Application of Neural Network Classifiers to Electrocardiographic Body Surface Mapping

by

Tim Freeman, B.Sc.

Submitted in fulfilment of the requirements
for the degree of

Masters in Science

Department of Electrical Engineering and Computer Science
University of Tasmania

April 1998

This thesis contains no material which has been accepted for the award of any other higher degree or graduate diploma in any tertiary institution and to the best of the candidate's knowledge and belief, this thesis contains no material previously published or written by another person, except when due reference is made in the text of the thesis.

T. W. Freeman

Tim Freeman

Abstract

This thesis examines the capabilities of artificial neural networks for classifying electrocardiographic body surface mapping data. In particular it examines the diagnostic detection of myocardial infarctions and coronary artery disease. An overview of patten recognition, neural networks, electrocardiography, and electrocardiographic body surface mapping is presented followed by a detailed description of the experiments and analysis conducted.

The experimental analysis in this thesis is divided into three sections. Firstly, a range of feed-forward artificial neural network architectures and training techniques are used to classify the body surface mapping data with the aim of identifying patients with myocardial infarctions, coronary artery disease, and normal heart function. In this initial study a number of pre-processing techniques are also explored.

Secondly, a range of traditional classification techniques (linear regression, k-nearest-neighbour, and inductive learning) are applied to the same problems and compared with the neural network results. When classifying myocardial infarction it was found that artificial neural networks perform as well but no better than traditional classification techniques. This outcome provides some interesting insights into the nature of the classification problem and the information content of body surface maps. However, attempting to separate patients with coronary artery disease from patients with normal heart function neural networks were found to perform much better than traditional classification techniques.

The third experimental section examines the bayesian equivalence of neural network outputs and how these probabilistic properties may be used to deal with diagnostic uncertainty. Apart from examining the theoretical connection between network outputs and *a posteriori* probabilities, a number of experiments are conducted to show how this information can be used to provide the physician with some important information about the classification certainty.

Acknowledgments

I would like to extend my thanks to Dr Tony Adams for his help and support throughout this research, also to the members of the Neural Network Research Group for their insights suggestions and fascinating diversions.

I would also like to thank Dr Peter Vamplew for taking on a supervisory role during the final stages of the write-up. Thanks for your patience, suggestion and input.

To the members of the Information Systems Department and in particular Brian Marriott, thanks for giving me time to finish this journey.

To my parents for their support, encouragement, and love.

To my wife Sarah for continuing to believe in me, particularly in those times when I found it hard to believe in myself. I continue to own you so much.

To my son Liam for his contribution to this thesis (wefpm43po efw[p'k,,dewpoEWK <LjmePOIWEJMfcwho) and for reminding me of what life is about (http://tfreeman.infosys.utas.edu.au/liam/).

Contents

1. Introduction

This thesis examines the capabilities of artificial neural networks for classifying electrocardiographic body surface mapping data. In particular it examines the diagnostic detection of myocardial infarction and coronary artery disease. An overview of pattern recognition, neural networks, electrocardiography, and electrocardiographic body surface mapping is presented followed by a detailed description of the experiments and analysis conducted.

1.1 Background

Electrocardiography is an extremely useful diagnostic tool used by physicians for assessing heart disorders. As the human heart beats the heart muscle generates an electrical field that can be observed on the body surface. An electrocardiogram (ECG) is a recording of this electrical activity and can be used by a physician or cardiologist in the assessment of a patient's heart function.

The study of ECGs and their clinical significance began around the turn of the century when it was discovered that the electrical activity of the heart could be detected using a galvanometer connected to two electrodes placed on or in the body. It was soon discovered that the electrical field generated by the heart was complex and could not be observed in its entirety with one pair of electrodes. As a result much research was conducted in the 1920s and 1930s in an attempt to determine the most appropriate number and location of electrode pairs that provided a sufficient summary of the heart's activity. By the late 1930s numerous configurations had been put forward for recording ECGs. In 1938 in an attempt to rationalise the procedure of ECG recording a joint committee of the American Heart Association and the Cardiac Society of Great Britain and Ireland assessed the current approaches to ECG lead placement and recording and developed the first standard (American Heart Journal 1938). This underwent refinement and in 1954 the American Heart Association defined the 12-lead ECG standard now being used by cardiologists worldwide.

One of the major assumptions in cardiac research up until the mid 1950s was that the electrical field generated by the heart at any instance could be summarised as an electrical dipole located on the heart surface. As a consequence many researches considered that the 12-lead approach was inappropriate and contained redundant data and believed that a far more logical ECG need only record the X,Y and Z components of the heart's electrical dipole (Williams 1914, Mann 1938, Wilson et al. 1938). In 1956 the single dipole concept was challenged by Nelson (Nelson 1957) who showed (by the use of an electrocardiographic belt strapped around the human thorax) that the potential distribution observed at any instance on the chest surface was generated by more than one dipole. A single vector could not represent the electrical output of the heart and thus a more comprehensive representation was required.

A number of researches in the 1960s followed up Nelson's research and began to develop techniques for mapping the surface potentials of the entire thorax in the hope that this would provide a more complete representation of the heart's behaviour (Tacacardi 1957-1963, Amirov 1961, Horan et al. 1963). The recording of such body surface maps was a time consuming task as maps were constructed from a grid of between 50 and 600 ECGs. Complicating the matter further, researchers were limited by recording equipment having to record ECGs in batches (from anywhere between 2 to 20 ECGs at a time). Although this research was extremely fruitful, the use of such body surface mapping techniques for clinical diagnosis was technically impractical.

In the 1970s the problem of recording body surface maps (BSMs) was simplified with the use of computing technology, making it possible to record and store all electrode potentials simultaneously and quickly display the resulting surface potential maps (Taccardi et al. 1976, Kilpatrick et al. 1979, Spach et al. 1979, Heringa et al. 1981, and Yajima et al. 1983). As a consequence research into body surface mapping expanded and the 1980s saw a move towards the use of BSMs in clinical diagnosis.

1.2 Traditional Analysis of BSMs

A number of techniques are currently being used for analysing and classifying electrocardiographic body surface maps. Most of these to date have been focused on traditional statistical and pattern recognition techniques and no attempt has been made to apply neural networks to this problem domain.

1.3 Thesis Objectives

The principle aim of this thesis is two fold. Firstly, a range of neural network techniques will be applied to the problem of classifying BSM data. Secondly, a number of traditional classification techniques will be applied to the same classification problems to provide a comparative benchmark for the neural networks results.

1.4 Outline of Thesis

This thesis is divided into four broad sections: background, experimental design, results and final conclusions. Each of the chapters is summarised below:

1.4.1 Background Chapters

Chapter 2 Introduces the concept of *pattern recognition* and the various approaches that may be used to construct, train, and implement a pattern recognition system. A number of classification techniques which will be used in this study are described.

Chapter 3 Presents a number of *artificial neural network* architectures and training procedures. It serves to provide the reader with a background to the history of neural networks from its earliest beginnings to recent developments in the field.

Chapter 4 Introduction to *electrocardiography* and as background to the following chapter which discusses the specifics of electrocardiographic body surface mapping.

Chapter 5 Introduces *body surface mapping* and presents the range of techniques used for recording and analysing BSMs.

1.4.2 Experimental Design

Chapter 6 Describes the data acquisition system used and the data sets constructed followed by a detailed description of the classification problems and classification techniques applied.

1.4.3 Results and Discussion

Chapter 7 Initial Experiments - application of multilayer perceptrons (MLPs) to the classification problems. These initial experiments highlight the key challenges associated with classifying myocardial infarctions and coronary artery disease using electrocardiographic body surface mapping data.

Chapter 8 A number of alternative feature extraction and classification techniques are considered.

Chapter 9 Investigates a number of techniques for dealing with misclassification in neural networks and applies these techniques to the BSM data.

1.4.4 Conclusions

Chapter 10 Presents the final conclusions.

2. Pattern Recognition

This thesis is primarily concerned with assessing the classification ability of neural networks when applied to electrocardiographic body surface mapping data. As will be shown in this chapter, the field of classification is part of the broader field of pattern recognition. Therefore, it is important to understand the issues related to the problem of pattern recognition and how these impact upon the design and implementation of any classifier. This chapter will provide the reader with an understanding of what is involved in the process of pattern recognition and the various approaches that may be used to construct, train, and implement a pattern recognition system. A number of classification techniques which will be used in this study are described, although a discussion of neural networks has been left until Chapter 3.

2.1 Pattern Recognition Model

Pattern recognition is the task of categorising objects or events (Young and Calvert 1974) and the objective of a pattern recognition device is to observe an object or event and assign a category (Figure 2-1).

Figure 2-1: Process of pattern recognition

This process of pattern recognition can be broken down into three sub-processes (Duda and Heart 1973, Young and Calvert 1974); sensing, feature extraction, and classification (Figure 2-2). The sensing process is analogous to the human sensory system, and serves to extract observations from the environment associated with the object or event. The feature extraction process involves reducing this set of observations to a set of features that are relevant to the classification problem being considered. The final stage is the classification of the extracted features.

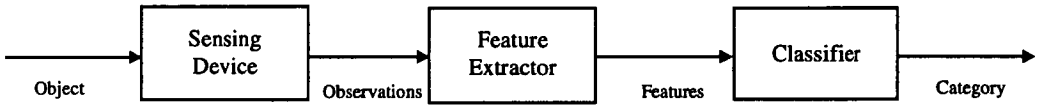


Figure 2-2: Sub-Processes of Pattern Recognition

It should be realised that the distinction between these three phases may not be as clearly defined. As we will see, the boundaries between these steps are often blurred, particularly in relation to feature extraction and classification, as many classification techniques exhibit some degree of feature extraction capability. The boundary between sensing and feature extraction on the other hand, is more clearly defined. Even so, part of the feature extraction process may identify observations that are either redundant or irrelevant, in which case the feature extraction process may result in modification of the sensing phase. The next three sections will discuss these three processes in more detail.

2.1.1 Sensing

The sensing phase of pattern recognition is principally a task of *measuring* a set of *attributes* associated with the object or event concerned and translating these into a representation that can be manipulated by the feature extractor and classifier.

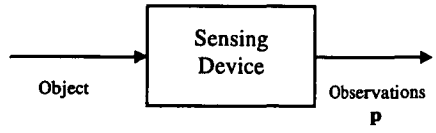


Figure 2-3: Sensing phase

For the most part, feature extraction and classification techniques tend to operate in a numeric domain, and as such the process of sensing involves constructing a numeric representation of the object or event being classified. The standard approach used (Tou 1973) is to construct a *pattern vector,* which encapsulates the measurements of the object. For example, if k measurements are associated with a particular object, then the object can be represented by the pattern vector **p:**

$$\mathbf{p}=\begin{pmatrix} p_1 \\ p_2 \\ \cdot \\ \cdot \\ \cdot \\ p_k \end{pmatrix} \qquad (2\text{-}1)$$

The approach used when translating attributes into a pattern vector will vary depending on the attribute type. Continuous attributes (eg. voltage measurements, distances and weights) can be represented by a single element p_i in the pattern vector. Discrete attributes (e.g. colour, sex and age) may require some simple coding. For example, if the attribute is a person's sex then this could be coded as $p_i = 0$ for male and $p_i = 1$ for female.

In the case of attributes with spatial or temporal relationships these may be represented as a set of measurements. For example, if the attribute being represented is some form of continuous signal f(t) (such as an acoustic signal) this signal could be sampled at discrete points; $t_1, t_2, \ldots, t_k$, and a pattern vector may be formed to represent this signal; $p_1=f(t_1), p_2=f(t_2), \ldots, p_k=f(p_k)$.

2.1.2 *Feature Extraction*

The primary aim of the feature extraction phase is to *reduce* the number of measurements associated with a pattern to a set of features that are relevant to the particular classification task being considered.

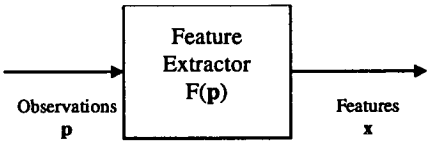


Figure 2-4: Feature extraction phase.

This process is often described as a functional transformation (Figure 2-4), where a *pattern vector* **p** is translated or transformed into the *feature vector* **x**:

$$\mathbf{x} = F(\mathbf{p}) \qquad (2\text{-}2)$$

where **p** is of dimension K and **x** is of dimension M and M<K. There are a range of techniques that can be used to construct the F(**p**) transformation. A number of these will be discussed below.

2.1.2.1 *Feature Selection*

One of the simplest techniques used for feature extraction is the *feature selection* technique. This technique involves selecting M attributes from the K attributes in the pattern vector **p** to form the feature vector **x**. This approach in effect discards (K-M) attributes from the pattern vector **p**. From a mathematical point of view the feature selection process can be described by:

$$\mathbf{x} = \mathbf{S}\mathbf{p} \qquad (2\text{-}3)$$

where **S** is a matrix of dimension MxK and each of the M rows of **S** consist of a one-element and (K-1) zero-elements. The aim of this approach is to select attributes from **p** that provide the best discrimination between classes, without losing any information which may be useful in the classification process.

There are a number of approaches that can be used to select attributes. The simplest of these is manual selection. Using this approach a set of known patterns and associated classifications are analysed to determine which attributes provide the greatest degree of discrimination between classes (eg difference of class means). For example, if we consider the simple case in Figure 2-5, where the pattern vector **p** consists of two attributes p_1 and p_2, it is clear that attribute p_1 provides better discrimination between classes A and B than does attribute p_2.

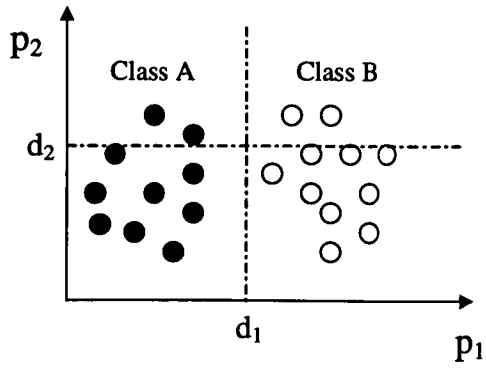


Figure 2-5: Illustration of attribute selection

An alternative to this approach is to consider every possible combination of M attributes from the pattern vector set, and select the set which provides the lowest level of total discrimination error. However this exhaustive approach tends to be somewhat computationally expensive (Young and Calvert 1973).

2.1.2.2 Linear Feature Extraction

Another feature extraction approach is the use of a linear transform:

$$\mathbf{x} = \mathbf{T}\mathbf{p} \tag{2-4}$$

where each feature x_j is a linear combination of the pattern attributes:

$$x_j = t_{j1}p_1 + t_{j2}p_2 + \ldots + t_{jk}p_k \tag{2-5}$$

The best known and most useful feature extraction scheme for constructing linear feature extractors is the Karhunen-Loeve transform. This technique is extremely useful for optimising the linear transform T.

The Karhunen-Loeve transform provides a technique for determining an optimised set of orthonormal basis vectors appropriate for the patterns being considered. This technique makes no assumptions about the underlying process associated with the patterns being presented, and is ideal for removing redundancy in patterns associated with apparently random processes.

To construct a set of orthogonal basis vectors a covariance matrix **R** of dimension KxK is calculated given a set of known patterns {**p**} of dimension K. As a result the covariance matrix **R** will consist of a set of ordered orthonormal basis vectors $[\beta_1, \beta_2, \beta_3, \ldots, \beta_k]$. Given these basis vectors, a pattern **p** may be represented as a linear sum:

$$\mathbf{p} = e_1\beta_1 + e_2\beta_2 + \ldots + e_k\beta_k \tag{2-6}$$

where {e} is a set of eigenvalues defined by:

$$e_i = \mathbf{p} \cdot \beta_i \tag{2-7}$$

One key feature of the expansion in (2-6) is that features are of decreasing significance and as such $e_1\beta_1$ is the most significant term and $e_k\beta_k$ the least significant term. Therefore an approximation of $\mathbf{p}_M$ may be constructed from M terms:

$$\mathbf{p}_M = e_1\beta_1 + e_2\beta_2 + \ldots + e_M\beta_M \quad (2\text{-}8)$$

and therefore a feature vector $\mathbf{x}$ can be constructed using the eigenvalues of this expansion:

$$\mathbf{x} = [e_1, e_2, \ldots, e_M] \quad (2\text{-}9)$$

Therefore returning to the linear transformation described in (2-4) a transformation matrix T_{KL} may be constructed from the M eigenvectors:

$$T_{KL} = [\beta_1, \beta_2, \ldots, \beta_M] \quad (2\text{-}10)$$

Which can therefore be used to translate a pattern vector into a feature vector:

$$\mathbf{x} = T_{KL}\,\mathbf{p} \quad (2\text{-}11)$$

2.1.3 Classification

The main aim of the classification phase is to translate a feature vector into a classification. As illustrated in figure 2-6 this requires the construction of a decision function D(**x**), which translates a feature vector **x** into a *classification vector* **y**. Where the dimension of **y** is N for an N-class classification problem.

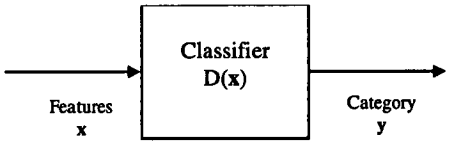


Figure 2-6: Classification Process

The approach used for deriving or constructing a decision function varies depending on the classification technique used, but in broad terms this function is derived based on prior knowledge of the relationship between features and their associated classifications. Thus the construction of a decision function will require some sample of the population $[\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3, \ldots, \mathbf{x}_s]$ and an associated set of known classifications for each of these cases $[\mathbf{y}_1,\mathbf{y}_2,\mathbf{y}_3, \ldots, \mathbf{y}_s]$. The following sections will describe a number of classification techniques that may be used to construct a decision function.

2.1.3.1 Template Matching

One of the simplest techniques that can be used for constructing a decision function is template matching. This technique assigns a class to a pattern **x** by searching the set of known patterns $[\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3, \ldots, \mathbf{x}_s]$ and finding a known pattern identical to **x**. If a match $\mathbf{x} = \mathbf{x}_i$ is found, then **x** is classified as belonging to the class designated by $\mathbf{y}_i$. Although this may be a useful technique, it does require a large sample population

and is unable to classify patterns that do not match a known pattern. A more generalised technique is the nearest neighbour classification technique.

2.1.3.2 Nearest Neighbour

As the name suggests the nearest neighbour technique assigns a classification to **x** based on the closest known pattern [$\mathbf{x}_1$,$\mathbf{x}_2$,$\mathbf{x}_3$, ... , $\mathbf{x}_s$]. The technique calculates the distance between **x** and all known patterns and selects the pattern closest to **x**:

This is illustrated by the example in Figure 2-7 where patterns are described by the two features (x_1,x_2) and the objective is to assign patterns to class A or B. In this particular case $\mathbf{x}_b$ is the closest known pattern to **x** and therefore **x** will be classified as belonging to class B.

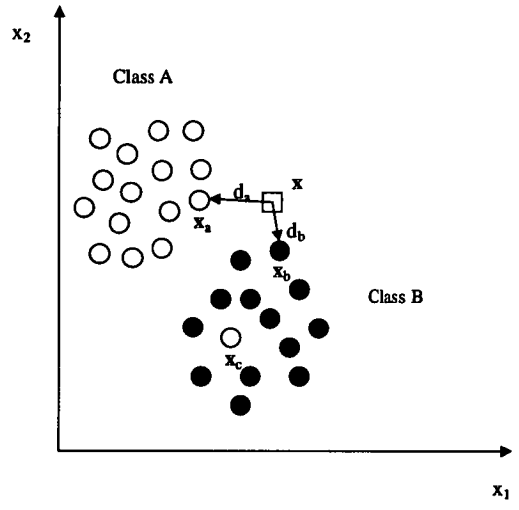


Figure 2-7: Nearest neighbour technique

Although this technique is a significant improvement over the template-matching algorithm it is prone to problems. In many cases there will be overlap in classes, which can present some classification problems for the nearest neighbour technique. If we consider Figure 2-7 again and in particular the rogue pattern x_c which is assigned to class A but in actual fact is closer to class B. Given the nearest neighbour algorithm any pattern **x** which is close to x_c will be incorrectly classified as belonging to class A despite the fact that this region is populated by a higher proportion of pattern from class B. An extension to the nearest neighbour algorithm, which overcomes this problem, is the k-nearest neighbour algorithm.

2.1.3.3 k-nearest neighbour

As the name suggests the k-nearest neighbour technique selects the *k* nearest patterns from the known pattern set [$\mathbf{x}_1$,$\mathbf{x}_2$,$\mathbf{x}_3$, ... , $\mathbf{x}_s$] and assigns the pattern **x** to the class most frequently represented amoung the *k* nearest samples, in effect taking a vote on the class to which **x** most likely belongs.

An example of this technique is given in Figure 2-8 in which the nearest five (k=5) examples are used to classify the pattern **x**. In this case, 3 out of the 5 samples

belong to class A and therefore x will be classified as belonging to class A. The advantage of this technique becomes apparent when we consider the rogue pattern x_c. Since all neighbouring samples belong to class B then any **x** close to x_c will still be classified as belonging to class B. Clearly the selection of a set of k samples will resolve localised aberrations and provide an average classification.

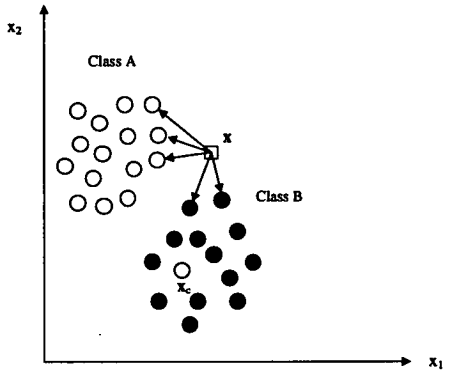


Figure 2-8: Illustration of k-nearest neighbour technique

Determining the most appropriate *k* for a particular problem is difficult and will depend on the particular classification problem being considered and the size of the sample population used (Duda and Hart 1973). In practice the value of *k* must be determined by experimental testing as the size and nature of the sample population will very much determine the most appropriate value.

An extension to the *k*-nearest neighbour technique used in this thesis involved incorporating the distance measure into the voting scheme (Tou and Gonzalez 1974). This approach was a slight variation on the above and avoided the possibility of ties when voting. Rather than selecting the *k*-nearest sample overall, the *k*-nearest samples from *each* class were selected (this is illustrated in Figure 2-9 where *k*=5 samples are selected from each class). The average distance from the feature vector **x** and the *k* sample cases in each class is calculated and the class closest on average is selected.

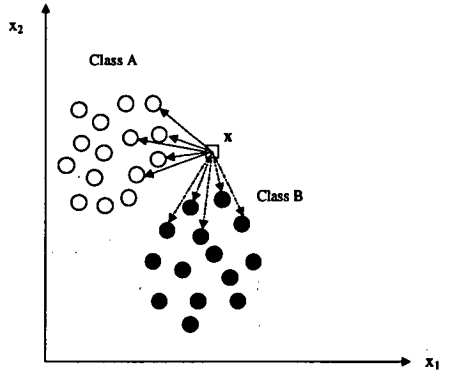


Figure 2-9: Alternative k-nearest neighbour technique

2.1.3.4 Linear Discrimination

Fisher (1936) originally proposed the concept of using linear discriminant functions for pattern classification. While this was a statistically oriented approach, it forms the basis of the early work in neural networks (McCulloch and Pitt 1943, Rosenblatt 1957, 1962).

Fisher proposed that given the feature vector input **x** of dimension M, a decision function d(**x**) could be constructed from a linear combination of the components of **x**:

$$d(\mathbf{x}) = w_1x_1 + x_2x_2 + \ldots + w_Mx_M + w_0 \qquad (2\text{-}12)$$

or:

$$d(\mathbf{x}) = \mathbf{w}^t\mathbf{x} + w_0 \qquad (2\text{-}13)$$

where **w** is called the *weight vector* and w_0 the *threshold weight*. This function in effect creates a decision boundary along the plane d(**x**)=0 in M-dimensional space, and divides the space into two regions. One region is define by those feature vectors where d(**x**)<0 and the other region is defined by those feature vectors x where d(**x**)>0. Thus, a single linear discriminator may be used to discriminate between two categories. This is best illustrated in Figure 2-10 (for a 2-dimensional feature space). In this case classes A and B can be separated by constructing a linear discriminant function which forms a decision boundary between the two classes. The slope of the decision boundary is determined by the weight vector **w** (the decision boundary will be perpendicular to the weight vector) and the offset of the decision boundary from the origin is determined by the $w_0/|w|$ (note that if **w** is normalised then the offset will simply be w_0).

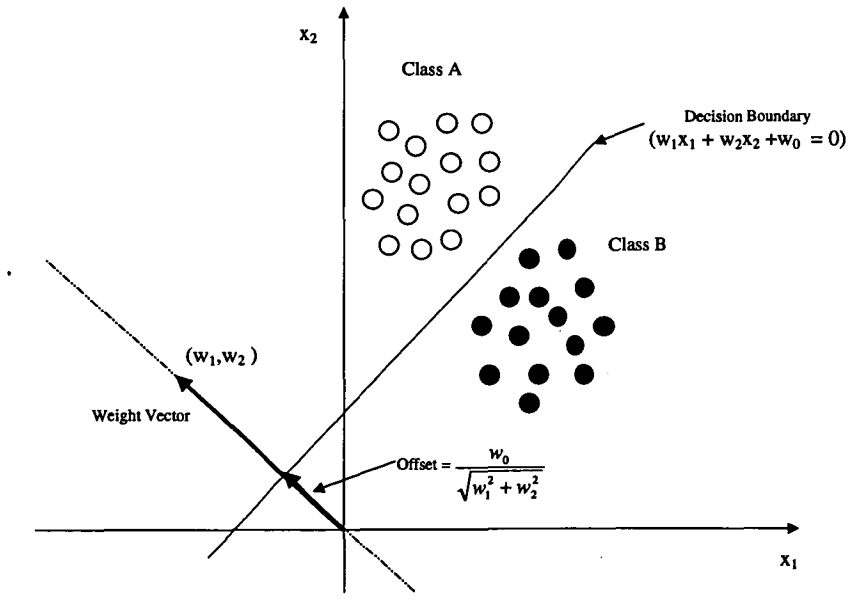


Figure 2-10: Example of linear discriminator

On analysis of the linear discriminant function we find that d(**x**) actually performs a mapping of feature vectors **x** onto the weight vector **w** in effect reducing the two-dimensional feature vectors to a one-dimensional measure. In effect d(**x**) is a distance measure which indicates the distance that **x** is from the decision boundary[1]. This is illustrated in Figure 2-11. Therefore, if d(**x**)>0 then **x** belongs to class A, if d(**x**)<0 then **x** belongs to class B.

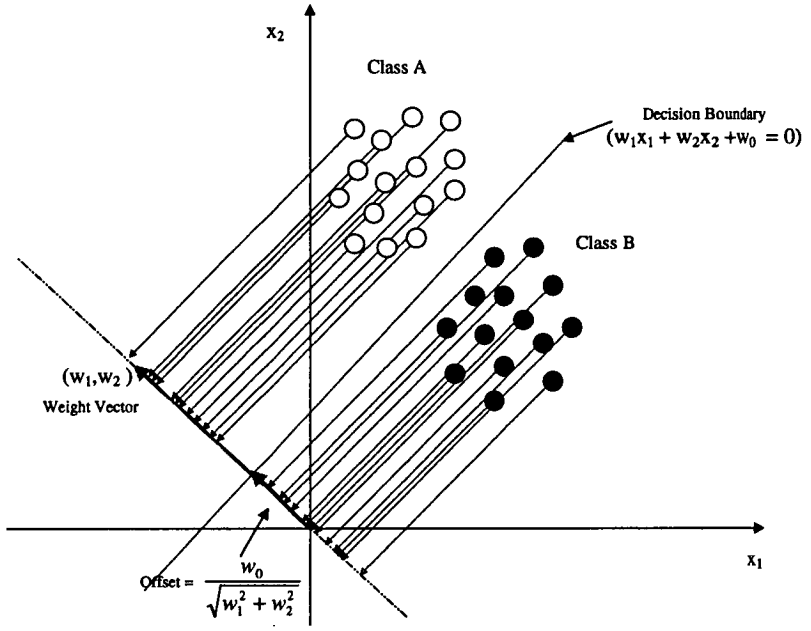


Figure 2-11: Illustration of linear discrimination mapping.

Although a single linear discriminant function is only appropriate for a two-class problem this classification technique can also be extended to multi-class problems by using a set of linear discriminant functions. For an N-class problem a set of N linear discriminant functions [$d_1(\mathbf{x})$, $d_2(\mathbf{x})$, ... ,$d_N(\mathbf{x})$] may be used to discriminate between classes [c_1, c_2, ... ,c_N], where;

$d_1(\mathbf{x})$ discriminates between c_1 and not-c_1
$d_2(\mathbf{x})$ discriminates between c_2 and not-c_2

.

$d_N(\mathbf{x})$ discriminates between c_N and not-c_N

2.1.3.5 Finding the optimum linear discriminator

The application of linear discriminant functions to pattern classification was well described by Highleyman (1962), who posed the problem of finding the optimal linear discriminant. Highleyman proposed that a plausible technique of determining the linear discriminant function was to apply some form of gradient descent

---

[1] Note that this distance measure will be scaled according to the magnitude of the weight vector – d(**x**) = distance(**x**) . |**w**|

procedure, using a know set of examples, to minimise classification error. Interestingly this technique was identical to that proposed by Rosenblatt (1958) in relation to perceptron learning. Considering the similarity of these techniques a discussion of the gradient descent algorithm will be presented in Chapter 3.

A number of other statistical approaches also exist for finding the optimum linear discriminator. A particular example is linear regression (R^2). However, it should be realised that these techniques are iterative error minimisation techniques similar in approach to the gradient descent algorithm (Duda and Hart 1973).

2.1.3.6 Decision Tree Inference

The decision tree classification scheme is very different to those classification techniques considered thus far, which have focused very much on the geometric proximity of classes in feature space and the use of either distance measures or linear functions.

A decision tree is probably easiest described by a simple example. For the most part decision trees construct decision boundaries that are based on individual features rather than a functional combination of features. This in effect means that all decision boundaries are orthogonal planes. Using a combination of these boundaries more complex decision regions can be constructed. A simple example of this is given in figure 2-12. This is the same classification problem used to illustrate linear discrimination in the previous section, but in this case decision boundaries are constructed using individual features, rather than functional combinations of features.

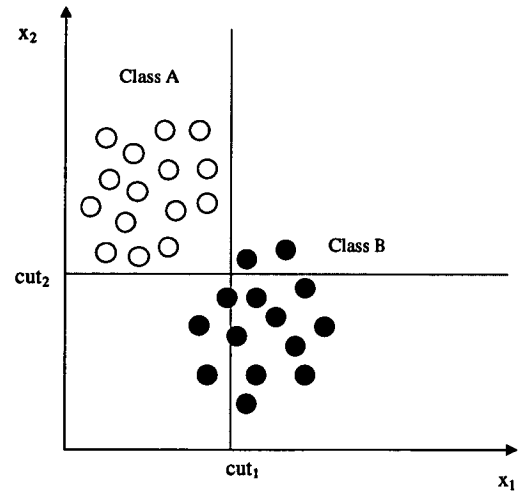


Figure 2-12: Example decision boundaries for an inductive decision tree

This approach may at first seem somewhat simplistic but is very powerful in practice as a decision tree can be constructed using a series of these decision boundaries. And in this particular case a decision tree could be constructed like that given in fig 2-13 that is capable of classifying all cases of classes A and B. It is noted that although these decision boundaries do not completely separate all cases, when used in combination a 100% classification performance can be achieved.

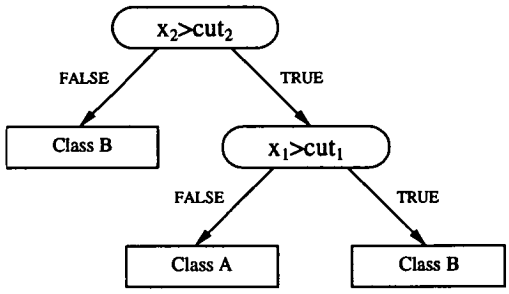


Figure 2-13: Decision tree associated with decision boundaries given in Figure 2-12.

A number of heuristic algorithms have been developed for determining optimum decision trees. All such heuristics require some form of training set consisting of a set of examples and associated classifications. One of the earliest systems developed for constructing optimum decision trees was the Concept Learning System framework (CLS – Hunt Marin and Stone 1966). CLS attempted to analyse all possible decision trees but restricted this search process by attempting to minimise a look-ahead cost function. The cost criteria used consisted of two components; the complexity of the decision tree and the misclassification cost associated with the decision tree.

Quinlan (1979, 1983) further refined CLS with the development of ID3. ID3 abandoned the cost driven look-ahead of CLS and considered a more information driven evaluation function where decision nodes were added. The basic structure of ID3 is iterative. A subset of the training set called the *window* is chosen at random and a decision tree formed from it, where the tree correctly classifies all objects in the window. All other objects in the training set are then classified in the tree. If the tree give the correct answer for all of these objects then it is correct for the entire training set and the process terminates. If not, a selection of incorrectly classified objects is added to the window and the process continues. This approach tends to produce a more generalised decision tree, rather than just focussing on classifying the particular training set at hand, and it is more likely to correctly classify new objects.

The formation of the decision tree by ID3 tends to be a somewhat more directed approach than that used by CLS. Given an arbitrary collection of training set objects ID3 examines each of the features in the feature vector and selects that feature which provides the best discrimination between the classes concerned. Having selected this feature a decision node is created using this feature and the collection of objects in the training set are partitioned using this decision criteria. Having partitioned the collection of example objects subsequent nodes are iteratively added to the decision tree using a similar feature selection criteria.

A lot of the early work with decision trees particularly in relation to CLS and ID3 was focussed on classification problems where the feature attributes were of a discrete nature. Quinlan extended ID3 to handle continuous features with the development of C4.5 (Quinlan 1993). Considering the nature of electrocardiographic body surface mapping data C4.5 was considered a more appropriate inductive decision tree algorithm for use in this thesis. Further to this the inductive decision tree algorithm MML (Wallace 1990, Wallace and Patrick 1993) will also be considered. Although MML is based on a similar approach to C4.5 there are a

number of improvements that may prove useful when considering the classification of the electrocardiographic data.

2.2 Concluding Comments

This chapter has presented a general background to the field of pattern recognition and has presented a number of classification techniques. The following chapter will focus on the specific treatment of neural networks and present a number of techniques which can be used for pattern classification.

3. Neural Networks

This chapter introduces a number of artificial neural network architectures and training procedures. It serves to provide the reader with a background to the history of neural networks from its earliest beginnings to recent developments in the field.

3.1 The Perceptron

The concept of artificial neurons was originally proposed by McCulloch and Pitt (1943) and is the basic building block of any artificial neural network. This basic artificial neuron model consists of four elements; a set of *synapes*, a *summing unit*, a *threshold* parameter, and an *activation function* (Figure 3-1).

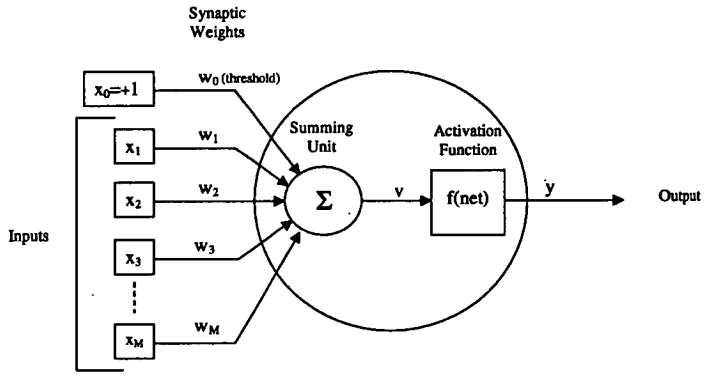


Figure 3-1: Model of a basic artificial neuron

In this model, the *inputs* (x_1,x_2, ... ,x_M) are connected to the summing unit by a set of synapses. Each synapse is characterised by a *weight*, which determines the degree of influence of the associated inputs. The summing unit sums the input signals, weighted by the respective synapses and produces a single result:

$$v=(w_1x_1+w_2x_2+ \ldots +w_Mx_M)+w_0 \qquad (3\text{-}1)$$

This is in effect a linear discriminant function (as described in chapter 2) which forms a decision plane separating two regions in the input space. One region is defined by those input patterns that produce an output of $v>0$, the other region is defined by those input patterns that produce an output of $v<0$. The orientation of the decision plane will be determined by the input weights (w_1,w_2,w_3, ... ,w_M) and will be offset by the threshold parameter w_0.

An artificial neuron model also incorporates a non-linear activation function f(v) which squashes or limits the permissible amplitude range of the output signal to some finite range. The simplest example of an activation function is the *step function* (Figure 3-2), often referred to as the *threshold function*.

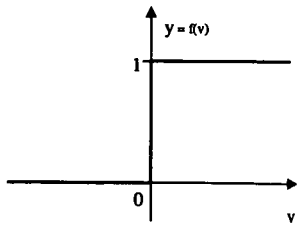


Figure 3-2: A simple activation function – the step function.

The step function produces a binary output of 0 or 1 depending on the value of v:

$$\text{if } v>0 \text{ then } y=1 \tag{3-3}$$

$$\text{if } v<=0 \text{ then } y=0 \tag{3-4}$$

A training algorithm for this neuron model was proposed by Rosenblatt in 1958. Rosenblatt proposed that a single McCulloch-Pitt neuron could be trained to classify between two classes A and B using an iterative training procedure know as the *perceptron learning algorithm.*

This algorithm is a supervised training procedure that uses a set of *training examples* $[\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_s]$ associated with known classifications or *desired outputs* $[d_1, d_2, \ldots, d_s]$ where s is the size of the example set and $d_j = 0$ when $\mathbf{p}_i$ belongs to class A and $d_j = 1$ when $\mathbf{p}_i$ belongs to class B. Given the neuron model presented in Figure 3-1, if we define $w_i(t)$, where $0 \leq i \leq M$ to be the weight associated with input i at time t, then the perceptron learning algorithm may be described as follows :

Perceptron Learning Algorthim

1. Initialise the weights and threshold – set $w_i(0)$ to a random value in the range [0,1].
2. Randomly select an example from the training set $[\mathbf{p}_j, d_j]$ and apply this to the input of the perceptron; $\mathbf{x}_i(t)=\mathbf{p}_j$.
3. Calculate the perceptron output; $y(t) = f[\Sigma w_i(t)x_i(t)]$
4. Adjust the weights according to the following procedure (for all i):

if	$y(t) = d_j$	then	$w_i(t+1) = w_i(t)$
if	$y(t) = 0$ and $dj = 1$	then	$w_i(t+1) = w_i(t) + x_i(t)$
if	$y(t) = 0$ and $dj = 1$	then	$w_i(t+1) = w_i(t) + x_i(t)$

5. Increment time t and repeat 2 to 5.

Widrow and Hoff (1960) introduced a more generalised form of weight update proposing a learning rule know as the Widrow-Hoff delta rule, which calculated the difference between the weighted sum and the required output.

$$\Delta(t) = d_j - v(t) \qquad (3\text{-}5)$$

This was described by Widrow and Hoff as the *output error* and is used to adjust the network weights in combination with a constant η which controls the learning rate (where $0 < \eta \leq 1$):

Widrow-Hoff Weight Adjustment Procedure

Weights adjusted according to the following procedure (for all i):

$$\Delta(t) = d_j - v(t)$$

$$wi(t+1) = wi(t) + \eta \,.\, \Delta(t) \,.\, xi(t)$$

The learning rate η was introduced to reduce the rate of individual changes imposed upon the weights and was found to provide a more stable learning process and weights were more likely to settle.

It is worth noting that the output error in the Widrow-Hoff delta rule is calculated with respect to the sum of the network inputs $v(t)$ rather than with respect to the perceptron output $y(t)$ and therefore the activation function is not used during the training process. The principle reason for this is that the activation function output $y(t)$ provides no indication of the magnitude of error associated with the output as the activation function removes this detail. Therefore the intermediate sum $v(t)$ must be used (Beale and Jackson 1990).

With the findings of Rosenblatt (1958) and the modification proposed by Widrow and Hoff (1960) the research into perceptrons and neural network concepts began to flourish. However, there was one major problem with the perceptron model, it was only capable of solving problems where the classes concerned were linearly separable. In their book *Perceptrons,* Minsky and Papert (1969) presented what is know as the XOR problem as an example of a non-linearly separable problem which although somewhat simple could not be solved using perceptrons.

3.2 Multilayer Perceptrons (MLPs)

The development of the *back-propagation* algorithm for training neural networks made up of layers of artificial neurons came to light in 1986. Rumelhart, Hinton and Williams (1986) published an algorithm that would train a multi-layered perceptron using a gradient decent technique to train the network weights and reduce the network error.

3.2.1 Architecture

Multilayer perceptrons are simply collections of individual neurons arranged in consecutive layers where the output of one layer becomes the input to the next layer. The simplest form of this architecture is a network consisting of a set of input nodes and two layers of neurons; one layer described as the hidden layer the other described as the output layer. Note that the input nodes x_0 and h_0 are set to a fixed value of +1 and provide a threshold input to the hidden and output layers respectively.

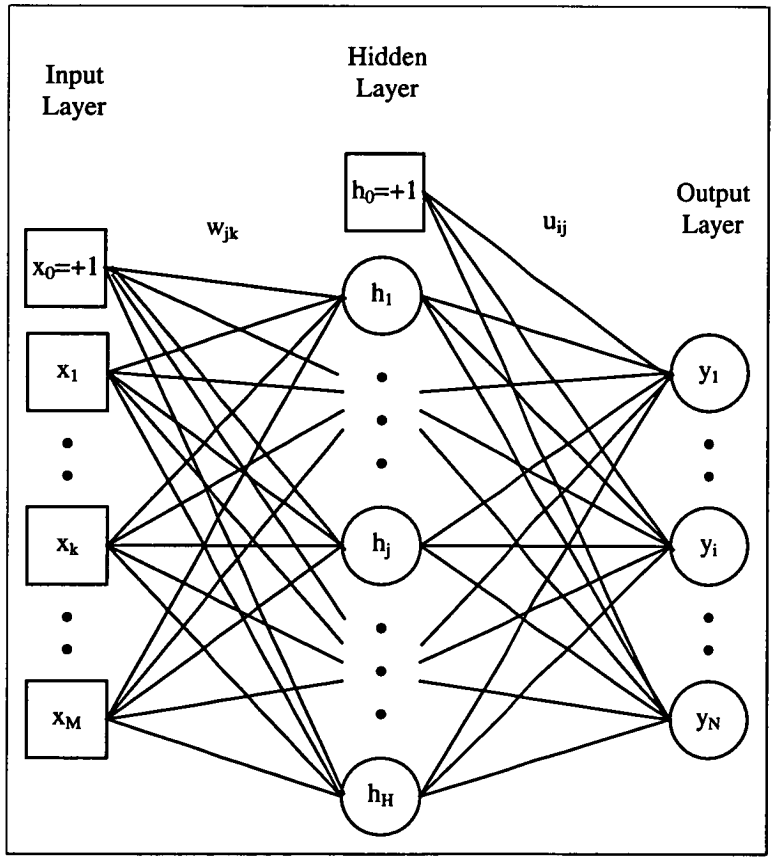


Figure 3-3: Multilayer perceptron with one hidden layer.

The structure depicted in Figure 3-3 is described as a *feed-forward* network. Input nodes feed forward into the hidden layer and the hidden layer neurons feed forward into the output layer. Also note that the standard multilayer perceptron model is a *fully-connected* architecture in that all inputs are connected to every hidden layer neuron and all hidden layer neurons are connected to every output layer neuron.

Another key aspect of this architecture is the neuron activation function. Unlike the step function originally used in the Rosenblatt perceptron model, the multilayer perceptron model utilised a new form of activation function described as the *sigmoid* function:

$$f(sum) = \frac{1}{1+e^{-sum}} \tag{3-6}$$

A plot of this function is provided in Figure 3-4:

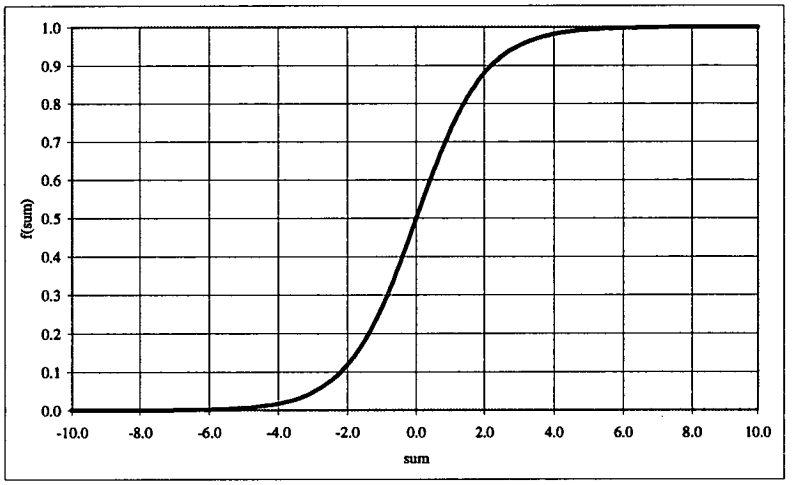


Figure 3-4: Sigmoid Activation Function

This function is central to the back-propagation algorithm and is key to simplifying the weight adjustment algorithm. As will be show in the next section, the back-propagation requires the calculation of the derivative of the activation function and in the case of the sigmoid function this derivative is easily determined:

$$f'(sum) = f(sum)[1 - f(sum)] \tag{3-7}$$

3.2.2 Back-propagationTraining

The back-propagation training algorithm involves a supervised training procedure, using a training data set to adjust the weights of a feed-forward neural network in order to minimise the output error of that network. To describe the specifics of this training procedure, consider the feed-forward network in Figure 3-4. This network has *M* input nodes, a single hidden layer containing *H* neurons, and an output layer containing *N* neurons. The input and hidden layers are connected by *M*x*H* weights, where w_{jk} represents a connection between the x_k input node and the h_j hidden neuron. The hidden and output layers are connected by *H*x*N* weights, where u_{ij} represents a connection between the h_j hidden neuron and the y_i output neuron. Consider also a training dataset containing *S* training examples:

$$[(\mathbf{p}_1,\mathbf{d}_1), \ldots (\mathbf{p}_e,\mathbf{d}_e) \ldots (\mathbf{p}_S,\mathbf{d}_S)]$$

Where $(\mathbf{p}_e,\mathbf{d}_e)$ represent a single training example; $\mathbf{p}_e = (p_{e1},p_{e2}, \ldots, p_{eM})$ represents an input pattern, and $\mathbf{d}_e = (d_{e1},d_{e2}, \ldots . d_{eN})$ represents the desired output for pattern $\mathbf{p}_e$.

The back-propagation training procedure is an iterative process of presenting the training patterns ($\mathbf{p}_e$) to the network one at a time and adjusting the network weights in relation to the error between the network output and the desired output ($\mathbf{d}_e$). As with perceptron training, the network weights are initialised to random values in the range [0,1] (or –1 to 1). If we consider the presentation of the training example ($\mathbf{p}_e$,$\mathbf{d}_e$), then the weight adjustment is done as follows:

Pattern Presentation and Weight Adjustment

1. The pattern $\mathbf{p}_e$ is applied to the network and the network output determined:

   The input pattern $\mathbf{p}_e$ is applied to the network:

$$x_k = p_{ek} \tag{3-8}$$

   The hidden layer is calculated:

$$net_j = \sum_{c=1}^{M} w_{jc} x_c \tag{3-9}$$

$$h_j = f(net_j) \tag{3-10}$$

   The output layer is calculated:

$$net_i = \sum_{b=1}^{H} u_{ib} h_b \tag{3-11}$$

$$y_i = g(net_i) \tag{3-12}$$

2. The output error is calculated:

$$E = \frac{1}{2} \sum_{a=1}^{N} (y_a - d_a)^2 \tag{3-13}$$

3. The network weights are adjusted using the *gradient decent rule*:

$$\Delta w = -\eta \frac{\partial E}{\partial w} \tag{3-14}$$

$$w = w + \Delta w \tag{3-15}$$

3.2.3 *Back-propagation Process*

The weight update procedure (3-14) is dependent on the differential $\delta E/\delta w$. The nature of the differential is dependent on the location of the weights in the MLP (the derivation of these differentials is provided in Appendix A).

The process of back-propagation involves propagation of the mean-squared error (MSE) back through the network. For those weights connected to the output layer (u_{ij}) the back-propagation process involves calculating the total output error (E), propagating this back through the y_i output neuron and calculating $\delta E/\delta u_{ij}$ which can then be used to update the weight. This is diagrammatically illustrated by Figure 3-5.

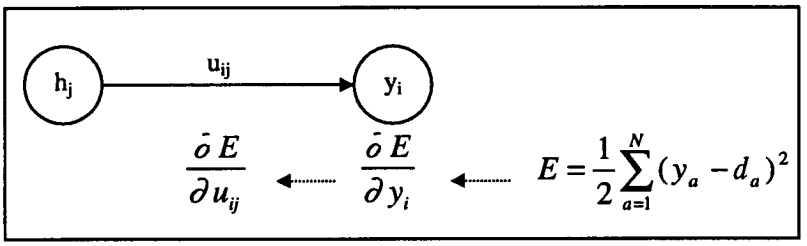


Figure 3-5: Adjusting the output layer weights.

This differential is calculated as follows:

$$\delta E/\delta u_{ij} = (y_i - d_i) \,.\, y_i \,.\, (1 - y_i) \,.\, h_j \qquad (3\text{-}16)$$

and therefore u_{ij} is updated using this differential:

$$u_{ij}(t+1) = u_{ij}(t) + \eta \,.\, \delta E/\delta u_{ij} \qquad (3\text{-}17)$$

The calculation of the error in the hidden layer weights (w_{jk}) is similar. However the error will be associated with all u_{ij} weights connected to the hidden neuron *j*. This back-propagation process is illustrated by Figure 3-6.

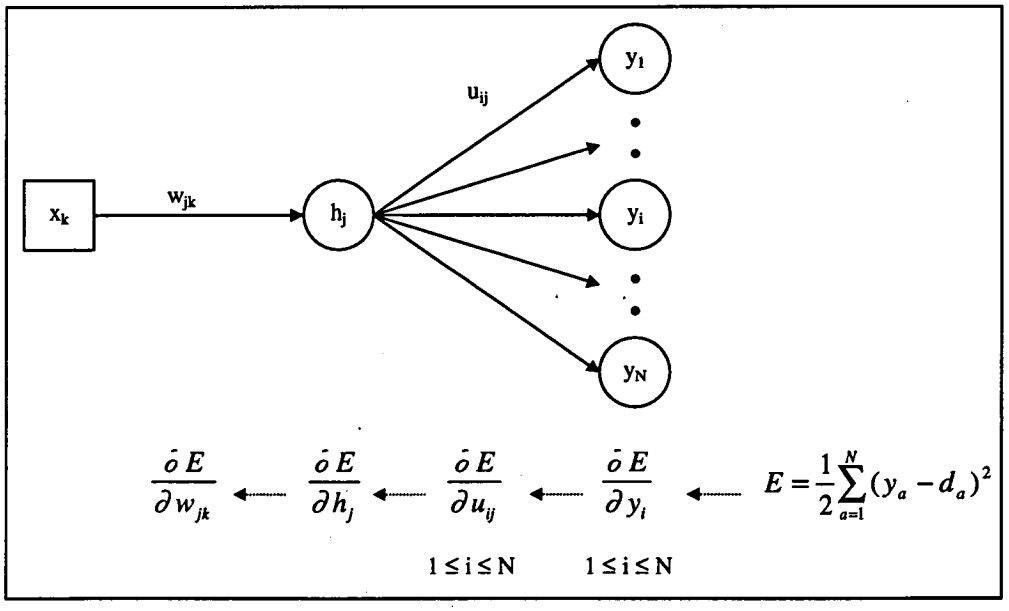


Figure 3-6: Adjusting the hidden layer weights.

The hidden weight differential is calculated as follows:

$$\delta E/\delta w_{jk} = \delta E/\delta h_j \,.\, h_j \,.\, (1 - h_j) \,.\, x_k \qquad (3\text{-}18)$$

where the cumulative error on hidden neuron *j* is:

$$\delta E/\delta h_j = \Sigma \,[\, (y_i - d_i) \,.\, y_i \,.\, (1 - y_i) \,.\, u_{ij} \,] \qquad (3\text{-}19)$$

Given the similarity of (3-16) and (3-18) it is clear that the back-propagation process can be used in an MLP with more than one hidden layer, and as such may be used with any number of hidden layers (see Appendix A for details).

3.2.4 Training Process

As already mentioned the back-propagation training process is an iterative process. Given a training set $[(\mathbf{p}_1,\mathbf{d}_1), \ldots (\mathbf{p}_e,\mathbf{d}_e) \ldots (\mathbf{p}_S,\mathbf{d}_S)]$ each pattern $\mathbf{p}_e$ in the training set is presented to the MLP and then the mean-squared error between the network output $\mathbf{y}(t)$ and the desired output $\mathbf{d}_e$ is propagated back through the network and each network weight is updated according to the weight update equations presented in section 3.2.3. This process is the repeated for different training set pairs until all training set pairs have been presented to the network. One complete presentation of the entire training set during the training process is called an *epoch*. In most cases the training process will involve repeated presentation of the training set before the training process is complete.

The measure generally used for assessing training progress is the overall training set error (3-20). There are other possible measures that can be used for measuring training progress (see section 3.2.5), but the overall training set error is the most common.

$$E_{overall} = \sum_{e=1}^{S}\left[\frac{1}{2}\sum_{i=1}^{N}(y_{ei} - d_{ei})^2\right] \qquad (3\text{-}20)$$

The ideal aim would be to achieve $E_{overall} = 0$, but in reality this is often not possible and the best approach is to continue presenting the training set until $E_{overall}$ settles to some minimum.

3.2.5 Stopping Criteria

A number of different techniques can be used to stop the training process. Three common measures are used:

Error – checking the overall training error ($E_{overall}$). As suggested in section 3.2.4 the training error can be monitored and training stopped when the error settles to some minimum. A common approach is to stop training when the *change* in $E_{overall}$ per epoch is less than some predefined value; $| E_{overall}(e) - E_{overall}(e\text{-}1) | / E_{overall}(e) < \tau$, where $E_{overall}(e)$ is the overall training error at the end of epoch e.

Correctness – checking the classification accuracy. This involves stopping the training process once a certain percentage of the training examples are classified correctly. Similar to the error measure; $| Correct(e) - Correct(e-1) | / Correct(e) < \tau$, where *Correct(e)* is the percentage of training examplesclassified correctly at the end of epoch *e*.

Epoch Limit – stopping after a set number of epochs.

Patience – this approach is used in conjunction with the Error or Correctness measures but allows the network to continue training without improvement for a set number of epochs.

3.2.6 Batch and Pattern Presentation

The weight update procedure presented in section 3.2.2 and 3.2.3 involves updating the weights after the presentation of *each* $[(\mathbf{p}_1,\mathbf{d}_1), \ldots (\mathbf{p}_e,\mathbf{d}_e) \ldots (\mathbf{p}_S,\mathbf{d}_S)]$. As the technique suggests this approach is described as *pattern presentation*:

$$w(t+1) = w(t) + \Delta w(t) \tag{3-21}$$

An alternative to this approach is *batch presentation*. This approach updates the weights in the network after *all* patterns in the training set $[(\mathbf{p}_1,\mathbf{d}_1), \ldots (\mathbf{p}_e,\mathbf{d}_e) \ldots (\mathbf{p}_S,\mathbf{d}_S)]$ have been presented to the network.. This technique sums the weight adjustments:

$$\Delta W = \Delta W + \Delta w(t) \tag{3-22}$$

and adjusts the network weights at the end of each epoch:

$$w(e+1) = w(e) + \Delta W \tag{3-23}$$

There has been much debate in the literature concerning which weight update approach is the best (Hildebrandt 1992). Using the standard gradient descent update procedure it has been show that pattern presentation (Hildebrandt 1992) will converge quicker than batch presentation. However, if speed of convergence is not an issue, batch presentation does tend to provide a more stable convergence.

3.2.7 Quickprop Algorithm

The Quickprop algorithm (Fahlman 1988) is an alternative weight update procedure which adds an extra term to the gradient descent algorithm. The standard gradient descent algorithm:

$$\Delta w(t) = - \eta \, . \, \delta E/\delta w(t) \tag{3-24}$$

make the assumption that the relationship between network error *E* and the weight *w* is a linear relationship. However, Fahlman (1988) suggests that the relationship between $\delta E/\delta w(t)$ tends to be of a quadratic or higher order relationship. Given these

observations Falhman proposed an alternative weight update procedure used in combination with a batch presentation approach:

$$\Delta w(e) = - s(e) \,.\, \Delta w(e-1) - \eta \,.\, \delta E/\delta w(e) \qquad (3\text{-}25)$$

where:

$$s(e) = \delta E/\delta w(e) \,/\, [\delta E/\delta w(e-1) - \delta E/\delta w(e)] \qquad (3\text{-}26)$$

The additional term $s(e) \,.\, \Delta w(e-1)$ is loosely based on Newton's method and aims to locate the minimum of the postulated quadratic relationship between E and w. To provide a degree of stability a maximum growth rate term (μ) is defined to restrict the maximum value of the quadratic term. If $|s(e) \,.\, \Delta w(e-1)| > \mu$ then the term is limited to $-\mu$ or $+\mu$ depending on the magnitude.

The advantages of the quickprop algorithm primarily relate in increasing the speed of training and has been reported as providing speed improvements (in terms of epochs) of anywhere between 2 and 10 fold (Shiffmann et. al., 1992, Fahlmann 1988).

3.3 Committee Networks

Committee networks are constructed by combining a group of neural networks together to form a single classifier. The input vector is connected to all network inputs and the outputs of each network are combined to form a single output vector (Figure 3-7). It has been found (Vamplew and Adams 1993; Baxt 1993) that if each individual network is trained using different starting weights, then the resulting committee network will perform better (with respect to the percentage of correct classifications) than any one individual network.

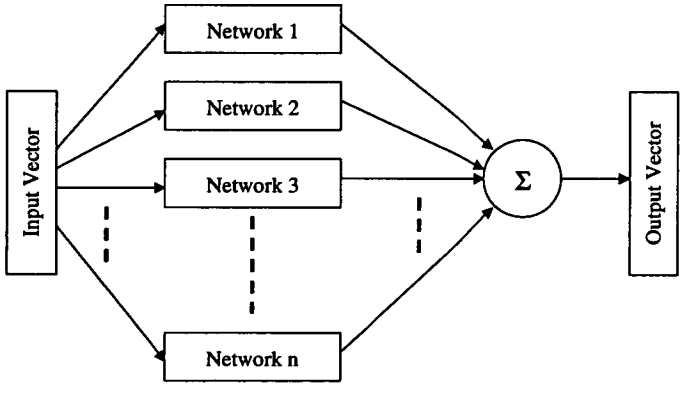


Figure 3-7: An example of a committee based network.

3.4 Cascade Correlation

The cascade correlation algorithm (Fahlman and Lebiere 1991) differs in many ways from the neural network techniques presented thus far. The algorithm begins with a single layer network and automatically trains and adds new hidden units (neurons) one by one, creating a multilayer structure.

The algorithm starts by creating a single layer of output neurons, one for each class output, and connecting these to the inputs and the threshold input (+1). An example of this structure is presented in Figure 3-8. Note that the nomenclature and symbols is slightly different to that used for perceptrons and MLPs, the principle reason for this is to simplify the representation. This architecture is trained using a training set to minimise the total output error using a maximum epoch limit and patience stopping criteria. If at the end of this phase all training examples are correctly classified then then the training process is stopped. Otherwise the training algorithm moves into a second phase.

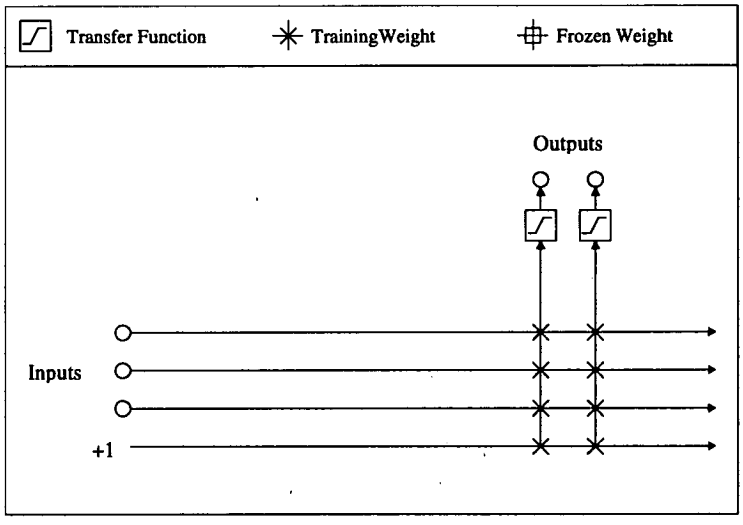


Figure 3-8: Starting architecture – a single layer network.

In the second phase a set of candidate units are inserted into the network and connected to the input nodes (Figure 3-9), but for the moment are not connected to the output units. During this training phase the output weights are frozen and the candidate unit input weights are trained. These units are trained to correlate with the output error (the reason for this approach being that once the candidate is connected then the overall network error will be reduced by the candidate output). As with the output training phase, candidates are training using an epoch training limit and a patience stopping criteria.

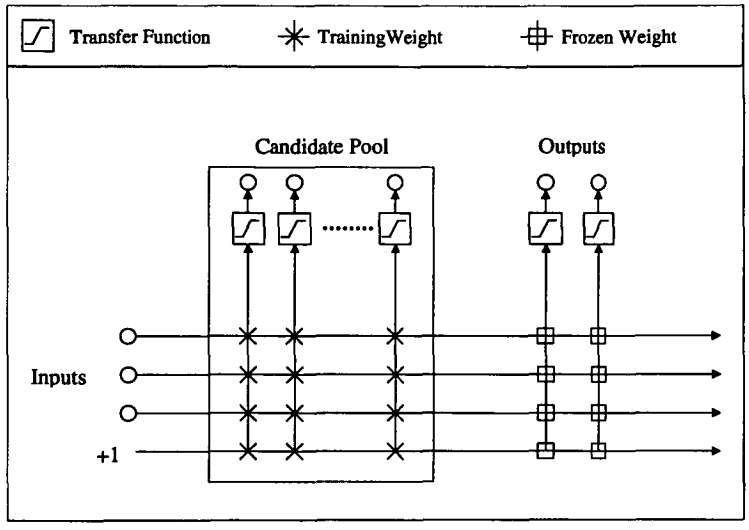


Figure 3-9: Candidate training candidate training phase.

At the end of the candidate training phase the best of the candidate units is selected and the other candidates are discarded. The output of this candidate is connected to the output units, the candidate input weights are frozen and the output weights are trained yet again (Figure 3-10).

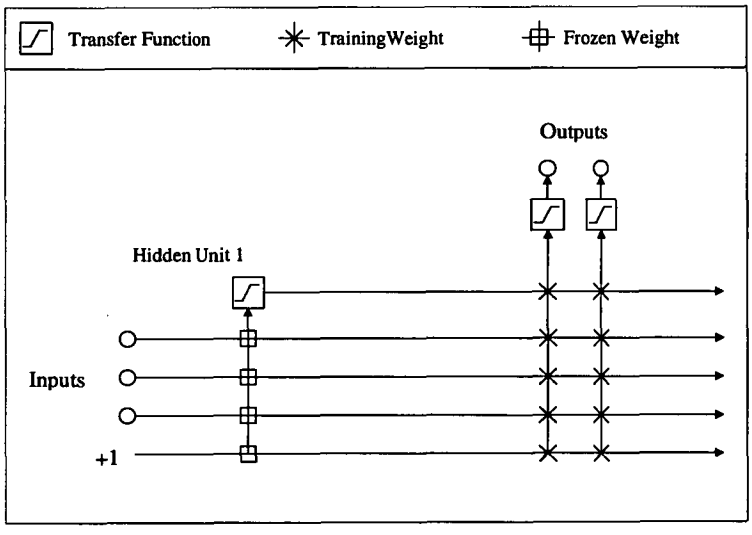


Figure 3-10: Output layer training phase.

This two phase training procedure is continued until no improvement in the overall training error is achieved. As is illustrated in Figure 3-11, the algorithm creates a cascade of hidden units during the course of training and new hidden units are not only connected to the input nodes, but also to all previously installed hidden units.

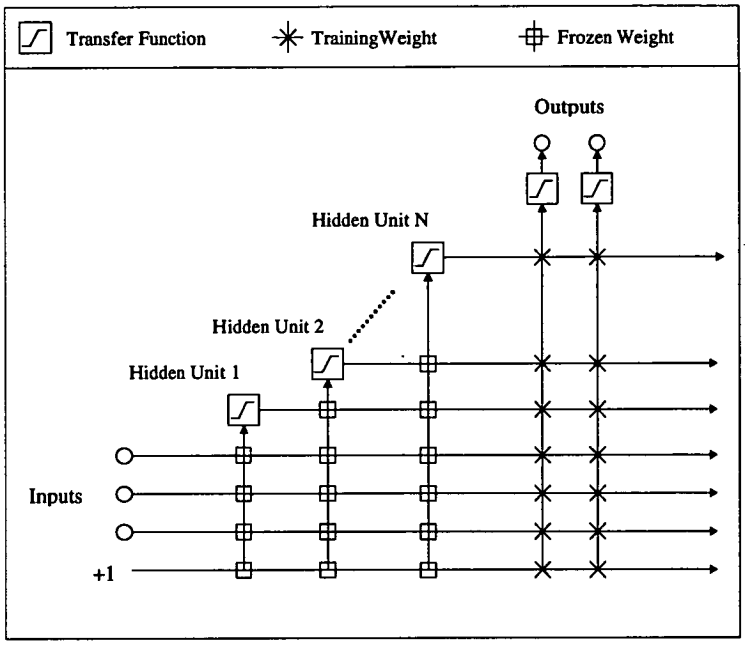


Figure 3-11: Output layer training phase.

Weight adjustment is performed using quickprop (Fahlmann 1988) and a candidate pool of 10 candidates is generally used (Waugh 1995) with sigmoid activation functions.

3.5 Concluding Comments

A number of neural network classification techniques have been presented in this chapter. These techniques will be applied to the classification of the electrocardiographic body surface maps (refer to chapters 6 through 9 for a detailed description of the experiments conducted.

4 The Heart and Standard Electrocardiography

This chapter aims to provide the reader with an understanding of the following topics:

- the physiology of the heart.
- the electrical behaviour of the heart.
- the electrocardiogram and its relation to heart function.
- coronary artery disease and myocardial infarction.
- 12-lead electrocardiography.
- detection of coronary artery disease and myocardial infarctions using electrocardiograms.

This chapter serves as an introduction to electrocardiography and as background to the following chapter which discusses the specifics of electrocardiographic body surface mapping.

4.1 The Heart

The human circulatory system is the transport system that maintains cell function throughout the body by performing the following tasks:

- Supplies substances absorbed by the digestive system.
- Supplies oxygen absorbed by the lungs.
- Returns carbon dioxide to the lungs.
- Returns other products of metabolism to the kidneys.
- Functions in the regulation of body temperature.
- Distributes hormones and other agents to regulate cell function.

The blood, the carrier of all these substances, is pumped through the body by the heart, which is not one, but two pumps operating in series. These two blood pumps form the left and right sides of the heart. The left side is responsible for pumping blood through the body, and on return, it is pumped through the lungs by the right side of the heart before beginning the cycle again (4-1).

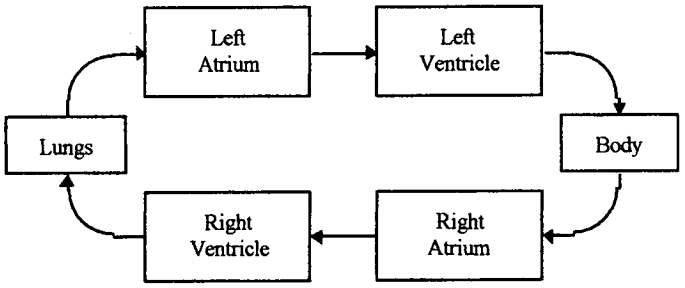


Figure 4-1: Circulatory System

A cross-section of the heart and blood flow is illustrated in Figure 4-2. The heart consists of four chambers: the left and right atria and the left and right ventricles. The pumping action of the chambers is achieved by contraction and relaxation of special muscle tissue in the chamber walls known as myocardial tissue. Contraction

and relaxation is associated with electrical changes in the muscle tissue. When contracting, the electrical change in the myocardial tissue is termed *depolarisation*, and when relaxing, *repolarisation.*

The rhythmic depolarisation and repolarisation of the myocardial tissue throughout the heart is regulated by the cardiac conduction system and is sourced by the sinoatrial (SA) node, the heart's natural pacemaker. This is located at the junction of the superior vena cava and the right atrium. The atria are directly depolarised by the SA node, whereas the atrioventricular (AV) node triggers the ventricles. Impulses from the SA node source pass via the internodal tracts, traversing around the wall of the right atrium, into the AV node. Then, after a slight delay, from the AV node down the Bundle of His into the Purkinje fibres, which activate the myocardial tissue in the ventricles.

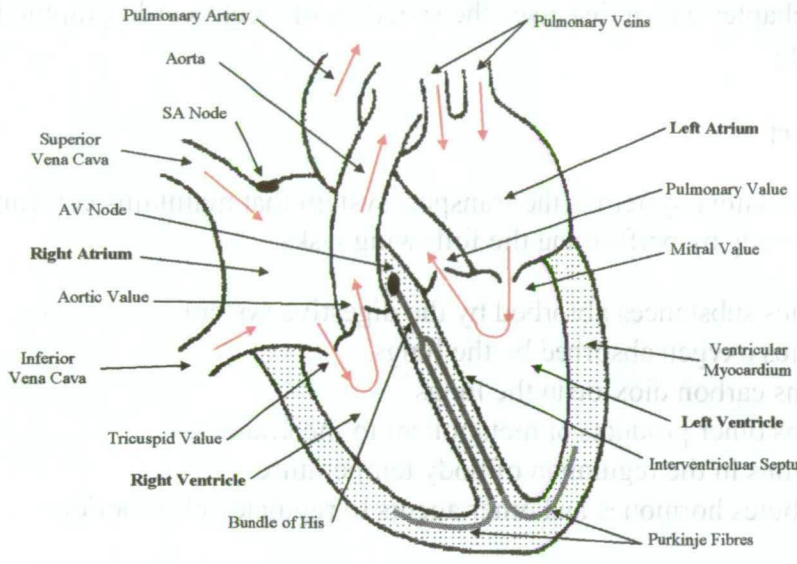


Figure 4-2: Components of the Heart (Ganong 1983)

The heart valves restrict the backward flow of blood. The tricuspid and mitral values allow blood to be pumped from the atria into the ventricles, but restrict blood flowing backwards from the ventricles into the atria. Similarly the aortic and pulmonary values allow blood to flow from the ventricles into the aorta and pulmonary artery respectively, but restrict backward flow into the ventricles.

The following section describes how all of the above mentioned components work together during the cardiac cycle.

4.1.1 The Cardiac Cycle

Although the heart is divided into two separate halves, both pumps actually function concurrently. The sequence of events in the cardiac cycle can be divided into three phases; contraction of the atria (*atrial systole*), then contraction of the ventricles (*ventricular systole*), and finally a relaxation phase (*diastole*) in which all four chambers relax. Figure 4-3 summarises the events that occur in these three phases.

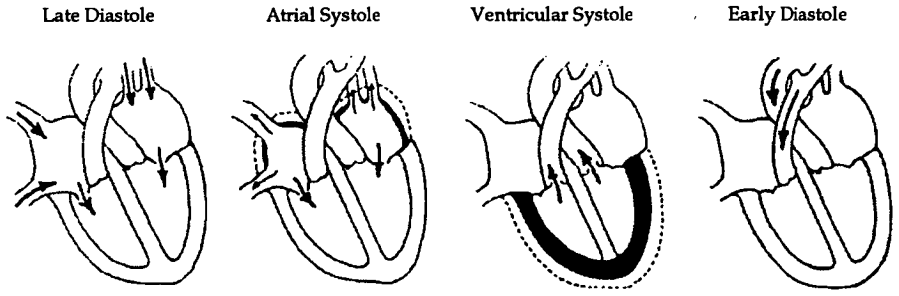


Figure 4-3: Events in the Cardiac Cycle (Ganong 1983)

4.1.1.1 Late Diastole

Late in the diastole phase, the mitral and tricuspid values between the atria and ventricles are open and the aortic and pulmonary valves are closed. Blood flows into the heart throughout diastole, filling the atria and ventricles.

4.1.1.2 Atrial Systole

Depolarisation initiated in the SA node spreads radially through the atria causing the atrial myocardial tissue to contract, narrowing the vena cava and pulmonary veins, and propelling blood into the ventricles. This depolarisation also propagates down the internodal tracts and converges on the AV node, where it is delayed by the slow conduction rate of the tissue in the AV node.

4.1.1.3 Ventricular Systole

Having passed though the AV node depolarisation quickly spreads to the ventricles via the Bundle of His and Purkinje fibre. Depolarisation of the ventricular muscle starts on the left side of the interventricular septum and moves first to the right side of the septum, down to the apex of the heart, and then along the ventricle walls. As the ventricles begin to contract the mitral and tricuspid (AV) valves close and the pressure in the left and right ventricles rises sharply until it exceeds the pressure in the aorta and pulmonary arteries. At this point, the pulmonary and aortic valves open and the blood in the ventricles is ejected into the arteries.

4.1.1.4 Early Diastole

Once the ventricular muscle is fully contracted, the ventricular pressure begins to drop and the aortic and pulmonary valves close. As the ventricles begin to relax the ventricular pressure drops even further and the AV valves begin to open, permitting the ventricles to fill (late diastole phase) starting the cardiac cycle again.

4.1.2 Coronary Arteries

Blood is supplied to the heart muscle tissue by the left and right coronary arteries. These arteries originate from aortic sinuses located immediately above the aortic valve, and wrap around the outside of the heart as illustrated in Figure 4-4.

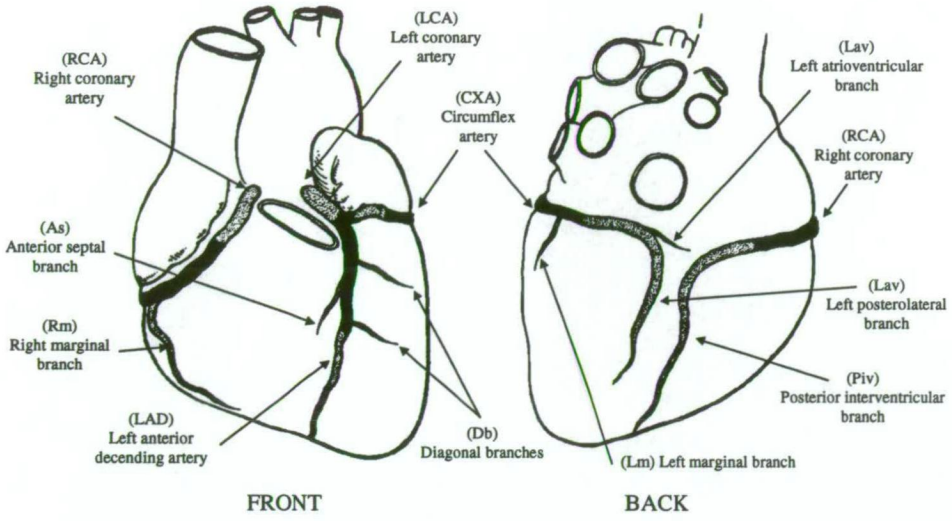


Figure 4-4: Coronary Ateries (Hunt et al.,1983)

The right coronary artery (RCA) passes from the aorta along a groove between the right atrium and ventricle. Just before the RCA wraps around the back of the heart the right marginal branch (Rm) arises and runs down the side of the right ventricle towards the apex. As the RCA wraps around the back of the heart it gives off its largest branch, the posterior interventricular branch (PIV), which runs down to the apex in the groove between the right and left ventricles.

The left coronary artery (LCA) passes down the front of the heart between the right ventricle and left atrium until it reaches the groove between the right and left ventricles. On reaching the grove between the ventricles the LCA gives off the left anterior descending branch (LAD) which continues down the ventricular groove to the apex. The LDA also gives off a number of other branches: the anterior septal branch (As) and the diagonal branches (Db), which branch out over the surface of the left ventricle. After the LAD branches the remainder of the LCA becomes the circumflex artery (CXA). The CXA wraps around the back of the heart along the groove between the left atrium and ventricle. On the back of the heart the CXA gives off the left marginal branch (Lm) and the left atrioventricular branch (Lav) before giving off its largest branch, the left posterolateral branch.

In summary, the right and left coronary arteries supply blood to the right and the left sides of the heart respectively although the balance of supply does tend to vary. For example, in 85-90% of people the post-inferior (back) aspect of the heart is supplied by the right coronary artery; this is referred to as right coronary dominance. However, in 10-15% of people the circumflex artery is dominant and supplies much of the post-inferior aspect of the heart. In this case the right coronary artery is small and the left coronary artery is responsible for the supply of blood to most of the heart.

At the apex there is some profusion of blood between the LAD and the PIV. There is also profusion of blood between the CXA and the Rm.

4.2 Coronary Artery Disease

Coronary artery disease is characterised by a narrowing of the coronary arteries. This is usually caused by focal deposits of fat, complex carbohydrates and platelets on the internal lining of the coronary arteries. These deposits can damage the arterial lining leading to the formation of plaques, scarring and calcification. As plaque deposits form the blood flow to the associated myocardial tissue is restricted.

The myocardial tissue will continue to function normally providing the restricted coronary artery can maintain sufficient blood flow to meet the oxygen demands of the associated myocardial tissue. However, if the oxygen demand exceeds that which the restricted coronary artery is capable of delivering, the myocardial tissue will become oxygen deficient and the patient will experience chest pain commonly referred to as angina. Angina is initially experienced when a patient physically exerts himself or herself, raising their heart rate and thus their myocardial oxygen demand. Initially rest relieves angina pain as this decreases the myocardial oxygen requirement.

Myocardial oxygen deficiency, or ischemic hypoxia as it is referred to medically, will not damage the myocardial tissue if sufficient oxygen levels are re-established quickly. Unfortunately, as coronary artery disease progresses, simply resting a patient may not be sufficient. The restriction of blood flow may be compounded by the formation of thrombosis (blood clots) around plaque deposits. If a thrombosis forms there is a risk of the artery becoming completely occluded and the myocardial tissue may be permanently damaged.

4.3 Myocardial Ischemia, Injury and Infarction

If a coronary artery becomes permanently occluded the myocardial tissue supplied by the artery will initially experience an oxygen deficiency, then injury, and finally cellular death. These three states are medically described as myocardial ischemia, myocardial injury and myocardial infarction.

The initial state of myocardial ischemia is generally associated with the onset of angina, as described in the previous section, and provided blood supply is re-established the myocardial tissue will return to normal function almost immediately. If ischemia progresses myocardial injury will begin to occur. At this stage the damage is still reversible and if the blood supply is re-established the myocardial tissue will recover. Finally, if myocardial injury is allowed to progress the myocardial tissue will be irreversibly damaged resulting in a myocardial infarction. These three stages are summarised in Table 4-1.

Myocardial State	*Problem*	*Tissue Damage*
ischemia	oxygen deficiency	none
injury	cellular injury	reversible
infarction	cellular death	irreversible

Table 4-1: Myocardial stages of damage due to coronary artery occlusion

The major concern with myocardial infarction is its impact on heart function. Since infarcted myocardial tissue is dead this in effect increases the load on the remaining myocardial tissue. If an infarct continues to spread there is a high risk of heart failure and the patient could die. Therefore it is important diagnostically that the location and extent of myocardial damage can be determined, as this will assist the physician in providing the most appropriate treatment.

For clinical purposes myocardial damage is broadly described as occurring on one of three surfaces of the heart; the anterior surface, the inferior surface, or posterior surface. The anterior aspect of the heart refers to the front of the heart. The inferior aspect refers to the bottom or basal aspect of the heart. The posterior aspect refers to the back of the heart. The majority of myocardial ischemia, injury and infarction tend to occur in the myocardium of the left ventricle. Cases of right ventricular damage are uncommon and tend to be brought on by other cardiac complications. Therefore anterior damage generally refers to damage in the front of the left ventricle, inferior damage generally refers to damage in the basal aspect of the left ventricle and posterior damage generally refers to damage toward the back of the left ventricle. This is illustrated in the cross-sectional diagram in Figure 4-5.

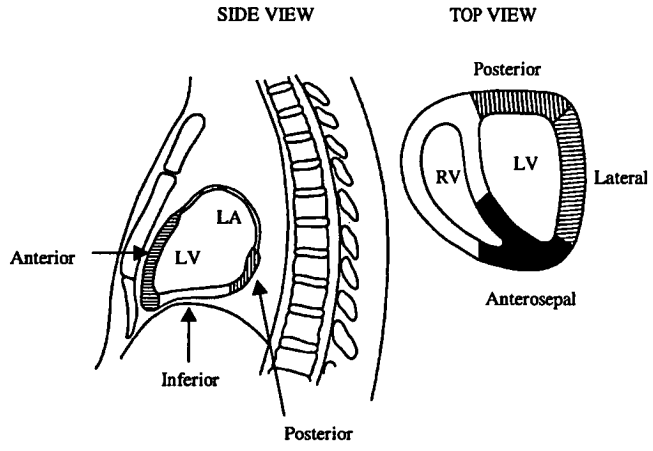


Figure 4-5: Cross-section of the left ventricle walls (Droste 1997)

For an initial diagnosis, classifying myocardial damage as anterior, inferior or posterior is generally sufficient, but more specific classifications are often useful when determining the impact of the damage and in particular when classifying the location of myocardial infarcts. Table 4-2 summarises the seven more specific classes assigned to myocardial infarcts. Four are anterior types; anterior, anteroseptal, apical anterior, and anterolateral. Two are inferior types; inferolateral, interior, and posterior infarcts are often assigned to a separate class.

Location of Infarction	*Cause*
Large Anterior	Occlusion of the LAD
Anteroseptal	Occlusion of peripheral parts of LAD As
Apical anterior	Occlusion of the peripheral parts of LAD
Anterolateral	Occlusion of peripheral parts of LAD and inclusion of Db
Inferolateral	Occulusion of the Lm
Large Inferior	Occlusion of the RCA or CXA
Posterior	Occlusion of peripheral parts of CXA, especially of Lav

Table 4-2: Classification of left ventricular infarcts with respect to location.

Each of these infarct types is generally caused by the occlusion of specific coronary arteries. A large anterior walled infarct is usually brought about by an occlusion of the LAD close to its origin. As such the infarct will tend to impact significant proportion of the anterior surface of the left ventricle. The approximate location of such an infarct is illustrated in Figure 4-6.

In many cases a large anterior walled infarct will tend to spread into most of the anterior surface and may well impact the peripheral aspects of the anterior surface, namely anteroseptal, apical anterior, and anterolateral. However, the peripheral anterior infarcts tend to be caused by occlusion of peripheral branches of the LCA. An anteroseptal infarct is generally caused by occlusion of the As branches. An apical anterior is caused by occlusion of the latter sections of the LAD. An anterolateral infarct is generally caused by occlusion of the Db branches.

On the other hand inferior and posterior infarcts tend to be caused by occlusion of the arteries wrapped around the back of the heart. A large inferior infarct will tend to be caused by occlusion of peripheral aspects of RCA or CXA. An inferolateral tends to be caused by occlusion of the LM branch. Finally a posterior infarct is generally caused by occlusion of peripheral parts of CXA especially the Lav.

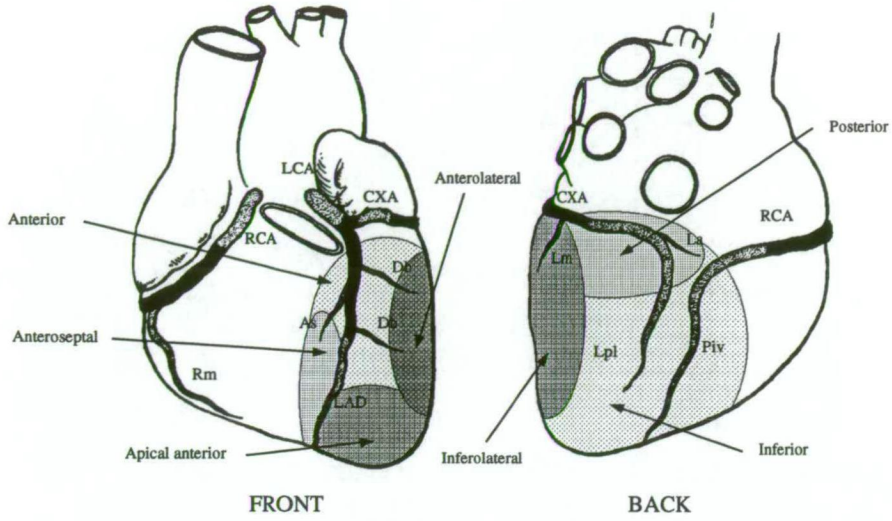


Figure 4-6: Infarct Locations (Hunt et al., 1983)

This is a somewhat simplistic description of the regions of myocardial damage and the associated arteries. Although it serves as a good starting point it must be understood that arterial dominance and arterial layout may differ significantly from patient to patient. Therefore one cannot always be guaranteed that having identified the location of the myocardial damage the corresponding arteries described above are guaranteed to be directly responsible. When myocardial damage is present in a region of the myocardium it is also important to be aware that this damage may not be present in only one of the three states described previously (injury, ischemia and infarction). For example if an infarct is present in a region in the myocardium it will most probably be surrounded firstly by ischemic tissue, which will in turn be surrounded by injured tissue. Therefore from a diagnostic point of view when observing a patient with myocardial infarction the symptoms associated with injury and ischemia may well be present along with the symptoms of the infarct. Similarly if myocardial ischemia is present and has not progressed to the point of infarction then the diagnostic symptoms of both ischemia and injury will both be present.

Therefore although the cause, impact and progression of myocardial damage is clearly understood, in the clinical setting care must be taken when analysing diagnostic data, particularly of a non invasive nature as the specific arterial dominance and arterial layout may differ to a certain extent from the clinical norm. As such since electrocardiography is a non-invasive technique care needs to be taken when drawing conclusions about the observations made.

The identification and classification of myocardial damage is extremely important in the clinical setting for the purposes of determining treatment. Electrocardiography is extremely useful as an initial assessment tool in determining the location and extent of myocardial damage. The following section will describe the features of the electrocardiogram and how they relate to heart function and then discuss how electrocardiograms can be used to detect and diagnose myocardial ischemia, injury and infarction.

4.4 Electrocardiography

Electrocardiography is an extremely useful diagnostic tool used by physicians for the assessment of heart disorders. As the human heart beats the heart muscle generates an electrical field that can be observed on the body surface. An electrocardiogram (ECG) is a recording of this electrical activity and can be used by a physician or cardiologist in the assessment of a patient's heart function.

The study of ECGs and their clinical significance began around the turn of the century when it was discovered that the electrical activity of the heart could be detected using a galvanometer connected to two electrodes placed on or in the body. It was soon discovered that the electrical field generated by the heart was complex and could not be observed in its entirety with one pair of electrodes. As a result much research was conducted in the 1920s and 1930s in an attempt to determine the most appropriate number and location of electrode pairs that provided a sufficient summary of the heart's activity. By the late 1930s numerous configurations had been put forward for recording ECGs. In 1938 in an attempt to rationalise the procedure of ECG recording a joint committee of the American Heart Association and the Cardiac Society of Great Britain and Ireland assessed the current approaches to ECG lead placement and recording and developed the first standard (American Heart Journal 1938). This underwent refinement and in 1954 the American Heart Association defined the 12-lead ECG standard (Circulation 1954) now being used by cardiologists worldwide.

The following sections will present a description of ECG features and the 12-lead ECG configuration. This will be followed by a description of how 12-lead ECGs can be used to detect the location and extent of myocardial damage.

4.4.1 The Electrocardiogram

The electrical potential of the myocardial tissue in the atria and ventricles produces electrical forces in the heart. Because body fluids are good conductors these electrical forces can be observed on the surface of the body. Therefore as the heart moves though the cardiac cycle the electrical potentials on the surface of the body will change according to the electrical potential of the heart. An electrocardiogram (ECG) is a simple graphical representation of such electrical forces recorded by two electrodes on the surface of the body.

The names of the various waves of the ECG are shown in Figure 4-7. There are three distinct features of this signature: the P wave, QRS waves (commonly called the QRS complex), and the T wave. The P wave is produced by atrial depolarisation, the QRS complex by ventricular depolarisation and the T wave by ventricular repolarisation. Atrial depolarisation is normally submerged in the QRS complex. Two other features often referred to in the ECG are the PQ segment and the ST segment: located between the PQ and ST waves respectively.

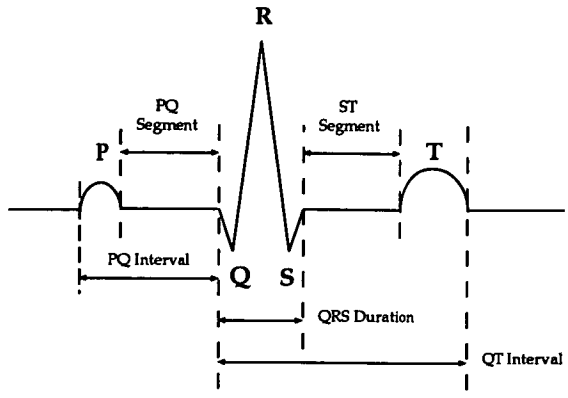


Figure 4-7: Features of the ECG (Hampton 1986)

The QRST waves all occur during the systolic phase of the cardiac cycle and the remainder of the ECG wave (after the T wave and before the Q wave) constitutes the diastolic phase. The magnitude of these waves tends to be of the order of 1-2mV. The average timing of various intervals of the ECG signature are summarised in Table 4-3.

Interval	Duration (ms)
PQ Interval	120 - 200
QRS Duration	80 - 100
QT Interval	400 - 430
ST Segment	320 - 350

Table 4-3: Average duration of ECG Intervals (Ganong 1983)

4.4.2 Bipolar leads

As mentioned in section 4.4 the earliest recordings of ECGs were performed by measuring the changing voltage potential between two electrodes placed on the surface of the human body. This type of lead placement and measurement is known as *bipolar lead* configuration. In this type of configuration one electrode is referred to as the *reference electrode*, and the other as the *active electrode*. Therefore the voltage signal recorded with a bipolar lead is a relative signal with respect to the *reference electrode.*

Despite the relative nature of bipolar lead configuration the primary features of the ECG (as illustrated in Figure 4-7) can still be observed. It is important to note though that the magnitude and sign of these features will vary depending on where the recording electrodes are placed on the body surface. The electrical potential generated by the heart is not a single voltage source but a somewhat more complex electrical dipole created by the potential difference between the polarised and depolarised myocardial tissue. This electrical field can be summarised as a single electrical vector (or cardiac vector) that continually changes magnitude and direction during the cardiac cycle (Figure 4-8).

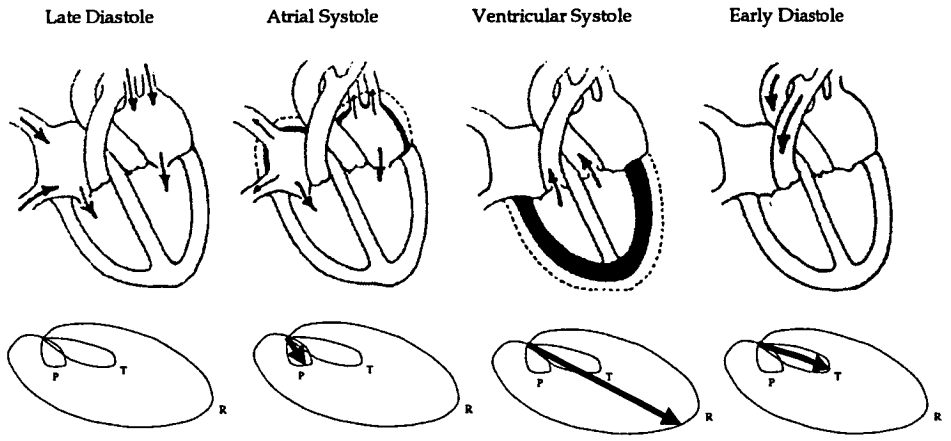


Figure 4-8: Direction and magnitude of the cardiac vector during the cardiac cycle (Winsor 1969)

Therefore, a bipolar ECG reading measures some component of this cardiac vector with respect to the two electrodes concerned (see Figure 4-9). Consequently, if the cardiac vector is pointing towards the active electrode then the ECG reading at that instant will be positive, if pointing towards the reference electrode then the reading will be negative. Note also that if the cardiac vector is perpendicular to the electrode axis then the reading will be zero.

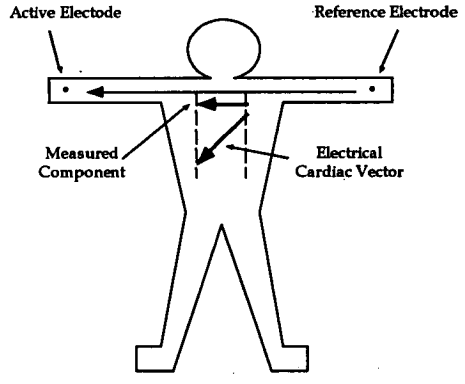


Figure 4-9: Component measurement of cardiac vector.

It is important to appreciate that this type of ECG recording only measures the potential difference between the two electrodes and therefore the voltages being observed are simply relative to the reference electrode. However, the information provided by bipolar lead configurations is extremely useful and as we will see in section 4.4.4 a number of bipolar lead configurations were adopted in the 12-lead ECG standard.

4.4.3 Unipolar Leads

Much of the early research into ECG lead configuration was concerned with finding the optimal set of bipolar leads to provide the most effective diagnostic information. One of the on-going problems which plagued lead design was the fact that measurements were relative to one or a number of reference electrodes and therefore did not provide any absolute measurement of voltages on the body surface.

Wilson, Macleod and Barker solved this problem (Wilson, Macleod and Barker 1932) when they proposed a technique for constructing a zero-potential-centre-terminal, now commonly know as the Wilson centre terminal. The concept of the zero-potential-centre-terminal was based on earlier work by William Einthoven (Einthoven 1913), one of the founding researchers in electrocardiography. Einthoven proposed that the summation of three electrodes placed in a triangle on the body surface would result in a zero potential on the basis of the following assumptions:

- body is a large conducting medium
- medium is homogeneous and resistive
- source of potential is a dipole
- diploe is at the centre of the medium
- dipole undergoes no change in position during the cardiac cycle

Thus Einthoven proposed that the summation of electrodes placed on the right arm (VR), left arm (VL) and right leg (VF) would effectively result in a zero potential throughout the cardiac cycle:

$$\text{Einthoven's Triangle: } VR+VL+VF = 0$$

Based on this work Wilson, Macleod and Barker suggested that a summation of electrodes placed on the right arm, left arm and right leg could be used as an absolute reference electrode. Although there are many assumptions associated with this concept and there have been many publications over the years challenging the zero potential of the centre terminal, the Wilson centre terminal has been found to not vary more than 0.3 of a millivolt over the entire cardiac cycle. It is therefore considered to be an extremely useful reference.

Using this centre terminal Wilson, Macleod and Barker demonstrated the clinical usefulness of using the terminal as a reference electrode in combination with active electrodes place in a range of locations on the body surface. As a result the term *unipolar lead* was coined to describe this technique. Although somewhat of a misnomer, since two poles are still involved (the Wilson centre terminal and the active electrode), it serves as a good description for the technique.

The following section will describe how the use of lead combinations was rationalised with the development of the 12-lead ECG standard.

4.4.4 12-Lead ECG

To construct a more complete picture of heart behaviour it is necessary to record more than one ECG. As mentioned in the introduction, the American Heart Association defined the 12-Lead ECG standard in 1954. This standard requires the placement of nine electrodes on the patient; six on the left side of the chest (V1 to V6), two on the left and right arms (VR and VL), and one on the left leg (VF) as illustrated in Figure 4-10. Using these electrodes, three bipolar leads and nine unipolar leads can be constructed.

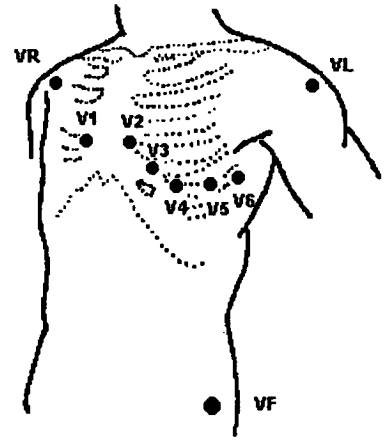


Figure 4-10: Lead placement for 12-lead ECG recording

The three bipolar leads I,II, and III, often referred to as the standard limb leads, record the differences in potential between the three electrodes attached to the extremities. Lead I measures the potential difference between VL and VR where VR is used as the reference electrode. Lead II measures the potential difference between VF and VR where VR is used as the reference electrode. Lead III measures the potential difference between VF and VL, where VL is used as the reference electrode. Figure 4-11 illustrates the vectors along which each lead measures the cardiac dipole.

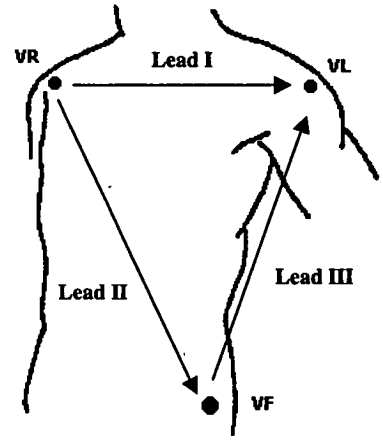


Figure 4-11: Bipolar lead configuration.

The nine unipolar leads consist of six leads measured with respect to the Wilson centre terminal (V_{Wilson}) and three *augmented* leads. The three augmented leads; aVR, aVL and aVF, although not measured with respect to V_{Wilson} are still effectively unipolar leads. Lead aVR measures the potential difference between the active electrode VR and a reference electrode constructed by averaging VL and VF. Lead aVL measures the potential difference between the active electrode VL and a reference electrode constructed by averaging VR and VF. Lead aVF measures the potential difference between the active electrode VL and a reference electrode constructed by averaging VR and VL.

Lead	Measurement	Result (substituting VR+VL+VF=0)
aV_R	VR-(VL+VF)/2	3/2 VR
aV_L	VL-(VR+VF)/2	3/2 VL
aV_F	VF-(VR-VL)/2	3/2 VR

Table 4-4: Augmented Leads

The averaged electrodes are constructed by placing a resistor pair in series between the two electrodes and then tapping the voltage between the two resistors in effect measures an average voltage of the electrode pair. Considering the mathematics associated with these leads and applying Einthoven's Triangle (see Table 4-4) these leads are effectively unipolar apart from a 50% increase in magnitude; aVR = 3/2 VR, aVL = 3/2 VL and aVF = 3/2 VF. Figure 4-12 illustrates the vector along which each lead measures the cardiac dipole.

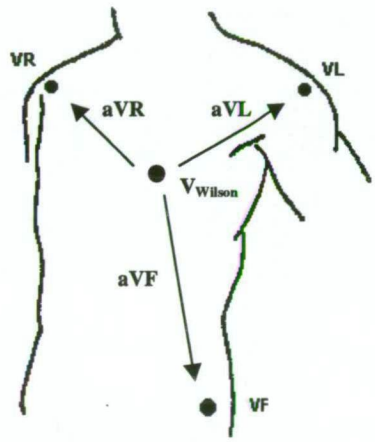


Figure 4-12: Augmented lead configuration (Hampton 1986).

Finally the unipolar leads V_1, V_2, V_3, V_4, V_5 and V_6 are constructed by measuring V1, V2, V3, V4, V5, and V6 with respect to V_{Wilson} (ie VR+VL+VF). These leads effectively measure the cardiac dipole along vectors radiating out of the chest surface as illustrated in Figure 4-13.

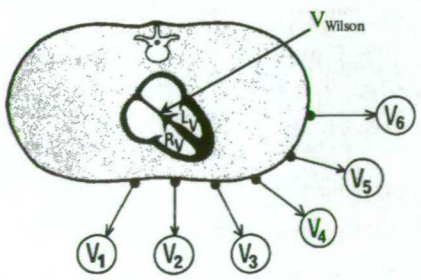


Figure 4-13: Unipolar lead configuration (Hampton 1986).

Therefore, in summary the 12-lead ECG configuration consists of nine electrodes (VR, VL, VF, V1, V2, V3, V4, V5, and V6) placed on the body surface from which 12 lead configurations are derived (I, II, III, aVR, aVL, aVF, V_1, V_2, V_3, V_4, V_5 and V_6). The American Heart Association selected this configuration because it provided the most effective and complete diagnostic information.

The following two sections will describe how this lead configuration can be used to detect myocardial damage and coronary artery disease. Although there are a range of other disorders that can be detected using this configuration they will not be presented as they are outside the scope of the diagnostic problems considered in this thesis.

4.4.5 Detection of Myocardial Ischemia, Injury, and Infarction

When coronary arteries are restricted or occluded, as described in section 4.3, then the associated myocardial tissue is impacted. It has already been described how this can result in myocardial damage. As myocardial damage occurs the electrical behaviour of the associated myocardial tissue begins to change and causes distinct changes in the ECG features. These distinctive electrical changes are useful indicators of myocardial damage.

In the case of myocardial infarction the heart muscle is dead and the tissue cannot be either polarised or depolarised. Consequently as the heart polarises and depolarises, the resulting electrical dipole generated by the heart will be significantly different from the norm. Similarly, injured and ischemic tissue tends to be extremely slow at depolarising.

There are distinct changes in the ECG features associate with myocardial ischemia, injury and infarction. Myocardial ischemia is associated with an inversion of the ECG T wave. Myocardial injury is associated with an elevation of the ST segment, whilst myocardial infarction is associated with a deepening of the Q wave. This is summarised in Table 4-5.

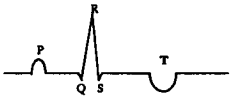
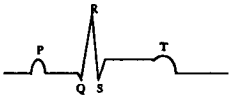
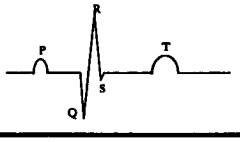
Myocardial State	Problem	Tissue Damage	ECG Change
Ischemia	Oxygen deficiency	none	T-Inversion 
Injury	Cellular injury	reversible	ST-Elevation 
Infarction	Cellular death	irreversible	Large Q Wave 

Table 4-5: Myocardial damage and associate ECG changes.

As stated in section 4.3 an infarction will in general be surrounded by ischemic and injured tissue and therefore all three ECG features (T inversion, ST elevation and deepened Q wave) will be present when an infarction occurs. These features will tend to be apparent at different stages dependent on the passage of time and how the heart has recovered. In the early stages, up to a day after the infarct has occurred, a large Q wave and elevated ST segment will be observed indicating infarcted and injured tissue. In the weeks and months that follow as the patient begins to recover a large Q wave will be present in the ECG, the ST segment elevation will decrease and the T wave will invert indicating the transition of some injured tissue into a state of ischemia. About two years following the onset of the infarct the ST segment and T wave will return somewhat to normal but with the continued presence of the deepened Q wave. This indicates the recovery of ischemic and injured tissue leaving mainly infarcted tissue.

Clearly the information provided by the ECG not only assists in identifying the state of a patient's heart but is also extremely useful for monitoring recovery. Note also that from a diagnostic point of view, ECGs can also be used to ascertain approximately when an infarct occurred.

Electrocardiography is not only useful for detecting myocardial damage, but also extremely useful in identifying the location of the damage. Since ECG leads provide information relative to the vector along which the measurement is being made (as highlighted in sections 4.4.2, 4.4.3, and 4.4.4) then the leads that measure along a vector perpendicular to the damaged surface will exhibit changes in the ECG features.

In relation to 12-lead ECG analysis the location of myocardial damage is determined on the basis of where signal changes occur. A summary of the types of infarction (see section 4.3) and the associated leads of diagnostic significance is given in Table 4-6. This is not definitive but provides a good indication as to where the changes may be observed.

Location of Infarction	*Lead s of diagnostic significance*
Large Anterior	V_1, V_2, V_3, V_4
Anteroseptal	V_2, V_3
Apical anterior	I, III, V_3, V_4
Anterolateral	I, III, V_3, V_4, V_5
Inferolateral	II, III, aV_L
Large Inferior	aV_F, I, II, III, V_4
Posterior	V_3, V_4

Table 4-6: Infarct locations and associated leads (Droste 1997)

In broad terms anterior damage is detected using V_1 through to V_6 as these leads measure cardiac behaviour perpendicular to the anterior surface. Similarly, inferior damage is detected using leads I, II, aV_F and aV_L as these are located perpendicular to the inferior surface. Posterior damage is generally detected by V_3 and V_4, as these

leads are located on the opposing surface of the heart. In some cases additional back leads have been found useful in locating posterior damage.

It is important to note that although discrimination between anterior, inferior and posterior damage is generally clear, the classification into sub-classes is often difficult as the rotational position of the heart inside the chest tends to vary from patient to patient. Therefore care must be taken when analysing ECG data and a physician needs to be aware of the analytical limitations of the technique.

4.4.6 Detection of Coronary Artery Disease

The detection of coronary artery disease using ECG data is much more difficult. If ischemia or injury has not occurred then a patient with coronary artery disease can present normal electrocardiograms. There has been significant debate concerning this problem. Some researchers have suggested that the detection of coronary artery disease is possible using ECG data but that 12-lead ECGs provide insufficient information and more complete electrocardiograms are required (eg electrocardiographic body surface mapping) to detect coronary artery disease. This is a complex issue, but should be considered carefully in light of the heart's physiology. The reality is that until ischemia or injury occurs there is no change in the electrical behaviour of the heart. As such, it is difficult to argue that a lack of information is the root cause of this diagnostic problem, as it is unclear from a physiological point of view what electrical changes we are trying to identify in an effort to detect coronary artery disease.

One of the other aspects of heart function which makes coronary artery disease difficult to identify is collateral supply. The heart has a certain degree of redundancy in the arterial supply of blood to the myocardium. As such it is quite possible for an arterial occlusion to exist without any myocardial damage occurring. In such a case the neighbouring arteries may well provide sufficient blood to the impacted region and thus allow the myocardial tissue to continue functioning normally. Such a patient will exhibit no detectable change in the ECG signatures particularly if the patient's heart function is observed at rest.

One of the proven techniques for detecting the presence of coronary artery disease is to observe the ECG signatures while the patient exercises (ie using an exercise bicycle or treadmill). As the patient's heart is worked harder the oxygen demands of the myocardial tissue will increase and in many cases the restricted arteries and collateral supply are unable to meet the demand. In effect the myocardial tissue will tend to progress into a state of mild ischemia which can be observed by the associated ECG changes (ie T wave inversion).

4.5 Concluding Comments

This chapter has provided a detailed description of heart physiology, its electrical behaviour, and how coronary artery disease and myocardial damage impact heart function. It has provided an introduction to electrocardiography and how it may be used to detect and locate myocardial damage. It has presented not only the benefits of 12-lead ECG analysis but also highlighted a number of limitations of this diagnostic technique.

Although this background is not directly related to electrocardiographic body surface mapping techniques it serves as an important starting point in understanding many of the issues associated with ECG analysis.

The following chapter will extend this material by presenting the specific background associated with electrocardiographic body surface mapping. Although this technique is more extensive with respect to data gathering, the principle relationships between electrocardiographic behaviour and heart function remain the same.

5. Electocardiographic Body Surface Mapping

This chapter presents the historical background to the field of body surface mapping as well as presenting the range of techniques used for recording and analysing BSMs. As such this chapter is divided into three sections. Section 5.1 provides a brief historical background, section 5.2 presents a summary of the BSM data recording techniques currently used by researchers, and section 5.3 describes the standard techniques used to analyse and classify BSMs.

5.1 History

One of the major assumptions in cardiac research up until the mid 1950s was that the electrical field generated by the heart at any instance could be summarised as an electrical dipole located on the heart surface. As a consequence many researches considered that the 12-lead approach was inappropriate and contained redundant data and believed that a far more logical ECG need only record the X,Y and Z components of the heart's electrical dipole (Williams 1914, Mann 1938, Wilson et al. 1938). In 1956 the single dipole concept was challenged by Nelson (Nelson 1957) who showed (by the use of an electrocardiographic belt strapped around the human thorax) that the potential distribution observed at any instance on the chest surface was generated by more than one dipole. A single vector could not represent the electrical output of the heart and thus a more comprehensive representation was required.

A number of researches in the 1960s followed up Nelson's research and began to develop techniques for mapping the surface potentials of the entire thorax in the hope that this would provide a more complete representation of the heart's behaviour (Tacacardi 1957-1963, Amirov 1961, Horan et al. 1963). The recording of such body surface maps was a time consuming task as maps were constructed from a grid of between 50 and 600 ECGs. Complicating the matter further, researchers were limited by recording equipment having to record ECGs in batches (from anywhere between 2 to 20 ECGs at a time). Although this research was extremely fruitful, the use of such body surface mapping techniques for clinical diagnosis was technically impractical.

In the 1970s the problem of recording body surface maps (BSMs) was simplified with the use of computing technology, making it possible to record and store all electrode potentials simultaneously and quickly display the resulting surface potential maps (Taccardi et al. 1976, Kilpatrick et al. 1979, Spach et al. 1979, Heringa et al. 1981, and Yajima et al. 1983). As a consequence research into body surface mapping expanded and the 1980s saw a move towards the use of BSMs in clinical diagnosis.

5.2 BSM Construction

The results of a body surface map recording are generally displayed as a series of iospotential maps as shown in Figure 5-1.

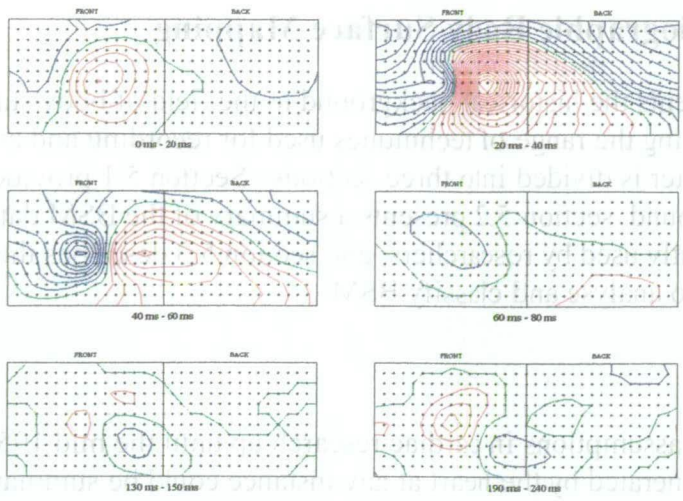


Figure 5-1: Isopotential body surface map
(contour intervals – 10μV, red indicates positive potential,
blue indicates negative potential, and green indicates zero potential).

The construction of BSMs is a four stage process (Figure 5-2) which involves the placement of electrodes on the patient, data aquisition, data preparation, and finally the calculation and display of a series of isopotential BSMs. This first section will describe the methods used in each of these stages.

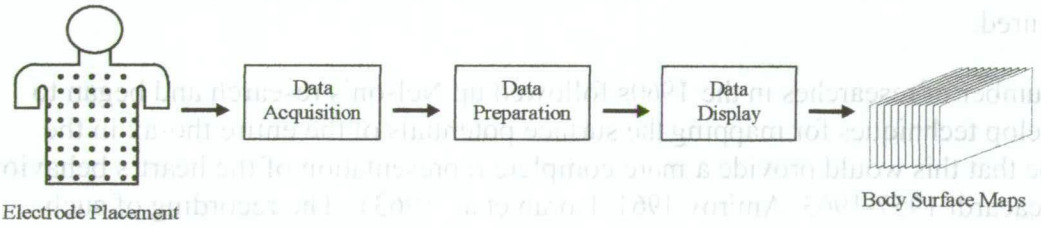


Figure 5-2: Four stages of BSM construction.

Although there are some moves towards the standardisation of BSM recording and display (Horacek 1985, Teramachi 1985) it is important to note that at present there are a range of BSM systems in use with significant differences in approach. As such, this section will attempt to describe the range of techniques used.

5.2.1 Electrode Systems

One of the critical factors in the performance of any body surface mapping systems in the design of the electrode array which will record all the relevant cardiac potentials on the thoracic surface. The most common approach used in the literature is the comprehensive lead system.

Comprehensive lead systems use an electrode array which measures all or most of the thoracic surface. Typically this array would consist of somewhere between 50 and 240 electrodes evenly spaced over the thoracic surface (Figure 5-3).

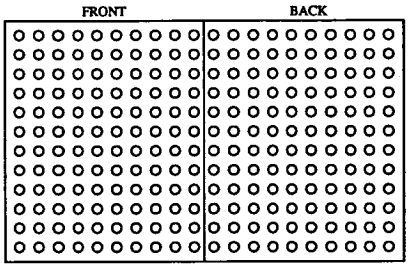


Figure 5-3: Typical Comprehensive Lead System (240 electrodes)

When compared with the lead placement of other systems (Figure 5-4), such as standard 12-lead ECG and vectorcardiography, it is clear that this approach covers far more of the thoracic surface. As such it is reasoned that a comprehensive system will provide more diagnostic information, and also provides some level of redundancy, allowing for the possibility of noisy leads.

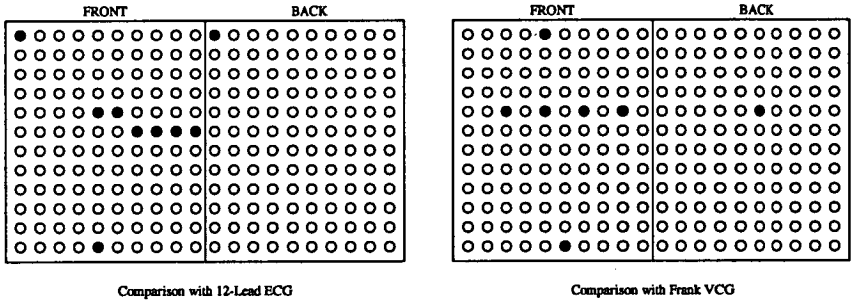


Figure 5-4: Comparison of comprehensive lead placement with standard 12-lead ECG and vectorcardiography

The optimum number of leads to use in a comprehensive lead system is unknown. Increasing the number of leads does provide greater map detail, as well as providing physical redundancy to compensate for the possibility of noisy leads. However, increasing the number of leads also increases the complexity of the recording equipment, and as such can impact upon the recording accuracy (Barr and Spach 1983). In general whilst up to 240 electrodes are manageable in a research environment, it should also be noted that little qualitative difference has been found in maps constructed from 85 to 121 electrodes (Yamada et al. 1978).

An alternative to comprehensive lead systems is subset lead systems. Subset lead systems attempt to reduce the number of electrodes needed to record a body surface map, without degrading the information content. The argument generally given is that most comprehensive systems tend to record a lot of redundant information, which can be calculated by a small set of carefully selected leads.

Barr (Barr et al. 1971and Spach,Barr 1971) and Lux (Lux et al. 1978 and Lux et al. 1979) have done a significant amount of work on determining the optimum lead set that can be used without a reduction in information content (Barr - 24 electrodes, Lux - 32 electrode).

Although the reasoning for the use of a subset lead system is sound it does require more careful placement of electrodes, which for clinical purposes tends to be impractical. The advantage of a comprehensive lead system is that lead placement is not critical, and normalisation of electrode positions can be achieved during the post-

processing phase. Therefore comprehensive lead systems tend to be more widely used in clinical settings.

Ranges of approaches have been used to place leads on patient. The most common being strips of electrodes running vertically on the torso and attached to the skin with double-sided adhesive tape. Other methods include belts of electrodes running around the torso (Yajima et al. 1983), electrodes mounted on a vest (Liebman et al. 1981 and Walker 1983) and electrodes held on by suction (Reek et al. 1984). Some systems use active electrodes, which eliminate the need for conductive jelly when the electrodes are applied. Others use jelly, which tends to increase the time taken to apply the electrodes, but allows a simpler electrode construction.

As with any electrocardiogram, the electrode signals must be measured as a potential difference with respect to some reference. The typical approach used in BSM mapping is to record all electrodes with respect to the Wilson centre terminal voltage (refer to the previous chapter for a description). This reference can be constructed by physically setting up a Wilson centre terminal (ie attaching VL, VR, and VF). Alternatively the electrodes can be recorded relative to an ankle lead and then all resulting signals are offset during the post-processing phase by a subtracting a centre terminal constant calculated from electrodes positioned at VL, VR, VF. This second approach is commonly used and has been shown to have no adverse affect on the resulting isopotential contours (Taccardi 1962).

5.2.2 Data Aquisition

The next crucial aspect of any BSM system is the recording hardware, which measures and records the electrode signals. In most cases purpose build hardware is used to amplify, filter and digitise electrode signals. This hardware is usually interfaced to some form of standard computing equipment, which stores the resulting digital data.

The electrode signals require amplification in order to obtain a signal, which can be sampled accurately by the analog-to-digital sampling hardware. Considering that the maximum electrode signal that would be observed is ~10mV, and most analog-to-digital conversion hardware samples in a range of ±5 volts, an amplifier gain of the order of 10^3 is required to achieve maximum discrimination of the electrode signal.

As well as amplification, the electrode signals should be filtered to remove muscle and background electrical noise. The American Heart Association recommend the use of a band-pass filter of 0.05 - 100 Hz, although a number of BSM systems have used 0 - 250 Hz as it has been found that significant information pertaining to myocardial scarring tends to appear at these higher frequencies.

The sampling rate of the analog-to-digital conversion is very much dependant on the maximum frequency content of the amplified signals. For general sampling theory it is recognised that the sampling rate should be twice that of the maximum frequency that will be observed. This will in effect be determined by the band-pass filter being applied to the signal, and as such the sampling frequency should be at least twice the maximum frequency of the band-pass filter. Thus a sampling rate of 200-500 Hz is common, and a number of systems use a sampling rate of up 1000Hz.

Another important issue pertaining to analog-to-digital conversion is the word length of each sample, as this will impact upon the effective voltage discrimination. If an 8-bit A-to-D converter is used, with an amplifier gain of the order of 10^3, then this will permit a voltage discrimination of approximately 40uV. Alternatively a 10-bit A-to-D converter will provide a resolution of 10uV, and a 12-bit A-to-D converter will provide a resolution of 2.5uV. The issue of resolution may become significant when considering electrode signals on the patient's back, or when considering signals in the P-R segment which may be lower than 120uV.

In summary, the selection of data acquisition hardware will often depend upon the type of problems the system is to be used on. Ideally, a high sample rate and long sample word length are best, but this is often practically impossible, due to the demands on hardware. For example, if 150 electrodes are sampled at 1000Hz simultaneously, a data rate of 150 000 samples per second is required, which in some cases is beyond the performance capability of current high resolution A-to-D converting hardware. Therefore, there are often practical restrictions on the hardware.

5.2.3 Data Preparation

In preparing BSM data for display and analysis, there are a number of post-processing techniques used to address some problems encountered during the recording procedure. The two most common problems found when recording are the presence of noise and DC-offsets in the recorded data. Both need to be minimised, particularly in situations where the data is being used for a comparative study.

The reduction of noise, due to muscle tremors and environmental conditions (ie. power signals), can be achieved firstly by careful construction of the band-pass filter hardware, but inevitably the effects of such noise are observed to some degree in the resulting data. The most common approach used to deal with noise during the post-processing phase is to conduct a point-by-point averaging of a number of heart cycles. This approach will attenuate the noise in proportion to the square root of the number of waveforms averaged.

The environmental conditions and imperfections in the signal amplifiers generally cause the presence of a DC-offset in recorded data. As with noise, DC-offset can be reduced by careful construction and tuning of recording hardware, but post-processing is still necessary. The standard approach used is to sample a section of the T-P segment on the heart cycle, and subtract this constant from each point of the sampled signal. Note that this should be done for each electrode recording. The reason for selection of the T-P segment of the heart cycle is that the heart is generally at rest during this phase.

One final adjustment that may be required during the post-processing phase is the calculation of lead signals relative to the Wilson-centre-terminal. This is only required if the electrode signals have been recorded relative to a non Wilson-centre-terminal (ie. an ankle or hip lead - as mentioned in section 5.2.1).

5.2.4 Data Display

The most common way to display BSM data is in the form of a series of isopotential maps (Figure 5-1). Such maps are usually displayed on a standard map grid of 20 by 12, or 32 by 12, and therefore the recorded BSM data usually requires spatial translation.

Two common issues arise when translating post-processed BSM data into this form. Firstly as a result of bad lead contact a number of lead signals may be unusable and therefore must be approximated based on the signal observed on surrounding leads. Secondly if electrodes are administered using electrode belts or an electrode jacket then the relative location of electrodes will be determined by the size and shape of the patient's torso.

Therefore spatial translation is generally achieved by performing a spline interpolation in both the X and Y-axis of the map grid, thus translating the electrode data from the actual electrode grid onto the standard map grid. This is performed for each sampling instance, thus building up a set of time series maps for the entire heart cycle.

It is also important in this process of standardisation to consider the issues of temporal normalisation. Firstly, it is commonplace to start the map sequence at the onset of the QRS complex, which is generally selected by the operator of the BSM system. Secondly, there are a number of researchers who believe that it is also important to time-normalise BSM data, as the duration of a heart cycle will vary from patient to patient.

5.3 Data Analysis

A number of techniques are currently being used for analysing and classifying electrocardiographic maps. This section will provide the reader with an overview of such techniques.

The aim with any analysis technique is to assign a diagnosis given the BSM data. As will be shown there are a number of qualitative and quantitative techniques that can be used to perform such analysis. In essence though, all these techniques rely on the previous experience of the observer or the comparative use of know a sample population.

All BSM data referred to in the following section is assumed to be spatially and temporally normalised. The following algebraic terminology will be used when referring to any BSM data. Although any BSM map is a set of isopotential readings distributed over a two dimensional grid, a BSM map can be described algebraically as a single vector. For example, a BSM map with dimensions X by Y can be represented by a vector A with dimension M, where M=X*Y, which may be described as a set of M values:

$$A = [a_1, a_2, \ldots, a_M] \qquad (5\text{-}1)$$

Note that any BSM vector which makes no reference to time only describes a single map instance. A complete BSM can therefore be represented by the vector function A(t), which may be described by a set of M functions, where t=0 refers to agreed starting point, which in general is the onset of the QRS complex.

$$A(t) = [a_1(t), a_2(t), \ldots, a_M(t)] \qquad (5\text{-}2)$$

This terminology will be used in the following sections when describing the mathematical manipulation of BSM data.

5.3.1 Visual Inspection

One of the simplest techniques used to analyse body surface maps is visual inspection. A skilled physician would usually perform such an assessment. This approach typically involves observing the position and magnitude of positive and negative extremes and how such features change over a series of maps. The rational for this approach is similar to the pattern recognition approach used by physicians in standard 12-lead electrocardiography. Although a somewhat qualitative approach which focuses on high level potential patterns, studies have shown that this technique is highly accurate (Pham-Huy et al. 1981).

There are two ways BSM maps can be presented to the observer for visual inspection. One option is to present a single map which the observer can move forwards and backwards in time. Thus the user can only view one map time instance at a time, and can slide the observation instance backwards and forwards through all the map readings. The advantage of this technique is that it provides the observer with a complete presentation of all the BSM maps. Alternatively the BSM data can be presented to the observer as a set of averaged maps which attempt to capture the most significant features of the QRST signature. For example a set of six maps may be constructed as described in Table 5-1.

Average Map	*Temporal Aspect*	*Calculation*
Map 1	0 to 19ms	$m_i = \sum_{t=0}^{19} a_i(t)/20$
Map 2	20 to 39ms	$m_i = \sum_{t=20}^{39} a_i(t)/20$
Map 3	40 to 59ms	$m_i = \sum_{t=40}^{59} a_i(t)/20$
Map 4	60 to 79ms	$m_i = \sum_{t=60}^{79} a_i(t)/20$
Map 5	130 to 149ms	$m_i = \sum_{t=130}^{149} a_i(t)/20$
Map 6	190 to 239ms	$m_i = \sum_{t=190}^{239} a_i(t)/50$

Table 5-1: Description of summary BSM maps

Maps 1 to 4 are selected to capture the QRS complex, map 5 will capture the ST segment, and map 6 will capture the T wave. An example of a set of such maps is provided in Figure 5-5.

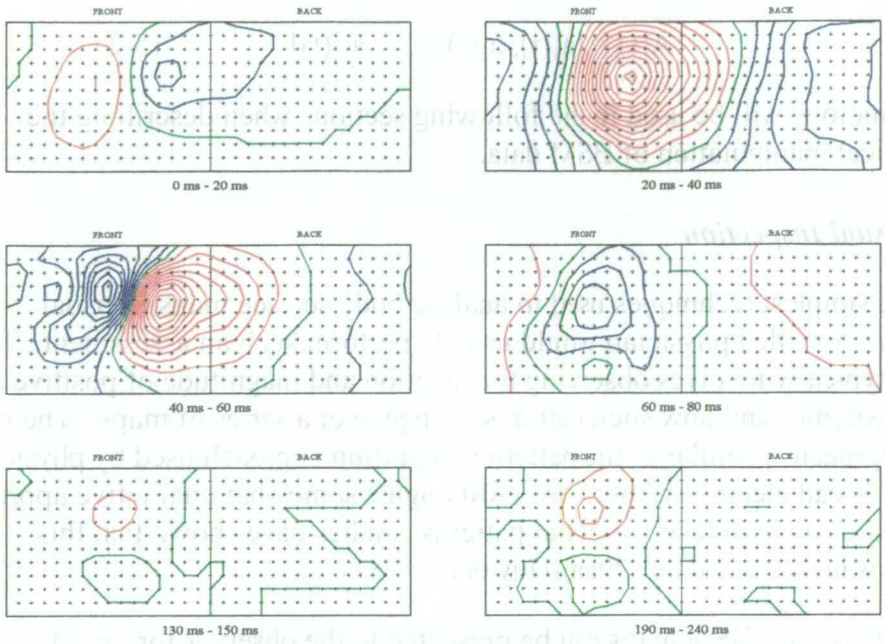


Figure 5-5: Example of a BSM map summary
(contour intervals – 10μV, red indicates positive potential,
blue indicates negative potential, and green indicates zero potential).

5.3.2 Difference Maps

Difference maps are specifically used for evaluating changes in heart behaviour after some form of operation or intervention (ie. angioplasty, bypass surgery, heart transplant, etc.). This technique therefore requires that two BSM readings be taken; one before the operation or intervention - B(t), and one after the event - A(t). Given this map information for a particular time instance in the heart cycle, the difference map is calculated by subtracting the map before intervention from the map after intervention for that time instance. Map subtraction is performed by subtracting the site potentials site-by-site, which may be described as follows:

$$d_i(t) = a_i(t) - b_i(t) \text{ for all } i, \text{where } 1 \le i \le M \quad (5\text{-}3)$$

Having constructed the difference maps, assessment is generally performed by visual inspection, although there is no practical reason why the statistical techniques described in the later sections of this chapter could not be applied to such maps.

As mentioned, this technique is generally used for assessing the impact of some form of surgical intervention and has been found to be useful for monitoring patient recovery. Difference mapping has also been used to study the changes occurring after acute myocardial infarction (Montague 1983, Montague 1984, Montague 1986). However this approach is not particularly practical in a clinical diagnostic setting as

body surface maps prior to an event are generally not available unless there has been some reason for carrying out mapping in the past.

5.3.3 Departure Maps

Departure maps, in contrast to difference maps, display the difference between a particular body surface map and a control group. This technique requires a control population consisting of either a set of normal body surface maps, or a set of abnormal body surface maps representing a particular condition.

A departure map is calculated as follows. Given a control population consisting of K body surface maps [$P_1(t)$,$P_2(t)$, . . . , $P_K(t)$], the average and standard deviation of each map point at each time instance is calculated:

$$\bar{x}_i(t) = \frac{\sum_{e=1}^{K} p_{ei}(t)}{K} \quad \text{for all } i, \text{ where } 1 \le i \le M \qquad (5\text{-}4)$$

$$s_i(t) = \sqrt{\frac{\sum_{e=1}^{K} \left(p_{ei}(t) - \bar{x}_i(t)\right)^2}{K^2}} \quad \text{for all } i, \text{ where } 1 \le i \le M \qquad (5\text{-}5)$$

The depature map for a given body surface map A(t) can then be calculated as follows:

$$d_i(t) = \frac{a_i(t) - \bar{x}_i(t)}{s_i(t)} \quad \text{for all } i, \text{ where } 1 \le i \le M \qquad (5\text{-}6)$$

Having constructed the departure maps, assessment is generally performed by visual inspection, although again there is no practical reason why the statistical techniques described in the later sections of this chapter could be applied to such maps.

5.3.4 Statistical Comparison

The analysis techniques described so far in this chapter; visual inspection, difference mapping, and departure mapping, are all qualitative analysis techniques. A number of statistical approaches have been used to provide a more quantitative approach to the analysis of body surface maps. These techniques aim to provide a measure of similarity between a map being analysed and a control group. These techniques can be used when comparing two maps before and after an event (ie difference mapping), or can be used to compare a particular BSM with a control group (ie departure mapping).

There are three statistical measures used in the literature for comparing body surface maps; *percent error*, *root-mean-squared error*, and *correlation coefficient*. All measures indicate similarity when close to zero. The further these measures deviate from zero the greater the difference between the two maps in question.

The *percentage error* (PE) is the sum square difference between two maps expressed as the percent of the sum square potential of one map:

$$PE_{AB} = \frac{\sum_{i=1}^{M}(a_i - b_i)^2}{\sum_{i=1}^{M} b_i^2} \qquad (5\text{-}7)$$

Lead error or *root-mean-squared error* (RMSE) in contrast to *percentage error* calculates the approximate error at any arbitrary individual surface site.

$$RMSE_{AB} = \sqrt{\frac{\sum_{i=1}^{M}(a_i - b_i)^2}{M}} \qquad (5\text{-}8)$$

The third measure calculates the correlation between two map patterns. The *correlation coefficient* (R) is calculated as

$$R_{AB} = \frac{\sum_{i=1}^{M}(a_i b_i)}{\sqrt{\sum_{i=1}^{M}(a_i^2 b_i^2)}} \qquad (5\text{-}9)$$

The correlation coefficient is somewhat different from the other statistical measures. PE and RMSE are specifically concerned with differences in map potentials, whereas the correlation coefficient provides a measure of similarity independent of map potentials and is more focused on the trends of isopotential features. The debate as to which is the better statistical measure is still open, but in recent years it would appear that the correlation coefficient (Horan et al., 1968) measure is more widely used.

5.3.5 Inverse Mapping

An alternative approach to BSM analysis is inverse mapping. The general aim of this technique is to translate body surface potentials, based on some mathematical model of the electrical behaviour of the torso, directly into voltage potentials on the surface of the heart (epicardial potentials). Having calculated the epicardial potentials, then the process of diagnosis is somewhat simplified, as the electrical behaviour of the heart can be observed directly.

Although this would seem an ideal solution to the problem of electrocardigraphic analysis there are many problems with the technique. The primary problem is that there are an unlimited number of cardiac sources that produce identical surface potential distributions (ie the mapping is not one-to-one).

Although methods to achieve this goal have been developed (Horan et al 1979, Barr et al 1978, and Colli-Franzone et al 1985) significant problems remain. Calculated epicardial values tend to be sensitive to small errors in the surface values (Colli-

Franzone et al 1985) and the technique tends to require detailed geometric measurement of the torso to improve performance (Choi et al 1981).

5.3.6 Data Reduction Techniques

One of the major problems when analysing BSMs is the volume of data associated with BSM recordings. If for example a BSM is constructed from 150 electrodes sampled at a rate of 1000Hz for 1 second, then the complete BSM recording would consist of 150 000 samples. One approach to reduce this is to use a lead subset of with fewer electrodes, as mentioned in section 5.2.1. Although lead subset approaches do not technically reduce the information content, they are more prone to error, as lead failure is more critical.

An alternative approach is to reduce the spatial and temporal redundancy using a numerical post-processing technique, thus taking advantage of the lead redundancy at the time of recording, but reducing the data size during post-processing. The primary aim when applying data reduction techniques is to ensure that the loss of information content is minimised and as such care needs to be taken when selecting a reduction technique.

One such reduction technique that has been clinically applied is the Karhunen-Loeve (K-L) expansion (Lux, Evans et al 1981, Evans, Lux et al 1981). Lux and Evans applied this technique in two phases, firstly removing the spatial redundancy in individual maps, and then removing the temporal redundancy.

The K-L expansion provides a technique for determining an optimised set of orthonormal basis vectors appropriate for the signals being represented. Since this technique makes no assumptions about the underlying process associated with the signal being represented, then it is ideal for removing redundancy in signals associated with apparent random processes.

Lux and Evans applied this expansion by firstly representing map instances as vectors with M dimensions (where M leads are associated with the map instance), for example,

$$A = [a_1, a_2, \ldots, a_M] \qquad (5\text{-}10)$$

then proposed that these maps can be represented by a linear sum of basis vectors,

$$A = \sum_{i=1}^{M} e_i \beta_i \qquad (5\text{-}11)$$

where $\{ \beta \}$ is a set of orthonormal basis vectors which span the M dimensional vector space, and the coefficient set {e} is unique for each A, and is defined by

$$e_i = A \cdot \beta_i \quad i = 1, 2, \ldots, M \qquad (5\text{-}12)$$

In practice, any orthonormal vector set of M vectors could be used for this purpose, and not all M basis vectors may be required for an acceptably accurate representation of A, particularly when dependency exists between dimension (ie spatial proximity). Therefore Lux and Evans proposed that,

$$A_N = \sum_{i=1}^{N} e_i \beta_i \quad N<M \qquad (5\text{-}13)$$

is an appropriate approximation of A with an acceptable degree of error,

$$E_N = |A_N - A| \qquad (5\text{-}14)$$

Using the Karhounen-Loeve (K-L) expansion Lux and Evans derived a set eigenvectors from a covariance matrix of map frames in 221 patients. On examination of reduced eigenvector sets Lux and Evans concluded that for a comprehensive lead configuration consisting of 192 leads, a single map instance could be adequately represented by using only the first 12 eignvectors. This in effect meant that a 16x12 BSM map (ie 192 leads) could be adequately represented by 12 eignvalues (rms error < 12uv).

Lux and Evans went on to apply this technique to the temporal features of the same BSM data with similar results. In this case, BSMs where reduced spatially using the 12 spatial eignvectors derived in the previous study, thus generating a set of 12 temporal signals of 300 samples in length (sampling 2msec intervals). Using the same population of 221 patients a 300x300 covariance matrix was constructed and 300 temporal eignvectors were calculated. After analysis Lux and Evans concluded that a single temporal signal could be adequately represented by 18 eigenvectors (rms error < 50uv). Thus each of the 12 temporal signals could be represented by 18 eigenvalues. Consequently a complete BSM recording (16x12 by 300 samples) could be represented by 216 coefficients, providing an impressive compression ratio of approximately 267:1.

The accuracy of maps reconstructed from the coefficients was found to be high. In 34 test cases (these cases were not included in the population used to compute the basis functions), average mean square errors were 6.4% for the QRS interval and 7.5% for the ST-T interval. A certain amount of smoothing and filtering does, however result.

It has been shown in a number of studies that the coefficients calculated using this technique can be used to classify patients into subgroups of diagnostic relevance. Studies to date have demonstrated that this approach can accurately detect patients with myocardial infarctions but not with normal 12 -lead electrocardiograms (Kilpatrick and Walker 1987). Another study also suggests that coronary artery disease can also be detected using this technique (Green et al 1987).

6. Experimental Design

This chapter will describe the data acquisition system used and the data sets constructed, followed by a detailed description of the classification problems. The chapter will finish by presenting the classification techniques applied as well as discussing the pre-processing techniques.

6.1 Data Acquisition

The data used for this study was provided by Dr David Kilpatrick from the Department of Medicine at the University of Tasmania. The recording of the electrocardiographic data was performed using an electrode jacket interfaced to an Apple Macintosh computer. This system was housed on a trolley to allow ease of transportation to the bedside. A description of the hardware and software is given below:

Electrode Jacket: The electrode jacket consists of a fixed array of 50 electrodes. The positions of these electrodes are shown in Figure 6-1. A further two electrodes: a reference electrode (electrode 0) and a neck electrode (electrode 51), are attached to the jacket by short lengths of ECG lead. When fitting the jacket on a patient, it is wrapped around the patient's torso with the second column of electrodes positioned over the mid-sternum. The amount of jacket overlap and the exact location of electrodes varies depending on the size of the patient's torso which is taken into account during the post-processing of the BSM data. To simplify electrode application, active electrodes are used to eliminate the need for conductive jelly. This does increase the chance of bad-lead contact, but is an accepted trade-off when recording body surface maps (studies have show that up to a 10% lead loss can occur without significant loss of information content – Barr el al. 1971).

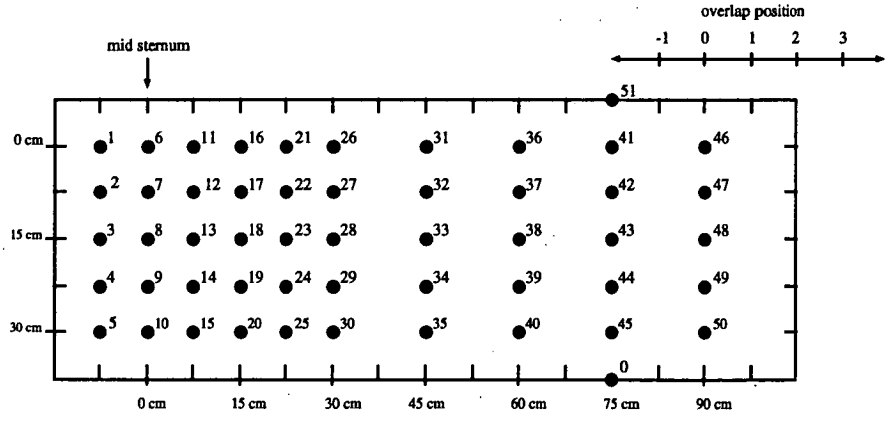


Figure 6-1: Electrode Jacket Configuration

Interfacing Hardware : The jacket is connected to the computer via interfacing hardware which provides the necessary lead amplification. The signals are then multiplexed into a 16-bit analog-to-digital converter. The lead amplifiers are all fully programmable, providing software control of individual amplifier gains and DC-offsets, thus simplifying calibration. Sampling timing is managed by the software using a simple hardware polling technique.

Sampling Software: The software for the BSM jacket allows the operator to record, view and save the BSM data. The recording software is designed to sample the ECG signals (all 51 leads simultaneously) and allows the operator to record the patient details and jacket overlap. Any bad leads are marked (ie any leads that appear excessively noisy or exhibit no signal) and the start of the heart beat signal is marked (ie the onset of the QRS signature). This procedure is described in more detail in the following section.

Recording Procedure: To record an electrocardiographic body surface map, the electrode jacket is fitted to the patient and the amount of overlap noted. The neck electrode is placed on the right side of the neck and the reference electrode is placed on the right anterior superior iliac spine (the right hip-bone). Once the patient is fitted the operator initiates the sampling software and enters the patient's details, jacket overlap, and sampling frequency required (in most cases the sampling frequency used is 1000Hz). By means of the interfacing hardware, the sampling software samples all 51 leads simultaneously for two or three heart beats and then displays a summarised signal for the operator to inspect. The operator selects an appropriate heart beat signature from the two or three that have been recorded at which point the software displays all 51 lead signals for the selected duration. Any leads which appear to be excessively noisy or exhibit no signal are marked as 'bad leads' by the operator (if there are too many bad leads the operator may choose to discard the recording and attempt to record another sample). Once all bad leads have been marked the software displays a single ECG lead summary allowing the operator to mark the onset of the QRS signature. At the conclusion of this procedure the sampling software calculates and displays the BSM as a number of iso-potential maps (normally averaged over 20ms time segments). If satisfied with the quality of the BSM recording the operator saves this data for future reference or processing.

BSM File Format : The BSM data file saved by the sampling software consists of a header record followed by a 51 lead records for each ECG lead (Figure 6-2). The header record contains information such as the sampling rate, jacket overlap, and QRS onset. The lead records are fixed length arrays of 1024 by 16-bit elements. The first element of a lead record indicates if the lead is bad (ie noisy or no signal) and the following 1023 elements constitute the actual sampled data for that particular lead (note - the sample is always padded out to 1023 elements). Although this data is in an extremely raw format this ensures that sufficient data is stored in these files to allow a wide range of post-processing to be performed on the data.

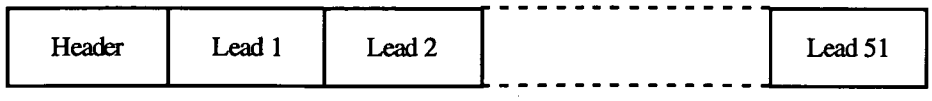


Figure 6-2: BSM File Format

6.2 Data Preparation

All data used for this study is recorded in the file format described in the previous section. To allow meaningful comparison of BSM data, all recordings are standardised using post-processing routines provided by the Department of Medicine. These routines perform the following tasks:

- remove DC-drift from all lead signals.
- replace bad lead signals with an interpolation of surrounding leads.
- construct a reference electrode by simulating a Wilson centre terminal.
- map data onto a standard 32x12 grid for each time instance.

Although the amplifiers in the BSM system are calibrated with appropriate offsets, it is not possible to avoid small changes in the baseline lead outputs. It is therefore necessary to remove this when standardising the data so that realistic comparisons can be made between lead signal levels. Similarly, bad lead signals are unavoidable, particularly with this system as no conductive jelly is used on jacket electrodes.

With regard to standardising the BSM data it is essential that the data be referenced with respect to a Wilson centre terminal. This is simulated by identifying the electrodes that are located closest to the standard VR, VL, VF leads and then calculating $V_{Wilson} = VR+VL+VF$. This reference voltage is calculated for each sample instance and subtracted from all lead values for that instance.

The final issue that needs to be resolved is the electrode placement. The electrode jacket consists of a fixed array of electrodes but the positioning of these electrodes varies depending on the circumference of the patient's chest. To account for this, the jacket overlap is used to calculate the true location of the electrodes. By means of a spline interpolation method the lead data for each time instance is mapped onto a standard 32x12 grid evenly distributed around the thoracic surface. The 32x12 grid was historically chosen to provide compatibility with BSM formats in previous studies (Evans et al., 1981; Lux et al., 1981; Green et al., 1987). This makes it possible to compare BSM maps from different patients.

These post-processing routines are not only capable of producing 32x12 maps for every instance (ie every 1ms) of the heart signature, they also can extract sequences of maps averaged over longer time periods. These proved most useful in producing the various data sets required for this study.

6.3 Datasets

The data used for this study was divided into four classes. Patients were classified as having either an inferior myocardial infarction (IMI), an anterior myocardial infarction (AMI), coronary artery disease (CAD), or normal heart function.

The IMI and AMI data was gathered from patients admitted to the coronary care unit with acute myocardial infarction. The diagnosis was based on a history of myocardial ischemic pain, changes in the standard 12-lead electrocardiogram, and enzyme levels. The diagnosis was confirmed when possible by echocardiography, arteriography, and nuclear imaging. All patients with bundle branch block or previous acute myocardial infarction were excluded.

The CAD data was gathered from patients undergoing cardiac catheterisation who had normal or near normal 12 lead electrocardiograms. Electrocardiograms were analysed independently by two experienced cardiologists. Patients were included if the cardiologists agreed that all ECGs were normal. Patients were excluded if the

coronary angiography was performed for heart disease unrelated to ischemic heart disease or if the patient had a past history of myocardial infarction. The normal patient data was gathered from healthy first year medical students. Students were screened by 12-lead electrocardiogram to verify normality.

These four patient groups were used to construct training and testing sets. The breakdown of patient numbers in these data sets is given in Table 6-1.

Classes	Training Set	Testing Sets
anterior	268	107
inferior	289	118
normal	96	56
CAD	103	100

Table 6-1: Class sizes

Further sub-classes were defined within the IMI and AMI groups. Each BSM was assigned to one of the four sub-classes depending on the time that had elapsed since the onset of infarction until when the BSM was recorded (ie 6, 24, 48 hours and 2 years). This information is important, as the electrocardiographic data will change as a patient's heart begins to recover from an infarction. The sub-class breakdown of the IMI and AMI data is described in Table 6-2 and Table 6-3.

Sub-Classes	Elapsed Time	Training Set	Testing Sets
anterior.6	6 hours	96	29
anterior.24	24 hours	102	29
anterior.48	48 hours	44	29
anterior.fu	2 years	26	20
Total		268	107

Table 6-2: Anterior myocardial infarction (AMI) class breakdown

Sub-Classes	Elapsed Time	Training Sets	Testing Sets
inferior.6	6 hours	93	30
inferior.24	24 hours	82	30
inferior.48	48 hours	52	30
inferior.fu	2 years	62	28
Total		289	118

Table 6-3: Inferior myocardial infarction (IMI) class breakdown

It is important to note that these sub-classes are not true classes, as some patients will recover quicker than others and the extent of infarction may vary. As a consequence there is significant overlap in these classes, and patients can be assigned to different sub-classes and yet exhibit the same electrocardiographic behaviour. Similar to this, some of the follow-up patients (inferior.fu, aterior.fu) may exhibit normal electrocardiograms. To use these during training may confuse classification training. These issues will be discussed in more detail in the following chapters.

6.4 Classification Problems

The primary classification problem presented by the data sets is differentiating between the four major classes; anterior, inferior, normal and CAD. As such the primary focus of the following chapters will be to construct a classifier capable of categorising any body surface map into one of these four categories.

This is by no means a new problem and in fact has been the focus of much research (see chapters 4 and 5). It is clear from chapter 4 that myocardial damage is detectable using electrocardiographic techniques and as such the separation of normal patients from those with an anterior or inferior myocardial infarction should be a relatively simple classification problem. However there will be some overlap between theses groups particularly with respect to those patients in the follow-up categories (inferior.fu, anterior.fu) as many of these patients may well exhibit normal electrocardiograms.

The classification problem which will present much more of a challenge is the separation of patients with normal heart function from those with coronary artery disease. The reason that this is such a difficult classification problem is because patients with coronary artery disease in most cases present normal electrocardiograms (as discussed in chapter 4). From a physiological point of view a patient with coronary artery disease does not exhibit any physiological changes in the myocardial tissue and therefore the myocardium should generate electrical activity which is no different from that of a normal patient (Lipman et al., 1984).

The detection of coronary artery disease using BSM data has primarily focussed on analysing the cardiographic data as a patient exercises (Fox et al., 1979; Wada et al., 1981; Simoons and Block 1981; Yanowitz 1982; Kutota et al.,1984; Ikeda et al., 1986). The primary reason for exercising the patient is to increase the myocardial oxygen demands to a point where ischemia begins to occur (since the restricted arteries cannot supply sufficient oxygen). This technique has been quite successful although in many cases it has been found that collateral supply (supply form other arteries) can compensate for the increased oxygen demand (Hill et al., 1983; Murvis 1985; Murvis et al., 1986).

The patients with coronary artery disease in this study are recorded at rest, and as such are somewhat more difficult to classify. Green and Lux (1987) presented experimental results that suggest body surface mapping could be used to detect coronary artery disease in such cases. Attempts have been made previously to reproduce these results using this data set (Bell 1993), but without success. It should be highlighted from the outset that the results obtained by Green and Lux (1987) were only in relation to a training set and that no testing set was used to verify the results (Bell 1993).

With these issues in mind, four classification problems were considered. Firstly, the complete problem was considered, that is attempting to separate all four classes; anterior, inferior, normal, and coronary artery disease (Problem 1). Secondly, due to the conflicting nature of different classes, three further problems were considered; the separation of anterior, inferior and normal patients (Problem 2), the separation of

anterior, inferior and normal patients but not using the follow-up classes during training (Problem 3), and the separation of CAD and normal patients (Problem 4). These four separate classification problems were considered in an attempt to identify the nature of these classes and what impact the similarity of classes had on classification performance. These four problem sets are summarised in Table 6-5.

Classification Problem	Description	Training Set
Problem 1	Separation of *anterior*, *inferior*, *normal* and *CAD* classes	anterior (anterior.6, anterior.24, anterior.48, anterior.fu), inferior (inferior.6, inferior.24, inferior.48, inferior.fu), normal, and CAD
Problem 2	Separation of *anterior*, *inferior*, and *normal* classes	anterior (anterior.6, anterior.24, anterior.48, anterior.fu), inferior (inferior.6, inferior.24, inferior.48, inferior.fu), and normal
Problem 3	Separation of *anterior*, *inferior*, and *normal* classes (not using follow-up cases during training or classification)	anterior (anterior.6, anterior.24, anterior.48), inferior (inferior.6, inferior.24, inferior.48), normal, and CAD
Problem 4	Separation of *normal* and *CAD* classes	normal, and CAD

Table 6-5: Summary of experiments

In the following chapters a range of neural network and traditional classification techniques will be applied to each of these classification problems.

6.5 Feature Extraction Techniques

The primary problem in preparing the data sets for classification was the volume of data associated with individual BSM recordings. Having applied the post-processing routines (as described in section 6.2) a single BSM recording consists of 1000 map instances each with a dimension of 32x12, resulting in a total of 384000 individual attributes.

This number of attributes was somewhat impractical for classifier training. All the classification techniques considered (see section 6.5) although capable of theoretically managing this number of attributes would take an impractical amount of time to train. Therefore some form of data reduction transform was required to reduce the attribute set to a more manageable size, hopefully without removing relevant information or the discriminatory potential of the BSM data.

Two data reduction techniques were considered in this study. The first was the Karhunen-Loeve transform (KLT), whilst the second was an adaptation of the data subset usually used in the manual analysis of BSM data.

6.5.1 Karhunen-Loeve Transform

The Karhunen-Loeve Transform (KLT) technique had been applied to the data sets in a previous study (Walker et al., 1987) based on the work by Lux and Evans (Lux et al. 1981). This eigenvector set consisted of 6 spatial eigenvectors { β } and 12 temporal eigenvectors { λ }. Spatial reduction was achieved by reducing each map instance, A, to a set of 6 spatial eigenvalues using the 6 spatial eigenvectors{ β };

$$e_i = A \cdot \beta_i \quad \text{for } 1 \le i \le 6 \qquad (6\text{-}1)$$

thus constructing a set of spatial eigenvalue sequences from t=0 to t=999;

$$e_i(t) = A(t) \cdot \beta_i \quad \text{for } 1 \le i \le 6 \qquad (6\text{-}2)$$

Temporal reduction was achieved by reducing each spatial eigenvalue sequence to a set of 12 coefficients using the 12 temporal eigenvectors { λ };

$$v_{ij} = e_i(t) \cdot \lambda_j \quad \text{for } 1 \le i \le 6 \text{ and } 1 \le j \le 12 \qquad (6\text{-}3)$$

This procedure generated a set of 72 coefficients for each BSM recording. Refer to Figure 6-3 for a diagrammatic description of the technique.

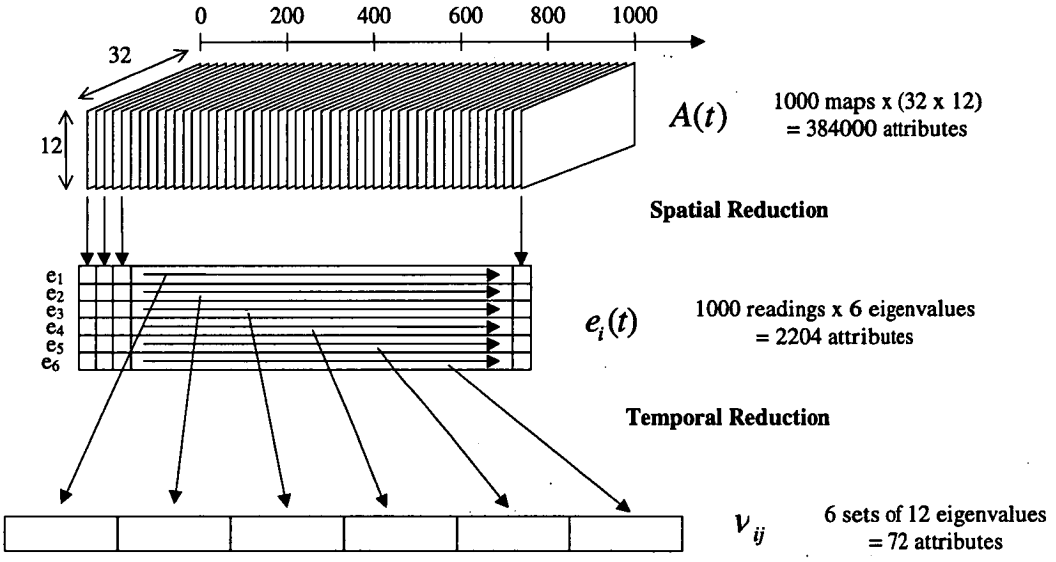


Figure 6-3: Spatial and temporal eigenvector reduction technique.

This data reduction technique has been evaluated by Pigot (Pigot et al. 1987). Maps were reproduced from coefficients and found to have a 0.95-0.96 correlation to the original maps, thus suggesting that information loss was low.

6.5.2 Logical Transform

The second data reduction technique was proposed by the author. Having considered the data analysis techniques used in the literature (see chapter 5), it was clear that many researchers, when classifying BSMs by visual inspection, were using an extremely summarised form of the BSM data (Murvis 1981).

The effectiveness of such qualitative approaches suggested that the information content of averaged maps was diagnostically significant, and therefore should be considered as a data reduction technique.

As described in chapter 5 one technique used for visual inspection is to create a set of six averaged maps; four summarising the QRS complex, one summarising the ST segment, and one summarising the T wave. This results in a set of six 32x12 maps, which represents 2204 attributes (see Table 6-4).

Average Map	*Temporal Aspect*	*Calculation*
Map 1	0 to 19ms	$m_i = \sum_{t=0}^{19} a_i(t)/20$
Map 2	20 to 39ms	$m_i = \sum_{t=20}^{39} a_i(t)/20$
Map 3	40 to 59ms	$m_i = \sum_{t=40}^{59} a_i(t)/20$
Map 4	60 to 79ms	$m_i = \sum_{t=60}^{79} a_i(t)/20$
Map 5	130 to 149ms	$m_i = \sum_{t=130}^{149} a_i(t)/20$
Map 6	190 to 239ms	$m_i = \sum_{t=190}^{239} a_i(t)/50$

Table 6-4: Description of summary BSM maps

In effect this map averaging is a temporal reduction, and although effective, the set of 2204 attributes was found to be too large for use during the training process. Therefore for the purposes of reducing this attribute set to a more manageable size each averaged 32x12 map was spatially reduced to an 8x4 grid (ie splitting the 32x12 maps into 4x3 blocks and averaging the content to produce one attribute per block). The result is a set of six 8x4 maps, representing 192 attributes (see Figure 6-4).

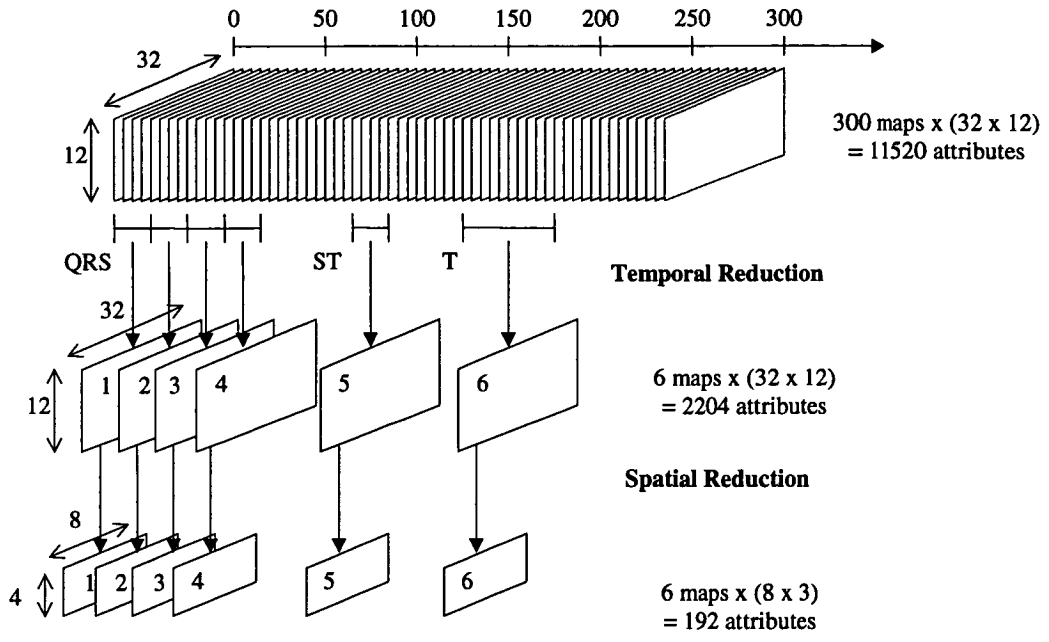


Figure 6-4: Logical reduction technique.

To explore this data reduction technique further two variations on this approach were also considered to determine the diagnostic significance of specific subsets of this attribute set. The first variation utilised only the first four maps thus considering the QRS features. The second variation utilised the last two maps thus only considering the ST segment and T wave features.

The primary reason for considering these attribute subsets is to ascertain which approach is more appropriate for classifying myocardial infarction and coronary artery disease.

6.6 Classification Techniques

A number of neural network classifiers and alternative classification techniques where considered in this study. These are described in the following three sections.

6.6.1 Neural Network Techniques

Six neural networks classification techniques were considered (refer to chapter 3 for more details concerning each of these approaches):

- multilayered perceptrons (MLPs) trained using the back-propagation training algorithm,
- multilayered perceptrons (MLPs) training using the quickprop training agorithm,
- cascade correlation networks,
- committees of MLPs trained using back-propagation training,
- committees of MLPs trained using quickprop training, and
- committees of cascade correlation networks.

Each of these techniques was tested in conjunction with each of the feature extraction techniques described in section 6.5.

6.6.2 Alternative Classification Techniques

Four alternative classification techniques where also considered (refer to chapter 2 concerning each of these approaches):

- linear discriminant analysis,
- k-nearest neighbour, and
- two inductive learning techniques: C4.5 and MML.

Each of these techniques was tested in conjunction with each of the feature extraction techniques described in section 6.5.

6.6.3 Traditional BSM Classification Techniques

As highlighted in chapter 5 there are many traditional techniques used for analysing and classifying BSM data. Having considered the literature the most appropriate traditional approach for this particular classification problem was departure mapping. This technique utilised the training set as control groups. The control groups were used to classify patients according to distance from class means (departure mapping – see section 5.3.3) or by using a standard k-nearest neighbour classification technique. It is worth noting that this technique uses the complete data set and is different to the k-nearest neighbour approach mentioned in the previous section. Two distance measures were used in conjunction with these approaches: the root-mean-squared error (RMSE) and correlation coefficient (R) (see section 5.3.4).

Therefore, four traditional classification techniques were considered:

- k-nearest neighbour using the root-mean-squared error (RMSE) as distance measure,
- k-nearest neighbour using the correlation coefficient (R) as distance measure,
- nearest class mean using the root-mean-squared error (RMSE) as distance measure, and
- nearest class mean using the correlation coefficient (R) as distance measure.

These approaches did not use any feature extraction techniques. Therefore all distance measures were calculated using the complete BSM attribute set (384000 attributes – see sections 6.2 and 6.5.1). This is possible as no training is required.

6.7 Experiment Referencing

For the purposes of clarity and consistency the following referencing will be used when referring to any experiment. An experiment is described according to three possible parameters; the classification problem being attempted, the data reduction technique being used, and the type of classifier being applied. Therefore, an experiment will be referenced according to the following descriptor:

<feature extraction technique><classification problem>.<classification technique>

The descriptors used for the feature extraction techniques are outlined in Table 6-6.

Feature extraction technique	Description
E	KLT eigenvector data reduction technique (see section 6.5.1) 72 attributes
L	Logical data reduction technique (see section 6.5.2) 192 attributes
qrs	First four maps of the logical reduction technique (see section 6.5.2) 128 attributes
st	Last two maps of the logical reduction technique (see section 6.5.2) 64 attributes
C	Complete BSM recording – no data reduction is performed (see section 6.6.3). Note this approach was only used in conjunction with the traditional BSM classification techniques 384000 attributes

Table 6-6: Data reduction techniques.

The descriptors used for the classification problem will be simply a numeric index (ie 1, 2, 3, 4) referring to the particular classification problem being considered (see section 6.4). The descriptors used for the classification techniques are outlined in Table 6-7.

Classifier	*Description*
bp	MLP trained using backpropagation
qp	MLP trained using quickprop
cas	cascade correlation network
cbp	committee of MLPs trained using backpropagation
cqp	committee of MLPs trained using quickprop
ccas	committee of cascor networks
knn	k-nearest neighbour
linreg	linear discriminant function
C4.5	inductive learning technique C4.5
MML	inductive learning technique MML

Table 6-7: Experimental referencing for classification techniques.

Further to this, the four traditional BSM classification techniques will be referenced as follows (Table 6-8):

Classifier	*Description*
knn-rmse	k-nearest neighbour using the root-mean-squared error (RMSE) as distance measure
knn-cc	k-nearest neighbour using the correlation coefficient (R) as distance measure
ncm-rmse	nearest class mean using the root-mean-squared error (RMSE) as distance measure
ncm-cc	nearest class mean using the correlation coefficient (R) as distance measure

Table 6-7: Experimental referencing for traditional BSM classification techniques.

Some example experiment references are given below:

E1.bp	MLP trained using back-propagation applied to problem 1 (separation of all four classes; anterior, inferior, normal and CAD), using the eigenvector data reduction technique.
L2.ccas	committee of cascor networks applied to problem 2 (separation of the classes anterior, inferior and normal), using the logical data reduction technique.
qrs4.linreg	linear discriminant function applied to problem 4 (separation of the classes normal and CAD) using the first four maps of the logical data reduction technique.
C2.ncm-cc	nearest class mean using the correlation coefficient (R) as distance measure applied to classification problem 3.

6.8 Experiments

Each of the classification techniques described in section 6.6 were applied to the four classification problems outlined in section 6.4. These experiments were conducted in two phases.

The first set of experiments considered the use of multilayered perceptrons trained using the standard back-propagation algorithm utilising the KLT feature extraction technique. The results of these initial experiments are presented and discussed in chapter 7. These experiments serve as a preliminary investigation into the nature of the classification problems and highlight some of the classification challenges.

The second set of experiments provided a comparative assessment of each classification technique applied to the four classification problems. The results of these experiments are presented and discussed in chapter 8. These experiments provide some insights into how alternative neural network classifiers perform and provide comparisons with alternative and traditional classification techniques. The aim of this second set of experiments is to identify not only the most appropriate classifiers for BSM classification, but also to identify the most effective feature extraction technique to use in conjunction with these classifiers. This resulted in forty-four classification experiments being applied to each of the four classification problems;

- 24 neural network experiments
  (6 neural network techniques – *bp, qp, cas, cbp, cqp, ccas*
  using 4 different feature extraction techniques – *E, L, qrs, st*),
- 16 alternative classification experiments
  (4 alternative classification techniques – *knn, linreg, C4.5, MML*
  using 4 different feature extraction techniques – *E, L, qrs, st*), and
- 4 traditional BSM classification experiments
  (4 traditional BSM classification
  techniques – *knn-rmse, knn-cc, ncm-rmse, ncm-cc*
  using the complete BSM recordings - *C*).

7. Application of Multilayer Perceptrons to BSM Data Classification

This chapter presents the first set of classification experiments attempted in this study. Multilayer perceptrons (MLPs) trained using the traditional backpropagation training algorithm were applied to the four classification problems described in chapter 6. These initial experiments highlight the key challenges associated with classifying myocardial infarctions and coronary artery disease using electrocardiographic body surface mapping data and MLP classifiers.

The chapter is divided into six sections. The first section describes the method used to select the MLP architecture and training parameters. The four sections that follow provide detailed presentation and discussion of the results obtained for each of the classification problems. The final section discusses the key problems with classifying these data sets and provides some background to the following chapter and the reasons for considering a number of alternative feature extraction and classification techniques.

7.1 Initial Experiment (Problem 1)

The principle problem considered in the initial experiments was the complete four-class problem (problem 1). These experiments involved training MLPs to classify patients into one of four categories; anterior, inferior, normal or CAD, depending on the patients heart condition (see chapter 6 for a detailed description of the data sets and classes).

The KLT data reduction technique was used for the purposes of feature extraction, reducing each patient's BSM recording down to a set of 72 coefficients. The KLT data reduction technique was used initially, as previous studies (Walker et al., 1987; Lux, Evans et al., 1981; Evans, Lux et al.,1981) had shown it to be an effective technique, capable of reducing data volume and maintaining data content.

7.1.1 Initial Experiments

Having selected the four-class classification problem and feature extraction technique (KLT), a number of initial MLP training runs were conducted to determine the most appropriate MLP network architecture and training parameters.

7.1.1.1 Selecting Network Architecture

A number of experiments were conducted to determine the most appropriate network architecture. Clearly the feature extraction technique (KLT) defined the number of MLP inputs (ie 72 coefficients) and the number of classes associated with the classification problem defined the number of MLP outputs (ie 4 classes). What was unknown initially was the number of hidden nodes and hidden layers required.

To assist in this process a number of MLPs with a single hidden layer were constructed ranging from network architectures of 72:2:4 up to 72:120:4. Each of these networks was trained on the training set with the backpropagation algorithm

using a mean-squared-error error measure. All neurons utilised a sigmoid transfer function and a bias input. Network weights were initialised to random starting values in the range –1.0 to 1.0.

Network weights were updated after each pattern presentation using a learning rate of 0.05. The training period varied depending on the size of the network, but in general most networks completed training in less than 20000 epochs. Training was stopped when the change in the average training set error (mean-squared-error) was less than 0.1%.

After each MLP network had completed training, the test set was classified using the network. It was observed that the number of test set examples classified correctly increased progressively as the number of hidden nodes was increased. This reached a maximum at approximately 72 hidden nodes after which the addition of hidden nodes provided no statistical improvement in the classification performance. In fact, it was observed that large networks (>100 hidden nodes) tended to be slower to train and often performance degraded.

A number of networks with more than one hidden layer were also considered in the hope that this would improve the classification performance further. A range of architectures with two, three and four hidden layers were tested but were not found to provide any improvement.

The final network architecture selected was 72:72:4 using sigmoid transfer functions and bias inputs on both the hidden and output layers.

7.1.1.2 Training Parameters

Having selected the network architecture some further experimentation was conducted to determine the most appropriate training parameters. Since standard back-propagation was being used in these initial experiments, the only training parameters that required configuration were the network learning rate and the stopping criteria.

A number of learning rates were applied during the training phase ranging from 0.01 up to 0.5. Using a training rate of 0.01 a high variation in classification performance was observed. A number of networks were initialised with different starting weights, trained using a learning rate of 0.01 and then applied to the testing set. It was observed that the percentage of testing examples correctly classified varied dramatically (standard deviation of greater than ±8%). When the learning rate was increased it was noted that this variation in network performance decreased and stabilised at approximately 0.1. It was also observed that using a learning rate of anything above 0.25 resulted in an unstable training process and in many cases the training error tended to fluctuate dramatically and did not settle to a minimum.

7.1.1.3 Stopping Criteria

A number of criteria were considered for determining when to stop MLP training.

The initial stopping criteria considered involved monitoring the average training error at the end of each epoch. Training was stopped when the change in the error per epoch was below 0.1%. Although this training method worked well, it was noted that over-training or over-fitting was occurring. In the latter stages of the training process it was noted that although the training error continued to decrease the number of correctly classified training patterns did not increase. It was also noted that at this point in training the testing set error reached a minimum and began to increase again. This suggested that over-training was occurring.

An alternative stopping criteria using a patience technique (Fahlman 1991) was found to be more effective. The patience technique stopped training when a network's *patience measure* had not improved for a specified number of epochs (termed the *patience limit*).

Two *patience measures* were considered; the average training set error, and the percentage of correctly classified training patterns. It was found that the percentage of correctly classified training patterns was a more effective patience measure, stopping training before the observed over-fitting started to occur.

A *patience limit* of at least 1000 epochs was found to be appropriate for this problem. If the patience limit was set lower than 250 epochs network training was stopped too soon. Increasing the patience limit above 1000 epochs did not provide any improvement.

7.1.1.4 Pattern verses Batch Presentation

Both batch and pattern presentation techniques were considered. Pattern presentation was found to perform consistently better than the batch presentation technique.

7.1.1.5 Imbalance in training class sizes

Another issue that was found to influence training was the difference in training class sizes; anterior – 268, inferior – 289, normal – 96, and CAD – 103. It was observed that if the 756 patients were presented for each epoch the networks would be biased towards the anterior and inferior classes. It was found that presenting classes evenly avoided this issue and provided the networks with a 'balanced' training (see appendix B).

7.1.2 Results

Having experimented with a number of possible network architectures and training parameters, the final set of parameters used to construct and train the MLPs applied to problem 1 are outlined in Table 7-1. To assess the repeatability of experiments, 20 training runs were executed using networks with different random starting weights (assigned in the range of -1.0 to 1.0). Networks were trained to binary targets; for a particular target vector all elements were set to zero except for the output element associated with the assigned class for that pattern, which was set to one.

Parameter	Setting
network architecture	72:72:4
transfer function	sigmoid (hidden and output layer)
training technique	back-propagation
training limit	20000 epochs
training runs	20
learning rate	0.1
error measure	MSE
patience measure	percentage correct
patience limit	1000 epochs
presentation type	pattern presentation
data selection	weighted sequential

Table 7-1: Training parameters used for experiment E1.bp

The patience measure halted training runs, on average, at 3940 epochs (±1330 epochs). At the end of the training phase the average mean squared error of network outputs was 0.113 (±0.010) for the training set and 0.178 (±0.010) for the testing set. The resulting networks classified training set examples with an accuracy of 71.0% (±3.2%) and the testing set examples with an accuracy of 55.5% (±2.3%). The classification performance of these networks is summarised in Table 7-2.

%	Percentage Classified Correctly				
Data set	Anterior	Inferior	Normal	CAD	Total
Training	83.4±2.3	82.0±2.3	59.8±16.3	58.9±5.1	71.0±3.2
Testing	65.9±4.6	67.2±4.8	30.7±10.7	58.2±9.7	55.5±2.3

Table 7-2: Classification results - experiment E1.bp
(all figures indicate percentage of patients classified correctly)

The selection of the best training run on the basis of training performance was considered, but it was found not to provide statistically significant or consistent improvements in test set results and in some cases was actually less than the mean result. Therefore it was not considered appropriate to discuss these results in terms of the best network.

The classification performance for the anterior and inferior classes is significantly higher than for the normal and CAD classes for both training and testing results. In particular, the classification of normal patients is extremely poor, especially for the testing set [30.7% (±10.7%)]. The reason for this observation becomes clearer when the breakdown of the classification results is considered in more detail. The classification breakdown in Table 7-3 describes how the MLP networks classified each of the classes. For example, in relation to the training set, of those patients with anterior infarctions 83.4% (±2.3%) were classified correctly, and of the remaining incorrect classifications; 5.3% (±1.5%) were classified as inferior infarctions, 4.5% (±2.2%) were classified as normal, and 6.8% (±1.6%) were classified as having coronary artery disease. Thus all greyed cells in Table 7-3 indicate correct classifications and all other cells indicate incorrect classifications.

%	Training Set Classification Breakdown				Testing Set Classification Breakdown			
Class	anterior	inferior	normal	CAD	Anterior	inferior	normal	CAD
anterior	83.4±2.3	5.3±1.5	4.5±2.2	6.8±1.6	65.9±4.6	10.7±3.0	10.5±4.2	13.0±3.2
inferior	2.8±0.6	82.0±2.3	4.7±1.8	10.5±2.7	7.2±1.9	67.2±4.8	6.9±3.6	18.7±4.4
normal	10.9±3.9	8.2±2.1	59.8±16.3	21.1±13.2	19.7±5.5	7.8±4.7	30.7±10.7	41.8±12.6
CAD	9.4±1.7	17.3±2.6	14.5±5.2	58.9±5.1	4.8±3.9	14.3±4.8	22.7±9.0	58.2±9.7

Table 7-3: Classification breakdown - experiment E1.b
(greyed cells indicate correct classifications, all other cells are incorrect classifications)

It is clear from Table 7-3 that the MLP networks are capable of correctly classifying a significant proportion of the anterior and inferior infarctions. With respect to the training set more than 80% of anterior and inferior infarctions are consistently classified correctly. Although the networks do not achieve this classification performance when applied to the testing set, it is clear that the networks have managed to identify discriminating features in the data, correctly classifying more that 60% of testing set infarctions.

In contrast, the classification of the normal and CAD classes is somewhat poor. Overall the percentage of normal and CAD examples correctly classified is significantly less than that achieved in relation to anterior and inferior infarctions. It would appear that a significant number of normal patients are being incorrectly classified as having coronary artery disease (21.1%±13.2% in the training set, and 41.8%±12.6% in the testing set). Further to this it is observed that when applied to the testing set, more normal patients are incorrectly classified as having coronary artery disease (41.8%±12.6%) than those classified correctly (30.7%±10.7%). It is also observed that although the networks perform reasonably well when classifying CAD patients in the testing set (58.2%±9.7%) a significant number of examples are incorrectly classified as normal (22.7%±9.0%). These observations would suggest that the networks are having difficulty discriminating between normal patients and those with CAD.

%	Training Set				Testing Set			
Class	anterior	inferior	normal	CAD	anterior	inferior	Normal	CAD
anterior.6	72.0±5.0	9.8±2.9	7.0±3.5	11.1±3.6	64.7±6.8	10.3±5.5	12.2±6.2	12.8±5.6
anterior.12	86.7±3.5	4.3±1.6	5.6±3.0	3.4±2.2	72.1±7.5	4.1±3.6	8.3±4.8	15.5±6.2
anterior.48	91.6±2.4	3.6±1.7	1.8±2.0	3.0±2.0	76.7±5.0	5.9±2.7	8.6±4.2	8.8±2.8
anterior.fu	83.3±4.8	3.5±3.2	3.7±3.9	9.6±3.7	50.0±8.9	22.5±5.6	12.8±8.0	14.8±6.8
inferior.6	3.6±1.4	80.5±4.6	6.5±2.8	9.4±4.4	9.3±4.7	64.0±8.2	11.3±6.3	15.3±7.6
inferior.12	1.2±0.9	84.9±3.6	2.9±2.3	11.0±3.1	10.8±3.6	72.2±7.2	4.5±4.1	12.5±6.3
inferior.48	2.2±1.3	89.7±3.2	3.1±2.1	5.0±2.9	6.0±3.1	74.2±7.0	5.2±5.1	14.7±6.6
inferior.fu	4.1±2.2	72.7±3.9	6.5±3.4	16.7±4.6	2.5±3.0	58.6±6.6	6.6±4.7	32.3±7.1
normal	10.9±3.9	8.2±2.1	59.8±16.3	21.1±13.2	19.7±5.5	7.8±4.7	30.7±10.7	41.8±12.6
CAD	9.4±1.7	17.3±2.6	14.5±5.2	58.9±5.1	4.8±3.9	14.3±4.8	22.7±9.0	58.2±9.7

Table 7-4: Expanded classification breakdown – experiment E1.bp

The expanded classification breakdown (Table 7-4) provides further insight into this experiment. When comparing percentages of correctly classified examples in the anterior and inferior sub-classes it is noted that the MLPs consistently perform better when classifying patients 12 to 48 hours after the onset of the infarction. This is observed in both the training and testing set results. In relation to follow-up patients (anterior.fu and inferior.fu) the MLPs clearly have difficulty classifying these sub-classes. Based on the testing set results, only 50% (±8.9%) of follow-up patients with anterior infarcts (anterior.fu) were correctly classified. Similarly, of those follow-up patients with inferior infarcts (inferior.fu) only 58.6% (±6.6%) were correctly classified and a significant proportion were incorrectly classified as having coronary artery disease (32.3%±7.1%).

If the infarct classification results for the testing set are ranked, a trend becomes apparent (Table 7-5). The MLPs perform best when classifying 48 hour infarcts, followed by 12 hour infarcts, followed by 6 hour infarcts, and perform least effectively when classifying follow-up patients.

Class	% Correctly Classified
anterior.48	76.7±5.0
inferior.48	74.2±7.0
inferior.12	72.2±7.2
anterior.12	72.1±7.5
anterior.6	64.7±6.8
inferior.6	64.0±8.2
inferior.fu	58.6±6.6
anterior.fu	50.0±8.9

Table 7-5: Ranked testing set results for infarcts – experiment E1.bp

A similar ranking of training set infarct classification results is observed (Table 7-6). The MLPs perform best when classifying 48 hour infarct, followed by 12 hour infarcts, but the follow-up and 6 hour infarct patient rankings are not the same as for the testing set, in particular, the classification of *anterior.fu* patients is better than for *anterior.6* patients.

Class	% Correctly Classified
anterior.48	91.6±2.4
inferior.48	89.7±3.2
anterior.12	86.7±3.5
inferior.12	84.9±3.6
anterior.fu	83.3±4.8
inferior.6	80.5±4.6
inferior.fu	72.7±3.9
anterior.6	72.0±5.0

Table 7-6: Ranked training set results for infarcts – experiment E1.bp

7.1.3 Discussion

The aim of this experiment (*E1.bp*) was to train an MLP to discriminate between anterior infarctions, inferior infarctions, coronary artery disease and patients with normal heart function, based on the electrocardiographic body surface mapping data of each patient. Clearly this objective has not been completely achieved. Although the MLPs do exhibit a certain degree of discriminatory capability, a significant number of patients (in both the training and testing data sets) were classified incorrectly.

A number of observations may be made when analysing the results from this experiment. Firstly MLPs are capable of correctly classifying a significant proportion of patients with anterior and inferior infarctions. Secondly, the MLPs have difficulty differentiating patients with coronary artery disease from those patients with normal hearts. Thirdly, the effectiveness of infarct classification would appear to depend on the age of the infarct being classified.

The most significant downfall observed in these results is the difficulty MLPs have in separating patients with coronary artery disease from normal patients. Superficially we could conclude that the fault lies with deficiencies in the classification capabilities of MLPs. It should be realised that the reasons for this outcome may not necessarily be associated with a failure of the classifier. As has been highlighted in chapter 5 the detection of coronary artery disease using electrocardiographic data is not possible unless the patient is exercised at the time the recording is made. Since the BSM data used was of patients who were resting, this would suggest that patients with coronary artery disease might exhibit electrocardiographic behaviour that is the same as patients with normal hearts.

This raises an important point which needs to be considered when attempting to improve the classification performance of MLPs or any classifier applied to this problem. The poor performance of the classifier may be the result of a lack of information in the original data, rather than inadequacies in the classifier being applied to the problem. Alternatively the feature extraction technique used may be removing the key features required to perform the classification.

So in summary, there are three possible reasons for the poor classification performance observed;

- the classifier (MLP) is unable to identify the discriminating features,
- the feature extraction technique (KLT) is inadvertently removing the discriminating features, or
- the discriminating features do not exist in the BSM data,

or a combination of these issues may contribute to the classification performance.

Therefore the remainder of this chapter will attempt to explore this problem further by considering its various aspects. This will then be followed in chapter 8 by an examination of alternative neural network and traditional classification techniques in combination with different feature extraction techniques. The overall objective of each of the experiments that follow (in this chapter and chapters 8) is to explore this problem from a range of perspectives in an attempt to more accurately understand the classification problem and to determine the most appropriate classification techniques to apply.

7.2 Second Experiment (Problem 2)

Since in the initial experiments it would appear that the MLPs could not discriminate between CADs and normals, a number of problem subsets were considered in an attempt to understand the classification problem further. The first problem subset considered was the classification of all classes except for CAD. In this problem (problem 2) the classifiers were trained to separate patients into three classes; anterior infarctions, inferior infarctions and normals. Twenty MLP networks were separately trained on this classification problem using the parameters listed in Table 7-7. Similar testing of possible optimum architecture and training parameters was conducted and those used in the first experiment were found to be the most effective.

Parameter	Setting
network architecture	72:72:3
transfer function	sigmoid (hidden and output layer)
training technique	back-propagation
training limit	20000 epochs
training runs	20
learning rate	0.1
patience measure	Percentage
patience limit	1000 epochs
presentation type	pattern presentation
data selection	weighted sequential

Table 7-7: Training parameters used for experiment E2.bp

The patience measure halted training runs, on average, at 3530 epochs (±920 epochs). At the end of the training phase the average mean squared error of network outputs was 0.1033 (±0.0889) for the training set and 0.1898 (±0.0732) for the testing set. The resulting networks classified training set examples with an accuracy of 84.0% (±6.8%) and the testing set examples with an accuracy of 69.6% (±5.2%). The classification performance of these networks is summarised Table 7-8.

%	Percentage Classified Correctly			
Data set	anterior	inferior	normal	total
Training	85.2±3.6	88.9±1.7	77.7±18.3	84.0±6.8
Testing	71.6±3.7	77.6±4.2	59.7±15.5	69.6±5.2

Table 7-8: Network classification results - experiment E2.bp

These results are an improvement over the classification results for the first experiment (E1.bp). Given the classification breakdown in Table 7-9 a comparison was made between these results and the results obtained in the first experiment (Table 7-3).

%	Training Set			Testing Set		
Class	anterior	inferior	normal	anterior	inferior	normal
anterior	85.2±3.6	7.1±4.5	7.6±3.0	71.6±3.7	14.7±5.1	13.7±5.1
inferior	3.8±1.9	88.9±1.7	7.3±2.4	9.8±2.3	77.6±4.2	12.6±4.1
normal	11.6±9.5	10.6±9.3	77.7±18.3	26.3±8.2	14.0±10.3	59.7±15.5
CAD*	16.0±3.2	34.6±9.6	49.4±11.9	8.5±5.5	34.3±12.3	57.2±14.5

Table 7-9: Classification breakdown - experiment E2.bp
(* CAD class was not used during training)

The difference between these results was calculated (difference of means – Blalock 1960) and is presented in table 7-10 (a significance level of 0.0001 was selected). A number of observations can be made. Firstly the *E2.bp* networks are performing significantly better than the *E1.bp* networks when classifying normal patients in the testing set [E1.bp – 30.7% (±10.7%), E2.bp – 58.7% (±15.5%), p<0.0001]. Secondly, the classification of inferior infarcts has improved for both training and testing data sets. Thirdly, when the networks are applied to the CAD classes it is noted that a large proportion (training set - 49.4±11.9, testing set - 57.2±14.5) are classified as normal, suggesting that the normal and CAD classes are similar, as was suggested from the results obtained in the first experiment.

p	Training Set			Testing Set		
Class	anterior	inferior	normal	anterior	inferior	normal
anterior	0.0753	0.1117	0.0009	0.0002	0.0061	0.0414
inferior	0.0391	p<0.0001	0.0006	0.0005	p<0.0001	p<0.0001
normal	0.7688	0.2850	0.0029	0.0063	0.0243	p<0.0001
CAD*	p<0.0001	p<0.0001	p<0.0001	0.0223	p<0.0001	p<0.0001

Table 7-10: Comparison of classification breakdowns
between experiment E1.bp (Table 7-3) and E2.bp (Table 7-9)
(*CAD class was not used during training)

On examination of the expanded classification breakdown (Table 7-11) it is noted that the classification of infarction sub-classes exhibits similar trends to those observed in the first experiment. In relation to both the training and testing data sets the MLPs perform best when classifying 48 hour infarcts, followed by 12 hour infarcts, and perform least effectively when classifying 6 hour infarcts and follow-up patients (although the overall ranking of 6 hour and follow-up patients is not as clear).

%	Training Set			Testing Set		
Class	anterior	inferior	normal	anterior	inferior	normal
anterior.6	74.6±7.8	13.8±9.3	11.6±4.7	69.3±6.4	15.5±6.0	15.2±6.1
anterior.12	87.9±3.8	5.6±1.8	6.5±3.9	79.8±8.2	6.0±7.3	14.1±7.3
anterior.48	92.2±3.4	4.3±2.9	3.5±2.7	80.9±5.3	7.9±3.6	11.2±5.5
anterior.fu	86.3±5.5	4.8±6.3	8.8±4.2	56.5±8.1	29.2±8.8	14.2±9.0
inferior.6	5.6±4.2	86.5±4.0	7.8±3.4	13.0±7.0	71.3±7.8	15.7±6.7
inferior.12	2.0±1.9	92.0±1.9	6.0±2.8	13.7±2.8	80.0±6.4	6.3±4.8
inferior.48	2.0±1.7	93.7±2.2	4.3±2.3	7.2±3.4	84.3±4.7	8.5±5.3
inferior.fu	5.6±2.4	83.4±5.0	11.0±4.7	5.4±2.6	74.6±7.2	20.0±6.6
normal	11.6±9.5	10.6±9.3	77.7±18.3	26.3±8.2	14.0±10.3	59.7±15.5
CAD*	16.0±3.2	34.6±9.6	49.4±11.9	8.5±5.5	34.3±12.3	57.2±14.5

Table 7-11: Expanded classification breakdown – experiment E2.bp
(*CAD class was not used during training)

When these expanded results (Table 7-11) are compared with the first experiment (Table 7-4), as detailed if Table 7-12, it is noted that the improvement in inferior infarct classification performance is specifically associated with the inferior.48 and inferior.fu classes (as these increases are significant for both training and testing set).

p	Training Set			Testing Set		
Class	anterior	inferior	normal	anterior	inferior	normal
anterior.6	0.2299	0.0867	0.0016	0.0381	0.0082	0.1408
anterior.12	0.3176	0.0238	0.4304	0.0044	0.3176	0.0066
anterior.48	0.5338	0.3710	0.0339	0.0162	0.0606	0.1100
anterior.fu	0.0810	0.4291	0.0004	0.0237	0.0085	0.6152
inferior.6	0.0611	0.0001	0.2060	0.0642	0.0077	0.0436
inferior.12	0.1088	p<0.0001	0.0006	0.0087	0.0011	0.2215
inferior.48	0.6861	p<0.0001	0.1011	0.2625	p<0.0001	0.0577
inferior.fu	0.0516	p<0.0001	0.0017	0.0029	p<0.0001	p<0.0001
normal	0.7688	0.2850	0.0029	0.0063	0.0243	p<0.0001
CAD*	p<0.0001	p<0.0001	p<0.0001	0.0223	p<0.0001	p<0.0001

Table 7-12: Comparison of expanded classification breakdown
between experiment E1.bp (Table 7-4) and experiment E2.bp (Table 7-11)
(*CAD class was not used during training)

7.2.1 Discussion

It becomes clear when comparing the results of this experiment (*E2.bp*) with the first experiment (*E1.bp*) that the MLP classifiers were having difficultly separating normal patients from CAD patients. When the MLPs were only required to classify patients into three classes (anterior, inferior, and normal) the classification performance subsequently improved particularly in relation to normal patients. Although the CAD patients were not used during training, they were presented to the network after training to see what classification might be assigned. It is noted that when the CAD classes (for both training and testing sets) are classified by the *E2.bp* MLPs a significant proportion of the CAD patients are classified as normal.

When comparing the distribution of normal patient classifications with CAD patients it is noted that the proportion of incorrect classifications is different. In particular, in both the training and testing set results approximately 34% [training – 34.6% (±9.6%), testing set – 34.3% (±12.3%)] of CAD patients are classified as having inferior infarcts. This is significantly different when compared to the normal classes [training set - 10.6% (±9.3%) normals classified as inferior infarcts, testing set – 14.0% (±10.3%) normals classified as inferior infarcts]. This may suggest that some CAD patients are presenting electrocardiographic features that are similar to inferior infarcted patients. This is also supported by the noted improvement in the inferior infarct classification performance with the removal of the need to assign a CAD classification. Although this is a somewhat interesting observation this will require further analysis and consideration before a firm conclusion can be made (see chapter 8).

With regard to misclassification there are two other aspects of the classification breakdown which are of interest. When examining the proportion of infarcted patients misclassified in the testing set results (Table 7-11), two classes of misclassification are of concern. Firstly, 20% (±6.6%) of inferior follow-up patients are misclassified as normal. Secondly, 29.9% (±8.8%) of anterior follow-up patients are misclassified as inferior infarcts. This degree of misclassification suggests that the MLPs are having difficultly correctly classifying follow-up infarcts. Therefore, as with the CAD/normal patient problem observed in the previous experiment, the use of follow-up infarcts in the training process may degrade the effectiveness of training and consequently the overall MLP classification performance. Having identified this issue a second version of the experiment (E3.bp) was conducted where follow-up infarct patients were excluded from the training process to ascertain whether this would improve the overall three-class classification performance.

7.3 Third Experiment (Problem 3)

As described in section 7.2.1 there are a significant number of follow-up infarcts misclassified by the MLP networks with respect to the three-class problem (problem 2). Therefore, the third experiment considered involved training MLP networks to classify patients into three classes (anterior, inferior and normal) as for Problem 2. However in this problem follow-up infarct patients were not used during the training process. The objective of this experiment was to see if the MLP classification performance could be improved over those results obtained in the second experiment (E2.bp).

Twenty MLP networks were separately trained on this classification problem using the parameters listed in Table 7-13. Similar testing of possible optimum architecture and training parameters was conducted as with the first experiment.

Parameter	Setting
network architecture	72:72:3
transfer function	sigmoid (hidden and output layer)
training technique	back-propagation
training limit	20000 epochs
training runs	20
learning rate	0.1
patience measure	Percentage
patience limit	1000 epochs
presentation type	pattern presentation
data selection	weighted sequential

Table 7-13: Training parameters used for experiment E3.bp

The patience measure halted training runs, on average, at 4030 epochs (±1120 epochs). At the end of the training phase the average mean squared error of network outputs was 0.0954 (±0.0942) for the training set and 0.1757 (±0.0808) for the testing set. The resulting networks classified training set examples with an accuracy of 81.8% (±6.4%) and the testing set examples with an accuracy of 69.9% (±5.4%). The classification performance of these networks is summarised Table 7-14.

%	Percentage Classified Correctly			
Data set	anterior	inferior	normal	total
training	82.4±2.5	85.6±1.7	77.5±18.1	81.8±6.4
testing	70.9±4.0	75.0±4.9	63.8±17.1	69.9±5.4

Table 7-14: Network classification results - experiment E3.bp

The final testing set classification performance of 69.9% (±5.4%) is not significantly different from that obtained in the second experiment [69.9% (±5.2%)] since a difference of means test fails with p=0.86.

%	Training Set			Testing Set		
Class	anterior	inferior	Normal	anterior	Inferior	normal
anterior	82.4±2.5	7.5±3.5	10.0±3.5	70.9±4.0	15.0±4.3	14.1±4.6
inferior	5.3±2.6	85.6±1.7	9.1±2.6	10.3±2.9	75.0±4.9	14.6±5.0
normal	3.0±10.7	9.5±8.0	77.5±18.1	7.7±12.1	8.5±7.5	63.8±17.1
CAD*	16.0±3.2	34.6±9.6	49.4±11.9	10.0±6.8	28.0±11.1	62.0±15.2

Table 7-15: Classification breakdown - experiment E3.bp
(* anterior.fu, inferior.fu, and CAD classes where not used during training)

When comparing the results in Table 7-15 (E3.bp) with those in Table 7-9 (E2.bp) using a difference of means test, there are few significant differences between the two experiments (Table 7-16). Two significant changes however are noted. Firstly, the correct classification of training set inferiors has degraded [E2.bp – 88.9% (±1.7%) down to E3.bp – 85.6% (±1.7%)] although no similar change is noted in relation to the correct classification of testing set inferiors. Secondly, the misclassification of testing set normals as anterior infarcts has decreased [E2.bp – 26.3% (±8.2%) down to E3.bp – 7.7% (±12.1%)].

p	Training Set			Testing Set		
Class	anterior	inferior	Normal	anterior	Inferior	normal
anterior	0.0086	0.7614	0.0289	0.5787	0.8456	0.8009
inferior	0.0498	p<0.0001	0.0325	0.5596	0.0869	0.1856
normal	0.0125	0.6980	0.9732	p<0.0001	0.0680	0.4434
CAD*	1.0000	1.0000	1.0000	0.4593	0.1054	0.3254

Table 7-16: Comparison of classification breakdown
between experiment E2.bp (Table 7-9) and experiment E3.bp (Table 7-15)

An analysis of the expanded classification breakdown provides some further insights in the difference between experiment E2.bp and E3.bp. When comparing Table 7-17 (E3.bp) with Table 7-11 (E2.bp) a number of significant differences become apparent (see Table 7-18).

%	Training Set			Testing Set		
Class	anterior	inferior	normal	anterior	inferior	normal
anterior.6	78.7±6.8	11.6±8.2	9.7±3.9	69.3±5.4	14.1±5.1	16.6±7.2
anterior.12	90.1±2.8	4.7±1.5	5.2±3.0	80.2±6.6	7.2±6.2	12.6±5.8
anterior.48	94.1±2.2	3.8±1.7	2.2±1.8	80.9±5.1	8.3±3.7	10.9±4.7
anterior.fu*	66.7±8.2	10.2±5.3	23.1±10.2	53.2±9.5	30.2±8.3	16.5±7.8
inferior.6	4.5±5.7	88.9±4.9	6.7±2.9	12.2±6.4	70.8±8.2	17.0±8.0
inferior.12	2.0±1.8	92.2±1.8	5.8±2.3	13.8±3.5	77.5±4.9	8.7±5.0
inferior.48	2.7±2.7	94.1±2.7	3.2±2.0	7.8±6.2	82.5±6.6	9.7±5.6
inferior.fu*	12.0±3.7	67.2±7.3	20.8±6.6	7.5±3.2	69.3±9.5	23.2±8.8
normal	3.0±10.7	9.5±8.0	77.5±18.1	7.7±12.1	8.5±7.5	63.8±17.1
CAD*	17.4±4.4	30.7±8.5	52.0±12.3	10.0±6.8	28.0±11.1	62.0±15.2

Table 7-17: Expanded classification breakdown – experiment E3.bp
(* anterior.fu, inferior.fu, and CAD classes where not used during training)

The only changes in the classification of infarcts relate to follow-up patients in the training set. Clearly the number of training set follow-up infarcts correctly classified by MLPs in experiment E3.bp has degraded, and consequently the percentage of

associated incorrect classification has increased. This is to be expected, as none of the follow-up training examples are used during training. It is interesting to note that this change in classification performance is not observed in the testing set results. Further to this the only significant changes in the testing set results pertain to the decrease in the proportion of normal patients misclassified as having anterior infarcts.

p	Training Set			Testing Set		
Class	Anterior	inferior	normal	anterior	inferior	normal
anterior.6	0.0921	0.4439	0.1831	1.0000	0.4432	0.5217
anterior.12	0.0496	0.1023	0.2569	0.8693	0.5882	0.4875
anterior.48	0.0489	0.5215	0.0898	1.0000	0.7374	0.8575
anterior.fu*	p<0.0001	0.0069	p<0.0001	0.2564	0.7205	0.4050
inferior.6	0.5026	0.1064	0.2901	0.7151	0.8483	0.5903
inferior.12	1.0000	0.7409	0.8112	0.9230	0.1846	0.1393
inferior.48	0.3459	0.6195	0.1238	0.7141	0.3395	0.5015
inferior.fu*	p<0.0001	p<0.0001	p<0.0001	0.0325	0.0603	0.2129
normal	0.0125	0.6980	0.9732	p<0.0001	0.0680	0.4434
CAD*	0.2694	0.1926	0.5117	0.4593	0.1054	0.3254

Table 7-18: Comparison of expanded classification breakdown between experiment E2.bp (Table 7-11) and experiment E3.bp (Table 7-17)

7.3.1 Discussion

The advantages of not using the infarcted follow-up classes during training would appear to be clear. Although it does not provide any significant improvement in the testing set classification performance, it does reduce the number of normal patients incorrectly classified as having anterior infarctions.

The other interesting aspect of these results is the differences between the E2.bp and E3.bp results in relation to the follow-up patients. Changes with respect to both correct and incorrect classification of training set follow-up infarcts are noted. In particular the number of correctly classified anterior.fu and inferior.fu patients decreases in this third experiment [anterior.fu; E2.bp - 86.3% (±5.5%) down to E3.bp - 66.7% (±8.2%), inferior.fu; E2.bp – 83.4% (±5.0%) down to E3.bp – 67.2% (±7.3%)]. Interestingly a significant drop in performance is not observed in relation to the testing set follow-up infarcts. This may well suggest that the extra discriminatory information gained by the inclusion of follow-up infarcts during training (E2.bp) is specific to the training set patterns and does not generalise to the testing set. This would therefore suggest that follow-up infarcts should not be included in the training process, and therefore the training approach in the third experiment (E3.bp) is the most appropriate for the three-class problem.

7.4 Fourth Experiment (Problem 4)

The final experiment in this initial series returns to address the problem of diferentiating between patients with CAD and normal patients. This was initially observed in the first experiment (E1.bp – see section 7.1). This section considers an experiment (E4.bp) which focuses specifically on attempting to train MLPs to discriminate between CAD patients and normal patients. This is a two-class problem, and the training parameters are described in Table 7-19. Twenty MLP networks with different starting weights were trained using the CAD and normal training classes.

Parameter	Setting
network architecture	72:72:2
transfer function	sigmoid (hidden and output layer)
training technique	back-propagation
training limit	20000 epochs
training runs	20
learning rate	0.1
patience measure	percentage
patience limit	1000 epochs
presentation type	pattern presentation
data selection	weighted sequential

Table 7-19: Training parameters used for experiment E4.bp

The patience measure halted training runs, on average, at 1790 epochs (±450 epochs). At the end of the training phase the average mean squared error of network outputs was 0.4231 (±0.0243) for the training set and 0.4746 (±0.0124) for the testing set (which is somewhat high when compared to the other experiments). The resulting networks classified training set examples with an accuracy of 62.8% (±0.9%) and the testing set examples with an accuracy of 63.9% (±2.3%). The classification performance of these networks is summarised in Table 7-20.

%	Percentage Classified Correctly		
data set	normal	CAD	total
training	56.4±5.9	69.2±5.9	62.8±0.9
testing	57.7±7.7	70.2±5.1	63.9±2.3

Table 7-20: Network classification results - experiment E4.bp

These results would suggest that the MLPs are managing to classify more than 60% of training and testing set patients correctly. However, the individual classification results for each run (see Table 7-21) reveal a number of problems. Although all networks manage to achieve overall classification results which are greater than 60% for both the training and testing sets many of these networks are biased toward one particular class.

%	Percentage Classified Correctly					
	training data set			testing data set		
run	normal	CAD	total	normal	CAD	total
1	58.0	66.9	62.5	66.7	73.3	70.0
2	58.0	68.5	63.3	56.7	70.0	63.3
3	53.4	73.1	63.3	53.3	73.3	63.3
4	51.2	75.7	63.4	43.3	76.7	60.0
5	58.8	65.6	62.2	63.3	66.7	65.0
6	52.7	73.1	62.9	50.0	70.0	60.0
7	55.7	68.9	62.3	53.3	73.3	63.3
8	56.5	68.2	62.3	60.0	66.7	63.3
9	55.0	68.5	61.7	60.0	70.0	65.0
10	55.0	70.2	62.6	60.0	66.7	63.3
11	54.2	69.5	61.9	60.0	66.7	63.3
12	50.4	76.4	63.4	50.0	73.3	61.7
13	58.0	66.6	62.3	66.7	66.7	66.7
14	79.4	46.6	63.0	80.0	53.3	66.7
15	59.5	70.5	65.0	53.3	76.7	65.0
16	53.4	69.5	61.5	60.0	70.0	65.0
17	55.7	69.8	62.8	56.7	70.0	63.3
18	55.0	68.9	61.9	56.7	73.3	65.0
19	51.2	74.4	62.8	50.0	76.7	63.3
20	57.3	72.5	64.9	53.3	70.0	61.7

Table 7-21: Classification results for each training run - experiment E4.bp

Some extreme examples of this classification bias are observed in runs 12 and 14. Run 12 performs very well when classifying CAD patients (training set – 76.4%, testing set – 73.3%), but performs poorly when classifying normal patients (training set – 50.4%, testing set – 50.0%). Clearly run 12 is biased toward the CAD class. Similarly, run 14 performs very well when classifying normal patients (training set – 79.4%, testing set – 80%), but performs poorly when classifying CAD patients (training set – 46.6%, testing set – 53.3%). This suggests that run 14 is biased toward the normal class. Since this is a two-class problem, then a random guess could achieve a classification performance of 50%. In light of this fact the poor results achieved by runs 12 and 14 would appear to be no better than a guess, which is a somewhat concerning outcome.

The problem of classification bias is not only restricted to extreme cases. For example, run 15 achieved the best overall classification performance for the training set (65%). However, when examining the testing set results it is noted that the network is biased toward the CAD class (53.3% of normals classified correctly, 76.7% of CADs classified correctly).

Given the instability of these results a different technique was considered for selecting the best network. As just illustrated, selecting the network that performs best overall with respect to the training set may still be biased toward one of the two classes. Ideally, the best classifier would be one whose classification performance is not biased. Therefore, a network ranking approach was considered that did not rank networks according to the best training set classification performance, but with respect to the difference in class classification performance.

%	Percentage Classified Correctly					
	training data set			testing data set		
run	normal (N_{TR})	CAD (C_{TR})	difference (N_{TR}-C_{TR})	normal (N_{TS})	CAD (C_{TS})	difference (N_{TS}-C_{TS})
12	50.4	76.4	-26.0	50.0	73.3	-23.3
4	51.2	75.7	-24.6	43.3	76.7	-33.3
19	51.2	74.4	-23.3	50.0	76.7	-26.7
6	52.7	73.1	-20.4	50.0	70.0	-20.0
3	53.4	73.1	-19.7	53.3	73.3	-20.0
16	53.4	69.5	-16.1	60.0	70.0	-10.0
11	54.2	69.5	-15.3	60.0	66.7	-6.7
20	57.3	72.5	-15.2	53.3	70.0	-16.7
10	55.0	70.2	-15.2	60.0	66.7	-6.7
17	55.7	69.8	-14.1	56.7	70.0	-13.3
18	55.0	68.9	-13.9	56.7	73.3	-16.7
9	55.0	68.5	-13.6	60.0	70.0	-10.0
7	55.7	68.9	-13.1	53.3	73.3	-20.0
8	56.5	68.2	-11.7	60.0	66.7	-6.7
15	59.5	70.5	-11.0	53.3	76.7	-23.3
2	58.0	68.5	-10.5	56.7	70.0	-13.3
1	58.0	66.9	-8.9	66.7	73.3	-6.7
13	58.0	66.6	-8.5	66.7	66.7	0.0
5	58.8	65.6	-6.8	63.3	66.7	-3.3
14	79.4	46.6	32.8	80.0	53.3	26.7

Table 7-22: Training runs ranked with respect to the difference in class classification performance for the training set - experiment E4.bp

The difference in class classification performance for each *E4.bp* runs is presented in Table 7-22. Runs are ranked according to differences in class classification performance on the training set.

These results reveal some distinct trends. The majority of networks tend to bias the CAD class and only one network is biased toward the normal class (run 14). Further to this a correlation exists between the training set bias (N_{TR}-C_{TR}) and the testing set bias (N_{TS}-C_{TS}). Statistically this relationship reveals a correlation coefficient of R=0.8858. In short, this suggests that if a network yields a balanced classification performance with respect to the training set then the same network would provide a similar balanced classification performance for the testing set.

Given these results the best network could be selected by choosing the network with the minimum training set bias $|N_{TR}-C_{TR}|$. In the case of this experiment this selection criteria would select run 5 as the most balanced network with a training set bias of 6.8%. Clearly this is not the best network with respect to testing set bias. However the testing set bias for run 5 is still very low at 3.3% and achieves an above average classification performance with respect to the testing set (65% classified correctly overall, 63.3% of normals classified correctly, and 66.7% of CAD patients classified correctly)

At first this approach may appear somewhat arbitrary, but as will be seen in the next chapter, the approach was found to work reliably for a number of other neural network classifiers applied to the same problem.

7.4.1 Summary

In summary the following findings were presented in this chapter:

***Problem 1** (seperating patients into anterior, inferior, normal and CAD classes)*

Testing set results (percentage correct):
anteriors - 65.9±4.6%, inferiors – 67.2±4.8%, normals – 30.7±10.7%, CADs – 58.2±9.7%.

MLPs are capable of correctly classifying a significant proportion of patients with anterior and inferior infarctions. However, the MLPs have difficulty differentiating patients with coronary artery disease from those patients with normal hearts (41.8±12.6% of normals incorrectly classified as CADs).

***Problem 2** (classifying patients into anterior, inferior, and normal classes)*

Testing set results (percentage correct):
anteriors – 71.6±3.7%, inferiors – 77.6±4.2%, normals – 59.7±15.5%.

When the MLPs were only required to classify patients into three classes the classification performance subsequently improved particularly in relation to the normal patients. Further to this it was noted that when the CAD patients were presented to the network a significant proportion of the CAD patients were classified as normal. This further suggested that the CAD patients exhibited very similar BSMs to normal patients.

***Problem 3** (classifying patients into anterior, inferior, and normal classes - not using the follow-up infarct patients during training)*

Testing set results (percentage correct):
anteriors – 70.9±4.0%, inferiors – 75.0±4.9%, normals – 63.8±17.1%.

These results were very similar to those achieved for problem 2. However the results would suggest that the inclusion of the follow-up patients in problem 2 was biasing network training. Removing these patients from the training process would appear to improve the classification of normal patients.

***Problem 4** (classifying patients into normal and CAD classes)*

Testing set results (percentage correct):
nromals – 57.7±7.7%, CADs – 70.2±5.1%, overall – 63.9±2.3%.

This classification problem was found to be very difficult. It was observed that a number of networks tended to bias toward correctly classifying CAD patients and subsequently degrading the classification performance in relation to normal patients. Although the overall classification performance was greater than 60% the classification of normal patients was often no better than a guess (50%). A network selection criteria was devised which selected an unbiased network from the 20 training runs. The final network classified 63.3% of normal patients correctly and 66.7% of CAD patients correctly (with respect to the testing set).

8. Alternative Classification Techniques Applied to BSM Data Classification

This chapter presents the second set of experiments conducted in this study. Having applied multilayer perceptrons (MLPs) to the four BSM classification problems, a number of alternative feature extraction and classification techniques were considered.

These experiments were conducted for a number of reasons. Firstly to determine whether the classification performance in the initial experiments could be improved upon. Secondly, to provide a comparison between neural network classification techniques and a number of alternative classification techniques. Thirdly, to determine which feature extraction techniques are most appropriate to use in conjunction with these classifiers.

As described in chapter 6, six neural network classification techniques were considered:

- MLPs trained using back-propagation (*bp*),
- MLPs trained using quickprop (*qp*),
- cascade correlation networks (*cas*),
- committees of MLPs trained using back-propagation (*cbp*),
- committees of MLPs trained using quickprop (*cbp*),
- committees of cascade correlation networks (*ccas*).

And four alternative classification techniques:

- k-nearest neighbour (*knn*),
- linear regression (*linreg*),
- two inductive learning techniques (*C4.5, MML*).

These ten classification techniques were used in conjunction with four different feature extraction techniques:

- KLT transform (72 attributes) (*E*)
- logical reduction of the QRS, ST, and T features into six summary maps (192 attributes) (*L*).
- logical reduction of the QRS complex into four summary maps (128 attributes) (*qrs*).
- logical reduction of the ST segment and T wave features into two summary maps (64 attributes) (*st*).

Further to this, four traditional BSM classification techniques were applied to each classification problem without applying feature extraction;

- k-nearest neighbour using root-mean-squared error (RMSE) as distance measure (*knn-rmse*),
- k-nearest neighbour using correlation coefficient (R) as distance measure (*knn-cc*),
- nearest class mean using root-mean-squared error (RMSE) as distance measure (*ncm-rmse*), and
- nearest class mean using correlation coefficient (R) as distance measure (*ncm-cc*).

Thus, a total of forty-four classification experiments were applied to each of the four classification problems;

- 24 neural network experiments (6 neural network techniques – *bp, qp, cas, cbp, cqp, ccas* using 4 different feature extraction techniques – *E, L, qrs, st*),
- 16 alternative classification experiments (4 alternative classification techniques – *knn, linreg, C4.5, MML* using 4 different feature extraction techniques – *E, L, qrs, st*), and
- 4 traditional BSM classification experiments (4 traditional BSM classification techniques – *knn-rmse, knn-cc, ncm-rmse, ncm-cc* using the complete BSM recordings - *C*).

A detailed description of these classification techniques is provided in chapter 2 and 3. Other specific aspects of data acquisition and data reduction (ie feature extraction) and classification problems are provided in chapter 6. The following four sections (8.1, 8.2, 8.3, and 8.4) provide a detailed presentation and discussion of the results obtained for each classification problem.

8.1 Problem 1

As described in chapters 6 and 7 this problem involved the classification of patients into one of four categories: anterior myocardial infarction, inferior myocardial infarction, coronary artery disease, and normal patients. The next section describes the training of each type of classifier and this is followed by a description of the results obtained.

8.1.1 Training Details

The training set data described in chapter 7 was used to train all classifiers. Four training and testing data sets were prepared using the four feature extraction techniques (E, L, qrs, st). Each classification technique was applied to these data sets (except for the traditional BSM classification techniques, which use the complete BSM recordings). The classifier training details are presented in the following sections. The CTAS tools (Classifier Training and Analysis Suite – see Appendix B) were used to train all MLP networks. Committee networks and k-nearest neighbour classifiers were also constructed using the CTAS tools. The TasCas simulator (Waugh 1993) was used to train cascade-correlation networks. The regression tools in the statistical package S (Becker and Chambers 1984) were used for constructing linear regression discriminant functions. The programs C4.5 (Quinlan 1993) and MML (Wallace 1990) were use for constructing inductive learning decision trees. All results of these experiments were compiled using the CTAS analysis tools (see Appendix B).

8.1.1.1 MLPs trained using back-propagation (bp)

The training of MLPs with back-propagation was very similar to that described for the initial experiments in chapter 7. The architecture of the MLP networks used was determined by the feature extraction technique used. Trial runs were performed to select the most appropriate number of hidden nodes. The network architecture used in the final training runs were; 72:72:4 when using the KLT transform (*E*), 192:120:4 when using the six map logical reduction technique (*L*), 128:120:4 when using the four map QRS logical reduction technique (*qrs*), and 64:64:4 when using the two map ST logical reduction technique. All neurons used sigmoid transfer functions and a bias input was assigned to both hidden and output layer nodes.

A suitable learning rate (η) was found to be 0.1. A patience limit of 1000 epochs was used in conjunction with a percentage correct patience measure (see chapter 7 for further details). Network weights were initialised to random starting values in the range –1.0 to 1.0 and weight updates were performed after each pattern presentation. For each experiment twenty networks with different starting weights were trained on this problem.

8.1.1.2 MLPs training using quickprop (qp)

The network architectures used for MLPs trained with quickprop were the same as the architectures used for those trained with backprop (see section 8.1.1.1).

A suitable learning rate (η) was found to be 0.1 using a maximum growth rate of 1.75. A patience limit of 1000 epochs was used in conjunction with a percentage correct patience measure. Network weights were initialised to random starting values in the range –1.0 to 1.0 and weight updates were performed at the end of each epoch (batch presentation) as required by the quickprop algorithm. For each experiment twenty networks with different starting weights were trained on this problem.

8.1.1.3 Cascade-Correlation Networks (cas)

The TasCas (Waugh 1994) cascade-correlation network simulator was used for all cascade-correlation experiments. This simulator (see Appendix B) was based on the work of Falhman (1991). The networks started with a perceptron style architecture and hidden layers were added during training. Weight adjustment is performed using the quickprop algorithm and the following training parameters (Table 8-1):

training parameter	hidden layer (candidate) training	ouput layer training
learning rate	1.0	0.35
maximum growth rate	1.75	1.75
patience period	50 epochs	50 epochs
patience percentage	3%	1%
epoch limit	500 epochs	500 epochs

Table 8-1 : Cascade-Correlation Traning Parameters

The patience measure was applied to the training error. Training was stopped when the error did not improve by the patience percentage within the patience period set (hidden layer 3% in 50 epochs, output layer 1% in 50 epochs). A standard limit (Waugh 1994) of 20 hidden nodes was set as the maximum number of hidden nodes that could be added to any network architecture. On average networks stopped training after 14 (±3) hidden nodes were added to the architecture.

*8.1.1.4 Committees of MLPs trained using back-propagation (*cbp*)*

Committee networks were constructed by combining the twenty individual networks from the corresponding individual training runs for MLPs trained using back-propagation (section 8.1.1.1). To classify a patient the patient attributes are presented to each trained MLP network and the output vectors are summed together and averaged to produce a single combined output vector (note – averaging is not necessary but was applied to for the purposes of applying thresholds – see chapter 9).

8.1.1.5 Committees of MLPs trained using quickprop (cbp)

Similar to the committee of MLPs trained using back-propagation, the quickprop committee networks were constructed using the twenty networks trained in individual training runs for the MLPs trained using quickprop (section 8.1.1.2).

8.1.1.6 Committees of cascade correlation networks (ccas)

Similar to the other committee networks these networks were constructed from the twenty cascade correlation networks trained in the individual cascade correlation training runs (section 8.1.1.3).

8.1.1.7 k-nearest neighbour (knn)

The training set was used as a control group for the nearest neighbour classifiers. A euclidean distance measure was used. The most appropriate k value was found to be 5 for this particular classification problem. When classifying a particular patient, the five closest examples from each training set class were selected and the average euclidean distance for each class calculated. The patient was assigned to the closest class. All knn classifiers were constructed using the CTAS tools.

8.1.1.8 Linear regression (linreg)

Linear discriminant functions were calculated using linear regression applied to the training set data. This statistical package S (Becker and Chambers 1984) was used to perform the regression analysis. The discriminant functions generated were of the form:

$$C = Ax + B$$

where x is the patient pattern being classified, and C is the output vector of the classifier. The regression analysis generated the transformation matrix A and the constant vector B. The dimensions of these parameters was determined by the number of attributes generated by the feature extraction techniques and the number of classes, which in this case was four classes.

8.1.1.9 *C4.5* (C4.5)

The inductive learning program C4.5 (Quinlan 1993) was applied using the default parameter settings. After some initial trails it was identified that C4.5 was performing well with respect to classifying the training data set, but was performing poorly when classifying the testing data set. It was found that activating a pruning phase (-p argument) improved the classification performance on the testing set.

8.1.1.10 *MML* (MML)

The inductive learning program MML (Wallace 1990) was applied using the default parameter settings. This program did not suffer from the same overtraining problem observed when using C4.5. (as MML uses a degree of decision tree pruning by default). No adjustment was made to the default parameters.

8.1.1.11 Traditional BSM classification techniques (knn-rmse, knn-cc, ncm-rmse, ncm-cc)

The CTAS tools were not used for constructing these classifiers, although the classifier output vectors were analysed using the CTAS tools. Purpose designed programs written in C were designed to calculate distance measures and classify patients. The CTAS tools could not be used to construct these classifiers as the complete data sets could not be loaded by the CTAS tools (1137 patients with 384000 attributes represents over 400 million attribute values). The purpose-designed programs overcame this problem by loading patient data as required.

The training set was used as a control group for constructing the k-nearest neighbour (knn) and nearest class mean (ncm) classifiers.

For k-nearest neighbour (knn) classifiers k=5 was found to provide an optimum classification performance for both training and testing data sets. The root-mean-squared error (RMSE) and correlation coefficients (cc) measures were calculated for each patient with respect to each class. Output classification vectors were calculated using an average of these measures (k=5), which were used to classify the patients. A patient was assigned to a class based on the class associated with the minimum value in the output vector.

For nearest class mean (ncm) classifiers an average mapping was calculated for each class (constructing a mean attribute vector consisting of 38400 attributes). The root-mean-squared error (RMSE) and correlation coefficients (cc) measures were calculated for each patient with respect to these class means. Output classification vectors were calculated using these measures, which were used to classify the patients. A patient was assigned to a class based on the class associated with the minimum value in the output vector.

8.1.2 Results

The classification results were analysed and prepared using the CTAS analysis tools *performance* and *breakdown* (see appendix B). In this chapter only the summarised classification performance results are presented. The more detailed classification breakdown results are provided in appendix C for reference and will be referred in the discussion. The classification results for these experiments are summarised in Tables 8-2, 8-3, 8-4, 8-5 and 8-6.

Note that where a number of training runs were performed in an experiment (*bp*, *qp*, and *cas* – 20 networks trained in each case) the averaged results are presented. The selection of the best training run on the basis of training performance was considered, but it was found not to provide statistically significant or consistent improvements in test set results and in some cases was actually less than the mean results (see chapter 7 for details).

%	Training Set					Testing Set				
experiment	anterior	inferior	normal	cad	total	anterior	inferior	normal	cad	total
E1.knn	97.6	97.9	97.7	99.3	98.1	76.6	66.7	33.3	80.0	64.2
E1.linreg	89.9	89.1	61.8	57.7	74.6	83.0	78.5	30.0	80.0	67.9
E1.C4.5	98.5	98.2	100.0	96.4	98.3	60.2	58.3	26.7	43.3	47.1
E1.MML	82.2	86.0	67.9	56.4	73.1	70.6	46.3	16.7	36.7	42.6
E1.bp	83.4±2.3	82.0±2.3	59.8±16.3	58.9±5.1	71.0±3.2	65.9±4.6	67.2±4.8	30.7±10.7	58.2±9.7	55.5±2.3
E1.qp	77.0±6.0	84.1±5.5	31.0±23.5	29.5±23.3	55.4±8.7	67.5±6.1	79.1±8.5	25.3±20.2	32.5±26.5	51.1±7.4
E1.cas	96.2±1.6	85.1±2.3	44.1±19.6	24.4±13.9	62.4±3.3	86.3±1.5	78.7±1.0	40.0±13.2	37.0±22.2	60.5±3.4
E1.cbp	91.1	89.9	75.6	71.2	81.9	76.1	76.0	26.7	63.3	60.5
E1.cqp	85.2	88.8	42.0	47.9	66.0	77.0	83.0	30.0	43.3	58.3
E1.ccas	95.8	85.9	39.7	25.9	61.8	84.7	77.6	30.0	53.3	61.4

Table 8-2: Classification results for E1 experiments (using KLT feature extraction).

%	Training Set					Testing Set				
experiment	anterior	inferior	normal	cad	total	anterior	inferior	normal	cad	total
L1.knn	96.8	98.7	95.4	99.0	97.5	69.7	73.4	36.7	76.7	64.1
L1.linreg	91.4	96.5	74.1	64.9	81.7	73.6	80.2	36.7	63.3	63.5
L1.C4.5	99.0	98.9	100.0	95.4	98.3	65.7	76.6	26.7	50.0	54.7
L1.MML	92.2	92.9	100.0	83.9	92.3	72.8	85.4	13.3	70.0	60.4
L1.bp	90.5±2.6	90.0±2.5	7.8±23.0	56.1±32.5	61.1±9.9	74.2±6.3	71.9±6.6	3.2±9.6	53.3±32.6	50.6±6.0
L1.qp	65.8±8.9	72.4±9.0	33.2±23.9	58.9±12.9	57.5±8.4	53.4±8.2	62.3±7.9	25.0±17.6	65.2±15.3	51.5±6.7
L1.cas	98.6±0.6	90.8±2.3	67.7±8.3	31.6±10.5	72.2±2.3	89.1±1.7	77.2±1.8	38.7±3.1	33.3±9.4	59.6±1.9
L1.cbp	96.4	97.4	26.7	90.5	77.8	78.7	80.4	0.0	80.0	59.8
L1.cqp	71.1	83.8	47.3	74.4	69.2	55.0	71.9	40.0	83.3	62.5
L1.ccas	97.9	91.8	66.4	34.1	72.5	90.2	77.3	40.0	46.7	63.6

Table 8-3: Classification results for L1 experiments (using QRST six map feature extraction).

%	Training Set					Testing Set				
experiment	anterior	inferior	normal	cad	total	anterior	inferior	normal	cad	total
qrs1.knn	96.8	98.5	95.4	98.0	97.2	74.4	70.2	36.7	76.7	64.5
qrs1.linreg	88.2	89.7	66.4	59.7	76.0	72.3	82.9	50.0	63.3	67.1
qrs1.C4.5	98.5	97.7	100.0	96.1	98.1	74.0	61.0	30.0	40.0	51.2
qrs1.MML	96.0	92.5	98.5	76.4	90.8	74.8	63.5	36.7	36.7	52.9
qrs1.bp	86.1±2.8	85.0±5.1	29.7±36.5	41.3±34.2	60.5±11.3	70.6±7.6	67.0±11.4	15.3±20.9	38.8±32.9	47.9±5.6
qrs1.qp	67.2±5.5	70.4±9.7	29.4±23.2	46.5±20.4	53.4±9.8	55.6±7.2	61.4±9.0	22.8±21.2	54.2±20.9	48.5±7.4
qrs1.cas	96.5±0.8	77.5±3.8	54.0±18.6	19.7±7.6	61.9±5.8	86.6±2.3	59.0±1.6	35.3±12.8	20.0±7.1	50.2±3.3
qrs1.cbp	92.5	93.9	61.1	78.7	81.5	77.8	78.0	13.3	70.0	59.8
qrs1.cqp	68.4	84.4	45.8	69.8	67.1	54.9	72.1	40.0	70.0	59.3
qrs1.ccas	96.3	80.3	56.5	23.0	64.0	84.4	62.5	23.3	40.0	52.6

Table 8-4: Classification results for qrs1 experiments (using four map QRS feature extraction).

%	Training Set					Testing Set				
experiment	anterior	inferior	normal	cad	total	anterior	inferior	normal	cad	total
st1.knn	95.2	97.9	96.2	99.7	97.2	53.0	71.2	0.0	66.7	47.7
st1.linreg	78.6	79.3	53.4	52.5	66.0	72.0	68.0	26.7	66.7	58.3
st1.C4.5	97.8	96.7	100.0	96.1	97.6	58.5	66.2	26.7	56.7	52.0
st1.MML	95.0	85.9	100.0	84.3	91.3	61.4	71.3	26.7	56.7	54.0
st1.bp	72.3±6.3	70.2±3.8	54.5±10.7	50.9±11.1	62.0±1.8	62.3±9.3	66.6±3.4	31.3±8.8	40.7±13.3	50.2±3.6
st1.qp	78.2±3.3	76.0±5.2	59.2±8.1	58.7±7.0	68.0±2.0	69.3±4.9	69.6±3.1	31.0±12.0	46.5±10.4	54.1±2.8
st1.cas	93.1±1.7	77.4±1.9	31.7±12.4	20.5±9.3	55.7±3.1	91.9±2.2	70.0±2.2	20.7±11.5	21.7±10.1	51.1±2.9
st1.cbp	85.2	72.3	64.1	64.6	71.6	80.0	68.6	36.7	53.3	59.7
st1.cqp	88.7	83.1	69.5	72.1	78.4	85.5	72.0	23.3	56.7	59.4
st1.ccas	93.9	77.8	29.0	26.2	56.7	93.3	70.6	6.7	26.7	49.3

Table 8-5: Classification results for st1 experiments (using two map ST feature extraction).

%	Training Set					Testing Set				
experiment	anterior	inferior	normal	cad	total	anterior	inferior	normal	cad	total
C1.knn-rmse	80.1	73.8	35.1	67.9	64.2	76.2	72.4	36.7	76.7	65.5
C1.knn-cc	77.2	75.3	29.8	77.1	64.8	80.4	73.4	16.7	90.0	65.1
C1.ncm-rmse	77.1	72.5	30.5	66.6	61.7	76.6	66.7	33.3	80.0	64.2
C1.ncm-cc	73.1	70.7	16.8	74.8	58.8	71.8	66.6	26.7	86.7	62.9

Table 8-6: Classification results for traditional techniques applied to problem 1.

8.1.3 Initial Observations

8.1.3.1 KLT feature extraction (E)

When comparing the performance of classifiers using the KLT feature extraction technique (Table 8-2) a number of initial observations can be made. Firstly, the best testing set classification result is achieved using the linear regression classifier (67.9% testing set patients classified correctly), followed by k-nearest neighbour classifier (64.2%) and the committee of cascade correlation networks (61.4%). The worst testing set classification results were obtained using C4.5 (47.1%) and MML (42.6%).

It is also noted that the committee-based neural network classifiers perform consistently better than the average performance of the individual networks:

- E1.bp – 55.5% (±2.3) combined in a committee achieves (E1.cbp) 60.5%,
- E1.qp - 55.5% (±2.3) combined in a committee (E1.cbp) achieves 60.5%, and
- E1.cas – 60.5% (±3.4%) combined in a committee (E1.ccas) achieves 61.4%.

It is also observed that all classifiers perform poorly when classifying normal patients in the testing set. On closer examination of the performance breakdowns (Appendix C) it is apparent that all classifiers consistently misclassify these normal patients as having coronary artery disease as was found in initial experiments (chapter 7).

8.1.3.2 Logical data reduction (L)

When the logical data reduction technique is used in place of the KLT feature extraction technique (Table 8-3) similar classification results are observed. For this set of experiments the best testing set results are achieved using the k-nearest neighbour (*L1.knn* – 64.1%) followed by the linear regression classifier (L1.linreg – 63.5%) and the committee of cascade correlation networks (63.6%). One major difference between these results (Table 8-3) and the KLT results (Table 8-2) is the classification performance of the inductive learning techniques. Both the C4.5 and MML testing set classification results improved significantly (*E1.C4.5* – 47.1%, *E1.MML* – 42.6%, compared with *L1.C4.5* – 54.7%, and *L1.MML* – 60.4%). As with the KLT results, the committee-based networks consistently perform better than the average performance of the individual networks, and all classifiers still misclassified a significant proportion of testing set normals as having coronary artery disease.

8.1.3.3 QRS logical data reduction (qrs)

Similar results are observed when using the *qrs* (Table 8-4) feature extraction techniques. When compared with the logical reduction technique results (Table 8-3) a number of observations can be made. Firstly, the linear regression and k-nearest neighbour classifiers perform better when using the *qrs* feature extraction technique (*L1.knn* – 64.1% and *L1.linreg* – 63.5%, compared with *qrs1.knn* – 64.5% and *qrs1.linreg* – 67.1%). Secondly, the committee MLPs trained using back-propagation (bp) achieved the same test set classification performance (*L1.cbp* and *qrs1.cbp* – 59.8%).

8.1.3.4 ST logical data reduction (st)

The results obtained when using the *st* (Table 8-5) feature extraction technique are not as good overall as those achieved with the other three feature extraction techniques. Some classifiers manage to achieve results comparable with those achieved using the qrs feature extraction technique (*st1.C4.5*, *st1.MML*, *st1.cbp* and *st1.cqp*), but apart from these the classification results were not as good.

8.1.3.5 Traditional BSM classification

The traditional BSM classification techniques (Table 8-6) performed comparatively well when applied to this classification problem. All four classification approaches achieved testing set classification performances comparable with the best results achieved in the other experiments. As was found in the other experiments, all these classifiers consistently misclassify normal patients suggesting that discriminating between normal and CAD patients is difficult.

8.1.4 Comparing Results

To provide a more complete comparison the results described in the previous section were ranked according to testing set classification performance (Table 8-7) providing a more encompassing understanding of the results. A summary of this ranking is also provided in Table 8-8. This table provides a quick reference to positioning of experiments within the ranking.

%	Training Set					Testing Set				
experiment	anterior	inferior	normal	cad	total	anterior	inferior	normal	cad	total
E1.linreg	89.9	89.1	61.8	57.7	74.6	83.0	78.5	30.0	80.0	67.9
qrs1.linreg	88.2	89.7	66.4	59.7	76.0	72.3	82.9	50.0	63.3	67.1
C1.knn-rmse	80.1	73.8	35.1	67.9	64.2	76.2	72.4	36.7	76.7	65.5
C1.knn-cc	77.2	75.3	29.8	77.1	64.8	80.4	73.4	16.7	90.0	65.1
qrs1.knn	96.8	98.5	95.4	98.0	97.2	74.4	70.2	36.7	76.7	64.5
E1.knn	97.6	97.9	97.7	99.3	98.1	76.6	66.7	33.3	80.0	64.2
C1.ncm-rmse	77.1	72.5	30.5	66.6	61.7	76.6	66.7	33.3	80.0	64.2
L1.knn	96.8	98.7	95.4	99.0	97.5	69.7	73.4	36.7	76.7	64.1
L1.ccas	97.9	91.8	66.4	34.1	72.5	90.2	77.3	40.0	46.7	63.6
L1.linreg	91.4	96.5	74.1	64.9	81.7	73.6	80.2	36.7	63.3	63.5
C1.ncm-cc	73.1	70.7	16.8	74.8	58.8	71.8	66.6	26.7	86.7	62.9
L1.cqp	71.1	83.8	47.3	74.4	69.2	55.0	71.9	40.0	83.3	62.5
E1.ccas	95.8	85.9	39.7	25.9	61.8	84.7	77.6	30.0	53.3	61.4
E1.cas	96.2±1.6	85.1±2.3	44.1±19.6	24.4±13.9	62.4±3.3	86.3±1.5	78.7±1.0	40.0±13.2	37.0±22.2	60.5±3.4
E1.cbp	91.1	89.9	75.6	71.2	81.9	76.1	76.0	26.7	63.3	60.5
L1.MML	92.2	92.9	100.0	83.9	92.3	72.8	85.4	13.3	70.0	60.4
L1.cbp	96.4	97.4	26.7	90.5	77.8	78.7	80.4	0.0	80.0	59.8
qrs1.cbp	92.5	93.9	61.1	78.7	81.5	77.8	78.0	13.3	70.0	59.8
st1.cbp	85.2	72.3	64.1	64.6	71.6	80.0	68.6	36.7	53.3	59.7
L1.cas	98.6±0.6	90.8±2.3	67.7±8.3	31.6±10.5	72.2±2.3	89.1±1.7	77.2±1.8	38.7±3.1	33.3±9.4	59.6±1.9
st1.cqp	88.7	83.1	69.5	72.1	78.4	85.5	72.0	23.3	56.7	59.4
qrs1.cqp	68.4	84.4	45.8	69.8	67.1	54.9	72.1	40.0	70.0	59.3
E1.cqp	85.2	88.8	42.0	47.9	66.0	77.0	83.0	30.0	43.3	58.3
st1.linreg	78.6	79.3	53.4	52.5	66.0	72.0	68.0	26.7	66.7	58.3
E1.bp	83.4±2.3	82.0±2.3	59.8±16.3	58.9±5.1	71.0±3.2	65.9±4.6	67.2±4.8	30.7±10.7	58.2±9.7	55.5±2.3
L1.C4.5	99.0	98.9	100.0	95.4	98.3	65.7	76.6	26.7	50.0	54.7
st1.qp	78.2±3.3	76.0±5.2	59.2±8.1	58.7±7.0	68.0±2.0	69.3±4.9	69.6±3.1	31.0±12.0	46.5±10.4	54.1±2.8
st1.MML	95.0	85.9	100.0	84.3	91.3	61.4	71.3	26.7	56.7	54.0
qrs1.MML	96.0	92.5	98.5	76.4	90.8	74.8	63.5	36.7	36.7	52.9
qrs1.ccas	96.3	80.3	56.5	23.0	64.0	84.4	62.5	23.3	40.0	52.6
st1.C4.5	97.8	96.7	100.0	96.1	97.6	58.5	66.2	26.7	56.7	52.0
L1.qp	65.8±8.9	72.4±9.0	33.2±23.9	58.9±12.9	57.5±8.4	53.4±8.2	62.3±7.9	25.0±17.6	65.2±15.3	51.5±6.7
qrs1.C4.5	98.5	97.7	100.0	96.1	98.1	74.0	61.0	30.0	40.0	51.2
E1.qp	77.0±6.0	84.1±5.5	31.0±23.5	29.5±23.3	55.4±8.7	67.5±6.1	79.1±8.5	25.3±20.2	32.5±26.5	51.1±7.4
st1.cas	93.1±1.7	77.4±1.9	31.7±12.4	20.5±9.3	55.7±3.1	91.9±2.2	70.0±2.2	20.7±11.5	21.7±10.1	51.1±2.9
L1.bp	90.5±2.6	90.0±2.5	7.8±23.0	56.1±32.5	61.1±9.9	74.2±6.3	71.9±6.6	3.2±9.6	53.3±32.6	50.6±6.0
qrs1.cas	96.5±0.8	77.5±3.8	54.0±18.6	19.7±7.6	61.9±5.8	86.6±2.3	59.0±1.6	35.3±12.8	20.0±7.1	50.2±3.3
st1.bp	72.3±6.3	70.2±3.8	54.5±10.7	50.9±11.1	62.0±1.8	62.3±9.3	66.6±3.4	31.3±8.8	40.7±13.3	50.2±3.6
st1.ccas	93.9	77.8	29.0	26.2	56.7	93.3	70.6	6.7	26.7	49.3
qrs1.qp	67.2±5.5	70.4±9.7	29.4±23.2	46.5±20.4	53.4±9.8	55.6±7.2	61.4±9.0	22.8±21.2	54.2±20.9	48.5±7.4
qrs1.bp	86.1±2.8	85.0±5.1	29.7±36.5	41.3±34.2	60.5±11.3	70.6±7.6	67.0±11.4	15.3±20.9	38.8±32.9	47.9±5.6
st1.knn	95.2	97.9	96.2	99.7	97.2	53.0	71.2	0.0	66.7	47.7
E1.C4.5	98.5	98.2	100.0	96.4	98.3	60.2	58.3	26.7	43.3	47.1
E1.MML	82.2	86.0	67.9	56.4	73.1	70.6	46.3	16.7	36.7	42.6

Table 8-7: Ranking of experiments for problem 1.
(Ranked with respect to percentage of testing set patients classified correctly)

Experiment	Ranking	% Correct (testing set)	E	L	qrs	st	C	*delta
E1.linreg	1	67.9	linreg					0.0
qrs1.linreg	2	67.1			linreg			0.8
C1.knn-rmse	3	65.5					knn-rmse	2.4
C1.knn-cc	4	65.1					knn-cc	2.8
qrs1.knn	5	64.5			knn			3.4
E1.knn	6	64.2	knn					3.7
C1.ncm-rmse	7	64.2					ncm-rmse	3.8
L1.knn	8	64.1		knn				3.8
L1.ccas	9	63.6		ccas				4.3
L1.linreg	10	63.5		linreg				4.4
C1.ncm-cc	11	62.9					ncm-cc	5.0
L1.cqp	12	62.5		cqp				5.4
E1.ccas	13	61.4	ccas					6.5
E1.cas	14	60.5±3.4	cas					7.4
E1.cbp	15	60.5	cbp					7.4
L1.MML	16	60.4		MML				7.5
L1.cbp	17	59.8		cbp				8.1
qrs1.cbp	18	59.8			cbp			8.1
st1.cbp	19	59.7				cbp		8.2
L1.cas	20	59.6±1.9		cas				8.3
st1.cqp	21	59.4				cqp		8.5
qrs1.cqp	22	59.3			cqp			8.6
E1.cqp	23	58.3	cqp					9.6
st1.linreg	24	58.3				linreg		9.6
E1.bp	25	55.5±2.3	bp					12.4
L1.C4.5	26	54.7		C4.5				13.2
st1.qp	27	54.1±2.8				qp		13.8
st1.MML	28	54.0				MML		13.9
qrs1.MML	29	52.9			MML			15.0
qrs1.ccas	30	52.6			ccas			15.3
st1.C4.5	31	52.0				C4.5		15.9
L1.qp	32	51.5±6.7		qp				16.4
qrs1.C4.5	33	51.2			C4.5			16.7
E1.qp	34	51.1±7.4	qp					16.8
st1.cas	35	51.1±2.9				cas		16.8
L1.bp	36	50.6±6.0		bp				17.3
qrs1.cas	37	50.2±3.3			cas			17.7
st1.bp	38	50.2±3.6				bp		17.7
st1.ccas	39	49.3				ccas		18.6
qrs1.qp	40	48.5±7.4			qp			19.4
qrs1.bp	41	47.9±5.6			bp			20.0
st1.knn	42	47.7				knn		20.2
E1.C4.5	43	47.1	C4.5					20.8
E1.MML	44	42.6	MML					25.3

Table 8-8: Summary of ranked results for problem 1.
(* delta – difference between experiment performance and best result – E1.linreg)

8.1.5 Discussion

The following discussion will focus on the results presented in Table 8-8. There are a number of key features of these results, which are presented below:

8.1.5.1 Best Classification Results

The best classification result was achieved using a linear regression classifier in conjunction with the KLT feature extraction technique (E1.linreg – 67.9% correct). It is also noted that the linear regression and k-nearest neighbour classifiers are all ranked in the top ten results (and within 5% of the best result) except when used in conjunction with the *st* feature extraction technique.

It is also noted that all four traditional BSM classification techniques are ranked in the top ten experiments and achieved classification results on the testing set within 5% of the best classification result.

With respect to neural network classifiers only one neural network technique is ranked in the top 10 experiments (L1.ccas – ranked 9th). This feature of the ranked results would suggest that the neural network techniques considered do not perform as well as some more traditional and alternative classification techniques.

These results would suggest that linear discriminant functions training using linear regression are most suited to this classification problem (used in conjunction with the *E* or *qrs* feature extraction techniques). It is also noted that this approach is also better than the traditional BSM classification techniques, although it is worth noting that the k-nearest neighbour classifiers perform as well as the traditional BSM classification techniques (when used in conjunction with the *E*, *L* or *qrs* feature extraction techniques).

8.1.5.2 Neural Network Results

The neural network techniques did not perform as well as linear regression or k-nearest neighbour classifiers (apart from *L1.ccas*). Despite this result a number of observations can be made.

All committee-based MLPs trained with back-propagation and quickprop perform better than the corresponding individual MLP networks. All committee-based MLPs achieve classification results within 10% of the best classification result, but are not ranked in the top ten results. This result suggests that the use of committees of MLP networks will consistently perform better than individual neural network classifiers using the same architecture. This supports previous claims in the literature (Vamplew and Adams 1993; Baxt 1993) and is a useful result to note.

It is also noted that individual cascade-correlation classifiers produce comparable results to MLPs when used in conjunction with the *E* and *L* feature extraction techniques. It is also observed that committees of cascade-correction networks perform better than the individual cascade-correction networks. Interestingly, it is noted that cascade correlation networks perform poorly when used in conjunction with *qrs* and *st* feature extraction techniques.

8.1.5.3 Inductive Learning Results

The inductive learning techniques C4.5 and MML perform comparatively poorly, but perform slightly better when used in conjunction with the *L* feature extraction technique.

8.1.5.4 Feature Extraction Techniques

When comparing feature extraction techniques the *E*, *L* and *qrs* techniques would appear to provide similar performance. Experiments using the *st* feature extraction technique appear to consistently perform worse than the when using the other techniques.

Selecting the best feature extraction technique is difficult, as different classifiers appear to perform better with different feature extraction techniques. For example, *linreg* performs best when using *E* feature extraction, *knn* performs best when using *qrs* feature extraction, and *ccas* performs best when using *L* feature extraction.

This aside it is interesting to note the poor performance of the *st* technique. This may suggest that key discriminating features associated with this classification problem are located in the QRS aspects of the BSM data. This should not suggest that the *st* features are not important, and it would be pertinent to note that the classifiers using the *L* feature extraction technique (apart from knn and linreg) perform better than those using the *qrs* feature extraction technique. This would suggest that although the *qrs* feature provides good discrimination, the inclusion of the *st* features improves discrimination, and therefore the *L* feature extraction technique will allow improved discrimination over the qrs and st techniques.

Whether KLT feature extraction (*E*) is better than the logical data reduction technique (*L*) it is difficult to determine without further analysis. It is important to note that the *E1.linreg* experiment produced the best result, and as such would suggest that the combination would be best suited to this classification problem.

8.1.5.5 Final Comments

When comparing the results of these experiments with the original experiment in chapter 7 (E1.bp – ranked 25th) it is clear that a number of approaches perform significantly better. In particular, committee-based neural networks (*cbp*, *cqp* and *ccas*), k-nearest neighbour *(knn)* and linear regression (*linreg*) classifiers all perform consistently better than the orignial *E1.bp* experiment, as do the traditional BSM classification techniques.

Despite this improvement in classification performance it is clear that all of the classifiers in this experiment still have difficultly discriminating between normal and CAD patients. On examination of the results in Table 8-7 it is noted that all classifiers perform poorly when classifying normal patients in the testing set. On closer examination of the classification breakdowns in appendix C (sections C.1 to C.44) it is noted that the classifiers tend to incorrectly classify normal patients in the testing set as having coronary artery disease. This outcome is consistent with the results obtained in chapter 7.

On closer examination of the results for the top two classifiers (*E1.linreg* – section C.2 and *qrs1.linreg* – section C.32) a number of observations are made.

As just mentioned, although these classifiers perform significantly better than the original neural network classifier tested (*E1.bp*), both classifiers still incorrectly classify a large proportion of normal patients in the testing set as having coronary artery disease (*E1.linreg* – 63.3%, *qrs1.linreg* – 43.3%). Thus these classifiers would appear to still have difficulty solving this particular problem.

As highlighted in chapter 7, the *E1.bp* classifiers exhibited a trend in the classification of infarct sub-classes: the MLPs performed best when classifying 48 hour infarcts, followed by 12 hour infarcts, followed by 6 hour infarcts, and perform least effectively when classifying follow-up patients. Similar trends are observed in the *E1.linreg* and *qrs1.linreg* classification results but the precise ranking is not so clear (Tables 8-9 and 8-10).

Class	%Correctly Classified
anterior.12	96.6
inferior.48	93.3
anterior.6	86.2
inferior.6	80.0
inferior.12	80.0
anterior.48	79.3
anterior.fu	70.0
inferior.fu	60.7

Table 8-9: Ranked testing set results for infarcts – experiment E1.linreg

Class	%Correctly Classified
inferior.48	90.0
inferior.6	86.7
inferior.12	83.3
anterior.12	79.3
anterior.6	72.4
anterior.48	72.4
inferior.fu	71.4
anterior.fu	65.0

Table 8-10: Ranked testing set results for infarcts – experiment qrs1.linreg

It is clear that both classifiers have the most difficulty classifying follow-up patients. Apart from this observation it is difficult to draw any firm conclusions about the ranking of the other sub-classes. One observation worth noting is that these classifiers do not appear to have the same difficulty classifying 6 hour infarcts, as was observed in the *E1.bp* experiments.

In conclusion, the best results (*E1.linreg*) in this set of experiments are a significant improvement over the original results (*E1.bp*). This aside, the *E1.linreg* classifier still has difficulty separating normal and CAD patients.

8.2 Problem 2

Having applied the various classification techniques to *problem 1* (section 8.1) the same techniques were then applied to *problem 2*. As has been stated previously, the aim in *problem 2* is to construct a classifier to assign patients to one of three classes: anterior infarction, inferior infarction, or normal depending on the predicted state of the patient's heart.

All forty-four classification approaches were applied to this problem using the same training parameters as described in section 8.1.1. The only difference in classifier construction was the number of outputs (three instead of four). The MLP architecture used for *bp* and *qp* experiments was similar to that used in *problem 1*; 72:72:3 (*E2.bp, E2.qp*), 192:120:3 (*L2.bp, L2.qp*), 128:120:3 (*qrs2.bp, qrs2.qp*) and 64:64:3 (*st2.bp, st2.qp*).

The results of these experiments are presented and discussed in the following sections. The detailed classification breakdowns are provided in appendix C for reference.

8.2.1 Results

The classification results were analysed and prepared using the CTAS analysis tool *performance* and *breakdown* (see appendix B). The classification results for these experiments are summarised in Tables 8-11, 8-12, 8-13, 8-14 and 8-15. Note that where a number of training runs were performed in an experiment (*bp*, *qp*, and *cas* – 20 networks were trained in each case) the averaged results are presented. The more detailed classification breakdown results are provided in appendix C for reference and will be referred in the discussion.

%	Training Set				Testing Set			
experiment	anterior	inferior	normal	total	anterior	inferior	normal	total
E2.knn	96.0	97.5	100.0	97.9	70.7	59.1	93.3	74.3
E2.linreg	87.6	91.5	90.1	89.7	81.7	80.3	90.0	84.0
E2.C4.5	97.7	97.5	100.0	98.4	72.4	70.1	63.3	68.6
E2.MML	88.8	93.9	62.6	81.8	77.4	71.9	33.3	60.9
E2.bp	85.2±3.6	88.9±1.7	77.7±18.3	84.0±6.8	71.6±3.7	77.6±4.2	59.7±15.5	69.6±5.2
E2.qp	76.0±4.0	86.9±2.1	78.9±3.6	80.6±1.4	63.9±4.5	79.4±3.3	71.2±7.8	71.5±3.2
E2.cas	97.7±0.9	90.8±2.1	81.4±1.6	90.0±0.8	89.1±1.2	82.8±0.9	64.3±4.0	78.7±1.4
E2.cbp	92.1	94.5	84.7	90.4	74.5	84.8	66.7	75.3
E2.cqp	81.8	90.2	84.7	85.6	70.2	83.0	80.0	77.7
E2.ccas	96.4	92.1	85.5	91.3	88.9	82.0	63.3	78.1

Table 8-11: Classification results for E2 experiments.

%	Training Set				Testing Set			
experiment	anterior	inferior	normal	total	anterior	inferior	normal	total
L2.knn	96.8	98.4	99.2	98.1	67.6	69.1	96.7	77.8
L2.linreg	96.2	97.8	97.7	97.2	75.7	82.8	73.3	77.3
L2.C4.5	99.0	99.6	100.0	99.5	75.8	87.7	60.0	74.5
L2.MML	96.2	98.6	100.0	98.3	72.4	85.2	63.3	73.6
L2.bp	91.3±1.6	93.4±1.3	88.6±2.7	91.1±1.1	71.6±5.0	75.5±3.7	59.8±9.3	69.0±3.1
L2.qp	78.6±3.9	86.5±3.2	85.4±3.7	83.5±2.4	65.0±6.0	76.0±3.5	68.2±7.6	69.7±2.8
L2.cas	100.0±0.1	98.9±0.7	99.2±1.1	99.4±0.6	93.2±0.8	80.3±1.8	60.7±2.0	78.0±1.0
L2.cbp	98.0	98.3	98.5	98.3	76.6	82.9	70.0	76.5
L2.cqp	83.6	91.5	90.1	88.4	70.7	80.3	76.7	75.9
L2.ccas	100.0	99.5	100.0	99.8	92.4	82.7	66.7	80.6

Table 8-12: Classification results for L2 experiments.

%	Training Set				Testing Set			
experiment	anterior	inferior	normal	total	anterior	inferior	normal	total
qrs2.knn	96.5	97.4	99.2	97.7	69.7	65.9	96.7	77.4
qrs2.linreg	91.2	93.2	95.4	93.3	73.5	82.1	76.7	77.4
qrs2.C4.5	98.5	98.9	100.0	99.1	74.1	77.9	63.3	71.8
qrs2.MML	97.4	98.2	99.2	98.3	77.4	75.3	60.0	70.9
qrs2.bp	90.3±2.3	89.3±1.8	88.1±2.9	89.2±1.2	71.6±4.8	69.1±5.2	62.2±8.6	67.6±3.3
qrs2.qp	72.7±4.4	81.5±2.9	81.9±3.7	78.7±2.4	58.7±4.7	69.8±5.2	63.3±8.2	63.9±4.2
qrs2.cas	97.2±0.4	87.6±1.5	89.8±3.3	91.5±1.3	82.6±1.1	69.3±2.4	65.7±1.5	72.5±0.8
qrs2.cbp	96.1	95.5	96.2	95.9	74.9	75.5	70.0	73.5
qrs2.cqp	77.3	85.6	88.6	83.8	63.5	74.5	73.3	70.4
qrs2.ccas	97.0	92.2	93.9	94.4	80.4	75.2	66.7	74.1

Table 8-13: Classification results for qrs2 experiments.

%	Training Set				Testing Set			
experiment	anterior	inferior	normal	total	anterior	inferior	normal	total
st2.knn	94.5	97.1	99.2	96.9	47.4	68.6	90.0	68.7
st2.linreg	79.1	76.7	86.3	80.7	69.0	63.8	80.0	70.9
st2.C4.5	99.0	96.9	100.0	98.6	65.7	73.9	33.3	57.7
st2.MML	98.5	96.5	100.0	98.3	61.0	79.2	43.3	61.2
st2.bp	75.7±7.3	74.5±3.9	72.0±8.7	74.1±1.4	63.0±8.1	69.1±3.0	45.2±11.3	59.1±3.6
st2.qp	79.8±4.9	77.7±4.0	76.2±7.6	77.9±3.6	70.2±5.7	71.5±2.9	49.2±12.5	63.6±4.1
st2.cas	94.2±2.2	82.1±1.3	69.6±4.4	82.0±1.5	88.8±2.7	71.1±1.9	47.3±5.7	69.1±1.7
st2.cbp	88.2	78.0	80.9	82.4	77.5	72.0	50.0	66.5
st2.cqp	91.4	85.3	85.5	87.4	80.9	75.4	46.7	67.7
st2.ccas	93.5	83.5	76.3	84.5	88.7	70.7	63.3	74.2

Table 8-14: Classification results for st2 experiments.

%	Training Set				Testing Set			
experiment	anterior	inferior	normal	total	anterior	inferior	normal	total
C2.knn-rmse	83.4	88.5	76.3	82.7	83.9	88.1	80.0	84.0
C2.knn-cc	81.1	86.1	77.1	81.4	86.9	88.0	73.3	82.7
C2.ncm-cc	81.0	86.1	72.5	79.9	82.1	83.7	73.3	79.7
C2.ncm-rmse	81.7	84.8	72.5	79.7	81.7	82.1	70.0	77.9

Table 8-15: Classification results for traditional techniques applied to problem 2.

8.2.2 Ranking Results

To provide a more complete comparison, the results described in the previous section were ranked according to testing set classification performance (Table 8-16) providing a more encompassing understanding of the results. The results were ranked in descending order, according to the testing set classification performance achieved. A summary in Table 8-17 provides a quick reference to positioning of experiments within this ranking.

%	Training Set				Testing Set			
experiment	anterior	inferior	normal	total	anterior	inferior	normal	total
E2.linreg	87.6	91.5	90.1	89.7	81.7	80.3	90.0	84.0
C2.knn-rmse	83.4	88.5	76.3	82.7	83.9	88.1	80.0	84.0
C2.knn-cc	81.1	86.1	77.1	81.4	86.9	88.0	73.3	82.7
L2.ccas	100.0	99.5	100.0	99.8	92.4	82.7	66.7	80.6
C2.ncm-cc	81.0	86.1	72.5	79.9	82.1	83.7	73.3	79.7
E2.cas	97.7±0.9	90.8±2.1	81.4±1.6	90.0±0.8	89.1±1.2	82.8±0.9	64.3±4.0	78.7±1.4
E2.ccas	96.4	92.1	85.5	91.3	88.9	82.0	63.3	78.1
L2.cas	100.0±0.1	98.9±0.7	99.2±1.1	99.4±0.6	93.2±0.8	80.3±1.8	60.7±2.0	78.0±1.0
C2.ncm-rmse	81.7	84.8	72.5	79.7	81.7	82.1	70.0	77.9
L2.knn	96.8	98.4	99.2	98.1	67.6	69.1	96.7	77.8
E2.cqp	81.8	90.2	84.7	85.6	70.2	83.0	80.0	77.7
qrs2.knn	96.5	97.4	99.2	97.7	69.7	65.9	96.7	77.4
qrs2.linreg	91.2	93.2	95.4	93.3	73.5	82.1	76.7	77.4
L2.linreg	96.2	97.8	97.7	97.2	75.7	82.8	73.3	77.3
L2.cbp	98.0	98.3	98.5	98.3	76.6	82.9	70.0	76.5
L2.cqp	83.6	91.5	90.1	88.4	70.7	80.3	76.7	75.9
E2.cbp	92.1	94.5	84.7	90.4	74.5	84.8	66.7	75.3
L2.C4.5	99.0	99.6	100.0	99.5	75.8	87.7	60.0	74.5
E2.knn	96.0	97.5	100.0	97.9	70.7	59.1	93.3	74.3
st2.ccas	93.5	83.5	76.3	84.5	88.7	70.7	63.3	74.2
qrs2.ccas	97.0	92.2	93.9	94.4	80.4	75.2	66.7	74.1
L2.MML	96.2	98.6	100.0	98.3	72.4	85.2	63.3	73.6
qrs2.cbp	96.1	95.5	96.2	95.9	74.9	75.5	70.0	73.5
qrs2.cas	97.2±0.4	87.6±1.5	89.8±3.3	91.5±1.3	82.6±1.1	69.3±2.4	65.7±1.5	72.5±0.8
qrs2.C4.5	98.5	98.9	100.0	99.1	74.1	77.9	63.3	71.8
E2.qp	76.0±4.0	86.9±2.1	78.9±3.6	80.6±1.4	63.9±4.5	79.4±3.3	71.2±7.8	71.5±3.2
st2.linreg	79.1	76.7	86.3	80.7	69.0	63.8	80.0	70.9
qrs2.MML	97.4	98.2	99.2	98.3	77.4	75.3	60.0	70.9
qrs2.cqp	77.3	85.6	88.6	83.8	63.5	74.5	73.3	70.4
L2.qp	78.6±3.9	86.5±3.2	85.4±3.7	83.5±2.4	65.0±6.0	76.0±3.5	68.2±7.6	69.7±2.8
E2.bp	85.2±3.6	88.9±1.7	77.7±18.3	84.0±6.8	71.6±3.7	77.6±4.2	59.7±15.5	69.6±5.2
st2.cas	94.2±2.2	82.1±1.3	69.6±4.4	82.0±1.5	88.8±2.7	71.1±1.9	47.3±5.7	69.1±1.7
L2.bp	91.3±1.6	93.4±1.3	88.6±2.7	91.1±1.1	71.6±5.0	75.5±3.7	59.8±9.3	69.0±3.1
st2.knn	94.5	97.1	99.2	96.9	47.4	68.6	90.0	68.7
E2.C4.5	97.7	97.5	100.0	98.4	72.4	70.1	63.3	68.6
st2.cqp	91.4	85.3	85.5	87.4	80.9	75.4	46.7	67.7
qrs2.bp	90.3±2.3	89.3±1.8	88.1±2.9	89.2±1.2	71.6±4.8	69.1±5.2	62.2±8.6	67.6±3.3
st2.cbp	88.2	78.0	80.9	82.4	77.5	72.0	50.0	66.5
qrs2.qp	72.7±4.4	81.5±2.9	81.9±3.7	78.7±2.4	58.7±4.7	69.8±5.2	63.3±8.2	63.9±4.2
st2.qp	79.8±4.9	77.7±4.0	76.2±7.6	77.9±3.6	70.2±5.7	71.5±2.9	49.2±12.5	63.6±4.1
st2.MML	98.5	96.5	100.0	98.3	61.0	79.2	43.3	61.2
E2.MML	88.8	93.9	62.6	81.8	77.4	71.9	33.3	60.9
st2.bp	75.7±7.3	74.5±3.9	72.0±8.7	74.1±1.4	63.0±8.1	69.1±3.0	45.2±11.3	59.1±3.6
st2.C4.5	99.0	96.9	100.0	98.6	65.7	73.9	33.3	57.7

Table 8-16: Ranking of experiments applied to problem 2.
(Ranked with respect to percentage of testing set patients classified correctly)

Experiment	Ranking	%Correct (testing set)	E	L	qrs	st	C	*delta
E2.linreg	1	84.0	linreg					0.0
C2.knn-rmse	2	84.0					knn-rmse	0.0
C2.knn-cc	3	82.7					knn-cc	1.3
L2.ccas	4	80.6		ccas				3.4
C2.ncm-cc	5	79.7					ncm-cc	4.3
E2.cas	6	78.7±1.4	cas					5.3
E2.ccas	7	78.1	ccas					5.9
L2.cas	8	78.0±1.0		cas				6.0
C2.ncm-rmse	9	77.9					ncm-rmse	6.1
L2.knn	10	77.8		knn				6.2
E2.cqp	11	77.7	cqp					6.3
qrs2.knn	12	77.4			knn			6.6
qrs2.linreg	13	77.4			linreg			6.6
L2.linreg	14	77.3		linreg				6.7
L2.cbp	15	76.5		cbp				7.5
L2.cqp	16	75.9		cqp				8.1
E2.cbp	17	75.3	cbp					8.7
L2.C4.5	18	74.5		C4.5				9.5
E2.knn	19	74.3	knn					9.7
st2.ccas	20	74.2				ccas		9.8
qrs2.ccas	21	74.1			ccas			9.9
L2.MML	22	73.6		MML				10.4
qrs2.cbp	23	73.5			cbp			10.5
qrs2.cas	24	72.5±0.8			cas			11.5
qrs2.C4.5	25	71.8			C4.5			12.2
E2.qp	26	71.5±3.2	qp					12.5
st2.linreg	27	70.9				linreg		13.1
qrs2.MML	28	70.9			MML			13.1
qrs2.cqp	29	70.4			cqp			13.6
L2.qp	30	69.7±2.8		qp				14.3
E2.bp	31	69.6±5.2	bp					14.4
st2.cas	32	69.1±1.7				cas		14.9
L2.bp	33	69.0±3.1		bp				15.0
st2.knn	34	68.7				knn		15.3
E2.C4.5	35	68.6	C4.5					15.4
st2.cqp	36	67.7				cqp		16.3
qrs2.bp	37	67.6±3.3			bp			16.4
st2.cbp	38	66.5				cbp		17.5
qrs2.qp	39	63.9±4.2			qp			20.1
st2.qp	40	63.6±4.1				qp		20.4
st2.MML	41	61.2				MML		22.8
E2.MML	42	60.9	MML					23.1
st2.bp	43	59.1±3.6				bp		24.9
st2.C4.5	44	57.7				C4.5		26.3

Table 8-17: Summary of ranked results for problem 2.
(* delta – difference between experiment performance and best result – E2.linreg)

8.2.3 Discussion

The following discussion will focus on the results presented in Table 8-17.

When comparing the results in Table 8-16 with those obtained for *problem 1* (Table 8-7) one clear improvement is noted. The classification of normal patients in the testing set is significantly improved. The distinct features of these results are discussed below.

8.2.3.1 Best Classification Results

The best classification result was achieved by the *E2.linreg* experiment (84%) and by the *C2.knn-rmse* experiment (also achieving 84% on the testing data set). All four traditional BSM classification techniques are ranked in the top ten experiments and achieved classification results on the testing set within 6.1% (*C1.ncm-rmse*) of the best classification result.

With respect to neural network classifiers the only neural network technique ranked in the top ten experiments are cascade correlation networks (*E2.cas*, *E2.ccas*, *L2.cas*, *L2.ccas*). However, these experiments still produce results inferior to the top three techniques (*E2.linreg*, *C2.knn-rmse*, and *C2.knn-cc*).

8.2.3.2 Neural Network Results

As found for *problem 1*, all committee-based MLPs trained with back-propagation and quickprop perform better than the corresponding individual MLP networks. However, the committee-based MLP results were more spread through the ranking (compared with *problem 1*). In particular it was noted that MLP classifier performed poorly when used in conjunction with the *qrs* and *st* feature extraction techniques.

8.2.3.3 Inductive Learning Results

Inductive learning techniques C4.5 and MML perform comparatively poorly, but perform slightly better when used in conjunction with the *L* feature extraction technique.

8.2.3.4 Feature Extraction Techniques

It is clear from this ranking that the *E* and *L* feature extraction techniques perform consistantly better than the *qrs* and *st* feature extraction techniques. Similar to the trends observed with *problem 1* these results would suggest yet again that discriminating features are present in both the QRS and ST aspects of the BSM data.

8.2.3.5 Final Comments

These results are a significant improvement on the results observed in the initial experiment (E2.bp – ranked 31st) and clearly suggest that linear regression techniques (*E2.linreg*) and traditional BSM classification techniques perform best

when applied to this classification problem (C2.knn-rmse, C2.knn-cc). Yet again, these results would suggest that neural network classification techniques do not provide any improvements and in fact do not perform as well.

The best feature extraction technique would appear to be the KLT technique (as supported by the E2.linreg result). However, the logical data reduction technique produces similar results when used in conjunction with the *cas* and *ccas* classifiers. Comparatively, classifiers using the *qrs* and *st* feature extraction techniques do not appear to perform as well. This would suggest that discriminating features are lost when using these techniques and that discriminating information is to be found in both the QRS and ST segments.

The need to use information in both the QRS and ST features of the BSM data is not an unusual outcome as it is clear from the literature (see chapter 4) that changes in both the QRS and ST features will be observed when myocardial infarction occurs.

In conclusion, there are number of observations worth noting when reflecting on the classification breakdown for the best classifier (E2.linreg – see appendix C, section C.42). With respect to the classification of infarction sub-classes, the *E2.linreg* classifier performs least effectively when classifying follow-up patients which was the same outcome observed in *problem 1*. This is clearly a distinct trend in the classification profile and would suggest that discriminating between these patients is difficult.

A second observation that is made relates to the classification of CAD patients. It was observed in the initial experiments (E2.bp) that although a significant proportion of patients were classified as normal, a large proportion were classified as having inferior infarcts. In the *E2.linreg* experiment a similar outcome is observed. The CAD patients, although not used during training, were presented to the *E2.linreg* classifier after training. The majority of CAD patients are classified as normal (training set – 70.8%, testing set – 76.7%). This is to be expected as these patients are clearly very similar to normal patients (as noted when exploring *problem 1*). What is interesting to note is that a significant proportion of the remaining CAD patients are classified as having inferior infarcts (training set 19%, testing set – 23.3%). This supports the observations made in chapter 7 and would suggest that some CAD patients exhibit features that are similar to inferior infarct patients.

Having explored the features of problem 2, problem 3 was considered to see if the classification could be improved further by eliminating the follow-up patients from the training phase.

8.3 Problem 3

Having applied the various classification techniques to *problem 2* (section 8.2) the same techniques were then applied to *problem 3*. As has been stated previously, the aim in *problem 3* is to construct a classifier to assign patients to one of three classes: anterior infarction, inferior infarction, or normal depending on the predicted state of the patient's heart. In this particular problem, infarcted follow-up patients were not used during training in an attempt to improve the overall classification performance of classifiers.

All forty-four classification approaches were applied to this problem using the same training parameters as described in section 8.1.1. The MLP architecture used for *bp* and *qp* experiments was similar to that used in *problem 1*; 72:72:3 (*E3.bp, E3.qp*), 192:120:3 (*L3.bp, L3.qp*), 128:120:3 (*qrs3.bp, qrs3.qp*) and 64:64:3 (*st3.bp, st3.qp*).

The results of these experiments are presented and discussed in the following sections. The detailed classification breakdowns are provided in appendix C for reference.

8.3.1 Results

The classification results were analysed and prepared using the CTAS analysis tool *performance* and *breakdown* (see appendix B). The classification results for these experiments are summarised in Tables 8-18, 8-19, 8-20, 8-21 and 8-22. Note that where a number of training runs were performed in an experiment (*bp*, *qp*, and *cas* – 20 networks were trained in each case) the averaged results are presented. The more detailed classification breakdown results are provided in appendix C for reference and will be referred in the discussion.

%	Training Set				Testing Set			
experiment	anterior	inferior	normal	total	anterior	inferior	normal	total
E3.knn	91.8	90.7	100.0	94.2	81.7	76.9	76.7	78.4
E3.linreg	81.7	86.8	89.3	85.9	79.3	78.5	86.7	81.5
E3.C4.5	88.6	85.5	100.0	91.4	73.6	55.7	56.7	62.0
E3.MML	80.6	81.8	70.2	77.5	75.8	69.3	53.3	66.2
E3.bp	82.4±2.5	85.6±1.7	77.5±18.1	81.8±6.4	70.9±4.0	75.0±4.9	63.8±17.1	69.9±5.4
E3.qp	76.7±5.0	82.9±3.6	81.6±5.0	80.4±2.6	66.7±4.4	75.8±4.7	75.0±9.1	72.5±3.1
E3.cas	93.3±0.7	88.0±0.8	87.3±1.9	89.5±0.7	91.1±1.6	79.7±2.6	64.7±5.4	78.5±1.9
E3.cbp	89.3	89.5	84.7	87.9	74.5	82.1	73.3	76.6
E3.cqp	82.6	86.7	84.7	84.7	73.6	81.3	83.3	79.4
E3.ccas	93.3	87.1	87.0	89.1	88.6	78.3	73.3	80.1

Table 8-18: Classification results for E3 experiments.

%	Training Set				Testing Set			
experiment	anterior	inferior	normal	total	anterior	inferior	normal	total
L3.knn	92.0	90.6	99.2	93.9	78.7	79.5	76.7	78.3
L3.linreg	86.6	89.1	97.0	90.9	72.8	71.6	73.3	72.6
L3.C4.5	92.0	84.4	100.0	92.1	79.7	77.3	60.0	72.3
L3.MML	85.7	85.7	100.0	90.5	63.2	77.4	73.3	71.3
L3.bp	85.2±1.8	84.5±2.4	88.9±3.0	86.2±1.5	76.0±3.7	69.9±3.8	64.3±8.5	70.1±3.1
L3.qp	78.8±4.4	82.7±1.9	89.8±2.9	83.8±2.4	69.8±6.1	70.9±3.6	71.0±6.0	70.6±3.3
L3.cas	93.6±1.2	83.7±0.7	99.8±0.3	92.4±0.5	91.7±2.1	72.9±2.2	64.7±4.8	76.4±1.6
L3.cbp	90.8	88.0	99.2	92.7	81.7	76.0	76.7	78.1
L3.cqp	82.1	86.8	97.0	88.6	76.6	78.6	83.3	79.5
L3.ccas	92.3	84.7	100.0	92.3	89.4	74.7	70.0	78.1

Table 8-19: Classification results for L3 experiments.

%	Training Set				Testing Set			
experiment	anterior	inferior	normal	total	anterior	inferior	normal	total
qrs3.knn	92.2	93.4	99.2	95.0	80.0	82.2	70.0	77.4
qrs3.linreg	84.2	88.6	96.2	89.7	74.6	70.0	76.7	73.8
qrs3.C4.5	91.3	90.9	100.0	94.1	73.6	78.1	70.0	73.9
qrs3.MML	89.3	87.6	99.2	92.1	84.2	77.1	40.0	67.1
qrs3.bp	84.3±2.5	83.4±2.3	89.0±3.3	85.6±1.6	71.6±4.3	65.3±4.2	62.0±8.9	66.3±3.3
qrs3.qp	71.9±3.7	77.8±4.6	83.0±4.0	77.6±2.4	58.9±6.2	64.9±5.6	64.8±8.3	62.8±3.7
qrs3.cas	92.3±0.5	77.6±2.5	95.3±2.2	88.4±1.4	83.9±1.1	52.8±2.0	64.3±4.2	67.0±1.4
qrs3.cbp	90.5	89.5	97.0	92.3	80.3	68.5	73.3	74.0
qrs3.cqp	75.2	82.9	89.3	82.5	60.5	67.7	73.3	67.2
qrs3.ccas	90.0	82.8	97.7	90.2	83.9	59.9	66.7	70.1

Table 8-20: Classification results for qrs3 experiments.

%	Training Set				Testing Set			
experiment	anterior	inferior	normal	total	anterior	inferior	normal	total
st3.knn	84.4	79.1	97.7	87.1	69.6	73.6	73.3	72.2
st3.linreg	69.3	70.0	87.0	75.5	68.2	58.6	80.0	68.9
st3.C4.5	80.5	81.8	100.0	87.4	60.6	71.1	53.3	61.7
st3.MML	76.6	78.5	98.5	84.5	60.6	71.3	56.7	62.8
st3.bp	72.6±5.6	70.0±4.5	70.0±4.8	70.9±2.4	64.8±8.4	67.4±3.3	45.2±11.4	59.1±3.7
st3.qp	74.0±4.3	73.2±4.1	73.5±10.1	73.6±3.7	67.9±5.9	68.8±2.3	49.5±13.6	62.1±3.7
st3.cas	85.6±1.3	77.7±1.8	69.2±4.3	77.5±1.1	87.6±2.3	71.3±1.6	41.7±4.8	66.8±1.7
st3.cbp	82.6	73.5	80.9	79.0	79.3	67.8	46.7	64.6
st3.cqp	82.4	78.3	80.2	80.3	78.9	69.5	50.0	66.1
st3.ccas	81.6	79.8	74.8	78.8	88.3	70.5	53.3	70.7

Table 8-21: Classification results for st3 experiments.

%	Training Set				Testing Set			
experiment	anterior	inferior	normal	total	anterior	inferior	normal	total
C3.knn-cc	81.8	83.1	83.2	82.7	85.6	81.0	80.0	82.2
C3.knn-rmse	83.2	84.3	83.2	83.6	81.8	81.0	80.0	80.9
C3.ncm-rmse	80.6	83.3	80.9	81.6	81.7	76.9	76.7	78.4
C3.ncm-cc	80.0	80.9	80.2	80.4	80.4	77.6	73.3	77.1

Table 8-22: Classification results for traditional techniques applied to problem 3.

8.3.2 Ranking Results

To provide a more complete comparison, the results described in the previous section were ranked according to testing set classification performance (Table 8-23) providing a more encompassing understanding of the results. The results were ranked in ascending order, according to the testing set classification performance achieved. A summary in Table 8-24 provides a quick reference to positioning of experiments within this ranking.

%	Training Set				Testing Set			
experiment	anterior	inferior	normal	total	anterior	inferior	normal	total
C3.knn-cc	81.8	83.1	83.2	82.7	85.6	81.0	80.0	82.2
E3.linreg	81.7	86.8	89.3	85.9	79.3	78.5	86.7	81.5
C3.knn-rmse	83.2	84.3	83.2	83.6	81.8	81.0	80.0	80.9
E3.ccas	93.3	87.1	87.0	89.1	88.6	78.3	73.3	80.1
L3.cqp	82.1	86.8	97.0	88.6	76.6	78.6	83.3	79.5
E3.cqp	82.6	86.7	84.7	84.7	73.6	81.3	83.3	79.4
E3.cas	93.3±0.7	88.0±0.8	87.3±1.9	89.5±0.7	91.1±1.6	79.7±2.6	64.7±5.4	78.5±1.9
E3.knn	91.8	90.7	100.0	94.2	81.7	76.9	76.7	78.4
C3.ncm-rmse	80.6	83.3	80.9	81.6	81.7	76.9	76.7	78.4
L3.knn	92.0	90.6	99.2	93.9	78.7	79.5	76.7	78.3
L3.cbp	90.8	88.0	99.2	92.7	81.7	76.0	76.7	78.1
L3.ccas	92.3	84.7	100.0	92.3	89.4	74.7	70.0	78.1
qrs3.knn	92.2	93.4	99.2	95.0	80.0	82.2	70.0	77.4
C3.ncm-cc	80.0	80.9	80.2	80.4	80.4	77.6	73.3	77.1
E3.cbp	89.3	89.5	84.7	87.9	74.5	82.1	73.3	76.6
L3.cas	93.6±1.2	83.7±0.7	99.8±0.3	92.4±0.5	91.7±2.1	72.9±2.2	64.7±4.8	76.4±1.6
qrs3.cbp	90.5	89.5	97.0	92.3	80.3	68.5	73.3	74.0
qrs3.C4.5	91.3	90.9	100.0	94.1	73.6	78.1	70.0	73.9
qrs3.linreg	84.2	88.6	96.2	89.7	74.6	70.0	76.7	73.8
L3.linreg	86.6	89.1	97.0	90.9	72.8	71.6	73.3	72.6
E3.qp	76.7±5.0	82.9±3.6	81.6±5.0	80.4±2.6	66.7±4.4	75.8±4.7	75.0±9.1	72.5±3.1
L3.C4.5	92.0	84.4	100.0	92.1	79.7	77.3	60.0	72.3
st3.knn	84.4	79.1	97.7	87.1	69.6	73.6	73.3	72.2
L3.MML	85.7	85.7	100.0	90.5	63.2	77.4	73.3	71.3
st3.ccas	81.6	79.8	74.8	78.8	88.3	70.5	53.3	70.7
L3.qp	78.8±4.4	82.7±1.9	89.8±2.9	83.8±2.4	69.8±6.1	70.9±3.6	71.0±6.0	70.6±3.3
qrs3.ccas	90.0	82.8	97.7	90.2	83.9	59.9	66.7	70.1
L3.bp	85.2±1.8	84.5±2.4	88.9±3.0	86.2±1.5	76.0±3.7	69.9±3.8	64.3±8.5	70.1±3.1
E3.bp	82.4±2.5	85.6±1.7	77.5±18.1	81.8±6.4	70.9±4.0	75.0±4.9	63.8±17.1	69.9±5.4
st3.linreg	69.3	70.0	87.0	75.5	68.2	58.6	80.0	68.9
qrs3.cqp	75.2	82.9	89.3	82.5	60.5	67.7	73.3	67.2
qrs3.MML	89.3	87.6	99.2	92.1	84.2	77.1	40.0	67.1
qrs3.cas	92.3±0.5	77.6±2.5	95.3±2.2	88.4±1.4	83.9±1.1	52.8±2.0	64.3±4.2	67.0±1.4
st3.cas	85.6±1.3	77.7±1.8	69.2±4.3	77.5±1.1	87.6±2.3	71.3±1.6	41.7±4.8	66.8±1.7
qrs3.bp	84.3±2.5	83.4±2.3	89.0±3.3	85.6±1.6	71.6±4.3	65.3±4.2	62.0±8.9	66.3±3.3
E3.MML	80.6	81.8	70.2	77.5	75.8	69.3	53.3	66.2
st3.cqp	82.4	78.3	80.2	80.3	78.9	69.5	50.0	66.1
st3.cbp	82.6	73.5	80.9	79.0	79.3	67.8	46.7	64.6
st3.MML	76.6	78.5	98.5	84.5	60.6	71.3	56.7	62.8
qrs3.qp	71.9±3.7	77.8±4.6	83.0±4.0	77.6±2.4	58.9±6.2	64.9±5.6	64.8±8.3	62.8±3.7
st3.qp	74.0±4.3	73.2±4.1	73.5±10.1	73.6±3.7	67.9±5.9	68.8±2.3	49.5±13.6	62.1±3.7
E3.C4.5	88.6	85.5	100.0	91.4	73.6	55.7	56.7	62.0
st3.C4.5	80.5	81.8	100.0	87.4	60.6	71.1	53.3	61.7
st3.bp	72.6±5.6	70.0±4.5	70.0±4.8	70.9±2.4	64.8±8.4	67.4±3.3	45.2±11.4	59.1±3.7

Table 8-23: Ranking of experiments applied to problem 3.
(Ranked with respect to percentage of testing set patients classified correctly)

Experiment	Ranking	%Correct (testing set)	E	L	qrs	st	C	* delta
C3.knn-cc	1	82.2					knn-cc	0.0
E3.linreg	2	81.5	linreg					0.7
C3.knn-rmse	3	80.9					knn-rmse	1.3
E3.ccas	4	80.1	ccas					2.1
L3.cqp	5	79.5		cqp				2.7
E3.cqp	6	79.4	cqp					2.8
E3.cas	7	78.5±1.9	cas					3.7
E3.knn	8	78.4	knn					3.8
C3.ncm-rmse	9	78.4					ncm-rmse	3.8
L3.knn	10	78.3		knn				3.9
L3.cbp	11	78.1		cbp				4.1
L3.ccas	12	78.1		ccas				4.1
qrs3.knn	13	77.4			knn			4.8
C3.ncm-cc	14	77.1					ncm-cc	5.1
E3.cbp	15	76.6	cbp					5.6
L3.cas	16	76.4±1.6		cas				5.8
qrs3.cbp	17	74.0			cbp			8.2
qrs3.C4.5	18	73.9			C4.5			8.3
qrs3.linreg	19	73.8			linreg			8.4
L3.linreg	20	72.6		linreg				9.6
E3.qp	21	72.5±3.1	qp					9.7
L3.C4.5	22	72.3		C4.5				9.9
st3.knn	23	72.2				knn		10.0
L3.MML	24	71.3		MML				10.9
st3.ccas	25	70.7				ccas		11.5
L3.qp	26	70.6±3.3		qp				11.6
qrs3.ccas	27	70.1			ccas			12.1
L3.bp	28	70.1±3.1		bp				12.1
E3.bp	29	69.9±5.4	bp					12.3
st3.linreg	30	68.9				linreg		13.3
qrs3.cqp	31	67.2			cqp			15.0
qrs3.MML	32	67.1			MML			15.1
qrs3.cas	33	67.0±1.4			cas			15.2
st3.cas	34	66.8±1.7				cas		15.4
qrs3.bp	35	66.3±3.3			bp			15.9
E3.MML	36	66.2	MML					16.0
st3.cqp	37	66.1				cqp		16.1
st3.cbp	38	64.6				cbp		17.6
st3.MML	39	62.8				MML		19.4
qrs3.qp	40	62.8±3.7			qp			19.4
st3.qp	41	62.1±3.7				qp		20.1
E3.C4.5	42	62.0	C4.5					20.2
st3.C4.5	43	61.7				C4.5		20.5
st3.bp	44	59.1±3.7				bp		23.1

Table 8-24: Summary of ranked results for problem 3.
(* delta – difference between experiment performance and best result – C3.knn-cc)

8.3.3 Discussion

When comparing these results (8-24) with those obtained in relation to problem 2 (Table 8-17) a number of interesting observations can be made. Firstly, the best results achieved for *problem 3* are significantly less than those achieved for *problem 2*. This would suggest that removing the follow-up patients from the training process actually degrades classifier performance. It is noted that all traditional BSM classifier results for *problem 3* (C3.knn-rmse – 80.9%, C3.knn-cc – 82.2%, C3.ncm-rmse – 78.4%, and C3.ncm-cc – 77.1%) are lower than those achieved in *problem 2* (C2.knn-rmse – 84%, C2.knn-cc – 82.7%, C2.ncm-rmse – 79.9%, and C2.ncm-cc – 79.7%), and similarly *E3.linreg* (81.5%) does not perform as well as *E2.linreg* (84%).

There are a number of classification approaches that improve with the removal of follow-up infarct classes from training, but the trends are unclear and still do not perform better than the best results obtained for *problem 2*. In short, these results would suggest that no overall advantage is achieved by excluding the follow-up patients from the classifier training or control groups.

The performance trends of feature extraction techniques are very similar to the trends observed for *problem 2*. Classifiers perform comparatively well when used in conjunction with the *E* and *L* feature extraction techniques, and their performance is degraded when the *qrs* and *st* feature extraction techniques are used.

Overall, the classification techniques *linreg (*using KLT feature extraction), *knn-cc* and *knn-rmse* produce the best classification results for this particular problem. The same overall outcome was observed in *problem 2,* and further to this, the results obtained in *problem 2* were better than those achieved in *problem 3*.

8.4 Problem 4

The final problem attempted using the forty-four classification techniques was *problem 4*. As has been stated previously, the aim in *problem 4* is to construct a classifier to separate patients with normal hearts from those with coronary artery disease. This is therefore a two-class classification problem.

All forty-four classification approaches were applied to this problem using the same training parameters as described in section 8.1.1. All classifiers were constructed with a two-class output.

The MLP architecture used for *bp* and *qp* experiments was similar to that used in *problem 1*; 72:72:2 (*E4.bp, E4.qp*), 192:120:2 (*L4.bp, L4.qp*), 128:120:2 (*qrs4.bp, qrs4.qp*) and 64:64:2 (*st4.bp, st4.qp*).

The results of these experiments are presented and discussed in the following sections. The detailed classification breakdowns are provided in appendix C for reference.

8.4.1 Results

The classification results were analysed and prepared using the CTAS analysis tool *performance* and *breakdown* (see appendix B). The classification results for these experiments are summarised in Tables 8-25, 8-26, 8-27, 8-28 and 8-29. Note that where a number of training runs were performed in an experiment (*bp*, *qp*, and *cas* – 20 networks were trained in each case) the averaged results are presented. The more detailed classification breakdown results are provided in appendix C for reference and will be referred in the discussion.

%	Training Set			Testing Set		
experiment	normal	cad	total	normal	cad	total
E4.knn	97.71	100	98.85	33.33	83.33	58.33
E4.linreg	70.99	72.46	71.73	33.33	73.33	53.33
E4.C4.5	100	99.02	99.51	33.33	90	61.67
E4.MML	99.24	60.33	79.78	63.33	43.33	53.33
E4.bp	56.4±5.9	69.2±5.9	62.8±0.9	57.7±7.7	70.2±5.1	63.9±2.3
E4.qp	57.3±7.1	68.3±7.2	62.8±0.8	58.0±10.0	68.7±6.7	63.3±2.9
E4.cas	84.2±4.2	60.6±4.3	72.4±1.3	52.3±2.4	68.8±4.4	60.6±2.5
E4.cbp	41.22	80.66	60.94	40	80	60
E4.cqp	41.22	80.33	60.77	40	80	60
E4.ccas	82.44	66.23	74.34	53.33	73.33	63.33

Table 8-25: Classification results for E4 experiments.

%	Training Set			Testing Set		
experiment	normal	cad	total	normal	cad	total
L4.knn	96.95	100	98.47	36.67	83.33	60
L4.linreg	83.21	81.64	82.42	46.67	76.67	61.67
L4.C4.5	100	98.69	99.34	30	80	55
L4.MML	90.84	81.64	86.24	43.33	73.33	58.33
L4.bp	65.6±6.9	67.0±7.6	66.3±3.3	56.2±13.0	71.3±7.9	63.8±6.9
L4.qp	61.5±5.2	69.0±5.9	65.2±2.9	58.3±13.1	74.7±6.9	66.5±5.6
L4.cas	88.0±3.5	62.4±5.1	75.2±1.7	56.7±5.1	46.8±5.6	51.8±3.9
L4.cbp	57.25	89.84	73.54	13.33	93.33	53.33
L4.cqp	61.83	96.07	78.95	26.67	90	58.33
L4.ccas	85.5	70.49	77.99	46.67	60	53.33

Table 8-26: Classification results for L4 experiments.

%	Training Set			Testing Set		
experiment	normal	cad	total	normal	cad	total
qrs4.knn	96.18	100	98.09	36.67	83.33	60
qrs4.linreg	76.34	74.43	75.38	63.33	80	71.67
qrs4.C4.5	100	98.36	99.18	30	83.33	56.67
qrs4.MML	95.42	69.51	82.46	60	56.67	58.33
qrs4.bp	73.9±10.5	65.7±13.3	69.8±7.9	62.0±18.5	63.5±10.2	62.8±9.2
qrs4.qp	70.1±8.0	61.3±7.7	65.7±3.3	66.7±9.5	66.3±8.0	66.5±4.3
qrs4.cas	84.3±4.0	60.2±4.6	72.3±1.9	55.5±5.8	48.0±4.6	51.8±3.0
qrs4.cbp	84.73	98.03	91.38	40	86.67	63.33
qrs4.cqp	74.81	88.2	81.5	50	83.33	66.67
qrs4.ccas	83.97	67.21	75.59	53.33	50	51.67

Table 8-27: Classification results for qrs4 experiments.

%	Training Set			Testing Set		
experiment	normal	cad	total	normal	cad	total
st4.knn	98.47	100	99.24	3.33	70	36.67
st4.linreg	66.41	66.56	66.48	26.67	66.67	46.67
st4.C4.5	100	99.67	99.84	30	86.67	58.33
st4.MML	92.37	79.67	86.02	23.33	60	41.67
st4.bp	82.1±7.4	64.0±9.3	73.1±1.9	48.2±13.6	46.8±12.2	47.5±5.2
st4.qp	79.5±6.2	76.5±6.4	78.0±2.4	48.0±15.4	58.7±8.7	53.3±6.4
st4.cas	85.1±3.3	67.1±6.0	76.1±2.9	33.5±8.1	62.7±7.9	48.1±5.1
st4.cbp	92.37	67.21	79.79	40	46.67	43.33
st4.cqp	92.37	81.97	87.17	63.33	60	61.67
st4.ccas	96.18	78.69	87.44	23.33	73.33	48.33

Table 8-28: Classification results for st4 experiments.

%	Training Set			Testing Set		
experiment	normal	cad	total	normal	cad	total
C4.ncm-cc	19.9	86.2	53.0	26.7	93.3	60.0
C4.knn-rmse	36.6	76.4	56.5	36.7	80.0	58.3
C4.ncm-rmse	32.1	80.0	56.0	33.3	83.3	58.3
C4.knn-cc	31.3	87.5	59.4	16.7	96.7	56.7

Table 8-29: Classification results of traditional techniques applied to problem 4.

8.4.2 Ranking Results

The initial ranking of these experiments is detailed in Table 8-30. Note that a different ranking criteria was chosen compared with the criteria used in the previous problems. It is observed that many of the classifiers are biased toward either normal or CAD patients. Consequently ranking classifiers on the basis of overall testing set performance was found to be misleading, as many classifiers achieve a good overall classification performance, but achieve this by strongly biasing one particular class (a good example being experiment *E4.C4.5* – 33.3% of normals correctly classified, 90% of CADs correctly classified, resulting in an overall classification performance 61.7%).

To eliminate this biasing problem, it was found to be more effective to rank experiments according to the minimum normal/CAD classification performance, thus focusing on the worst class in each experiment. As a result we find that experiments which produce balanced results (ie classifying well with respect to normals and CADs) are positioned higher on the ranking.

It will be noted that the ranking provided in Table 8-30 presents the averaged results for the *bp*, *qp*, and *cas* experiments. As will be recalled from the initial experiments (*E4.bp*) a selection criteria was developed for selecting the best classifiers in these multiple run experiments (see chapter 7). This criteria selected the best training run on the basis of balanced training classification (ie the minimum $|N_{TR}-C_{TR}|$, where N_{TR} is percentage of training set normals classified correctly and C_{TR} is percentage of training set CADs classified correctly).

This selection criteria was applied to all multiple run experiments (*bp*, *qp*, *cas*) and it was found that in all cases networks were selected that performed well on the testing set. These selected networks are presented in the ranking provided in Table 8-31 and a summary ranking is provided in Table 8-32.

%	Training Set			Testing Set			
experiment	normal	cad	total	normal	cad	total	*minimum (normal,cad) (testing set)
qrs4.qp	70.1±8.0	61.3±7.7	65.7±3.3	66.7±9.5	66.3±8.0	66.5±4.3	66.3
qrs4.linreg	76.3	74.4	75.4	63.3	80.0	71.7	63.3
qrs4.bp	73.9±10.5	65.7±13.3	69.8±7.9	62.0±18.5	63.5±10.2	62.8±9.2	62.0
st4.cqp	92.4	82.0	87.2	63.3	60.0	61.7	60.0
L4.qp	61.5±5.2	69.0±5.9	65.2±2.9	58.3±13.1	74.7±6.9	66.5±5.6	58.3
E4.qp	57.3±7.1	68.3±7.2	62.8±0.8	58.0±10.0	68.7±6.7	63.3±2.9	58.0
E4.bp	56.4±5.9	69.2±5.9	62.8±0.9	57.7±7.7	70.2±5.1	63.9±2.3	57.7
qrs4.MML	95.4	69.5	82.5	60.0	56.7	58.3	56.7
L4.bp	65.6±6.9	67.0±7.6	66.3±3.3	56.2±13.0	71.3±7.9	63.8±6.9	56.2
E4.ccas	82.4	66.2	74.3	53.3	73.3	63.3	53.3
E4.cas	84.2±4.2	60.6±4.3	72.4±1.3	52.3±2.4	68.8±4.4	60.6±2.5	52.3
qrs4.cqp	74.8	88.2	81.5	50.0	83.3	66.7	50.0
qrs4.ccas	84.0	67.2	75.6	53.3	50.0	51.7	50.0
st4.qp	79.5±6.2	76.5±6.4	78.0±2.4	48.0±15.4	58.7±8.7	53.3±6.4	48.0
qrs4.cas	84.3±4.0	60.2±4.6	72.3±1.9	55.5±5.8	48.0±4.6	51.8±3.0	48.0
L4.cas	88.0±3.5	62.4±5.1	75.2±1.7	56.7±5.1	46.8±5.6	51.8±3.9	46.8
st4.bp	82.1±7.4	64.0±9.3	73.1±1.9	48.2±13.6	46.8±12.2	47.5±5.2	46.8
L4.linreg	83.2	81.6	82.4	46.7	76.7	61.7	46.7
L4.ccas	85.5	70.5	78.0	46.7	60.0	53.3	46.7
L4.MML	90.8	81.6	86.2	43.3	73.3	58.3	43.3
E4.MML	99.2	60.3	79.8	63.3	43.3	53.3	43.3
qrs4.cbp	84.7	98.0	91.4	40.0	86.7	63.3	40.0
E4.cbp	41.2	80.7	60.9	40.0	80.0	60.0	40.0
E4.cqp	41.2	80.3	60.8	40.0	80.0	60.0	40.0
st4.cbp	92.4	67.2	79.8	40.0	46.7	43.3	40.0
L4.knn	97.0	100.0	98.5	36.7	83.3	60.0	36.7
qrs4.knn	96.2	100.0	98.1	36.7	83.3	60.0	36.7
C4.knn-rmse	36.6	76.4	56.5	36.7	80.0	58.3	36.7
st4.cas	85.1±3.3	67.1±6.0	76.1±2.9	33.5±8.1	62.7±7.9	48.1±5.1	33.5
E4.C4.5	100.0	99.0	99.5	33.3	90.0	61.7	33.3
E4.knn	97.7	100.0	98.9	33.3	83.3	58.3	33.3
E4.linreg	71.0	72.5	71.7	33.3	73.3	53.3	33.3
C4.ncm-rmse	32.1	80.0	56.0	33.3	83.3	58.3	33.3
st4.C4.5	100.0	99.7	99.8	30.0	86.7	58.3	30.0
qrs4.C4.5	100.0	98.4	99.2	30.0	83.3	56.7	30.0
L4.C4.5	100.0	98.7	99.3	30.0	80.0	55.0	30.0
L4.cqp	61.8	96.1	79.0	26.7	90.0	58.3	26.7
st4.linreg	66.4	66.6	66.5	26.7	66.7	46.7	26.7
C4.ncm-cc	19.9	86.2	53.0	26.7	93.3	60.0	26.7
st4.ccas	96.2	78.7	87.4	23.3	73.3	48.3	23.3
st4.MML	92.4	79.7	86.0	23.3	60.0	41.7	23.3
C4.knn-cc	31.3	87.5	59.4	16.7	96.7	56.7	16.7
L4.cbp	57.3	89.8	73.5	13.3	93.3	53.3	13.3
st4.knn	98.5	100.0	99.2	3.3	70.0	36.7	3.3

Table 8-30: Ranking of all experiments applied to problem 4.
(* minimum classification result achieved on normal or CAD classes in the testing set)
(Ranked according to minimum classification result achieved on testing set classes)

%	Training Set			Testing Set			
experiment	normal	cad	total	normal	cad	total	minimum (normal, cad) (testing set)
*E4.qp	61.8	62.3	62.1	70	73.3	71.7	70
*qrs4.qp	65.7	65.3	65.5	70	66.7	68.3	66.7
*qrs4.bp	66.4	66.2	66.3	66.7	66.7	66.7	66.7
qrs4.linreg	76.3	74.4	75.4	63.3	80	71.7	63.3
*L4.bp	66.4	65.6	66	63.3	70	66.7	63.3
*E4.bp	58.8	65.6	62.2	63.3	66.7	65	63.3
st4.cqp	92.4	82	87.2	63.3	60	61.7	60
qrs4.MML	95.4	69.5	82.5	60	56.7	58.3	56.7
E4.ccas	82.4	66.2	74.3	53.3	73.3	63.3	53.3
*E4.cas	76.3	68.5	72.4	53.3	70	61.7	53.3
qrs4.cqp	74.8	88.2	81.5	50	83.3	66.7	50
*L4.qp	67.2	66.9	67	50	70	60	50
*qrs4.cas	72.5	64.9	68.7	50	60	55	50
*L4.cas	82.4	72.1	77.3	56.7	50	53.3	50
qrs4.ccas	84	67.2	75.6	53.3	50	51.7	50
L4.linreg	83.2	81.6	82.4	46.7	76.7	61.7	46.7
L4.ccas	85.5	70.5	78	46.7	60	53.3	46.7
L4.MML	90.8	81.6	86.2	43.3	73.3	58.3	43.3
E4.MML	99.2	60.3	79.8	63.3	43.3	53.3	43.3
qrs4.cbp	84.7	98	91.4	40	86.7	63.3	40
E4.cbp	41.2	80.7	60.9	40	80	60	40
E4.cqp	41.2	80.3	60.8	40	80	60	40
st4.cbp	92.4	67.2	79.8	40	46.7	43.3	40
L4.knn	97	100	98.5	36.7	83.3	60	36.7
qrs4.knn	96.2	100	98.1	36.7	83.3	60	36.7
*st4.qp	81.7	82.3	82	36.7	63.3	50	36.7
C4.knn-rmse	36.64	76.39	56.52	36.67	80	58.33	36.67
C4.ncm-rmse	32.06	80	56.03	33.33	83.33	58.33	33.33
E4.C4.5	100	99	99.5	33.3	90	61.7	33.3
E4.knn	97.7	100	98.9	33.3	83.3	58.3	33.3
E4.linreg	71	72.5	71.7	33.3	73.3	53.3	33.3
*st4.bp	75.6	74.4	75	33.3	53.3	43.3	33.3
st4.C4.5	100	99.7	99.8	30	86.7	58.3	30
qrs4.C4.5	100	98.4	99.2	30	83.3	56.7	30
L4.C4.5	100	98.7	99.3	30	80	55	30
*st4.cas	80.2	79	79.6	30	56.7	43.3	30
L4.cqp	61.8	96.1	79	26.7	90	58.3	26.7
st4.linreg	66.4	66.6	66.5	26.7	66.7	46.7	26.7
C4.ncm-cc	19.85	86.23	53.04	26.67	93.33	60	26.67
st4.ccas	96.2	78.7	87.4	23.3	73.3	48.3	23.3
st4.MML	92.4	79.7	86	23.3	60	41.7	23.3
C4.knn-cc	31.3	87.54	59.42	16.67	96.67	56.67	16.67
L4.cbp	57.3	89.8	73.5	13.3	93.3	53.3	13.3
st4.knn	98.5	100	99.2	3.3	70	36.7	3.3

Table 8-31: Ranking of all experiments applied to problem 4.
(* best bp,qp,cas classifiers – see chapter 7 for discussion of selection criteria)
(Ranked according to minimum classification result achieved on testing set classes)

Experiment	Ranking	minimum (normal,cad) (testing set)	E	L	qrs	st	C	#delta
*E4.qp	1	70	qp					0
*qrs4.qp	2	66.7			qp			3.3
*qrs4.bp	3	66.7			bp			3.3
qrs4.linreg	4	63.3			linreg			6.7
*L4.bp	5	63.3		bp				6.7
*E4.bp	6	63.3	bp					6.7
st4.cqp	7	60				cqp		10
qrs4.MML	8	56.7			MML			13.3
E4.ccas	9	53.3	ccas					16.7
*E4.cas	10	53.3	cas					16.7
qrs4.cqp	11	50			cqp			20
*L4.qp	12	50		qp				20
*qrs4.cas	13	50			cas			20
*L4.cas	14	50		cas				20
qrs4.ccas	15	50			ccas			20
L4.linreg	16	46.7		linreg				23.3
L4.ccas	17	46.7		ccas				23.3
L4.MML	18	43.3		MML				26.7
E4.MML	19	43.3	MML					26.7
qrs4.cbp	20	40			cbp			30
E4.cbp	21	40	cbp					30
E4.cqp	22	40	cqp					30
st4.cbp	23	40				cbp		30
L4.knn	24	36.7		knn				33.3
qrs4.knn	25	36.7			knn			33.3
*st4.qp	26	36.7				qp		33.3
C4.knn-rmse	27	36.67					knn-rmse	33.33
C4.ncm-rmse	28	33.33					ncm-rmse	36.67
E4.C4.5	29	33.3	C4.5					36.7
E4.knn	30	33.3	knn					36.7
E4.linreg	31	33.3	linreg					36.7
*st4.bp	32	33.3				bp		36.7
st4.C4.5	33	30				C4.5		40
qrs4.C4.5	34	30			C4.5			40
L4.C4.5	35	30		C4.5				40
*st4.cas	36	30				cas		40
L4.cqp	37	26.7		cqp				43.3
st4.linreg	38	26.7				linreg		43.3
C4.ncm-cc	39	26.67					ncm-cc	43.33
st4.ccas	40	23.3				ccas		46.7
st4.MML	41	23.3				MML		46.7
C4.knn-cc	42	16.67					knn-cc	53.33
L4.cbp	43	13.3		cbp				56.7
st4.knn	44	3.3				knn		66.7

Table 8-32: Summary of ranked results for problem 4.
(* best bp,qp,cas classifiers – see chapter 7 for discussion of selection criteria)
(# delta – difference between experiment performance and best result – E4.qp)

8.4.3 Discussion

The ranked results for this particular problem are significantly different to those observed for the previous three problems considered. One of the key differences relates to the ranking of the traditional BSM classification techniques. Although these techniques were found to be most suitable for separating myocardial infarct from normals (*problems 2 and 3*), the traditional BSM classification techniques (C4.knn-rmse, C4.knn-cc, C4.ncm-rmse and C4.ncm-cc) rank poorly when attempting to separate patients with coronary artery disease from normal patients. On closer examination it is noted that the traditional BSM techniques are strongly biased toward the CAD patients and do not even achieve a classification performance of 50% with respect to normal patients (test set results: C4.knn-rmse – 36.7%, C4.knn-cc – 33.3%, C4.ncm-rmse – 26.7% and C4.ncm-cc – 16.7%).

Since this is a two-class problem it is important that any classifier can correctly classify at least 50% of both classes in the testing set. If not achieved then this suggests that the network is simply biasing one class over the other. On closer examination of these results it is noted that those classifiers ranked below the 15th result (qrs4.ccas) classify less than 50% of the normal patients in the testing set correctly and are biased toward the CAD class (apart from E4.MML which is biased toward the normal class). Further to this, classifiers ranked 11th through 15th only manage to classify 50% of one class correctly. This would therefore suggest that only those classifiers ranked in the top ten have managed to identify discriminating features.

Of these classifiers in the top ten, six techniques manage to classify more than 60% of testing set patients correctly and only three techniques (*E4.qp*, *qrs4.qp*, and *qrs4.bp*) correctly classify more than 65% of testing set patients correctly. It is also interesting to note that the top three techniques are neural network techniques, and only two non neural network techniques are ranked in the top ten experiments (*qrs4.linreg*, and *qrs4.MML*).

Given these results a number of conclusions can be drawn. Firstly, the problem of discriminating between normal and CAD patients is difficult. Secondly, traditional BSM classification techniques perform poorly when applied to this problem. Thirdly, neural networks perform significantly better than traditional BSM classification techniques and other alternative classification techniques.

It is worth noting that one extension to neural network training, which has contributed to this result, is the use of multiple training runs and the selection of the best network.

8.5 Summary

A number of conclusions can be draw from this second set of experiments:

1. Traditional BSM classification techniques or linear regression used in conjunction with KLT feature extraction are the most appropriate classification techniques for separating patients with myocardial infarctions from normal patients.

2. Discriminating between normal patients and patients with coronary artery disease is difficult and cannot be achieved using traditional BSM classification techniques.

3. It is possible using neural networks to discriminate between normal patients and patients with coronary artery disease. The best result achieved was 71% (experiment *E4.qp*: normals – 70.0% correct, CADs – 71.7% correct).

4. The inductive learning techniques C4.5 and MML performed poorly on all BSM classification problems. This would suggest that the problems, although linearly separable cannot be easily separated using orthogonal decision planes (as used by these inductive learning techniques).

9. Improving Classification Reliability

Having explored a range of approaches for classifying BSM data in chapters 7 and 8 the following outcomes have been achieved thus far:

1. Attempting to construct a classifier to separate all four classes; anterior infarctions, inferior infarctions, normals and CADs is a difficult task (*problem 1*). The best classification result achieved was for experiment *E1.linreg* (testing set result – 67.9%: anteriors – 83%, inferiors – 78.5%, normal – 30% and CADs – 80%). It was noted that all classifiers applied to this problem had difficulty separating normals and CADs and tended to classify most normal patients as having CAD. On examining this problem more closely it was discovered that a far more effective way to discriminate between classes was to consider this classification problem as two separate problems; the discrimination of infarcts from normals (*problems 2 and 3*) and the discrimination of normals and CADs (*problem 4*). Focusing on these problems separately improved the classification performance of classifiers significantly.

2. When applying the classifiers to the infarct/normal classification problem (*problem 2 and 3*) the classification results improved significantly over those obtained for the initial problem particularly in relation to the classification of normals. The best classification result achieved was for experiment *E2.linreg* (testing set result – 84%: anteriors – 81.7%, inferiors – 80.3% and normals – 90%).

3. The separation of normal patients and patients with coronary artery disease (CAD) was found be a difficult problem (*problem 4*). Many classifiers applied to this problem did not manage to discriminate between these classes and tended to bias strongly toward one class. However, the neural network classifiers considered performed significantly better than other classifiers. The best classification result achieved was for experiment *E4.qp* (testing set result – 71.7%: normals – 70% and CADs - 73.3%).

Although these results are encouraging it is clear that no technique applied to these classification problems achieved a 100% classification performance. In all cases a number of patients were incorrectly classified. As highlighted in chapter 7 there are three possible reasons for these less than optimum results;

- the classifiers applied are unable to identify the discriminating features,
- the feature extraction techniques used are inadvertently removing discriminating features, or
- the discriminating features required are not present in the BSM data.

Every effort has been made thus far to consider a range of different classification and feature extraction techniques in an attempt to determine which combination of techniques are most suited to these classification problems. As such the first two issues identified above have been explored extensively. The issue unaddressed thus far is the possibility that the discriminating features required to achieve a 100% classification performance are not present in the BSM data.

This chapter aims to explore this issue. Section 9.1 illustrates how a lack of discriminating features will result in misclassification. Section 9.2 explores the bayesian equivalence of neural network and possible bayesian solutions to the issue of misclassification. Section 9.3 then proposes two techniques for dealing with misclassifications in neural networks and the remainder of the chapter presents the practical application of these techniques to BSM classification problems.

9.1 Misclassification

It is important to appreciate when tackling any classification problem that given a specific feature space, completely separating two classes may not be possible. At first this may be a difficult concept to grasp, and one may be convinced that with a better classifier a 100% classification performance may be achieved, but the reality is that in some cases this is not possible.

A simple example that illustrates this point is given in Figure 9-1 where a set of objects are classified as being in either class A or B. In this particular example objects are described by the features X_1 and X_2 and consequently can be presented in a two dimensional feature space (Figure 9-1).

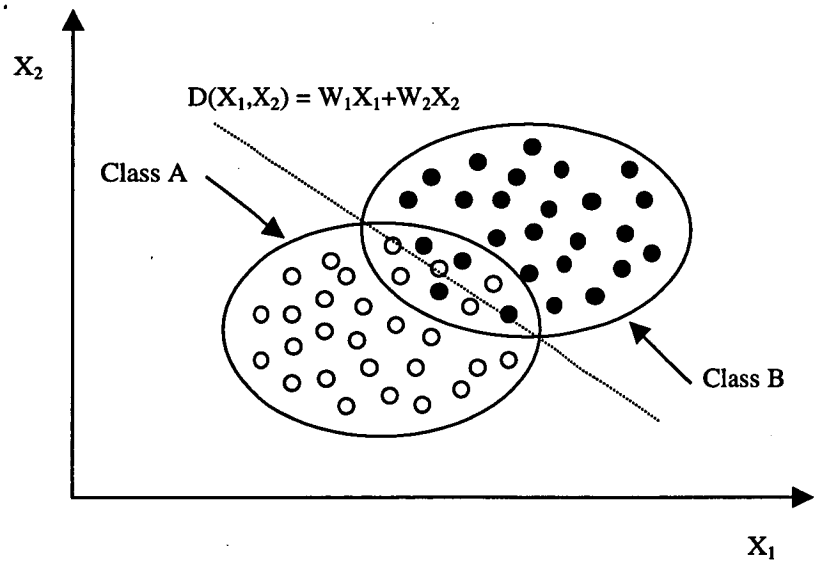


Figure 9-1: Example of class overlap.

It is clear from the outset that these two classes overlap in the (X_1,X_2) feature space.. This overlap presents a number of problems. Firstly, it is possible in the overlapping region for an object belonging to class A to have identical feature values to an object from class B. If this is the case, then no classifier can be constructed to successfully classify such cases. Secondly, no linear discriminator $D(X_1,X_2)$ or higher order function can be constructed which is capable of separating these classes with a 100% success rate. The reality is that given this feature space some degree of misclassification is inevitable.

Clearly the only way that misclassification can be resolved it to identify additional features that would achieve clear separation of these classes. For example in the case above there may exist a feature X_3 which can be added to the current feature space

(thus creating a three dimensional space) which allows clear separation of these classes. However, if no such features exist then the fact remains that this is the best solution and that some degree of misclassification is inevitable. Every effort can be made to construct a classifier which will optimise the number of correct classifications, but misclassification will be unavoidable.

When using classification tools for clinical diagnostic testing the possibility of misclassification is a serious problem, particularly if a classifier is being used to diagnose a life threatening disease. Therefore, it is crucial that the possibility of misclassification is minimised.

9.2 Dealing with Misclassifications when using Neural Networks

If a classification performance of 100% cannot be achieved and misclassification is inevitable, then a different strategy needs to be considered for dealing with misclassification. The following three sections propose a strategy for dealing with class boundary overlap and misclassification when using neural network classifiers.

Sections 9.2.1, 9.2.2, 9.2.3 examine a bayesian approach for dealing with misclassifications and sections 9.2.4 and 9.3 examine how these techniques may be applied to neural network classifiers.

9.2.1 Identifying Class Boundary Overlap

A simple example of overlapping class boundaries is illustrated in Figure 9-2. The objective in this classification problem is to separate objects described by the feature vector x into classes A or B, which are normally distributed around $\overline{a}$ and $\overline{b}$ respectively. These distributions are described by the *a priori* probability functions P(x|A) and P(x|B), where P(x|A) is the probability of feature vector x given that the object described by x is known to be in class A, and similarly for P(x|B). Clearly, these two classes overlap, resulting in ambiguous examples such as x_1 and x_2 which could belong to either class.

Intuitively, the optimum solution to this problem would be to separate examples around the point x_s (point of equal probability), thus minimising the number of misclassifications. If $x<x_s$ then x is in class A. If $x>x_s$ then x is in class B. Although this is the optimum solution, it will not achieve 100% classification, as some examples from class A will still satisfy the class B criterion ($x>x_s$) and vice versa.

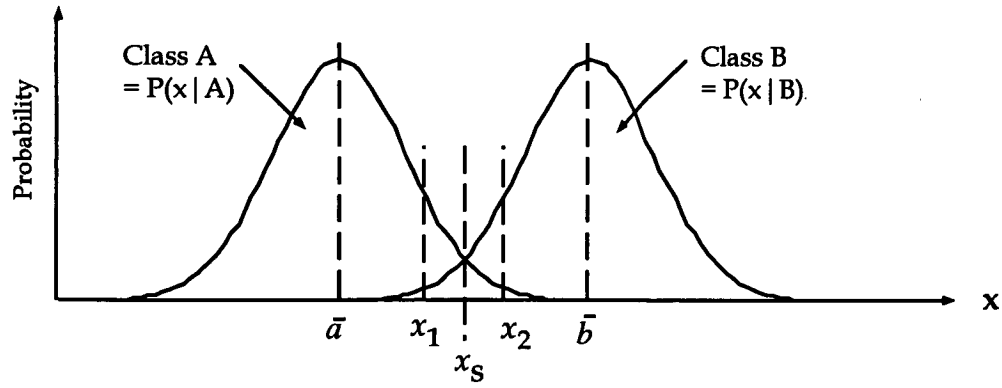


Figure 9-2: Class overlap example.

The only way to improve the classification accuracy in this example is to eliminate or reject the ambiguous examples by introducing a third class — an uncertain class (Figure 9-3). The classification rules would then be:

Rule	Classification
$x < x_a$	class A
$x > x_b$	class B
$x_a \geq x \geq x_b$	uncertain

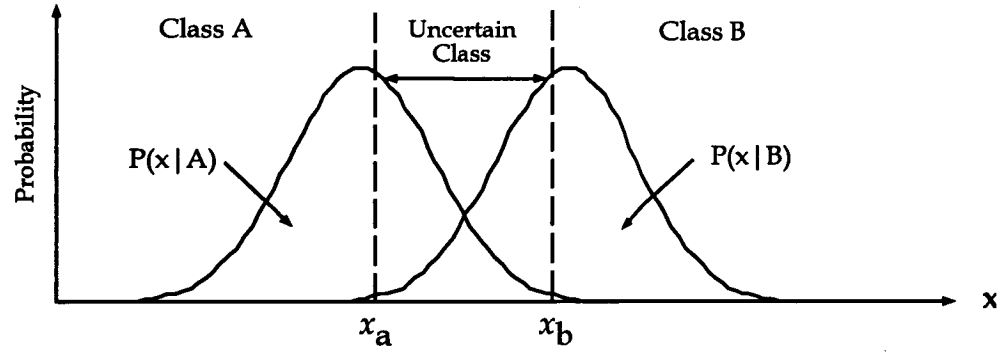


Figure 9-3: Inclusion of an 'uncertain' class.

This will not guarantee 100% classification performance, as classes A and B are continuous distributions in the feature space x and some examples will still be misclassified irrespective of the x_a and x_b chosen. However, this strategy will eliminate the majority of misclassifications, and the classification performance for the remaining 'certain' or 'accepted' examples (where $x<x_a$ and $x>x_b$) will be significantly improved. Such a scheme is useful in classification problems like the BSM data where the correctness of the classification is crucial.

This is a rather intuitive and informal solution to the problem of class boundary overlap. However, the example does clarify why class boundary overlap causes misclassifications and how the problem could be solved.

9.2.2 Bayesian Classifiers

An examination of Bayes decision theory provides a more complete treatment of class boundary overlap problem. It is important to understand how the problem of class boundary overlap can be resolved using Bayesian classifiers, as this will provide the theoretical background needed to resolve class boundary overlap when using neural network classifiers.

Bayes decision theory is a fundamental statistical approach to the problem of pattern recognition. This theory proposes that the optimum solution to any classification problem is the Bayesian discriminant function or Bayesian classifier.

To illustrate how a Bayesian classifier is constructed, consider a classification problem where the pattern vector x must be classified into one of N classes: C_1, C_2, $C_3, \ldots, C_N$. And suppose that the probability of each class $P(C_i)$ and the *a priori* probability $P(x|C_i)$ are known. Given this information the Bayesian discriminant function can be constructed using the following classification rule:

An object described by feature vector x is assigned to class i if

$$P(C_i \mid x) > P(C_j \mid x) \text{ for all } i \neq j \text{ and } 1 \leq j \leq N \tag{9-1}$$

where $P(C_i|x)$ is the probability of the object belonging to class i — this is the *a posteriori* class probability, or Bayesian probability, and is defined in terms of *a priori* probabilities by *Bayes Rule*:

$$P(C_i \mid x) = \frac{P(x \mid C_i)P(C_i)}{\sum_{all\,j} P(x \mid C_j)P(C_j)} \tag{9-2}$$

Applying (9-2) to the previous example problem results in the *a posteriori* probabilities in Figure 9-4. If (9-1) is applied to this, the following classification rules are derived:

Rule	Classification
P(A|x)>P(B|x)	class A
P(B|x)>P(A|x)	class B

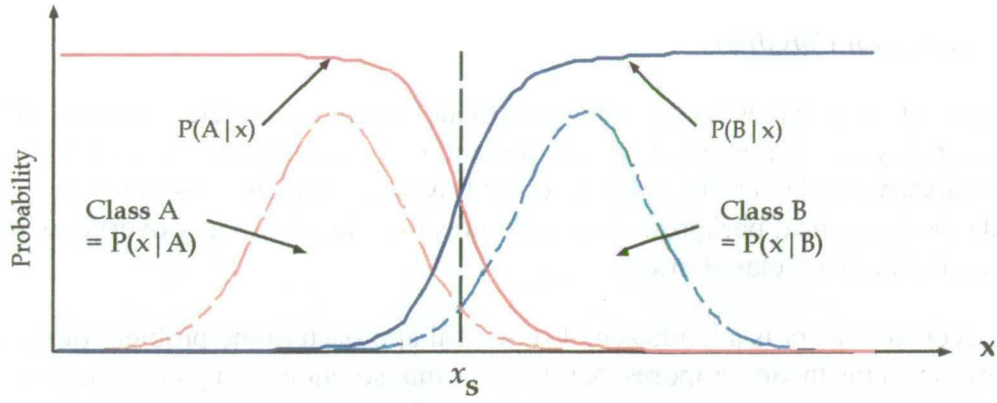


Figure 9-4: Bayesian Solution.

The conditions that satisfy these rules are:

$$P(A|x)>P(B|x) \text{ when } x<x_s$$
$$P(B|x)>P(A|x) \text{ when } x>x_s$$

Interestingly this is identical to the intuitive classification solution suggested in Figure 9-2. The next step is to use the *a posteriori* probabilities to identify possible misclassifications. One solution is to apply a rejection rule.

9.2.3 Chow's Rejection Rule

Although Bayesian classifiers provide the optimum solution to any classification problem, misclassifications are still going to occur when class boundaries overlap. Chow (1957) proposed a technique for dealing with misclassifications when using Bayesian classifiers. This technique still uses the Bayesian discriminant function to determine the classification, but adds a further restriction — a rejection rule. The classification is only accepted if the *a posteriori* probability for the classification selected is greater than some pre-defined threshold *t*. If the *a posteriori* probability is less than this threshold, then it is rejected and the classification is withheld. This in effect creates an uncertain class (Figure 9-5).

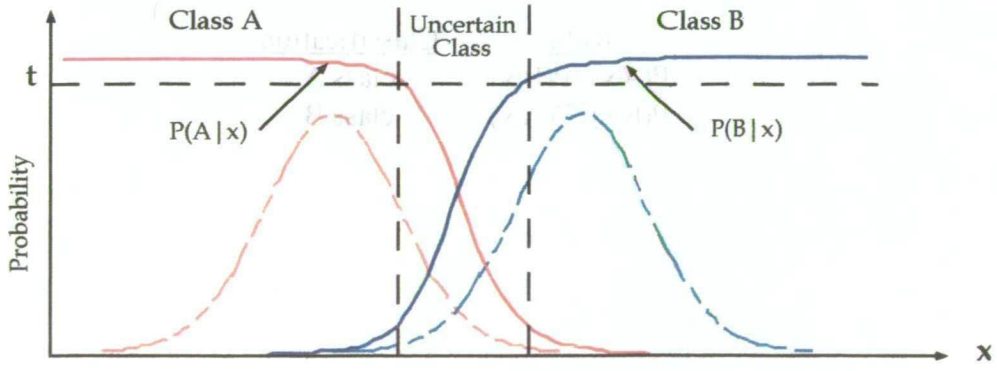


Figure 9-5: Chow's Rejection Rule.

It is possible to select a threshold *t* that isolates the majority of the misclassifications, as was achieved with the less formal technique previously described. As a result the

examples now assigned to classes A and B are more likely to be correct and, therefore, the classification accuracy will be improved.

The selection of the threshold t is determined by the classification accuracy required. As t is increased the classification accuracy will increase, since more ambiguous examples will be classified as uncertain. The down side of increasing the threshold is that fewer examples will be assigned a classification.

This technique does not claim to reject only misclassified examples. In identifying classifications with *a posteriori* probabilities below an acceptable threshold, it is inevitable that some correct classifications will be assigned to the uncertain class.

9.2.4 Bayesian Equivalence of Neural Networks

The rejection rule just described is designed for use with Bayesian classifiers, not neural networks. Therefore, it is important to understand the Bayesian properties of neural networks before applying a similar rejection rule to such classifiers.

In practice, Bayesian classifiers are created by using training data to estimate the *a priori* probabilities. This involves fitting specific parametric distributions (guassian or guassian mixture) to the training examples. Having estimated the *a priori* probabilities, the *a posteriori* probabilities can be calculated (9-2) and a classification can be assigned using the Bayesian discriminant function (9-1).

As described in chapter 3, neural networks solve classification problems using a very different technique. During training the weights of a network are adjusted using the back-propagation technique to minimise the overall error between the actual network outputs and the desired outputs for each training example. Once trained, a classification can be assigned using the following rule:

An object described by feature vector x is assigned to class i if:

$$O_i(x,\Theta) > O_j(x,\Theta) \text{ for all } i \neq j \text{ and } 1 \leq j \leq N \qquad (9\text{-}3)$$

where, for an N-class problem, $O_i(x,\Theta)$ is the output of the network for class i, and Θ represents the parameters or weights of the network (Figure 9-6).

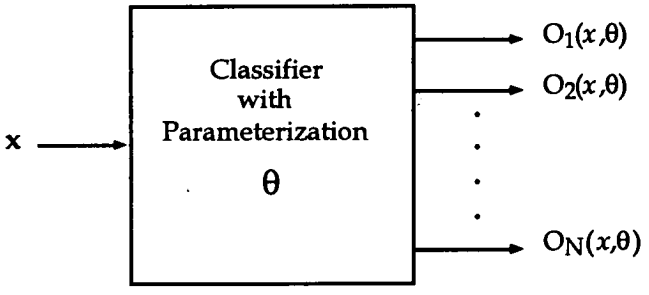


Figure 9-6: General N-class neural network classifier

In an attempt to establish a connection between Bayesian and neural network classifiers, a number of theoretical papers have shown that neural networks trained using the mean squared error criterion yield outputs that approximate *a posteriori* probabilities (9-4):

$$O_i(x,\Theta) \approx P(C_i|x) \tag{9-4}$$

This relationship was established as early as 1973 by Duda and Hart (1973) who provided a derivation for the two-class case when using a single-layer network with linear outputs. Many recent papers have extended this proof for multi-layer networks applied to the more general N-class case (Bourland and Wellekens 1989; Gish 1990; Hampshire and Pearlmutter 1990; Ruck *et al.* 1990; Shoemaker 1991; Wan 1990; White 1989).

This suggests that during the training process neural network classifiers are attempting to estimate the *a posteriori* probabilities directly from the training data, as opposed to the somewhat indirect method used by Bayesian classifiers.

Such derivations do require a number of conditions to be met before assuming that network outputs approximate Bayesian probabilities. Hampshire and Pearlmutter (1990) identify three necessary conditions:

- the network must be sufficiently parameterised by Θ to model the *a posteriori* probability functions,
- the training set must contain an asymptotically large number of statistically independent training samples, and
- the network is trained to binary targets (that is target vectors are constructed by assigning a value of 1.0 to the output associated with the class and 0.0 to all other outputs).

Therefore, provided these conditions are met, then the output of a neural network classifier should approximate *a posteriori* probabilities and therefore a rejection rule could be used to reject possible misclassifications.

9.3 Utilising the Bayesian Equivalence of Neural Networks

Given the theoretical premise that neural network outputs approximate Bayesian probabilities, a number of experiments were conducted to assess the possibility of using this information to improve the reliability of the neural network classification of the BSM data. Two approaches were considered: the thresholding and binning techniques.

9.3.1 Thresholding Technique

The first technique considered was a direct application of Chow's rejection rule by applying a threshold t to the network to reject possible misclassifications:

$$\text{given } O_i(x,\Theta) > O_j(x,\Theta) \text{ for all } i \neq j \text{ and } 1 \leq j \leq N$$

Rule	Classification
$O_i(x,\Theta) \geq t$	class i
$O_i(x,\Theta) < t$	uncertain

If the neural network outputs do approximate Bayesian probabilities, then it threshold should identify possible misclassifications as uncertain.

9.3.2 Binning Technique

The second technique considered was a variation on the thresholding technique. Instead of using one threshold and either assigning or rejecting a classification on the basis of the maximum network output, a classification was assigned to one of five categories:

$$\text{given } O_i(x,\Theta) > O_j(x,\Theta) \text{ for all } i \neq j \text{ and } 1 \leq j \leq N, \text{ then}$$

Rule	Category
$0.0 \leq O_i(x,\Theta) < 0.2$	$B_{0.0\text{-}0.2}$
$0.2 \leq O_i(x,\Theta) < 0.4$	$B_{0.2\text{-}0.4}$
$0.4 \leq O_i(x,\Theta) < 0.6$	$B_{0.4\text{-}0.6}$
$0.6 \leq O_i(x,\Theta) < 0.8$	$B_{0.6\text{-}0.8}$
$0.8 \leq O_i(x,\Theta) \leq 1.0$	$B_{0.8\text{-}1.0}$

If the network outputs do approximate Bayesian probabilities, these categories should provide an indication of certainty. For example, if a classification were assigned to $B_{0.0\text{-}0.2}$ then likelihood of this classification being correct would be less than if it were assigned to the $B_{0.8\text{-}1.0}$. As such it is postulated that misclassification would be low in categories $B_{0.8\text{-}1.0}$ and $B_{0.6\text{-}0.8}$ and conversely misclassification would be high in categories $B_{0.2\text{-}0.4}$ and $B_{0.0\text{-}0.2}$.

9.4 Experiments

The remainder of this chapter will present the application of the thresholding and binning techniques to the neural network experiments documents in chapters 7 and 8. These techniques will firstly be applied to the initial experiments considered in chapter 7 (E1.bp, E2.bp, E3.bp, and E4.bp) and then these results are compared with some of the alternative neural network and feature extraction techniques considered in chapter 8.

9.5 Problem 1

The first neural networks tested using the thresholding and binning techniques were the MLPs created in experiment *E1.bp*. As documented in chapter 7, the *E1.bp* experiment consisted of twenty training runs where each MLP was trained using the back-propagation algorithm to classify patients into one of the four BSM classses.

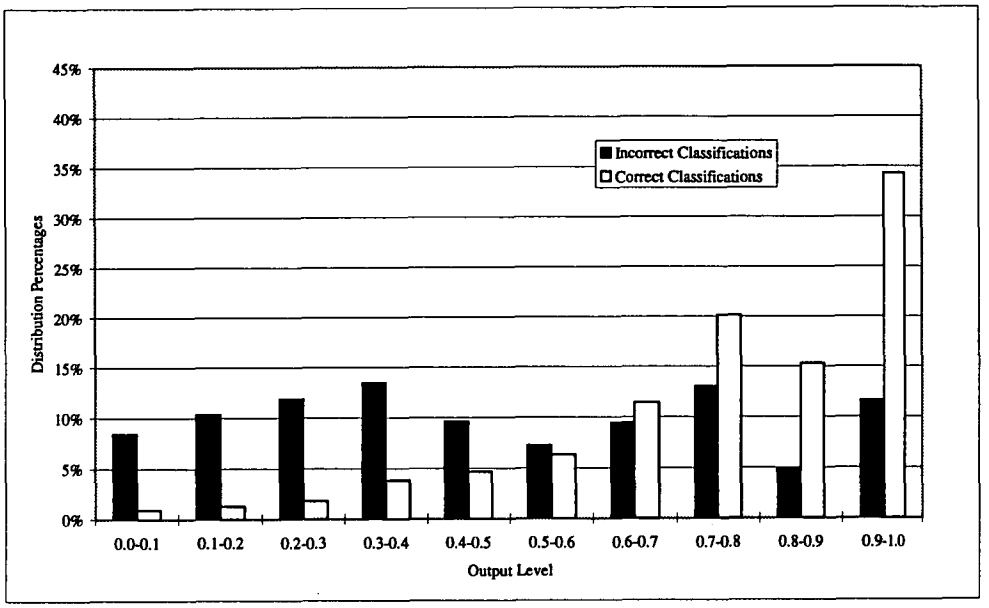
9.5.1 Preliminary Analysis of Network Outputs

Before applying the thresholding or binning techniques to the E1.bp experiment a preliminary analysis was conducted to determine whether network outputs provided any discrimination between correct and incorrect classifications. Clearly the premise on which these techniques are operating is that if a classification is unclear or uncertain then the network output should be significantly lower than for classification that is more certain. If this is true, then the thresholding and binning techniques many well prove useful.

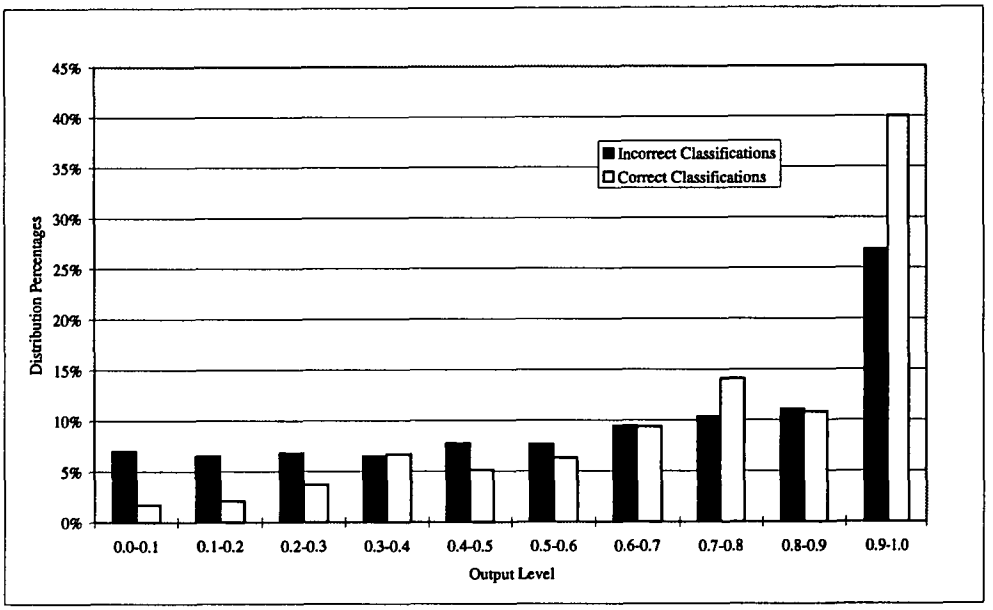
An initial analysis of network outputs shows that there is indeed a clear distinction between correct and incorrect classifications with respect to network outputs. The classifications in the E1.bp experiments were divided into correct and incorrect classifications and the network output levels for each classification were noted. An examination of the distribution of these network output levels shows a degree of difference between the output levels of those classifications that are incorrect as opposed to these classifications that are correct. A plot of correct and incorrect output levels for training and testing sets is given in Graphs 9-1 and 9-2.

On examination of the distribution of training set examples (Graph 9-1) a number of features are noted when comparing the number of correct and incorrect classifications in each grouping. For all classifications with output levels below 0.6 more patients are classified incorrectly than correctly. However when output levels are higher than 0.6 more patients are classified correctly than incorrectly. It is important to note that some correct classifications still present very low output values but overall the distribution of correct classifications are skewed toward high output levels. These results would suggest that the application of a threshold will provide some degree of discrimination between correct and incorrect classifications.

A similar distribution of output levels was observed for testing set results (Graph 9-2). Although the difference between the correct and incorrect distributions is not as clear as for the training set similar trends are observed. Of those classifications with outputs below 0.6 it is noted that more classifications are classified incorrectly than correctly (apart from classifications between 0.3 and 0.4). Similarly, it is observed that for classifications with outputs above 0.6 more are classified correctly than incorrectly (apart from classifications between 0.8 and 0.9). As observed for the training set, the distribution of outputs associated with correct classifications are skewed toward high output levels, however the differences between the correct and incorrect distributions are not as clear as those observed for the training set. However the testing set distributions still suggest that the application of a threshold will eliminate more incorrect classifications than correct classifications and therefore improve the classification performance for the remaining classifications.



Graph 9-1: Distribution of network output (E1.bp – training set)



Graph 9-2: Distribution of network outputs (E1.bp – testing set)

9.5.2 Thresholding Outputs

The thresholding technique was tested on the *E1.bp* experiments by applying several different threshold levels to the classification results and observing the impact on performance. Two measures were used to observe the impact of the threshold on classifier performance: *percentage classified* and *percentage correct.*

The *percentage classified* indicates the proportion of the classifications with output levels greater than or equal to the threshold applied. This measure indicates how many of the examples in the data set were assigned a classification. Therefore, for any threshold applied, (*100% - percentage classified*) indicates the proportion of examples rejected by the threshold and deemed uncertain.

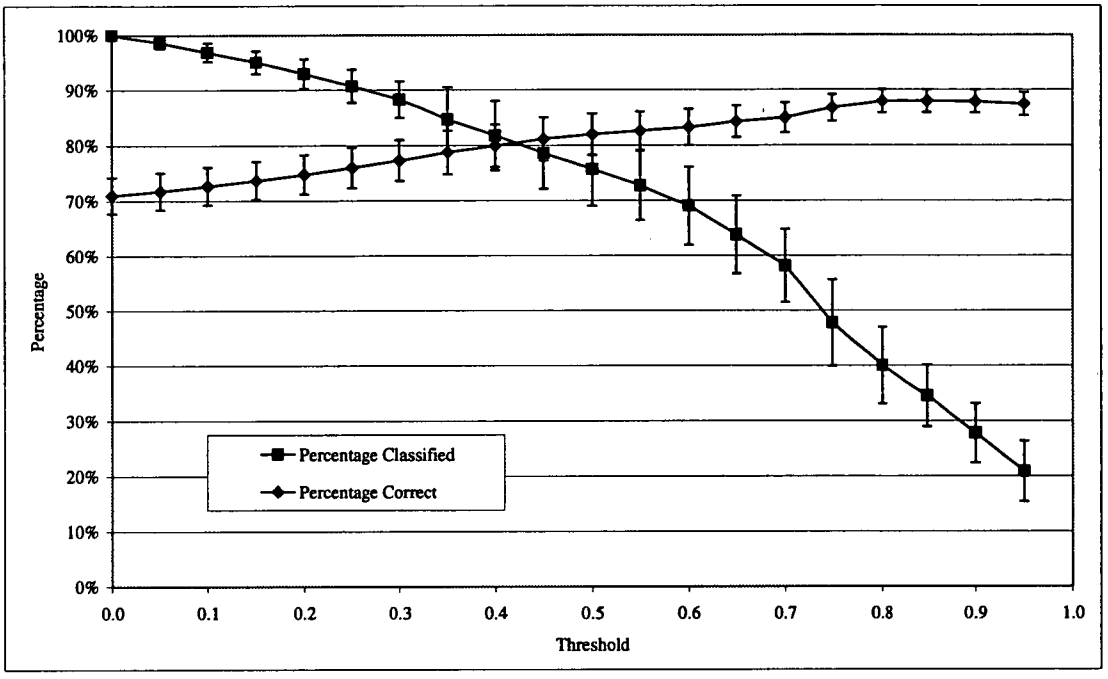
The *percentage correct* indicates the revised classification performance achieved for those classifications with outputs above the threshold - that is the percentage correct for the 'accepted' or 'certain' classifications.

When a threshold of t=0 is applied, then no classifications will be rejected. Therefore, the *percentage classified* will be 100% and the percentage correct will be the same as for the original experiment. As the threshold is increased it is hoped that more incorrect than correct classifications are rejected. If this is achieved, then the *percentage correct* measure should increase as the threshold is increased. This increase in performance will obviously result in some classifications being rejected, and therefore the *percentage classified* will progressively decrease as the threshold is increased.

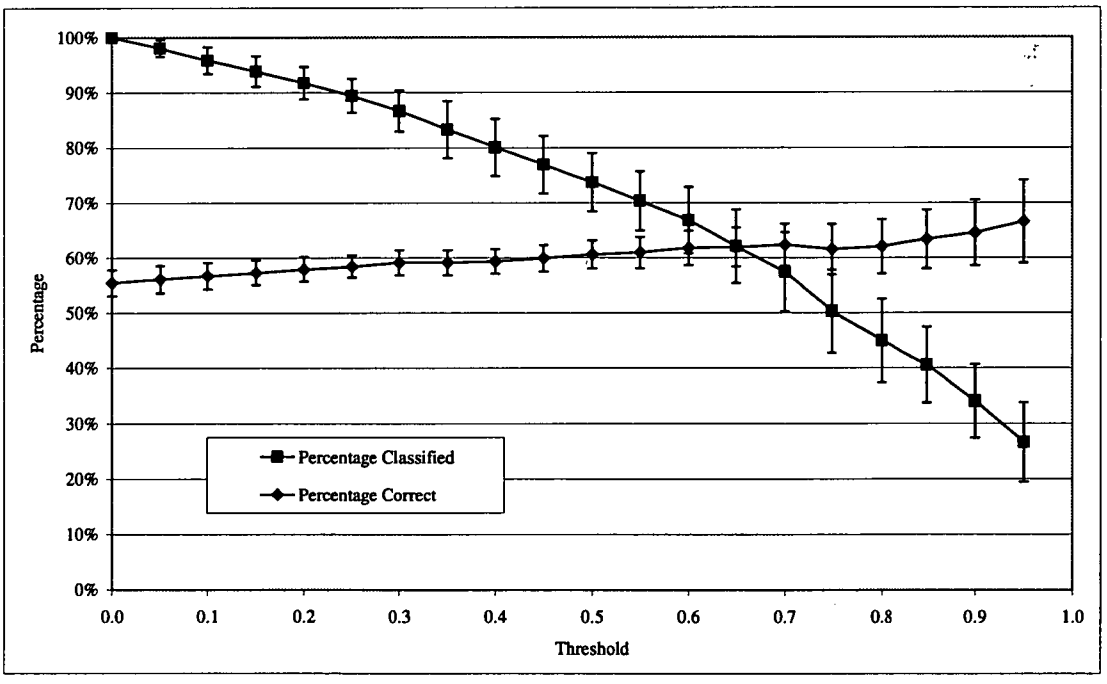
The thresholding technique was applied to the *E1.bp* MLPs starting with a threshold of t=0.0 and progressively increasing the threshold in 0.05 increments up until t=0.95. A threshold of t=1.0 was not applied, as this threshold simply rejected all classifications (percentage classified = 0%). Since twenty networks were originally trained in the original *E1.bp* experiment, the thresholds were applied to each network. The results presented are an indication of the average performance at each threshold value.

The thresholding results achieved when applied to experiment *E1.bp* are summarised in Graphs 9-3 and 9-4. As proposed, the percentage of correct classifications increases and the percentage classified decreases as the threshold is increased.

These results are encouraging and would suggest that the application of a threshold to network outputs provide a degree of discrimination between correct and incorrect classifications. For example, if a threshold of t=0.6 is applied to the training set classifications the classification performance (percentage correct) increases from 71.0% (±3.2%) (for t=0) to 83.3% (±3.3%) and the percentage of cases classified is 69.0% (±7.0%) [that is 31% (±7%) of classification were rejected or deemed uncertain]. Similarly, if a threshold of t=0.6 is applied to the testing set classifications the classification performance (percentage correct) increases from 55.5% (±2.3%) (for t=0) to 62.0% (±3.0%) and the percentage of cases classified is 67.0% (±6.0%) [that is 33% (±6%) of classifications are rejected or deemed uncertain].



Graph 9-3: Thresholding E1.bp classifers (training set).



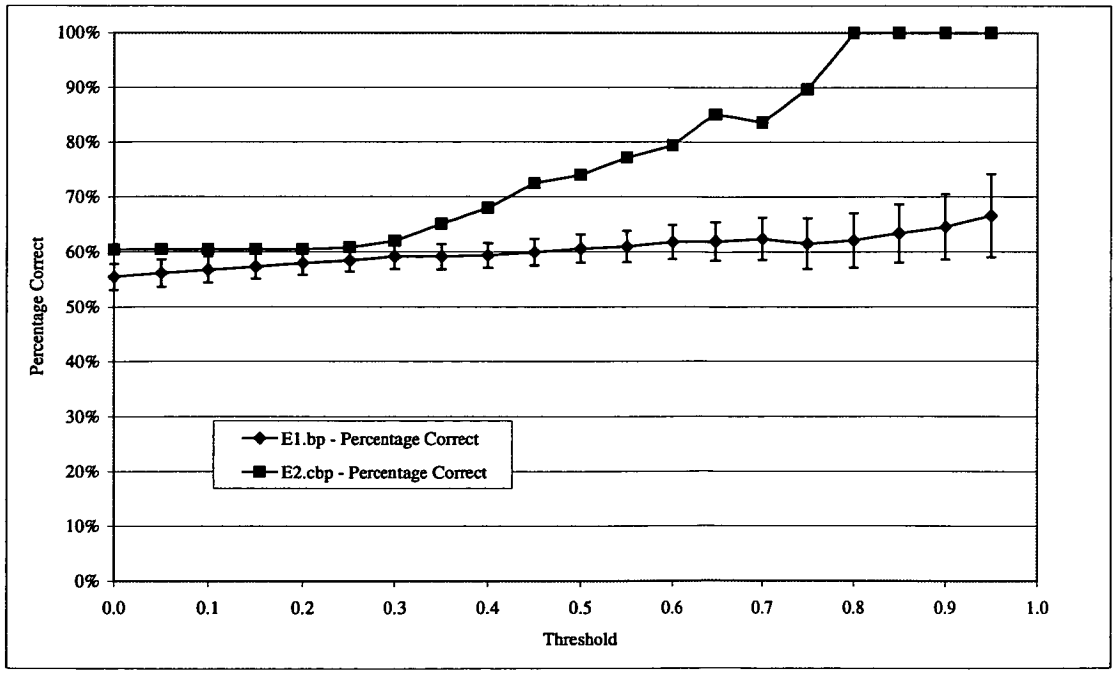
Graph 9-4:Thresholding E1.bp classifiers (testing set).

9.5.3 Comparing experiments E1.bp and E1.cbp

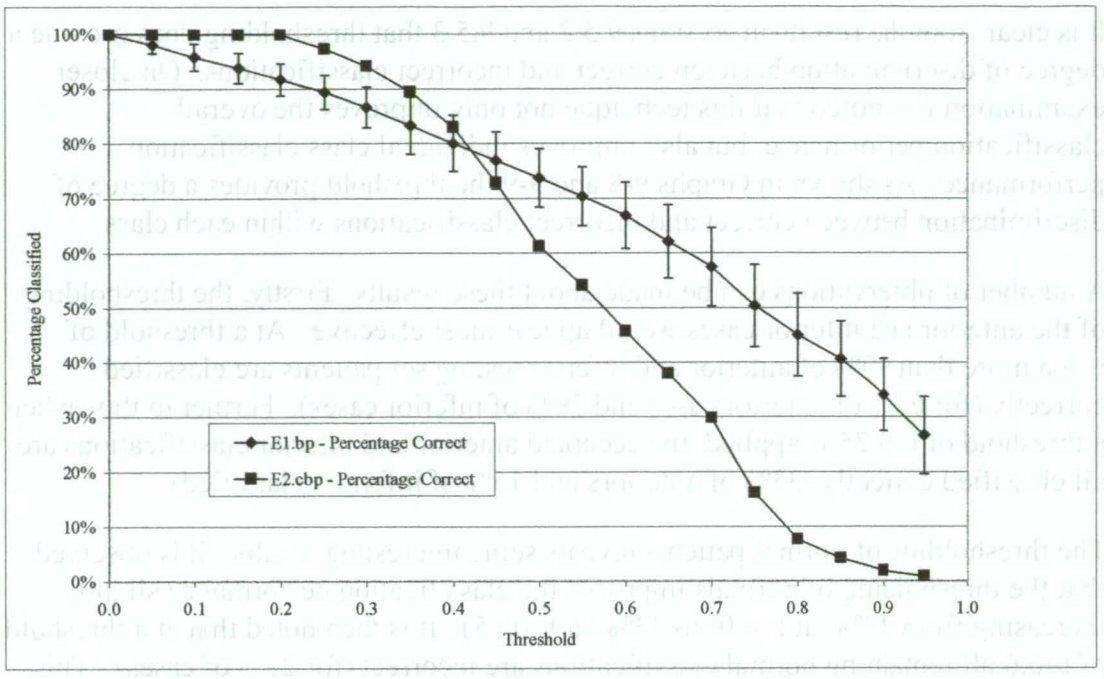
The thresholding performance of the E1.bp experiments was compared with the committee based classsifier E1.cbp. Since committee-based networks were found to perform better than single networks it was anticipated that the outputs of committee based networks would provide more discrimination between correct and incorrect classifications. This was found to be true.

Graph 9-5 and 9-6 provide a comparison of the thresholding response of experiments *E1.bp* and *E1.cbp*. It is observed that the *percentage correct* results gained when thresholding the *E1.cbp* network are significantly greater than those for the *E1.bp* networks. However, it is noted that although the *E1.cbp* rejects less classifications using low threshold levels (0.0 to 0.4), for higher thresholds the *E1.cbp* achieves higher classification results by rejecting a greater proportion of classifications. To understand this relationship more clearly it is easier to compare these results parametrically as the relationship between the *percentage correct* and *percentage classified* is important. The objective of a good rejection rule is to reject possible misclassifications but at the same time minimise the number of classifications rejected.

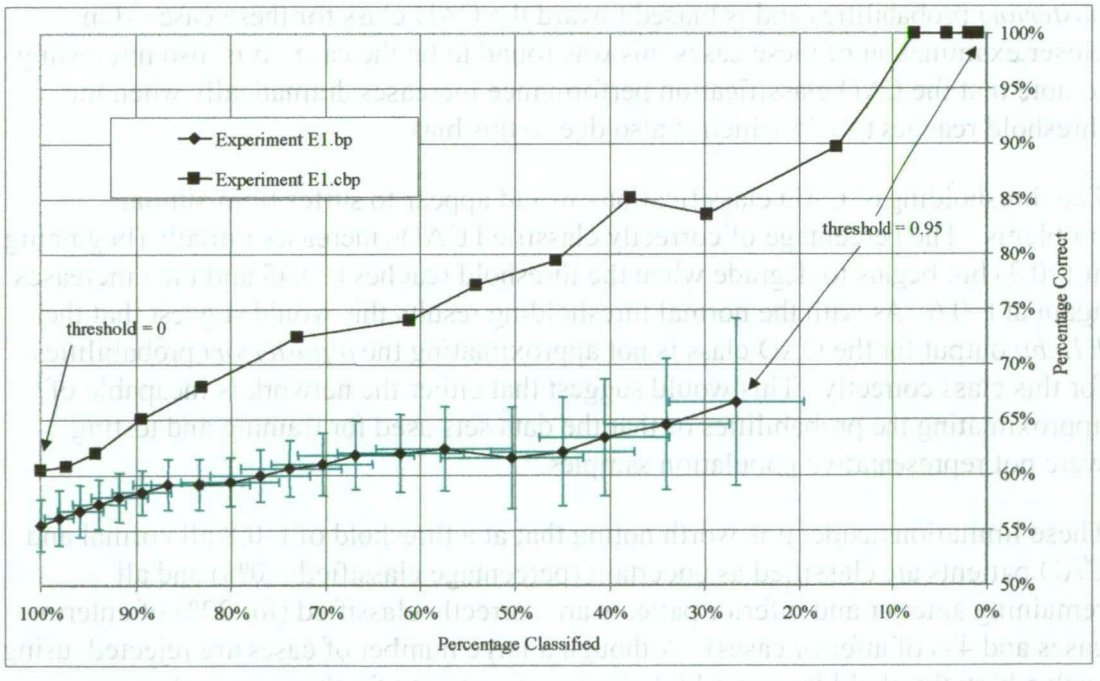
Given the parametric comparison in Graph 9-7 it is clear that the *E1.cbp* network performs better than the *E1.bp* networks for any given level of rejection. For example, if a threshold is set to reject 50% of classifications the *E1.cbp* network classifies more than 75% of the remaining patients correctly (t=between 0.55 and 0.6) whereas the *E1.bp* networks classify on average less than 65% of patients correctly (t=0.75). It was thus concluded that when using MLPs in conjunction with a rejection threshold, committee-based networks outperform single networks. This was found to be true not only in this particular case, but for all committee-based networks considered.



Graph 9-5: Comparing E1.cbp with E1.bp (percentage correct - testing set)



Graph 9-6: Comparing E1.cbp with E1.bp (percentage classified - testing set)



Graph 9-7: Comparison of thresholding for experiments E1.bp and E1.cbp (testing set).

9.5.4 Thresholding Class Results

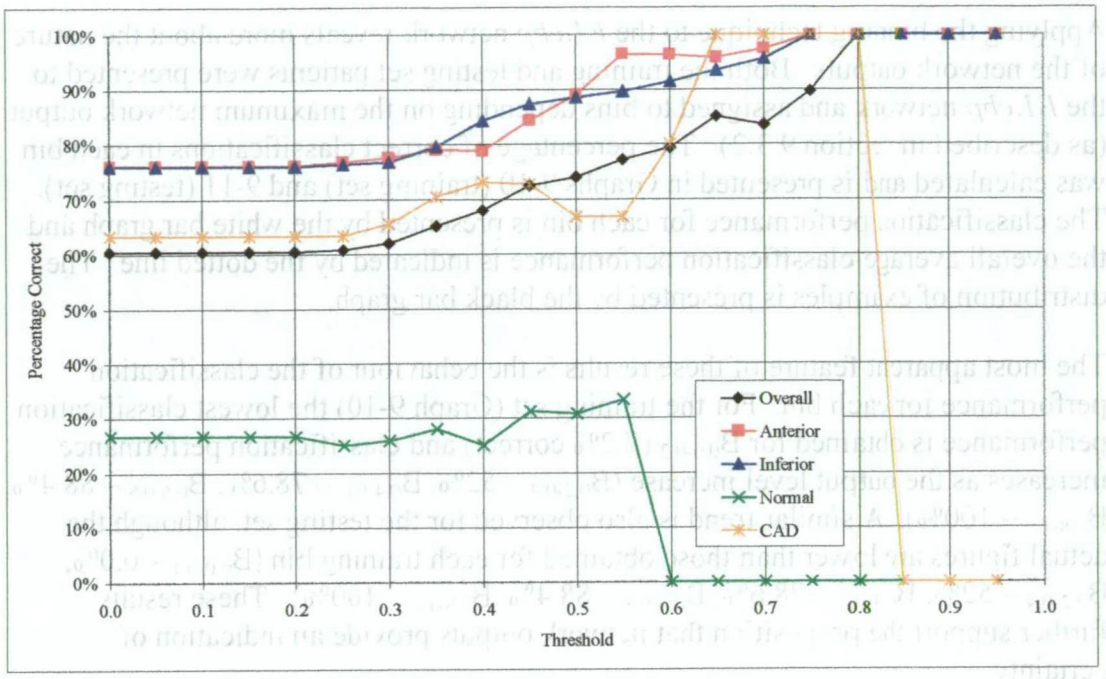
It is clear from the results in sections 9.5.2 and 9.5.3 that thresholding does provide a degree of discrimination between correct and incorrect classifications. On closer examination it is noted that this technique not only improves the overall classification performance, but also improves individual class classification performance. As shown in Graphs 9-8 and 9-9 the threshold provides a degree of discrimination between correct and incorrect classifications within each class.

A number of observations can be made about these results. Firstly, the thresholding of the anterior and inferior cases would appear most effective. At a threshold of t=0.6 more than 90% of anterior and inferior testing set patients are classified correctly (for 67% of anterior cases and 59% of inferior cases). Further to this, when a threshold of t=0.75 is applied, the accepted anterior and inferior classifications are all classified correctly (33% of anteriors and 15% of inferiors classified).

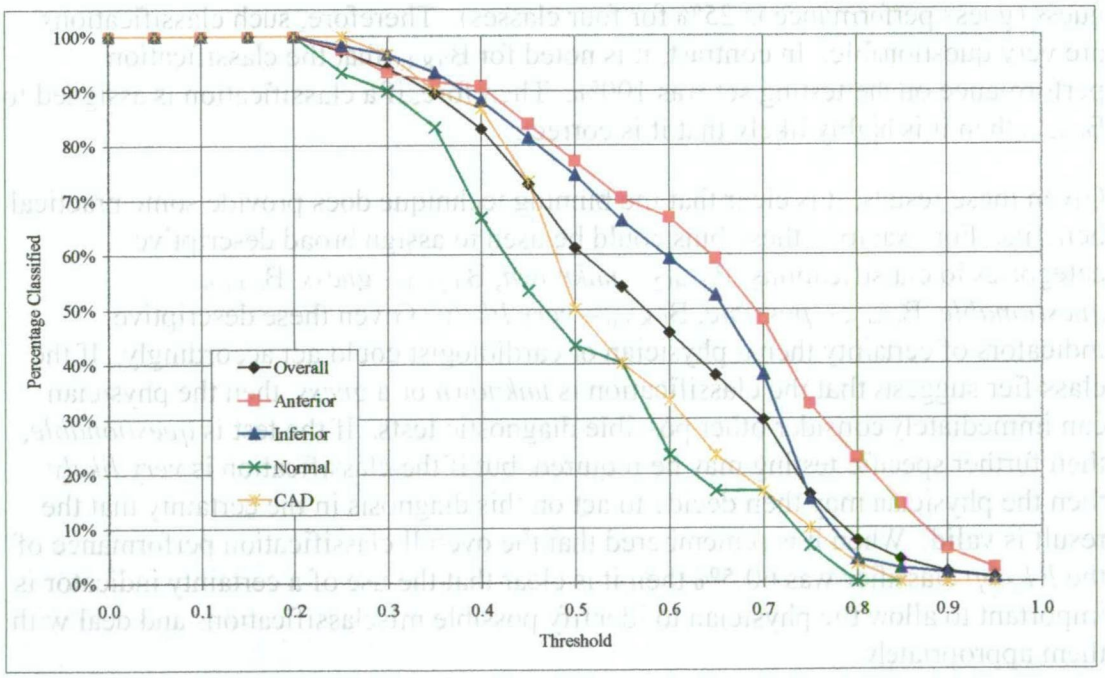
The thresholding of normal patients reveals some interesting results. It is observed that the thresholding of normals improves the classification performance slightly, increasing from 27% (at t=0.0) to 33% (at t=0.55). It is then noted that at a threshold of t=0.6 all remaining normal classifications are incorrect (for 23% of cases). This outcome is somewhat unexpected, as one would assume that if network outputs were approximating *a posteriori* probabilities that the percentage of correctly classified normals would continue to increase as the threshold is increased. Since this is not the case, then this would suggest that the network has not approximated the *a posteriori* probabilities and is biased toward the CAD class for these cases. On closer examination of these cases this was found to be the case. It is also interesting to note that the CAD classification performance increases dramatically when the threshold reaches t=0.55 which is also due to this bias.

The thresholding of CAD classifications would appear to suffer from similar problems. The percentage of correctly classified CADs increases initially (beginning at t=0.3) but begins to degrade when the threshold reaches t=0.45 and then increases again at t=0.6. As with the normal thresholding results this would suggest that the *E1.cbp* output for the CAD class is not approximating the *a posteriori* probabilities for this class correctly. This would suggest that either the network is incapable of approximating the probabilities or that the data sets used for training and testing were not representative population samples.

These limitations aside, it is worth noting that at a threshold of t=0.8 all normal and CAD patients are classified as uncertain (percentage classified = 0%) and all remaining anterior and inferior patients are correctly classified (for 22% of anterior cases and 4% of inferior cases). Although a large number of cases are rejected, using such a high threshold insures a high degree of certainty for those cases that are accepted. The possible uses of this outcome in the diagnostic setting will be discussed later in the chapter.



Graph 9-8: Thresholding results for experiment E1.cbp (testing set- percentage correct)



Graph 9-9: Thresholding results for experiment E1.cbp (testing set – percentage classified)

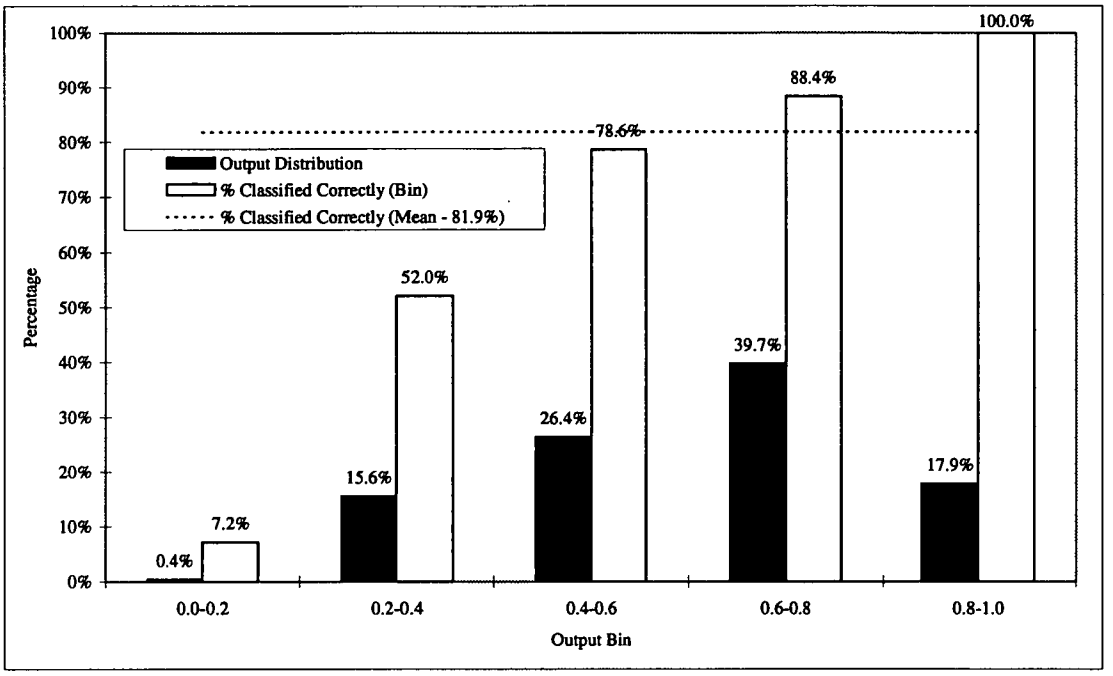
9.5.5 Binning Outputs

Applying the binning technique to the *E1.cbp* network reveals more about the nature of the network outputs. Both the training and testing set patients were presented to the *E1.cbp* network and assigned to bins depending on the maximum network output (as described in section 9.3.2). The percentage of correct classifications in each bin was calculated and is presented in Graphs 9-10 (training set) and 9-11 (testing set). The classification performance for each bin is presented by the white bar graph and the overall average classification performance is indicated by the dotted line. The distribution of examples is presented by the black bar graph.

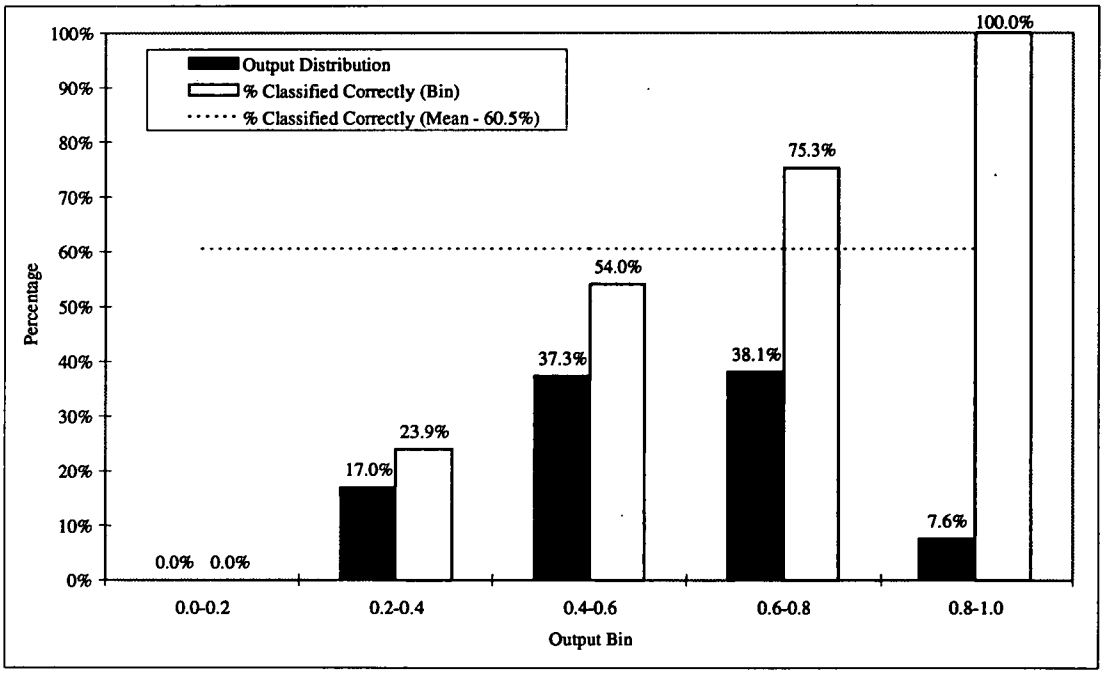
The most apparent feature of these results is the behaviour of the classification performance for each bin. For the training set (Graph 9-10) the lowest classification performance is obtained for $B_{0.0\text{-}0.2}$ (7.2% correct) and classification performance increases as the output level increase ($B_{0.2\text{-}0.4}$ – 52%, $B_{0.4\text{-}0.6}$ – 78.6%, $B_{0.6\text{-}0.8}$ – 88.4%, $B_{0.8\text{-}1.0}$ – 100%). A similar trend is also observed for the testing set, although the actual figures are lower than those obtained for each training bin ($B_{0.0\text{-}0.1}$ – 0.0%, $B_{0.2\text{-}0.4}$ – 52%, $B_{0.4\text{-}0.6}$ – 78.6%, $B_{0.6\text{-}0.8}$ – 88.4%, $B_{0.8\text{-}1.0}$ – 100%). These results further support the proposition that network outputs provide an indication of certainty.

A number of other specific observations can be made. It is noted for $B_{0.2\text{-}0.4}$ that the classification performance on the testing set was 23.9%. Since this is a four class problem then this would suggest that these classifications would be no better than a guess (guess performance is 25% for four classes). Therefore, such classifications are very questionable. In contrast, it is noted for $B_{0.8\text{-}1.0}$ that the classification performance on the testing set was 100%. Therefore, if a classification is assigned to $B_{0.8\text{-}1.0}$ then it is highly likely that it is correct.

Given these results, it is clear that the binning technique does provide some practical benefits. For example, these bins could be used to assign broad descriptive categories to classifications: $B_{0.0\text{-}0.2}$ – *unknown*, $B_{0.2\text{-}0.4}$ – *guess*, $B_{0.4\text{-}0.6}$ – *questionable*, $B_{0.6\text{-}0.8}$ – *possible*, $B_{0.8\text{-}1.0}$ – *very likely*. Given these descriptive indicators of certainty then a physician or cardiologist could act accordingly. If the classifier suggests that the classification is *unknown* or a *guess*, then the physician can immediately consider other possible diagnostic tests. If the test is *questionable*, then further specific testing may be required, but if the classification is *very likely* then the physician may then decide to act on this diagnosis in the certainty that the result is valid. When it is remembered that the overall classification performance of the *E1.cbp* classifier was 60.5% then it is clear that the use of a certainty indicator is important to allow the physician to identify possible misclassifications and deal with them appropriately.



Graph 9-10: Binning results for E1.cbp (training set).



Graph 9-11: Binning results for E1.cbp (testing set).

9.5.6 Binning Classes Results

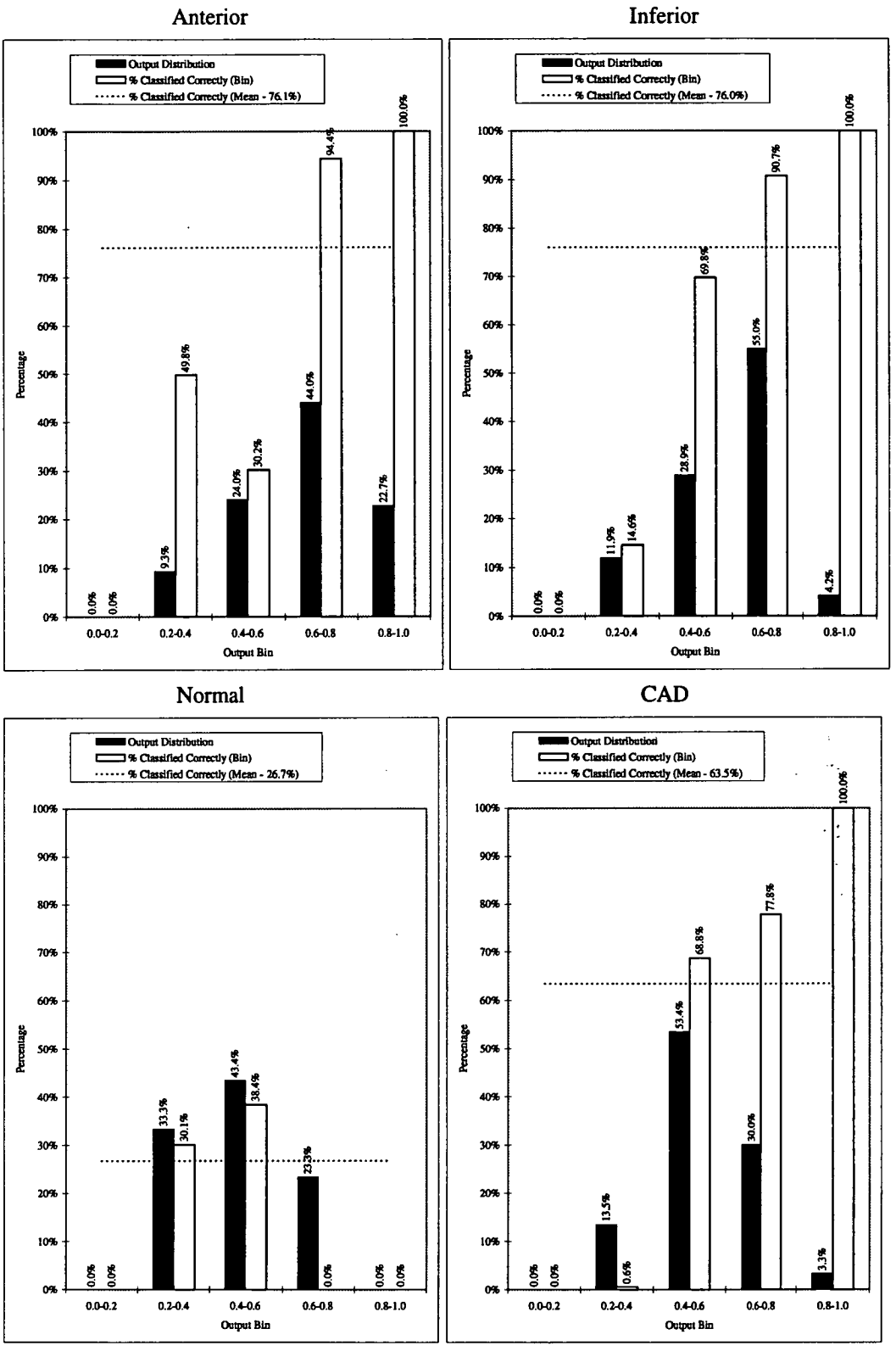
The binning technique was also applied to the individual class results for experiment *E1.cbp*. Graph 9-12 presents the binning of anterior, inferior, normal and CAD patients in the testing set.

For inferior and CAD patients the classification performance is observed to increase as higher output levels are considered. However, for patients with anterior infarcts it is noted that the percentage of correct classifications for $B_{0.2\text{-}0.4}$ (49.8% correct) is higher than that achieved for $B_{0.4\text{-}0.6}$ (24% correct). For the normal patients the overall output levels are clearly low compared to the other classes, and as observed in the thresholding experiment all $B_{0.6\text{-}0.8}$ classifications are incorrect. As mentioned in the thresholding experiments, this result for the normal class suggests that the network output is not approximating *a posteriori* probabilities for the normal classifications.

The classification of anterior and inferior infarctions in $B_{0.6\text{-}0.8}$ and $B_{0.8\text{-}1.0}$ is a significant improvement over the mean classification performance for these classes. For $B_{0.6\text{-}0.8}$, 94.4% of anterior infarcts are classified correctly and 90.7% of inferior infacts are classified correctly. For $B_{0.8\text{-}1.0}$, all anterior and inferior infarcts are classified correctly.

In relation to the classification distributions a number of observations can be made. For the anterior and inferior patients more than half the cases present output levels that are above 0.6 (anterior – 66.7% of cases above 0.6, inferior – 59.2% of cases above 0.6). In contrast, the opposite is observe for the CAD and normal classes (76.7% of normal cases and 66.9% of CAD cases are associated with network output levels are below 0.6). This would suggest in broad terms that the network is more confident about anterior and inferior classification than normal and CAD classifications which is understandable considering the clear overlap in the CAD and normal classes observed in chapter 7 and 8.

Overall, it is interesting to note the differences in the classification performance for each class in relation to each bin. Although the classification of anterior and inferior infarctions is above 90% for $B_{0.6\text{-}0.8}$, a similar result is not achieved for CAD (77.8%) and normal (0%) patients. This would suggest that when interpreting the certainty of a classification this should be considered with respect to the classification assigned. This would suggest that the use of different bin groupings for each class may be more appropriate.



Graph 9-12: Binning results for classes (E1.cbp - testing set)

9.6 Problem 2/3

In essence problems 2 and 3 are very similar. Both examine the three-class problem (separating anterior, inferior and normal classes), the difference being that problem 2 uses the follow-up infarct patient during training whereas problem 3 omits follow-up patients from the training process. Apart from this distinction in training, the classifiers are still trained to assign patients to one of three classes.

The thresholding and binning techniques were applied to experiments *E2.bp*, *E2.cbp*, *E3.bp*, and *E3.cbp*. As found for *problem 1*, the committee-based neural networks provided better discrimination between correct and incorrect classifications than the individual networks applied to the same problem. Further to this it was found that overall the *E3.cbp* network produced the best thresholding and binning results. Therefore, the results presented in the following section are in relation to experiment *E3.cbp*.

9.6.1 Thresholding Outputs

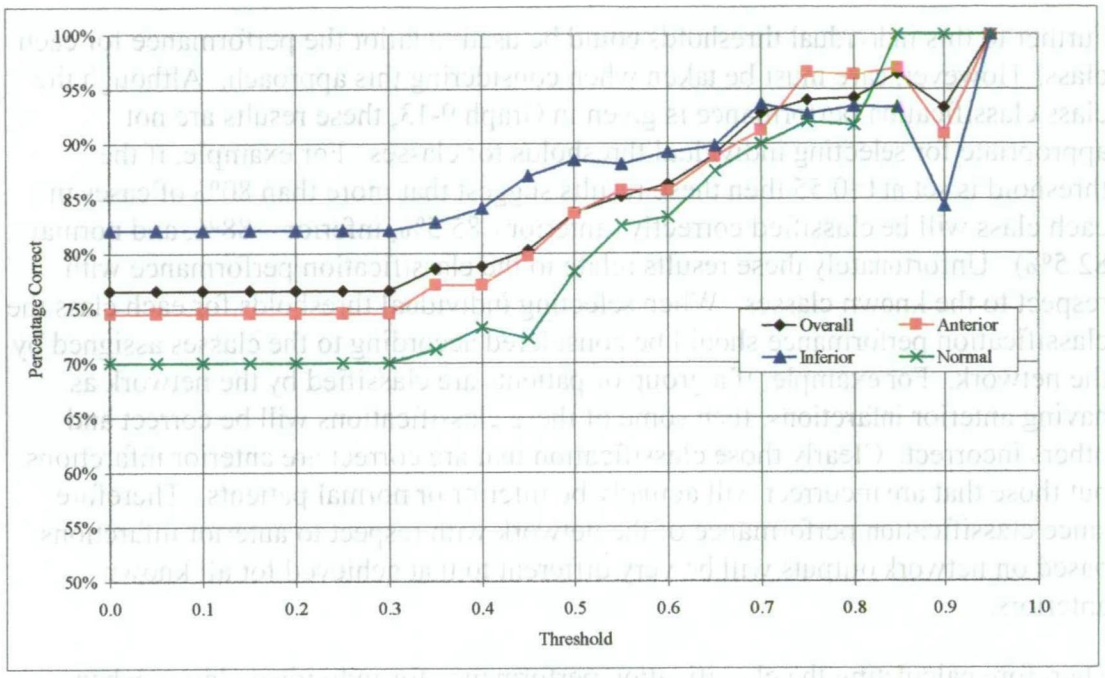
The thresholding results for experiment *E3.cbp* are presented in Graphs 9-13 and 9-14. These results present the overall and individual class thresholding results for the testing set.

As found when thresholding the *E1.cbp* network, the application of the threshold to the *E3.cbp* network improved the classification performance. It is noted that thresholding not only improved the overall classification performance but also improved the classification performance for individual classes (Graph 9-13).

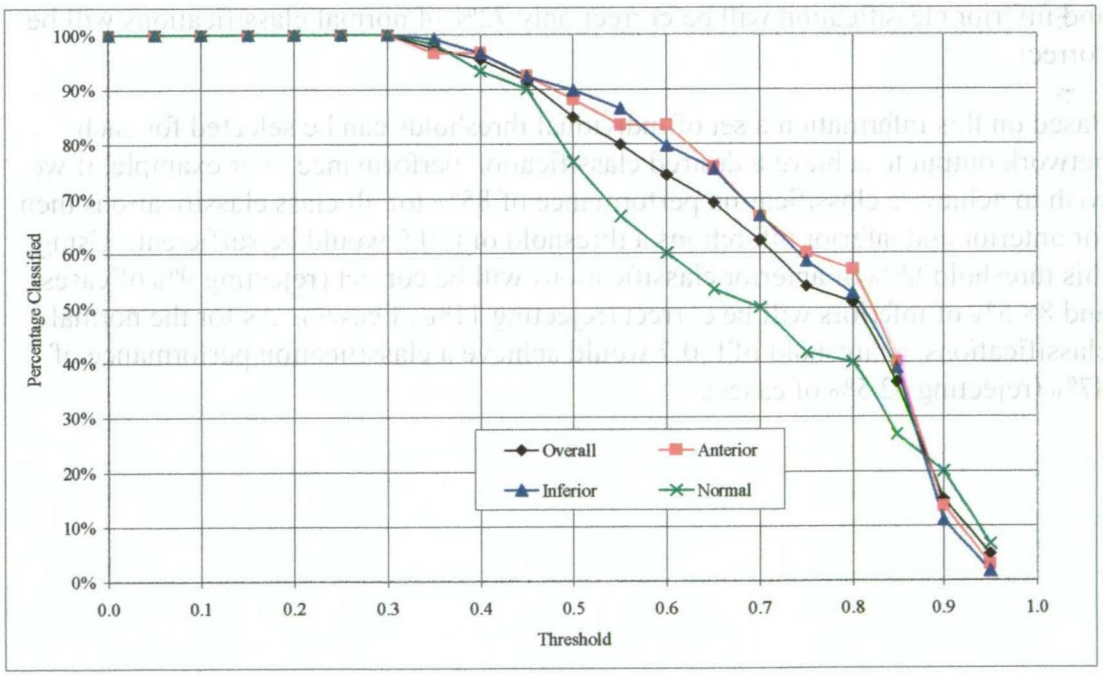
Although direct comparisons cannot be drawn between the *E1.cbp* and *E3.cbp* thresholding results, a number of observations can be made. In both thresholding experiments the classification performance for the anterior and inferior infarct classes was improved by the applying a threshold. However, the thresholding of the normal class results produced very different results in each case.

For experiment *E1.cbp* the application of a threshold did not improve the percentage of normal patients classified correctly (Graph 9-8) and the output distribution suggested that for normal patients the network outputs did not approximate *a posteriori* probabilities. By comparison the application of a threshold to the *E3.cbp* network did improve the percentage of normal patients classified correctly. This would suggest that the *E3.cbp* network is managing to approximate the *a posteriori* probabilities for the normal class.

Given this improvement, it is clear that thresholding the *E3.cbp* network is extremely useful and allows the reliability of the network to be improved significantly. For example, if a classification performance of 80% is required, then setting the threshold at t=0.45 would achieve this result (80.5% classified correctly with 8% of classifications rejected).



Graph 9-13: Percentage correct thresholding results for experiment E3.cbp based on known class (testing set)

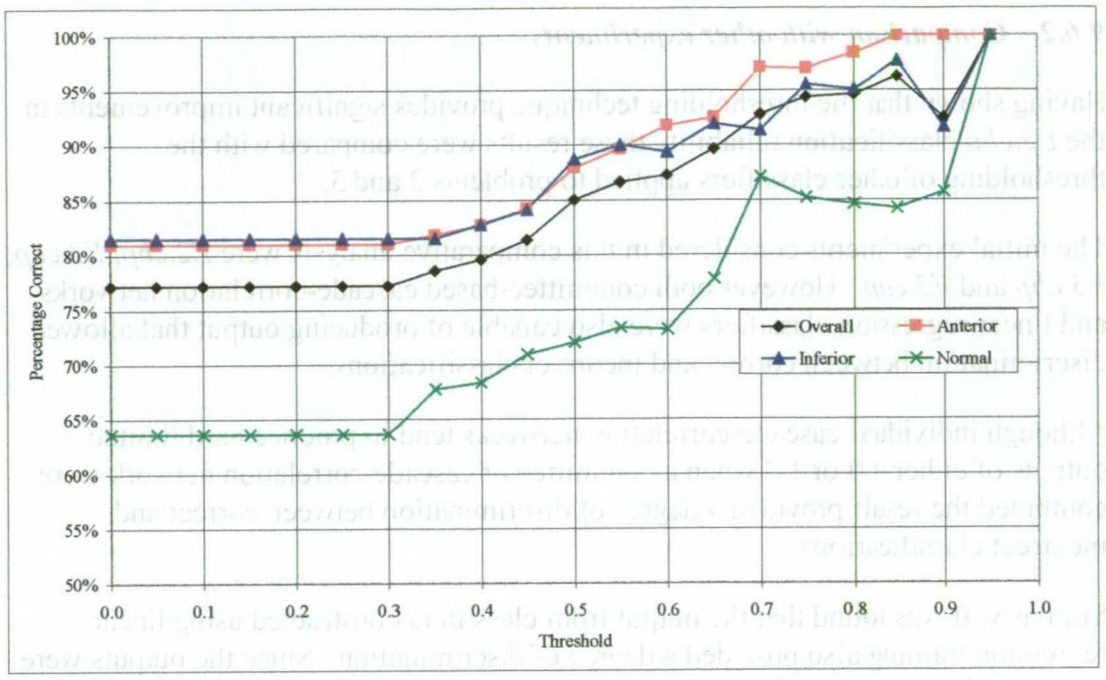


Graph 9-14: Percentage classified thresholding results for experiment E3.cbp based on known class (testing set)

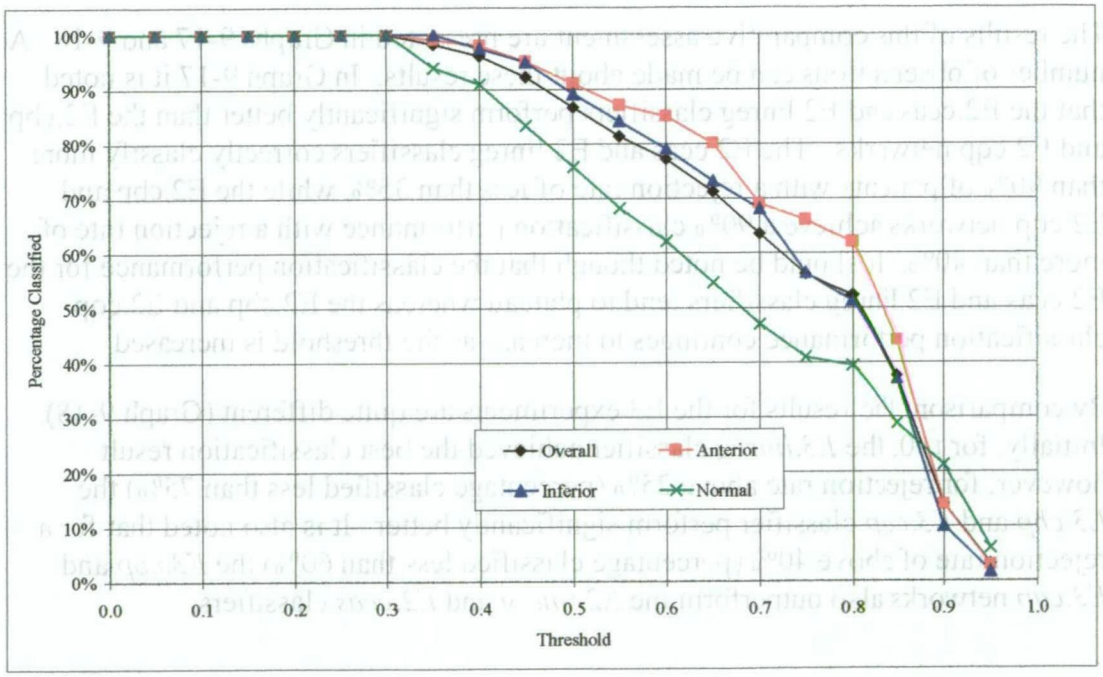
Further to this individual thresholds could be used to tailor the performance for each class. However, care must be taken when considering this approach. Although the class classification performance is given in Graph 9-13, these results are not appropriate for selecting individual thresholds for classes. For example, if the threshold is set at t=0.55 then these results suggest that more than 80% of cases in each class will be classified correctly (anterior - 85.5%, inferior – 88%, and normal 82.5%). Unfortunately these results relate to the classification performance with respect to the known classes. When selecting individual thresholds for each class the classification performance should be considered according to the classes assigned by the network. For example, if a group of patients are classified by the network as having anterior infarctions, then some of these classifications will be correct and others incorrect. Clearly those classification that are correct are anterior infarctions, but those that are incorrect will actually be inferior or normal patients. Therefore since classification performance of the network with respect to anterior infarctions based on network outputs will be very different to that achieved for all known anteriors.

Therefore calculating the classification performance for individual classes when thresholding should be calculated with respect to the classifications assigned by the network. In the case of the *E3.cbp* experiment these performance figures were calculated for each class and are presented in Graphs 9-15 and 10 –16. It is clear from these results that a threshold setting of t=0.55 is not sufficient to achieve an 80% classification performance for all classes. Although more than 85% of anterior and inferior classification will be correct only 72% of normal classifications will be correct.

Based on this information a set of individual thresholds can be selected for each network output to achieve a desired classification performance. For example, if we wish to achieve a classification performance of 85% for all class classifications then for anterior and inferior infarctions a threshold of t=0.5 would be sufficient. Using this threshold 88% of anterior classifications will be correct (rejecting 9% of cases) and 88.5% of inferiors will be correct (rejecting 11% of cases). As for the normal classifications, a threshold of t=0.7 would achieve a classification performance of 87% (rejecting 52.5% of cases).



Graph 9-15: Thresholding results for experiment E3.cbp based on network classification(testing set)



Graph 9-16: Thresholding results for experiment E3.cbp based on network classification(testing set)

9.6.2 Comparison with other experiments

Having shown that the thresholding technique provides significant improvements in the *E3.cbp* classification reliability these results were compared with the thresholding of other classifiers applied to problems 2 and 3.

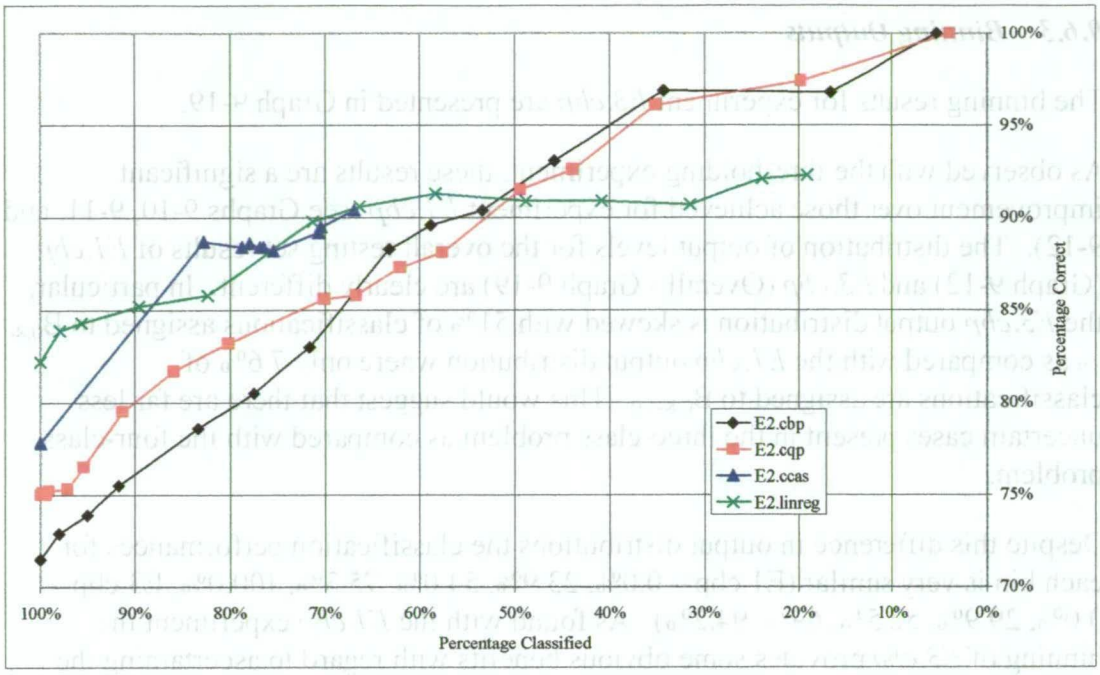
The initial experiments considered in this comparative analysis were *E2.cbp*, *E2.cbp*, *E3.cbp* and *E3.cqp*. However both committee-based cascade-correlation networks and linear regression classifiers were also capable of producing output that allowed discrimination between correct and incorrect classifications.

Although individual cascade-correlation networks tend to produce hard limited outputs of either 0.0 or 1.0 when a committee of cascade-correlation network were combined the result provided a degree of discrimination between correct and incorrect classifications.

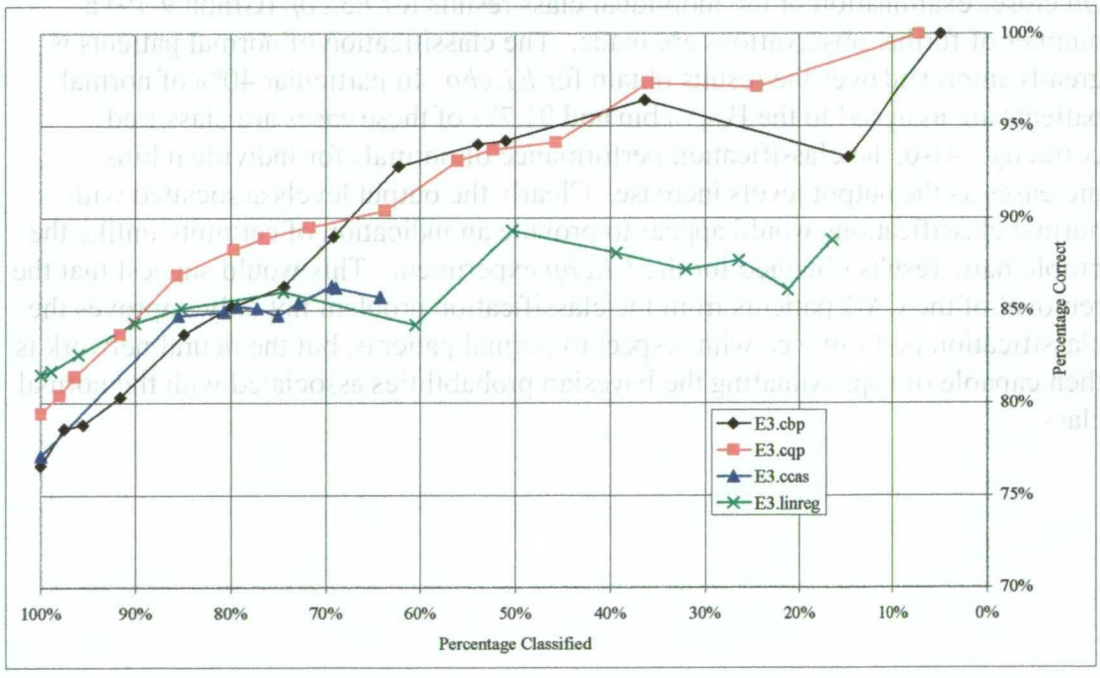
Similarly, it was found that the output from classifiers constructed using linear regression training also provided a degree of discrimination. Since the outputs were not limited by a transfer function it was necessary to normalise the outputs of the linear regression classifiers to produce an output in the range of 0.0 to 1.0. Apart from this simple modification, no other changes were made to the linear regression classifiers.

The results of this comparative assessment are presented in Graphs 9-17 and 9-18. A number of observations can be made about these results. In Graph 9-17 it is noted that the E2.ccas and E2.linreg classifiers perform significantly better than the E2.cbp and E2.cqp networks. The E2.ccas and E2.linreg classifiers correctly classify more than 90% of patients with a rejection rate of less than 35%, while the E2.cbp and E2.cqp networks achieve a 90% classification performance with a rejection rate of more than 40%. It should be noted though that the classification performance for the E2.ccas and E2.linreg classifiers tend to plateau whereas the E2.cbp and E2.cqp classification performance continues to increase as the threshold is increased.

By comparison, the results for the E3 experiments are quite different (Graph 9-18). Initially, for t=0, the *E3.linreg* classifier achieved the best classification result however, for rejection rate above 25% (percentage classified less than 75%) the *E3.cbp* and *E3.cqp* classifier perform significantly better. It is also noted that for a rejection rate of above 40% (percentage classified less than 60%) the *E3.cbp* and *E3.cqp* networks also outperform the *E2.linreg* and *E2.ccas* classifiers.



Graph 9-17: Parametric comparison of thresholding for E2 experiments (testing set)



Graph 9-18: Parametric comparison of thresholding for E3 experiments (testing set)

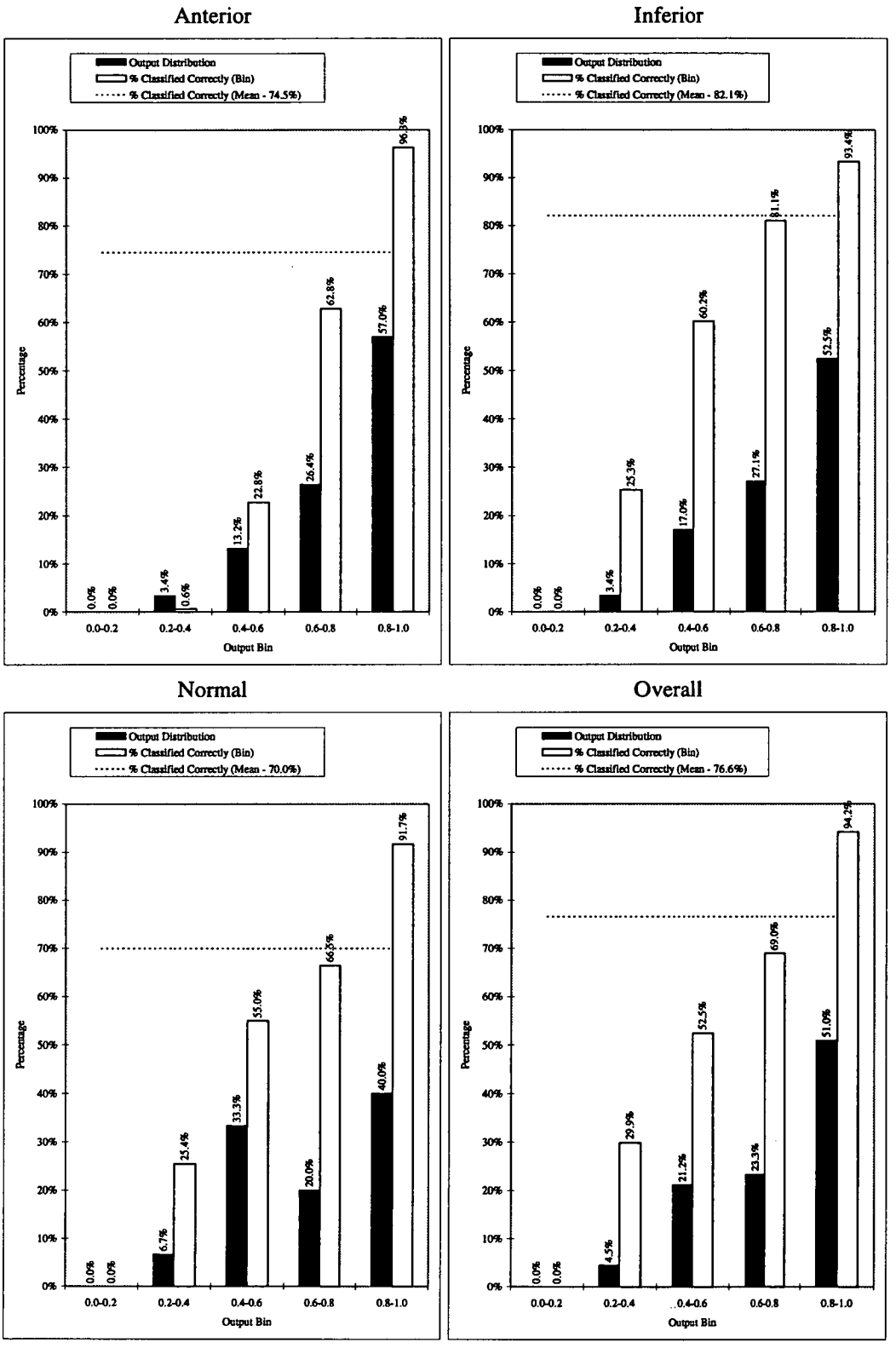
9.6.3 Binning Outputs

The binning results for experiment *E3.cbp* are presented in Graph 9-19.

As observed with the thresholding experiment, these results are a significant improvement over those achieved for experiment *E1.cbp* (see Graphs 9-10, 9-11, and 9-12). The distribution of output levels for the overall testing set results of *E1.cbp* (Graph 9-12) and *E3.cbp* (Overall - Graph 9-19) are clearly different. In particular, the *E3.cbp* output distribution is skewed with 51% of classifications assigned to $B_{0.8\text{-}1.0}$ as compared with the *E1.cbp* output distribution where only 7.6% of classifications are assigned to $B_{0.8\text{-}1.0}$. This would suggest that there are far less uncertain cases present in the three-class problem as compared with the four-class problem.

Despite this difference in output distributions the classification performances for each bin is very similar (E1.cbp – 0.0%, 23.9%, 54.0%, 75.3%, 100.0%; E3.cbp – 0.0%, 29.9%, 52.5%, 69%, 94.2%). As found with the *E1.cbp* experiment the binning of *E3.cbp* provides some obvious benefits with regard to ascertaining the certainty of a classification. Although a 100% classification performance is not achieved for $B_{0.8\text{-}1.0}$ the result of 94.2% (for 51% of cases) is still significantly higher than the mean classification performance of 76.6%.

On closer examination of the individual class results for *E3.cbp* (Graph 9-19) a number of further observations are made. The classification of normal patients is greatly improved over the results obtain for *E1.cbp*. In particular 40% of normal patients are assigned to the $B_{0.8\text{-}1.0}$ bin and 91.7% of these cases are classified correctly. Also, the classification performance of normals for individual bins increases as the output levels increase. Clearly the output levels associated with normal classifications would appear to provide an indication of certainty unlike the problematic results obtained for the *E1.cbp* experiment. This would suggest that the removal of the CAD patients from the classification problem not only improves the classification performance with respect to normal patients, but the neural network is then capable of approximating the bayesian probabilities associated with the normal class.



Graph 9-19: Binning results for classes (E3.cbp - testing set)

9.7 Problem 4

The thresholding and binning techniques were finally applied to the somewhat more difficult problem of discrimination normal patients from patients with coronary artery disease (problem 4). The initial assessment of these techniques was conducted using the best network from experiment *E4.bp* (see chapter 7 for details). The *E4.cbp* experiment was not considered for this initial test since the *E4.bp* network performed significantly better (see chapter 8 for details).

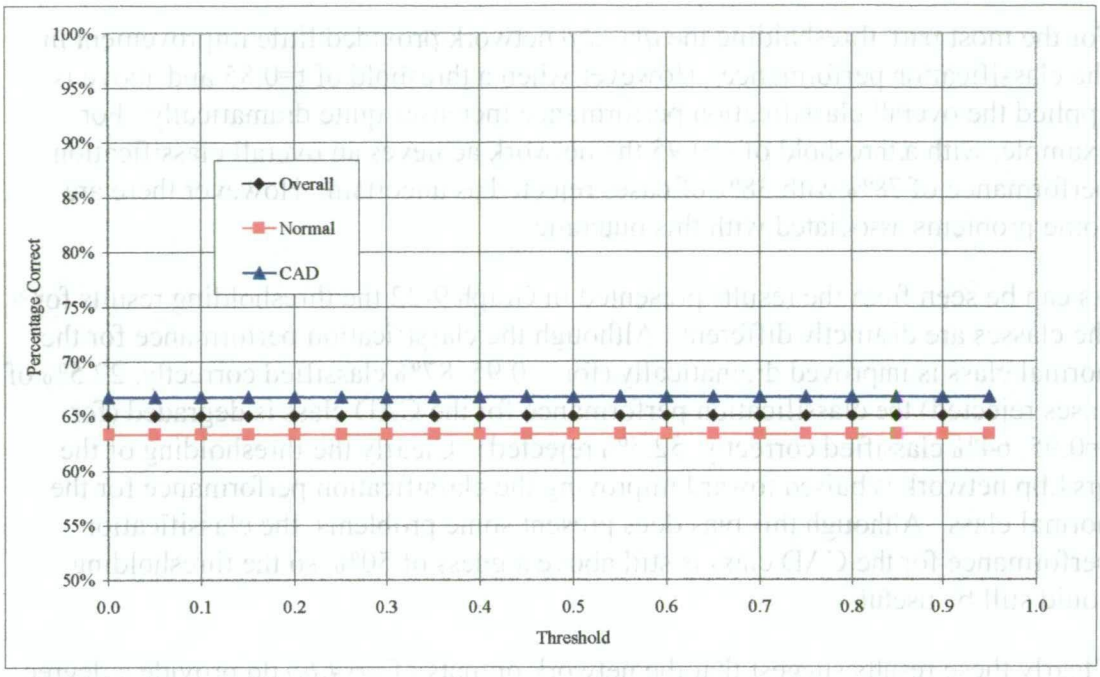
9.7.1 Thresholding Experiments

The thresholding results for experiment *E4.bp* were found to be dramatically different from those observed in the previous thresholding tests. As presented in Graphs 9-20 and 9-21 it was found that the application of a threshold to the *E4.bp* network outputs had no impact on the classification performance. It was observed that even at a threshold of t=0.95 no classifications were rejected. On closer examination it was found that for all classifications the associated network outputs were greater than 0.98, which clearly was why the thresholding did not reject any classifications. This is due to the fact that the output vectors generated by the network were extremely polarised. When the network classified a patient as normal, then the normal output was very near to 1.0 and the CAD output was close to 0.0. Similarly, when the network classified a patient as having CAD the normal output was very near to 0.0 and the CAD output was very near to 1.0. Interestingly the average sum of the output vector was found to be 0.995±0.005 suggesting that the network outputs were still normalised.

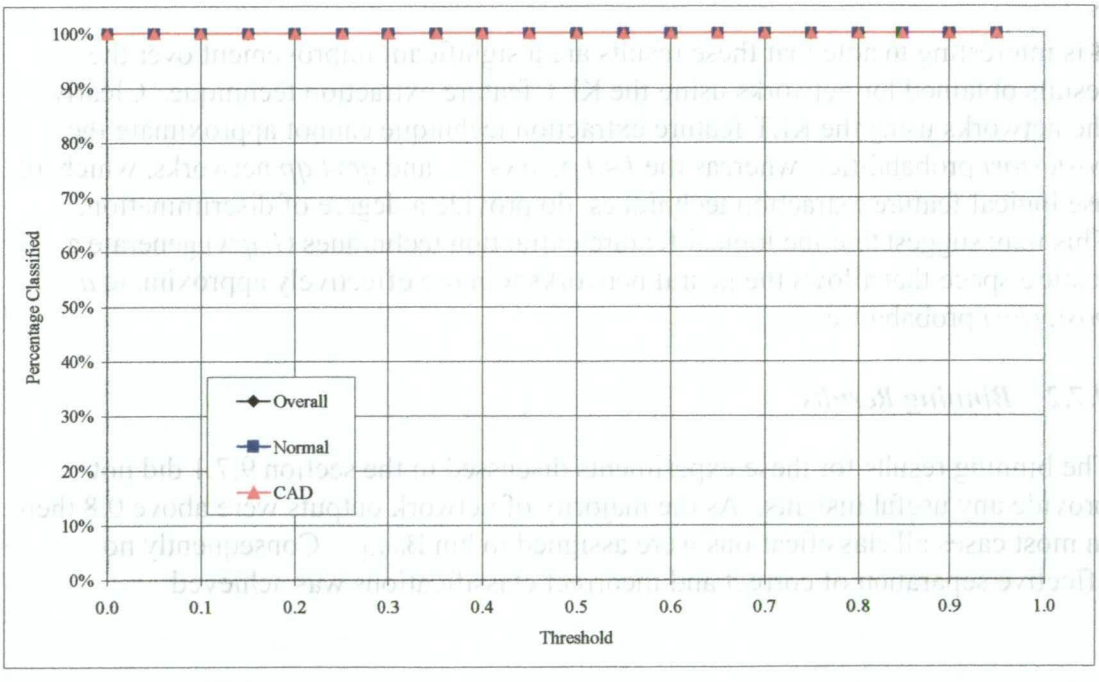
This result clearly suggests that the *E4.bp* network was unable to approximate the *a posteriori* probabilities. With an average classification performance of 65% it is clear that these two classes are overlapping in the feature space, but no discrimination was possible between correct and incorrect classifications.

At first it was considered that this might be an aberrant result. So the thresholding technique was applied to a number of other neural network classifiers applied to problem 4. To begin with thresholding was applied to the *E4.qp*, *E4.cas*, *E4.cbp*, *E4.cqp*, *E.ccas* networks (which constituted all the neural network techniques applied to problem 4 using the KLT feature extraction technique). All these experiments presented the same problem. Network outputs were extremely polarised and the threshold provided no discrimination between correct and incorrect classifications.

To explore this problem further all remaining neural networks applied to this problem were considered (*L4.bp, L4.qp, L4.cas, L4.cbp, L4.cqp, L4.ccas, qrs4.bp, qrs4.qp, qrs4.cas, qrs4.cbp, qrs4.cqp, qrs4.ccas, st4.bp, st4.qp, st4.cas, st4.cbp, st4.cqp, and st4.ccas*). For the majority of these experiment the same outcome was observed. However for experiments *L4.bp*, *qrs4.bp*, and *qrs4.qp* a degree of discrimination between correct and incorrect classifications was possible. However there are some problems associated with these results that need to be considered with care. To highlight the problems the thresholding of the *qrs4.bp* network is presented in Graphs 9-22 and 9-23.



Graph 9-18: Thresholding results for experiment E4.bp (testing set)



Graph 9-19: Thresholding results for experiment E4.bp (testing set)

For the most part, thresholding the *qrs4.bp* network provided little improvement in the classification performance. However when a threshold of t=0.85 and above is applied the overall classification performance increases quite dramatically. For example, with a threshold of t=0.95 the network achieves an overall classification performance of 78% with 38% of cases rejected as uncertain. However there are some problems associated with this outcome.

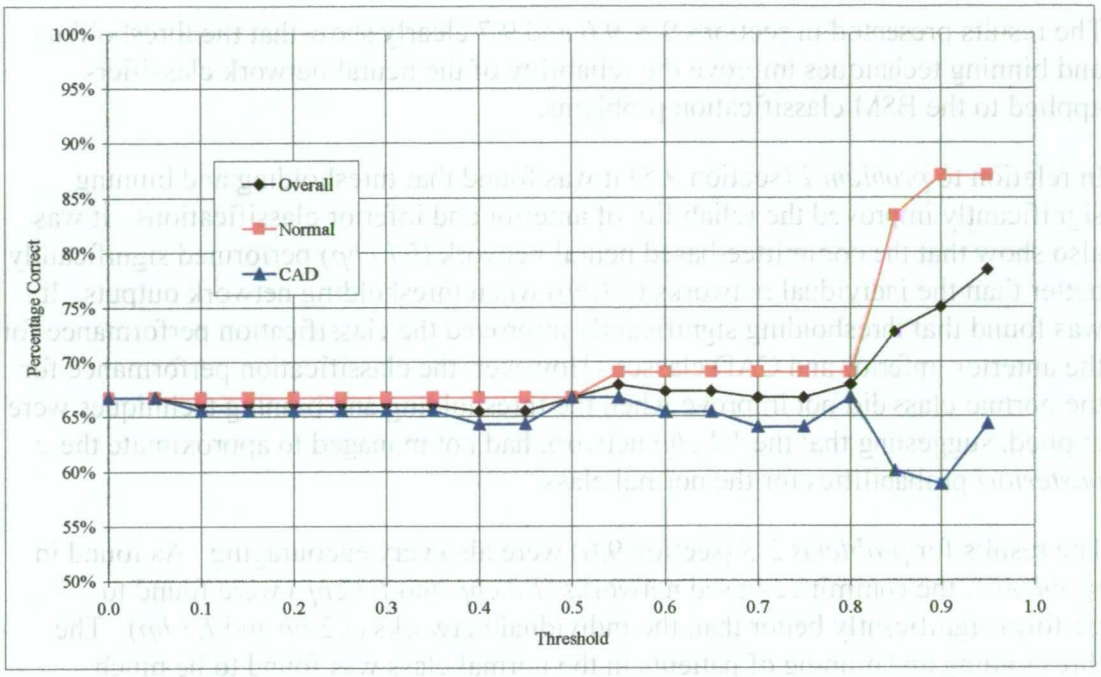
As can be seen from the results presented in Graph 9-22 the thresholding results for the classes are distinctly different. Although the classification performance for the normal class is improved dramatically (for t=0.95, 87% classified correctly, 22.5% of cases rejected) the classification performance for the CAD class is degraded (for t=0.95, 64% classified correctly, 52.5% rejected). Clearly the thresholding of the qrs4.bp network is baised toward improving the classification performance for the normal class. Although this bias does present some problems, the classification performance for the CAD class is still above a guess of 50%, so the thresholding could still be useful.

Clearly these results suggest that the network outputs of *qrs4.bp* do provide a degree of discrimination between the correct and incorrect classifications (and a similar result was observed for the *L4.bp* and *qrs4.qp* networks). Although producing somewhat biased results the thresholding did improve the overall classification performance.

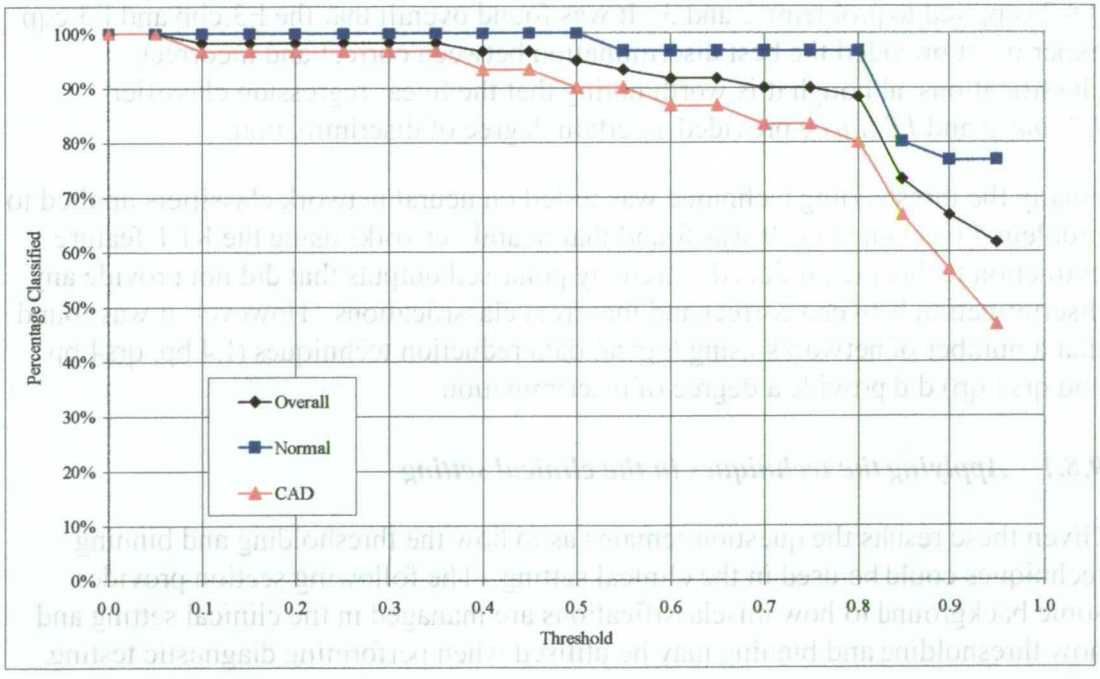
It is interesting to note that these results are a significant improvement over the results obtained for networks using the KLT feature extraction technique. Clearly, the networks using the KLT feature extraction technique cannot approximate the *a posteriori* probabilities, whereas the *L4.bp*, *qrs4.bp*, and *qrs4.qp* networks, which all use logical feature extraction techniques, do provide a degree of discrimination. This may suggest that the logical feature extraction techniques (*L,qrs*) generate a feature space that allows the neural networks to more effectively approximate *a posteriori* probabilities.

9.7.2 Binning Results

The binning results for these experiments discussed in the section 9.7.1 did not provide any useful insights. As the majority of network outputs were above 0.8 then in most cases all classifications were assigned to bin $B_{0.8\text{-}1.0}$. Consequently no effective separation of correct and incorrect classifications was achieved.



Graph 9-22: Thresholding results for experiment qrs4.bp (testing set)



Graph 9-23: Thresholding results for experiment qrs4.bp (testing set)

9.8 Concluding Comments

The results presented in sections 9.5, 9.6 and 9.7 clearly show that the thresholding and binning techniques improve the reliability of the neural network classifiers applied to the BSM classification problems.

In relation to *problem 1* (section 9.5) it was found that thresholding and binning significantly improved the reliability of anterior and inferior classifications. It was also show that the committee-based neural network (*E1.cbp*) performed significantly better than the individual networks (*E1.bp*) when thresholding network outputs. It was found that thresholding significantly improved the classification performance for the anterior, inferior and CAD classes. However, the classification performance for the normal class did not improve when the thresholding and binning techniques were applied, suggesting that the *E1.cbp* network had not managed to approximate the *a posteriori* probabilities for the normal class.

The results for *problems 2/3* (section 9.6) were also very encouraging. As found in *problem 1*, the committee-based networks (*E2.cbp* and *E3.cbp*) were found to perform significantly better than the individual networks (*E2.bp* and *E3.bp*). The thresholding and binning of patients in the normal class was found to be much improved over the results observed for *problem 1*. Given these results it was shown that individual output thresholds could be applied to a network to achieve a desired classification performance for all classes (section 9.6.1). Further to this, the thresholding technique was also applied to a range of different classifiers (section 9.6.2) applied to problems 2 and 3. It was found overall that the E3.cbp and E3.cqp experiment provided the best discrimination between correct and incorrect classifications, although it is worth noting that the linear regression classifier *E2.linreg* and *E3.linreg* provided a certain degree of discrimination.

Finally the thresholding technique was tested on neural network classifiers applied to problem 4 (section 9.7). It was found that neural networks using the KLT feature extraction technique produced extremely polarised outputs that did not provide any discrimination between correct and incorrect classifications. However, it was found that a number of networks using logical data reduction techniques (L4.bp, qrs4.bp and qrs4.qp) did provide a degree of discrimination.

9.8.1 Applying the techniques in the clinical setting

Given these results the question remains as to how the thresholding and binning techniques could be used in the clinical setting. The following section provides some background to how misclassifications are managed in the clinical setting and how thresholding and binning may be utilised when performing diagnostic testing.

9.8.1.1 Background - Misclassification in the clinical setting

When using classification tools in the clinical setting the issue of misclassification and class overlap needs to be treated with care, as peoples lives may be at stake. The key issues involved can be illustrated by considering a simple example in Figure 9-7.

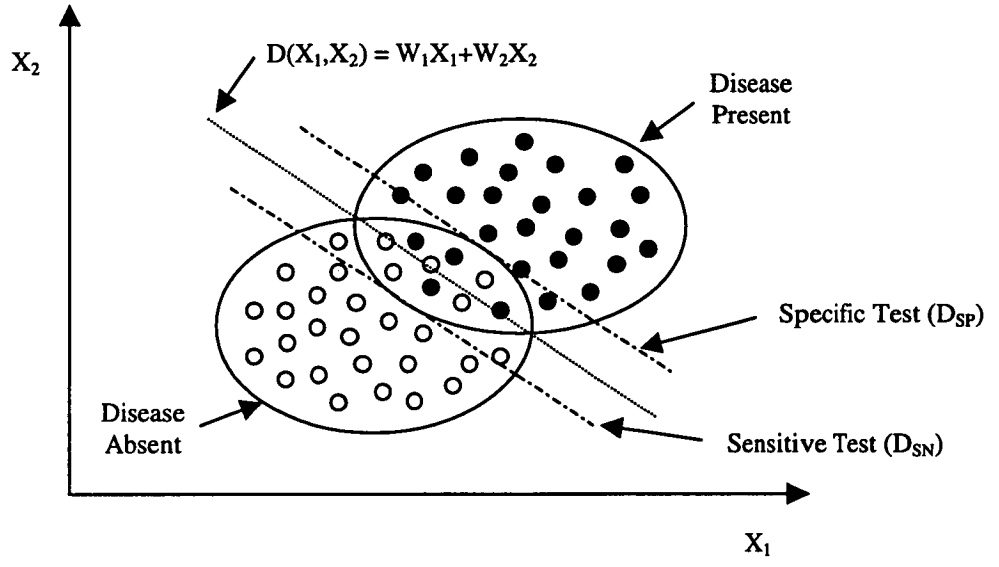


Figure 9-7: Diagnostic testing.

Illustrated in Figure 9-7 is a diagnostic test for detecting the presence or absence of a disease using the observations X_1 and X_2. With a degree of geometric argument it can be shown that the optimum classification performance can be obtained if the linear discriminator $D(X_1,X_2)$ is used, but unfortunately some patients will be misclassified.

When using classifiers for diagnostic testing, the possibility of misclassification is not desirable. Clearly an overall classification performance of 100% is not possible, but it is possible to set up the test so that a 100% classification performance is possible for at least one class. This is illustrated in Figure 9-7. If the discriminating boundary of D_{SN} is used then we can confidently conclude that those patients who are classified as not having the disease in question do not have the disease. Such a test is described as being *sensitive,* that is it is extremely sensitive to the possibility of disease. Alternatively if the discriminating boundary of D_{SP} is used then we can confidently conclude that those patients who are classified as having the disease in question do have the disease. Such a test is described as being *specific*, in that it is specifically focused on classifying the disease in question.

When applying a diagnostic test, the decision to use a sensitive or specific test will depend on the purpose of the test. The process of diagnosis (Griner 1981) requires three essential steps. The first step is to list all the possible diagnostic hypotheses based on a patient's presenting symptoms. The second step is to attempt to reduce the number of hypothesis by ruling out specific diseases. The third step is to critically test the hypotheses remaining to confirm the disease presence. Thus, when

attempting to rule out a particular disease, then a sensitive test would be appropriate, but when attempting to critically test a hypothesis then a specific test should be used.

Apart from diagnostic purpose another influencing issue is risk. If for example, not detecting the disease could result in a patient dying then it is probably appropriate to apply a sensitive test. In such a case the risk of misclassifying a patient as having the disease when they do not is far more desirable than misclassifying a patient who does have the disease.

9.8.1.2 Utilising thresholding and binning in the clinical setting

Given the diagnostic testing model presented in Figure 9-7 it becomes clear that the neural network thresholding technique will provide some distinct benefits. Since thresholding can be used to identify those patient classification that are uncertain (this is seen as particular effective with respect to the myocardial infarction cases), then this information would prove extremely useful for improving the sensitivity or specificity of any neural network based diagnostic test.

For example, if a more sensitive test is required, then a threshold could be applied to the neural network and all those cases that are rejected by the threshold could be classified as having the disease. That is, if the neural network is uncertain as to whether a particular patient does not have a particular disease, then it could be patient can be assigned as having the disease, for the sake of avoiding the risk of misclassification.

Alternatively, if a more specific test is required, then a threshold could be applied to the neural network and all those cases that are rejected by the threshold could be classified as not having the disease. This would ensure a higher classification performance with respect to those patient classified with disease.

Finally, it is worth noting that the thresholding concept and classification uncertainty raise a number of issues about how such diagnostic test are used. It is clear from the discussion of diagnostic process in section 9.8.1.1 (Griner 1981) that medical diagnostic testing is focused on diagnostic models for tests that provide a TRUE/FALSE outcome. Given that thresholding provides a tri-state result, this would suggest that some work may need to be done to intergrate such tests into current medical testing.

10. Conclusion

The experimental analysis in this thesis was divided into three sections.

Firstly, a range of feed-forward artificial neural network architectures and training techniques were used to classify the body surface mapping data with the aim of identifying patients with myocardial infarctions, coronary artery disease, and normal heart function.

Secondly, a range of traditional classification techniques (linear regression, k-nearest-neighbour, and inductive learning) were applied to the same problems and compared with the neural network results.

Thirdly, the bayesian equivalence of neural network outputs was examined and a number of approaches were considered as to how this information may be used to deal with diagnostic uncertainty. Apart from examining the theoretical connection between network outputs and *a posteriori* probabilities, a number of experiments were conducted to show how this information could be used to provide the physician with some important information about the classification certainty.

A summary of these results is presented in the following sections.

10.1 Comparison of Neural Network and Traditional Classification Techniques

Having explored a range of approaches for classifying BSM data in chapters 7 and 8 the following outcomes were observed:

Attempting to construct a classifier to separate all four classes; anterior infarctions, inferior infarctions, normals and CADs is a difficult task (*problem 1*). The best classification result achieved was for experiment *E1.linreg* (testing set result – 67.9%: anteriors – 83%, inferiors – 78.5%, normal – 30% and CADs – 80%). It was noted that all classifiers applied to this problem had difficulty separating normals and CADs and tended to classify most normal patients as having CAD. On examining this problem more closely it was discovered that a far more effective way to discriminate between classes was to consider this classification problem as two separate problems; the discrimination of infarcts from normals (*problems 2 and 3*) and the discrimination of normals and CADs (*problem 4*). Focusing on these problems separately improved the classification performance of classifiers significantly.

When applying the classifiers to the infarct/normal classification problem (*problem 2 and 3*) the classification results improved significantly over those obtained for the initial problem particularly in relation to the classification of normals. The best classification result achieved was for experiment *E2.linreg* (testing set result – 84%: anteriors – 81.7%, inferiors – 80.3% and normals – 90%).

The separation of normal patients and patients with coronary artery disease (CAD) was found be a difficult problem (*problem 4*). Many classifiers applied to this problem did not manage to discriminate between these classes and tended to bias

strongly toward one class. However, the neural network classifiers considered performed significantly better than other classifiers. The best classification result achieved was for experiment *E4.qp* (testing set result – 71.7%: normals – 70% and CADs - 73.3%).

In summary the following conclusions were made:

- Traditional BSM classification techniques or linear regression used in conjunction with KLT feature extraction are the most appropriate classification techniques for separating patients with myocardial infarctions from normal patients.

- Discriminating between normal patients and patients with coronary artery disease is difficult and cannot be achieved using traditional BSM classification techniques.

- It is possible using neural networks to discriminate between normal patients and patients with coronary artery disease. The best result achieved was 71% (experiment *E4.qp*: normals – 70.0% correct, CADs – 71.7% correct).

- The inductive learning techniques C4.5 and MML performed poorly on all BSM classification problems. This would suggest that the problems, although linearly separable cannot be easily separated using orthogonal decision planes (as used by these inductive learning techniques).

10.2 Improving Neural Network Classification Reliability

As presented in chapter 9, two techniques were considered for improving the classification reliability of neural network classifiers applied to the BSM data. The central aim of this exploration was to devise a technique for gauging the certainty of a classification and the utilise this information in the classification process.

The application of a *rejection threshold* and output *binning techniques* were applied to a number of the neural networks applied to the BSM data. It was found that the thresholding and binning techniques improve the reliability of the neural network classifiers applied to the BSM classification problems.

In relation to the four class problem (*problem 1* – separating anterior, inferior, normal, and CAD patients) it was found that thresholding and binning significantly improved the reliability of anterior and inferior classifications. It was also show that the committee-based neural network (*E1.cbp*) performed significantly better than the individual networks (*E1.bp*) when thresholding network outputs. It was found that thresholding significantly improved the classification performance for the anterior, inferior and CAD classes. However, the classification performance for the normal class did not improve when the thresholding and binning techniques were applied, suggesting that the *E1.cbp* network had not managed to approximate the *a posteriori* probabilities for the normal class.

The results for the three class problems (*problems 2/3* – seperating anterior, inferior, and normal classes) were also very encouraging. As found in *problem 1*, the

committee-based networks (*E2.cbp* and *E3.cbp*) were found to perform significantly better than the individual networks (*E2.bp* and *E3.bp*). The thresholding and binning of patients in the normal class was found to be much improved over the results observed for *problem 1*. Given these results it was shown that individual output thresholds could be applied to a network to achieve a desired classification performance for all classes. Further to this, the thresholding technique was also applied to a range of different classifiers applied to problems 2 and 3. It was found overall that the E3.cbp and E3.cqp experiment provided the best discrimination between correct and incorrect classifications, although it is worth noting that the linear regression classifier *E2.linreg* and *E3.linreg* provided a certain degree of discrimination.

Finally the thresholding technique was tested on neural network classifiers applied to two class problem (*problem 4* – separating normal and CAD patients). It was found that neural networks using the KLT feature extraction technique produced extremely polarised outputs that did not provide any discrimination between correct and incorrect classifications. However, it was found that a number of networks using logical data reduction techniques (L4.bp, qrs4.bp and qrs4.qp) did provide a degree of discrimination.

10.3 Final Comments

The most appropriate classification technique for classifying electrocardiographic body surface maps clearly depends on the classification problem being considered.

When attempting to discriminate between myocardial infarction and normal patients it is clear that the problem is linearly separable. As such it is appropriate in this situation to use either a linear discriminant function or k-nearest neighbour classification technique. With respect to neural network classifiers, no advantage is gained by using such techniques.

However, when considering the more difficult problem of discriminating between patients with coronary artery disease and patients with normal heart function the problem is very different. In this case neural network techniques were found to be far superior to traditional classification techniques and traditional classification techniques had great difficultly separating these classes.

References

Amirov, R. Z. (1961). "Cardiotoposcopy." Kardiologia (USSR) **2:** 55-61.

Barmann, F. and F. Biegler-Konig (1992). "On a Class of Efficient Learning Algorithms for Neural Networks." Neural Networks **5:** 139-144.

Barnard, E. (1992). "Comments on "Bayes Statistical Behavior and Valid Generalization of Pattern Classifying Neural Networks." IEEE Transactions on Neural Networks **3(6):** 1026-1028.

Barr, R. C., M. S. Spach, et al. (1971). "Selection of the Number and Positions of Measuring Locations for Electrocardiography." IEEE Transactions in Bio-Medical Engineering **BME-18(2):** 1-11.

Barr, R. C. and M. S. Spach (1978). "Inverse calculation of QRS-T epicardial potentials from body surface potential distributions fro normal and ectopic beats in the intect dog." Circulation Research **42:** 661-675.

Barr, R. C. and M. S. Spach (1983). "Construction and Interpretation of Body Surface Maps." Progress in Cardiovascular Diseases **26(1):** 33-42.

Barschdorff, D., S. Ester, et al. (1990). "Neural Network based multi sensor heart sound analysis." Computers in Cardiology: 303-306.

Baum, E. B. and F. Wilczek (1988). "Supervised Learning of Probability Distributions by Neural Networks." Neural Information Processing Systems: 52-61.

Baxt, W. G. (1990). "Use of an Artifical Neural Network for Data Analysis in Clinical Decision-Making: The Diagnosis of Acute Coronary Occlusion." Neural Computation(2): 480-489.

Baxt, W. G. (1991). "Use of an artificial neural network for the diagnosis of myocardial infarction." Annals of Internal Medicine **115(11):** 843-848.

Baxt, W. G. (1992). "Analysis of the clinical variables driving decision in an artifical neural network trained to identify the presence of myocardial infaction." Annals of Emergency Medicine **21(12):** 1439-1444.

Baxt, W. G. (1992). "Improving the accuracy of an artificial neural network using multiple differently trained networks." Neural Computations(4): 772-780.

Baxt, W. G. (1994). "A neural network trained to identify the presence of myocardial infarction bases some decisions on clinical associations that differ from accepted clinical teaching." Medical Decision Making **14(3):** 217-222.

Baxt, W. G. (1994). "Complexity, chaos and human physiology: the justification for non-linear neural computational analysis." Cancer-Lett **77(2-3):** 85-93.

References

Baxt, W. G. and H. White (1995). "Bootstrapping Confidence Intervals for Clinical Input Variable Effects in a Network Trained to Indentify the Presence of Acute Myocardial Infaction." Neural Computation **7:** 624-638.

Beale, R. and T. Jackson (1990). Neural Computing: An Introduction, IOP Publishing Ltd.

Bonelli, P. and A. Parodi (1991). An efficient classifier system and its experimental comparison with two representative learning methods on three medical domains. GA Proceedings, 4th Internaltional Conference July 1991.

Bortolan, G. and J. L. Willems (1993). "Diagnostic ECG classification based on neural networks." Journal of Electrocardiology **26 Suppl:** 75-79.

Bromley, J. and J. S. Denker (1993). "Improving Rejection Performance on Handwritten Digits by Training with "Rubbish"." Neural Computation(5): 367-370.

Browner, W. S. (1992). "Myocardial infarction prediction by artificial neural networks [letter;comment]." Annals of Internal Medicine **116(8):** 701-702.

Burrascano, P. (1991). "A Norm Selection Criterion for the Generalized Delta Rule." IEEE Transactions on Neural Networks **2(1):** 125-130.

Carter, J. H. (1992). "Myocardial infarction prediction by artificial neural networks [letter;comment]." Annals of Internal Medicine **116(8):** 701-702.

Cathers, I. (1992). A Neural Network Approach to Phonocardiography. Australian Conference on Neural Networks, Canberra, Australia.

Chen, K., K. J. Cios, et al. (1990). "Neural Networks in the detection of early acute myocardial ischemia from canine echocardiographic radio frequency data." Computers in Cardiology: 299-302.

Cheng, B. and D. M. Titterington (1994). "Neural Networks : A Review from a Statistical Perspective." Statistical Science **9(1):** 2-54.

Chi, J. and J. Marwan (1991). "A Comparison of MLP and ID3-derived Approaches for ECG Classification." ACNN: 263-266.

Choi, J. and T. C. Pilkington (1981). "Effects of geometrical uncertainties on electrocardiography." IEEE Transactions in Biomedical Engineering **28:** 325-344.

Chow, C. K. (1957). "An Optimum Character Recognition System Using Decision Functions." IRE Transactions on Electronic Computera **December:** 247-254.

Chow, C. K. (1965). "Statistical Independence and Threshold Functions." IEEE Transactions on Electronic Computers **EC-14:** 66-68.

Chow, C. K. (1970). "On Optimum Recognition Error and Reject Tradeoff." IEEE Transactions on Information Theory **IT-16(1):** 41-46.

Colli-Franzone, P., L. Guerri, et al. (1985). "A mathematical procedure for solving the inverse potential problem in electrocardiography. Analysis of the time-space accuracy from in vitro experimental data." Mathematic in Biomedical Science **77:** 353-396.

Cordella, L. P., C. De Stefano, et al. (1995). "A Method for Improving Classification Reliability of Multilayer Perceptrons." IEEE Transactions on Neural Networks **6(5):** 1140-1147.

Cox, J. R., F. M. Nolle, et al. (1972). "Digital Analysis of the Electroencephalogram, the Blood Pressure Wave, and the Electrocardiogram." Proceedings of the IEEE **60(10):** 1137-1164.

Dassen, W. R. M., R. G. A. Mulleneere, et al. (1990). "A Neural Network to Differentiate Wide-QRS Tachycardias." Computers in Cardiology: 287-290.

Dassen, W. R. M., R. G. A. Mulleneers, et al. (1993). "Proposal for an Editorial Policy for Publication Desribing the Use of Artificial Neural Networks in Electrocardiography." Journal of Electrocardiography **26(3):** 241-244.

Dassen, W. R., R. G. Mulleneers, et al. (1993). "Artificial neural networks and ECG interpretation. Use and abuse." Journal of Electrocardiology **26 Suppl:** 61-65.

Devijver, P. A. and J. Kittler (1982). Pattern Recognition: A Statistical Approach, Prentice-Hall.

Diamond, G. A. (1986). "Reverend Bayes' Silent Majority: An alternative factor affecting sensitivity and specificity of exercise electrocardiography." The American Journal of Cardiology **57:** 1175-1180.

Dowe, D. L. and N. Krusel (1993). A Decision Tree Model of Bushfire Activity. Melbourne, Monash University, Clayton, Victoria, Australia.

Droste, C. and M. van Planta (1997). Memorix, Clinical Medicine. London, Chapman and Hall Medical.

Duda, R. O. and P. E. Hart (1973). Pattern Classification and Scene Analysis, John Wiley.

Eaton, H. A. C. and T. L. Oliver (1992). "Learning Coefficient Dependence on Training Set Size." Neural Networks **5:** 283-288.

Edenbrandt, L., B. Heden, et al. (1993). "Neural networks for analysis of ECG complexes." Journal of Electrocardiology **26 Suppl:** 74.

Einthoven, W., G. Fahr, et al. (1913). "Uber die Richtung und die Manifeste Grosse der Potential Schwankungen im Menschlichen Herzen and uber den Einfluss der Herzlage auf die Form des Electrokardiogramms." Arch. ges Physiol. **150:** 275.

El-Jaroudi, A. and J. Makhoul (1990). A new error criterion for posterior probability estimation with neural nets. Proceedings of International Joint Conference on Neural Networks, San Diego, CA, IEEE.

Essen, R. V., W. Merx, et al. (1980). "QRS Mapping in the Evaluation of Acute Anterior Myocardial Infaction." Circulation **62(2):** 266-276.

Evans, A. K., R. L. Lux, et al. (1981). "Redundancy Reduction for Improved Display and Analysis of Body Surface Maps, II. Temporal Compression." Circulation Research **49(1):** 197-203.

Evans, S. J., H. Hastings, et al. (1994). "Differentiation of beats of ventricular and sinus origin using a self-training neural network." PACE - Pacing Clinical Electrophysiology **17(4(1)):** 611-626.

Everitt, B. (1974). Cluster Analysis, Heinemann Educational Books.

Fahlman, S. E. (1988). An Empirical Study of Learning Speed in Back-Propagation Networks, School of Computer Science, Carnegie Mellon University.

Fahlman, S. E. and C. Lebiere (1991). The Cascade-Correlation Learning Architecture, School of Computer Science, Carnegie Mellon University.

Fahlman, S. E. (1991). The Recurrent Cascade-Correlation Architecture, School of Computer Science, Carnegie Mellon University.

Fisher, R. A. (1936). "The use of multiple measurements in taxonomic problems." Annals of Eugenics 7(2): 179-188.

Flowers, N. C., L. G. Horan, et al. (1976). "Anterior Infarctional Changes Occuring during Mid and late Ventricular Activation Detectable by Surface Mapping Techniques." Circulation **54(6):** 906-913.

Fox, K. M., A. Selwyn, et al. (1979). "Relation between the precordial projection of S-T segment changes after exercise and coronary angiographic findings." American Journal of Cardiology **44:** 1068-1075.

Fuchs, R. M., S. C. Achuff, et al. (1982). "Electrocardiographic Localization of Coronary Artery Narrowing: Studies During Myocardial Ischemia and Infaction in Patients with One-vessel Disease." Circulation **66(6):** 1168-1176.

Fujita, H., T. Katafuchi, et al. (1992). "Application of artificial neural network to computer-aided diagnosis of coronary artery disease in myocardial SPECT bull's-eye images." Journal of Nuclear Medicine **33(2):** 272-276.

Fukunaga, K. (1972). Introduction to Statistical Pattern Recognition. New York, Academic Press.

Furlong, J. W., M. E. Dupuy, et al. (1991). "Neural network analysis of serial cardiac enzyme data. A clinical application of artifical machine intelligence." American Journal of Clinical Pathology **96(1):** 134-141.

Ganong, W. F. (1983). Review of Medical Physiology, Lange Medical Publications.

Gish, H. (1990). A Probabilistic Approach to the Understanding and Training of Neural Network Classifiers. IEEE Conference on Acoustic Speech and Signcal Processing.

Goldman, L., M. Weinberg, et al. (1982). "A computer-derived protocol to aid is the diagnosis of emergency room patients with acute chest pain." The New England Journal of Medicine **307(10):** 588-596.

Goldman, L., E. F. Cook, et al. (1988a). "A computer protocol to predict myocardial infarction in emergency department patients with chest pain." The New England Journal of Medicine **318(13):** 797-803.

Goldman, L., E. F. Cook, et al. (1988b). "Letter to the editor." The New England Journal of Medicine **319:** 792.

Green, L. S., R. L. Lux, et al. (1987). "Detection and localization of coronary artery disease with body surface mapping in patients with normal electrocardiograms." Circulation **76(6):** 1290-1297.

Griner, P. F., R. J. Mayewski, et al. (1981). "Selection and interpretation of diagnositic tests and procedures." Annals of Internal Medicine **94(4):** 553-600.

Guerriere, M. R. J. and A. S. Detsky (1991). "Neural Networks: What Are They?" Annals of Internal Medicine **115(11):** 906-907.

Hampshire, J. B. and B. Pearlmutter (1990). Equivalence Proof for Multi-Layer Perceptron Classifiers and the Bayesian Discriminant Function. Proceedings of the 1990 Connectionist Summer School, Morgan Kaufmann, San Mateo, CA.

Hampton, J. R. (1986). The ECG made easy, Churchill Livingstone.

Hampton, J. R. (1986). The ECG in practice, Churchill Livingstone.

Harrison, R. F., S. J. Marshall, et al. (1991). The Early Diagnosis of Heart Attacks : A Neurocomputational Approach. International Joint Conference of Neural Networks.

Heden, B., L. Edenbrandt, et al. (1994). "Artificial neural network for the electrocardiographic diagnosis of healed myocardial infactions." American Journal of Cardiology **74(1):** 5-8.

Heringa, A., G. J. H. Uijen, et al., Eds. (1981). A 64-Channel System for Body Surface Potential Mapping. Electrocardiology. Budapest.

Hildebrandt, T. H. (1992). Why Batch Learning is Slower than Per-Sample Learning, Lehigh University, Bethlehem, PA 18015.

Horan, L. G., N. C. Flower, et al. (1963). "Body surface potential distribution: Comparison of naturally and artificially produced signals as analyzed by digital computer." Circulation Research **13:** 373-387.

Horan, L. G. and N. C. Flowers (1968). "The principle of waveform correlation in electrocardiographic research." Journal of Electrocardiology **1:** 43-50.

Horan, L. G., R. C. Hand, et al. (1976). "The multipolar content of the human electrocardiogram." Annals of Biomedical Engineering **4:** 280-301.

Hu, Y. H., J. Tompkins, et al. (1992). "Artifical Neural Networks for ECG Arrhythmia Monitoring." IJCNN: II-987-991.

Hu, Y. H., W. J. Tompkins, et al. (1993). "Application of artifical neural networks for ECG signal detection and classification." Journal of Electrocardiology **26 Suppl:** 66-73.

Hunt, F., C. Carlisle, et al. (1983). Coronary Care Workbook: A Handbook for Coronary Care Nurses, Excerpta Medica.

Ikeda, K., I. Kubota, et al. (1985). "Detection of local abnormalities in ventricular activation sequences by body surface isochrone mapping with previous myocardial infactions." Circulation **72(4):** 801-809.

Ikeda, K., S. Kawashima, et al. (1986). "Non-invasive detection of coronary artery disease by body surface electrocardiographic mapping after dipyridamole infusion." Journal of Electrocardiology **19:** 213-224.

Irino, T. and H. Kawahara (1990). "A method for designing neural networks using nonlinear multivariate analysis: application to speaker-independent vowel recognition." Neural Computation(2): 386-397.

James, M. (1985). Classification Alogorithms. London, Collins.

Jensch, P., L. Kohler, et al. (1990). "Extended Coronary Trees by Neural Network Based Fusion of DSA-Sequences and ECT-Images." Computers in Cardiology: 307-310.

Kanaya, F. and S. Miyake (1991). "Bayes Statistical Behavior and Valid Generalization of Pattern Classifying Neural Networks." IEEE Transactions on Neural Networks **2(4):** 471-475.

Kendall, M. G. (1957). A Course in Multivariate Analysis, Chales Griffin.

Kilpatrick, D., P. Duffin, et al. (1979). "Apparatus for the Rapid Measurement of Electrocardoigraphic Body Surface Potentials." Med. Biol. Eng. Comput. **17:** 257-260.

Kilpatrick, D. and S. J. Walker (1987). "A validation of derived epicardial potential distributions by prediction of the coronary artery involved in acute myocardial infarction in humans." Circulation **76(6):** 1282-1289.

Kornreich, F., T. J. Montague, et al. (1990). "Knowledge-Based Systems for Classification of Body Surface Potential Maps." Computers in Cardiology: 479-482.

Kubota, I., H. Igarashi, et al. (1984). "Attentuation of exercise-induced R wave increase after diltiazem in effort angina pectoris." Japanese Heart Journal **25:** 937-945.

LeCun, Y. (1985). "Une procedure d'apprentissage pour reseau a seuil assymetrique." Cognitiva 85: 599-604.

Lee, T. H., E. F. Cook, et al. (1985). "Acute chest pain in the emergency ward: Identification and examination of low-risk patients." Ach. Intern. Med. **145:** 65-69.

Lee, T. H., G. W. Rouan, et al. (1987a). "Sensitivity of routine clinical criteria for diagnosing myocardial infarctions within 24 hours of hospitalizations." Annals of Internal Medicine **106(2):** 181-186.

Lee, T. H., G. W. Rouan, et al. (1987b). "Clinical characteristics and natural history of patients with acute myocardial infarction sent home from the emergency room." American Journal of Cardiology **60:** 219-224.

Leibman, J., C. W. Thomas, et al. (1981). "Electrocaradiographic body surface potential maps of the QRS of normal children." Journal of Electrocardiology **14(3):** 249-260.

Lette, J., B. W. Colletti, et al. (1994). "Artificial intelligence vesus logistic regression statistical modelling to predict cardiac complications after noncardiac surgery." Clinical Cardiology **17(11):** 609-614.

Lippmann, R. P. (1987). An Introduction to Computing with Neural Networks. IEEE ASSP Magazine: 4-22.

Lux, R. L., C. R. Smith, et al. (1978). "Limited Lead Selection for Estimation of Body Surface Potential Maps in Electrocardiography." IEEE Transactions in Bio-Medical Engineering **BME-25(3):** 270-276.

Lux, R. L., M. J. Burgess, et al. (1979). "Clinically Practical Lead Systems for Improved Electrocardiography: Comparison with Precordial Grids and Conventional Lead Systems." Circulation **59(2):** 356-363.

Lux, R. L., A. K. Evans, et al. (1981). "Redundancy Reduction for Improved Display and Analysis of Body Surface Potential Maps, I. Spatial Compression." Circulation Research **49(1):** 186-196.

Mann, H. C. (1938). "The Monocardiograph." American Heart Journal **15:** 681-699.

McAuliffe, J. D. (1993). "Data compression of the exercise ECG using a Kohonen neural network." J Electrocardiol **26 Suppl:** 80-9.

McClelland, J. H., N. D. Danieley, et al. (1990). "Neural Networks for Improving Automation of Ventricular Activation Mapping." Computers in Cardiology: 295-298.

McCulloch, W. S. and W. Pitts (1943). "A logical calculus of the ideas immanent in nervous activity." Bulletin of Mathematical Biophysics 5: 115-133.

McMechan, S. R., C. M. Cullen, et al. (1994). "Discriminant function analysis of body surface potential maps in acute myocardial infarction." J Electrocardiol **27 Suppl:** 117-20.

Miller, J. W., R. Goodman, et al. (1991). Objective Functions for Probability Estimation. International Joint Conference on Neural Networks.

Minsky, M. L. and S. A. Papert (1969). Perceptrons : An Introduction to Computational Geometry, MIT Press.

Mirvis, D. M. (1981). "Body Surface Distributions of Repolarization Potentials After Acute Myocardial Infarction. II. Relationship Between Isopotential Mapping and ST-segment Potenial Summation Metods." Ciculation **63(3):** 623-631.

Mirvis, D. M. (1987). "Current status of body surface electrocardiographic mapping." Circulation **75(4):** 684-688.

Mirvis, D. M., Ed. (1988). Body Surface Electrocardiographic Mapping. Boston, Kluwer Academic Publishers.

Miyake, S. and F. Kanaya (1991). "A Neural Network Approach to a Bayesian Statistical Decision Problem." IEEE Transactions on Neural Networks **2(5):** 538-540.

Montague, T. J., E. R. Smith, et al. (1981). "Isointegral Analysis of Body Surface Maps: Surface Distribution and Temporal Variability in Normal Subjects." Circulation **63(5):** 1166-1172.

Montague, T. J., E. R. Smith, et al. (1983). "Body surface electrocardiographic mapping in inferior myocardial infarction. Manifestations of left and right ventricular involvement." Circulation **67:** 665-673.

Montague, T. J., E. R. Smith, et al. (1984). "Temporal evolution of body surface map patterns following acute inferior myocardial infarction." Journal of Electrocardiology **17:** 319-328.

Montague, T. J., M. J. Gardner, et al. (1986). "Electrocardiographic and ventriculographic recovery patterns in Q wave myocardial infaction." Journal of American College Cardiology **58:** 1173-1180.

Moskowitz, A. J., B. J. Kuipers, et al. (1988). "Dealing with uncertainty, risks, and trade-offs in clinical decisions. A cognitive science approach." Annals of Internal Medicine **108:** 435-449.

Murray, R. G., M. P. Watts, et al. (1981). "On-line computer-assisted exercise mapping." Cardiology **68 Suppl(2):** 133-140.

Nelson, C. V. (1957). "Human thorax potentials." Annals of the New York Academy of Science **65:** 1014-1050.

Ng, K. and R. P. Lippmann (1991). "A Comparitive Study of the Practical Characteristics of Neural Networks and Conventional Pattern Classifiers." Advances in Neural Networks **3:** 970-976.

Parker, D. B. (1985). Learning-logic: Casting the cortex of the human brain in silicon. Cambridge, MA, Center for Computational Research in Economics and Management Science, MIT.

Periyalwar, S. S., S. T. Nugent, et al. (1989). "Two-dimensional Fourier spectrum of QRST integral maps in classification of patients prone to ventricular arrhythmia." IEEE Trans Biomed Eng **36(4):** 493-6.

Pham-Huy, H., R. M. Gulraiani, et al. (1981). "A comparitive evaluation of three different approaches for detecting body surface isopotential map abnormalities in patients with myocardial infaction." Journal of Electrocardiology **14:** 43-56.

Pigot, L. and D. Kilpatrick (1990). The application of eigenvector analysis to the reduction of electrocardiographic body surface potential data, Department of Medicine, University of Tasmania.

Pozen, M. W., J. K. Stechmiller, et al. (1977). "Prognistic efficacy of early clinical categorization of myocardial infarction patients." Circulation **56:** 816-819.

Pozen, M. W., R. B. D'Agostino, et al. (1980). "The usefulness of a predictive instrument to reduce inappropriate admissions to the coronary care unit." Annals of Internal Medicine **92:** 238-242.

Pozen, M. W., R. B. D'Agostino, et al. (1984). "A predictive instrument to improve coronary-care-unit admission practices in acute ischemic heart disease: A prospective multicenter clinical trial." The New England Journal of Medicine **310(20):** 1273-1278.

Quinlan, J. R. (1986). "Induction of Decision Trees." Machine Learning **1:** 81-86.

Quinlan, J. R. (1993). C4.5 : Programs for Machine Learning, Morgan Kaufmann Publishers.

Reek, E. J., C. A. Grimbergen, et al. (1984). A low cost 64 channel microcomputer based data aquisition system for bedside registration of body surface maps, Department of Medical Physics and Biophysics, University of Nijmegen, The Netherlands.

Rembold, C. M. and D. Watson (1988). "Posttest Probability Calculation by Weight: A simple form of Bayes' Theorem." Annals of Internal Medicine **108:** 155-120.

Richard, M. D. and R. P. Lippmann (1991). "Neural Network Classifiers Estimate Bayesian a posteriori Probabilities." Neural Computation **3:** 461-483.

Ripley, B. D. (1992). Statistical Aspects of Neural Networks, Department of Statistics, University of Oxford.

Ripley, B. D. (1993). Networks and Chaos - Statistical and Probabilistic Aspects, Chapman & Hall.

Rosenblatt, F. (1958). "The Perceptron: A probabilistic model for information storage and organization in the brain." Psychological Review 65: 386-408.

Ruck, D. W., S. K. Rogers, et al. (1990). "The Multilayer Perceptron as an Approximation to a Bayes Optimal Discriminant Function." IEEE Transactions on Neural Networks **1(4):** 296-298.

Rumelhart, D. E. and J. L. McClelland (1986). Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Cambridge, MA, MIT Press.

Schiffmann, W., M. Joost, et al. (1992). Synthesis and Performance Analysis of Multilayer Neural Network Architectures, Institute of Physics, University of Koblenz.

Schiffmann, W., M. Joost, et al. (1992). Optimization of the Backpropagation Algorithm for Training Multilayer Perceptrons, Institute of Physics, University of Koblenz.

Shoemaker, P. A. (1991). "A Note on Least-Squares Learning Procedures and Classification by Neural Network Models." IEEE Transactions on Neural Networks **2(1):** 158-160.

Shufflebarger, C. M. (1992). "What is a neural network? [editorial; comment]." Annals of Emergency Medicine **21(12):** 1461-1462.

Simoons, M. L. and P. Bloack (1981). "Toward the optimal lead system and optimal criteria for exercise electrocardiography." American Journal of Cardiology **47:** 1366-1374.

Sox, H. C. (1986). "Probability theory in the use of diagnositic tests." Annals of Internal Medicine **104:** 60-66.

Spach, M. S., R. C. Barr, et al. (1979). "Isopotential Body Surface Mapping in Subjects of All Ages: Emphasis on Low Level Potentials with Analysis of the Method." Circulation **59(4):** 805-821.

Specht, D. F. and P. D. Shapiro (1991). Generalization accuracy of probabilisitic neural networks compared with back-propagation networks. International Joint Conference on Neural Networks.

Squires, C. S. and J. W. Shavlik (1991). Experimental Analysis of Aspects of the Cascade-Correlation Learning Architecture, Computer Science Department, University of Wisconsin-Madison.

Strand, E. M. and W. T. Jones (1990). "Tracking Ventricluar Arrhythmias with an Artifical Neural Network." Computer In Cardiology: 291-294.

Szolovits, P., R. S. Patil, et al. (1988). "Artifical intelligence in medical diagnosis." Annals of Internal Medicine **108:** 80-87.

Taccardi, B. (1958). "La distribution spatiale des potentiels cardiaques." Acta Cardiol **13:** 173-189.

Taccardi, B. (1960). "La distribuzione delle linee equipotenziali sul torace del cane narcotizzato." Atti Acad Med Lomb **15:** 176-178.

Taccardi, B. (1962). "Distrubution of heart potentials on dog's thoracic surface." Circulation Research **11:** 862-869.

Taccardi, B. (1963). "Distribution of heart potentials on the thoracic surface of normal human subjects." Circulation Research **12:** 341-352.

Taccardi, B., L. DeAmbroggi, et al. (1976). Body surface mapping of heart potentials. The Theoretical Basis of Electrocardiology. C. V. Nelson and D. B. Geselowitz. Oxford, Clarendon Press.

Tou, J. T. and R. C. Gonzalez (1974). Pattern Recognition Principles, Addison-Wesley.

Traven, H. G. C. (1991). "A Neural Network Approach to Statistical Pattern Classification by "Semiparametric" Estimation of Probability Density Functions." IEEE Transactions on Neural Networks **2(3):** 336-377.

Ultsch, A. and G. Halmans (1991). Data Normalization with Self-Organizing Feature Maps. International Joint Conference on Neural Networks.

Vamplew, P. and A. Adams (1991). Real World Problems in Backpropagation, Department of Computer Science, University of Tasmania. Technical Report R91-4.

van Dam, R. T. and A. e. van Oosterom (1986). Electrocardiographic Body Surface Mapping, Martinus Nijhoff Publishers.

Vogt, L., M. Sigmund, et al. (1990). "Non-invasive detection of acute rejection after cardiac transplantation by means of body surface potential mapping." Computers in Cardiology: 487-490.

Wada, M., K. Kaneko, et al. (1981). "Exercise stress body surface isopotential map in patients with coronary artery disease: Comparison with coronary angiographic and stress myocardial perfusion scintigraphic findings." Japanese Circulation Journal **45:** 1203-1207.

Walker, S. J., P. S. Lavercombe, et al. (1983). "A Body Surface Mapping System with Immediate Interactive Data Processing." IEEE Computers in Cardiology: 305-308.

Walker, S. J., A. J. Bell, et al. (1987). "Spatial distribution and prognostic significance of ST segment potential determined by body surface mapping in patients with acute inferior myocardial infarction." Circulation **76(2):** 289-297.

Wallace, C. S. (1990). Classification by minimum-message-length inference. Advances in Computing and Information - ICCI 1990, Niagara Falls.

Wallace, C. S. and J. D. Patrick (1993). "Coding Decision Trees." Machine Learning(11): 7-22.

Wan, E. A. (1990). "Neural Network Classification: A Bayesian Interpretation." IEEE Transactions on Neural Networks **1(4):** 303-305.

Werbos, P. J. (1974). Beyond regression: New tools for prediction and analysis in the behavioral sciences, Harvard University.

White, H. (1989). "Learning in Artifical Neural Networks: A Statistical Perspective." Neural Computation 1: 425-464.

Widrow, B. and M. E. Hoff (1960). Adaptive switching circuits. IRE WESCON Convention Record.

Williams, H. B. (1914). "On the cause of phase difference frequently observed between homogeneous peaks of the electrocardiogram." American Journal of Physiology 35: 292-300.

Wilson, F. N., A. C. Macleod, et al. (1932). "Electrocardiographic leads which record potential variations of the heart at a single point." Proceedings of the Society of Experimental Biological Medicine **29:** 1011-1012.

Wilson, F. N. and F. D. Johnston (1938). "The Vectorcardiogram." American Heart Jorunal **16:** 14-28.

Winsor, T. (1969). "The Electrocardiogram in Myocardial Infarction." Clinical Symposia **20:** 3-29.

Yajima, K., S. Kinoshita, et al. (1983). "Body Surface Potential Mapping System equipped with a Microprocessor for the Dynamic Observation of Potential Patterns." Med. Biol. Eng. Comput. **21:** 83-90.

Yamada, K., J. Toyama, et al. (1978). "Body Surface Isopotential Maps:Clinical Application to the Diagnosis of Myocardial Infarction." Japanese Heart Journal **19(1):** 28-45.

Yang, T. F., B. Devine, et al. (1993). "Deterministic logic versus software-based artificial neural networks in the diagnosis of atrial fibrillation." J Electrocardiol **26 Suppl:** 90-4.

Yang, T. F., B. Devine, et al. (1994). "Artificial neural networks for the diagnosis of atrial fibrillation." Med Biol Eng Comput **32(6):** 615-9.

Yang, T. F., B. Devine, et al. (1994). "Use of artificial neural networks within deterministic logic for the computer ECG diagnosis of inferior myocardial infarction." J Electrocardiol **27 Suppl:** 188-93.

Yanowitz, F. G., G. M. Vincent, et al. (1982). "Application of body surface mapping to exercise testing: S-T80 isoarea mapping in patients with coronary artery disease." American Journal of Cardiology **50:** 1109-1113.

Young, T. Y. (1974). Classification Estimation and Pattern Recognition, Elsevier.

Young and Calvert (1974). Chapter 7 - Pattern Recogniton of Electrocardiograms and Vectorcardiograms. Classification, Estimation and Pattern Recognition: 277-310.

Yuille, A. L., S. M. Smirnakis, et al. (1995). "Bayesian Self-Organization Driven by Prior Probability Distributions." Nueral Computation **7:** 580-593.

Zelicoff, A. P. (1992). "Myocardial infarction prediction by artificial neural networks [letter;comment]." Annals of Internal Medicine **116(8):** 701-702.

Zhu, K., P. D. Noakes, et al. "Training Neural Networks for ECG Feature Recognition." : 137-140.

A. Backpropagation

This appendix provides a detailed description and derivation of the backpropagation algorithm developed by Rumelhart and McClelland. The derivation will consider a neural network with one hidden layer. This will then be generalised for any number of hidden layers, and the final section of this appendix will provide a detailed description of how this generalised algorithm can be expressed in a matrix model of feed-forward neural networks.

This matrix model of a neural network forms the theoretical basis of the neural network object model used to implement the software tools used in this thesis. The implementation of these software tools is described in detail in Appendix B.

A.1 Backpropagation Training

The backpropagation training algorithm involves a supervised training procedure using a training data set to adjust the weights of a feed-forward neural network to minimise the output error of that network. To describe the specifics of this training procedure, consider the feed-forward network in Figure A-1. This network has *M* input nodes, a single hidden layer containing *H* neurons, and an output layer containing *N* neurons. The input and hidden layers are connected by *M*x*H* weights, where w_{jk} represents a connection between the x_k input node and the h_j hidden neuron. The hidden and output layers are connected by *H*x*N* weights, where u_{ij} represents a connection between the h_j hidden neuron and the y_i output neuron. Consider also a training dataset containing *S* training examples:

$$[(P_1,T_1), \ldots (P_e,P_e) \ldots (P_S,T_S)]$$

Where (P_e,T_e) represent a single training example; $P_e = (p_{e1},p_{e2}, \ldots, p_{eM})$ represents an input pattern, and $T_e = (t_{e1},t_{e2}, \ldots . t_{eN})$ represents the desired output for pattern P_e.

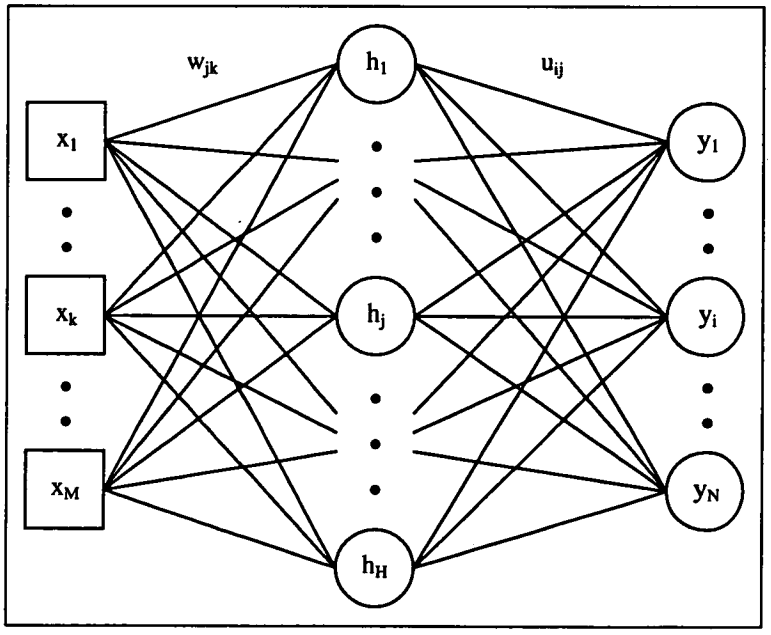


Figure A-1: Feed-Forward Neural Network (single hidden layer).

Appendix A

The backpropagation training procedure is an iterative process of presenting the training patterns (P_e) to the network one at a time and adjusting the network weights in relation to the error between the network output and the desired output (T_e). If we consider the presentation of the training example (P_e,T_e), then the weight adjustment is done as follows:

Pattern Presentation and Weight Adjustment

1. The pattern P_e is applied to the network and the network output determined:

   The input pattern P_e is applied to the network:

$$x_k = p_{ek} \tag{A-1}$$

   The hidden layer is calculated:

$$net_j = \sum_{c=1}^{M} w_{jc} x_c \tag{A-2}$$

$$h_j = f(net_j) \tag{A-3}$$

   The output layer is calculated:

$$net_i = \sum_{b=1}^{H} u_{ib} h_b \tag{A-4}$$

$$y_i = g(net_i) \tag{A-5}$$

2. The output error is calculated :

$$E = \frac{1}{2} \sum_{a=1}^{N} (y_a - t_a)^2 \tag{A-6}$$

3. The network weights are adjusted using the gradient decent rule:

$$\Delta w = -\eta \frac{\partial E}{\partial w} \tag{A-7}$$

$$w = w + \Delta w \tag{A-8}$$

The activation function on the hidden layer neurons is a sigmoid activation function:

$$f(net_j) = \frac{1}{1+e^{-net_j}} \tag{A-9}$$

The activation function for the output layer may be the same or simply a linear transfer function:

$$g(net_i) = \frac{1}{1+e^{-net_i}}, \text{ or } g(net_i) = net_i \tag{A-10}$$

The training examples are repeatedly applied to the network until the overall network error (A-9) reduces to an "acceptable level" (please refer to chapter 3 for more specific details concerning stopping procedures for training).

$$E_{overall} = \sum_{e=1}^{S}\left[\frac{1}{2}\sum_{i=1}^{N}(y_{ei} - t_{ei})^2\right] \tag{A-11}$$

You will note that although the above definition is somewhat straight forward, the specifics of how the differential $\partial E/\partial w$ is calculated in equation (A-7) has been omitted. The next two sections provide a derivation of $\partial E/\partial w_{jk}$ for the hidden layer weights and $\partial E/\partial u_{ij}$ for the output layer weights.

A.2 Derivation of Weight Adjustment for Output Layer

For the purposes of this derivation, consider the output layer weight u_{ij} which is the weight connecting the hidden layer neuron h_j to the output layer neuron y_i (Figure A-2).

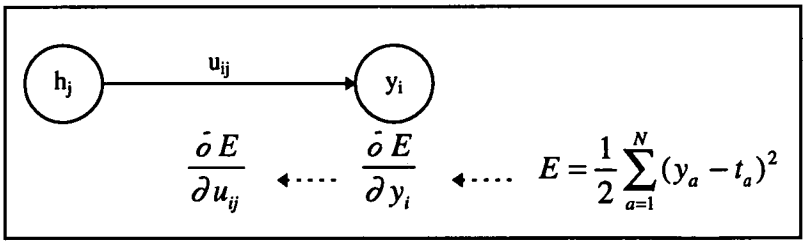


Figure A-2: Output Layer Weight

The objective in this derivation is to determine $\partial E/\partial u_{ij}$ given the network error E. This differential can be expressed as follows:

$$\frac{\partial E}{\partial u_{ij}} = \frac{\partial E}{\partial y_i} \cdot \frac{\partial y_i}{\partial net_i} \cdot \frac{\partial net_i}{\partial u_{ij}} \tag{A-12}$$

Expanding these differentials, we get the following:

$$\begin{aligned}\frac{\partial E}{\partial y_i} &= \frac{\partial}{\partial y_i}\left[\frac{1}{2}\sum_{a=1}^{N}(y_a - t_a)^2\right] \\ &= (y_i - t_i)\end{aligned} \tag{A-13}$$

Let $\delta_i = (y_i - t_i)$, therefore (A-14)

$$\frac{\partial E}{\partial y_i} = \delta_i \tag{A-15}$$

$$\begin{aligned}\frac{\partial y_i}{\partial net_i} &= \frac{\partial}{\partial net_i}[g(net_i)] \\ &= g'(net_i)\end{aligned} \tag{A-16}$$

$$\begin{aligned}\frac{\partial net_i}{\partial u_{ij}} &= \frac{\partial}{\partial u_{ij}}\left[\sum_{b=1}^{H} u_{ib}x_b\right] \\ &= h_j\end{aligned} \tag{A-17}$$

Substituting (A-13), (A-15), and (A-16) into (A-12), then:

$$\frac{\partial E}{\partial u_{ij}} = \delta_i \cdot g'(net_i) \cdot h_j \tag{A-18}$$

Therefore (A-18) can be used to calculate update for u_{ij} using the gradient decent rule (A-7):

$$\begin{aligned}\Delta u_{ij} &= -\eta\frac{\partial E}{\partial u_{ij}} \\ &= -\eta\left[\delta_i \cdot g'(net_i) \cdot h_j\right]\end{aligned} \tag{A-19}$$

where, if $g(net_i) = net_i$, then $g'(net_i) = 1$

or, if $g(net_i) = \frac{1}{1+e^{-net_i}}$, then $g'(net_i) = y_i(1 - y_i)$

A.3 Derivation of Weight Adjustment for Hidden Layer

For the purposes of this derivation, consider the output layer weight w_{jk} which is the weight connecting the input node x_k to the hidden layer neuron h_j (Figure A-3).

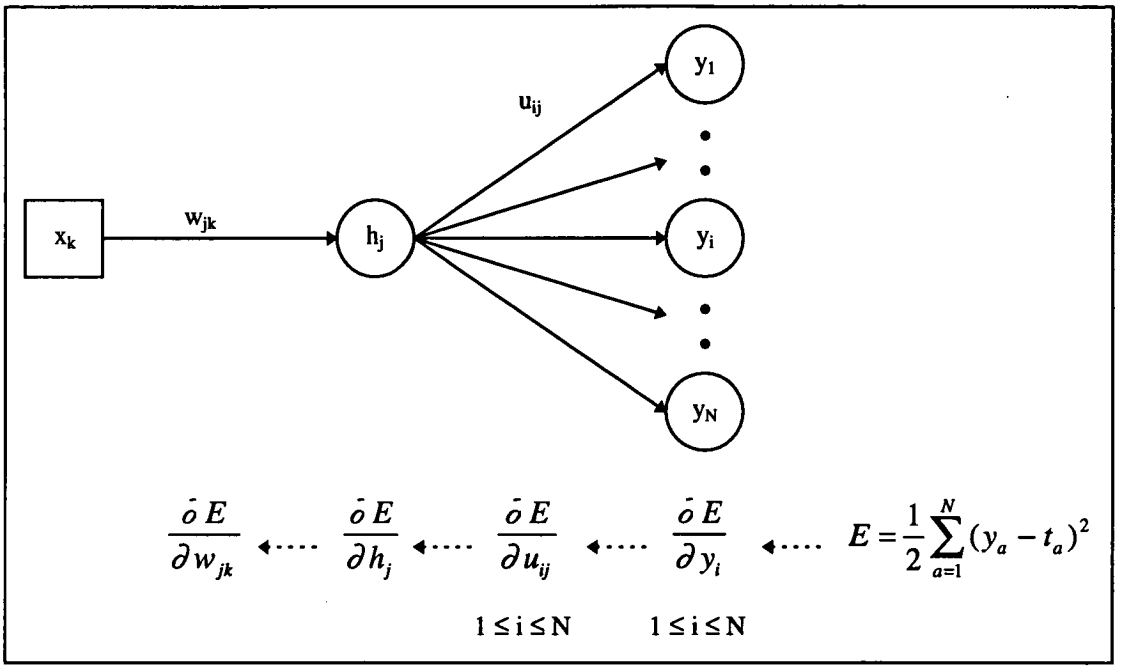


Figure A-3: Hidden Layer Weight and Output Layer Weights

The objective in this derivation is to determine $\partial E/\partial w_{jk}$ given the network error E. This differential can be expressed as follows:

$$\frac{\partial E}{\partial w_{jk}} = \frac{\partial E}{\partial h_j} \cdot \frac{\partial h_j}{\partial net_j} \cdot \frac{\partial net_j}{\partial w_{jk}} \tag{A-20}$$

Expanding these differentials, we get the following:

$$\begin{aligned}
\frac{\partial E}{\partial h_j} &= \frac{\partial}{\partial h_j}\left[\frac{1}{2}\sum_{a=1}^{N}(y_a - t_a)^2\right] \\
&= \sum_{a=1}^{N}\left[\frac{\partial}{\partial y_a}\left\{\frac{1}{2}(y_a - t_a)^2\right\} \cdot \frac{\partial y_a}{\partial net_a} \cdot \frac{\partial net_a}{\partial h_j}\right] \\
&= \sum_{a=1}^{N}\left[(y_a - t_a) \cdot \frac{\partial}{\partial net_a}\{g(net_a)\} \cdot \frac{\partial}{\partial h_j}\left\{\sum_{b=1}^{H} u_{ab} h_b\right\}\right]
\end{aligned}$$

$$\begin{aligned}
\frac{\partial E}{\partial h_j} &= \sum_{a=1}^{N}\left[(y_a - t_a) \cdot g'(net_a) \cdot u_{aj}\right] \\
&= \sum_{a=1}^{N}\left[\delta_a \cdot g'(net_a) \cdot u_{aj}\right]
\end{aligned} \tag{A-21}$$

$$\text{let } \delta_j = \sum_{a=1}^{N}\left[\delta_a \cdot g'(net_a)\cdot u_{aj}\right], \text{ therefore} \tag{A-22}$$

$$\frac{\partial E}{\partial h_j} = \delta_j \tag{A-23}$$

$$\begin{aligned}\frac{\partial h_j}{\partial net_j} &= \frac{\partial}{\partial net_j}\left[f(net_j)\right] \\ &= f'(net_j)\end{aligned} \tag{A-24}$$

$$\begin{aligned}\frac{\partial net_j}{\partial w_{jk}} &= \frac{\partial}{\partial w_{jk}}\left[\sum_{c=1}^{M} w_{jc}x_c\right] \\ &= x_k\end{aligned} \tag{A-25}$$

Substituting (A-23), (A-24), and (A-25) into (A-20), then:

$$\frac{\partial E}{\partial w_{jk}} = \delta_j \cdot f'(net_j)\cdot x_k \tag{A-26}$$

Therefore (A-26) can be used to calculate update for w_{jk} using the gradient decent rule (A-7):

$$\begin{aligned}\Delta w_{jk} &= -\eta\frac{\partial E}{\partial w_{jk}} \\ &= -\eta\left[\delta_j \cdot f'(net_j)\cdot x_k\right]\end{aligned} \tag{A-27}$$

where, if $f(net_j) = \dfrac{1}{1+e^{-net_j}}$, then $f'(net_j) = h_j(1-h_j)$

A.4 Generalised Weight Adjustment

In light of the previous two sections, it is possible to describe the weight update of a network with any number of hidden layers. The results of the previous two sections are summarised by (A-18) and (A-26). That is:

$$\frac{\partial E}{\partial u_{ij}} = \delta_i \cdot g'(net_i) \cdot h_j \qquad \{A\text{-}18\}$$

$$\text{where, } \delta_i = \frac{\partial E}{\partial y_i} = (y_i - t_i) \qquad \{A\text{-}13\}$$

$$\frac{\partial E}{\partial w_{jk}} = \delta_j \cdot f'(net_j) \cdot x_k \qquad \{A\text{-}25\}$$

$$\text{where, } \delta_j = \frac{\partial E}{\partial h_j} = \sum_{a=1}^{N} \left[\delta_a \cdot g'(net_a) \cdot u_{aj} \right] \qquad \{A\text{-}22\}$$

Clearly (A-18) and (A-25) are identical in form, and (A-22) provides an insight into how the error from one neural layer (δ_a) may be back-propagated to the previous neural layer (δ_j). Therefore, a generalised form of these results can be derived.

Consider the neural layers described in Figure A-4. Neuron m_j is connected to neuron n_i via the weight v_{ij}. In addition to this, this model also proposes the addition of an input bias b_i which is connects a unit input node to neuron n_i.

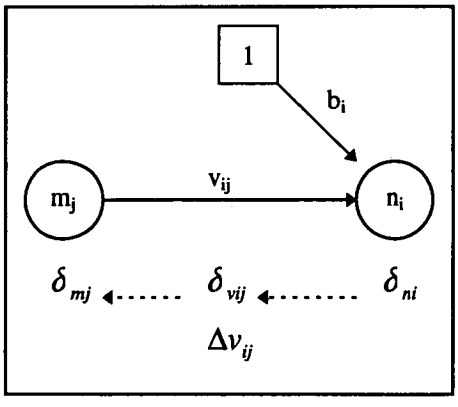


Figure A-4: Generalised relationship between two neural layers.

Therefore the output from the n_i is defined as:

$$net_{ni} = \left(\sum_j m_j v_{ij} \right) + b_i \tag{A-28}$$

$$n_i = f(net_{ni}) \tag{A-29}$$

If the n_i neurons form the output layer, then the error differential δ_{ni} is defined as follows:

$$\delta_{ni} = \frac{\partial E}{\partial n_i} = t_i - n_i \tag{A-30}$$

If the ni neurons do not form the output layer, then the error differential δ_{ni} is calculated using the error differential from the next layer (using same form as equation for δ_{mj} below).

Generalising (A-18) and (A-25), the error differential δ_{vij} is defined as:

$$\delta_{vij} = \frac{\partial E}{\partial v_{ij}} = \delta_{ni} \cdot f'(net_{ni}) \cdot m_j \tag{A-31}$$

Therefore the update for v_{ij} is defined as:

$$\Delta v_{ij} = -\eta \delta_{vij} \tag{A-32}$$

Similarly, the error differential and update for b_i is:

$$\delta_{bi} = \delta_{ni} \cdot f'(net_{ni}) \cdot 1 \tag{A-33}$$

$$\Delta b_i = -\eta \delta_{bi} \tag{A-34}$$

Finally, generalising (A-22) the error differential δ_{mj} is defined as:

$$\delta_{mj} = \frac{\partial E}{\partial m_j} = \sum_i \left(\delta_{ni} \cdot f'(net_{ni}) \cdot v_{ij} \right) \tag{A-35}$$

where

$$\text{If } f(net_{ni}) = net_{ni}, \text{ then } f'(net_{ni}) = 1 \tag{A-36}$$

$$\text{and, if } f(net_{ni}) = \frac{1}{1+e^{-net_{ni}}}, \text{ then } f'(net_{ni}) = n_i(1-n_i) \tag{A-37}$$

This generalised model now provides a framework for the definition of the generalised matrix model of backpropagation described in the next section.

A.5 Generalised Matrix Model of Backpropagation

This matrix model of feed-forward networks and backpropagation was developed to simplify the construction and analysis of networks used in this thesis. This model provides a framework for the Classification Object Library (COL) described in Appendix B.

Before describing this model, two new vector operators will be defined to simplify the description and implementation. These binary operators are the component-by-component multiply, indicated by a % symbol, and the component-by-component cross-multiply, indicated by a & symbol. These operators are defined as follows:

component-by-component multiply

$$X\%Y = \begin{bmatrix} x_1 \\ \cdot \\ \cdot \\ x_n \end{bmatrix} \% \begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 y_1 \\ \cdot \\ \cdot \\ x_n y_n \end{bmatrix}$$

component-by-component cross-multiply

$$X\&Y = \begin{bmatrix} x_1 \\ \cdot \\ \cdot \\ x_m \end{bmatrix} \& \begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 y_1 & \cdot & \cdot & x_1 y_n \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ x_m y_1 & \cdot & \cdot & x_m y_n \end{bmatrix}$$

Also note that this model makes no distinction between vectors and matrices, since a vector can be expressed as a matrix with a single column. This simplifies the model as only one class, *matrix*, needs to be defined when implementing this model.

All of the above in perspective, a generalised matrix model of a feed-forward neural network may be described by Figure A-5, where X_0 is the input for the network and X_n is the output.

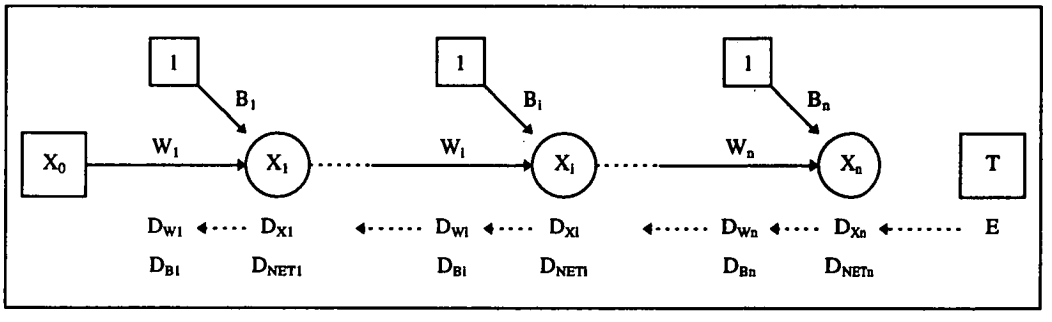


Figure A-5: Generalised Matrix Model

If N_i describes the number of neurons in each layer, then each layer is expressed as:

$$L_i = (W_i, B_i, X_i, D_{Wi}, D_{Bi}, D_{Xi}, D_{NETi})$$

where W_i is the input weights, B_i is the input bias, X_i is the layer output, and $D_{Wi}, D_{Bi}, D_{Xi}, D_{NETi}$ refer to the following:

$$D_{Wi} = \frac{\partial E}{\partial W_i}, D_{Bi} = \frac{\partial E}{\partial B_i}, D_{Xi} = \frac{\partial E}{\partial X_i}, D_{NETi} = \frac{\partial E}{\partial NET_i}$$

and the dimensions of these parameters are:

Variable	Type	Dimensions (r,c)
W_i, D_{Wi}	matrix	(N_i , N_{i-1})
X_i, B_i, D_{Xi}, D_{Bi}, D_{NETi}	vector	(Ni , 1)
X_{i-1}	vector	(N_{i-1} , 1)

The network may be initialised by applying the randomise function R{}:

$$D_{Wi} = 0, \text{ for all i}$$

$$D_{Bi} = 0, \text{ for all i}$$

$$R\{W_i\} \text{ for all i}$$

$$R\{B_i\} \text{ for all i}$$

The network can then be iteratively trained using the training dataset [(P_1,T_1), . (P_e,P_e) . (P_S,T_S)], the pattern P_e is applied to the network input and feed forward through each layer (L_1 through to L_n):

$$X_0 = P_e$$

$$NET_i = W_i X_{i-1} + B_i$$

$$X_i = F_i\{NET_i\}$$

Applying (A-30) the output error differential can be calculated:

$$D_{Xn} = X_n - T_e$$

and the error may be propagated back through L_n through to L_1 using (A-31), (A-33), and (A-35):

$$D_{NETi} = D_{Xi} \% F_i'\{NET_i\}$$

$$D_{Wi} = D_{Wi} + (D_{NETi} \& X_{i-1})$$

$$D_{Bi} = D_{Bi} + D_{NETi}$$

$$D_{X(i-1)} = W_i^T D_{NETi}$$

where:

$F_i'\{NET_i\} = 1$, if F_i is a linear activation functions,

or $F_i'\{NET_i\} = X_i \%(1 - X_i)$, if F_i is a sigmoid activation function.

Finally, the weight matrices and bias matrices can be used using (A-32), and (A-34):

$$W_i = W_i - \eta D_{Wi}$$

$$B_i = B_i - \eta D_{Bi}$$

$$D_{Wi} = 0$$

$$D_{Bi} = 0$$

where η is the learning rate defined for the training session.

Appendix A

B. Classification Tools

B.1 Introduction

This appendix describes the software tools developed to assist in performing and analysing the classification experiments conducted in this thesis. The first section will provide a general overview of the tools developed (CTAS - Classifier Training and Analysis Suite) and the underlying object model used to encapsulate specific aspects of the classifier training and analysis process (COL - Classifier Object Model). The second section documents the specific tools in CTAS and how they are used. The third section provides a detailed description of COL and the various object classes developed to simplify the development of CTAS. The fourth and final section provides some practical examples of how to use CTAS to train and analyse feed forward neural networks and how CTAS may be used to analyse classification results generated by third party classification programs.

CTAS and COL were implemented in ANSI C/C++ and rely heavily on the object-oriented aspects of C++. The CTAS tools are designed for use in a command line environment and have been tested under a number of user environments (DOS 5.0-6.0, NT-Shell, and a range of UNIX environments; SunOS, Solaris, and AIX).

B.2 Overview

The Classifier Training and Analysis Suite (CTAS) was originally developed to create, train, and analyse the performance of feed-forward neural networks. The aim was to make these tools flexible and generic, capable of constructing and manipulating a range of network architectures and configurations. In relation to training, the tools needed to allow the user to define the network architecture, training parameters, and training procedures to be used. Similarly the analysis tools needed to be flexible and configurable.

B.2.1 CTAS Tools

The final set of tools developed consisted of one training program, *trainer*, and four analysis programs; *performance*, *breakdown*, *threshold*, *binning*, *compare*. Their function is briefly described below:

trainer allows the user to construct and train a feed-forward neural network. The user can specify the network architecture to be used; number of input and output nodes, number of neurons in each layer, number of layers, bias inputs, and activation function to be used. The user can also specify the dataset to used for training, training parameters, stopping criteria, and the number of runs to be performed. The program's output is configurable, providing a range of intermediate and final results. The final networks from each run may be stored as well as the final output vectors for the specified training and testing datasets. The network and vectors output generated by trainer can be analysed by any of the following analysis tools.

performance analyses the classification of the training and testing datasets and calcuates the classification performance (percentage correct) for individual classes in each dataset and the overall classification performance for each dataset.

breakdown provides a detailed breakdown of the classification performance of both the training and testing datasets. Each dataset is broken down into class groups, and for each class group the program calculates what percentage of that group was classified into each of the possible class outputs by the classifier being considered.

threshold allows the user to test a range of thresholding techniques. The program calculates a set of percentage correct and percentage classified results for the threshold applied. This program produces overall thresholding results for the training and testing dataset as well as providing a breakdown of thresholding results for each class group in the datasets being considered. For more details concerning the background to thresholding and its benifits in relation to classifier reliability, please refer to chapter 10.

binning similar to threshold, this program allows the user to test the output binning technique described in chapter 10. The user can specify the number of bins required and the program will calculate the percentange correct and percentage classified in each bin. As with threshold, binning generates overall results for each dataset as well as results for each class group in the datasets.

compare compares the output levels of patterns classified correctly with patterns classified incorrectly. This program is useful for determining if there is any significant difference between output levels. As with threshold and binning, compare produces overall dataset results as well as results for each class group in the datasets.

B.2.2 Conducting an Experiment

Figure 1 provides a conceptual model of how the CTAS tools are used together. All the programs described above derive their configuration parameters from an experiment definition file, which defines all the dataset, classifier, training, and analysis parameters for an experiment. This approach of centralising experiment parameters was chosen because it provides a useful means of documenting and tracking the specifics of each experiment being conducted. To conduct an experiment, the user must construct an experiment definition file, and prepare training and testing dataset for the experiment. The user can then run *trainer*, which will generate a set of training results, which in turn can be analysed with any one of the analysis tools described above.

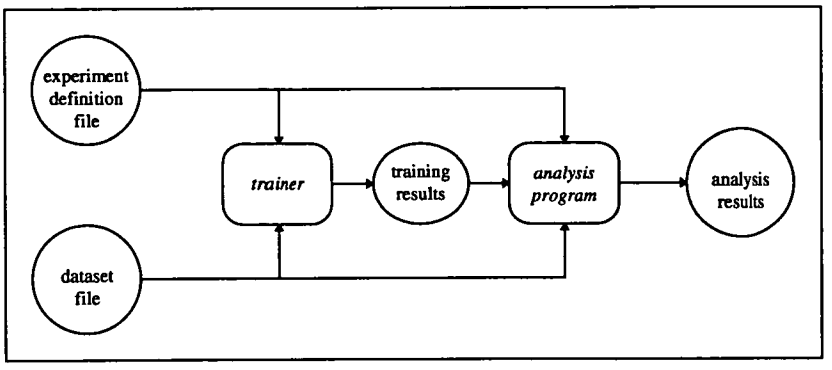


Figure 1: Conceptual model of CTAS training and analysis.

An experiment definition file name must be of the form *<experiment-name>.exp*. To execute the training phase, *trainer* is run, passing the *experiment-name* as a command line argument (note all CTAS tools assume that the experiment files have the extension *.exp*):

```
trainer <experiment-name>
```

trainer will load the experiment parameters, load the datasets, create the network, and then proceed to train the network with the training set. During training, *trainer* provides feedback of training progress via standard output. When the training phase is complete, trainer will output the training results in one of two possible forms; either by creating a file containing the network configurations and final weights (with the file name *<experiment-name>.networks*), or by creating a file containing the final classifer output vectors for each pattern in the training set followed by the classifier output vectors for the testing dataset (with the file name *<experiment-name>.vectors*). The experiment definition file will define which type of output is required.

When using a network file an analysis program loads and constructs the network and then calculates the output vectors by applying this network to the specified dataset, this approach is far more flexible with regard to analysis, as datasets other than the training and testing datasets used in the training phase can be used in the analysis phase. Alternatively, when using output vector files, an analysis program simply matches the output vectors with the patterns in the training and testing datasets (note that output vector files are sequentially ordered to match the order of the examples in the training and testing datasets, with the training output vectors first followed by the testing output vectors).

These training results can then be analysed using any of the analysis tools. The user can simply run any one of the analyse tools passing the *experiment-name* as a command line argument, for example:

```
performance <experiment-name>
```

After loading the experiment definition file, datasets, and training results, the analysis tools calculate the anlaysis results and output then via standard output.

Please note that the user can specify or override any of the experiment parameters on the command line, for example:

```
<tool-name> <experiment-name> [<override parameters>]
```

This provides some extra flexiblity, particularly if the user is wishing to trial a number of experiments, where only one parameter is varied between each experiment. For example, the user may wish to trial an experiment by the name of *experiment1* using a range of different learning rates for training the network, this could be done as follows:

```
trainer experiment1 -T10.01
performance experiment1
trainer experiment1 -T10.5
performance experiment1
trainer experiment1 -T10.1
performance experiment1
```

The dataset format used by all CTAS tools is an adaptation of Quinlan's C4.5 data format (Quinlan 1993) with some extensions. For a more complete discription of how to construct experiment files and how to use the CTAS tools, please refer to the next section,

B.2.3 Using 3rd Party Classifiers

There are two reason for designing the analysis tools to handle output vector files. Firstly, in some cases, the size of a network file can be excessively large, and it may not be completely neccessary, with regard to the experiments being conducted, to store this large network file. In such cases, it may be far more space efficient to simply store the output vectors, which still provide the analysis tools with sufficient information for calculating results. The second reason for designing the analysis tools to handle output vector files is to allow these tools to be used to analyse the the classification results of third party classification tools. Conceptually this is descibed in Figure 2.

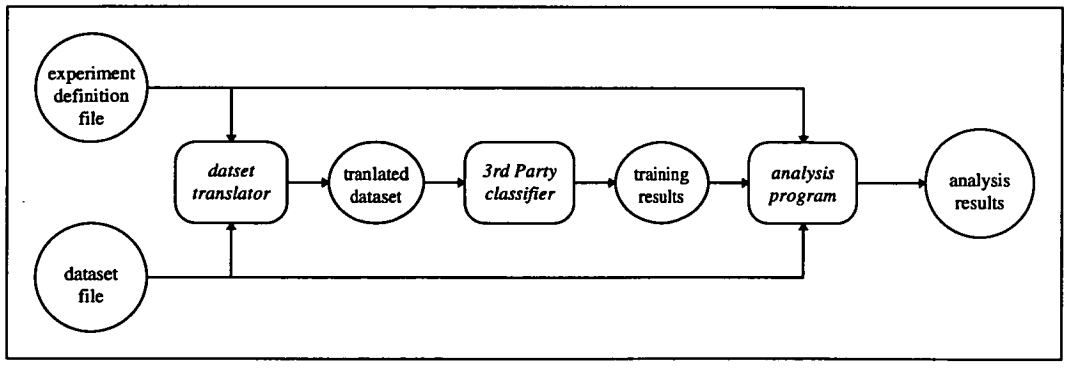


Figure 2 : Conceptual model of how CTAS is used with 3rd Party classifiers.

Since a number of classification techniques use different dataset formats a number of dataset translation programs where created to translate from CTAS dataset format to a number of different dataset formats:

C. Detailed Classification Breakdowns

C.1 Experiment E1.knn

Training Data

	anterior	inferior	normal	CAD
anterior	97.6	0.2	1.0	1.2
inferior	0.5	97.9	0.5	1.1
normal	0.0	0.0	97.7	2.3
CAD	0.3	0.3	0.0	99.3

Testing Data

	anterior	inferior	normal	CAD
anterior	76.6	6.3	6.9	10.2
inferior	7.6	66.7	5.9	19.9
normal	3.3	3.3	33.3	60.0
CAD	0.0	3.3	16.7	80.0

Training Data

	anterior	inferior	normal	CAD
anterior.6	99.0	0.0	0.0	1.0
anterior.12	99.0	1.0	0.0	0.0
anterior.48	100.0	0.0	0.0	0.0
anterior.fu	92.3	0.0	3.8	3.8
inferior.6	0.0	100.0	0.0	0.0
inferior.12	0.0	98.8	0.0	1.2
inferior.48	1.9	96.2	1.9	0.0
inferior.fu	0.0	96.8	0.0	3.2
normal.50	0.0	0.0	97.7	2.3
CAD.50	0.3	0.3	0.0	99.3

Testing Data

	anterior	inferior	normal	CAD
anterior.6	72.4	3.4	13.8	10.3
anterior.12	89.7	3.4	3.4	3.4
anterior.48	79.3	3.4	10.3	6.9
anterior.fu	65.0	15.0	0.0	20.0
inferior.6	3.3	76.7	6.7	13.3
inferior.12	10.0	76.7	6.7	6.7
inferior.48	13.3	63.3	6.7	16.7
inferior.fu	3.6	50.0	3.6	42.9
normal.50	3.3	3.3	33.3	60.0
CAD.50	0.0	3.3	16.7	80.0

C.2 Experiment E1.linreg

Training Data

	anterior	inferior	normal	CAD
anterior	89.9	2.3	4.9	2.8
inferior	2.0	89.1	4.4	4.6
normal	4.6	7.6	61.8	26.0
CAD	7.9	10.5	23.9	57.7

Testing Data

	anterior	inferior	normal	CAD
anterior	83.0	8.8	6.4	1.7
inferior	6.0	78.5	5.1	10.5
normal	3.3	3.3	30.0	63.3
CAD	0.0	6.7	13.3	80.0

Training Data

	anterior	inferior	normal	CAD
anterior.6	85.4	3.1	6.2	5.2
anterior.12	94.1	3.9	2.0	0.0
anterior.48	95.5	2.3	0.0	2.3
anterior.fu	84.6	0.0	11.5	3.8
inferior.6	1.1	95.7	2.2	1.1
inferior.12	0.0	91.5	7.3	1.2
inferior.48	1.9	98.1	0.0	0.0
inferior.fu	4.8	71.0	8.1	16.1
normal.50	4.6	7.6	61.8	26.0
CAD.50	7.9	10.5	23.9	57.7

Testing Data

	anterior	inferior	normal	CAD
anterior.6	86.2	6.9	6.9	0.0
anterior.12	96.6	0.0	3.4	0.0
anterior.48	79.3	3.4	10.3	6.9
anterior.fu	70.0	25.0	5.0	0.0
inferior.6	6.7	80.0	10.0	3.3
inferior.12	10.0	80.0	6.7	3.3
inferior.48	0.0	93.3	0.0	6.7
inferior.fu	7.1	60.7	3.6	28.6
normal.50	3.3	3.3	30.0	63.3
CAD.50	0.0	6.7	13.3	80.0

C.3 Experiment E1.C4.5

Training Data

	anterior	inferior	normal	CAD
anterior	98.5	1.3	0.0	0.3
inferior	0.0	98.2	0.8	1.0
normal	0.0	0.0	100.0	0.0
CAD	0.7	1.6	1.3	96.4

Testing Data

	anterior	inferior	normal	CAD
anterior	60.2	14.0	3.0	22.8
inferior	10.3	58.3	8.5	22.9
normal	13.3	13.3	26.7	46.7
CAD	10.0	16.7	30.0	43.3

Training Data

	anterior	inferior	normal	CAD
anterior.6	96.9	2.1	0.0	1.0
anterior.12	97.1	2.9	0.0	0.0
anterior.48	100.0	0.0	0.0	0.0
anterior.fu	100.0	0.0	0.0	0.0
inferior.6	0.0	100.0	0.0	0.0
inferior.12	0.0	97.6	0.0	2.4
inferior.48	0.0	100.0	0.0	0.0
inferior.fu	0.0	95.2	3.2	1.6
normal.50	0.0	0.0	100.0	0.0
CAD.50	0.7	1.6	1.3	96.4

Testing Data

	anterior	inferior	normal	CAD
anterior.6	69.0	13.8	3.4	13.8
anterior.12	75.9	3.4	0.0	20.7
anterior.48	75.9	13.8	3.4	6.9
anterior.fu	20.0	25.0	5.0	50.0
inferior.6	13.3	53.3	6.7	26.7
inferior.12	6.7	63.3	3.3	26.7
inferior.48	3.3	66.7	16.7	13.3
inferior.fu	17.9	50.0	7.1	25.0
normal.50	13.3	13.3	26.7	46.7
CAD.50	10.0	16.7	30.0	43.3

C.4 Experiment E1.MML

Training Data

	anterior	inferior	normal	CAD
anterior	82.2	5.1	3.5	9.1
inferior	4.1	86.0	1.6	8.3
normal	14.5	6.9	67.9	10.7
CAD	14.1	14.4	15.1	56.4

Testing Data

	anterior	inferior	normal	CAD
anterior	70.6	11.4	7.3	10.7
inferior	11.1	46.3	12.7	29.9
normal	16.7	10.0	16.7	56.7
CAD	10.0	16.7	36.7	36.7

Training Data

	anterior	inferior	normal	CAD
anterior.6	84.4	5.2	3.1	7.3
anterior.12	77.5	6.9	1.0	14.7
anterior.48	86.4	4.5	2.3	6.8
anterior.fu	80.8	3.8	7.7	7.7
inferior.6	5.4	83.9	2.2	8.6
inferior.12	3.7	82.9	2.4	11.0
inferior.48	5.8	88.5	1.9	3.8
inferior.fu	1.6	88.7	0.0	9.7
normal.50	14.5	6.9	67.9	10.7
CAD.50	14.1	14.4	15.1	56.4

Testing Data

	anterior	inferior	normal	CAD
anterior.6	72.4	10.3	10.3	6.9
anterior.12	72.4	0.0	10.3	17.2
anterior.48	72.4	10.3	3.4	13.8
anterior.fu	65.0	25.0	5.0	5.0
inferior.6	16.7	46.7	10.0	26.7
inferior.12	6.7	53.3	20.0	20.0
inferior.48	6.7	56.7	6.7	30.0
inferior.fu	14.3	28.6	14.3	42.9
normal.50	16.7	10.0	16.7	56.7
CAD.50	10.0	16.7	36.7	36.7

C.5 Experiment E1.bp

Training Data

	anterior	inferior	normal	CAD
anterior	83.4±2.3	5.3±1.5	4.5±2.2	6.8±1.6
inferior	2.8±0.6	82.0±2.3	4.7±1.8	10.5±2.7
normal	10.9±3.9	8.2±2.1	59.8±16.3	21.1±13.2
CAD	9.4±1.7	17.3±2.6	14.5±5.2	58.9±5.1

Testing Data

	anterior	inferior	normal	CAD
anterior	65.9±4.6	10.7±3.0	10.5±4.2	13.0±3.2
inferior	7.2±1.9	67.2±4.8	6.9±3.6	18.7±4.4
normal	19.7±5.5	7.8±4.7	30.7±10.7	41.8±12.6
CAD	4.8±3.9	14.3±4.8	22.7±9.0	58.2±9.7

Training Data

	anterior	inferior	normal	CAD
anterior.6	72.0±5.0	9.8±2.9	7.0±3.5	11.1±3.6
anterior.12	86.7±3.5	4.3±1.6	5.6±3.0	3.4±2.2
anterior.48	91.6±2.4	3.6±1.7	1.8±2.0	3.0±2.0
anterior.fu	83.3±4.8	3.5±3.2	3.7±3.9	9.6±3.7
inferior.6	3.6±1.4	80.5±4.6	6.5±2.8	9.4±4.4
inferior.12	1.2±0.9	84.9±3.6	2.9±2.3	11.0±3.1
inferior.48	2.2±1.3	89.7±3.2	3.1±2.1	5.0±2.9
inferior.fu	4.1±2.2	72.7±3.9	6.5±3.4	16.7±4.6
normal.50	10.9±3.9	8.2±2.1	59.8±16.3	21.1±13.2
CAD.50	9.4±1.7	17.3±2.6	14.5±5.2	58.9±5.1

Testing Data

	anterior	inferior	normal	CAD
anterior.6	64.7±6.8	10.3±5.5	12.2±6.2	12.8±5.6
anterior.12	72.1±7.5	4.1±3.6	8.3±4.8	15.5±6.2
anterior.48	76.7±5.0	5.9±2.7	8.6±4.2	8.8±2.8
anterior.fu	50.0±8.9	22.5±5.6	12.8±8.0	14.8±6.8
inferior.6	9.3±4.7	64.0±8.2	11.3±6.3	15.3±7.6
inferior.12	10.8±3.6	72.2±7.2	4.5±4.1	12.5±6.3
inferior.48	6.0±3.1	74.2±7.0	5.2±5.1	14.7±6.6
inferior.fu	2.5±3.0	58.6±6.6	6.6±4.7	32.3±7.1
normal.50	19.7±5.5	7.8±4.7	30.7±10.7	41.8±12.6
CAD.50	4.8±3.9	14.3±4.8	22.7±9.0	58.2±9.7

C.6 Experiment E1.qp

Training Data

	anterior	inferior	normal	CAD
anterior	77.0±6.0	11.1±5.1	6.2±5.4	5.8±5.4
inferior	6.2±4.1	84.1±5.5	3.5±3.0	6.1±5.5
normal	24.0±19.2	24.7±13.4	31.0±23.5	20.2±15.3
CAD	19.0±14.0	37.1±17.1	14.3±12.1	29.5±23.3

Testing Data

	anterior	inferior	normal	CAD
anterior	67.5±6.1	16.0±6.8	8.4±6.7	8.2±7.3
inferior	10.1±4.9	79.1±8.5	3.2±3.3	7.6±8.2
normal	30.0±22.5	17.8±11.7	25.3±20.2	26.8±19.1
CAD	14.8±14.3	37.0±18.8	15.7±14.5	32.5±26.5

Training Data

	anterior	inferior	normal	CAD
anterior.6	58.7±8.5	21.1±10.9	7.9±8.0	12.3±10.9
anterior.12	82.2±6.7	7.6±3.1	6.7±6.1	3.5±3.5
anterior.48	86.8±5.0	7.4±3.2	3.6±4.1	2.2±2.8
anterior.fu	80.2±6.7	8.3±5.7	6.3±5.3	5.2±5.6
inferior.6	9.4±7.8	81.9±7.2	3.9±3.5	4.8±4.3
inferior.12	3.8±2.7	88.8±4.9	2.2±2.3	5.1±5.1
inferior.48	4.2±3.7	88.2±4.1	3.8±3.0	3.7±3.2
inferior.fu	7.5±4.2	77.6±10.5	4.3±4.4	10.6±11.1
normal.50	24.0±19.2	24.7±13.4	31.0±23.5	20.2±15.3
CAD.50	19.0±14.0	37.1±17.1	14.3±12.1	29.5±23.3

Testing Data

	anterior	inferior	normal	CAD
anterior.6	62.4±6.7	16.7±10.0	11.9±9.8	9.0±8.0
anterior.12	74.7±8.8	8.6±6.9	6.4±7.1	10.3±9.7
anterior.48	77.2±6.8	8.6±5.3	8.4±6.7	5.7±5.0
anterior.fu	55.5±8.8	30.0±9.2	6.8±6.0	7.8±9.1
inferior.6	14.7±9.1	73.2±10.4	5.8±6.0	6.3±8.2
inferior.12	12.0±4.0	81.0±6.6	2.0±3.2	5.0±6.4
inferior.48	7.2±4.7	84.2±7.7	2.8±3.4	5.8±7.4
inferior.fu	6.4±7.5	78.0±15.3	2.1±2.6	13.4±14.5
normal.50	30.0±22.5	17.8±11.7	25.3±20.2	26.8±19.1
CAD.50	14.8±14.3	37.0±18.8	15.7±14.5	32.5±26.5

C.7 Experiment E1.cas

Training Data

	anterior	inferior	normal	CAD
anterior	96.2±1.6	0.8±0.4	2.4±1.5	0.6±0.4
inferior	10.6±2.2	85.1±2.3	1.9±1.0	2.4±1.5
normal	43.5±14.9	2.4±0.9	44.1±19.6	10.0±7.4
CAD	44.5±10.4	7.7±2.4	23.5±12.9	24.4±13.9

Testing Data

	anterior	inferior	normal	CAD
anterior	86.3±1.5	6.8±1.2	5.3±2.0	1.7±1.2
inferior	12.6±1.9	78.7±1.0	4.0±1.8	4.6±2.5
normal	27.7±11.6	6.7±1.5	40.0±13.2	25.7±12.0
CAD	25.7±11.1	1.7±3.1	35.7±21.6	37.0±22.2

Training Data

	anterior	inferior	normal	CAD
anterior.6	94.8±1.9	0.9±0.7	2.6±1.6	1.7±1.2
anterior.12	98.5±1.0	0.7±0.4	0.6±0.8	0.2±0.4
anterior.48	96.1±1.0	1.6±1.0	1.8±0.9	0.5±0.9
anterior.fu	95.4±4.1	0.0±0.0	4.6±4.1	0.0±0.0
inferior.6	6.6±2.0	92.2±2.0	1.0±0.8	0.3±0.5
inferior.12	5.5±2.3	91.8±2.3	2.6±1.3	0.1±0.4
inferior.48	8.8±2.1	90.2±2.2	0.0±0.0	1.0±1.3
inferior.fu	21.6±5.7	66.1±5.1	4.2±2.7	8.1±5.0
normal.50	43.5±14.9	2.4±0.9	44.1±19.6	10.0±7.4
CAD.50	44.5±10.4	7.7±2.4	23.5±12.9	24.4±13.9

Testing Data

	anterior	inferior	normal	CAD
anterior.6	90.3±1.4	3.4±0.0	5.5±2.8	0.7±1.4
anterior.12	96.9±2.9	1.4±1.7	1.7±3.2	0.0±0.0
anterior.48	87.9±3.9	2.8±1.4	8.3±2.3	1.0±2.2
anterior.fu	70.0±4.5	19.5±4.7	5.5±2.7	5.0±3.2
inferior.6	10.3±1.0	82.0±2.2	5.3±1.6	2.3±1.5
inferior.12	15.3±2.2	79.3±2.9	2.7±2.0	2.7±2.5
inferior.48	5.7±2.1	93.7±2.3	0.3±1.0	0.3±1.0
inferior.fu	18.9±7.2	60.0±4.7	7.9±5.9	13.2±7.5
normal.50	27.7±11.6	6.7±1.5	40.0±13.2	25.7±12.0
CAD.50	25.7±11.1	1.7±3.1	35.7±21.6	37.0±22.2

C.8 Experiment E1.cbp

Training Data

	anterior	inferior	normal	CAD
anterior	91.1	3.9	1.0	4.0
inferior	1.4	89.9	3.6	5.2
normal	9.2	6.1	75.6	9.2
CAD	6.9	12.8	9.2	71.1

Testing Data

	anterior	inferior	normal	CAD
anterior	76.1	8.4	3.4	12.0
inferior	9.2	76.0	0.8	14.0
normal	26.7	3.3	26.7	43.3
CAD	3.3	16.7	16.7	63.3

Training Data

	anterior	inferior	normal	CAD
anterior.6	83.3	9.4	2.1	5.2
anterior.12	93.1	3.9	2.0	1.0
anterior.48	95.5	2.3	0.0	2.3
anterior.fu	92.3	0.0	0.0	7.7
inferior.6	2.2	90.3	4.3	3.2
inferior.12	0.0	92.7	1.2	6.1
inferior.48	1.9	94.2	3.8	0.0
inferior.fu	1.6	82.3	4.8	11.3
normal.50	9.2	6.1	75.6	9.2
CAD.50	6.9	12.8	9.2	71.1

Testing Data

	anterior	inferior	normal	CAD
anterior.6	72.4	6.9	10.3	10.3
anterior.12	79.3	0.0	0.0	20.7
anterior.48	82.8	6.9	3.4	6.9
anterior.fu	70.0	20.0	0.0	10.0
inferior.6	10.0	76.7	3.3	10.0
inferior.12	16.7	80.0	0.0	3.3
inferior.48	6.7	90.0	0.0	3.3
inferior.fu	3.6	57.1	0.0	39.3
normal.50	26.7	3.3	26.7	43.3
CAD.50	3.3	16.7	16.7	63.3

C.9 Experiment E1.cqp

Training Data

	anterior	inferior	normal	CAD
anterior	85.2	8.4	2.0	4.4
inferior	4.5	88.8	2.4	4.3
normal	15.3	16.8	42.0	26.0
CAD	9.8	28.5	13.8	47.9

Testing Data

	anterior	inferior	normal	CAD
anterior	77.0	12.3	2.6	8.1
inferior	10.9	83.0	0.8	5.3
normal	26.7	13.3	30.0	30.0
CAD	6.7	33.3	16.7	43.3

Training Data

	anterior	inferior	normal	CAD
anterior.6	71.9	14.6	3.1	10.4
anterior.12	91.2	6.9	1.0	1.0
anterior.48	93.2	4.5	0.0	2.3
anterior.fu	84.6	7.7	3.8	3.8
inferior.6	5.4	90.3	2.2	2.2
inferior.12	1.2	95.1	0.0	3.7
inferior.48	1.9	92.3	5.8	0.0
inferior.fu	9.7	77.4	1.6	11.3
normal.50	15.3	16.8	42.0	26.0
CAD.50	9.8	28.5	13.8	47.9

Testing Data

	anterior	inferior	normal	CAD
anterior.6	69.0	13.8	6.9	10.3
anterior.12	86.2	3.4	0.0	10.3
anterior.48	82.8	6.9	3.4	6.9
anterior.fu	70.0	25.0	0.0	5.0
inferior.6	16.7	80.0	3.3	0.0
inferior.12	13.3	86.7	0.0	0.0
inferior.48	10.0	86.7	0.0	3.3
inferior.fu	3.6	78.6	0.0	17.9
normal.50	26.7	13.3	30.0	30.0
CAD.50	6.7	33.3	16.7	43.3

C.10 Experiment E1.ccas

Training Data

	anterior	inferior	normal	CAD
anterior	95.8	0.8	2.3	1.1
inferior	9.1	85.9	2.7	2.3
normal	52.7	1.5	39.7	6.1
CAD	51.8	5.9	16.4	25.9

Testing Data

	anterior	inferior	normal	CAD
anterior	84.7	7.6	6.4	1.2
inferior	12.9	77.6	3.4	6.1
normal	26.7	6.7	30.0	36.7
CAD	20.0	0.0	26.7	53.3

Training Data

	anterior	inferior	normal	CAD
anterior.6	92.7	0.0	5.2	2.1
anterior.12	99.0	1.0	0.0	0.0
anterior.48	95.5	2.3	0.0	2.3
anterior.fu	96.2	0.0	3.8	0.0
inferior.6	5.4	92.5	1.1	1.1
inferior.12	3.7	91.5	4.9	0.0
inferior.48	9.6	90.4	0.0	0.0
inferior.fu	17.7	69.4	4.8	8.1
normal.50	52.7	1.5	39.7	6.1
CAD.50	51.8	5.9	16.4	25.9

Testing Data

	anterior	inferior	normal	CAD
anterior.6	89.7	3.4	6.9	0.0
anterior.12	93.1	3.4	3.4	0.0
anterior.48	86.2	3.4	10.3	0.0
anterior.fu	70.0	20.0	5.0	5.0
inferior.6	10.0	80.0	6.7	3.3
inferior.12	13.3	80.0	3.3	3.3
inferior.48	6.7	93.3	0.0	0.0
inferior.fu	21.4	57.1	3.6	17.9
normal.50	26.7	6.7	30.0	36.7
CAD.50	20.0	0.0	26.7	53.3

C.11 Experiment L1.knn

Training Data

	anterior	inferior	normal	CAD
anterior	96.8	0.8	0.6	1.9
inferior	0.0	98.7	1.3	0.0
normal	0.8	0.8	95.4	3.1
CAD	0.3	0.7	0.0	99.0

Testing Data

	anterior	inferior	normal	CAD
anterior	69.7	10.2	6.0	14.1
inferior	5.9	73.4	4.3	16.4
normal	3.3	3.3	36.7	56.7
CAD	0.0	6.7	16.7	76.7

Training Data

	anterior	inferior	normal	CAD
anterior.6	99.0	1.0	0.0	0.0
anterior.12	98.0	2.0	0.0	0.0
anterior.48	97.7	0.0	2.3	0.0
anterior.fu	92.3	0.0	0.0	7.7
inferior.6	0.0	100.0	0.0	0.0
inferior.12	0.0	100.0	0.0	0.0
inferior.48	0.0	98.1	1.9	0.0
inferior.fu	0.0	96.8	3.2	0.0
normal.50	0.8	0.8	95.4	3.1
CAD.50	0.3	0.7	0.0	99.0

Testing Data

	anterior	inferior	normal	CAD
anterior.6	69.0	6.9	10.3	13.8
anterior.12	79.3	6.9	3.4	10.3
anterior.48	65.5	6.9	10.3	17.2
anterior.fu	65.0	20.0	0.0	15.0
inferior.6	3.3	76.7	6.7	13.3
inferior.12	6.7	93.3	0.0	0.0
inferior.48	10.0	70.0	3.3	16.7
inferior.fu	3.6	53.6	7.1	35.7
normal.50	3.3	3.3	36.7	56.7
CAD.50	0.0	6.7	16.7	76.7

C.12 Experiment L1.linreg

Training Data

	anterior	inferior	normal	CAD
anterior	91.4	2.8	1.3	4.5
inferior	0.7	96.5	1.5	1.3
normal	0.8	3.1	74.0	22.1
CAD	4.9	9.5	20.7	64.9

Testing Data

	anterior	inferior	normal	CAD
anterior	73.6	13.1	9.4	3.8
inferior	2.6	80.2	8.6	8.6
normal	10.0	3.3	36.7	50.0
CAD	0.0	3.3	33.3	63.3

Training Data

	anterior	inferior	normal	CAD
anterior.6	86.5	3.1	2.1	8.3
anterior.12	88.2	5.9	2.9	2.9
anterior.48	90.9	2.3	0.0	6.8
anterior.fu	100.0	0.0	0.0	0.0
inferior.6	0.0	98.9	0.0	1.1
inferior.12	1.2	95.1	1.2	2.4
inferior.48	0.0	100.0	0.0	0.0
inferior.fu	1.6	91.9	4.8	1.6
normal.50	0.8	3.1	74.0	22.1
CAD.50	4.9	9.5	20.7	64.9

Testing Data

	anterior	inferior	normal	CAD
anterior.6	79.3	10.3	10.3	0.0
anterior.12	75.9	10.3	13.8	0.0
anterior.48	79.3	6.9	3.4	10.3
anterior.fu	60.0	25.0	10.0	5.0
inferior.6	3.3	86.7	10.0	0.0
inferior.12	0.0	90.0	3.3	6.7
inferior.48	3.3	83.3	3.3	10.0
inferior.fu	3.6	60.7	17.9	17.9
normal.50	10.0	3.3	36.7	50.0
CAD.50	0.0	3.3	33.3	63.3

C.13 Experiment L1.C4.5

Training Data

	anterior	inferior	normal	CAD
anterior	99.0	0.0	0.8	0.3
inferior	0.0	98.9	0.0	1.1
normal	0.0	0.0	100.0	0.0
CAD	1.6	1.0	2.0	95.4

Testing Data

	anterior	inferior	normal	CAD
anterior	65.6	4.2	6.8	23.3
inferior	4.5	76.5	2.6	16.4
normal	20.0	3.3	26.7	50.0
CAD	23.3	10.0	16.7	50.0

Training Data

	anterior	inferior	normal	CAD
anterior.6	97.9	0.0	1.0	1.0
anterior.12	98.0	0.0	2.0	0.0
anterior.48	100.0	0.0	0.0	0.0
anterior.fu	100.0	0.0	0.0	0.0
inferior.6	0.0	98.9	0.0	1.1
inferior.12	0.0	100.0	0.0	0.0
inferior.48	0.0	100.0	0.0	0.0
inferior.fu	0.0	96.8	0.0	3.2
normal.50	0.0	0.0	100.0	0.0
CAD.50	1.6	1.0	2.0	95.4

Testing Data

	anterior	inferior	normal	CAD
anterior.6	79.3	3.4	6.9	10.3
anterior.12	79.3	0.0	0.0	20.7
anterior.48	69.0	3.4	10.3	17.2
anterior.fu	35.0	10.0	10.0	45.0
inferior.6	0.0	93.3	0.0	6.7
inferior.12	0.0	80.0	6.7	13.3
inferior.48	0.0	90.0	0.0	10.0
inferior.fu	17.9	42.9	3.6	35.7
normal.50	20.0	3.3	26.7	50.0
CAD.50	23.3	10.0	16.7	50.0

C.14 Experiment L1.MML

Training Data

	anterior	inferior	normal	CAD
anterior	92.2	2.2	1.3	4.3
inferior	1.2	92.9	1.1	4.8
normal	0.0	0.0	100.0	0.0
CAD	4.6	8.5	3.0	83.9

Testing Data

	anterior	inferior	normal	CAD
anterior	72.8	6.4	11.4	9.4
inferior	2.6	85.4	5.1	7.0
normal	23.3	10.0	13.3	53.3
CAD	10.0	6.7	13.3	70.0

Training Data

	anterior	inferior	normal	CAD
anterior.6	86.5	3.1	3.1	7.3
anterior.12	90.2	2.0	2.0	5.9
anterior.48	100.0	0.0	0.0	0.0
anterior.fu	92.3	3.8	0.0	3.8
inferior.6	0.0	95.7	1.1	3.2
inferior.12	1.2	93.9	0.0	4.9
inferior.48	1.9	98.1	0.0	0.0
inferior.fu	1.6	83.9	3.2	11.3
normal.50	0.0	0.0	100.0	0.0
CAD.50	4.6	8.5	3.0	83.9

Testing Data

	anterior	inferior	normal	CAD
anterior.6	69.0	10.3	6.9	13.8
anterior.12	79.3	3.4	6.9	10.3
anterior.48	82.8	6.9	6.9	3.4
anterior.fu	60.0	5.0	25.0	10.0
inferior.6	0.0	93.3	6.7	0.0
inferior.12	0.0	90.0	10.0	0.0
inferior.48	6.7	86.7	0.0	6.7
inferior.fu	3.6	71.4	3.6	21.4
normal.50	23.3	10.0	13.3	53.3
CAD.50	10.0	6.7	13.3	70.0

C.15 Experiment L1.bp

Training Data

	anterior	inferior	normal	CAD
anterior	90.5±2.6	4.3±1.7	0.4±1.0	4.8±3.4
inferior	4.2±1.9	90.0±2.5	0.4±1.1	5.4±3.5
normal	27.1±16.3	21.9±15.7	7.8±23.0	43.2±29.6
CAD	19.6±15.8	22.8±18.2	1.5±4.5	56.1±32.5

Testing Data

	anterior	inferior	normal	CAD
anterior	74.2±6.3	13.1±4.2	1.4±4.2	11.2±7.2
inferior	12.7±4.2	71.9±6.6	1.0±3.0	14.5±8.7
normal	37.2±20.6	15.3±10.8	3.2±9.6	44.3±27.5
CAD	14.0±15.4	30.0±19.0	2.7±9.0	53.3±32.6

Training Data

	anterior	inferior	normal	CAD
anterior.6	88.3±3.4	5.9±2.2	0.4±1.2	5.4±4.2
anterior.12	91.3±2.3	4.2±2.3	0.4±1.2	4.1±3.1
anterior.48	92.2±2.4	4.4±2.4	0.1±0.5	3.3±2.8
anterior.fu	90.2±5.9	2.9±2.7	0.6±1.8	6.3±6.0
inferior.6	1.9±1.3	93.3±2.7	0.4±1.3	4.4±3.1
inferior.12	2.4±1.9	91.0±3.8	0.5±1.7	6.1±4.1
inferior.48	3.7±1.8	94.0±2.7	0.1±0.4	2.1±2.3
inferior.fu	8.8±4.7	81.6±4.7	0.4±1.4	9.2±6.2
normal.50	27.1±16.3	21.9±15.7	7.8±23.0	43.2±29.6
CAD.50	19.6±15.8	22.8±18.2	1.5±4.5	56.1±32.5

Testing Data

	anterior	inferior	normal	CAD
anterior.6	80.2±9.6	8.3±5.5	2.2±6.9	9.3±7.1
anterior.12	77.4±7.4	10.7±5.2	0.3±1.5	11.6±7.8
anterior.48	76.0±6.3	11.0±5.6	1.4±4.0	11.6±7.9
anterior.fu	63.2±9.9	22.5±8.0	1.8±5.8	12.5±10.7
inferior.6	10.5±5.8	73.0±9.4	1.7±5.1	14.8±9.8
inferior.12	11.7±3.9	75.2±7.8	0.7±2.3	12.5±8.6
inferior.48	8.2±3.7	77.2±9.1	0.8±3.0	13.8±10.3
inferior.fu	20.4±9.0	62.1±9.1	0.7±2.4	16.8±11.0
normal.50	37.2±20.6	15.3±10.8	3.2±9.6	44.3±27.5
CAD.50	14.0±15.4	30.0±19.0	2.7±9.0	53.3±32.6

C.16 Experiment L1.qp

Training Data

	anterior	inferior	normal	CAD
anterior	65.8±8.9	9.5±5.3	8.4±7.3	16.3±8.6
inferior	10.1±8.6	72.4±9.0	5.3±5.0	12.3±5.3
normal	17.1±14.8	12.2±7.2	33.2±23.9	37.5±13.2
CAD	12.2±8.1	16.5±8.7	12.4±9.9	58.9±12.9

Testing Data

	anterior	inferior	normal	CAD
anterior	53.4±8.2	13.5±8.2	11.4±9.6	21.7±12.0
inferior	13.8±8.0	62.3±7.9	5.7±6.1	18.2±6.6
normal	23.5±16.0	11.5±13.5	25.0±17.6	40.0±17.3
CAD	5.8±7.9	12.8±8.0	16.2±13.7	65.2±15.3

Training Data

	anterior	inferior	normal	CAD
anterior.6	52.0±11.0	13.8±8.4	9.8±9.5	24.4±13.0
anterior.12	76.5±8.3	6.2±3.0	6.8±6.5	10.5±6.2
anterior.48	79.4±7.3	5.7±4.1	7.0±6.7	7.8±5.2
anterior.fu	55.2±11.6	12.1±8.3	10.0±9.1	22.7±12.6
inferior.6	9.4±10.4	75.6±11.0	5.0±4.9	10.1±5.1
inferior.12	5.1±6.0	76.9±10.0	4.8±5.6	13.2±6.1
inferior.48	8.8±8.1	79.7±10.8	3.8±3.8	7.6±5.5
inferior.fu	17.1±12.3	57.3±8.5	7.4±7.2	18.2±8.6
normal.50	17.1±14.8	12.2±7.2	33.2±23.9	37.5±13.2
CAD.50	12.2±8.1	16.5±8.7	12.4±9.9	58.9±12.9

Testing Data

	anterior	inferior	normal	CAD
anterior.6	62.4±7.9	10.3±6.1	10.5±10.0	16.7±10.1
anterior.12	57.2±8.8	10.0±9.6	9.7±8.4	23.1±12.4
anterior.48	65.0±8.9	6.0±5.7	12.4±10.0	16.6±8.8
anterior.fu	29.0±12.0	27.5±15.0	13.0±12.9	30.5±19.2
inferior.6	13.0±10.7	62.8±11.9	5.5±6.1	18.7±8.6
inferior.12	10.8±4.2	70.2±5.6	2.8±4.5	16.2±7.3
inferior.48	12.3±7.7	64.8±10.0	7.2±7.3	15.7±8.8
inferior.fu	19.1±12.8	51.4±9.9	7.1±8.4	22.3±9.0
normal.50	23.5±16.0	11.5±13.5	25.0±17.6	40.0±17.3
CAD.50	5.8±7.9	12.8±8.0	16.2±13.7	65.2±15.3

C.17 Experiment L1.cas

Training Data

	anterior	inferior	normal	CAD
anterior	98.6±0.6	0.5±0.3	0.6±0.6	0.3±0.1
inferior	7.5±1.9	90.8±2.3	1.0±0.3	0.7±0.5
normal	24.7±7.4	1.1±0.7	67.7±8.3	6.5±4.1
CAD	34.4±9.0	3.9±1.2	30.0±7.4	31.6±10.5

Testing Data

	anterior	inferior	normal	CAD
anterior	89.1±1.7	6.6±1.2	3.2±0.8	1.1±1.0
inferior	13.6±1.3	77.2±1.8	5.8±1.0	3.4±1.3
normal	36.7±3.9	6.7±2.1	38.7±3.1	18.0±5.0
CAD	20.3±7.4	5.0±1.7	41.3±5.2	33.3±9.4

Training Data

	anterior	inferior	normal	CAD
anterior.6	97.3±0.8	1.2±0.4	0.4±0.7	1.0±0.5
anterior.12	98.5±1.5	0.9±1.2	0.6±1.0	0.0±0.0
anterior.48	98.6±1.1	0.0±0.0	1.4±1.1	0.0±0.0
anterior.fu	100.0±NaN	0.0±0.0	0.0±0.0	0.0±0.0
inferior.6	3.0±2.1	95.8±2.4	0.1±0.3	1.1±1.0
inferior.12	5.2±1.6	92.7±2.1	1.2±0.0	0.9±0.8
inferior.48	3.5±1.4	96.5±1.4	0.0±0.0	0.0±0.0
inferior.fu	18.2±5.1	78.1±5.5	2.7±1.0	1.0±0.8
normal.50	24.7±7.4	1.1±0.7	67.7±8.3	6.5±4.1
CAD.50	34.4±9.0	3.9±1.2	30.0±7.4	31.6±10.5

Testing Data

	anterior	inferior	normal	CAD
anterior.6	92.1±1.6	4.1±1.4	3.8±1.0	0.0±0.0
anterior.12	94.1±2.7	3.1±2.9	2.8±2.1	0.0±0.0
anterior.48	89.3±3.3	8.6±2.3	1.7±1.7	0.3±1.0
anterior.fu	81.0±4.9	10.5±1.5	4.5±1.5	4.0±4.4
inferior.6	4.0±2.0	91.7±2.2	3.3±0.0	1.0±1.5
inferior.12	9.3±3.6	86.3±2.8	0.3±1.0	4.0±1.3
inferior.48	10.0±1.5	86.3±1.8	2.7±2.0	1.0±1.5
inferior.fu	31.1±3.6	44.6±4.9	16.8±2.8	7.5±4.1
normal.50	36.7±3.9	6.7±2.1	38.7±3.1	18.0±5.0
CAD.50	20.3±7.4	5.0±1.7	41.3±5.2	33.3±9.4

C.18 Experiment L1.cbp

Training Data

	anterior	inferior	normal	CAD
anterior	96.4	0.8	0.0	2.9
inferior	1.2	97.4	0.0	1.4
normal	6.1	4.6	26.7	62.6
CAD	4.6	4.9	0.0	90.5

Testing Data

	anterior	inferior	normal	CAD
anterior	78.7	9.3	0.0	12.0
inferior	8.6	80.4	0.0	11.1
normal	23.3	3.3	0.0	73.3
CAD	6.7	13.3	0.0	80.0

Training Data

	anterior	inferior	normal	CAD
anterior.6	95.8	2.1	0.0	2.1
anterior.12	94.1	1.0	0.0	4.9
anterior.48	95.5	0.0	0.0	4.5
anterior.fu	100.0	0.0	0.0	0.0
inferior.6	1.1	98.9	0.0	0.0
inferior.12	0.0	97.6	0.0	2.4
inferior.48	1.9	98.1	0.0	0.0
inferior.fu	1.6	95.2	0.0	3.2
normal.50	6.1	4.6	26.7	62.6
CAD.50	4.6	4.9	0.0	90.5

Testing Data

	anterior	inferior	normal	CAD
anterior.6	82.8	6.9	0.0	10.3
anterior.12	82.8	3.4	0.0	13.8
anterior.48	79.3	6.9	0.0	13.8
anterior.fu	70.0	20.0	0.0	10.0
inferior.6	3.3	86.7	0.0	10.0
inferior.12	10.0	76.7	0.0	13.3
inferior.48	6.7	86.7	0.0	6.7
inferior.fu	14.3	71.4	0.0	14.3
normal.50	23.3	3.3	0.0	73.3
CAD.50	6.7	13.3	0.0	80.0

C.19 Experiment L1.cqp

Training Data

	anterior	inferior	normal	CAD
anterior	71.1	8.4	6.7	13.8
inferior	4.1	83.8	2.8	9.4
normal	5.3	9.2	47.3	38.2
CAD	7.2	12.8	5.6	74.4

Testing Data

	anterior	inferior	normal	CAD
anterior	55.0	13.1	10.6	21.3
inferior	9.5	71.8	2.5	16.2
normal	3.3	3.3	40.0	53.3
CAD	0.0	6.7	10.0	83.3

Training Data

	anterior	inferior	normal	CAD
anterior.6	58.3	12.5	8.3	20.8
anterior.12	80.4	4.9	3.9	10.8
anterior.48	84.1	4.5	6.8	4.5
anterior.fu	61.5	11.5	7.7	19.2
inferior.6	1.1	90.3	2.2	6.5
inferior.12	0.0	86.6	2.4	11.0
inferior.48	3.8	90.4	1.9	3.8
inferior.fu	11.3	67.7	4.8	16.1
normal.50	5.3	9.2	47.3	38.2
CAD.50	7.2	12.8	5.6	74.4

Testing Data

	anterior	inferior	normal	CAD
anterior.6	69.0	10.3	10.3	10.3
anterior.12	62.1	3.4	6.9	27.6
anterior.48	69.0	3.4	10.3	17.2
anterior.fu	20.0	35.0	15.0	30.0
inferior.6	6.7	73.3	3.3	16.7
inferior.12	3.3	83.3	3.3	10.0
inferior.48	10.0	70.0	3.3	16.7
inferior.fu	17.9	60.7	0.0	21.4
normal.50	3.3	3.3	40.0	53.3
CAD.50	0.0	6.7	10.0	83.3

C.20 Experiment L1.ccas

Training Data

	anterior	inferior	normal	CAD
anterior	97.9	0.3	0.8	1.1
inferior	5.9	91.8	1.1	1.2
normal	26.0	0.0	66.4	7.6
CAD	41.0	3.0	22.0	34.1

Testing Data

	anterior	inferior	normal	CAD
anterior	90.2	6.8	3.0	0.0
inferior	12.1	77.3	5.3	5.2
normal	33.3	3.3	40.0	23.3
CAD	13.3	0.0	40.0	46.7

Training Data

	anterior	inferior	normal	CAD
anterior.6	94.8	1.0	2.1	2.1
anterior.12	99.0	0.0	1.0	0.0
anterior.48	97.7	0.0	0.0	2.3
anterior.fu	100.0	0.0	0.0	0.0
inferior.6	1.1	96.8	0.0	2.2
inferior.12	2.4	95.1	1.2	1.2
inferior.48	3.8	96.2	0.0	0.0
inferior.fu	16.1	79.0	3.2	1.6
normal.50	26.0	0.0	66.4	7.6
CAD.50	41.0	3.0	22.0	34.1

Testing Data

	anterior	inferior	normal	CAD
anterior.6	93.1	3.4	3.4	0.0
anterior.12	96.6	0.0	3.4	0.0
anterior.48	86.2	13.8	0.0	0.0
anterior.fu	85.0	10.0	5.0	0.0
inferior.6	3.3	93.3	3.3	0.0
inferior.12	6.7	90.0	0.0	3.3
inferior.48	10.0	86.7	0.0	3.3
inferior.fu	28.6	39.3	17.9	14.3
normal.50	33.3	3.3	40.0	23.3
CAD.50	13.3	0.0	40.0	46.7

C.21 Experiment st1.knn

Training Data

	anterior	inferior	normal	CAD
anterior	95.2	1.5	1.0	2.3
inferior	0.0	97.9	0.5	1.6
normal	1.5	0.8	96.2	1.5
CAD	0.0	0.3	0.0	99.7

Testing Data

	anterior	inferior	normal	CAD
anterior	53.0	6.3	6.3	34.4
inferior	0.9	71.2	4.3	23.6
normal	6.7	6.7	0.0	86.7
CAD	3.3	0.0	30.0	66.7

Training Data

	anterior	inferior	normal	CAD
anterior.6	97.9	0.0	0.0	2.1
anterior.12	99.0	0.0	0.0	1.0
anterior.48	95.5	2.3	0.0	2.3
anterior.fu	88.5	3.8	3.8	3.8
inferior.6	0.0	100.0	0.0	0.0
inferior.12	0.0	100.0	0.0	0.0
inferior.48	0.0	98.1	1.9	0.0
inferior.fu	0.0	93.5	0.0	6.5
normal.50	1.5	0.8	96.2	1.5
CAD.50	0.0	0.3	0.0	99.7

Testing Data

	anterior	inferior	normal	CAD
anterior.6	72.4	3.4	6.9	17.2
anterior.12	62.1	0.0	3.4	34.5
anterior.48	72.4	6.9	0.0	20.7
anterior.fu	5.0	15.0	15.0	65.0
inferior.6	0.0	90.0	3.3	6.7
inferior.12	0.0	83.3	3.3	13.3
inferior.48	0.0	90.0	0.0	10.0
inferior.fu	3.6	21.4	10.7	64.3
normal.50	6.7	6.7	0.0	86.7
CAD.50	3.3	0.0	30.0	66.7

C.22 Experiment st1.linreg

Training Data

	anterior	inferior	normal	CAD
anterior	78.6	4.6	8.2	8.5
inferior	6.2	79.3	8.4	6.0
normal	4.6	8.4	53.4	33.6
CAD	11.1	9.5	26.9	52.5

Testing Data

	anterior	inferior	normal	CAD
anterior	72.0	4.2	12.8	11.0
inferior	7.7	68.0	11.9	12.3
normal	3.3	10.0	26.7	60.0
CAD	10.0	6.7	16.7	66.7

Training Data

	anterior	inferior	normal	CAD
anterior.6	75.0	1.0	11.5	12.5
anterior.12	79.4	5.9	6.9	7.8
anterior.48	90.9	0.0	6.8	2.3
anterior.fu	69.2	11.5	7.7	11.5
inferior.6	2.2	90.3	5.4	2.2
inferior.12	6.1	86.6	4.9	2.4
inferior.48	3.8	90.4	5.8	0.0
inferior.fu	12.9	50.0	17.7	19.4
normal.50	4.6	8.4	53.4	33.6
CAD.50	11.1	9.5	26.9	52.5

Testing Data

	anterior	inferior	normal	CAD
anterior.6	79.3	3.4	10.3	6.9
anterior.12	79.3	0.0	6.9	13.8
anterior.48	79.3	3.4	13.8	3.4
anterior.fu	50.0	10.0	20.0	20.0
inferior.6	10.0	80.0	10.0	0.0
inferior.12	6.7	73.3	10.0	10.0
inferior.48	0.0	86.7	13.3	0.0
inferior.fu	14.3	32.1	14.3	39.3
normal.50	3.3	10.0	26.7	60.0
CAD.50	10.0	6.7	16.7	66.7

C.23 Experiment st1.C4.5

Training Data

	anterior	inferior	normal	CAD
anterior	97.8	0.3	1.0	1.0
inferior	1.2	96.7	1.0	1.1
normal	0.0	0.0	100.0	0.0
CAD	1.3	2.0	0.7	96.1

Testing Data

	anterior	inferior	normal	CAD
anterior	58.4	7.2	8.5	25.8
inferior	10.3	66.2	8.7	14.8
normal	23.3	13.3	26.7	36.7
CAD	23.3	3.3	16.7	56.7

Training Data

	anterior	inferior	normal	CAD
anterior.6	97.9	1.0	0.0	1.0
anterior.12	93.1	0.0	3.9	2.9
anterior.48	100.0	0.0	0.0	0.0
anterior.fu	100.0	0.0	0.0	0.0
inferior.6	1.1	98.9	0.0	0.0
inferior.12	3.7	92.7	2.4	1.2
inferior.48	0.0	100.0	0.0	0.0
inferior.fu	0.0	95.2	1.6	3.2
normal.50	0.0	0.0	100.0	0.0
CAD.50	1.3	2.0	0.7	96.1

Testing Data

	anterior	inferior	normal	CAD
anterior.6	72.4	0.0	10.3	17.2
anterior.12	79.3	3.4	6.9	10.3
anterior.48	62.1	10.3	6.9	20.7
anterior.fu	20.0	15.0	10.0	55.0
inferior.6	3.3	93.3	3.3	0.0
inferior.12	16.7	66.7	3.3	13.3
inferior.48	3.3	83.3	3.3	10.0
inferior.fu	17.9	21.4	25.0	35.7
normal.50	23.3	13.3	26.7	36.7
CAD.50	23.3	3.3	16.7	56.7

C.24 Experiment st1.MML

Training Data

	anterior	inferior	normal	CAD
anterior	94.9	0.5	1.3	3.2
inferior	0.7	85.9	1.4	12.1
normal	0.0	0.0	100.0	0.0
CAD	2.3	7.2	6.2	84.3

Testing Data

	anterior	inferior	normal	CAD
anterior	61.4	6.7	12.3	19.6
inferior	7.7	71.2	5.2	15.8
normal	33.3	6.7	26.7	33.3
CAD	10.0	10.0	23.3	56.7

Training Data

	anterior	inferior	normal	CAD
anterior.6	93.7	0.0	2.1	4.2
anterior.12	92.2	2.0	1.0	4.9
anterior.48	97.7	0.0	2.3	0.0
anterior.fu	96.2	0.0	0.0	3.8
inferior.6	0.0	95.7	1.1	3.2
inferior.12	1.2	95.1	1.2	2.4
inferior.48	0.0	96.2	0.0	3.8
inferior.fu	1.6	56.5	3.2	38.7
normal.50	0.0	0.0	100.0	0.0
CAD.50	2.3	7.2	6.2	84.3

Testing Data

	anterior	inferior	normal	CAD
anterior.6	75.9	0.0	10.3	13.8
anterior.12	69.0	0.0	6.9	24.1
anterior.48	75.9	6.9	6.9	10.3
anterior.fu	25.0	20.0	25.0	30.0
inferior.6	3.3	90.0	3.3	3.3
inferior.12	3.3	83.3	3.3	10.0
inferior.48	10.0	86.7	0.0	3.3
inferior.fu	14.3	25.0	14.3	46.4
normal.50	33.3	6.7	26.7	33.3
CAD.50	10.0	10.0	23.3	56.7

C.25 Experiment st1.bp

Training Data

	anterior	inferior	normal	CAD
anterior	72.3±6.3	5.3±2.0	9.6±4.1	12.9±5.4
inferior	8.4±3.9	70.2±3.8	8.7±4.2	12.7±4.4
normal	15.9±5.1	6.9±3.1	54.5±10.7	22.7±8.9
CAD	17.6±6.5	10.3±2.9	21.3±8.6	50.9±11.1

Testing Data

	anterior	inferior	normal	CAD
anterior	62.3±9.3	7.9±2.9	14.2±6.3	15.7±6.0
inferior	11.0±3.1	66.6±3.4	9.0±5.8	13.4±5.2
normal	20.8±9.2	16.2±3.7	31.3±8.8	31:7±9.7
CAD	18.7±7.8	12.2±3.2	28.5±11.5	40.7±13.3

Training Data

	anterior	inferior	normal	CAD
anterior.6	70.0±9.3	2.7±1.7	11.5±5.5	15.8±7.8
anterior.12	72.7±6.8	5.8±2.9	9.9±6.2	11.6±5.2
anterior.48	82.8±7.1	2.7±2.1	7.3±4.9	7.2±3.7
anterior.fu	63.5±10.1	9.8±5.2	9.6±4.1	17.1±9.3
inferior.6	1.7±1.5	91.4±2.9	4.0±2.0	2.8±1.8
inferior.12	3.9±2.9	82.4±4.6	6.7±3.5	7.0±4.4
inferior.48	5.0±4.5	78.7±4.8	7.6±4.7	8.8±5.3
inferior.fu	23.1±10.0	28.2±8.3	16.4±11.1	32.3±11.2
normal.50	15.9±5.1	6.9±3.1	54.5±10.7	22.7±8.9
CAD.50	17.6±6.5	10.3±2.9	21.3±8.6	50.9±11.1

Testing Data

	anterior	inferior	normal	CAD
anterior.6	67.2±13.5	3.8±2.2	10.5±6.5	18.4±13.2
anterior.12	73.1±11.0	2.6±3.1	9.1±5.0	15.2±7.3
anterior.48	64.5±11.1	8.8±4.3	16.2±8.7	10.5±5.4
anterior.fu	44.2±13.4	16.2±6.7	20.8±13.9	18.8±9.5
inferior.6	3.7±3.0	84.2±4.8	6.3±4.5	5.8±3.1
inferior.12	5.3±3.2	82.2±4.6	4.0±3.6	8.5±2.5
inferior.48	4.2±3.9	79.5±8.0	8.2±6.6	8.2±5.9
inferior.fu	30.7±9.4	20.5±5.4	17.7±13.2	31.1±15.6
normal.50	20.8±9.2	16.2±3.7	31.3±8.8	31.7±9.7
CAD.50	18.7±7.8	12.2±3.2	28.5±11.5	40.7±13.3

C.26 Experiment st1.qp

Training Data

	anterior	inferior	normal	CAD
anterior	78.2±3.3	5.7±1.9	6.8±2.7	9.2±2.6
inferior	5.1±2.6	76.0±5.2	8.9±3.5	10.0±3.7
normal	12.0±3.2	7.4±3.6	59.2±8.1	21.4±7.1
CAD	10.2±1.7	10.6±3.1	20.6±6.8	58.7±7.0

Testing Data

	anterior	inferior	normal	CAD
anterior	69.3±4.9	7.0±2.3	10.9±4.7	12.8±4.0
inferior	10.3±2.5	69.6±3.1	7.8±2.3	12.3±3.3
normal	14.8±8.7	14.8±6.1	31.0±12.0	39.3±8.5
CAD	13.0±6.3	14.5±3.5	26.0±11.3	46.5±10.4

Training Data

	anterior	inferior	normal	CAD
anterior.6	73.5±4.3	3.5±3.2	9.0±4.9	14.0±4.9
anterior.12	83.5±3.4	3.2±1.9	6.4±2.5	6.9±2.4
anterior.48	91.9±2.8	2.2±1.7	3.2±2.6	2.7±2.3
anterior.fu	64.0±8.4	14.0±4.4	8.7±4.2	13.3±7.7
inferior.6	1.3±1.1	91.9±3.7	4.1±3.1	2.7±2.4
inferior.12	2.7±2.1	87.6±5.4	5.0±2.9	4.8±3.1
inferior.48	2.9±2.7	84.4±6.5	7.9±4.1	4.8±3.4
inferior.fu	13.5±6.4	40.1±8.6	18.8±10.2	27.6±10.3
normal.50	12.0±3.2	7.4±3.6	59.2±8.1	21.4±7.1
CAD.50	10.2±1.7	10.6±3.1	20.6±6.8	58.7±7.0

Testing Data

	anterior	inferior	normal	CAD
anterior.6	75.3±6.7	2.1±2.5	10.2±5.2	12.4±5.0
anterior.12	78.8±5.9	3.1±3.1	6.7±5.5	11.4±5.2
anterior.48	76.0±5.8	5.5±4.0	8.6±5.9	9.8±5.0
anterior.fu	47.0±11.8	17.2±6.8	18.2±10.4	17.5±9.4
inferior.6	3.3±2.4	87.8±5.1	4.7±3.6	4.2±3.1
inferior.12	5.0±3.6	85.0±4.3	3.2±3.2	6.8±3.6
inferior.48	3.5±2.7	83.8±5.6	5.0±4.0	7.7±6.0
inferior.fu	29.5±5.5	21.8±6.8	18.2±7.8	30.5±7.5
normal.50	14.8±8.7	14.8±6.1	31.0±12.0	39.3±8.5
CAD.50	13.0±6.3	14.5±3.5	26.0±11.3	46.5±10.4

C.27 Experiment st1.cas

Training Data

	anterior	inferior	normal	CAD
anterior	93.1±1.7	1.7±0.8	3.7±1.6	1.4±1.2
inferior	17.6±1.7	77.4±1.9	3.3±1.5	1.7±1.0
normal	52.7±11.1	6.7±1.7	31.7±12.4	8.9±5.6
CAD	51.0±10.1	7.9±1.1	20.7±8.5	20.5±9.3

Testing Data

	anterior	inferior	normal	CAD
anterior	91.9±2.2	3.0±1.0	3.3±1.1	1.7±1.4
inferior	22.9±1.6	70.0±2.2	3.9±1.6	3.1±1.5
normal	41.0±14.1	11.0±2.6	20.7±11.5	27.3±14.4
CAD	47.7±10.9	13.7±2.3	17.0±7.5	21.7±10.1

Training Data

	anterior	inferior	normal	CAD
anterior.6	94.6±2.7	1.5±0.7	2.9±1.8	1.0±1.0
anterior.12	93.6±1.5	1.9±1.1	3.2±1.2	1.3±0.6
anterior.48	94.8±1.5	0.9±1.5	3.6±1.5	0.7±1.0
anterior.fu	89.6±4.9	2.7±3.0	5.0±3.5	2.7±3.5
inferior.6	8.1±1.5	88.8±2.2	2.3±2.3	0.9±0.8
inferior.12	13.9±2.7	84.4±2.7	1.2±0.5	0.5±0.6
inferior.48	10.8±1.8	87.3±2.0	1.7±1.0	0.2±0.6
inferior.fu	37.7±5.3	49.0±5.8	7.9±4.0	5.3±3.3
normal.50	52.7±11.1	6.7±1.7	31.7±12.4	8.9±5.6
CAD.50	51.0±10.1	7.9±1.1	20.7±8.5	20.5±9.3

Testing Data

	anterior	inferior	normal	CAD
anterior.6	98.3±2.3	0.3±1.0	0.7±1.4	0.7±1.4
anterior.12	98.3±1.7	0.0±0.0	1.7±1.7	0.0±0.0
anterior.48	95.2±1.7	3.8±1.0	0.3±1.0	0.7±1.4
anterior.fu	76.0±6.6	8.0±4.0	10.5±4.2	5.5±5.2
inferior.6	15.3±2.2	81.3±2.7	3.3±2.1	0.0±0.0
inferior.12	15.3±1.6	81.7±2.2	0.0±0.0	3.0±1.8
inferior.48	15.0±2.2	84.3±2.1	0.7±1.3	0.0±0.0
inferior.fu	46.1±4.9	32.9±5.7	11.8±5.1	9.3±5.3
normal.50	41.0±14.1	11.0±2.6	20.7±11.5	27.3±14.4
CAD.50	47.7±10.9	13.7±2.3	17.0±7.5	21.7±10.1

C.28 Experiment st1.cbp

Training Data

	anterior	inferior	normal	CAD
anterior	85.2	3.0	4.3	7.5
inferior	6.5	72.3	8.7	12.6
normal	13.7	4.6	64.1	17.6
CAD	11.5	5.9	18.0	64.6

Testing Data

	anterior	inferior	normal	CAD
anterior	80.0	6.3	5.1	8.5
inferior	10.5	68.6	6.8	14.0
normal	20.0	16.7	36.7	26.7
CAD	13.3	10.0	23.3	53.3

Training Data

	anterior	inferior	normal	CAD
anterior.6	86.5	1.0	4.2	8.3
anterior.12	84.3	4.9	6.9	3.9
anterior.48	93.2	2.3	2.3	2.3
anterior.fu	76.9	3.8	3.8	15.4
inferior.6	1.1	94.6	3.2	1.1
inferior.12	3.7	87.8	6.1	2.4
inferior.48	1.9	80.8	7.7	9.6
inferior.fu	19.4	25.8	17.7	37.1
normal.50	13.7	4.6	64.1	17.6
CAD.50	11.5	5.9	18.0	64.6

Testing Data

	anterior	inferior	normal	CAD
anterior.6	82.8	3.4	3.4	10.3
anterior.12	89.7	0.0	3.4	6.9
anterior.48	82.8	6.9	3.4	6.9
anterior.fu	65.0	15.0	10.0	10.0
inferior.6	3.3	86.7	6.7	3.3
inferior.12	3.3	86.7	3.3	6.7
inferior.48	3.3	83.3	6.7	6.7
inferior.fu	32.1	17.9	10.7	39.3
normal.50	20.0	16.7	36.7	26.7
CAD.50	13.3	10.0	23.3	53.3

C.29 Experiment st1.cqp

Training Data

	anterior	inferior	normal	CAD
anterior	88.7	3.3	2.6	5.5
inferior	2.6	83.1	4.7	9.6
normal	11.5	3.8	69.5	15.3
CAD	8.5	8.2	11.1	72.1

Testing Data

	anterior	inferior	normal	CAD
anterior	85.5	4.2	3.8	6.4
inferior	14.0	72.0	2.6	11.4
normal	3.3	16.7	23.3	56.7
CAD	10.0	20.0	13.3	56.7

Training Data

	anterior	inferior	normal	CAD
anterior.6	86.5	2.1	3.1	8.3
anterior.12	92.2	1.0	1.0	5.9
anterior.48	95.5	2.3	2.3	0.0
anterior.fu	80.8	7.7	3.8	7.7
inferior.6	0.0	95.7	3.2	1.1
inferior.12	2.4	96.3	0.0	1.2
inferior.48	0.0	90.4	5.8	3.8
inferior.fu	8.1	50.0	9.7	32.3
normal.50	11.5	3.8	69.5	15.3
CAD.50	8.5	8.2	11.1	72.1

Testing Data

	anterior	inferior	normal	CAD
anterior.6	82.8	3.4	6.9	6.9
anterior.12	93.1	0.0	0.0	6.9
anterior.48	86.2	3.4	3.4	6.9
anterior.fu	80.0	10.0	5.0	5.0
inferior.6	3.3	93.3	3.3	0.0
inferior.12	6.7	86.7	3.3	3.3
inferior.48	3.3	90.0	0.0	6.7
inferior.fu	42.9	17.9	3.6	35.7
normal.50	3.3	16.7	23.3	56.7
CAD.50	10.0	20.0	13.3	56.7

C.30 Experiment st1.ccas

Training Data

	anterior	inferior	normal	CAD
anterior	93.9	0.8	3.9	1.5
inferior	17.3	77.8	2.9	2.1
normal	58.8	3.8	29.0	8.4
CAD	54.1	7.5	12.1	26.2

Testing Data

	anterior	inferior	normal	CAD
anterior	93.3	2.1	3.4	1.2
inferior	24.2	70.6	2.6	2.6
normal	36.7	10.0	6.7	46.7
CAD	56.7	10.0	6.7	26.7

Training Data

	anterior	inferior	normal	CAD
anterior.6	93.7	1.0	4.2	1.0
anterior.12	94.1	2.0	2.9	1.0
anterior.48	95.5	0.0	4.5	0.0
anterior.fu	92.3	0.0	3.8	3.8
inferior.6	8.6	90.3	0.0	1.1
inferior.12	13.4	84.1	0.0	2.4
inferior.48	11.5	86.5	1.9	0.0
inferior.fu	35.5	50.0	9.7	4.8
normal.50	58.8	3.8	29.0	8.4
CAD.50	54.1	7.5	12.1	26.2

Testing Data

	anterior	inferior	normal	CAD
anterior.6	100.0	0.0	0.0	0.0
anterior.12	96.6	0.0	3.4	0.0
anterior.48	96.6	3.4	0.0	0.0
anterior.fu	80.0	5.0	10.0	5.0
inferior.6	16.7	80.0	3.3	0.0
inferior.12	13.3	83.3	0.0	3.3
inferior.48	16.7	83.3	0.0	0.0
inferior.fu	50.0	35.7	7.1	7.1
normal.50	36.7	10.0	6.7	46.7
CAD.50	56.7	10.0	6.7	26.7

C.31 Experiment qrs1.knn

Training Data

	anterior	inferior	normal	CAD
anterior	96.8	0.8	0.6	1.9
inferior	0.0	98.5	0.9	0.6
normal	0.8	0.0	95.4	3.8
CAD	0.7	1.3	0.0	98.0

Testing Data

	anterior	inferior	normal	CAD
anterior	74.4	11.0	5.2	9.4
inferior	8.4	70.2	4.3	17.1
normal	3.3	3.3	36.7	56.7
CAD	0.0	10.0	13.3	76.7

Training Data

	anterior	inferior	normal	CAD
anterior.6	99.0	1.0	0.0	0.0
anterior.12	98.0	2.0	0.0	0.0
anterior.48	97.7	0.0	2.3	0.0
anterior.fu	92.3	0.0	0.0	7.7
inferior.6	0.0	98.9	0.0	1.1
inferior.12	0.0	98.8	0.0	1.2
inferior.48	0.0	98.1	1.9	0.0
inferior.fu	0.0	98.4	1.6	0.0
normal.50	0.8	0.0	95.4	3.8
CAD.50	0.7	1.3	0.0	98.0

Testing Data

	anterior	inferior	normal	CAD
anterior.6	69.0	6.9	13.8	10.3
anterior.12	82.8	6.9	3.4	6.9
anterior.48	75.9	10.3	3.4	10.3
anterior.fu	70.0	20.0	0.0	10.0
inferior.6	6.7	66.7	6.7	20.0
inferior.12	13.3	86.7	0.0	0.0
inferior.48	10.0	66.7	3.3	20.0
inferior.fu	3.6	60.7	7.1	28.6
normal.50	3.3	3.3	36.7	56.7
CAD.50	0.0	10.0	13.3	76.7

C.32 Experiment qrs1.linreg

Training Data

	anterior	inferior	normal	CAD
anterior	88.2	6.8	2.1	2.9
inferior	4.1	89.7	2.8	3.3
normal	3.8	4.6	66.4	25.2
CAD	7.2	12.8	20.3	59.7

Testing Data

	anterior	inferior	normal	CAD
anterior	72.3	17.5	5.2	5.1
inferior	6.7	82.9	6.1	4.3
normal	3.3	3.3	50.0	43.3
CAD	0.0	6.7	30.0	63.3

Training Data

	anterior	inferior	normal	CAD
anterior.6	81.2	8.3	3.1	7.3
anterior.12	85.3	9.8	2.9	2.0
anterior.48	86.4	9.1	2.3	2.3
anterior.fu	100.0	0.0	0.0	0.0
inferior.6	2.2	91.4	3.2	3.2
inferior.12	2.4	92.7	1.2	3.7
inferior.48	3.8	94.2	1.9	0.0
inferior.fu	8.1	80.6	4.8	6.5
normal.50	3.8	4.6	66.4	25.2
CAD.50	7.2	12.8	20.3	59.7

Testing Data

	anterior	inferior	normal	CAD
anterior.6	72.4	20.7	6.9	0.0
anterior.12	79.3	13.8	6.9	0.0
anterior.48	72.4	10.3	6.9	10.3
anterior.fu	65.0	25.0	0.0	10.0
inferior.6	6.7	86.7	6.7	0.0
inferior.12	16.7	83.3	0.0	0.0
inferior.48	0.0	90.0	0.0	10.0
inferior.fu	3.6	71.4	17.9	7.1
normal.50	3.3	3.3	50.0	43.3
CAD.50	0.0	6.7	30.0	63.3

C.33 Experiment qrs1.C4.5

Training Data

	anterior	inferior	normal	CAD
anterior	98.5	0.0	0.5	1.0
inferior	0.3	97.7	0.5	1.5
normal	0.0	0.0	100.0	0.0
CAD	0.7	3.0	0.3	96.1

Testing Data

	anterior	inferior	normal	CAD
anterior	74.0	7.3	4.7	14.0
inferior	7.6	61.0	11.0	20.5
normal	13.3	10.0	30.0	46.7
CAD	20.0	16.7	23.3	40.0

Training Data

	anterior	inferior	normal	CAD
anterior.6	96.9	0.0	2.1	1.0
anterior.12	97.1	0.0	0.0	2.9
anterior.48	100.0	0.0	0.0	0.0
anterior.fu	100.0	0.0	0.0	0.0
inferior.6	1.1	93.5	2.2	3.2
inferior.12	0.0	98.8	0.0	1.2
inferior.48	0.0	100.0	0.0	0.0
inferior.fu	0.0	98.4	0.0	1.6
normal.50	0.0	0.0	100.0	0.0
CAD.50	0.7	3.0	0.3	96.1

Testing Data

	anterior	inferior	normal	CAD
anterior.6	75.9	10.3	6.9	6.9
anterior.12	79.3	3.4	3.4	13.8
anterior.48	75.9	10.3	3.4	10.3
anterior.fu	65.0	5.0	5.0	25.0
inferior.6	10.0	56.7	23.3	10.0
inferior.12	3.3	63.3	3.3	30.0
inferior.48	10.0	66.7	10.0	13.3
inferior.fu	7.1	57.1	7.1	28.6
normal.50	13.3	10.0	30.0	46.7
CAD.50	20.0	16.7	23.3	40.0

C.34 Experiment qrs1.MML

Training Data

	anterior	inferior	normal	CAD
anterior	96.0	1.3	0.8	2.0
inferior	3.9	92.5	1.3	2.3
normal	0.8	0.8	98.5	0.0
CAD	5.6	15.7	2.3	76.4

Testing Data

	anterior	inferior	normal	CAD
anterior	74.8	4.7	3.8	16.7
inferior	8.6	63.5	10.1	17.9
normal	13.3	10.0	36.7	40.0
CAD	10.0	20.0	33.3	36.7

Training Data

	anterior	inferior	normal	CAD
anterior.6	91.7	3.1	2.1	3.1
anterior.12	92.2	2.0	1.0	4.9
anterior.48	100.0	0.0	0.0	0.0
anterior.fu	100.0	0.0	0.0	0.0
inferior.6	6.5	86.0	1.1	6.5
inferior.12	3.7	92.7	2.4	1.2
inferior.48	3.8	96.2	0.0	0.0
inferior.fu	1.6	95.2	1.6	1.6
normal.50	0.8	0.8	98.5	0.0
CAD.50	5.6	15.7	2.3	76.4

Testing Data

	anterior	inferior	normal	CAD
anterior.6	62.1	6.9	6.9	24.1
anterior.12	79.3	6.9	0.0	13.8
anterior.48	82.8	0.0	3.4	13.8
anterior.fu	75.0	5.0	5.0	15.0
inferior.6	13.3	56.7	13.3	16.7
inferior.12	3.3	66.7	10.0	20.0
inferior.48	3.3	73.3	10.0	13.3
inferior.fu	14.3	57.1	7.1	21.4
normal.50	13.3	10.0	36.7	40.0
CAD.50	10.0	20.0	33.3	36.7

C.35 Experiment qrs1.bp

Training Data

	anterior	inferior	normal	CAD
anterior	86.1±2.8	6.7±3.2	2.1±2.6	5.1±4.5
inferior	5.5±2.2	85.0±5.1	2.7±3.9	6.7±5.9
normal	21.8±20.1	21.8±18.6	29.7±36.5	26.7±29.9
CAD	18.4±14.8	30.0±20.8	10.3±16.0	41.3±34.2

Testing Data

	anterior	inferior	normal	CAD
anterior	70.6±7.6	14.5±5.1	4.8±6.2	10.1±8.7
inferior	14.3±4.2	67.0±11.4	5.1±7.3	13.6±12.0
normal	32.5±22.2	19.0±13.0	15.3±20.9	33.2±29.0
CAD	15.8±13.9	31.2±22.6	14.2±20.8	38.8±32.9

Training Data

	anterior	inferior	normal	CAD
anterior.6	79.9±4.5	10.6±4.1	2.8±3.7	6.6±6.1
anterior.12	85.5±3.2	7.1±4.2	2.6±3.5	4.8±4.5
anterior.48	90.1±3.2	4.9±3.2	1.1±2.2	3.9±3.7
anterior.fu	88.8±5.4	4.2±4.4	1.7±3.1	5.2±5.6
inferior.6	5.7±2.4	82.4±7.4	3.6±5.0	8.3±7.8
inferior.12	2.9±1.9	85.9±7.1	3.3±4.8	7.9±7.3
inferior.48	4.9±2.1	91.0±3.3	1.3±2.2	2.8±3.1
inferior.fu	8.5±5.0	80.9±6.4	2.7±4.5	7.8±7.4
normal.50	21.8±20.1	21.8±18.6	29.7±36.5	26.7±29.9
CAD.50	18.4±14.8	30.0±20.8	10.3±16.0	41.3±34.2

Testing Data

	anterior	inferior	normal	CAD
anterior.6	69.7±10.9	15.7±7.4	6.0±8.4	8.6±8.8
anterior.12	71.6±8.3	11.6±6.1	4.1±5.8	12.8±11.2
anterior.48	73.6±7.3	14.7±4.9	3.4±5.6	8.3±7.7
anterior.fu	67.5±10.7	16.2±8.5	5.5±7.6	10.8±10.3
inferior.6	15.8±7.9	58.8±17.0	7.8±11.5	17.5±16.4
inferior.12	14.2±3.9	72.3±11.9	2.3±4.0	11.2±10.3
inferior.48	12.8±6.2	69.3±11.2	5.7±9.1	12.2±11.8
inferior.fu	14.3±7.1	67.5±12.6	4.5±6.7	13.7±12.6
normal.50	32.5±22.2	19.0±13.0	15.3±20.9	33.2±29.0
CAD.50	15.8±13.9	31.2±22.6	14.2±20.8	38.8±32.9

C.36 Experiment qrs1.qp

Training Data

	anterior	inferior	normal	CAD
anterior	67.2±5.5	14.0±5.9	6.7±7.3	12.2±5.7
inferior	13.6±9.5	70.4±9.7	4.4±4.0	11.6±5.2
normal	19.5±17.0	20.3±9.9	29.4±23.2	30.8±14.8
CAD	14.7±12.5	26.1±12.5	12.7±11.2	46.5±20.4

Testing Data

	anterior	inferior	normal	CAD
anterior	55.6±7.2	18.7±8.6	10.1±10.6	15.6±7.5
inferior	16.6±8.0	61.4±9.0	5.4±5.6	16.7±7.6
normal	29.7±21.3	19.0±12.7	22.8±21.2	28.5±18.3
CAD	8.5±13.4	21.5±11.7	15.8±14.8	54.2±20.9

Training Data

	anterior	inferior	normal	CAD
anterior.6	57.6±6.7	17.2±6.3	8.0±9.4	17.3±8.6
anterior.12	76.0±4.0	10.3±4.3	5.1±4.9	8.5±4.7
anterior.48	80.8±4.6	9.0±5.3	4.5±5.7	5.7±3.6
anterior.fu	54.2±10.3	19.4±10.8	9.2±11.4	17.1±9.3
inferior.6	14.9±10.1	68.2±10.3	4.8±4.3	12.1±6.8
inferior.12	9.8±8.0	73.0±9.7	4.0±4.5	13.2±6.4
inferior.48	11.9±8.8	77.2±10.8	3.1±2.8	7.8±4.1
inferior.fu	17.8±11.7	63.1±11.4	5.6±6.0	13.4±7.1
normal.50	19.5±17.0	20.3±9.9	29.4±23.2	30.8±14.8
CAD.50	14.7±12.5	26.1±12.5	12.7±11.2	46.5±20.4

Testing Data

	anterior	inferior	normal	CAD
anterior.6	63.6±6.0	13.8±7.9	9.5±11.1	13.1±8.3
anterior.12	59.5±8.9	12.9±7.0	8.3±8.8	19.3±9.1
anterior.48	69.1±9.3	10.2±5.2	9.8±9.1	10.9±5.9
anterior.fu	30.2±9.5	37.8±19.5	12.8±16.3	19.2±12.9
inferior.6	16.2±8.3	56.0±13.6	6.5±7.3	21.3±11.5
inferior.12	13.2±5.3	71.7±9.3	3.0±5.5	12.2±7.0
inferior.48	16.5±9.8	61.7±9.5	5.5±4.7	16.3±8.6
inferior.fu	20.5±11.8	56.1±12.0	6.6±7.3	16.8±8.7
normal.50	29.7±21.3	19.0±12.7	22.8±21.2	28.5±18.3
CAD.50	8.5±13.4	21.5±11.7	15.8±14.8	54.2±20.9

C.37 Experiment qrs1.cas

Training Data

	anterior	inferior	normal	CAD
anterior	96.5±0.8	1.5±0.5	1.0±0.6	1.0±0.5
inferior	18.9±3.3	77.5±3.8	2.4±1.0	1.1±0.7
normal	39.3±21.2	0.6±0.6	54.0±18.6	6.1±4.7
CAD	45.9±16.2	7.2±2.2	27.2±11.0	19.7±7.6

Testing Data

	anterior	inferior	normal	CAD
anterior	86.6±2.3	7.2±0.6	4.4±1.7	1.8±0.9
inferior	36.0±3.0	59.0±1.6	2.5±1.1	2.5±1.4
normal	40.3±12.8	4.7±2.2	35.3±12.8	19.7±2.8
CAD	29.3±17.1	12.7±1.3	38.0±13.5	20.0±7.1

Training Data

	anterior	inferior	normal	CAD
anterior.6	91.6±1.4	4.5±1.3	1.6±0.7	2.4±1.0
anterior.12	97.8±1.1	1.5±1.2	0.7±0.4	0.0±0.0
anterior.48	96.8±2.9	0.0±0.0	1.8±2.0	1.4±1.5
anterior.fu	99.6±1.2	0.0±0.0	0.0±0.0	0.4±1.2
inferior.6	16.5±5.0	78.5±5.0	3.2±1.3	1.8±1.4
inferior.12	18.2±3.8	79.5±4.1	1.6±1.0	0.7±0.6
inferior.48	11.2±2.2	87.5±2.2	1.3±1.5	0.0±0.0
inferior.fu	30.0±5.3	64.5±5.3	3.5±1.7	1.9±1.4
normal.50	39.3±21.2	0.6±0.6	54.0±18.6	6.1±4.7
CAD.50	45.9±16.2	7.2±2.2	27.2±11.0	19.7±7.6

Testing Data

	anterior	inferior	normal	CAD
anterior.6	89.3±2.4	6.9±0.0	3.4±1.5	0.3±1.0
anterior.12	95.5±2.2	3.4±2.2	0.7±1.4	0.3±1.0
anterior.48	91.0±3.2	2.8±1.4	4.1±2.1	2.1±1.7
anterior.fu	70.5±4.7	15.5±1.5	9.5±4.2	4.5±2.7
inferior.6	42.0±5.2	53.3±2.6	2.3±2.6	2.3±3.3
inferior.12	42.3±3.3	55.3±2.7	0.7±1.3	1.7±1.7
inferior.48	15.7±2.6	79.7±1.0	0.0±0.0	4.7±2.2
inferior.fu	43.9±4.8	47.5±3.6	7.1±3.2	1.4±2.4
normal.50	40.3±12.8	4.7±2.2	35.3±12.8	19.7±2.8
CAD.50	29.3±17.1	12.7±1.3	38.0±13.5	20.0±7.1

C.38 Experiment qrs1.cbp

Training Data

	anterior	inferior	normal	CAD
anterior	92.5	3.9	0.6	3.1
inferior	2.6	93.8	0.3	3.3
normal	3.8	7.6	61.1	27.5
CAD	6.2	14.4	0.7	78.7

Testing Data

	anterior	inferior	normal	CAD
anterior	77.8	11.1	1.7	9.4
inferior	10.2	78.0	0.8	11.0
normal	20.0	13.3	13.3	53.3
CAD	6.7	16.7	6.7	70.0

Training Data

	anterior	inferior	normal	CAD
anterior.6	88.5	7.3	0.0	4.2
anterior.12	92.2	5.9	0.0	2.0
anterior.48	93.2	2.3	2.3	2.3
anterior.fu	96.2	0.0	0.0	3.8
inferior.6	2.2	94.6	1.1	2.2
inferior.12	1.2	92.7	0.0	6.1
inferior.48	3.8	96.2	0.0	0.0
inferior.fu	3.2	91.9	0.0	4.8
normal.50	3.8	7.6	61.1	27.5
CAD.50	6.2	14.4	0.7	78.7

Testing Data

	anterior	inferior	normal	CAD
anterior.6	75.9	10.3	6.9	6.9
anterior.12	79.3	6.9	0.0	13.8
anterior.48	75.9	17.2	0.0	6.9
anterior.fu	80.0	10.0	0.0	10.0
inferior.6	6.7	76.7	0.0	16.7
inferior.12	13.3	76.7	0.0	10.0
inferior.48	10.0	80.0	3.3	6.7
inferior.fu	10.7	78.6	0.0	10.7
normal.50	20.0	13.3	13.3	53.3
CAD.50	6.7	16.7	6.7	70.0

C.39 Experiment qrs1.cqp

Training Data

	anterior	inferior	normal	CAD
anterior	68.4	10.1	6.6	14.9
inferior	6.7	84.4	1.9	7.1
normal	5.3	12.2	45.8	36.6
CAD	6.6	16.1	7.5	69.8

Testing Data

	anterior	inferior	normal	CAD
anterior	54.9	20.3	9.4	15.3
inferior	12.7	72.1	1.7	13.5
normal	13.3	10.0	40.0	36.7
CAD	0.0	13.3	16.7	70.0

Training Data

	anterior	inferior	normal	CAD
anterior.6	59.4	16.7	4.2	19.8
anterior.12	78.4	7.8	3.9	9.8
anterior.48	81.8	4.5	6.8	6.8
anterior.fu	53.8	11.5	11.5	23.1
inferior.6	7.5	84.9	1.1	6.5
inferior.12	3.7	86.6	1.2	8.5
inferior.48	5.8	88.5	1.9	3.8
inferior.fu	9.7	77.4	3.2	9.7
normal.50	5.3	12.2	45.8	36.6
CAD.50	6.6	16.1	7.5	69.8

Testing Data

	anterior	inferior	normal	CAD
anterior.6	65.5	13.8	10.3	10.3
anterior.12	58.6	13.8	3.4	24.1
anterior.48	65.5	13.8	13.8	6.9
anterior.fu	30.0	40.0	10.0	20.0
inferior.6	10.0	63.3	0.0	26.7
inferior.12	16.7	76.7	3.3	3.3
inferior.48	10.0	73.3	3.3	13.3
inferior.fu	14.3	75.0	0.0	10.7
normal.50	13.3	10.0	40.0	36.7
CAD.50	0.0	13.3	16.7	70.0

C.40 Experiment qrs1.ccas

Training Data

	anterior	inferior	normal	CAD
anterior	96.3	1.3	1.6	0.8
inferior	15.6	80.3	3.5	0.7
normal	38.9	0.0	56.5	4.6
CAD	49.2	5.6	22.3	23.0

Testing Data

	anterior	inferior	normal	CAD
anterior	84.4	8.1	5.5	2.1
inferior	34.0	62.5	1.8	1.7
normal	36.7	3.3	23.3	36.7
CAD	20.0	13.3	26.7	40.0

Training Data

	anterior	inferior	normal	CAD
anterior.6	90.6	5.2	1.0	3.1
anterior.12	99.0	0.0	1.0	0.0
anterior.48	95.5	0.0	4.5	0.0
anterior.fu	100.0	0.0	0.0	0.0
inferior.6	15.1	79.6	4.3	1.1
inferior.12	15.9	81.7	2.4	0.0
inferior.48	5.8	90.4	3.8	0.0
inferior.fu	25.8	69.4	3.2	1.6
normal.50	38.9	0.0	56.5	4.6
CAD.50	49.2	5.6	22.3	23.0

Testing Data

	anterior	inferior	normal	CAD
anterior.6	89.7	6.9	3.4	0.0
anterior.12	93.1	6.9	0.0	0.0
anterior.48	89.7	3.4	3.4	3.4
anterior.fu	65.0	15.0	15.0	5.0
inferior.6	43.3	56.7	0.0	0.0
inferior.12	36.7	63.3	0.0	0.0
inferior.48	16.7	80.0	0.0	3.3
inferior.fu	39.3	50.0	7.1	3.6
normal.50	36.7	3.3	23.3	36.7
CAD.50	20.0	13.3	26.7	40.0

C.41 Experiment E2.knn

Training Data

	anterior	inferior	normal
anterior	96.0	0.2	3.7
inferior	0.0	97.5	2.5
normal	0.0	0.0	100.0

Testing Data

	anterior	inferior	normal
anterior	70.6	6.3	23.0
inferior	6.7	59.0	34.2
normal	3.3	3.3	93.3

Training Data

	anterior	inferior	normal
anterior.6	99.0	0.0	1.0
anterior.12	99.0	1.0	0.0
anterior.48	97.7	0.0	2.3
anterior.fu	88.5	0.0	11.5
inferior.6	0.0	100.0	0.0
inferior.12	0.0	98.8	1.2
inferior.48	0.0	96.2	3.8
inferior.fu	0.0	95.2	4.8
normal.50	0.0	0.0	100.0
CAD.50	0.0	0.3	99.7

Testing Data

	anterior	inferior	normal
anterior.6	65.5	3.4	31.0
anterior.12	86.2	3.4	10.3
anterior.48	75.9	3.4	20.7
anterior.fu	55.0	15.0	30.0
inferior.6	3.3	63.3	33.3
inferior.12	10.0	70.0	20.0
inferior.48	10.0	60.0	30.0
inferior.fu	3.6	42.9	53.6
normal.50	3.3	3.3	93.3
CAD.50	0.0	3.3	96.7

C.42 Experiment E2.linreg

Training Data

	anterior	inferior	normal
anterior	87.6	3.1	9.3
inferior	1.8	91.5	6.7
normal	2.3	7.6	90.1

Testing Data

	anterior	inferior	normal
anterior	81.7	7.2	11.1
inferior	6.0	80.3	13.7
normal	0.0	10.0	90.0

Training Data

	anterior	inferior	normal
anterior.6	82.3	6.2	11.5
anterior.12	94.1	3.9	2.0
anterior.48	93.2	2.3	4.5
anterior.fu	80.8	0.0	19.2
inferior.6	2.2	94.6	3.2
inferior.12	0.0	92.7	7.3
inferior.48	1.9	98.1	0.0
inferior.fu	3.2	80.6	16.1
normal.50	2.3	7.6	90.1
CAD.50	10.2	19.0	70.8

Testing Data

	anterior	inferior	normal
anterior.6	82.8	6.9	10.3
anterior.12	89.7	3.4	6.9
anterior.48	79.3	3.4	17.2
anterior.fu	75.0	15.0	10.0
inferior.6	6.7	83.3	10.0
inferior.12	10.0	76.7	13.3
inferior.48	0.0	93.3	6.7
inferior.fu	7.1	67.9	25.0
normal.50	0.0	10.0	90.0
CAD.50	0.0	23.3	76.7

C.43 Experiment E2.C4.5

Training Data

	anterior	inferior	normal
anterior	97.7	1.3	1.0
inferior	1.9	97.5	0.7
normal	0.0	0.0	100.0

Testing Data

	anterior	inferior	normal
anterior	72.4	12.3	15.3
inferior	8.6	70.1	21.3
normal	23.3	13.3	63.3

Training Data

	anterior	inferior	normal
anterior.6	94.8	2.1	3.1
anterior.12	96.1	2.9	1.0
anterior.48	100.0	0.0	0.0
anterior.fu	100.0	0.0	0.0
inferior.6	2.2	96.8	1.1
inferior.12	3.7	96.3	0.0
inferior.48	0.0	100.0	0.0
inferior.fu	1.6	96.8	1.6
normal.50	0.0	0.0	100.0
CAD.50	27.2	23.6	49.2

Testing Data

	anterior	inferior	normal
anterior.6	79.3	3.4	17.2
anterior.12	75.9	10.3	13.8
anterior.48	79.3	10.3	10.3
anterior.fu	55.0	25.0	20.0
inferior.6	3.3	76.7	20.0
inferior.12	6.7	70.0	23.3
inferior.48	10.0	76.7	13.3
inferior.fu	14.3	57.1	28.6
normal.50	23.3	13.3	63.3
CAD.50	13.3	30.0	56.7

C.44 Experiment E2.MML

Training Data

	anterior	inferior	normal
anterior	88.8	7.8	3.5
inferior	4.6	93.9	1.5
normal	14.5	22.9	62.6

Testing Data

	anterior	inferior	normal
anterior	77.4	17.5	5.1
inferior	13.6	71.8	14.5
normal	33.3	33.3	33.3

Training Data

	anterior	inferior	normal
anterior.6	87.5	9.4	3.1
anterior.12	88.2	7.8	3.9
anterior.48	90.9	2.3	6.8
anterior.fu	88.5	11.5	0.0
inferior.6	4.3	93.5	2.2
inferior.12	3.7	92.7	3.7
inferior.48	5.8	94.2	0.0
inferior.fu	4.8	95.2	0.0
normal.50	14.5	22.9	62.6
CAD.50	23.9	40.7	35.4

Testing Data

	anterior	inferior	normal
anterior.6	65.5	24.1	10.3
anterior.12	86.2	13.8	0.0
anterior.48	82.8	17.2	0.0
anterior.fu	75.0	15.0	10.0
inferior.6	10.0	73.3	16.7
inferior.12	16.7	73.3	10.0
inferior.48	10.0	80.0	10.0
inferior.fu	17.9	60.7	21.4
normal.50	33.3	33.3	33.3
CAD.50	13.3	46.7	40.0

C.45 Experiment E2.bp

Training Data

	anterior	inferior	normal
anterior	85.2±3.6	7.1±4.5	7.6±3.0
inferior	3.8±1.9	88.9±1.7	7.3±2.4
normal	11.6±9.5	10.6±9.3	77.7±18.3

Testing Data

	anterior	inferior	normal
anterior	71.6±3.7	14.7±5.1	13.7±5.1
inferior	9.8±2.3	77.6±4.2	12.6±4.1
normal	26.3±8.2	14.0±10.3	59.7±15.5

Training Data

	anterior	inferior	normal
anterior.6	74.6±7.8	13.8±9.3	11.6±4.7
anterior.12	87.9±3.8	5.6±1.8	6.5±3.9
anterior.48	92.2±3.4	4.3±2.9	3.5±2.7
anterior.fu	86.3±5.5	4.8±6.3	8.8±4.2
inferior.6	5.6±4.2	86.5±4.0	7.8±3.4
inferior.12	2.0±1.9	92.0±1.9	6.0±2.8
inferior.48	2.0±1.7	93.7±2.2	4.3±2.3
inferior.fu	5.6±2.4	83.4±5.0	11.0±4.7
normal.50	11.6±9.5	10.6±9.3	77.7±18.3
CAD.50	16.0±3.2	34.6±9.6	49.4±11.9

Testing Data

	anterior	inferior	normal
anterior.6	69.3±6.4	15.5±6.0	15.2±6.1
anterior.12	79.8±8.2	6.0±7.3	14.1±7.3
anterior.48	80.9±5.3	7.9±3.6	11.2±5.5
anterior.fu	56.5±8.1	29.2±8.8	14.2±9.0
inferior.6	13.0±7.0	71.3±7.8	15.7±6.7
inferior.12	13.7±2.8	80.0±6.4	6.3±4.8
inferior.48	7.2±3.4	84.3±4.7	8.5±5.3
inferior.fu	5.4±2.6	74.6±7.2	20.0±6.6
normal.50	26.3±8.2	14.0±10.3	59.7±15.5
CAD.50	8.5±5.5	34.3±12.3	57.2±14.5

C.46 Experiment E2.qp

Training Data

	anterior	inferior	normal
anterior	76.0±4.0	9.7±1.9	14.3±3.1
inferior	3.0±1.4	86.9±2.1	10.1±1.7
normal	8.6±2.7	12.5±2.2	78.9±3.6

Testing Data

	anterior	inferior	normal
anterior	63.9±4.5	15.8±3.4	20.3±3.9
inferior	7.4±2.1	79.4±3.3	13.1±3.0
normal	17.8±7.0	11.0±5.5	71.2±7.8

Training Data

	anterior	inferior	normal
anterior.6	56.4±8.5	18.6±5.4	24.9±7.4
anterior.12	83.7±3.7	7.2±1.6	9.1±3.5
anterior.48	86.4±3.4	8.0±2.3	5.7±3.0
anterior.fu	77.3±5.1	5.2±3.0	17.5±5.2
inferior.6	4.1±2.3	85.9±3.3	9.9±3.0
inferior.12	1.5±1.5	90.4±3.4	8.1±2.6
inferior.48	1.6±1.1	92.5±3.0	5.9±2.9
inferior.fu	4.8±3.6	78.8±3.4	16.5±4.1
normal.50	8.6±2.7	12.5±2.2	78.9±3.6
CAD.50	9.1±2.1	33.6±2.9	57.2±3.6

Testing Data

	anterior	inferior	normal
anterior.6	61.4±7.4	16.9±7.1	21.7±5.6
anterior.12	70.0±6.1	7.8±5.0	22.2±5.5
anterior.48	74.8±5.4	9.0±3.5	16.2±4.4
anterior.fu	49.2±10.0	29.8±7.8	21.0±9.4
inferior.6	11.2±3.7	72.3±6.4	16.5±5.6
inferior.12	10.8±3.3	82.7±4.2	6.5±4.0
inferior.48	5.8±3.1	85.7±4.1	8.5±4.4
inferior.fu	2.0±2.4	77.0±6.3	21.1±6.3
normal.50	17.8±7.0	11.0±5.5	71.2±7.8
CAD.50	3.7±3.0	32.8±6.1	63.5±6.5

C.47 Experiment E2.cas

Training Data

	anterior	inferior	normal
anterior	97.7±0.9	0.5±0.3	1.8±1.0
inferior	6.9±1.8	90.8±2.1	2.3±0.7
normal	16.0±2.1	2.6±1.6	81.4±1.6

Testing Data

	anterior	inferior	normal
anterior	89.1±1.2	4.7±0.7	6.2±0.9
inferior	11.1±1.2	82.8±0.9	6.1±0.5
normal	23.7±3.1	12.0±2.7	64.3±4.0

Training Data

	anterior	inferior	normal
anterior.6	96.7±0.4	1.0±0.0	2.3±0.4
anterior.12	99.4±0.5	0.5±0.5	0.1±0.3
anterior.48	98.0±1.2	0.5±0.9	1.6±1.0
anterior.fu	96.9±2.9	0.0±0.0	3.1±2.9
inferior.6	5.8±1.6	93.8±1.6	0.4±0.5
inferior.12	2.7±1.3	95.9±2.1	1.5±1.2
inferior.48	3.5±1.9	96.5±1.9	0.0±0.0
inferior.fu	15.6±4.0	76.9±3.7	7.4±2.3
normal.50	16.0±2.1	2.6±1.6	81.4±1.6
CAD.50	28.2±1.3	23.2±1.4	48.6±1.3

Testing Data

	anterior	inferior	normal
anterior.6	89.0±1.4	3.4±0.0	7.6±1.4
anterior.12	89.7±2.2	6.2±2.1	4.1±2.1
anterior.48	89.3±1.0	3.4±0.0	7.2±1.0
anterior.fu	88.5±3.2	5.5±1.5	6.0±2.0
inferior.6	14.0±2.0	82.3±1.5	3.7±1.0
inferior.12	10.7±1.3	83.0±1.8	6.3±1.0
inferior.48	1.7±1.7	95.0±1.7	3.3±0.0
inferior.fu	18.2±3.4	70.7±2.7	11.1±1.9
normal.50	23.7±3.1	12.0±2.7	64.3±4.0
CAD.50	16.0±2.0	14.3±2.1	69.7±1.8

C.48 Experiment E2.cbp

Training Data

	anterior	inferior	normal
anterior	92.1	3.6	4.2
inferior	1.8	94.5	3.7
normal	9.2	6.1	84.7

Testing Data

	anterior	inferior	normal
anterior	74.5	12.7	12.8
inferior	9.2	84.8	6.0
normal	26.7	6.7	66.7

Training Data

	anterior	inferior	normal
anterior.6	84.4	8.3	7.3
anterior.12	94.1	3.9	2.0
anterior.48	97.7	2.3	0.0
anterior.fu	92.3	0.0	7.7
inferior.6	2.2	95.7	2.2
inferior.12	0.0	97.6	2.4
inferior.48	1.9	94.2	3.8
inferior.fu	3.2	90.3	6.5
normal.50	9.2	6.1	84.7
CAD.50	10.2	29.5	60.3

Testing Data

	anterior	inferior	normal
anterior.6	72.4	13.8	13.8
anterior.12	82.8	0.0	17.2
anterior.48	82.8	6.9	10.3
anterior.fu	60.0	30.0	10.0
inferior.6	10.0	80.0	10.0
inferior.12	13.3	86.7	0.0
inferior.48	10.0	86.7	3.3
inferior.fu	3.6	85.7	10.7
normal.50	26.7	6.7	66.7
CAD.50	3.3	36.7	60.0

C.49 Experiment E2.cqp

Training Data

	anterior	inferior	normal
anterior	81.8	9.5	8.7
inferior	1.3	90.2	8.5
normal	6.1	9.2	84.7

Testing Data

	anterior	inferior	normal
anterior	70.2	14.0	15.8
inferior	10.1	83.0	7.0
normal	13.3	6.7	80.0

Training Data

	anterior	inferior	normal
anterior.6	61.5	17.7	20.8
anterior.12	90.2	5.9	3.9
anterior.48	90.9	6.8	2.3
anterior.fu	84.6	7.7	7.7
inferior.6	3.2	89.2	7.5
inferior.12	0.0	95.1	4.9
inferior.48	1.9	94.2	3.8
inferior.fu	0.0	82.3	17.7
normal.50	6.1	9.2	84.7
CAD.50	7.9	29.2	63.0

Testing Data

	anterior	inferior	normal
anterior.6	72.4	13.8	13.8
anterior.12	72.4	6.9	20.7
anterior.48	75.9	10.3	13.8
anterior.fu	60.0	25.0	15.0
inferior.6	13.3	76.7	10.0
inferior.12	13.3	86.7	0.0
inferior.48	10.0	90.0	0.0
inferior.fu	3.6	78.6	17.9
normal.50	13.3	6.7	80.0
CAD.50	0.0	33.3	66.7

C.50 Experiment E2.ccas

Training Data

	anterior	inferior	normal
anterior	96.4	1.1	2.6
inferior	5.7	92.1	2.2
normal	13.7	0.8	85.5

Testing Data

	anterior	inferior	normal
anterior	88.9	4.7	6.4
inferior	11.1	82.0	6.9
normal	23.3	13.3	63.3

Training Data

	anterior	inferior	normal
anterior.6	94.8	1.0	4.2
anterior.12	99.0	1.0	0.0
anterior.48	95.5	2.3	2.3
anterior.fu	96.2	0.0	3.8
inferior.6	5.4	94.6	0.0
inferior.12	2.4	95.1	2.4
inferior.48	1.9	98.1	0.0
inferior.fu	12.9	80.6	6.5
normal.50	13.7	0.8	85.5
CAD.50	27.9	20.0	52.1

Testing Data

	anterior	inferior	normal
anterior.6	89.7	3.4	6.9
anterior.12	89.7	6.9	3.4
anterior.48	86.2	3.4	10.3
anterior.fu	90.0	5.0	5.0
inferior.6	13.3	83.3	3.3
inferior.12	10.0	83.3	6.7
inferior.48	3.3	93.3	3.3
inferior.fu	17.9	67.9	14.3
normal.50	23.3	13.3	63.3
CAD.50	13.3	13.3	73.3

C.51 Experiment L2.knn

Training Data

	anterior	inferior	normal
anterior	96.8	0.8	2.5
inferior	0.0	98.4	1.6
normal	0.8	0.0	99.2

Testing Data

	anterior	inferior	normal
anterior	67.6	8.9	23.5
inferior	4.2	69.0	26.7
normal	3.3	0.0	96.7

Training Data

	anterior	inferior	normal
anterior.6	99.0	1.0	0.0
anterior.12	98.0	2.0	0.0
anterior.48	97.7	0.0	2.3
anterior.fu	92.3	0.0	7.7
inferior.6	0.0	100.0	0.0
inferior.12	0.0	98.8	1.2
inferior.48	0.0	98.1	1.9
inferior.fu	0.0	96.8	3.2
normal.50	0.8	0.0	99.2
CAD.50	0.0	0.3	99.7

Testing Data

	anterior	inferior	normal
anterior.6	65.5	6.9	27.6
anterior.12	79.3	6.9	13.8
anterior.48	65.5	6.9	27.6
anterior.fu	60.0	15.0	25.0
inferior.6	3.3	73.3	23.3
inferior.12	3.3	90.0	6.7
inferior.48	6.7	70.0	23.3
inferior.fu	3.6	42.9	53.6
normal.50	3.3	0.0	96.7
CAD.50	0.0	6.7	93.3

C.52 Experiment L2.linreg

Training Data

	anterior	inferior	normal
anterior	96.2	2.5	1.3
inferior	0.4	97.8	1.8
normal	0.8	1.5	97.7

Testing Data

	anterior	inferior	normal
anterior	75.7	17.8	6.4
inferior	5.1	82.8	12.1
normal	10.0	16.7	73.3

Training Data

	anterior	inferior	normal
anterior.6	91.7	6.2	2.1
anterior.12	93.1	3.9	2.9
anterior.48	100.0	0.0	0.0
anterior.fu	100.0	0.0	0.0
inferior.6	0.0	100.0	0.0
inferior.12	0.0	97.6	2.4
inferior.48	0.0	100.0	0.0
inferior.fu	1.6	93.5	4.8
normal.50	0.8	1.5	97.7
CAD.50	13.4	18.0	68.5

Testing Data

	anterior	inferior	normal
anterior.6	65.5	20.7	13.8
anterior.12	86.2	10.3	3.4
anterior.48	86.2	10.3	3.4
anterior.fu	65.0	30.0	5.0
inferior.6	3.3	86.7	10.0
inferior.12	3.3	96.7	0.0
inferior.48	6.7	80.0	13.3
inferior.fu	7.1	67.9	25.0
normal.50	10.0	16.7	73.3
CAD.50	6.7	10.0	83.3

C.53 Experiment L2.C4.5

Training Data

	anterior	inferior	normal
anterior	99.0	0.3	0.8
inferior	0.4	99.6	0.0
normal	0.0	0.0	100.0

Testing Data

	anterior	inferior	normal
anterior	75.8	13.5	10.6
inferior	7.1	87.7	5.1
normal	33.3	6.7	60.0

Training Data

	anterior	inferior	normal
anterior.6	96.9	1.0	2.1
anterior.12	99.0	0.0	1.0
anterior.48	100.0	0.0	0.0
anterior.fu	100.0	0.0	0.0
inferior.6	0.0	100.0	0.0
inferior.12	0.0	100.0	0.0
inferior.48	0.0	100.0	0.0
inferior.fu	1.6	98.4	0.0
normal.50	0.0	0.0	100.0
CAD.50	24.6	23.3	52.1

Testing Data

	anterior	inferior	normal
anterior.6	86.2	3.4	10.3
anterior.12	82.8	10.3	6.9
anterior.48	79.3	10.3	10.3
anterior.fu	55.0	30.0	15.0
inferior.6	0.0	100.0	0.0
inferior.12	0.0	93.3	6.7
inferior.48	0.0	93.3	6.7
inferior.fu	28.6	64.3	7.1
normal.50	33.3	6.7	60.0
CAD.50	23.3	23.3	53.3

C.54 Experiment L2.MML

Training Data

	anterior	inferior	normal
anterior	96.2	2.5	1.3
inferior	0.4	98.6	1.0
normal	0.0	0.0	100.0

Testing Data

	anterior	inferior	normal
anterior	72.4	10.2	17.5
inferior	3.5	85.2	11.3
normal	16.7	20.0	63.3

Training Data

	anterior	inferior	normal
anterior.6	91.7	5.2	3.1
anterior.12	93.1	4.9	2.0
anterior.48	100.0	0.0	0.0
anterior.fu	100.0	0.0	0.0
inferior.6	0.0	98.9	1.1
inferior.12	0.0	98.8	1.2
inferior.48	0.0	100.0	0.0
inferior.fu	1.6	96.8	1.6
normal.50	0.0	0.0	100.0
CAD.50	18.4	22.3	59.3

Testing Data

	anterior	inferior	normal
anterior.6	82.8	6.9	10.3
anterior.12	82.8	6.9	10.3
anterior.48	69.0	6.9	24.1
anterior.fu	55.0	20.0	25.0
inferior.6	0.0	96.7	3.3
inferior.12	3.3	90.0	6.7
inferior.48	0.0	93.3	6.7
inferior.fu	10.7	60.7	28.6
normal.50	16.7	20.0	63.3
CAD.50	20.0	13.3	66.7

C.55 Experiment L2.bp

Training Data

	anterior	inferior	normal
anterior	91.3±1.6	4.1±1.1	4.6±1.4
inferior	3.1±0.6	93.4±1.3	3.5±1.0
normal	5.6±2.8	5.8±2.4	88.6±2.7

Testing Data

	anterior	inferior	normal
anterior	71.6±5.0	14.4±2.9	14.0±4.5
inferior	11.8±2.6	75.5±3.7	12.7±1.8
normal	23.5±9.4	16.7±5.9	59.8±9.3

Training Data

	anterior	inferior	normal
anterior.6	89.6±2.6	5.9±1.8	4.5±2.2
anterior.12	90.2±2.5	5.0±1.8	4.8±1.9
anterior.48	93.6±2.7	2.8±1.9	3.5±2.6
anterior.fu	91.5±5.0	2.7±3.2	5.8±4.6
inferior.6	1.5±1.0	96.2±1.8	2.3±1.5
inferior.12	1.6±1.6	93.8±2.2	4.6±1.8
inferior.48	3.6±1.4	95.2±2.3	1.2±1.6
inferior.fu	5.8±2.1	88.3±3.5	5.9±2.2
normal.50	5.6±2.8	5.8±2.4	88.6±2.7
CAD.50	17.6±2.4	29.6±2.9	52.7±2.7

Testing Data

	anterior	inferior	normal
anterior.6	77.1±5.7	8.4±4.7	14.5±5.4
anterior.12	74.3±9.0	11.2±6.7	14.5±6.6
anterior.48	73.4±5.5	12.1±5.4	14.5±5.4
anterior.fu	61.8±8.1	25.8±5.5	12.5±8.4
inferior.6	9.5±3.4	76.3±4.9	14.2±4.7
inferior.12	10.2±4.7	79.0±6.8	10.8±4.9
inferior.48	8.0±4.6	80.2±5.8	11.8±4.7
inferior.fu	19.5±6.5	66.4±9.2	14.1±6.3
normal.50	23.5±9.4	16.7±5.9	59.8±9.3
CAD.50	7.0±5.0	27.7±8.0	65.3±8.3

C.56 Experiment L2.qp

Training Data

	anterior	inferior	normal
anterior	78.6±3.9	8.4±2.3	13.0±4.1
inferior	5.2±1.6	86.5±3.2	8.3±2.3
normal	6.9±2.1	7.6±2.6	85.4±3.7

Testing Data

	anterior	inferior	normal
anterior	65.0±6.0	14.5±3.5	20.5±6.1
inferior	10.4±2.2	76.0±3.5	13.6±3.4
normal	20.3±7.4	11.5±6.5	68.2±7.6

Training Data

	anterior	inferior	normal
anterior.6	72.2±7.1	12.2±4.3	15.6±6.4
anterior.12	83.0±3.3	7.0±1.6	10.0±3.1
anterior.48	87.3±3.2	4.4±2.1	8.3±2.6
anterior.fu	71.9±6.7	10.0±5.6	18.1±7.7
inferior.6	3.5±2.9	89.7±4.1	6.8±2.3
inferior.12	2.4±2.1	88.7±3.9	9.0±3.3
inferior.48	3.9±1.4	91.2±3.8	4.9±3.5
inferior.fu	11.0±3.6	76.5±5.5	12.5±4.6
normal.50	6.9±2.1	7.6±2.6	85.4±3.7
CAD.50	13.2±2.5	25.4±3.0	61.4±4.0

Testing Data

	anterior	inferior	normal
anterior.6	67.6±5.5	11.9±5.3	20.5±7.7
anterior.12	72.1±9.6	11.0±5.6	16.9±6.6
anterior.48	74.7±4.7	8.1±2.9	17.2±4.9
anterior.fu	45.8±10.2	27.0±7.3	27.2±11.9
inferior.6	6.2±3.2	78.8±7.0	15.0±5.7
inferior.12	9.2±3.9	78.8±5.1	12.0±4.9
inferior.48	9.5±3.0	80.2±6.5	10.3±6.6
inferior.fu	16.8±5.2	66.1±4.9	17.1±4.7
normal.50	20.3±7.4	11.5±6.5	68.2±7.6
CAD.50	2.8±3.4	26.8±5.9	70.3±5.7

C.57 Experiment L2.cas

Training Data

	anterior	inferior	normal
anterior	100.0±0.1	0.0±0.0	0.0±0.1
inferior	1.1±0.7	98.9±0.7	0.0±0.0
normal	0.8±1.1	0.0±0.0	99.2±1.1

Testing Data

	anterior	inferior	normal
anterior	93.2±0.8	4.2±1.2	2.6±0.7
inferior	17.4±1.7	80.3±1.8	2.3±0.3
normal	31.0±1.5	8.3±2.2	60.7±2.0

Training Data

	anterior	inferior	normal
anterior.6	99.9±0.3	0.0±0.0	0.1±0.3
anterior.12	100.0±NaN	0.0±0.0	0.0±0.0
anterior.48	100.0±NaN	0.0±0.0	0.0±0.0
anterior.fu	100.0±NaN	0.0±0.0	0.0±0.0
inferior.6	0.3±0.7	99.7±0.7	0.0±0.0
inferior.12	1.3±1.3	98.7±1.3	0.0±0.0
inferior.48	2.1±0.6	97.9±0.6	0.0±0.0
inferior.fu	0.5±0.7	99.5±0.7	0.0±0.0
normal.50	0.8±1.1	0.0±0.0	99.2±1.1
CAD.50	28.4±1.4	16.2±1.4	55.4±1.6

Testing Data

	anterior	inferior	normal
anterior.6	91.4±2.8	0.0±0.0	8.6±2.8
anterior.12	100.0±NaN	0.0±0.0	0.0±0.0
anterior.48	95.9±2.6	2.8±2.6	1.4±1.7
anterior.fu	85.5±4.2	14.0±4.4	0.5±1.5
inferior.6	5.0±1.7	91.7±1.7	3.3±0.0
inferior.12	13.0±1.8	83.7±1.8	3.3±0.0
inferior.48	7.3±1.3	90.0±0.0	2.7±1.3
inferior.fu	44.3±4.8	55.7±4.8	0.0±0.0
normal.50	31.0±1.5	8.3±2.2	60.7±2.0
CAD.50	20.3±3.1	16.3±3.1	63.3±4.7

C.58 Experiment L2.cbp

Training Data

	anterior	inferior	normal
anterior	98.0	1.3	0.8
inferior	1.3	98.3	0.4
normal	0.8	0.8	98.5

Testing Data

	anterior	inferior	normal
anterior	76.6	12.3	11.1
inferior	8.6	82.9	8.5
normal	23.3	6.7	70.0

Training Data

	anterior	inferior	normal
anterior.6	94.8	3.1	2.1
anterior.12	97.1	2.0	1.0
anterior.48	100.0	0.0	0.0
anterior.fu	100.0	0.0	0.0
inferior.6	0.0	100.0	0.0
inferior.12	0.0	100.0	0.0
inferior.48	1.9	98.1	0.0
inferior.fu	3.2	95.2	1.6
normal.50	0.8	0.8	98.5
CAD.50	12.5	25.9	61.6

Testing Data

	anterior	inferior	normal
anterior.6	82.8	6.9	10.3
anterior.12	82.8	3.4	13.8
anterior.48	75.9	13.8	10.3
anterior.fu	65.0	25.0	10.0
inferior.6	6.7	86.7	6.7
inferior.12	3.3	86.7	10.0
inferior.48	6.7	86.7	6.7
inferior.fu	17.9	71.4	10.7
normal.50	23.3	6.7	70.0
CAD.50	3.3	13.3	83.3

C.59 Experiment L2.cqp

Training Data

	anterior	inferior	normal
anterior	83.6	7.0	9.4
inferior	2.8	91.5	5.7
normal	4.6	5.3	90.1

Testing Data

	anterior	inferior	normal
anterior	70.6	15.6	13.7
inferior	8.6	80.3	11.1
normal	16.7	6.7	76.7

Training Data

	anterior	inferior	normal
anterior.6	82.3	10.4	7.3
anterior.12	88.2	5.9	5.9
anterior.48	90.9	0.0	9.1
anterior.fu	73.1	11.5	15.4
inferior.6	1.1	94.6	4.3
inferior.12	0.0	91.5	8.5
inferior.48	1.9	96.2	1.9
inferior.fu	8.1	83.9	8.1
normal.50	4.6	5.3	90.1
CAD.50	10.8	25.2	63.9

Testing Data

	anterior	inferior	normal
anterior.6	69.0	13.8	17.2
anterior.12	82.8	6.9	10.3
anterior.48	75.9	6.9	17.2
anterior.fu	55.0	35.0	10.0
inferior.6	3.3	86.7	10.0
inferior.12	3.3	86.7	10.0
inferior.48	10.0	80.0	10.0
inferior.fu	17.9	67.9	14.3
normal.50	16.7	6.7	76.7
CAD.50	0.0	20.0	80.0

C.60 Experiment L2.ccas

Training Data

	anterior	inferior	normal
anterior	100.0	0.0	0.0
inferior	0.5	99.5	0.0
normal	0.0	0.0	100.0

Testing Data

	anterior	inferior	normal
anterior	92.4	5.9	1.7
inferior	14.8	82.7	2.5
normal	30.0	3.3	66.7

Training Data

	anterior	inferior	normal
anterior.6	100.0	0.0	0.0
anterior.12	100.0	0.0	0.0
anterior.48	100.0	0.0	0.0
anterior.fu	100.0	0.0	0.0
inferior.6	0.0	100.0	0.0
inferior.12	0.0	100.0	0.0
inferior.48	1.9	98.1	0.0
inferior.fu	0.0	100.0	0.0
normal.50	0.0	0.0	100.0
CAD.50	28.9	13.1	58.0

Testing Data

	anterior	inferior	normal
anterior.6	93.1	0.0	6.9
anterior.12	100.0	0.0	0.0
anterior.48	96.6	3.4	0.0
anterior.fu	80.0	20.0	0.0
inferior.6	3.3	93.3	3.3
inferior.12	10.0	86.7	3.3
inferior.48	6.7	90.0	3.3
inferior.fu	39.3	60.7	0.0
normal.50	30.0	3.3	66.7
CAD.50	16.7	13.3	70.0

C.61 Experiment st2.knn

Training Data

	anterior	inferior	normal
anterior	94.5	1.5	4.0
inferior	0.0	97.1	2.9
normal	0.0	0.8	99.2

Testing Data

	anterior	inferior	normal
anterior	47.4	6.3	46.3
inferior	0.9	68.6	30.5
normal	3.3	6.7	90.0

Training Data

	anterior	inferior	normal
anterior.6	96.9	0.0	3.1
anterior.12	97.1	0.0	2.9
anterior.48	95.5	2.3	2.3
anterior.fu	88.5	3.8	7.7
inferior.6	0.0	100.0	0.0
inferior.12	0.0	100.0	0.0
inferior.48	0.0	98.1	1.9
inferior.fu	0.0	90.3	9.7
normal.50	0.0	0.8	99.2
CAD.50	0.0	0.3	99.7

Testing Data

	anterior	inferior	normal
anterior.6	72.4	3.4	24.1
anterior.12	55.2	0.0	44.8
anterior.48	62.1	6.9	31.0
anterior.fu	0.0	15.0	85.0
inferior.6	0.0	86.7	13.3
inferior.12	0.0	83.3	16.7
inferior.48	0.0	90.0	10.0
inferior.fu	3.6	14.3	82.1
normal.50	3.3	6.7	90.0
CAD.50	3.3	0.0	96.7

C.62 Experiment st2.linreg

Training Data

	anterior	inferior	normal
anterior	79.1	6.3	14.6
inferior	7.3	76.7	15.9
normal	6.1	7.6	86.3

Testing Data

	anterior	inferior	normal
anterior	69.0	5.9	25.1
inferior	7.8	63.8	28.4
normal	10.0	10.0	80.0

Training Data

	anterior	inferior	normal
anterior.6	72.9	3.1	24.0
anterior.12	79.4	6.9	13.7
anterior.48	90.9	0.0	9.1
anterior.fu	73.1	15.4	11.5
inferior.6	3.2	86.0	10.8
inferior.12	6.1	79.3	14.6
inferior.48	3.8	88.5	7.7
inferior.fu	16.1	53.2	30.6
normal.50	6.1	7.6	86.3
CAD.50	14.4	13.4	72.1

Testing Data

	anterior	inferior	normal
anterior.6	79.3	3.4	17.2
anterior.12	79.3	0.0	20.7
anterior.48	72.4	0.0	27.6
anterior.fu	45.0	20.0	35.0
inferior.6	6.7	76.7	16.7
inferior.12	6.7	70.0	23.3
inferior.48	0.0	80.0	20.0
inferior.fu	17.9	28.6	53.6
normal.50	10.0	10.0	80.0
CAD.50	6.7	13.3	80.0

C.63 Experiment st2.C4.5

Training Data

	anterior	inferior	normal
anterior	99.0	0.2	0.8
inferior	1.7	96.9	1.4
normal	0.0	0.0	100.0

Testing Data

	anterior	inferior	normal
anterior	65.7	14.4	19.9
inferior	7.9	73.9	18.2
normal	36.7	30.0	33.3

Training Data

	anterior	inferior	normal
anterior.6	99.0	0.0	1.0
anterior.12	97.1	1.0	2.0
anterior.48	100.0	0.0	0.0
anterior.fu	100.0	0.0	0.0
inferior.6	1.1	98.9	0.0
inferior.12	2.4	95.1	2.4
inferior.48	0.0	100.0	0.0
inferior.fu	3.2	93.5	3.2
normal.50	0.0	0.0	100.0
CAD.50	21.6	28.9	49.5

Testing Data

	anterior	inferior	normal
anterior.6	82.8	3.4	13.8
anterior.12	79.3	10.3	10.3
anterior.48	75.9	13.8	10.3
anterior.fu	25.0	30.0	45.0
inferior.6	0.0	86.7	13.3
inferior.12	0.0	90.0	10.0
inferior.48	6.7	86.7	6.7
inferior.fu	25.0	32.1	42.9
normal.50	36.7	30.0	33.3
CAD.50	16.7	16.7	66.7

C.64 Experiment st2.MML

Training Data

	anterior	inferior	normal
anterior	98.5	0.5	1.0
inferior	2.5	96.5	1.0
normal	0.0	0.0	100.0

Testing Data

	anterior	inferior	normal
anterior	61.0	20.7	18.2
inferior	7.7	79.2	13.2
normal	36.7	20.0	43.3

Training Data

	anterior	inferior	normal
anterior.6	97.9	0.0	2.1
anterior.12	96.1	2.0	2.0
anterior.48	100.0	0.0	0.0
anterior.fu	100.0	0.0	0.0
inferior.6	1.1	98.9	0.0
inferior.12	2.4	95.1	2.4
inferior.48	0.0	100.0	0.0
inferior.fu	6.5	91.9	1.6
normal.50	0.0	0.0	100.0
CAD.50	21.3	32.1	46.6

Testing Data

	anterior	inferior	normal
anterior.6	79.3	13.8	6.9
anterior.12	79.3	10.3	10.3
anterior.48	65.5	13.8	20.7
anterior.fu	20.0	45.0	35.0
inferior.6	0.0	93.3	6.7
inferior.12	13.3	80.0	6.7
inferior.48	6.7	93.3	0.0
inferior.fu	10.7	50.0	39.3
normal.50	36.7	20.0	43.3
CAD.50	36.7	23.3	40.0

C.65 Experiment st2.bp

Training Data

	anterior	inferior	normal
anterior	75.7±7.3	5.7±1.8	18.6±7.5
inferior	10.8±4.5	74.5±3.9	14.6±5.1
normal	19.8±8.1	8.2±2.4	72.0±8.7

Testing Data

	anterior	inferior	normal
anterior	63.0±8.1	12.2±3.0	24.8±8.2
inferior	15.1±5.6	69.1±3.0	15.8±5.0
normal	31.0±12.0	23.8±7.0	45.2±11.3

Training Data

	anterior	inferior	normal
anterior.6	73.3±10.1	4.0±1.7	22.7±10.1
anterior.12	74.9±7.4	6.0±2.0	19.1±7.8
anterior.48	83.4±6.5	2.5±2.1	14.1±6.4
anterior.fu	71.2±9.8	10.4±5.6	18.5±9.2
inferior.6	2.0±1.8	92.2±2.8	5.8±2.4
inferior.12	5.2±5.1	84.8±6.0	10.1±5.1
inferior.48	7.4±4.3	79.9±5.9	12.7±5.8
inferior.fu	28.7±10.5	41.3±6.3	30.0±11.3
normal.50	19.8±8.1	8.2±2.4	72.0±8.7
CAD.50	32.3±8.4	21.8±3.3	45.9±8.8

Testing Data

	anterior	inferior	normal
anterior.6	68.8±11.0	6.6±4.1	24.7±9.8
anterior.12	76.4±11.3	6.2±4.8	17.4±10.1
anterior.48	61.4±9.9	11.7±5.3	26.9±10.0
anterior.fu	45.5±14.6	24.2±7.5	30.2±12.6
inferior.6	5.5±4.3	84.3±5.1	10.2±5.0
inferior.12	8.2±4.0	83.5±5.0	8.3±4.7
inferior.48	7.2±5.3	82.5±4.1	10.3±4.8
inferior.fu	39.5±15.0	26.2±6.9	34.3±13.7
normal.50	31.0±12.0	23.8±7.0	45.2±11.3
CAD.50	33.5±10.4	18.3±6.1	48.2±10.6

C.66 Experiment st2.qp

Training Data

	anterior	inferior	normal
anterior	79.8±4.9	5.5±2.3	14.7±3.7
inferior	6.9±3.1	77.7±4.0	15.4±4.2
normal	15.2±3.9	8.6±5.2	76.2±7.6

Testing Data

	anterior	inferior	normal
anterior	70.2±5.7	10.3±4.1	19.6±5.6
inferior	12.7±2.9	71.5±2.9	15.8±2.2
normal	30.0±11.5	20.8±9.0	49.2±12.5

Training Data

	anterior	inferior	normal
anterior.6	78.2±6.2	3.7±3.5	18.1±5.5
anterior.12	81.4±4.0	4.9±2.4	13.7±3.3
anterior.48	90.7±5.2	1.6±2.0	7.7±4.6
anterior.fu	69.0±9.9	11.9±5.4	19.0±7.9
inferior.6	2.0±1.8	91.6±4.9	6.4±4.1
inferior.12	2.4±2.1	88.0±4.8	9.6±4.3
inferior.48	2.5±3.4	87.2±5.2	10.3±4.4
inferior.fu	20.9±8.7	43.8±6.1	35.3±10.0
normal.50	15.2±3.9	8.6±5.2	76.2±7.6
CAD.50	26.7±4.3	21.1±4.3	52.2±5.5

Testing Data

	anterior	inferior	normal
anterior.6	76.4±8.7	4.0±4.9	19.7±7.6
anterior.12	82.2±5.9	5.9±3.6	11.9±6.4
anterior.48	74.5±7.0	7.2±3.9	18.3±5.5
anterior.fu	47.5±9.9	24.0±11.2	28.5±10.4
inferior.6	5.3±3.2	88.2±4.3	6.5±4.0
inferior.12	6.5±4.3	83.8±4.4	9.7±4.7
inferior.48	2.8±2.6	88.2±3.7	9.0±4.0
inferior.fu	36.1±7.5	25.7±6.8	38.2±7.1
normal.50	30.0±11.5	20.8±9.0	49.2±12.5
CAD.50	27.5±7.6	22.5±7.7	50.0±8.3

C.67 Experiment st2.cas

Training Data

	anterior	inferior	normal
anterior	94.2±2.2	1.5±0.9	4.3±1.9
inferior	12.9±1.8	82.1±1.3	5.0±1.3
normal	23.4±4.3	7.0±1.5	69.6±4.4

Testing Data

	anterior	inferior	normal
anterior	88.8±2.7	5.5±1.6	5.7±1.7
inferior	18.5±3.2	71.1±1.9	10.4±2.1
normal	43.0±7.1	9.7±4.1	47.3±5.7

Training Data

	anterior	inferior	normal
anterior.6	95.3±2.4	1.4±0.9	3.3±2.1
anterior.12	94.2±1.9	2.8±1.1	2.9±1.2
anterior.48	93.9±2.7	1.1±1.5	5.0±2.8
anterior.fu	93.5±4.6	0.8±2.3	5.8±3.5
inferior.6	8.8±2.3	88.4±2.6	2.8±1.1
inferior.12	9.6±3.0	88.2±2.3	2.2±1.2
inferior.48	6.7±1.0	91.2±1.3	2.1±0.6
inferior.fu	26.5±4.3	60.6±3.8	12.9±4.6
normal.50	23.4±4.3	7.0±1.5	69.6±4.4
CAD.50	35.7±3.7	15.3±2.1	48.9±4.1

Testing Data

	anterior	inferior	normal
anterior.6	96.6±3.1	0.7±1.4	2.8±3.0
anterior.12	97.2±2.1	1.7±1.7	1.0±2.2
anterior.48	92.8±1.9	7.2±1.9	0.0±0.0
anterior.fu	68.5±6.3	12.5±4.6	19.0±3.0
inferior.6	14.3±4.5	79.0±3.7	6.7±3.0
inferior.12	10.0±3.0	84.3±1.5	5.7±2.1
inferior.48	12.7±1.3	80.3±1.0	7.0±1.8
inferior.fu	37.1±10.5	40.7±6.0	22.1±7.3
normal.50	43.0±7.1	9.7±4.1	47.3±5.7
CAD.50	42.3±5.8	12.3±3.7	45.3±5.0

C.68 Experiment st2.cbp

Training Data

	anterior	inferior	normal
anterior	88.2	2.8	9.0
inferior	9.1	78.0	12.8
normal	14.5	4.6	80.9

Testing Data

	anterior	inferior	normal
anterior	77.5	7.2	15.3
inferior	14.0	72.0	14.0
normal	30.0	20.0	50.0

Training Data

	anterior	inferior	normal
anterior.6	88.5	1.0	10.4
anterior.12	84.3	3.9	11.8
anterior.48	95.5	2.3	2.3
anterior.fu	84.6	3.8	11.5
inferior.6	2.2	93.5	4.3
inferior.12	2.4	92.7	4.9
inferior.48	7.7	80.8	11.5
inferior.fu	24.2	45.2	30.6
normal.50	14.5	4.6	80.9
CAD.50	30.5	17.7	51.8

Testing Data

	anterior	inferior	normal
anterior.6	89.7	3.4	6.9
anterior.12	96.6	3.4	0.0
anterior.48	69.0	6.9	24.1
anterior.fu	55.0	15.0	30.0
inferior.6	3.3	90.0	6.7
inferior.12	6.7	90.0	3.3
inferior.48	3.3	90.0	6.7
inferior.fu	42.9	17.9	39.3
normal.50	30.0	20.0	50.0
CAD.50	26.7	20.0	53.3

C.69 Experiment st2.cqp

Training Data

	anterior	inferior	normal
anterior	91.4	2.5	6.0
inferior	5.6	85.3	9.1
normal	10.7	3.8	85.5

Testing Data

	anterior	inferior	normal
anterior	80.9	9.7	9.4
inferior	13.2	75.4	11.4
normal	30.0	23.3	46.7

Training Data

	anterior	inferior	normal
anterior.6	90.6	2.1	7.3
anterior.12	91.2	2.0	6.9
anterior.48	95.5	2.3	2.3
anterior.fu	88.5	3.8	7.7
inferior.6	0.0	97.8	2.2
inferior.12	0.0	98.8	1.2
inferior.48	0.0	96.2	3.8
inferior.fu	22.6	48.4	29.0
normal.50	10.7	3.8	85.5
CAD.50	27.9	17.0	55.1

Testing Data

	anterior	inferior	normal
anterior.6	82.8	3.4	13.8
anterior.12	93.1	3.4	3.4
anterior.48	82.8	6.9	10.3
anterior.fu	65.0	25.0	10.0
inferior.6	3.3	96.7	0.0
inferior.12	6.7	86.7	6.7
inferior.48	0.0	93.3	6.7
inferior.fu	42.9	25.0	32.1
normal.50	30.0	23.3	46.7
CAD.50	26.7	23.3	50.0

C.70 Experiment st2.ccas

Training Data

	anterior	inferior	normal
anterior	93.5	1.3	5.2
inferior	11.5	83.5	5.0
normal	18.3	5.3	76.3

Testing Data

	anterior	inferior	normal
anterior	88.7	4.2	7.1
inferior	19.8	70.7	9.5
normal	23.3	13.3	63.3

Training Data

	anterior	inferior	normal
anterior.6	91.7	2.1	6.2
anterior.12	93.1	2.9	3.9
anterior.48	93.2	0.0	6.8
anterior.fu	96.2	0.0	3.8
inferior.6	7.5	90.3	2.2
inferior.12	8.5	90.2	1.2
inferior.48	5.8	92.3	1.9
inferior.fu	24.2	61.3	14.5
normal.50	18.3	5.3	76.3
CAD.50	31.1	13.8	55.1

Testing Data

	anterior	inferior	normal
anterior.6	96.6	0.0	3.4
anterior.12	100.0	0.0	0.0
anterior.48	93.1	6.9	0.0
anterior.fu	65.0	10.0	25.0
inferior.6	16.7	80.0	3.3
inferior.12	10.0	83.3	6.7
inferior.48	13.3	80.0	6.7
inferior.fu	39.3	39.3	21.4
normal.50	23.3	13.3	63.3
CAD.50	36.7	13.3	50.0

C.71 Experiment qrs2.knn

Training Data

	anterior	inferior	normal
anterior	96.5	0.8	2.8
inferior	0.0	97.4	2.6
normal	0.8	0.0	99.2

Testing Data

	anterior	inferior	normal
anterior	69.7	11.0	19.3
inferior	7.6	65.9	26.5
normal	3.3	0.0	96.7

Training Data

	anterior	inferior	normal
anterior.6	97.9	1.0	1.0
anterior.12	98.0	2.0	0.0
anterior.48	97.7	0.0	2.3
anterior.fu	92.3	0.0	7.7
inferior.6	0.0	97.8	2.2
inferior.12	0.0	98.8	1.2
inferior.48	0.0	96.2	3.8
inferior.fu	0.0	96.8	3.2
normal.50	0.8	0.0	99.2
CAD.50	0.0	1.0	99.0

Testing Data

	anterior	inferior	normal
anterior.6	65.5	6.9	27.6
anterior.12	79.3	6.9	13.8
anterior.48	69.0	10.3	20.7
anterior.fu	65.0	20.0	15.0
inferior.6	6.7	60.0	33.3
inferior.12	10.0	83.3	6.7
inferior.48	10.0	66.7	23.3
inferior.fu	3.6	53.6	42.9
normal.50	3.3	0.0	96.7
CAD.50	0.0	6.7	93.3

C.72 Experiment qrs2.linreg

Training Data

	anterior	inferior	normal
anterior	91.2	5.1	3.7
inferior	3.6	93.2	3.2
normal	3.1	1.5	95.4

Testing Data

	anterior	inferior	normal
anterior	73.4	18.8	7.8
inferior	8.5	82.1	9.4
normal	0.0	23.3	76.7

Training Data

	anterior	inferior	normal
anterior.6	85.4	8.3	6.2
anterior.12	86.3	9.8	3.9
anterior.48	93.2	2.3	4.5
anterior.fu	100.0	0.0	0.0
inferior.6	2.2	96.8	1.1
inferior.12	2.4	92.7	4.9
inferior.48	1.9	96.2	1.9
inferior.fu	8.1	87.1	4.8
normal.50	3.1	1.5	95.4
CAD.50	12.5	19.3	68.2

Testing Data

	anterior	inferior	normal
anterior.6	65.5	24.1	10.3
anterior.12	79.3	13.8	6.9
anterior.48	69.0	17.2	13.8
anterior.fu	80.0	20.0	0.0
inferior.6	6.7	83.3	10.0
inferior.12	16.7	83.3	0.0
inferior.48	0.0	86.7	13.3
inferior.fu	10.7	75.0	14.3
normal.50	0.0	23.3	76.7
CAD.50	3.3	20.0	76.7

C.73 Experiment qrs2.C4.5

Training Data

	anterior	inferior	normal
anterior	98.5	0.8	0.8
inferior	0.6	98.9	0.6
normal	0.0	0.0	100.0

Testing Data

	anterior	inferior	normal
anterior	74.1	8.9	17.0
inferior	12.0	77.9	10.2
normal	30.0	6.7	63.3

Training Data

	anterior	inferior	normal
anterior.6	95.8	2.1	2.1
anterior.12	98.0	1.0	1.0
anterior.48	100.0	0.0	0.0
anterior.fu	100.0	0.0	0.0
inferior.6	1.1	97.8	1.1
inferior.12	1.2	97.6	1.2
inferior.48	0.0	100.0	0.0
inferior.fu	0.0	100.0	0.0
normal.50	0.0	0.0	100.0
CAD.50	27.5	27.2	45.2

Testing Data

	anterior	inferior	normal
anterior.6	79.3	6.9	13.8
anterior.12	72.4	10.3	17.2
anterior.48	89.7	3.4	6.9
anterior.fu	55.0	15.0	30.0
inferior.6	16.7	70.0	13.3
inferior.12	6.7	86.7	6.7
inferior.48	6.7	83.3	10.0
inferior.fu	17.9	71.4	10.7
normal.50	30.0	6.7	63.3
CAD.50	13.3	20.0	66.7

C.74 Experiment qrs2.MML

Training Data

	anterior	inferior	normal
anterior	97.4	2.0	0.5
inferior	1.2	98.2	0.6
normal	0.8	0.0	99.2

Testing Data

	anterior	inferior	normal
anterior	77.4	11.5	11.1
inferior	9.3	75.3	15.4
normal	20.0	20.0	60.0

Training Data

	anterior	inferior	normal
anterior.6	92.7	5.2	2.1
anterior.12	97.1	2.9	0.0
anterior.48	100.0	0.0	0.0
anterior.fu	100.0	0.0	0.0
inferior.6	2.2	96.8	1.1
inferior.12	1.2	97.6	1.2
inferior.48	0.0	100.0	0.0
inferior.fu	1.6	98.4	0.0
normal.50	0.8	0.0	99.2
CAD.50	24.9	29.2	45.9

Testing Data

	anterior	inferior	normal
anterior.6	72.4	13.8	13.8
anterior.12	86.2	10.3	3.4
anterior.48	75.9	6.9	17.2
anterior.fu	75.0	15.0	10.0
inferior.6	10.0	73.3	16.7
inferior.12	6.7	86.7	6.7
inferior.48	10.0	73.3	16.7
inferior.fu	10.7	67.9	21.4
normal.50	20.0	20.0	60.0
CAD.50	10.0	33.3	56.7

C.75 Experiment qrs2.bp

Training Data

	anterior	inferior	normal
anterior	90.3±2.3	4.6±1.2	5.2±1.5
inferior	5.4±1.0	89.3±1.8	5.3±1.2
normal	5.2±2.1	6.7±1.9	88.1±2.9

Testing Data

	anterior	inferior	normal
anterior	71.6±4.8	14.0±3.6	14.4±4.0
inferior	15.3±2.9	69.1±5.2	15.5±4.5
normal	22.3±6.3	15.5±7.2	62.2±8.6

Training Data

	anterior	inferior	normal
anterior.6	85.4±3.7	8.6±2.3	6.0±2.2
anterior.12	88.9±2.3	5.7±1.8	5.4±2.4
anterior.48	94.2±2.5	1.8±1.5	4.0±2.3
anterior.fu	92.7±4.7	2.1±2.6	5.2±4.3
inferior.6	5.4±2.8	88.7±3.7	6.0±1.9
inferior.12	3.0±1.9	88.9±3.2	8.1±2.7
inferior.48	5.1±2.4	93.0±2.6	1.9±1.6
inferior.fu	8.3±3.4	86.5±3.3	5.2±2.4
normal.50	5.2±2.1	6.7±1.9	88.1±2.9
CAD.50	16.0±2.5	29.9±3.2	54.1±4.0

Testing Data

	anterior	inferior	normal
anterior.6	71.4±5.5	11.9±6.2	16.7±6.8
anterior.12	75.3±8.5	9.8±5.5	14.8±6.3
anterior.48	72.8±6.3	15.3±5.7	11.9±5.8
anterior.fu	66.8±9.0	19.0±8.6	14.2±7.9
inferior.6	14.7±4.8	63.7±7.4	21.7±8.5
inferior.12	15.7±4.7	73.2±6.8	11.2±4.6
inferior.48	13.0±3.6	72.2±8.1	14.8±6.5
inferior.fu	18.0±6.3	67.5±8.6	14.5±4.6
normal.50	22.3±6.3	15.5±7.2	62.2±8.6
CAD.50	8.0±5.6	28.5±6.4	63.5±8.8

C.76 Experiment qrs2.qp

Training Data

	anterior	inferior	normal
anterior	72.7±4.4	10.6±1.9	16.7±3.7
inferior	7.7±1.5	81.5±2.9	10.7±2.0
normal	7.1±2.5	11.0±2.3	81.9±3.7

Testing Data

	anterior	inferior	normal
anterior	58.7±4.7	16.7±3.7	24.6±4.0
inferior	13.3±2.6	69.8±5.2	17.0±3.5
normal	19.2±7.0	17.5±7.2	63.3±8.2

Training Data

	anterior	inferior	normal
anterior.6	61.6±6.2	15.9±3.4	22.6±5.1
anterior.12	80.2±3.4	8.1±2.0	11.7±3.1
anterior.48	85.6±3.5	5.1±2.6	9.3±3.2
anterior.fu	63.5±7.2	13.3±4.3	23.3±6.6
inferior.6	8.9±1.8	79.6±4.2	11.5±3.6
inferior.12	4.8±1.6	83.5±3.3	11.8±2.4
inferior.48	6.3±2.1	86.6±3.8	7.1±2.9
inferior.fu	10.9±3.3	76.5±4.1	12.7±3.4
normal.50	7.1±2.5	11.0±2.3	81.9±3.7
CAD.50	10.6±2.2	29.5±2.7	59.9±3.3

Testing Data

	anterior	inferior	normal
anterior.6	62.9±4.5	15.5±4.8	21.6±5.0
anterior.12	63.6±6.9	11.2±5.5	25.2±7.2
anterior.48	70.0±5.7	10.7±4.9	19.3±5.3
anterior.fu	38.2±8.8	29.2±7.6	32.5±8.0
inferior.6	12.5±4.1	66.2±7.7	21.3±6.4
inferior.12	14.3±4.1	72.7±7.4	13.0±4.8
inferior.48	12.2±3.7	72.0±8.7	15.8±6.8
inferior.fu	14.1±4.1	68.2±5.7	17.7±3.1
normal.50	19.2±7.0	17.5±7.2	63.3±8.2
CAD.50	2.2±2.6	28.3±8.3	69.5±9.2

C.77 Experiment qrs2.cas

Training Data

	anterior	inferior	normal
anterior	97.2±0.4	2.2±0.4	0.6±0.3
inferior	11.0±1.6	87.6±1.5	1.4±0.5
normal	8.2±3.4	2.0±1.7	89.8±3.3

Testing Data

	anterior	inferior	normal
anterior	82.6±1.1	10.8±1.1	6.6±0.7
inferior	27.6±1.9	69.3±2.4	3.1±0.9
normal	22.3±3.7	12.0±3.1	65.7±1.5

Training Data

	anterior	inferior	normal
anterior.6	93.6±1.9	4.1±1.1	2.3±1.2
anterior.12	96.5±0.8	3.3±0.5	0.2±0.4
anterior.48	98.6±1.1	1.4±1.1	0.0±0.0
anterior.fu	100.0±NaN	0.0±0.0	0.0±0.0
inferior.6	9.6±2.2	87.3±2.0	3.1±1.0
inferior.12	9.9±2.2	89.9±2.3	0.2±0.5
inferior.48	7.3±1.9	92.7±1.9	0.0±0.0
inferior.fu	17.4±3.3	80.3±3.1	2.3±1.1
normal.50	8.2±3.4	2.0±1.7	89.8±3.3
CAD.50	24.2±0.6	22.4±0.6	53.3±0.8

Testing Data

	anterior	inferior	normal
anterior.6	73.1±2.6	16.6±1.4	10.3±2.2
anterior.12	94.5±1.7	5.5±1.7	0.0±0.0
anterior.48	88.6±2.2	2.8±1.4	8.6±1.7
anterior.fu	74.0±2.0	18.5±3.2	7.5±2.5
inferior.6	26.0±2.5	69.7±2.3	4.3±2.1
inferior.12	31.7±1.7	65.7±3.0	2.7±2.0
inferior.48	15.3±2.7	84.0±2.5	0.7±1.3
inferior.fu	37.5±5.6	57.9±5.9	4.6±2.8
normal.50	22.3±3.7	12.0±3.1	65.7±1.5
CAD.50	10.7±2.0	19.0±1.5	70.3±1.8

C.78 Experiment qrs2.cbp

Training Data

	anterior	inferior	normal
anterior	96.1	1.8	2.1
inferior	2.7	95.5	1.8
normal	3.1	0.8	96.2

Testing Data

	anterior	inferior	normal
anterior	74.9	14.0	11.1
inferior	14.4	75.5	10.1
normal	23.3	6.7	70.0

Training Data

	anterior	inferior	normal
anterior.6	92.7	6.2	1.0
anterior.12	96.1	1.0	2.9
anterior.48	95.5	0.0	4.5
anterior.fu	100.0	0.0	0.0
inferior.6	2.2	95.7	2.2
inferior.12	0.0	95.1	4.9
inferior.48	3.8	96.2	0.0
inferior.fu	4.8	95.2	0.0
normal.50	3.1	0.8	96.2
CAD.50	11.1	27.2	61.6

Testing Data

	anterior	inferior	normal
anterior.6	75.9	3.4	20.7
anterior.12	89.7	6.9	3.4
anterior.48	69.0	20.7	10.3
anterior.fu	65.0	25.0	10.0
inferior.6	13.3	73.3	13.3
inferior.12	16.7	76.7	6.7
inferior.48	13.3	73.3	13.3
inferior.fu	14.3	78.6	7.1
normal.50	23.3	6.7	70.0
CAD.50	3.3	26.7	70.0

C.79 Experiment qrs2.cqp

Training Data

	anterior	inferior	normal
anterior	77.3	9.1	13.6
inferior	7.4	85.6	7.0
normal	4.6	6.9	88.5

Testing Data

	anterior	inferior	normal
anterior	63.4	16.9	19.7
inferior	12.7	74.5	12.7
normal	20.0	6.7	73.3

Training Data

	anterior	inferior	normal
anterior.6	69.8	12.5	17.7
anterior.12	85.3	7.8	6.9
anterior.48	88.6	4.5	6.8
anterior.fu	65.4	11.5	23.1
inferior.6	7.5	86.0	6.5
inferior.12	4.9	85.4	9.8
inferior.48	5.8	90.4	3.8
inferior.fu	11.3	80.6	8.1
normal.50	4.6	6.9	88.5
CAD.50	9.2	30.5	60.3

Testing Data

	anterior	inferior	normal
anterior.6	65.5	13.8	20.7
anterior.12	72.4	3.4	24.1
anterior.48	75.9	10.3	13.8
anterior.fu	40.0	40.0	20.0
inferior.6	10.0	73.3	16.7
inferior.12	16.7	73.3	10.0
inferior.48	10.0	80.0	10.0
inferior.fu	14.3	71.4	14.3
normal.50	20.0	6.7	73.3
CAD.50	0.0	23.3	76.7

C.80 Experiment qrs2.ccas

Training Data

	anterior	inferior	normal
anterior	97.0	2.0	1.0
inferior	6.6	92.2	1.2
normal	6.1	0.0	93.9

Testing Data

	anterior	inferior	normal
anterior	80.4	11.9	7.7
inferior	22.3	75.2	2.6
normal	16.7	16.7	66.7

Training Data

	anterior	inferior	normal
anterior.6	92.7	4.2	3.1
anterior.12	95.1	3.9	1.0
anterior.48	100.0	0.0	0.0
anterior.fu	100.0	0.0	0.0
inferior.6	5.4	91.4	3.2
inferior.12	7.3	92.7	0.0
inferior.48	3.8	96.2	0.0
inferior.fu	9.7	88.7	1.6
normal.50	6.1	0.0	93.9
CAD.50	20.3	21.0	58.7

Testing Data

	anterior	inferior	normal
anterior.6	69.0	17.2	13.8
anterior.12	93.1	6.9	0.0
anterior.48	89.7	3.4	6.9
anterior.fu	70.0	20.0	10.0
inferior.6	23.3	73.3	3.3
inferior.12	23.3	73.3	3.3
inferior.48	6.7	93.3	0.0
inferior.fu	35.7	60.7	3.6
normal.50	16.7	16.7	66.7
CAD.50	10.0	16.7	73.3

C.81 Experiment E3.knn

Training Data

	anterior	inferior	normal
anterior	91.8	1.5	6.7
inferior	2.5	90.7	6.8
normal	0.0	0.0	100.0

Testing Data

	anterior	inferior	normal
anterior	81.7	8.8	9.5
inferior	6.7	76.8	16.4
normal	10.0	13.3	76.7

Training Data

	anterior	inferior	normal
anterior.6	100.0	0.0	0.0
anterior.12	98.0	2.0	0.0
anterior.48	100.0	0.0	0.0
anterior.fu	69.2	3.8	26.9
inferior.6	0.0	100.0	0.0
inferior.12	0.0	98.8	1.2
inferior.48	1.9	96.2	1.9
inferior.fu	8.1	67.7	24.2
normal.50	0.0	0.0	100.0
CAD.50	6.2	23.6	70.2

Testing Data

	anterior	inferior	normal
anterior.6	75.9	3.4	20.7
anterior.12	93.1	3.4	3.4
anterior.48	82.8	3.4	13.8
anterior.fu	75.0	25.0	0.0
inferior.6	3.3	86.7	10.0
inferior.12	10.0	83.3	6.7
inferior.48	10.0	76.7	13.3
inferior.fu	3.6	60.7	35.7
normal.50	10.0	13.3	76.7
CAD.50	0.0	16.7	83.3

C.82 Experiment E3.linreg

Training Data

	anterior	inferior	normal
anterior	81.7	6.7	11.6
inferior	3.6	86.8	9.6
normal	3.8	6.9	89.3

Testing Data

	anterior	inferior	normal
anterior	79.3	9.2	11.5
inferior	4.3	78.5	17.1
normal	3.3	10.0	86.7

Training Data

	anterior	inferior	normal
anterior.6	87.5	4.2	8.3
anterior.12	92.2	4.9	2.9
anterior.48	93.2	2.3	4.5
anterior.fu	53.8	15.4	30.8
inferior.6	1.1	95.7	3.2
inferior.12	0.0	93.9	6.1
inferior.48	1.9	98.1	0.0
inferior.fu	11.3	59.7	29.0
normal.50	3.8	6.9	89.3
CAD.50	9.5	15.4	75.1

Testing Data

	anterior	inferior	normal
anterior.6	89.7	3.4	6.9
anterior.12	86.2	3.4	10.3
anterior.48	86.2	0.0	13.8
anterior.fu	55.0	30.0	15.0
inferior.6	3.3	80.0	16.7
inferior.12	3.3	83.3	13.3
inferior.48	0.0	90.0	10.0
inferior.fu	10.7	60.7	28.6
normal.50	3.3	10.0	86.7
CAD.50	6.7	6.7	86.7

C.83 Experiment E3.C4.5

Training Data

	anterior	inferior	normal
anterior	88.6	5.8	5.6
inferior	6.5	85.5	8.0
normal	0.0	0.0	100.0

Testing Data

	anterior	inferior	normal
anterior	73.6	11.9	14.5
inferior	8.6	55.7	35.7
normal	30.0	13.3	56.7

Training Data

	anterior	inferior	normal
anterior.6	94.8	3.1	2.1
anterior.12	98.0	1.0	1.0
anterior.48	100.0	0.0	0.0
anterior.fu	61.5	19.2	19.2
inferior.6	1.1	96.8	2.2
inferior.12	2.4	95.1	2.4
inferior.48	0.0	100.0	0.0
inferior.fu	22.6	50.0	27.4
normal.50	0.0	0.0	100.0
CAD.50	21.6	25.2	53.1

Testing Data

	anterior	inferior	normal
anterior.6	69.0	10.3	20.7
anterior.12	86.2	10.3	3.4
anterior.48	79.3	6.9	13.8
anterior.fu	60.0	20.0	20.0
inferior.6	13.3	53.3	33.3
inferior.12	6.7	50.0	43.3
inferior.48	0.0	76.7	23.3
inferior.fu	14.3	42.9	42.9
normal.50	30.0	13.3	56.7
CAD.50	13.3	26.7	60.0

C.84 Experiment E3.MML

Training Data

	anterior	inferior	normal
anterior	80.6	10.6	8.9
inferior	4.9	81.8	13.4
normal	18.3	11.5	70.2

Testing Data

	anterior	inferior	normal
anterior	75.8	15.3	8.9
inferior	17.0	69.3	13.7
normal	26.7	20.0	53.3

Training Data

	anterior	inferior	normal
anterior.6	84.4	10.4	5.2
anterior.12	86.3	8.8	4.9
anterior.48	97.7	0.0	2.3
anterior.fu	53.8	23.1	23.1
inferior.6	3.2	90.3	6.5
inferior.12	1.2	91.5	7.3
inferior.48	3.8	90.4	5.8
inferior.fu	11.3	54.8	33.9
normal.50	18.3	11.5	70.2
CAD.50	19.3	33.1	47.5

Testing Data

	anterior	inferior	normal
anterior.6	82.8	6.9	10.3
anterior.12	86.2	13.8	0.0
anterior.48	79.3	10.3	10.3
anterior.fu	55.0	30.0	15.0
inferior.6	16.7	80.0	3.3
inferior.12	20.0	66.7	13.3
inferior.48	13.3	73.3	13.3
inferior.fu	17.9	57.1	25.0
normal.50	26.7	20.0	53.3
CAD.50	10.0	40.0	50.0

C.85 Experiment E3.bp

Training Data

	anterior	inferior	normal
anterior	82.4±2.5	7.5±3.5	10.0±3.5
inferior	5.3±2.6	85.6±1.7	9.1±2.6
normal	13.0±10.7	9.5±8.0	77.5±18.1

Testing Data

	anterior	inferior	normal
anterior	70.9±4.0	15.0±4.3	14.1±4.6
inferior	10.3±2.9	75.0±4.9	14.6±5.0
normal	27.7±12.1	8.5±7.5	63.8±17.1

Training Data

	anterior	inferior	normal
anterior.6	78.7±6.8	11.6±8.2	9.7±3.9
anterior.12	90.1±2.8	4.7±1.5	5.2±3.0
anterior.48	94.1±2.2	3.8±1.7	2.2±1.8
anterior.fu	66.7±8.2	10.2±5.3	23.1±10.2
inferior.6	4.5±5.7	88.9±4.9	6.7±2.9
inferior.12	2.0±1.8	92.2±1.8	5.8±2.3
inferior.48	2.7±2.7	94.1±2.7	3.2±2.0
inferior.fu	12.0±3.7	67.2±7.3	20.8±6.6
normal.50	13.0±10.7	9.5±8.0	77.5±18.1
CAD.50	17.4±4.4	30.7±8.5	52.0±12.3

Testing Data

	anterior	inferior	normal
anterior.6	69.3±5.4	14.1±5.1	16.6±7.2
anterior.12	80.2±6.6	7.2±6.2	12.6±5.8
anterior.48	80.9±5.1	8.3±3.7	10.9±4.7
anterior.fu	53.2±9.5	30.2±8.3	16.5±7.8
inferior.6	12.2±6.4	70.8±8.2	17.0±8.0
inferior.12	13.8±3.5	77.5±4.9	8.7±5.0
inferior.48	7.8±6.2	82.5±6.6	9.7±5.6
inferior.fu	7.5±3.2	69.3±9.5	23.2±8.8
normal.50	27.7±12.1	8.5±7.5	63.8±17.1
CAD.50	10.0±6.8	28.0±11.1	62.0±15.2

C.86 Experiment E3.qp

Training Data

	anterior	inferior	normal
anterior	76.7±5.0	7.7±2.9	15.5±3.2
inferior	3.7±1.3	82.9±3.6	13.4±3.4
normal	8.2±3.4	10.1±2.9	81.6±5.0

Testing Data

	anterior	inferior	normal
anterior	66.7±4.4	12.5±3.9	20.8±3.9
inferior	7.8±2.8	75.8±4.7	16.5±4.6
normal	16.7±7.5	8.3±4.7	75.0±9.1

Training Data

	anterior	inferior	normal
anterior.6	64.6±11.2	12.9±6.8	22.5±7.6
anterior.12	85.5±4.2	5.8±1.9	8.7±3.1
anterior.48	88.0±3.4	5.5±2.4	6.6±2.8
anterior.fu	68.8±8.3	6.7±4.0	24.4±7.7
inferior.6	3.5±1.8	85.4±4.6	11.1±4.2
inferior.12	1.6±1.4	89.5±3.7	8.9±3.3
inferior.48	2.2±1.5	89.0±4.6	8.7±5.0
inferior.fu	7.6±3.9	67.7±5.9	24.7±5.7
normal.50	8.2±3.4	10.1±2.9	81.6±5.0
CAD.50	10.0±2.1	28.2±3.8	61.9±4.1

Testing Data

	anterior	inferior	normal
anterior.6	63.6±7.7	11.7±7.4	24.7±7.6
anterior.12	73.4±5.8	4.3±4.2	22.2±5.4
anterior.48	77.8±4.3	6.4±3.5	15.9±4.3
anterior.fu	52.0±7.3	27.5±9.2	20.5±6.5
inferior.6	8.5±4.3	72.3±7.6	19.2±7.7
inferior.12	12.2±4.4	80.3±6.2	7.5±5.4
inferior.48	5.2±3.4	81.5±5.8	13.3±5.3
inferior.fu	5.2±5.0	68.9±7.6	25.9±7.4
normal.50	16.7±7.5	8.3±4.7	75.0±9.1
CAD.50	4.8±3.6	24.0±6.6	71.2±8.0

C.87 Experiment E3.cas

Training Data

	anterior	inferior	normal
anterior	93.3±0.7	0.0±0.0	6.7±0.7
inferior	7.3±1.0	88.0±0.8	4.7±0.8
normal	12.7±1.9	0.0±0.0	87.3±1.9

Testing Data

	anterior	inferior	normal
anterior	91.1±1.6	3.0±1.5	5.9±1.2
inferior	12.2±1.7	79.7±2.6	8.1±1.6
normal	23.3±4.7	12.0±3.4	64.7±5.4

Training Data

	anterior	inferior	normal
anterior.6	100.0±NaN	0.0±0.0	0.0±0.0
anterior.12	100.0±NaN	0.0±0.0	0.0±0.0
anterior.48	100.0±NaN	0.0±0.0	0.0±0.0
anterior.fu	73.1±3.0	0.0±0.0	26.9±3.0
inferior.6	0.0±0.0	100.0±0.0	0.0±0.0
inferior.12	0.0±0.0	100.0±NaN	0.0±0.0
inferior.48	0.0±0.0	100.0±NaN	0.0±0.0
inferior.fu	29.2±4.1	51.9±3.3	18.9±3.1
normal.50	12.7±1.9	0.0±0.0	87.3±1.9
CAD.50	29.4±1.0	17.5±1.8	53.1±1.9

Testing Data

	anterior	inferior	normal
anterior.6	89.7±0.0	3.4±0.0	6.9±0.0
anterior.12	97.6±1.6	0.0±0.0	2.4±1.6
anterior.48	94.1±1.6	2.4±1.6	3.4±0.0
anterior.fu	83.0±5.1	6.0±5.4	11.0±5.4
inferior.6	10.0±2.6	86.3±2.8	3.7±2.8
inferior.12	12.3±1.5	83.7±1.8	4.0±1.3
inferior.48	2.7±1.3	94.0±1.3	3.3±0.0
inferior.fu	23.9±7.7	54.6±9.5	21.4±5.8
normal.50	23.3±4.7	12.0±3.4	64.7±5.4
CAD.50	5.3±3.1	8.7±3.1	86.0±4.7

C.88 Experiment E3.cbp

Training Data

	anterior	inferior	normal
anterior	89.3	4.6	6.1
inferior	2.5	89.5	8.0
normal	9.2	6.1	84.7

Testing Data

	anterior	inferior	normal
anterior	74.5	12.3	13.2
inferior	9.3	82.1	8.6
normal	23.3	3.3	73.3

Training Data

	anterior	inferior	normal
anterior.6	88.5	8.3	3.1
anterior.12	94.1	3.9	2.0
anterior.48	97.7	2.3	0.0
anterior.fu	76.9	3.8	19.2
inferior.6	0.0	96.8	3.2
inferior.12	0.0	97.6	2.4
inferior.48	1.9	94.2	3.8
inferior.fu	8.1	69.4	22.6
normal.50	9.2	6.1	84.7
CAD.50	12.5	25.2	62.3

Testing Data

	anterior	inferior	normal
anterior.6	72.4	13.8	13.8
anterior.12	82.8	3.4	13.8
anterior.48	82.8	6.9	10.3
anterior.fu	60.0	25.0	15.0
inferior.6	10.0	83.3	6.7
inferior.12	16.7	83.3	0.0
inferior.48	3.3	86.7	10.0
inferior.fu	7.1	75.0	17.9
normal.50	23.3	3.3	73.3
CAD.50	3.3	30.0	66.7

C.89 Experiment E3.cqp

Training Data

	anterior	inferior	normal
anterior	82.6	5.4	12.0
inferior	1.3	86.7	12.0
normal	6.9	8.4	84.7

Testing Data

	anterior	inferior	normal
anterior	73.6	11.4	15.0
inferior	5.9	81.2	12.9
normal	10.0	6.7	83.3

Training Data

	anterior	inferior	normal
anterior.6	69.8	11.5	18.7
anterior.12	92.2	3.9	3.9
anterior.48	95.5	2.3	2.3
anterior.fu	73.1	3.8	23.1
inferior.6	0.0	90.3	9.7
inferior.12	0.0	95.1	4.9
inferior.48	1.9	90.4	7.7
inferior.fu	3.2	71.0	25.8
normal.50	6.9	8.4	84.7
CAD.50	7.5	25.6	66.9

Testing Data

	anterior	inferior	normal
anterior.6	69.0	13.8	17.2
anterior.12	82.8	3.4	13.8
anterior.48	82.8	3.4	13.8
anterior.fu	60.0	25.0	15.0
inferior.6	6.7	80.0	13.3
inferior.12	10.0	86.7	3.3
inferior.48	3.3	83.3	13.3
inferior.fu	3.6	75.0	21.4
normal.50	10.0	6.7	83.3
CAD.50	0.0	23.3	76.7

C.90 Experiment E3.ccas

Training Data

	anterior	inferior	normal
anterior	93.3	0.0	6.7
inferior	6.9	87.1	6.0
normal	13.0	0.0	87.0

Testing Data

	anterior	inferior	normal
anterior	88.6	3.0	8.4
inferior	11.2	78.3	10.5
normal	16.7	10.0	73.3

Training Data

	anterior	inferior	normal
anterior.6	100.0	0.0	0.0
anterior.12	100.0	0.0	0.0
anterior.48	100.0	0.0	0.0
anterior.fu	73.1	0.0	26.9
inferior.6	0.0	100.0	0.0
inferior.12	0.0	100.0	0.0
inferior.48	0.0	100.0	0.0
inferior.fu	27.4	48.4	24.2
normal.50	13.0	0.0	87.0
CAD.50	28.2	12.1	59.7

Testing Data

	anterior	inferior	normal
anterior.6	89.7	3.4	6.9
anterior.12	96.6	0.0	3.4
anterior.48	93.1	3.4	3.4
anterior.fu	75.0	5.0	20.0
inferior.6	10.0	86.7	3.3
inferior.12	10.0	83.3	6.7
inferior.48	3.3	93.3	3.3
inferior.fu	21.4	50.0	28.6
normal.50	16.7	10.0	73.3
CAD.50	3.3	3.3	93.3

C.91 Experiment L3.knn

Training Data

	anterior	inferior	normal
anterior	92.0	1.7	6.3
inferior	2.9	90.6	6.5
normal	0.8	0.0	99.2

Testing Data

	anterior	inferior	normal
anterior	78.7	9.7	11.6
inferior	8.4	79.5	12.1
normal	13.3	10.0	76.7

Training Data

	anterior	inferior	normal
anterior.6	100.0	0.0	0.0
anterior.12	97.1	2.9	0.0
anterior.48	97.7	0.0	2.3
anterior.fu	73.1	3.8	23.1
inferior.6	0.0	100.0	0.0
inferior.12	0.0	100.0	0.0
inferior.48	1.9	96.2	1.9
inferior.fu	9.7	66.1	24.2
normal.50	0.8	0.0	99.2
CAD.50	8.9	24.6	66.6

Testing Data

	anterior	inferior	normal
anterior.6	75.9	3.4	20.7
anterior.12	89.7	6.9	3.4
anterior.48	79.3	3.4	17.2
anterior.fu	70.0	25.0	5.0
inferior.6	3.3	90.0	6.7
inferior.12	10.0	83.3	6.7
inferior.48	16.7	73.3	10.0
inferior.fu	3.6	71.4	25.0
normal.50	13.3	10.0	76.7
CAD.50	0.0	16.7	83.3

C.92 Experiment L3.linreg

Training Data

	anterior	inferior	normal
anterior	86.6	4.4	9.0
inferior	5.5	89.1	5.4
normal	1.5	1.5	96.9

Testing Data

	anterior	inferior	normal
anterior	72.8	16.5	10.6
inferior	10.4	71.5	18.1
normal	16.7	10.0	73.3

Training Data

	anterior	inferior	normal
anterior.6	93.7	3.1	3.1
anterior.12	95.1	2.9	2.0
anterior.48	100.0	0.0	0.0
anterior.fu	57.7	11.5	30.8
inferior.6	1.1	97.8	1.1
inferior.12	0.0	98.8	1.2
inferior.48	0.0	100.0	0.0
inferior.fu	21.0	59.7	19.4
normal.50	1.5	1.5	96.9
CAD.50	13.4	16.1	70.5

Testing Data

	anterior	inferior	normal
anterior.6	65.5	17.2	17.2
anterior.12	82.8	10.3	6.9
anterior.48	93.1	3.4	3.4
anterior.fu	50.0	35.0	15.0
inferior.6	3.3	80.0	16.7
inferior.12	6.7	86.7	6.7
inferior.48	10.0	76.7	13.3
inferior.fu	21.4	42.9	35.7
normal.50	16.7	10.0	73.3
CAD.50	3.3	13.3	83.3

C.93 Experiment L3.C4.5

Training Data

	anterior	inferior	normal
anterior	92.0	3.4	4.6
inferior	4.4	84.4	11.2
normal	0.0	0.0	100.0

Testing Data

	anterior	inferior	normal
anterior	79.7	5.1	15.3
inferior	6.2	77.3	16.5
normal	33.3	6.7	60.0

Training Data

	anterior	inferior	normal
anterior.6	99.0	1.0	0.0
anterior.12	96.1	1.0	2.9
anterior.48	100.0	0.0	0.0
anterior.fu	73.1	11.5	15.4
inferior.6	0.0	100.0	0.0
inferior.12	0.0	98.8	1.2
inferior.48	0.0	100.0	0.0
inferior.fu	17.7	38.7	43.5
normal.50	0.0	0.0	100.0
CAD.50	21.6	17.7	60.7

Testing Data

	anterior	inferior	normal
anterior.6	93.1	0.0	6.9
anterior.12	82.8	6.9	10.3
anterior.48	82.8	3.4	13.8
anterior.fu	60.0	10.0	30.0
inferior.6	0.0	93.3	6.7
inferior.12	0.0	90.0	10.0
inferior.48	0.0	90.0	10.0
inferior.fu	25.0	35.7	39.3
normal.50	33.3	6.7	60.0
CAD.50	10.0	16.7	73.3

C.94 Experiment L3.MML

Training Data

	anterior	inferior	normal
anterior	85.7	5.8	8.5
inferior	5.9	85.7	8.5
normal	0.0	0.0	100.0

Testing Data

	anterior	inferior	normal
anterior	63.1	5.0	31.9
inferior	8.7	77.4	13.9
normal	16.7	10.0	73.3

Training Data

	anterior	inferior	normal
anterior.6	95.8	3.1	1.0
anterior.12	93.1	1.0	5.9
anterior.48	100.0	0.0	0.0
anterior.fu	53.8	19.2	26.9
inferior.6	0.0	100.0	0.0
inferior.12	2.4	97.6	0.0
inferior.48	0.0	100.0	0.0
inferior.fu	21.0	45.2	33.9
normal.50	0.0	0.0	100.0
CAD.50	15.7	17.4	66.9

Testing Data

	anterior	inferior	normal
anterior.6	79.3	0.0	20.7
anterior.12	75.9	0.0	24.1
anterior.48	72.4	0.0	27.6
anterior.fu	25.0	20.0	55.0
inferior.6	0.0	93.3	6.7
inferior.12	3.3	86.7	10.0
inferior.48	6.7	86.7	6.7
inferior.fu	25.0	42.9	32.1
normal.50	16.7	10.0	73.3
CAD.50	16.7	13.3	70.0

C.95 Experiment L3.bp

Training Data

	anterior	inferior	normal
anterior	85.2±1.8	5.6±1.5	9.2±1.9
inferior	7.8±1.2	84.5±2.4	7.7±1.7
normal	6.4±2.0	4.7±2.1	88.9±3.0

Testing Data

	anterior	inferior	normal
anterior	76.0±3.7	10.4±2.6	13.6±3.0
inferior	15.4±2.2	69.9±3.8	14.6±3.0
normal	23.8±7.6	11.8±5.4	64.3±8.5

Training Data

	anterior	inferior	normal
anterior.6	90.7±2.1	4.1±1.6	5.2±2.0
anterior.12	92.0±2.2	3.5±1.7	4.6±1.8
anterior.48	93.6±2.9	2.8±2.0	3.5±2.3
anterior.fu	64.4±5.9	12.1±4.9	23.5±6.1
inferior.6	2.3±1.1	95.0±2.4	2.7±1.4
inferior.12	0.7±0.8	95.4±2.1	4.0±1.9
inferior.48	3.9±1.5	94.4±2.0	1.6±1.4
inferior.fu	24.4±4.0	53.1±7.2	22.5±5.4
normal.50	6.4±2.0	4.7±2.1	88.9±3.0
CAD.50	18.3±1.8	23.6±3.7	58.1±4.0

Testing Data

	anterior	inferior	normal
anterior.6	78.8±6.7	7.8±4.5	13.4±4.7
anterior.12	78.8±6.3	7.4±4.0	13.8±5.5
anterior.48	76.4±4.9	8.1±4.7	15.5±4.4
anterior.fu	70.0±7.7	18.5±5.9	11.5±7.3
inferior.6	9.8±4.9	76.0±4.0	14.2±4.8
inferior.12	13.5±5.7	75.8±6.1	10.7±4.8
inferior.48	10.3±3.0	73.7±7.0	16.0±5.9
inferior.fu	28.0±5.3	54.3±5.8	17.7±5.8
normal.50	23.8±7.6	11.8±5.4	64.3±8.5
CAD.50	6.7±3.8	24.7±9.0	68.7±8.4

C.96 Experiment L3.qp

Training Data

	anterior	inferior	normal
anterior	78.8±4.4	7.2±2.7	14.0±2.5
inferior	7.0±1.3	82.7±1.9	10.3±1.5
normal	5.1±1.5	5.1±2.2	89.8±2.9

Testing Data

	anterior	inferior	normal
anterior	69.8±6.1	12.7±3.3	17.5±4.5
inferior	12.3±2.4	70.9±3.6	16.8±2.7
normal	18.3±6.1	10.7±5.0	71.0±6.0

Training Data

	anterior	inferior	normal
anterior.6	78.7±8.1	9.5±4.7	11.7±4.8
anterior.12	86.7±3.1	5.3±2.3	8.0±2.3
anterior.48	90.2±2.9	3.4±2.6	6.4±2.3
anterior.fu	59.6±7.1	10.6±5.1	29.8±6.4
inferior.6	2.8±1.8	92.2±3.4	5.1±2.4
inferior.12	1.3±1.3	91.6±3.0	7.0±2.5
inferior.48	4.6±2.0	91.5±2.7	3.8±1.8
inferior.fu	19.1±4.5	55.4±6.0	25.5±6.0
normal.50	5.1±1.5	5.1±2.2	89.8±2.9
CAD.50	13.7±2.4	21.2±3.2	65.2±2.2

Testing Data

	anterior	inferior	normal
anterior.6	72.6±6.0	9.5±4.2	17.9±5.2
anterior.12	75.5±7.6	9.0±4.8	15.5±6.8
anterior.48	78.1±5.5	7.9±3.6	14.0±4.8
anterior.fu	53.0±12.7	24.5±8.0	22.5±11.8
inferior.6	8.7±4.5	77.8±7.0	13.5±5.4
inferior.12	10.5±3.7	76.8±5.7	12.7±5.5
inferior.48	8.8±3.0	75.2±5.7	16.0±5.0
inferior.fu	21.2±6.7	53.7±8.1	25.0±6.4
normal.50	18.3±6.1	10.7±5.0	71.0±6.0
CAD.50	1.8±2.0	16.5±6.0	81.7±5.5

C.97 Experiment L3.cas

Training Data

	anterior	inferior	normal
anterior	93.6±1.2	1.0±0.7	5.5±0.9
inferior	14.5±0.7	83.7±0.7	1.8±0.4
normal	0.2±0.3	0.0±0.0	99.8±0.3

Testing Data

	anterior	inferior	normal
anterior	91.7±2.1	2.4±0.9	5.9±1.9
inferior	19.8±1.8	72.9±2.2	7.3±1.0
normal	28.7±3.1	6.7±4.2	64.7±4.8

Training Data

	anterior	inferior	normal
anterior.6	100.0±NaN	0.0±0.0	0.0±0.0
anterior.12	100.0±NaN	0.0±0.0	0.0±0.0
anterior.48	100.0±NaN	0.0±0.0	0.0±0.0
anterior.fu	74.2±4.9	3.8±3.0	21.9±3.5
inferior.6	0.0±0.0	100.0±0.0	0.0±0.0
inferior.12	0.0±0.0	100.0±NaN	0.0±0.0
inferior.48	0.0±0.0	100.0±NaN	0.0±0.0
inferior.fu	57.9±2.9	35.0±2.9	7.1±1.8
normal.50	0.2±0.3	0.0±0.0	99.8±0.3
CAD.50	32.7±1.2	8.1±1.6	59.2±1.2

Testing Data

	anterior	inferior	normal
anterior.6	91.7±2.3	0.7±1.4	7.6±2.6
anterior.12	100.0±NaN	0.0±0.0	0.0±0.0
anterior.48	93.1±NaN	3.4±0.0	3.4±0.0
anterior.fu	82.0±7.1	5.5±4.2	12.5±6.4
inferior.6	3.3±3.0	93.3±3.0	3.3±0.0
inferior.12	17.7±4.0	80.0±4.9	2.3±1.5
inferior.48	11.0±1.5	86.7±1.5	2.3±1.5
inferior.fu	47.1±4.5	31.8±3.7	21.1±3.4
normal.50	28.7±3.1	6.7±4.2	64.7±4.8
CAD.50	22.3±3.0	11.0±3.7	66.7±1.5

C.98 Experiment L3.cbp

Training Data

	anterior	inferior	normal
anterior	90.8	2.0	7.2
inferior	6.9	88.0	5.1
normal	0.8	0.0	99.2

Testing Data

	anterior	inferior	normal
anterior	81.7	8.1	10.3
inferior	13.8	76.0	10.2
normal	20.0	3.3	76.7

Training Data

	anterior	inferior	normal
anterior.6	95.8	3.1	1.0
anterior.12	98.0	1.0	1.0
anterior.48	100.0	0.0	0.0
anterior.fu	69.2	3.8	26.9
inferior.6	1.1	98.9	0.0
inferior.12	0.0	98.8	1.2
inferior.48	3.8	96.2	0.0
inferior.fu	22.6	58.1	19.4
normal.50	0.8	0.0	99.2
CAD.50	16.1	20.7	63.3

Testing Data

	anterior	inferior	normal
anterior.6	82.8	6.9	10.3
anterior.12	86.2	3.4	10.3
anterior.48	82.8	6.9	10.3
anterior.fu	75.0	15.0	10.0
inferior.6	6.7	86.7	6.7
inferior.12	10.0	80.0	10.0
inferior.48	10.0	80.0	10.0
inferior.fu	28.6	57.1	14.3
normal.50	20.0	3.3	76.7
CAD.50	3.3	23.3	73.3

C.99 Experiment L3.cqp

Training Data

	anterior	inferior	normal
anterior	82.1	5.8	12.1
inferior	5.3	86.8	8.0
normal	2.3	0.8	96.9

Testing Data

	anterior	inferior	normal
anterior	76.6	12.7	10.7
inferior	10.3	78.6	11.1
normal	13.3	3.3	83.3

Training Data

	anterior	inferior	normal
anterior.6	85.4	7.3	7.3
anterior.12	88.2	5.9	5.9
anterior.48	93.2	2.3	4.5
anterior.fu	61.5	7.7	30.8
inferior.6	1.1	96.8	2.2
inferior.12	0.0	96.3	3.7
inferior.48	3.8	94.2	1.9
inferior.fu	16.1	59.7	24.2
normal.50	2.3	0.8	96.9
CAD.50	11.5	19.7	68.9

Testing Data

	anterior	inferior	normal
anterior.6	75.9	6.9	17.2
anterior.12	82.8	6.9	10.3
anterior.48	82.8	6.9	10.3
anterior.fu	65.0	30.0	5.0
inferior.6	6.7	86.7	6.7
inferior.12	6.7	86.7	6.7
inferior.48	10.0	76.7	13.3
inferior.fu	17.9	64.3	17.9
normal.50	13.3	3.3	83.3
CAD.50	0.0	6.7	93.3

C.100 Experiment L3.ccas

Training Data

	anterior	inferior	normal
anterior	92.3	1.0	6.7
inferior	12.9	84.7	2.4
normal	0.0	0.0	100.0

Testing Data

	anterior	inferior	normal
anterior	89.4	2.1	8.4
inferior	16.5	74.7	8.7
normal	26.7	3.3	70.0

Training Data

	anterior	inferior	normal
anterior.6	100.0	0.0	0.0
anterior.12	100.0	0.0	0.0
anterior.48	100.0	0.0	0.0
anterior.fu	69.2	3.8	26.9
inferior.6	0.0	100.0	0.0
inferior.12	0.0	100.0	0.0
inferior.48	0.0	100.0	0.0
inferior.fu	51.6	38.7	9.7
normal.50	0.0	0.0	100.0
CAD.50	30.2	4.6	65.2

Testing Data

	anterior	inferior	normal
anterior.6	89.7	0.0	10.3
anterior.12	100.0	0.0	0.0
anterior.48	93.1	3.4	3.4
anterior.fu	75.0	5.0	20.0
inferior.6	0.0	96.7	3.3
inferior.12	13.3	83.3	3.3
inferior.48	10.0	86.7	3.3
inferior.fu	42.9	32.1	25.0
normal.50	26.7	3.3	70.0
CAD.50	20.0	3.3	76.7

C.101 Experiment st3.knn

Training Data

	anterior	inferior	normal
anterior	84.4	5.4	10.2
inferior	6.7	79.1	14.2
normal	1.5	0.8	97.7

Testing Data

	anterior	inferior	normal
anterior	69.6	6.3	24.1
inferior	3.5	73.6	22.9
normal	20.0	6.7	73.3

Training Data

	anterior	inferior	normal
anterior.6	100.0	0.0	0.0
anterior.12	100.0	0.0	0.0
anterior.48	95.5	2.3	2.3
anterior.fu	42.3	19.2	38.5
inferior.6	1.1	98.9	0.0
inferior.12	0.0	100.0	0.0
inferior.48	0.0	98.1	1.9
inferior.fu	25.8	19.4	54.8
normal.50	1.5	0.8	97.7
CAD.50	17.7	9.2	73.1

Testing Data

	anterior	inferior	normal
anterior.6	82.8	3.4	13.8
anterior.12	79.3	0.0	20.7
anterior.48	86.2	6.9	6.9
anterior.fu	30.0	15.0	55.0
inferior.6	0.0	93.3	6.7
inferior.12	3.3	90.0	6.7
inferior.48	0.0	93.3	6.7
inferior.fu	10.7	17.9	71.4
normal.50	20.0	6.7	73.3
CAD.50	16.7	3.3	80.0

C.102 Experiment st3.linreg

Training Data

	anterior	inferior	normal
anterior	69.3	6.8	23.9
inferior	7.4	70.0	22.6
normal	6.1	6.9	87.0

Testing Data

	anterior	inferior	normal
anterior	68.1	8.8	23.0
inferior	9.7	58.6	31.7
normal	10.0	10.0	80.0

Training Data

	anterior	inferior	normal
anterior.6	77.1	1.0	21.9
anterior.12	82.4	6.9	10.8
anterior.48	90.9	0.0	9.1
anterior.fu	26.9	19.2	53.8
inferior.6	2.2	82.8	15.1
inferior.12	2.4	81.7	15.9
inferior.48	3.8	86.5	9.6
inferior.fu	21.0	29.0	50.0
normal.50	6.1	6.9	87.0
CAD.50	14.1	8.5	77.4

Testing Data

	anterior	inferior	normal
anterior.6	82.8	3.4	13.8
anterior.12	79.3	3.4	17.2
anterior.48	65.5	3.4	31.0
anterior.fu	45.0	25.0	30.0
inferior.6	6.7	80.0	13.3
inferior.12	0.0	70.0	30.0
inferior.48	0.0	70.0	30.0
inferior.fu	32.1	14.3	53.6
normal.50	10.0	10.0	80.0
CAD.50	6.7	3.3	90.0

C.103 Experiment st3.C4.5

Training Data

	anterior	inferior	normal
anterior	80.5	7.7	11.8
inferior	6.4	81.8	11.9
normal	0.0	0.0	100.0

Testing Data

	anterior	inferior	normal
anterior	60.6	12.6	26.9
inferior	10.5	71.1	18.4
normal	30.0	16.7	53.3

Training Data

	anterior	inferior	normal
anterior.6	100.0	0.0	0.0
anterior.12	95.1	3.9	1.0
anterior.48	100.0	0.0	0.0
anterior.fu	26.9	26.9	46.2
inferior.6	0.0	98.9	1.1
inferior.12	1.2	97.6	1.2
inferior.48	0.0	100.0	0.0
inferior.fu	24.2	30.6	45.2
normal.50	0.0	0.0	100.0
CAD.50	22.3	15.7	62.0

Testing Data

	anterior	inferior	normal
anterior.6	75.9	0.0	24.1
anterior.12	72.4	0.0	27.6
anterior.48	69.0	10.3	20.7
anterior.fu	25.0	40.0	35.0
inferior.6	3.3	93.3	3.3
inferior.12	10.0	83.3	6.7
inferior.48	0.0	90.0	10.0
inferior.fu	28.6	17.9	53.6
normal.50	30.0	16.7	53.3
CAD.50	20.0	13.3	66.7

C.104 Experiment st3.MML

Training Data

	anterior	inferior	normal
anterior	76.6	4.8	18.6
inferior	6.5	78.5	15.0
normal	0.8	0.8	98.5

Testing Data

	anterior	inferior	normal
anterior	60.6	10.6	28.9
inferior	4.5	71.2	24.3
normal	23.3	20.0	56.7

Training Data

	anterior	inferior	normal
anterior.6	89.6	0.0	10.4
anterior.12	88.2	0.0	11.8
anterior.48	97.7	0.0	2.3
anterior.fu	30.8	19.2	50.0
inferior.6	1.1	95.7	3.2
inferior.12	2.4	96.3	1.2
inferior.48	0.0	96.2	3.8
inferior.fu	22.6	25.8	51.6
normal.50	0.8	0.8	98.5
CAD.50	17.0	14.4	68.5

Testing Data

	anterior	inferior	normal
anterior.6	75.9	0.0	24.1
anterior.12	69.0	6.9	24.1
anterior.48	72.4	10.3	17.2
anterior.fu	25.0	25.0	50.0
inferior.6	0.0	96.7	3.3
inferior.12	0.0	80.0	20.0
inferior.48	0.0	83.3	16.7
inferior.fu	17.9	25.0	57.1
normal.50	23.3	20.0	56.7
CAD.50	20.0	10.0	70.0

C.105 Experiment st3.bp

Training Data

	anterior	inferior	normal
anterior	72.6±5.6	8.9±3.5	18.5±4.9
inferior	14.7±2.9	70.0±4.5	15.3±3.8
normal	20.6±5.2	9.4±3.5	70.0±4.8

Testing Data

	anterior	inferior	normal
anterior	64.8±8.4	11.5±3.5	23.7±7.3
inferior	16.5±3.6	67.4±3.3	16.1±3.5
normal	35.2±13.1	19.7±4.7	45.2±11.4

Training Data

	anterior	inferior	normal
anterior.6	81.7±5.4	3.0±1.7	15.3±5.2
anterior.12	77.1±6.9	6.4±3.4	16.5±6.1
anterior.48	86.0±6.0	3.5±2.4	10.5±5.6
anterior.fu	45.6±12.9	22.7±9.1	31.7±9.5
inferior.6	3.2±2.1	91.6±3.9	5.2±2.6
inferior.12	6.3±3.2	83.5±6.0	10.2±5.7
inferior.48	8.7±3.9	79.7±6.6	11.5±4.6
inferior.fu	40.5±8.5	25.1±5.1	34.4±8.5
normal.50	20.6±5.2	9.4±3.5	70.0±4.8
CAD.50	34.7±5.2	18.4±4.0	46.8±4.2

Testing Data

	anterior	inferior	normal
anterior.6	75.7±12.7	4.7±3.8	19.7±13.4
anterior.12	79.1±10.1	4.7±3.1	16.2±9.1
anterior.48	65.0±13.3	10.2±5.6	24.8±10.4
anterior.fu	39.5±9.7	26.5±8.8	34.0±8.6
inferior.6	5.0±2.2	87.7±3.7	7.3±3.7
inferior.12	10.2±4.9	80.8±5.6	9.0±5.3
inferior.48	7.2±3.5	80.0±4.9	12.8±5.0
inferior.fu	43.6±11.2	21.1±6.7	35.4±10.6
normal.50	35.2±13.1	19.7±4.7	45.2±11.4
CAD.50	34.3±8.3	17.8±5.4	47.8±9.0

C.106 Experiment st3.qp

Training Data

	anterior	inferior	normal
anterior	74.0±4.3	7.6±2.6	18.4±3.8
inferior	11.4±3.6	73.2±4.1	15.5±5.4
normal	16.8±6.4	9.7±5.1	73.5±10.1

Testing Data

	anterior	inferior	normal
anterior	67.9±5.9	11.8±3.5	20.3±5.6
inferior	14.0±2.9	68.8±2.3	17.2±3.1
normal	31.8±11.8	18.7±6.9	49.5±13.6

Training Data

	anterior	inferior	normal
anterior.6	80.4±5.4	2.9±1.3	16.7±6.0
anterior.12	81.6±5.6	5.3±3.2	13.1±5.3
anterior.48	89.8±5.6	1.7±1.9	8.5±5.2
anterior.fu	44.2±7.8	20.6±7.1	35.2±7.7
inferior.6	1.2±1.1	92.3±5.0	6.5±5.1
inferior.12	3.0±3.0	87.2±5.5	9.8±6.1
inferior.48	4.2±3.6	84.3±5.6	11.4±5.3
inferior.fu	36.9±9.5	28.9±4.8	34.2±9.5
normal.50	16.8±6.4	9.7±5.1	73.5±10.1
CAD.50	30.1±7.2	17.8±4.1	52.0±8.8

Testing Data

	anterior	inferior	normal
anterior.6	77.1±8.4	3.3±2.3	19.7±8.9
anterior.12	79.5±7.1	6.2±4.2	14.3±6.0
anterior.48	72.9±8.5	9.1±3.7	17.9±8.2
anterior.fu	42.2±9.7	28.5±8.7	29.2±10.6
inferior.6	5.0±2.5	88.0±4.9	7.0±4.5
inferior.12	6.5±3.6	83.2±4.1	10.3±6.1
inferior.48	5.3±3.4	82.8±4.3	11.8±4.7
inferior.fu	39.1±8.7	21.1±5.5	39.8±10.0
normal.50	31.8±11.8	18.7±6.9	49.5±13.6
CAD.50	32.5±10.4	19.2±4.7	48.3±11.3

C.107 Experiment st3.cas

Training Data

	anterior	inferior	normal
anterior	85.6±1.3	4.3±1.1	10.2±1.3
inferior	15.3±1.5	77.7±1.8	7.0±1.0
normal	27.0±4.6	3.7±2.2	69.2±4.3

Testing Data

	anterior	inferior	normal
anterior	87.6±2.3	6.1±1.8	6.3±1.3
inferior	22.1±1.5	71.3±1.6	6.6±1.9
normal	56.0±4.4	2.3±2.1	41.7±4.8

Training Data

	anterior	inferior	normal
anterior.6	94.0±1.9	0.5±0.8	5.5±1.7
anterior.12	97.7±1.4	0.9±1.1	1.4±0.9
anterior.48	96.4±2.3	0.2±0.7	3.4±2.1
anterior.fu	54.2±5.8	15.4±3.4	30.4±6.1
inferior.6	7.8±2.5	89.6±2.5	2.6±1.0
inferior.12	4.5±2.0	92.1±2.5	3.4±1.2
inferior.48	4.4±2.3	92.3±2.7	3.3±1.5
inferior.fu	44.4±3.1	36.9±2.7	18.7±3.1
normal.50	27.0±4.6	3.7±2.2	69.2±4.3
CAD.50	43.1±2.7	10.2±1.2	46.7±2.1

Testing Data

	anterior	inferior	normal
anterior.6	95.2±3.5	0.3±1.0	4.5±3.1
anterior.12	99.0±1.6	0.7±1.4	0.3±1.0
anterior.48	93.1±3.1	3.4±0.0	3.4±3.1
anterior.fu	63.0±7.5	20.0±6.7	17.0±5.1
inferior.6	9.7±2.3	88.7±2.2	1.7±1.7
inferior.12	11.0±3.3	85.0±3.1	4.0±2.5
inferior.48	5.0±3.4	88.0±2.7	7.0±2.3
inferior.fu	62.9±4.6	23.6±4.3	13.6±6.1
normal.50	56.0±4.4	2.3±2.1	41.7±4.8
CAD.50	45.0±6.4	12.0±1.6	43.0±5.9

C.108 Experiment st3.cbp

Training Data

	anterior	inferior	normal
anterior	82.6	5.7	11.8
inferior	14.2	73.5	12.3
normal	16.0	3.1	80.9

Testing Data

	anterior	inferior	normal
anterior	79.3	8.4	12.3
inferior	17.4	67.8	14.8
normal	33.3	20.0	46.7

Training Data

	anterior	inferior	normal
anterior.6	92.7	1.0	6.2
anterior.12	88.2	3.9	7.8
anterior.48	95.5	2.3	2.3
anterior.fu	53.8	15.4	30.8
inferior.6	4.3	93.5	2.2
inferior.12	4.9	90.2	4.9
inferior.48	5.8	82.7	11.5
inferior.fu	41.9	27.4	30.6
normal.50	16.0	3.1	80.9
CAD.50	31.1	17.0	51.8

Testing Data

	anterior	inferior	normal
anterior.6	86.2	3.4	10.3
anterior.12	96.6	3.4	0.0
anterior.48	79.3	6.9	13.8
anterior.fu	55.0	20.0	25.0
inferior.6	3.3	86.7	10.0
inferior.12	13.3	86.7	0.0
inferior.48	6.7	80.0	13.3
inferior.fu	46.4	17.9	35.7
normal.50	33.3	20.0	46.7
CAD.50	26.7	16.7	56.7

C.109 Experiment st3.cqp

Training Data

	anterior	inferior	normal
anterior	82.4	6.6	11.0
inferior	10.6	78.2	11.2
normal	11.5	8.4	80.2

Testing Data

	anterior	inferior	normal
anterior	78.9	10.6	10.6
inferior	16.6	69.5	13.9
normal	30.0	20.0	50.0

Training Data

	anterior	inferior	normal
anterior.6	90.6	1.0	8.3
anterior.12	91.2	3.9	4.9
anterior.48	97.7	2.3	0.0
anterior.fu	50.0	19.2	30.8
inferior.6	0.0	97.8	2.2
inferior.12	0.0	96.3	3.7
inferior.48	1.9	86.5	11.5
inferior.fu	40.3	32.3	27.4
normal.50	11.5	8.4	80.2
CAD.50	27.9	15.4	56.7

Testing Data

	anterior	inferior	normal
anterior.6	89.7	3.4	6.9
anterior.12	89.7	6.9	3.4
anterior.48	86.2	6.9	6.9
anterior.fu	50.0	25.0	25.0
inferior.6	3.3	90.0	6.7
inferior.12	10.0	86.7	3.3
inferior.48	6.7	83.3	10.0
inferior.fu	46.4	17.9	35.7
normal.50	30.0	20.0	50.0
CAD.50	26.7	16.7	56.7

C.110 Experiment st3.ccas

Training Data

	anterior	inferior	normal
anterior	81.6	5.1	13.3
inferior	11.8	79.8	8.4
normal	22.1	3.1	74.8

Testing Data

	anterior	inferior	normal
anterior	88.3	5.9	5.9
inferior	19.1	70.5	10.4
normal	40.0	6.7	53.3

Training Data

	anterior	inferior	normal
anterior.6	91.7	0.0	8.3
anterior.12	97.1	1.0	2.0
anterior.48	95.5	0.0	4.5
anterior.fu	42.3	19.2	38.5
inferior.6	5.4	92.5	2.2
inferior.12	2.4	93.9	3.7
inferior.48	3.8	94.2	1.9
inferior.fu	35.5	38.7	25.8
normal.50	22.1	3.1	74.8
CAD.50	39.0	9.2	51.8

Testing Data

	anterior	inferior	normal
anterior.6	96.6	0.0	3.4
anterior.12	100.0	0.0	0.0
anterior.48	96.6	3.4	0.0
anterior.fu	60.0	20.0	20.0
inferior.6	13.3	86.7	0.0
inferior.12	10.0	80.0	10.0
inferior.48	6.7	83.3	10.0
inferior.fu	46.4	32.1	21.4
normal.50	40.0	6.7	53.3
CAD.50	43.3	13.3	43.3

C.111 Experiment qrs3.knn

Training Data

	anterior	inferior	normal
anterior	92.2	2.4	5.4
inferior	2.5	93.4	4.1
normal	0.8	0.0	99.2

Testing Data

	anterior	inferior	normal
anterior	80.0	11.4	8.6
inferior	9.2	82.2	8.6
normal	13.3	16.7	70.0

Training Data

	anterior	inferior	normal
anterior.6	100.0	0.0	0.0
anterior.12	98.0	2.0	0.0
anterior.48	97.7	0.0	2.3
anterior.fu	73.1	7.7	19.2
inferior.6	0.0	100.0	0.0
inferior.12	0.0	100.0	0.0
inferior.48	1.9	96.2	1.9
inferior.fu	8.1	77.4	14.5
normal.50	0.8	0.0	99.2
CAD.50	8.9	33.8	57.4

Testing Data

	anterior	inferior	normal
anterior.6	75.9	3.4	20.7
anterior.12	89.7	6.9	3.4
anterior.48	79.3	10.3	10.3
anterior.fu	75.0	25.0	0.0
inferior.6	6.7	86.7	6.7
inferior.12	10.0	83.3	6.7
inferior.48	16.7	76.7	6.7
inferior.fu	3.6	82.1	14.3
normal.50	13.3	16.7	70.0
CAD.50	0.0	26.7	73.3

C.112 Experiment qrs3.linreg

Training Data

	anterior	inferior	normal
anterior	84.2	6.5	9.3
inferior	6.5	88.6	4.8
normal	2.3	1.5	96.2

Testing Data

	anterior	inferior	normal
anterior	74.6	18.2	7.2
inferior	15.2	70.0	14.8
normal	16.7	6.7	76.7

Training Data

	anterior	inferior	normal
anterior.6	85.4	7.3	7.3
anterior.12	88.2	8.8	2.9
anterior.48	97.7	2.3	0.0
anterior.fu	65.4	7.7	26.9
inferior.6	3.2	93.5	3.2
inferior.12	1.2	93.9	4.9
inferior.48	3.8	96.2	0.0
inferior.fu	17.7	71.0	11.3
normal.50	2.3	1.5	96.2
CAD.50	9.8	20.3	69.8

Testing Data

	anterior	inferior	normal
anterior.6	69.0	24.1	6.9
anterior.12	89.7	6.9	3.4
anterior.48	89.7	6.9	3.4
anterior.fu	50.0	35.0	15.0
inferior.6	16.7	70.0	13.3
inferior.12	23.3	76.7	0.0
inferior.48	10.0	83.3	6.7
inferior.fu	10.7	50.0	39.3
normal.50	16.7	6.7	76.7
CAD.50	6.7	6.7	86.7

C.113 Experiment qrs3.C4.5

Training Data

	anterior	inferior	normal
anterior	91.3	2.4	6.3
inferior	2.6	90.9	6.5
normal	0.0	0.0	100.0

Testing Data

	anterior	inferior	normal
anterior	73.6	12.4	14.0
inferior	7.6	78.1	14.3
normal	10.0	20.0	70.0

Training Data

	anterior	inferior	normal
anterior.6	97.9	0.0	2.1
anterior.12	98.0	2.0	0.0
anterior.48	100.0	0.0	0.0
anterior.fu	69.2	7.7	23.1
inferior.6	0.0	100.0	0.0
inferior.12	2.4	97.6	0.0
inferior.48	0.0	100.0	0.0
inferior.fu	8.1	66.1	25.8
normal.50	0.0	0.0	100.0
CAD.50	16.1	30.2	53.8

Testing Data

	anterior	inferior	normal
anterior.6	79.3	6.9	13.8
anterior.12	79.3	17.2	3.4
anterior.48	75.9	10.3	13.8
anterior.fu	60.0	15.0	25.0
inferior.6	6.7	70.0	23.3
inferior.12	13.3	73.3	13.3
inferior.48	3.3	83.3	13.3
inferior.fu	7.1	85.7	7.1
normal.50	10.0	20.0	70.0
CAD.50	20.0	23.3	56.7

C.114 Experiment qrs3.MML

Training Data

	anterior	inferior	normal
anterior	89.3	5.5	5.2
inferior	7.3	87.6	5.0
normal	0.8	0.0	99.2

Testing Data

	anterior	inferior	normal
anterior	84.2	5.9	9.9
inferior	11.8	77.1	11.1
normal	43.3	16.7	40.0

Training Data

	anterior	inferior	normal
anterior.6	87.5	4.2	8.3
anterior.12	99.0	0.0	1.0
anterior.48	97.7	2.3	0.0
anterior.fu	73.1	15.4	11.5
inferior.6	3.2	87.1	9.7
inferior.12	6.1	91.5	2.4
inferior.48	3.8	96.2	0.0
inferior.fu	16.1	75.8	8.1
normal.50	0.8	0.0	99.2
CAD.50	26.9	21.3	51.8

Testing Data

	anterior	inferior	normal
anterior.6	86.2	0.0	13.8
anterior.12	82.8	10.3	6.9
anterior.48	82.8	3.4	13.8
anterior.fu	85.0	10.0	5.0
inferior.6	10.0	70.0	20.0
inferior.12	20.0	80.0	0.0
inferior.48	10.0	83.3	6.7
inferior.fu	7.1	75.0	17.9
normal.50	43.3	16.7	40.0
CAD.50	16.7	16.7	66.7

C.115 Experiment qrs3.bp

Training Data

	anterior	inferior	normal
anterior	84.3±2.5	6.7±1.9	8.9±1.7
inferior	8.3±1.5	83.4±2.3	8.3±2.0
normal	5.5±2.1	5.5±2.3	89.0±3.3

Testing Data

	anterior	inferior	normal
anterior	71.6±4.3	13.2±3.2	15.1±3.0
inferior	17.8±3.5	65.3±4.2	16.9±4.5
normal	26.0±7.4	12.0±6.8	62.0±8.9

Training Data

	anterior	inferior	normal
anterior.6	87.0±2.7	7.6±2.5	5.4±1.9
anterior.12	91.2±2.4	4.2±2.0	4.6±1.6
anterior.48	94.9±3.1	1.8±2.2	3.3±2.7
anterior.fu	64.2±7.8	13.3±6.4	22.5±6.2
inferior.6	5.2±2.0	88.5±3.5	6.3±2.9
inferior.12	2.0±1.5	91.8±2.6	6.3±2.1
inferior.48	4.3±1.8	92.9±2.4	2.8±1.9
inferior.fu	21.9±4.4	60.3±6.4	17.8±5.2
normal.50	5.5±2.1	5.5±2.3	89.0±3.3
CAD.50	17.1±2.1	28.7±3.5	54.2±3.8

Testing Data

	anterior	inferior	normal
anterior.6	70.5±6.2	10.7±5.2	18.8±6.5
anterior.12	75.7±7.5	11.0±3.9	13.3±6.3
anterior.48	74.5±5.4	10.5±4.3	15.0±4.8
anterior.fu	65.8±8.6	20.8±9.1	13.5±8.1
inferior.6	16.3±6.6	60.3±8.7	23.3±7.9
inferior.12	18.8±5.2	72.5±6.8	8.7±4.8
inferior.48	14.3±3.8	69.5±6.8	16.2±6.3
inferior.fu	21.8±6.5	58.7±8.5	19.5±7.8
normal.50	26.0±7.4	12.0±6.8	62.0±8.9
CAD.50	6.8±3.6	26.5±6.4	66.7±7.1

C.116 Experiment qrs3.qp

Training Data

	anterior	inferior	normal
anterior	71.9±3.7	9.9±2.6	18.3±2.9
inferior	9.8±2.1	77.8±4.6	12.4±3.6
normal	7.8±2.5	9.2±3.2	83.0±4.0

Testing Data

	anterior	inferior	normal
anterior	58.9±6.2	14.6±4.6	26.5±5.2
inferior	14.9±3.2	64.9±5.6	20.3±4.6
normal	20.5±9.2	14.7±7.3	64.8±8.3

Training Data

	anterior	inferior	normal
anterior.6	64.8±7.2	12.7±4.0	22.5±5.1
anterior.12	81.3±3.1	6.9±1.9	11.8±3.0
anterior.48	86.8±2.9	3.5±3.0	9.7±2.9
anterior.fu	54.6±4.6	16.3±5.6	29.0±5.4
inferior.6	8.9±3.1	79.8±5.6	11.3±3.9
inferior.12	4.8±2.2	85.1±3.1	10.1±2.7
inferior.48	6.3±1.5	87.0±5.1	6.6±4.1
inferior.fu	19.0±4.0	59.4±7.7	21.6±7.4
normal.50	7.8±2.5	9.2±3.2	83.0±4.0
CAD.50	11.6±2.0	26.8±3.4	61.6±3.6

Testing Data

	anterior	inferior	normal
anterior.6	62.4±6.6	14.5±6.9	23.1±6.2
anterior.12	62.4±8.7	11.4±5.8	26.2±6.4
anterior.48	73.6±5.3	7.9±5.1	18.4±5.3
anterior.fu	37.0±12.1	24.8±10.7	38.2±10.5
inferior.6	15.5±4.5	60.7±8.5	23.8±7.5
inferior.12	13.5±3.4	70.8±7.9	15.7±6.4
inferior.48	12.7±3.3	69.3±7.1	18.0±6.2
inferior.fu	17.9±6.4	58.6±7.4	23.6±7.3
normal.50	20.5±9.2	14.7±7.3	64.8±8.3
CAD.50	4.5±2.4	23.8±8.9	71.7±8.9

C.117 Experiment qrs3.cas

Training Data

	anterior	inferior	normal
anterior	92.3±0.5	1.6±0.7	6.1±1.0
inferior	20.3±2.0	77.6±2.5	2.0±1.0
normal	3.7±1.7	1.0±1.1	95.3±2.2

Testing Data

	anterior	inferior	normal
anterior	83.9±1.1	8.4±1.0	7.8±1.0
inferior	42.7±2.2	52.8±2.0	4.5±1.0
normal	30.7±4.2	5.0±1.7	64.3±4.2

Training Data

	anterior	inferior	normal
anterior.6	94.2±1.9	3.1±2.1	2.7±0.7
anterior.12	98.6±0.7	0.4±0.5	1.0±0.6
anterior.48	99.5±0.9	0.5±0.9	0.0±0.0
anterior.fu	76.9±2.4	2.3±1.9	20.8±3.5
inferior.6	9.8±3.3	88.1±4.0	2.2±1.4
inferior.12	10.7±3.3	88.5±3.8	0.7±1.1
inferior.48	8.3±2.4	91.5±2.6	0.2±0.6
inferior.fu	52.6±1.6	42.4±1.6	5.0±1.5
normal.50	3.7±1.7	1.0±1.1	95.3±2.2
CAD.50	27.5±1.0	17.2±2.0	55.3±1.7

Testing Data

	anterior	inferior	normal
anterior.6	82.4±2.9	6.9±2.2	10.7±2.9
anterior.12	94.8±2.3	4.8±2.3	0.3±1.0
anterior.48	88.3±1.7	1.7±1.7	10.0±1.0
anterior.fu	70.0±0.0	20.0±0.0	10.0±0.0
inferior.6	48.7±4.5	44.3±5.0	7.0±2.3
inferior.12	45.7±2.6	54.3±2.6	0.0±0.0
inferior.48	26.7±3.7	72.7±3.3	0.7±1.3
inferior.fu	49.6±3.0	40.0±2.7	10.4±1.9
normal.50	30.7±4.2	5.0±1.7	64.3±4.2
CAD.50	12.3±3.7	14.3±3.7	73.3±3.7

C.118 Experiment qrs3.cbp

Training Data

	anterior	inferior	normal
anterior	90.5	3.5	6.0
inferior	7.7	89.5	2.8
normal	2.3	0.8	96.9

Testing Data

	anterior	inferior	normal
anterior	80.3	9.8	9.9
inferior	17.0	68.5	14.5
normal	23.3	3.3	73.3

Training Data

	anterior	inferior	normal
anterior.6	95.8	4.2	0.0
anterior.12	97.1	2.0	1.0
anterior.48	100.0	0.0	0.0
anterior.fu	69.2	7.7	23.1
inferior.6	3.2	94.6	2.2
inferior.12	1.2	96.3	2.4
inferior.48	3.8	96.2	0.0
inferior.fu	22.6	71.0	6.5
normal.50	2.3	0.8	96.9
CAD.50	11.8	25.6	62.6

Testing Data

	anterior	inferior	normal
anterior.6	79.3	6.9	13.8
anterior.12	86.2	6.9	6.9
anterior.48	75.9	10.3	13.8
anterior.fu	80.0	15.0	5.0
inferior.6	13.3	63.3	23.3
inferior.12	20.0	76.7	3.3
inferior.48	13.3	76.7	10.0
inferior.fu	21.4	57.1	21.4
normal.50	23.3	3.3	73.3
CAD.50	6.7	20.0	73.3

C.119 Experiment qrs3.cqp

Training Data

	anterior	inferior	normal
anterior	75.2	9.7	15.1
inferior	9.5	82.9	7.6
normal	5.3	5.3	89.3

Testing Data

	anterior	inferior	normal
anterior	60.5	12.3	27.2
inferior	17.0	67.7	15.3
normal	16.7	10.0	73.3

Training Data

	anterior	inferior	normal
anterior.6	70.8	10.4	18.7
anterior.12	85.3	6.9	7.8
anterior.48	90.9	2.3	6.8
anterior.fu	53.8	19.2	26.9
inferior.6	7.5	86.0	6.5
inferior.12	3.7	89.0	7.3
inferior.48	5.8	90.4	3.8
inferior.fu	21.0	66.1	12.9
normal.50	5.3	5.3	89.3
CAD.50	10.8	23.9	65.2

Testing Data

	anterior	inferior	normal
anterior.6	65.5	6.9	27.6
anterior.12	65.5	6.9	27.6
anterior.48	75.9	10.3	13.8
anterior.fu	35.0	25.0	40.0
inferior.6	16.7	60.0	23.3
inferior.12	16.7	73.3	10.0
inferior.48	13.3	76.7	10.0
inferior.fu	21.4	60.7	17.9
normal.50	16.7	10.0	73.3
CAD.50	3.3	20.0	76.7

C.120 Experiment qrs3.ccas

Training Data

	anterior	inferior	normal
anterior	90.0	0.8	9.2
inferior	14.1	82.8	3.1
normal	2.3	0.0	97.7

Testing Data

	anterior	inferior	normal
anterior	83.9	8.4	7.7
inferior	34.0	59.9	6.1
normal	30.0	3.3	66.7

Training Data

	anterior	inferior	normal
anterior.6	93.7	3.1	3.1
anterior.12	97.1	0.0	2.9
anterior.48	100.0	0.0	0.0
anterior.fu	69.2	0.0	30.8
inferior.6	5.4	91.4	3.2
inferior.12	4.9	93.9	1.2
inferior.48	5.8	94.2	0.0
inferior.fu	40.3	51.6	8.1
normal.50	2.3	0.0	97.7
CAD.50	24.6	14.4	61.0

Testing Data

	anterior	inferior	normal
anterior.6	86.2	3.4	10.3
anterior.12	89.7	10.3	0.0
anterior.48	89.7	0.0	10.3
anterior.fu	70.0	20.0	10.0
inferior.6	36.7	53.3	10.0
inferior.12	36.7	63.3	0.0
inferior.48	20.0	80.0	0.0
inferior.fu	42.9	42.9	14.3
normal.50	30.0	3.3	66.7
CAD.50	10.0	13.3	76.7

C.121 Experiment E4.knn

Training Data

	normal	CAD
normal	97.7	2.3
CAD	0.0	100.0

Testing Data

	normal	CAD
normal	33.3	66.7
CAD	16.7	83.3

C.122 Experiment E4.linreg

Training Data

	normal	CAD
normal	71.0	29.0
CAD	27.5	72.5

Testing Data

	normal	CAD
normal	33.3	66.7
CAD	26.7	73.3

C.123 Experiment E4.C4.5

Training Data

	normal	CAD
normal	100.0	0.0
CAD	1.0	99.0

Testing Data

	normal	CAD
normal	33.3	66.7
CAD	10.0	90.0

C.124 Experiment E4.MML

Training Data

	normal	CAD
normal	99.2	0.8
CAD	39.7	60.3

Testing Data

	normal	CAD
normal	63.3	36.7
CAD	56.7	43.3

C.125 Experiment E4.bp

Training Data

	normal	CAD
normal	56.4±5.9	43.6±5.9
CAD	30.8±5.9	69.2±5.9

Testing Data

	normal	CAD
normal	57.7±7.7	42.3±7.7
CAD	29.8±5.1	70.2±5.1

C.126 Experiment E4.qp

Training Data

	normal	CAD
normal	57.3±7.1	42.7±7.1
CAD	31.7±7.2	68.3±7.2

Testing Data

	normal	CAD
normal	58.0±10.0	42.0±10.0
CAD	31.3±6.7	68.7±6.7

C.127 Experiment E4.cas

Training Data

	normal	CAD
normal	84.2±4.2	15.8±4.2
CAD	39.4±4.3	60.6±4.3

Testing Data

	normal	CAD
normal	52.3±2.4	47.7±2.4
CAD	31.2±4.4	68.8±4.4

C.128 Experiment E4.cbp

Training Data

	normal	CAD
normal	41.2	58.8
CAD	19.3	80.7

Testing Data

	normal	CAD
normal	40.0	60.0
CAD	20.0	80.0

C.129 Experiment E4.cqp

Training Data

	normal	CAD
normal	41.2	58.8
CAD	19.7	80.3

Testing Data

	normal	CAD
normal	40.0	60.0
CAD	20.0	80.0

C.130 Experiment E4.ccas

Training Data

	normal	CAD
normal	82.4	17.6
CAD	33.8	66.2

Testing Data

	normal	CAD
normal	53.3	46.7
CAD	26.7	73.3

C.131 Experiment L4.knn

Training Data

	normal	CAD
normal	96.9	3.1
CAD	0.0	100.0

Testing Data

	normal	CAD
normal	36.7	63.3
CAD	16.7	83.3

C.132 Experiment L4.linreg

Training Data

	normal	CAD
normal	83.2	16.8
CAD	18.4	81.6

Testing Data

	normal	CAD
normal	46.7	53.3
CAD	23.3	76.7

C.133 Experiment L4.C4.5

Training Data

	normal	CAD
normal	100.0	0.0
CAD	1.3	98.7

Testing Data

	normal	CAD
normal	30.0	70.0
CAD	20.0	80.0

C.134 Experiment L4.MML

Training Data

	normal	CAD
normal	90.8	9.2
CAD	18.4	81.6

Testing Data

	normal	CAD
normal	43.3	56.7
CAD	26.7	73.3

C.135 Experiment L4.bp

Training Data

	normal	CAD
normal	65.6±6.9	34.4±6.9
CAD	33.0±7.6	67.0±7.6

Testing Data

	normal	CAD
normal	56.2±13.0	43.8±13.0
CAD	28.7±7.9	71.3±7.9

C.136 Experiment L4.qp

Training Data

	normal	CAD
normal	61.5±5.2	38.5±5.2
CAD	31.0±5.9	69.0±5.9

Testing Data

	normal	CAD
normal	58.3±13.1	41.7±13.1
CAD	25.3±6.9	74.7±6.9

C.137 Experiment L4.cas

Training Data

	normal	CAD
normal	88.0±3.5	12.0±3.5
CAD	37.6±5.1	62.4±5.1

Testing Data

	normal	CAD
normal	56.7±5.1	43.3±5.1
CAD	53.2±5.6	46.8±5.6

C.138 Experiment L4.cbp

Training Data

	normal	CAD
normal	57.3	42.7
CAD	10.2	89.8

Testing Data

	normal	CAD
normal	13.3	86.7
CAD	6.7	93.3

C.139 Experiment L4.cqp

Training Data

	normal	CAD
normal	61.8	38.2
CAD	3.9	96.1

Testing Data

	normal	CAD
normal	26.7	73.3
CAD	10.0	90.0

C.140 Experiment L4.ccas

Training Data

	normal	CAD
normal	85.5	14.5
CAD	29.5	70.5

Testing Data

	normal	CAD
normal	46.7	53.3
CAD	40.0	60.0

C.141 Experiment st4.knn

Training Data

	normal	CAD
normal	98.5	1.5
CAD	0.0	100.0

Testing Data

	normal	CAD
normal	3.3	96.7
CAD	30.0	70.0

C.142 Experiment st4.linreg

Training Data

	normal	CAD
normal	66.4	33.6
CAD	33.4	66.6

Testing Data

	normal	CAD
normal	26.7	73.3
CAD	33.3	66.7

C.143 Experiment st4.C4.5

Training Data

	normal	CAD
normal	100.0	0.0
CAD	0.3	99.7

Testing Data

	normal	CAD
normal	30.0	70.0
CAD	13.3	86.7

C.144 Experiment st4.MML

Training Data

	normal	CAD
normal	92.4	7.6
CAD	20.3	79.7

Testing Data

	normal	CAD
normal	23.3	76.7
CAD	40.0	60.0

C.145 Experiment st4.bp

Training Data

	normal	CAD
normal	82.1±7.4	17.9±7.4
CAD	36.0±9.3	64.0±9.3

Testing Data

	normal	CAD
normal	48.2±13.6	51.8±13.6
CAD	53.2±12.2	46.8±12.2

C.146 Experiment st4.qp

Training Data

	normal	CAD
normal	79.5±6.2	20.5±6.2
CAD	23.5±6.4	76.5±6.4

Testing Data

	normal	CAD
normal	48.0±15.4	52.0±15.4
CAD	41.3±8.7	58.7±8.7

C.147 Experiment st4.cas

Training Data

	normal	CAD
normal	85.1±3.3	14.9±3.3
CAD	32.9±6.0	67.1±6.0

Testing Data

	normal	CAD
normal	33.5±8.1	66.5±8.1
CAD	37.3±7.9	62.7±7.9

C.148 Experiment st4.cbp

Training Data

	normal	CAD
normal	92.4	7.6
CAD	32.8	67.2

Testing Data

	normal	CAD
normal	40.0	60.0
CAD	53.3	46.7

C.149 Experiment st4.cqp

Training Data

	normal	CAD
normal	92.4	7.6
CAD	18.0	82.0

Testing Data

	normal	CAD
normal	63.3	36.7
CAD	40.0	60.0

C.150 Experiment st4.ccas

Training Data

	normal	CAD
normal	96.2	3.8
CAD	21.3	78.7

Testing Data

	normal	CAD
normal	23.3	76.7
CAD	26.7	73.3

C.151 Experiment qrs4.knn

Training Data

	normal	CAD
normal	96.2	3.8
CAD	0.0	100.0

Testing Data

	normal	CAD
normal	36.7	63.3
CAD	16.7	83.3

C.152 Experiment qrs4.linreg

Training Data

	normal	CAD
normal	76.3	23.7
CAD	25.6	74.4

Testing Data

	normal	CAD
normal	63.3	36.7
CAD	20.0	80.0

C.153 Experiment qrs4.C4.5

Training Data

	normal	CAD
normal	100.0	0.0
CAD	1.6	98.4

Testing Data

	normal	CAD
normal	30.0	70.0
CAD	16.7	83.3

C.154 Experiment qrs4.MML

Training Data

	normal	CAD
normal	95.4	4.6
CAD	30.5	69.5

Testing Data

	normal	CAD
normal	60.0	40.0
CAD	43.3	56.7

C.155 Experiment qrs4.bp

Training Data

	normal	CAD
normal	73.9±10.5	26.1±10.5
CAD	34.3±13.3	65.7±13.3

Testing Data

	normal	CAD
normal	62.0±18.5	38.0±18.5
CAD	36.5±10.2	63.5±10.2

C.156 Experiment qrs4.qp

Training Data

	normal	CAD
normal	70.1±8.0	29.9±8.0
CAD	38.7±7.7	61.3±7.7

Testing Data

	normal	CAD
normal	66.7±9.5	33.3±9.5
CAD	33.7±8.0	66.3±8.0

C.157 Experiment qrs4.cas

Training Data

	normal	CAD
normal	84.3±4.0	15.7±4.0
CAD	39.8±4.6	60.2±4.6

Testing Data

	normal	CAD
normal	55.5±5.8	44.5±5.8
CAD	52.0±4.6	48.0±4.6

C.158 Experiment qrs4.cbp

Training Data

	normal	CAD
normal	84.7	15.3
CAD	2.0	98.0

Testing Data

	normal	CAD
normal	40.0	60.0
CAD	13.3	86.7

C.159 Experiment qrs4.cqp

Training Data

	normal	CAD
normal	74.8	25.2
CAD	11.8	88.2

Testing Data

	normal	CAD
normal	50.0	50.0
CAD	16.7	83.3

C.160 Experiment qrs4.ccas

Training Data

	normal	CAD
normal	84.0	16.0
CAD	32.8	67.2

Testing Data

	normal	CAD
normal	53.3	46.7
CAD	50.0	50.0