FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

BACHELOR THESIS

Glejdis Shkëmbi

Machine Learning Tools for Diagnosis of Heart Arrhythmia

Department of Theoretical Computer Science and Mathematical Logic

Supervisor of the bachelor thesis: Mgr. Marta Vomlelová, Ph.D.
Study programme: Computer Science
Study Branch: General Computer Science

Prague 2020

I declare that I carried out this bachelor thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

In…...... date........... signature

I would like to express my sincere gratitude to my advisor, Mgr. Marta Vomlelová, Ph.D. Your guidance, professionalism and valuable suggestions throughout the making of this thesis played an essential role. I feel privileged to have had the opportunity to take this journey with your knowledge and support.

A very special thanks goes to my dear family for their unlimited support, their encouragement and kindness throughout all my years of study. I will be entirely grateful for believing in me.

Lastly, I thank my friends for their willingness to share their most sincere feedback and advice with me. I could not have asked for better friends.

Title: Machine Learning Tools for Diagnosis of Heart Arrhythmia

Author: Glejdis Shkëmbi

Department / Institute: Department of Theoretical Computer Science and Mathematical Logic

Supervisor of the bachelor thesis: Mgr. Marta Vomlelová, Ph.D., Department of Theoretical Computer Science and Mathematical Logic

Abstract: Electrocardiogram (ECG) is considered to be the most reliable, efficient and low-cost tool used in the healthcare industry to diagnose cardiac arrhythmia. However, visual representation of ECG signals manually by medical workers is intricate and time-consuming, and may lead to human mistakes and inaccuracy in heartbeat recognition. In this paper, different machine learning techniques for the classification of five classes of ECG heartbeats using Discrete Wavelet Transform (DWT) features are compared. In particular, the significant role of statistical features of DWT coefficients in distinguishing between different heartbeat classes is highlighted. Performances of the models have been evaluated using the online MIT-BIH arrhythmia database. The obtained results indicate the reliability of the machine learning-based approaches for diagnoses of cardiac arrhythmia from ECG signals.

Keywords: Electrocardiogram (ECG); Discrete Wavelet Transform (DWT); Support Vector Machine (SVM); Random Forest; Heart Arrhythmia.

Contents

List of Abbreviations

AFIB - Atrial Fibrillation

AFL - Atrial Flutter

APC - Atrial Premature Contraction

AV - Atrioventricular Node

CART - Classification and Regression Tree

CARET - Classification and Regression Training

CP - Complexity Parameter

DCT - Discrete Cosine Transform

DWT - Discrete Wavelet Transform

ECG - Electrocardiography

F - Fusion beat

FN - False Negatives

FP - False Positives

ICA - Independent Component Analysis

LA - Left Atrium

LBBB - Left Bundle Branch Block

LV - Left Ventricle

MI - Myocardial Infarction

ML - Machine Learning

MSPCA - Multiscale Principal Component Analysis

N - Normal beat

NE - Non-Ectopic beat

PCA - Principal Component Analysis

PSVT - Paroxysmal Supraventricular Tachycardia

PVC - Premature Ventricular Complex

RA - Right Atrium

RBBB - Right Bundle Branch Block

RF - Random Forest

RV - Right Ventricle

S - Supraventricular ectopic beat

SA - Sinoatrial block

SBR - Sinus bradycardia

SNR - Signal-to-Noise Ratio

SVM - Support Vector Machine

SVTA - Supraventricular Tachyarrhythmia

TN - True Negatives

TP - True Positives

U - Unknown beat

V - Ventricular beat

VFB - Ventricular Fibrillation

VFL - Ventricular Flutter

VT - Ventricular Tachycardia

Introduction

According to the World Health Organisation (Moran, et al. 2018), cardiovascular diseases including heart disease and stroke are the leading cause of deaths worldwide, killing 17.62 million people in 2016. Unfortunately, due to the present lifestyle factors, this number continues to grow with an alarming rate all around the world. The heart is one of the most important organs in our body and the first organ developed in embryogenesis within three weeks. Beating approximately 72 times per minute, it pumps blood throughout our body supplying oxygen and nutrients to the tissues and removing carbon dioxide and other wastes (Thaler, 2015). Each time our heart beats, it creates energy in the form of electrical currents. In order to be able to recognise and diagnose different heart diseases, and measure the health of a patient's heart, it is important to be able to collect this data.

Electrocardiography (ECG) is the most reliable and low-cost method that registers the electrical activity of the heart, recording it as waveforms and generating a graph of voltage versus time. It is through the changes in the normal electrical patterns shown by the ECG which enable us to diagnose many different cardiac disorders (Thaler, 2015). Electrodes, placed on the surface of the chest, are able to detect the small changes in voltage during depolarization and repolarization of cardiomyocytes in each heart beat (Gacek & Pedrycz, 2012). Thaler (2015) recognizes the disturbance in the electrical flow through the heart, called arrhythmia, as the most common natural cause of sudden death in young people nowadays. However, interpretation of ECG signals by cardiologists is complicated and time consuming, and may lead to errors due to beat misclassification (Desai, et al., 2016). A wrong diagnosis of a patient will not only waste time and money, but it could lead to the sudden death of the patient. Therefore, several machine learning (ML) techniques are proposed for contributing to the clinical applications .

Machine learning is a branch of artificial intelligence that allows computers to learn directly from data and experience, and widely used to replace human decision making (Boutaba, et al. 2018). *"We are actually living in the data age"* says Jiawein

Han in his 2012 book *"Data Mining: Concepts and Techniques"*. Thousands of terabytes of data are being collected every day from businesses, society, science, engineering and medicine, and due to the increasing technological advances in all fields, analysing such data using various machine learning techniques plays an essential role in every aspect of our everyday life (Han, 2012). In the healthcare industry, these data collected from electronic patient records and diagnosis of diseases can be used in order to accurately predict the presence or absence of heart related diseases (Ramalingam, et al., 2018). As a result, significantly reducing the workload of cardiologists and enabling them to focus more on treatment rather than diagnosis.

As stated in many research papers, one of the biggest problems in building a computer-assisted detection of different arrhythmia types is the selection of the correct and appropriate machine learning technique which will successfully detect arrhythmia types. In addition, another problem stands in extracting the proper feature selection that will be used in training the classifier. In this paper, the ECG heartbeat signals, acquired from the online MIT-BIH arrhythmia database, are classified using four machine learning techniques, namely CART, C5.0, Random Forest and Support Vector Machine. In the first step, the Discrete Wavelet Transform is applied to decompose ECG signals into numerous wavelets at different frequency bands. Next, in order to have a better performance in heartbeat classification, statistical features are extracted from these frequency bands. In this study, only five types of ECG heartbeats are being analysed: Normal beats, Right Bundle Branch Blocks beats (RBBB), Left Bundle Branch Blocks beats (LBBB), Premature Ventricular Contractions (PVC) and Atrial Premature Contractions (APC) beats.

Outline

The Introductory Chapter provides a general sense of the topic by emphasizing the need of such machine learning approaches in the classification of arrhythmia types and the available tools to help us build accurate and successful arrhythmia diagnosis systems. Then the focus is shifted in the biggest problems faced when building such classifiers.

Chapter 1: The Electrocardiographic Interpretation describes the main theoretical characteristics of the cardiac muscle and the normal electrocardiogram. Moreover, we provide a brief explanation to all types of arrhythmias and their electrocardiographic interpretation, giving, in particular, a deeper explanation to the five types of heartbeats that will be analysed in this thesis.

Chapter 2: Materials and Methodology presents the dataset used, describes the methods used in data pre-processing, gives a brief explanation to DWT and how it will be used in this thesis, and shows the feature extraction methodology chosen in this study. Next, we analyse the machine learning classifiers used and the main reasons why we chose such classifiers. Lastly, ten-fold cross validation is introduced and evaluation metrics that are used to give us some easily comparable numbers are described.

Chapter 3: Experiments gives a full description of the proposed procedure, the experimental results and discussions by comparing the performance results obtained by the suggested system in this paper with the results obtained in other studies.

Conclusion summarises the main points made in this paper, and suggests one final model with the highest results for arrhythmia classification among all other methods used in our experiments.

Goals

The goals of this thesis are:

- to review current research conducted on MIT-BIH arrhythmia database
- to analyze the DWT (Discrete Wavelet Transform) as the only pre-processing method for arrhythmia classification on MIT-BIH arrhythmia database

1. The Electrocardiographic Interpretation

In order to build a successful ECG signal classifier, first, it is important to understand what causes arrhythmias, and how they can be detected. Therefore, in this chapter, we will provide a brief introduction on the electrocardiographic interpretation of the cardiac muscle, what are the main characteristics of a normal electrocardiogram, the main types of arrhythmias and how they can be recognised using ECG signal.

1.1. The Cardiac Muscle

The cardiac muscle is an involuntary muscle constructing the main tissue of the heart wall, and forming a thick middle layer between its outer and inner layer (Saxton, 2018). Cardiomyocytes are one of the main cells contained in the cardiac muscle. They are specialised in generating electrical potentials during contraction. At rest, cardiomyocytes are negatively charged inside with respect to their outside; that is, they are electrically polarized with an electrical membrane potential of about −90 mV (Gacek & Pedrycz, 2012). By letting ions pass in and out of the cell, the membrane pumps make sure that ions are well-distributed in order to keep their internal negativity. However, the cells can lose this electrical polarity through a fundamental electrical event of the heart called depolarization. The arrival of an electrical impulse that causes positively charged ions to cross the cell membrane leading to *depolarization* (Thaler, 2015). This produces a wave of electricity that is then transmitted across the entire heart. After depolarization is complete, the heart muscle cells return to their rest state by reversing the flow of ions, a process called *repolarization* (Thaler, 2015). The movement of ions, causing depolarization and repolarization of the cardiac cells, is in the center of the heart's electrical activity (Gacek & Pedrycz, 2012). Moreover, all of the different waves recorded on ECG are a presentation of these two processes (see Figure 1.1.2). These waves are characterized by three main attributes: duration, amplitude and configuration. Configuration refers to the shape of a wave (Thaler, 2015).

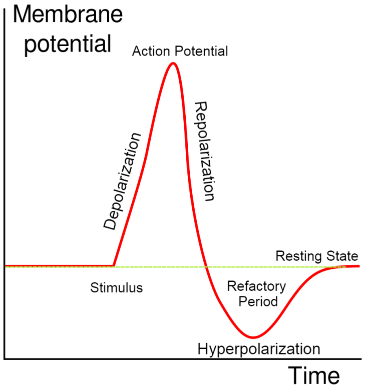


Figure 1.1.1 A labeled diagram of an action potential

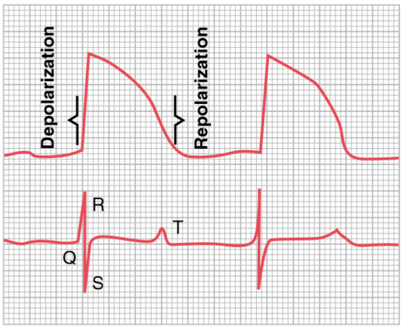


Figure 1.1.2 Above, depolarization and the repolarization. Below, Electrocardiogram recorded at the same time (Source: Thaler, 2015)

1.2. Characteristics of the Normal Electrocardiogram

The normal electrocardiogram (see Figure 1.2.1) is composed of the P wave, the QRS complex, and the T wave. The QRS complex consists of three separate waves: the Q wave, the R wave and the S wave (Hall, 2011). T waves are caused by the repolarization of the ventricle. The ventricle is the lower chamber of the heart that pumps blood from the heart into the lungs and the circulation system (Gacek & Pedrycz, 2012). It is composed of the right ventricle (RV) and the left ventricle (LV) (see Figure 1.2.1). The P wave and the waves of the QRS complex are both depolarization waves. Depolarization of the atria causes the P wave. The atria, which are called the upper two heart chambers, receive the blood that comes back from the body to the heart (Gacek & Pedrycz, 2012). Normally, the P wave lasts 0.12 seconds and with a maximum voltage of 0.25mV. If a P wave exceeds the values given above, then it is considered to be abnormal (Gacek & Pedrycz, 2012). The QRS complex, being the largest group of waves on the ECG, corresponds exactly to the depolarization of ventricles (see Figure 1.2.1). The normal duration of the QRS complex is considered to be 0.12 seconds. If the QRS complex lasts more than 0.12 seconds, then this shows signs of bundle branch block, pre-excitation syndromes, or premature ventricular contraction, and as we will see later on this paper, all three of them will appear in our chosen dataset (Gacek & Pedrycz, 2012).

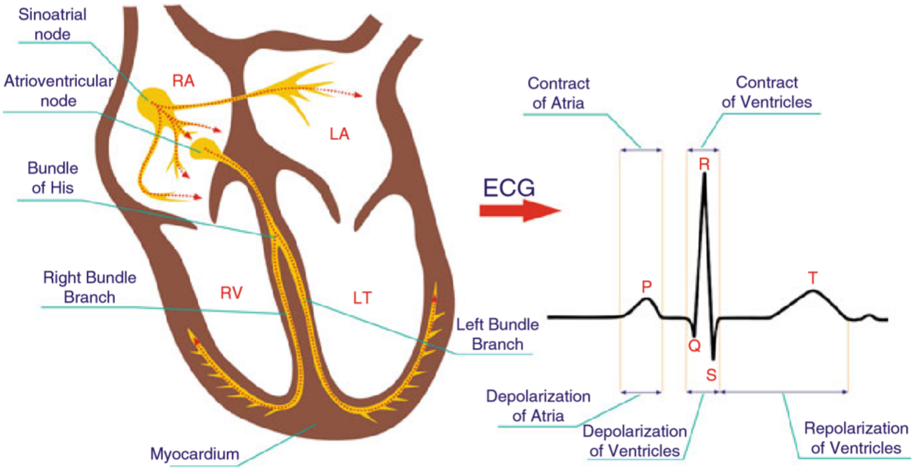


Figure 1.2.1 Propagation of the depolarization wave in the heart muscle (Source: Gacek & Pedrycz, 2012)

As it is shown in the heartbeats of Figure 1.2.2, the P-Q interval measures the time from the beginning of atrial depolarization to the start of ventricular depolarization. This occurs while the PQ segment measures the time from the end of atrial depolarization to the start of ventricular depolarization. The P-Q interval, which extends between the beginning of the P wave and the beginning of the QRS complex, normally lasts between 0.12 to 0.20 seconds (Gacek & Pedrycz, 2012). The ST segment, extending from the end of the QRS complex to the beginning of the T wave, measures the time between the end of ventricular depolarization and the start of ventricular repolarization (Hall, 2011). As it can be observed in both Figure 1.2.1 and Figure 1.2.2, the Q-T interval includes the QRS complex and the T wave, and represents the duration of the ventricular action potential and repolarisation. The Q-T interval's length is influenced directly by the heart rate. The faster the heart rate is, the shorter this interval will appear in the ECG signal (Gacek & Pedrycz, 2012). Moreover, extension of the duration of this interval more than 0.44 seconds corresponds to the increased risk of polymorphic tachycardia, which can cause unexpected cardiac death (Gacek & Pedrycz, 2012). In order to calculate the heart rate, the R-R interval is used. It represents one cardiac cycle (Gacek & Pedrycz, 2012).

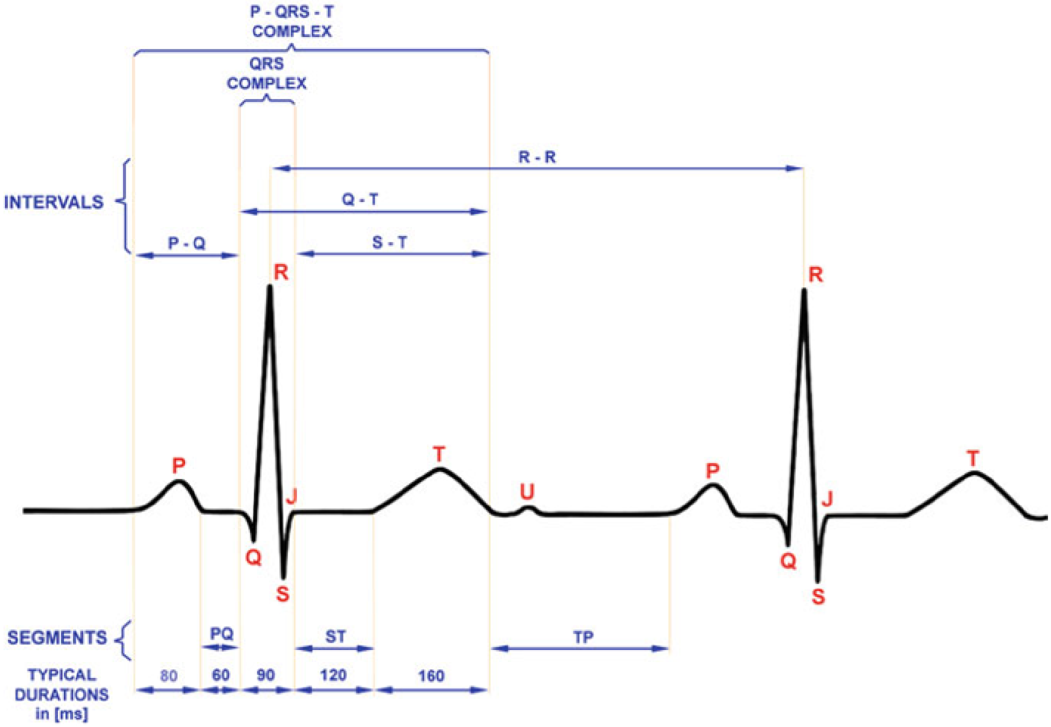


Figure 1.2.2 Normal electrocardiogram for two beats of the heart (Source: Gacek & Pedrycz, 2012)

1.3. Arrhythmias and Their Electrocardiographic Interpretation

ECG signal is a series of waves, where each individual heartbeat shows how the electrical activity of the heart is evolving with time from a certain view called a *lead* (Thaler, 2015). Each lead measures the voltage changes between electrodes at different positions in our body. Leads I, II, and III are called *bipolar leads*, and leads aVR, aVL, aVF, V1 to V6 are called *unipolar leads* (Thaler, 2015). By observing the electrocardiographic signal, we are able to detect any disorder of the heart rhythm or any abnormal change in the depolarization-repolarization pattern, which suggests a risk of cardiac arrhythmia (Gacek & Pedrycz, 2012). Therefore, an accurate real-time analysis and classification of ECG signals in clinical settings comes of significant help to doctors in identifying arrhythmias. Moreover, the diagnosis of an arrhythmia is one of the most important things an ECG can do, and it is considered to be the most reliable and cost-effective tool (Desai, et al, 2016). The purpose of this section is to discuss the physiology of cardiac arrhythmias and their diagnosis by electrocardiology.

Abnormality	P wave	P:QRS ratio	QRS regularity	QRS shape	QRS rate	Rhythm
Occasional				Normal		Supraventricular
				Abnormal		Ventricular
Sustained	Present	P:QRS = 1:1	Regular	Normal	Normal	Sinus rhythm
					>= 150/min	Atrial tachycardia
			Slightly irregular	Normal	Normal	Sinus arrhythmia
					Slow	Atrial escape
		More P waves than QRS complexes	Regular	Normal	Fast	Atrial tachycardia with block
					Slow	Second degree heart block
				Abnormal	Slow	Complete heart block
	Absent		Regular	Normal	Fast	Junctional tachycardia
					Slow	Junctional escape
				Abnormal	Fast	Junctional tachycardia with bundle branch block or ventricular tachycardia
			Irregular	Normal	Any speed	Atrial fibrillation
				Abnormal	Any speed	Atrial fibrillation and bundle branch block
		QRS complexes absent				Ventricular fibrillation

Table 1.3.1 Recognizing ECG abnormalities (Source: Hampton, 2013)

The resting heart beats usually 60-100 times per minute, and this normal cardiac rhythm is called *normal sinus rhythm* (NSR), while anything else is called an *arrhythmia* (Thaler, 2015). The term arrhythmia includes any abnormality in the conduction, rate or regularity of the cardiac electrical impulses (Thaler, 2015). Referring to Thaler (2015), an arrhythmia can be a single abnormal beat, an extended pause between beats, or a rhythm disturbance that can continue during the whole life of the patient. However, it is important to note that not every arrhythmia is dangerous to human's health. For instance, a heart rate between 35 to 40 beats per minute is considered normal in well-trained athletes. Therefore, a precise analysis of the ECG signal is crucial in preventing wrong diagnosis.

Hampton (2013), in his book *"The ECG Made Easy",* categorizes abnormal rhythms as:

- bradycardias, which have a slow and sustained rhythm;
- tachycardias, which also have a sustained rhythm, but faster in speed;
- fibrillation, where the activation of the atria or ventricles is completely out of order.

↓ Origin, Type →	Tachycardia	Bradycardia	Error of conduction
Supraventricular arrhythmias (Narrow QRS complexes)	Sinus tachycardia	Sinus bradycardia	AV block I
	Premature atrial contraction / premature impulse from the AV node	SA block / sinus block	AV block II
	Junctional tachycardia		AV block III
	Supraventricular tachycardia		
	Atrial flutter		
	Atrial fibrillation		
Ventricular arrhythmias (Broad QRS complexes)	Premature ventricular contraction	Ventricular escape	Right bundle branch block
	Ventricular tachycardia		Left bundle branch block
	Ventricular fibrillation		

Table 1.3.2 Arrhythmias overview

In reference to Hampton (2013) and bearing in mind the abnormal heartbeats that appear in our dataset, arrhythmias are categorized into two major groups based on their origin and type (see Table 1.3.2): supraventricular arrhythmias and ventricular arrhythmias.

1.3.1. Narrow QRS Complexes: Supraventricular Arrhythmias

Supraventricular Arrhythmias are the arrhythmias that originate in the atria or the atrioventricular node (AV) (Thaler, 2015). *Atrioventricular node* is an electrical gate at the meeting point of the atria and the ventricles (see Figure 1.2.1).

Supraventricular rhythms are characterised by narrow QRS complexes, less than 120 ms (Hampton, 2013). Narrow complex tachycardia may be a sign of *sinus tachycardia, premature atrial contraction* or *premature impulse from the AV node, junctional tachycardia, atrial flutter* and *atrial fibrillation* (see Table 1.3.1). *Sinus rhythm* is identified by one P wave per QRS complex, by a normal QRS shape and a normal QRS rate (see Table 1.3.1). *Premature atrial contraction* appears as a P wave and QRS complex happening earlier than expected. Moreover, it is characterized by a R-R interval that is prolonged after a premature beat.

Sinus tachycardia seems to have a normal electrocardiogram. However, while a normal heart rate is 72 beats per minute, for sinus tachycardia the heart rate is expected to be 150 beats per minute (Hall, 2011). In *junctional tachycardia*, we observe that the P waves are not present. Additionally, a normal QRS shape with a rate of 150-180 beats per minute is noticed (Hampton, 2013). In contrast, *atrial flutter* is characterized by a P wave at a rate of 300 beats per minute. Furthermore, the atrial depolarization happens very fast such that we are not able to see the P waves separated by a flat baseline (Thaler, 2015). This results in a great number of atrial impulses that want to pass through the AV node to generate QRS complexes.

Nevertheless, not all atrial impulses succeed to pass through the AV node. The most that they can do is hit the refractory node, and this is called the *AV block* (Thaler, 2015). The first-degree AV block causes a prolonged P-Q interval and all P waves are followed by QRS complexes. As mentioned earlier, a normal P-Q interval lasts between 0.12 and 0.20 seconds, while the PR interval in a patient with first degree AV block is prolonged more than 0.20 seconds (Hall, 2011). Therefore, the *first degree AV block* is more of a postponement of conduction from the atria to the ventricles, rather than an obstruction of conduction. In addition, in the atrioventricular node, we notice an atrial P wave, but no QRS-T wave (Hall, 2011). As shown in Table 1.3.1, the *second degree AV block* is characterized by a regular QRS complex with a normal shape and slow rate. The *third degree AV block*, also called a complete AV block, is characterised by a P wave that is completely disassociated from the QRS-T complexes.

Atrial fibrillation, known as the most irregular rhythm and a completely chaotic atrial activity, has a QRS complex rate typically varying between 120 beats per minute and 190 beats per minute (Hampton, 2013). However, the QRS complexes have a normal shape and a normal T wave (see Table 1.3.1). Same as with junctional tachycardia, atrial fibrillation has no P wave. On the other hand, in the groups of slow rhythms, known under the name of bradycardia, is included sinus bradycardia and SA block (see Table 1.3.2). *Sinus bradycardia* happens when the rhythm of the heart slows down below 60 beats per minute. It is the most common rhythm disturbance observed in the early stages of an acute myocardial infarction (Thaler, 2015).

1.3.2. Broad QRS Complexes: Ventricular Arrhythmias

Ventricular rhythms are identified by wide QRS complexes with duration more than 120 ms (Hampton, 2013). Very wide QRS complexes, which are greater than 160 ms, are usually a sign of ventricular tachycardia. Furthermore, such rhythms have no visible P waves. If the ventricular muscle depolarizes with a high frequency, the rhythm is called ventricular tachycardia. The QRS complexes are seen to be wide, duration 280 ms and with a very abnormal shape, and T waves are difficult to be identified (Hampton, 2013). Moreover, the P waves are absent.

The *premature ventricular contractions* (PVC) are the most widespread ventricular arrhythmias. The QRS complexes in PVC are usually commonly prolonged and have a high voltage (Hall, 2011). In addition, the QRS duration must be at least 0.12 seconds in most of the leads of an ECG signal in order to make the diagnosis of PVC (Thaler, 2015). Typically, PVCs do not have the same amplitude of the R peak, and the T wave that precedes the QRS is placed higher. Moreover, there is no P wave after the QRS complex (Carvalho, et al., 2011). The main characteristic of PVC is its premature occurrence, which is clearly displayed by the R-R interval in Figure 1.3.1.

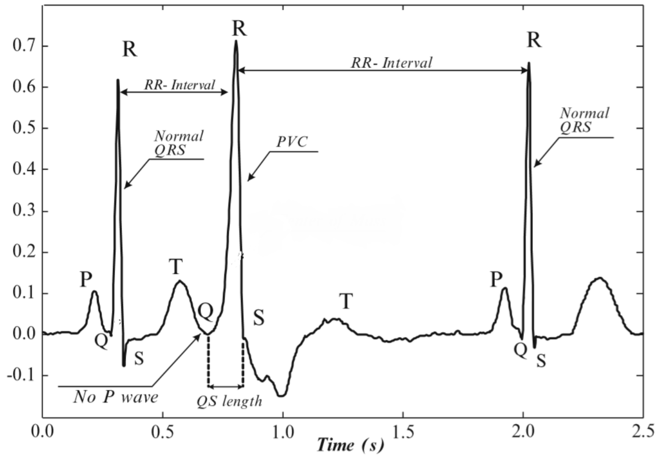


Figure 1.3.1 Premature ventricular contraction (Source: Carvalho, et al., 2011)

Ventricular fibrillation (FVB) is known to be the arrhythmia that causes the greatest number of sudden deaths in adults (Thaler, 2015). The main identifier of FVB is the absence of QRS complexes. *Escape beats* are called the rescuing beats that come into action when sinus arrest occurs. Sinus arrest, shown in the ECG as a flat line without any electrical activity, happens when the sinus node stops firing (Thaler, 2015). The escape beats included in our dataset are: *atrial escape beat, junctional escape beat, supraventricular escape beat* and *ventricular escape beat.*

Ventricular escape is most often observed in the cases when then complete heart block interrupts the conduction between the atria and ventricles (Hampton, 2013). On the ECG signal, after a pause, a single abnormal wide QRS complex is observed. One could diagnose a bundle branch block by observing the width and configuration of the QRS complexes.

The *right bundle branch block* (RBBB) is caused by a delay in the right ventricular depolarization which is represented in the ECG signal by a wider QRS complex beyond 0.12 seconds. In RBBB, with the left ventricle depolarizing we see the initial R and S waves, while with the late depolarization of the right ventricle, we notice a second R wave, called R′ (Thaler, 2015). Some other commonly used ECG criterias to diagnose RBBB include: a QRS complex in the shape of letter “M” and an inverted T wave appears in leads V1 and V2, a broad S wave in leads V5 and V6, either S wave lasts longer than R wave or S wave duration is beyond 40 msec in V6 (Thaler, 2015).

On the other hand, in the *left bundle branch block* (LBBB), it is the left ventricular decomposition that is delayed. Similarly to the RBBB, the QRS complex lasts longer than 0.12 seconds. In addition, RBBB is characterised by a broad R wave that lasts longer than 30 msec (Wyngaarden, et al., 2004). Some other commonly used ECG criterias to diagnose LBBB are: deep and broad S wave in leads V1 and V2, while leads V5 and V6 are characterized by broad and clumsy R wave, and, moreover, the ST segment goes higher than 5mm. You may refer to Figure 1.3.2 for a visual representation of such characteristics.

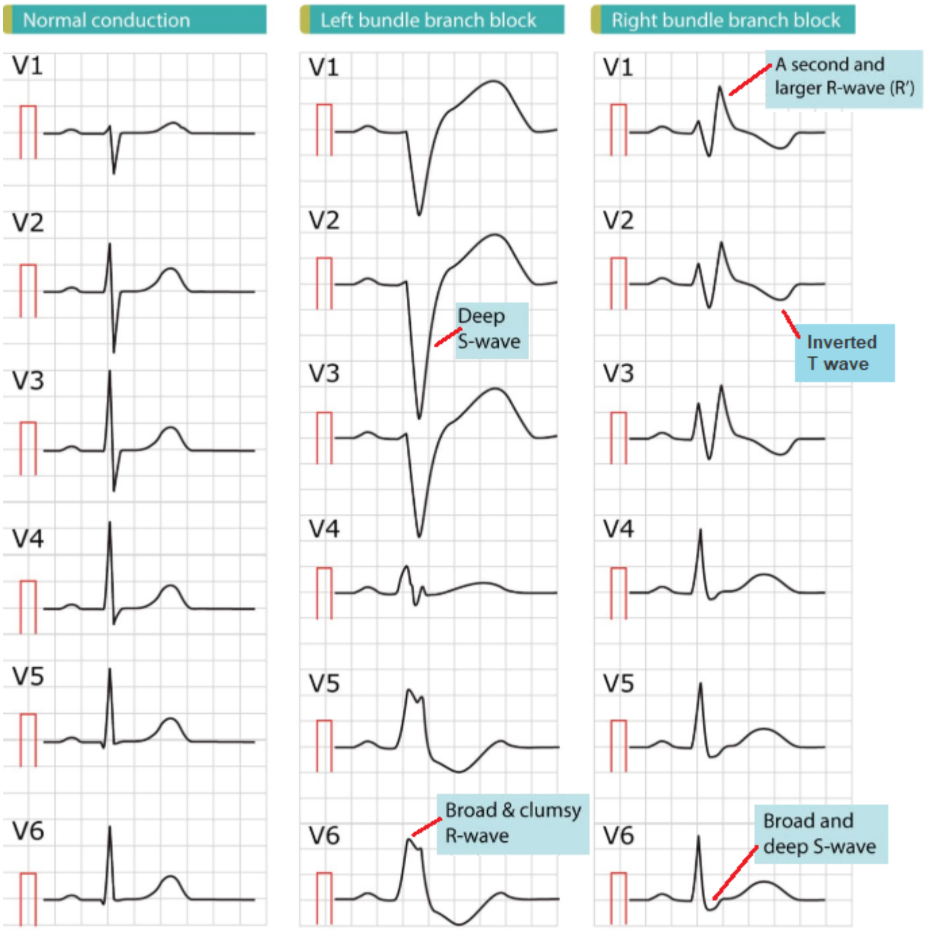


Figure 1.3.2 ECG signals showing the difference between Normal, LBBB and RBBB beats (Source: ECG and ECHO Learning)

An overview of the main characteristics of the beats that are being considered in this thesis are displayed in Table 1.3.3.

APC	PVC	RBBB	LBBB
- Early and abnormal P wave	- No P wave after QRS complex	- Long S wave (>= 40 ms) in leads V5-V6	- R wave >= 30 ms, deep and broad in leads V5-V6
- Early and narrow QRS complex	- Wide (>= 0.12 s) and prolonged QRS complex	- Wide (>= 0.12 s) QRS complex with "M" shape	- Wide (>= 0.12 s) QRS complex
- Prolonged RR interval	- Higher T wave after QRS complex	- Inverted T wave in leads V1-V2	- ST segment higher than 5mm
			- Deep and broad S wave in leads V1-V2

Table 1.3.3 Recognizing main characteristics of APC, PVC, RBBB and LBBB beats

2. Materials and Methodology

2.1. Dataset Used

In this thesis, the MIT-BIH arrhythmia database, which is free and publicly available on PhysioNet is used (Goldberger, et al., 2003). The database contains 48 records, each with two-channel ECG signals (Moody & Mark, 2001). As described in Moody & Mark (2001), each record has a duration of 30 minutes selected from 24 hours of recordings after studying 47 different patients at the Boston's Beth Israel Hospital (BIH) Arrhythmia Laboratory between 1975 and 1979. The continuous ECG signals are band-pass filtered at 0.1–100Hz and then digitized at 360 Hz (Moody & Mark, 2001). Subjects include 22 females of age 23 to 89 and 25 males of age 32 to 89. Approximately 40% of these recordings were obtained from outpatients, and 60% of these recordings were obtained from inpatients. Two records, 201 and 202, are from the same male subject (Moody & Mark, 2001). The database contains annotations for both timing information and beat class information verified by experts independently.

Overall, there are 112,646 labelled beats, all annotated by at least two cardiologists (Moody & Mark, 2001). First, a simple QRS detector generated an initial set of labels for each beat, tagging each detected event as a normal beat. Then, for each record, the cardiologists added additional beat labels for those beats that the detector missed. Moreover, they deleted false detections where it was necessary, and checked the labels for all abnormal beats (Moody & Mark, 2001). The dataset includes 15 different heartbeat classes, where *Normal Beat* is the group with the highest amount of data, and *Supraventricular Premature Beat* with only two samples, being the class with the smallest amount of data. More specifically, some of the main beats, that our dataset includes, are: 75052 labeled normal beats, 8075 left bundle branch block (LBBB) beats, 7259 right bundle branch block (RBBB) beats, 7130 PVC beats, 7028 paced beats, 2546 atrial premature contraction (APC) beats, 229 nodal (junctional) escape beats, 106 ventricular escape beats, 83 nodal (junctional) premature beats, 16 atrial escape beats, 2 supraventricular premature beats and 33 are unclassified beats (Moody & Mark, 2001).

2.2. Data Pre-processing

Objective evaluation of the classifiers proposed in machine learning research is essential. Therefore, using all the data of the MIT-BIH arrhythmia dataset in the classification process seems the right path to follow. However, this results in being time consuming. Hence, the most recent scientific papers use only a subset of this database, extracting different heartbeat classes to use in their study. In Alickovic & Subasi (2016), five different heartbeat classes were selected with a total of 1800 heartbeats: 1000 normal beats, 300 left bundle branch block beats, 200 right bundle branch block beats, 100 atrial premature contraction heartbeats and 200 premature ventricular contraction heartbeats. Nayak, et al. (2016) and Desai, et al. (2016) both categorize the entire dataset into five arrhythmia classes based on the ANSI/AAMI EC57:1998 standard. In their experiments, they use 110,093 heartbeats, consisting of 82.3% non-ectopic (NE), 2.7% supraventricular ectopic (S), 7% ventricular ectopic (V), 1.6% fusion (F) and 6.4% unknown beats (U). In Acır (2006), 4 types of beats are used: normal beats, LBBB beats, premature ventricular contraction beats (PVC) and non-conducted P-wave.

In this paper, 3600 heart beats were randomly selected out of 112,646 labelled beats contained in the chosen database, from which: 2000 Normal heartbeats out of 75052, 600 LBBB heartbeats out of 8075, 400 RBBB heartbeats out of 7259, 400 PVC heartbeats out of 7130 and 200 APC heartbeats out of 2546. In each heartbeat class, half of the beats were used for training of the classifiers and half of the beats for testing and evaluation of our models. In particular, 1000 Normal beats, 300 LBBB beats, 200 RBBB beats, 200 PVC beats and 100 APC beats were used to form the training set, and the same amount of beats were used to form the test set. Every ECG heartbeat is a matrix using one lead and having a window of length 320 data points. 320 data points corresponds to approximately 0.889 seconds.

2.3. Discrete Wavelet Transform (DWT)

Wavelet transform is the decomposition of a signal into components called wavelets. The key idea of the wavelet transform is the multiresolution decomposition of signals and images. The most commonly used method in multiresolution decomposition includes creating an approximation component using a scaling function (a low-pass filter) and the detail components using wavelet functions (high-pass filters) (Sundararajan, 2015). Wavelets are functions that separate data into various frequency components and are the foundation for representing images in various degrees of resolution (Gonzales & Woods, 2008). There are different versions of wavelet analysis; one of them is the Discrete Wavelet Transform. DWT decomposes the signal into wavelets at distinct frequency bands having distinct resolutions.

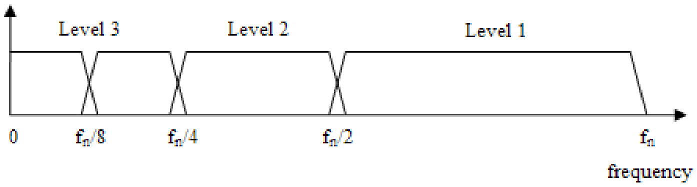


Figure 2.3.1 Frequency domain representation of the DWT

It decomposes the signal into approximation A using low-pass filter, which attenuates the high frequency part of the initial signal and passes only the low frequencies, corresponding to the smooth parts of the image, and into detail D using high-pass filter, which does the opposite of the low-pass filter, passing only the high frequency part of the initial signal and corresponding to the detailed parts of the image (Gonzales & Woods, 2008). The first down sampled high-pass filters' and low-pass filters' produce detail D1 and approximation A1. In the next level of decomposition, in the same way, A1 is decomposed into detail D2 and approximation A2. Repeatedly, this procedure is performed in this way until no more sub-sampling is possible. The two level decomposition of a signal s by the DWT is shown in Figure 2.3.2. Even though, the DWT are not capable of distinguishing the noise coefficients from the signal coefficients at low SNRs, it is still a good choice as it is capable of saving the significant phase information of ECG heartbeat signals (Alickovic & Abdulhamit, 2016). DWT involves forward and inverse transforms (see Figure 2.3.2 and Figure 2.3.3). Discrete Wavelet Transform is being used by all of the most recent papers in the process of feature extraction from ECG signals.

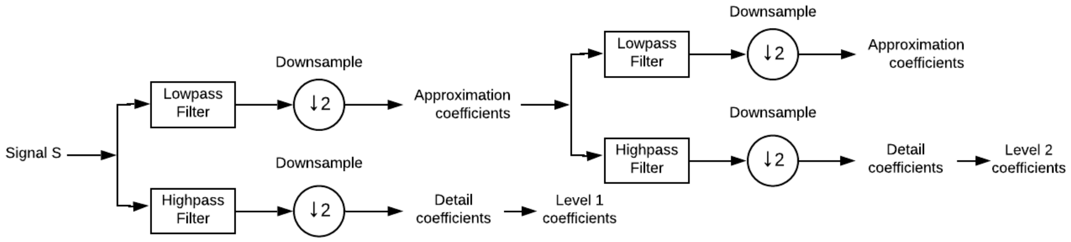


Figure 2.3.2 A two level decomposition of a signal s by the forward DWT

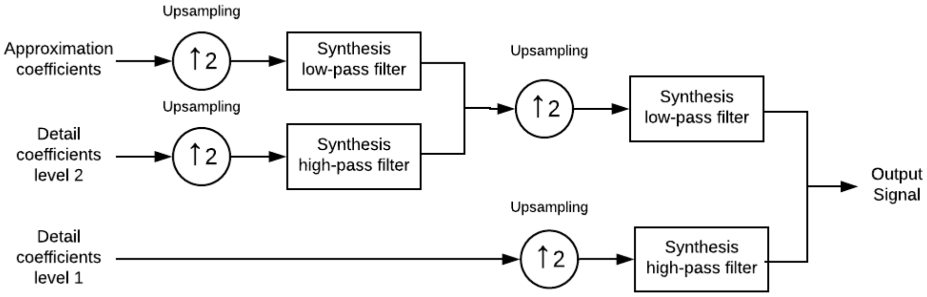

Upsampling
Approximation coefficients
↑2
Synthesis low-pass filter
Upsampling
Detail coefficients level 2
↑2
Synthesis high-pass filter
Upsampling
↑2
Synthesis low-pass filter
Output Signal
Upsampling
Detail coefficients level 1
↑2
Synthesis high-pass filter

Figure 2.3.3 A two level inverse DWT

The first level of DWT is the Haar wavelet invented by the Hungarian mathematician Alfred Haar. The Haar Wavelet Transform is well-known for being simple and fast in computation (Stanković & Falkowski 2003). There are two types of coefficients produced by Haar Wavelet Transform, the approximation coefficients, which are calculated by averaging the two adjacent samples, and the detail coefficients, acquired by subtracting two adjacent samples (Stanković & Falkowski 2003). Moreover, the Haar Wavelet Transform involves forward and inverse transforms. The forward transform requires two main steps: computation of the scaling coefficients, achieved by adding two adjacent sample values and dividing by 2, and computation of the wavelet coefficients, achieved by subtracting two adjacent sample values and dividing by 2.

On the other hand, the computation of the inverse transform requires simply addition and subtraction. In other words, for an input signal of length 2n, where n is the number of levels, the Haar wavelet transform joins together the input values. It saves their difference and passes their sum. Daubechies wavelet family, invented by the Belgian mathematician Ingrid Daubechies, is considered to be the most practical wavelet family due to its orthogonal abilities. An N-th order Daubechies wavelet is denoted as dbN, where Daubechies db1 is the same as Haar Transform (Malik & Verma, 2012).

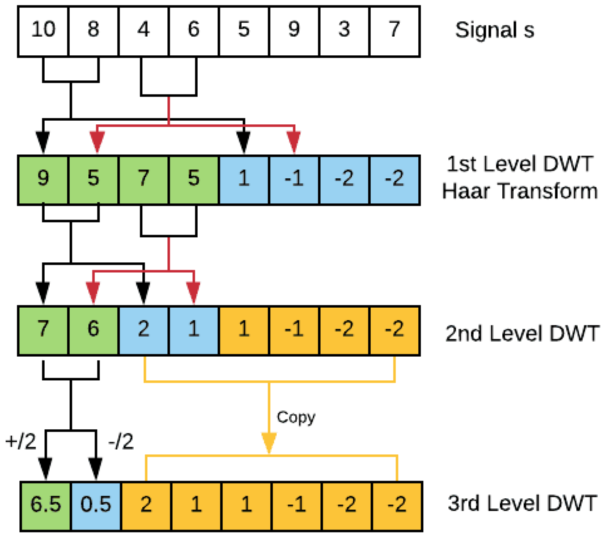


Figure 2.3.4 Example of a 3-level forward DWT signal decomposition

The Figure 2.3.4 displays a visual representation of how the DWT algorithm works. In order to make the explanation as simple to understand, we decided to take a simple example of a signal and perform 3rd level DWT. The green squares represent the approximation coefficients and the blue squares represent the detail coefficients. As mentioned above, the approximation coefficients are obtained by adding two adjacent sample values and dividing by 2 (ie. (10+8)/2 = 9), while the detail coefficients are obtained by subtracting the two adjacent sample values and dividing by 2 (ie. (10-8)/2 = 1). We proceed in this way, taking two-by-two samples and performing the simple calculations above. As a result of this transformation, the first level of discrete wavelet transform is obtained. In addition, to acquire the second level of DWT, the same calculations are performed, but only on the first half of the output transformed signal. The second half is a copy of the values of the signal in the 1st level of DWT, represented by the yellow colored boxes. We proceed in this way recursively, depending on how many levels of DWT decomposition we wish to perform.

In many recent research papers, it is stated that the 6th level detail coefficients are considered to be the most significant ones. Therefore, we will be performing up to 6th level DWT decomposition of the ECG signal. Moreover, Figure 2.3.5 shows the plot of the original signal of record 100 from MIT-BIH arrhythmia database and the output of DWT for level 1, 2, 4, 5 and 6.

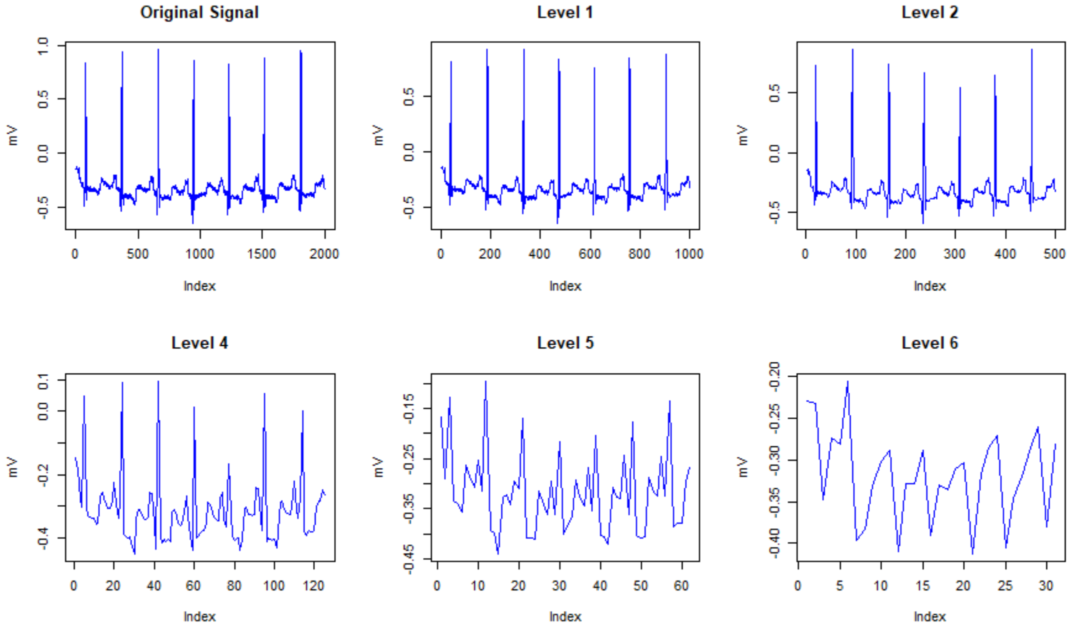


Figure 2.3.5 Plotting original signal of record 100 and the output of DWT

2.4. Feature Extraction

In both papers of Nayak, et al (2016) and Desai, et al (2015), the DWT multiresolution analysis with Daubechiesdb4 mother wavelets is being used in the ECG signal denoising process. They reconstruct the wavelet coefficients in the detailed sub-bands of 3th, 4th, 5th, 6th, 7th, 8th and 9th level to obtain a denoised ECG signal. The first two levels of detailed coefficients are set to zero, because the ECG above 45Hz does not contain any important information. Moreover, each cardiac beat consists of 200 samples, and the DWT is used to decompose them into four sub-bands. In addition, the feature extraction is performed at the QRS-complex frequency range from 3rd level detail and 4th level detail coefficients.

Acir (2006) analyses three different techniques in feature extraction. The first method uses the amplitude of the raw samples as input vectors in the recognition procedure, the second calculates the Discrete Cosine Transform coefficients in the original ECG data, and the third involves the DWT of ECG data using Daubechies-2 wavelet. For each feature vector, only the 2nd, 3rd and 4th levels of wavelet approximation coefficients are calculated, resulting in the 4th level being the most significant one.

In the study of Alickovic and Abdulhamit (2016), the DWT analysis is also used in the process of feature extraction. To reconstruct the original signal, they rebuild the approximations and details results using Daubechies-4 wavelet filters.

In order to accurately classify the heartbeat from an ECG signal, dimensionality reduction of the feature extraction is needed. The wavelet coefficients of the DWT provide us with a good picture of the distribution in time and frequency domain of the ECG signal. However, to reduce its dimensionality, in this thesis, statistical indices are used to indicate the distribution of the ECG signals over time and frequency as follows (Alickovic & Abdulhamit, 2016):

1. The average of the absolute values of the wavelet coefficients in each sub-band.
2. The mean of the values of the coefficients in each sub-band.
3. The standard deviation of the wavelet coefficient in each sub-band.
4. The ratio of the mean of the absolute values of each two adjacent sub-bands.

In the implementation of this thesis, we are considering up to the 6th level of detail of the DWT, and the inputs of the classifiers were the statistics calculated on the frequency bands A6 and D1-D6. In total there are 27 feature vectors.

2.5. Machine Learning Classifiers

2.5.1. Decision Tree Classifiers

Classification and Regression Tree and C5.0

CART analysis is a well-known decision tree technique for constructing predictors from the data. C5.0 algorithm is also known as divide and conquer, because it uses the features to divide the data into smaller and smaller subsets of similar classes (Lantz, 2013). The CART implementation is very similar to C5.0. The predictors are obtained by recursively partitioning the data and fitting a single predictor model within each partition (Loh, 2011). The root node of the tree represents the entire dataset. The algorithm starts at the root node, and, at each step, it picks one feature which helps us better in predicting the target class. It then uses this feature to make a split, forming in this way the first set of tree branches. The algorithms recursively proceeds in this way until a stopping criterion is reached. This happens when (Lantz, 2013):

- All of the examples at the node belong to the same class
- There are no features left to identify between examples
- The tree has reached the predefined size limit

Decision Tree classifiers are based on identifying the best split from which the greatest information is obtained, and then using that feature to split the data. In order to identify the best split, it is needed to measure the purity of the target variable within the subsets, which can be accomplished by using *entropy* (Lantz, 2013):

$$E(S) = \sum_{i=1}^{c} - p_i \log_2 p_i$$

,

where S represents the segment of data, c the number of different classes and p the probability of the values being part of class i. Now, using entropy, the Information Gain of a feature F can be calculated as shown below (Lantz, 2013):

InfoGain(F) = Entropy(S1) - Entropy(S2).

The higher the information gain, the more homogenous the group that is created after the split on this feature is.

Decision Tree classifier is a binary recursive partitioning technique, because each node in the decision tree stands for a group of examples. This node can be split only into two child nodes, making the original node the parent. This process of binary partitioning happens recursively, over and over again. In this way, each parent node is divided into two children nodes, and each children node may be partitioned into additional children nodes (Lewis, 2000). However, there is the risk of model overfitting as the tree grows bigger and bigger making overly specific decisions. Therefore, there is the need to prune the tree.

Pruning is the process of reducing the size of a decision tree so that it would generalize better to new unseen data (Lantz, 2013). There are two main types of pruning: pre-pruning and post-pruning. In pre-pruning, we allow the tree to grow until it reaches some predefined number of decisions or if there is only a small number of examples in the node. In post-pruning, first we grow a very large tree and then reduce the size using pruning criteria (Lantz, 2013). Some main advantages of decision tree analysis include (Lewis, 2000 and Lantz, 2013):

- Work with highly skewed numerical data and ordinal and non-ordinal categorical data.
- Work well with missing variables.
- Compared to the complexity of the algorithm, it requires relatively little input from the user.
- Require little knowledge of statistics to be interpreted.

Random Forest

A Random Forest (RF) is an ensemble method based on decision trees and bagging. This method was introduced first by Leo Breiman (2001), and then developed by Leo Brieman and Adele Cutler. RF combines the principles of bagging and random feature selection (see Figure 2.5.1). First, it generates an ensemble of trees, and, then, using voting, it allows these trees to vote for the most popular class (Lantz, 2013).

In order to grow these ensembles, for the k-th tree, this method generates a random vector Θ_k with the same distribution for all the trees in the forest, but independent of all the previous random vectors Θ_i for i = 1,...,k-1, and grows a tree using the training set and the generated random vector Θ_k. This results in a classifier h(x, Θ_k), where x is the input vector (Breiman, 2001). All trees are allowed to grow to the largest extent possible, without pruning, to result in low bias trees (Alickovic & Subasi, 2016). After a large number of trees is obtained, using voting, these trees select the most popular class (Breiman, 2001). Random Forest classifiers are mostly known for the following advantages (Lantz, 2013):

- Performs well on most problems.
- Can work with noisy and missing data.
- Can handle both categorical and continuous features.

- There is little risk of overfitting.
- Selects a subset of features. Therefore, it can be used on data with a very large number of features.

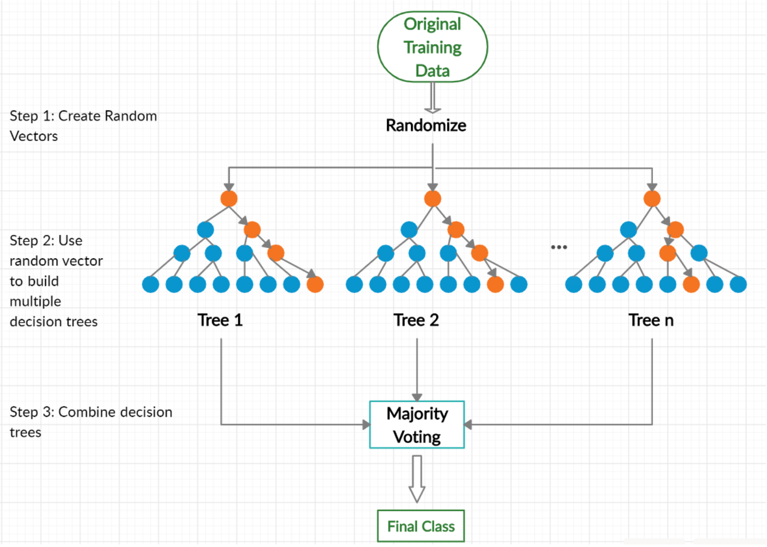

Original Training Data
Step 1: Create Random Vectors
Randomize
Step 2: Use random vector to build multiple decision trees
...
Tree 1
Tree 2
Tree n
Step 3: Combine decision trees
Majority Voting
Final Class

Figure 2.5.1 Random Forest algorithm

2.5.2. Support Vector Machine Classifier

Support Vector Machine is widely used for solving classification problems (Goodfellow et al., 2016). One can imagine SVM as determining a boundary between different points of data representing examples plotted in a multidimensional space according to their feature values (Lantz, 2013). Its main objective is to create a hyperplane $\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x} + b = 0$ that creates reasonably homogenous partitions of the examples on each side of this boundary. SVM is able to model very complex relationships by combining some features from both the instance-based nearest neighbor learning and linear regression (Lantz, 2013). Like linear regression, SVM is driven by a linear function. However, it does not provide probabilities, but instead it identifies an output class. An example will be classified in the positive class if $\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x} + b \geq 0$, and will be classified in the negative class if $\boldsymbol{\omega}^{T}\boldsymbol{x} + b < 0$ (Goodfellow et al., 2016).

$$f(\boldsymbol{x}) = \begin{cases} 1 & if \quad \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x} + b \geq 0 \\ -1 & if \quad \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x} + b < 0 \end{cases}$$

At the heart of SVM classifiers is a quadratic optimization problem that tries to maximize the margin between the decision boundary and the training data in the feature space. The subset of examples that lie close to the decision boundary in the plane are called support vectors. In Figure 2.5.2, we can see a visual representation of the architecture of SVM, M is the number of support vectors and p is the input dimension. However, data in real life is not always uniformly separable. Therefore, it is necessary to apply different kernel transformations for nonlinear mapping to a higher dimensional feature space (labeled by K(.) in the Figure 2.5.2), which are a key innovation associated with SVMs (Acir, 2006).

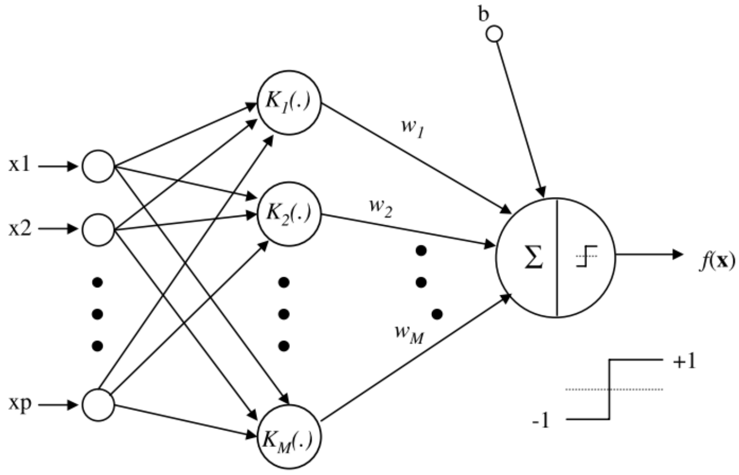


Figure 2.5.2 Visual representation of the architecture of SVM (Source: Acir, 2006)

The kernel exploited the fact that many machine learning algorithms can be written as dot product between examples. Therefore, it can be written as the following (Goodfellow et al., 2016):

$$\boldsymbol{\omega}^T \boldsymbol{x} + b = b + \sum_{i=1}^{m} y_i \alpha_i \boldsymbol{x}^T \boldsymbol{x}^{(i)} ,$$

where is a training example and is a vector of coefficients. By replacing x with the output of a given feature function and the dot product with a kernel function, we can make predictions using the function (Goodfellow et al., 2016):

$$f(\boldsymbol{x}) = b + \sum_{i} y_i \alpha_i k(\boldsymbol{x}, \boldsymbol{x}^{(i)}) .$$

In this way the input data are mapped to a higher dimensional space in which the separating plane is constructed to maximize the margin. If we go back to the lower dimensional data space, this hyperplane becomes a nonlinear separating function (see Figure 2.5.2). The prediction kernel-based function shown above, roughly corresponds to pre-processing the data by applying to all input values and then learning a linear model in the new transformed space (Goodfellow et al., 2016).

2.6. Ten-fold Cross-validation

The ten-fold cross-validation randomly separates the data into ten equally sized subsets called folds. Evidence suggests that taking a greater number than ten for the folds does not add many benefits (Lantz, 2013). One fold is used as the model evaluation and the remaining nine folds to train the classifier. A detailed description of k-fold cross validation and other cross-validation methods can be found in Arlot and Celisse (2010). In this thesis, the 10-fold cross validation is used in model parameter tuning. A detailed description of this process may be found in Chapter 3.

2.7. Evaluation Metrics

Different methods are used in the evaluation of the classifiers. I will be considering three different statistical approaches: ROC curve, F-measure and the overall accuracy. The F-measure and accuracy is calculated using a multi-class confusion matrix. The confusion matrix is a table that describes the performance of our classification model, by categorizing predictions on whether the predicted value is the same as the true value (Lantz, 2013).

The first statistical index used in this paper to measure the performance of the classifiers is ROC curve. The ROC curve (receiver operating characteristic curve) shows the performance of the classifier at all classification thresholds. This graph is obtained by plotting true positives (sensitivity) in the x-axis and false positives (1 - specificity) in the y-axis. Sensitivity measures how often a test correctly gives a positive result for patients who suffer from a certain arrhythmia type for which they are being tested. On the other hand, specificity measures how often a test correctly gives a negative result for people who do not suffer from the arrhythmia type that they are being tested for. The question *"How specific is the test?"* tells us *"How* *many heartbeats that are not of type A were correctly confirmed as not being of type A?".* The question *"How sensitive is the test?"* tells us *"How many heartbeats that actually correspond to class A were correctly identified as so?".* True Positive Rate (TPR) is a synonym for recall and is defined as:

$$Sensitivity = TPR = \frac{TP}{TP + FN},$$

while False Positive Rate (FPR) is defined as follows:

$$1 - Specificity = FPR = 1 - \frac{TN}{FP + TN} = \frac{FP}{FP + TN}.$$

TP stands for true positives, TN stands for true negatives, FN stands for false negatives and FP stands for false positives. The True Positives show the number of predictions belonging to class A which were correctly classified in the class A. By using the confusion matrix, the True Negatives for a particular class A can be calculated by summing the values in every row and column excluding the row and column of the class A, and the False Positives of class A by summing all the values in the column of class A excluding the True Positive value. The same holds for False Negatives, but instead of taking the column of class A, we sum the rows. The points in the ROC curve display the true positive rate at different false positive thresholds (Lantz, 2013). The closer the curve is to the left upper corner of the graph, the better the model is at identifying positive values. We will be calculating the AUC value (area under the ROC curve) which calculates the entire two-dimensional area under the ROC curve. The AUC ranges from 0.5, for a random classifier with no predictive value, to 1.0 for a perfect classifier (Lantz, 2013).

The second statistical approach used in performance evaluation is F-measure. The F-measure integrates precision and recall using the harmonic mean and is defined as follows:

$$F - measure = \frac{2 \times Precision \times Recall}{Recall + Precision},$$

where

$$Precision = \frac{TP}{TP + FP}$$

and

$$Recall = \frac{TP}{TP + FN}.$$

Precision shows the proportion of positive predictions that were actual positives, while recall shows the proportion of actual positives that were predicted correct. In evaluation of our model we aim for both a high recall and a high precision, therefore, calculating F-measure will provide us with such data. Unlike the 2-class classification, in the multi-class classification the true positives and false negatives for each class are calculated in the confusion matrix.

Due to the fact that our data is imbalanced, the weighted F-measure, weighted precision and weighted recall for each classifier will be calculated (see Table 4.1.6). In order to find the weighted precision, we weigh the precision of each class by the number of samples from that class. The example below demonstrates how to find the weighted precision of rpart classifier using the data from Table 3.2.6:

$$Weighted\ Precision = \frac{0.79\times 1000 + 0.36\times 100 + 0.28\times 200 + 0.47\times 200 + 0.78\times 300}{1800} \approx 0.67$$

The most used statistical approach in model evaluation is the overall accuracy, which is described by the formula below:

$$Accuracy = \frac{Number\ of\ \ correct\ predictions}{Total\ number\ of\ \ predictions} = \frac{TP + TN}{TP + TN + FP + FN}.$$

3. Experiments

3.1. Overall Procedure of ECG Signal Classification

In this thesis, 3600 randomly selected heartbeats belonging to five arrhythmia classes, namely N (normal heart beat), LBBB (Left Bundle Branch Block), RBBB (Right Bundle Branch Block), PVC (Premature Ventricular Complex) and APC (Atrial Premature Contraction) from MIT-BIH arrhythmia database were considered. This dataset is split into two halves, where 1800 beats are used for training the classifiers and the other 1800 beats are being used in the evaluation process. The training dataset contains 1000 normal beats, 300 RBBB beats, 200 LBBB beats, 200 PVC beats and 100 APC beats, and the same holds for the test set. A rectangular window of 320 data points is used to extract heartbeats.

The considered ECG signals were subject to 6 level sub-band decomposition using DWT. This aims to extract feature vectors from each ECG signal segment, which are significant in model training and testing. In order to reduce dimensionality of the feature extraction composed of the set of the wavelet coefficients of DWT, statistical indices were used on the frequency bands A6 and D1-D6 to denote the time-frequency distribution. A detailed explanation of these statistical tools used, may be found in 2.4.

As a result, the system is trained using one channel and two matrices, each composed of 1800 rows and 27 feature vectors generated from DWT decomposition and the application of statistical indices, and evaluated on 28 columns including the annotation and 27 feature vectors. The classification is performed using four machine learning techniques: CART, C5.0, random forest and support vector machine using quadratic kernel as it resulted in a better overall performance than other kernel functions.

In addition, the 10-fold cross validation is used to tune model parameters in order to help us find the best combinations of algorithm parameters for our classification problem. In order to achieve such tuning, we use the CARET package of R, which stands for Classification and Regression Training. This package provides a great facility to tune ML algorithm parameters.

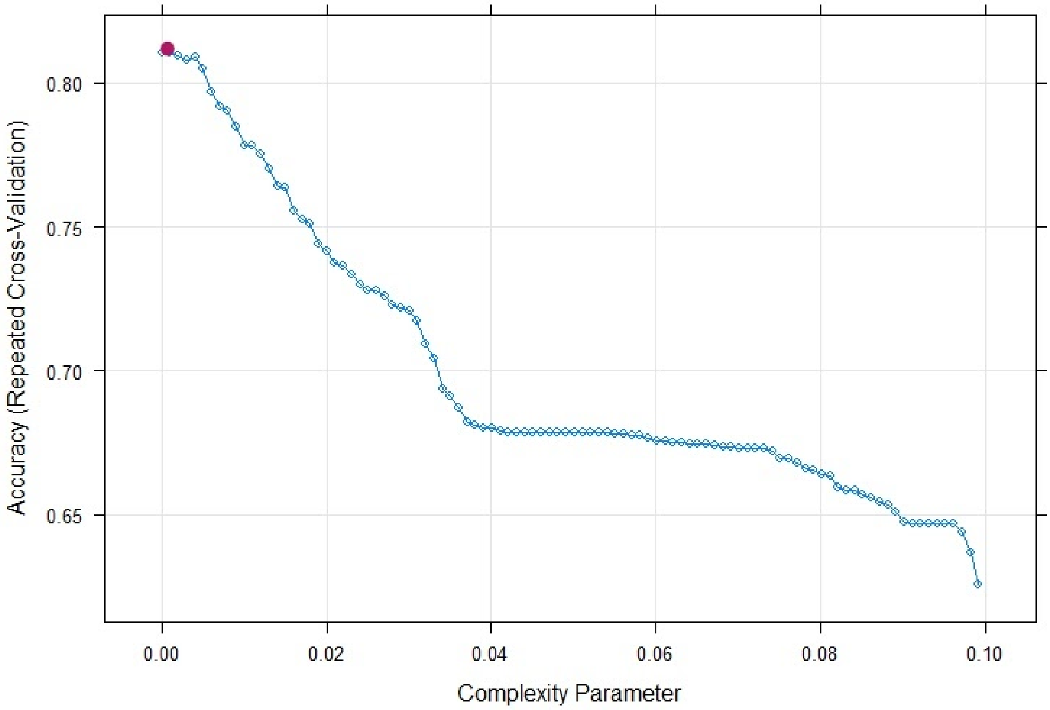


Figure 3.1.1 rpart - cross validation tuning parameters

For the CART classifier, 10-fold cross validation is repeated ten times by setting the *method* parameter of *trainControl* function to *"repeatedcv"* and the *repeats* parameter to 10, such that each subset is used for training at least once, then the average performance on all 10 folds is outputted (Alickovic & Abdulhamit, 2016). Our model is tuned by tuning the *complexity parameter* (cp) or *tree pruning*. The *tuneLength* parameter, which tells the algorithm to try different default values for the main parameter and allows the system to tune the algorithm automatically, is set to 100. Therefore, the total number of combinations that will be evaluated is 100. Figure 3.1.1 shows us how the change in cp value changes the model accuracy. As displayed in the graph with a colored node, for cp equal to 0.001, the highest accuracy of 81.05% and a kappa score of 0.696 is obtained. The standard deviation of cross validation for CART was estimated to be 0.03. As our CART classification model, we choose the final model used by the ten-fold cross validation. The final decision tree is displayed in Figure 3.1.2.

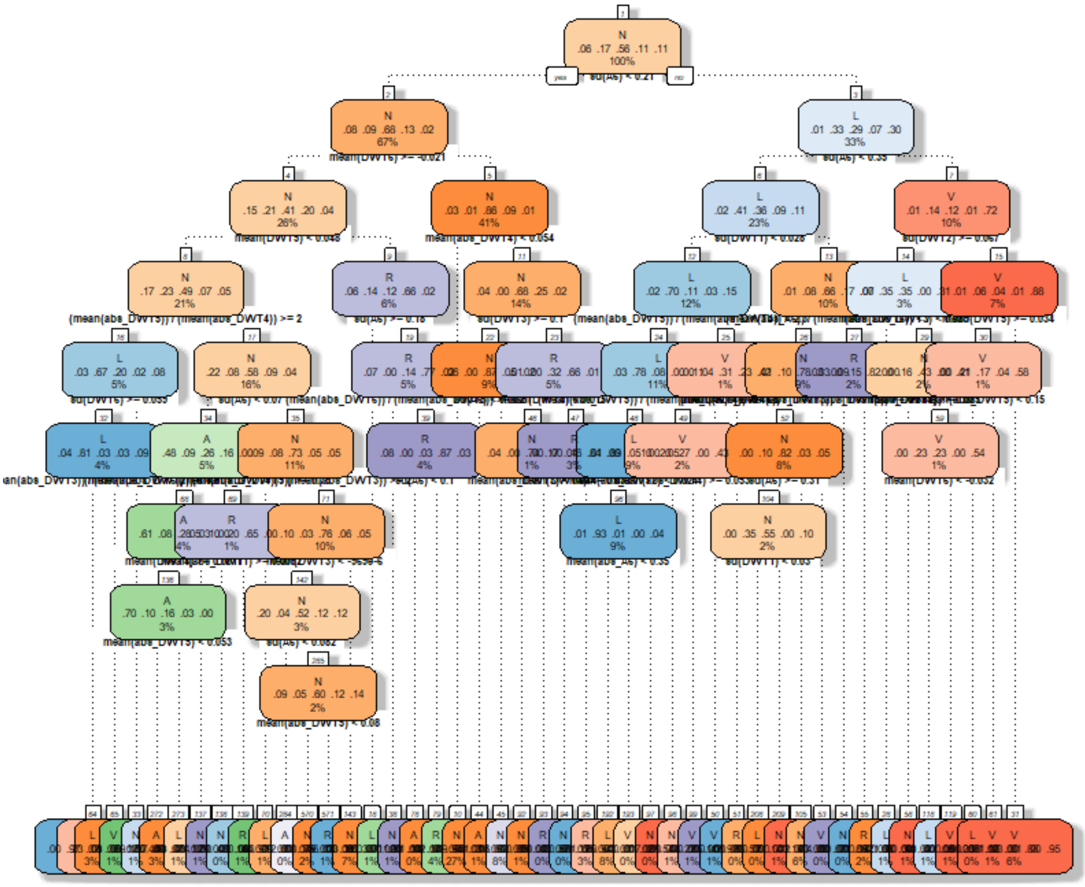


Figure 3.1.2 The final decision tree resulted from ten-fold cross validation

As stated in Scornet (2017), random forest classifier is controlled by four parameters: *mtry* representing the number of variables randomly sampled as candidates at each split, *nodesize* controlling the maximal number of observations in each cell, *maxnodes* representing the tree depth and *ntree* representing the number of trees to grow. Therefore, for random forest, 10-fold cross validation is used to search for the best number mtry, the best maxnodes of the maximum number of terminal nodes trees in the forest and the best ntrees. For RF, we decide to define a grid to tune the model by setting the *search* parameter of *TrainControl* function to "*grid*". Each axis of the grid represents a parameter and each point represents a specific combination of parameters. Accuracy was used to select the optimal model using the largest value. First, we search for the best mtry, by setting a ntree to 500 and nodesize to 10. Figure 3.1.3 shows how the accuracy changes with respect to the number of variables used in the model, mtry score. Clearly, it can be seen that the highest accuracy is obtained by mtry equal to 10 represented with a colored node, which provides us with an accuracy of 91.2% and kappa value of 0.86. Next, we look for the best maxnodes value. This is performed in several iterations, where the range is expanded by 10 in each iteration in order to achieve the highest accuracy. Finally, we search for the best ntrees score, by looking at the vector of values [250, 300, 350, 400, 450, 500, 550, 600, 800, 1000, 2000]. Training our model on ntree equal to 800, a maximal accuracy of 92.8% and a kappa score of 0.88 is obtained. As a result, the highest accuracy score is obtained with a value of maxnodes equals 56, mtry equals 10 and ntree equals 800. The random forest model is trained with such parameter values and the *randomForest* standard library of R is used.

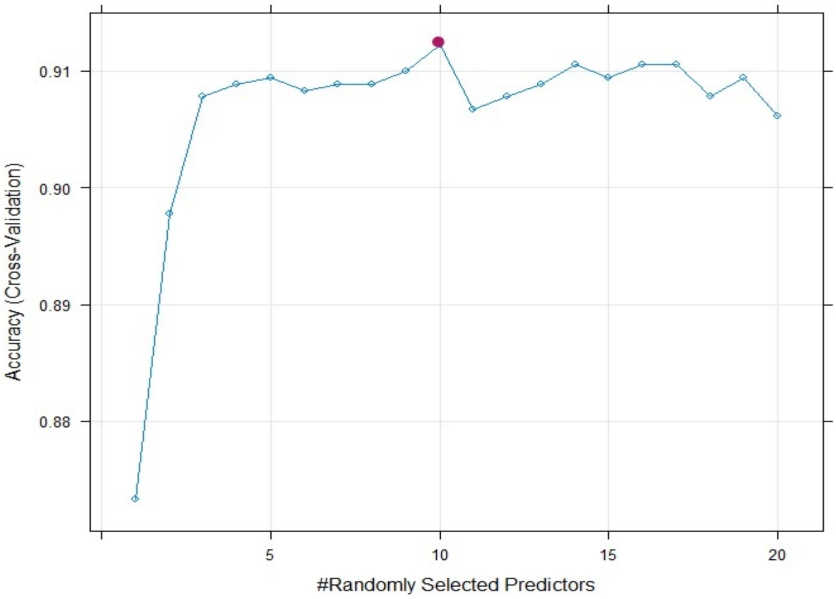


Figure 3.1.3 Plot of accuracy by number of variables used in model for Random Forest

In SVM, first, the models are trained using different kernel functions including linear, polynomial of degree 2 and 3, and radial. Then, the SVM model with radial kernel function is tuned to find the best cost and gamma values, by looking at the vector [0.001, 0.005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04] for the gamma value, and at the vector [0.5, 1, 5, 10] for the cost value. Finally, 10-fold cross validation is run using the best cost and gamma, 10 and 0.035 respectively. This results in the lowest error rate of 0.076.

The benefits of performing ten-fold cross-validation can be seen in all three models where such procedure is applied. In CART classifier, the accuracy of the model before ten-fold cross-validation was 64.5%, and the AUC value was 0.728, while, after using ten-fold cross-validation to select the best cp value, an accuracy of 67.1% and AUC value of 0.766 is achieved. In addition, the accuracy of the RF model before and after using ten-fold cross validation for parameter tuning is 71.3% and 89.4% respectively. Lastly, the accuracy of SVM using a radial basis kernel increases from 71.6% to 93.8%.

Each prediction model was tested and evaluated on the test set using the evaluation metrics described in 2.7. The experimental results of each classifier can be found in 3.2. When we finish with parameter tuning, training and testing the tree model, it comes naturally to all of us to ask which of the variables have the most predictive power in our model. Variables with high importance impact the most the result of our classifier. The more the model depends on a variable to make a prediction, the higher the importance of this variable is for the model. In the random forest, the importance of each variable is measured using two tools (James, et al., 2017). The first tool is the mean decrease of accuracy, which tells us how much the accuracy of a given variable is decreased when we exclude this variable from the prediction process. The second is the mean decrease of impurity of a given variable that results from splits over that variable. For classification problems, Gini index is used to calculate the node impurity. Therefore, the most important variables that have the highest predictive power on our decision tree models are being plotted. The results are shown in 3.2.

The proposed system has been implemented using R codes and standard libraries including: *rpart* to build the CART decision tree, *rpart.plot* to display a visual representation of the tree, *gmodels* to compute CrossTable, *C50* to build the C5.0 model, *e1071* to build the SVM classifier, *multiROC* to find the area under the curve value and *pROC* to plot the ROC graphs for each classifier. Figure 3.1.4 gives a visual representation of the proposed procedure for the ECG heartbeat classification.

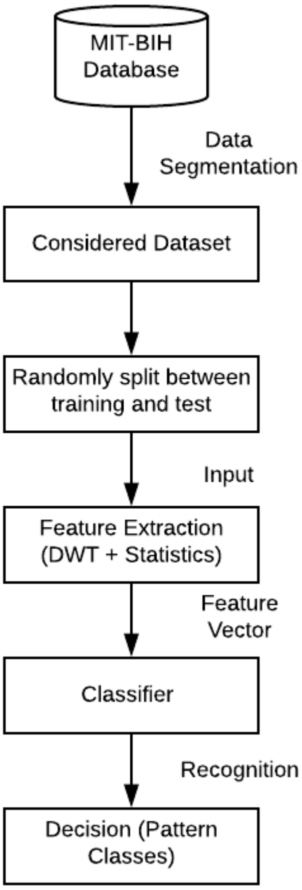


Figure 3.1.4 Procedure for ECG heartbeat classification used in this study

3.2. Experimental Results

In this experiment, DWT and statistical indices are used to extract the feature vectors from each ECG signal segment. Then, four different machine learning techniques were used in the classification. In this section, we will be reporting and analysing the results obtained from all our experiments.

Tables 3.2.1 - 3.2.4 provide the confusion matrix of five classes for four machine learning classifiers, where rows represent the classifier outputs and the columns represent the gold standard. The correct predictions of each classifier fall on the diagonal of the confusion matrix, highlighted in a grey background, while all the other values outside of the diagonal count the cases of incorrect prediction. Moreover, comparing the results from all four classifiers, the best correct predictions are displayed in bold.

Based on the results in the confusion matrices below, it can be seen that SVM classifier performs the best in classifying APC, LBBB, RBBB and PVC beats, while Random Forest performs slightly better in classifying Normal beats. RF manages to predict correctly 989 Normal beats out of 1000 Normal beats used for the testing in total, misclassifying 4 beats as LBBB and 7 beats as PVC. On the other hand, SVM wrongly classifies only 9 RBBB heartbeats, achieving in this way a high accuracy in RBBB classification, which can be seen in Table 3.2.5 and SVM ROC graph of Figure 3.2.1. In addition, SVM only correctly classifies 67 out of 100 APC heartbeats; however, this number is still higher compared to the other classifiers. CART classifier has the lowest number of correctly classified APC heartbeats, only 36, which is less than half of the total APC heartbeats in the training set. If C5.0 and RF are compared, it can be observed that the latter perform better in classifying Normal beats, LBBB beats and PVC beats. However, C5.0 achieves a higher number of correctly classified APC and RBBB heartbeats. These results can be observed also in the ROC curves of Figure 3.2.1.

Classes	APC	LBBB	Normal	RBBB	PVC
APC	36	3	20	12	3
LBBB	16	235	26	6	14
Normal	34	32	787	81	116
RBBB	5	20	57	94	12
PVC	9	10	110	7	55

Table 3.2.1 Confusion Matrix for classification using CART

Classes	APC	LBBB	Normal	RBBB	PVC
APC	61	3	15	2	4
LBBB	6	262	20	1	22
Normal	29	19	922	27	37
RBBB	4	2	16	165	2
PVC	0	14	27	5	135

Table 3.2.2 Confusion Matrix for classification using C5.0

Classes	APC	LBBB	Normal	RBBB	PVC
APC	60	0	0	6	0
LBBB	4	257	4	0	5
Normal	32	37	**989**	42	43
RBBB	4	0	0	152	0
PVC	0	6	7	0	152

Table 3.2.3 Confusion Matrix for classification using Random Forest

Classes	APC	LBBB	Normal	RBBB	PVC
APC	**67**	0	5	3	2
LBBB	1	**275**	7	0	5
Normal	24	19	974	6	12
RBBB	7	0	6	**191**	0
PVC	1	6	8	0	**181**

Table 3.2.4 Confusion Matrix for classification using SVM

The area under the ROC curve has been widely used in computer-based medical decision making to help define the accuracy of the model. One of the advantages of using AUC is its visual accessibility from a ROC plot, as shown in Figure 3.2.1. Each plot represents one machine learning technique used to classify the ECG signals, and each color represents one heartbeat class. As mentioned previously, the closer the curve is to the left upper corner of the graph, the better the classifier performs for this specific beat class.

As it can be seen in the plots of the Figure 3.2.1, SVM classifier performs the best in classifying correctly all the beats, except the APC beats who have a lower performance as shown represented by the blue line. On the other hand, if we look at the ROC of the Random Forest, we notice that it performs also very well in classifying the beats. CART classifier seems to be performing very poorly, especially in classifying PVC beats and APC beats. As it can be seen from the plot of the upper left corner of the Figure 3.2.1, the green curve representing the PVC beats is very close to the diagonal. The same can be observed for the blue line, representing APC beats. In addition, we observe that all C5.0, RF and SVM classifiers score the lowest performance in classifying correctly the APC beats compared to the classification of the other beats. One explanation for this could be the small number of examples of such heartbeat class in our database. Therefore, resulting in a small number of APC heartbeats in our training data of only 100 compared to 1000 examples from the Normal heartbeat class.

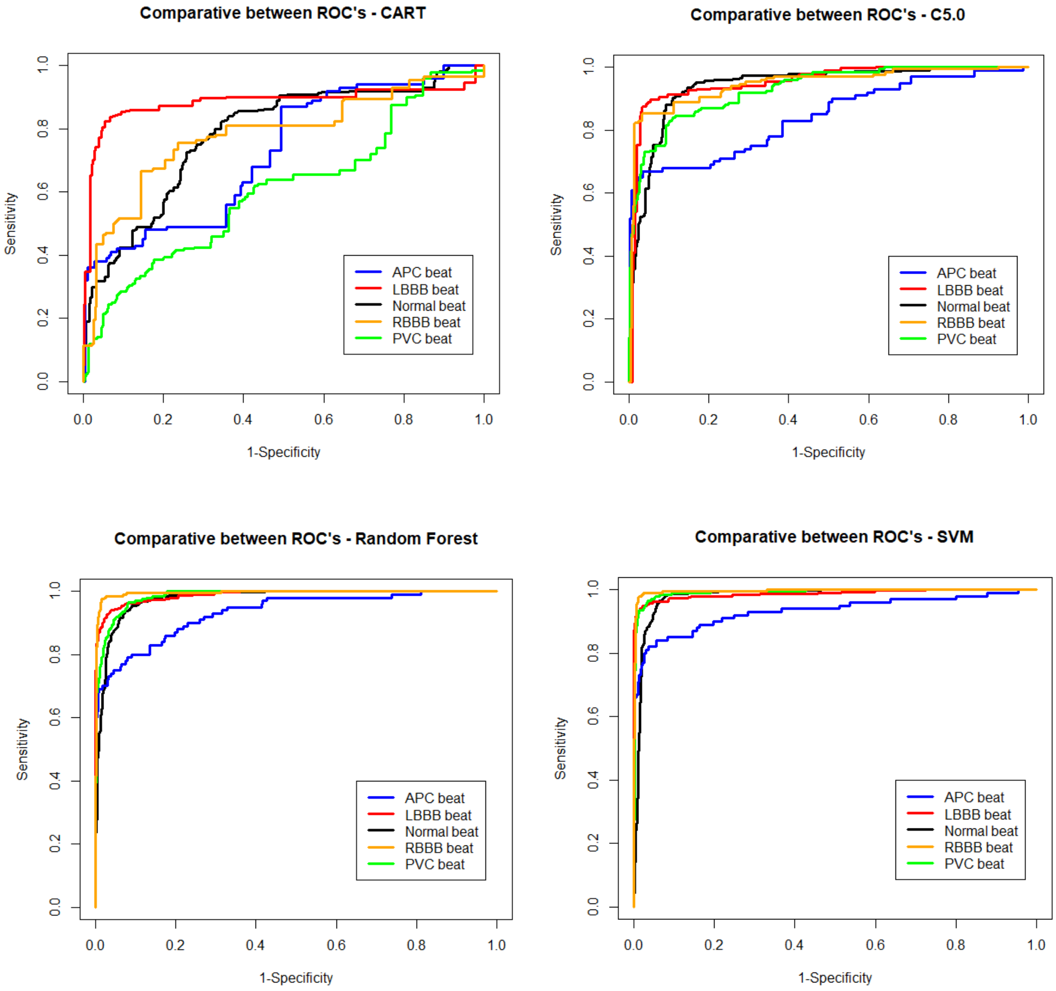
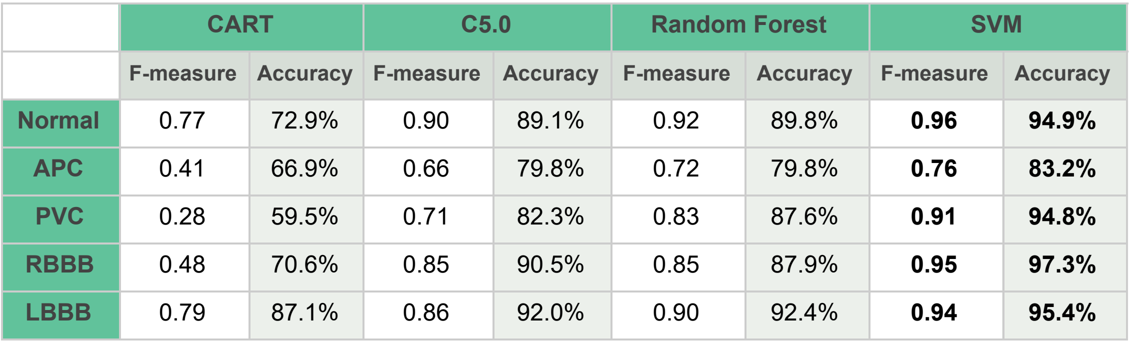


Figure 3.2.1 Receiver operating characteristics (ROC) for each classifier and each class

Table 3.2.5 displays the results of F-measure and accuracy obtained from all the classifiers for each heartbeat class, in which highlighted in bold are displayed the best results obtained by our classifier. The performance of SVM results in the best accuracy and highest F-measure for each heartbeat class. SVM classifies with the highest accuracy RBBB beats achieving 97.3%, and, as we observed in the ROC plots above, SVM has lower classification accuracy for APC beats of 83.2%. In addition, the highest F-measure is achieved by SVM for the recognition of Normal heartbeats. The same holds for RF classifiers in classification of APC beats. However, RF classifies LBBB beats with the highest accuracy. C5.0 and RF classify APC beats with the same accuracy, while the Random Forest model scores a slightly higher F-measure value. On the other hand, the lowest F-measure of only 0.28 is scored by CART for PVC beats, which also has the lowest accuracy of 59.5%.



	CART		C5.0		Random Forest		SVM	
	F-measure	Accuracy	F-measure	Accuracy	F-measure	Accuracy	F-measure	Accuracy
Normal	0.77	72.9%	0.90	89.1%	0.92	89.8%	**0.96**	**94.9%**
APC	0.41	66.9%	0.66	79.8%	0.72	79.8%	**0.76**	**83.2%**
PVC	0.28	59.5%	0.71	82.3%	0.83	87.6%	**0.91**	**94.8%**
RBBB	0.48	70.6%	0.85	90.5%	0.85	87.9%	**0.95**	**97.3%**
LBBB	0.79	87.1%	0.86	92.0%	0.90	92.4%	**0.94**	**95.4%**

Table 3.2.5 Comparison of the obtained results of accuracy and F-measure from all classifiers for each class

In addition, the performance of the models in terms of precision and recall for each heartbeat class is given in Table 3.2.6. Both RF and SVM seem to perform the best. SVM performs very well in classifying the PVC and LBBB beats, scoring both the highest precision and recall. However, RF also achieves the same recall as SVM for PVC and LBBB beats. Moreover, RF has an almost perfect precision for the Normal beat, scoring 0.99, which means that only 1% of normal heartbeats that were predicted resulted in wrong predictions. On the other hand, SVM wins over RF in precision for all the other classes. A score of 0.97 in recall from RF for the RBBB beats means that 97% of total RBBB beats were correctly classified by RF. Comparing the performance of the CART model and SVM classifier, we notice that there is a significant increase in terms of precision and recall in classification of PVC, APC and RBBB heartbeats. Overall, SVM is the best classifier in terms of precision and recall. These values were reflected also in the F-measure column of Table 3.2.5.

	CART		C5.0		Random Forest		SVM	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
Normal	0.79	0.75	0.92	0.89	**0.99**	0.87	0.97	**0.94**
APC	0.36	0.49	0.61	0.72	0.60	**0.91**	**0.67**	0.87
PVC	0.28	0.29	0.68	0.75	0.76	**0.92**	**0.91**	**0.92**
RBBB	0.47	0.50	0.83	0.87	0.76	**0.97**	**0.96**	0.94
LBBB	0.78	0.79	0.87	0.84	0.86	**0.95**	**0.92**	**0.95**

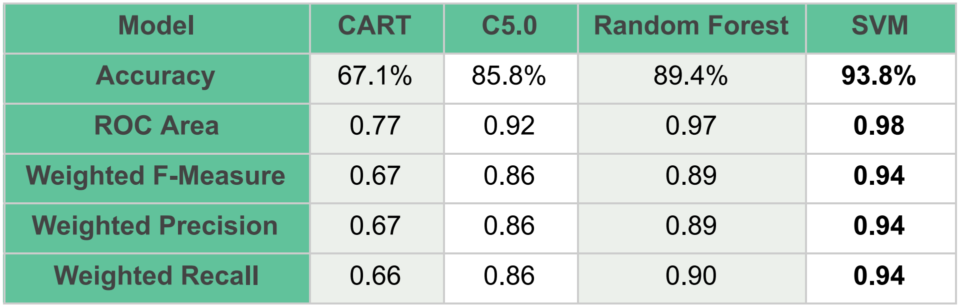
Table 3.2.6 Comparison of the obtained results of precision and recall from all classifiers for each class

Table 3.2.7 displays the specificity and sensitivity score of each classifier for each beat class. It can be seen that for SVM, the sensitivity for RBBB beats is 0.96. This means that the model will correctly classify as RBBB beats 96% of the RBBB beats, but will return a negative result for 4% of the beats that should have resulted as RBBB. A model that is highly sensitive will not generate many false negative results. On the other hand, that RF scores a perfect specificity value for APC and RBBB beats, and an almost perfect specificity score for PVC and LBBB heartbeat classes. This means that RF does not incorrectly classify any beat as RBBB when the beat does not belong to RBBB. The same holds true for APC beats. SVM classifier scores and almost perfect specificity score for APC, PVC, RBBB and LBBB heartbeat classes. Moreover, this classifier achieves the highest specificity value for the Normal beats compared to the other models.

	CART		C5.0		Random Forest		SVM	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
Normal	0.79	0.67	0.92	0.86	**0.99**	0.81	0.97	**0.92**
APC	0.36	0.98	0.61	0.99	0.60	**1.00**	**0.67**	0.99
PVC	0.28	0.92	0.68	0.97	0.76	**0.99**	**0.91**	**0.99**
RBBB	0.47	0.94	0.83	0.99	0.76	**1.00**	**0.96**	0.99
LBBB	0.78	0.96	0.87	0.97	0.86	**0.99**	**0.92**	**0.99**

Table 3.2.7 Comparison of the obtained results of sensitivity and specificity from all classifiers for each class

As seen in Table 3.2.8, CART classifier results in the lowest accuracy of 67.1%, the lowest F-measure of 0.67 and the lowest ROC area of 0.77. In addition, CART classifier has the lowest Precision and Recall of only 0.67 and 0.66 respectively. C5.0 gave a significantly higher accuracy, ROC area and F-measure than CART. Accuracy of CART and C5.0 are still noticeably lower than the results of the Random Forest classifier. The accuracy for Random Forest scores 89.4%, while C5.0 has a 85.8% accuracy. The best performance results on overall accuracy, ROC area, F-measure, precision and recall were obtained with SVM which were 93.8%, 0.98, 0.94, 0.94 and 0.94 respectively. A significant improvement in F-measure is achieved by the SVM model compared to the other classifiers. Taking into consideration all the results reported above, we consider SVM as our best model in heart arrhythmia classification.



Model	CART	C5.0	Random Forest	SVM
Accuracy	67.1%	85.8%	89.4%	**93.8%**
ROC Area	0.77	0.92	0.97	**0.98**
Weighted F-Measure	0.67	0.86	0.89	**0.94**
Weighted Precision	0.67	0.86	0.89	**0.94**
Weighted Recall	0.66	0.86	0.90	**0.94**

Table 3.2.8 Comparison of the obtained overall results from all classifiers

As it can be seen from Figure 3.2.2 and Figure 3.2.3, sd(A6) is ranked as the variable with the highest weight in the prediction of both CART and RF models. The top 5 most important variables for CART classifier are (see Figure 3.2.2): *sd(A6), sd(DWT2), mean(abs_DWT1), sd(DWT1),* and *mean(abs_DWT2)* .

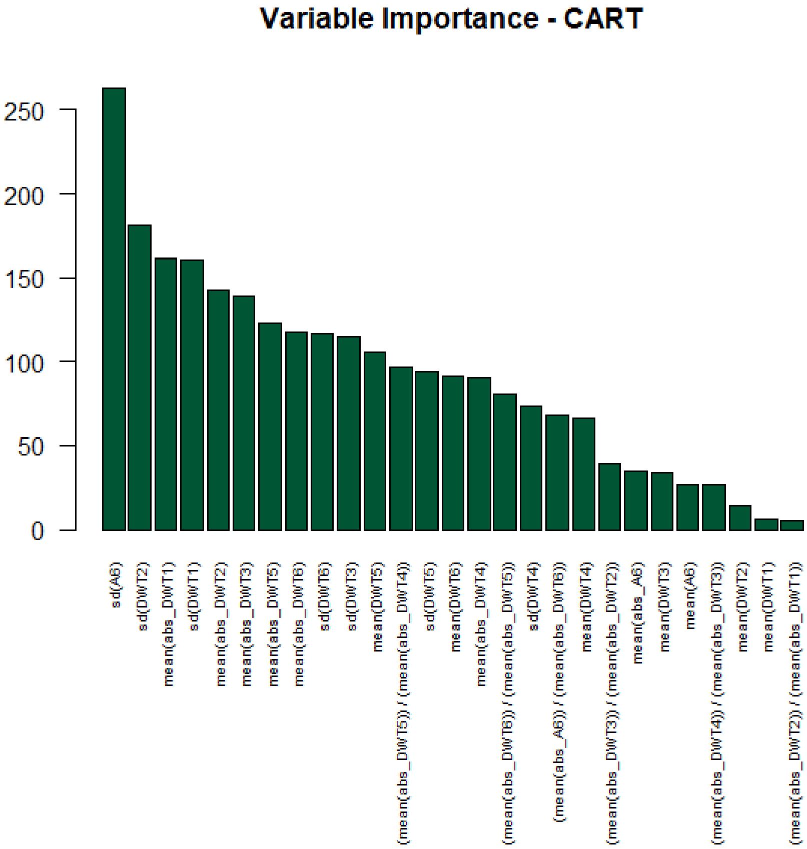


Figure 3.2.2 Barplot of variable importance in CART model

The top 5 most important features for RF classifier based on Gini impurity are (see Figure 3.2.3): *sd(A6), mean(DWT6), (mean(abs_DWT6))/(mean(abs_DWT5)), (mean(abs_DWT5))/(mean(abs_DWT4))* and *sd(DWT6)*. This means that node splits based on these five features on average result in a large decrease of node impurity. On the other hand, the top 5 most important variables for RF classifier based on accuracy are: *sd(A6), mean(DWT6), sd(DWT6), (mean(abs_DWT6))/(mean(abs_DWT5)), (mean(abs_DWT5))/(mean(abs_DWT4))*. In other words, node splits based on *sd(A6)* or on any of the other top 5 features mentioned above result in a large decrease of accuracy. This proves what it is stated in 2.3 that the 6th level detail coefficients of DWT are considered to be the most significant ones. However, it is important to note that the feature importance from RF is calculated based on the training dataset, not on the predictions made on the test dataset. Therefore, it does not indicate the true predictive power of the model.

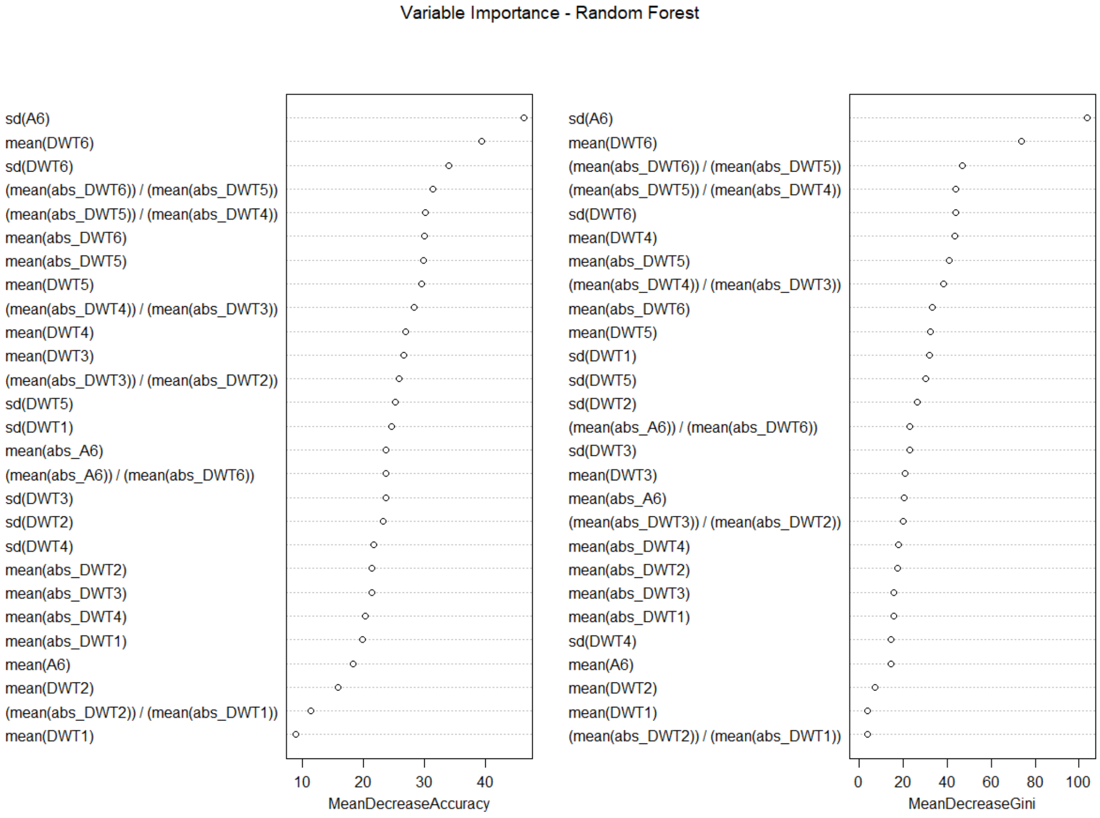


Figure 3.2.3 Variable importance for Random Forest

3.3. Discussion

In this part, we will be comparing the performance results obtained by the suggested system in this paper with the results obtained in other studies. Although it is important to note that other studies analyze different arrhythmia types, use different databases, pre-processing techniques and performance evaluations.

Alickovic and Abdulhamit (2016) conducted four experiments in total, two for the MIT-BIH arrhythmia database and the other using St.-Petersburg Institute of Cardiological Technics 12-lead arrhythmia database. For each database two experiments were performed, one of them was conducted using DWT and statistical indices without applying MSPCA de-noising and the other experiment with MSPCA de-noising. They apply three ML classifiers: CART, C 4.5 and Random Forest. For the MIT-BIH database, without MSPCA de-noising, they achieve classification accuracy of 79% on CART, 80.4% on C4.5 and 85.3% on Random Forest on five different ECG heartbeat types (N, APC, PVC, RBBB, LBBB).

Desai, et al, (2016) performed 9 level sub band DWT decomposition, used ICA for dimensionality reduction and employed SVM classifier with 10-fold cross validation on the ECG signals from MIT-BIH arrhythmia database. Five classes of cardiac arrhythmias (NE, S, V, F, U) are detected with a classification accuracy of 95.24% on SVM linear kernel, 98.42% on SVM polynomial kernel, 97.8% on SVM RBF kernel and the highest accuracy of 98.49% scored using SVM quadratic kernel.

Nayak, et al. (2016) applied PCA instead of ICA and using SVM achieved the following results: 93.13% accuracy on SVM linear kernel, 97.29% accuracy on SVM polynomial kernel, 96.46% accuracy on SVM RBF kernel and the highest accuracy of 97.48% scored using SVM quadratic kernel.

Acir, et al. (2006) introduce a ECG beat recognition system that uses SVM classifier designed with a perturbation method for input dimension reduction, and using DCT for feature extraction and selection, they achieve an accuracy of 96.5%, while using DWT for feature extraction and selection, they achieve an accuracy of 94.0%. They considered only four types of heartbeats: N, LBBB, PVC and non-conducted P-wave.

Obtained results in this study allow us to point out the following:

- The statistical features of DWT coefficients play a significant part in the ECG signal heartbeats classification, as they provide a great characterization and a very good distinction between ECG signal classes.
- The results of our classifiers show that using DWT as the only pre-processing and feature extraction technique, we achieve notable accomplishments and high accuracy in ECG heartbeat classification.
- Random Forest and C5.0 classifiers proposed in this paper produce significant performance compared to the results found in literature.
- According to the results in this study and comparing the performance of Support Vector Machine-based models in ECG heartbeat classification in other studies, we conclude that SVM with DWT as the only pre-processing method results in high performance, while additional pre-processing and

feature extraction techniques like PCA and ICA seem to increase accuracy further.

- Based on the achieved performance of this study, the utilization of Support Vector Machine approach in diagnostic systems results to be simple, efficient, easy and practical.

Conclusion

This thesis is focused on reviewing current research conducted on the classification of ECG signal and on analyzing the Discrete Wavelet Transform as the only preprocessing method for arrhythmia classification on MIT-BIH arrhythmia database. Five heartbeat classes were considered: Normal (N), Premature Ventricular Contraction (PVC), Atrial Premature Contraction (APC), Right Bundle Branch Block (RBBB) and Left Bundle Branch Block (LBBB). Using DWT as the only preprocessing method and applying four different machine learning techniques, namely: Classification and Regression Tree (CART), C5.0, Random Forest and Support Vector Machine (SVM) classifiers, the highest accuracy is achieved on the SVM model. It can classify the five most frequent arrhythmia types with an accuracy rate of 93.8%, AUC value of 0.98 and F-measure score of 0.94.

In order to show the effectiveness and efficiency of our proposed models, we performed a comparison of the results obtained in this study with the results achieved in other research papers. In comparison, other SVM techniques designed with a perturbation method used by other research papers, achieved only a slightly better accuracy of 94% (Acir et al., 2006). There are other papers using Independent Component Analysis (ICA) or Principal Component Analysis (PCA) in addition to DWT, and achieving a somewhat better accuracy of a maximum score of 97.48%. It is important to note that for the other classifiers we still got a slightly better result in terms of accuracy, AUC value and F-measure score than comparable results reported in similar research conducted.

The positive results obtained in this study, encourage for future further design and evaluation of the proposed models in diagnoses of cardiovascular diseases. With this work, we succeed in an efficient, simple and practical machine learning-based approach in heart arrhythmia detection and diagnosis from ECG signal.

Currently, there is a high number of people suffering from cardiovascular diseases around the world, and this number is rising dramatically. Therefore, early detection of heart diseases is crucial to improve the quality of life. The development of medical decision support systems using machine learning-based methodology in diagnoses of heart arrhythmia can help the healthcare industry with considerable reliability and precision.

Future Work

First, we have conducted experiments only using four machine learning techniques and DWT for feature extraction. Even though the performance has exceeded our expectations, we point out that it is possible that different systems or different ML techniques can result in better performance. Moreover, as our dataset is unbalanced, having more examples of rare classes may result in a more accurate heartbeat classification. Regarding the classifier tuning, in this study, we only perform tuning using ten-fold cross validation. Other tuning methods which involve an optimization algorithm may be helpful. Moreover, as one of the goals in this study was to show the advantages of Discrete Wavelet Transform as the only pre-processing method for arrhythmia classification, the use of different de-noising techniques could be implemented to increase classification performance. Finally, as our approach may be considered simple, it could be helpful to use more advanced approaches such as neural networks in classification. However, this is out of the scope of this thesis.

Bibliography

Adam Gacek and Witold Pedrycz. ECG signal processing, classification, and interpretation: a comprehensive framework of computational intelligence. London : Springer, 2012.

Ali Isin and Selen Ozdalili. Cardiac Arrhythmia Detection Using Deep Learning. Procedia Computer Science, vol. 120, 2017, pp. 268–275. doi:10.1016/j.procs.2017.11.238.

Anthony Saxton and Bruno Bordoni. Anatomy, Thorax, Cardiac Muscle. StatPearls. 4 December 2018. [Accessed 17 January 2020]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK535355/

Ary L. Goldberger, Amaral A.N. Luis, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, Chung-Kang Peng, and H. Eugene Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals (2003). Circulation. 101(23):e215-e220.

Brett Lantz. Machine Learning with R: Learn How to Use R to Apply Powerful Machine Learning Methods and Gain an Insight into Real-World Applications. Packt Publ., 2013.

C. Gurudas Nayak, et al. Identification of Arrhythmia Classes Using Machine-Learning Techniques. *International Journal of Biology and Biomedicine.* 2016. Vol: 2 pp: 48-53

D. Sundararajan. Discrete Wavelet Transform: a Signal Processing Approach. John Wiley & Sons Singapore Pte. Ltd., 2015.

ECG and ECHO Learning. Left Bundle Branch Block (LBBB): ECG Criteria, Causes, Management. *ECG & ECHO,* Retrieved from: ecgwaves.com/topic/left-bundle-branch-block-lbbb-ecg-criteria-treatment/.

Emina Alickovic and Subasi Abdulhamit. Medical Decision Support System for Diagnosis of Heart Arrhythmia using DWT and Random Forests Classifier. *Journal of Medical Systems.* 2016. Vol. 40, no. 4. doi:10.1007/s10916-016-0467-8.

Erwan Scornet. Tuning Parameters in Random Forests. *ESAIM: Proceedings and Surveys*, vol. 60, 2017, pp. 144–162., doi:10.1051/proc/201760144.

Gareth James, et al. An Introduction to Statistical Learning: with Applications in R. Springer, 2017.

George B. Moody GB, Roger G. Mark. The impact of the MIT-BIH Arrhythmia Database. *IEEE Engineering in Medicine and Biology Magazine.* 20(3):45-50 (May-June 2001). (PMID: 11446209)

Ian Goodfellow, et al. Deep Learning. MIT Press, 2016.

James B. Wyngaarden, et al. Cecil Textbook of Medicine. Saunders, 2004.

John E. Hall. Guyton and Hall Textbook of Medical Physiology E-Book. 12th ed., Saunders Elsevier, 2011.

John Hampton. The ECG Made Easy. 8th ed., Churchill Livingstone/Elsevier, 2013.

Jiawei Han, et al. Data Mining: Concepts and Techniques. Elsevier, Morgan Kaufmann, 2012.

José L. Rojo-Álvarez, et al. A Review of Kernel Methods in ECG Signal Classification. *ECG Signal Processing, Classification and Interpretation,* Nov. 2011, pp. 195–217., doi:10.1007/978-0-85729-868-3_9.

L. Brent Mitchell, Overview of Arrhythmias - Cardiovascular Disorders. *Merck Manuals Professional Edition*. [Accessed 14 January 2020]. Available from: http://www.merckmanuals.com/professional/cardiovascular-disorders/arrhythmias-and-conduction-disorders/overview-of-arrhythmias

Leo Breiman. Machine Learning. October 2001. Vol. 45, no. 1p. 5–32. doi:10.1023/a:1010933404324.

Malcolm S. Thaler. The only EKG book you'll ever need. Philadelphia : Wolters Kluwer, 2015.

MIT-BIH Arrhythmia Database. MIT-BIH Arrhythmia Database v1.0.0. 24 February 2005. [Accessed November 2019]. Available from: http://www.physionet.org/content/mitdb/1.0.0/

Moody GB, Mark RG. The impact of the MIT-BIH Arrhythmia Database. IEEE Eng in Med and Biol 20(3):45-50 (May-June 2001). (PMID: 11446209)

Moran AE, Wood DA and Narula J. The 2000-2016 WHF Global Atlas of CVD: Take Two. *Global Heart.* 2018. 13(3):139–41. doi:: http://doi.org/10.1016/j.gheart.2018.09.512

Nurettin Acir. A support vector machine classifier algorithm based on a perturbation method and its application to ECG beat recognition systems. Expert Systems with Applications. 2006. Vol. 31, no. 1p. 150–158., doi:10.1016/j.eswa.2005.09.013.

Paulo De Carvalho, et al. Model-Based Atrial Fibrillation Detection. *ECG Signal Processing, Classification and Interpretation*, Nov. 2011, pp. 99–133., doi:10.1007/978-0-85729-868-3_5.

Radomir S. Stanković, and Bogdan J. Falkowski. The Haar Wavelet Transform: Its Status and Achievements. *Computers & Electrical Engineering,* vol. 29, no. 1, 2003, pp. 25–44., doi:10.1016/s0045-7906(01)00011-8.

Rafael C. Gonzalez and Richard E. Woods. Digital Image Processing. 3rd ed., Prentice Hall, 2008.

Raouf Boutaba, et al. A Comprehensive Survey on Machine Learning for Networking: Evolution, Applications and Research Opportunities. *Journal of Internet Services and Applications*, vol. 9, no. 1, 2018, doi:10.1186/s13174-018-0087-2.

Roger J. Lewis. An Introduction to Classification and Regression Tree (CART) Analysis. *Annual Meeting of the Society for Academic Emergency Medicine.* San Francisco, California, 2000.

Sonam Malik and Vikram Verma. Comparative analysis of DCT, Haar and Daubechies Wavelet for Image Compression. *International Journal of Applied Engineering Research.* 2012. Vol. 7, no.11, ISSN 0973-4562

Sylvain Arlot and Alain Celisse. A Survey of Cross-Validation Procedures for Model Selection. *Statistics Surveys*, vol. 4, no. 0, 2010, pp. 40–79. doi:10.1214/09-ss054.

Usha Desai, Roshan Joy Martis, C.Gurudas Nayak, Sarika K. and Seshikala G. Machine intelligent diagnosis of ECG for arrhythmia classification using DWT, ICA and SVM techniques. *2015 Annual IEEE India Conference (INDICON).* 29 March 2016. doi:10.1109/indicon.2015.7443220.

V.V .Ramalingam, et al. Heart Disease Prediction Using Machine Learning Techniques : a Survey. *International Journal of Engineering & Technology*. vol. 7, no. 2.8, 2018, p. 684., doi:10.14419/ijet.v7i2.8.10557.

Wei-Yin Loh. Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery.* 2011. Vol. 1, no. 1p. 14–23. doi:10.1002/widm.8.

List of Figures

List of Tables