

Selecting the best rhythmical and morphological features for a QRS complex classification system

Bachelor Thesis

presented by

cand. el. Robert Menges



INSTITUTE OF BIOMEDICAL ENGINEERING
PROF. DR. RER. NAT. OLAF DÖSSEL
KARLSRUHE INSTITUTE OF TECHNOLOGY
2013

Supervisor: Dipl.-Ing. Gustavo Lenis

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Arbeit selbstständig und ohne unzulässige fremde Hilfsmittel angefertigt habe. Wörtlich oder inhaltlich übernommene Stellen sind als solche kenntlich gemacht und die verwendeten Literaturquellen im Literaturverzeichnis vollständig angegeben. Die „Regeln zur Sicherung guter wissenschaftlicher Praxis im Karlsruher Institut für Technologie (KIT)“ in ihrer gültigen Form wurden beachtet.

Karlsruhe, den 13.11.2013

Acknowledgments

First of all, I would like to thank Prof. Dr. rer. nat. Olaf Dössel that I got the chance to write my bachelor thesis at the Institute of Biomedical Engineering at the KIT. A big thank is also dedicated to my supervisor Gustavo Lenis who always helped me whenever needed and from whom I learned a lot. I professionally and personally developed in our cooperative work definitely. Another thank to all the members of the biosignal group. Interesting talks in the meetings helped me in proceeding with this thesis. Furthermore, I would like to thank my parents and brothers for supporting me all the time, especially in the final phase of writing. Finally, I thank all the students of the Diplomandenraum. During the time of this thesis, I had a lot of fun working on the 4th floor.

Contents

1	Introduction	1
1.1	Motivation and problem	1
1.2	Structure	1
2	Fundamentals	3
2.1	Medical background	3
2.1.1	Anatomy and physiology of the human heart	3
2.1.2	Electrical activity of the human heart	4
2.1.3	Electrocardiogram (ECG)	6
2.1.4	Ectopic beats	9
2.2	Mathematical background	12
2.2.1	Support vector machine (SVM)	12
2.2.2	Receiver-operating characteristic (ROC)	16
2.2.3	Classification tree	19
2.2.4	Gini Diversity Index (GDI)	21
2.2.5	Information Gain Ratio (IGR)	21
2.2.6	Basics in statistics	23
3	State of the art	25
4	Feature evaluation	27
4.1	Evaluation using ROC	28
4.2	Evaluation using GDI	30
4.3	Evaluation using IGR	32
4.4	Evaluation using statistical methods	38
4.5	Equalization of data	46
4.6	Feature dependency	47
5	Choice of new features	51
5.1	Final feature rankings	51
5.2	Choice of new features	58
6	Training a new support vector machine	59
6.1	Feature normalization	59
6.2	Training and Testing	60
7	Results and discussion	63
7.1	Results	63
7.2	Discussion	64

8	Summary and future prospects.....	67
8.1	Summary	67
8.2	Future prospects	68
A	Appendix A.....	69
	References	75

1

Introduction

The task of this bachelor thesis is to optimize an algorithm which automatically detects and classifies normal and ectopic beats in electrocardiogram (ECG) signals. The analysis should be done by evaluating the features used for the classification. This first passage gives the reader a motivation for the research project and describes the problems dealt with. In the second part, the structure and way of proceeding are presented.

1.1 Motivation and problem

Ectopic beats are a common cause for cardiac arrhythmia which in particular cases can even lead to sudden death. Especially atrial fibrillation, as the most frequent cardiac arrhythmia, can be initiated by beats which do not originate from the sinus node. In order to improve life quality of patients suffering from cardiac arrhythmias and prolong their lives, it is important to find these ectopic beats. As the ECG is one of the most important diagnosing methods in clinical daily practice, the correct detection and classification of ectopic beats in ECG signals would make a big contribution to the diagnosis and monitoring of cardiac arrhythmias. To apply this process to clinical routine, it is required to develop an automatic detection and classification of all beats. In previous works, these algorithms were already implemented using a support vector machine (SVM) as a classifier [1]. The SVM is able to classify the beats as normal or ectopic using 55 rhythmical and morphological features for every QRS complex. Every beat class should ideally differ in its feature values in order to ensure a good separability.

This thesis is about evaluating these features and deciding if a reduction of features can improve the classification performance. The idea is that some features do not yield any further information, e.g. as they are based on the same attribute. They may be redundant and should not be used by the SVM anymore. Features that are influenced by noise processes are also not ideal for the classification. Furthermore, too many features may even confuse the classifier instead of supporting the classification task. These are the reasons why it is important to reduce the number of features.

A better and more reliable classification may help physicians in the future to make clearer decisions concerning a therapy for cardiac patients.

1.2 Structure

First in this thesis, fundamental anatomical and physiological knowledge, on which the work is based, is presented. To entirely understand this thesis, mathematical principles

are also explained. The following chapter three summarizes the contents of other research projects and publications which dealt with similar tasks. To distinguish between normal and ectopic beats in the ECG, several features have already been developed. Based on their values, the classes of beats get separated by the SVM. To improve the classification results, these features are evaluated. Therefore, different methods for feature analysis are implemented. The feature evaluation is described in detail in chapter four. To get the best SVM, two parameters have to be optimized: C and γ . That is why a closer look at the process of training and testing of the SVM is necessary. Chapter five is about the intensive examination of this topic. This leads to the results presented in chapter six. They include a final choice based on the previous evaluation which features are finally used to train the new SVM. Furthermore, the classification results of the new SVM are shown here. A final discussion of the results in context is also given. In the end, the entire work is summarized. Possible future prospects relating to feature evaluation are pointed out.

2

Fundamentals

2.1 Medical background

This section is to inform the reader about the anatomical and physiological principles one has to know to entirely understand the research project.

2.1.1 Anatomy and physiology of the human heart

The heart is the central organ of the human circulatory system. It is located in the left half of the thorax, its size is comparable to a human fist and it weights about 300 g. It is divided into two halves, left and right. Each is subdivided into left atrium (LA) and left ventricle (LV) on the one hand and right atrium (RA) and right ventricle (RV) on the other hand. The atria are separated by the interatrial septum while the ventricles are separated by the interventricular septum. The function of the heart is to maintain the blood flow by pumping oxygenated blood from the lungs through the body and deoxygenated blood back to the lungs. The heart is able to keep up a stable circulation of the blood and to adapt its pumping function to outside influences [2].

The cardiac cycle can be split into two phases. First, the contraction of the heart builds up a certain pressure and ejects blood in the systemic circle. This process is called the systole. The left part of the heart pumps oxygen-rich blood into the aorta which is the main transport path through the body. The right part on the other hand pumps oxygen-deficient blood to the lungs via the left and right pulmonary arteries. Second, the systole is followed by a relaxing phase, named diastole. In that time, the four chambers are filled with blood again and made to be ready for the next contraction. Deoxygenated blood comes into the RA via the superior and inferior vena cava. The LA is meanwhile filled with oxygenated blood delivered by the pulmonary veins from the lungs. To allow blood to flow from the atria to the ventricles, the atrioventricular valves exist, in detail the tricuspid valve at the right side and the mitral valve at the left side. They can be seen as inner connections of atria and ventricles. The papillary muscles which are connected to the atrioventricular valves via the chordae tendineae prevent the valves to flip back in the atria and because of this avoid a disturbance of the blood flow. The outer connections of the heart to the circulatory system are also built of two valves, called semilunar valves. The deoxygenated blood leaves the RV through the pulmonary valve into the pulmonary circulation while the aortic valve represents the gate for the oxygenated blood to the body. The four valves are taken together to the so called valvular plane and have a great impact on the direction of the blood flow in the heart from the atria to the ventricles and further on to the aorta respectively the pulmonary arteries [2, 3]. The anatomy of the

heart is graphically shown in figure 2.1.

The wall of the heart is composed of three muscular layers - the epicardium, the endocardium and the myocardium. The exterior is built of a thin membrane, the epicardium. It is strongly connected to the myocardium which mainly consists of cardiac muscle cells and which makes the main contribution to the pumping function. The endocardium covers the inner surface of the heart. The whole organ is encased by the pericardium, a sack of tissue, which defines the form of the heart, closes it off from other organs and stabilizes its position in the thorax [2].

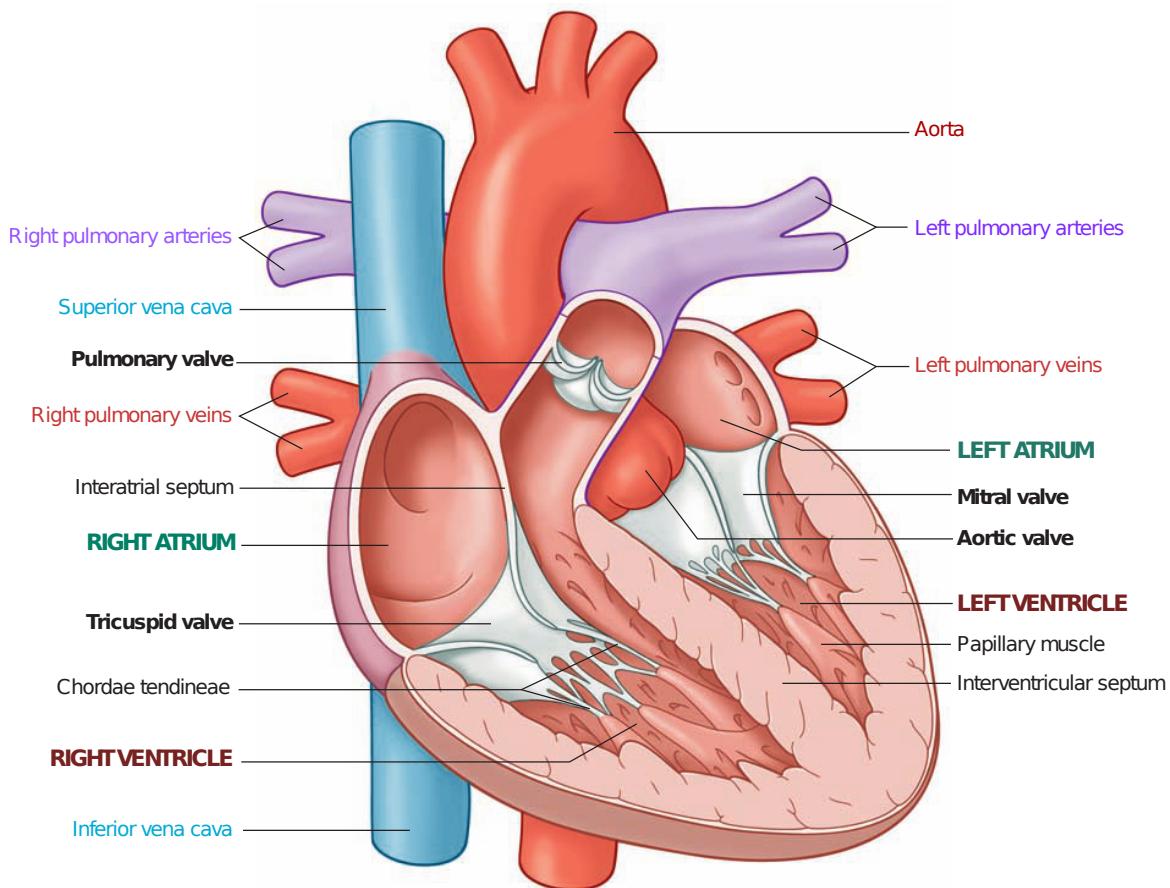


Figure 2.1. Schematic structure of the human heart adapted from [4]. One has to mention that the terms right and left are in medical practice always seen of the patient's perspective.

2.1.2 Electrical activity of the human heart

As already mentioned in 2.1.1, cardiac muscle cells are the principal components of the heart. One special attribute is that they are electrically excitable. In normal case they are at a resting potential of about -90 mV. The change of this potential is mainly caused by ion flows. The reason for the resting potential is the balance of the dominating potassium ions between intra- and extracellular space. For further details have a look at [2].

To get the heart muscle contracted, the muscle cells have to get depolarized. Therefore a

natural pacemaker, the sinoatrial node (SA node or sinus node), exists which is located in the right atrium. These spontaneous active pacemaker cells do not stay at the resting potential but instead depolarize independently and if a threshold potential is reached pass an electrical pulse on to the surrounding cells. This so called action potential (AP) spreads out through the atria and reaches the atrioventricular node (AV node) at the border between RA and RV. The AV node bunches the excitations of the atria and assures that the contraction of the atria is completely finished before it transmits the stimulation to the bundle of His. The bundle of His represents the conducting tissue which ensures that the AP can propagate to the ventricles. This is very important since the valvular plane between atria and ventricles acts as an electrical insulator. The bundle of His is then split into left and right bundle branches and the excitation is brought to the ventricular myocardium via the vascular bundles and Purkinje fibres. After every stimulation the cells stay in a refractory phase, i.e. they cannot be excited again for a short time. This avoids the excitation to return and always leads the triggered contraction in a prearranged direction from the atria via the bundle of His down to the apex and then up again via the Purkinje fibres to the ventricles [2]. The conducting system is pictured in figure 2.2.

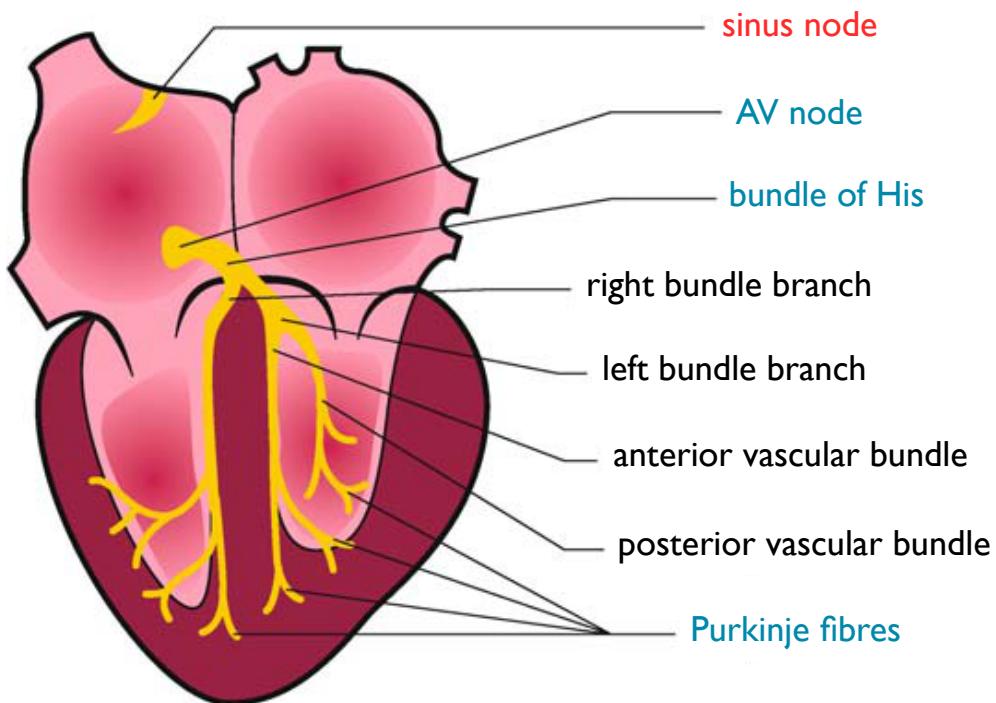


Figure 2.2. Conducting system of the heart and its components. Adapted from [2].

The SA node acts under normal circumstances as primary generator producing 60 to 80 beats per minute in rest. If the stimulus conduction is somehow abnormal, interrupted or the function of the SA node falls completely out, there is a hierarchy of pacemakers which may take over its work. In this case the AV node would fill in its place supplying 40 to 50 beats per minute to keep the heart beating. As last instances the bundle of His and the bundle branches are known as tertiary pacemakers. Their characteristic frequency is

at 30 to 40 beats per minute [2].

But sometimes irregularities may also cause a premature excitation of parts of the myocardial tissue which does not fit into the normal heart rhythm. The regions provoking extra beats besides the SA node are called ectopic foci. A more detailed description of these ectopic beats is given in section 2.1.4.

2.1.3 Electrocardiogram (ECG)

The electrocardiogram (ECG) is a very important measurement system in clinical practice concerning heart monitoring and the diagnosis of cardiac diseases. The electrical activity of the heart, which initiates the mechanical contraction, is therefore measured as electrical potential differences on the body surface. Not every single cell activity is measured, but the sum of all action potentials of the specific area. Since the current from an excited cell to a neighboring unexcited cell creates the electric field of a current dipole, an electric field strength vector arises. Summarizing the vectors of every single current dipole, the electric summation vector results following the principle of vector addition. It is defined from excited to unexcited cells. The more cardiac cells are reached by the excitation front, the more dipoles are included in the summation and the electric summation vector gets stronger as the single current dipoles show in the same direction. This implies that if an excitation proceeds straight through a large myocardial area, the summation vector is relatively large, e.g. during the excitation of the ventricles. In general, regions with many excitable cells dominate the strength and direction of the summation vector. In figure 2.3 the electric summation vector is shown in correspondence to the heart excitation.

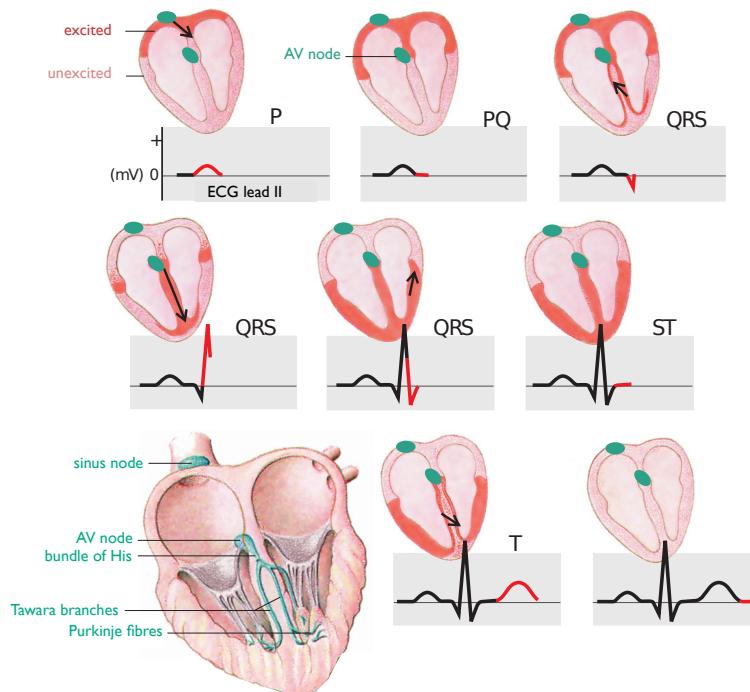


Figure 2.3. Generation of ECG, corresponding heart excitation and electric summation vector. Adapted from [3].

The ECG measuring procedure itself is very easy and popular in clinics as the electrodes are placed directly on the skin, neither causing any injuries nor requiring a big effort for the physician.

Three different systems combined compose the commonly used standard 12-lead-ECG. There are six limb leads which are called after their developers Einthoven and Goldberger. Einthoven designed a system which includes three leads between the left arm, right arm and left foot. The scheme is presented in figure 2.4. Connecting these leads to a reference potential, the Goldberger leads yield a different representation of the excitation building the three other limb leads (figure 2.5). The fact that the Einthoven leads measure the potential difference directly between two electrodes assigns them to the class of bipolar leads. The Goldberger leads instead are called unipolar leads, measuring between one positive electrode and a reference point with potential zero. Finally, to complete the 12-lead-ECG, Wilson's six chest leads are used. They are also unipolar leads and fixed around the chest as seen in figure 2.6 [5].

Einthoven

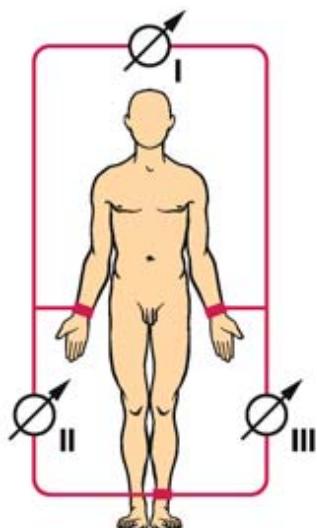


Figure 2.4. Einthoven leads I, II, III. Adapted from [2].

A normal ECG signal is composed of different parts as it describes the whole excitation process of the heart. A schematic ECG image of a complete heart beat is shown in figure 2.7. The signal is divided into P wave, QRS complex and T wave. Sometimes even a U wave can be noticed. From left to right, the P wave displays the depolarization of the atria. The PQ segment represents the time the AV node and the bundle of His need to convey the action potential to the ventricular myocardium. The now proceeding propagation through the ventricles can be discovered by observing the QRS complex which is subdivided into Q, R and S peaks. The interval between the S peak and the T wave is called ST segment. Directly following the ST segment, the T wave shapes the last part of the signal and delineates the repolarization of the cardiac cells (see figure 2.3). Since the Purkinje fibres often show a very long AP duration, their repolarization is sometimes noticeable in an appearing U wave [2, 3].

Goldberger

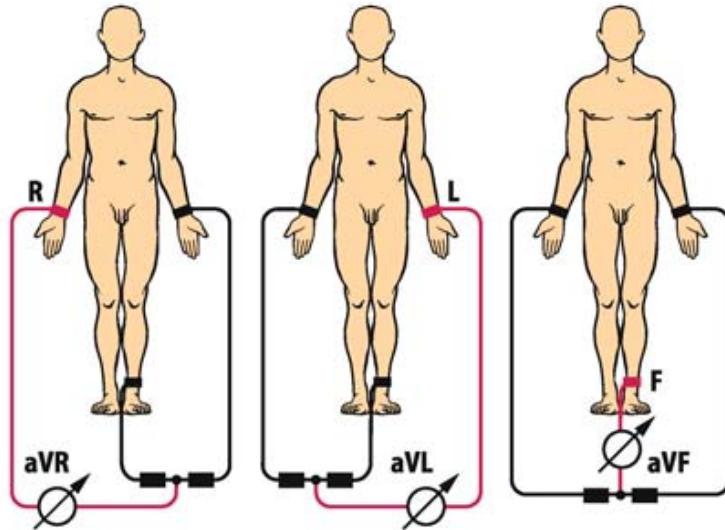


Figure 2.5. Goldberger leads aVR, aVL, aVF. Adapted from [2].

Wilson

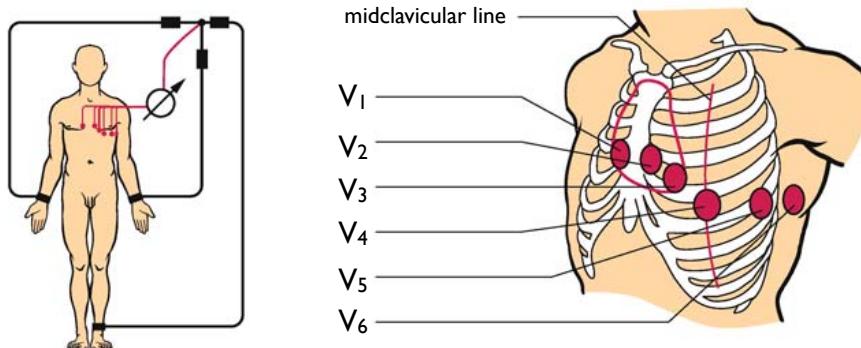


Figure 2.6. Wilson leads V₁-V₆ and their position on the chest. Adapted from [2].

The ECG may provide information about the heart's position, its frequency, the excitation rhythm and its origin, as well as about impulse conduction, excitation degeneration and disturbances. Concerning contraction or pumping power the ECG cannot directly make diagnosing points and the physician has to apply other methods [3].

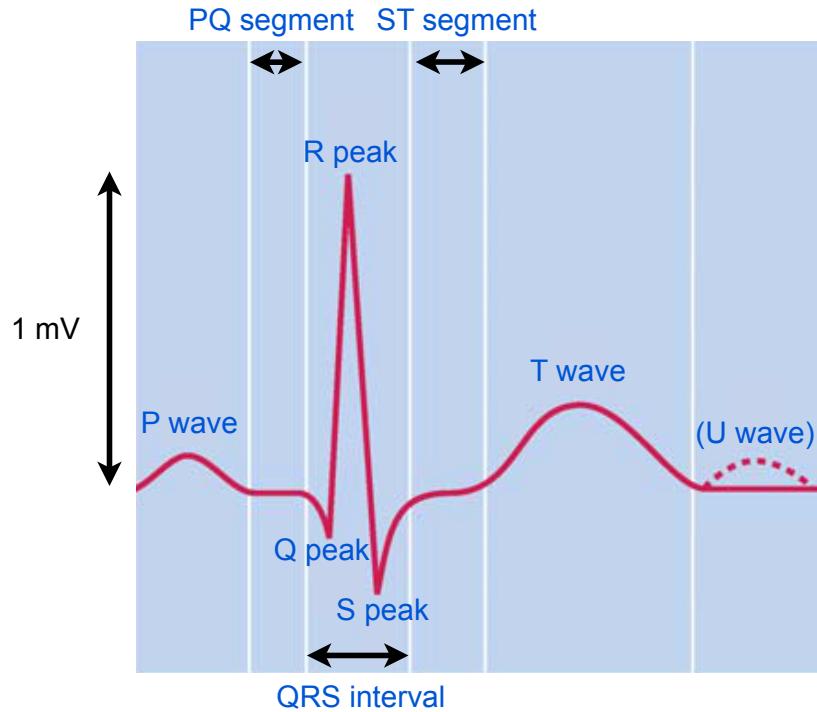


Figure 2.7. Schematic ECG of a complete heart beat with designation of the different parts. Adapted from [2].

2.1.4 Ectopic beats

As mentioned before, ectopic excitations may lead to extra beats which do not fit into the normal sinus rhythm. Physicians distinguish between many different types of these so called extrasystoles. For this research project only two of them are of interest. If the ectopic beat originates from atrial tissue, the beat is classified as supraventricular extrasystole (SVES). In contrast, an excitation starting in ventricular regions is assigned to the class of ventricular extrasystoles (VES).

SVES are characterized by a premature aberrated P wave, which often seems to have an abnormal shape. The detailed morphology of the P wave depends on the exact starting point in the atria. Since its source is located in the atria above the AV node, the excitation is conducted via the conducting system to the ventricular myocardium as usual and hence the QRS complex looks morphologically normal (see figure 2.9). Only aberrant conducted action potentials may lead to a misshapen QRS complex. An AV node resting in refractory state may even block the excitation and prevent a QRS complex completely in rare cases. A compensatory break is not noticeable because the SA node is indeed disturbed but can go on working directly after the SVES, it is comparable to a reset. Nevertheless, the RR interval between the SVES and the beat before is shortened, just as the RR interval to the next beat is prolonged. There are many reason why SVES may appear. Examples are a spontaneous sympathetic nervous system activation or vibrations caused by the influx of the blood into the atria. Most of the time, SVES do not pose a threat to the patient. In most cases a therapy is not needed [2, 5].

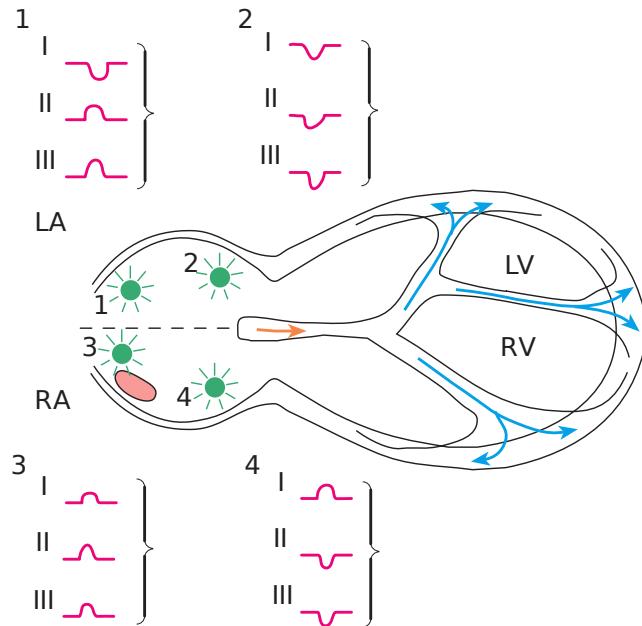


Figure 2.8. SVES with different origins, possible pathways and respective P wave morphologies. Adapted from [5].

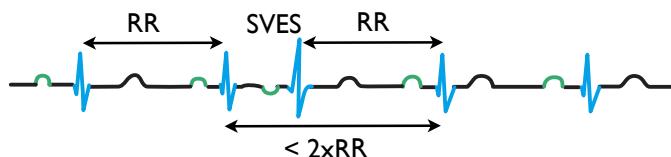


Figure 2.9. ECG signal of typical SVES. Adapted from [5].

In contrast, the fundamental electrocardiographic symptoms of VES are a deformed, premature occurring QRS complex and a missing P wave in consequence of a missing atrial excitation. As distinct from SVES, a compensatory break is indicative for VES. The sinus rhythm is not troubled but the following normal sinus pulse meets a still refractory ventricular myocardium and that is why the next contraction can only be triggered by the next but one stimulus. In figures 2.10 and 2.11 two examples of VES are shown. Regarding the morphology of the extrasystoles, one separates between monomorphic and polymorphic extrasystoles. On the one hand, if they are from identical shape, i.e. their origin lies in the same area of the ventricles, the term monomorphic is used. On the other hand, if their QRS complexes differ from each other, that implies that their origin is located in different ventricular regions, one speaks of polymorphic extrasystoles [5]. In general, VES are often provoked by abnormal Purkinje cells [6]. As they seem to be one cause for cardiac arrhythmia and an increased risk of sudden death, a therapy is highly

recommended, especially, in case the patient is suffering from heart conditions and the VES appear very frequent.

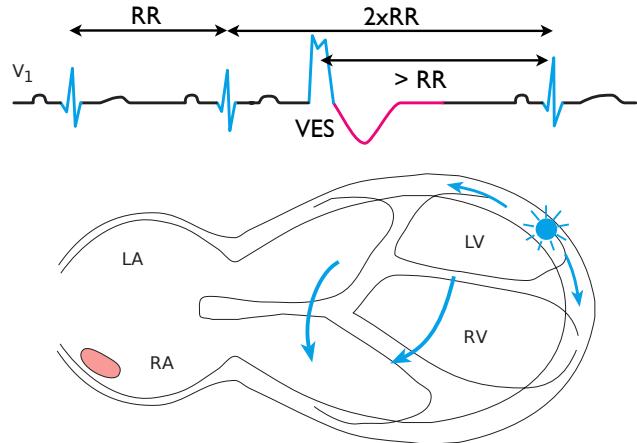


Figure 2.10. VES with origin in LV, corresponding ECG signal and proportion of RR intervals. Adapted from [5].

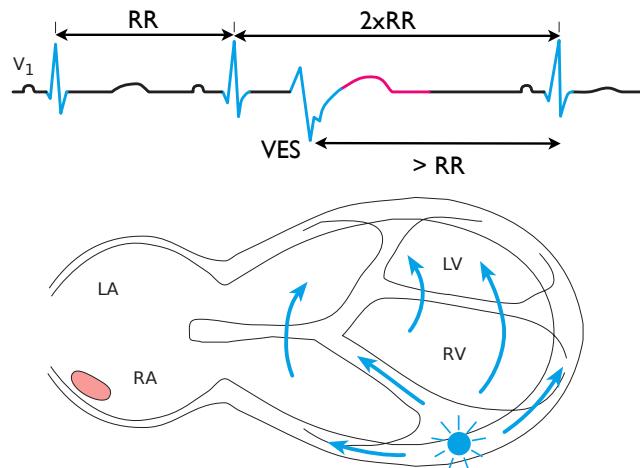


Figure 2.11. VES with origin in RV, corresponding ECG signal and proportion of RR intervals. Adapted from [5].

2.2 Mathematical background

Besides the medical fundamentals, it is also important to know how the applied mathematical methods basically work. In this section the mathematical principles are explained. It is to say to the nomenclature, that vectors are bold print and variables are printed curvishly.

2.2.1 Support vector machine (SVM)

Concerning classification problems, the task of a classifier is to assign unknown data instances characterized by different attributes (features) to different classes. A support vector machine (SVM) works principally by creating a hyperplane which acts as a boundary and thereby separates the classes (see figure 2.12). The optimal hyperplane is given by the maximization of the distance between the hyperplane itself and the nearest data points, called support vectors. Their relevance and usage is explained in detail in section 2.2.1.1. The SVM is a supervised learning method which is descended from the group of linear classifiers. By applying a so called kernel trick, SVMs are also able to classify in a highly nonlinear way (see section 2.2.1.2 for details). The learning procedure is split into two parts: first one has to train the SVM using a training data set where the features and corresponding classes are known. After that an unknown testing data set is given to the SVM whose task is now to predict the classes [7].

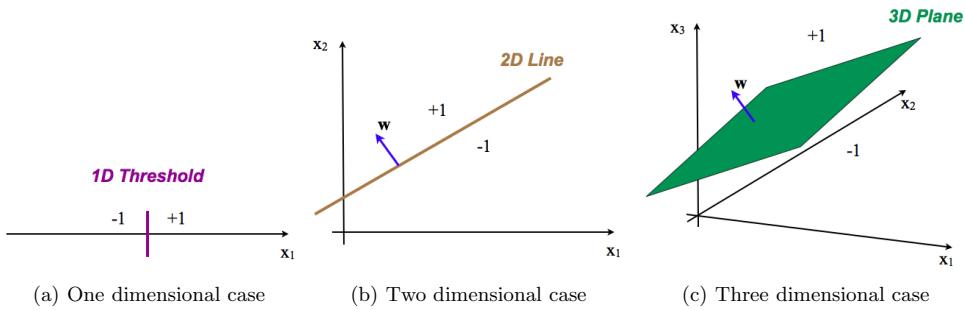


Figure 2.12. Principle of SVM for a) one-dimensional b) two-dimensional c) three-dimensional cases [1].

2.2.1.1 The linearly separable case

Considering the linear separation case first, we make use of a data set consisting of data points of the form

$$\mathcal{L} = \{(\mathbf{x}_i, y_i) : i = 1, 2, \dots, n\} \quad (2.1)$$

where \mathbf{x}_i represents the feature vector and y_i the possible classification results: $y_i \in \{-1; +1\}$. n is the number of data points.

The boundary that separates the classes in the feature space is given by the term

$$\langle \mathbf{x}, \mathbf{w} \rangle + b = x_1 \cdot w_1 + x_2 \cdot w_2 + \dots + x_n \cdot w_n + b = 0 \quad , \quad (2.2)$$

that represents a hyperplane in the linear case. \mathbf{w} is hereby the weight vector of the classifier (in linear case a normal vector to the hyperplane) and b stands for the bias. The classification itself is done as following:

$$z = \begin{cases} +1, & \text{if } \langle \mathbf{x}, \mathbf{w} \rangle + b \geq 0 \\ -1, & \text{else} \end{cases} \quad (2.3)$$

The distance between any point in the feature space and the hyperplane is then calculated by

$$D = \frac{\langle \mathbf{x}, \mathbf{w} \rangle + b}{\|\mathbf{w}\|} \quad . \quad (2.4)$$

The principle of the SVM is to separate the classes using the hyperplane that maximizes the distance between the hyperplane itself and the nearest points in the feature space. The distance between the hyperplane and those nearest points is called the margin. The data points which build the margin are called support vectors. They are the reason for the SVM's name. A maximal large margin assures the best possible classification for the particular case. An example for a linear separable data set and the optimal hyperplane can be seen in figure 2.13.

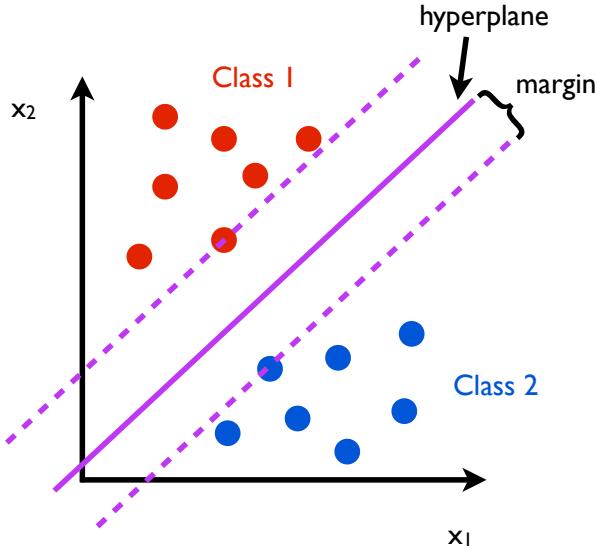


Figure 2.13. Example data points and separating hyperplane with margin

The margin is described by

$$\rho = \min_i \left\{ y_i \cdot \left(\frac{\langle \mathbf{x}_i, \mathbf{w} \rangle + b}{\|\mathbf{w}\|} \right) \right\} \quad . \quad (2.5)$$

By imposing the constraint $\rho = \frac{1}{\|\mathbf{w}\|}$ the optimization problem can be formulated as [8, 9]

$$\min_{\mathbf{w}, b} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 \right\} \quad (2.6)$$

with subject to

$$y_i \cdot (\langle \mathbf{x}_i, \mathbf{w} \rangle + b) \geq 1 \quad \text{over all } i \quad . \quad (2.7)$$

The solution is thus based on the saddle point $(\mathbf{w}_0, \alpha_0, b_0)$ of the Lagrange function

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i ([\langle \mathbf{x}_i, \mathbf{w} \rangle + b] y_i - 1) \quad , \quad (2.8)$$

where the α_i stand for the Lagrange multipliers. The partial differentials of the so called Lagrangian must satisfy the following conditions at the saddle point:

$$\left. \frac{\partial L(\mathbf{w}, b, \alpha)}{\partial \mathbf{w}} \right|_{\mathbf{w}_0} = 0 \quad , \quad (2.9)$$

$$\left. \frac{\partial L(\mathbf{w}, b, \alpha)}{\partial b} \right|_{b_0} = 0 \quad . \quad (2.10)$$

Furthermore, the Karush-Kuhn-Tucker (KKT) conditions are important conditions to be satisfied for this kind of optimization problem. The KKT conditions can be written as [7]

$$1 - y_i \cdot (\langle \mathbf{x}_i, \mathbf{w} \rangle + b) \leq 0 \quad \text{over all } i \quad (2.11)$$

$$\alpha_i \cdot (1 - y_i \cdot (\langle \mathbf{x}_i, \mathbf{w} \rangle + b)) = 0 \quad \text{over all } i \quad (2.12)$$

$$\alpha_i \geq 0 \quad \text{over all } i \quad (2.13)$$

The next task, while satisfying the mentioned conditions, is to minimize (2.8) with respect to \mathbf{w} and b as well as to maximize (2.8) with respect to $\alpha_i > 0$ ([8]).

After different mathematical procedures, the result for the separating rule, based on the optimal hyperplane, comes out as

$$f(\mathbf{x}, \alpha_s, b) = \text{sign} \left(\sum_{s=1}^S y_s \alpha_s (\langle \mathbf{x}, \mathbf{x}_s \rangle) + b \right) \quad , \quad (2.14)$$

where S is the number of support vectors and \mathbf{x}_s describes the support vectors themselves [7].

Applying different conditions while solving the optimization problem, a noticeable aspect appears: the hyperplane is only determined by the support vectors and the complexity of the SVM is therefore only determined by the number of support vectors and not by the dimensionality of the feature space. Adding data points has no impact on the separating hyperplane as long as they do not lead to a smaller margin (see [8] for details).

In case the data set contains overlapping classes, it is often recommendable to introduce a soft margin. This is to accept some classification errors. Therefore a nonnegative slack variable ξ_i is introduced. In order to get a maximal margin despite of the possible errors a parameter C which has to be tuned by the user is created. It provides information about a trade-off between maximization of the margin and minimization of the classification error. Including those two additional parameters the optimization problem 2.17 and equation 2.18 can be written as

$$\min_{\mathbf{w}, b} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i \xi_i \right\} \quad (2.15)$$

with subject to

$$y_i \cdot (\langle \mathbf{x}_i, \mathbf{w} \rangle + b) \geq 1 - \xi_i \quad \text{over all } i, \text{ for } \xi_i \geq 0 . \quad (2.16)$$

ξ_i can be seen as a measure of the classification error which assumes a value between zero and one for the case the object lies within the margin but is still classified correctly and does not produce a classification error. If $\xi_i > 1$ the instance is misclassified [9]. An example is shown in figure 2.14.

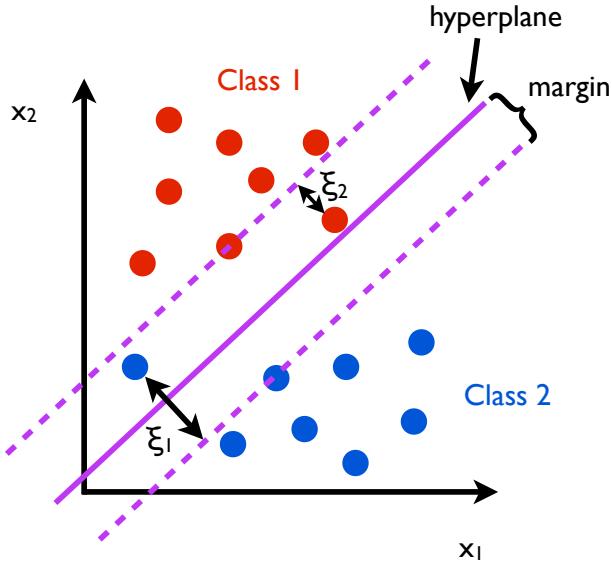


Figure 2.14. Example data points, separating hyperplane with margin and possible errors ξ_1 and ξ_2 . ξ_1 leads to a classification error ($\xi_1 > 1$) while ξ_2 still leads to a correct classification ($0 < \xi_2 < 1$).

2.2.1.2 The nonlinear separation

If the instances are not linearly separable, it may become difficult to create a separating curve in the input space. The SVM solves this problem by mapping the input vectors \mathbf{x}_i to a high dimensional (sometimes even infinite) feature space using a nonlinear transformation $\phi(\mathbf{x}_i)$. The transformation to higher dimensions simplifies the nonlinear problem to a linear case in the feature space (see figure 2.15).

The optimization problem is then modified to the following term:

$$\min_{\mathbf{w}, b} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 \right\} \quad (2.17)$$

with subject to

$$y_i \cdot (\langle \phi(\mathbf{x}_i), \mathbf{w} \rangle + b) \geq 1 \quad \text{over all } i . \quad (2.18)$$

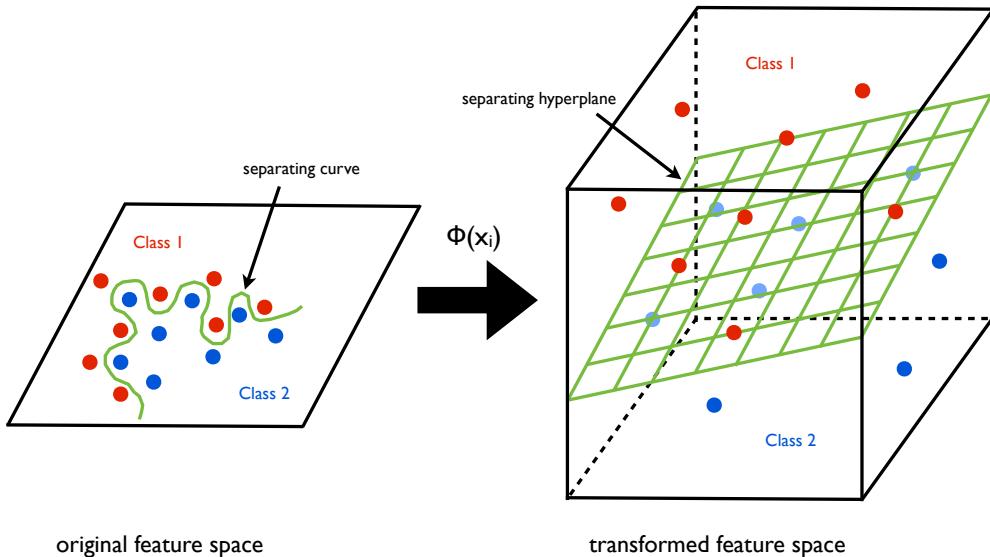


Figure 2.15. Schematic structure of principle of SVMs for nonlinear case

The transformation is achieved by using a kernel function of the form $K(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$. In this case the Gaussian radial basis function (RBF) kernel which is given by

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right) \quad (2.19)$$

is used [1]. The hyperplane is then calculated in the high dimensional feature space. The parameter σ has to be chosen by the user so that an optimal combination of σ and C is found concerning the correct rate by testing every possible combination of those two parameters. The general functionality of the SVM for a nonlinear separable data set is presented schematically in figure 2.15.

2.2.2 Receiver-operating characteristic (ROC)

The receiver-operating characteristic (ROC) is a method to visualize and evaluate a classifier's performance with respect to a single feature. It was actually developed in World War II to check how far radar equipment was able to separate noise from signals clearly originating from enemy planes. In our case, the ROC is applied to evaluate the performance of a feature in a classification task, so that one can make a point about the strength of the single features.

The classification results can be evaluated using four possible outcomes. The instance may be classified as true positive (TP) if a positive event is correctly classified as positive. False positive (FP) is the outcome if the negative instance is mistakenly classified as positive. True negatives (TN) represent examples where the instance is correctly classified as negative. False negatives (FN) are incorrectly categorized as negatives. They are often presented in a confusion matrix (figure 2.16).

		<u>true class</u>	
		positive	negative
<u>assigned class</u>	positive	TP	FP
	negative	FN	TN

Figure 2.16. Confusion matrix

These results are used to build specific statistical values. In case of ROC, the sensitivity (2.20) and specificity (2.21) are of importance as they are calculated and used to extract the ROC plot.

$$Sensitivity = \frac{TP}{TP + FN} \quad (2.20)$$

$$Specificity = \frac{TN}{TN + FP} \quad (2.21)$$

The sensitivity can also be seen as true positive rate, i.e. the number of TP divided by the total number of actual positives, as well as the specificity can be seen as true negative rate respectively number of TN divided by the total amount of actual negatives. The term 1-specificity is consequentially seen as false positive rate, or in other words the number of FP divided by the total number of actual negatives.

For the sake of completeness, the correct rate (CR), the positive predictive value (PPV) and negative predictive value (NPV) are also introduced here. The CR provides information about the number of correctly classified beats in relation to the total number of beats. The PPV declares how many of the positive classified instances are really positive. In contrast, the NPV gives the relation of all correctly classified negative objects to all objects classified as negative.

$$CR = \frac{TP + TN}{TP + FP + TN + FN} \quad (2.22)$$

$$PPV = \frac{TP}{TP + FP} \quad (2.23)$$

$$NPV = \frac{TN}{TN + FN} \quad (2.24)$$

To create an ROC plot, a threshold, which divides the underlying set into two subsets, is shifted over the whole range of the chosen feature. For classifiers using multiple features, the analysis has to be done individually for every attribute. Every instance of one side of the split is classified as positive, all the rest on the other side as negative. The sensitivity and specificity are computed for every threshold and lead to as many data points as tested thresholds in a two-dimensional coordinate system. One can find the sensitivity on the y-axis while 1-specificity is plotted on the x-axis. In other words, the true positive rate is plotted versus the false positive rate of the classifier for the chosen feature. This implies that a discrete classifier (i.e. with a fixed threshold) only produces one single point in the ROC space [10].

Characteristically, a bended curve from the lower left corner to the upper right corner results (see figure 2.17). The optimum that can be achieved is a step function. In worst case, a diagonal line from down left to top right comes out where besides the number of TP, the number of FP equally increases.

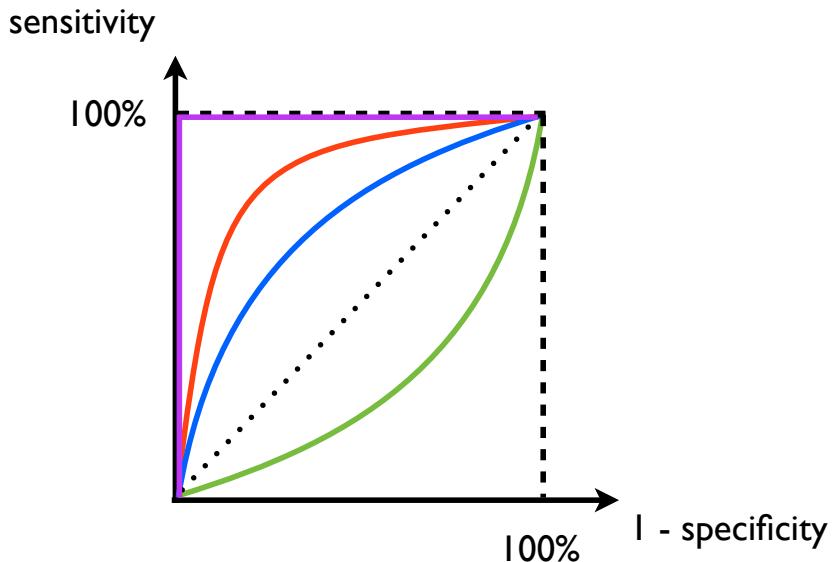


Figure 2.17. Typical ROC plots are shown here. The optimum is given by the violet curve while the dotted diagonal line stands for random classification. The red curve represents a better performance than the blue one. The green one has an AUC of < 0,5 and results in the blue one by logical inversion.

For comparing different analyses it is often desirable to introduce a single value as a measure of the performance of the feature. Therefore, the area under the ROC curve (AUC) is calculated. Trapezoidal integration can be used for approximation ([11]). Consequentially, a maximal value of 1 respectively minimal value of 0.5 is possible. Curves lying beneath the diagonal line and so resulting in an AUC of smaller than 0.5 can be mirrored at the diagonal line by inverting the logical assignments, i.e. positives switch to be negatives and

vice versa. As a result this leads to AUC values of $AUC_{inverted} = 1 - AUC_{normal}$ (also see example in figure 2.2.2).

The key advantage of the ROC is the fact that the analysis is independent of the real feature value, it gives the user an information about the classifier's performance for convenient choice of a feature value. One has to notice that the ROC analysis investigates different features separately, so that you get no information about the relationship among them.

2.2.3 Classification tree

Decision trees play an important role in the domain of decision theory and machine learning. In case of classification they are also called classification trees. Their main advantages are a robustness against outliers and their easily understandable decisions [12]. They are based on the principle of partitioning an underlying set into subsets which are associated with a specific class. They consist of decision nodes which are interconnected by so called branches or links. The nodes are subdivided into one root node, various sub-nodes or connecting nodes and different terminal nodes which are allocated to a certain class. The root node always contains all instances, the subsets put together result in the underlying set again (in case of hard decision limits). If two sub-nodes arise from every node, one speaks of a binary tree. Every path through the tree can be seen as a conjunction of feature tests, while the tree as a whole stands for a disjunction of these conjunctions [12]. Every instance is assigned to either the left or the right following node by a certain decision criterion and so is conducted through the tree structure until it stops its way at a terminal node. Therefore, a criterion when to stop splitting is needed. The classification is thus done in several steps and depends on the terminal node the instance is assigned to. The splits dividing the underlying set are declared by different criteria. The goal of the split is to increase the purity of the set in the following node. A schematic structure is presented in figure 2.18. The node t_1 represents the root node while t_2 and t_3 are connecting nodes and $t_4 - t_7$ terminal nodes in this case. The instances of the decision nodes $t_1 - t_3$ are separated by the splits s_i concerning feature x_i . The structure can be extended to the desired depth by continuing this scheme in the same line.

The scheme can easily be understood by regarding figure 2.19 that visualizes the splitting process. One has to notice that in this example, t_3 is also separated by a split based on feature x_2 . It is also possible to split t_3 with any other feature x_i . Features may also be selected more than once, if they yield the best separation for several nodes. In this case, t_4 and t_7 represent the red class while t_5 and t_6 are assigned to the blue class.

The construction of a classification tree fundamentally proceeds by splitting a feature space into different areas. The steps of tree-constructing are composed of three elements:

1. The selection of the splits
2. The criterion when to stop splitting
3. The assignments which class should be represented by which terminal node

The class assignment is thus the simplest task, the hard work has to be done in solving the problems of the first two steps [13]. The basic concept of choosing a split is to create "purer" subsets as the parental set [13]. The stop criterion is mentioned here for the sake of completeness to illustrate the way of constructing but will not be discussed in this thesis. For detailed information see [13] or [14].

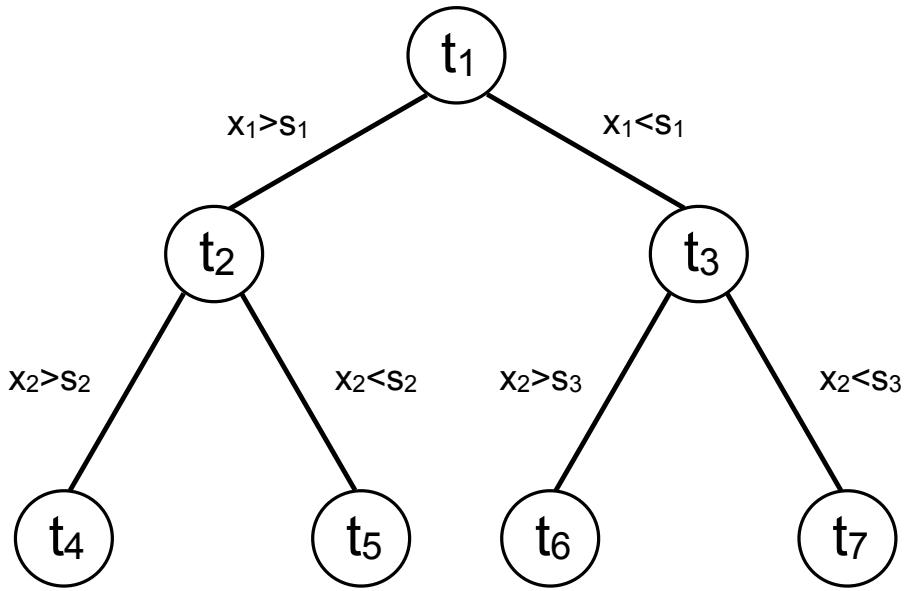


Figure 2.18. Schematic structure of classification tree

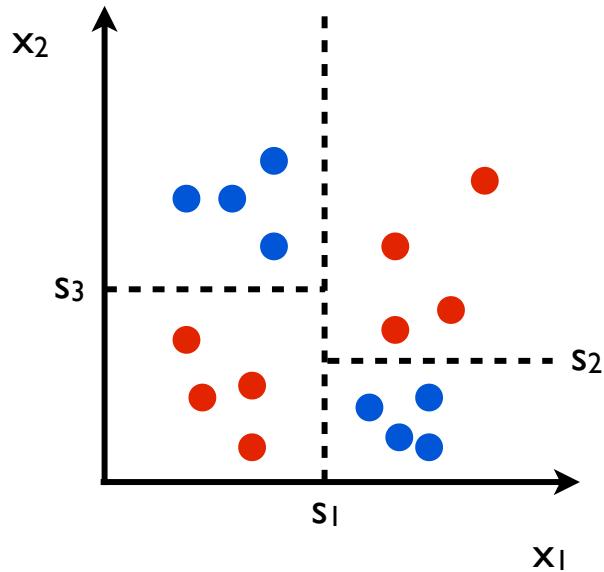


Figure 2.19. Visualization of splitting process of classification tree. t_4 and t_7 stand for the red class, t_5 and t_6 for the blue class (cf. figure 2.18).

Two of the commonly used splitting criteria are described in the following sections since they are used later on in this thesis.

2.2.4 Gini Diversity Index (GDI)

The Gini Diversity Index (GDI) is a widely used split criterion in decision trees. It is defined for a node t_i as [13]

$$GDI(t_i) = \sum_{j \neq k} p(w_j|t_i)p(w_k|t_i) . \quad (2.25)$$

The indices j and k represent different classes while $p(w_j|t_i)$ stands for the conditional probability of an object assigned to class j under the condition that is located in node t_i . Equation 2.25 can also be transformed to the following term where J represents the number of classes [13]:

$$GDI(t_i) = \sum_{j=1}^J 1 - p^2(w_j|t_i) . \quad (2.26)$$

To create now the best split value for the set to divide, every attribute, which is represented by the attribute vector \mathbf{x}_i , is regarded separately. The N feature values are sorted in ascending order as shown in 2.27.

$$\mathbf{x}_i = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} \quad \text{with} \quad x_1 \leq x_2 \leq \dots \leq x_N \quad (2.27)$$

After that, every possible separation of the set is tested, i.e. a split is set between two feature values, calculated by the formula of the arithmetic mean:

$$s_i = x_i + \frac{x_{i+1} - x_i}{2} . \quad (2.28)$$

As a consequence the N -dimensional feature vector leads to a $(N - 1)$ -dimensional split vector. Once all splits s_i are built, the GDI for every split value is calculated. The split corresponding to the smallest GDI is chosen as best split for the examined attribute. To get the best split for the node, it is necessary to repeat the procedure calculating a possible best split for every feature. Subsequently, the choice of the best splitting feature for the node is made by reference to the feature which yields the minimal GDI.

Exemplarily applied on a self-created test data set, the best split calculated via GDI would be placed as one can see in figure 2.20. On the left side, the split achieves a completely pure set and results in a GDI of zero. The instances on the right side are almost of one class, only two ectopic beats are still mixed in. Therefore, a GDI of 0.0384 follows. The sum of both values leads to a total GDI of 0.0384 for this split.

2.2.5 Information Gain Ratio (IGR)

The information gain ratio (IGR) is an entropy-based criterion for attribute selection. It is composed of two different values: the information gain (IG) and the split information (SI) [14]. The entropy as basic principle is calculated in the following way:

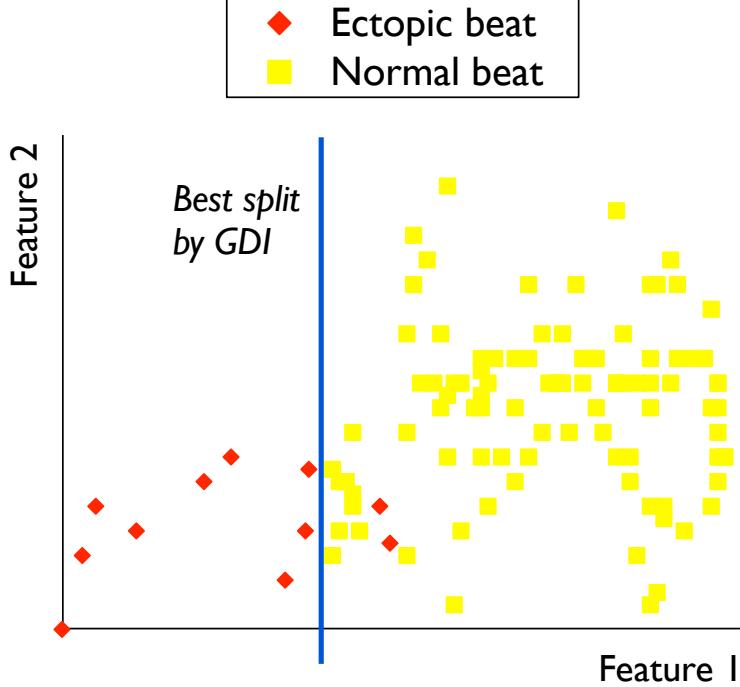


Figure 2.20. A self-created test data set is shown in this figure. The blue line indicates the best split that results of the GDI calculation.

$$H(W) = \sum_{j=1}^J -p_j \cdot \log_2(p_j) \quad , \quad (2.29)$$

where W represents a specific set containing J values and p_j is the probability for an object of W to be part of class j [15]. After partitioning W by a split attaining the values of feature X the new information content is computed by

$$H_X(W) = \sum_{k=1}^K \frac{|W_k|}{|W|} \cdot H(W_k) \quad . \quad (2.30)$$

W_k stands here for the objects of W that attain the value k of feature X while $|W|$ respectively $|W_k|$ describe the cardinality of the particular sets. Feature X consists of K possible values. Since the IG is defined by

$$IG(X) = H(W) - H_X(W) \quad , \quad (2.31)$$

it can be seen as a measurement of the information gained by splitting W using feature X [14].

As a result of the fact that the IG show a strong bias in favor of features with many outcomes, a type of a normalization is introduced to minimize the impact of that bias:

$$SI(X) = \sum_{j=1}^J -\frac{|W_j|}{|W|} \cdot \log_2 \left(\frac{|W_j|}{|W|} \right) \quad . \quad (2.32)$$

In contrast to the IG, which estimates the entropy reduction, the SI depicts the outcoming entropy generated by the split.

Once the IG and SI are calculated, the IGR can be specified by

$$IGR(X) = \frac{IG(X)}{SI(X)} . \quad (2.33)$$

In the end, the feature that achieves the highest IGR is chosen.

To get around the problem that the IGR gets unstable as a consequence of a very small SI, only tests with an IG above the average IG over all tests are taken. Its main conclusion is to give a proportion of useful information for the classification task which is produced by a split on W choosing the values of attribute X as separating borders [14].

2.2.6 Basics in statistics

Since a feature analysis using probabilistic parameters is implemented, some fundamentals are introduced in this section.

A random variable $X(\xi)$ is defined as a real function based on the sample space of a random experiment. It maps the elementary events ξ to a real number $x(\xi)$. For reasons of simplification and better readability, ξ is left out. To keep in mind that $X(\xi)$ represents a function and not just a number, it is written with a cursive capital X . Its realizations are written with a cursive small x . X is characterized by its probability density function $f_X(x)$. It gives the probability per interval width Δx that X lies in that small interval around x [16]:

$$f_X(x) \approx \frac{P(x - \frac{\Delta x}{2} < X \leq x + \frac{\Delta x}{2})}{\Delta x} . \quad (2.34)$$

Furthermore, $f_X(x)$ possesses the following properties:

$$f_X(x) \geq 0 , \quad (2.35)$$

$$\int_{-\infty}^{\infty} f_X(x) dx = 1 , \quad (2.36)$$

$$\int_a^b f_X(x) dx = P(a < X \leq b) . \quad (2.37)$$

The first moment or mean of a random variable X is defined as

$$\mu_X = E\{X\} = \int_{-\infty}^{\infty} x f_X(x) dx \quad (2.38)$$

and describes the center of mass of $f_X(x)$.

In general, the m -th central moment of X is calculated by

$$E\{(x - E\{X\})^m\} = \int_{-\infty}^{\infty} (x - E\{X\})^m f_X(x) dx . \quad (2.39)$$

Consequentially the second central moment, named variance, is given by

$$\sigma_X^2 = E\{(x - E\{X\})^2\} = \int_{-\infty}^{\infty} (x - E\{X\})^2 f_X(x) dx . \quad (2.40)$$

The variance provides information about the width of the probability density function of X and can be seen as a measure of dispersion. The wider $f_X(x)$, the higher is σ_X^2 . The square root of the variance results in the standard deviation:

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{E\{(x - E\{X\})^2\}} . \quad (2.41)$$

Since they are important for this thesis, statistic values based on the third and fourth central moments are also introduced. The skewness, built of the third central moment in relation to the cube of the standard deviation, is a measure of the symmetry of the distribution:

$$\varrho_X = \frac{E\{(x - E\{X\})^3\}}{\sigma_X^3} . \quad (2.42)$$

A skewness greater than zero indicates a distribution that is inclined to the right, a negative skewness indicates a distribution that is biased to the left. For a completely symmetric distribution around zero, one has $\varrho_X = 0$. The other way round cannot be applied generally, for a symmetric distribution the case $\varrho_X = 0$ is not always valid.

To complete this section, a statistic value related to the fourth central moment is explained. The quotient of the fourth central moment and the standard deviation raised to the power of four is called the kurtosis:

$$\kappa_X = \frac{E\{(x - E\{X\})^4\}}{\sigma_X^4} . \quad (2.43)$$

It gives a description of the peakness of $f_X(x)$. $\kappa_X < 0$ characterizes a flat (platykurtic) distribution, $\kappa_X > 0$ is a sign for a steep (leptokurtic) distribution. Additionally, it is a measure for the discrepancy between an unimodal and a Gaussian distribution. The term

$$\varepsilon_X = \frac{E\{(x - E\{X\})^4\}}{\sigma_X^4} - 3 \quad (2.44)$$

is known as excess kurtosis and, caused by the added subtraction of the kurtosis of a Gaussian distribution, can be seen as a kind of normalized kurtosis so that the Gaussian distribution has an excess kurtosis of zero [15, 16].

3

State of the art

This chapter is intended to give the reader an overview about other research projects and publications concerning classification tasks and their evaluation. The problem of feature evaluation is not often the main goal of researches, most of the time the task is to evaluate the whole classifier itself.

Concerning detection and classification of heartbeats in ECG signals, two main references are important to mention.

Krasteva et al. developed a method to automatically detect and classify QRS complexes in ECG recordings ([17]). The process of ECG analysis in their work consists of preprocessing filtration, heartbeat detection and classification. For this task, they used different rhythmical and morphological features. They are mainly based on RR intervals and the width of the QRS complex. The classification of the beats is thus done by choosing class-specific threshold values of the features. Their algorithms resulted a sensitivity of 92.2% and a specificity of 96%.

Lenis also detected and classified the heartbeats. A number of 55 different rhythmical and morphological features were used to distinguish between normal beats, ventricular and supraventricular extrasystoles ([1]). As classifier, he used an SVM to assign every beat to one of these three classes. He trained one SVM using all 55 developed features and a second one only using 19 scores from a previously conducted PCA from the original feature space. The first one achieved a correct rate of 98.21%, a sensitivity of 92.46% and a PPV of 94.52%. The second one led to a correct rate of 97.87%, a sensitivity of 90.59% and a PPV of 93.3%.

The following publications are related to the methods used in this thesis.

Bradley investigated the use of the AUC in the evaluation of machine learning algorithms ([11]). Zweig and Campbell published a paper discussing the ROC as a fundamental evaluation tool in clinical medicine ([18]). From their point of view, the method plays an important role in making clinical decisions as it is important to know how accurate the diagnosing performance of clinical tests are. Both groups came to the conclusion that the ROC gives an unbiased view on the accuracy of the classifier.

Breiman et al. developed a classification tree that uses the GDI to select the attributes which are used to separate the classes ([13]). This method chooses the attributes first that achieve the highest gain of purity. In the course of time, it has found a use in many classification tree applications.

In his book "C4.5: Programs for Machine Learning", Quinlan et al. also explained an

algorithm to create decision trees ([14]). They introduced the IGR as their favorite tool to choose features in classification tasks.

4

Feature evaluation

The main task of this thesis is to evaluate the features for a QRS complex classification system. In this case, an SVM was used as classifier. The features on which the classification is based have been introduced by G. Lenis in his work [1]. One can find a list of all features in appendix A. His algorithms create 55 features for every QRS complex that are given to the SVM to distinguish between normal beats, ventricular and supraventricular extrasystoles in ECG signals. The goal is to reduce the number of features that the SVM uses to classify. The assumption is that the actual number of 55 features for every heart beat confuses the classifier and leads to worse classification results. The computational time is another aspect which is influenced by the number of attributes. The more features have to be calculated, the more time is needed by the SVM to analyze the data. This plays an important role in real-world applications. To find out which features yield much information and which features can be left out without worsening the classification results, all features are tested with different evaluation methods.

After that, a quality ranking of the features is created. This is to bundle all the information of the different analyses and to decide which features are appropriate to train the new SVM in the end. To get the features ranked, a points system is introduced. As there are 55 features in total, the best feature of the respective analysis gains 55 points, the worst 1 point. For every method, one ranking is created. Finally, the different rankings are weighted to one total ranking on which the feature choice was based. A point value of 55 in average would represent a feature that "wins" all analyses, a point value of 1 would indicate a feature that comes in last place in every single ranking. The process of feature evaluation is presented in this chapter.

It is important to mention that for the analysis of the rhythmical features (1 to 20) the classes of ventricular and supraventricular extrasystoles are merged together and therefore separated from the class of normal beats. In case of the morphological features (21 to 55), the classes of normal beats and of supraventricular extrasystoles are put together as their QRS complexes look very similar (see section 2.1.4). They get divided from the class of ventricular extrasystoles. Furthermore, during the process of feature calculation, the positions of the R peaks are taken from annotations by physicians to eliminate errors in the peak detection.

4.1 Evaluation using ROC

First, an evaluation using the ROC is conducted. As described previously in section 2.2.2, the ROC is actually a method to evaluate a classifier's performance. In this thesis, it is used to investigate the goodness of the features. Therefore, the analysis is conducted 55 times, one time for every feature. As they are regarded separately, one can make a point about the strength of the single features. Since the result is one AUC value for every feature, a first ranking can be created (see table 4.1). It shows that the features 48 (skewness of amplitude distribution), 33 (maximum of cross correlation) and 31 (linear correlation coefficient) are the top three winners of this evaluation. Since their AUC values of better than 0.95 are very good, these attributes should lead to a good classification. The features 43 (first coefficient of linear combination of Hermite basis functions), 41 (center of mass for activity interval) and 52 (first PCA score) yield here the worst results. Their AUC values of nearly 0.5 indicate a quite random classification ability.

Also if they are plotted against each other, the two worst features do not seem to be easily separable (see figure 4.2). In contrast, the best two features are plotted against each other in figure 4.1. One can see that a good separability is given. The ROC curve of the best feature (feature 48) is also shown in figure 4.3. An AUC of 0.9607 is achieved here. The worst AUC of 0.5041 comes out for feature 43. Its corresponding ROC curve is plotted in figure 4.4.

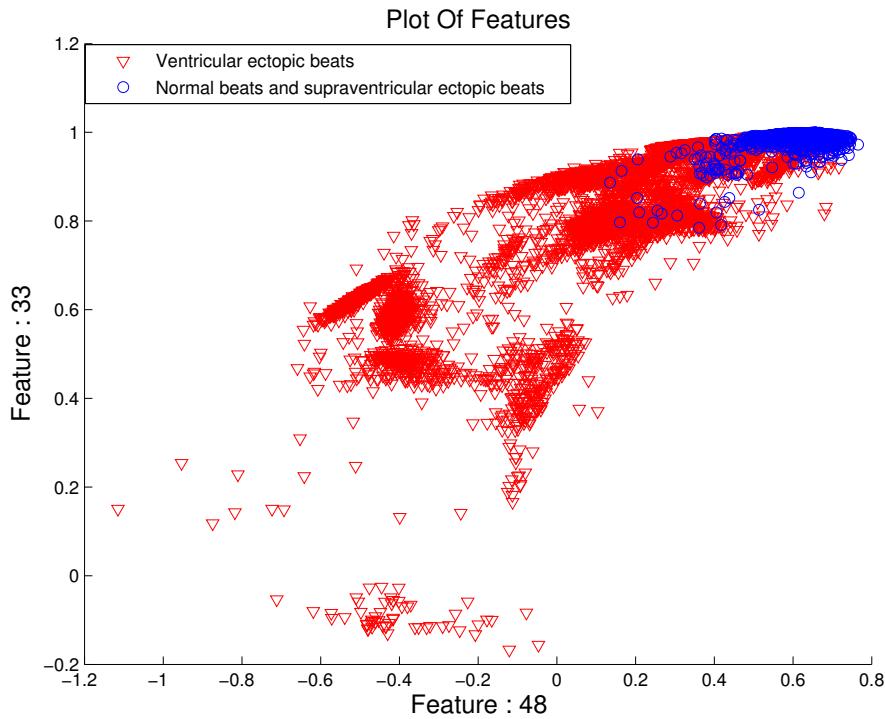


Figure 4.1. The best two features of the ROC analysis are plotted in this figure. One can separate the classes relatively clear.

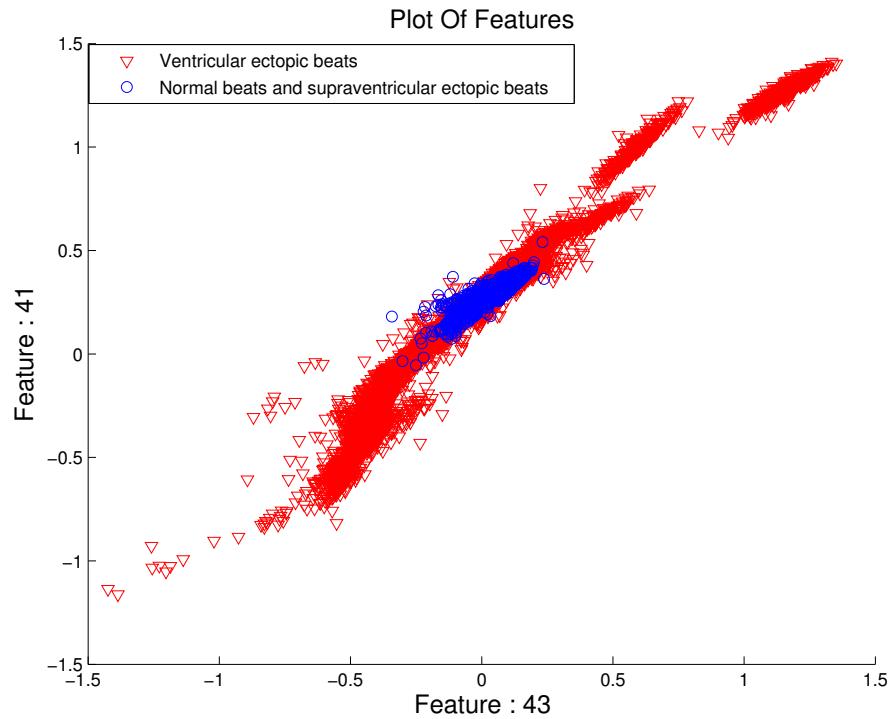


Figure 4.2. The worst two features of the ROC analysis are plotted in this figure. The classes are overlapping and therefore not separable.

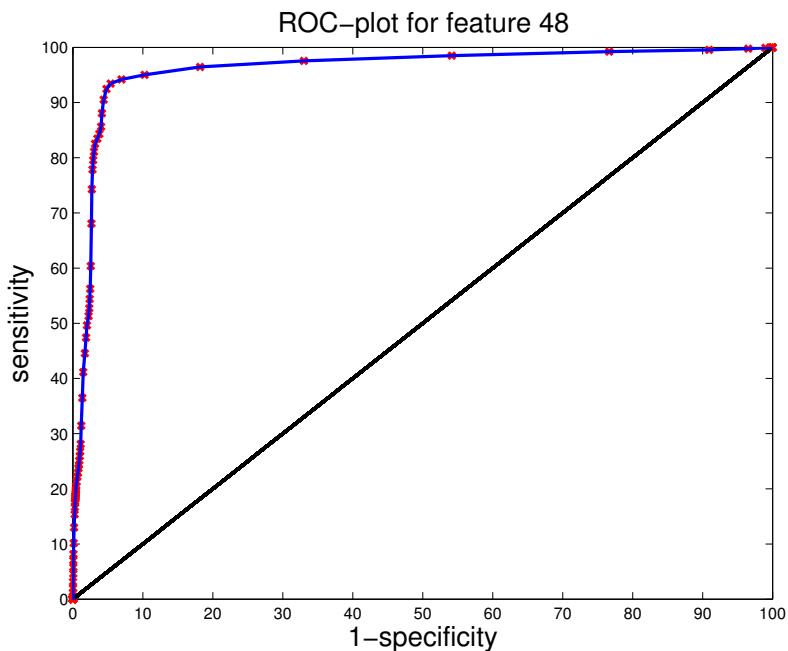


Figure 4.3. The ROC curve of the best feature (feature 48) of the ROC analysis is plotted concerning the separation of ventricular ectopic beats from normal and supraventricular ectopic beats. Its shape is close to the optimal step function. The achieved AUC value is 0.9607

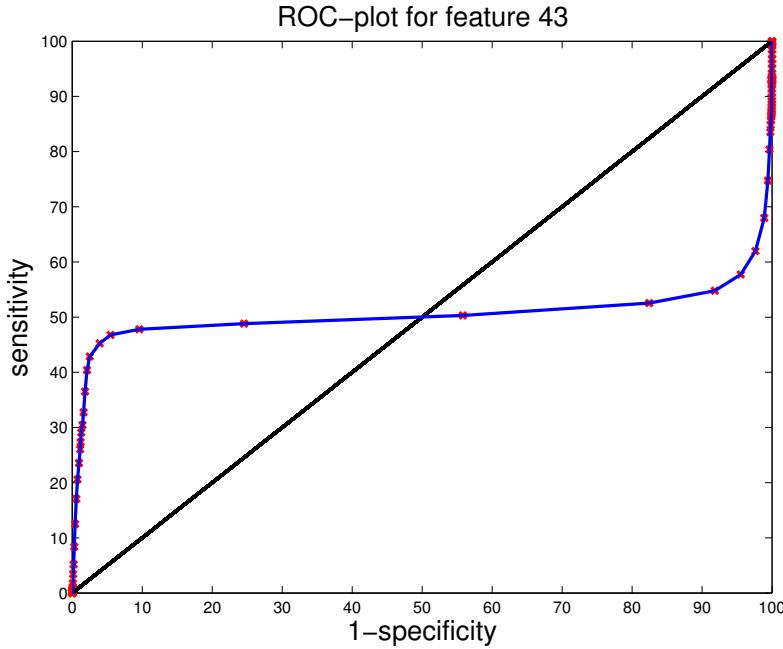


Figure 4.4. The ROC curve of the worst feature (feature 43) of the ROC analysis is plotted. Again ventricular ectopic beats are separated from normal beats and supraventricular ectopic beats. The corresponding AUC value is 0.5041 which is indicative of a nearly random classification.

4.2 Evaluation using GDI

Commonly used as split criterion in decision trees, the GDI separates the underlying set trying to increase the purity of the following subsets. A classification tree would choose the feature first which separates the underlying set the best. Therefore, the GDI analysis provides information about the strength of the features.

The outcomes of the feature evaluation using the GDI are listed in table 4.2.

One can see that feature 31, which stands for the linear correlation coefficient, yields the best results. The skewness of the amplitude distribution (feature 48) and the maximum of the cross correlation (feature 33) follow on places two and three. The last three attributes are in this case the quotient of the RR intervals of the past beat (feature 12), the quotient of the instantaneous heart rates of the past beat (feature 14) and the RR interval to the next beat (feature 2). It is noticeable that the first three features are equivalent to the winners of the ROC evaluation. The two best features are again plotted against each other in figure 4.5. The two worst ones can be seen in figure 4.6.

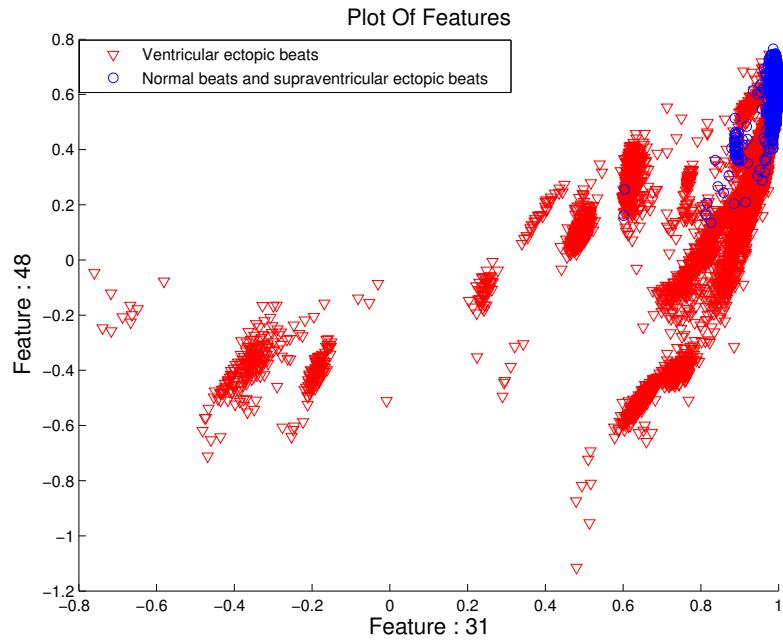


Figure 4.5. The best two features of the GDI analysis are plotted in this figure. A good separability is again noticeable.

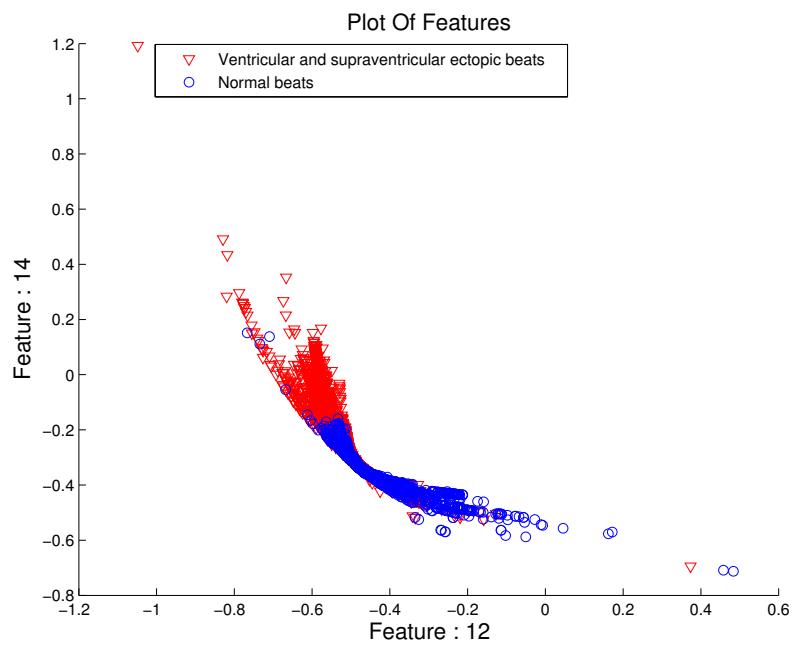


Figure 4.6. The worst two features of the GDI analysis are plotted in this figure. It is not possible to create a border that separates the classes well.

4.3 Evaluation using IGR

The IGR analysis is composed of the calculation of the IG and the SI (see section 2.2.5). Thereby, the IG has the problem that it induces a bias in favor of features with many outcomes. To avoid this problem, the SI acts as a kind of normalization. Additionally, to reduce the impact of that bias, an interval nesting is introduced during this work. After normalizing every feature by subtracting the mean and dividing by the standard deviation, its whole range is apportioned into 101 intervals (see figure 4.7). The number of 101 is chosen so that there are 50 intervals on the positive and 50 intervals on the negative side as well as one interval around zero.

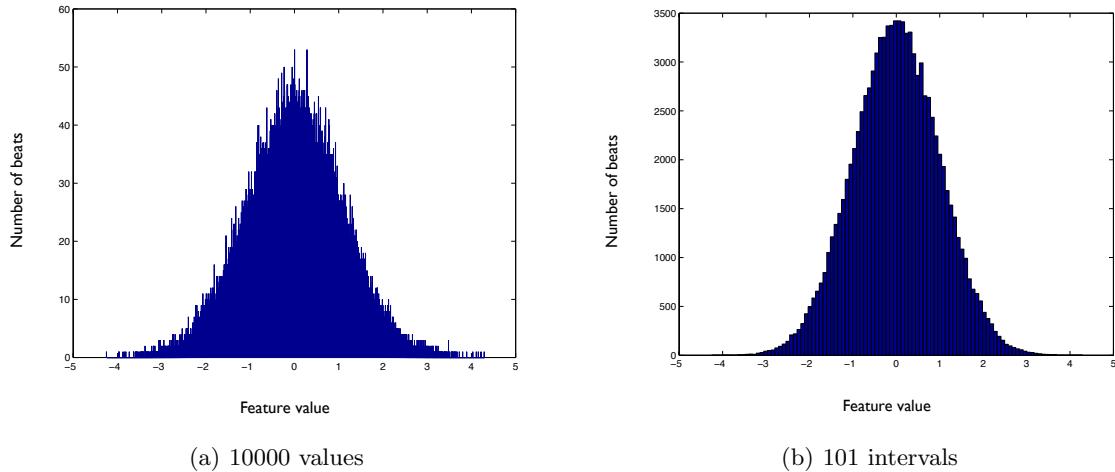


Figure 4.7. Since the IGR is used as evaluation tool, an interval nesting is conducted. After a normalization, all occurring feature values are apportioned into 101 intervals. This is to avoid a bias of the IGR to features with many outcomes and therefore achieve a better stratification in the feature ranking.

The evaluation by the IGR yields as result the ranking shown in table 4.3. The IGR leads to similar results as the previous two analyses. The best features are in this case the linear correlation coefficient (feature 31) and the maximum of the cross correlation (feature 33). Additionally, feature 44, which represents the second coefficient of the linear combination of the Hermite basis functions, seems to be a very useful attribute. The worst results of the IGR evaluation come out for the third and fourth PCA score (features 54 and 55) and the skewness of the QRS complex (feature 36).

As it is a quality criterion of good features in the IGR analysis to have only one class in most of the intervals, two histograms of the best and worst feature of the IGR analysis are presented in figure 4.8 and figure 4.9. The normal and ectopic beats are in these plots marked in different colors. Red stands for the ectopic beats, blue for the normal beats. On the one hand, it is striking that for the best feature (feature 31) the feature values for the different classes are located in different ranges. On the other hand, regarding the worst feature in this case (feature 55), most of the beats lie in the same feature value range. This makes the classes almost impossible to separate. This form of presentation seems to be very appropriate to visualize and emphasize the results of the IGR analysis.

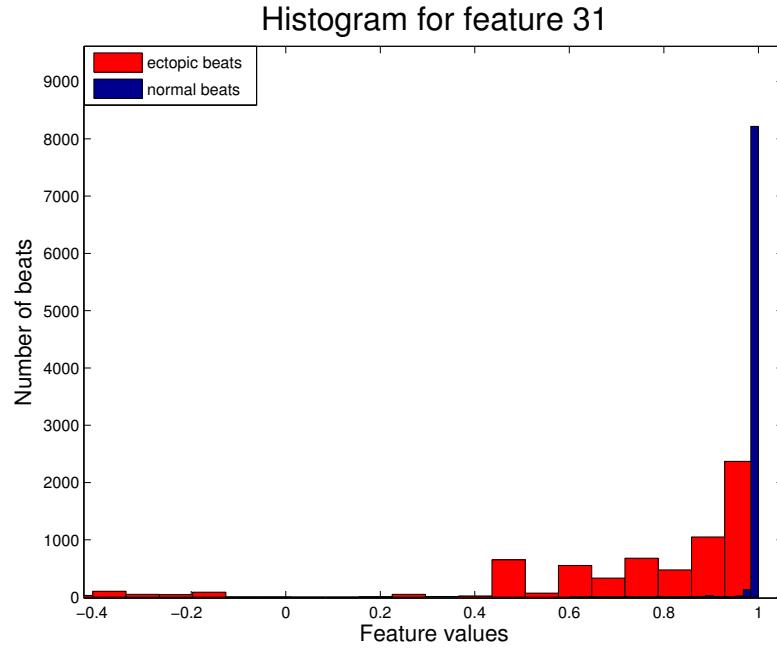


Figure 4.8. The values of the best feature of the IGR analysis are plotted in this histogram. The values of the different classes strongly differ from each other. This ensures a good separability. The red bins correspond to the ectopic beats, the blue bins to the normal beats.

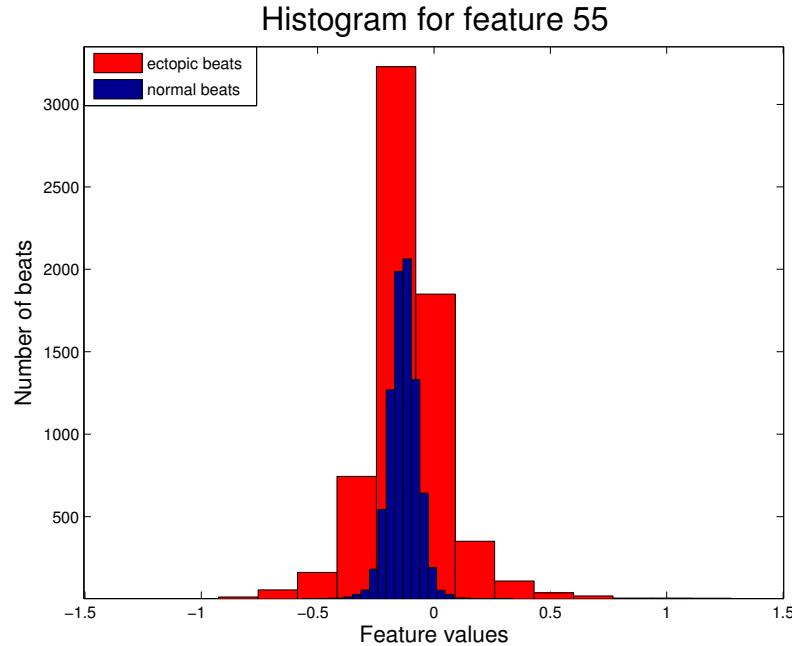


Figure 4.9. The values of the worst feature of the IGR analysis are plotted in this histogram. The red bins represent the ectopic beats, the blue bins the normal beats. One can see that the classes are located in the same intervals and therefore hard to separate.

Table 4.1. The results of the ROC analysis are presented in this table. The features and their corresponding AUC value as well as the achieved points according to the introduced points system are listed.

Achieved points	AUC value	Feature number
1	0.5041	43
2	0.5051	41
3	0.5109	52
4	0.5140	27
5	0.5150	55
6	0.5346	23
7	0.5372	54
8	0.5442	22
9	0.5578	29
10	0.5924	53
11	0.6022	26
12	0.6056	36
13	0.6107	39
14	0.6515	51
15	0.6527	32
16	0.6585	46
17	0.6850	37
18	0.6993	45
19	0.7133	28
20	0.7208	25
21	0.7489	40
22	0.7831	24
23	0.7848	4
24	0.7882	34
25	0.7886	2
26	0.7909	47
27	0.8221	30
28	0.8299	35
29	0.8311	38
30	0.8691	21
31	0.8705	12
32	0.8744	14
33	0.8786	15
34	0.8793	16
35	0.8827	6
36	0.8844	13
37	0.8901	11
38	0.8913	5
39	0.8979	19
40	0.8990	18
41	0.9037	20
42	0.9123	1
43	0.9137	3
44	0.9161	42
45	0.9205	17

Achieved points	AUC value	Feature number
46	0.9244	9
47	0.9245	10
48	0.9272	44
49	0.9319	7
50	0.9323	8
51	0.9438	49
52	0.9440	50
53	0.9533	31
54	0.9558	33
55	0.9607	48

Table 4.2. The results of the GDI analysis are presented in this table. The features and their corresponding GDI value as well as the achieved points according to the introduced points system are listed.

Achieved points	GDI value	Feature number
1	0.4998	12
2	0.4987	14
3	0.4976	2
4	0.4963	4
5	0.4951	1
6	0.4925	26
7	0.4920	37
8	0.4916	36
9	0.4915	55
10	0.4914	20
11	0.4898	54
12	0.4888	39
13	0.4878	40
14	0.4871	25
15	0.4863	51
16	0.4833	35
17	0.4781	23
18	0.4753	22
19	0.4749	3
20	0.4733	53
21	0.4696	24
22	0.4644	29
23	0.4606	13
24	0.4589	19
25	0.4549	18
26	0.4529	46
27	0.4459	43
28	0.4444	9
29	0.4444	15
30	0.4435	27

Achieved points	GDI value	Feature number
31	0.4430	6
32	0.4414	16
33	0.4408	41
34	0.4389	5
35	0.4378	32
36	0.4378	52
37	0.4371	28
38	0.4361	8
39	0.4272	11
40	0.4210	47
41	0.4207	21
42	0.4179	10
43	0.4172	34
44	0.4129	7
45	0.4124	17
46	0.3862	30
47	0.3816	45
48	0.3472	38
49	0.2522	44
50	0.2127	42
51	0.1664	50
52	0.1633	49
53	0.1386	33
54	0.1216	48
55	0.0994	31

Table 4.3. The results of the IGR analysis are presented in this table. The features and their corresponding IGR value as well as the achieved points according to the introduced points system are listed.

Achieved points	IGR value	Feature number
1	0.0266	55
2	0.0344	54
3	0.0419	36
4	0.0479	37
5	0.0536	26
6	0.0567	25
7	0.0597	40
8	0.0618	51
9	0.0635	53
10	0.0665	23
11	0.0696	39
12	0.0746	2
13	0.0757	4
14	0.0778	12
15	0.0785	24

Achieved points	IGR value	Feature number
16	0.0808	14
17	0.0840	22
18	0.0893	1
19	0.0895	13
20	0.0900	19
21	0.0913	18
22	0.0930	5
23	0.0948	28
24	0.0949	15
25	0.0957	6
26	0.0961	3
27	0.0962	16
28	0.0974	32
29	0.0983	35
30	0.1002	20
31	0.1018	11
32	0.1022	9
33	0.1056	8
34	0.1062	10
35	0.1125	7
36	0.1182	21
37	0.1198	17
38	0.1203	34
39	0.1227	47
40	0.1239	45
41	0.1267	38
42	0.1343	42
43	0.1386	49
44	0.1408	30
45	0.1417	46
46	0.1432	50
47	0.1443	29
48	0.1447	43
49	0.1486	41
50	0.1501	27
51	0.1536	52
52	0.1600	48
53	0.1608	44
54	0.2001	33
55	0.2300	31

4.4 Evaluation using statistical methods

Additionally to the three evaluation methods described above, an analysis based on statistical properties of the features is implemented. It is assumed that the SVM can handle compact and symmetrical distributed attributes easier. This idea will be put to the test later. An optimal case is shown in figure 4.10. The SVM might set hard decision borders at both sides of the distribution to separate those beats from others. The purpose of the investigation of the compactness is to examine if the beats are easily separable. The aim is that the normal beats are concentrated compactly around zero while the ectopic beats may be located somewhere else in the feature space. This would lead to an easily separable case. The fact that the compactness of the normal beats seems to be important is the reason why for this statistical evaluation only normal beats are considered. Furthermore, the error rate of the SVM is proportional to the number of support vectors [8]. The less support vectors there are, the higher is the correct rate. Therefore, it is better to have compact classes that lead to less support vectors. The correspondence is shown in equation 4.1.

$$E\{P(\text{Error})\} \leq \frac{E\{\text{number of support vectors}\}}{\text{total number of training examples}} \quad (4.1)$$

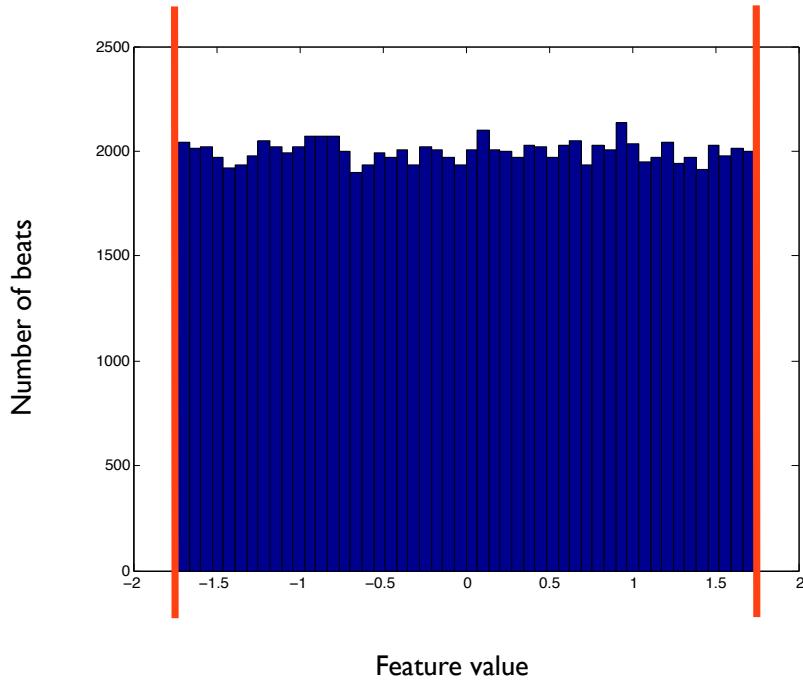


Figure 4.10. An optimal distribution of the normal beats is shown. The SVM could separate the class clearly with the borders displayed in red. The ectopic beats can be located somewhere else in this feature space.

Another example is given in figure 4.11 which shows a three-dimensional histogram of two features. The normal beats of those two example features are distributed following

a uniform distribution. As they lie concentrated in the middle of the feature space, they can easily be separated from other beats using only one radial function in the hyperspace. This small number of support vectors would lead to a higher correct rate (see equation 4.1).

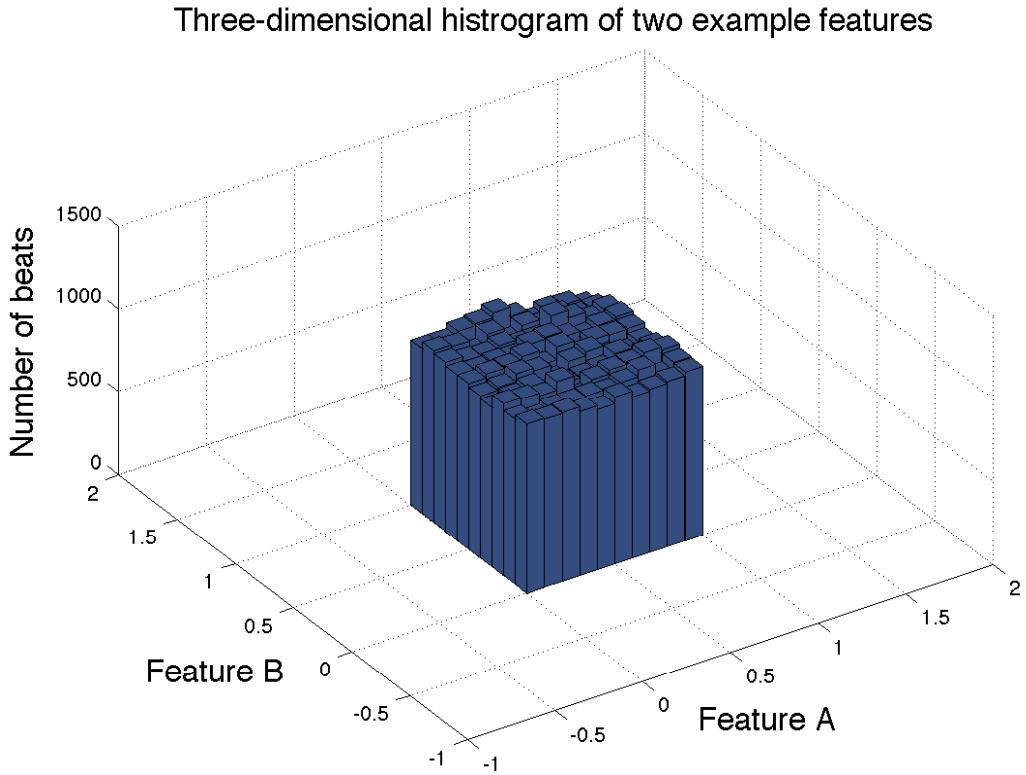


Figure 4.11. The distribution of the normal beats of two equally distributed features are exemplarily given in this histogram. This case would be quite optimal as one could separate those beats in a hyperspace with only one radial function. This would lead to a small number of support vectors and therefore to a low error rate (cf. equation 4.1).

First of all, a normalization of every feature is conducted by subtracting the mean and dividing by the standard deviation:

$$X_{norm} = \frac{x - \mu_X}{\sigma_X} . \quad (4.2)$$

The features are hereby seen as random variables X with realizations x .

In order to examine the symmetry of the features, the skewness is calculated. Also using the points system described previously in the beginning of section 4, a first statistical ranking can be created (see table 4.4). The skewness of a perfectly symmetric distribution takes on the value of zero. The most symmetric distribution of normal beats can be found regarding the mean of the absolute of the Fourier transform between 10Hz to 20Hz (feature 39). But also the values for the attributes second and third PCA score (features

53 and 54) are symmetrically distributed. Whereas, the linear correlation coefficient (feature 31), the maximum of the cross correlation (feature 33) and the second coefficient of the linear combination of the Hermite basis functions (feature 44) show an asymmetrical shape.

In contrast, the calculation of the kurtosis is used to extract information about the peakness of the distribution of the features. It shows how steep or flat the distribution is shaped. As its calculation includes every value to the power of four, the kurtosis is also very sensitive to outliers and rises very fast for great values. A wide distribution that includes many outliers is not desirable. A quite optimal case, as shown in figure 4.10, leads to a kurtosis of about 1.8. The feature ranking built of the results of the kurtosis is given in table 4.5. One can see that the features 24 (maximum of first derivative), 54 (third PCA score) and 16 (quotient of moving average of quotient of RR intervals and quotient of RR intervals) head the ranking. Clearly the worst results show in this case the features 48 (skewness of the amplitude distribution), 33 (maximum of cross correlation) and 31 (linear correlation coefficient).

Finally, a third statistical analysis is introduced. The idea of investigating the compactness of the features is realized by calculating the difference between the quantile at 97.5% and the quantile at 2.5% (see figure 4.12). This is to leave out the extreme outliers but consider 95% of the distribution. The smaller the difference between those two quantiles, the stronger are the feature values concentrated around zero. The results are as presented in table 4.6. The attributes yielding the best outcomes are the features 31 (linear correlation coefficient), 33 (maximum of cross correlation) and 44 (second coefficient of linear combination of Hermite basis functions). Whereas a large difference between the quantiles comes out for the features 5 (difference of two consecutive RR intervals), 14 (quotient of instantaneous heart rates of past beat) and 17 (quotient of quotient of current RR intervals and quotient of past RR intervals).

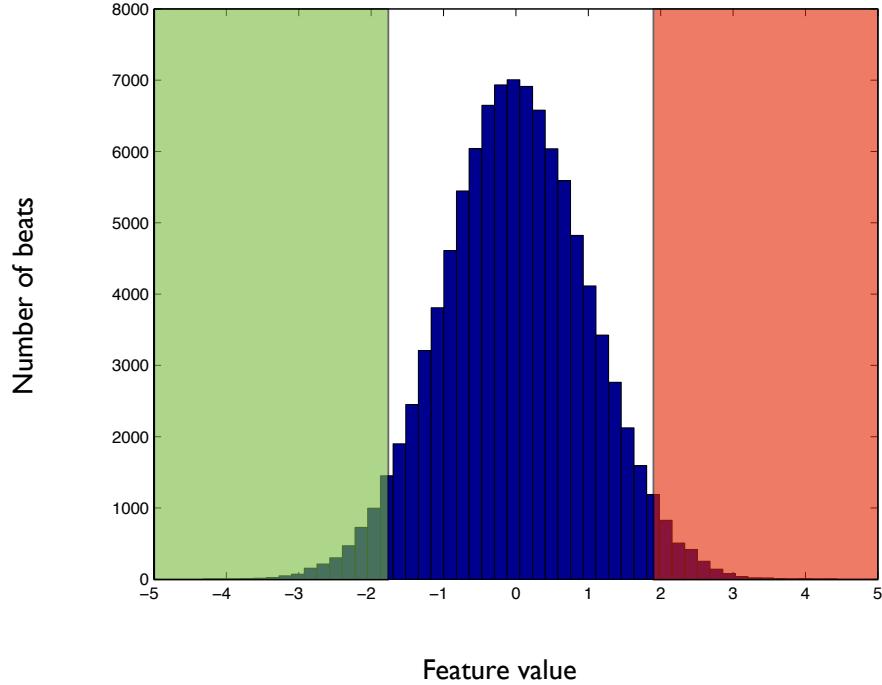


Figure 4.12. The evaluation method of the difference of two quantiles is visualized in this figure. The red part shows the proportion above the quantile at 97.5%. In contrast, the green part indicates the proportion below the quantile at 2.5%. The difference of those two quantiles should be as small as possible to ensure a good compactness around zero. Additionally, using the quantiles has the advantage of leaving out the extreme outliers.

Table 4.4. The results of the calculation of the skewness are presented in this table. The features and their corresponding skewness as well as the achieved points according to the introduced points system are listed. A perfectly symmetric feature would achieve a skewness of 0. In this analysis, only normal beats are included.

Achieved points	Skewness	Feature number
1	38.2464	31
2	21.5815	33
3	8.0610	44
4	7.5965	32
5	6.9537	48
6	6.8249	37
7	6.6004	28
8	6.4836	35
9	6.1241	47
10	5.8249	50
11	5.1220	30
12	5.0905	20
13	4.2908	12
14	4.2306	34
15	4.0288	36
16	3.5653	13
17	3.2107	4
18	3.0862	8
19	2.9412	9
20	2.8071	6
21	2.5514	19
22	2.5456	21
23	2.4602	18
24	2.0085	26
25	1.9758	29
26	1.9462	5
27	1.8449	1
28	1.7478	43
29	1.7075	10
30	1.6633	38
31	1.6597	42
32	1.6582	51
33	1.5980	7
34	1.5269	23
35	1.4951	2
36	1.3556	14
37	1.1889	25
38	0.9937	40
39	0.9578	15
40	0.8512	24
41	0.8015	55
42	0.7481	11
43	0.7213	49
44	0.6836	46
45	0.5702	45

Achieved points	Skewness	Feature number
46	0.5525	17
47	0.5115	3
48	0.5018	22
49	0.3382	16
50	0.3268	52
51	0.3096	41
52	0.3007	27
53	0.2760	54
54	0.2579	53
55	0.1174	39

Table 4.5. The results of the calculation of the kurtosis are presented in this table. The features and their corresponding kurtosis as well as the achieved points according to the introduced points system are listed. The kurtosis of an ideal distributed feature take on the value 1.8. In this analysis, only normal beats are included.

Achieved points	Kurtosis	Feature number
1	2193.05228	31
2	875.1102	33
3	178.3898	48
4	169.9890	32
5	129.3630	28
6	119.9266	44
7	116.0980	47
8	111.6415	36
9	92.9778	46
10	90.9572	37
11	82.1166	35
12	80.2210	45
13	74.9971	50
14	68.3999	30
15	59.9490	34
16	46.6422	51
17	37.4240	55
18	36.8030	20
19	31.2219	12
20	27.6080	26
21	24.0599	43
22	23.1704	21
23	22.3733	13
24	21.7371	53
25	20.6586	8
26	20.1381	42
27	18.4922	4
28	18.3355	6
29	17.6542	9
30	17.4503	39

Achieved points	Kurtosis	Feature number
31	16.7782	3
32	16.4372	29
33	15.0203	19
34	15.0067	38
35	14.3081	18
36	14.0554	27
37	14.0429	52
38	13.5849	1
39	12.4952	5
40	12.4630	49
41	11.7253	23
42	11.6999	7
43	10.8307	14
44	10.6387	11
45	10.4801	2
46	10.2241	41
47	9.9099	10
48	9.8397	15
49	9.7083	17
50	9.6563	40
51	9.6172	22
52	9.6114	25
53	9.1990	16
54	8.4555	54
55	8.2131	24

Table 4.6. The results of the difference of the quantiles at 97.5% and 2.5% are presented in this table. The features and the corresponding difference as well as the achieved points according to the introduced points system are listed. The smaller the difference, the compacter are the feature values located around zero. This leads to a better separability. In this analysis, only normal beats are included.

Achieved points	Quantile _{97.5%} – Quantile _{2.5%}	Feature number
1	4.2887	17
2	4.2294	14
3	4.2153	5
4	4.1666	41
5	4.1461	22
6	4.1314	11
7	4.1249	3
8	4.1233	18
9	4.1221	16
10	4.1176	1
11	4.1149	19
12	4.0930	10
13	4.0776	4
14	4.0762	15
15	4.0651	7

Achieved points	Quantile _{97.5%} – Quantile _{2.5%}	Feature number
16	4.0512	9
17	4.0455	43
18	4.0206	27
19	4.0078	6
20	3.9921	24
21	3.9885	12
22	3.9634	13
23	3.9545	25
24	3.9317	2
25	3.8712	39
26	3.8605	52
27	3.8315	23
28	3.8182	21
29	3.8036	8
30	3.7907	40
31	3.7523	29
32	3.7145	38
33	3.6749	54
34	3.6577	20
35	3.5897	55
36	3.5867	49
37	3.5006	26
38	3.3810	53
39	3.3808	32
40	3.3772	42
41	3.3348	51
42	3.2821	28
43	3.2508	48
44	3.2358	36
45	3.2039	46
46	3.1890	47
47	3.1731	30
48	3.1429	34
49	3.0855	37
50	2.9728	35
51	2.9519	45
52	2.8946	50
53	2.7135	44
54	1.5836	33
55	0.9217	31

4.5 Equalization of data

The goodness of a classification task is often evaluated by its achieved sensitivity and specificity. The ROC analysis e.g. is also based on those two values. But sometimes, there may appear some problems concerning the interpretation of the results only regarding the true positive and true negative rates. Sensitivity and specificity do not always contain enough information to decide whether it is a good or bad classification. The problem will be explained in the following short example.

A ground truth of 1000 beats exists, consisting of 990 normal beats and 10 extrasystoles. A classification is done which classifies 980 normal beats correctly, all 10 extrasystoles correctly, but 10 normal beats falsely as ectopic beats. This results in a sensitivity of 1 and a specificity of 0.9899 (see equations (4.3) - (4.6)).

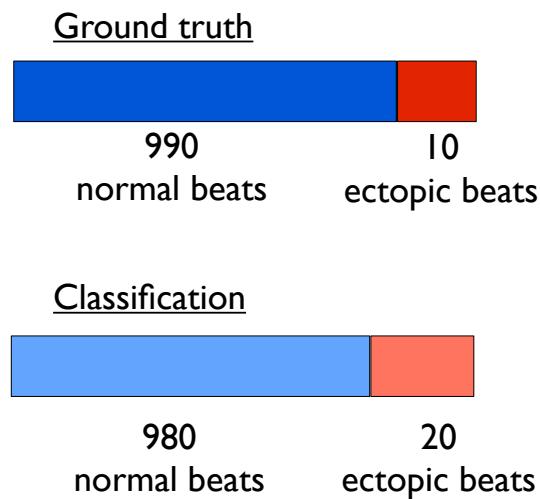


Figure 4.13. If the classes are unevenly distributed, there might appear misleading results if only the sensitivity and specificity are considered. They intuitively indicate better classification results than what they really are. An example to this problem is given in this figure.

$$Sensitivity = \frac{10}{10 + 0} = 1 \quad (4.3)$$

$$Specificity = \frac{980}{980 + 10} = 0.9899 \quad (4.4)$$

$$PPV = \frac{10}{10 + 10} = 0.5 \quad (4.5)$$

$$NPV = \frac{980}{980 + 0} = 1 \quad (4.6)$$

Although sensitivity and specificity indicate a good classification, one cannot be really satisfied. Only half of the beats classified as ectopic are really ectopic beats. This is pointed out in a PPV of 0.5. This may be irritating and not be seen in analyses like the ROC which are only based on sensitivity and specificity. That is why those characteristic values have to be treated with caution.

		<u>ground truth</u>	
		ectopic beats	normal beats
<u>classification</u>	ectopic beats	TP=10	FP=10
	normal beats	FN=0	TN=980

Figure 4.14. The confusion matrix of the example is given here.

To avoid the problems of such a prevalence of the classes, an equalization is conducted, i.e. the same number of ectopic and normal beats is used. Therefore, all 7505 ectopic beats of the used signals of the MIT-BIH-Arrhythmia-Database, as well as 7505 randomly chosen normal beats, are taken for the feature evaluation and also later for the training and test of the SVM.

4.6 Feature dependency

In this thesis, the features have been up to this point always regarded separately during the evaluation. But the situation may change if the features are regarded in relation to each other. Or in other words, an analysis of the feature dependency may yield different information. First approaches are also considered in this thesis.

In order to get some information about the relationship among the features, a second ROC algorithm is implemented. Therefore, the basic ROC analysis is extended. After the first run-through, the best feature is selected to separate the classes a first time. In this case, feature 48 is chosen to split the first time as it is the winner of the ROC analysis (cf. table 4.1). This first classification is done using the threshold that achieved the highest correct rate in the first analysis. In the following, the beats are compared with the annotations and separated into correctly and falsely classified beats. The falsely classified beats are again given to the ROC analysis. The idea is to find out whether the ranking changes when the underlying set has already been split one time and only falsely classified beats are regarded. This process is schematically shown in figure 4.15.

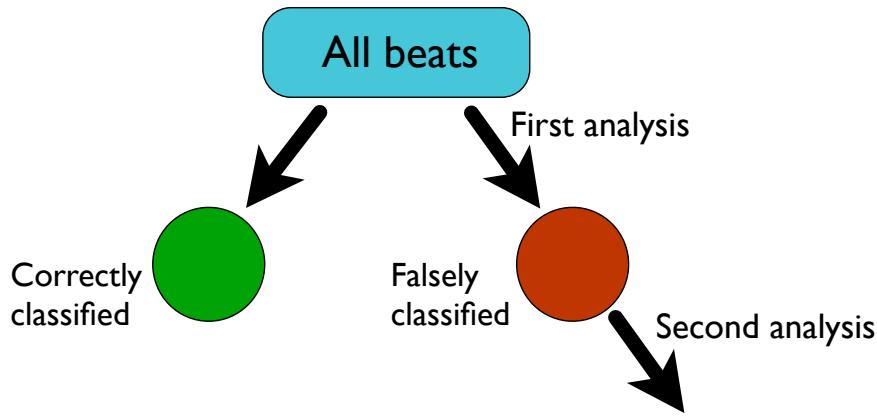


Figure 4.15. A first approach to investigate the dependency among the features is the implementation of an ROC algorithm that analyzes only beats that are falsely classified in a first run-through.

Since this is also implemented as a first approach, a ranking is also built based on these results. It is shown in table 4.7. The best results are produced by feature 48 (skewness of the amplitude distribution), feature 7 (quotient of moving average of the previous five and next five RR intervals and RR interval to previous beat) and feature 8 (difference of RR interval to previous beat and moving average of RR intervals). In contrast, bad results concerning this analysis come out for the features 53 (second PCA score), 49 (kurtosis of amplitude distribution) and 54 (third PCA score).

This pattern can be continued iteratively and seems to be a good approach to investigate the dependency between different features in future.

Secondly, the classification tree is another tool which may be used to investigate the dependency of the features. This is only barely used in this thesis. Hereto, a classification tree already implemented in Matlab is taken. Using the GDI as split criterion, it confirms the results of the GDI analysis (see section 4.2).

Further possibilities are presented in future prospects in chapter 8.

Table 4.7. The results of the second ROC analysis are presented in this table. The features are tested again using only the falsely classified beats from a first run-through. The features and their corresponding AUC value after the second run-through as well as the achieved points according to the introduced points system are listed.

Achieved points	AUC after second run-through	Feature number
1	0.5215	54
2	0.5245	49
3	0.5282	53
4	0.5383	26
5	0.5388	21
6	0.5607	36
7	0.5655	37
8	0.5769	55
9	0.5798	42
10	0.5950	40
11	0.6013	35
12	0.6041	46
13	0.6044	44
14	0.6137	31
15	0.6247	45
16	0.6426	33
17	0.6432	51
18	0.6767	34
19	0.6820	28
20	0.7379	50
21	0.7870	27
22	0.7902	4
23	0.7989	2
24	0.8030	32
25	0.8092	43
26	0.8276	29
27	0.8323	52
28	0.8489	41
29	0.8602	38
30	0.8621	30
31	0.8631	25
32	0.8690	16
33	0.8691	15
34	0.8717	23
35	0.8742	6
36	0.8787	11
37	0.8787	13
38	0.8802	24
39	0.8824	47
40	0.8829	12
41	0.8846	5
42	0.8900	14
43	0.8904	19
44	0.8917	18
45	0.8945	39

Achieved points	AUC after second run-through	Feature number
46	0.9039	22
47	0.9094	1
48	0.9123	20
49	0.9192	17
50	0.9268	3
51	0.9441	10
52	0.9459	9
53	0.9524	8
54	0.9531	7
55	0.9561	48

5

Choice of new features

After the process of evaluating the features that are used for the classification task, a choice which attributes will be taken to train the new SVM must be made. Therefore, it is necessary to merge all the information gained of the previous evaluation methods. This chapter is about creating a final quality ranking of the features and subsequently deciding which attributes are suitable to build the new classifier.

5.1 Final feature rankings

The analyses presented in chapter 4 yield a huge amount of information. To handle all this new knowledge, the single rankings shown in the specific section of each evaluation tool are combined to a total ranking.

First, the results of the analyses of ROC, GDI and IGR (see sections 2.2.2 to 2.2.5) as well as the ranking of the second ROC step (see section 4.6) are combined to one ranking. During that process, the achieved points are summed up for every feature and divided by four. As a result, one gets a ranking according to the average point score of every feature for these four evaluation tools. The list is given in table 5.1. One can see that, on average, the features 48 (skewness of the amplitude distribution), 7 (quotient of moving average of the previous five and next five RR intervals and RR interval to previous beat) and 33 (maximum of cross correlation) yield the highest point scores. In contrast, one can find the features 26 (maximum of second derivative), 55 (fourth PCA score) and 54 (third PCA score) at the end of the ranking.

Since the statistical analysis yielded three single rankings, they are also averaged and lead to a total ranking of the statistical evaluation. The calculation of the skewness, the calculation of the kurtosis and the difference of the quantiles at 97.5% and 2.5% are here included. These results are shown in table 5.2. They differ from the results of the first ranking. In this case, the attributes 54 (third PCA score), 49 (kurtosis of the amplitude distribution) and 50 (number of elements of highest bin of histogram of amplitude distribution referred to total number of elements) show the best results. The features 12 (quotient of the RR intervals of the past beat), 48 (skewness of amplitude distribution) and 32 (minimum of cross correlation) are at the bottom of the ranking.

In the end, those two main rankings are combined to a total feature ranking. In the next step, this will be used to decide which attributes will be taken to train the new SVM. Since the analyses ROC, GDI and IGR are the more reliable analyzing methods, one needs to put an emphasis on the ranking created by those methods. Furthermore, the guess that the symmetry and compactness may be important for the SVM is only a sug-

gestion and not a proven fact. Although the compactness of the values for the normal beats is center of this statistical evaluation, the problem may appear that features are too strongly concentrated. Small differences may already lead to a misclassification in this case. Besides, the features may be too rigid, so that the ectopic beats are projected to the same range where the normal beats are located.

That is why the ranking built of ROC, GDI and IGR is weighted with three to one to the ranking created by the statistical analyses. This final feature ranking can be seen in table 5.3. It is important to mention that not the averaged point scores are taken for the weighting but the new achieved point score in the two main rankings 5.1 and 5.2. The formula for the final weighting is given in equation 5.1. According to this conclusive ranking, feature 7 (quotient of moving average of the previous five and next five RR intervals and RR interval to previous beat), feature 49 (kurtosis of the amplitude distribution) and feature 17 (quotient of quotient of current RR intervals and quotient of past RR intervals) are the most useful attributes. An average score of 47 to 49 points indicate very good results. The worst point score yield the features 4 (the instantaneous heart rate to the next beat), 36 (skewness of the QRS complex) and 37 (kurtosis of the QRS complex). They achieve a score of about seven to eight points on average.

$$\text{total ranking} = \frac{3 \cdot \text{first ranking} + 1 \cdot \text{second ranking}}{4} \quad (5.1)$$

To visualize the performance of the best feature after the whole evaluation process, the ROC plot of feature 7 is presented in figure 5.1. The curve corresponds to an AUC of 0.9319. The histogram with the distribution of the different beat types is also shown in figure 5.2. An IGR value of 0.1125 results. To conclude, a GDI value of 0.4129 is achieved.

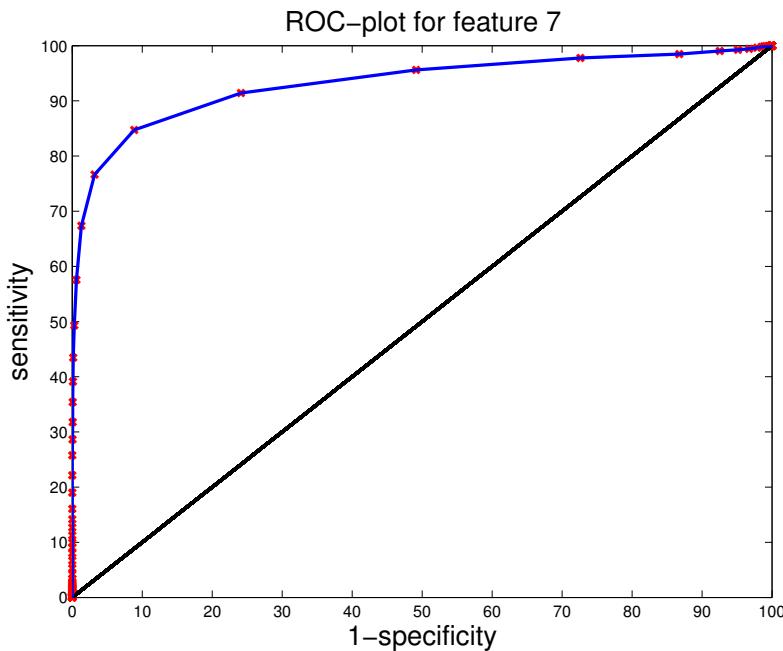


Figure 5.1. The ROC curve of the best feature is shown here. The AUC is 0.9319 .

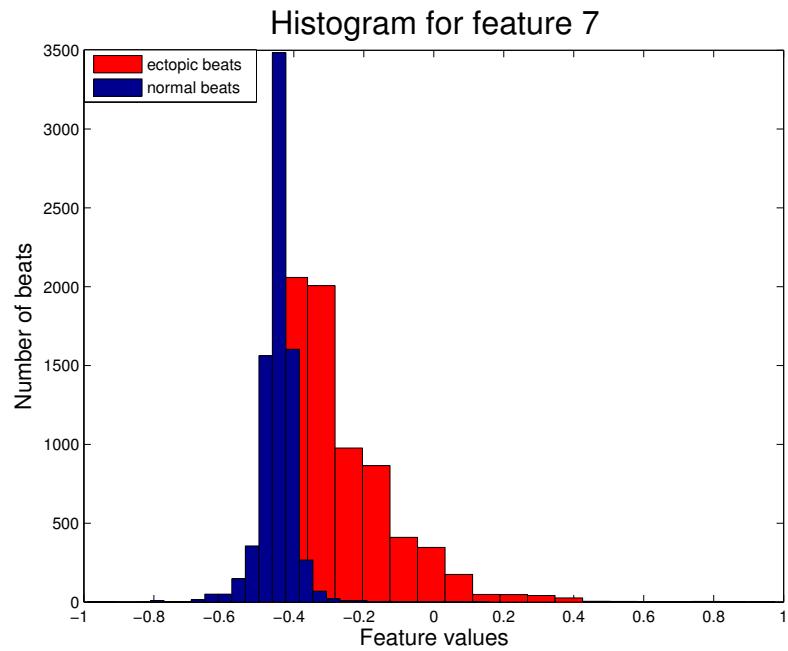


Figure 5.2. The histogram of the best feature is given. The red bins represent the ectopic beats, the blue bins stand for the normal beats. A good separation can be achieved as the values of the different types of beats are mainly located in different intervals.

Table 5.1. The point scores of the analyses ROC, GDI, IGR and the second step of the ROC are averaged and listed in a new ranking. The left column describes the new point scores achieved in this ranking.

New achieved points	Previous averaged points	Feature number
1	5.25	54
2	5.75	55
3	6.5	26
4	7.25	36
5	8.75	37
6	10.5	53
7	12.75	40
8	13.5	51
9	15.5	4
10	15.75	2
11	16.75	23
12	17.75	25
13	20.25	39
14	21	35
15	21.5	12
16	22.25	22
17	23	14
18	24	24
19	24.5	28
20	24.75	46
21	25.25	43
22	25.5	32
23	26	29
24	26.25	27
25	28	1
26	28	21
27	28	41
28	28.75	13
29	29.25	52
30	29.75	15
31	30	45
32	30.75	34
33	31.25	16
34	31.5	6
35	31.5	19
36	32.25	20
37	32.5	18
38	33.75	5
39	34.5	3
40	35.75	11
41	36	47
42	36.25	42
43	36.75	30
44	36.75	38
45	37	49

New achieved points	Previous averaged points	Feature number
46	39.5	9
47	40.75	44
48	42.25	50
49	43.5	8
50	43.5	10
51	44	17
52	44.25	31
53	44.25	33
54	45.5	7
55	54	48

Table 5.2. The point scores of the statistical analyses, i.e. the calculation of the skewness, the kurtosis and the difference of the quantiles at 97.5% and 2.5%, are averaged and listed in a new ranking. The left column describes the new point scores achieved in this ranking.

New achieved points	Previous averaged points	Feature number
1	15.67	32
2	17	48
3	17.67	12
4	18	28
5	19	4
6	19	31
7	19.33	33
8	20.33	13
9	20.67	44
10	20.67	47
11	21.33	9
12	21.33	20
13	21.67	19
14	21.67	37
15	22	18
16	22	43
17	22.33	6
18	22.33	36
19	22.67	5
20	23	35
21	24	8
22	24	21
23	24	30
24	25	1
25	25	50
26	25.67	34
27	27	14
28	27	26
29	28.33	3
30	29.33	10

New achieved points	Previous averaged points	Feature number
31	29.33	29
32	29.67	51
33	30	7
34	30.67	11
35	31	55
36	32	17
37	32	38
38	32.33	42
39	32.67	46
40	33.67	15
41	33.67	41
42	34	23
43	34.67	2
44	34.67	22
45	35.33	27
46	36	45
47	36.67	39
48	37	16
49	37.33	25
50	37.67	52
51	38.33	24
52	38.67	53
53	39.33	40
54	39.67	49
55	46.67	54

Table 5.3. The final feature ranking is presented in this table. The two main rankings are weighted with three to one in favor of the ranking of ROC, GDI and IGR.

Total weighted points	Feature number
7.25	37
7.5	36
8	4
9.25	26
10.25	55
12	12
14	51
14.5	54
15.25	28
15.5	35
16.75	32
17.5	53
18.25	2
18.5	40
18.75	23

Total weighted points	Feature number
19.5	14
19.75	43
21.25	25
21.5	39
23	13
23	22
24.75	1
24.74	46
25	21
25	29
26.25	24
29.25	27
29.5	19
29.75	6
30	20
30.5	34
30.5	41
31.5	18
32.5	15
33.25	5
33.25	47
34.25	52
34.75	45
36.5	3
36.75	16
37.25	9
37.5	44
38	30
38.5	11
40.5	31
41	42
41.5	33
41.75	48
42	8
42.25	38
42.25	50
45	10
47.25	17
47.25	49
48.75	7

5.2 Choice of new features

After combining all the information in one final feature ranking, one has to decide which features will be used in the further process, i.e. for training and testing the new SVM. To make this decision, a decision rule based on the final ranking is created. Basically, the border is set at 20 points, so that features which achieved a point score of less than 20 points will be left out. But in a small area around the border, the features are investigated in detail again. Since the ROC analysis is the most intuitive evaluation method, the AUC values of the features near the border are again compared. After this detailed view, a decision rule is finalized: Features with a point score of less than 22 points will not be used to train and test the new SVM. Subsequently, 19 features will be left out, 36 features will still be used.

Table 5.4. To decide which features will be used later on for the new classification, a decision rule is introduced. This is illustrated in the following table. Features achieving a total point score of less than 22 points will be left out, features achieving more than 22 points will be used. The whole table can be seen in table 5.3.

Total weighted points	Feature number
\vdots	\vdots
18.25	2
18.5	40
18.75	23
19.5	14
19.75	43
21.25	25
21.5	39
23	13
23	22
24.75	1
24.74	46
25	21
25	29
\vdots	\vdots

In order to check how far the statistical feature evaluation is useful, there is also made a choice of features based on a ranking which does not consider the statistical evaluation (see figure 5.1).

If one does not change the decision rule, the results change in the way that more features are chosen to be used for classification. In this case, the decision border is still at 22 points, but the number of the features that are used for classification changes to 28.

If one keeps the number of features constant, it comes out that the decision rule increases. Here, one uses again 36 features, but therefore has to adapt the decision rule so that all features with an averaged point score of less than 24.75 are left out.

6

Training a new support vector machine

As the SVM counts to the class of supervised learning algorithms, a training phase has to be initially conducted by the user. During this process, the SVM is taught how to classify the beats in later applications. Therefore, an optimization of two parameters (C and γ) is executed. After that, a test data set is given to the SVM to get characteristic values which describe its performance. This chapter deals with the process of training and testing an SVM in order to get the best possible SVM. Exemplarily, some steps of the whole process are presented.

6.1 Feature normalization

Before the different features are given to the SVM, a normalization is indispensable. It seems to be clear that e.g. the values for the RR intervals strongly differ from values for the PCA scores or the integral of the QRS complex. A problem which may appear is that features containing great values may dominate features consisting of small values. In addition, dealing with values of different magnitudes may cause numerical problems. Therefore, a linear scaling to the range $[-1, 1]$ is conducted [1]. The normalization is thus done by the following formula as recommended by [19]:

$$P_{i,n \text{ scaled}} = 2 \cdot \left(\frac{P_i - \max_n\{P_{i,n}\}}{\max_n\{P_i\} - \min_n\{P_i\}} \right) + 1 \quad (6.1)$$

This calculation is valid for a given feature P_i and a given beat number n . The parameter i represents the feature number in this case. Consequentially, one has $i \in \{0 \dots I\}$ and $n \in \{1 \dots N\}$, where I stands for the number of features and N for the total number of beats.

This leads to the fact, that the minimal feature value is projected on -1 while the maximal feature value is set to $+1$. The values lying in between are linearly scaled within this interval.

A different approach would be to normalize only using the amount of normal beats between the quantiles at 97.5% and 2.5%. This would project most of the normal beats to the interval $[-1, 1]$ but leave out the outliers. The ectopic beats would probably be located slightly out of this range. The most important goal of the normalization is that the values the SVM has to handle are nearly in the same magnitude and do not differ in powers of ten. That is why it should not be a problem that the ectopic beats and the outliers of the normal beats would be slightly out of the aimed interval.

This should only give a short outlook in further possibilities. In this thesis, the recommended standard normalization as shown in equation 6.1 is conducted.

6.2 Training and Testing

To teach a classifier how to divide data into different classes, a training process has to be executed. Based on annotated data, the classifier learns which attribute values are connected to which type of beat. Therefore, selected signals of the MIT-BIH-Arrhythmia-Database build the data set which is used to train and test the classifier. In total, 15010 QRS complexes were considered. Since the SVM is able to distinguish between normal beats, supraventricular and ventricular extrasystoles, the signals were chosen in that way that they only contain these three types of beats.

To get the best possible SVM, an optimization of two parameters has to be done. The user tries to find the combination of the parameters γ and C which yields the best classification results. As a Gaussian kernel function was used for the nonlinear transformation in this thesis, γ determines how this kernel function is exactly shaped. In this case, γ is directly related to the variance σ^2 and defined as [20]

$$\gamma = \frac{1}{2\sigma^2} . \quad (6.2)$$

The kernel function can consequentially be written as

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right) = \exp(-\gamma \cdot \|\mathbf{x}_i - \mathbf{x}_j\|^2) . \quad (6.3)$$

On the other hand, the parameter C describes the misclassification error which the user is willing to allow. For details concerning the SVM, see section 2.2.1.

For this classification task, the libsvm, an optimal algorithm implemented by the Department of Computer Science of the University of Taiwan, was used [20].

One has to remind, that a strong prevalence to the class of normal beats exists. That is the reason why all ectopic beats and the same number of randomly chosen normal beats were taken for training and testing the SVM (also see section 4.5). Before having started the training process, the data set was divided into three parts of the same size. Two thirds were used for training and one third for testing. In other words, 10006 beats were taken for training and 5004 beats for testing. This was conducted three times, so that every of the three parts was once used for testing, following the principle of cross validation. In the end, an averaged correct rate was calculated. The whole optimization process was done by testing every combination of γ and C . Usually, one chooses their values on a logarithmic scale to cover a larger range using smaller numbers. In the beginning, all combinations of $\log(\gamma) \in [-6, \dots, 6]$ and $\log(C) \in [-6, \dots, 6]$ were chosen as parameters for the SVM. The correct rates as primary results of the classification are displayed in a heatmap. In this kind of diagram, red areas correspond to high correct rates, blue areas stand for lower correct rates. For the first optimization step, the results are shown in figure 6.1. Located at $\log(\gamma) = -0.67$ and $\log(C) = 3.33$, a correct rate of 98.2545% was obtained in the first step.

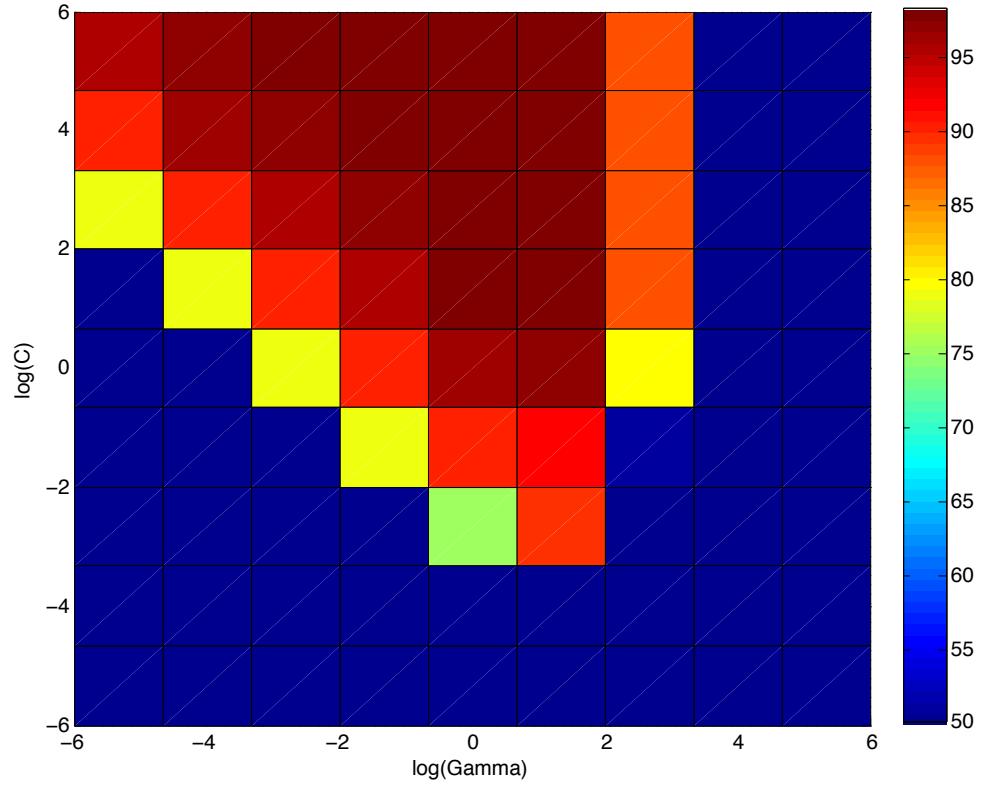


Figure 6.1. The correct rates are presented in a heatmap. The parameters $\log(\gamma) = -0.67$ and $\log(C) = 3.33$ lead to the best correct rate of 98.2545%.

After this first round, the user chooses a new value range for the two parameters in a smaller interval around the values that achieved the best correct rate so far. Every possible combination was then tested again. This procedure of narrowing down the best correct rate is repeated until the best results are obtained. In figure 6.2, the heatmap for the fourth optimization step is shown. A correct rate of 98.574% came out in this case for $\log(\gamma) = 0.044$ and $\log(C) = 1.429$.

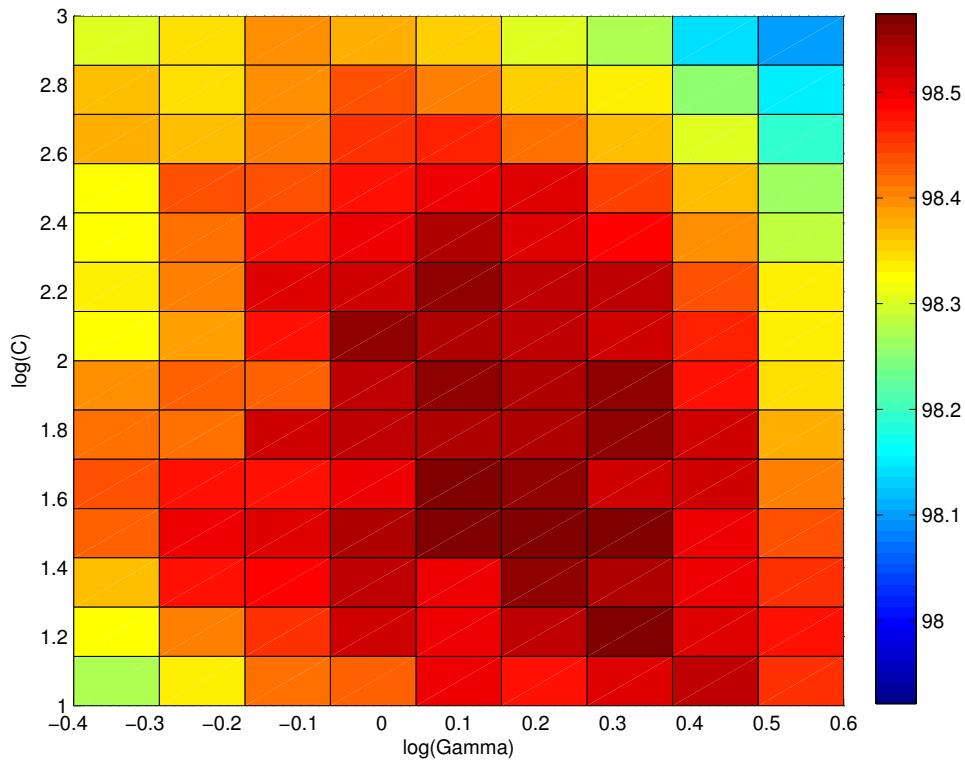


Figure 6.2. The achieved correct rates are visualized in a heatmap. The best correct rate of 98.574% is obtained for $\log(\gamma) = 0.044$ and $\log(C) = 1.429$ in this case.

Results and discussion

In this chapter, the classification results of the new SVM is presented. It only uses the features selected after the feature evaluation (see section 4 and 5.2). To check if the statistical feature evaluation was useful, two SVMs were additionally developed. For the purposes of comparing, their classification results are also given. The classification is done with respect to the three classes normal beat, ventricular ectopic beat and supraventricular ectopic beat.

7.1 Results

Using the 36 best features that are selected from table 5.3, the new SVM yields the following results:

- CR = 98.574%
- Sensitivity = 98.592%
- Specificity = 99.069%
- PPV = 99.062%
- NPV = 98.592% .

These values correspond to the following values for γ and C :

- $\log(\gamma) = 0.044$
- $\log(C) = 1.429$.

For details concerning the calculation, see equations 2.20 to 2.24 in section 2.2.2.

To check if the statistical evaluation brings a benefit to the later classification results, two SVMs were additionally developed using the features that were selected without the statistical analyses.

For the first alternative, the features were chosen by keeping the decision rule constant and therefore taking 28 features and leaving 27 out (also see section 5.2). The results are as follows:

- CR = 98.374%
- Sensitivity = 98.442%
- Specificity = 98.868%
- PPV = 98.864%
- NPV = 98.449% .

The best choice for the parameters γ and C are:

- $\log(\gamma) = -0.1$
- $\log(C) = 1.571$.

For the second one, the feature were selected by keeping the number of features constant and therefore changing the decision rule (also see section 5.2). This leads to the following classification results:

- CR = 98.521%
- Sensitivity = 98.390%
- Specificity = 99.041%
- PPV = 99.034%
- NPV = 98.396% .

The corresponding γ and C are:

- $\log(\gamma) = -0.15$
- $\log(C) = 2.129$.

In conclusion, three SVMs were trained and tested using 15010 beats of the MIT-BIH-Arrhythmia-Database, i.e. 7505 normal beats and 7505 ectopic beats. The results will be discussed in context in the following section.

7.2 Discussion

One can notice that the SVM using features that were selected including the statistical evaluation yields the best results. The important values for a classification task are very satisfying since the main goal – an improvement of the performance of the former SVM – is achieved. Especially, the sensitivity and the PPV show strong improvements. For comparison only, the old classification results are shown below [1]:

- CR = 98.2108%
- Sensitivity = 92.4551%
- Specificity = 99.2847%
- PPV = 94.5205%
- NPV = 98.9960% ,

with

- $\log(\gamma) = 1.25$
- $\log(C) = 0.7$.

One has to say, that the specificity and the NPV are not directly given in Lenis' thesis ([1]), but can be calculated of the given values. The prevalence of his used data set is 0.1177. It is calculated by

$$\text{Prevalence} = \frac{\text{Total number of ectopic beats}}{\text{Total number of beats}} . \quad (7.1)$$

Particularly, the increase of the PPV is very gratifying. It shows that 99.062% of the beats classified as ectopic are really ectopic beats. The increase of the sensitivity is also

very satisfying. 98.592% of the beats that are ectopic are classified correctly. The small decrease of the specificity and the NPV is caused by the prevalence in the training and test data sets. Lenis used an amount of 271151 beats in total, of which about 88% were normal beats ([1]). As there was a strong prevalence in favor of normal beats, the SVM learned more easily how to classify normal beats. In this thesis, the number of normal and ectopic beats used for training and testing is equal. That is why the values representing the classification of normal beats are slightly worse. In total, the classification results are very good as the correct rate as well as the results of the classification of ectopic beats increased significantly.

The results illustrate that the evaluation of the features that were used to classify QRS complexes was a useful task. Only using features that contain enough information and thus have a positive impact on the separation of beats lead in the end to better classification results. The computational time is also shortened noticeably. This is mainly caused by the removal of useless features so that the SVM has to handle less attributes. Furthermore, the SVM has been improved and updated in the last years in comparison to the one G. Lenis used in his work ([1]). This also has an impact on the calculation time. The first optimization step, which usually takes the most time, lasted about six hours (using cross validation, as described in section 6.2). Nevertheless, the comparison with the work of G. Lenis ([1]) concerning the computational time has to be treated with caution as a smaller number of beats was used for training and testing the SVM in the present thesis (see section 4.5).

All in all, the SVM is a very strong automatic classifier that is able to classify several classes in huge data sets regarding a multidimensional feature space. The very good classification results and the acceptable computational time make the SVM suitable for real-world applications.

Summary and future prospects

This chapter summarizes the most important aspects the thesis dealt with. The second part names some possibilities of how this work can be extended in future research projects.

8.1 Summary

In conclusion, this thesis was about evaluating features that are used in an automatic classification process to distinguish between normal and ectopic beats in ECG signals. Features are in this case characteristic values which correspond to different types of QRS complexes. 55 features were developed in a former research project ([1]).

To investigate the attributes, three analyzing methods were implemented. The ROC, GDI and IGR yielded first results. The ROC corresponds to sensitivity and specificity. The GDI measures the purity of subsets after splitting an underlying set. In contrast, the IGR is based on the entropy. Finally, a second ROC analysis was introduced. In this case, the feature evaluation is done using only the falsely classified beats from the first run-through. This was to extract some information about the dependency among the features.

Additionally, a statistical investigation was conducted, based on the guess that symmetrical and concentrated features are easier for the SVM to handle. Therefore, the attributes were scaled by subtracting the mean and dividing by the standard deviation. To ensure a good separability, the normal beats should ideally lie around zero whereas the ectopic beats may be located somewhere else in the feature space. That is why only normal beats were used for this evaluation. In the following, the skewness and kurtosis of the features were calculated. Furthermore, the difference of the quantiles at 97.5% and 2.5% was built to investigate how strongly the feature values are concentrated around zero.

Every analyzing method finally led to a feature ranking. A point system was introduced which assigned a point score to every feature based on its achieved position in every ranking. The worst feature received one point, the best one 55 points. Since there was one ranking for every evaluation method, there were a total of seven single feature rankings. In order to bundle all the new information, the results of the methods ROC, GDI, IGR and the ROC in the second step were merged to one ranking. The same procedure was done including three statistical evaluation methods so that two main rankings remained. Since features may also be too rigid, i.e. the values for the different classes are projected to the same numbers and therefore are not separable, the statistical evaluation was weighted less in total. The final ranking was thus calculated by weighting the two rankings with three to one in favor of the first ranking consisting of ROC, GDI and IGR.

One has to notice that for the feature evaluation the same number of normal and ectopic

beats was used. The reason for this was that a strong prevalence to one class may lead to deceiving results regarding only sensitivity and specificity as characteristic values of the classification task. Only the PPV and NPV may show the difficulties. Using the equalization, the worse results can often be identified directly by looking at the sensitivity and specificity. All in all, 7505 ectopic and randomly chosen normal beats of selected signals of the MIT-BIH-Arrhythmia-Database were taken. The new SVM was trained and tested using the same amount of beats.

It came out that in the end 19 features were left out and 36 features still remained to be used for the new classifier. The new SVM obtained a correct rate of 98.574%. To check if the statistical feature evaluation brings a benefit to the classification results, two other SVMs were trained using features that were chosen without the statistical feature evaluation. The first one used on the one hand the same decision rule to choose the features but on the other hand a different number of features (28). A correct rate of 98.374% was achieved. In the second case the decision rule for the feature choice was changed but the number of features (19) stayed constant. For this setting, the classifier yielded a correct rate of 98.521%.

Although all three results were very satisfying, one can see that the SVM using features that were selected additionally using the statistical evaluation obtained the best results.

8.2 Future prospects

The main approach to continue this work can be evaluating the dependency of the features. A first approach was already implemented in this thesis by evaluating the ROC after the beats were already separated one time. This "second step analysis" can also be extended to other methods than the ROC. Furthermore, a more detailed view on the classification tree may be useful. The way it chooses the features to split the sets may deliver new findings about the information content of the features. Another problem would be how to bundle the gained information to be able to compare the strength of the features. Regarding the features that are left out after the evaluation presented in this thesis, one should investigate if it is possible to modify these attributes in a way that they improve their values for the classification task. In this case, they may be used again.

Other aspects would be to see how far the classification results change if the decision rule of the feature choice is modified. The proportion of three to one of the two main rankings during the process of building a final ranking can also be discussed.

Finally, adding features extracted of other waves of the ECG may provide new information. This thesis was about evaluating attributes based on the QRS complex but especially the P and T wave include much information about ectopic beats and cardiac arrhythmia. Developing features coming from these parts of the ECG and evaluating them using the presented (and future) methods may also contribute to a better classification of ectopic beats in ECG signals.

A

Appendix A

Table A.1. Description of rhythmical features and assignment to feature numbers corresponding to [1].

Feature number	Feature
1	RR interval to previous beat
2	RR interval to next beat ($\hat{=}$ (Feature 1) $^{-1}$)
3	Instantaneous heart rate to previous beat
4	Instantaneous heart rate to next beat ($\hat{=}$ (Feature 3) $^{-1}$)
5	Difference of two consecutive RR intervals
6	Difference of two consecutive heart rates
7	Quotient of moving average of the previous five and next five RR intervals and RR interval to previous beat
8	Difference of RR interval to previous beat and moving average of RR intervals
9	Quotient of moving average of the previous five and next five heart rates and instantaneous heart rate to previous beat
10	Difference of instantaneous heart rate to previous beat and moving average of heart rates
11	Quotient of RR interval of next beat and RR interval of previous beat
12	Quotient of RR intervals of past beat
13	Quotient of instantaneous heart rate of next beat and instantaneous heart rate of previous beat
14	Quotient of instantaneous heart rates of past beat
15	Difference of moving average of quotient of RR intervals and quotient of RR intervals
16	Quotient of moving average of quotient of RR intervals and quotient of RR intervals
17	Quotient of quotient of current RR intervals and quotient of past RR intervals
18	Difference of moving average of quotient of heart rates and quotient of heart rates
19	Quotient of moving average of quotient of heart rates and quotient of heart rates
20	Quotient of quotient of current heart rates and quotient of past heart rates

Table A.2. Description of morphological features and assignment to feature numbers corresponding to [1].

Feature number	Feature
21	Minimum of amplitude
22	Maximum of amplitude
23	Minimum of first derivative
24	Maximum of first derivative
25	Minimum of second derivative
26	Maximum of second derivative
27	Integral of QRS complex
28	Integral of positive part of QRS complex
29	Integral of negative part of QRS complex
30	Energy of QRS complex
31	Linear correlation coefficient of beat and mean beat
32	Minimum of cross correlation between beat and mean beat
33	Maximum of cross correlation between beat and mean beat
34	Center of mass (QRS complex seen as density function)
35	Variance
36	Skewness
37	Kurtosis
38	Maximum of absolute of Fourier transform (5Hz to 10Hz)
39	Mean of absolute of Fourier transform (10Hz to 20Hz)
40	Maximum of absolute of Fourier transform (20Hz to 50Hz)
41	Center of mass for activity interval (based on Hermite basis functions and wavelet transform)
42	Variance for activity interval (based on Hermite basis functions and wavelet transform)
43	First coefficient of linear combination of Hermite basis functions
44	Second coefficient of linear combination of Hermite basis functions
45	Third coefficient of linear combination of Hermite basis functions
46	Median of amplitude distribution of QRS complex
47	Variance of amplitude distribution of QRS complex
48	Skewness of amplitude distribution of QRS complex
49	Kurtosis of amplitude distribution of QRS complex
50	Number of elements of highest bin of histogram of amplitude distribution referred to total number of elements
51	Position of highest bin of histogram of amplitude distribution
52	First PCA score
53	Second PCA score
54	Third PCA score
55	Fourth PCA score

List of Figures

2.1 Anatomy of the heart	4
2.2 Conducting system of the heart	5
2.3 Excitation of the heart	6
2.4 Einthoven leads	7
2.5 Goldberger leads	8
2.6 Wilson leads	8
2.7 Schematic ECG	9
2.8 Example for SVES	10
2.9 Typical ECG of SVES	10
2.10 Example for VES 1	11
2.11 Example for VES 2	11
2.12 Principle of SVM in three dimensions	12
2.13 Separation of SVM by margin	13
2.14 Separation of SVM by soft margin	15
2.15 Transformation of SVM	16
2.16 Confusion matrix	17
2.17 Schematic ROC plot	18
2.18 Structure of classification tree	20
2.19 Visualization of splitting process by classification tree	20
2.20 Example for GDI	22
4.1 Best two features of ROC analysis	28
4.2 Worst two features of ROC analysis	29
4.3 ROC of the best feature of ROC analysis	29
4.4 ROC plot of worst feature of ROC analysis	30
4.5 Best two features of GDI analysis	31
4.6 Worst two features of GDI analysis	31
4.7 Interval nesting in IGR analysis	32
4.8 Best two features of IGR analysis in histogram	33
4.9 Worst feature of IGR analysis in histogram	33
4.10 Optimal distribution for feature	38
4.11 Three-dimensional histogram of two example features	39
4.12 Difference of quantiles	41
4.13 Problem of prevalence	46
4.14 Problem of prevalence	47
4.15 Feature dependency investigation via second ROC analysis	48

5.1	ROC plot of best feature	52
5.2	Histogram of best feature.....	53
6.1	Visualization of correct rates of first optimization step in heatmap	61
6.2	Visualization of correct rates of fourth optimization step in heatmap	62

List of Tables

4.1	The results of the ROC analysis are presented in this table. The features and their corresponding AUC value as well as the achieved points according to the introduced points system are listed.	34
4.2	The results of the GDI analysis are presented in this table. The features and their corresponding GDI value as well as the achieved points according to the introduced points system are listed.	35
4.3	The results of the IGR analysis are presented in this table. The features and their corresponding IGR value as well as the achieved points according to the introduced points system are listed.	36
4.4	The results of the calculation of the skewness are presented in this table. The features and their corresponding skewness as well as the achieved points according to the introduced points system are listed. A perfectly symmetric feature would achieve a skewness of 0. In this analysis, only normal beats are included.	42
4.5	The results of the calculation of the kurtosis are presented in this table. The features and their corresponding kurtosis as well as the achieved points according to the introduced points system are listed. The kurtosis of an ideal distributed feature take on the value 1.8. In this analysis, only normal beats are included.	43
4.6	The results of the difference of the quantiles at 97.5% and 2.5% are presented in this table. The features and the corresponding difference as well as the achieved points according to the introduced points system are listed. The smaller the difference, the compacter are the feature values located around zero. This leads to a better separability. In this analysis, only normal beats are included.	44
4.7	The results of the second ROC analysis are presented in this table. The features are tested again using only the falsely classified beats from a first run-through. The features and their corresponding AUC value after the second run-through as well as the achieved points according to the introduced points system are listed.	49
5.1	The point scores of the analyses ROC, GDI, IGR and the second step of the ROC are averaged and listed in a new ranking. The left column describes the new point scores achieved in this ranking.	54

5.2	The point scores of the statistical analyses, i.e. the calculation of the skewness, the kurtosis and the difference of the quantiles at 97.5% and 2.5%, are averaged and listed in a new ranking. The left column describes the new point scores achieved in this ranking.	55
5.3	The final feature ranking is presented in this table. The two main rankings are weighted with three to one in favor of the ranking of ROC, GDI and IGR.....	56
5.4	To decide which features will be used later on for the new classification, a decision rule is introduced. This is illustrated in the following table. Features achieving a total point score of less than 22 points will be left out, features achieving more than 22 points will be used. The whole table can be seen in table 5.3.	58
A.1	Description of rhythmical features and assignment to feature numbers corresponding to [1].	69
A.2	Description of morphological features and assignment to feature numbers corresponding to [1].	70

References

- [1] G. Lenis, “Automatic detection and classification of ectopic beats in the ECG using a Support-Vector-Machine,” Studienarbeit, Institut für Biomedizinische Technik, Karlsruher Institut für Technologie (KIT), Karlsruhe, 2010.
- [2] R. F. Schmidt, G. Thews, and F. Lang, *Physiologie des Menschen*. Berlin; Heidelberg; New York: SpringerMedizinVerlagHeidelberg, 2005.
- [3] S. Silbernagl and A. Despopoulos, *Taschenatlas Physiologie*. Thieme, 2007.
- [4] G. S. Thompson, *Understanding anatomy & physiology*. F.A. Davis Company, 2013.
- [5] H. P. Schuster and H. J. Trappe, *EKG-Kurs für Isabel*. 2005.
- [6] M. Gertsch, *Das EKG: Auf einen Blick und im Detail*. Heidelberg: Springer Medizin Verlag, 2008.
- [7] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and other kernel-based learning methods*. Cambridge: Press Syndicate of the University of Cambridge, 2006.
- [8] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer, 2000.
- [9] S. Abe, *Support Vector Machines for Pattern Classification*. London, Dordrecht, Heidelberg, New York: Springer, 2010.
- [10] T. Fawcett, “ROC graphs: Notes and Practical Considerations for Data Mining Researchers,” 2003.
- [11] A. E. Bradley, “The use of the area under the ROC curve in the evaluation of machine learning algorithms,” 1997.
- [12] T. M. Mitchell, *Machine Learning*. McGraw-Hill Companies, Inc., 1999.
- [13] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and Regression Trees*. Chapman & Hall/CRC, 1984.
- [14] J. R. Quinlan, *C 4.5 Programs for Machine Learning*. Morgan Kaufmann Publishers, Inc., 1993.
- [15] A. Steland, *Basiswissen Statistik : Kompaktkurs für Anwender aus Wirtschaft, Informatik und Technik*. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg, 2010.
- [16] F. Puente León, *Messtechnik: Systemtheorie für Ingenieure und Informatiker*. Berlin, Heidelberg: Springer, 2011.
- [17] V. Krasteva, I. Jekova, and I. Christov, “Automatic detection of premature atrial contractions in the electrocardiogram,” 2006.
- [18] M. H. Zweig and G. Campbell, “Receiver-Operating Characteristic (ROC) Plots: A Fundamental Evaluation Method in Clinical Medicine,” 1993.
- [19] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, “A Practical Guide to Support Vector Classification,” 2010.

- [20] C.-C. Chang and C.-J. Lin, “LIBSVM: a library for support vector machines, Department of Computer Science of the University of Taiwan,” Nov. 2013. software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.