Ambulatory ECG ST Segment Analysis for Detection of Ischemia

KIRIKO KAMIYA

MSc by Research in Pattern Analysis and Neural Networks

ASTON UNIVERSITY

August 2004

ASTON UNIVERSITY

Ambulatory ECG ST Segment Analysis for Detection of Ischemia

KIRIKO KAMIYA

MSc by Research in Pattern Analysis and Neural Networks, 2004

Thesis Summary

Some changes from normality of the ST segment of an electrocardiogram (ECG) are known to be indicative of myocardial ischemia, or lack of oxygen in the heart muscles. Ischemia can lead to myocardial infarction which could be fatal. However, little has been achieved in developing techniques for automatically identifying abnormality in ST segment which is caused by ischemia. This is because ECG is sensitive to the subject's posture and movement, noise in the leads and other factors, and the origin of the abnormality is difficult to be determined. This project investigates methods for modeling and classifying ST segments in ECG in order to detect ischemia automatically.

Keywords: Pattern analysis, ECG, Ischemia, ST segment, Signal Processing

Acknowledgements

Dr David Evans, thank you for not giving up on me despite my stubbornness! Thank you for the "question". Prof David Lowe, Prof David Saad, Prof Ian Nabney, Dr Manfred Opper, thank you for the invaluable theory course in MSc PANN. Everybody at NCRG, thank you for your company and intellectual inputs. Ms Vicky Bond, thank you for your super efficient administration work. Mr Alex Brulo, thank you for fishing me out of numerous computer-related problems. Fellow MSc PANN students, Christophe, Youssef, Thomas, Youenn, Dharmesh and Pierre, PhD students Dan and Ben, thank you for your company, intellectual inputs and moral support. Thank you, Phillip, for your love and support. And finally, thank you Daddy, for the financial support, and the whole family for the encouragement and entertaining e-mails that lifted my spirits whenever I was in need.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In modern medicine, there is high interest in non-invasive diagnosis and treatment of diseases. This includes minimising the need for major surgery by using indirect measurement methods and key-hole surgery, which reduces the cost of diagnosis and treatment and minimise the risk of infection, complication, scarring and mental and physical stress for the patient. Although more invasive, direct measurements of physiological signals give more accurate values, the benefits of non-invasive measurements are motivating research into this area. This project focuses on one such non-invasive method for diagnosing diseases of the heart.

The activity of the heart muscle is monitored by an electrocardiogram (ECG) which consists of electrodes attached to the subject's skin non-invasively, and some kind of recording device.

The heart consists of four chambers, the left and right atria and the left and right ventricles. It is known by the physicians that ischemia, or the lack of oxygen to a muscle, is observed through recordings of the electrical potential in the muscle during recovery from contraction. In the case of the heart, the repolarisation from the atria are obscured by the contraction signals from the ventricles. Therefore, observation of the repolarisation of the ventricles, the ST segment, gives indication of the oxygen saturation in the heart muscles, or the myocardia.

Figure 1.1 shows a schematic drawing of a heart beat on the ECG. A heart beat trace on the ECG consists of P, Q, R, S and T waves[1]. The P wave occurs at the contraction of the atria. The Q, R and S waves are referred to as the QRS complex and is the contraction of the ventricles. The section extending from the S wave to T wave is the ST segment. The T wave is the recovery of the ventricles. The most flat part in the P-Q interval is the iso-electric level and is the equilibrium level of the electric potential in the heart muscle.

The ECG consists of 9 electrodes, making up 12 leads. The positions of these electrodes are shown in Figure 1.2. Electrodes are placed on the patient's left and right arms, left leg and six locations along the ribs. The electrodes on the limbs can be placed nearer the torso on the shoulders and thigh for ambulatory recordings and leads using these electrodes are labeled as modified. The 12 leads are

[1]The existence of a further U wave is still debated.

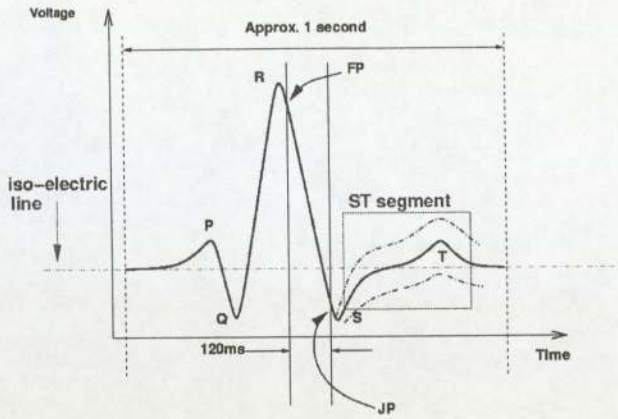


Figure 1.1: Schematic drawing of a heartbeat on ECG.

the signals recorded on each of these electrodes with respect to a reference level, and 3 signals recorded as potential between a pair of electrodes. The leads are numbered from v1 to v6 on the chest, AVR for the right arm, AVL for the left arm and AVF for the left leg. The lead consisting of AVR and AVL is the standard lead I (SLI), AVR and AVF make up standard lead II (SLII) and AVL and AVF make up standard lead III (SLIII) [2] [14].

Myocardial ischemia often precedes myocardial infarct, or heart attack, which can be fatal. It is of high clinical interest to identify abnormality in the ST segment resulting from ischemia as this may allow time for clinicians to administer preventative measures for patients who are likely to suffer myocardial infarct.

Analysis of ambulatory ECG is ideal for the purpose of this project. ECG is non-invasive and ambulatory recordings have a better chance of observing ischemic signals which can be quite rare at early stages of the disease and often missed at short recordings of a few minutes at check-ups at hospitals.

The aim of this project is to automatically distinguish between ischemic and non-ischemic ST episodes in ambulatory ECG recordings. First, this chapter describes the background and motivation for this project, and gives some descriptions for terminology which are specific to this project.

Literature survey of the work done in the area so far is given in Chapter 2. Chapter 3 explains in detail the Long Term ST Database (LTSTDB) which is used in this project. In Chapter 4 a different approach using Neural Networks is described. Finally, Chapter 5 gives the conclusion for this project.

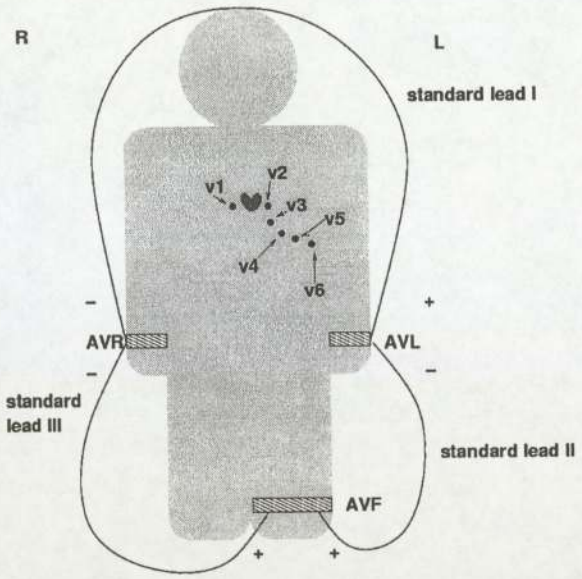


Figure 1.2: Sketch of positions of the electrodes around the body.

1.1 Quantifying results

The quality of the classification is quantified by showing the classification results as confusion matrices and by calculating statistical quality measures [9].

A confusion matrix is presented in the form:

$$\begin{matrix} Normal\ as\ normal & false\ positive \\ false\ negative & ST\ as\ ST \end{matrix}$$

An ideal confusion matrix will have values only on the diagonal, i.e. zero false positive and false negative. Especially in clinical applications, it is vital to minimise false negatives, sometimes at the expense of a high false positive. However, if automated clinical diagnosis systems "cry wolf" too often, this could also have adverse effect on the use of automated systems.

The four most common measures are

$$sensitivity = \frac{positive\ classification}{(positive\ classification + false\ negatives)}, \tag{1.1}$$

$$specificity = \frac{negative\ classification}{(negative\ classification + false\ positives)}, \tag{1.2}$$

$$positive\ predictive\ value\ (PPV) = \frac{positive\ classification}{(positive\ classification + false\ positives)}, \tag{1.3}$$

$$negative\ predictive\ value\ (NPV) = \frac{negative\ classification}{(negative\ classification + false\ negatives)}. \tag{1.4}$$

Of particular interest are the sensitivity (se) and the specificity (sp) as they are used in many papers in Chapter 2 to quantify their results.

1.2 Terminology

Here, some terminology specific to the project is explained.

PhysioNet The name of an internet site providing biomedical signals and processing software, belonging to the Research Resource for Complex Physiologic Signals [4].

PhysioBank A bank of digital biomedical recordings which can be accessed free of charge through PhysioNet [4].

LTSTDB Long Term ST Database. Available from PhysioBank.

ESTDB European ST Database. Predecessor to LTSTDB.

Aristotle A computer program for beat classification.

SEMIA A computer program which allows experts to semi-automatically label interesting events in ECG recordings.

PhysioNet/ Computers in Cardiology Challenges Hosted by PhysioNet and Computers in Cardiology conferences, participants attempt to solve real clinical problems.

1.3 The Challenge

This project is based on the Physionet/ Computers in Cardiology Challenge 2003, "Distinguishing Ischemic from non-Ischemic ST Changes". The Challenge asks a question, "Is it possible to tell the difference between transient ST changes in the ECG that are due to myocardial ischemia, and those that are not?". The answer has long been "no". However, the hosts say, it may be possible to establish inferential associations between specific features of the ECG to myocardial ischemia, in particular, the ST segment. Since the ST segment measurements can easily be affected by change in heart rate, conduction pattern, position of the subject and noise, transient changes in ST segment readings are suggestive of ischemia, but insufficient for a diagnosis. Participants are invited to produce a novel approach into detecting ischemia from ECG alone. The algorithms developed need not be able to detect the ST change events, but need to be able to classify the ST events annotated by experts as ischemic, non-ischemic or unknown.

1.4 Method

The approach to this project is given in Figure 1.3. First, data will be preprocessed and reformed into manageable formats. Then some kind of feature extraction will be used to reduce the dimensionality of the data. The classification will be performed using artificial neural network based method.

In order to classify the abnormal beats as ischemic or non-ischemic, their characteristics need to be quantified as a measure of deviation from normality. In order to obtain a model of normality, the first step is to attempt to define normality by classifying beats as normal or otherwise.

Once a model of normality is constructed and beats compared against it, the change in the difference measure through time can be analysed to obtain classification of abnormal beats.

In this project, the tasks involved in determining a model of "normality" were tackled, from preprocessing up to classification of beats into normal and episodes.

Results obtained from experiments will be compared against existing work by other researchers. Algorithms will be developed on a small set of data first, then applied on a larger number of data sets to determine the quality of the algorithms.

Algorithms will be developed using MATLAB and NETLAB toolbox [10].

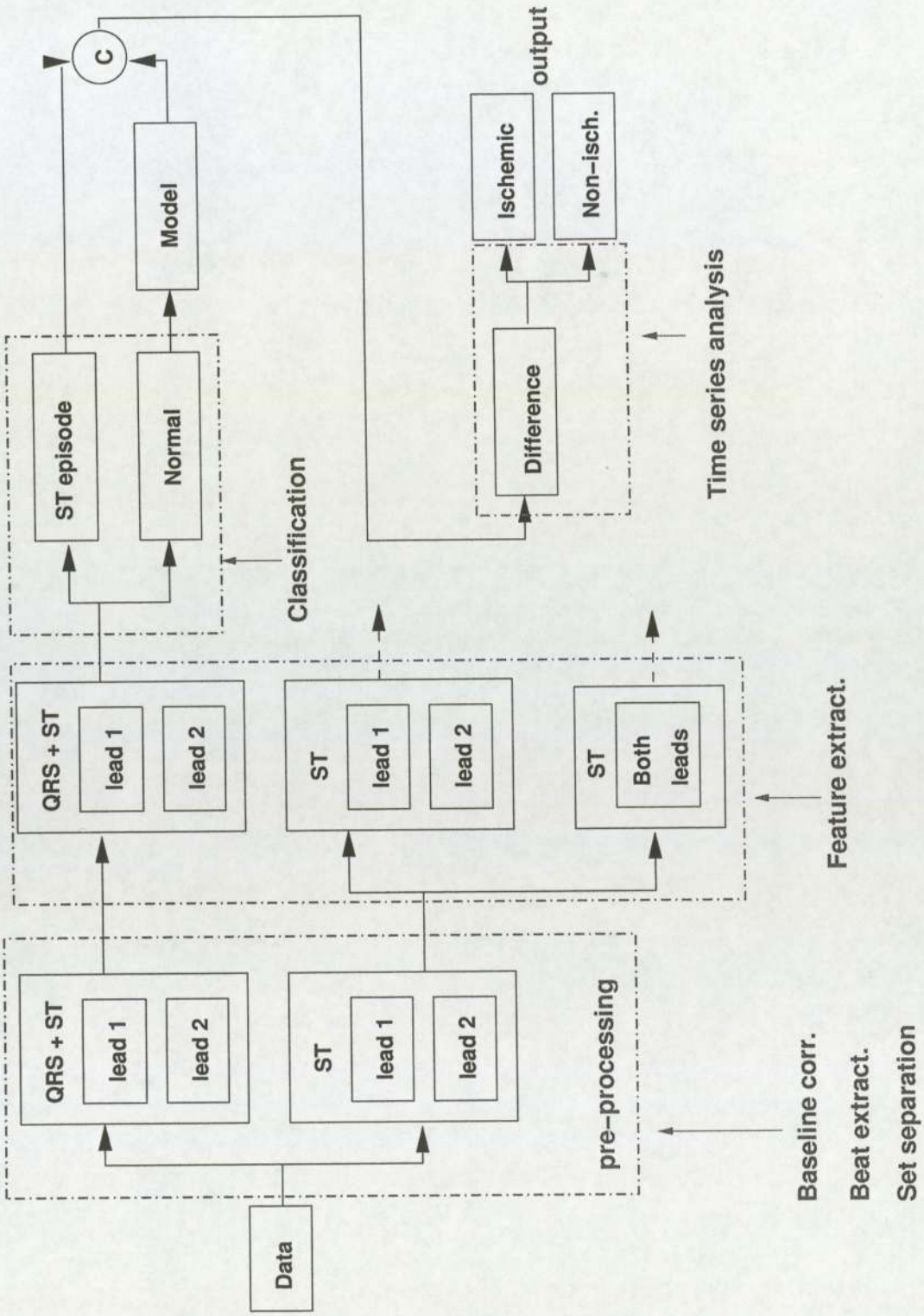


Figure 1.3: Flow chart showing the approach to the project.

Chapter 2

Literature Survey

There have been a variety of approaches to the problem of automatic detection of ischemia. A number of researchers have used principal component analysis (PCA) (or KLT[1])[2] as a first step of the analysis. After that, some have used time series approach, some have used rule based methods and others have used neural networks. A selection of papers are described in brief in the following sections in chronological order.

2.1 Jager *et al.* 1992

According to Jager *et al.* it is possible to distinguish ischemic ST change from non-ischemic ST change if observation of axis shift is available. Axis shift is where the electrical axis of the heart changes, for example, due to the subject changing posture. A detailed description of axis shift in the heart is given in Boutkan [2] and Schamroth [14]. By projecting the ST segments from the European ST databse onto the feature space of the first five KLT components of the ST segment, Jager's team were able to detect ischemic episodes.

Whereas this project is concerned with distinguishing between ischemic and heart-rate related ST episodes, Jager's group's work dealt with distinguishing between ischemic and axis shift ST episodes. They argue that this is possible as ischemic ST episodes are characterised by smooth changes in ST segment level, but ST episodes due to axis shift show sudden changes in ST segment level.

The ECG recordings from the ESTDB are reformatted into manageable data form which then can be used for feature extraction using KLT. There are two data sets; one consisting of 16 readings of ST segments from FP+40 ms to FP+160 ms taken at 8 ms intervals and another of 16 readings of QRS complex from FP−96 ms to FP+24 ms taken at the same interval.

The FP, or fiducial point, is defined as the centre of mass of deflection of the QRS complex. Since the R wave is the most dominant feature in the QRS complex, the fiducial point occurs close to the R wave peak, and can be used to locate a beat within a length of recording.

[1] Karhunen-Loeve Transform
[2] Since PCA and KLT are mathematically the same, these names are used interchangeably.

The first 5 KL coefficients are used for the rest of the analysis.

Noisy beats are excluded from analysis. These beats are automatically found by analysis of the residual error after normalisation to unit variance. A beat whose ST or QRS KL coefficients differing from 15 beats preceding it by more than 25% are excluded as noise.

The set of feature vectors, or KL coefficients, are smoothed using 15 point moving average, re-sampled to 0.5 Hz and further smoothed using 9 point moving average. Axis shifts are detected by detecting stepwise change in ST segment KL coefficients.

To account for slow drift of ST reference level, the mean feature vector is corrected to a new value given as a function of a time constant, τ, the current mean feature vector, $\overline{s}(k)$, and the current feature vector value, $s(k)$. The time constant is chosen by the author. This correction takes place if consecutive mean feature vectors differ more than a certain amount, chosen by the author empirically.

An ST episode is detected by analysis of the distance of a beat from the mean in the feature space. A non-ischemic ST episode is identified if an axis shift coincides with the onset or the end of the episode, and the amplitude of the ECG reading at the extrema of the episode does not exceed 300 μV. All other episodes are identified as ischemic.

In this method, the group achieved gross ischemic episode sensitivity of 85.2% and gross ischemic duration sensitivity of 78.0%. Ischemic episode sensitivity is where all the episodes are weighted equally. Ischemic duration sensitivity is where all minutes of the episodes are weighted equally. This compares to their earlier work in 1991 of 82.4% and 72.5% for the respective measures which were obtained using only the first 2 KL components [5].

2.2 Stamkopoulos *et al.* 1998

Since PCA can only identify linear correlations between random variables, Stamkopoulos argued that for a highly non-linear system such as the ECG and the heart, it is more beneficial to use a non-linear method, a non-linear PCA (NLPCA).

NLPCA is achieved by minimising the mean squared error:

$$J = E||\mathbf{x} - \mathbf{g}(\mathbf{h}(\mathbf{x}))||^2,$$

where the function $\mathbf{h}(\mathbf{x})$ maps $\mathbf{x}$ to the feature space and function $\mathbf{g}(\mathbf{x})$ maps $\mathbf{x}$ back to the real space. There are actually infinitely many solutions to this problem. Further constraints limiting the solutions to contours generated by $\mathbf{h}$ and surface generated by $\mathbf{g}$ makes the problem unique. The solution to this problem is achieved by a 4-layer multiple layer perceptron (MLP) autoassociative network.

This autoassocaitive network is two back-to-back 2 layer multiple layer perceptrons (MLP). The input, the first layer and the second layer make up the first MLP representing the function $\mathbf{h}$. The dimension of the second layer is necessarily smaller than that of the input. This second layer is the data mapped into the feature space. The second layer, the third layer and the forth layer make up the second MLP representing the function $\mathbf{g}$ which maps the features back onto the data space. The

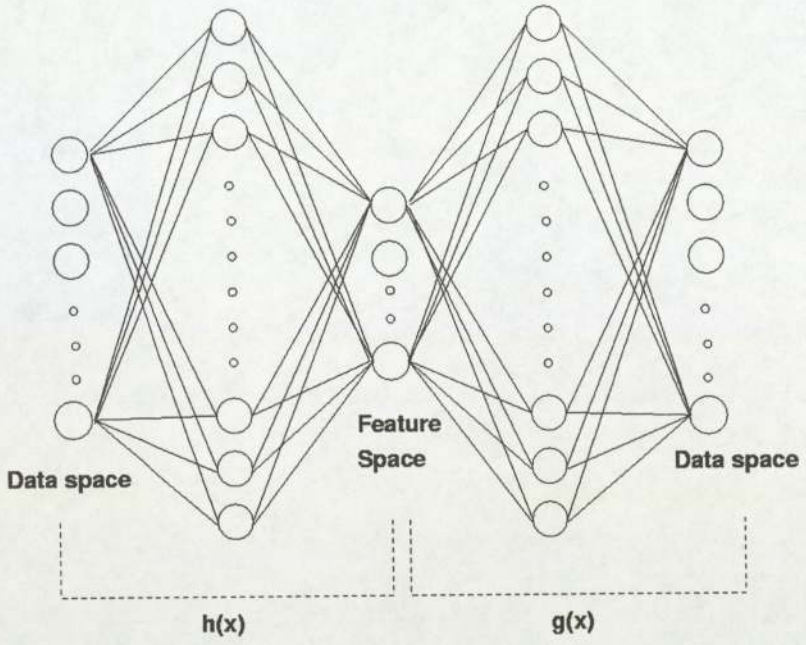


Figure 2.1: Sketch of 4 layer autoassociative MLP for implementing NLPCA.

dimension of the forth layer is necessarily the same as the input layer. The hidden layers, the first and the third, are arbitrarily large.

After correcting the changes in isoelectric level, the input to the neural network is constructed. The beginning of the ST segment is defined as 60 ms[3] after the peak of the R wave and lasts a set length of 160 ms. At a sampling frequency of 250 Hz this gives rise to 40 data points. These samples are compared against a reference template of the average of the 10 first normal ST segments in the record and the difference is used for analysis. The input vector is further reduced in size by taking the average values every 2 samples, resulting in 20 dimensional data going into the neural network.

Classification is achieved by training a Radial Basis Function (RBF) neural network on just normal beats. The decision threshold is set so that 80% of the normal beats are classified correctly. The author achieved average sensitivity to normal beats of 79.32% and 75.19% to abnormal (ischemic) beats. The author notes that if there are more than 80% or fewer than 1% ischemic beats in the data, the sensitivity drops. These results were obtained using 2 principal components, but it is noted that, although 1 principal component analysis produced noticeably inferior results, increasing the number of principal components to 3 or 4 did not improve the results significantly. There is no mention of abnormal, non ischemic beats [15].

[3] 40 ms if there is sign of tachycardia.

2.3 Laguna *et. al* 1999

This group of researchers used KLT to improve the signal to noise ratio of the data. This piece of work is not specifically on the detection of ischemic ST episode, but has lead to work by other researchers in that area.

A KL_n coefficient time series, $kl_n(i)$ is used in the analysis where each value in the time series is the KL coefficient for that beat in the KL order of interest. The subscript n refers to the coefficient number and i represents time as the position of that beat in a whole series of beats in the data.

The records to be used for analysis were carefully chosen from a number of databases, and the final selection amounted to 105 fifteen minute recordings. The data set is constructed from ST segments and T waves, defined as a 600 ms window from 85 ms after the fiducial point of a beat to 240 ms before the next fiducial point.

This data was preprocessed by cubic spline subtraction method to correct for baseline wander, and manual rejection of apparently noisy beats. Six different methods of further preprocessing the data were compared. The methods were combinations of filtering and rule-based heart rate correction (Bazette's formula).

The covariance of the data was calculated using only the ST-T complexes whose length extend to that element of the covariance matrix. Elements of the covariance matrix lacking a full data set were estimated using only the available data.

By inspection of the eigenvalues of the covariance matrix, n was chosen to be 2.

The author found the estimation of $kl_n(i)$ obtained from the inner product of the KLT basis with the data is too noisy for use in beat-by-beat analysis. Instead, the $kl_n(i)$ is estimated by assuming each ST-T complex is made up of a recurring deterministic part and some uncorrelated noise. This analysis performed on a beat-by-beat basis produces a smoother $kl_n(i)$. The error minimisation algorithm requires a choice of a convergence factor μ and a time constant τ_{mse} which is chosen empirically by the author on the basis of trade off between convergence rate and stability.

The author comments that, another researcher, Gracia, achieved sensitivity to ischemic episodes of 81% using this technique with 4 KLT components in 1998 [7].

2.4 Papaloukas *et al.* 2002

Papaloukas *et al.* used PCA as the feature extraction method, then used a Multiple Layer Perceptron (MLP) with Bayesian approach as the classifier. The author has produced a 4-stage ischemic episode detection method earlier in 2001, where the beat classification was performed using a knowledge-based method. This paper presents a novel method to replace that knowledge-based beat classification method in hope to minimise the adverse effect of noise which caused the Positive Predictive Value (PPV) to be low.

The 4 stages of the new ischemia detector is as follows:

1. QRS detection, filtering, J point detection,

2. Neural Network beat classification,

3. Sliding adaptive window,

4. Merging.

The following procedures were carried out on each lead separately. In the merging stage, not described in this paper, the information from each lead is combined to aid clinicians to understand the information easily.

The data set consisted of 100 sample points from J-point to 400 ms after the J-point. J-point, or junction point (JP) is where QRS complex meets the ST segment, i.e. the beginning of the ST segment and therefore the end of the QRS complex.

ST-T complexes whose lengths were shorter than this window were padded out with zeros. This data dimension was reduced to 4 using PCA. These features became the input to a two layer MLP with 10 hidden units and one output unit. This number of hidden units was chosen empirically by comparing different model complexities. Sensitivity and specificity were used to quantify the quality of classification for each model complexity. The number of principal components were chosen so that 95% of the total variance of the data was represented by the chosen principal components.

11 hours worth of ECG recordings were used for the training and testing of the neural network. Out of this, 2.5% were used in training and the rest in testing. The data were chosen to include a similar number of normal and abnormal beats.

The author achieved aggregate average of 86% sensitivity to episodes. It is, however, noted that the lowest sensitivity value over all the records in ESTDB used for testing was 0%, while the highest is 100%. There is no mention of non-ischemic episodes nor of the fact that the class priors in the real data is different from that of his training and testing sets [12].

2.5 Tasoulis *et al.* 2002

This piece of work also starts with the use of PCA and performs classification using a neural network.

100 samples of the ST-T complex corresponding to 400 ms of data at 250 Hz are reduced to 5 dimensions using PCA. The J-point is taken at 60 ms after the R wave peak, but this is adjusted according to heart rate. The number of principal components was chosen so that over 98% of the total signal energy were represented by the principal components. A wavelet decomposition method was used to reduce the noise in the principal components.

The k-window clustering algorithm is used to determine the number of clusters in the feature space. Then a 4-layer neural network is used as local predictors for each cluster. For testing, each test

data point is assigned to a cluster based on a distance measure from each cluster. Then, that data point is classified by the neural network responsible for that cluster.

The k-window clustering algorithm works in two stages. First, a box in n dimension, where n is the dimension of the feature space, is moved across the space until the number of data points in the box reaches a certain stopping criteria. Then the box is expanded until the number of new data points included at each enlargement step is less than a certain limit. Then the expanded box is the boundary for a cluster. This expansion takes place in each dimension separately. The process of finding the number of clusters present is automated by starting with a sufficiently large number of boxes. Then, at the end of the process, overlapping boxes are merged.

On average, sensitivity of 78.4% was achieved over 6 testing records, over five different training schemes for the neural networks. The lowest sensitivity fell just below 60% and the best was over 90% [18].

The work above all used European ST Database as at least one of the sources of their data. European ST Database is very similar to the Long Term ST Database used in this project.

2.6 Zimmerman *et al.* 2003

Work by Zimmerman *et al.* was one of the entries for the Challenge. The data used in this paper is the same as for this project. Only the first lead is used for analysis. If the first lead is very noisy, the analysis is switched to the next lead.

The J-points were obtained from the 16 beat average annotation file. The author then extracts 8 J points before and after an event of interest. The events are marked up by experts as ischemic, heart rate related, axis shift or conduction change. 100 samples following the J-points are extracted as the data to be analysed. The data is then embedded into a reconstructed phase space according to the equation

$$\mathbf{x}_n = [x_n \; x_{n-\tau} \; \cdots \; x_{n-(d-1)\tau}],$$

where $n = (1 + (d-1)\tau) \ldots N$. The parameters d and τ were chosen empirically by the author to be 6 and 5 respectively. Embedding is done per J-point number (-8 to 7, total of 16) per channel per event type.

Gaussian Mixture Model (GMM) is trained for each embedding. Classification is achieved by maximising the log-likelihood of the data in the GMM.

The training set is constructed in such a way as to balance the number of events across the event types. That is to say, ischemic episodes, non-ischemic episodes, axis shift and conduction change happened at a roughly the same frequency in this training set. The events were assumed to have been detected.

A maximum average sensitivity of 80.6% was achieved for the validation set. The test set, as tested by the Challenge organiser, was found to be 63.8%, but the author attributes this to problems in translating the codes from one program language to another [20].

Researcher	Feature extraction		Classification			other
	PCA (n)	NLPCA (n)	rule	MLP	RBF	
Jager	yes (5)		yes			
Stamkopoulos		yes (2)			yes	
Laguna/ Gracia	yes (4)					
Papaloukas	yes (4)			yes		
Tasoulis	yes (5)			yes		k-window
Zimmerman						GMM and embedding

Table 2.1: Summary of literature survey. (n) for the PCA and NLPCA columns are the number of principal components chosen by the authors.

2.7 Summary of the chapter

Here, various techniques used by the researchers mentioned above are summarised in Table 2.1.

It can be seen that PCA is a popular and well established method for feature extraction, and 4 or 5 principal components are often used.

It is decided that this project will take a similar direction to that of Papaloukas *et al.* because the method is most simple and principled. The number of hidden units for the classifier network is empirically fixed in the paper by Papaloukas, but this project will choose a suitable model complexity based on the data to give a more tailored classifier for a given data set. Regularisation will be achieved by early stopping as opposed to the Bayesian technique employed by Papaloukas.

It is important to stress that none of the authors mentioned have taken into account the fact that in a piece of ECG recording, there will be far more normal beats than abnormal beats. Using neural network based methods and not being careful with class priors can produce misleading results.

Chapter 3

Database

The predecessor to the Long Term ST Database (LTSTDB) was the European ST Database (ESTDB) which contained ninety 2-hour recordings of ECG with predominantly ischemic ST episodes. Although this collection of records contributed greatly to the advancement in automated ischemia detection, there were a couple of shortcomings.

Firstly, the European database did not actively include non-ischemic events as they have no clinical importance. The lack of understanding of non-ischemic events lead to a high false-positive rate in the automated detectors. Secondly, the 2-hour excerpts were too short to observe more complicated ST changes with mixed causes. A longer set of recordings were necessary to observe such patterns with ischemic episodes nested within non-ischemic episodes or the "slow drift" ST deviation [6].

In contrast, LTSTDB contains eighty-six records lasting between 21 and 24 hours with many ischemic as well as non-ischemic ST events. The recordings are of eighty people, of which forty-three are available from PhysioBank as the training set and the remainder are held by the PhysioNet for testing the programs produced by the competitors.

3.1 Structure of LTSTDB

There are two types of data in LTSTDB; signal data and annotation data. Each of these are described here in turn. Along with the data are two other types of information; the header file, which can be read automatically by computer programs, and a file containing the summary of ST events in a patient record.

There are 34 records with 2 lead recordings and 9 records with 3 lead recordings. Of the 2 lead records, 11 are omitted from analysis as their leads are unknown. Another 1 is omitted as it has neither ischemic nor non-ischemic ST episodes. The project also looks at only the 2 lead recordings as a starting point.

Of the remaining 22 records, another 1 is omitted as the record has inconsistent labeling, where one lead is showing ischemic episode and the other is showing non-ischemic episode at the same time.

It was not possible to establish whether the labelling was inconsistent and wrong, or inconsistent but still valid, therefore this record was omitted from analysis.

Due to problems with computer hardware breaking down, further 4 records were lost, and another 2 had slightly different record format to the rest, making reading them slightly difficult. These six were again eliminated from analysis to save time. In total, 15 two-lead records were analysed for this project.

All of the 15 records used for analysis in this project have either ischemic episodes or non-ischemic episodes, but not both. The non-ischemic ST episodes in this database are all heart-rate related, that is, due to a change in heart rate.

The 2 lead records are recorded in a certain combination of leads:

lead 0	lead 1
ML2	MV2
MLIII	v4
v4	MLIII
v5	v2

where M stands for Modified. The positions of the electrodes forming these leads can be seen in Figure 1.2.

Each recording lasts up to 24 hours amounting to around 100 000 beats per lead per record. Typically, around 95% of these beats are normal and the remaining 5% are labeled as some kind of ST episode. See Table B.1 for the actual number of beats and ratio of episodes in each record.

3.1.1 Signal

The analogue ECG recordings were digitised at 250 Hz and amplitude scaling of 200 Analogue-to-Digital Converter (ADC) units per mV. The resolution of the digitisation is 12 bits over the range of $\pm$ 10 mV. The records were provided by research groups in the US, Slovenia and Italy [13]. The signals were meticulously annotated and validated by these three research groups in turn.

3.1.2 Annotation

There are two types of annotation used in this project; one for the entire beat, and the other for the ST segment. The first type are the automatic classification by ARISTOTLE, (`.ari`) and a manually corrected version of this (`.atr`). These annotations give the positions of the fiducial points of each beat, therefore the "positions" of each beat within the length of the recording. The classification provided by these annotations are not used in this project.

The second type are the ST segment measurements based on 16-second moving average (`.16a`), locations of ST episodes according to different standards, A (`.sta`), B (`.stb`) and c (`.stc`), the ST reference functions and deviation (`.stf`) and KLT projection of QRS complex and ST segments (`.klt`). The `.16a` annotations also give the positions of the J-points.

The heart rate is the number of heart beats per length of time. It is often given as beats per minute.

J-point is defined as a certain amount of time after the fiducial point (usually 120ms), but this interval varies with heart rate. As heart rate increases, the heart beat becomes shorter, therefore the interval between the fiducial point and J-point becomes shorter. The threshold heart rate values corresponding to different FP-JP intervals are summarised in Table A.3.

The ST level is the amplitude of the ST segment measured a certain time, determined by the heart rate, after the J-point. The amplitude of the ST segment is measured from the isoelectric level. The isoelectric level is the most flat part in the P-Q interval of that beat [6].

A significant ST shift is where a significant and sudden step-change in ST level is observed with simultaneous step-change in QRS morphology. These can be due to axis shift or conduction change [6].

Detailed account of the contents of the database and its structure are given in Appendix A.

3.2 Summary of the chapter

In this chapter, a brief description of the database used in this project, Long Term ST Database (LTSTDB) was given. The LTSTDB consists of digitised ECG signals and automatic and manual annotations. Of the 43 records available from PhysioBank, 21 will be used for analysis in this project. The detailed description of each data and annotation files in LTSTDB is given in Appendix A.

Chapter 4

Artificial Neural Networks Based Method

4.1 Motivation

The approach in this chapter is to attempt to classify normal and ST episode beats from the morphology of the beat only. Once successful classification is achieved, the normal class can be used to create a model of a normal beat. Then the record will be compared against this normal model, and the deviation will be analysed to give classification of ischemic or non-ischemic ST episode.

The experiments were conducted on per-patient, per-lead basis, therefore taking into account the fact that patient records were recorded on different combination of leads.

4.2 Data pre-processing

By inspection of the signals provided by the Challenge, it was decided that baseline removal was needed, but the signal was of good enough quality not to require bandpass filtering.

The baseline removal algorithm was based on that by Cardionetics [11]. Instead of detecting the R waves, the fiducial points given by the LTSTDB were used to locate each beat. Looking backwards from the fiducial point, the Q wave was chosen as the first minima in a sample window of 30 samples [19]. If the signal is noisy and has huge amplitude changes near the Q wave, it is possible that these noise can be erroneously picked up as the Q wave. In order to guard against such scenarios, the Q wave detector replaces any detected points which deviate from the mean by more than 4 standard deviations with the mean. The amount of deviation to tolerate, 4 standard deviations, was chosen empirically. This cannot take into account other problems such as slow drift of the baseline. Other methods that could have been employed was to check that each detected Q wave was within sensible range every few beats. Once the Q waves were detected a cubic spline was fitted through these Q waves and subtracted from the records.

Following the baseline correction, the data was separated into normal and ST episode sets using the `sta` annotation standard (see Appendix A.2.8). Then, training sets, validation sets and testing sets were randomly selected in the ratios 1/2, 1/3 and 1/6 respectively. In this process, beats were selected completely at random thereby breaking any time connection between the beats.

In practice, this is not the optimum way to separate out the data sets. Since beats in a given ST episode will share some similarities, a classifier trained on the dataset above will learn some peculiarity about a certain episode. In the validation and testing sets, it will find episodes which it has already "seen". In practice, a small section of the patient's recording will be used to as the training set, and the remainder will be testing sets. A practical classifier needs to be able to generalise across all episodes.

The ratio of normal to ST episode beats selected for each sets were consistent with the ratio of that patient record. Of course, in order to separate the beats in this way, normal and ST episode beats must already be labelled. The aim is, from these pre-labelled data, produce a classifier that can successfully distinguish between normal and ST episode beats such that for any new data obtained for the particular patient, it will be possible to label it automatically. In reality, the portion of data available to build the classifier with will be much smaller.

For initial examination, 2 records were chosen at random, one with ischemic episodes and one with non-ischemic episodes. The records used in the analysis are s20011,which has non-ischemic episodes, and s20021, which has ischemic episodes. s20011 has 14 episodes in lead 0 and 6 episodes in lead 1. s20021 has 20 episodes in lead 0 and 26 in lead 1 (see Table B.2 for a summary of episodes).

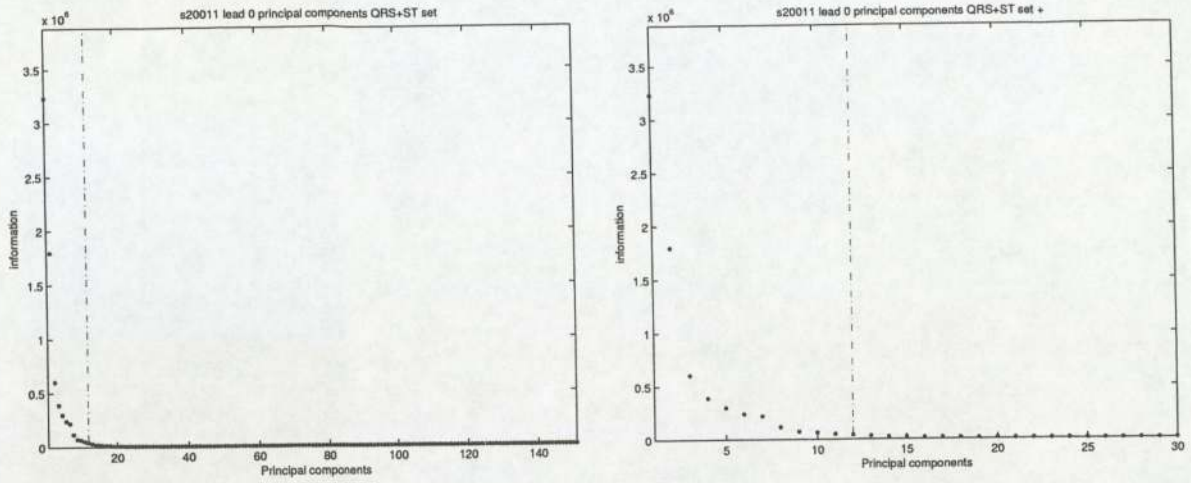
The records were then reconstructed into two multi dimensional data sets; one with 151 dimensions including the QRS complex and the ST segment and the other with 91 dimensions including the ST segment only. The first data set is chosen as 60 samples before the fiducial point, the fiducial point and 90 samples after the fiducial point. The second data set is chosen as the fiducial point and 90 samples after the fiducial point. Since the ST episodes do not necessarily occur at the same time in both leads, only one lead was considered at a time. In cases where both leads were used in the analysis, the ST episode beats were selected only from sections of the recording where both leads had ST episodes.

There are around 50 000 normal and 6 000 episode beats in the training set and around 30 000 normal and 4 000 episode beats in the validation set.

4.3 Feature extraction

Principal Component Analysis (PCA) was used to perform feature extraction on the data sets. PCA has been used in ECG analysis in the past and has shown to be a reliable method as described in Chapter 2. It is also very simple to implement, and therefore a good starting block onto which further methods can be built on.

PCA decomposes the data matrix into orthogonal components of the same dimensionality as the



(a) Eigenvalues

(b) Close-up

Figure 4.1: Eigenvalues of covariance matrices for s20011 lead 0 QRS + ST set.

data. This allows for the n components corresponding to the n largest eigenvalues of the covariance matrix of the data to be selected by inspection. Data can be projected into a space spanned by this smaller number of principal components for further analysis or for visualisation if $n = 2$ or 3.

First, the data is normalised to zero mean and unit variance. Then the covariance of the data is calculated;

$$\mathbf{C} = \mathbf{x}\mathbf{x}^T,$$

where $\mathbf{C}$ is the covariance matrix of the data matrix $\mathbf{x}$.

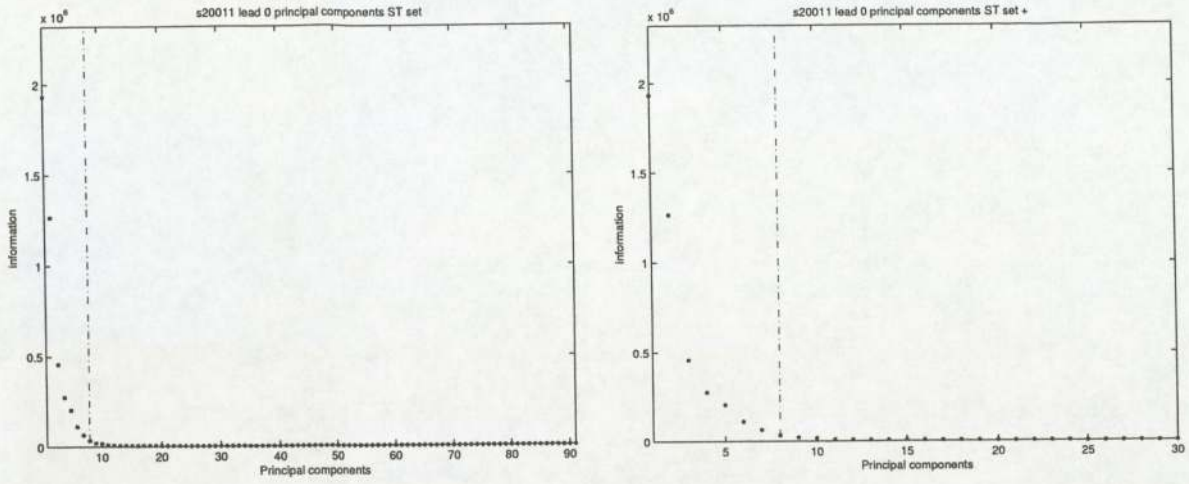
The n most dominant, principal components are chosen by selecting the eigenvectors of $\mathbf{C}$ corresponding to the n largest eigenvalues.

The approximate dimensionality of the data can be found by choosing n where the value of the eigenvalue suddenly drops.

Figure 4.1 show the eigenvalues of the covariance matrices of s20011 lead 0 QRS + ST set whole and close up. The dash-dot line is where the sum of eigenvalues to that point reaches 95% of the total information contained in the covariance matrix. For this set, 7 and 15 eigenvalues are chosen, as interesting points on either side of the dash-dot line.

Figure 4.2 shows the eigenvalues for s20011 lead 0 ST set. For this set, 5 eigenvalues are chosen, although this is just short of the dash-dot line. This is the same number of eigenvalues chosen by Jager *et al.* as described in Section 2.1 and is also as that chosen in the KLT data included in LTSTDB as described in detail in Appendix A.

An initial inspection of the data is possible by projecting the data using a pair of principal component as shown in Figure 4.3. The normal data points are shown in circles and abnormal points in crosses. In all of these principal directions, normal and abnormal beats appear overlapped. This



(a) Eigenvalues

(b) Close-up

Figure 4.2: Eigenvalues of covariance matrices for s20011 lead 0 ST set.

means that either they are acutally inseperable or that the boundary between the classes is a complex multidimensional manifold which cannot be seen in 2D.

4.4 Classification

A Multi-Later Perceptron (MLP) with a logistic output layer was trained as a classifier. The detailed mathematical description of MLP can be found in the book by Bishop [1]. The scaled conjugate gradient algorithm is used to train the network. A sketch of the schematic of an MLP is shown in Figure 4.4. The number of input units is determined by the number of principal components selected at feature extraction, nPC. The number of hidden units is determined from the data. There is only one output unit, giving 0 to 1 coding.

Early stopping was used to prevent over fitting of the model. This is a simple method but it is easy to implement and the processing can be done in batches requiring little attention until the results are produced. It is also a good method when there are abundant data samples as in this case.

The MLP was trained with various model complexities for a sufficiently long time so that the increase in generalisation error can be clearly observed. The total number of iterations were empirically chosen to be 3 000, and the training and validating errors were evaluated every 100 iterations.

The minimum validation errors for each model complexity were compared and the one giving the least error was chosen as the optimum model complexity. The optimum number of iterations was obtained by referring back to the log of generalisation error during the model training. An example of the training and validation errors is shown in Figure 4.5. The first figure show the training error in stars and validation error in circles. The training error never rises, but the validation error stops

decreasing after a while, then starts to increase. This is when overfitting starts to occur.

4.4.1 Classification results

Tables 4.1, 4.2 and 4.3 summarise the classification results for records s20011 and s20021. The tables in this project have the form:

record	lead	Conf mat		*se*	*sp*
sXYYYZ	0/1	normal as normal	false positive	*se*	*sp*
		false negative	ST as ST		

or in some cases,

record	lead	Conf mat		$\frac{se}{sp}$	change
sXYYYZ	0/1	normal as normal	false positive	*se*	change in *se* w.r.t. other tables
		false negative	ST as ST	*sp*	change in *sp* w.r.t. other tables.

Comparing against the average result obtained by Papaloukas *et al.* of 86% sensitivity, the ST segment set (Table 4.3) gives considerably worse classification. The QRS + ST data set with 7 principal components perform better, but still not good enough. The QRS + ST data set with 15 principal components perform better or equally well, except for s20021 lead1.

Understanding why beats are wrongly classified can help identify methods to improve classification.

Uncertainty

A simplistic explanation to why beats are classified wrongly is that beats just before or after an ST episode-normal boundary are hard to identify as one or the other. Such beats, then, are expected to be classified with high uncertainty.

To verify this, the positions of the false negatives and the false positives are shown in Figure 4.6 with respect to the positions of the actual ST episodes for s20011 lead 0. The whole recording is represented in the graph, starting at the top left hand corner and finishing in the bottom right hand corner. There is a top-hat shape blip in the plot where there is an episode. Everywhere else is flat. The stars are where there are false positives and the circles are where there are false negatives. The amplitude is chosen arbitrarily with no significance except for to show the three types of information clearly.

The close up plot is of the first ST episode in s20011 lead 0. The false classifications are *not* necessarily at the class boundaries.

To determine the effect of the uncertain classifications on the quality measures, data points that were uncertainly classified were eliminated from the confusion matrices. The classifier is given labeling such that ST episodes are 1 and normal beats are 0. Therefore, uncertain beats are those whose network output lay between 0.33 and 0.67.

Of interest are the bold-font validation sensitivities and the changes corresponding to them. The changes are with respect to the original set of results in Table 4.1 to Table 4.3.

Some of the data sets show rise in sensitivity as the result of eliminating uncertain classifications from analysis. For example, s20011 lead 0 training set for QRS + ST set with 15 principal components

record	lead	Training set				Validation set			
		Conf mat		*se*	*sp*	Conf mat		*se*	*sp*
s20011	0	42887	680	73.8%	98.4%	28517	534	**71.1%**	98.2%
		1538	4323			1130	2779		
	1	47393	161	81.6%	99.7%	31510	149	**80.6%**	99.5%
		350	1550			245	1021		
s20021	0	43062	56	83.9%	99.9%	28667	61	**80.8%**	99.8%
		134	696			107	449		
	1	42261	103	60.9%	99.8%	28138	118	**54.8%**	99.6%
		612	955			470	569		

Table 4.1: Classification result for QRS + ST data set with 7 PCs.

record	lead	Training set				Validation set			
		Conf mat		*se*	*sp*	Conf mat		*se*	*sp*
s20011	0	43119	448	88.2%	99.0%	28609	442	**83.2%**	98.5%
		693	5168			658	3251		
	1	47433	121	88.7%	99.8%	31546	113	**87.2%**	99.6%
		215	1685			162	1104		
s20021	0	43069	49	82.2%	99.9%	28689	39	**84.4%**	99.9%
		484	682			87	469		
	1	42245	119	66.9%	99.7%	28102	154	**60.9%**	99.5%
		519	1048			406	633		

Table 4.2: Classification result for QRS + ST data set with 15 PCs.

record	lead	Training set				Validation set			
		Conf mat		*se*	*sp*	Conf mat		*se*	*sp*
s20011	0	42812	772	66.4%	98.2%	28554	511	**65.0%**	98.2%
		1970	3894			1367	2543		
	1	47415	158	76.0%	99.7%	31535	140	**73.2%**	99.6%
		456	1446			340	927		
s20021	0	43064	79	74.0%	99.8%	28684	59	**74.9%**	99.8%
		217	616			140	418		
	1	42308	75	50.6%	99.8%	28199	71	**49.5%**	99.8%
		776	794			527	516		

Table 4.3: Classification result for ST segment data set with 5 PCs.

record	lead	Training set				Validation set			
		Conf mat		*se* *sp*	change	Conf mat		*se* *sp*	change
s20011	0	42122	286	76.8%	↗ 3.0%	28018	245	**74.6%**	↗ 3.5%
		1144	3788	99.3%	↗ 0.9%	841	2470	99.1%	↗ 1.0%
	1	47274	88	83.7%	↗ 2.1%	31432	91	**83.3%**	↗ 2.6%
		277	1425	99.8%	↗ 0.2%	187	932	99.7%	↗ 0.2%
s20021	0	43009	33	85.6%	↗ 1.7%	26437	41	**83.1%**	↗ 2.3%
		111	658	99.9%	↗ 0.1%	87	427	99.9%	↗ 0.1%
	1	42135	41	60.6%	↘ **0.3%**	28013	59	**56.4%**	↗ 1.6%
		550	847	99.9%	↗ 0.1%	411	531	99.8%	↗ 0.2%

Table 4.4: Classification result for QRS + ST data set with 7 PCs without uncertain classifications. Change with respect to Table 4.1.

record	lead	Training set				Validation set			
		Conf mat		*se* *sp*	change	Conf mat		*se* *sp*	change
s20011	0	42769	221	91.3%	↗ 3.2%	28394	286	**86.1%**	↗ 2.9%
		466	4907	99.5%	↗ 0.5%	502	3108	99.0%	↗ 0.5%
	1	47319	66	90.9%	↗ 2.2%	31479	82	**88.8%**	↗ 1.6%
		160	1591	99.9%	↗ 0.1%	131	1037	99.7%	↗ 0.1%
s20021	0	43020	28	83.7%	↗ 1.6%	28663	26	**85.9%**	↗ 1.6%
		127	654	99.9%	0	74	451	99.9%	0
	1	42086	63	67.6%	↗ 0.8%	28008	98	**63.0%**	↗ 2.1%
		463	968	99.9%	↗ 0.1%	350	597	99.7%	↗ 0.2%

Table 4.5: Classification result for QRS + ST data set with 15 PCs without uncertain classifications. Change with respect to Table 4.2.

record	lead	Training set				Validation set			
		Conf mat		*se* *sp*	change	Conf mat		*se* *sp*	change
s20011	0	41806	285	68.5%	↗ 2.1%	27895	201	**66.8%**	↗ 1.8%
		1483	3230	99.3%	↗ 1.1%	1057	2127	99.3%	↗ 1.0%
	1	47249	81	77.7%	↗ 1.6%	31417	85	**75.0%**	↗ 1.9%
		379	1318	99.8%	↗ 0.2%	285	857	99.7%	↗ 0.2%
s20021	0	42998	41	75.8%	↗ 1.9%	28637	29	**77.6%**	↗ 2.6%
		179	562	99.9%	↗ 0.1%	110	380	99.9%	↗ 0.1%
	1	42228	28	47.6%	↘ **3.0%**	28106	40	**47.8%**	↘ **1.6%**
		729	661	99.9%	↗ 0.1%	496	454	99.8%	↗ 0.1%

Table 4.6: Classification result for ST data set with 5 PCs without uncertain classifications. Change with respect to Table 4.3.

record	lead	Validation set			
		Conf mat		*se*	change
				sp	
s20011	0	25565	3500	94.2%	↗ 29.2%
		228	3682	88.0%	↘ 10.2%

Table 4.7: Classification result for ST data set with 5 PCs with decision boundary altered to 0.1. Change with respect to relevant values in Table 4.3.

in Table 4.5 show rise in sensitivity of 3.2%. However, there are some others that resulted in the sensitivity being lowered by this operation such as s20021 lead 1 training set for ST set with 5 principal components in Table 4.6.

The ST set in particular (Table 4.6) shows considerably lower sensitivity than results obtained by Papaloukas, although this is similar to the data set used by Papaloukas *et al.*

Eliminating uncertain classifications from analysis does not contribute significantly in improving classification overall.

ROC curve

Another way to consider is by inspection of the Receiver Operating Characteristic (ROC) curve [9]. The ROC curve plots sensitivity against (1-specificity).

Figure 4.7 shows the ROC curve for s20011 lead 0 ST set with 5 principal components. An ideal ROC curve will approach very close to the top left corner of the graph. It is possible to move the decision boundary of the classifier such that better classification is achieved.

For example, the current decision boundary of 0.5 which is shown as the circle can be moved to the star which places the decision boundary at 0.1.

Table 4.7 shows dramatic increase in sensitivity at a relatively small cost in specificity. The change is with respect to the original set of results as shown in Table 4.3.

However the problem with adjusting the decision boundary using the ROC curve is that there is no principled manner in which to decide how much you can move the decision boundary before the outcome becomes unjustified.

4.4.2 Different data set

When one of the classes in a classification problem has a far larger data set than the other, that is to say, when the class priors are very asymmetric, the neural network output is biased in such a way that the majority class is classified correctly even if it is at the expense of the minority class. Since in the real world, normal beats far outnumbers the episode beats, when the data set is constructed in the same ratio as the real world, the classifier cannot classify the minority class, which, unfortunately is the more important in this case [17].

Another, more principled method to obtain better classification is by changing the selection of the data set. Instead of making the ratio of normal beats to ST episode beats the same as that in the real

record	lead	Training set				Validation set			
		Conf mat		*se* *sp*	change	Conf mat		*se* *sp*	change
s20011	0	5866	98	66.2%	↘ 0.2%	3820	90	66.2%	↗ 1.2%
		1983	3881	98.3%	↗ 0.1%	1323	2587	97.7%	↘ 0.5%
	1	1895	7	68.6%	↘ 7.5%	1260	7	69.1%	↘ 4.0%
		598	1304	99.6%	0	391	876	99.4%	↘ 0.1%
s20021	0	832	1	77.7%	↗ 3.7%	555	3	77.1%	↗ 2.2%
		186	647	99.9%	↗ 0.1%	128	430	99.5%	↘ 0.3%
	1	1568	2	48.8%	↘ 1.8%	1035	8	49.2%	↘ 0.3%
		804	766	99.9%	0	530	513	99.2%	↘ 0.5%

Table 4.8: Classification result for ST data set with 5 PCs with 50:50 data set. Change is with respect to Table 4.3.

record, they are forced to 50:50.

Now, to take into account the fact that this ratio of 50:50 is very different from that of a typical real record, the output of the classifier is adjusted [16].

The real class priors are

$$P(C_{ST}) \sim 0.05,$$
$$P(C_{NM}) \sim 0.95,$$

where the subscripts stand for "ST episode" and "normal" respectively. The data set skews this to

$$P'(C_{ST}) = 0.5,$$
$$P'(C_{NM}) = 0.5.$$

The true, compensated posterior probability becomes

$$P(C_{ST} \mid \mathbf{x}) = \frac{A}{(A+B)}, \tag{4.1}$$

where

$$A = \frac{(Network\ output).P(C_{ST})}{P'(C_{ST})},$$
$$B = \frac{(1-(Network\ output)).(1-P(C_{ST}))}{(1-P'(C_{ST}))}.$$

Table 4.8 summarises the outcome of the classification using the 50:50 data set. The change is with respect to Table 4.3. There is limited improvement in the sensitivity. Some data sets, for example, s20011 lead 0, show a reduction in both sensitivity and specificity. However, this is the preferred method as this is more principled than adjusting the decision boundary using the ROC curve.

record	lead	Uncompensated			Compensated			
		Conf mat		*se* *sp*	Conf mat		*se* *sp*	change
s20011	0	1531	426	93.3%	1793	164	65.4%	↘ 27.9%
		131	1826	78.2%	677	1280	91.6%	↗ 13.4%
	1	599	36	96.9%	624	11	68.6%	↘ 28.2%
		20	615	94.3%	193	422	98.3%	↗ 3.9%
s20021	0	254	25	94.6%	274	5	69.9%	↘ 24.7%
		15	264	91.0%	84	195	98.2%	↗ 7.2%
	1	200	324	78.1%	446	78	52.3%	↘ 25.8%
		115	409	38.2%	250	274	85.1%	↗ 47.9%

Table 4.9: Comparison between uncompensated and compensated output of the classifier for ST set with 5 PCs in the 50:50 test set. Change is with respect to the uncompensated values.

L. Tarassenko [16] gives a warning, however, that Equation 4.1 is not a fix-all method. If the smaller class prior is of the same order as the error in the network, this method will not work. In such cases, novelty detection methods instead of classification methods should be used.

Table 4.9 shows the effect of compensation using Equation 4.1 using the 50:50 test set. The change in the far right hand column is for the compensated sensitivity and specificity with respect to the uncompensated values. The specificity values rise a significant amount, but the sensitivity drops a far larger amount after the compensation, in general. This means that it is possible to classify between normal and episode beats with over 90% sensitivity and specificity provided there are equal number of samples in each class. Since the authors in Chapter 2 did not take into account the skew in the class priors, the result obtained in this project before adjusting the class priors should be compared with the results of the other authors. Clearly, the results obtained in this project are comparable or better than the other authors.

4.5 Improving the classification

There are a few methods by which the classification achieved above can be improved. One approach is to improve the input to the classifier, and the other to improve the classifier itself.

4.5.1 Additional features

In addition to the principal components, it is possible to include the information about the maximum value of the ST segment and the minimum value of the ST segment for each beat as features. This is because, it is expected that the PCA is good at extracting information about the ST segment along the time axis, but not in the measurement amplitude [3]. If this is the case, adding the information about the amplitude should improve the classification.

The results are summarised in Table 4.10. There are some data set whose classification improved, but in others the classification got worse. Therefore, it is inconclusive whether or not this method is useful in improving the classification. It is suspected that by adding the peak and trough information to

record	lead	Training set				Validation set			
		Conf mat		*se* *sp*	change	Conf mat		*se* *sp*	change
s20011	0	5654	210	73.0%	↗ 6.6%	3776	134	72.3%	↗ 7.3%
		1585	4279	96.4%	↘ 1.8%	1083	2827	96.6%	↘ 1.7%
	1	1808	94	62.2%	↘ 13.8%	1207	65	61.0%	↘ 12.2%
		719	1183	95.1%	↘ 4.6%	494	773	95.3%	↘ 4.3%
s20021	0	817	16	70.1%	↘ 3.8%	551	7	75.6%	↗ 0.7%
		249	584	98.1%	↘ 1.7%	136	422	98.7%	↘ 1.0%
	1	1565	5	52.2%	↗ 1.7%	1036	7	53.9%	↘ 4.4%
		750	820	99.7%	↘ 0.1%	481	562	99.3%	↘ 0.4%

Table 4.10: Classification result for ST data set with 5 PCs with peaks and troughs. Change is with respect to Table 4.3.

record	Training set				Validation set			
	Conf mat		*se* *sp*	change	Conf mat		*se* *sp*	change
s20011	1609	6	88.2%	↗ 17.0%	1072	7	87.7%	↗ 20.0%
	191	1426	99.6%	↗ 0.7%	118	961	99.8%	↗ 0.5%
s20021	815	1	89.3%	↗ 28.5%	543	1	91.0%	↗ 27.1%
	75	741	99.9%	↗ 0.1%	58	486	99.8%	0

Table 4.11: Classification result for ST data set with 5 PCs with data fusion. Change is with respect to Table 4.3.

the records whose sensitivity dropped as a result, for example, s20011 lead 1, the additional information provided inter-class similarity rather than intra-class similarity.

4.5.2 Data fusion

Data fusion is where information from different sources are concatenated. In this case, principal components from both leads are used as inputs to the MLP. The MLP is modified as show in Figure 4.8. Of course, this is only possible for records which have episodes in both leads and those that have episodes in both leads with portions of, if not all of, the episodes occurring at the same time.

This is expected to give the MLP more information about the similarity between ST episode beats and help it distinguish the difference between ST episode beats and normal beats.

The result of this experiment is summarised in Table 4.11. The change is with respect to the average over the leads in Table 4.3. Although the change in specificity is small, sensitivity rise greatly for all cases. This method produces results comparable to those by other researchers. However, the problem with this method is that it is inapplicable to records with episodes in only one lead, or episodes in the leads not overlapping.

4.5.3 All records

Classification using ST segment data set with 5 principal components was repeated on all records. Here, the interesting records are shown in Table 4.12. The results for all records are summarised in

record	lead	Training				Validating (50:50)			
		Conf mat		*se*	*sp*	Conf mat		*se*	*sp*
s20161	0	13052	462	64.0%	96.6%	8599	371	64.0%	95.9%
		4866	8648			3232	5738		
	1	1047	2	82.5%	99.8%	694	5	83.5%	99.3%
		184	865			115	584		
s20241	0	9404	159	84.4%	98.3%	6263	114	83.7%	98.2%
		1488	8075			1041	5336		
	1	4371	22	97.8%	99.5%	9214	17	97.7%	99.4%
		96	4297			68	2863		
s20321	0	122	0	100.0%	100.0%	82	0	100.0%	100.0%
		0	122			0	82		
	1	239	0	42.3%	100.0%	159	1	50.0%	99.4%
		138	101			80	80		
s20521	0	1049	3	18.3%	99.7%	697	3	15.7%	99.6%
		859	193			590	110		
	1	65	0	20.0%	100.0%	41	2	16.3%	95.3%
		52	13			36	7		

Table 4.12: 50:50 data set with compensation experiment for selected records.

Appendix C. This experiment gave an aggregate average sensitivity of 70.5%, which is lower than the 86% achieved by Papaloukas *et al.*.

The outcome is very varied, from as low as 15.7% for s20521 lead 0 to 100.0% for s20321 lead 0 sensitivity. Specificity is high for all cases, hardly falling far below 99%. Again, the strong influence of the asymmetry of the class priors can be seen.

The results obtained using the test set are summarised in Table C.2. The left half of the table shows the results when the test set has the same class priors as the real records. The network is tested with real class priors because when the classifier is used on a real data set, it will not be possible to construct a 50:50 set.

The aggregate average sensitivity is 39.4% which is very low compared to the 86% by Papaloukas *et al.*. This is also a reduction in sensitivity of 31.1% from the validation set. This reduction is not surprising, as the validation set has the class priors skewed to 50:50, whereas the test set has the same class priors as the whole record. Although the overall average sensitivity is low in the test set, there are some records and leads showing over 86% sensitivity, such as s20021 lead 0 of 100%, s20101 lead 1 of 99.2%. The lowest sensitivity of 0% was obtained in s20241 lead 1 and s20521 lead 1. The latter is not surprising, as neither the training nor the validation showed very high sensitivity in this record. However, the s20241 is unexpected, as the training sensitivity is 97.8% and the validation sensitivity is 97.7%, both of which are high.

For comparison, results obtained using 50:50 prior test set is given in the right half of Table C.2. The aggregate average sensitivity is 69.8%, which is much higher than that for the test set with real class priors.

The data fusion experiment repeated on all records where there are episodes in both leads at the same time are summarised in Table E.1. Record s20521 has episodes in both leads but it was found

the episodes do not occur at the same time. As it was established using just s20011 and s20021, data fusion produces very good results with all validation sensitivities over 89%. The aggregate average sensitivity for this method is 97.3%.

The results obtained using the test set are summarised in Table E.2. As with the per-lead case, the left half show the result from the test set with real class prior. The aggregate average sensitivity was 75.6%, which is still lower than the single lead results obtained by Papaloukas *et al.* This is result is owing to records s20011 and s20021 whose sensitivities are 3.7% and 1.1% respectively. All other records showed 100% sensitivity.

The right half of Table E.2 is the test results obtained using 50:50 class prior test set. The aggregate average sensitivity is 84.5%.

The 50:50 test set helps to show the generalisation ability of the network. Since the network has not "seen" the test set in its training, if the network is over-fitted to the training and validation set, the classification results for the test set with 50:50 class priors would be considerably lower than that of the training set or the validation set. The average sensitivity for the per-lead experiment using the 50:50 set is only 0.6% lower than the that of the training set, so the generalisation is good in this case. For the data fusion case, the training set shows 12.8% higher sensitivity than the 50:50 test set, so the generalisation is slightly worse than the per-lead experiment.

4.6 Discussion and analysis

4.6.1 Analysis of results

Clearly, the data fusion method with 50:50 data set outperforms any other method attempted in this project. There are two possible explanations for this. One is that with the introduction of information from two leads, the classifier was provided with more detailed information on discriminating between normal and ST episode beats. The other is that ST episodes seen in both leads are better characterised than those occurring only in one lead.

The next highest sensitivities were obtained when the 50:50 classification was not compensated. This is still a "wrong" set of results, as the decision boundary drawn by the classifier is biassed due to the skewing of the class priors. However the fact that the sensitivities are high before compensation is a significant finding because it shows that it is actually possible to distinguish between normal and ST episode beats solely from beat morphology, but this ability falls considerably when taking into account the class priors of the real world. Stamkopoulos, Laguna, Papaloukas and Tasoulis who all used some kind of automatic classification method using neural networks, as described in the literature survey in Chapter 2, do not take into account the asymmetry of the class priors. Their results may show lower sensitivities if they compensated the network decision boundary taking into account the class priors in the real-life records.

Since sensitivity is a function of the positive classification of ST episode beats and false negatives, as shown in Equation 1.1, a rise in sensitivity represents a decrease in false negatives. The number

of false negatives is important in clinical problems, therefore, the sensitivity gives the quality of the classification.

Similarly, a rise in specificity means a reduction in false positives. Since the number of false positives are not as crucial, a rise in specificity is of secondary importance. This is not to say that specificity can be neglected. When the sensitivity is raised by changing the decision boundary using the ROC curve, for example, the specificity is compromised. Instead, what is required is for both false negatives and false positives to be reduced, in other words, to raise both sensitivity and specificity. This is required to raise confidence in automated systems such as the detection of ischemia using ECG, so that the system only alerts humans when there is a real abnormality, and not "cry wolf" too often.

The first set of experiments in Section 4.4.1 show that QRS + ST set performs better than the ST set. Although it is widely known that myocardial ischemia is observed in the ST segment, this suggests that there are some observable features in the QRS complex that appear when there is ischemia. Since the ST segment is studied because the recovery signal of the atria are obscured by the ventricular contraction, it could be that abnormality in atrial recovery due to ischemia affects ventricular contraction signal as well. The remainder of the experiments were done using the ST set in order to compare the results with existing work by other researchers.

4.6.2 Specific cases

There are some interesting results in Table 4.12. For example, there is the extremely good result of 100% sensitivity and specificity of lead 0 s20321. Notice the relatively small number of data points in the training and the validation sets. This occurs when there are only a small number of ST episode beats in a record, as, to make the class priors 50:50, the number of normal beats are bound by the number of episode beats, even though there are abundant beats available in each record. This size of the data set could be responsible for the results.

On the other hand, there is lead 1 of s20521 which has equally small data sets but perform extremely poorly. In fact, both leads in s20521 performs very poorly regardless of the size of the data sets. On further inspection of both leads of s20521 using the same method to identify uncertain classifications as in Section 4.4.1 and studying the effects of uncertain classifications, it became apparent that most of false positives and positive classifications (ST as ST) were uncertain. This makes the sensitivity even lower. The expected cause of this is that there is more similarity between classes than within classes, so the network cannot capture the class boundary correctly. What can also be noted is that this is the only record in the analysis where there are episodes in both leads, but not at the same time. There may be something unusual about this record which has not been understood so far.

For the data fusion experiment, records s20011 and s20021 produced very low sensitivities in the test sets of 3.7% and 1.1% respectively. These records show relatively low sensitivity in the training set and the validation set compared to the rest of the records. In addition, these records do not show good results in the per-lead analysis either, although there are worse records in the whole experiment. In

these records, the ST episodes may not be clearly characterised. The features used in the experiments were insufficient to describe the inter-class dissimilarity in s20011 and s20021.

4.7 Summary of the chapter

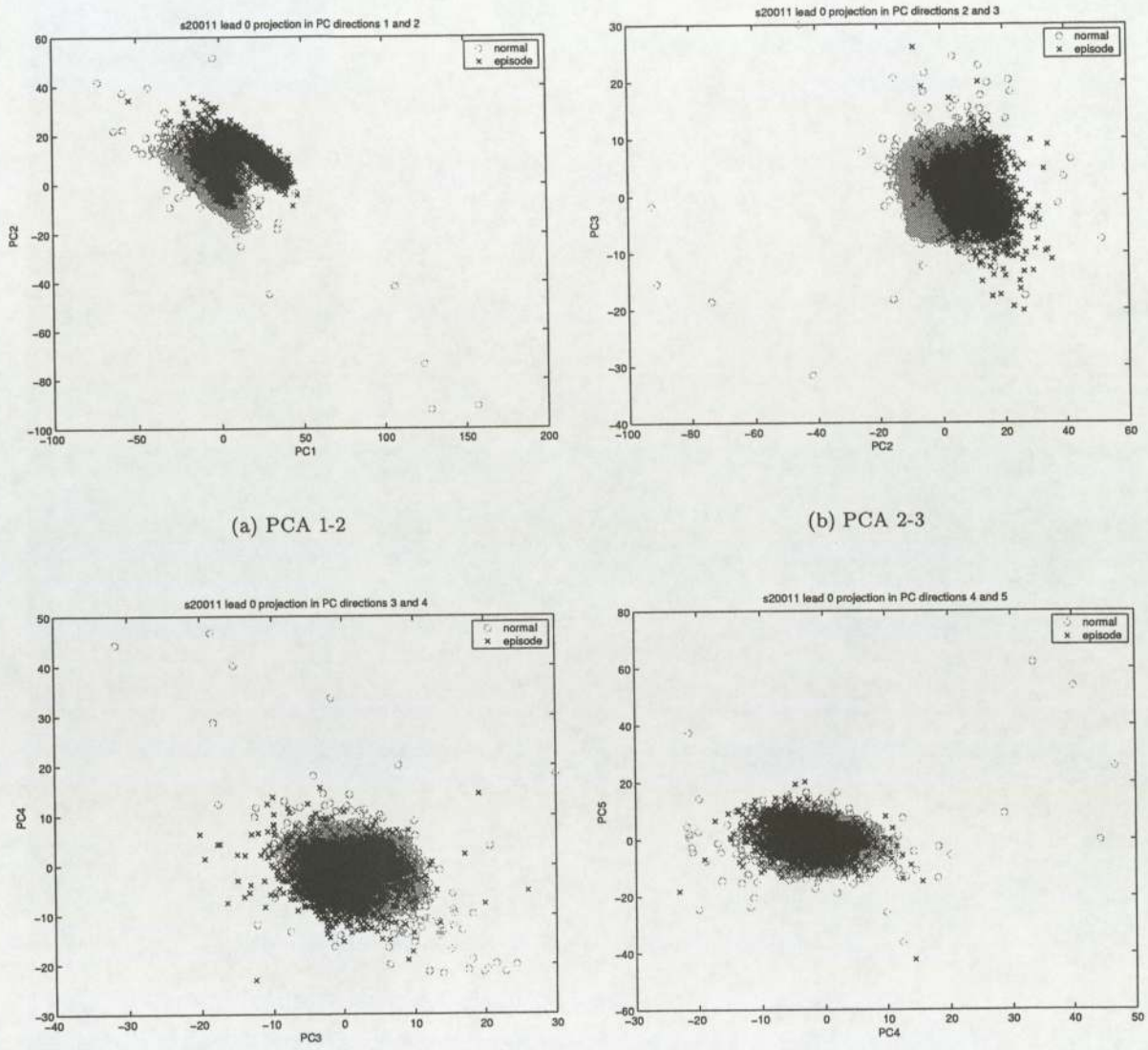
This chapter described the methods used for classifying normal and ST episode beats using MLPs. A detailed description of pre-processing of the data, feature extraction and training and validating of the classifier was given.

The classification results were quantified using sensitivity and specificity values and by showing the results as confusion matrices. The results were compared for different data set and different class priors.

The algorithms were developed first by considering two records only, then expanded to incorporate all 21 records chosen to be analysed for this project.

Attempts were made to improve the classification by taking into account uncertain classifications, ROC curves, additional information and data fusion technique.

The outcomes were analysed and the origins of the results were discussed.



(a) PCA 1-2

(b) PCA 2-3

(c) PCA 3-4

(d) PCA 4-5

Figure 4.3: Projection of data into 2D. Normal points in circles and abnormal points in crosses.

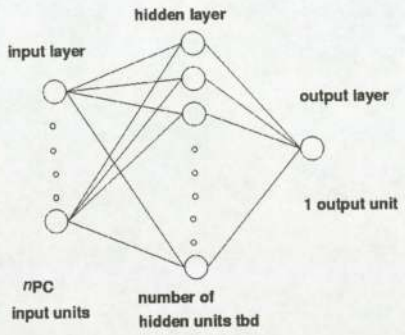
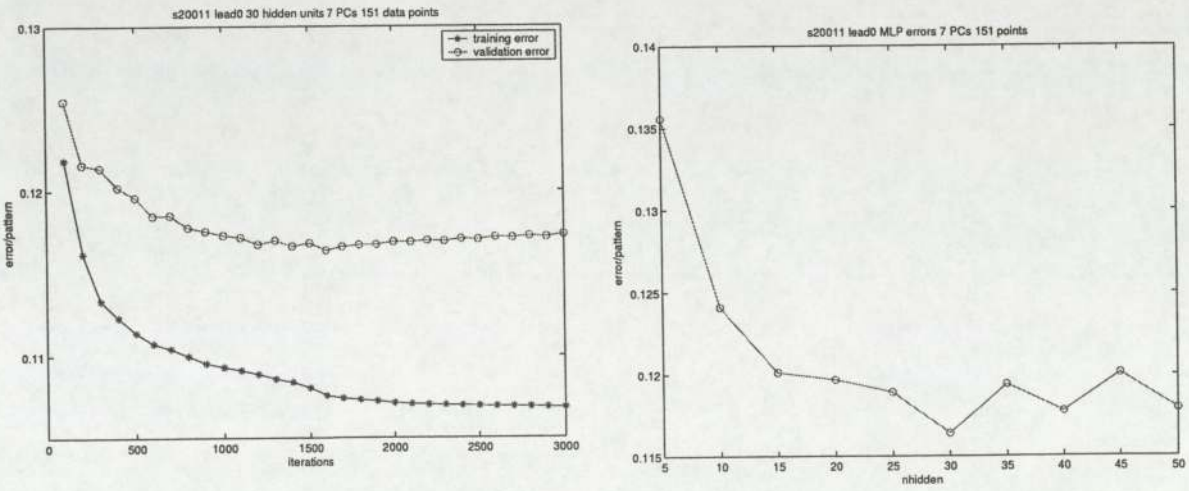


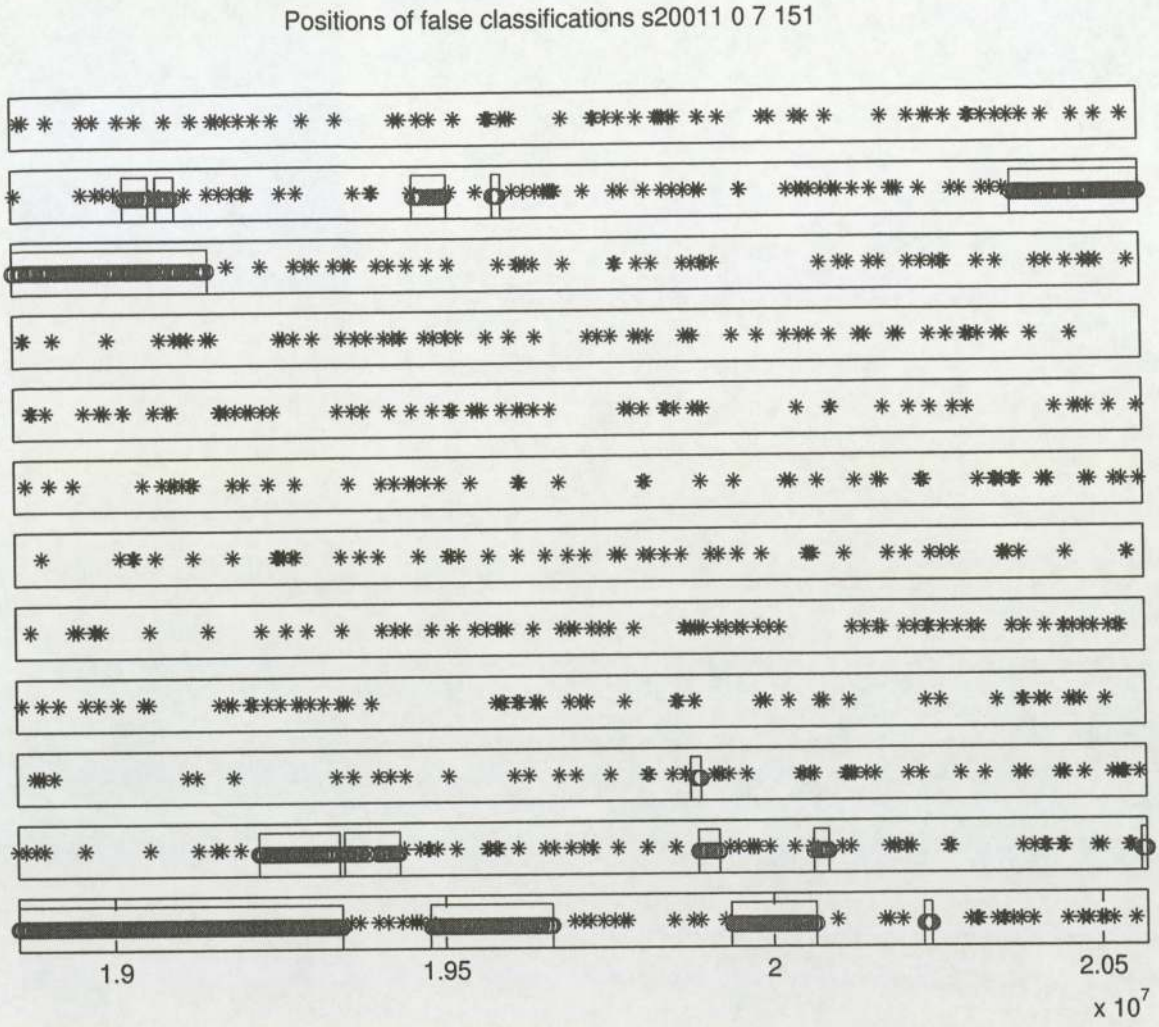
Figure 4.4: Sketch of MLP used for classification.



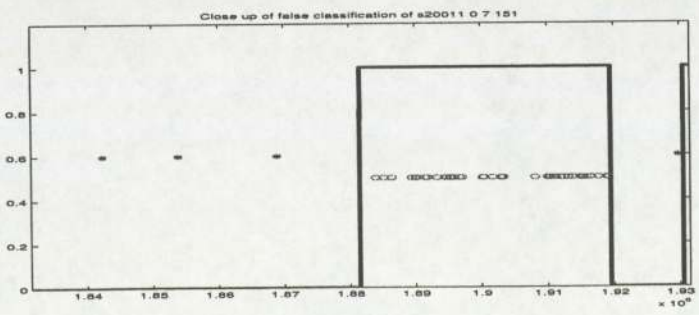
(a) 30 hidden units

(b) All hidden units

Figure 4.5: Errors during training MLP. Error is compare per model complexity, then for all model complexity. Crosses for training set and circles for validation set.



(a) Positions of false classifications



(b) Close up

Figure 4.6: Positions of false classifications in s20011 lead 0. Stars are false positives and circles are false negatives.

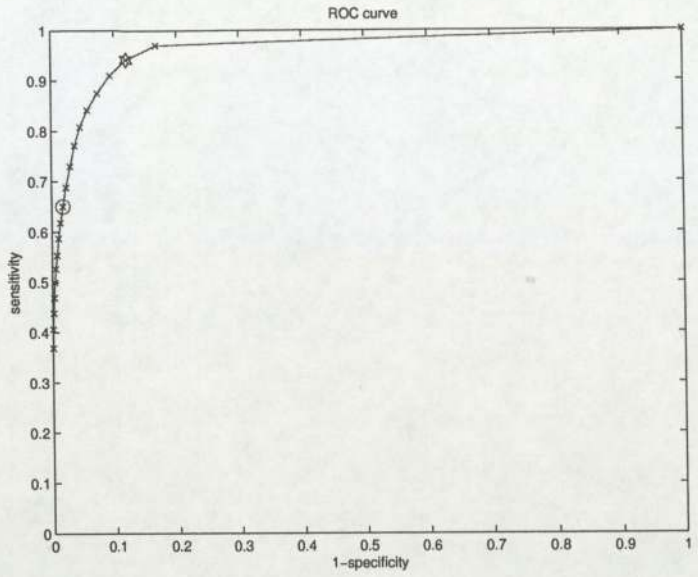


Figure 4.7: ROC curve for s20011 lead 0 ST set with 5 principal components.

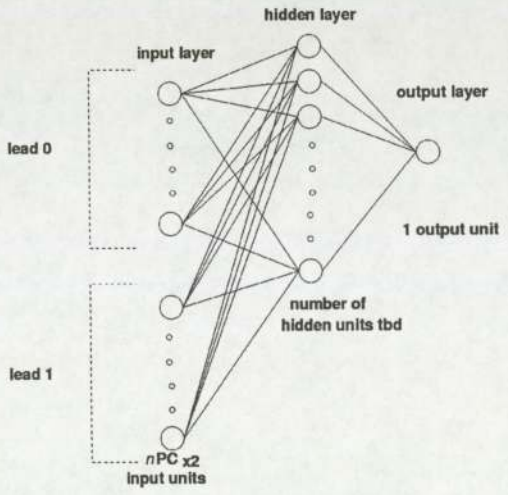


Figure 4.8: Sketch of MLP for classification using data fusion.

Chapter 5

Finale

5.1 Summary of the project

This project, based on the PhysioNet-Computers in Cardiology Challenge 2003, has considered the problem of distinguishing ischemic ST episodes from non-ischemic ST episodes from the analysis of ECG alone. Although the Challenge states that the detection of ST episodes is not necessary, this project has considered the problem of detecting ST episodes as the first step of analysis to make a generalised, principled approach to the problem.

The research methodology was constructed on the basis of existing work on the field. The experimental results were compared against these work to assess their quality.

The data was pre-processed using cubic spline subtraction method to remove baseline wander. Data sets were created by selecting QRS complexes and ST segments. Two data sets were used and compared; QRS + ST set and ST set.

Feature extraction was performed using PCA. The number of principal components to be used for analysis were chosen by inspection of the eigenvalues.

MLPs with logistic outputs were trained as classifiers. The early stopping method was employed to avoid over-fitting. An MLP was trained per model complexity and the errors compared in order to choose the optimum model complexity and iteration number.

Two methods to take into account the asymmetry of the class priors were compared. One was to construct the data set such that the class priors are the same as that in the real world. The other is to construct the data set such that the class priors are 50:50 but to skew the output of the classifier according to the real class priors. The difference between the network training class priors (50:50) and the real class priors ($\sim$ 95:5) were not taken into account by the work by other researchers as described in Chapter 2.

Then, in order to improve the classification results obtained, additional information was used as features going into the classifier. Peaks and troughs of each ST segment were added.

Finally, features from both leads were concatenated and put into the classifier.

5.2 Conclusion

This project has established that it is possible to separate normal beats from ST episode beats with over 80% sensitivity from beat morphology in ECG alone. The sensitivity can be raised to an average of 97.3% if two leads are available for analysis using the data fusion technique. However, the sensitivity can fall as low as 16.3% if the class prior asymmetry of a real record is taken into account.

When the asymmetry of the class prior are taken into account by testing the classifier with data whose class priors are not 50:50, the sensitivity falls as low as 0%, although some records still show sensitivity of 100%. The aggregate average sensitivity is 39.4%, which leads to the conclusion that this method is not effective in general.

When the asymmetry of the class priors are taken int account for the data fusion experiment, the aggregate average sensitivity fell to 75.6%, although the majority of the records showed 100% sensitivity.

It was also decided that QRS + ST set performs better under the same condition compared to the ST segment set. However, since there are much research into classification using just the ST segments, detailed experiments were carried out on the ST segment set to compare results.

The method of skewing the class priors before classification was rejected because the class boundaries will be drawn to maximise classification of the majority class even if that is at the expense of the minority class. Outputs of the classifier can be improved by adjusting the decision boundary with the aid of an ROC curve, but there is no principled manner in which to perform this. The method of skewing the class priors after classification was adopted, as it leads to show that classification was possible solely on the basis of beat morphology.

This project has shown that ST episode detection can be achieved using neural networks solely on the beat morphology, but that the sensitivity to episodes is significantly reduced when taking into account the skewness of the class priors in the real world. This issue has not been addressed by the researchers mentioned in Chapter 2 and can be an interesting issue to consider for future work in heart beat classification.

This project has not taken into account the numerous ST shifts which are also annotated in the LTSTDB. An ST shift is an instantaneous event where the ST level shifts for more than $50\,\mu$V. Because ischemia is only recognised in episodes, the shifts were excluded from analysis in this project. The positions or the existence of shifts may be incorporated into analysis to attempt to improve classification.

5.3 Ongoing and future work

For future work, a model of the normal beats can be created on the basis of this classification. All beats in the records then can be compared against the model to quantify the deviation. Further analysis of the deviation is expected to reveal a satisfactory classification of the types of ST episodes.

There are also different approaches that can be investigated.

5.3.1 Novelty detection

As mentioned in Section 4.4.2, Equation 4.1 is only effective if the smaller class prior is still larger than the error in the network output. There are some records used in this project where the class prior of the ST episode beats are of the order 10^{-3} (see Table B.1). Tarassenko shows that this could be too small for Equation 4.1 to work [16]. The detection of ischemia in ambulatory ECG may be better approached from the point of view of novelty detection rather than classification.

5.3.2 PCA revisited

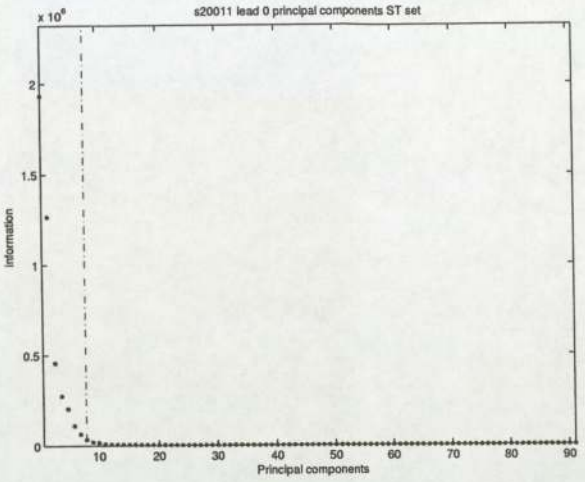
In Chapter 2 it can be seen that many researchers chose 4 or 5 principal components because over 95% of information is contained within these components. A close inspection of eigenvalues of covariance matrices of records in LTSTDB reveals that this is not always the case.

It was found that there are two types of records in LTSTDB, a "flat" one where, as in Chapter 2, most of the information is contained in the first 4 or 5 components, and a "upright" one where up to 20 components contain significant information.

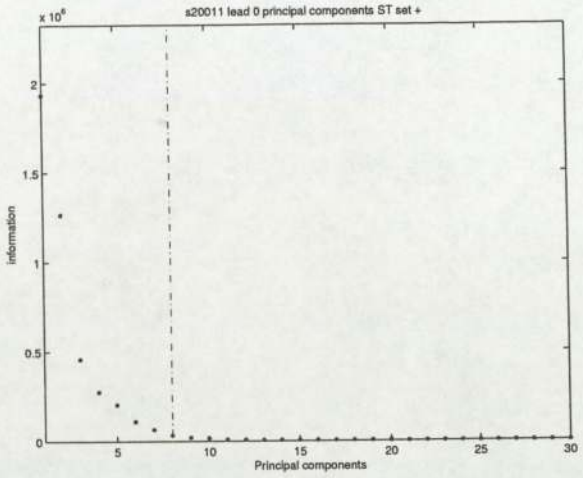
In Figure 5.1 the top plots for s20011 is a typical "flat" behaviour, and the bottom plots for s20141 is a typical "upright" behaviour.

The ST eigenvalues and normal eigenvalues are plotted separately in Figure 5.2. There is some similarity between the two types of records. The ST set for s20011 and s20141 have a similar shape, except that s20141 lifts off further from the x-axis than s20011. Similarly for the normal set, s20011 and s20141 both hug the y-axis. The effects of these behaviours are not seen in the classification results.

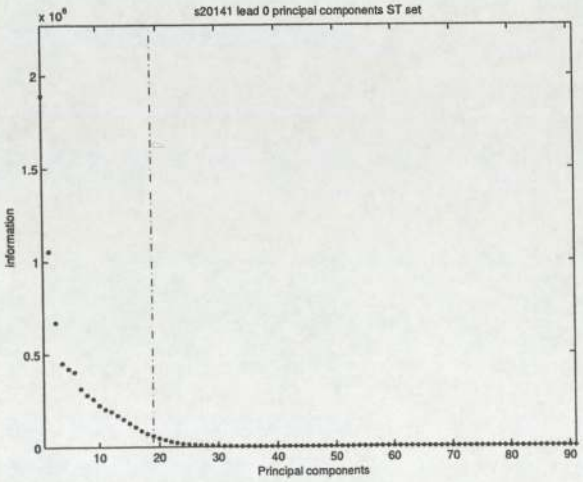
Future research may benefit from the knowledge that there are two different types of ST segment data.



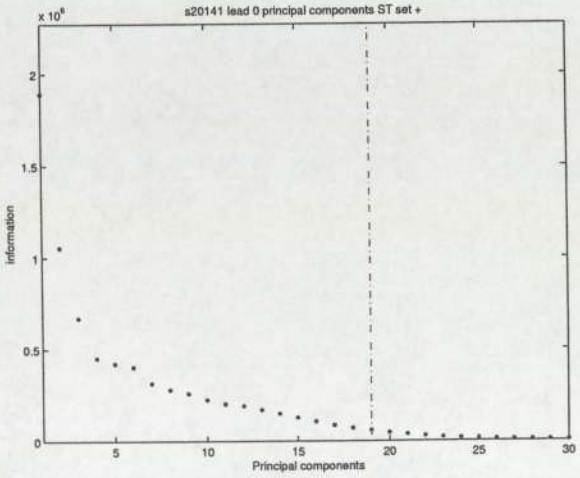
(a) s20011 PCs



(b) s20011 close up

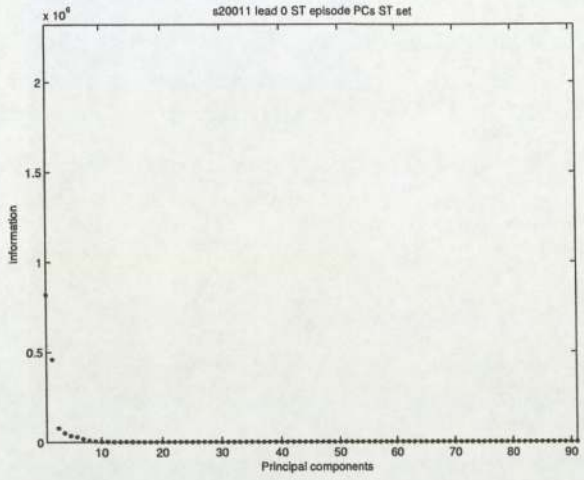


(c) s20141 PCs

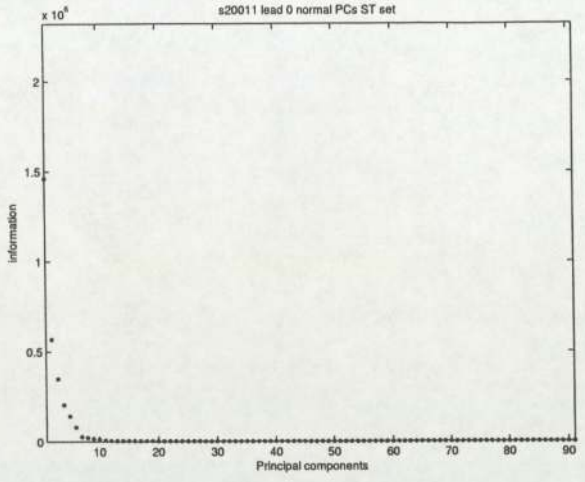


(d) s20141 close up

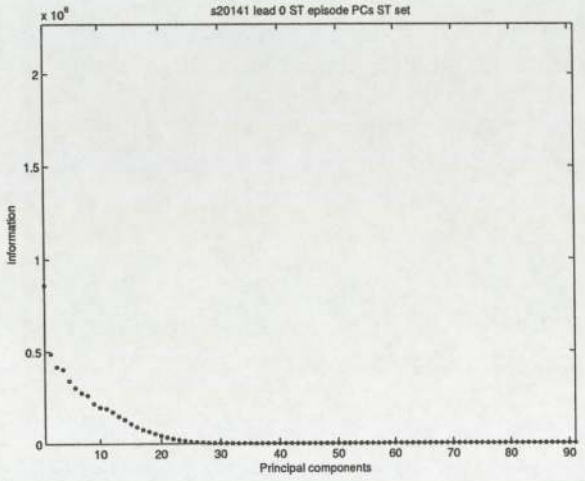
Figure 5.1: Different types of behavior in eigenvalues.



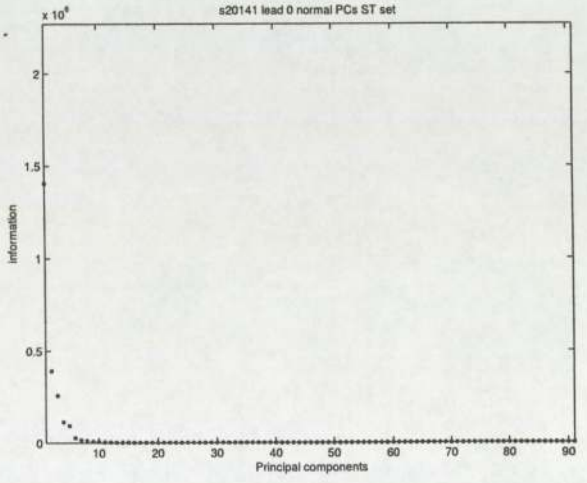
(a) s20011 STs



(b) s20011 normals



(c) s20141 STs



(d) s20141 normals

Figure 5.2: Different types of behavior in eigenvalues for ST and normal sets separately.

Bibliography

[1] C. M. Bishop. *Neural Networks for Pattern Recognition.* Oxford University Press, 1995.

[2] J. Boutkan. *ABC of the ECG -a Guide to Electrocardiography.* The Macmillan Press ltd, 1968.

[3] D. Evans. Personal communications. Conversation and e-mails.

[4] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000 (June 13). Circulation Electronic Pages: http://circ.ahajournals.org/cgi/content/full/101/23/e215.

[5] F. Jager, R. G. Mark, G. B. Moody, and S. Divijak. Analysis of transient ST segment changes during ambulatory monitoring using the Karhunen-Loève Transform. 1992.

[6] F. Jager, A. Taddei G. B. Moody, M. Emdin, G. Antolic, R. Dorn, A. Smrdel, C. Marchesi, and R. G. Mark. Long term ST database: a reference for the development and evaluation of automated ischemia detectors and for the study of the dynamics of myocardial ischaemia. *Computers in Cardiology*, 2003.

[7] P. Laguna, G. B. Moody, J. Garcia, A. L. Goldberger, and R. G. Mark. Analysis of the ST-T complex of the electrocardiogram using the Karhunen–Loeve transform: adaptive monitoring and alternans detection. *Medical And Biological Engineering And Computing*, 37:175–189, Mar 1999.

[8] G. B. Moody and R. G. Mark. Development and evaluation of a 2-lead ECG analysis program. *Computers in Cardiology*, 1982.

[9] R. F. Mould. *Introductory Medical Statistics.* IOP Publisher Ltd., 3rd edition, 1998.

[10] I. T. Nabney. NETLAB*: Algorithms for Pattern Recognition.* Springer, 2002.

[11] I. T. Nabney and D. J. Evans. Cardionetics institue of bioinformatics (CIB). ECG research group.

[12] C. Papaloukas, D. I. Fotiadis, A. Likas, and L. K. Michalis. An ischemia detection method based on artificial neural networks. *Artificial Intelligence in Medicine*, 2002.

[13] Contributors to the Long-Term ST Database. www.physionet.org/physiobank/database/ltstdb/thanks.sht:

[14] L. Schamroth. *Introduction to Electrocariography.* Blackwell Scientific Publications, fifth edition, unknown.

[15] T. Stamkopoulos, K. Diamantaras, N. Maglaveras, and M. Strintzis. ECG analysis using nonlinear PCA neural networks for ischemia detection. *IEEE Transaction on Signal Processing*, 46(11), November 1998.

[16] L. Tarassenko. *A Guide to Neural Computing Applications.* Arnold, 1998.

[17] L. Tarassenko, A. Nairac, N. Townsend, and P. Cowley. Novelty detection for the identification of abnormalities. 2000.

[18] D. K. Tasoulis, L. Vladutu, V. P. Plagianakos, A. Bezerianos, and M. N. Vrahatis. On-line neural network training for automatic ischemia episode detection. 2004.

[19] D. Woodcock. Personal communication. Conversations and e-mails.

[20] M. W. Zimmerman, R. J. Povinelli, M. T. Johnson, and K. M. Ropella. A reconstructed phase space approach for distinguishing ischemic from non-ischemic ST changes using holter ECG data. *Computers in Cardiology*, 2003.

Appendix A

Data Description

A.1 LTSTDB structure

The signals in the Long Term ST Database (LTSTDB) are given names according to a consistent scheme throughout the databases in PhysioBank [4]. These names are referred to as *records*. The records have the format

`sXYYYZ`

where `X` is the number of leads, `YYY` is the patient number and `Z` is the record number of the patient.

The records for which data is available on Physiobank is shown in Table A.1. There are 86 records in LTSTDB out of which 43 in Table A.1 are used for training. PhysioNet holds the remaining 43 records for testing purpose.

s20011	s20281	s20581
s20021	s20301	s20601
s20031	s20321	s20621
s20041	s20341	s20641
s20061	s20361	s30661
s20081	s20381	s30681
s20101	s20401	s30701
s20121	s20421	s30721
s20141	s20441	s30741
s20161	s20461	s30742
s20181	s20481	s30761
s20201	s20501	s30781
s20221	s20521	s30801
s20241	s20541	
s20261	s20561	

Table A.1: Records used for the project.

The signals and annotations stored in the Physiobank are separated into several files with various extensions indicating the contents as shown in Table A.2. Annotation files are `.16a`, `.ari`, `.atr`,

extension	format	description
.dat	binary	all the signals
.16a	binary	ST segments based on 16-second moving average
.ari	binary	automatically generated beat annotation
.atr	binary	manually corrected beat annotation
.cnt	text	number of ST episodes in a record
.hea	text	header of the .dat
.klt.zip	zipped text	KLT[a] basis of ST and QRS signals
.sta	binary	ST episode $V_{min} = 75\mu V$, $T_{min} = 30s$
.stb	binary	ST episode $V_{min} = 100\mu V$, $T_{min} = 30s$
.stc	binary	ST episode $V_{min} = 100\mu V$, $T_{min} = 60s$
.stf	text	ST level, approx. baseline, ST deviation

Table A.2: Physiobank data scheme.

[a]Karhunen-Loeve Transform

.sta, .stb and .stc. Data is stored according to the convention:

$$[hh : mm : ss.sss\ dd/mm/yyyy]\quad time\quad anntype\quad subtype\quad chan\quad num\quad aux$$

where *time* is given in sample numbers, *anntype* is the annotation, *chan* is the lead and *aux* is the information characteristic of that type of annotation.

Signal files are .dat, .klt and .stf and in which the data is stored according to the convention:

$$time\quad data1\quad data2\quad data3\cdots$$

and again, *time* is given in sample numbers.

A.2 File contents

This section explains the contents of the files and how they were preprocessed as documented by the creators [6].

A.2.1 .dat

The analogue ECG recordings were digitised at 250 Hz and amplitude scaling of 200 ADC units per milli-volt. Baseline wander has been corrected for using cubic spline approximation and subtraction technique. The signal was low-pass-filtered through a 6th order Butterworth filter with cutoff frequency at 55 Hz.

The beats were located automatically by ARISTOTLES ECG processing software [8].

Abnormal beats and noisy beats and their neighbours were rejected. A noisy beat is defined as a beat whose KLT coefficients varied from that of preceding 15 beats by more than 1 standard deviation, or if the normalised residual error of the reconstructed beat exceeded 25%. The resulting time series were smoothed, re-sampled and smoothed again. The ST segment is later replaced by an ST level function in .stf

code	description
[]	Positions of the isoelectric level and J-point (linear interpolation)
DMY	Positions of the isoelectric level and J-point (manual)
n	Lead number
aa	ST amplitude at point J+80(60) msec in μV (ST level)
bb	ST amplitude at point J+0 msec in μV
cc	ST amplitude at point J+20 msec in μV
dd	ST amplitude at point J+40 msec in μV
ee	ST amplitude at point J+60 msec in μV
ff	ST amplitude at point J+80 msec in μV
gg	ST amplitude at point J+100 msec in μV
hh	ST amplitude at point J+120 msec in μV
ISO	Position of the isoelectric level prior to ARISTOTLE's fiducial point in msec
JP	Position of the J-point after ARISTOTLE's fiducial point in msec
J80(60)	Position of the point of measurement after ARISTOTLE's fiducial point in msec J+80 msec, if HR < 100 bpm J+72 msec, if 100 bpm <= HR < 110 bpm J+64 msec, if 110 bpm <= HR < 120 bpm J+60 msec, if HR >= 120 bpm
NL	Number of beats left to current beat included into average beat
NR	Number of beats right to current beat included into average beat

Table A.3: Annotation for .16a file.

A.2.2 .16a

This file contains the ST measurements based on a 16-beat moving average. A sample is shown here.

$$[09:25:27.752\ 07/01/1993]\quad 21938\quad s\quad 0\quad 0\quad 0\ \cdots$$

$$\cdots\ ST0 + 98, +56, +51, +50, +70, +98, +132, +169, 96, 36, 116, 0, 10$$

The final section, the aux field is in the form

$$[\text{DMY}]\ \text{ST}\ n \pm aa\ \pm bb\ \pm cc\ \pm dd\ \pm ee\ \pm ff\ \pm gg\ \pm hh\ \pm ISO\ \pm JP\ \pm J80(60)\ \pm NL\ \pm NR.$$

The meaning of each of these values are explained in Table A.3[1]. The "J-point" is the starting point of the ST segment which, using ARISTOTLES, is simply 120 ms[2] after the fiducial point.

Entries with "DMY" preceding the aux field have had the J-point and or the isoelectric points manually corrected. Manual correction took place every 20 minutes or at the beginning, the extrema and the end of a significant ST episode. The corrected values were obtained using a 16-second moving average window.

The various ST amplitudes are measured from the isoelectric level.

A.2.3 .ari

This file contains the QRS complex annotation and the fiducial points as found by ARISTOTLES. The annotation scheme is shown in Table A.4.

[1] Problems with LaTeX for some of the symbols.
[2] This value depends on the heart rate.

Beat annotation

Code	Description
N	Normal beat (displayed as "?" by the Chart-O-Matic, pschart, and psfd)
L	Left bundle branch block beat
R	Right bundle branch block beat
B	Bundle branch block beat (unspecified)
A	Atrial premature beat
a	Aberrated atrial premature beat
J	Nodal (junctional) premature beat
S	Supraventricular premature or ectopic beat (atrial or nodal)
V	Premature ventricular contraction
r	R-on-T premature ventricular contraction
F	Fusion of ventricular and normal beat
e	Atrial escape beat
j	Nodal (junctional) escape beat
n	Supraventricular escape beat (atrial or nodal)
E	Ventricular escape beat
/	Paced beat
f	Fusion of paced and normal beat
Q	Unclassifiable beat
?	Beat not classified during learning

Non-beat annotation

[	Start of ventricular flutter/fibrillation
!	Ventricular flutter wave
]	End of ventricular flutter/fibrillation
x	Non-conducted P-wave (blocked APC)
(	Waveform onset
)	Waveform end
p	Peak of P-wave
t	Peak of T-wave
u	Peak of U-wave
‘	PQ junction
’	J-point
^	(Non-captured) pacemaker artifact
|	Isolated QRS-like artifact
~	Change in signal quality
+	Rhythm change
s	ST segment change
T	T-wave change
*	Systole
D	Diastole
=	Measurement annotation
“	Comment annotation
@	Link to external data

Table A.4: Annotation scheme.

A.2.4 .atr

This file contains the *true* QRS complex annotation as found by experts. The records were re-scanned by experts using two different systems (Marquette Holter and Zymed Holter scanners) and manually corrected for false classification of beats by the scanners. Resulting streams of annotations from both scanners were merged using the WFDB utility software (BXB) and any discrepancy between the annotations of the two experts were corrected for manually.

A.2.5 .cnt

This file contains the summary of number of events in each records. The number of occurance of each type of ST shift and ST episode, combined episodes and durations are given per standard per lead.

A.2.6 .hea

The header file contains information about how the signals were recorded and digitised along with a brief clinical description of the patient. It also includes the ADC gain and bias values.

A.2.7 KLT

There are 5 KLT coefficient time series and a Mahalanobis distance measure for ST segment and QRS complex, separately. KLT coefficient time series can be obtained by assigning the KLT coefficient of a beat to that beat [7]. Deviation of this time series from its baseline show ST changes.

The Mahalanobis distance, r, is defined as:

$$r^2 = (\mathbf{x} - \mathbf{m_x})' \mathrm{C_x^{-1}} (\mathbf{x} - \mathbf{m_x})$$

where $\mathbf{m_x}$ is the mean vector and $\mathrm{C_x}$ is the covariance matrix of vector $\mathbf{x}$.

A.2.8 .sta, .stb and .stc

These three files are essentially the same, but adhere to different annotation standards. The annotation codes are given in Table A.5. These files were produced by experts using SEMIA as an editing tool. The annotations were further reviewed and corrected independently by three groups of experts in Ljubljana, Pisa and Cambridge (US). The global reference is chosen near the beginning of the recording where the ST level is stable for at least 5 minutes. The local references are placed at certain intervals in the non-ischemic section of the data.

`.stc` is a subset of `.stb`, and `.stb` is a subset of `.sta`.

A.2.9 .stf

This file contains the ST level function, linear approximation of ST reference function and ST deviation.

code	description
GRSTn	Global reference
LRST$n\pm ll$	Local reference
s[cc]st$n$	Significant ST shift
([rt]st$n\pm dd$	Beginning of significant ST episode
a[rt]st$n\pm dd$	Extrema of significant ST episode
[rt]st$n$)$\pm dd$	End of significant ST episode
noi$n\pm dd$	Noise
(urd$n$	Beginning of unreadable interval
urd$n$)	End of unreadable interval

[cc] Type of ST shift (none: axis shift, cc: conduction change)
[rt] Type of ST episode (none: ischemic, rt: heart-rate related)
n Lead number
ll ST level in μV
dd ST deviation in μV

Table A.5: Annotation for .sta, .stb and .stc files.

The ST level function is the `.16a aux` field re-sampled at 0.5 Hz and smoothed using 7-point moving average. The reference function is the linear interpolation between local reference points annotated in `.sta`. The ST deviation is simply the difference between the ST level function and the reference function.

Appendix B

Record Description

Table B.1 summarises the number of ST episode and normal beats there are in each record used for analysis in this project. ST refers to ST episodes, and NM refers to normal beats. The RATIO is $\frac{ST}{(ST+NM)}$.

Table B.1: Number of beats in each record.

record (lead)	Number of beats		Ratio
	ST	NM	
s20011 (ML2)	11737	88316	0.117
s20011 (MV2)	3789	96263	0.038
s20021 (MLIII)	1676	87284	0.019
s20021 (v4)	3149	85811	0.268
s20061 (ML2)	27848	93078	0.211
s20101 (ML2)	3117	74964	0.040
s20101 (MV2)	2854	75227	0.037
s20121 (ML2)	9356	76170	0.109
s20141 (ML2)	25134	91540	0.215
s20141 (MV2)	17190	99484	0.147
s20161 (MLIII)	24640	59092	0.294
s20161 (v4)	2102	81630	0.025
s20201 (ML2)	3698	87774	0.040
s20221 (ML2)	1629	117552	0.014

Table B.1: Number of beats in each record.

record (lead)	Number of beats		Ratio
	ST	NM	
s20241 (ML2)	18999	73440	0.206
s20241 (MV2)	8800	83639	0.095
s20301 (v4)	16012	90751	0.150
s20321 (v4)	247	91663	0.003
s20321 (MLIII)	481	91429	0.005
s20481 (v4)	7672	84105	0.084
s20521 (v4)	2108	73833	0.028
s20521 (MLIII)	130	75811	0.002
s20541 (v4)	10895	104331	0.095
s20541 (MLIII)	6562	108664	0.057

The number of episodes and shifts in each record are summarised in Table B.2. The number of ischemic episodes are given for each annotation standard as described in Table A.5.

Table B.2: ST shifts and episodes for all records.

		ST episodes						ST shifts					
		Ischemic			HR related			Axis			Conduction		
Record	Lead	a	b	c	a	b	c	a	b	c	a	b	c
	0	0	0	0	14	3	3	0	0	0	0	0	0
s20011	1	0	0	0	6	1	1	7	7	7	0	0	0
	tot	0	0	0	20	4	4	7	7	7	0	0	0
	0	20	11	5	0	0	0	2	2	2	0	0	0
s20021	1	26	21	7	0	0	0	37	37	37	0	0	0
	tot	46	32	12	0	0	0	39	39	39	0	0	0
	0	5	5	5	6	3	2	0	0	0	0	0	0
s20031	1	5	5	5	7	3	3	0	0	0	0	0	0
	tot	10	10	10	13	6	5	0	0	0	0	0	0
	0	26	17	13	0	0	0	5	5	5	0	0	0
s20041	1	30	21	16	0	0	0	6	6	6	0	0	0
	tot	56	38	29	0	0	0	11	11	11	0	0	0

Table B.2: ST shifts and episodes for all records.

		ST episodes						ST shifts					
		Ischemic			HR related			Axis			Conduction		
Record	Lead	a	b	c	a	b	c	a	b	c	a	b	c
	0	0	0	0	26	9	4	0	0	0	0	0	0
s20061	1	0	0	0	0	0	0	0	0	0	0	0	0
	tot	0	0	0	26	9	4	0	0	0	0	0	0
	0	33	10	2	0	0	0	7	7	7	0	0	0
s20081	1	0	0	0	24	2	0	9	9	9	0	0	0
	tot	33	10	2	24	2	0	16	16	16	0	0	0
	0	1	1	1	0	0	0	2	2	2	0	0	0
s20101	1	2	1	1	0	0	0	0	0	0	0	0	0
	tot	3	2	2	0	0	0	2	2	2	0	0	0
	0	9	2	1	0	0	0	15	15	15	0	0	0
s20121	1	0	0	0	0	0	0	8	8	8	0	0	0
	tot	9	2	1	0	0	0	23	23	23	0	0	0
	0	0	0	0	44	24	5	0	0	0	0	0	0
s20141	1	0	0	0	48	27	7	10	10	10	0	0	0
	tot	0	0	0	92	51	12	10	10	10	0	0	0
	0	59	41	32	0	0	0	16	16	16	0	0	0
s20161	1	10	3	3	0	0	0	2	2	2	0	0	0
	tot	69	44	35	0	0	0	18	18	18	0	0	0
	0	0	0	0	0	0	0	74	74	74	0	0	0
s20181	1	36	12	7	0	0	0	20	20	20	0	0	0
	tot	36	12	7	0	0	0	94	94	94	0	0	0
	0	0	0	0	7	1	1	20	20	20	0	0	0
s20201	1	0	0	0	0	0	0	2	2	2	0	0	0
	tot	0	0	0	7	1	1	22	22	22	0	0	0
	0	0	0	0	4	3	2	70	70	70	0	0	0
s20221	1	0	0	0	0	0	0	95	95	95	0	0	0
	tot	0	0	0	4	3	2	165	165	165	0	0	0
	0	0	0	0	10	7	5	0	0	0	0	0	0
s20241	1	0	0	0	15	6	2	0	0	0	0	0	0
	tot	0	0	0	25	13	7	0	0	0	0	0	0
	0	17	13	11	0	0	0	18	18	18	0	0	0
s20261	1	13	8	5	0	0	0	13	13	13	0	0	0

Table B.2: ST shifts and episodes for all records.

		ST episodes						ST shifts					
		Ischemic			HR related			Axis			Conduction		
Record	Lead	a	b	c	a	b	c	a	b	c	a	b	c
	tot	30	21	16	0	0	0	31	31	31	0	0	0
	0	14	5	3	0	0	0	0	0	0	0	0	0
s20281	1	2	0	0	0	0	0	0	0	0	0	0	0
	tot	16	5	3	0	0	0	0	0	0	0	0	0
	0	41	27	21	0	0	0	44	44	44	0	0	0
s20301	1	0	0	0	0	0	0	8	8	8	0	0	0
	tot	41	27	21	0	0	0	52	52	52	0	0	0
	0	2	0	0	0	0	0	0	0	0	0	0	0
s20321	1	3	3	3	0	0	0	2	2	2	0	0	0
	tot	5	3	3	0	0	0	2	2	2	0	0	0
	0	8	4	3	0	0	0	0	0	0	0	0	0
s20341	1	0	0	0	0	0	0	2	2	2	0	0	0
	tot	8	4	3	0	0	0	2	2	2	0	0	0
	0	2	2	2	0	0	0	0	0	0	0	0	0
s20361	1	2	2	2	0	0	0	0	0	0	0	0	0
	tot	4	4	4	0	0	0	0	0	0	0	0	0
	0	1	1	1	3	1	0	0	0	0	0	0	0
s20381	1	1	1	1	0	0	0	0	0	0	0	0	0
	tot	2	2	2	3	1	0	0	0	0	0	0	0
	0	9	8	4	0	0	0	1	1	1	0	0	0
s20401	1	7	7	3	0	0	0	1	1	1	0	0	0
	tot	16	15	7	0	0	0	2	2	2	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0
s20421	1	20	17	11	0	0	0	6	6	6	0	0	0
	tot	20	17	11	0	0	0	6	6	6	0	0	0
	0	8	5	1	0	0	0	0	0	0	0	0	0
s20441	1	14	8	7	0	0	0	4	4	4	0	0	0
	tot	22	13	8	0	0	0	4	4	4	0	0	0
	0	4	4	3	0	0	0	0	0	0	0	0	0
s20461	1	4	2	0	0	0	0	0	0	0	0	0	0
	tot	8	6	3	0	0	0	0	0	0	0	0	0
	0	7	1	1	0	0	0	1	1	1	0	0	0
s20481	1	0	0	0	0	0	0	1	1	1	0	0	0

Table B.2: ST shifts and episodes for all records.

		ST episodes						ST shifts					
		Ischemic			HR related			Axis			Conduction		
Record	Lead	a	b	c	a	b	c	a	b	c	a	b	c
	tot	7	1	1	0	0	0	2	2	2	0	0	0
	0	0	0	0	0	0	0	112	112	112	0	0	0
s20501	1	0	0	0	0	0	0	120	120	120	0	0	0
	tot	0	0	0	0	0	0	232	232	232	0	0	0
	0	0	0	0	8	4	2	12	12	12	0	0	0
s20521	1	0	0	0	1	0	0	2	2	2	0	0	0
	tot	0	0	0	9	4	2	14	14	14	0	0	0
	0	0	0	0	11	9	6	0	0	0	281	281	281
s20541	1	0	0	0	5	4	3	0	0	0	148	148	148
	tot	0	0	0	16	13	9	0	0	0	429	429	429
	0	8	5	5	0	0	0	0	0	0	0	0	0
s20561	1	0	0	0	0	0	0	0	0	0	0	0	0
	tot	8	5	5	0	0	0	0	0	0	0	0	0
	0	2	2	1	0	0	0	4	4	4	0	0	0
s20581	1	1	1	1	0	0	0	1	1	1	0	0	0
	tot	3	3	2	0	0	0	5	5	5	0	0	0
	0	1	0	0	0	0	0	0	0	0	0	0	0
s20601	1	3	1	1	0	0	0	1	1	1	0	0	0
	tot	4	1	1	0	0	0	1	1	1	0	0	0
	0	0	0	0	19	13	10	5	5	5	0	0	0
s20621	0	0	0	0	17	12	6	1	1	1	0	0	0
	tot	0	0	0	36	25	16	6	6	6	0	0	0
	0	0	0	0	5	3	2	0	0	0	0	0	0
s20641	1	0	0	0	5	4	3	0	0	0	0	0	0
	tot	0	0	0	10	7	5	0	0	0	0	0	0
	0	21	13	11	0	0	0	3	3	3	0	0	0
	1	19	12	9	0	0	0	4	4	4	0	0	0
s30661	2	8	6	6	0	0	0	1	1	1	0	0	0
	tot	48	31	26	0	0	0	8	8	8	0	0	0
	0	26	11	8	0	0	0	0	0	0	0	0	0
	1	11	7	6	0	0	0	0	0	0	0	0	0
s30681	2	30	17	11	0	0	0	0	0	0	0	0	0

Table B.2: ST shifts and episodes for all records.

		ST episodes						ST shifts					
		Ischemic			HR related			Axis			Conduction		
Record	Lead	a	b	c	a	b	c	a	b	c	a	b	c
	tot	67	35	25	0	0	0	0	0	0	0	0	0
s30701	0	3	0	0	0	0	0	1	1	1	0	0	0
	1	2	2	2	0	0	0	2	2	2	0	0	0
	2	4	4	2	0	0	0	1	1	1	0	0	0
	tot	9	6	4	0	0	0	4	4	4	0	0	0
s30721	0	0	0	0	2	1	1	0	0	0	0	0	0
	1	1	1	1	0	0	0	0	0	0	7	7	7
	2	1	0	0	0	0	0	0	0	0	5	5	5
	tot	2	1	1	2	1	1	0	0	0	12	12	12
s30741	0	17	5	2	0	0	0	0	0	0	0	0	0
	1	11	3	1	0	0	0	0	0	0	0	0	0
	2	18	9	6	0	0	0	0	0	0	0	0	0
	tot	46	17	9	0	0	0	0	0	0	0	0	0
s30742	0	18	12	8	0	0	0	0	0	0	0	0	0
	1	18	3	2	0	0	0	0	0	0	0	0	0
	2	19	11	7	0	0	0	0	0	0	0	0	0
	tot	55	26	17	0	0	0	0	0	0	0	0	0
s30761	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	25	8	6	0	0	0	0	0	0	0	0	0
	2	7	1	0	0	0	0	0	0	0	0	0	0
	tot	32	9	6	0	0	0	0	0	0	0	0	0
s30781	0	0	0	0	10	2	1	0	0	0	0	0	0
	1	1	1	1	1	0	0	4	4	4	0	0	0
	2	1	1	1	0	0	0	3	3	3	0	0	0
	tot	2	2	2	11	2	1	7	7	7	0	0	0
s30801	0	0	0	0	0	0	0	1	1	1	0	0	0
	1	4	0	0	0	0	0	0	0	0	0	0	0
	2	6	3	2	0	0	0	2	2	2	0	0	0
	tot	10	3	2	0	0	0	3	3	3	0	0	0

Appendix C

Results with ST set

Here, results for all records are displayed. 50:50 data set with compensation using Equation 4.1 is shown in Table C.1. Those records with only one set of results have episodes in only one of the leads.

Table C.2 shows the test set results.

Table C.1: 50:50 data set with compensation experiment for all records.

		Training				Validating (50:50)			
record	lead	Conf mat		*se*	*sp*	Conf mat		*se*	*sp*
s20011	0	5766	98	66.2%	98.3%	3826	90	60.2%	97.7%
		1983	3881			1323	2587		
	1	1895	7	68.6%	99.6%	1260	7	69.1%	99.4%
		598	1304			391	876		
s20021	0	832	1	77.7%	99.9%	555	3	77.1%	99.5%
		186	647			128	430		
	1	1568	2	48.8%	99.9%	1035	8	49.2%	99.2%
		804	766			530	513		
s20061	0	15351	214	82.4%	98.6%	10170	201	81.7%	98.1%
		2735	12830			1896	8475		
s20101	0	1555	3	85.4%	99.8%	1028	11	86.9%	98.9%
		228	1330			136	903		
	1	1422	4	92.9%	99.7%	945	6	93.9%	99.4%
		101	1325			58	893		
s20121	0	4630	48	83.8%	99.0%	3068	47	84.3%	98.5%
		760	3918			489	2626		
s20141	0	14244	320	80.2%	97.8%	9431	282	79.6%	97.1%
		2887	11677			1978	7735		

Table C.1: 50:50 data set with compensation experiment for all records.

		Training				Validating (50:50)			
record	lead	Conf mat		*se*	*sp*	Conf mat		*se*	*sp*
	1	8538	41	97.5%	99.5%	5693	28	97.0%	99.5%
		217	8362			171	5550		
s20161	0	13052	462	64.0%	96.6%	8599	371	64.0%	95.9%
		4866	8648			3232	5738		
	1	1047	2	82.5%	99.8%	694	5	83.5%	99.3%
		184	865			115	584		
s20201	0	2214	5	61.3%	99.8%	1464	4	64.6%	99.7%
		859	1360			520	948		
s20221	0	813	0	70.1%	100.0%	542	1	66.1%	99.8%
		243	570			184	359		
s20241	0	9404	159	84.4%	98.3%	6263	114	83.7%	98.2%
		1488	8075			1041	5336		
	1	4371	22	97.8%	99.5%	9214	17	97.7%	99.4%
		96	4297			68	2863		
s20301	0	8791	207	56.0%	97.7%	5785	210	54.7%	96.5%
		3961	5037			2717	3278		
s20321	0	122	0	100.0%	100.0%	82	0	100.0%	100.0%
		0	122			0	82		
	1	239	0	42.3%	100.0%	159	1	50.0%	99.4%
		138	101			80	80		
s20481	0	3970	34	57.3%	99.2%	2627	33	56.9%	98.8%
		1711	2293			1146	1514		
s20521	0	1049	3	18.3%	99.7%	697	3	15.7%	99.6%
		859	193			590	110		
	1	65	0	20.0&	100.0%	41	2	16.3%	95.3&
		52	13			36	7		
s20541	0	6833	98	70.6%	98.6%	4535	74	71.2%	98.4%
		2040	4891			1326	3283		
	1	3261	18	82.4%	99.5%	2161	24	82.5%	98.9%
		576	2703			382	1803		

Table C.2: Test results for 50:50 data set with compensation experiment for ST set for all records. WARNING: RESULTS HERE ARE NOT CORRECT

		Testing (real priors)				Testing (50:50)			
record	lead	Conf mat		*se*	*sp*	Conf mat		*se*	*sp*
s20011	0	20854	580	15.5%	97.3%	1793	164	65.4%	91.6%
		1654	303			677	1280		
	1	21222	2119	10.2%	90.9%	624	11	69.6%	98.3%
		570	65			193	442		
s20021	0	13	21131	100.0%	0.1%	274	5	69.9%	98.2%
		0	279			84	195		
	1	16708	4070	41.0%	80.4%	446	78	52.3%	85.1%
		309	215			250	274		
s20061	0	25618	960	36.1%	96.4%	5104	95	82.1%	98.2%
		3323	1876			932	4267		
s20101	0	11379	6668	86.3%	63.1%	516	4	85.4%	99.2%
		71	449			76	444		
	1	5996	12196	99.2%	33.0%	474	2	93.3%	99.6%
		4	472			32	444		
s20121	0	18452	7	4.6%	100.0%	1537	22	84.9%	98.6%
		1488	71			236	1323		
s20141	0	23576	1347	74.1%	94.6%	4702	146	80.6%	97.0%
		1257	3591			939	3909		
	1	23797	276	81.6%	98.9%	2846	14	97.3%	99.5%
		526	2334			78	2782		
s20161	0	15301	624	63.5%	96.1%	4314	180	63.5%	96.0%
		1642	2852			1642	2852		
	1	19586	24	0.3%	99.9%	347	3	82.6%	99.1%
		349	1			61	289		
s20201	0	21747	0	13.1%	100.0%	724	0	59.5%	100.0%
		629	95			293	431		
s20221	0	28728	1	8.1%	100.0%	272	0	68.0%	100.0%
		158	14			87	185		
s20241	0	13860	4091	50.1%	77.2%	3122	68	84.6%	97.9%
		1593	1597			490	2700		
	1	20220	0	0.0%	100.0%	1456	11	96.7%	99.3%

Table C.2: Test results for 50:50 data set with compensation experiment for ST set for all records. WARNING: RESULTS HERE ARE NOT CORRECT

		Testing (real priors)				Testing (50:50)			
record	lead	Conf mat		*se*	*sp*	Conf mat		*se*	*sp*
		1467	0			48	1419		
s20301	0	23475	11	4.5%	100.0%	2925	111	54.8%	96.3%
		2899	137			1373	1663		
s20321	0	19361	2853	100.0%	87.2%	39	2	95.1%	95.1%
		0	41			2	39		
	1	19353	2750	21.0%	87.6%	81	0	51.9%	100.0%
		64	17			39	42		
s20481	0	20530	72	21.4%	99.7%	1312	18	56.9%	98.6%
		1046	284			573	757		
s20521	0	17886	12	2.6%	99.9%	348	4	21.3%	98.9%
		343	9			277	75		
	1	18393	38	0.0%	99.8%	21	1	9.1%	95.5%
		22	0			20	2		
s20541	0	26646	144	27.6%	99.5%	2278	37	70.7%	98.4%
		1676	639			678	1637		
	1	18392	7879	85.5%	70.0%	1083	10	80.1%	99.1%
		158	935			217	876		

Appendix D

Results with QRS + ST set

For comparison, the classification experiment using 50:50 training set and real class prior validation set was conducted for QRS + ST set as well. In preliminary investigation in Section 4.4.1 it was suspected that this larger set performs better than a set with just the ST segment. Here, in Table D.1 it is shown that most records perform significantly better in this data set. However, the worst reduction in sensitivity is a staggering 76.1% in s20021 lead 0 while the maximum improvement is 65.4%. The reduction in sensitivity may be due to the fact that, for this record, the QRS complex does not contain any more information than the ST segment on the abnormality of the recovery of the heart muscles. Worse still, the addition of the QRS complex can be giving more information on the similarity of the normal and abnormal beats, resulting in worse classification.

Table D.1: QRS + ST 50:50 data set with compensation experiment for all records.

	Training				Validating			
record	Conf mat		*se*	*sp*	Conf mat		*se*	*sp*
s20011	5801	60	85.3%	99.0%	3893	16	99.0%	99.6%
	861	5000			38	3872		
	1880	20	79.5%	98.9%	1254	12	99.2%	99.1%
	390	1510			10	1256		
s20021	830	0	86.1%	100.0%	556	0	89.6%	100.0%
	115	715			58	498		
	1564	3	64.4%	99.8%	1015	24	66.7%	97.7%
	558	1009			346	693		
s20061	15369	184	89.7%	98.8%	10175	192	88.5%	98.1%
	1604	13949			1196	9171		
s20101	1558	0	98.3%	100.0%	1033	6	97.6%	99.4%
	27	1531			25	1014		

Table D.1: QRS + ST 50:50 data set with compensation experiment for all records.

record	Training				Validating			
	Conf mat		*se*	*sp*	Conf mat		*se*	*sp*
	1424	1	94.6%	99.9%	947	4	95.4%	99.6%
	77	1348			44	907		
s20121	4662	14	93.1%	99.7%	3096	18	92.2%	99.4%
	321	4355			243	2871		
s20141	14217	342	87.5%	97.7%	9357	349	86.8%	96.4%
	1820	12739			1280	8426		
	8533	34	98.0%	99.6%	5682	30	96.9%	99.5%
	175	8392			179	5533		
s20161	13147	358	76.7%	97.3%	8673	292	77.4%	96.7%
	3151	10354			2025	6940		
	1047	0	86.4%	100.0%	698	1	97.3%	99.9%
	142	905			80	2875		
s20201	2210	8	73.9%	99.6%	1458	10	73.1%	99.3%
	580	1638			395	1073		
s20221	811	0	78.2%	100.0%	542	1	71.6%	99.8%
	177	634			154	389		
s20241	9452	106	91.6%	98.9%	6277	100	90.5%	98.4%
	807	8751			607	5770		
	4385	4	98.8%	99.9%	2910	17	98.2%	99.4%
	54	4335			52	2875		
s20301	8875	114	83.6%	98.7%	5866	124	83.0%	97.9%
	1472	7517			1021	4969		
s20321	121	0	100.0%	100.0%	81	0	100.0%	100.0%
	0	121			0	81		
	238	0	42.3%	100.0%	160	0	85.6%	100.0%
	22	216			23	137		
s20481	3977	26	67.8%	99.4%	2631	29	62.7%	98.9%
	1290	2713			991	1669		
s20521	1043	8	34.3%	99.2%	695	4	31.3%	99.4%
	690	361			480	219		
	65	0	60.0&	100.0%	41	2	39.5%	95.3%
	26	39			26	17		
s20541	6873	51	81.6%	99.3%	4552	56	79.3%	98.8%

Table D.1: QRS + ST 50:50 data set with compensation experiment for all records.

record	Training				Validating			
	Conf mat		*se*	*sp*	Conf mat		*se*	*sp*
	1276	5648			954	3654		
	3270	6	84.3%	99.8%	2178	7	84.3%	99.7%
	515	2761			242	1842		

Table D.2: Test results for 50:50 data set with compensation experiment for QRS + ST set for all records. WARNING: RESULTS HERE ARE NOT CORRECT

record	lead	Testing (real priors)				Testing (50:50)			
		Conf mat		*se*	*sp*	Conf mat		*se*	*sp*
s20011	0	608	20815	100.0%	2.8%	1947	9	99.0%	99.5%
		0	1956			19	1937		
	1	23326	4	0.2%	100.0%	630	4	99.5%	99.4%
		633	1			3	631		
s20021	0	21129	9	22.4%	100.0%	277	0	92.4%	100.0%
		215	62			21	256		
	1	7480	13294	90.3%	36.0%	394	130	67.7%	75.2%
		51	473			169	355		
s20061	0	1924	24643	90.6%	7.2%	5096	98	89.1%	98.1%
		489	4705			568	4626		
s20101	0	17990	51	46.2%	99.7%	515	5	97.7%	99.0%
		280	240			12	508		
	1	14727	3461	73.9%	81.0%	475	1	92.6%	99.8%
		124	352			35	441		
s20121	0	18042	411	50.2%	97.8%	1542	17	92.2%	98.9%
		763	769			122	1437		
s20141	0	24527	385	63.9%	98.5%	4670	171	87.7%	96.5%
		1750	3091			595	4246		
	1	21287	2781	99.8%	88.4%	2849	9	97.6%	99.7%
		5	2853			69	2789		
s20161	0	15444	478	76.9%	97.0%	4350	141	76.9%	96.9%
		1039	3452			1039	3452		

Table D.2: Test results for 50:50 data set with compensation experiment for QRS + ST set for all records. WARNING: RESULTS HERE ARE NOT CORRECT

		Testing (real priors)				Testing (50:50)			
record	lead	Conf mat		*se*	*sp*	Conf mat		*se*	*sp*
	1	19546	58	65.0%	99.7%	350	0	89.1%	100.0%
		513	953			38	312		
s20201	0	21739	1	40.2%	100.0%	719	5	71.8%	99.3%
		433	291			204	520		
s20221	0	28718	0	38.6%	100.0%	272	0	74.3%	100.0%
		167	105			70	202		
s20241	0	15301	2646	55.1%	85.3%	3130	60	91.3%	98.1%
		1361	1672			279	2911		
	1	8424	13671	100.0%	38.1%	1454	12	97.5%	99.2%
		0	79			37	1429		
s20301	0	20011	3469	55.1%	85.2%	2980	53	84.0%	98.3%
		1361	1672			485	2538		
s20321	0	22045	165	97.6%	99.3%	41	0	97.6%	100.0%
		1	40			1	40		
	1	18270	152	4.5%	99.2%	79	0	88.6%	100.0%
		21	1			9	70		
s20481	0	20525	65	30.8%	99.7%	1315	15	64.3%	98.9%
		921	409			475	855		
s20521	0	10925	6968	49.1%	61.1%	348	4	36.9%	98.9%
		179	173			222	130		
	1	18393	38	0.0%	99.8%	22	0	40.9%	100.0%
		22	0			13	9		
s20541	0	25669	1115	64.1%	95.8%	2290	22	80.4%	99.0%
		830	1482			454	1858		
	1	24529	1732	90.1%	93.4%	1089	4	83.7%	99.6%
		108	985			178	915		

Appendix E

Data Fusion Results

Table E.1: Data fusion results with 50:50 data sets.

	Training			Validating (50:50)			
record	Conf mat		*se*	*sp*	Conf mat		*se*	*sp*
s20011	1609	6	88.2%	99.6%	1072	7	89.1%	99.4%
	191	1426			228	961		
s20021	815	1	90.8%	99.9%	543	1	89.3%	99.8%
	75	741			58	486		
s20101	1266	2	100.0%	99.8%	844	1	100.0%	99.9%
	0	1268			0	845		
s20141	7158	1	100.0%	100.0%	4772	3	100.0%	99.9%
	3	7162			1	4776		
s20161	1400	1	99.7%	99.9%	699	1	100.0%	99.9%
	3	1048			0	700		
s20241	4217	1	100.0%	100.0%	2811	1	100.0%	100.0%
	2	4220			1	2813		
s20321	122	0	100.0%	100.0%	81	0	100.0%	100.0%
	0	122			0	81		
s20541	1906	0	99.9%	100.0%	1273	0	100.0%	100.0%
	2	1908			0	1273		

Table E.2: Data fusion test results with 50:50 data sets. WARNING: RESULTS HERE ARE NOT CORRECT

	Testing (real priors)				Testing (50:50)			
record	Conf mat		*se*	*sp*	Conf mat		*se*	*sp*
s20011	15607	0	3.7%	100.0%	540	0	75.4%	100.0%
	520	20			133	407		
s20021	0	14130	1.1%	0.0%	0	273	0.7%	0.0%
	270	3			271	2		
s20101	7711	131	100.0%	98.3%	417	7	100.0%	98.3%
	0	424			0	424		
s20141	16463	205	100.0%	98.8%	2361	25	100.0%	99.0%
	0	2386			0	1386		
s20161	8957	47	100.0%	99.5%	347	4	100.0%	98.9%
	0	351			0	351		
s20241	13667	175	100.0%	98.7%	1387	19	100.0%	98.6%
	0	1406			0	1406		
s20321	10084	59	100.0%	99.4%	42	0	100.0%	100.0%
	0	42			0	42		
s20541	18022	119	100.0%	99.3%	631	5	100.0%	99.2%
	0	636			0	636		