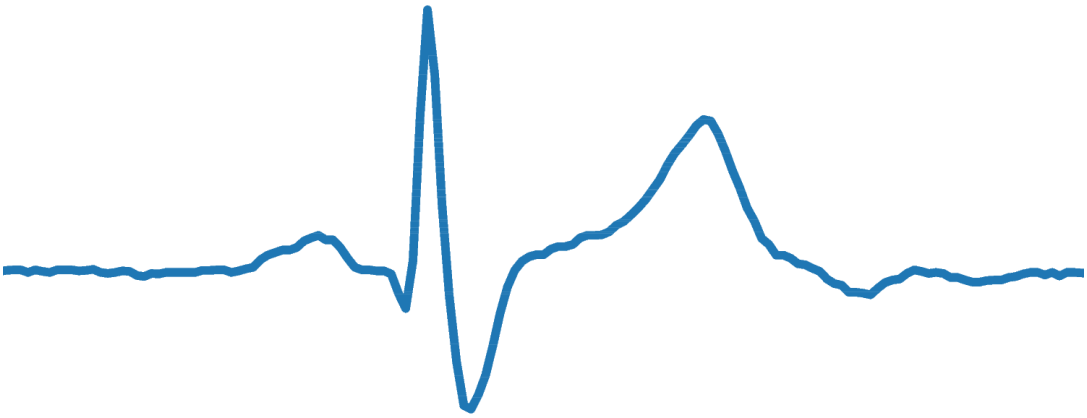
Heartbeat Classification of Electrocardiograms

Interpretive Shapelet Transformation and Classification of Multivariate, Multi-label and Multi-class Electrocardiograms



Project Group:
deis1025f18

Supervisor:
Manfred Jaeger
Stefan Schmid

Summary

This thesis researches the benefit of classifying heart arrhythmias using shape-based features from Electrocardiograms (ECGs). We explore this by developing an interpretive heartbeat classification system for identifying heart arrhythmias based on shapelet transformation of the ECG data.

This study is a continuation of our previous semester project in which we analyzed the Aalborg University Electrocardiograms (AAU-ECG) data set, containing ECG records labeled with heart arrhythmias, using shape-based cluster analysis. We evaluated the clustering using the cluster homogeneity as the quality measure with regards to the heart arrhythmia labels. The results showed that shape-based clustering, on average, could not produce homogeneous clusters. However, we observed a tendency of different leads being good indicators of different heart arrhythmias based on their shape. We leverage this knowledge and use all the ECG leads to train the shapelet based classification system.

This master thesis builds upon work done by researchers in the field of shapelet classification. Ye and Keogh are the authors of the original shapelet paper [1] in which they describe shapelets as a subsequence that can define class membership. Later work by Lines et al. [2] presents the Binary Shapelet Transform (BST) method of transforming the data into a feature vector of distances to shapelets which enables the use of a wider range of classifiers in conjunction with shapelets. We use the BST method to extract shapelets from the data and use them for the transformation. The paper [3] by Rakthanmanon and Keogh introduces Fast Shapelets Search (FSS); an upfront pre-filtering of the shapelet search space by selecting the top k best shapelets based on a heuristic quality measure. We apply the FSS method and evaluate what the tradeoffs are when using FSS compared to the standard full evaluation of the shapelet candidates.

Our heartbeat classification system consists of the following steps; preprocessing, shapelet extraction, shapelet transformation and classification. The preprocessing step reduces the dimensionality of the ECG time series, correct each ECG for baseline wander and applies noise filtering. Then the BST method is used to extract shapelets from a subset of the data set which are then used to transform the data into a feature vector before classification. Finally, we train a heterogeneous ensemble of classifiers on the transformed data. We test our method by applying it on the AAU-ECG and Massachusetts Institute of Technology – Beth Israel Hospital (MIT-BIH) data sets that both contain heart arrhythmia labeled records. The MIT-BIH is often used by researchers to compare their classification methods. We use the MIT-BIH data set to compare our approach to previous work within ECG classification. The AAU-ECG data contains records from Danish patients who underwent ECG recordings at the Copenhagen General Practitioners Laboratory from 2001 to 2015. The Marquette™ 12SL™ ECG analysis program (Marquette) system digitally manages the

AAU-ECG records. The AAU-ECG record contains both heart arrhythmia labels produced by Marquette as well as corrections of these labels made by a doctor reviewing them.

Following the above-mentioned motivations, we define three research questions. The first research question concerns the performance of our approach applied to the AAU-ECG data set. We train our classifier using the doctor's labels as the ground truth and compare the performance of our model against the predictions of the Marquette system. This comparison is one of the fundamental contributions as the Marquette system relies on knowledge-based predefined descriptors of heart arrhythmias, whereas our approach only uses shapelet learned from the ECG waveform. The results show that we on average cannot outperform the Marquette classification system. However, we outperform the system on four diagnoses which are related to left ventricular hypertrophy arrhythmias.

Our second research question concerns the comparison of the performance of our approach on the MIT-BIH data set to ECG classification methods from previous work using the inter-patient scheme. The results demonstrate that our approach has comparable or better performance on three out of the four heartbeat classes, namely normal, ectopic and fusion beats but worse on the supraventricular beats. The normal, ectopic and fusion heartbeat types can be discriminated by studying a single heartbeat; however, the supraventricular heartbeats requires temporal features of the previous heartbeat to be detected. We do not include these temporal features as our feature vector only contains information from a single heartbeat.

Finally, the last research question concerns the performance of using heuristic approximation techniques. We compare the FSS heuristic method to the standard exhaustive search method applied on the MIT-BIH data set. This evaluation shows that the quality of the produced shapelets appears to increase linearly as a function of the heuristic pre-selection size. The run time analysis of the method shows that when using 10% for the heuristic pre-selection size, the shapelet extraction is more than one order of magnitude faster and three times faster when using a ratio of 50%. We propose a novel window constraint method and compare it against the state-of-the-art distance method. We find that the window method improves run time and accuracy.

We conclude that the shapelet based classification approach on ECG data shows promise as it performs comparably or better for heart arrhythmias reflected in the morphology of a single heartbeat than previous work within ECG classification. We also conclude that the approach can, in fact, identify some heartbeat types better than the descriptor based approach leveraged by Marquette.

AALBORG UNIVERSITY

STUDENT REPORT

Cassiopeia
Department of Computer Science
Selma Lagerlöfsvej 300
9220 Aalborg East
Phone: 9940 9940
Fax: 9940 9798
http://www.cs.aau.dk/

Title:
Heartbeat Classification of Electrocardiograms

Theme:
Machine Intelligence

Project Period:
February 2018 - June 2018

Project Group:
deis1025f18

Participants:
Carsten Vestergaard Risager
Kaj Printz Madsen
Morten Brodersen Jensen

Supervisor:
Manfred Jaeger
Stefan Schmid

Number of pages: 94
Appendix: A - D
Date of Completion: June 8, 2018

Abstract:

We propose an automatic and interpretive heartbeat classification approach for identifying heart arrhythmias based on learned shapelets from annotated electrocardiograms (ECG).

The heartbeat classification approach consists of ECG signal preprocessing, shapelet extraction, shapelet transformation, and classification. The preprocessing step involves removal of baseline wander, noise filtering and dimension reduction of the multi-lead ECG signals. The binary shapelet transform is used to extract discriminative subsequences from the ECG, which are used to transform the ECG data into feature vectors. Finally, we train a heterogeneous ensemble of classifiers on the shapelet transformed data set.

We evaluate the performance of the approach on two data sets. The MIT-BIH data set for comparison with previous work within ECG classification following the inter-patient scheme and the AAMI recommendations. As well as AAU-ECG, a real-world multi-labeled ECG data set consisting of 413,151 ECG records where the performance is tested against the industry-leading knowledge-based Marquette 12SL ECG analysis program (Marquette).

The MIT-BIH experiments show that shapelets improves the recall metric of normal and ventricular ectopic heartbeats as well as the precision of fusion beats. In addition, our approach achieves the highest global performance for the four classes with an overall accuracy of 94.3%. For the AAU-ECG data set, the knowledge-based Marquette, in general, surpasses the performance of the learn-based shapelet approach. However, our approach has good discrimination power for right and left bundle branch block, associated with significant cardiovascular mortality, and outperforms Marquette on four diagnoses related to left ventricular hypertrophy.

Preface

This study is written by three master thesis students from the Department of Computer Science at Aalborg University (AAU). The study is a continuation of the pre-specialization semester project by the same students. The project theme is Machine Intelligence with a focus on shape-based classification of electrocardiograms. The project took place during the spring semester of 2018 from February 1st until June 8th.

We would like to thank both our supervisors Manfred Jaeger and Stefan Schmid for providing excellent guidance during the project. We also want to extend our gratitude to Claus Graff Associate Professor from Department of Health Science and Technology for his expert domain knowledge and for providing access to the AAU-ECG electrocardiogram data set.

Contents

Introduction

The Electrocardiogram (ECG) is a recording of the heart's electrical activity and is used by cardiologists to diagnose heart arrhythmias. Heart arrhythmias are a group of conditions in which the heartbeat is irregular, too fast or too slow and are reflected in the morphology of the ECG as abnormal heart activity. Heart arrhythmias are a significant threat and are a subgroup of the cardiovascular diseases which are the most common causes of death worldwide [4]. Due to the high mortality rate of heart diseases early and precise discrimination of heart arrhythmias is vital for detecting heart diseases and choosing appropriate treatment for patients.

Medical experts in clinical settings commonly use knowledge-based systems that use predefined rules and feature descriptors to assist in heart arrhythmia diagnosis of patients [5]. We want to explore what valuable information emerge purely from the ECG waveform to avoid the potential bias from said rules and feature descriptors. A recent promising Time Series Classification (TSC) approach that satisfies this criterion is shapelet based classification [2]. Shapelets are subsequences derived from time series that are defined by their ability to define class membership. Shapelets are learned from labeled training data and do not place any assumptions or restrictions on the structure of the data.

In medical applications, interpretability and the decision process behind the diagnosis given to a patient are of high priority. Health-care practitioners prefer methods where they can understand the contributions of specific features leading to a diagnosis [6, p. 1721]. Shapelets offer a new method for medical practitioners to interpret the correlation between diagnoses and the patterns on the ECG that discriminate the diagnoses.

In collaboration with the Faculty of Medicine at Aalborg University, we are granted access to a data set provided by the Danish health-care system comprised of 974,333 ECG records. Each record is labeled with multiple diagnosis statements by the Marquette™ 12SL™ ECG analysis program (Marquette) [5] followed by a review and potential correction by a doctor. The ECG data contained in the records consist of 12 leads represented as time series making each ECG record multivariate. The data set includes almost 10,000 times the number of unique patients than the Massachusetts Institute of Technology – Beth Israel Hospital (MIT-BIH) data set [7] commonly used in the ECG literature.

Inspired by the recent surge in the success of using shapelets for TSC [8], we propose a method of transforming the multi-labeled, multi-class and multivariate Aalborg University Electrocardiograms (AAU-ECG) data set using shapelet transformation whereby a heterogeneous ensemble of classifiers is trained and evaluated on the transformed data set. The main objective of this master thesis is to address the question of how well shapelet transformed ECG data sets using an ensemble of classifiers can predict diagnoses from ECG

waveforms. Based on the above motivations we construct the three following research questions to guide the project:

- *Can the shapelet transformation classification more accurately predict the doctor's diagnosis compared to the knowledge-based Marquette 12SL ECG analysis program?*
- *Can the shapelet transformation classification approach outperform previous work within heartbeat classification using electrocardiograms?*
- *What are the trade-offs between the run time and shapelet quality when using shapelet heuristic approximating techniques?*

To explore the questions stated above, we conduct shapelet transformation of the AAU-ECG data set before the classification and compare the results against the Marquette predicted statements. Secondly, we compare our proposed method performance with state-of-the-art multivariate time series classification algorithms and previous work within ECG classification on the MIT-BIH arrhythmia data set following the inter-patient scheme. Finally, we explore how the shapelet candidate approximation heuristic called Fast Shapelets Search (FSS) affects the classification result, by gradually increasing heuristically approximated search space on the MIT-BIH data set. The main contributions of the report are the following:

- An approach that uses Binary Shapelet Transform (BST) with an heterogeneous ensemble of classifier for a real-word, multi-labeled, multi-class and multivariate ECG data set which after filtering consisting of 413,151 12-lead ECGs records each attached with a subset of 87 unique diagnosis statements.
- We propose a novel and domain-specific window constraint optimization of the distance calculation for the BST.
- We evaluate our approach against the, currently in medical practice used, knowledge-based analysis program Marquette on the AAU-ECG data set.
- An evaluation of our approach compared to previous work within ECG classification following the inter-patient scheme and Association for the Advancement of Medical Instrumentation (AAMI) recommendations, as well as state-of-the-art multivariate TSC algorithms on the MIT-BIH data set.
- We explore what the trade-offs are between shapelet quality and runtime using the FSS to extract the shapelet candidates in the BST algorithm.

We find that our approach achieves comparable or improved results compared to previous work within ECG classification following the inter-patient scheme on the MIT-BIH data set for three out of four classes. The performance of our approach improves the recall of identifying normal (99.4%) and ventricular ectopic heartbeats (86.6%) as well as the precision of fusion beats (50.7%). Also, our approach achieves the highest global performance for the four classes with an overall accuracy of 94.3%. The limitation of the proposed method on the MIT-BIH data set is the low recall (1.31%) of identifying supraventricular ectopic

heartbeats where the discriminatory feature for this heartbeat type is not reflected in the morphology of a single heartbeat.

We find that the learn-based shapelet approach on the AAU-ECG data set, on average, is surpassed in performance by the knowledge-based Marquette analysis program. However, we improve the recall of identifying diagnosis statements related to left ventricular hypertrophy arrhythmias compared to Marquette with an increase of respectively 14.5%, 9.17%, 29.5% and 62.4%. Also, our approach achieves good discriminating power on bundle branch block arrhythmias.

We also find that our window constraint optimization for the ECG domain reduces the run-time of our approach by a factor of three while achieving improved accuracy. We find that while the FSS heuristic improves the runtime complexity of the shapelet extraction algorithm, it also reduces the overall accuracy of the MIT-BIH data set from 94.3% to 91.0%.

This master thesis is structured as follows; in Chapter 1 we present the required background knowledge of the ECG and the AAU-ECG data set. Next, in Chapter 2 we describe the related work in regards to time series and ECG classification. We elaborate on our approach to performing shapelet-based classification in Chapter 3. In Chapter 4 we present our conducted experiments, evaluate the results, and interesting shapelets are analyzed. We discuss some of our choices made throughout the project in Chapter 5. Finally, we conclude on our project and propose future work in Chapter 6.

1 Background

In this chapter, relevant background knowledge of the domain is presented as well as notational definitions. We describe basic knowledge regarding Electrocardiograms (ECGs) in Section 1.1 and present the Aalborg University Electrocardiograms (AAU-ECG) data set in Section 1.2. Finally we briefly demonstrate the rule-based system Marquette™ 12SL™ ECG analysis program (Marquette) in Section 1.3 and notational definitions in Section 1.4.

1.1 The Electrocardiogram

The ECG is a recording of the heart's electrical activity over a period. The measurement is done by strategic placement of electrodes on the patient's skin that measure the electrical changes in polarisation that occurs as the heart contracts and relaxes. The contraction causes depolarization, and the subsequent relaxation causes repolarization of the tissue. These changes in polarisation are what can be seen as waves on the ECG [9, p. 3].

Figure 1.1 depicts a section of an ECG containing a heartbeat. The figure has annotations symbolizing the waves *P, Q, R, S* and *T* together with important intervals, for instance, the *ST Segment*.

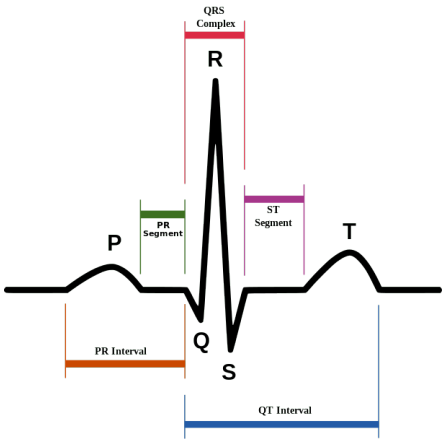


Figure 1.1: *The ECG of a heartbeat. Source: Ref. [10]*

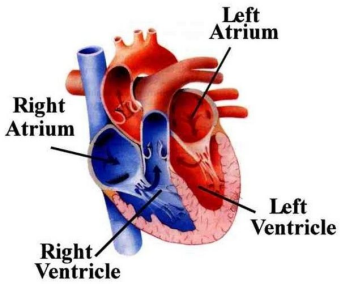


Figure 1.2: *A heart with annotated chambers. Source: Ref. [11]*

The depiction of a heart, seen in Figure 1.2, has two sections blue and red. The blue part

of the heart takes in oxygen-depleted blood and pumps it into the lungs to be oxygenated, and the red section pumps oxygenated blood into the body. We will now briefly explain what phenomena of the beating heart that causes the most important waves and intervals seen in Figure 1.1 based on [9, p. 25-31]. The *P* wave represents depolarization of the right and left atrium tissue. Stimulation of the right atrium pumps oxygen-depleted blood into the right ventricle and stimulation of the left atrium pumps oxygen-rich blood into the left ventricle. The *PR Interval* represents the duration of the time blood pumps into the ventricles and the *PR Segment* is the time in between atrium and ventricle depolarization. The *QRS Complex* is made up of three waves *Q*, *R*, and *S* which is the duration of time in which the right and left ventricles to depolarize. This depolarization contracts the ventricles pumping the oxygen-depleted blood into the lungs and the oxygen-rich blood around the body. The *T* wave represents repolarization of the ventricles which makes them relax allowing the influx of new blood. The *QT Interval* is the total time it takes for the ventricles first to depolarize and then repolarize. The *ST Interval* is the duration of time between depolarization and repolarization of the ventricles. Methods of classifying heartbeats often use features such as the waves and intervals as input to supervised learning methods [7, p. 150].

1.1.1 The 12-lead ECG

The heart is a complex 3-dimensional organ, which makes it difficult to measure its electrical charges adequately [12, p. 37]. The 12-lead ECG configuration solves this by measuring the overall magnitude of the hearts electrical potential from 12 different angles.

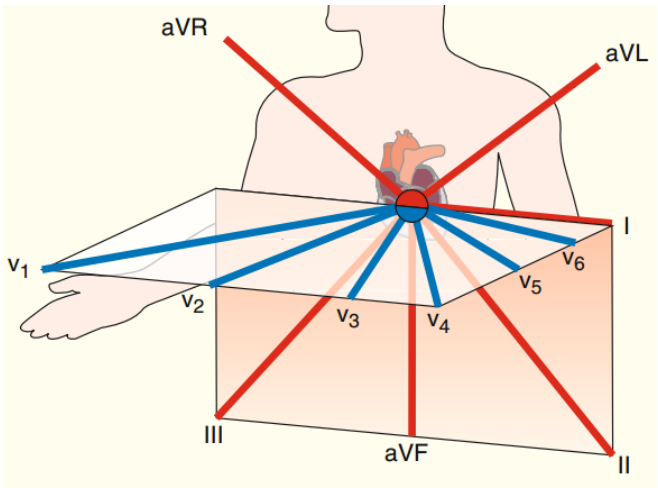


Figure 1.3: *The 12 leads and the angle at which they view the heart. Source: Ref. [9, p. 24].*

The measurement creates 12 different ECG readings also called leads as seen in Figure 1.3. These 12 leads are derived from 10 electrodes strategically placed on the surface of the body [12, p. 37-45]. The six blue electrodes on the chest view the heart in the horizontal

plane, and each produces their own lead, for example, electrode *V1* produces lead *V1*. The placement makes them particularly good at recording the polarity changes that occur in the ventricular tissue of the heart.

The remaining four electrodes produce the last six red leads as seen in Figure 1.3. The four electrodes are placed on the limbs, one on each, such that they view the heart in the vertical plane. The electrode on the right leg acts as ground and is not included in the ECG recording [12, p. 38-39]. The remaining three electrodes creates the basis for the three standard limb leads *I*, *II* and *III* as well as the three augmented limb leads *aVF*, *aVR* and *aVL*. The leads can further be grouped by the angle at which they measure the heart:

- **Lateral leads:** *I, aVL, V5* and *V6* as they measure the side or lateral part of the heart.
- **Inferior leads:** *II, III* and *aVF* measures the heart from below or posteriorly.
- **Septal leads:** *V1* and *V2* measures the heart's inner area or the area opposite the lateral leads.
- **Anterior leads:** *V3* and *V4* measures the front or anterior part of the heart.
- **Lead aVR:** Are not part of any of the groups, but it can be argued that it measures the inner part of the heart.

In practice, and in the AAU-ECG data set we have acquired, the 12-lead ECG only need measurements from two of the three standard limb electrodes, as all limb leads can be derived from just two electrodes using Einthoven's triangle [12, p. 40]. This practice complies with the recommendations of the American Heart Association [13, p. 1313-1314].

1.2 The AAU-ECG Data Set

We have, in collaboration with the Faculty of Medicine at Aalborg University, acquired access to a study population comprised of patients who underwent ECG recordings at the Copenhagen General Practitioners' Laboratory at the request of their general practitioners from 2001 to 2015. We will from now on refer to this data set as AAU-ECG. The ECGs were digitally recorded and analyzed using the Marquette. The study population consists of 974,333 ECG records and a total of 450,232 unique patients where 55% has been recorded once, 21% twice, 11% three times and 13% has four or more records.

The ECG record contains information that uniquely identifies the record, descriptive statements about the record analysis result, the ECG data in the form of leads modeled as time series as well as features such as patients' heart rate. The information that identifies a given record is the patient's ID and the date it was recorded. The patient ID is a sequence of numbers that uniquely identifies that patient in the data set and together with the date it uniquely identifies a single record. The dates are anonymized as they are shifted some unknown amount in time such that the time intervals between records are kept consistent. The statements on the records are assigned by the Marquette, where some of these

statements represent heart arrhythmias. We will be using the statements defining the heart arrhythmias to represent labels in our classification problem. The ECG data, on the record, come in the form of eight leads namely: *I*, *II*, *V1*, *V2*, *V3*, *V4*, *V5*, and *V6*. The lead data is recorded with a standard 12 lead ECG measurement procedure explained in Section 1.1.1. We use Einthoven's triangle to calculate the remaining four leads *III*, *aVF*, *aVL* and *aVR*. The leads are time series that has a raw and a median component. The raw lead is a result of a 10-second measurement sampled at 500 Hertz, resulting in 5,000 data points and the median lead created by the Marquette using the raw lead. What follows is a description of the median leads and diagnosis statements.

1.2.1 The Median

The median lead is a smaller time series representation of a raw lead that contains a single heartbeat with 600 data points. Marquette, in short, creates a median lead by first segmenting the raw lead into individual heartbeats and then calculates the median lead as the average of these heartbeats. We present a median in Figure 1.4 from lead *II*, refer to Appendix B to see examples of all the median leads. We will train the classification model using these medians as per recommendation by our domain expert Claus Graff [14].

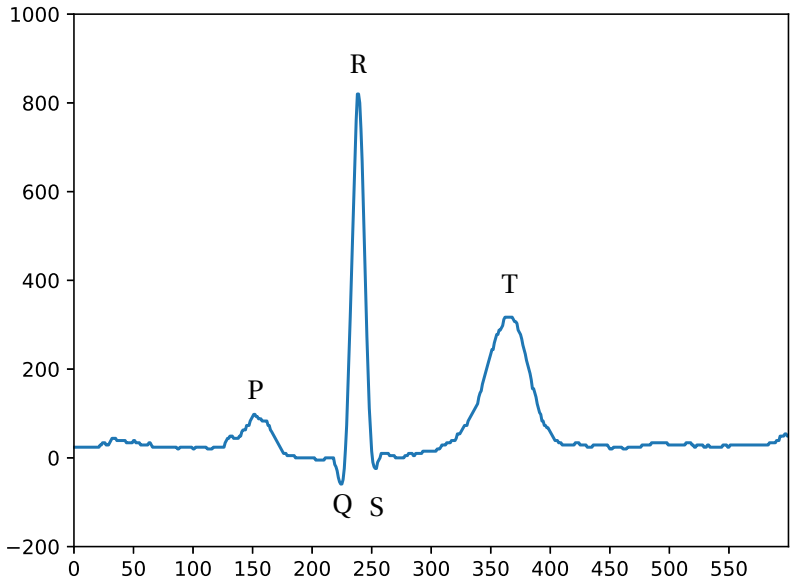


Figure 1.4: *An example of a median for lead* II *where waves are marked.*

We now present a short description of the Median lead creation process that we base on the Marquette manual [5]. The first step in the process is to examine each of the 12 leads

to determine a representation of the QRS complex. The algorithm determines the QRS complex representation by sliding across the 12 raw lead's data, and once it matches a specific criterion, a QRS complex has been detected, and it is then saved as a template, one for each lead [5, p. 3-6].

The template is used to find matching QRS complexes in the given lead. If another QRS complex is detected which do not match the previous template, a new template is created and used for later template matching. This process continues for the full length of the raw leads. Following this, the algorithm determines what heartbeat contains the most information, which can be any heartbeat with at least three matching QRS complexes.

This heartbeat is referenced as a *primary beat*. The creation of the median lead's QRS complex uses the median values of all the beats matching the primary beat. The rest of the median is created in a similar fashion using the beats matching the primary beat and the QRS complex as a base. The creation process of the median includes various steps for noise reduction like the application of high and low pass filters on the data. The algorithm aligns each median such that the beginning of the Q wave has an amplitude of 0 on the y-axis.

1.2.2 The Statements

The Marquette system automatically generates a list of statements for each record it analyses. These statements are manually verified by a doctor on the AAU-ECG data set, and any changes in what the doctor assess to be the correct statements are added to a list of diagnoses separate from the diagnose statements predicted by the Marquette system.

The statements are organized into groups, and they vary in their application. For instance, the group *Technical Problems* cover statements about technical problems that occurred during the recording of the ECG whereas the group *Rhythm* is a type of heart arrhythmia.

The statements groups in the data set can be seen in Table 1.1. The frequency count is made based on the statements assigned by the doctor. The Marquette system have six heart arrhythmia groups each with their own set of heart arrhythmia statements. We have marked these groups with bold in the table. There are two types of heart arrhythmia statements, modifier, and regular statements. Regular statements refer to a single heart arrhythmia diagnosis while the modifiers are attached to regular statements to tweak their meaning. These modifier statements are included in Table 1.1 whereas following tables will not include them. We will limit our analysis to the regular statements as per recommendation by Claus Graff [14].

We will classify heart arrhythmias, and we will therefore only include records with statements from any of the groups marked with bold. Further, we have excluded records that have a *Technical Problem* statement because such a statement implies that a problem occurred during the ECG recording which likely ruins its value to us. We will use the median leads as the basis for our classification analysis, following advice by Claus Graff [14].

	Statements	
Groups	**Unique**	**Total**
Rhythm	69	1,225,657
Infarction	13	221,385
QRS Axis and Voltage	9	113,060
Intraventricular Conduction	16	104,392
Repolarization Abnormalities	46	173,265
Chamber Hypertrophy or Enlargement	18	285,424
Names	5	55
ECG Classification	4	970,719
Technical Problems	3	1165
Miscellaneous	10	315,970
Total	**193**	**3,411,092**

Table 1.1: *Statement distribution of the entire data set.*

This decision necessitates that we delimit the considered heart arrhythmia groups by excluding the *Rhythm* group. The *Rhythm* group is the largest of the diagnosis groups, but this arrhythmia diagnosis is based on an examination of the dynamic differences between multiple heartbeats. This dynamic is not reflected in the median leads, and we will, therefore, exclude this statement group. The remaining five heart arrhythmia groups comprises statements that should be detectable by analyzing the median leads.

	Statements	
Groups	**Unique**	**Total**
Infarction	11	112,409
QRS Axis and Voltage	9	113,060
Repolarization Abnormalities	40	173,265
Intraventricular Conduction	14	103,652
Chamber Hypertrophy or Enlargement	13	266,622
Total	**87**	**750,237**

Table 1.2: *Heart arrhythmia statement groups excluding modifier statements.*

The statement groups we will use is shown in Table 1.2. The statements do not include modifiers as they are not a diagnosis but additional information on the ECG. The data set is multi-labeled which means that records can have any number of statements attached to them from the five statement groups. The data set now consists of 413,151 ECG records

with a total of 211,391 patients where 60% has been recorded once, 20% twice, 9% three times and 11% have four or more records. We show a list of every statement along with their description and frequency count in Appendix A and a histogram in Appendix C. Statement 542 belonging to the group *Chamber Hypertrophy or Enlargement* is the most numerous statement in the data set. We also note that the statement group *Repolarization Abnormalities* is made up of many statements, but many of these occur quite infrequently in the data set. The statements with low total frequency could prove to be problematic as it is hard to draw any substantial conclusions when the available data is low.

One of the research goals of this master thesis is to compare our prediction performance vs. the Marquette predictions. We analyze how often the Marquette and the doctor agrees on diagnoses, disagrees, as well as in what cases the doctor adds or removes a statement. We have identified the five cases that might occur, (I) the diagnoses are identical, (II) there are no statements in common, (III) the doctor adds a statement, (IV) the doctor removes a statement or (V) the doctor and machine agree partially on some of the diagnoses.

We list the cases in Table 1.3 and the frequency they occur in the AAU-ECG. These numbers show that the doctor makes changes on about 50% of the records which inspires some confidence regarding our research question stating whether or not we can perform better than the machine predictions. In the following section, we will describe the rule-based analysis program used to predict the diagnosis statements for the AAU-ECG data set, to help the doctors assigning a diagnosis statement to each ECG record.

Type	Occurrences	% of total
(I) Same	224,131	50%
(II) Changed	48,110	11%
(III) Added	92,731	20%
(IV) Removed	49,225	11%
(V) Partial	33,354	8%

Table 1.3: *The types of disagreement and agreement between the Marquette and doctor predictions.*

1.3 Marquette 12SL ECG Analysis Program

The knowledge-based Marquette system is a computerized ECG analysis program that offers means of recording, analyzing and presenting 12-lead ECGs to medical practitioners. The records in the AAU-ECG data set are all created and analyzed by the Marquette system.

The Marquette labels ECG records using a rule-based system with a tree-like structure that relies on feature descriptors of heart arrhythmia to classify the ECG records. The system can detect rhythm and morphology abnormalities in the ECG. Rhythm abnormalities occur across several heartbeats, whereas the morphology abnormalities are local to a single

heartbeat. We will not include rhythm based abnormalities detection descriptors as they are not detectable by analyzing the median.

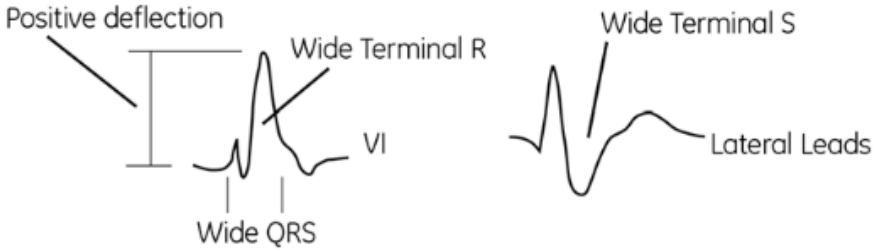


Figure 1.5: *Feature descriptor related to the diagnosis of Right Bundle Branch Block (RBBB (440)). Source: Ref. [5, c. 6 p. 7]*

.

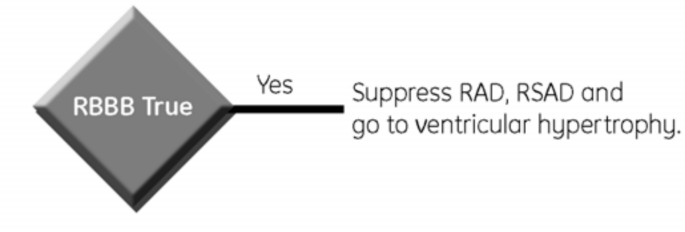


Figure 1.6: *Rule governing the diagnosis of RBBB (440). Source: Ref. [5, c. 6 p. 7]*

.

In Figure 1.5 and Figure 1.6, we show the descriptors related to the RBBB (440) arrhythmia. The picture shows that a positive and wide *QRS* complex, as well as a wide *R* wave, must be present in lead *V1*. Additionally, any of the lateral leads *I, aVL, V5* and *V6* must contain a wide *S* wave. The Figure 1.6 shows how the system will suppress the analysis of statement 380 *RAD* and 382 *RSAD* and will instead test for the presence of Voltage criteria for left ventricular hypertrophy (LVH (540)) when the RBBB (440) feature descriptors are present. This example is relatively simple in comparison to for instance the *ST* segment elevation diagnosis that has two whole pages of rules [5, c. 6, p. 15-16]. For interested readers, we refer to [5, chapters 6-8] for the major categories of the Marquette statement rules. We will now introduce some notational definitions which we will use throughout the remainder of the report.

1.4 Classification Notation

The objective of classification is to construct a classifier using labeled instances that map unseen instances to a label. The classifier is evaluated based on its ability to generalize from the train instances to correctly predict new instances. We use this section to describe

the commonly used notation throughout the report, where an extended version of the notation can be seen in Table 1.4.

We denote a time series data set of n observation as $X = \{T_1, T_2, \cdots, T_n\}$, where each time series has a dimensionality of m is denoted as $T = \{t_1, t_2, \cdots, t_m\}$. The observations are labeled with a subset Y of the total label set L containing discrete values.

We redefine the notation for the data set to also include the time series and their associated labels, $X = \{(T_1, Y_1), (T_2, Y_2), \cdots, (T_n, Y_n)\}$. The data set X is split into two disjoint set a train- and a test data set. A classifier e is trained on the train set whereas the test set is used to evaluate how well the classifier e has learned the relationship between samples and class labels. Multiple classifiers can be combined to form ensembles of classifiers. We denote the ensemble classifier as, $E = \{e_1, e_2, ..., e_f\}$ where f is number of classifiers in the ensemble.

Parameter	**Description**	**Extra**
T	A time series.	$T = \{t_1, t_2, \cdots, t_m\}$
$\|T\|$, m	The dimensions of a time series.	
C	Set of classes in the data set.	$C = \{c_1, c_2, \cdots, c_d\}$
$\|C\|$, d	Amount of classes in the data set.	
Y_x	A subset of predicted labels of the time series T_x.	$Y_x \subseteq C$
X	A data set of time series.	$X = \{(T_1, Y_1), (T_2, Y_2), \cdots, (T_n, Y_n)\}$
$\|X\|$, n	Amount of observations in the data set.	
E	Set of classifiers or an ensemble.	$E = \{e_1, e_2, \cdots, e_f\}$
$\|E\|$, f	Amount of classifiers.	
S	Set of shapelets.	$S = \{s_1, s_2, \cdots, s_g\}$
$\|S\|$, g	Amount of shapelets. in S	
s_x	A continuous subsequence of a time series, a shapelet or shapelet candidate.	$s_x \subseteq T_x$
w	The window used for the *Windowed constraint*	$0 < w < 0.5 \cdot \|T\|$
k	The maximum number of shapelets to find	

Table 1.4: *Lookup table of the notation used in this report.*

2 Related Work

In this chapter, we review the literature in the domain of time series classification with a focus on classification of ECGs. The content of the chapter mostly reflect the challenges that arise when classifying the AAU-ECG data set. These challenges are addressed in Section 2.1 regarding time series classification, where the challenge of handling multiple leads and labels are seen in Section 2.1.1 and Section 2.1.2 respectively. The shapelet-based classification of time series in Section 2.2 and the classification of ECGs in Section 2.3.

2.1 Time Series Classification

A large experimental analysis of 18 state-of-the-art Time Series Classification (TSC) algorithms was conducted by Bagnall et al. in [8]. The experiments were evaluated on the UCR TSC archive [15] consisting of 85 data sets with time series from different domains. The result of their experiments shows that only nine of the proposed algorithms were significantly more accurate than the benchmark classifiers, the 1-Nearest Neighbor (1-NN) using Dynamic Time Warping (DTW) and Rotation Forest (RotF). One of the most successful algorithms within the study where the transformation of data sets with Shapelet Transform (ST) [16] after which the Heterogeneous Ensembles of Standard Classification Algorithms (HESCA) [17] where used for classification. Within the ECG problem domain, the shapelet-based approaches achieved the best performance on the ECG data sets compared to Vector, Interval, Elastic, Dictionary and The Collective of Transformation-Based Ensembles (COTE) [18] based algorithms. The study only considered univariate TSC problems and their primary concern was accuracy. The best performing algorithm, COTE, where hugely computationally intensive making it unsuitable for large data sets.

2.1.1 Multivariate Time Series Classification

A time series is multivariate when it consists of two or more signals or channels. In the ECG domain, each channel of an ECG is called a lead. The aim of Multivariate Time Series Classification (MTSC) is to classify observations, using information using multiple channels.

Three MTSC algorithms adopting the ST where proposed by Bostrom and Bagnall in [19], the algorithms are:

1. **Independent shapelet**: Univariate shapelets are found on a channel and evaluated against the same channel in another time series.
2. **Multidimensional dependent shapelet**: The multivariate shapelet spans across all time series channels, and is evaluated by sliding it across a time series.
3. **Multidimensional independent shapelet**: A multivariate shapelet, but the minimum distance between a time series and a shapelet is found independently for each channel.

Each of the three distance calculation methods results in multiple distances across the channels, which are concatenated into a single feature vector as it achieved better results than using the distances as multivariate features. Of their three proposed algorithms, only the multidimensional dependent shapelet was not significantly worse than the baselines of three multivariate 1-NN DTW classifiers used in the paper. The three ST-based algorithms were restrained to only search for shapelet for one hour, and accordingly, did not process the whole data like the DTW-based algorithms.

The 1-NN classifier using the distance method DTW, has been shown to be a good classifier regarding univariate time series classifications [8]. Shokooshi-Yekta et al. mentioned in [20] two modification to the distance method, which allowed this classifier to be applied to multivariate time series. Given two observations A and B, the independent DTW applies the DTW algorithm on each channels individual, and add the distance scores together, as seen in Equation (2.1) on time series of two channels subscripted with a 0 or a 1. The dependent DTW, seen in Equation (2.2), is similar to the original DTW algorithm with the differences of using the cumulative distance across all the channels at each data point.

$$DTW_I = DTW(A_0, B_0) + DTW(A_1, B_1) \tag{2.1}$$

$$DTW_D = DTW(\{A_0, A_1\}, \{B_0, B_1\}) \tag{2.2}$$

As it was shown in [20] that DTW_I and DTW_D significantly outperforms each other on different data sets. They proposed a scheme to dynamically choose, on an instance base, which of DTW_I or DTW_D did the correct classification using a DTW-based 1-NN classifier. This scheme was dubbed DTW_A and it uses a score-based approach together with a threshold learned on the train data, to decide which prediction of DTW_I or DTW_D to use. In ECG research, the multivariate classification problem is common, as ECG records often include multiple channels that each depict the heart's activity from different angles. Chazal et al. proposed a method for incorporating the information of two leads in [21]. They extracted the same features from each channel and used them as input to two Linear Discriminants (LDs). To combine the results from the two classifiers, the *Unweighted Bayesian Product*

[22] is used as seen in Equation (2.3).

$$\hat{P}(c|x) = \frac{\prod_{e=1}^{|E|} P_e(c|x)}{\sum_{i=1}^{|C|} \prod_{e=1}^{|E|} P_e(C_i|x)} \tag{2.3}$$

where: c = a class,
x = an observation,
E = set of classifiers,
C = set of classes.
$P_e(c|x)$ = estimated posterior probability of the eth classifier

To classify a given observation x, the final posterior probability $\hat{P}(c|x)$ between each class c and the observation x is calculated, and the observation is assigned to the class with the highest score.

2.1.2 Multi-labeled Classification

In a multi-labeled classification problem, each observation can have multiple labels, in contrast to the one-to-one relation between labels and observations of single-labeled problems. We will use the extensive experimental comparison of multi-label methods that was conducted by Madjarov et al. in [23] as a basis for this section. Madjaroc et al. arrange the methods into three groups:

Problem Transformation Methods: Transforming the problem from multi-labeled into one or more single-labeled problems or regression problems, allowing standard classification methods to be used.

Algorithm Adaption Methods: Either adapt or extend existing classification models to handle a multi-labeled problem.

Ensemble Methods: The method, proposed in [23], involves building an ensemble of classifiers using either the methods of problem transformation or algorithm adaptions.

Madjaroc et al. experimented with a total of 12 multi-learning methods with a distribution of three algorithm adaptation, five problem transformations, and four ensemble methods [23]. The two best-performing methods were the random forest ensemble of predictive clustering trees [24] and the problem transformation HOMER [25]. Each tree in the random forest ensemble makes a multi-label prediction of an observation's labels, where a voting scheme decides the final result. The HOMER method is a label power-set method, which tries to combine set of labels into a single label. It uses a divide-and-conquer style to split the multi-label set into smaller problems, where a classifier is constructed at each node.

An interesting result from [23] is that one of the most straightforward problem transformations, binary relevance, was the third best performing method. It uses a one-vs-all strategy, converting the multi-label problem into a single-label binary problem for each unique labels in the label set.

2.1.3 Ensemble-based Classification

Within TSC it has been shown that an ensemble classifier scheme can improve accuracy [18]. An ensemble is a collection of classifiers which are combined to produce a single prediction. The key idea of ensemble construction is that the ensemble should be diverse [26]. Diversity can be achieved by building an ensemble using classifiers from different families of algorithms called heterogeneous ensembles or by changing the training data or training scheme for an ensemble of classifiers from the same family of algorithms called homogeneous ensembles.

The classifier typically used in conjunction with ST is HESCA [17, 18, 27]. HESCA is a heterogeneous ensemble of eight diverse classifiers from different families being probabilistic, tree-based and kernel-based models. The eight classifiers of HESCA can be seen in Table 2.1 with each classifiers parameter settings as specified in [8].

Algorithm	**Type of Model**	**Parameters**	**CV**
Bayesian Network	Probabilistic		
C4.5 Decision Tree	Tree		
K-Nearest Neighbour	Kernel	Initial K = 100	10 Folds
Naive Bayes	Probabilistic		
Random Forest	Tree	500 Trees	
Rotation Forest	Tree	20 Trees	
Support Vector Machine Linear	Kernel		
Support Vector Machine Quadratic	Kernel		

Table 2.1: *Table of the classifiers and parameters employed in the Heterogeneous Ensembles of Standard Classification Algorithms with settings from [8]. CV: Cross Validation*

In [17], Large et al. identified 11 homogeneous ensembles and showed that Random Forest (RandF) and RotF are not significantly worse then HESCA however for the remaining nine homogeneous classifiers HESCA where significantly more accurate. They concluded that it is better to ensemble different classifiers from different families of algorithms then using extra computational time on tuning a single classifier.

2.2 Shapelets

Shapelet-based classifiers have in recent years been shown to produce promising results in the time series classification domain [8]. Ye and Keogh first introduced the shapelet concept in [1] as a new data mining primitive. The authors describe shapelets as subsequences derived from a set of time series, each of which is selected based on its power to define class membership. They are phase independent, meaning that shapelets can occur at any point in a time series [8]. A time series is classified by the presence or absence of one or more shapelets somewhere in the time series.

In the first papers that describe the shapelet classification method, shapelets were tightly coupled with a Decision Tree (DT) classifier [1, 3, 28]. Later work, by Hills et al. in [2], decouples the shapelet extraction method from the classification process using the proposed transformation scheme called ST. The ST method transforms the data into a feature vector using the minimum distance between shapelets and each time series. This decoupling means that the transformed data is free to be used as input to any classifier unlike previous shapelet classification methods [1, 3, 28]. Bostrom and Bagnall propose the state of the art ST-based algorithm Binary Shapelet Transform (BST) in [16]. The BST enforce an equal distribution of shapelets for each class. This method makes it more suitable for multiclass classification problems, as classes that are associated with low-quality shapelets are still ensured to have the same number of shapelets as any other class.

Shapelets has another attractive feature in addition to producing good classification results; they enable a new way of interpreting the classification results. Shapelets allows direct interpretation of the correlation between a classification result and an input sample by studying the shapelets for the class.

The downside of using a shapelet based classification approach is its high computational complexity. The complexity of the shapelet extraction algorithm is $O(n^2m^4)$ [3], where n is the amount of time series and m length of the time series. This downside has motivated much research into improving the run time of the shapelet search algorithm [1, 3, 16, 28]. Methods of reducing the number of distance calculations performed using early abandoning of distance calculations is used in [1, 16]. Ye and Keogh [1] present a way to prune shapelet candidates, an upper bound of the shapelet quality is calculated, and the shapelet is pruned if this bound cannot possibly exceed the quality of the best-so-far shapelet. These optimizations make the average run time faster on most data sets, but the worst case run time remains the same. Mueen et al. in [28] achieve a reduction in time complexity to $O(n^2m^3)$ by caching statistics regarding the distance calculation, hence, trading memory for computational speed. Rakthanmanon and Keogh [3] propose the Fast Shapelets Search (FSS) method that reduces the worst case complexity by decreasing the shapelet candidate search space. The technique prefilters the candidate shapelet search space by selecting the top k best shapelets based on a heuristic quality measure. The FSS algorithm has a time complexity of $O(n^2km^2)$, where k is the number of top shapelets to prefilter based on the heuristic quality measure.

The quality of a shapelet is defined by the quality measure used. The most commonly used quality measure is information gain [1, 3, 28]. The quality measure quantifies the information gained by splitting the data set into two disjoint data sets using the minimum distance from a shapelet to all time series and an optimal splitting point.

An alternative quality measure is presented in [2], which is based on a hypothesis test concerning how the distribution of distances from the shapelet to all the time series of the same class differs. The authors used the F-statistic of a fixed effect Analysis of Variance (ANOVA) but mentioned that they could have used other alternative approaches [2]. For an overview and some more information of the shapelet-based related work mentioned in this section, see the Table 2.2.

Paper	Preprocess	Pruning	Speedup	Quality Measure	Classifier
Ye and Keogh [1]	None	Distance, Candidates.		Information gain	Shapelet-based DT.
Mueen et al. [28]	None	Distance, Candidates.	Cache statistics.	Information gain	Shapelet-based DT.
Rakthanmanon and Keogh [3]	None	Distance, Candidates.	Distance, Candidates.	Information gain	Shapelet-based DT.
Lines et al. [2]	ST	Distance, Candidates.		Information gain, or F-statistic.	HESCA
Bostrom and Bagnal [16]	BST	Distance, Candidates.	Change distance evaluate order.	Not mentioned	HESCA

Table 2.2: *Overview of the shapelet-based related work.*

2.3 ECG Classification

The following text and Section 2.3.1 are modified versions from our previous work [29]. ECG classification is a subset of TSC with a large research field [7, 30], where the focus is to diagnostice each individual hearbeat in a ECG signal. The classification of ECG-based heartbeat methods tries to identify heart diseases by detecting abnormalities in the electrical signal produced by the heart [30]. Luz et al. have in [7] surveyed ECG-based heartbeat classification for arrhythmia detection. This section are based on their findings along with our own investigating of existing studies in the literature.

Luz et al. divide the steps involved in ECG classification into: Preprocessing of the ECG signal, heartbeat segmentation techniques, feature extraction, and classification. Preprocessing of the ECG signal consist of noise reduction of the measured electrical signal into a digital form whereafter different normalization techniques often are used. Examples of contamination of the ECG signal are power line interference and baseline wandering. A common method to correct for the wandering baseline is to isolate it using two median filters and subtract it from the original ECG [7, 21, 31]. Afterward, the power line interference and high-frequency noise can be removed with a low-pass filter [21]. Additionally wavelet

transform [32] and nonlinear Bayesian filters [33] have shown good results in noise reduction while preserving the ECG signal properties. The z-normalization with zero mean and with a standard deviation of one are a commonly used normalization [30].

Heartbeat segmentation techniques are concerned with the segmentation of a heartbeat in the ECG signal relative to detected fiducial points like the R peaks or QRS complexes. To increase the accuracy of the heartbeat segmentation, some algorithms furthermore includes the detection of P wave and T wave associated with the heartbeats [7]. Adaptive detection threshold techniques are widely used for identifying fiducial points as a result of their simplicity and reasonable results [7], and was used in [34, 35]. Other approaches of heartbeat segmentation presented in the literature utilize neural network [36], wavelet transform [37] or filter banks [38].

An essential activity in the classification of ECGs is to extract the relevant features from the segmented heartbeats, such as the amplitude and time intervals of the different waves and complexes [39]. Examples of features which can be extracted from a segmented heartbeat are the P wave, QRS width and QT interval, which can be seen in Figure 1.1. One of the most commonly used features is the cardiac rhythm, measured as the RR interval; the time between two heartbeats' R peaks [7].

We will now explore ECG databases used in previous work and how they are used for heartbeat classification.

2.3.1 Electrocardiogram Databases

Various databases contain cardiac cycles are freely available for ECG arrhythmia classification. The Massachusetts Institute of Technology – Beth Israel Hospital (MIT-BIH) Arrhythmia Database [40] is widely used within ECG classification publications. The data set is recommended by the Association for the Advancement of Medical Instrumentation (AAMI) [41] for creating reproducible and comparable experiments. The MIT-BIH arrhythmia data set consists of 48 records each with a duration of 30 minutes and a sample rate of 360 Hz. Each record consists of two ECG leads: Lead A, which for a majority of the records is a modification of lead II, and lead B that is one of the modified leads V1, V2, V4 or V5. Each heartbeat is independently labeled by at least two cardiologists such that, each heartbeat belongs to one of 15 possible beat types. Table 2.3 shows a comparison between the MIT-BIH and the AAU-ECG data set acquired from the public health sector in Denmark.

Database	Records	Leads	Sample rate	Duration	Attached
MIT-BIH [40]	48 ECG	II and V1	360 Hz	30 min	15 beat annotations
AAU-ECG	974,333 ECG	I, II, V1, V2, V3, V4, V5 and V6	500 Hz	10 sec	12 median leads for each record 193 statements

Table 2.3: *Comparison of the MIT-BIH data set recommended by AAMI together with our AAU-ECG data set.*

By the AAMI recommendation, the 15 beat types of MIT-BIH are divided into the following five groups: Normal beat (N), supraventricular ectopic beat (S), ventricular ectopic beat (V), fusion beats between diagnosis from the V and the N group (F) and unknown beat type (Q). The mapping from the 15 original heartbeat types to the five superclasses for arrhythmias are shown in Table 2.4.

Group Symbol	Group Description	Original Symbol	Original Description
N	Any heartbeat not categorized as SVEB, VEB or Q	L	Left bundle branch block beat
		N	Normal beat
		R	Right bundle branch block beat
		e	Atrial escape beat
		j	Nodal (junctional) escape beat
S	Supraventricular ectopic beat (SVEB)	A	Atrial premature beat
		J	Nodal (junctional) premature beat
		S	Supraventricular premature beat
		a	Aberrated atrial premature beat
V	Ventricular ectopic beat (VEB)	E	Ventricular escape beat
		V	Premature ventricular contraction
F	Fusion beat	F	Fusion of ventricular and normal beat
Q	Unknown beat	P	Paced beat
		U	Unclassifiable beat
		f	Fusion of paced and normal beat

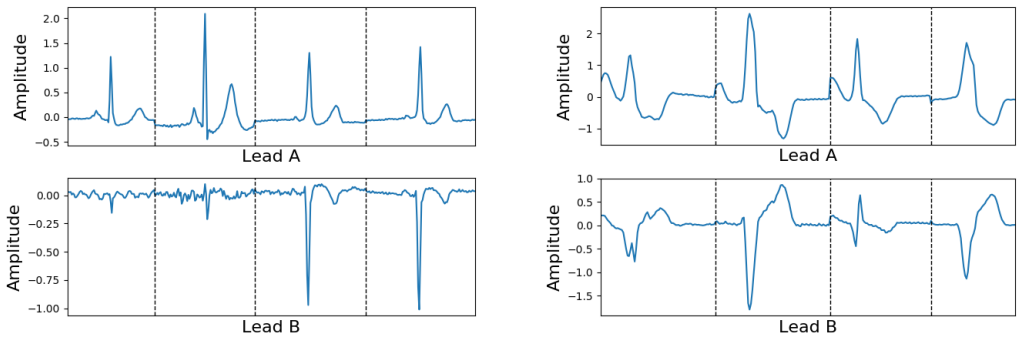
Table 2.4: *The grouping of diagnosis recommended by the AAMI standard.*

2.3.2 Inter-patient Scheme

Within ECG classification, there are two schemes for evaluating arrhythmia classification models [21]; the intra-patient scheme where heartbeats from the same patient are allowed during both the training and test phase opposed by the inter-patient scheme where heartbeats from the same person cannot be used for both test and training. By not mixing the heartbeats in the train and test phase, a more realistic evaluation of the performance of a classification model can be conducted as the evaluation is not biased. The bias occurs in the intra-patient scheme as a result of the classification models tends to learn the particularities of the individual patient's heartbeats during training, which have been shown to give close to 100% accuracy on the MIT-BIH data set [42–44]. However when the proposed methods were evaluated with the inter-patient scheme the performances of the accuracy dropped up to 22.4% [30].

In a realistic scenario where the classification model would be used in a clinical setting, the model should be trained and tested on different patients to learn the discriminative features of the different arrhythmia, instead of learning a given patients heartbeats. In

Figure 2.1 four different patients' heartbeat for lead A and lead B, separated by the dashed lines, are displayed for the N and the V group. The heartbeats in each group have been annotated as the same beat type despite their differences, which illustrates the complexity of the inter-patient scheme.



(a) *Normal Beats (N)* **(b)** *Ventricular Ectopic Beats (V)*

Figure 2.1: *Illustration of heartbeats from different patients diagnosed with the same beat type. Four different patients' lead A and lead B for both the beat group N and V are displayed.*

To report unbiased results aligned with a clinical point of view and for literature comparison, it is recommended by Luz et al. in [30] to follow the AAMI specifications on the MIT-BIH data set with the inter-patient division scheme proposed by Chazal et al. in [21] which can be seen in Table 2.5. The DS1 set is used for training the classification model, and the DS2 set is used for evaluation of the performance.

Data Set Name	Record Number
DS1 - Train	{101, 106, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223, 230}
DS2 - Test	{100, 103, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233, 234}

Table 2.5: *The division of the MIT-BIH data set records following the inter-patient scheme proposed by Chazal et al. [21].*

2.3.3 Automatic Heartbeat Classification

We present the literature of ECG classification which follows the AAMI recommendations and adhere to the inter-patient scheme. An overview and further information of the classifiers mentioned in this section can be found in Table 2.6.

Chazal et al. [21] was an advocate of the inter-patient scheme, and they proposed the most commonly used inter-patient data set split on the MIT-BIH data set. Following we will present previous work within ECG classification following the inter-patient scheme. The best results obtained from Chazal et al. was achieved by extracting two feature sets containing the same 26 features from the two leads of the MIT-BIH data set and used as

input to two weighted LD classifiers. The feature sets contained features like RR-intervals, heartbeat intervals and the morphology of the ECG.

Another weighted LD classifier approach was presented by Llamedo et al. in [45]. They used a sequential floating feature selection algorithm to find the combinations of features, from a more extensive feature set, and the parameter for the classifier model which optimized either the recall or the precision of the classification on the MIT-BIH data set. With the use of only eight features, they were able to outperform Chazal et al. on the V class and have comparable performance on the three remaining classes on the MIT-BIH data set.

A one-versus-one classification scheme was proposed by Zhang et al. in [31], where 6 one-versus-one L_2 regularized Support Vector Machines (SVMs) was build on the four classes of the MIT-BIH data set, excluding the Q class. To handle the imbalances of the data set, the geometric mean between the predicted sensitivities of the negative and positive class are used. The predictions from the six classifiers were combined into a single prediction using a majority voting scheme.

Chen et al. [46] proposed a new feature extracting methods, by applying a random projection matrix on each heartbeat, transforming it into a 30×300 matrix. This matrix had each column normalized, and the discrete cosine transformation was applied to each row to represent the matrix as 30 features. These features were used in classification together with three features of weighted RR intervals. A SVM with a radial basis function kernel was used in the classification of the features, where the parameters to the kernel were learned through 10-fold cross validation on the train set. They tested both the intra-patient and inter-patient scheme, where they achieved a Overall Accuracy (ACC) of 98.46% and 93.1% respectively on the MIT-BIH data set. This further illustrates the importance of not following the intra-patient scheme, as it leads to optimistic results and might not be realistic in real-world practice when the classifiers are trained and tested on different patients.

Authors	**Features**	**# Features**	**Preprocessing**	**Classifiers**
Chazal et al. [21]	Inter-beat intervals, heartbeat intervals, and morphology.	26	Remove baseline wander, and low freq. filter	Weighted LDs
Llamedo et al. [45]	Inter-beat intervals, heartbeat intervals, 2-D CVG loop, and discrete wavelet transform.	8	Discrete wavelet transform	Weighted LDs
Zhang et al. [31]	Inter-beat intervals, heartbeat intervals, and morphology.	46	Remove baseline wander, and low freq. filter	SVMs
Chen et al. [46]	Weighted inter-beat intervals, and random projection matrix.	33	Remove baseline wander, and band-pass filter.	SVM

Table 2.6: *The ECG classifiers which follows the AAMI recomendatiosn and adhere to the inter-patient scheme.*

3 Methodology

We will now present an overview of our methodology. The Figure 3.1 shows the different steps of our methodology each of which we explain in greater detail in the following sections.

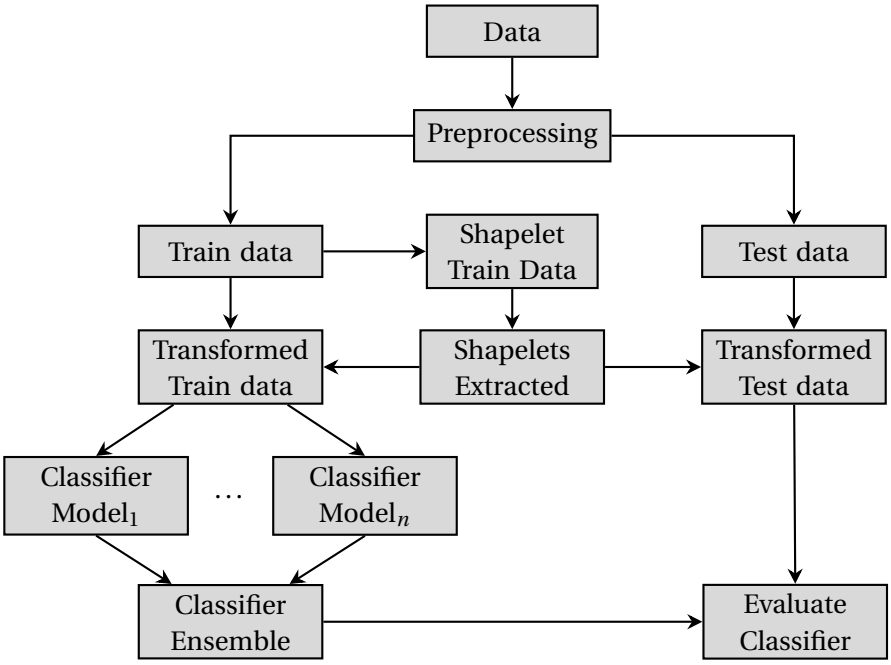


Figure 3.1: *The steps of our methodology.*

Two data sets are used in this project: The AAU-ECG data set and the MIT-BIH arrhythmia data set [40] described in Section 1.2 and Section 2.3.1 respectively.

The data preprocessing step contains various methods we apply to the raw data to prepare it for shapelet extraction, transformation, and later classification. We apply heartbeat segmentation, noise reduction and baseline wander removal methods to the MIT-BIH data set and use the Piecewise Aggregate Approximation (PAA) algorithm on both data sets to reduce the dimensionality of the time series [47]. The preprocessing steps are explained in details in Section 3.2.

The second step splits each data set into two disjoint data sets: A train- and a test data set used in the classification. To handle the multi-labeled AAU-ECG data set we transform it into multiple binary data sets. We present the data split step in Section 3.1.

In the following step we use a modified version of the BST algorithm from [16] to extract shapelets on a smaller subset of the train data, the *Shapelet Train Data* from Figure 3.1. The BST uses the extracted shapelet to transform the train- and test data set. The transformation creates a new feature vector for each time series using the minimum distance between the time series and all the extracted shapelets. The BST algorithm and our modifications to it are explained in Section 3.3.1.

The final step is classification, in which we use the feature vectors to train the HESCA ensemble classifier. The HESCA ensemble contains a diverse set of classifiers and employs a weighting scheme that is learned from the results of each classifier model and applied as the final classifier. Finally, we evaluate the trained classifier using the transformed test data set and four statistic indices, described in Section 3.5: Accuracy, precision, recall and false positive rate. The classification step is detailed in Section 3.4.

3.1 Data Sets

We classify the MIT-BIH and AAU-ECG data sets both of which contain arrhythmia labeled ECG time series. The AAU-ECG dataset is described in Section 1.2. The MIT-BIH data set is widely used in the ECG classification literature [48] and is unique since it consist of beat types for each of the five groups recommended by AAMI [7].

The MIT-BIH Arrhythmia data set consist of a total of 48 ECG records from 47 different patients. As described in Section 2.3.1 each record is multivariate and consists of two leads with a duration of 30 minutes giving a total of approximately 100,000 heartbeats. Heartbeats are labeled by two independent cardiologists such that each heartbeat is annotated with a single label from one of 15 possible types of heartbeat labels. We follow the AAMI recommendations which exclude four of the 48 records from the data set due to paced beats [49].

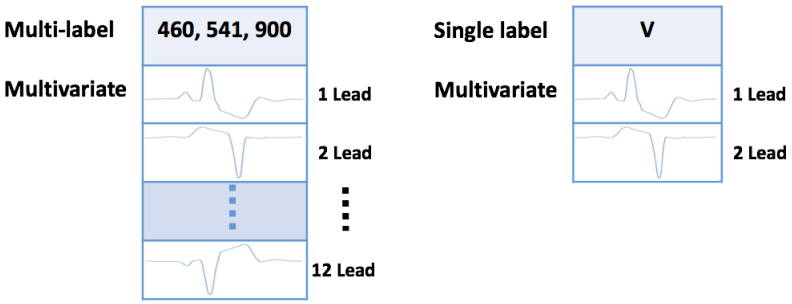


Figure 3.2: *Comparison of a AAU-ECG record and a MIT-BIH record.*

Figure 3.2 illustrates an ECG record for, respectively, the AAU-ECG data set with multi-labeled diagnosis statements and 12 leads compared to the MIT-BIH data set with single

label and two leads. The AAU-ECG data set contains 87 classes whereas the MIT-BIH contains four different classes.

Following the inter-patient scheme proposed in [21] to allow unbiased literature comparison we divide the 44 ECG records from the MIT-BIH data set into the training set (DS1) and the testing set (DS2) each consisting of 22 records as shown in Table 3.1. We discard the Q class as the class is marginally represented in the data set due to the AAMI recommendation of discarding paced beats. The Q class is represented in the data set with only 7 and 8 heartbeats in the train and test data set, respectively, and of no help for further classification purposes [50, 51]

Dataset Name	N	S	V	F	Q	# Beats	# Records
DS1 - Train	45,644	943	3,788	415	8	50,998	22
DS2 - Test	44,221	1,837	3,220	388	7	49,666	22
Total	89,865	2,780	7,008	803	15	100,664	44

Table 3.1: *Distribution of the classes in MIT-BIH data split.*

The AAU-ECG and MIT-BIH data set are partitioned into a training and a test set. For AAU-ECG 70% of the data are used for training and 30% for testing. For the MIT-BIH data set, we follow the standard train test split presented in [21] where roughly 50% of the data are used for training and 50% for testing. We perform shapelet extraction by using a fraction of the ECG records taken exclusively from the training set for both the AAU-ECG and MIT-BIH data sets. We use random sampling to extract an equal amount of records of each class to extract the shapelets from.

3.1.1 Binary Data Set Transformation

As the AAU-ECG data set is a multi-labeled data set, we need to be able to train a classifier on such a problem. We have chosen to adapt our data to the classifier using a problem transformation method called binary relevance [23]. With this transformation, we can use the same classifier on both the AAU-ECG and the MIT-BIH data set. We use the binary relevance method as it is compatible with the BST method, due to its simplicity and it has shown promising results in the experimental comparison of multi-label methods [23].

The binary relevance method follows a one-vs-all approach, where $|E|$ binary classifier is built for each unique class in the data set. In this setting, a single class is used as the positive class where the remaining classes are grouped into the negative class. Hence, we have a binary classification problem for each of the classes in the AAU-ECG data set. The AAU-ECG data set is also multi-labeled. We handle this multi-labeled aspect by treating each record having the given class in its label set as a positive sample and the remaining records are used as negative samples.

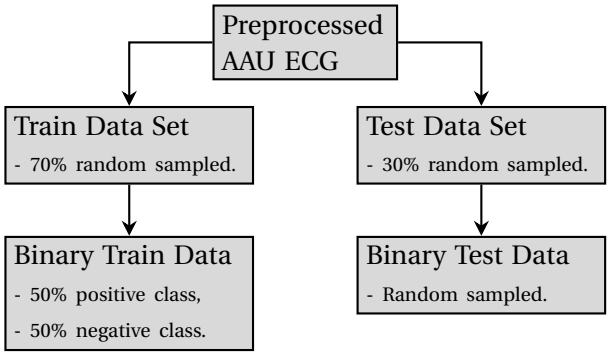


Figure 3.3: *The binary relevance method for data transformation used for each class in the AAU-ECG data set.*

A problem that occurs as a result of applying the binary relevance transformation method is that the distribution of the negative and positive observations in the train set gets skewed as we combine many classes into the negative class. We solve this problem by balancing the train set. We change the distribution of the train data set to consist of 50% observation from the positive class, and then randomly sample the remaining 50% negative observations from the other classes. The Figure 3.3 illustrate how we create the binary train and test set for a class in the AAU-ECG data set. We create the binary test set by randomly sampling observations from the test data set and then assigning the classified class as the positive class, where the remaining classes are combined into a single negative class.

3.1.2 Selection of Diagnoses

The BST algorithm we use to extract shapelets has a rather high time complexity of $O(n^2 m^4)$. To accommodate for this, we reduce the dimensionality m of each observation n using the PAA algorithm. For AAU-ECG we also reduce the number of observations used to extracts shapelets by limiting our focus to a subset of the 87 total classes in the data set.

The diagnosis selection is made in collaboration with associate professor Claus Graff [14], and they are listed in Table 3.2. The selection of diagnoses fall into four diagnosis groups Intraventricular Conduction (IC), Chamber Hypertrophy or Enlargement (CHE), Infarction (INF) and Repolarization Abnormalities (RA). From a medical point of view, these groups are critical as IC and INF carry a high risk of death, and they are all interrelated. The diagnoses of IC and INF are often preceded by diagnoses from CHE and RA. For example, diagnoses in the arrhythmia group RA are most often caused by ischemia, which is when an inadequate amount of blood is supplied to parts of the heart, leading to oxygen-deprived tissue. If untreated, ischemia may eventually lead to the death of the oxygen-deprived area in which case the diagnosis is then classified as an infarct (INF) diagnosis [14]. Infarct diagnoses are among the heart diseases that carry the highest mortality risk and as such the early detection and treatment of the preceding ischemia diagnosis is vital [14].

Group	Name	Description	Count
IC	RBBB (440)	Right bundle branch block	25,532
	IRBBB (445)	Incomplete right bundle branch block	20,780
	LBBB (460)	Left bundle branch block	13,283
	ILBBB (465)	Incomplete left bundle branch block	2,638
CHE	LVH (540)	Voltage criteria for left ventricular hypertrophy	729
	LVH2 (541)	Left ventricular hypertrophy	20,766
	QRSV (542)	Minimal voltage criteria for LVH	73,314
	LVH3 (548)	Moderate voltage criteria for LVH	11,599
INF	SMI (700)	Septal infarct	21,310
	AMI (740)	Anterior infarct	24,146
	LMI (760)	Lateral infarct	3,682
	IMI (780)	Inferior infarct	47,987
	IPMI (801)	Inferior-posterior infarct	823
	ASMI (810)	Anteroseptal infarct	8,845
	ALMI (820)	Anterolateral infarct	2,079
RA	NST (900)	Nonspecific ST abnormality	43,550
	NT (1140)	Nonspecific T wave abnormality	27,801
	NSTT (1141)	Nonspecific ST and T wave abnormality	22,152
	LNGQT (1143)	Prolonged QT	18827
	AT (1150)	T wave abnormality, consider anterior ischemia	3,844
	LT (1160)	T wave abnormality, consider lateral ischemia	11,387
	IT (1170)	T wave abnormality, consider inferior ischemia	7,532
	ALT (1180)	T wave abnormality, consider anterolateral ischemia	4,078

Table 3.2: *The 23 selected diagnoses of the AAU-ECG data set.*

3.2 Preprocessing

In this section, we present the preprocessing and segmentation steps applied to the AAU-ECG and MIT-BIH data sets. We do apply the preprocessing to the data before we do shapelet extraction and subsequent transformation. First, we remove low- and high-frequency noise from the ECG records in the MIT-BIH data set followed by a segmentation of each heartbeat. We do not apply these steps to the AAU-ECG data set as the median, described here Section 1.2, only contains a single heartbeat and the Marquette system applies noise filtering when creating the median. Finally, the dimensionality reduction method PAA is applied on both data sets to reduce the dimensionality of the time series.

The noise contamination mainly comes from three sources [52]; power line interference, frequency waves less than 1 Hz and high-frequency noise. The power line interference referred to as 50/60-Hz buzz and high-frequency noise, typically caused by muscle activity, cause small disturbances in the ECG signal. The frequency waves less then 1 Hz causing baseline wander is an occurrence that happens due to patient respiration.

To eliminate the high and low frequency noise of the ECG signals in the MIT-BIH data set we apply Butterworth Bandpass Filter (BPF) [53]. We chose this method for its high performance and simplicity compared to more complex filters [53]. The leads are filtered with a fifth order BPF with a low-frequency cutoff of 0.05 Hz and a high-frequency cutoff of 150 Hz as a first step to remove baseline wander and muscle activity. We also apply two median filters to remove low-frequency baseline wander [21]. First, each ECG signal is processed with a median filter of 200 ms width to remove the QRS and P waves followed by a median filter of 600 ms width to remove the T waves. We then extract the baseline wander signal from the output of the second median filter. Finally, we subtract the baseline wander signal from the original signal. The result is a signal free from baseline wander an example of which can be seen in Figure 3.4 where the filters are applied to lead *V1*.

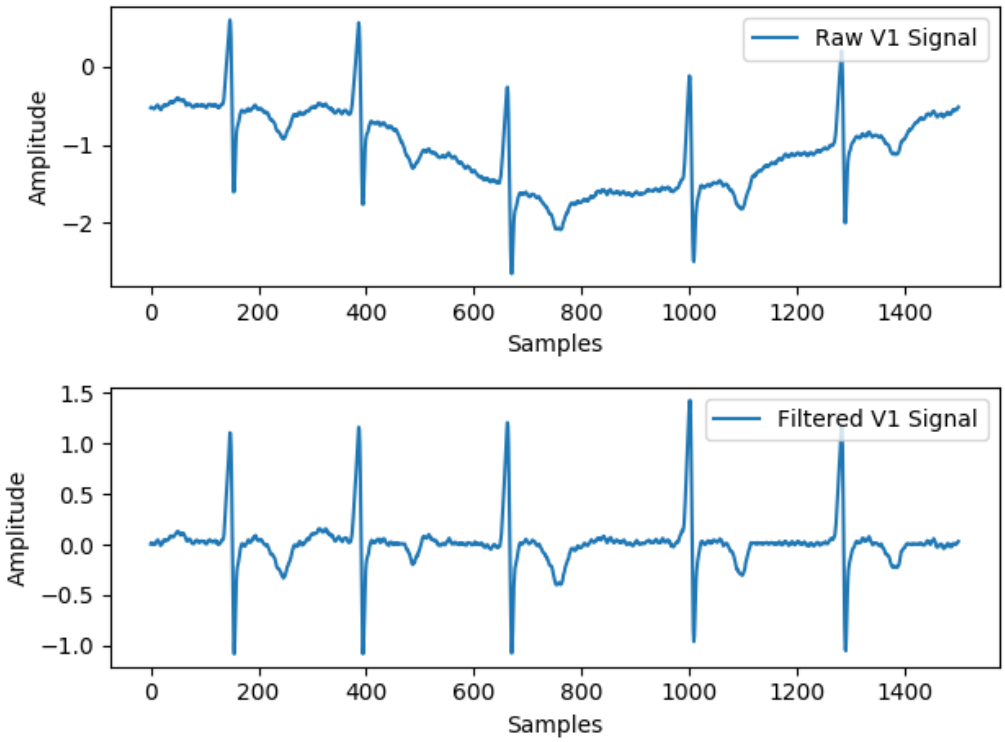


Figure 3.4: *Example of filtering a subsection of the raw lead V1 signal from the MIT-BIH database for removing baseline wander and high frequency noise.*

For the heartbeat segmentation of the ECG signals, the labeled R peaks by the cardiologist found in the annotations files of the MIT-BIH data base are used. A left and a right window are calculated for the current R peak by taking half of the distance from the preceding and succeeding R peak, respectively, to the current R peak.

3.2.1 Piecewise Aggregate Approximation

The BST, explained in Section 3.3.1, is a computationally expensive algorithm, where especially the time series dimensionality affect its runtime. Hence, to give the BST a greater number of time series from which to extract shapelets, the dimensions of each ECG is reduced before applying the BST algorithm. We use the PAA dimension reduction algorithm [47] which can handle time series of an arbitrary length and does a good job of maintaining the shape of the original time series.

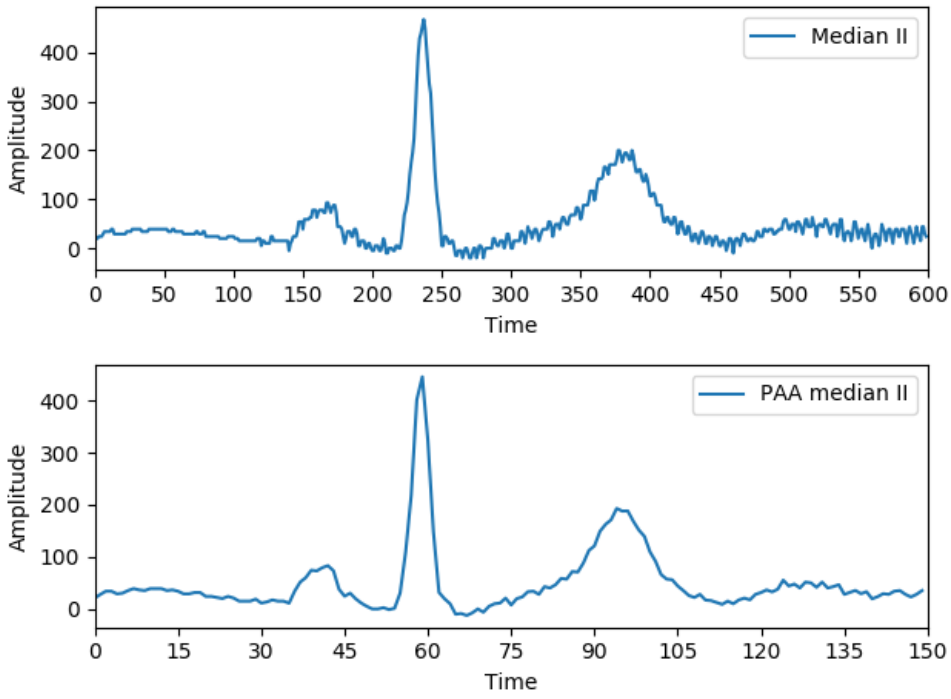


Figure 3.5: *A median lead II before and after reducing the dimensions by a factor of 4 using PAA.*

A positive side effect of using PAA on ECG signals is that the nature of its calculations smooths the time series. This effect is beneficial as it mitigates some of the electrical noise, that the noise filtering did not remove, from the ECG as seen in Figure 3.5. This noise may cause problems during the later shapelet generation, as it can have a significant impact on the z-normalization of a somewhat level candidate shapelet in the BST.

The PAA algorithm reduces the dimensions of a time series T by first partitioning it into b disjoint and equal sized subsequences, also called buckets. Then the mean value of each subsequence is used as a new dimension in the new time series: $\overline{T} = \overline{t_1}, \cdots, \overline{t_b}$, where $1 < b < m$ and m is the length of the original time series. The kth subsequence is calculated as

seen in Equation (3.1) [47].

$$\overline{t_k} = \frac{b}{|T|} \sum_{i=\text{start}}^{\frac{|T|}{b}} t_i, \quad \text{where start} = \frac{|T|}{b}(k-1)+1 \tag{3.1}$$

We reduce the dimensions of all our ECGs by a factor of 4, hence, we have a dimensionality of $b = \frac{600}{4} = 150$ on the AAU-ECG data set. The effect of reducing the dimensions from 600 to 150 can be seen in Figure 3.5.

3.3 Shapelet Transformation

We will now elaborate on the shapelet extraction, and subsequent transformation of the data sets using them. The method is often called *shapelet transformation*, but the majority of the algorithm deals with the extraction of shapelets. Our shapelet transformation approach is based on the BST algorithm from [16] with some modifications to improve run time and enable it to handle multivariate data.

3.3.1 Binary Shapelet Transform

We transform our data using a modified version of the state of the art ST-based algorithm, the BST algorithm, before we use it as input to our classifier. We choose to use a BST as it has been shown to achieve promising results on ECG-based data sets [8] and BST together with the HESCA ensemble was the best performing shapelet-based algorithm in a recent large experimental analysis of state of the art univariate TSC algorithms [8]. The BST algorithm has also been shown to perform better on multi-class problems compared to the original ST algorithm [16]. The BST is used to both extract the shapelets from the shapelet data set and transform the train- and test data sets seen in Figure 3.1. The algorithm can be seen in Algorithm 3.1, where two main steps are essential: The extraction of shapelets on Line 2 and the transformation of the data into feature vectors using the minimum distances from a time series to each shapelet on Line 6.

Algorithm 3.1 binaryShapeletTransform(X, $min \in \mathbb{Z}$, $max \in \mathbb{Z}$, $numSh \in \mathbb{Z}$)

Input: X: A list of time series,
min: Min shapelet length,
max: Max shapelet length,
$numSh$: number of shapelets to use for transformation.
Output: The binary shapelet transformed data set.

```
1: # find the best shapelets with lengths between min and max.
2: S = binaryShapeletSelection(X, min, max, numSh)
3: featureVectors = [|X|][|S|];
4: for i = 0 to |X| do
5:     for j = 0 to |S| do
6:         featureVectors[i][j] = minDist(X.get(i), S.get(j));
7:     end for
8: end for
9: return featureVectors;
```

The algorithm is similar to the original ST algorithm [2]. However, it requires that each class contribute with an equal amount of shapelets and it uses a more efficient distance method in both extracting the shapelets, and the transformation further explained in Section 3.3.3. The equal distribution of shapelet between the classes is advantageous, as we have many classes, represented as diagnosis statements, and the number of ECGs per class is unbalanced, why some classes could otherwise produce more shapelets then others.

The selection of shapelets from Algorithm 3.1 on Line 2 can be seen in Algorithm 3.2. For each time series T, all possible shapelet candidates between min and max length are found on Line 6. Using the minimum distance from a shapelet to all the time series together with the classes these time series belong to, the quality of a shapelet is calculated on Line 11. Information gain or F-statistics are the most commonly used quality measures. The number of shapelets to find $numSh$ is taken equally from each class in the loop on Line 21, where shapelets with a higher quality score are prioritized.

Algorithm 3.2 binaryShapeletSelection(X, $min \in \mathbb{Z}$, $max \in \mathbb{Z}$, $numSh \in \mathbb{Z}$)

Input: X: A list of time series,
min: Min shapelet length,
max: Max shapelet length,
$numSh$: number of shapelets to extract.

Output: A list of the $numSh$ highest quality shapelets with an equal distribution of classes.

```
1:  shapeletMap = {};   # key = class, value = list of shapelets.
2:  topKShapelets = {};
3:  for all T in X do
4:      S = {};
5:      # get shapelet candidates in data set T with length min to max.
6:      candList = shapeletCandidates(T, min, max);
7:      for all s in candList do
8:          # find min dists between candidate s and each series in X.
9:          D_s = minDists(s, X);
10:         # calc quality of shapelet s using distances.
11:         quality = calcQuality(s, D_s, X);
12:         s.setQuality(quality);
13:         S.add(s);
14:     end for
15:     # remove candidates that are similar to candidates of higher quality.
16:     S = removeSelfSimilar(S)
17:     class = T.getClass();
18:     shapeletMap.get(class).add(S);
19: end for
20: numClasses = |X.classes()|;
21: for all key in shapeletMap do
22:     temp = sortByQuality(shapeletMap.get(key));   # sort quality descending
23:     topKShapelets.add(temp.sublist(0, numSh / numClasses));
24: end for
25: return topKShapelets;
```

The BST algorithm is computationally expensive, with a time complexity of $O(n^2 m^4)$, where n is the number of observations and m is the dimensionality of each time series. This complexity makes it unfeasible to apply it on large data sets like ours. The complexity analysis of the algorithm is presented in Section 3.6.

In the literature, speed up improvements for ST and most shapelet-based algorithms can be divided into two categories: (I) Improve the time it takes to find the minimum distance between a shapelet and a time series (Algorithm 3.2 Line 9 and Algorithm 3.1 Line 6), (II) reduce the amount of shapelet candidates evaluated (Algorithm 3.2 Line 6). We use the state-of-the-art minimum distance *improved online subsequence distance* used in the

standard BST [16] together with a novel windowed constraint on the distance subsequence calculations. Further details about our distance method are provided in Section 3.3.3 and Section 3.3.4 respectively. To find the candidate shapelets in Algorithm 3.2 Line 6, we proposed to use the heuristic shapelet search which we name FSS from the *Fast Shapelet* classifier introduced in [3]. The FSS approximate a smaller subset of good shapelet candidates, instead of exhaustively evaluating all shapelet candidates as done in the original ST [2] and BST [16]. More details regarding the FSS can be found in Section 3.3.5.

3.3.2 Measuring The Quality of a Shapelet

The shapelet extraction approach in the original shapelet paper [1] used information gain to asses the quality of a shapelet. In [54] Hills et al. evaluated three alternative similarity measures in the context of shapelet transformed data and concluded that the F-statistic should be the default measure of choice as the F-statistic were significantly faster and more accurate than information gain in their experimental results on synthetic data sets. However in [2] Lines et al. concluded that information gain slightly outperforms the F-statistic. Hence, we will use both the information gain and the F-statistic to measure the quality of a shapelet.

The information gain, in the setting of shapelets, evaluate how much information is gained by using a splitting strategy to divide data set into two using the shapelet and the minimum distance from the shapelet to all the time series. It uses the entropy of a data set X containing two classes A and B, which can be seen in Equation (3.2).

$$I(X) = -p(A)log(p(A)) - p(B)log(p(B)) \tag{3.2}$$

where $p(A)$ is the portion of objects in class A and $p(B)$ is the portion of objects in class B. We use a 1-vs-all strategy, where the class the shapelet candidate originate from is the A class and the B class is the negative class containing all the remaining classes.

A splitting strategy divides X into two disjoint subsets, X_1 and X_2. The information remaining in the entire data set after splitting is defined by the weighted average entropy, $\hat{I}$, of each subset. The total entropy of X after splitting is shown in Equation (3.3).

$$\hat{I}(X) = \frac{|X_1|}{|X|} I(X_1) + \frac{|X_2|}{|X|} I(X_2) \tag{3.3}$$

Given a splitting strategy sp which divides X into to subsets X_1 and X_2 the information gained for the given split is difined in Equation (3.4).

$$IG(sp) = I(X) - \hat{I}(X) \tag{3.4}$$

In our case, the splitting strategy is defined as the tuple $sp =< s, \gamma >$, where a shapelet s is used to calculate the minimum distance between all time series and a distance threshold γ

is used to split the time series based on their distance to the shapelet. The information gain used as a quality measure of a shapelet s is the optimization problem of finding the split of all possible splits Γ that achieves the highest information gain as seen in Equation (3.5).

$$MaxIG(s) = \max_{\gamma \in \Gamma} IG(< s, \gamma >) \tag{3.5}$$

Lines et al. was the first to propose the use of the F-statistic as a quality measure for shapelets [2]. The F-statistic is original a statistical hypothesis test, but Lines et al. only used the F-value to see how the distribution of different classes' distances between time series and a shapelet differs. The intuition is that a good shapelet should be close to other time series of the same class and far from time series of different classes.

To asses the list of distances from all time series to a shapelet $D_s = \{d_1, d_2 \cdots, d_n\}$, we first group time series of the same class together, where D_i is the group for the ith class. The first step in calculating the F-value is to find the average model variability MS_M. We take the sum of squares of the model SS_M, also known as the sum of squares between groups, and average it by the number of values that are free to vary, the degree of freedom DF_M for the model seen in Equation (3.6).

$$MS_M = \frac{SS_M}{DF_M} = \frac{\left(\sum_{i=1}^{|C|} n_i \left(\bar{D}_i - \bar{D}\right)^2\right)}{|C| - 1} \tag{3.6}$$

, where $\bar{D}$ is the mean of all the distances in D, the mean of the group D_i is $\bar{D}_i$ and n_i is the amount of distances in group D_i.

The next step is to find the average residual variability MS_R, which is the average error or variation within each group. This is done by the residual sum of squares or sum of squares within group SS_R and average by the residual degree of freedom DF_R.

$$MS_R = \frac{SS_R}{DF_R} = \frac{\left(\sum_{i=1}^{|C|} \sum_{d_j \in D_i} \left(d_j - \bar{D}_i\right)^2\right)}{n - |C|} \tag{3.7}$$

Then the F-values is a ratio of how good the model is compared to how much error there is.

$$F - \text{value} = \frac{MS_M}{MS_R} \tag{3.8}$$

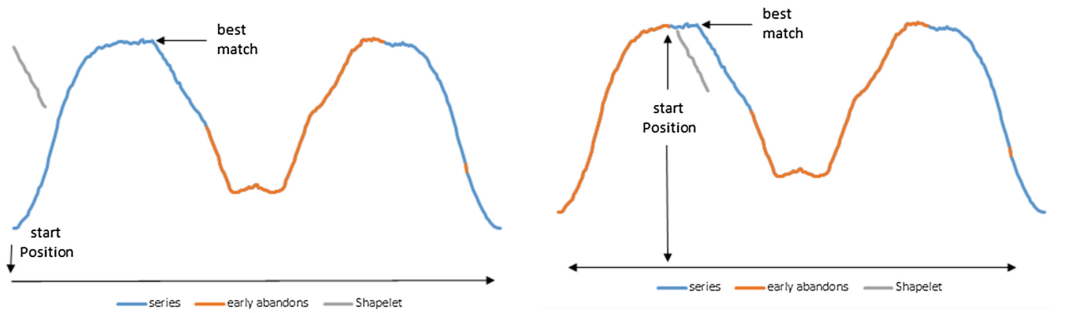
A high-quality shapelet has high F-value, meaning it has a high average variation of the distance between the different group of classes and a low variation of distances within the classes. In our case, we are doing this binary for the class of the shapelet like in the information gain.

3.3.3 Improved Online Subsequence Distance

The improved online subsequence distance measurement was proposed as the distance method used in BST [16] and is an improved version of the distance measure used in the original shapelet paper [1]. It improves the likelihood for early abandoning the distance calculations between a shapelet and a subsequence of a time series, by trying to find a good match to the shapelet early. The distance method used three optimizations to improve the original distance method used with ST from [2]:

1. New evaluation order between a shapelet and a time series
2. Reorder the distance evaluation order between a subsequence and shapelet.
3. Caching repeating calculation.

The improved online distance method changes the way it slides across a time series by starting the distance calculation from the time a shapelet was found and alternating moving one step to the right and one step to the left of the start position. The reason is that there is a high chance that a good match to a shapelet is within its vicinity. The effect of changing the start position can be seen in Figure 3.6. In the case of Figure 3.6, more distance calculation are being early abandoned, the orange part, when starting at the time a shapelet was found seen in Figure 3.6b, compared to previous method used in the ST algorithm [2] seen in Figure 3.6a.



(a) *Euclidean distance early abandon.*

(b) *Early abandon using improved online subsequence distance.*

Figure 3.6: *The difference in early distance abandons by starting at time 0 and slide across the time series (a), and start at the time the shapelet was found and alternating move a step to the right or left from that position (b). Source: Ref. [16]*

Another optimization used, is to reorder the evaluation order between a shapelet and a subsequence, by calculating the distance between the highest valued points first, as they tend to result in more substantial differences and makes the pruning happen earlier.

The last improvement is to cache the previous accumulating sum and squared sum of a time series' subsequence. The sums are updated by only calculating the sums at the last

and first positions of each new subsequence after each step in sliding across the time series, by subtracting the previous sum and add the new. These two sums are used to approximate the standard derivation for the local z-normalization of subsequences when sliding across the time series. As the sliding window are only moving a single step at a time, most of the repeating calculation of the same sum and the squared sum can be reused.

3.3.4 Windowed Constrained Optimization

When working with heartbeat classification of ECG medians, we can use the knowledge that they contain a single heartbeat thus all phenomena like p-waves should appear only once. Leveraging this knowledge, we can ensure that an abnormality found at one place in the median waveform should, if present, be found in the same vicinity on other time series.

With this in mind, we can reduce the search area by only performing distances calculations on a window surrounding the shapelet candidate. We define the window w as a fraction of the whole time series, where $w \in [0, 0.5]$. The window defines how much of the time series we are evaluating left of the start position and right of the end position of where the shapelet was found. The length of the windows to one of the sides are defined as $\tau = mw$. The windowed constraint is the function $W(T, |s|, p, \tau)$ seen in Equation (3.9). The function takes a time series $T = \{t_0, t_1, \ldots, t_{m-1}\}$, the length of the shapelet $|s|$, the position the shapelet was found p, and the length of the window τ as parameters and returns a subsequences of a length at most $2\tau + |s|$ enclosing the shapelet.

$$W(T, |s|, p, \tau) = \begin{cases} \{t_{p-\tau}, t_{p-\tau+1}, \ldots, t_p, \ldots, t_{p+|s|+\tau-1}, t_{p+|s|+\tau}\}, & \text{if } p-\tau>0 \text{ and } p+|s|+\tau<m \\ \{t_0, \ldots, t_p, \ldots, t_{p+|s|+\tau-1}, t_{p+|s|+\tau}\}, & \text{if } p-\tau\leq 0 \text{ and } p+|s|+\tau<m \\ \{t_{p-\tau}, t_{p-\tau+1}, \ldots, t_p, \ldots, t_{m-1}\}, & \text{if } p-\tau>0 \text{ and } p+|s|+\tau\geq m \\ \{t_0, \ldots, t_p, \ldots, t_{m-1}\}, & \text{if } p-\tau\leq 0 \text{ and } p+|s|+\tau\geq m \end{cases} \tag{3.9}$$

There are four cases in Equation (3.9):

1. The window of length τ is extracted at each side of the shapelet candidate.
2. There are enough samples to the right but not enough to the left of t_p, then take what remains.
3. There are enough samples to the left but not enough to the right of $t_{p+|s|}$, then take what remains.
4. The window cannot be extracted from either side, and we take the remaining from both sides.

Using the windowed constrained optimization, the last case can only occur on problems were the window and max length of the shapelet does not satisfy $2\tau + |s| < m$. When it is

satisfied, a time series of length $|T'| = 2\tau + |s|$ is compared to the shapelet in the worst case (1) and in the best case when the shapelet is found at the start or end of the time series only a length of $|T'| = \tau + |s|$ is compared.

3.3.5 Fast Shapelets Search

The FSS is the shapelet extracting method used in the *Fast Shapelet* classifier proposed by Rakthanmanon and Keogh in [3]. It uses a fast heuristic scoring system to find a subset of good shapelet candidates and only uses these for the later shapelet evaluation. The algorithm is centered around transforming each shapelet candidate, by applying the Symbolic Aggregate Approximation (SAX) algorithm [55], into a string representation. We refer to this string representation as a SAX word or simply a *word*. As a measure of shapelet candidate quality, a shapelet candidate's discriminating power is defined by how often other candidates from same classes have the same SAX word versus the *word* of candidates from different classes. The pseudo code for the FSS algorithm is seen in Algorithm 3.3. For all length of shapelet candidates a $saxList$ is created in Line 3. The list contains the SAX words of all shapelet candidates from the data set X having the given length len.

Algorithm 3.3 FSS(X, $min \in \mathbb{Z}$, $max \in \mathbb{Z}$, $iter \in \mathbb{Z}$, $topK \in \mathbb{Z}$)

Input: X: A list of time series,
min: Min shapelet length,
max: Max shapelet length,
$iter$: number of random projection iterations,
$topK$: number of shapelet candidates to use at each length.
Output: Returns the shapelet candidates with the topK best SAX scores.

```
1:  scoreList = {};
2:  for len = min to max do
3:      saxList = CreateSaxList(X, len);
4:      for i = 1 to iter do
5:          freqCount = RandProjection(saxList, X);
6:          scoreList = UpdateScore(scoreList, freqCount);
7:      end for
8:  end for
9:  scoreList = sortByScore(scoreList);    # sort scores descending.
10: scoreList = scoreList.subList(0, topK);
11: # find the candidates of the topK best scores.
12: return findCandidates(scoreList);
```

The algorithm has two steps for creating a SAX word: First, it reduces the dimensionality of the subsequence using PAA after which it transforms the time series into a symbolic representation. The symbolic representation is made using an alphabet of symbols

$A = \{\alpha_1, \alpha_2, \ldots, \alpha_h\}$, where the symbols are equally distributed across the h symbols using cutoff values from the Gaussian curve of the normalized time series' values. Data points that are between two cutoff values are assigned the same symbol. These cutoff values that splits a normalized Gaussian curve into h equal sizes can be found in a lookup table, like the one used in [55]. The selection of the alphabet cardinality h, also called *card size*, is a balancing act between how accurate the candidate shapelets are represented as SAX words versus the amount of the memory they take up. We choose seven as the card size as it the highest cardinality we can express using three bits per symbol, where the "000" bit sequence is omitted because of implementation details. Further, we will not include the PAA step of the SAX algorithm as we apply this dimension reduction to the entire data set already.

When the $saxList$ has been created, the next step is to score the shapelet candidates. Each shapelet candidates can be directly scored using the collision between the *words* in the $saxList$, however, Rakthanmanon and Keogh mentioned that doing may cause *false dismissals* [3]. This problem occurs when two almost identical subsequences are assigned to similar but slightly different SAX words as one or more of their values are barely at each side of the cutoff values. The algorithm handles this problem by applying random projection, see Line 5. The random projection masks different parts of the SAX words in each iteration of the loop at Line 2, and the collisions between identical SAX words together with their classes are kept track of by a collision table. The algorithm updates the scores of discriminating power using this collision table on Line 6. The score is based on how often the word collides with a word of the same class versus the number of collisions with words from other classes. The intuition is that a good shapelet candidate for a class has a similar SAX word compared to others from the same class and is dissimilar to SAX words from the other classes. After making the random projection and updating the scores $iter$ times, the $topK$ highest scoring shapelet candidates are selected and returned for shapelet evaluation. Rakthanmanon and Keogh conclude that the $iter$ parameter which determines the number of iterations of random projection to conduct does not affect the accuracy of their classifier above 10 or so iterations [3], hence, we will use the value of 10 as they do.

The Algorithm 3.3 is slightly altered from the approach used in [3], as they took the $topK$ candidates from each length. Our approach in Algorithm 3.3, on the other hand, does not care that there are equal amounts of shapelet candidates from each length, we want the best shapelet candidates regardless of the length. This fact also entails that our $topK$ parameter needs to be higher than the one used in [3], as we extract the shapelet candidates outside of the for the loop at Line 2.

3.3.6 Extension to Multivariate Time Series

The BST algorithm is by definition an algorithm for univariate time series, but our AAU-ECG data set is multivariate. Bostrom and Bagnall present three methods, described here Section 2.1.1, that when applied would make BST able to handle multivariate data [19].

The activity of the heart is reflected in the waveform of the ECG at roughly the same time on each lead. The propagation of the electrical signal that innervates the heart, causing it to contract, is recorded using lead at different angles which causes the leads to measure activity at slightly different times. There are, however, some diagnoses like RBBB (440) and Left Bundle Branch Block (LBBB (460)) that obstructs the electrical propagation through parts of the heart tissue, making some lead detect the electricity much later than others. Because of this, the distance method used between a shapelet and a time series across multiple channels cannot be strictly dependent in time like the multivariate dependent shapelet method proposed in [19]. On the other hand, the multivariate independent shapelet method from [19], combined with our window constraint, gives us the needed flexibility across the leads, which is why we have chosen to use it. With this new method, we now have a distance from each lead between the shapelet and a time series.

The multivariate independent shapelet method is only the second-best performer out of the three methods[19]. However, none of the tested data sets were ECG-based, and they imposed an artificial time limit of one hour to extract shapelets. Further, the as explained above we believe that the independent method in conjunction with the window constraint is well suited for ECG data.

Bostrom and Bagnall evaluate concatenating the shapelet distances into a single dimensional feature vector against ensemble trained each lead's distances. They found that the concatenation method performed best across multiple classifiers [19]. As we have 12 leads in the AAU-ECG data set, the concatenation method would increase the dimensionality of our feature vector from the shapelet transformation by a factor of 12. This increase in the feature vector dimensionality would have a significant effect on the run time of training the classifier. We could overcome this by drastically limiting the number of shapelets to produce shorter distances vectors. However, we have instead chosen to combine the 12 distances vectors from the leads into a single average distance, which allows us to use more shapelets in shapelet transformation of the time series.

3.4 Classification

The BST algorithm in the previous step, described in Section 3.3, transforms the multivariate time series into a univariate feature vector that can be classified by standard classifiers. We will now briefly present the classifier we have elected to use and our reasons for choosing it.

The classifier typically used in conjunction with ST is the heterogeneous ensemble HESCA that is made up of 8 different classifier models [17, 18, 27]. The HESCA together with ST was the best performing classifier using shapelet in the review of time series classification methods [8], which is why we have chosen to use it as the classifier. We will use a reduced version of the HESCA ensemble, only including the four best performing classifiers on our data sets. The reason is that we have seen that it improves the classification performance

on our data sets, and it improves the run time of the ensemble. The classifiers used in HESCA are listed in Table 3.3, where two of the classifiers themselves are ensemble classifiers of decision trees and the two others are kernel based classifiers.

We use the original implementation of the HESCA ensemble from the UEA & UCR Time Series Classification Repository [15], where the classifiers are implemented in the Waikato Environment for Knowledge Analysis (WEKA) Java library [56].

Algorithm	**Type**	**Parameters**
Random Forest	Tree	500 Trees
Rotation Forest	Tree	10 Trees
Support Vector Machine Linear	Kernel	
Support Vector Machine Quadratic	Kernel	

Table 3.3: *Table of the classifiers and parameters employed in our version of the Heterogeneous Ensembles of Standard Classification Algorithm.*

We will not include the support vector machine with a quadratic kernel when we train the binary classification problems of the AAU-ECG dataset. This decision is based on our experience of this model eliciting poor run time when training on a few problems that are hard to classify. The WEKA library implements the quadratic kernel support vector machine using the Sequential minimal optimization which is an algorithm that solves large quadratic problems by breaking the problem into a series of smaller problems. The algorithm has high run time when it struggles to break the problem into smaller problems [57] which is likely what happens on the hard to classify problems.

3.5 Evaluation Metric

We follow the AAMI metric evaluation recommendations when we evaluate test result of the classifier models. The recommendations specify the usage of four statistical indices, recall, precision, false positive rate (FPR) and ACC for the whole data set. Recall and precision are also sometimes referred to as sensitivity and positive predictivity respectively. The four statistical indices are derived from true positive (TP), true negative (TN), false positive (FP) and false negative (FN) as specified in Equations (3.10) to (3.13). TP is the number of samples that are correctly predicted, and FN is the number of samples that are not predicted as a given class but should have been. TN is the number of samples that the model correctly classifies as not belonging to a given class and FP is the number of samples incorrectly classified as belonging to a given class [58]. When working with a single-labeled and multi-class problem, a one-vs-all approach can be adopted. In this manner, the four values can be calculated for a class using a confusion matrix as seen for the class B in Figure 3.7 as follows:

TP: The intersection between the predicted class and the actual class, the diagonal of the matrix.

FP: The sum of the predicted class' column excluded the TP.

FN: The sum of the actual class' row excluded the TP.

TN: The sum of the remaining cells.

Predicted class

Actual class	A	B	C	D
A	TN_B	FP_B	TN_B	TN_B
B	FN_B	TP_B	FN_B	FN_B
C	TN_B	FP_B	TN_B	TN_B
D	TN_B	FP_B	TN_B	TN_B

Figure 3.7: *A confusion matrix of four classes where the TP, TN, FP and FN are seen for the class B.*

Recall measures the proportion of correctly classified samples of a class.

$$\text{Recall}(c) = \frac{TP_c}{TP_c + FN_c} \tag{3.10}$$

Precision is the fraction of correctly classified samples of a class out of all the samples predicted for that class:

$$\text{Precision}(c) = \frac{TP_c}{TP_c + FP_c} \tag{3.11}$$

FPR is the ratio of the incorrectly classified samples of a class over the total number of samples not classified as the class:

$$\text{FPR}(c) = \frac{FP_c}{TN_c + FP_c} \tag{3.12}$$

ACC measures the fraction of samples correctly classified divided by the total number of samples:

$$\text{ACC} = \frac{\sum_{c \in C} TP_c}{TP + TN + FP + FN}, \quad \text{where } C \text{ is the set of Classes in the data set.} \tag{3.13}$$

As both the AAU-ECG and especially the MIT-BIH data set are very imbalanced, the majority class will massively distort the ACC result. Thus most emphasis should be placed on the recall, precision, and FPR measures when we compare the methods used to do the ECG classification.

3.5.1 Multi-labeled Evaluation

To asses how well the Marquette performs compared to the doctors on the AAU-ECG data set, we need to handle multi-labeled data. One way of doing so is to change the way we count the TP, TN, FP, and FN. Given that the ith record have a set of predicted labels $\hat{Y}_i$ and a set of truth labels Y_i, a class c's values can be computed as seen in Equations (3.14) to (3.17).

$$\text{TP\_Multi}(c) = \sum_{i=1}^{n} \begin{cases} 1, & \text{if } c \in \hat{Y}_i \wedge c \in Y_i \\ 0, & \text{otherwise} \end{cases} \tag{3.14}$$

$$\text{FP\_Multi}(c) = \sum_{i=1}^{n} \begin{cases} 1, & \text{if } c \in \hat{Y}_i \wedge c \notin Y_i \\ 0, & \text{otherwise} \end{cases} \tag{3.15}$$

$$\text{FN\_Multi}(c) = \sum_{i=1}^{n} \begin{cases} 1, & \text{if } c \notin \hat{Y}_i \wedge c \in Y_i \\ 0, & \text{otherwise} \end{cases} \tag{3.16}$$

$$\text{TN\_Multi}(c) = \sum_{i=1}^{n} \begin{cases} 1, & \text{if } c \notin \hat{Y}_i \wedge c \notin Y_i \\ 0, & \text{otherwise} \end{cases} \tag{3.17}$$

With these new definitions, we can to use the statistical indices from Section 3.5 on the multi-labeled data from AAU-ECG.

3.6 Complexity Analysis

We will now analyze the time complexity of some of the important parts of our approach. The important parts are:

1. Dimension reduction of the data using PAA.
2. Extracting shapelets.
3. Transform data using shapelets.
4. Build the ensemble classifier.
5. Evaluate the ensemble classifier.
6. Classify using the baseline 1-NN$_{DTW}$.

The Table 3.4 shows the runtime complexity of the algorithms used in the steps mentioned above as well as explaining the notation used, and Figure 3.8 illustrates the overview of how

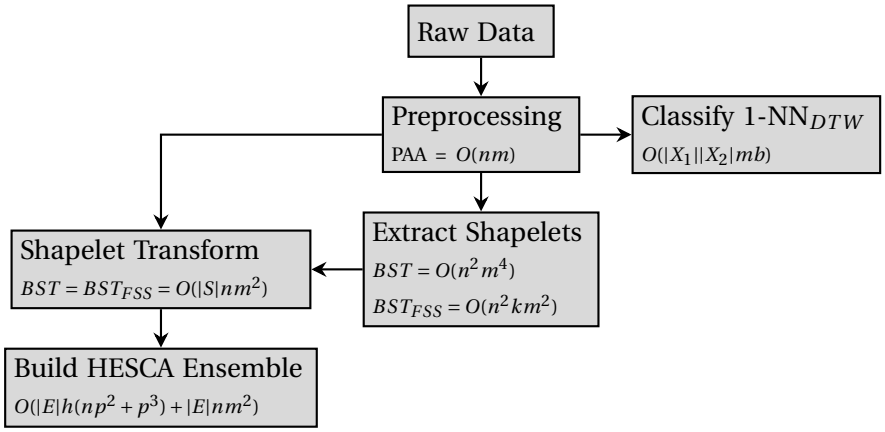


Figure 3.8: *The overview of some of the algorithms used and their complexity.*

we incorporate these algorithms in our approach. We will now further elaborate on these algorithms and their complexity.

The dimension reduction step of a single time series of length m using PAA can be done in a single parse across the time series in $O(m)$ time. Hence, transforming all n time series takes only $O(nm)$ time.

The extraction of shapelets is computationally expensive. Even with the different pruning methods, like the improved online subsequence distance described in Section 3.3.3, the runtime complexity does not improve compared to the brute force method of exhaustively evaluating all possible shapelet candidates. There is a total of $O(m^2)$ possible shapelet candidates from a single time series Line 6, resulting in $O(nm^2)$ total shapelet candidates to evaluate. The distance from a shapelet to all other time series on Line 9 takes $O(nm^2)$ time, which gives a final time complexity of $O(n^2m^4)$ for the brute force- and the BST algorithm to extract all the shapelet candidates.

Once the shapelets have been extracted, we use them to transform the data set into distance feature vectors. The transformation calculates the cross product of the minimum distances between the shapelets and every time series in the data set. The minimum distance between a shapelet and a time series in Algorithm 3.1 at Line 6 has a time complexity of $O(m^2)$ time, which gives a final time complexity of $O(|S|nm^2)$ from the cross product of the $|S|$ shapelets and n time series.

Testing and training an ensemble is upper bounded by the slowest classifier used, which in our case is the ensemble classifier RotF [59]. In the training phase, the RotF generates E classifiers each built on a rotation feature space of the original data set. The rotation feature space is created by multiplying the data set with a sparse rotation matrix. The rotation matrix is constructed by randomly splitting the feature vector into h disjoint subsequences and then applying Principal Component Analysis (PCA) on 75% of the features resulting in a length of $p = \frac{m \cdot 0.75}{h}$. The complexity of this construction is upper bounded by the num-

ber of splits h and the time it takes to apply PCA on each split. The application of PCA computes a covariance matrix in $O(np^2)$ time and the eigen-value decomposition is applied to that matrix in $O(p^3)$ time [60]. The run time complexity of applying PCA on the h subsequences is $O(h(np^2+p^3))$, which is faster than applying PCA on the whole feature vector of length m.

The E classifiers used in our RotF ensemble are all C4.5 decision trees [61]. The C4.5 decision tree has a build time complexity of $O(nm^2)$ [62] and building $|E|$ of such classifiers gives $O(|E|nm^2)$. This makes the final run time complexity of the RotF classifier $O(|E|h(np^2+p^3)+|E|nm^2)$. We omit analzing the time complexity of evaluation the HESCA classifiers as none of the has a high evaluation time.

The baselines 1-NN algorithms, on the other hand, does almost all of the work in the classification step. The algorithm classifies a data set by calculating the cross product of distances between the train and test set. The time complexity of the standard DTW algorithm is $O(m^2)$ [63], then the total complexity is $O(|X_1||X_2|m^2)$ classifying a train set X_1 and a test set X_2. We use an implementation of the DTW algorithm from the WEKA java library that uses a constrained warping path, the Sakoe-Chiba band, which reduces the complexity of the DTW algorithm to $O(mb)$), where b is the maximum warping the warping path are allowed to do.

Method	Complexity	Parameters
PAA dimension reduction.	$O(nm)$	n : Number of time series in the data set, m : Dimensionality of the time series.
BST extracting of shapelets.	$O(n^2m^4)$	
BST_{FSS} extracting of shapelets.	$O(n^2km^2)$	k : Number of shapelet candidates used.
Shapelet transform	$O(\|S\|nm^2)$	$\|S\|$: Number of shapelet used.
Built HESCA	$O(\|E\|h(np^2+p^3)+$ $\|E\|nm^2)$	$\|E\|$: Number of classifiers, h : Number of feature subsets, p : Length of each feature subsets for PCA.
Classify 1-NN_{DTW}	$O(\|X_1\|\|X_2\|mb))$	$\|X_1\|$ Size of the train data, $\|X_2\|$ Size of the test data, b : Sakoe-Chiba band size for DTW.

Table 3.4: *The worst case complexity of the different algorithm used in our approach.*

Given that our data sets, described in Section 3.1, has a large number of observations n and the dimension m is reduced in length using PAA, we can deduce that the real bottleneck is the extraction of shapelets from the data set. This bottleneck is also the reason that we have focused on improving the run time of the BST algorithm described in Section 3.3.1. One thing to notice is that the extracted shapelets can be reused to test different configurations

of classifiers and data sets used for classification.

Changing the shapelet search algorithm to the heuristic FSS described in Section 3.3.5 gives us the most substantial improvement in run time of our changes to the BST algorithm. The FSS heuristic reduces the number of shapelet candidates by pre-selecting the k best scoring candidates based on a collision-score. The slowest part of calculating this score is transforming all the shapelet candidates into the string representation of SAX words. The algorithm needs to parse through the m dimensions of all time series for each shapelet length $l = max - min$. This finally gives a complexity of $O(nml)$.

After the score has been calculated, the best scoring k shapelets are used for further evaluation, and this means that only $O(nk)$ candidate shapelets are evaluated where $k << m^2$. The worst-case complexity of evaluating k candidates shapelets is $O(n^2km^2)$. As m is always greater than l, the time complexity of evaluating the shapelets dominates the time complexity of scoring all the shapelet candidates.

We improve the evaluation of each shapelet candidate by using the improved online subsequence distance algorithm mentioned in Section 3.3.3 together with the windowed constraint in Section 3.3.4. These two methods do not improve the worst case complexity of the BST algorithm as no pruning may happen and a window taking up the whole time series can be used. However, in most cases, most distances calculations are abandoned and a window smaller than the time series is used, which speed up the average run time of the algorithm.

We calculate a more precise complexity, given that the number of shapelet candidates is constrained by the min and max candidate lengths, the total amount of shapelet candidates is reduced from $O(nm^2)$ to $O(nlm)$. Furthermore, we reduce the length of a time series using the windowed constraint. This gives a new max time series length $q = 2mw + max$, where $q \leq m$. This shorted length leads to an improvement from $O(nm^2)$ to $O(nmq)$ time for evaluating the distance between a shapelet and a time series. Using these optimizations, the BST now has a time complexity of $O(n^2m^2lq)$ and BSTFSS a time complexity of $O(n^2kmq)$.

Summary

We have in this chapter presented our approach to applying shapelet-based classification to the AAU-ECG and MIT-BIH data sets. For the preprocessing of the data mentioned in Section 3.2, standard noise reduction filters have been applied and the PAA algorithm is used to reduce the dimensions of the data to increase the speed up of the shapelet transform. The binary relevance method has been adopted to transform the AAU-ECG data set into a binary classification problem for each class, where we are focusing on 23 of essential classes in the data set. We will extract interpretative shapelets from the train data set using the state-of-the-art shapelet transformation BST from Section 3.3, together with two heuristics to speed up the otherwise slow algorithm. Finally, we have chosen in Section 3.4

to use a reduced version of the HESCA ensemble as our classifier for this project. We will follow the performance metrics recommended by AAMI: Accuracy, recall, and precision.

We also presented the performance metrics we use to evaluate the classification results in Section 3.5 and presented the time complexities of central algorithms used our approach in Section 3.6.

4 Experiments & Results

In this chapter, we report the experimental analysis conducted on the MIT-BIH and AAU-ECG data set following our methodology. We start by describing the experimental settings for the two data sets. We then report the results for the MIT-BIH data set compared with state-of-the-art methods followed by the results from the AAU-ECG where we compare the BST to the Marquette 12SL ECG analysis program. The MIT-BIH data set are evaluated with the metrics recommended by the AAMI standard, where we for the AAU-ECG data set report precision and recall. Finally, we present a qualitative evaluation of shapelets extracted from the AAU-ECG data set.

4.1 Preliminary Experiments

We highlight some of our preliminary experiments in this section. We use these experiments to find optimal values for parameters used in the main shapelet-transform and classification experiments. We first perform an experiment that compares the performance of basing our classification analysis on multiple leads versus single leads Section 4.1.1. Next, perform an experiment that finds the optimal size of the window constraint parameter w Section 4.1.2.

4.1.1 Performance of Single-lead Compared to Multi-lead

We want to test if including multiple leads in the analysis improves the accuracy compared to only using a single lead. Within ECG classification lead II is the most used single lead for diagnosing heart diseases [7]. To evaluate the performance of single-lead compared multi-lead, we conduct a preliminary experiment on the MIT-BIH data set. For the experiment, we use the 200 best shapelets from each class with a length between 10 and 50. The constrained window w is set to 0.05. For the single lead configuration, we use the MIT-BIH lead A which corresponds to lead II in a 12 lead ECG configuration. For the multi-lead configuration, we use both MIT-BIH lead A and B.

In Table 4.1 the AAMI performance metrics for the experiment are shown. The results demonstrate that the information from two leads enhances the classification performance of all classes except the *S* class. However, the low amount of samples of the *S* class together with the very low recall makes the difference in precision a result of few correctly classified

Method	ACC	N			S			V			F		
		Pre	Rec	FPR	Pre	Rec	FPR	Pre	Rec	FPR	Pre	Rec	FPR
BST Single-Lead	90.4	94.4	95.1	46.8	**28.6**	1.74	**0.17**	60.4	87.0	4.18	2.25	2.06	0.77
BST Multi-Lead	**93.4**	**95.2**	**98.1**	**42.1**	10.6	**2.01**	0.67	**83.2**	**87.6**	**1.29**	**40.8**	**36.10**	**0.44**

Table 4.1: *AAMI performance comparison of single lead versus multi-lead. The MIT-BIH lead A representing the lead II were used for the single-lead. For multi-lead MIT-BIH lead A and B where used. The bold numbers represent the best score of the two methods.*

samples. As the performance seems to be better using the information from multiple leads, we will in the following experiments use all leads.

4.1.2 Analysis of Windowed Constraint

We want to estimate the optimal size of the window size w of the window constraint. The window size is defined as a ratio of the time series. To do this, we conduct a preliminary experiment on the MIT-BIH data set. We conduct experiments by varying the size of the w as apply BST after which we measure the effect of the window using the classification performance. We extract the 100 best shapelets from each class with a length between 10 and 50. We use the overall test accuracy and Macro F1 score from the classification to evaluate the performance of the parameter w as well as the run time of the shapelet extracting step. The macro average F1 score is more robust measure than the accuracy against the class imbalance of the MIT-BIH data set and can be seen in Equation (4.1). It is the average across all the classes' harmonic mean of precision and recall. The windows size w is specified as a percentage of the time series.

$$\mathrm{F1}_{macro} = \frac{1}{|C|}\sum_{c\in C} 2\cdot\frac{\mathrm{precision}_c\cdot\mathrm{recall}_c}{\mathrm{precision}_c+\mathrm{recall}_c} \equiv \frac{1}{|C|}\sum_{c\in C}\frac{2\cdot tp_c}{2\cdot tp_c+fp_c+fp_c} \tag{4.1}$$

In Figure 4.2 the highest test accuracy (94.3%) and second highest macro F1 score (54.3%) is achieved with a window size of 5% and a run time for the shapelet extraction taking 4525 seconds seen in Figure 4.1. At a window size of 30% the second highest test accuracy (94.1%) and third highest macro F1 score (54.1%) is achieved, but with an increase of the shapelet extraction run time to 8983 seconds. The accuracy when using no window is 91.6% and the run time is 12656 seconds. Another thing worthy of note is that the run time is not linear as a function of the window size. This is because increasing the window size also makes the distance pruning more effective which leads to lower improvement overall from the window.

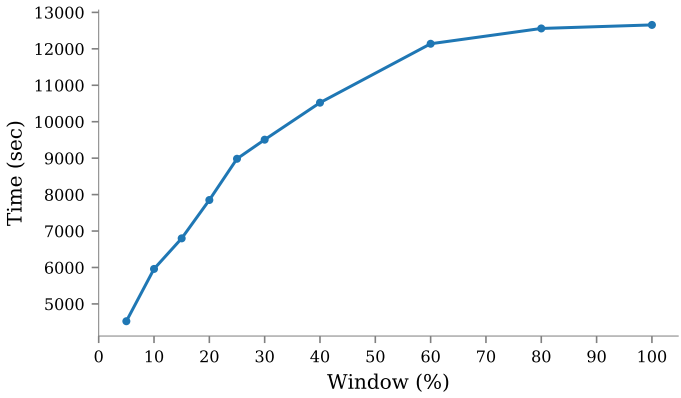


Figure 4.1: *The run time of the shapelet extraction with varying lengths of window w as a percentage of the total length of the time series.*

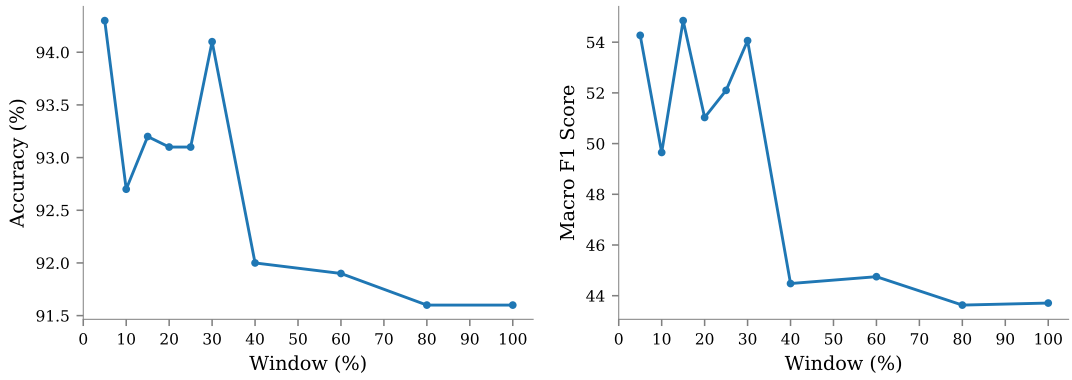


Figure 4.2: *The overall classification accuracy (left graph) and Macro F1 score (right graph) for the MIT-BIH test data set as a function of the size of window w.*

4.2 Experimental Settings

For the experimental analysis of the MIT-BIH data set we follow the inter-patient scheme presented in Section 3.1. We use the reduced HESCA settings for classification described in Section 3.4. For the shapelet transformation, the 200 best shapelets from each class are extracted with a minimum length of 10 and maximum length of 50. We set the windowed parameter w from Section 3.3.4 to 0.05 based on our findings in Section 4.1.2. For the BST_{FSS}, we reduce the shapelet candidates evaluated with 50%, which results in a reduction in the runtime of the algorithm by more than half of the full BST.

In addition to comparing our approach with previous work within ECG classification, we also include the DTW_D and DTW_I to examine our approach in relation to state-of-the-

art multivariate TSC algorithms. DTW_D and DTW_I are described in Section 2.1.1. The 1-NN DTW-based algorithms were not able to classify the MIT-BIH inter-patient scheme data set within a time limit of one week, because of the high computational complexity of the algorithm as described in Section 3.6. To reduce the classification time the test set DS2 was reduced by approximately $\frac{1}{5}$ to 10,000 randomly selected heartbeats leading to $50,998 \cdot 10,000$ distance comparisons between heartbeats. The Sakoe-Chiba band, b, were used in combination with the DTW and specified to 10% of the length of the time series. In Table 4.2 the parameters used for the MIT-BIH experiment can be seen.

Algorithm	Train / Test	Parameter	Description
BST	50,998/49,666	$w = 0.05$, $numSh = 200 \cdot \|C\|$, $min = 10$, $max = 50$.	Window size, Num shapelets extracted, Min shapelet len, Max shapelet len.
BST_{FSS}	50,998/49,666	$w = 0.05$, $numSh = 200 \cdot \|C\|$, $min = 10$, $max = 50$, $topK = 50\%$.	Window size, Num shapelets extracted, Min shapelet len, Max shapelet len, shapelet candidate ratio.
DTW_I	50,998/10,000	$b = 0.1$	Sakoe-Chiba band.
DTW_D	50,998/10,000	$b = 0.1$	Sakoe-Chiba band.

Table 4.2: *The experimental settings for the MIT-BIH data set.*

For the AAU-ECG data set we evaluate the BST against the predictions of the Marquette. To cope with the multiple diagnosis statements on each ECG records the binary relevance method is used to train 23 independently binary ensemble classifiers, one for each diagnosis statement. For each of the 23 binary classification problems of the AAU-ECG data set, we construct an individual train and test set consisting of two classes; the class in question and all other class represented as a single class referred to as the positive and negative class, respectively. We use random sampling to partition 70% of the AAU-ECG data set into an overall train set and 30% to an overall test set. For each of the binary classifiers, we use a subset of the overall train data set where we randomly select instances of the positive class representing 50% of the binary train set. For the remaining 50% representing the negative class we randomly selected from the other classes following the distribution of the overall train data set.

We use an upper limit of max 20,000 records in the binary train data set to be able to classify each of the 23 problems within a time limit of one day. The binary test data set likewise have a limit of 20,000 records. For the binary test data set we randomly select 20,000 records from the overall test data set following the distribution and creates the positive and negative class. 12 out of the 23 diagnosis statements was represented by less than 10,000 records in the overall train data set. In these cases, we extracted all the records for the given statement and then random sampled the same amount of the remaining classes. The eval-

uated was still performed on a test set of 20,000 random sampled records like the other classes. For the shapelet transformation, we extract the 50 best shapelets from each of the 23 classes where the length of the shapelet can be between 10 and 50 samples. The ST was conducted using the shapelets from the 23 classes as it was seen to give higher performance then only using shapelet from the class in question. Table 4.3 shows the parameters used at the AAU-ECG experiments.

Match	Algorithm	Train / Test	Parameter	Description
Precision	BST	20,000 / 20,000	$w = 0.25$, $numSh = 50 \cdot \lvert C \rvert$, $min = 10$, $max = 50$.	Window size, Num shapelets extracted, Min shapelet len, Max shapelet len.
Recall	BST	20,000 / 20,000	$w = 0.25$, $numSh = 50 \cdot \lvert C \rvert$, $min = 10$, $max = 50$.	Window size, Num shapelets extracted, Min shapelet len, Max shapelet len.

Table 4.3: *The experimental settings for the AAU-ECG data set.*

In the experiments, we adjust the confidence threshold of the HESCA ensemble for each diagnosis statement in the classification phase to respectively match the precision and the recall of the Marquette. These experiments are performed once for each of our 23 diagnosis statements that are in focus of this report. If our approach has a higher metric in both experiments of a diagnosis statement, our approach can be said to have outperformed the Marquette on that diagnosis statement.

4.3 Results

We now present the results of our experiments. The result of the MIT-BIH and AAU-ECG classification are found in Section 4.3.1 and Section 4.3.2 respectively. The results for both data sets are followed by a quantitative evaluation of the results and the AAU-ECG evaluation also contains a qualitative evaluation of interesting shapelets.

4.3.1 MIT-BIH Classification

We present the results from the MIT-BIH experiments, the BST and BST_{FSS} along with the results of the state-of-the-art multivariate time series classification algorithms DTW_I and DTW_D In addition we report the results from previous work within ECG classification following the inter-patient scheme. Information about the algorithms used by the previous

work can be found in Section 2.3.3. In Table 4.4 we report the overall accuracy, precision, recall and FPR following the AAMI recommendations.

Method	ACC	N			S			V			F		
		Pre	Rec	FPR	Pre	Rec	FPR	Pre	Rec	FPR	Pre	Rec	FPR
BST	**94.3**	94.7	**99.4**	46.1	49.0	1.31	**0.05**	91.5	**86.6**	0.58	**50.7**	19.8	**0.16**
BST_{FSS}	91.0	92.7	97.9	64.3	11.9	2.72	0.08	69.6	58.6	1.86	0.49	0.26	0.34
DTW_I	88.4	95.9	92.6	37.8	24.4	12.8	1.70	51.5	84.0	5.60	3.00	7.40	1.80
DTW_D	87.7	97.0	90.1	25.5	**78.8**	53.3	0.60	44.5	80.8	7.10	6.00	27.9	3.30
Chazal et al. [21]	81.9	99.2	86.9	6.00	38.5	75.9	4.60	81.9	77.7	1.20	8.60	89.4	7.50
Zhang et al. [31]	88.3	98.9	88.9	7.16	35.9	**79.1**	6.05	**92.8**	85.5	**0.46**	13.7	93.8	0.46
Llamedo et al. [45]	78.0	**99.5**	77.6	**3.32**	41.3	76.5	4.17	87.9	82.9	0.79	4.23	**95.4**	10.2
Chen et al. [46]	93.1	95.4	98.4	37.4	38.4	29.5	1.90	85.1	70.8	0.90	NaN	NaN	NaN

Table 4.4: *Comparison of the proposed methods, state-of-the-art multivariate time series classification algorithms and previous work following the AAMI recommendations and the inter-patient scheme in %.*

Table 4.5 displays the confusion matrixes of our experiments obtained by the BST and BST_{FSS} on the DS2 test set. We also present the confusion matrices for the DTW_I and DTW_D on the reduced DS2 test set with 10,000 heartbeats in total.

		Predicted			
		N	**S**	**V**	**F**
Actual	**N**	43949	19	204	49
	S	1786	24	27	0
	V	398	6	2790	26
	F	284	0	27	77

(a) *BST 49,666 heartbeats.*

		Predicted			
		N	**S**	**V**	**F**
Actual	**N**	43292	26	703	200
	S	1810	5	21	1
	V	1318	11	1887	4
	F	288	0	99	1

(b) *BST_{FSS} 49,666 heartbeats.*

		Predicted			
		N	**S**	**V**	**F**
Actual	**N**	8262	126	426	110
	S	236	49	51	47
	V	74	22	525	4
	F	42	4	17	5

(c) *DTW_I 10,000 heartbeats.*

		Predicted			
		N	**S**	**V**	**F**
Actual	**N**	8046	27	606	246
	S	149	204	16	14
	V	64	21	505	34
	F	36	7	6	19

(d) *DTW_D 10,000 heartbeats.*

Table 4.5: *The confusion matrices of our results on the MIT-BIH DS2 test set.*

Table 4.4 shows the results of our approach compared to the results of previous work following the inter-patient scheme described in Section 2.3.3. The BST algorithm achieves the highest ACC of 94.3% followed by the algorithm proposed by Chen et al. at 93.1% and BST_{FSS} at 91.0%.

The precision, recall, and FPR for class N of the BST algorithm are similar to the results

reported by Chen et al. The high performance of the N class in terms of recall and precision for both BST and Chen et al. is the main contributor to the high ACC as the N class represents 89% of the heartbeats in the test set.

The BST achieves the highest recall class V but with slightly lower precision than Zhang et al. The BST algorithm also achieves the highest precision of 50.7% on the F class and the lowest FPR of 0.16%. The previous work seems incapable of separating the F class from the two larger classes N and V resulting in a high FP error, hence the low precision. This fact might be because the F class contains heartbeats that are a fusion of the two classes N and V, making it similar to both of these classes. The algorithm from previous work, excluding Chen et al., has a higher recall on the F class than our proposed method.

Our approach has a low recall on class S compared to the results in previous work. The S class is related to heartbeats where different parts of the heart contracts prematurely. Determining whether or not a ECG contains premature waves requires knowledge of the prior heartbeats which is only reflected in the rhythm of an ECG. This information is not present in our approach, as we only use the distance from individual heartbeats to shapelets as features and do not use features from other heartbeats. However, the DTW_D baseline is one of the best performing algorithms for this class, with the highest precision of 78.8% and a recall of 53.3%. According to Associate Professor Claus Graff, the S class heartbeats are identical to the N class heartbeats. This fact affirms what we see in Table 4.5, where the BST algorithm predicts the heartbeats of the S class as the N class in 97.2% of the cases. An example of a premature beat from the S class can be seen in Figure 4.3

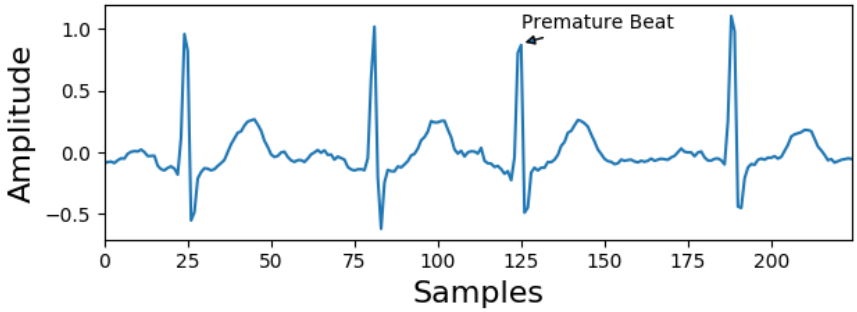


Figure 4.3: *Example of a premature beat from the MIT-BIH data set. The illustrated signal is a filtered and dimension reduced sub sequence from Record 209 Lead A.*

In general, we can see that our BST method performs comparably or better than previous work from the ECG literature on three out of the four classes and achieves the highest ACC of all the proposed methods on the MIT-BIH data set. The BST_{FSS} and the 1-NN-based approaches are outperformed by the full shapelet search BST algorithm on the MIT-BIH data set, except for the S class where the 1-NN algorithms perform better.

4.3.2 AAU-ECG Classification

The result of each AAU-ECG classification experiment is shown in Table 4.6, where the results are structured as follows:

1. The first row is the Marquette's scores.
2. The second row is the confidence threshold of our ensemble tuned towards matching the precision of Marquette.
3. The third row is the same as the second row, just with the confidence threshold tuned towards the recall.

See for example diagnosis Incomplete right bundle branch block (IRBBB (445)) where Marquette has 83.52% precision and 86.17% recall whereas our model achieves a precision of 83.56% and 12.53% recall when we tune the confidence threshold to match the precision of Marquette. When we tune the threshold to match the recall of Marquette our model has precision 36.70% and recall 87.27%. We mark the cases where we outperform Marquette as bold numbers.

Method	440		445		460		465		540		541		542		548	
	Pre	Rec	Pre	Rec	Pre	Rec	Pre	Rec	Pre	Rec	Pre	Rec	Pre	Rec	Pre	Rec
Marquette	95.96	97.90	83.52	86.17	90.46	93.89	49.79	70.32	1.44	60.49	33.56	54.34	40.99	30.13	11.58	25.67
Match Pre	95.43	79.69	83.56	12.53	90.44	87.54	28.57	3.39	2.95	**75.00**	33.61	**63.51**	41.00	**59.61**	11.52	**88.08**
Match Rec	83.09	96.03	36.70	87.27	87.13	93.95	7.07	66.10	**3.96**	62.50	**38.74**	54.64	**49.22**	32.81	**28.73**	25.50

Method	700		740		760		780		801		810		820		900	
	Pre	Rec	Pre	Rec	Pre	Rec	Pre	Rec	Pre	Rec	Pre	Rec	Pre	Rec	Pre	Rec
Marquette	77.58	87.19	79.59	86.91	62.21	85.33	81.71	88.41	68.53	82.02	79.14	89.08	68.10	85.86	69.94	50.02
Match Pre	66.67	1.51	81.82	1.57	62.50	5.62	81.36	15.47	50.00	11.11	62.50	2.42	68.75	22.92	70.00	35.20
Match Rec	23.97	87.15	22.93	86.61	9.94	88.76	47.56	88.59	3.48	83.33	18.02	90.34	18.98	85.42	65.69	50.28

Method	1140		1141		1143		1150		1160		1170		1180	
	Pre	Rec	Pre	Rec	Pre	Rec	Pre	Rec	Pre	Rec	Pre	Rec	Pre	Rec
Marquette	50.76	85.87	56.57	72.03	51.72	45.49	73.74	84.21	38.91	84.06	58.68	75.86	62.19	86.34
Match Pre	51.16	50.14	56.78	12.50	51.75	13.75	72.73	8.70	38.96	30.82	50.00	1.67	63.16	0.12
Match Rec	32.67	86.08	27.18	72.01	33.28	45.45	12.91	84.78	23.14	84.25	17.98	76.11	18.14	86.00

Table 4.6: *The results of comparing our proposed method with the Marquette predictions on the 23 diagnoses in % with the doctors' final diagnosis as the ground truth. The row "Marquette" is the 12SL predictions. "Match Pre" and "Match Rec" is our results when adjusting the threshold to match the 12SL precision or recall, respectively.*

Our model and Marquette both perform well on RBBB (440) and LBBB (460) diagnoses. These diagnoses are associated with increased cardiovascular risk [64] which is linked to high mortality. For the four statements 540, 541, 542 and 548 our method achieves better precision and recall than Marquette. Statement 541 is Left Ventricular Hypertrophy (LVH2 (541)) which indicates an increased mass of the heart's left ventricle and often occurs as a

reaction to cardiovascular disease or high blood pressure [65]. Statement 540, 542, 548 are different voltage criteria for LVH2 (541). The four statements combined represents 21.4% of the statements of the 23 diagnoses where statement 540 is the smallest representing only 0.14%. Our model's performance on statement 542 and 548 is considerably better than Marquette with an increase of recall of 29,5% and 62,4%, respectively. Furthermore, the precision of the two statements is 8.23% and 17.2% which higher than Marquette.

For Nonspecific ST Abnormality (NST (900)) the proposed method achieves a precision of 65.69% compared to 69.95% for Marquette. However, the Marquette achieves 14.8% higher recall compare to our model. For the remaining 16 diagnosis statements, the Marquette approach surpasses our model.

4.4 Qualitative Evaluation of AAU Results

The results we report in Section 4.3 show that classification based on shapelet transform can produce promising results. One of the key benefits of using shapelets, however, is that they offer a new way of interpreting the result of classification. Most classification methods act as a black box in regards to interpreting the correlation between an input sample and the classified label, whereas the shapelet-based approach is directly interpretive by studying the shapelets for the given class. To display this interpretability, we present an analysis of the shapelets found for a selected set of key diagnosis statements. The analysis is based on a comparison of the descriptors and rules the Marquette uses, as well as a reference to what the literature deems as the important discriminatory factors of a diagnosis.

We will analyze shapelets for RBBB (440), Anterior infarct (AMI (740)) and the four statements related to *Left Ventricular Hypertrophy*, namely LVH (540), LVH2 (541), Minimal voltage criteria for LVH, may be normal variant (QRSV (542)) and Moderate voltage criteria for LVH, may be normal variant (LVH3 (548)). We chose RBBB (440) as we achieve good discriminating power for the statement, the *Left Ventricular Hypertrophy* statements as we outperform Marquette on these classes and AMI (740) as our performance is low compared to Marquette. The figures shown throughout this section will be a selection of the leads which contains the phenomena that discriminate the particular diagnosis according to the Marquette descriptors and the medical literature. We present the shapelets by overlaying the best shapelets from a given class on top of the time series where the shapelet with most discriminatory power was found. Finally, we only show the best 20 out of a total of 50 shapelets found on the AAU-ECG data set. We refer to Appendix D for the full 12-lead configuration and their shapelets from the statements highlighted in this section.

4.4.1 Right Bundle Branch Block

The first shapelets we analyze are the ones for RBBB (440), shown on Figure 4.4, which displays lead *I* on the left and lead *V1* on the right. RBBB (440) is characterized by a *M*

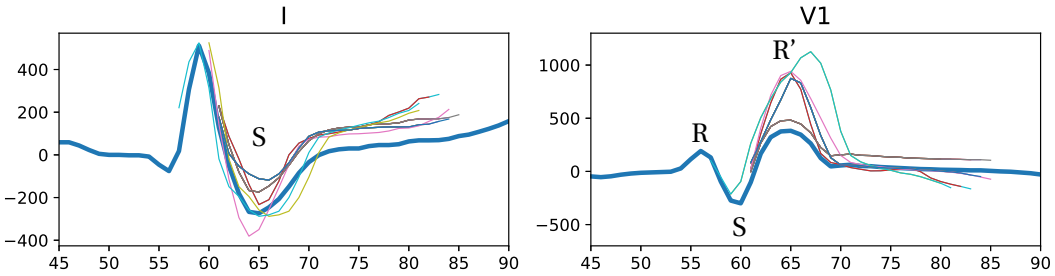


Figure 4.4: *RBBB (440) shapelets from lead* I *and* V1 *zoomed on the* QRS *complex.*

shaped *QRS* complex in leads *V1-3*, which is caused by a delayed *R* wave due to the right bundle block delaying the contraction of the right ventricle [66]. The *V1* figure shows the secondary *R* wave marked as *R'*. Marquette uses descriptors on lead *V1* as well as the lateral leads, namely *I, aVL, V5* and *V6*, to discriminate the RBBB (440). The descriptors for lead *V1* are a wide *R* wave and a positive wide *QRS* complex [5, p. 6-8]. The Figure 4.4 shows that the shapelets found on lead *V1* match the *V1* descriptors quite well, as they are all concentrated around the *R'* wave. The Marquette descriptor for the lateral leads is a wide and slurred *S* wave [5, p. 6-8] which also matches where the shapelets on lead *I* are concentrated. Refer to Appendix D to confirm the same tendency on the other leads and without zooming on the area of interest.

4.4.2 Left Ventricular Hypertrophy

The next shapelets we analyze are for the statements in the *Left Ventricular Hypertrophy* group. The Marquette system differentiates between ECG records that meet a voltage criteria of the *QRS* complex and records that exhibit other abnormalities that are associated with left ventricular hypertrophy.

The amplitude evaluation is measured on the *QRS* complex which, if matched correctly, leads to LVH (540), QRSV (542) or LVH3 (548) depending on the severity of the amplitude. The amplitude criteria evaluation uses the the *aVL, V1* and *V5* leads, where we shows the leads *aVL* and *V1* for each of the statements in Figures 4.5 to 4.7. After Marquette has analyzed the amplitude criteria it analyzes ECG for additional abnormalities on the lateral leads which, if matched correctly, results in a LVH2 (541) diagnosis.

The Amplitude Criteria

The Marquette system measures the amplitude of the *QRS* complex against age-dependent thresholds [5, p. 6-10]. A patient's age is a significant factor that affects the morphology and the amplitude of the *QRS* complex. In Section 5.2 we discuss what factors, like the patient

age, are considered when diagnosing heart arrhythmias. The Marquette system makes a choice between which of the three voltage criteria statements to select based on a point scoring system. The scoring is based on the number of lead *QRS* complexes exceed the threshold.

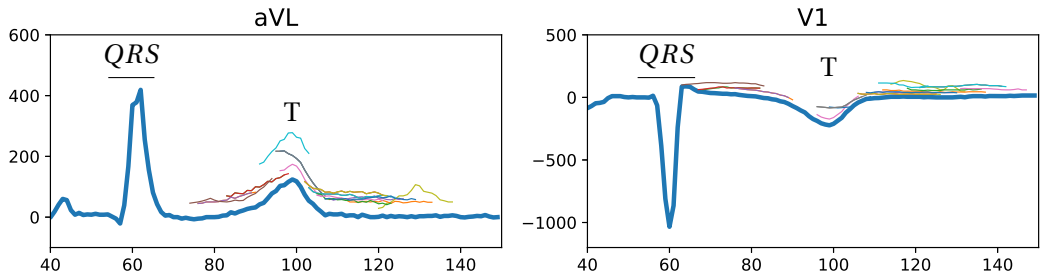


Figure 4.5: *QRSV (542) shapelets from lead* aVL *and* V1 *zoomed on the* QRS *complex and the* T *wave.*

If the score is between one and two, a low voltage score, the Marquette system adds the *minimal voltage criteria* statement with the acronym QRSV (542). The two leads, *aVL* and *V1*, left and right in Figure 4.5 shows that are found around the *T* wave and *ST* segments, instead of around the *QRS* complex that the Marquette descriptors use. The fact that we outperform the Marquette system on this diagnosis suggests that the morphology of the *T* wave and *ST* segments might have more discriminatory power.

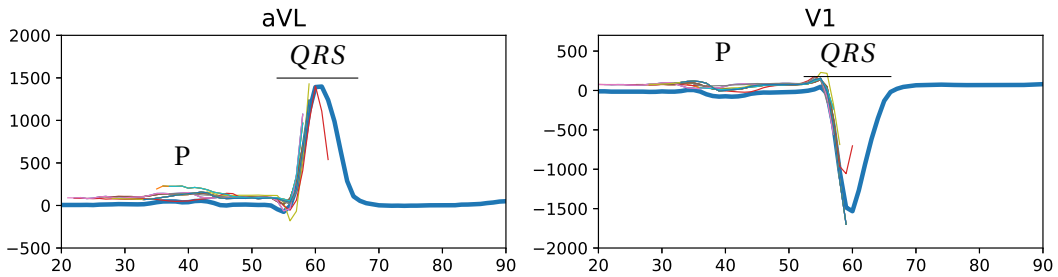


Figure 4.6: *LVH3 (548) shapelets from lead* aVL *and* V1 *zoomed on the* QRS *complex.*

Scores between three and four are used when the amplitude exceeds the threshold to a greater extent which makes Marquette assign the *moderate voltage criteria* diagnosis with the acronym LVH3 (548). In case the score is five or greater the system uses *Voltage criteria for left ventricular hypertrophy* with label LVH (540). Figure 4.6 shows the shapelets for LVH3 (548) are found in part on the *QRS* complex as well as around the *P* wave. This placement indicates that the part of the ECG leading up to the *QRS* complex carries discriminatory power regarding for LVH3 (548). The shapelets for LVH (540) in Figure 4.7 covers the entire sequence of the ECG from the *P* wave to the *T* wave. Again, it appears to be the case that parts of the ECG surrounding the *QRS* complex carries discriminatory

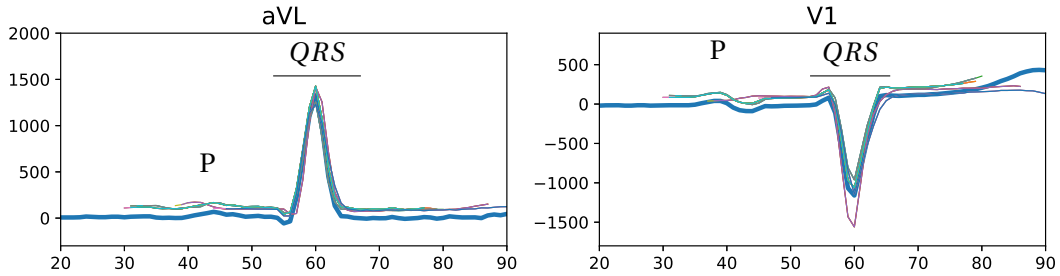


Figure 4.7: *LVH (540) shapelets from lead* aVL *and* V1 *zoomed on the* QRS *complex.*

power for left ventricular hypertrophy related diagnosis.

The Morphology Criteria

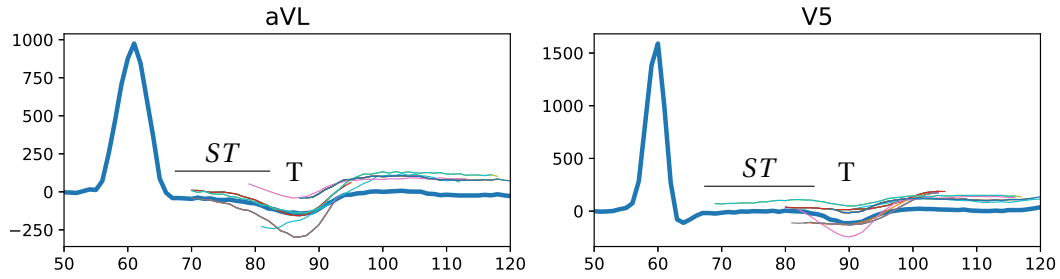


Figure 4.8: *LVH2 (541) shapelets from lead* aVL *and* V5 *zoomed on the* ST segment *and the* T *wave*

The Marquette descriptors that characterize diagnosis LVH2 (541) are a depression or a downwards slope of the *ST* segment in the lateral leads and a wide *R* wave in lead *V5* [5, p. 6-10,6-11]. On the left figure of Figure 4.8, it can be seen that the shapelets found on the lateral lead *aVL* are concentrated around the *T* wave while also covering most of the *ST* segment. The morphology of the shapelets that we find on the *ST* segment does slope downwards matching the Marquette descriptor. The shapelets we find on lead *V5* does not fit the Marquette descriptors as they are all found around the *T* wave. The concentration of the shapelets, for both leads, around the *T* wave and *ST* segment suggests that this carries a discriminatory power that is more meaningful than what the Marquette analysis program uses.

Edhouse et al. in [67] characterize LVH2 (541) by having the presence of a *left ventricular strain pattern*. The pattern is characterized by *ST* depression and *T* wave inversion in leads *I, aVL, V5-6, ST* segment elevation in *V1-3* which looking at Figure D.4 fits where some of the shapelets are found.

4.4.3 Anterior Infarction

The final shapelets we present are for AMI (740), a diagnosis for which the classification performance is low. Some of the descriptors Marquette uses for AMI (740) has to do with the amplitude of the *R* wave of the *V1-6* chest leads. The leftmost chest lead *V1* will under normal circumstances produces a small *R* wave and a large *S* wave, whereas for the rightmost chest lead *V6* this phenomenon is inverted such that the *R* wave is large and the *S* is small. This inversion of the waves is a continuum or a progression that happens from chest lead *V1* to *V6*. The *R* wave progression is illustrated in Figure 4.9 and shows that an abnormal progression can be see in a regression of the *R* wave amplitude from lead *V2* to *V3* as well as *V3* to *V4*.

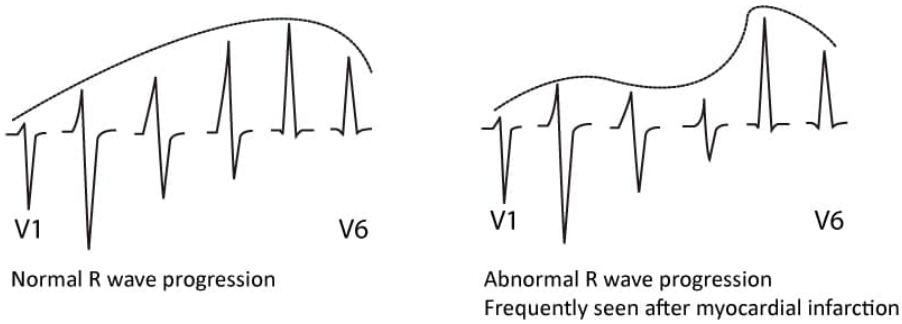


Figure 4.9: *The normal and abnormal progression of the* R *wave. Source: Ref. [68]*

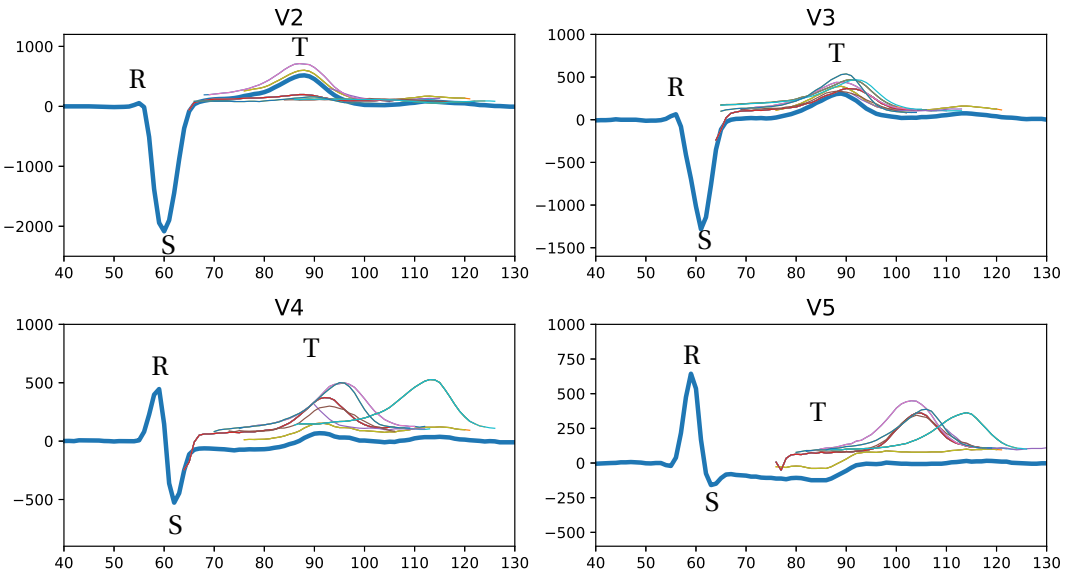
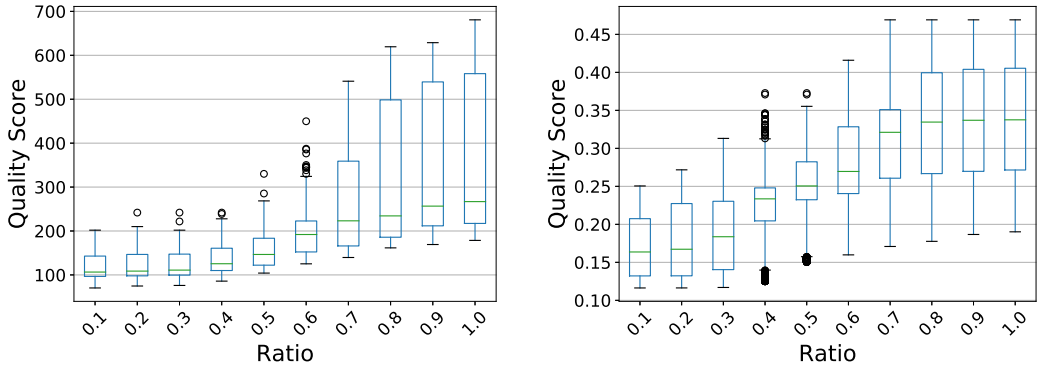


Figure 4.10: *AMI (740) shapelets from lead* V2-5 *zoomed on the* QRS *complex and the* T *wave.*

The Marquette specific descriptors for diagnosis AMI (740) are large *Q* waves by amplitude or duration in leads *V3-4*, a poor *R* wave progression in leads *V2-4* or an actual regression in the amplitude of *R* also in the *V2-4* leads [5, p. 6-12]. The Figure 4.10 illustrates the shapelets we find for this diagnosis on leads *V2-5*. Figure 4.10 shows that the shapelets are concentrated around the *T* wave. The position of the shapelets most likely causes the quite poor classification. The shapelets would be expected to be found around the *QRS* complex to capture the *R* wave progression changes as well as significant *Q* waves. Capturing an abnormal *R* wave progression is possible with the shapelet transform approach however it relies on classifiers that can handle multivariate correlation and not treat them as independent inputs. The HESCA ensemble used in this report, explained here Section 3.4, excludes classifiers that cannot handle dependent features like the *Naive-Bayes* [69, p. 265] and the *k-nearest-neighbor* [70, p. 5].

4.5 Quality of the Fast Shapelet Search

We conduct two tests to quantify what the trade-off is between the quality of the shapelets produced and the run time of FSS compared to the full shapelet search. First, we test what effect tweaking the *topK* parameter has on the shapelet quality. We then test the run time execution of the two algorithms in Section 4.5.1. Both experiments are conducted using the MIT-BIH data set.



(a) *F-statistic* **(b)** *Informaton gain*

Figure 4.11: *The box plots of the distribution of shapelets' quality scores found using different ratio of the total shapelet candidates for the parameter topK in the FSS algorithm.*

We use the F-statistic and information gain shapelet quality measures in the first experiment, where the results can be seen as a box plot of the distribution of found shapelets' quality measure in Figure 4.11a and Figure 4.11b, respectively. We use 200 training instances for each class in the MIT-BIH data set, resulting in a total of 800 training observations and 800 extracted shapelets. The circles on the box plots represent outliers, where a

value v is considered an outlier if $v > Q3 + 1.5 \cdot (Q3 - Q1)$ or $v < Q1 - 1.5 \cdot (Q3 - Q1)$. The $Q1$ and $Q3$ is the lower and upper bound of the box respectively, which encapsulate 50% if the data.

Ideally, the FSS algorithm would find the majority of the best shapelets at a low ratio like 0.1 of the total shapelet candidates where a higher ratio would only improve the quality by a small amount. However, as seen in Figure 4.11, this does not happen, as the distribution keeps getting markedly better as a function of the ratio. Hence, the FSS algorithm improves the run time of the shapelet extraction at the expense of finding lower quality shapelets. In both Figure 4.11a and Figure 4.11b it seems that most of the best shapelet are found around a ratio of 0.8 after which the quality plateaus.

4.5.1 Runtime Analysis of Fast Shapelet Search

In Section 3.6 the complexity of the full shapelet search, and the FSS was found to be $O(n^2m^4)$ and $O(n^2km^2)$ respectively, where k is the number of shapelet candidates used. We now perform run time experiments to see the actual time difference between using the full shapelet search and the FSS on different amount of training data. These experiments were performed in a Java environment on a shared server with following specifications:

- **OS:** x86_64 Ubuntu 16.04
- **CPU:** 8x Intel(R) Xeon(R) CPU E5420 @ 2.50GHz with 12M L2 Cache.
- **Memory:** 8x 4GB DDR2, HYNIX HYMP151F72CP4N3-Y5, clock = 667Mhz.

To our knowledge, the server was not used by others while we ran the test, but because it is a shared server, we cannot guarantee this. As such, the result of these experiments should be seen as a guide to the run time of the two algorithms. The parameters used for the experiments are the F-statistic quality measure, a shapelet length of range 10 to 50 and a window size $w = 0.25$. The results of the tests are plotting in Figure 4.12 and the discrete values can be seen in Table 4.7. As the run time of the full shapelet search exceeded a day at 1200+ trainings observation, we stopped the experiment.

The slope of the plots in Figure 4.12 increases when more training data is used. This increase occurs because more shapelet candidates need to be evaluated and each shapelet candidate is evaluated against more time series. As one might expect, the FSS algorithm seems to scale quadratically as a function of the number of observations which confirms the complexity of FSS $O(n^2km^2)$, where n is the number of observations, k is the number of heuristically preselected shapelet candidates and m is the dimensionality of the observations.

We would expect $FSS_{0.5}$ to be about twice as fast as the full shapelet search as the number of shapelet candidates is halved. This is, however, not the case as Figure 4.12 shows that it is, in fact, more than three times as fast. One possible reason for this could be that FSS evaluate high-quality shapelets early, making the pruning of candidates more efficient.

Search	# of Train Observations					
	200	400	800	1000	1200	1400
FULL	2171	8436	18770	33467	N/A	N/A
FSS$_{0.1}$	216	673	1341	2325	3608	5239
FSS$_{0.2}$	344	1103	2198	3738	5767	8634
FSS$_{0.3}$	482	1551	3081	5305	8322	12032
FSS$_{0.4}$	619	1999	4174	7091	11119	16406
FSS$_{0.5}$	746	2501	5136	9131	14275	20806

Table 4.7: *The seconds it takes to extract shapelets for the full shapelet search and the FSS with different $topK$ and different amount of training observations. The FSS$_{0.1}$ means FSS using 10% of the shapelet candidates.*

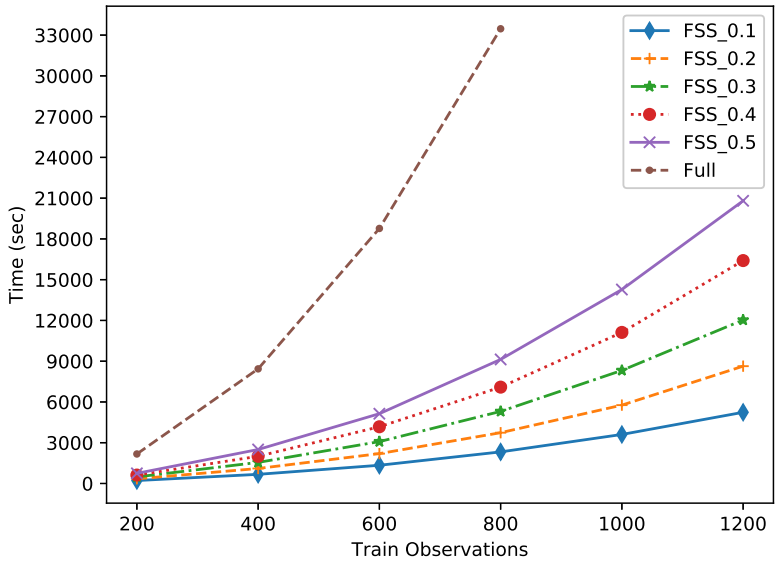


Figure 4.12: *A plot of the time it takes FSS with different ratios and the full shapelet search with varying amount of train data.*

4.5.2 Summary

We have in this chapter presented the experimental settings, the results of the experiments performed on the MIT-BIH and AAU-ECG and evaluated the results quantitatively and qualitatively. The BST approach performs as well or better than state-of-the-art classification algorithms following the inter-patient scheme except for the supraventricular ectopic heartbeats which discriminating characteristic is not reflected in the morphology of a single heartbeat. The qualitative evaluation of the AAU-ECG data set demonstrates

that the shapelets we find for RBBB (440) matches the descriptors used by Marquette. The shapelets we find the for left ventricular hypertrophy heartbeats are interesting as we outperform the Marquette classification and they also differ from the descriptors it uses. Finally, we conduct a quality measure and run time analysis of the fast shapelet search algorithm, which shows that a high ratio of the total shapelets candidates is required to obtain good quality shapelets.

5 Discussion

We will in this chapter discuss some of the choices and findings made throughout the project. We discuss the impact of using a single heartbeat for the transformation and classification, diagnosis influential factors such as age and heart rate and the potential bias in the AAU-ECG data set.

5.1 Shapelet Transform as Features

We have described how the ST algorithms transform ECG records into feature vectors of distances and the results in Section 4.3.1 demonstrates that it outperforms or is comparable with previous work within the classification of ECGs. The features extracted from the BST are well suited to diagnose heart arrhythmias reflected in the morphology of a single heartbeat. However, some diagnoses are only reflected in the morphology across multiple heartbeats like for instance the S class from the MIT-BIH data set. In these cases, the classifier does not have the needed information to classify these diagnoses accurately.

A standard way of accommodating for this, found in the literature, is to add extra features in the form of inter-beat interval features [21, 31, 45, 46], that extends across multiple heartbeats. Such features could be the RR-intervals from the current heartbeat to the previous and the next heartbeat.

We z-normalize shapelets to accommodate for differences that exist between ECGs and between patients. The normalization preserves the shape of the shapelet but removes information about the absolute amplitude. This fact might prove troublesome as some arrhythmias might require this information to be diagnosed accurately. For instance, the Marquette system classifies left ventricular hypertrophy arrhythmias using the real amplitude of the ECG as well as age-dependent criteria [5, p. 6-9]. However, the shapelet based transformation and classification seem capable of identifying amplitude related diagnosis statements without the absolute amplitude value. If further experiments show similar results the age-dependent feature that Marquette relies on could be replaced with a simplified model.

5.2 Diagnosis Influential Factors

The goal of this report is to explore how well a shape-based classification of ECGs can differentiate between classes of heart arrhythmias. Since we do shape-based classification, we intentionally ignore non-shape based factors that are normally important factors to consider when diagnosing heart arrhythmias. These factors are for example patient age and gender, taken medicine as well as blood pressure, body mass index (BMI), diabetes, and so on. [14]. These factors can influence the classification in two ways; it may change the prevalence of a diagnosis or diagnoses in general, or it may change the morphology of the ECG. According to associate professor Claus Graff [14] the most influential factor is the age of the patient, which is why we base the following discussion on what effect the age of a patient has on the ECG morphology as well as the prevalence diagnoses.

Rupali Khane et al. [71] explore the effects advancing age has on the prevalence heart arrhythmias. The authors study a population of patients between ages 54-74. To examine the causal connection between heart arrhythmia's and advancing ages, the authors separate the patients into groups based on ages after which they analyze the prevalence of heart arrhythmias like LVH (540) and RBBB (440). Rupali Kane et al. found that the prevalence of ECG arrhythmias is highly correlated with advancing age. The authors came to this conclusion by doing a hypothesis test to measure the statistic significance of the age which resulted in a p-value of 0.001. The inclusion of the factors such as the age would likely improve the classification performance. However, if we include such factors as part of the input to the classification model, we would no longer do purely shape-based classification of the ECGs. Still, if the goal is to achieve the highest classification performance, including such features could make a difference.

The age factor might also influence the morphology of the ECG in addition to increasing the prevalence of heart arrhythmias. Bachman et al. explore this connection in [72]. The authors examine the connection by studying the morphology changes that appear by comparing ECGs made ten years apart for the same patients. The authors examine the amplitude changes of the R,S,T waves, the duration changes of the PR segment, QRS complex duration and the QT duration as well as the frontal plane axis. They measure the statistical differences by performing a paired t-test, where they find that all variables have changed significantly with a statistical significance of $p < 0.001$. We can use this knowledge to incorporate age as part of our classification without changing the focus from a shape based classification by separating the shapelets into groups based on age segments. For instance, we could assign the shapelets extracted from a ECG of a patient aged 25 to the 20-30 aged shapelet segment group. Then we would train a classifier for each shapelet age segment and only use the given classifier that matches a new patient's age. We could use the same method of segmenting shapelets for each of the factors mentioned like gender, BMI, but this requires knowledge of whether or not these factors actually influence the morphology of the ECG and not just the prevalence of heart arrhythmias.

5.3 Heart Rate Influence on the ECG

The patient heart rate affects the morphology of the ECG. This fact presents a problem in general regarding classification of ECGs and in our case, as the extracted shapelets might not be representative for all patients due to the morphological changes caused by the heart rate.

To see the effect of the heart rate on the ECG, one could study segments that are used to diagnose heart arrhythmias like the duration of the *QT* interval or the *QRS* complex. We can do this because the AAU-ECG data set contains information about the patient heart rate as well as the duration of these segments. In Figure 5.1 we show the median interval durations where the x-axis represents segments of heart rates and the y-axis is the duration of either the *QT*, *PR* or the *QRS* intervals. The Figure 5.1 shows that only the *QT* duration changes as a function of the heart rate.

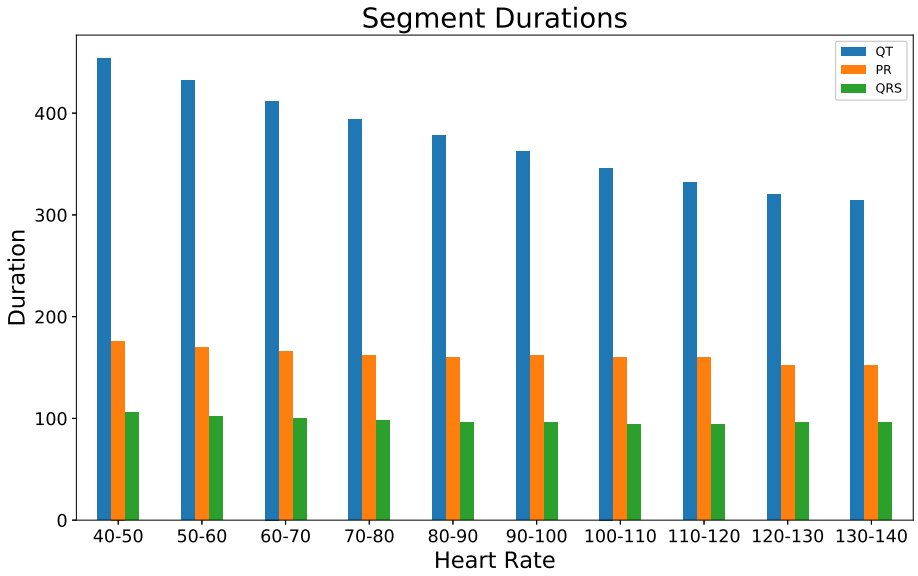


Figure 5.1: *The median QT, PR and QRS durations in intervals of heart rates from the AAU-ECG data set.*

The changes in *QT* segment duration should mainly affect the diagnoses in the RA group as they are all detected by studying phenomena around the *QT* segment [5, p. 6-22;6-25]. To handle this problem we could ensure that the shapelets extracted for diagnoses that affect the *QT* are divided into portions or "bins" based on the heart rate. This divisions or "binning" would enable us to use shapelets that are extracted from patient ECGs that match the heart rate of the patient being diagnosed.

5.4 Biased AAU-ECG Data Set

Our experiments in Section 4.3.2, regarding our approach against the Marquette system's predictions, is biased in favor of the Marquette system. The bias occurs as a result of the doctor's knowledge of the Marquette system's prediction when deciding the labels for an ECG, which might change their mind on which diagnosis statement the ECG should have compared to not knowing the predictions. The doctors also need to do an active action to change the predictions of the Marquette system, which might result in fewer changes from the doctor's side.

For an unbiased experiment, we would need a new data set of doctor labeled ECGs, where the doctor did not know the Marquette's predictions when assigning labels to the ECGs. Then, with access to the Marquette system, we could test if our approach or the Marquette performed better on this new data set.

6 Conclusion

We presented our approach to performing shapelet transformation and classification on a large multi-labeled ECG data set in Chapter 3, based on the idea that patients with same diagnosis statements would have similarities in their ECG waveforms. Experiments were performed on the MIT-BIH arrhythmia data set to compare our approach to previous work within ECG classification as well as the AAU-ECG data set consisting of 413,151 ECG records distributed on 211,391 patients after filtering, each attached with a subset of 87 diagnosis statements.

In the introduction, we derived a set of research questions based on the assumption that patients with the same diagnosis statements have shape similarities in their waveforms. We now reflect on these questions using the insight and results gained from our experiments.

- **Can the shapelet transformation classification approach outperform previous work within heartbeat classification using electrocardiograms?** The results of the experiments on the MIT-BIH data set in Section 4.3.1 demonstrate that our approach performs as well or better than previous work within ECG classification following the inter-patient scheme for the normal beats, ventricular ectopic beats, and fusion beats. Our approach achieves an improvement compared to the previously reported results as well as state-of-the-art multivariate time series classification algorithms with an overall accuracy of 94.3%, recall of 99.4% and 86.6% for the N and V class as well as the highest precision and lowest FPR for the F class of 50.7% and 0.16%, respectively. The main limitation of the proposed method on the MIT-BIH data set is the low recall, 1.31% for identifying supraventricular ectopic beats.
- **Can the shapelet transformation classification more accurately predict the doctor's diagnosis compared to the knowledge-based Marquette 12SL ECG analysis program:** The results in Section 4.3.2 shows that the shapelet-based approach, on average, cannot surpass the knowledge-based Marquette 12SL analysis program with regards to the performance metrics, precision, and recall. As discussed in Section 5.4 the performance of Marquette on the AAU-ECG data set can be biased towards Marquette. The performance of our approach on statements, 540, 541, 542, and 548 related to left ventricular hypertrophy outperforms Marquette. Especially statement 542 and 548 with a recall increase of 29.5% and 62.4% as well as a precision increase of 8.23% and 17.2%. For the diagnosis statements right and left bundle branch block, 440 and 460, the proposed method have a good discrimination power. As the shapelets are learned from labeled training data medical practitioners can use our

findings for left ventricular hypertrophy as well as right and left bundle branch block to interpret the shapelets that discriminate the six diagnosis statements.

- **What are the trade-offs between the run time and shapelet quality when using shapelet heuristic approximating techniques:** The preliminary experiments from Section 4.5 shows that using the FSS algorithm to extract shapelets is more than one order of magnitude faster than the full shapelet search when using 10% of the shapelet candidates and three time as fast when using 50% of the candidates. The quality of the shapelets found using FSS seems to be proportional to the ratio of shapelet candidates used, where 10% produced poor quality shapelets and 80% produced comparable results to full search. This result indicates that the performance of FSS is very dependent on the number of shapelet candidates used. The preliminary experiments of the window optimization show that our best performing window size is 5% and it achieves a higher accuracy of 94.3% compared to 91.6% when using no window. We achieve this performance increase while improving the run time from 12,656 to 5,425 seconds.

We conclude that the shapelet transform approach to classifying ECG data sets shows promising results as it performs comparably or better then previous work within ECG classification for heart arrhythmias reflected in the morphology of a single heartbeat. We also conclude that the learning-based approach can identify a subset of heartbeat types better than the knowledge-based approach leveraged by the Marquette 12SL ECG analysis program.

6.1 Future Work

In this section, we present improvements and new ideas for future work based on the insight we have gained regarding the shapelet-based transformation and classification.

6.1.1 Domain-Specific Knowledge

The focus of this research is to explore what information emerges from analyzing the ECG waveform. To explore this, we use a shapelet transformation and classification approach to classify ECG heart arrhythmia data sets.

The results of our analysis show promising; however, there are still some diagnosis types that prove troublesome for our approach like the supraventricular ectopic heartbeats in the MIT-BIH data set. The S class requires temporal features of previous heartbeats to classify the current heartbeat reliably.

In Section 5.2, we list some of the essential factors that medical professionals consider when they diagnose a patient with heart arrhythmia. We show that especially the age factor is exceedingly essential both regarding affecting changes in ECG morphology and the

prevalence of heart arrhythmias as the age increases. We could mix these factors with the current shapelet transform features merely by appending them to the feature vector, or we could keep them separate from the shape features by training individual HESCA classifiers. The second option would allow us to run these classifiers in parallel and analyze what diagnoses the domain-specific factors accurately classify and maybe combine them into a meta-classifier that leverage the strengths of both approaches.

6.1.2 Lead-based Ensemble

In our current approach, we extract shapelets from all the leads of the ECG, and the leads are all equally important in the classification stage. This fact has the advantage, that information across the leads are available in the shapelet extraction process. However, when looking into the ECG literature, the doctors and researcher use only a subset of the leads for most of the diagnosis.

Patri et al. in [73] used an alternative way of handling multivariate data, where the shapelets were extracted from each lead individually, and a classifier was trained on each lead. An ensemble scheme was used to combine the classifiers with a weighting scheme learned on the train data to decide the importance of the different channels for each class. By applying this method, only shapelets from the lead(s) which are best at discerning each class would be used. This scheme is arguably more interpretative than our current method, as each shapelet are found on a single lead, and the weighting scheme itself could be used to check the importance of a lead for a given diagnosis.

However, with this method, the information across leads in the shapelet extraction is lost. This might lower the classification performance on diagnoses where leads are dependent on each other.

Bibliography

[1] L. Ye and E. Keogh, "Time series shapelets: A new primitive for data mining," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2009, pp. 947–956.

[2] J. Lines, L. M. Davis, J. Hills, and A. Bagnall, "A shapelet transform for time series classification," in *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '12, Beijing, China: ACM, 2012, pp. 289–297, ISBN: 978-1-4503-1462-6. DOI: 10.1145/2339530.2339579. [Online]. Available: http://doi.acm.org/10.1145/2339530.2339579.

[3] T. Rakthanmanon and E. Keogh, "Fast shapelets: A scalable algorithm for discovering time series shapelets," in *Proceedings of the 2013 SIAM International Conference on Data Mining*, pp. 668–676. DOI: 10.1137/1.9781611972832.74. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9781611972832.74. [Online]. Available: https://epubs.siam.org/doi/abs/10.1137/1.9781611972832.74.

[4] M. Shanthi, P. Pekka, and N. B, *Global atlas on cardiovascular disease prevention and control*. World Health Organization in collaboration with the World Heart Federation and the World Stroke Organization, 2011, ISBN: 9789241564373. [Online]. Available: http://apps.who.int/iris/handle/10665/44701.

[5] *Marquette™ 12SL™ ECG Analysis Program Physician's Guide, Revision E*, GE Healthcare, 2008.

[6] R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm, and N. Elhadad, "Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2015, pp. 1721–1730. [Online]. Available: http://doi.acm.org/10.1145/2783258.2788613.

[7] E. J. da S. Luz, W. R. Schwartz, G. Cámara-Chávez, and D. Menotti, "ECG-based heartbeat classification for arrhythmia detection: A survey," *Computer Methods and Programs in Biomedicine*, vol. 127, no. Supplement C, pp. 144–164, 2016, ISSN: 0169-2607. DOI: https://doi.org/10.1016/j.cmpb.2015.12.008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0169260715003314.

[8] A. Bagnall, J. Lines, A. Bostrom, J. Large, and E. Keogh, "The great time series classification bake off: A review and experimental evaluation of recent algorithmic advances," *Data Mining and Knowledge Discovery*, vol. Online First, 2016.

[9] D. Avies, A. An, and A. Scott, *Starting to Read ECGs*, 2015.

[10] (Jun. 5, 2018). Electrocardiography, [Online]. Available: https://en.wikipedia.org/wiki/Electrocardiography#/media/File:SinusRhythmLabels.svg (visited on 06/05/2018).

[11] V. Nassehi and L. Shojai, "Modelling of blood flow through heart valves," *Biomedical Engineering and Technology*, vol. 6, 2011.

[12] M. S. Thaler, *The only EKG book you'll ever need*. Lippincott Williams & Wilkins, 2010.

[13] P. Kligfield, L. S. Gettes, J. J. Bailey, R. Childers, B. J. Deal, E. W. Hancock, G. van Herpen, J. A. Kors, P. Macfarlane, D. M. Mirvis, *et al.*, "Recommendations for the Standardization and Interpretation of the Electrocardiogram," *Circulation*, vol. 115, no. 10, pp. 1306–1324, 2007.

[14] C. Graff, *Associate Professor Department of Health Science and Technology, Aalborg University*, personal communication.

[15] W. V. Anthony Bagnall Jason Lines and E. Keogh. (Mar. 28, 2018). The uea & ucr time series classification repository, [Online]. Available: http://www.timeseriesclassification.com (visited on 03/28/2018).

[16] A. Bostrom and A. Bagnall, "Binary shapelet transform for multiclass time series classification," Sep. 2015, pp. 257–269, ISBN: 978-3-319-22728-3.

[17] J. Large, J. Lines, and A. Bagnall, "The heterogeneous ensembles of standard classification algorithms (HESCA): the whole is greater than the sum of its parts," *CoRR*, vol. abs/1710.09220, 2017. arXiv: 1710.09220. [Online]. Available: http://arxiv.org/abs/1710.09220.

[18] A. Bagnall, J. Lines, J. Hills, and A. Bostrom, "Time-series classification with cote: The collective of transformation-based ensembles," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 9, pp. 2522–2535, Sep. 2015, ISSN: 1041-4347. DOI: 10.1109/TKDE.2015.2416723.

[19] A. Bostrom and A. Bagnall, "A shapelet transform for multivariate time series classification," *CoRR*, vol. abs/1712.06428, 2017.

[20] M. Shokoohi-Yekta, B. Hu, H. Jin, J. Wang, and E. Keogh, "Generalizing dtw to the multi-dimensional case requires an adaptive approach," *Data Mining and Knowledge Discovery*, vol. 31, no. 1, pp. 1–31, Jan. 2017, ISSN: 1573-756X. DOI: 10.1007/s10618-016-0455-0. [Online]. Available: https://doi.org/10.1007/s10618-016-0455-0.

[21] P. De Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ecg morphology and heartbeat interval features," *IEEE transactions on biomedical engineering*, vol. 51, no. 7, pp. 1196–1206, 2004.

[22] M. Heckmann, F. Berthommier, and K. Kroschel, "Noise adaptive stream weighting in audio-visual speech recognition," *EURASIP Journal on Advances in Signal Processing*, vol. 2002, no. 11, p. 720 764, Nov. 2002, ISSN: 1687-6180. DOI: 10.1155/S1110865702206150. [Online]. Available: https://doi.org/10.1155/S1110865702206150.

[23] G. Madjarov, D. Kocev, D. Gjorgjevikj, and S. Džeroski, "An extensive experimental comparison of methods for multi-label learning," *Pattern recognition*, vol. 45, no. 9, pp. 3084–3104, 2012.

[24] D. Kocev, C. Vens, J. Struyf, and S. Džeroski, "Ensembles of multi-objective decision trees," in *European Conference on Machine Learning*, Springer, 2007, pp. 624–631.

[25] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Effective and efficient multilabel classification in domains with large number of labels," in *Proc. ECML/PKDD 2008 Workshop on Mining Multidimensional Data (MMD'08)*, sn, vol. 21, 2008, pp. 53–59.

[26] T. G. Dietterich, "An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization," *Machine Learning*, vol. 40, no. 2, pp. 139–157, Aug. 2000, ISSN: 1573-0565. DOI: 10.1023/A:1007607513941. [Online]. Available: https://doi.org/10.1023/A:1007607513941.

[27] J. Lines, S. Taylor, and A. Bagnall, "Hive-cote: The hierarchical vote collective of transformation-based ensembles for time series classification," in *2016 IEEE 16th International Conference on Data Mining (ICDM)*, Dec. 2016, pp. 1041–1046. DOI: 10.1109/ICDM.2016.0133.

[28] A. Mueen, E. Keogh, and N. Young, "Logical-shapelets: An expressive primitive for time series classification," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2011, pp. 1154–1162.

[29] M. Jensen, C. Risager, and K. Madsen, *Shape-based cluster analysis of electrocardiograms*, Jan. 2018.

[30] S. H. Jambukia, V. K. Dabhi, and H. B. Prajapati, "Classification of ECG signals using machine learning techniques: A survey," pp. 714–721, Mar. 2015.

[31] Z. Zhang, J. Dong, X. Luo, K.-S. Choi, and X. Wu, "Heartbeat classification using disease-specific feature selection," *Comput. Biol. Med.*, vol. 46, pp. 79–89, Mar. 2014, ISSN: 0010-4825. DOI: 10.1016/j.compbiomed.2013.11.019. [Online]. Available: http://dx.doi.org/10.1016/j.compbiomed.2013.11.019.

[32] B. N. Singh and A. K. Tiwari, "Optimal selection of wavelet basis function applied to ECG signal denoising," *Digital Signal Processing*, vol. 16, no. 3, pp. 275–287, 2006, ISSN: 1051-2004. DOI: https://doi.org/10.1016/j.dsp.2005.12.003. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1051200405001703.

[33] R. Sameni, M. B. Shamsollahi, C. Jutten, and G. D. Clifford, "A Nonlinear Bayesian Filtering Framework for ECG Denoising," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 12, pp. 2172–2185, Dec. 2007, ISSN: 0018-9294. DOI: 10.1109/TBME.2007.897817.

[34] P. Laguna, R. Jané, and P. Caminal, "Automatic Detection of Wave Boundaries in Multilead ECG Signals: Validation with the CSE Database," *Computers and Biomedical Research*, vol. 27, no. 1, pp. 45–60, 1994, ISSN: 0010-4809. DOI: https://doi.org/10.1006/cbmr.1994.1006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0010480984710068.

[35] J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985, ISSN: 0018-9294. DOI: 10.1109/TBME.1985.325532.

[36] Y. H. Hu, W. J. Tompkins, J. L. Urrusti, and V. X. Afonso, "Applications of artificial neural networks for ECG signal detection and classification.," English, vol. 26 Suppl, pp. 66–73, 1993, ISSN: 0022-0736. [Online]. Available: http://sfx.aub.aau.

dk/sfxaub?sid=google&auinit=YH&aulast=Hu&atitle=Applications+of+artificial+neural+networks+for+ECG+signal+detection+and+classification.&id=pmid:8189150.

[37] J. P. Martinez, R. Almeida, S. Olmos, A. P. Rocha, and P. Laguna, "A wavelet-based ECG delineator: evaluation on standard databases," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 4, pp. 570–581, Apr. 2004, ISSN: 0018-9294. DOI: 10.1109/TBME.2003.821031.

[38] V. X. Afonso, W. J. Tompkins, T. Q. Nguyen, and S. Luo, "ECG beat detection using filter banks," *IEEE Transactions on Biomedical Engineering*, vol. 46, no. 2, pp. 192–202, Feb. 1999, ISSN: 0018-9294. DOI: 10.1109/10.740882.

[39] S. Karpagachelvi, M. Arthanari, and M. Sivakumar, "Ecg feature extraction techniques-a survey approach," *arXiv preprint arXiv:1005.0957*, 2010.

[40] *MIT-BIH Arrhythmia Database*, https://www.physionet.org/physiobank/database/mitdb/, Accessed: 14-10-2017.

[41] ANSI/AAMI, *Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms, ANSI/AAMI/ISO EC57*, 2008.

[42] B. M. Asl, S. K. Setarehdan, and M. Mohebbi, "Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal," *Artificial Intelligence in Medicine*, vol. 44, no. 1, pp. 51–64, 2008, ISSN: 0933-3657. DOI: https://doi.org/10.1016/j.artmed.2008.04.007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0933365708000559.

[43] C. Ye, M. T. Coimbra, and B. V. K. V. Kumar, "Arrhythmia detection and classification using morphological and dynamic features of ecg signals," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, Aug. 2010, pp. 1918–1921. DOI: 10.1109/IEMBS.2010.5627645.

[44] S.-N. Yu and K.-T. Chou, "Integration of independent component analysis and neural networks for ecg beat classification," *Expert Systems with Applications*, vol. 34, no. 4, pp. 2841–2846, 2008, ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2007.05.006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417407001789.

[45] M. Llamedo and J. P. Martinez, "Heartbeat classification using feature selection driven by database generalization criteria," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 616–625, Mar. 2011, ISSN: 0018-9294. DOI: 10.1109/TBME.2010.2068048.

[46] S. Chen, W. Hua, Z. Li, J. Li, and X. Gao, "Heartbeat classification using projected and dynamic features of ecg signal," *Biomed. Signal Proc. and Control*, vol. 31, pp. 165–173, 2017.

[47] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra, "Dimensionality reduction for fast similarity search in large time series databases," *Knowledge and information Systems*, vol. 3, no. 3, pp. 263–286, 2001.

[48] G. B. Moody and R. G. Mark, "The impact of the mit-bih arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.

[49] A. for the Advancement of Medical Instrumentation and A. N. S. Institute, *Testing and Reporting Performance Results of Cardiac Rhythm and ST-segment Measurement Algorithms*, ser. ANSI/AAMI. The Association, 1999, ISBN: 9781570201165. [Online]. Available: https://books.google.dk/books?id=gzPdtgAACAAJ.

[50] M. Llamedo and J. P. Martínez, "Heartbeat classification using feature selection driven by database generalization criteria," vol. 58, pp. 616–25, Mar. 2011.

[51] T. Mar, S. Zaunseder, J. P. Martínez, M. Llamedo, and R. Poll, "Optimization of ecg classification by means of feature selection," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 8, pp. 2168–2177, Aug. 2011, ISSN: 0018-9294. DOI: 10.1109/TBME.2011.2113395.

[52] M. Blanco-Velasco, B. Weng, and K. E. Barner, "Ecg signal denoising and baseline wander correction based on the empirical mode decomposition," *Computers in Biology and Medicine*, vol. 38, no. 1, pp. 1–13, 2008, ISSN: 0010-4825. DOI: https://doi.org/10.1016/j.compbiomed.2007.06.003. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0010482507001114.

[53] G. Lenis, N. Pilia, A. Loewe, W. H. W. Schulze, and O. Dössel, "Comparison of baseline wander removal techniques considering the preservation of st changes in the ischemic ecg: A simulation study," in *Comp. Math. Methods in Medicine*, 2017.

[54] J. Hills, J. Lines, E. Baranauskas, J. Mapp, and A. Bagnall, "Classification of time series by shapelet transformation," *Data Mining and Knowledge Discovery*, vol. 28, no. 4, pp. 851–881, Jul. 2014, ISSN: 1573-756X. DOI: 10.1007/s10618-013-0322-1. [Online]. Available: https://doi.org/10.1007/s10618-013-0322-1.

[55] J. Lin, E. Keogh, L. Wei, and S. Lonardi, "Experiencing sax: A novel symbolic representation of time series," *Data Mining and knowledge discovery*, vol. 15, no. 2, pp. 107–144, 2007.

[56] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *SIGKDD Explorations*, vol. 11, no. 1, pp. 10–18, 2009.

[57] J. Platt, "Sequential minimal optimization: A fast algorithm for training support vector machines," Tech. Rep., Apr. 1998, p. 21. [Online]. Available: https://www.microsoft.com/en-us/research/publication/sequential-minimal-optimization-a-fast-algorithm-for-training-support-vector-machines/.

[58] E. J. D. S. Luz, T. M. Nunes, V. H. C. De Albuquerque, J. P. Papa, and D. Menotti, "Ecg arrhythmia classification based on optimum-path forest," *Expert Syst. Appl.*, vol. 40, no. 9, pp. 3561–3573, Jul. 2013, ISSN: 0957-4174. DOI: 10.1016/j.eswa.2012.12.063. [Online]. Available: http://dx.doi.org/10.1016/j.eswa.2012.12.063.

[59] J. J. Rodriguez, L. I. Kuncheva, and C. J. Alonso, "Rotation forest: A new classifier ensemble method," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 10, pp. 1619–1630, 2006.

[60] Q. Du and J. E. Fowler, "Low-complexity principal component analysis for hyperspectral image compression," *The International Journal of High Performance Computing Applications*, vol. 22, no. 4, pp. 438–448, 2008.

[61] J. R. Quinlan, *C4. 5: programs for machine learning*. Elsevier, 2014.

[62] J. Su and H. Zhang, "A fast decision tree learning algorithm," in *AAAI*, vol. 6, 2006, pp. 500–505.

[63] A. Mueen and E. Keogh, "Extracting optimal performance from dynamic time warping," in *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16, San Francisco, California, USA: ACM, 2016, pp. 2129–2130, ISBN: 978-1-4503-4232-2. DOI: 10.1145/2939672.2945383. [Online]. Available: http://doi.acm.org/10.1145/2939672.2945383.

[64] B. E. Bussink, A. G. Holst, L. Jespersen, J. W. Deckers, G. B. Jensen, and E. Prescott, "Right bundle branch block: Prevalence, risk factors, and outcome in the general population: Results from the copenhagen city heart study," *European Heart Journal*, vol. 34, no. 2, pp. 138–146, 2013. DOI: 10.1093/eurheartj/ehs291. eprint: /oup/backfile/content_public/journal/eurheartj/34/2/10.1093_eurheartj_ehs291/1/ehs291.pdf. [Online]. Available: http://dx.doi.org/10.1093/eurheartj/ehs291.

[65] A. H. Gradman and F. Alfayoumi, "From left ventricular hypertrophy to congestive heart failure: Management of hypertensive heart disease," *Progress in Cardiovascular Diseases*, vol. 48, no. 5, pp. 326–341, 2006, Hypertension 2006 Update, ISSN: 0033-0620. DOI: https://doi.org/10.1016/j.pcad.2006.02.001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0033062006000119.

[66] D. Vivas, M. J. Pérez-Vizcayno, R. Hernández-Antolín, A. Fernández-Ortiz, C. Bañuelos, J. Escaned, P. Jiménez-Quevedo, J. A. D. Agustín, I. Núñez-Gil, J. J. González-Ferrer, C. Macaya, and F. Alfonso, "Prognostic implications of bundle branch block in patients undergoing primary coronary angioplasty in the stent era," *The American Journal of Cardiology*, vol. 105, no. 9, pp. 1276–1283, 2010, ISSN: 0002-9149. DOI: https://doi.org/10.1016/j.amjcard.2009.12.044. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0002914909029191.

[67] J. Edhouse, R. K. Thakur, and J. M. Khalil, "Conditions affecting the left side of the heart," *BMJ*, vol. 324, no. 7348, pp. 1264–1267, 2002, ISSN: 0959-8138. DOI: 10.1136/bmj.324.7348.1264. eprint: https://www.bmj.com/content/324/7348/1264.1.full.pdf. [Online]. Available: https://www.bmj.com/content/324/7348/1264.1.

[68] *Pathological r-wave progression*. [Online]. Available: https://ecgwaves.com/wp-content/uploads/2016/10/x-R-progression.jpg (visited on 05/28/2018).

[69] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. New York, NY, USA: Cambridge University Press, 2008.

[70] K. S. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft, "When is "nearest neighbor" meaningful?" In *Proceedings of the 7th International Conference on Database Theory*, ser. ICDT '99, London, UK, UK: Springer-Verlag, 1999, pp. 217–235, ISBN: 3-540-65452-6. [Online]. Available: http://dl.acm.org/citation.cfm?id=645503.656271.

[71] R. Khane, A. D Surdi, and R. Shakar Bhatkar, "Changes in ecg pattern with advancing age," vol. 22, pp. 97–101, Dec. 2011.

[72] S. Bachman, D. Sparrow, and L. Smith, "Effect of aging on the electrocardiogram," *The American Journal of Cardiology*, vol. 48, no. 3, pp. 513–516, 1981, ISSN: 0002-9149. DOI: https://doi.org/10.1016/0002-9149(81)90081-3. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0002914981900813.

[73] O. P. Patri, A. B. Sharma, H. Chen, G. Jiang, A. V. Panangadan, and V. K. Prasanna, "Extracting discriminative shapelets from heterogeneous sensor data," in *Big Data (Big Data), 2014 IEEE International Conference on*, IEEE, 2014, pp. 1095–1104.

Glossary

1-NN 1-Nearest Neighbor. 13, 14, 42, 44, 50, 53

AAMI Association for the Advancement of Medical Instrumentation. 2, 19–22, 24, 25, 40, 46–48, 52

AAU-ECG Aalborg University Electrocardiograms. i, ii, 1–4, 6, 8, 10, 13, 19, 23–27, 30, 38–42, 45, 47, 50, 51, 54, 55, 62, 64, 66, 68, 88

ACC Overall Accuracy. 22, 40, 41, 52, 53

AMI (740) Anterior infarct. 55, 59, 60

ANOVA Analysis of Variance. 18

BPF Butterworth Bandpass Filter. 28

BST Binary Shapelet Transform. i, 2, 17, 18, 24–26, 29, 30, 32, 33, 35, 38, 39, 43–45, 47–53, 62, 64

CHE Chamber Hypertrophy or Enlargement. 26, 27

COTE The Collective of Transformation-Based Ensembles. 13

DT Decision Tree. 17, 18

DTW Dynamic Time Warping. 13, 14, 44, 49–53

ECG Electrocardiogram. i, ii, 1–10, 13, 14, 18–22, 24, 25, 27–31, 36, 39, 41, 47, 49–51, 53, 56, 57, 64–70, 84, 85

FN false negative. 40–42

FP false positive. 40–42, 53

FPR false positive rate. 40, 41, 52, 53, 68

FSS Fast Shapelets Search. i, ii, 2, 3, 17, 33, 37, 45, 60–62, 69

HESCA Heterogeneous Ensembles of Standard Classification Algorithms. 13, 16, 18, 24, 30, 39, 40, 44, 46, 49, 51, 60, 70

IC Intraventricular Conduction. 26, 27

INF Infarction. 26, 27

IRBBB (445) Incomplete right bundle branch block. 54

LBBB (460) Left Bundle Branch Block. 39, 54

LD Linear Discriminant. 14, 22

LVH (540) Voltage criteria for left ventricular hypertrophy. 11, 55–58, 65

LVH2 (541) Left Ventricular Hypertrophy. 54–56, 58

LVH3 (548) Moderate voltage criteria for LVH, may be normal variant. 55–57

Marquette Marquette™ 12SL™ ECG analysis program. i, ii, 1–4, 6–8, 10, 11, 27, 42, 47, 50, 51, 54–60, 63, 64, 67, 68

MIT-BIH Massachusetts Institute of Technology – Beth Israel Hospital. i, ii, 1–3, 19–25, 27, 28, 41, 45, 47–53, 60, 62, 64, 68, 69

MTSC Multivariate Time Series Classification. 13

NST (900) Nonspecific ST Abnormality. 55

PAA Piecewise Aggregate Approximation. 23, 26, 27, 29, 37, 38, 42–45

PCA Principal Component Analysis. 43, 44

QRSV (542) Minimal voltage criteria for LVH, may be normal variant. 55–57

RA Repolarization Abnormalities. 26, 27, 66

RandF Random Forest. 16

RBBB (440) Right Bundle Branch Block. 11, 39, 54–56, 63, 65

RotF Rotation Forest. 13, 16, 43, 44

SAX Symbolic Aggregate Approximation. 37, 38, 45

ST Shapelet Transform. 13, 14, 16–18, 30–33, 35, 39, 51, 64

SVM Support Vector Machine. 22

TN true negative. 40–42

TP true positive. 40–42

TSC Time Series Classification. 1, 2, 13, 16, 18, 30, 50

WEKA Waikato Environment for Knowledge Analysis. 40, 44

A Statements

This appendix lists all the diagnostic statements present in the data set. The table is sorted by the statement number. We use the following group abbreviations:

- INF - Infarction
- QAV - QRS Axis and Voltage
- IC - Intraventricular Conduction
- RA - Repolarization Abnormalities
- CHE - Chamber Hypertrophy and Enlargement

Stmt	Text	Group	Count
300	Ventricular pre-excitation, WPW pattern type A	IC	111
302	Ventricular pre-excitation, WPW pattern type B	IC	121
304	Wolff-Parkinson-White	IC	321
307	Dextrocardia	QAV	3
350	Right atrial enlargement	CHE	4918
360	Left atrial enlargement	CHE	15598
369	Biatrial enlargement	CHE	1083
372	Left axis deviation	QAV	66,255
380	Rightward axis	QAV	10837
383	Right axis deviation	QAV	1152
384	Right superior axis deviation	QAV	931
390	Indeterminate axis	QAV	468
391	Northwest axis	QAV	3
410	Low voltage QRS	QAV	30699
411	Pulmonary disease pattern	QAV	2712
435	Brugade pattern type 1	RA	35
440	**Right bundle branch block**	**IC**	**25532**
442	Right bundle branch block -or- Right ventricular hypertrophy	IC	4
445	**Incomplete right bundle branch block**	**IC**	**20780**
450	RSR' or QR pattern in V1 suggests right ventricular conduction delay	IC	1994
460	**Left bundle branch block**	**IC**	**13283**
465	**Incomplete left bundle branch block**	**IC**	**2638**
470	Left anterior fascicular block	IC	19756
471	Left posterior fascicular block	IC	661
480	*** Bifascicular block ***	IC	4880
482	Nonspecific intraventricular block	IC	5639
487	Nonspecific intraventricular conduction delay	IC	7932

Stmt	Text	Group	Count
520	Right ventricular hypertrophy	CHE	1109
530	R in aVL	CHE	39497
531	sokolow-Lyon	CHE	31635
533	Cornell product	CHE	60392
534	Romhilt-Estes	CHE	5612
540	**Voltage criteria for left ventricular hypertrophy**	**CHE**	**729**
541	**Left ventricular hypertrophy**	**CHE**	**20766**
542	**Minimal voltage criteria for LVH, may be normal variant**	**CHE**	**73314**
548	**Moderate voltage criteria for LVH, may be normal variant**	**CHE**	**11599**
570	Biventricular hypertrophy	CHE	370
700	**Septal infarct**	**INF**	**21310**
740	**Anterior infarct**	**INF**	**24146**
760	**Lateral infarct**	**INF**	**3682**
780	**Inferior infarct**	**INF**	**47987**
801	**Inferior-posterior infarct**	**INF**	**823**
802	Posterior infarct	INF	109
803	Larger R/S ratio in V1, consider early transition or posterior infarct	INF	1679
806	Consider right ventricular involvement in acute inferior infarct	INF	336
810	**Anteroseptal infarct**	**INF**	**8845**
820	**Anterolateral infarct**	**INF**	**2079**
821	** ** ACUTE MI / STEMI ** **	INF	1413
900	**Nonspecific ST abnormality**	**RA**	**43550**
901	Acute pericarditis	RA	37
902	ST elevation, consider early repolarization, pericarditis, or injury	RA	528
903	ST elevation, probably due to early repolarization	RA	1303
930	Anterior injury pattern	RA	57
940	Lateral injury pattern	RA	63
950	Inferior injury pattern	RA	45
961	Anterolateral injury pattern	RA	7
962	Inferolateral injury pattern	RA	2
963	ST elevation, consider inferior injury or acute infarct	RA	189
964	ST elevation, consider anterior injury or acute infarct	RA	320
965	ST elevation, consider lateral injury or acute infarct	RA	60
966	ST elevation, consider anterolateral injury or acute infarct	RA	71
967	ST elevation, consider inferolateral injury or acute infarct	RA	31
1000	Early repolarization	RA	1820
1001	Junctional ST depression, probably normal	RA	1128
1002	Junctional ST depression, probably abnormal	RA	71
1024	ST depression, consider subendocardial injury	RA	750
1040	Marked ST abnormality, possible septal subendocardial injury	RA	16
1050	Marked ST abnormality, possible anterior subendocardial injury	RA	304

Stmt	Text	Group	Count
1060	Marked ST abnormality, possible lateral subendocardial injury	RA	447
1070	Marked ST abnormality, possible inferior subendocardial injury	RA	660
1071	Marked ST abnormality, possible inferolateral subendocardial injury	RA	103
1080	Marked ST abnormality, possible anteroseptal subendocardial injury	RA	38
1081	Marked ST abnormality, possible anterolateral subendocardial injury	RA	220
1140	**Nonspecific T wave abnormality**	**RA**	**27801**
1141	**Nonspecific ST and T wave abnormality**	**RA**	**22152**
1142	Abnormal QRS-T angle, consider primary T wave abnormality	RA	3491
1143	**Prolonged QT**	**RA**	**18827**
1145	T wave abnormality, consider inferolateral ischemia	RA	2551
1150	**T wave abnormality, consider anterior ischemia**	**RA**	**3844**
1151	Marked T wave abnormality, consider anterior ischemia	RA	18
1160	**T wave abnormality, consider lateral ischemia**	**RA**	**11387**
1161	Marked T wave abnormality, consider lateral ischemia	RA	102
1170	**T wave abnormality, consider inferior ischemia**	**RA**	**7532**
1171	Marked T wave abnormality, consider inferior ischemia	RA	24
1172	Marked T wave abnormality, consider inferolateral ischemia	RA	108
1180	**T wave abnormality, consider anterolateral ischemia**	**RA**	**4078**
1181	Marked T wave abnormality, consider anterolateral ischemia	RA	726

B An ECG

This appendix section presents an example of the leads for a random ECG The leads are shown as sub figures, where the left column contains the limb leads *I*, *II* and *III* and the augmented leads *aVL*, *aVF* and *aVR* and the right column has the precordial leads *V1-6*.

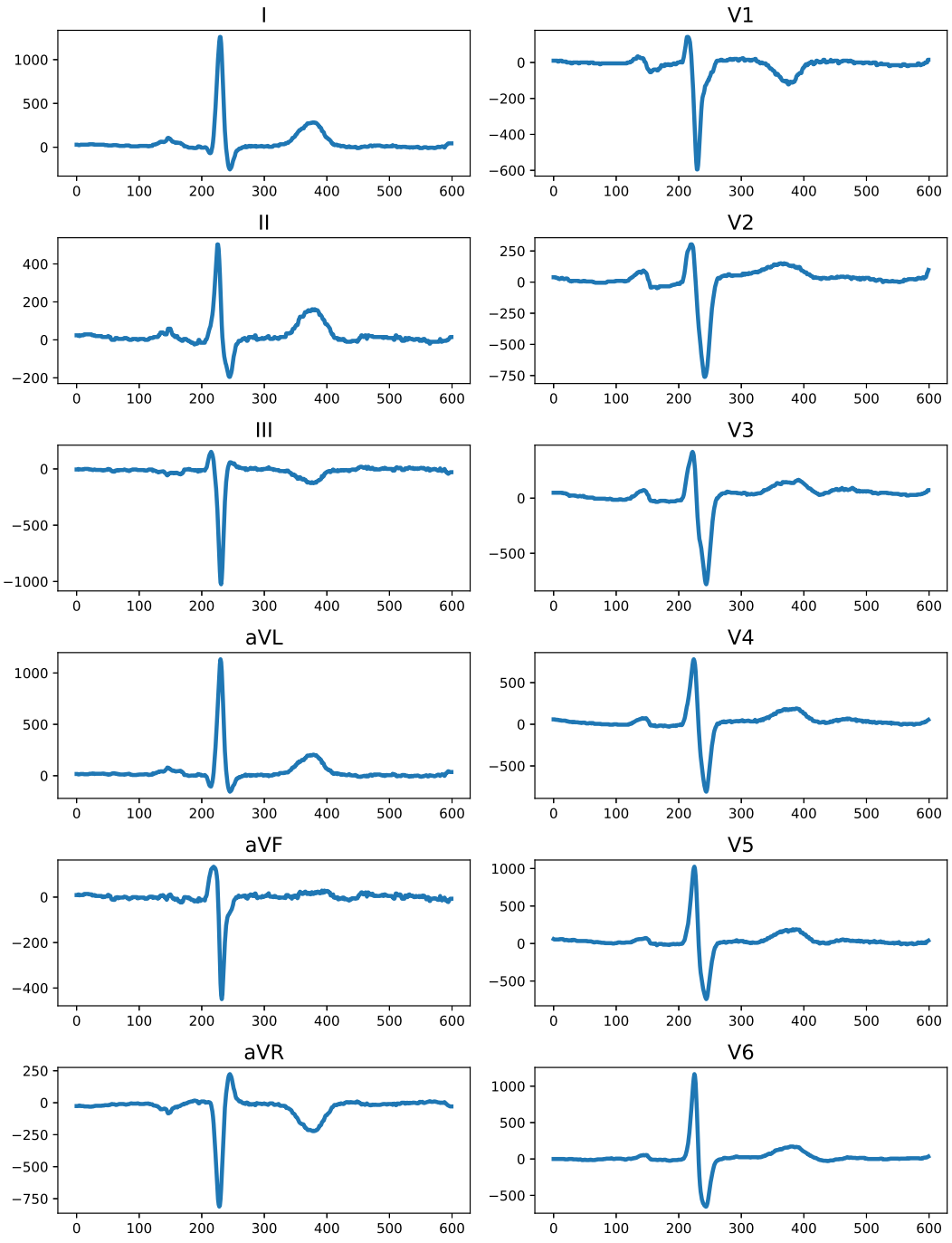
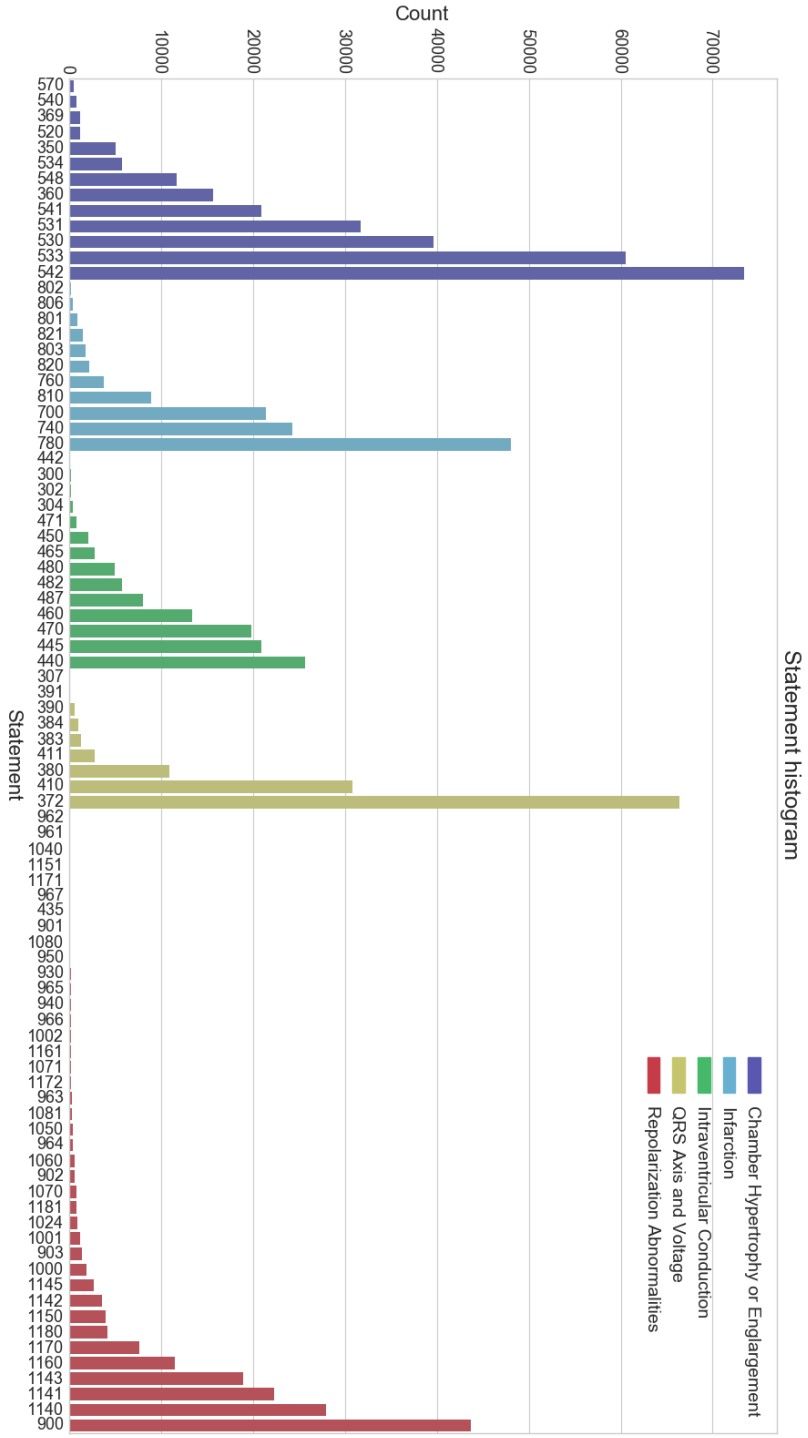


Figure B.1: *An example of the leads on a ECG*

C Histogram

Appendix C illustrates the distribution of our diagnosis statements in the data set. There is only shown the 76 remaining diagnosis statements after we have performed the filtering mentioned in Section 1.2.2.


Count
0
10000
20000
30000
40000
50000
60000
70000
Statement
570
540
369
520
350
534
548
360
541
531
530
533
542
802
806
801
821
803
820
760
810
700
740
780
442
300
302
304
471
450
465
480
482
487
460
470
445
440
307
391
390
384
383
411
380
410
372
962
961
1040
1151
1171
967
435
901
1080
950
930
965
940
966
1002
1161
1071
1172
963
1081
1050
964
1060
902
1070
1181
1024
1001
903
1000
1145
1142
1150
1180
1170
1160
1143
1141
1140
900
Statement histogram
Chamber Hypertrophy or Englargement
Infarction
Intraventricular Conduction
QRS Axis and Voltage
Repolarization Abnormalities

D Shapelets

This appendix presents examples of shapelets of few select diagnosis statements generated from the AAU-ECG dataset. The shapelets we display are from the using the full shapelet search method and using information-gain as the shapelet quality measure. We order the pictures as subplots each belonging to a particular lead for example *aVL*. The left column on shapelet picture contains the limb leads *I*, *II* and *III* and the augmented leads *aVL*, *aVF* and *aVR* and the right column has the precordial leads *V1*, *V2* and so on. We show the 20 best shapelets for each diagnosis statement.

We show shapelets for the following diagnosis statements:

- 440 - Right Bundle Branch Block
- 542 - Minimal Voltage criteria for Left Ventricular Hypertrophy
- 548 - Moderate Voltage criteria for Left Ventricular Hypertrophy
- 540 - Voltage criteria for Left Ventricular Hypertrophy
- 541 - Left Ventricular Hypertrophy
- 740 - Anterior infarct

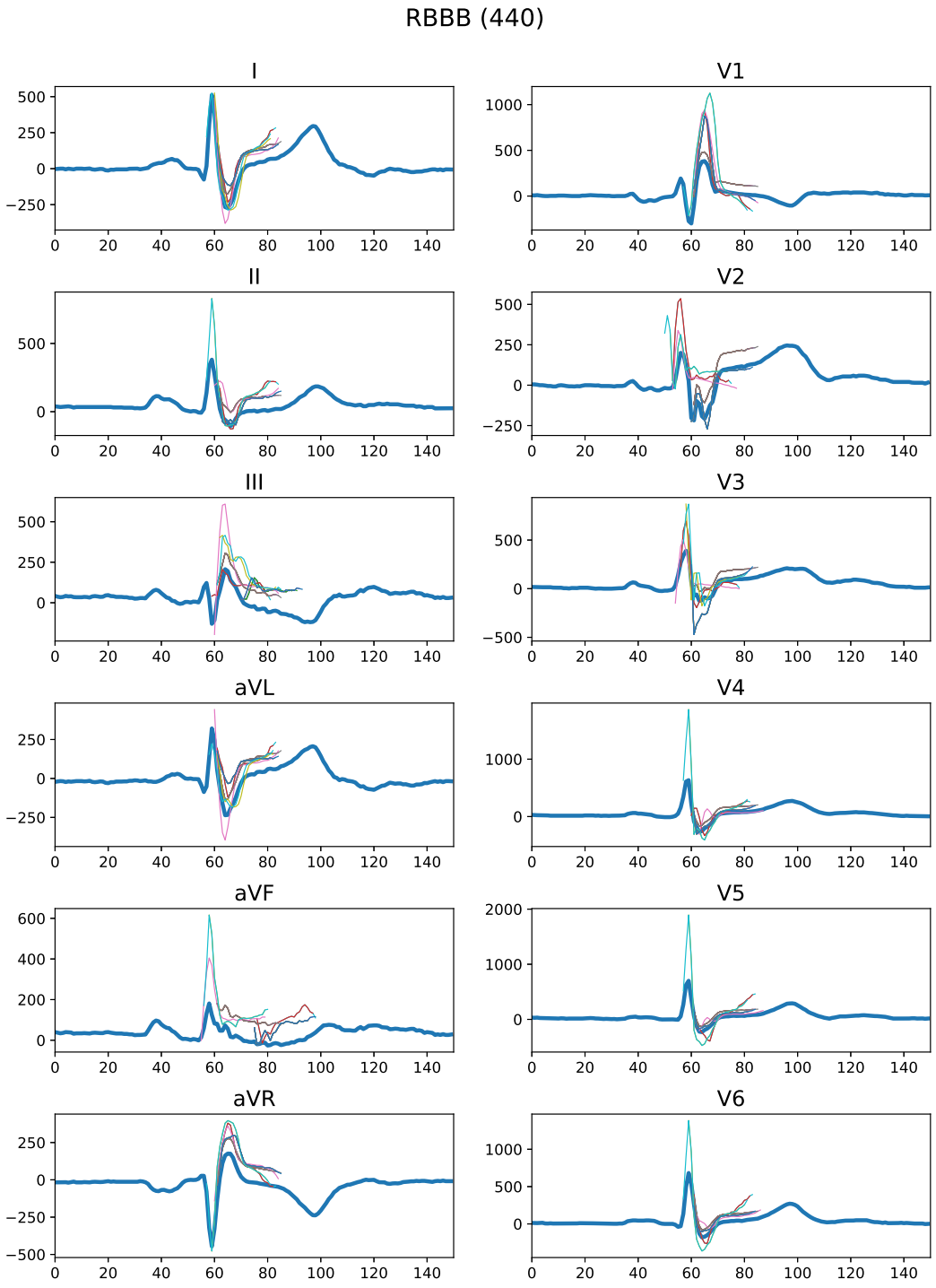


Figure D.1: *Shapelets for Right Bundle Branch Block (440)*

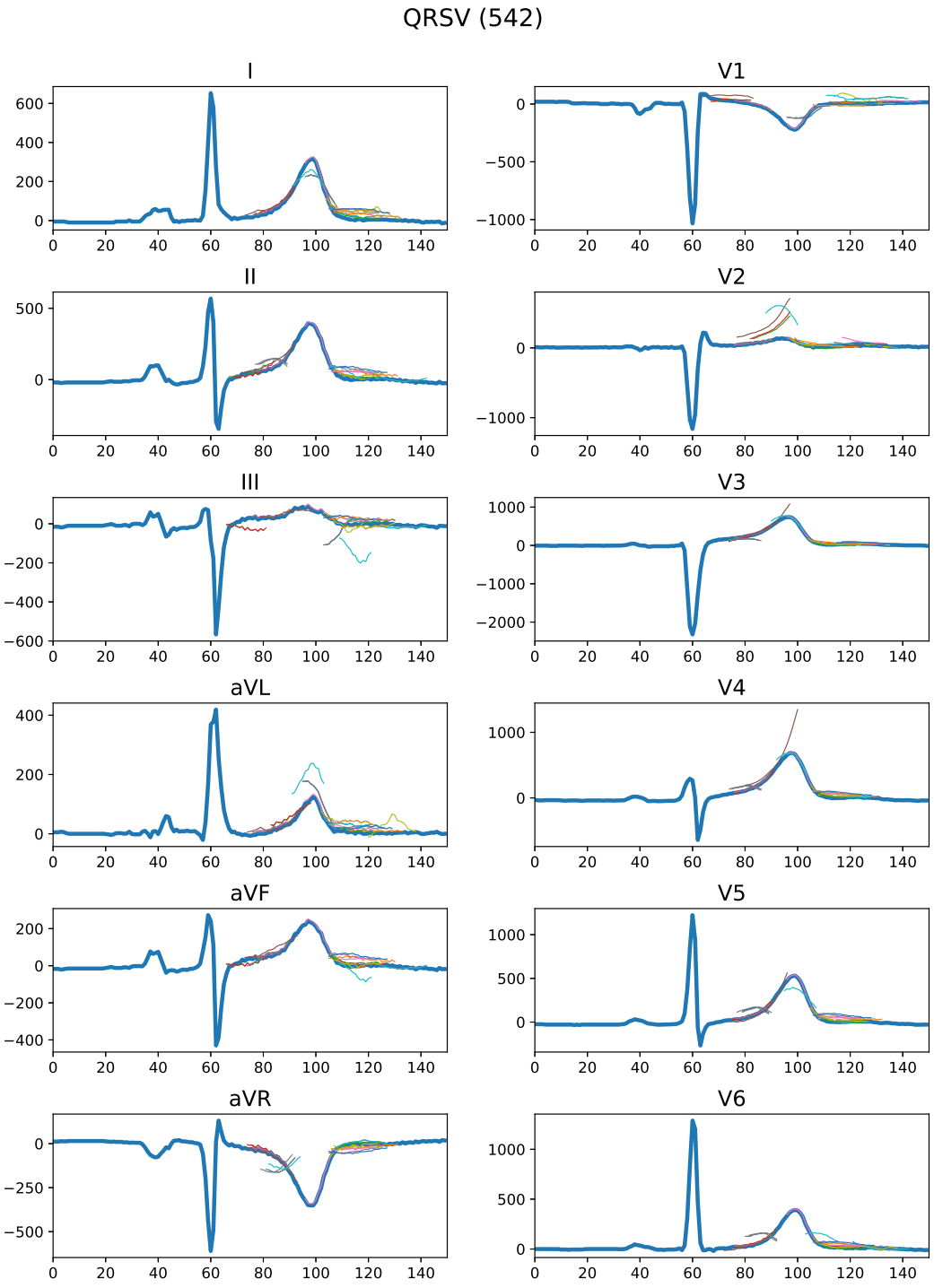


Figure D.2: *Shapelets for Minimal voltage criteria for LVH, may be normal variant (542)*

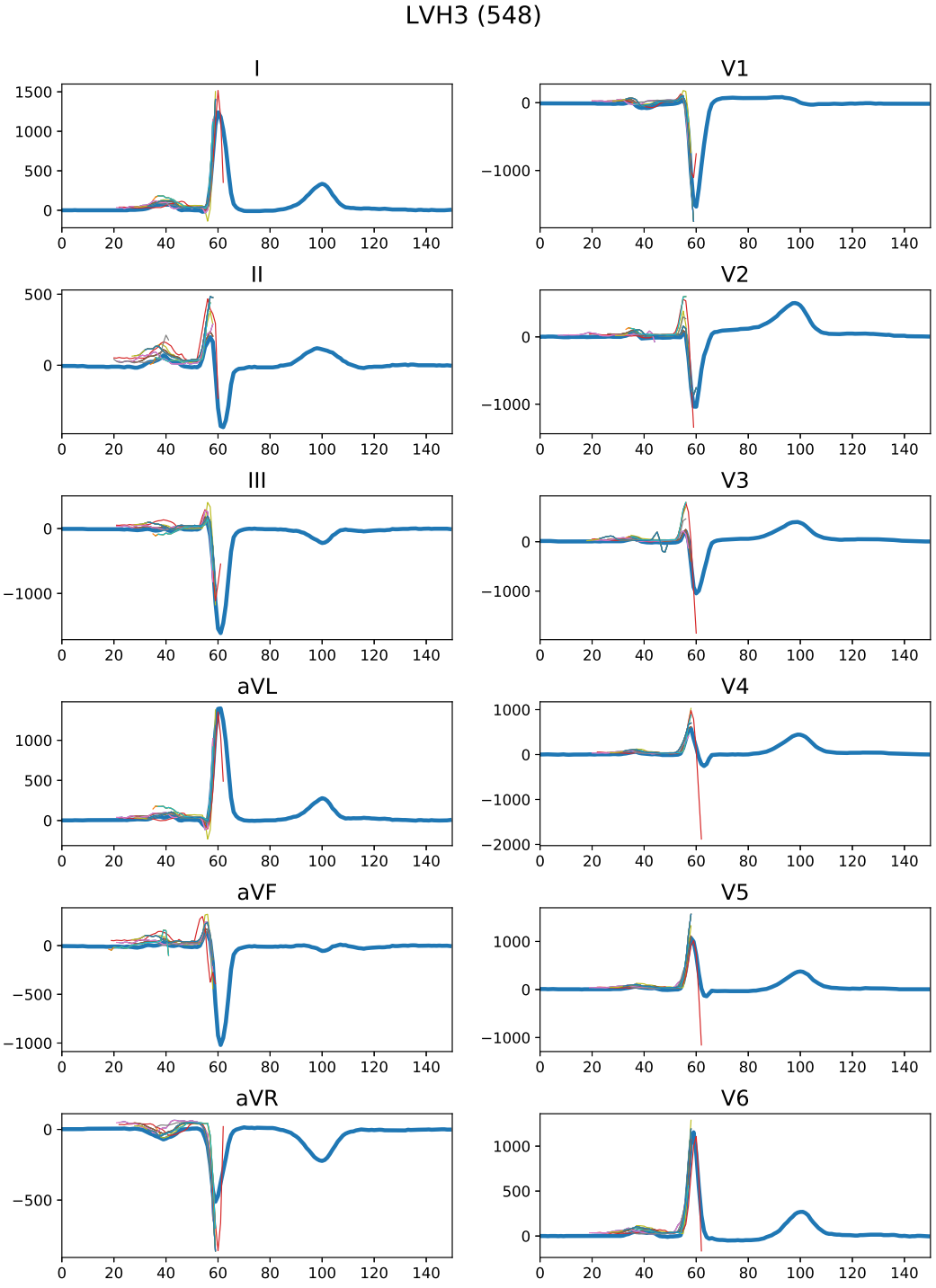


Figure D.3: *Shapelets for Moderate voltage criteria for LVH, may be normal variant (548)*

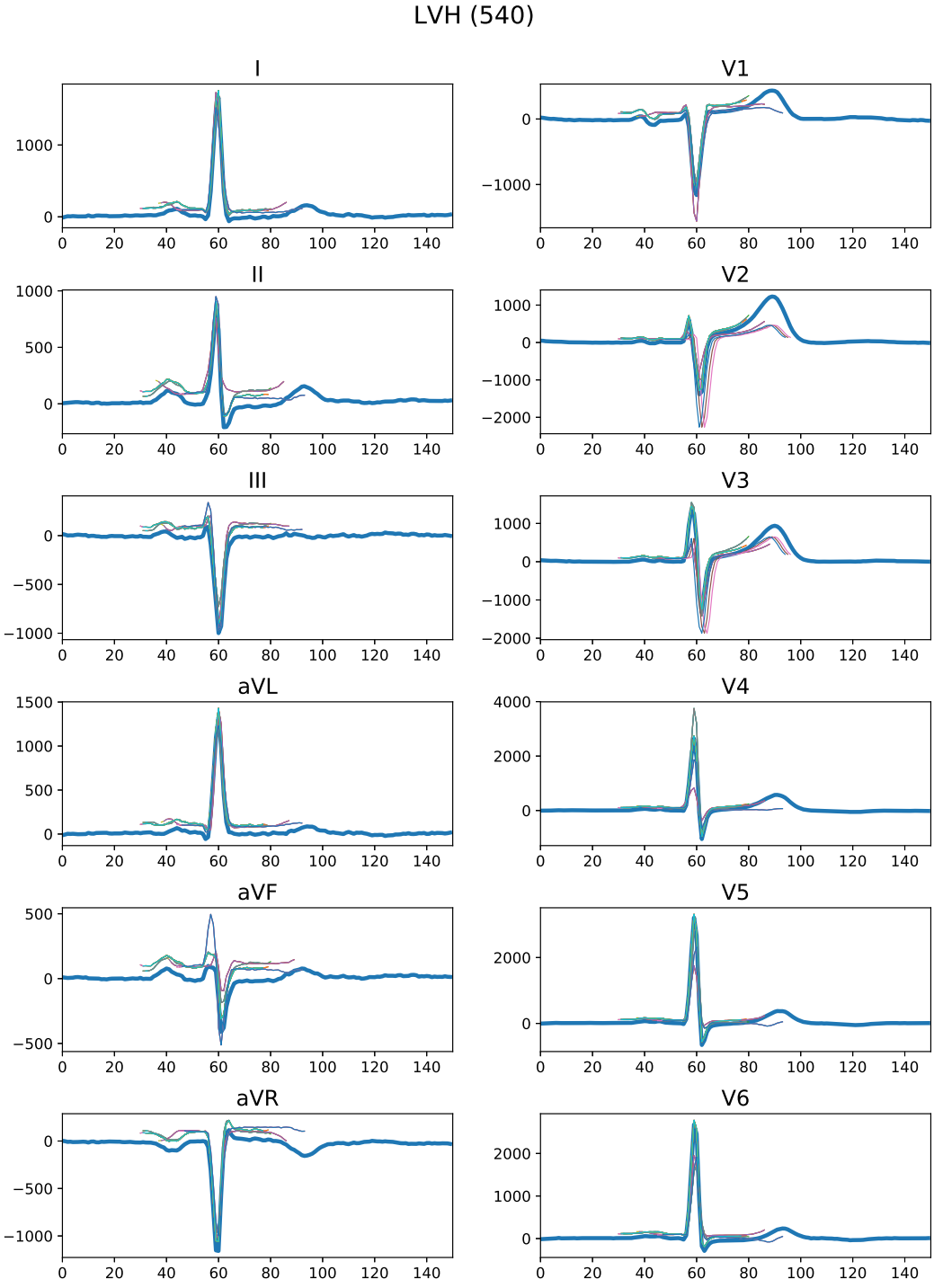


Figure D.4: *Shapelets for Voltage criteria for left ventricular hypertrophy (540)*

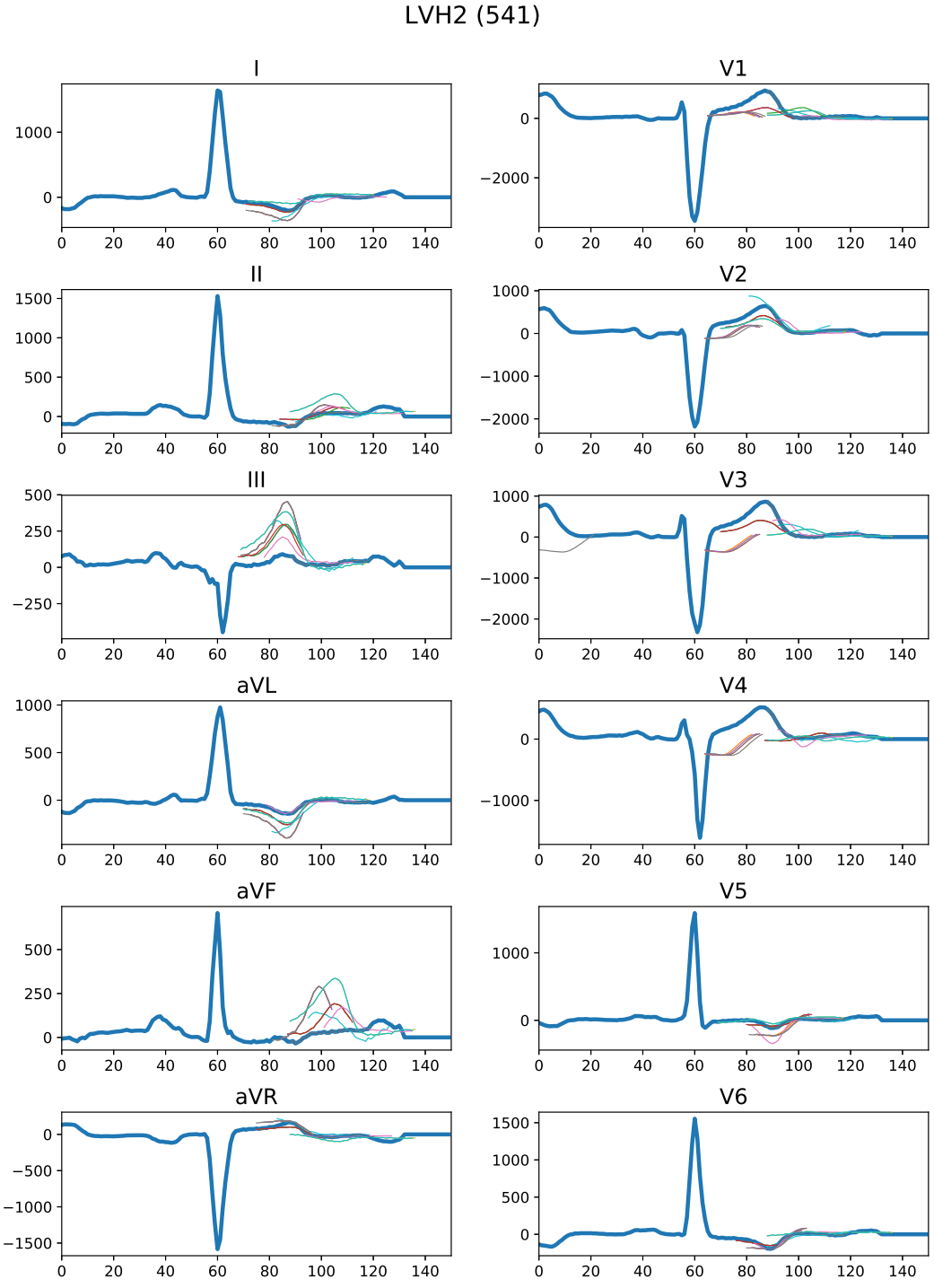


Figure D.5: *Shapelets for Left Ventricular Hypertrophy (541)*

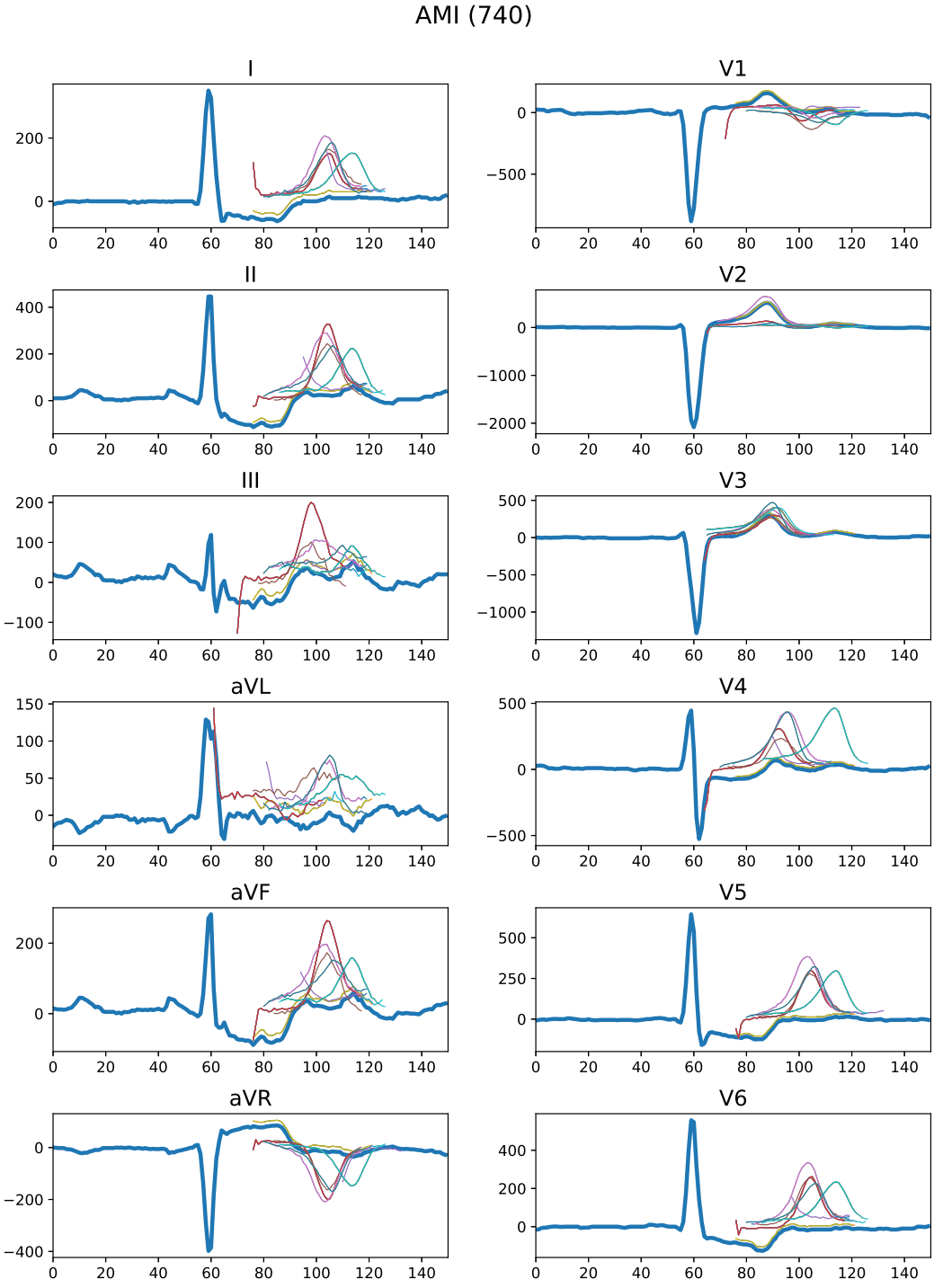


Figure D.6: *Shapelets for Anterior infarct (740)*