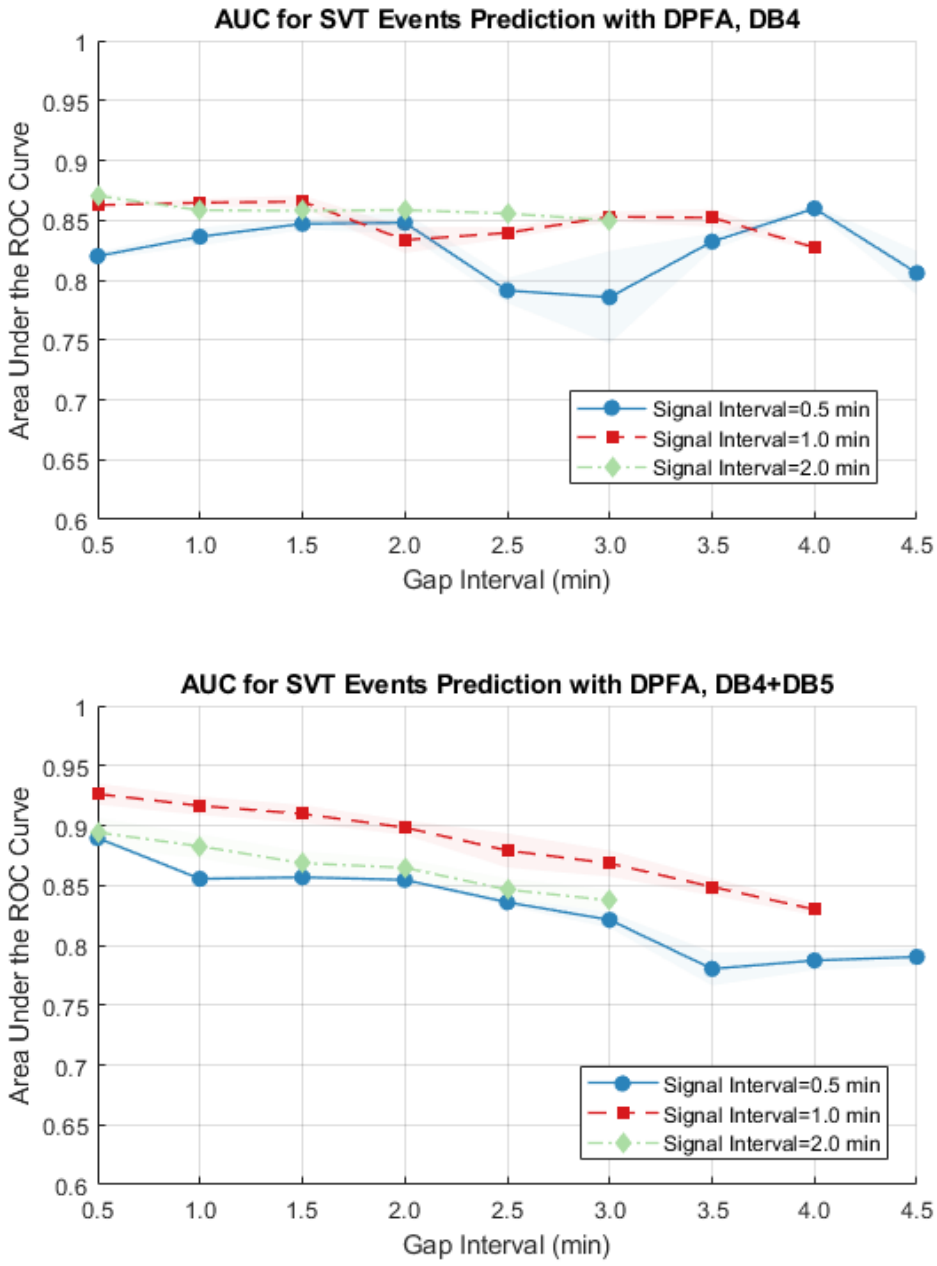


Figure 5.15: SVT prediction portable devices, AUC for SVT predictions.

Table 5.13: SVT prediction on data collected from portable devices, summary of results for SVT prediction with data from portable devices with DPFA.

Gap Interval (min)	**Prediction Interval (min)**	**All Arrhythmia (n=55)**		**DB4 DB5 (n=129)**	
		AUC Mean (STD)	ACC Mean (STD)	AUC Mean (STD)	ACC Mean (STD)
0.5	0.5	0.820(0.003)	0.974(0.001)	0.890(0.002)	0.904(0.019)
1	0.5	0.836(0.007)	0.959(0.007)	0.855(0.002)	0.875(0.007)
1.5	0.5	0.847(0.005)	0.946(0.003)	0.857(0.002)	0.909(0.003)
2	0.5	0.848(0.009)	0.966(0.002)	0.855(0.001)	0.882(0.007)
2.5	0.5	0.791(0.010)	0.977(0.010)	0.836(0.003)	0.885(0.011)
3	0.5	0.786(0.039)	0.979(0.002)	0.821(0.006)	0.838(0.019)
3.5	0.5	0.832(0.007)	0.953(0.025)	0.780(0.014)	0.864(0.017)
4	0.5	0.860(0.003)	0.976(0.004)	0.787(0.008)	0.862(0.004)
4.5	0.5	0.806(0.018)	0.974(0.003)	0.790(0.007)	0.842(0.048)
0.5	1	0.863(0.000)	0.961(0.014)	0.926(0.009)	0.934(0.005)
1	1	0.865(0.003)	0.970(0.004)	0.917(0.008)	0.936(0.007)
1.5	1	0.866(0.005)	0.959(0.007)	0.910(0.007)	0.925(0.008)
2	1	0.833(0.011)	0.972(0.001)	0.898(0.006)	0.927(0.005)
2.5	1	0.839(0.005)	0.974(0.017)	0.879(0.014)	0.921(0.010)
3	1	0.853(0.004)	0.965(0.007)	0.868(0.011)	0.902(0.025)
3.5	1	0.852(0.007)	0.948(0.014)	0.849(0.007)	0.908(0.015)
4	1	0.827(0.002)	0.975(0.001)	0.830(0.005)	0.906(0.007)
0.5	2	0.870(0.003)	0.968(0.016)	0.894(0.012)	0.919(0.002)
1	2	0.858(0.002)	0.973(0.001)	0.883(0.010)	0.903(0.001)
1.5	2	0.858(0.002)	0.960(0.014)	0.869(0.010)	0.912(0.012)
2	2	0.859(0.003)	0.952(0.006)	0.865(0.007)	0.877(0.034)
2.5	2	0.856(0.002)	0.955(0.013)	0.846(0.009)	0.878(0.026)
3	2	0.850(0.005)	0.949(0.008)	0.837(0.009)	0.868(0.041)

5.4.1 VA Detection on Benchmark Datasets

5.4.1.1 Data

Three publicly available data sets in DB1 with ventricular arrhythmia annotations were used in the evaluation of the proposed method. As all three databases have been examined by other algorithms, the proposed algorithm can be directly compared with the performance of existing algorithms. The MIT-BIH Arrhythmia Database (mitdb) contains 48 half-hour excerpts of two-channel ambulatory ECG recordings obtained from 47 human subjects*Goldberger et al.* (2000). The second database, the MIT-BIH Malignant Ventricular Arrhythmia Database (vfdb), contains 22 half-hour ECG recordings of subjects who experienced episodes of sustained VT, VFlutter, and VF *Goldberger et al.* (2000). Lastly, the Creighton University Ventricular Tachycardia Database (cudb) includes 35 8-minute long ECG recordings of human subjects who experienced episodes of sustained VT, ventricular flutter, or VF *Nolle et al.* (1986). The recordings from mitdb are sampled at 360 Hz, while those from vfdb and cudb are sampled at 250 Hz.

5.4.1.2 Method

The vfdb and cudb databases are sampled at 250 Hz, while the mitdb database is sampled at 360 Hz. Therefore, signals from mitdb were first re-sampled to 250 Hz. VA episodes including VT, ventricular flutter, and VF were extracted from the re-sampled signals according to the ground-truth annotations. VA episodes from a total of 28 patients were extracted. The signals from 22 (80%) patients were included in the training data set and the remaining 6 patients were grouped into the testing data set.

Signals were segmented into both 5-second and 2-second long episodes. A total of 1409 5-second episodes were in the training data set and 261 episodes were in the

testing data set. For the 2-second analysis, a total of 3667 episodes were in the training data set and 662 episodes were in the testing dataset. Non-VA data was partitioned in a similar way to ensure that the testing data set had a patient population disjoint from the training data set.

Five-fold cross validation was performed at the patient level for parameter tuning and to prevent over-fitting. The entire training data set was equally partitioned into 5 parts on the patient level. The first 4 parts were used as training data for generating the Markov models and the last part was the validation data set. This process was repeated five times. Average results for classification from all five experiments were used to assess the performance. The sensitivity, specificity, F1-score, and AUC were computed based on the training data set. The Markov model with the highest AUC over the training data set was then applied to the testing data set to obtain the final results.

5.4.1.3 Results

Performance was evaluated using 5-fold cross validation. Within the training data set, the best result had an AUC of 0.92 ± 0.05 and F1 score of 0.89 ± 0.03 for 5-second long episodes and AUC of 0.93 ± 0.03 and F1 score of 0.88 ± 0.04 for 2-second long episodes.

The parameters in the model with highest AUC in the training data were then applied to the testing data set. This testing set had a patient cohort separate from the training data.

When evaluated over the testing data set, the proposed algorithm correctly identified 243 of 261 (0.93 sensitivity) VA episodes and 227 of 261 (0.87 specificity) non-VA episodes for 5-second long signals. The AUC was 0.96 and the F1-score was 0.91 (Figure 5.16). For two-second long signals, the algorithm correctly identified 625 of 662 (0.94 sensitivity) VA episodes and 600 of 662 (0.91 specificity) non-VA episodes

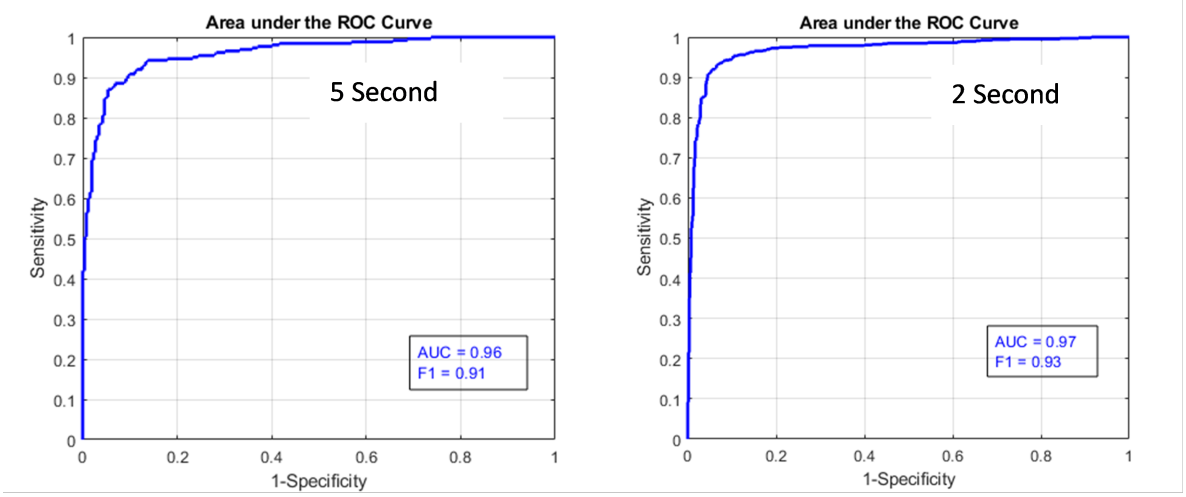


Figure 5.16: Ventricular arrhythmia detection on publicly available benchmark datasets, ROC curves for VA detection (5 seconds) and VA detection (2 seconds)

(Table 5.14) with an AUC of 0.97 and an F1-score of 0.93 (Figure 5.16).

Table 5.14: Ventricular arrhythmia detection on publicly available benchmark datasets, confusion matrix for VA vs Non-VA with DPFA.

	Annotation 5 seconds			**Annotation 2 Seconds**		
Prediction	**VA**	**Non-VA**	**Total**	**VA**	**Non-VA**	**Total**
VA	243	34	277	625	62	687
Non-VA	18	227	245	37	600	637
Total	261	261	552	662	662	1324

5.4.1.4 Discussion

The presented DPFA algorithm did not require any pre-processing or peak annotation of the signals. It was able to detect 5-second long VA episodes with a high AUC of 0.96 and F1-score of 0.91, and with an AUC of 0.97 and F1-score of 0.93 using 2-second long signals. Table 5.15 provides a performance comparison between our algorithm and other algorithms. Note that the precise conditions and set-up of these studies are not exactly the same.

There are multiple advantages of the proposed method over traditional feature extraction with machine learning algorithm based approaches. First, the proposed

Table 5.15: Ventricular arrhythmia detection on publicly available benchmark datasets, results comparison to other methods.

Author, (Year)	Data	Classification	Length(s)	Algorithm	Performance
Jekova, 2004	AHAVF vfdb	Non-shockable vs. Shockable (VT >180 bpm + VF)	10	preprocess, criteria based, bandpass digital filtration	Sen=0.96 Spec=0.94
Alonso, 2014	mitbih cudb vfdb	VF vs. Non-VF	8	preprocess, feature extraction, SVM	Sen=0.92 Spec=0.97 AUC=0.987
Tripathy, 2016	mitbih cudb vfdb	Non-shockable vs. Shockable (VF/VT)	5	variational mode decomposition, feature extraction, random forest	Sen=0.96 Spec=0.98 AUC=0.97
Acharya,2018	mitbih cudb vfdb	Non-shockable vs. Shockable (VFL, VT, VF)	2	CNN	Sen=0.95 Spec=0.91
DPFA	mitbih cudb vfdb	VFL/VT/VF vs. all others	2	DPFA with Automatically Generated States	Sen=0.94 Spec=0.91 AUC=0.97

algorithm did not rely on the efficacy of any pre-processing algorithms to remove noise and baseline wandering. Instead, the encoding algorithm uses filters and normalizes the signals. Most pre-processing algorithms require prior knowledge of noisy signals to build the best thresholds for filtering purpose. Our algorithm, by utilizing a word distribution algorithm, is more adaptive to different types of noisy signals. The second advantage was that the proposed algorithm did not require usage of an ECG peak annotation algorithm. Thirdly, this algorithm did not need extensive prior knowledge of the signals in order to build and extract features. Furthermore, even though recent novel algorithms using CNNs do not require pre-processing or feature extraction either, they still require longer training times and larger computational resources. Finally, the proposed model was flexible, robust, and adaptable to other types of arrhythmia like atrial fibrillation *Li et al.* (2018) and supraventricular tachycardia. It has potential applications to portable devices that could perform detection in real-time.

One limitation of the model was that it required a large number of good quality annotated signals for training. However, for severe types of arrhythmia with low prevalence such as VF, the number of annotated signals are limited.

Future work will utilize the proposed method to predict the onset of VA events

several minutes in advance with real-time data from portable ECG devices.

5.4.2 VT Prediction on Hospital Bedside Dataset

5.4.2.1 Data

Waveforms in DB2 are used for VT predictions. After excluding recordings with pacemakers, implantable cardioverter defibrillators (ICD) and ventricular assist devices (VAD), there are 152 lead II ECG waveforms in the database from patients diagnosed with VT.

5.4.2.2 Method

Each ECG recording could last up to several hours, we used 2 criteria to first filter out the intervals that might contain VT episodes. The first criterion is the heart rate should be above 100 bpm and the second criterion is average width corresponding to half length of the peaks should be wider than 40 ms. The second criterion is defined because VT episodes have wide QRS complex ($>$ 120 ms), with half length of the peaks $>$ 40 ms, the width QRS complex will likely to be at least $>$ 80 ms. We make this criterion lower than the 120 ms, to make sure that we capture all the possible VT episodes for manual review. Details about annotation algorithms in 3.3.1.

The last step of VT annotation requires manual review of the episodes which satisfied the two criteria. There is a total of 77 VT episodes after review, 59 of them were included in the training dataset and the remaining 18 episodes were included in the testing dataset. Five-fold cross validation was performed on training data on participant level for parameter tuning. The parameters with highest sensitivity was then applied on the testing set.

Table 5.16: VT prediction on hospital bedside dataset (DB2), result summary with DPFA.

Gap Interval (min)	Prediction Interval (min)	AUC	F1	Sensitivity	Specificity	Acc
0.5	2	0.78	0.55	0.67	0.84	0.84
1	2	0.77	0.50	0.56	0.84	0.79
1.5	2	0.76	0.50	0.50	0.85	0.72
2	2	0.75	0.52	0.33	0.86	0.76
2.5	2	0.75	0.48	0.44	0.85	0.72
3	2	0.76	0.52	0.56	0.85	0.81

5.4.2.3 Results and Discussion

The AUC for VT prediction is around 0.75, results on DB2 are shown in Table 5.16. The VT prediction results which are not as good as other types of arrhythmias like AFib or SVT. The major limitation of VT prediction is the limited number of data compares to other arrhythmia types. There are only 77 reviewed episodes obtained from 23 patients.

5.5 Bradycardia

5.5.1 Bradycardia Detection on Benchmark Datasets

Comparing to other types of arrhythmias, bradycardia does not cause as severe symptoms and have less research interests. Thus there are limited publicly available databases built for bradycardia events. Bradycardia events in DB1 is limited and experiments were not performed due to the small sample size.

5.5.2 Bradycardia Prediction on Hospital Bedside Dataset

Bradycardia in adults may cause complications in many kinds of chronic heart disease. On the other hand, bradycardia is also linked to various kinds of emergencies in situations such as when driving an automobile. It is thus crucial to monitor and

make automated urgent detection, preferably even to predict these bradycardia events in such situations. The aim of the study is to predict the onset of bradycardia events in patients with congestive heart failure (CHF), with the purpose to inform the development of an in-vehicle cardiac monitoring system.

Arrhythmia is a condition of improper beating of the heart, whether too fast or too slow or irregularly, which occurs when the electrical impulses that coordinate the heart beats are not working properly. Among various types of arrhythmia, tachycardia refers to a fast heart beat generally with a resting heart rate greater than 100 beats-per-minute (BPM), whereas bradycardia, on the other hand, refers to a slow heart beat commonly defined as having a heart rate of less than 60 BPM *Kusumoto et al.* (2019). Despite the general definition, a slow heart rate does not always indicate an underlying disease, since there is considerable variation in the resting heart rate among the healthy, asymptomatic population *Mangrum and DiMarco* (2000) such as the example of people with a family history of slow heart beats, or athletes and physically active individuals. However, bradycardia can cause inadequate blood flow to tissues and organs leading to symptoms such as fatigue, dizziness or heart failure *Mangrum and DiMarco* (2000). In more severe cases, certain symptomatic bradycardia could be caused by a serious underlying medical condition known as congestive heart failure (CHF), which is manifested by signs and symptoms of impaired cardiac ventricular performance. The ten-year mortality rate for CHF is 54.4% for men and 23.8% for women for the population with 25-74 years of age *Schocken et al.* (1992). Damage to the heart from CHF and other heart problems may affect the heart's electrical system, making the heart pump less efficiently and effectively and causing bradycardia. Upto 5 million patients experience CHF in United States and sudden cardiac death remains a significant problem in this population. Although most sudden deaths in CHF can be attributed to ventricular arrhythmia, bradycardia is another prominent cause and accounts for 42% of sudden deaths in hospital *Sanders et al.* (2004). In order to

prevent the negative effects of bradycardia for patients with life threatening cardiac disease such as CHF, it is crucial to not only make early detection, but also to predict the onset of bradycardia events.

Many studies on bradycardia prediction have focused on the detection or prediction of severe bradycardia including apnea bradycardia in premature newborns. Various techniques have employed statistical methods and machine learning methods such as nonparametric methods *Das et al.* (2019) and point process modeling *Gee et al.* (2016) to extract statistical features, deep sequential auto-encoders and online clustering *Hosseini and Sarrafzadeh* (2019), Kalman filter *Ghahjaverestan et al.* (2015b), principal components analysis (PCA) and hierarchical ascending classification (HAC) for clustering and classification *Pravisani et al.* (2003), hidden Markov model *Ghahjaverestan et al.* (2015a, 2021); *Altuve et al.* (2011); *Sadoughi et al.* (2021), system identification-based approach *Bakshi et al.* (2020) and fusion of different algorithms *Cruz et al.* (2006). However, when it comes to bradycardia prediction in adults there is limited research efforts, especially in patients with severe cardiac disease such as heart failure. Therefore, this study aims to achieve two goals: from a methodology perspective, to apply a new approach based on DPFA into the study of bradycardia prediction; from a medical perspective, to fill the gap in adult bradycardia prediction by developing such an algorithm for patients with underlying cardiac diseases. For those who are not familiar with the definition, a DPFA is an automaton with a finite number of states on a pre-determined finite alphabet consisting of possible deterministic transitions, such that at each time the transition from one state to the next is determined by a probability distribution on the alphabet depending only on the current state. For further details please refer to *Li et al.* (2019, 2021), algorithm related to bradycardia prediction is listed in Section 3.3.1.

In this part of the study, a DPFA based algorithm is proposed to monitor and predict bradycardia events, with the goal to create a real-time, prospective system

that can help clinical practice and prevent negative outcome due to sudden drop in heart rates in CHF patients.

5.5.2.1 Data

The analysis is based on lead II electrocardiogram (ECG) recordings. The ECG signals were extracted from a retrospective database from Michigan Medicine cardiac patients with bradycardia diagnosis. Recordings from patients with pacemakers, implantable cardioverter defibrillators (ICD), or ventricular assist devices (VAD) were further excluded. The recordings were sampled at 240 Hz.

There was a total of 2156 lead II ECG recordings from 1251 bradycardia patients. Among these 2156 recordings, 287 were from 173 CHF patients. After extracting bradycardia events with annotation criteria, a total of 95 bradycardia intervals from 47 patients. Among these events, a total of 21 bradycardia annotated events from 11 patients did not take beta-blockers. The patients without beta-blockers were separated out to test the severe bradycardia cases that were not related to beta-blockers. Beta-blockers are commonly used in treatment for heart failure and arrhythmia. They are thought to work by blocking the deleterious effects of adrenergic receptor stimulation.This results in a slowing of the heart rate and reduces the force at which blood is pumped around your body. Studies suggest that treatment to reduce resting heart rate in patients with heart failure to < 65 bpm is associated with a better prognosis. However, risk must increase with very low heart rates *Cullington et al.* (2012). Lead II ECG was used because it shows the p-wave and is most commonly used to record the rhythm strip. As a result, analyses based on hand-crafted clinical features mostly use lead II ECG signals. Table 5.17 summarizes demographics of the 47 patients included in the study and the sub-group with 11 patients with no beta-blockers. Among the 47 patients, 57% of them were males with an average of age of 64.7. All of them were cardiac arrhythmia patients with congestive heart failure. Among them 30%

Table 5.17: Bradycardia prediction on hospital bedside dataset(DB2), demographic table.

Demographics	**Value**	**Frequency (%)**	
		Total (n=47)	**Without Beta (n=11)**
Age	<= 50	11 (23%)	3 (27%)
	51-60	7 (15%)	1 (9%)
	61-70	10 (21%)	3 (27%)
	71-80	13 (28%)	3 (27%)
	>80	6 (13%)	1 (9%)
Gender	Female	20 (43%)	3 (27%)
	Male	27 (57%)	8 (73%)
Cardiac Arrhythmia	Yes	47 (100%)	11 (100%)
	No	0 (0%)	0 (0%)
Chronic Pulmonary Disease	Yes	14 (30%)	4 (36%)
	No	33 (70%)	7 (64%)
Congestive Heart Failure	Yes	47 (100%)	11 (100%)
	No	0 (0%)	0 (0%)
Hypertension Complicated	Yes	15 (32%)	2 (18%)
	No	32 (68%)	9 (82%)
Hypertension Uncomplicated	Yes	35 (74%)	8 (73%)
	No	12 (26%)	3 (27%)
Diabetes Complicated	Yes	11 (23%)	2 (18%)
	No	36 (77%)	9 (82%)
Diabetes Uncomplicated	Yes	17 (36%)	4 (36%)
	No	30 (64%)	7 (64%)
Pulmonary Circulation Disorders	Yes	18 (38%)	5 (45%)
	No	29 (62%)	6 (55%)

had chronic pulmonary disease, 32% had complicated hypertension and 74% of them had pulmonary circulation disorders.Among the 11 patients, 73% of them were males with an average of age of 63.5. Among them 36% had chronic pulmonary disease, 18% had complicated hypertension and 45% of them had pulmonary circulation disorders.

5.5.2.2 Method

In order to extract bradycardia events that have more clinical importance, several criteria were employed in terms of heart rate, medical diagnosis and medication.

Details about the bradycardia annotation algorithms are described in Section 3.3.4. Pre-processing, noise removal and pre-event extractions are described in Section 3.4.1, 3.4.2.2 and 3.4.3. Details of the DPFA algorithm is described in Section 3.2. Symbolization of the DPFA algorithm is described in Section 3.2.2, parameters are tuned with training and validation data and the optimal parameters are listed in 5.2.

The parameters a_1, b_1, a_2, b_2 were all tuned in the training step and set at

$$\begin{cases} a_1 &= 0.6 \\ b_1 &= 0.8 \\ a_2 &= 0.025 \\ b_2 &= 0.05 \end{cases} \tag{5.2}$$

For prediction of bradycardia in patients with or without beta blockers, five-fold cross-validation was used for model training, testing and validation. A total of 80 prediction intervals from 38 patients were used in training with the remaining 18 intervals from 9 patients were held out for testing. Within the training data set, 5-fold cross validation was performed for parameter tuning. Training, cross validation, and testing sets/folds were partitioned on the participant level, meaning that signals from the same participant were only included in one set/fold so as to prevent data leakage. The training-testing process was repeated for 10 times to avoid biased results due to randomness in training and testing data partition (Figure 5.17).

For prediction of bradycardia in patients without beta blockers, leave-five-out cross-validation was used for training and testing the model performance. The choice of leave-p-out instead of k-fold cross-validation was because of the limited sample size. There was a total of 11 bradycardia participants without beta blockers. For each iteration, 5 participants were chosen for the testing set and the remaining 6 participants were used to train the model (Figure 5.18). The choice of $p = 5$ ensures enough bradycardia cases in both the training and the testing datasets. For p smaller

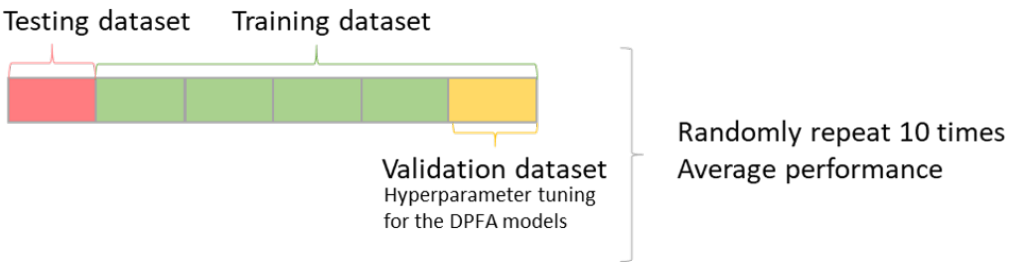


Figure 5.17: Bradycardia prediction on hospital bedside dataset (DB2), training scheme plot for five-fold cross validation.

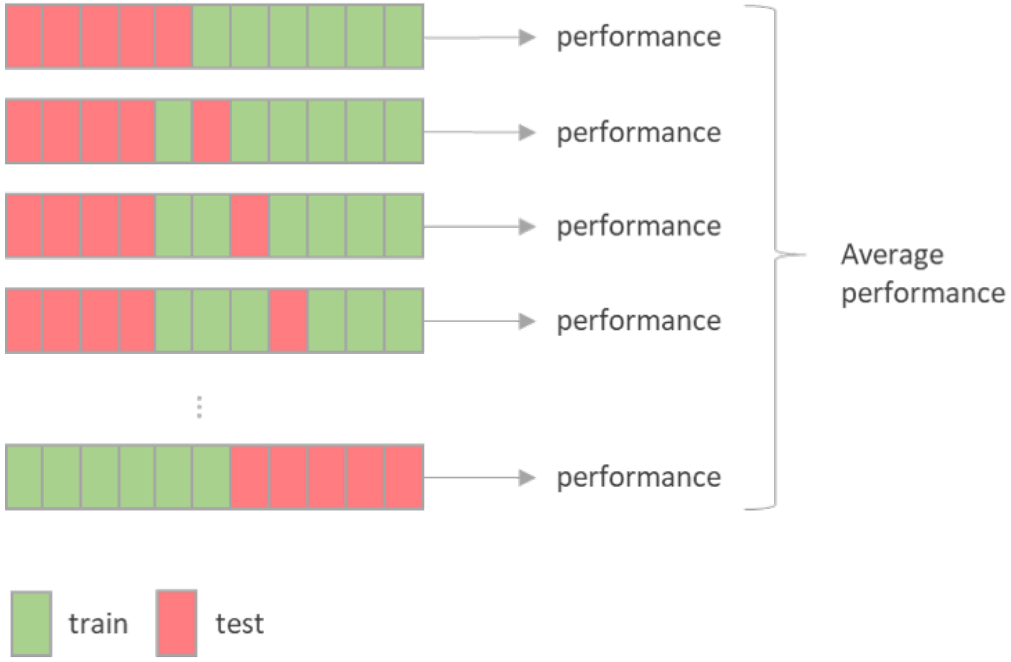


Figure 5.18: Bradycardia prediction on hospital bedside dataset (DB2), training scheme plot for leave-five-out cross-validation.

than 5, the results of the experiments were influenced by how the data was separated, causing a larger variance. The same issues arose for p larger than 6 by symmetry. Similar to five-fold cross-validation, the dataset was also split on the participant level to prevent data leakage. A total of ${}^{11}C_5 = 426$ iterations were performed to evaluate the models and the final result was the average of all the iterations.

5.5.2.3 Results

Different combinations of signal intervals (i.e., $t_{prediction} = 0.5, 1.0, 2.0$ minutes) and gap intervals (i.e., $t_{gap} = 0.5, 1.0, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5$ minutes) were used for prediction. These prediction intervals were tested using the DPFA algorithm.

Figure 5.19 shows a summary of the performance for the overall group including patients with beta-blockers. The overall average AUC is 0.67 with a standard deviation of 0.05. The average AUC is 0.67 ± 0.05 for $t_{prediction} = 0.5$, 0.68 ± 0.05 for $t_{prediction} = 1.0$ and 0.68 ± 0.05 for $t_{prediction} = 2.0$. The AUC is around 0.7 for all prediction intervals. For the bradycardia-positive DPFA, the algorithm generated an average of 607×3 transition states for the models using 0.5-minute long signals, 267×3 transitions states for the 1.0-minute long signal models, and 3867×3 transition states for the 2.0-minute long signal models. For the bradycardia-negative DPFA, the algorithm generated an average of 1949×3 transition states for the models using 0.5-minute long signals, 2483×3 transitions states for the 1.0-minute long signal models, and 3037×3 transition states for the 2.0-minute long signal models.

Figure 5.20 shows a summary of the performance for the group without beta-blockers. The overall average AUC is 0.72 with a standard deviation of 0.13. The average AUC is 0.71 ± 0.13 for $t_{prediction} = 0.5$, 0.72 ± 0.14 for $t_{prediction} = 1.0$ and 0.73 ± 0.12 for $t_{prediction} = 2.0$. The AUC is around 0.7 for all prediction intervals. For the bradycardia-positive DPFA, the algorithm generated an average of 19×3 transition states for the models using 0.5-minute long signals, 30×3 transitions states for the 1.0-minute long signal models, and 61×3 transition states for the 2.0-minute long signal models. For the bradycardia-negative DPFA, the algorithm generated an average of 58×3 transition states for the models using 0.5-minute long signals, 108×3 transitions states for the 1.0-minute long signal models, and 471×3 transition states for the 2.0-minute long signal models.

There are two major groups of arrhythmia detection algorithm based on ECG

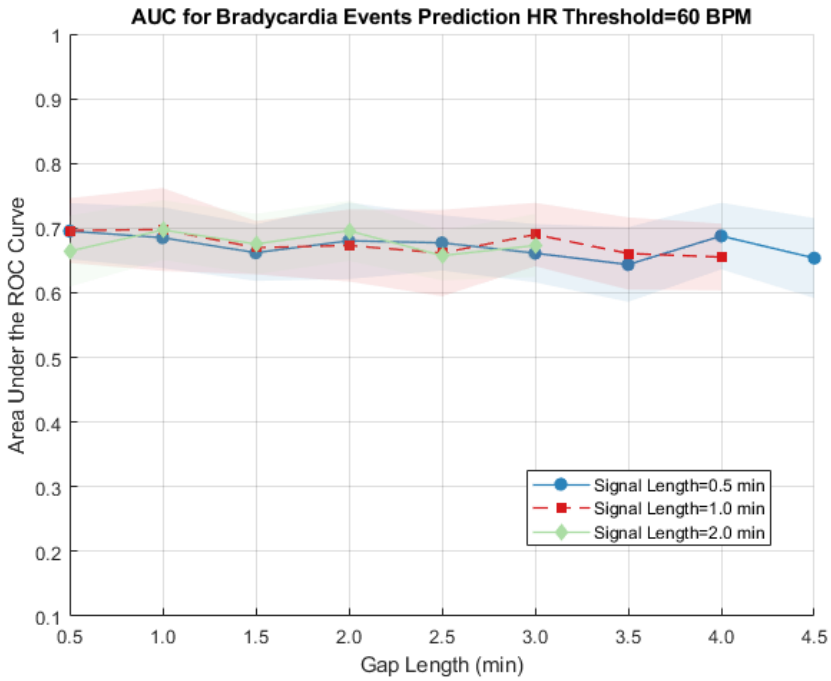


Figure 5.19: Bradycardia prediction on hospital bedside dataset (DB2), AUC for various signal lengths and gap intervals.

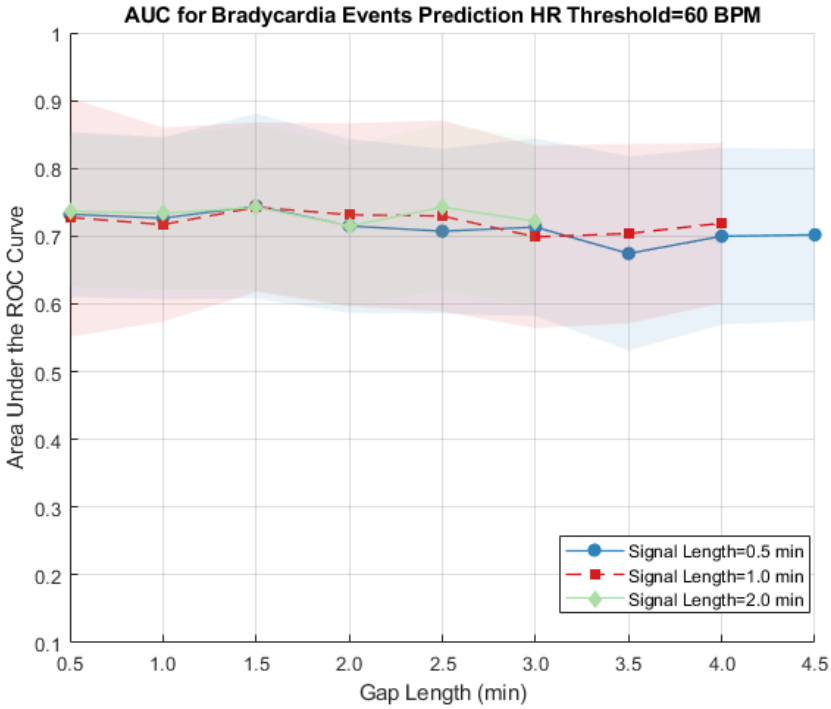


Figure 5.20: Bradycardia prediction on hospital bedside dataset (DB2), AUC for various signal lengths and gap intervals, without beta-blockers.

Table 5.18: Bradycardia prediction on hospital bedside dataset(DB2), result for bradycardia prediction with different gap intervals and lengths of signals, without beta-blocker.

Gap Interval (min)	**Prediction Interval (min)**	**AUC Mean (Std)**	**Sensitivity Mean (Std)**	**Specificity Mean (Std)**
0.5	0.5	0.732(0.12)	0.655(0.23)	0.637(0.20)
1	0.5	0.728(0.18)	0.565(0.25)	0.719(0.21)
1.5	0.5	0.737(0.11)	0.589(0.24)	0.696(0.21)
2	0.5	0.726(0.12)	0.765(0.26)	0.527(0.32)
2.5	0.5	0.717(0.14)	0.703(0.27)	0.596(0.23)
3	0.5	0.733(0.11)	0.577(0.25)	0.701(0.21)
3.5	0.5	0.744(0.14)	0.838(0.17)	0.510(0.28)
4	0.5	0.743(0.12)	0.811(0.20)	0.510(0.31)
4.5	0.5	0.743(0.12)	0.448(0.24)	0.778(0.17)
0.5	1	0.715(0.13)	0.835(0.18)	0.506(0.29)
1	1	0.731(0.13)	0.770(0.25)	0.544(0.31)
1.5	1	0.715(0.12)	0.561(0.26)	0.694(0.22)
2	1	0.707(0.12)	0.721(0.27)	0.571(0.25)
2.5	1	0.730(0.14)	0.767(0.24)	0.537(0.31)
3	1	0.743(0.12)	0.448(0.24)	0.778(0.17)
3.5	1	0.713(0.13)	0.751(0.28)	0.568(0.26)
4	1	0.699(0.13)	0.768(0.27)	0.512(0.33)
0.5	2	0.722(0.12)	0.538(0.24)	0.704(0.21)
1	2	0.674(0.14)	0.668(0.30)	0.561(0.28)
1.5	2	0.704(0.13)	0.752(0.26)	0.522(0.32)
2	2	0.700(0.13)	0.666(0.30)	0.621(0.26)
2.5	2	0.719(0.12)	0.736(0.28)	0.598(0.28)
3	2	0.702(0.13)	0.702(0.28)	0.595(0.27)

signals. The first group requires the extraction of hand-crafted features from time or frequency domain, followed by a machine learning classifier. The second group does not rely on hand-crafted features *Faust et al.* (2020), but instead uses deep learning methods such as convolutional neural networks (CNN) or long short-term memory (LSTM) recurrent neural networks (RNN) *Acharya et al.* (2017a); *Rajpurkar et al.* (2017); *Acharya et al.* (2017b).

In order to evaluate the performance of the DPFA algorithm relative to the other methods, the prediction experiments were performed with the same leave-five-out

training-testing scheme using the following algorithms:

1. Heart rate variability (HRV) features with random forest (RF). A total of 22 features were extracted from time-domain, frequency-domain and non-linear measurements. Time-domain features include: HR, RR, SDNN, SDHR, RMSSD, NN50, pNN50, SDSD, HRV TRI, and TINN. Frequency-domain features include: LF peak, HF peak, LF power, HF power, LFHF ratio, and VLF. Non-linear measurements include: SD1, SD2, Lorenz OC, Lorenz OC ratio, ApEN, and SampEn. See Table 5.19 for definition and full names of the variables *Shaffer and Ginsberg* (2017); *Sarkar et al.* (2008); *Hnatkova et al.* (1995); *Grant et al.* (2013).

   RF was used with grid search on number of trees, number of features at every split, maximum number of levels in trees, minimum number of samples required to split a node and minimum number of samples required at each leaf node.

2. 1D-CNN model: 1D-CNN model with 2 layers with ReLU activation followed by a dropout layer for regularization and a pooling layer. The learned features were then flattened and passed to a fully connected layer.

3. CNN-based ResNet: ResNets use a skipping mechanism to connect non-adjacent layers, thus making the features more robust to perturbations and improving the accuracy of deeper networks *Andreotti et al.* (2017). A 34 layers ResNet was applied to classify 30-seconds long single lead ECGs segments into 14 different cardiac disease classes with raw ECG data *Rajpurkar et al.* (2017). A ResNet with three residual blocks followed by a global average pooling layer and a softmax layer was used in this study for comparison.

   For the CNN-based models, architectures with fewer layers were chosen because of the limited sample size.

Table 5.19: Bradycardia prediction on hospital bedside dataset(DB2), table of features used in HRV with RF approach.

Parameter	**Description**
HR	Mean heart rates
RR	Mean of time elapsed between two successive R waves of the QRS signal
SDNN	Standard deviation of NN intervals
SDHR	Standard deviation of heart rates
RMSSD	Root mean square of successive RR intervals
NN50	Number of adjacent NN intervals that differ from each other by more than 50 ms requires a 2 min epoch
pNN50	Percentage of successive RR intervals that differ by more than 50 ms
SDSD	Standard deviation of successive RR interval differences
HRV TRI	Integral of the density of the RR interval histogram divided by its height
TINN	Baseline width of the RR interval histogram
LF Peak	Peak frequency of the low-frequency band (0.04-0.15 Hz)
HF Peak	Peak frequency of the high-frequency band (0.15-0.4 Hz)
LF Power	Absolute power of the low-frequency band (0.04-0.15 Hz)
HF Power	Peak frequency of the high-frequency band (0.15-0.4 Hz)
LFHF ratio	Ratio of LF-to-HF power
VLF	Absolute power of the very-low-frequency band (0.0033-0.04 Hz)
SD1	Poincaré plot standard deviation perpendicular the line of identity
SD2	Poincaré plot standard deviation along the line of identity
Lorenz OC	Number of points within the radius of normal sinus rhythm mask (80 ms)
Lorenz OC ratio	Ratio between number of points within the radius NSRmask to the total number of points
ApEn	Approximate entropy
SampEn	Sample entropy

The dataset in this study is imbalanced; in the total group there are 98 bradycardia intervals with around 1000 non-bradycardia intervals, in the group without beta-blockers there are 21 bradycardia cases with 371 non-bradycardia cases. A direct application of deep learning methods might be ineffective since these typically require much larger data sets and a more balanced class. In order to balance the two classes, each comparison method was then applied together with the synthetic minority over-sampling technique (SMOTE). SMOTE is a popular oversampling method proposed by Chawla in 2002, it can produce synthetic minority samples instead of duplicating samples *Chawla et al.* (2002). In the HRV with RF approach, the SMOTE method was applied after the HRV feature extraction step, in order to amplify the features extracted from the bradycardia group so as the achieve an equal number with the non-bradycardia group. In the CNN based method, SMOTE was applied to the raw training ECG to balance the number of the two classes.

In the total group including beta-blocker patients, the HRV with RF algorithm has an average AUC of 0.54 ± 0.03, sensitivity of 0.12 ± 0.07 and specificity of 0.97 ± 0.02 over all prediction intervals. There was a slight improvement after data augmentation with SMOTE, the algorithm achieved AUC of 0.57 ± 0.04, sensitivity of 0.24 ± 0.20 and specificity of 0.89 ± 0.04. Similarly in the group without beta-blocker patients using HRV with RF, data augmentation with SMOTE also resulted in with an average AUC of 0.55 ± 0.08 , sensitivity of 0.23 ± 0.20 and specificity of 0.87 ± 0.13. over all the prediction intervals, a slight improvement over the otherwise AUC of 0.5. However, there was not much improvement for CNN based methods, the AUC for the prediction intervals were still around 0.5. The algorithms failed to classify the minority class. This might due to the inherent requirement for larger datasets of deep learning algorithms in order to achieve better classification results.

5.5.2.4 Discussion

Despite active research in bradycardia prediction for premature infants, there have been only a limited number of studies in the adult population with CHF. In this study, lead II ECG signals were used to predict bradycardia events up to 5 minutes prior to the onset of the event. The proposed algorithm uses DPFA models to predict bradycardia events. The processed ECG is transformed to a probabilistic string of symbols representing broad morphology types for model generation. The generated models are then applied to the testing set for the prediction. An AUC of around 0.7 was achieved for all prediction intervals despite a relatively small sample size. To investigate impact of beta-blockers, analysis performed on patients without taking beta-blockers. With DPFA algorithm, there is a slightly increase in the prediction performance, however the variance has increased as well.

The proposed method does not rely on the extraction of hand-crafted features which requires extensive clinical knowledge. With a well-trained model, the algorithm has the ability to predict the events in a real-time prospective manner. This will allow time for medical attention for bradycardia events in patients with CHF. This will be a significant step toward the development of an in-vehicle arrhythmia prediction system.

One of limitations of the study is the small sample size. Although the entire bradycardia dataset consisted of 1251 bradycardia patients to start with, there were only 47 patients that met all the criteria for inclusion and for the beta-blocker excluded group there were only 11 patients. The strict criteria were imposed in order to focus on patients who were not only with bradycardia episodes but also had CHF, since it is also possible for healthy individuals such as athletes to have low heart rate but with no significant health concerns. Patients taking beta blockers were also excluded due to the medication effects on ECG signals. Although the goal was to include patients that would benefit most from a clinical alert system, this approach also led to a low number of patients thus limiting the ECG patterns that can be used for model

training. The leave-five-out training method was applied for algorithm evaluation, but a larger patient cohort could further help with algorithm training. The limited sample size also made it difficult to compare the performance the DFPA algorithm with any state-of-the-art deep learning algorithms, since the latter typically require much larger training dataset. In future studies, more patients with ECG signals that can be used for model training and evaluation will be enrolled.

5.6 Rapid Ventricular Rates with Low Activity

RVR are an important consequence of AFib *January et al.* (2014). A fast and irregular heart rate decreases time spent in diastole which impairs myocardial perfusion, ventricular diastolic filling, and cardiac output that in turn leads to greater symptom burden and reduced left ventricular systolic function *Skinner Jr et al.* (1964); *Kochiadakis et al.* (2002). However, there are many reasons for RVR during episodes of AF. In addition to inappropriate tachycardia due to the underlying disease, patients' heart rates may rise in a normal compensatory response to keep up with physiological demands during exercise or activities of daily living.

There is a need to not only detect and predict periods of AF with rapid ventricular rates but also distinguish periods of inappropriate tachycardia with AF from those associated with activity. Accelerometers are widely used in wearable sensors and provide important data that can be used to estimate patients' activity and movement within their natural setting *Godfrey et al.* (2008). They can provide important insights that are not available through investigating ECG alone. Therefore using raw accelerometry data in combination with ECG has the potential to identify and predict inappropriate tachycardia with AF. Accurate prediction of these episodes will allow just-in-time interventions for patients with AF.

Traditional algorithms have focused on the detection of AF episodes and are able to achieve this with high degree of accuracy *Teplitzky et al.* (2020). However, predic-

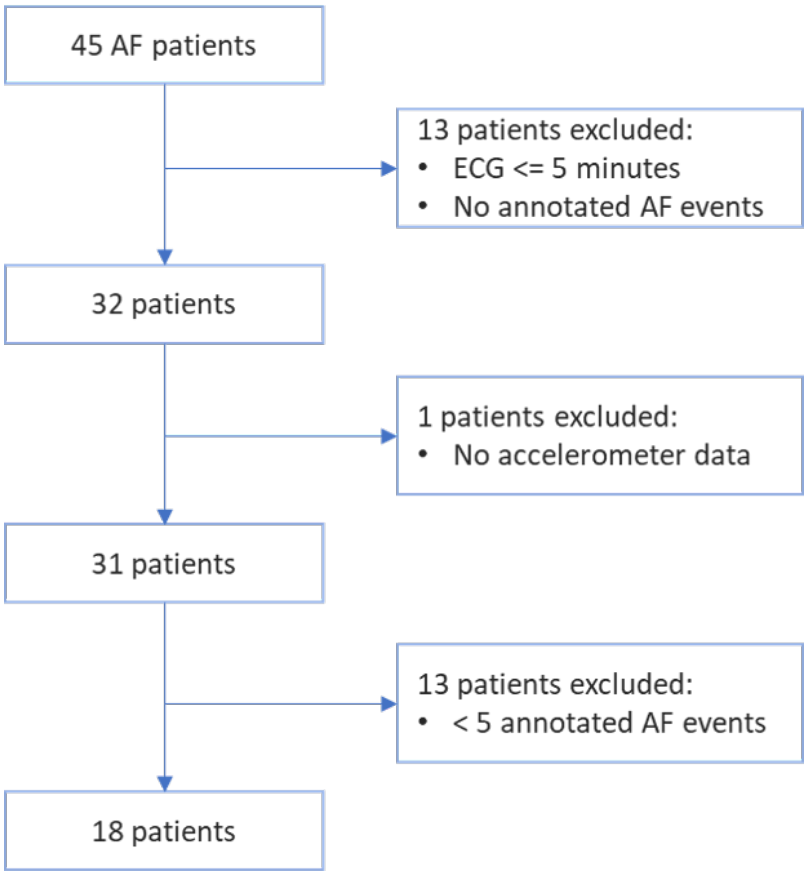
tion algorithms have not been able to achieve sufficient lag time to allow for medical or surgical interventions. There are several clinical scenarios where identifying rapid AF at low levels of activity can be of clinical value, including ensuring appropriate rate control for patients with AF, triggers for on-demand rhythm control approaches to improve patient care and facilitate monitoring system for AF that could be treated medically.

The aim of this part of study is to predict the onset of AF with RVR episodes associated with low activity using pre-event ECG signals. The proposed approach involves a DPFA-based algorithm that can predict episodes of AF with RVR that are associated with low levels of activity.

5.6.1 Data

A total of 45 patients with history of AF who presented to University of Michigan are recruited in the study. All patients wore an event recorder (Preventice solutions Inc) for up to three weeks. IRB approved the protocol and written informed consent was obtained. The ECG and accelerometer data were recorded continuously for up to 3 weeks. Thirteen patients are excluded because the length of the ECG is less than 5 minutes or there are no annotated AF events, 1 additional patient is excluded since no accelerometer data. Thirteen patients with AF episodes of duration less 30 seconds or less than 5 episodes of AF were further excluded. Finally a total of 18 patients are included in the prediction analysis. A flow chart of the patient inclusion/exclusion criteria is demonstrated in Figure 5.21. The characteristics of patients are summarized in Table 5.20.

Figure 5.21: RVR with low activity prediction, figure of patient flow.



5.6.2 Method

5.6.2.1 Accelerometer data

The raw accelerometer data are collected along three orthogonal axes in the device-specific frame of reference. The continuous accelerometer data are collected using a wireless monitoring device that adheres to patient's chest (Preventice Solutions, Inc) and sampled at 10 Hz. The accelerometer data were up-sampled to 256 Hz to match the ECG sampling rate. For the analysis, the accelerometer magnitude $am(t)$ consisting of the vector magnitude of the accelerometer data at each time point is

Table 5.20: RVR with low activity prediction, characteristics of patients.

Variable	**All Participants (n=45)**	**Participants in Prediction Analysis (n=18)**
Female	14 (31.82%)	5 (27.78%)
Age	66.36 (11.67%)	69.13 (7.31%)
BMI	31.30 (6.08%)	30.93 (5.66%)
Hypertension	26 (59.09%)	11 (61.11%)
History of Stroke	0 (0.00%)	0 (0.00%)
Diabetes	12 (27.27%)	5 (27.78%)
Coronary artery disease	11 (25.00%)	5 (27.78%)
Peripheral vascular disease	2 (4.54%)	2 (11.11%)
Beta Blockers	31 (70.45%)	13 (72.22%)
Calcium channel blockers	14 (31.82%)	4 (22.22%)
Antiarrhythmic drugs	9 (20.45%)	2 (11.11%)

used in analysis together with the synchronized ECG signal.

$$\begin{aligned} am(t) &= \sqrt{x(t)^2 + y(t)^2 + z(t)^2} \\ &= \text{ accelerometer magnitude at the time point } t. \end{aligned} \tag{5.3}$$

The data is aggregated using MAD which computes the deviation of $am(t)$ from its mean over the corresponding epoch, averaged over the length of the annotated AF signal *Bakrania et al.* (2016).

$$MAD = \frac{1}{n} \times \sum_{i=1}^{n} |am(t_i) - \overline{am}| \tag{5.4}$$

where:

$am(t_i) =$ accelerometer magnitude at the ith time point

$\overline{am} =$ mean accelerometer magnitude within the time period of interest

$n =$ length of the time period

The $MAD \leq 0.75$ quantile relative to the measurements from the entire group id des-

ignated as low activity and $MAD > 0.75$ quantile as high activity in order to account for inter-individual differences in activity levels within the study group. Threshold of 0.75 quantile for distinguishing activity levels is an arbitrary choice since there is no well-established guideline for low vs. high activity level.

5.6.2.2 ECG data

ECG is acquired using a single lead event recorder (Preventice Solutions, Inc) and sampled at 256 Hz. ECG data were collected with the same biopatch as the accelerometer. Continuous single channel ECG signals were collected while wearing the device. Arrhythmia classification was performed using the BeatLogic platform that is cleared to identify AF episodes *Teplitzky et al.* (2020). Our clinician has also reviewed 150 randomly selected AF events annotated by Beatlogic for further validation, out of which 139 (93%) were confirmed to be AF events. A total of 961 AF events were annotated by the algorithm. Most of the participants had less than 50 episodes of AF events throughout the 3 weeks with several participants with higher number of AF events. Figure 5.22 shows burden of AF for all participant and those included in the prediction algorithm analysis.

In terms of HR, RVR episodes are defined as having $HR > 110$ BPM, and non-RVR episodes with $HR \leq 110$ BPM. Pre-processing, noise removal, and prediction interval extraction were performed on the synchronized recordings of ECG and accelerometer magnitude signals.

In the signal pre-processing step, a second-order Butterworth band-pass filter was applied with cutoff frequencies of 0.5 and 40 Hz to the raw ECG signal for noise removal. In the next step, a double median filter with orders equal to 0.2 and 0.6 times the sampling frequency was applied to remove baseline wandering. Peak-detection algorithm was then applied to capture the R-peaks in the ECG signals. Noise section of the signals was detected with annotations from BeatLogic platform.

Events that occurred too close to previous events were excluded to avoid overlap of prediction intervals with arrhythmia events. Events that occurred within 8 minutes of a noisy signal were also excluded to ensure that the prediction interval is outside the noisy signal range. We used the ECG signal to predict episodes of AF with RVR and low activity level vs. all other AF episodes.

The gap interval t_{gap} represents the interval in minutes before the event that is used for prediction. The signal interval t_{signal} represents the length of the prediction signals. For example, prediction intervals with $t_{gap} = 1$ min, $t_{signal} = 2$ min are the signals that span from 3 to 1 minutes before the annotated events. Figure 5.23 shows an example of an annotated AF episode and prediction intervals.

5.6.2.3 DPFA Model

In this part study, we improved our algorithm and tested its ability to work with different types of synchronized physiological signals (i.e synchronized ECG with accelerometer magnitude signals) for arrhythmia events predictions. We used the DPFA algorithm for predicting AF episodes with RVR and low activity level vs. other AF episodes using ECG and accelerometer magnitude signals prior to the event.

Each DPFA is generated in two steps (see 3.2 for details of the algorithm):

First in the symbolization module, one begins with the training dataset which consists of annotated ECG and accelerometer magnitude signals. From these, the algorithm extracts windows that are indicative of imminent events and others that are not. The positive (i.e., AF with RVR with low activity) and negative (i.e., all the other regions) ECG and accelerometer magnitude signals are respectively combined and transformed into probabilistic strings. Here the probabilistic strings consist of an alphabet of nine symbols $\Sigma = \{\alpha_1\beta_1, \alpha_2\beta_1, \alpha_3\beta_1, \alpha_1\beta_2, \alpha_2\beta_2, \alpha_3\beta_2, \alpha_1\beta_3, \alpha_2\beta_3, \alpha_3\beta_3\}$ where α_i correspond to different activity levels and β_j correspond to different ECG

morphology types within each window, and the probabilities

$$\mathbf{p}(\alpha_i\beta_j) = \mathbf{p}(\alpha_i)\mathbf{p}(\beta_j) \tag{5.5}$$

are computed by assigning probabilities to α_i and β_j independently. The individual probabilities $\mathbf{p}(\alpha_i)$ and $\mathbf{p}(\beta_j)$ are obtained from the signal values x_t over discrete time windows t via soft-thresholding

$$\begin{cases} \mathbf{p}_1 &= \psi_1(x_t) \\ \mathbf{p}_2 &= \big(1 - \psi_1(x_t)\big) \cdot \psi_2(x_t) \\ \mathbf{p}_3 &= 1 - \mathbf{p}_1 - \mathbf{p}_2, \end{cases} \tag{5.6}$$

where the soft-thresholding functions are chosen to be piecewise linear functions $\psi_1, \psi_2 : [0,1] \to [0,1]$ of the form

$$\psi_j(x) = \begin{cases} 1 & \text{if } x > b_j \\ \frac{x-a_j}{b_j-a_j} & \text{if } a_j \leq x \leq b_j \\ 0 & \text{if } x < a_j. \end{cases} \tag{5.7}$$

The parameters a_1, b_1, a_2, b_2 were all tuned in the training step and set at

$$\begin{cases} a_1 &= 0.03 \\ b_1 &= 0.05 \\ a_2 &= 0.01 \\ b_2 &= 0.02 \end{cases} \text{ for activity } \alpha_i \text{ and } \begin{cases} a_1 &= 0.5 \\ b_1 &= 0.7 \\ a_2 &= 0.025 \\ b_2 &= 0.05 \end{cases} \text{ for ECG } \beta_j. \tag{5.8}$$

based on a grid search while running the inner loops of the nested cross validation.

Then the DPFA generation module constructs the positive (AF with RVR and low activity) DPFA M_+ and the negative (all other AF episodes) DPFA M_- respectively

from the positive and negative probabilistic strings, by first constructing the frequency prefix trees T_+ and T_-, followed by the largest suffix merging algorithm $T_+ \to M_+$ and $T_- \to M_-$. The frequency prefix trees are tree-like automata $T = \langle Q_0, \Sigma, \varepsilon, \text{Freq} \rangle$ with initial state ε whose state space Q_0 consists of strings in the alphabet Σ with non-zero frequency, and transition function given by concatenation of strings. The largest suffix merging algorithm selects those states $q \in Q_0$ of the frequency prefix tree T with sufficiently high frequency to be in the state space M, and then defines the transition state $T(q, \alpha_i\beta_j)$ to be the largest suffix of $q\alpha_i\beta_j$ that is itself contained in the state space of M. See 3.2 for further details.

The classification scheme contains a training phase and a testing phase. In the training phase, the algorithm learns the DPFA M_+ and M_- for the positive and negative classes respectively. Then in the testing phase, we classify a given synchronized ECG-accelerometer magnitude signal from the test dataset by comparing the goodness-of-fit with the DPFA M_+ and M_- to predict episodes of AF with RVR with low activity Figure 5.24.

5.6.3 Data Partition

A total of 961 episodes of AF events were annotated for the analysis. We used 5 fold nested cross validation for parameter tuning. Figure 5.25 shows how data was partitioned in the analysis. The 961 episodes were split into 5 folds, 4 folds were used for training and the remaining fold was used for testing. Within the inner loop, the training data had a total of 4 folds, 3 folds were used for parameter tuning and the remaining fold for validations. In the outer loop, the DPFA models were generated using all 4 folds of the training data with hyper-parameters tuned during the inner loop. Predictions were performed on the testing data in each split and then averaged.

5.6.4 Results

During the study period, we have recruited 45 patients with AF, of whom 31 had both ECG and accelerometer data. We excluded patients with AF episode of duration less 30 seconds and those with less than 5 episodes of AF. There was a total of 292 episodes of AF with RVR compared to 669 episodes of AF with controlled heart rate. There were 116 episodes in $>$ 75 percentile activity level in the RVR group and 124 episodes in the non-RVR group based on MAD relative to total group. The distribution of high activity based on heart rates above 110 BPM are summarized in Table 5.21. The 45 participants in the study had a median of 22 days of total wear time (range 2–31 days). The median overall AF burden was 38.4% (inter-quantile range (IQR) 4.9%-86.0%).

Among the 961 annotated AF events, 292 of them met the criterion for RVR episode, among which 176 episodes had low activity level and the remaining 116 episodes had high activity level based on MAD relative to the entire group activity level. For the rest of the annotated AF events, 669 of them were non-RVR episodes, among which 545 episodes had low activity and the remaining 124 episodes had high activity level. Table 5.21 summarizes the number of annotated AF episodes by activity and HR levels.

A total of 18 patients are included in the final DPFA model. They had a median of 22 days of total wear time (range 6–31 days). The median overall AF burden for the 18 patients was 19.7% (IQR 10.5%-62.3%). Two participants had an AF burden $>$ 90% over the duration of the study. The mean heart rate was 99.0 ± 19.4 BPM. The burden of AF for study participants is summarized in Figure 5.22.

Different combinations of signal intervals (i.e., $t_{signal} = 0.5, 1.0, 2.0$ min and gap intervals (i.e., $t_{gap} = 0.5, 1.0, 2.0, 2.5, 3.5, 4.0, 4.5$ min) up to 5 minutes before the AF events were used for prediction. For a given model based on various gap intervals, the threshold can be adjusted to optimize different parameters, i.e., a more sensitive

Table 5.21: AF episodes by activity level and HR level.

	Low Activity	High Activity	Total
RVR	176	116	292
Non-RVR	545	124	669
Total	721	240	961

model vs a more specific model. In Table 5.22, we report the threshold, sensitivity, specificity, precision, and other relevant data based on each gap interval.

5.6.4.1 Results

We limited the events to those that lasted for at least 30 seconds to investigate the effects of AF duration on predicting episodes of AF with RVR with low activity. As the time to the event increases, the mean prediction performance gradually declines while the variance increases. Using the MAD threshold relative to the whole group, the average AUC for 0.5 minute-long prediction intervals is 0.71 ± 0.04, 1 minute-long is 0.75 ± 0.03 and 2 minute-long is 0.77 ± 0.03 (Figure 5.26).The 2 minute-long prediction intervals also showed the highest AUC around 0.75 with smallest standard deviations. Prediction results for various gap intervals and prediction intervals are shown in Table 5.22.

5.6.4.2 DPFA model interpretation

In addition to the superior performance, another major advantage of our approach is the interpretability of the underlying DPFA models. Indeed the DPFA models could be understood by first calculating the relative frequencies of the states, and then identifying the most prevalent rhythm patterns represented by the states within the DPFA model constructed from each class. This could be achieved via standard techniques by first extracting the normalized leading eigenvector of the transition matrix of the DPFA model, and then taking the average of the relative frequencies across the five

Table 5.22: Prediction results for various gap intervals and prediction intervals.

Gap Interval (min)	**Prediction Interval (min)**	**AUC Mean (Std)**	**Sensitivity Mean (Std)**	**Specificity Mean (Std)**	**Accuracy Mean (Std)**
0.5	0.5	0.735(0.026)	0.552(0.049)	0.808(0.104)	0.806(0.027)
1	0.5	0.700(0.050)	0.546(0.075)	0.797(0.085)	0.811(0.035)
1.5	0.5	0.725(0.041)	0.545(0.098)	0.804(0.078)	0.810(0.024)
2	0.5	0.706(0.064)	0.527(0.111)	0.810(0.084)	0.799(0.027)
2.5	0.5	0.729(0.048)	0.521(0.103)	0.845(0.065)	0.781(0.062)
3	0.5	0.703(0.035)	0.511(0.115)	0.790(0.091)	0.758(0.088)
3.5	0.5	0.673(0.049)	0.527(0.121)	0.742(0.092)	0.793(0.059)
4	0.5	0.701(0.051)	0.521(0.115)	0.802(0.110)	0.718(0.068)
4.5	0.5	0.691(0.038)	0.520(0.145)	0.790(0.073)	0.785(0.060)
0.5	1	0.768(0.031)	0.541(0.041)	0.850(0.081)	0.779(0.109)
1	1	0.756(0.046)	0.524(0.074)	0.838(0.097)	0.841(0.019)
1.5	1	0.753(0.019)	0.545(0.092)	0.838(0.104)	0.826(0.021)
2	1	0.751(0.054)	0.488(0.094)	0.864(0.073)	0.828(0.043)
2.5	1	0.753(0.012)	0.585(0.102)	0.805(0.135)	0.827(0.040)
3	1	0.734(0.026)	0.569(0.116)	0.767(0.128)	0.783(0.106)
3.5	1	0.730(0.030)	0.573(0.108)	0.809(0.117)	0.784(0.103)
4	1	0.742(0.043)	0.567(0.101)	0.819(0.132)	0.775(0.117)
0.5	2	0.776(0.029)	0.547(0.087)	0.838(0.074)	0.836(0.047)
1	2	0.769(0.011)	0.584(0.141)	0.824(0.117)	0.818(0.044)
1.5	2	0.780(0.047)	0.572(0.119)	0.869(0.088)	0.814(0.071)
2	2	0.761(0.034)	0.603(0.090)	0.782(0.150)	0.826(0.038)
2.5	2	0.766(0.017)	0.572(0.081)	0.820(0.098)	0.790(0.076)
3	2	0.746(0.044)	0.586(0.114)	0.777(0.129)	0.779(0.124)

cross-validation folds. For example, by comparing the two DPFA models M_+ and M_- with signal length = 0.5 (mins) and gap length = 0.5 (mins), this approach yields the following five rhythm patterns as showing the most significant difference between M_+ and M_-: \$ff (which corresponds to the rhythm pattern $[\alpha_3\beta_2, \alpha_3\beta_2]$, where $\alpha_3\beta_2$ corresponds to a single window with activity level α_3 and ECG type β_2 as defined by equation (5.5,5.6,5.7,5.8)), \$ciic, \$ici, \$ffic, and \$cii. Here for ease of presentation we have switched to an alphabet consisting of single-letter symbols, the two alphabets correspond to each other as follows

old alphabet	$\alpha_1\beta_1$	$\alpha_2\beta_1$	$\alpha_3\beta_1$	$\alpha_1\beta_2$	$\alpha_2\beta_2$	$\alpha_3\beta_2$	$\alpha_1\beta_3$	$\alpha_2\beta_3$	$\alpha_3\beta_3$
new alphabet	a	b	c	d	e	f	g	h	i

and we use the \$ symbol to denote the empty string. Please see Figure 5.28 for a plot of the relative frequencies of all the states and the ten most significant rhythm patterns with signal length = 0.5 (mins) and gap length = 0.5 (mins). Please also see Table 5.23 for the five states with most difference between the relative frequencies in DPFA M_+ vs DPFA M_- for various gap intervals and prediction intervals.

5.6.5 Discussion

This study in patients with AF showed that our novel DPFA algorithm can predict the onset of AF with RVR associated with low levels of activity with AUC around 0.75 for intervals up to 4.5 minutes before the onset of the event. This is the first study to evaluate the performance of prediction algorithms using the ECG along with accelerometer data. Prediction of AF with rapid rates can result in personalized treatment options for prevention and management of AF episodes that are likely to be clinically significant. Our algorithm can help to distinguish between AF with RVR episodes that occur unexpectedly at low levels of activity, thus more likely to be clinically significant.

The machine learning techniques have shown impressive capability to analyze

Table 5.23: Five states with most difference between the relative frequencies in DPFA M_+ vs. DPFA M_- for various gap intervals and prediction intervals

Gap Interval (min)	**Prediction Interval (min)**	**State1**	**State2**	**State3**	**State4**	**State5**
0.5	0.5	$ff	$ciic	$ici	$ffic	$cii
1	0.5	$i	$ciic	$cii	$icii	$iffff
1.5	0.5	$iiiciici	$iiiciii	$ciic	$icci	$ciici
2	0.5	$ff	$cii	$ici	$cifi	$f
2.5	0.5	$ficific	$ific	$ificic	$iiiciii	$iffi
3	0.5	$fi	$cii	$ficific	$ciic	$icciic
3.5	0.5	$ficific	$iccifici	$iiiciicic	$iccii	$icific
4	0.5	$ff	$cii	$i	$fi	$icc
4.5	0.5	$cifici	$icif	$cific	$cifi	$fici
0.5	1	$iicciiiiciii	$iicciiiic	$iicciiiicii	$iicciiiici	$iicciiicii
1	1	$ificificifici	$ificificific	$iicifii	$iicifiic	$ciicicii
1.5	1	$ficifici	$cific	$ciici	$cif	$ficificificif
2	1	$ciic	$cii	$cific	$ciii	$iccii
2.5	1	$cific	$ciii	$icif	$ciic	$cicici
3	1	$ificiiic	$ii	$cificif	$ficif	$iccic
3.5	1	$i	$ic	$f	$if	$iiciici
4	1	$ificificifici	$icci	$iiccificii	$iiccifii	$ificificific
0.5	2	$ificificificifici	$iiiiciiiicii	$iiiiciiciicii	$iiiiciiiiciii	$ificific
1	2	$iiiiciicii	$ificific	$ifiic	$ificificificif	$ificificifici
1.5	2	$ificificifi	$ificificificific	$ii	$ficifici	$ificificificifici
2	2	$ici	$icifi	$icif	$ffff	$iccificifi
2.5	2	$iiccif	$ificificific	$iicc	$iiccic	$iiciiii
3	2	$iccii	$fici	$ficific	$ficif	$icif

massive amounts of data and are an effective method for classification of arrhythmias using ECG data. Despite being most commonly used in many areas, deep learning algorithms have the drawback that their architecture represents a "Black Box" *Topol* (2019). Their lack of transparency limits their clinical adoption, makes mechanistic interpretation difficult and reduces their trustworthiness. Further, the disclosure of meaningful details about medical treatment to patients requires the doctors to grasp the fundamental inner workings of the devices they use to some degree *Nagendran et al.* (2020). Explainability may also be required to justify the clinical validation of machine learning algorithms in prospective studies and randomized clinical trials *Doshi-Velez and Kim* (2017). Our DPFA model represents a novel and explainable algorithm. This allows others to validate our algorithm in a diverse population of patients and adopt it clinically if proven to be effective. Previous studies have used the features extracted from the accelerometer data to give an overall impression of patient activity over a period of time. However, these features can not fully capture the variability in daily activity. It also misses the features within the activity signal that can provide important physiological insights about the patient's condition *Yang and Hsu* (2010). Our approach shows that using the one-dimensional accelerometer magnitude from the tri-axial raw accelerometer data can provide important information that can augment the interpretation of the ECG and accelerometer signals beyond the traditional features extracted from the signals.

Most methods for physiological data analysis depend heavily on pre-processing. However, these methods tend to be less effective on noisy data, such as data collected in real-time or in outpatient settings. Therefore, it is desirable to introduce new methods that require minimal pre-processing to analyze such data, thus allowing these insights to be applied to automated clinical decision making. In the previous study, the DPFA method has showed to be particularly useful for the real-world noisy data. Our algorithm does not rely on peak detection, requires minimal pre-processing

and leads to good performance in the setting of noisy ECG signal. This provides a significant advantage over existing algorithms when it is necessary to perform rapid analysis in highly noisy environments. Thus DPFA algorithm can be useful for signals captured by portable devices which are prone to noise.

Prediction can only be helpful if it results in specific action that can impact clinical outcomes. In recent years, an increasing number of portable devices have been developed to monitor the physiological signals. Our algorithm provides a possibility for real-time prediction of clinically meaningful arrhythmias using ECG signal together with accelerometer signals with enough lag time for medical interventions.

This is the first algorithm that can identify features within the ECG that are capable of detecting and predicting periods of AF with RVR that are not associated with high activity levels. This allows clinicians to distinguish between clinically significant periods of AF and other periods. This has significant implications for the appropriate treatment of patients with AF. Identifying physiologically significant episodes of AF allows an opportunity to deliver just-in-time treatments that can be tailored for each individual while avoiding the side effects related to daily medications.

5.6.6 Limitations

The first limitation of our study is the sample size. Our algorithm was validated in a small subset of patients with AF. Although the number of patients included in the study was small, there were numerous episodes available for training our algorithm. The validity and generalizability of our algorithm needs to be tested in a wider more diverse group of patients with AF. We sought to mediate the impact of patient subtypes by leveraging a large number of episodes of AF in our patient population; however, future work incorporating medication status, co-morbidities and other factors would allow for measuring algorithm performance within specific patient subsets and ensure equal representation in the training dataset.

The aim of our project was to predict AF with rapid ventricular rates and low activity levels. Worsening AF symptom severity is associated with reduced daily activity *Semaan et al.* (2020). The second limitation of the study is a lack of patient-reported symptoms during periods of AF with RVR and low activity. Therefore, it is difficult to ascertain symptom severity during these episodes. Future studies with frequent momentary assessment of symptoms are needed to determine the relationship between symptoms and AF with RVR episodes and low activity.

The third limitation is the interpretability of the activity threshold. In the study, an arbitrary choice of 0.75 quantile MAD based on activity of the entire group is used for classifying activity level. However, participants might have large differences in activity level. To account for these differences, MAD threshold relative to individual participants is also used for analysis. The average AUC for 0.5 minute-long prediction intervals is 0.74±0.02, 1 minute-long is 0.75±0.01 and 2 minute-long is 0.78±0.01. The 2 minute-long prediction intervals showed most consistent results with AUC around 0.75 and smallest standard deviations. When t_{gap} is less than 1.5 minutes, the AUC ranges between 0.74 - 0.80. The AUC is above 0.69 for all prediction intervals. Figure 5.27 shows the prediction results based on activity events labeled based on participant level MAD. We did not distinguish between different types of activity. It is possible that certain forms of exercise (e.g., stationary cycling) may not be accounted for in the present analysis. The MAD threshold used in the study is a relative MAD threshold instead of an absolute MAD threshold. Studies have investigated different types of activities with their MAD thresholds. However these studies *Bakrania et al.* (2016) used accelerometer sensors attached to the hip or wrist regions while our sensors are attached to the chest which make our MAD measurements incomparable with these thresholds. There are several studies that investigate activity classification using chest mounted tri-axial accelerometer devices *Godfrey et al.* (2011); *Gjoreski et al.* (2014); *Purwar et al.* (2007); *Gao et al.* (2014) with signal processing and machine learning

techniques, however they did not state thresholds of activities with MAD metric. We tried to minimize this limitation by using both intra-individual and sample-level activity distributions for the threshold. However, in the future if accelerometer data can be collected with similar sensors and attached to wrists, the prediction results can be better generalized. Both the accelerometer data and ECG data are collected as continuous signals with timestamp when the device is attached to the participant, so the non-wear time information is available and automatically removed, since our study uses synchronized accelerometer and ECG signals. The non-wear time is unlikely to affect our results on prediction of RVR with low activity level. But for future studies it could be helpful to know the reason why the patient removed the device and if it was a consequence of potential events. To this end, we plan to standardize and automate the annotation process by intelligently mining the training data. We are hopeful that such efforts will continue to improve of the performance of the algorithm and further expand the scope of our approach.

Figure 5.22: RVR with low activity prediction, AF Burden for all participant and those included in the prediction of RVR with low activity.

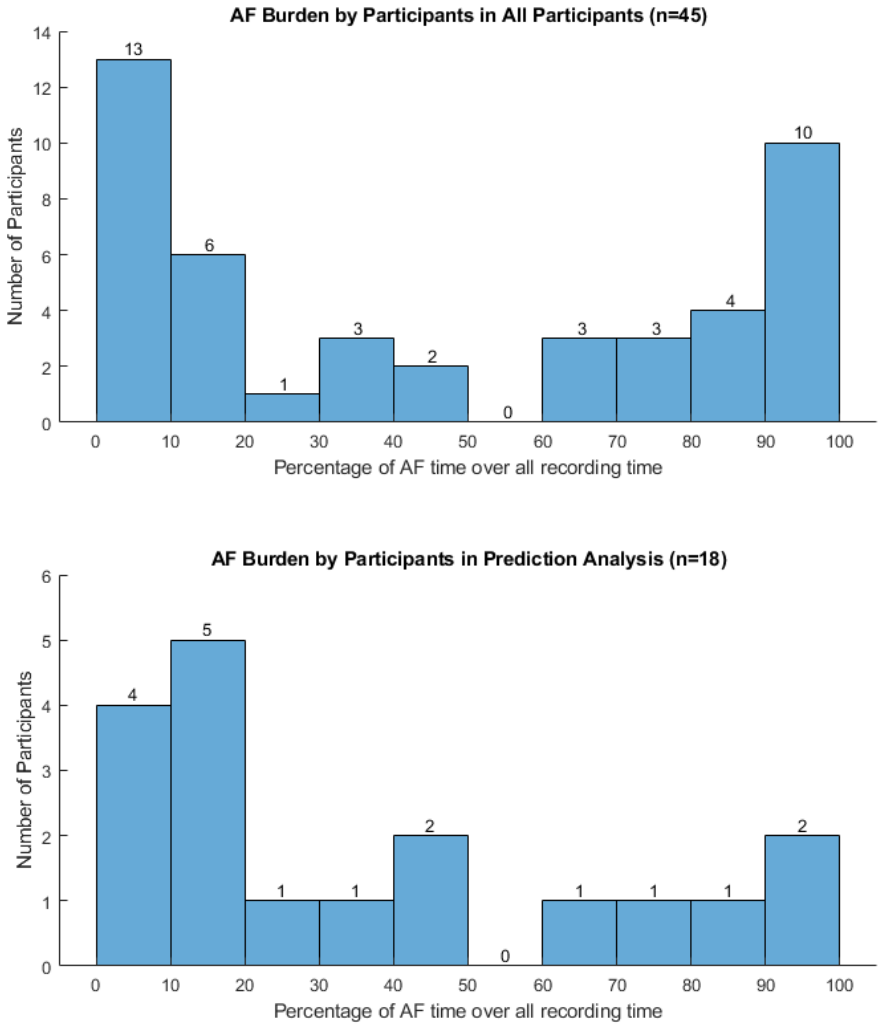


Figure 5.23: RVR with low activity prediction, prediction and gap intervals extractions.

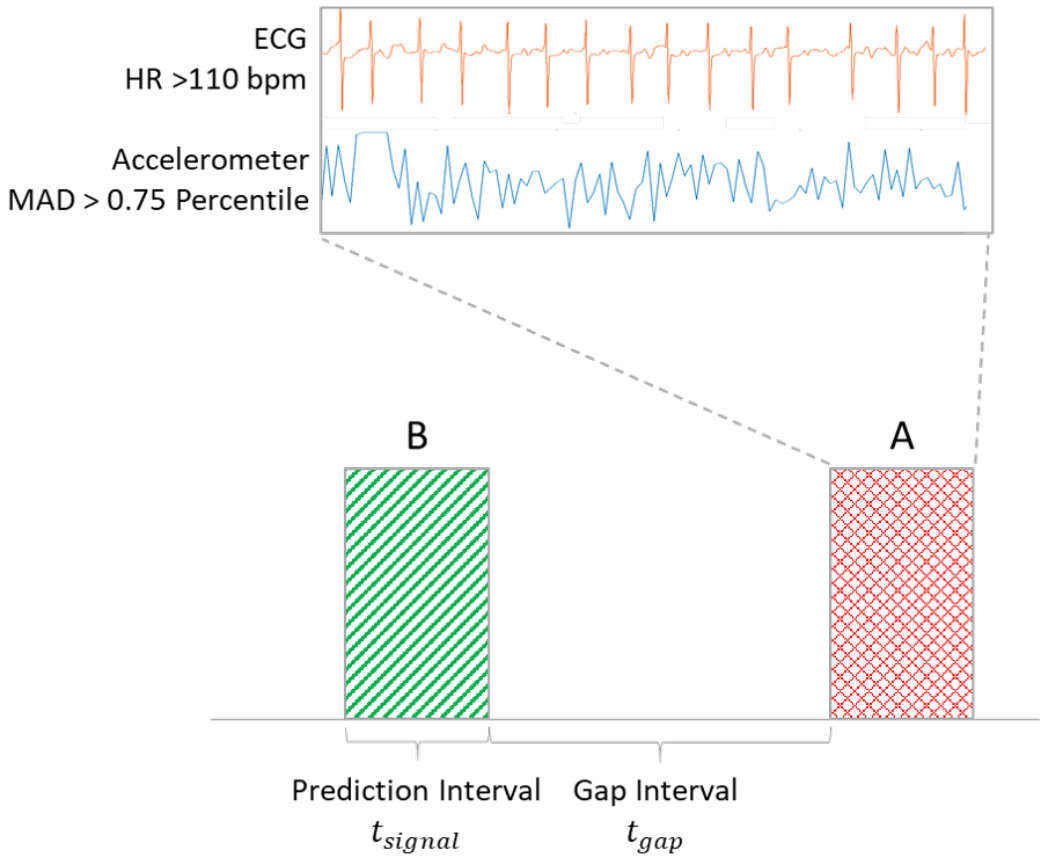


Figure 5.24: RVR with low activity prediction, classification of events using DPFA model.

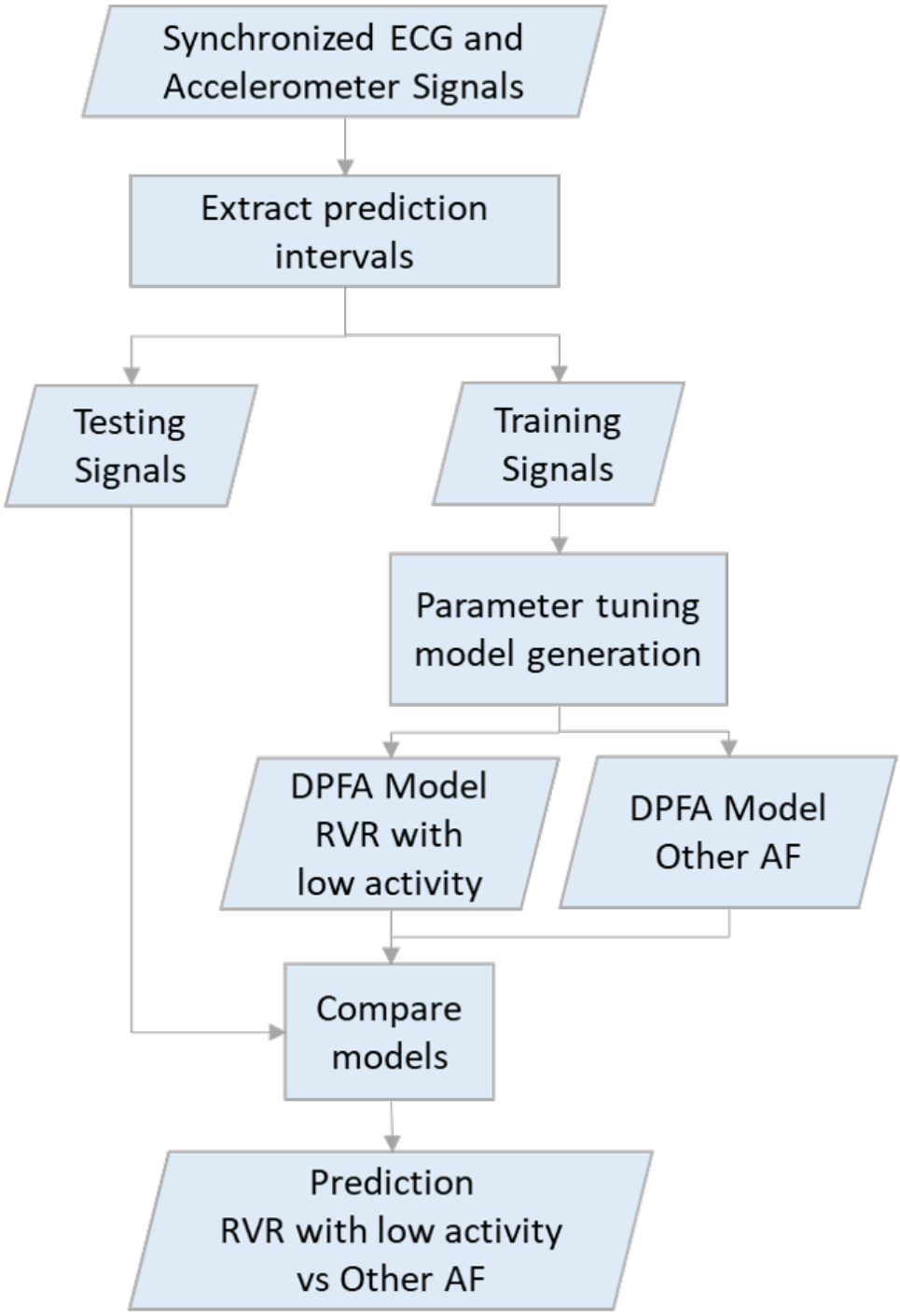


Figure 5.25: RVR with low activity prediction, five-fold nested cross validation.

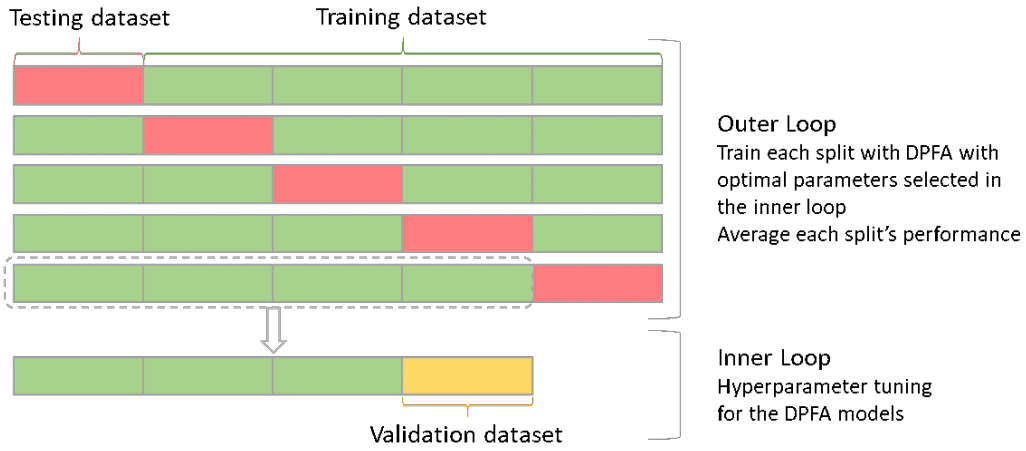
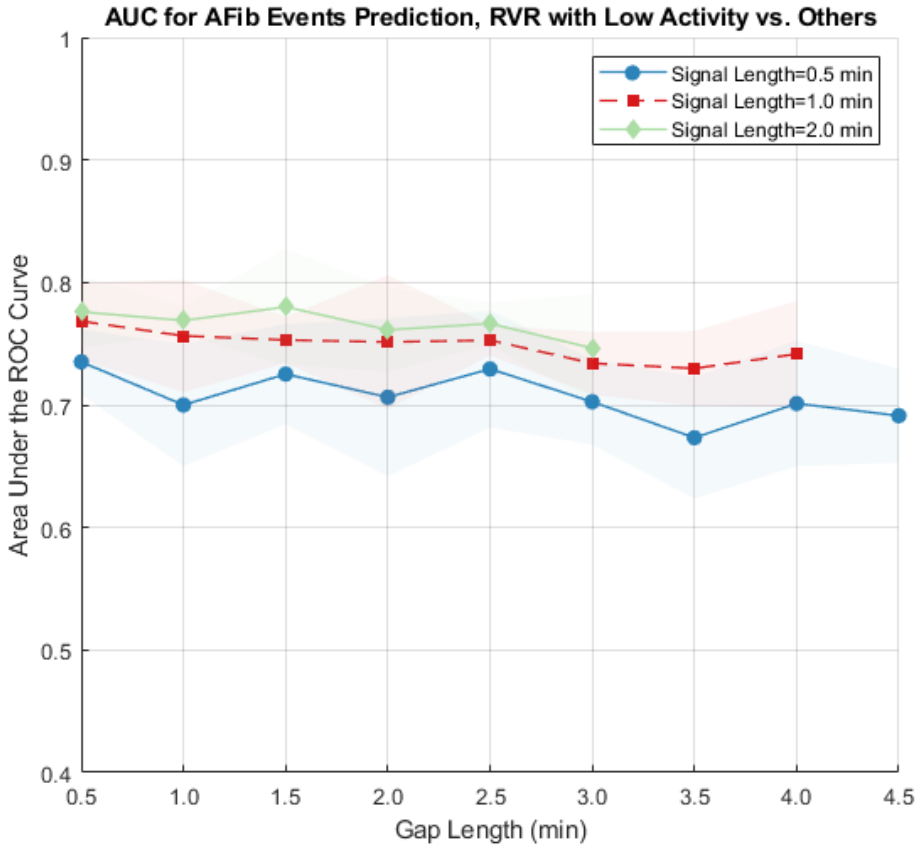
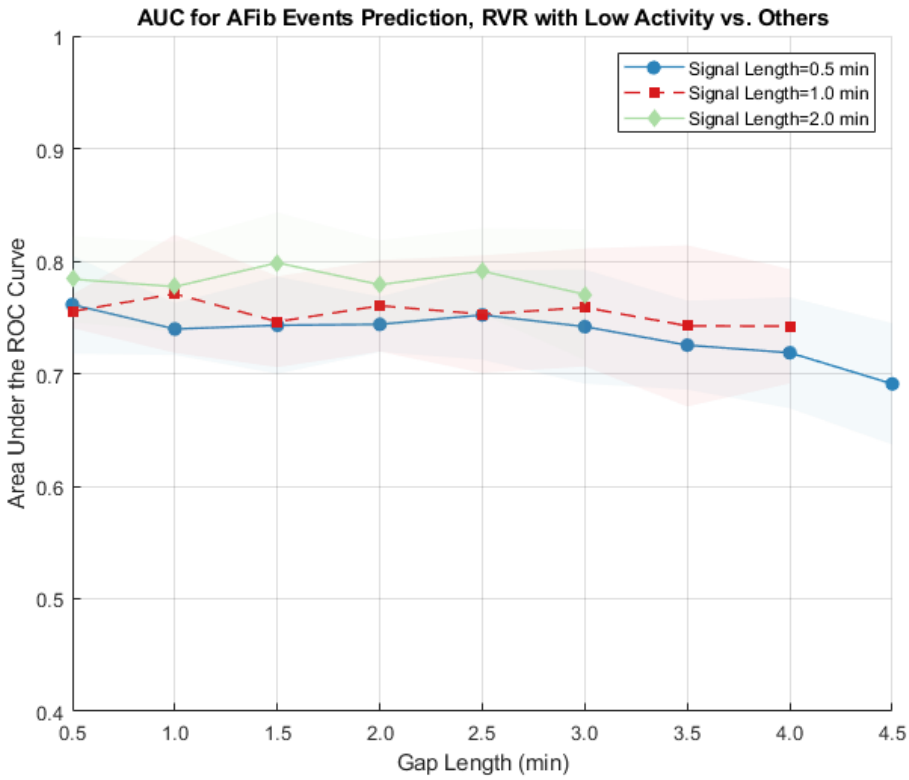


Figure 5.26: RVR with low activity prediction, AUC for RVR with low activity labeled by group level MAD vs. Others with various prediction intervals.



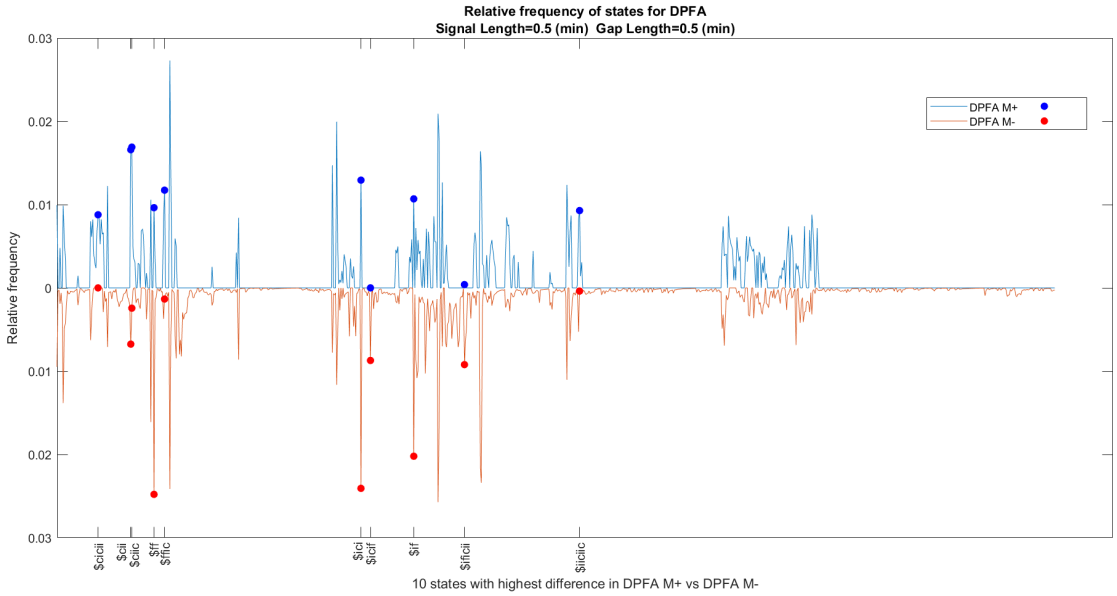
AF duration of at least 30 seconds, trained with nested cross validation, activity labeled by group level MAD.

Figure 5.27: RVR with low activity prediction, AUC for RVR with low activity labeled with participant level MAD vs. Others with various prediction intervals.



AF duration of at least 30 seconds, trained with nested cross validation, activity labeled by participant level MAD.

Figure 5.28: RVR with low activity, relative frequency of states for DPFA models.



Example, Signal Length = 0.5 (mins) , Gap Length= 0.5 (mins).

CHAPTER VI

Summary, Implications and Conclusion

In this paper a novel DPFA based method is presented to classify and predict cardiac arrhythmia events. The proposed method takes a probabilistic string extracted from physiological signals such as ECG in the training set as input, and via frequency analysis, constructs the underlying state space and transition probabilities of the DPFA model, directly from the input data. When applying the DPFA algorithm to classification and prediction problems, the decisions are based on comparing goodness-of-fit between the testing signal and various DPFA constructed from the training signals.

A collection of datasets have been constructed for this study, including publicly available benchmark datasets (DB1) for algorithm development and arrhythmia events detection; a retrospective database collected with data at hospital bed-side (DB2); prospective databases (DB4 and DB5) with signals collected from portable devices and database with AFib patients with signals with portable devices (DB6).

Detection and prediction experiments have been conducted on various cardiac arrhythmia types. The proposed algorithm has achieved an AUC of 0.95 in DB1 for AFib detection; for AFib prediction the algorithm has achieved greater than 0.80 in DB2 prediction with hospital bed-side data, AUC results are in the range of 0.79 to 0.87 for data collected from portable devices and 0.61 to 0.73 with data collected

while in-vehicle. The proposed algorithm has achieved an AUC of 0.95 in DB1 for SVT detection; for SVT prediction the algorithm has achieved greater than 0.75 in DB2 prediction with hospital bed-side data, AUC results in the range of 0.78 to 0.92 for data collected from portable devices. Prediction experiment for SVT has not been performed with in-vehicle data due to limited sample sizes. The proposed algorithm has achieved an AUC of 0.97 in DB1 for VA detection; for VT prediction the algorithm achieved around 0.75 in DB2 prediction with hospital bed-side data. VT prediction experiment has not been performed with data collected from portable devices due to limited sample sizes. The proposed algorithm has achieved an AUC around 0.71 to 0.74 for 2-minute-long signals for bradycardia prediction with hospital bed-side data. Bradycardia prediction experiment has not been performed with data collected from portable devices due to limited sample sizes. For prediction of RVR with low activity events the model has achieved an overall AUC around 0.70 to 0.80 depending on the lengths and gap sizes of the prediction intervals.

Comparing with other well-established methods, the proposed DPFA algorithm has achieved equal or better classification results, even though in some cases the advantage might not be statistically significant. In addition, the performance of the proposed DPFA algorithm is almost identical with or without any pre-processing on the data.

One of the limitations of our study is the sample size. Lack of events, especially in-vehicle arrhythmia events, has hampered the performance of the algorithm. Secondly, a lack of clinician reviewed annotations in databases has weaken the generalizability of the algorithms. To overcome the annotation issue, carefully reviewed and conducted automated annotation algorithms have been developed for various types of AFib events. For data collected from the portable devices, FDA cleared Preventice's annotations have been used as labels. In the future study, a dataset with a larger number of clinician annotated events will be helpful to further validate and compare

the monitoring system with existing methods. Also, in this thesis the prediction results have been presented with events upto 5 minutes prior to the events, in future studies a wider range of prediction intervals should be considered. The prediction of RVR with low activity level experiment uses synchronized ECG with accelerometer signals. In future studies, more types of physiological signals could be used since our algorithm could be adapted for different types of signals.

To conclude, The work in the thesis could be deployed as a cardiac arrhythmia monitoring and severe event prediction system which could alert patients and clinicians of an impending event, thereby enabling timely medical interventions.

BIBLIOGRAPHY

BIBLIOGRAPHY

Acharya, U. R., H. Fujita, O. S. Lih, Y. Hagiwara, J. H. Tan, and M. Adam (2017a), Automated detection of arrhythmias using different intervals of tachycardia ecg segments with convolutional neural network, *Information sciences*, *405*, 81–90.

Acharya, U. R., S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, A. Gertych, and R. San Tan (2017b), A deep convolutional neural network model to classify heartbeats, *Computers in biology and medicine*, *89*, 389–396.

Al-Khatib, S. M., et al. (2018), 2017 aha/acc/hrs guideline for management of patients with ventricular arrhythmias and the prevention of sudden cardiac death: a report of the american college of cardiology/american heart association task force on clinical practice guidelines and the heart rhythm society, *Journal of the American College of Cardiology*, *72*(14), e91–e220.

Altuve, M., G. Carrault, A. Beuchee, P. Pladys, and A. I. Hernández (2011), On-line apnea-bradycardia detection using hidden semi-markov models, in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4374–4377, IEEE.

Andrade, J. G., L. Macle, S. Nattel, A. Verma, and J. Cairns (2017), Contemporary atrial fibrillation management: a comparison of the current aha/acc/hrs, ccs, and esc guidelines, *Canadian Journal of Cardiology*, *33*(8), 965–976.

Andreao, R. V., B. Dorizzi, and J. Boudy (2006), Ecg signal analysis through hidden markov models, *IEEE Transactions on Biomedical engineering*, *53*(8), 1541–1549.

Andreotti, F., O. Carr, M. A. Pimentel, A. Mahdi, and M. De Vos (2017), Comparing feature-based classifiers and convolutional neural networks to detect arrhythmia from short segments of ecg, in *2017 Computing in Cardiology (CinC)*, pp. 1–4, IEEE.

Apandi, Z. F. M., R. Ikeura, and S. Hayakawa (2018), Arrhythmia detection using mit-bih dataset: A review, in *2018 International Conference on Computational Approach in Smart Systems Design and Applications (ICASSDA)*, pp. 1–5, IEEE.

Ashtiyani, M., S. N. Lavasani, A. A. Alvar, and M. Deevband (2018), Heart rate variability classification using support vector machine and genetic algorithm, *Journal of Biomedical Physics & Engineering*, *8*(4), 423.

Asl, B. M., S. K. Setarehdan, and M. Mohebbi (2008), Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal, *Artificial intelligence in medicine*, *44*(1), 51–64.

Bakrania, K., T. Yates, A. V. Rowlands, D. W. Esliger, S. Bunnewell, J. Sanders, M. Davies, K. Khunti, and C. L. Edwardson (2016), Intensity thresholds on raw acceleration data: Euclidean norm minus one (enmo) and mean amplitude deviation (mad) approaches, *PloS one*, *11*(10), e0164,045.

Bakshi, S., T. Feng, D. Chen, and W. Li (2020), Real-time bradycardia prediction in preterm infants using a dynamic system identification approach, *Journal of Engineering and Science in Medical Diagnostics and Therapy*, *3*(1).

Bazi, Y., N. Alajlan, H. AlHichri, and S. Malek (2013), Domain adaptation methods for ecg classification, in *2013 international conference on computer medical applications (ICCMA)*, pp. 1–4, IEEE.

Bousseljot, R., D. Kreiseler, and A. Schnabel (1995), Nutzung der ekg-signaldatenbank cardiodat der ptb über das internet.

Brown, P. F., P. V. Desouza, R. L. Mercer, V. J. D. Pietra, and J. C. Lai (1992), Class-based n-gram models of natural language, *Computational linguistics*, *18*(4), 467–479.

Calkins, H., et al. (2012), 2012 hrs/ehra/ecas expert consensus statement on catheter and surgical ablation of atrial fibrillation: recommendations for patient selection, procedural techniques, patient management and follow-up, definitions, endpoints, and research trial design: a report of the heart rhythm society (hrs) task force on catheter and surgical ablation of atrial fibrillation. developed in partnership with the european heart rhythm association (ehra), a registered branch of the european society of cardiology (esc) and the european cardiac arrhythmia society (ecas); and in collaboration with the american college of cardiology (acc), american heart association (aha), the asia pacific heart rhythm society (aphrs), and the society of thoracic surgeons (sts). endorsed by the governing bodies of the american college of cardiology foundation, the american heart association, the european cardiac arrhythmia society, the european heart rhythm association, the society of thoracic surgeons, the asia pacific heart rhythm society, and the heart rhythm society, *Europace*, *14*(4), 528–606.

Camm, A. J., et al. (2010), Guidelines for the management of atrial fibrillation: the task force for the management of atrial fibrillation of the european society of cardiology (esc), *European heart journal*, *31*(19), 2369–2429.

Ceylan, R., and Y. Özbay (2007), Comparison of fcm, pca and wt techniques for classification ecg arrhythmias using artificial neural network, *Expert Systems with Applications*, *33*(2), 286–295.

Chauhan, S., and L. Vig (2015), Anomaly detection in ecg time signals via deep long short-term memory networks, in *Data Science and Advanced Analytics (DSAA), 2015. 36678 2015. IEEE International Conference on*, pp. 1–7, IEEE.

Chawla, N. V., K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer (2002), Smote: synthetic minority over-sampling technique, *Journal of artificial intelligence research*, *16*, 321–357.

Coast, D. A., R. M. Stern, G. G. Cano, and S. A. Briller (1990), An approach to cardiac arrhythmia analysis using hidden markov models, *IEEE Transactions on biomedical Engineering*, *37*(9), 826–836.

Cruz, J., A. Hernández, S. Wong, G. Carrault, and A. Beuchee (2006), Algorithm fusion for the early detection of apnea-bradycardia in preterm infants, in *2006 Computers in Cardiology*, pp. 473–476, IEEE.

Cullington, D., K. M. Goode, A. L. Clark, and J. G. Cleland (2012), Heart rate achieved or beta-blocker dose in patients with chronic heart failure: which is the better target?, *European journal of heart failure*, *14*(7), 737–747.

Das, S., B. Moraffah, A. Banerjee, S. K. Gupta, and A. Papandreou-Suppappola (2019), Bradycardia prediction in preterm infants using nonparametric kernel density estimation, in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, pp. 1309–1313, IEEE.

Dorohonceanu, B., and C. G. Nevill-Manning (2000), Accelerating protein classification using suffix trees., in *ISMB*, vol. 2000, pp. 128–133.

Doshi-Velez, F., and B. Kim (2017), Towards a rigorous science of interpretable machine learning, *arXiv preprint arXiv:1702.08608*.

Elhaj, F. A., N. Salim, A. R. Harris, T. T. Swee, and T. Ahmed (2016), Arrhythmia recognition and classification using combined linear and nonlinear features of ecg signals, *Computer methods and programs in biomedicine*, *127*, 52–63.

Esposito, Y., A. Lemay, F. Denis, and P. Dupont (2002), Learning probabilistic residual finite state automata, in *International Colloquium on Grammatical Inference*, pp. 77–91, Springer.

Exarchos, T., M. Tsipouras, D. Nanou, C. Bazios, Y. Antoniou, and D. Fotiadis (2005), A platform for wide scale integration and visual representation of medical intelligence in cardiology: the decision support framework, in *Computers in Cardiology, 2005*, pp. 167–170, IEEE.

Faust, O., Y. Hagiwara, T. J. Hong, O. S. Lih, and U. R. Acharya (2018), Deep learning for healthcare applications based on physiological signals: a review, *Computer methods and programs in biomedicine*.

Faust, O., E. J. Ciaccio, and U. R. Acharya (2020), A review of atrial fibrillation detection methods as a service, *International Journal of Environmental Research and Public Health*, *17*(9), 3093.

Fuster, V., et al. (2006), Acc/aha/esc 2006 guidelines for the management of patients with atrial fibrillation, *Circulation*, *114*(7), e257–e354.

Gao, L., A. Bourke, and J. Nelson (2014), Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems, *Medical engineering & physics*, *36*(6), 779–785.

Ge, D., N. Srinivasan, and S. M. Krishnan (2002), Cardiac arrhythmia classification using autoregressive modeling, *Biomedical engineering online*, *1*(1), 5.

Gee, A. H., R. Barbieri, D. Paydarfar, and P. Indic (2016), Predicting bradycardia in preterm infants using point process analysis of heart rate, *IEEE Transactions on Biomedical Engineering*, *64*(9), 2300–2308.

Ghahjaverestan, N. M., S. Masoudi, M. B. Shamsollahi, A. Beuchee, P. Pladys, D. Ge, and A. I. Hernández (2015a), Coupled hidden markov model-based method for apnea bradycardia detection, *IEEE journal of biomedical and health informatics*, *20*(2), 527–538.

Ghahjaverestan, N. M., M. B. Shamsollahi, D. Ge, and A. I. Hernández (2015b), Switching kalman filter based methods for apnea bradycardia detection from ecg signals, *Physiological measurement*, *36*(9), 1763.

Ghahjaverestan, N. M., M. B. Shamsollahi, D. Ge, A. Beuchée, and A. I. Hernández (2021), Apnea bradycardia detection based on new coupled hidden semi markov model, *Medical & Biological Engineering & Computing*, *59*(1), 1–11.

Gjoreski, H., A. Rashkovska, S. Kozina, M. Lustrek, and M. Gams (2014), Telehealth using ecg sensor and accelerometer, in *2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 270–274, IEEE.

Godfrey, A., R. Conway, D. Meagher, and G. ÓLaighin (2008), Direct measurement of human movement by accelerometry, *Medical engineering & physics*, *30*(10), 1364–1386.

Godfrey, A., A. Bourke, G. Olaighin, P. Van De Ven, and J. Nelson (2011), Activity classification using a single chest mounted tri-axial accelerometer, *Medical engineering & physics*, *33*(9), 1127–1135.

Goldberger, A. L., et al. (2000), Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals, *Circulation*, *101*(23), e215–e220.

Gomes, P. R., F. O. Soares, J. Correia, and C. Lima (2010), Ecg data-acquisition and classification system by using wavelet-domain hidden markov models, in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pp. 4670–4673, IEEE.

Grant, C. C., C. Murray, D. C. Janse Van Rensburg, and L. Fletcher (2013), A comparison between heart rate and heart rate variability as indicators of cardiac health and fitness, *Frontiers in physiology*, *4*, 337.

Greenwald, S. D. (1986), The development and analysis of a ventricular fibrillation detector, Ph.D. thesis, Massachusetts Institute of Technology.

Greenwald, S. D., R. S. Patil, and R. G. Mark (1990), *Improved detection and classification of arrhythmias in noise-corrupted electrocardiograms using contextual information*, IEEE.

Hakkinen, J., and J. Tian (2001), N-gram and decision tree based language identification for written words, in *IEEE Workshop on Automatic Speech Recognition and Understanding, 2001. ASRU'01.*, pp. 335–338, IEEE.

Hannun, A. Y., P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng (2019), Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network, *Nature medicine*, *25*(1), 65.

Hnatkova, K., X. Copie, A. Staunton, and M. Malik (1995), Numeric processing of lorenz plots of rr intervals from long-term ecgs: comparison with time-domain measures of heart rate variability for risk stratification after myocardial infarction, *Journal of electrocardiology*, *28*, 74–80.

Hosseini, A., and M. Sarrafzadeh (2019), Unsupervised prediction of negative health events ahead of time, in *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pp. 1–4, IEEE.

Jadhav, S. M., S. L. Nalbalwar, and A. A. Ghatol (2011), Modular neural network based arrhythmia classification system using ecg signal data, *International Journal of Information Technology and Knowledge Management*, *4*(1), 205–209.

January, C. T., et al. (2014), 2014 aha/acc/hrs guideline for the management of patients with atrial fibrillation: a report of the american college of cardiology/american heart association task force on practice guidelines and the heart rhythm society, *Journal of the American College of Cardiology*, *64*(21), e1–e76.

Jiang, X., L. Zhang, Q. Zhao, and S. Albayrak (2006), Ecg arrhythmias recognition system based on independent component analysis feature extraction, in *TENCON 2006. 2006 IEEE Region 10 Conference*, pp. 1–4, IEEE.

John, R. M., U. B. Tedrow, B. A. Koplan, C. M. Albert, L. M. Epstein, M. O. Sweeney, A. L. Miller, G. F. Michaud, and W. G. Stevenson (2012), Ventricular arrhythmias and sudden cardiac death, *The Lancet*, *380*(9852), 1520–1529.

Johnson, A. E., et al. (2016), Mimic-iii, a freely accessible critical care database, *Scientific data*, *3*, 160,035.

Kannel, W. B., P. A. Wolf, E. J. Benjamin, and D. Levy (1998), Prevalence, incidence, prognosis, and predisposing conditions for atrial fibrillation: population-based estimates 1, *American Journal of Cardiology*, *82*(7), 2N–9N.

Kiranyaz, S., T. Ince, and M. Gabbouj (2016), Real-time patient-specific ecg classification by 1-d convolutional neural networks, *IEEE Transactions on Biomedical Engineering*, *63*(3), 664–675.

Kochiadakis, G., E. Skalidis, M. Kalebubas, N. Igoumenidis, S. Chrysostomakis, E. Kanoupakis, E. Simantirakis, and P. Vardas (2002), Effect of acute atrial fibrillation on phasic coronary blood flow pattern and flow reserve in humans, *European heart journal*, *23*(9), 734–741.

Kusumoto, F. M., et al. (2019), 2018 acc/aha/hrs guideline on the evaluation and management of patients with bradycardia and cardiac conduction delay: a report of the american college of cardiology/american heart association task force on clinical practice guidelines and the heart rhythm society, *Journal of the American College of Cardiology*, *74*(7), e51–e156.

Lankveld, T. A., S. Zeemering, H. J. Crijns, and U. Schotten (2014), The ecg as a tool to determine atrial fibrillation complexity, *Heart*, *100*(14), 1077–1084.

Li, Z., H. Derksen, J. Gryak, H. Ghanbari, P. Gunaratne, and K. Najarian (2018), A novel atrial fibrillation prediction algorithm applicable to recordings from portable devices, in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4034–4037, IEEE.

Li, Z., H. Derksen, J. Gryak, M. Hooshmand, A. Wood, H. Ghanbari, P. Gunaratne, and K. Najarian (2019), Markov models for detection of ventricular arrhythmia, in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1488–1491, IEEE.

Li, Z., H. Derksen, J. Gryak, C. Jiang, Z. Gao, W. Zhang, H. Ghanbari, P. Gunaratne, and K. Najarian (2021), Prediction of cardiac arrhythmia using deterministic probabilistic finite-state automata, *Biomedical Signal Processing and Control*, *63*, 102,200.

Luz, E. J. d. S., T. M. Nunes, V. H. C. De Albuquerque, J. P. Papa, and D. Menotti (2013), Ecg arrhythmia classification based on optimum-path forest, *Expert Systems with Applications*, *40*(9), 3561–3573.

Luz, E. J. d. S., W. R. Schwartz, G. Cámara-Chávez, and D. Menotti (2016), Ecg-based heartbeat classification for arrhythmia detection: A survey, *Computer methods and programs in biomedicine*, *127*, 144–164.

Macle, L., et al. (2016), 2016 focused update of the canadian cardiovascular society guidelines for the management of atrial fibrillation, *Canadian Journal of Cardiology*, *32*(10), 1170–1185.

Mahmoodabadi, S., A. Ahmadian, and M. Abolhasani (2005), Ecg feature extraction using daubechies wavelets, in *Proceedings of the fifth IASTED International conference on Visualization, Imaging and Image Processing*, pp. 343–348.

Mangrum, J. M., and J. P. DiMarco (2000), The evaluation and management of bradycardia, *New England Journal of Medicine*, *342*(10), 703–709.

Martis, R. J., U. R. Acharya, K. Mandana, A. K. Ray, and C. Chakraborty (2012), Application of principal component analysis to ecg signals for automated diagnosis of cardiac health, *Expert Systems with Applications*, *39*(14), 11,792–11,800.

Mei, Z., X. Gu, H. Chen, and W. Chen (2018), Automatic atrial fibrillation detection based on heart rate variability and spectral features, *IEEE Access*, *6*, 53,566–53,575.

Moody, G. (1983), A new method for detecting atrial fibrillation using rr intervals, *Computers in Cardiology*, pp. 227–230.

Myerburg, R. J. (2001), Sudden cardiac death: exploring the limits of our knowledge, *Journal of cardiovascular electrophysiology*, *12*(3), 369–381.

Myerburg, R. J., and M. J. Junttila (2012), Sudden cardiac death caused by coronary heart disease, *Circulation*, *125*(8), 1043–1052.

Nagendran, M., et al. (2020), Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies, *bmj*, *368*.

Naseri, E., A. Ghaffari, and M. Abdollahzade (2019), A novel ica-based clustering algorithm for heart arrhythmia diagnosis, *Pattern Analysis and Applications*, *22*(2), 285–297.

Nasiri, J. A., M. Naghibzadeh, H. S. Yazdi, and B. Naghibzadeh (2009), Ecg arrhythmia classification with support vector machines and genetic algorithm, in *Computer Modeling and Simulation, 2009. EMS'09. Third UKSim European Symposium on*, pp. 187–192, IEEE.

Ney, H., S. Martin, and F. Wessel (1997), Statistical language modeling using leaving-one-out, in *Corpus-based methods in Language and Speech processing*, pp. 174–207, Springer.

Nolle, F., F. Badura, J. Catlett, R. Bowser, and M. Sketch (1986), Crei-gard, a new concept in computerized arrhythmia monitoring systems, *Computers in Cardiology*, *13*, 515–518.

Oğul, H., and E. Ü. Mumcuoğlu (2006), Svm-based detection of distant protein structural relationships using pairwise probabilistic suffix trees, *Computational biology and chemistry*, *30*(4), 292–299.

Oh, S. L., E. Y. Ng, R. San Tan, and U. R. Acharya (2018), Automated diagnosis of arrhythmia using combination of cnn and lstm techniques with variable length heart beats, *Computers in biology and medicine*.

Olshansky, B., et al. (2004), The atrial fibrillation follow-up investigation of rhythm management (affirm) study: approaches to control rate in atrial fibrillation, *Journal of the American College of Cardiology*, *43*(7), 1201–1208.

Ortman, J. M., V. A. Velkoff, H. Hogan, et al. (2014), *An aging nation: the older population in the United States*, United States Census Bureau, Economics and Statistics Administration, US Department of Commerce.

Özbay, Y., R. Ceylan, and B. Karlik (2006), A fuzzy clustering neural network architecture for classification of ecg arrhythmias, *Computers in Biology and Medicine*, *36*(4), 376–388.

Ozcan, N. O., and F. Gurgen (2010), Fuzzy support vector machines for ecg arrhythmia detection, in *Pattern Recognition (ICPR), 2010 20th International Conference on*, pp. 2973–2976, IEEE.

Pan, J., and W. J. Tompkins (1985), A real-time qrs detection algorithm, *IEEE transactions on biomedical engineering*, (3), 230–236.

Passman, R., and R. A. Bernstein (2016), New appraisal of atrial fibrillation burden and stroke prevention, *Stroke*, *47*(2), 570–576.

Petrutiu, S., A. V. Sahakian, and S. Swiryn (2007), Abrupt changes in fibrillatory wave characteristics at the termination of paroxysmal atrial fibrillation in humans, *Europace*, *9*(7), 466–470.

Pravisani, G., A. Beuchee, L. Mainardi, and G. Carrault (2003), Short term prediction of severe bradycardia in premature newborns, in *Computers in Cardiology, 2003*, pp. 725–728, IEEE.

Purwar, A., D. U. Jeong, and W. Y. Chung (2007), Activity monitoring from real-time triaxial accelerometer data using sensor network, in *2007 International conference on control, automation and systems*, pp. 2402–2406, IEEE.

Rajpurkar, P., A. Y. Hannun, M. Haghpanahi, C. Bourn, and A. Y. Ng (2017), Cardiologist-level arrhythmia detection with convolutional neural networks, *arXiv preprint arXiv:1707.01836*.

Rizwan, A., A. Zoha, I. Mabrouk, H. Sabbour, A. Al-Sumaiti, A. Alomaniy, M. Imran, and Q. Abbasi (2020), A review on the state of the art in atrial fibrillation detection enabled by machine learning, *IEEE Reviews in Biomedical Engineering*.

Sadoughi, A., M. B. Shamsollahi, E. Fatemizadeh, A. Beuchée, A. I. Hernández, and N. M. Ghahjaverestan (2021), Detection of apnea bradycardia from ecg signals of preterm infants using layered hidden markov model, *Annals of biomedical engineering*, pp. 1–11.

Sanders, P., P. M. Kistler, J. B. Morton, S. J. Spence, and J. M. Kalman (2004), Remodeling of sinus node function in patients with congestive heart failure: reduction in sinus node reserve, *Circulation*, *110*(8), 897–903.

Sarkar, S., D. Ritscher, and R. Mehra (2008), A detector for a chronic implantable atrial tachyarrhythmia monitor, *IEEE Transactions on Biomedical Engineering*, *55*(3), 1219–1224.

Saul, L., and F. Pereira (1997), Aggregate and mixed-order markov models for statistical language processing, *arXiv preprint cmp-lg/9706007*.

Schijvenaars, B. J., G. van Herpen, and J. A. Kors (2008), Intraindividual variability in electrocardiograms, *Journal of Electrocardiology*, *41*(3), 190–196.

Schocken, D. D., M. I. Arrieta, P. E. Leaverton, and E. A. Ross (1992), Prevalence and mortality rate of congestive heart failure in the united states, *Journal of the American College of Cardiology*, *20*(2), 301–306.

Sedghamiz, H. (2014), Matlab implementation of pan tompkins ecg qrs detector.

Semaan, S., T. A. Dewland, G. H. Tison, G. Nah, E. Vittinghoff, M. J. Pletcher, J. E. Olgin, and G. M. Marcus (2020), Physical activity and atrial fibrillation: Data from wearable fitness trackers, *Heart rhythm*, *17*(5), 842–846.

Shaffer, F., and J. Ginsberg (2017), An overview of heart rate variability metrics and norms, *Frontiers in public health*, *5*, 258.

Shuai, W., X.-x. Wang, K. Hong, Q. Peng, J.-x. Li, P. Li, J. Chen, X.-s. Cheng, and H. Su (2016), Is 10-second electrocardiogram recording enough for accurately estimating heart rate in atrial fibrillation, *International journal of cardiology*, *215*, 175–178.

Skinner Jr, N. S., J. H. Mitchell, A. G. Wallace, and S. J. Sarnoff (1964), Hemodynamic consequences of atrial fibrillation at constant ventricular rates, *The American journal of medicine*, *36*(3), 342–350.

Sohinki, D., and O. A. Obel (2014), Current trends in supraventricular tachycardia management, *The Ochsner Journal*, *14*(4), 586–595.

Song, M. H., J. Lee, S. P. Cho, K. J. Lee, and S. K. Yoo (2005), Support vector machine based arrhythmia classification using reduced features, *International Journal of Control, Automation, and Systems*, *3*(4), 571–579.

Steinberg, B. A., et al. (2013), Use and associated risks of concomitant aspirin therapy with oral anticoagulation in patients with atrial fibrillation: insights from the outcomes registry for better informed treatment of atrial fibrillation (orbit-af) registry, *Circulation*, *128*(7), 721–728.

Taddei, A., G. Distante, M. Emdin, P. Pisani, G. Moody, C. Zeelenberg, and C. Marchesi (1992), The european st-t database: standard for evaluating systems for the analysis of st-t changes in ambulatory electrocardiography, *European heart journal*, *13*(9), 1164–1172.

Tan, J. H., Y. Hagiwara, W. Pang, I. Lim, S. L. Oh, M. Adam, R. San Tan, M. Chen, and U. R. Acharya (2018), Application of stacked convolutional and long short-term memory network for accurate identification of cad ecg signals, *Computers in biology and medicine*, *94*, 19–26.

Teplitzky, B. A., M. McRoberts, and H. Ghanbari (2020), Deep learning for comprehensive ecg annotation, *Heart rhythm*, *17*(5), 881–888.

Topol, E. J. (2019), High-performance medicine: the convergence of human and artificial intelligence, *Nature medicine*, *25*(1), 44–56.

Tsipouras, M. G., D. I. Fotiadis, and D. Sideris (2002), Arrhythmia classification using the rr-interval duration signal, in *Computers in Cardiology, 2002*, pp. 485–488, IEEE.

Van Gelder, I. C., et al. (2010), Lenient versus strict rate control in patients with atrial fibrillation, *New England Journal of Medicine*, *362*(15), 1363–1373.

Vidal, E., F. Thollard, C. De La Higuera, F. Casacuberta, and R. C. Carrasco (2005), Probabilistic finite-state machines-part i, *IEEE transactions on pattern analysis and machine intelligence*, *27*(7), 1013–1025.

Wang, T. J., et al. (2003), Temporal relations of atrial fibrillation and congestive heart failure and their joint influence on mortality: the framingham heart study, *Circulation*, *107*(23), 2920–2925.

Warrick, P., and M. N. Homsi (2017), Cardiac arrhythmia detection from ecg combining convolutional and long short-term memory networks, *Computing*, *44*, 1.

Wolf, P. A., R. D. Abbott, and W. B. Kannel (1991), Atrial fibrillation as an independent risk factor for stroke: the framingham study., *Stroke*, *22*(8), 983–988.

Yang, C.-C., and Y.-L. Hsu (2010), A review of accelerometry-based wearable motion detectors for physical activity monitoring, *Sensors*, *10*(8), 7772–7788.

Ye, C., M. T. Coimbra, and B. V. Kumar (2010), Arrhythmia detection and classification using morphological and dynamic features of ecg signals, in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pp. 1918–1921, IEEE.

Yildirim, Ö. (2018), A novel wavelet sequence based on deep bidirectional lstm network model for ecg signal classification, *Computers in biology and medicine*, *96*, 189–202.

Yu, S.-N., and K.-T. Chou (2009), Selection of significant independent components for ecg beat classification, *Expert Systems with Applications*, *36*(2), 2088–2096.

Zhang, C., G. Wang, J. Zhao, P. Gao, J. Lin, and H. Yang (2017), Patient-specific ecg classification based on recurrent neural networks and clustering technique, in *Biomedical Engineering (BioMed), 2017 13th IASTED International Conference on*, pp. 63–67, IEEE.

Zhao, Q., and L. Zhang (2005), Ecg feature extraction and classification using wavelet transform and support vector machines, in *Neural Networks and Brain, 2005. ICNN&B'05. International Conference on*, vol. 2, pp. 1089–1092, IEEE.