A FUZZY SUPPORT VECTOR MACHINE APPROACH FOR ECG ANALYSIS

by

N. Özlem ÖZCAN

B.S., Computer Engineering, METU, 2005

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in System and Control Engineering

Boğaziçi University

2010

A FUZZY SUPPORT VECTOR MACHINE APPROACH FOR ECG ANALYSIS

APPROVED BY:

Prof. Fikret Gürgen (Thesis Supervisor)	…….
Prof. Lale Akarun	…….
Assoc. Prof. Cengizhan Öztürk	…….

DATE OF APPROVAL: 25/01/2010

ACKNOWLEDGEMENTS

I would like to thank to the people who encouraged and supported me through this research.

First of all, I would like to thank my supervisor Prof. Fikret Gürgen, who proposed me this thesis. From the very beginning of the litreture survey and to the detailed tests, he supported me with his knowledge and experience. His encourage and guidance lead me to the succesful completion of this thesis study.

Secondly, I would like to thank my thesis examining committee, Prof. Lale Akarun and Assoc. Prof. Cengizhan Öztürk..

I am very grateful to my parents and my friends for their endless support and thrust. I want to thank to my father Prof. Mustafa Özcan who encouraged me during my thesis and my fiancée Ozan Şimşek who shared his valuable thoughts on the research methods. At the times when I was disappointed of some test trials, they encouraged me with their belief in my intention to accomplish this thesis study in the very best way.

Finally, I want to state appreciation to Fikret Küçükdeveci from TEPA. In the construction of the real life ECG dataset, he shared a part of their datasets with me and gave me chance to verify my work.

ABSTRACT

A FUZZY SUPPORT VECTOR MACHINE APPROACH FOR ECG ANALYSIS

In this thesis, fuzzy approach is used for ECG analysis. The ECG dataset in the UCI database is used. This dataset consists of inputs from normal and abnormal ECGs. All the anomalies are used to construct one class and normal ECG data is used to construct another class. The main purpose of the system is to detect anomalies correctly. In order to achieve this goal, a fuzzy support vector machine is constructed. Five different fuzzy membership functions are tested to reach the best performance: OCW, DTCM, DTOCM, CAR and FCM. Output of the fuzzy support vector machine system is compared to other classification methods. Results show that the fuzzy support vector machine outperforms other methods.

In order to interpret the classification model, rule base extraction methods are applied. C4.5, PART, RIPPER and ANFIS are the selected algorithms for ECG rule base generation. When accuracy is considered as performance metric, RIPPER method outperforms the other techniques. The advantage of using ANFIS is the membership function generation for the features in the dataset. The resulting membership functions are found to be consistent with medical knowledge.

ÖZET

EKG ANALİZİNDE BULANIK DESTEK VEKTÖR MAKİNESİ YAKLAŞIMI

Bu tezde, EKG analizinde bulanık yaklaşım kullanılmıştır. UCI veritabanındaki EKG veriseti kullanılmıştır. Bu veriseti, normal ve anormal EKGlerden oluşur. Tüm anormal EKGler bir sınıf ve normal EKGler bir sınıf oluşturacak şekilde düzenlendi. Sistemin esas amacı; anormallikleri doğru tespit etmektir. En iyi performansı elde edebilmek için beş bulanık üyelik fonksiyonu denendi: OCW, DTCM, DTOCM, CAR ve FCM. Sistemin çıktısı diğer sınıflandırma yöntemleri ile karşılaştırıldı. Sonuçlar bulanık destek vector makinesinin diğer yöntemlerden daha başarılı olduğunu gösterdi.

Sınıflandırma modelini tercüme etmek için kural tabanı çıkarımı metodları uygulandı. C4.5, PART, RIPPER ve ANFIS, seçilen EKG kural tabanı çıkarım algoritmalarıdır. Performans ölçeği olarak doğruluk alındığında, RIPPER yöntemi diğer tekniklerden daha başarılı oldu. ANFIS kullanmanın avantajı, verisetindeki özellikler için üyelik fonksiyonu üretmesidir. Nihayi üyelik fonksiyonlarının tıbbi bilgilerle tutarlı olduğu bulunmuştur.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

ABSTRACT iv

ÖZET v

LIST OF FIGURES ix

LIST OF TABLES x

LIST OF SYMBOLS/ABBREVIATIONS xi

1. INTRODUCTION 1

1.1. Motivation 4

1.2. Previous Work 6

1.3. Proposed Method 12

1.4. Dataset 17

1.5. Outline 18

2. DIMENSION REDUCTION 20

2.1. Feature Selection 20

2.1.1. Recursive Feature Elimination with Support Vector Machine 21

2.1.2. Correlation Based Feature Selection 22

2.2. Feature Extraction 24

2.2.1. Principal Component Analysis 24

2.2.2. Factor Analysis 25

3. FUZZY MEMBERSHIP FUNCTIONS 26

3.1. One Class Weighing 27

3.2. Distance To Class Mean 27

3.3. Distance To One Class Mean 28

3.4. Cardinality 29

3.5. Fuzzy C-Means Clustering 30
4. ECG ANALYSIS 32
4.1. ECG Classification 32
4.1.1. k-Nearest Neighbor 33
4.1.2. Multilayer Perceptron 33
4.1.3. Support Vector Machine 34
4.1.4. Fuzzy Support Vector Machine 36
4.2. Rule Base Development 37
4.2.1. C4.5 Decision Tree 38
4.2.2. Repeated Incremental Pruning to Produce Error Reduction 38
4.2.3. Partial Decision Tree 39
4.2.4. Adaptive Neuro-Fuzzy Inference System 39
5. EXPERIMENTAL RESULTS 41
5.1. Performance Metrics 42
5.2. Experimental Results for Dimensionality Reduction 43
5.2.1. Experimental Results for Feature Selection 44
5.2.2. Experimental Results for Feature Extraction 44
5.2.3. Empirical Feature Selection 45
5.3. Experimental Results for Fuzzy Membership Value Calculation 47
5.3.1. One Class Weighing 48
5.3.2. Distance To Class Mean 48
5.3.3. Distance To One Class Mean 50
5.3.4. Cardinality 50
5.3.5. Fuzzy C-Means Clustering 51
5.4. Experimental Results for ECG Classification 51
5.4.1. k-Nearest Neighbor 51
5.4.2. Multilayer Perceptron 56

5.4.3. Support Vector Machine 59
5.4.4. Fuzzy Support Vector Machine 62
5.4.5. Classifier Evaluation 63
5.5. Experimental Results for Rule Base Development 71
5.5.1. C4.5 Decision Tree 72
5.5.2. Repeated Incremental Pruning to Produce Error Reduction 74
5.5.3. Partial Decision Tree 76
5.5.4. Adaptive Neuro-Fuzzy Inference System 77
5.5.5. Rule Base Evaluation 78
6. EXPERIMENTS WITH REAL ECG DATA 83
7. CONCLUSION 84
APPENDIX 85
REFERENCES 115

LIST OF FIGURES

Figure 1.1 The placement of electrodes for the standard 12-lead ECG [1] . . . 1
Figure 1.2 Basic heart anatomy [2] 3
Figure 1.3 Waves in a normal ECG plot [3] 4
Figure 1.4 ECG output on scaled paper [4] 6
Figure 1.5 Main flow diagram of the proposed method 13
Figure 1.6 Flow diagram of the dimension reduction module 14
Figure 1.7 Flow diagram of the fuzzy membership value calculation module 15
Figure 1.8 Flow diagram of the classification module 16
Figure 1.9 Flow diagram of the rule base development module 17
Figure 4.1 ANFIS Structure a) Type-3 Reasoning b) Type-3 ANFIS [48] . . 40
Figure 5.1 EMP-ECG dataset 47
Figure 5.2 Illustration of DTCM method 49
Figure 5.3 Accuracy test results of k-NN 53
Figure 5.4 f-score test results of k-NN 54
Figure 5.5 Matthews correlation coefficient test results of k-NN 55
Figure 5.6 Accuracy test results of MLP 57
Figure 5.7 f-score test results of MLP 58
Figure 5.8 Matthews correlation coefficient test results of MLP 58
Figure 5.9 Accuracy test results of SVM 60
Figure 5.10 f-score test results of SVM 61
Figure 5.11 Matthews correlation coefficient test results of SVM 62
Figure 5.12 Accuracy test results of FSVM 64
Figure 5.13 f-score test results of FSVM 65
Figure 5.14 Matthews correlation coefficient test results of FSVM 66
Figure 5.15 ROC curve of CFS10 for FSVM-DTCM 70
Figure 5.16 CFS10 dataset 72
Figure 5.17 FIS Structure of ECG classification model of ANFIS 79
Figure 5.18 Membership functions from ECG FIS 82

LIST OF TABLES

Table 1.1	Orientations of the 12 Lead	2
Table 5.1	Feature list of EMP-ECG dataset	46
Table 5.2	Optimum cost and gamma values for SVM	59
Table 5.3	Classifier accuracy comparison	67
Table 5.4	Classifier f-score comparison	67
Table 5.5	Classifier Matthews correlation coefficient comparison	68
Table 5.6	Classifier precision comparison	68
Table 5.7	Classifier sensitivity or recall comparison	69
Table 5.8	Classifier specificity comparison	69
Table 5.9	C4.5 initial rule base	73
Table 5.10	C4.5 resulting rule base	74
Table 5.11	RIPPER initial rule base	75
Table 5.12	RIPPER resulting rule base	75
Table 5.13	PART initial rule base	76
Table 5.14	PART resulting rule base	77
Table 5.16	Rule base accuracy performance	78
Table 6.1	FSVM results of OWN-ECG	83

LIST OF SYMBOLS/ABBREVIATIONS

ACC	Accuracy
C4.5	C4.5 Decision Tree
CAR	Cardinality
CFS	Correlation Based Feature Selection
DTCM	Distance to Class Mean
DTOCM	Distance to One Class Mean
FA	Factor Analysis
FCM	Fuzzy C-Means
FN	False negative
FP	False positive
FSVM	Fuzzy Support Vector Machine
k-NN	k-Nearest Neighbors
MCC	Mathew's Correlation Coefficients
MIT-BIH	The MIT Arrhythmia Dataset
MLP	Multi Layer Perceptron
OCW	One Class Weighing
OWN-ECG	The Real Life ECG Dataset
PART	Partial Decision Tree
PCA	Principal Component Analysis
RFE-SVM	Recursive Feature Elimination with Support Vector Machine
RIPPER	Repeated Incremental Pruning to Produce Error Reduction
SVM	Support Vector Machine
TN	True Negative
TP	True positive
UCI-ECG	The UCI Arrhythmia Dataset

1. INTRODUCTION

ECG, electrocardiography, is a simple and common procedure in hospitals and health centers now. Despite of being simple to record, the analysis of ECG is of great importance. ECG is the recording of the heart's electrical activity. The recording is done by placing electrodes on the skin in specific coordinates. 10 electrodes are used for the standard 12-lead ECG. The electrodes are attached to an electrocardiograph which processes the inputs and outputs the electrical activity as a signal graph on a scaled paper. The placement of the electrodes is given in Figure 1.1:

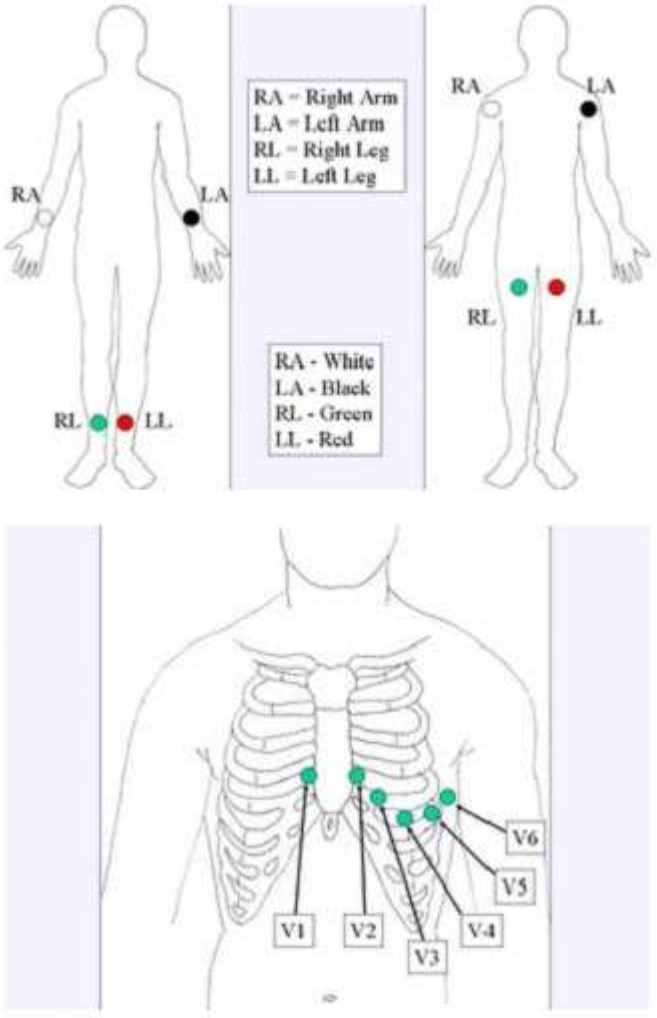


Figure 1.1 The placement of electrodes for the standard 12-lead ECG [1]

In this thesis, the standard 12-lead ECG is considered. Each lead represents the measurement from a different orientation. In the ECG output, all leads are plotted separately, Figure 1.4. The leads and their orientations are listed in Table 1.1:

Table 1.1 Orientations of the 12 Lead

Lead Name	**Orientation**
Lead I	RA(-) to LA(+)
Lead II	RA(-) to LF(+)
Lead III	LA(-) to LF(+)
Lead aVR	RA(+) to [LA and LF](-)
Lead aVL	LA(+) to [RA and LF](-)
Lead aVF	LF(+) to [RA and LA](-)
Lead V1	V1(+) to WCT(-)
Lead V2	V2(+) to WCT(-)
Lead V3	V3(+) to WCT(-)
Lead V4	V4(+) to WCT(-)
Lead V5	V5(+) to WCT(-)
Lead V6	V6(+) to WCT(-)
WCT: Wilson's Central Terminal is the average of the three limp leads (aVR, AVL, AVF).	

The heartbeat cycle is a two phase period: *diastole* and *systole*. It is triggered by a signal from the sinoatrial (SA) node. This signal is generated as the right atrium is filled with blood. The signal spreads across right and left atria causing them to contract. As a result of this action, blood is directed from atria to ventricles through open valves. This part is called *diastole*, the first phase of heartbeat cycle.

After diastole, the signal arrives at the atrioventricular (AV) node and slows to allow right and left ventricles to fill with blood. The signal moves along the bundle of His which is located in the walls of the ventricles and branches to the Purkinje fibers which surrounds the ventricles. This causes the ventricles to contract. Blood is directed to the lungs from the right ventricle and to the rest of the body from the left ventricle. This part is called *systole*, the second phase of the heartbeat. The walls of the ventricles relax as the signal passes and await the next signal. This cycle continues as signals come from the SA node. Heart rate is

determined via counting the number of SA signals per minute. The basic heart anatomy figure, Figure 1.2, shows the parts of the heart mentioned:

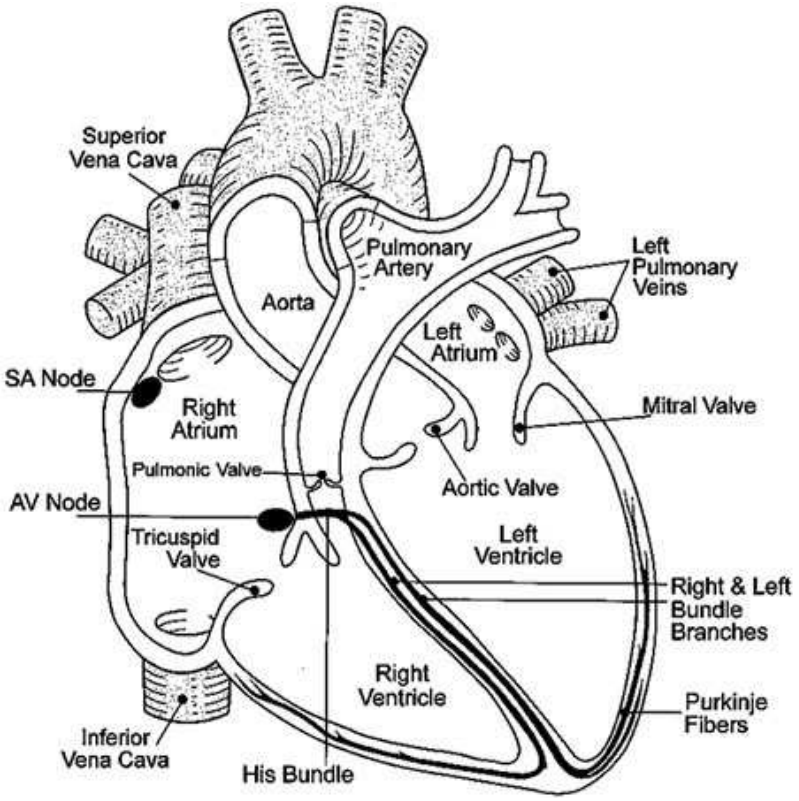


Figure 1.2 Basic heart anatomy [2]

The signal generated through the diastole and systole phases is captured and plotted on a scaled paper via an electrocardiograph. The plot consists of waves which are labeled by letters; P, Q, R, S, T, U. Each wave corresponds to a specific action during the cycle. P wave occurs when the atria contract. During the interval between P and Q waves, both ventricles are filled with blood. Q wave marks the signal travel through the bundle of His to the Purkinje fibers. The Purkinje fibers surround the ventricles and the signal causes them to contract. Left ventricle contracts before the right. R wave marks the contraction of the left ventricle while S wave marks the right. T and U waves occur in the relaxing process of the ventricles. The U wave is not mentioned in all ECG figures. The waves and their shape in a normal ECG can be seen in Figure 1.3:

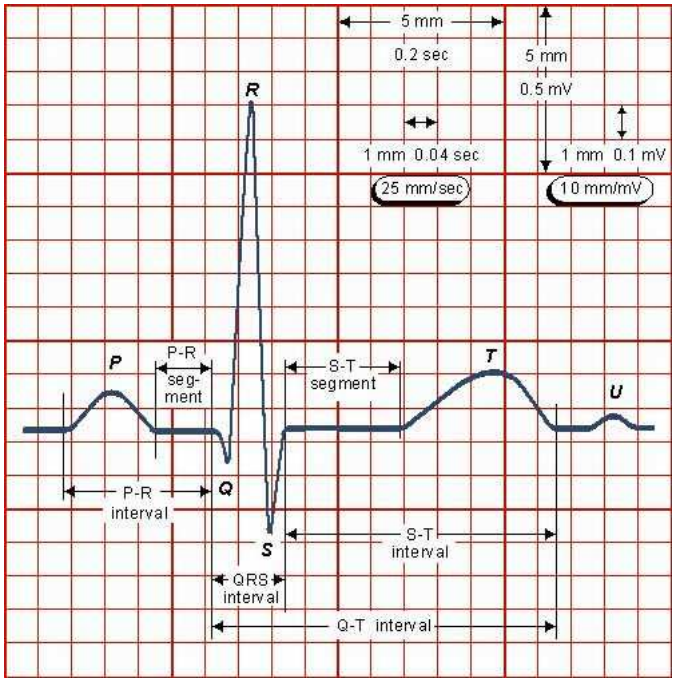


Figure 1.3 Waves in a normal ECG plot [3]

ECG output is plotted as signal strength versus time graph on a scaled paper. All leads are shown separately. Due to the orientation of the lead, all of the waves may not be considered in the plot. There are small squares (side-length: 1mm) and big squares (side-length: 5mm) made of 25 small squares in the scaled paper. In the standard ECG output, 10 small squares correspond to 1 mV in the vertical axis and 25 small squares correspond to 1 second in the horizontal axis, Figure 1.3. The scale is also noted on the output paper if it is a standard scale or not. Experts perform their calculations and analysis of the ECG according to this scale.

1.1. Motivation

Normal heart activity begins with a signal at the SA node. This signal travels through the heart giving rise to the stages of a heartbeat. These stages can be examined from the ECG output. As normal ECG signal shape and limits are known, anomalies can be detected by ECG analysis. The results of this analysis can lead to further tests and help in diagnosis.

ECG analysis emphasize on the amplitude, duration and regularity of the waves. Amplitude measurements are performed via counting the squares in vertical axis. In the standard scale, 1 mV is represented by 10 small squares. The amplitude of P wave in a normal ECG is below 0.25mV. This corresponds to 2.5 small squares in vertical direction. P wave is formed in consequence of the activation in right and left atria. If there exits an abnormal P wave in the ECG, the doctor may suspect a malfunction in atria. Amplitude also covers negative or positive direction of the wave. P, R, T and U waves are positive waves. Q and S waves are negative compared to the baseline of ECG. A normal ECG and ECG with an enormous P wave is given as an example in Figure 1.4. The change in P wave can be observed clearly in Leads I, II and III. The patient with the second ECG has right arterial enlargement.

Duration measurements are also performed by counting the squares but in horizontal axis. In the standard scale, 25 small squares correspond to 1 second. The duration of the QRS complex is at most 0.10 seconds equivalently 2.5 small squares horizontally. The QRS complex represents the activation of the ventricles. Varied QRS durations longer than 0.25 seconds, alarm a possible affection in the ventricles.

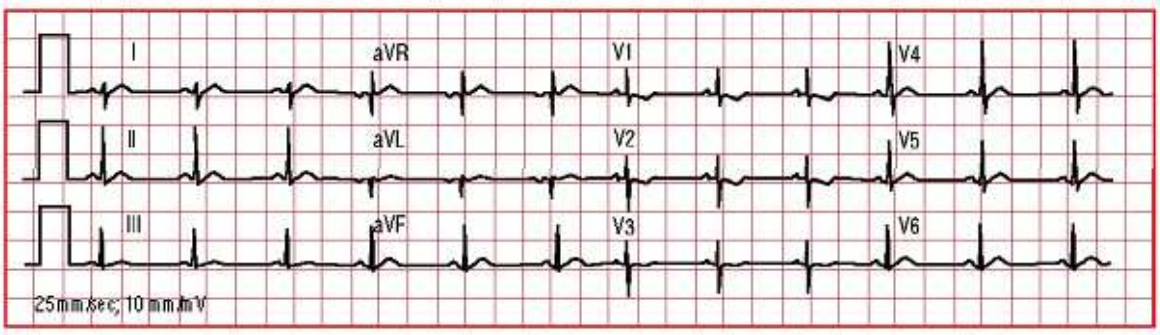
Regularity needs a diverse analysis compared to the former. The ECG signal is a series of waves and the order of the waves is determined. P wave occurs before QRS complex which is followed by T and U waves. Although not all leads can be monitored in all leads, each lead has a peculiar regularity. An extra wave or a missing wave can be treated as a symptom of anomaly.

Heart rate is another criterion that can be computed from ECG. The number of squares between two successive QRS complex is counted. If small squares are used, 1500 is divided by this count. If big squares are use, 300 is divided by this count. The result gives the heart rate of the patient. There are also limitations on normal heart rate. 60 or 90 beat per minute is accepted as normal rate.

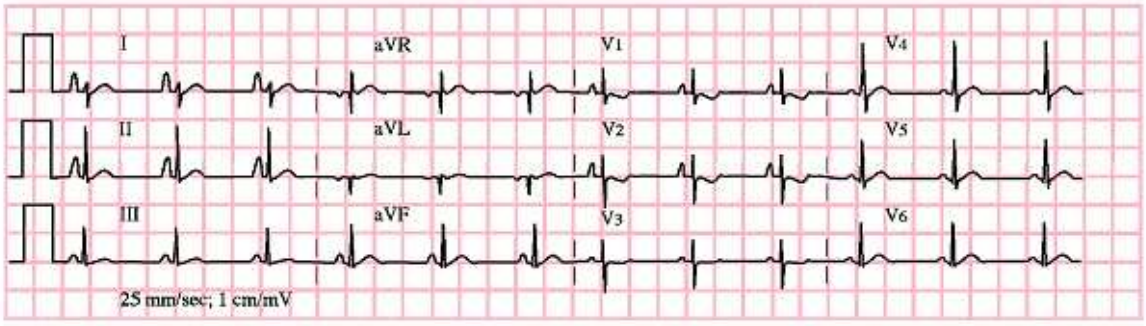
The cardiac analysis from the ECG paper is a challenging task. Experience and focused attention are the essential equipment. Hence, the judgment of a cardiologist is as successful as the limits of his experience. Another outcome of the experience dependency

is the development of automated ECG analysis tools. These tools provide contributory information for the cardiologist. They are capable of extracting common measurements from the ECG and handling complex features in short time. The outputs of these tools are attached to the ECG paper as additional information.

The aim of this research is to develop an advanced ECG analysis method. For the medical purposes, the developed method can be transformed to an automated analysis tool.



Normal ECG



Right Atrial Enlargement (RAE)

Figure 1.4 ECG output on scaled paper [4]

1.2. Previous Work

Besides cardiovascular diseases, heart attacks are the main cause of death around the world. Pre-monitoring or pre-diagnostic helps to prevent heart attacks and strokes. ECG plays a key role in this regard. Several and different combinations of techniques are applied on ECG for pattern recognition, classification and arrhythmia detection.

ECG analysis is performed for two types of data: signal, numeric. There are two commonly used databases: MIT-BIH [5] and UCI Arrhythmia [6]. MIT-BIH contains signal data of ECG whereas UCI Arrhythmia consists of numeric.

Neural networks are popular tools for ECG analysis. In a study of year 1998, several artificial neural network architectures are implemented [7]. The study presents a comparison between performance and training time of different architectures.

Specialized ECG analysis for ischemia detection is achieved in [8]. Ischemia is a restriction in blood supply, generally due to factors in the blood vessels, with resultant damage or dysfunction of tissue [9]. Ischemic cardiac beat detection relies on the detection of ST segment from the ECG. Nonlinear principal component analysis, implemented by a multilayer neural network, is used for extraction of this feature. European ST Database [10] is used and each ST segment is classified as normal or abnormal.

Classification techniques are applied to ECG morphological features in [11]. The selected features are mean R-peak value, mean power spectral density, area under QRS complex, energy of the signal, Q-S distance and autocorrelation value. Three neural network architectures are tested: multilayer perceptron (MLP), radial basis function (RBF) networks and self-organizing feature maps (SOFM). All classifiers are tested in absence and presence of additive Gaussian noise. The MLP network exhibits the best performance and is found to be more robust.

Machine learning schemes are used to classify arrhythmia from ECG data. Three machine learning techniques, OneR, C4.5 (decision tree) and Naïve Bayes, are applied on ECG data in [12]. The aim of the study was developing a non-invasive ECG hardware and embedded software for capturing, analyzing, diagnosing, and recommending remedies for homecare patients with heart condition sets. UCI Arrhythmia database is used. Experiments are performed on WEKA platform [13]. Accuracy and learning time are the performance criteria. High and stable accuracy and low learning time are reached with Naïve Bayes.

Wavelet transforms are applied in order to construct a multilead model [14]. Arrhythmia database of MIT-BIH is used. Classification is applied on features extracted from different scales of the wavelet decomposition of leads integrated in an RMS manner. Linear discriminant analysis (LDA) is used for dimensional reduction. Three classifiers: linear, quadratic and Mahalanobis distance are evaluated. Results show that the best performance is obtained by using LDA and a Mahalanobis distance classifier with a 28 feature subset.

Fuzzy and neuro-fuzzy approaches are also operated on ECG. One example is the fuzzy-Gaussian neural network (FGNN) [15]. This technique is applied for detection of ischemic heart disease. Only Lead V5 is used in the study. Principal component analysis (PCA) and discrete cosine transform (DCT) are used for feature extraction and FGNN is applied on the resulting features. High accuracy is obtained in the expense of high number of training epochs.

In [16], fuzzy knowledge is constructed using four non-linear parameters; spectral entropy, Poincaré plot geometry, largest Lyapunov exponent and detrended fluctuation analysis. MIT-BIH database is used in this study. Fuzzy rules are defined for the parameters. The proposed fuzzy method is implemented using Matlab fuzzy logic toolbox.

Adaptive fuzzy techniques are also applied on ECG data [17]. The study is built on MIT-BIH database. Feature detection is achieved with wavelet transform technique. Adaptive aspect of the proposed method consists of statistical learning of historical data and adapting threshold values and membership function boundaries according to the learning phase. Accuracy is evaluated as performance metric.

A study on ECG chaos features is carried out in [18]. Examples of the referred features can be spatial filling index (SFI), correlation dimension (D2), central tendency measure (CTM) and approximate entropy (ApEn). They are extracted by a platform called ECG Chaos Extractor, which is a previous study of the writers of [18]. The proposed method is applied on MIT-BIH database. Clustering and classification processes run on the extracted features on WEKA platform [13]. K-Means and EM (Expectation Maximization) are applied as clustering algorithms. For classification purposes, C4.5 and Bayesian

network algorithms are used. The results of the study show that EM clustering algorithm outperforms K-Means. But no apparent difference in efficiency between C4.5 and Bayesian Network algorithms can be stated.

Technique of support vector machines (SVM) is recently popular in classification scope. Applications of SVM on ECG classification are found in literature. SVM is applied with different kernels and optimization of model (C) and kernel (γ) parameters. In [19], SVM is applied after PCA. MIT-BIH database is used. RBF kernel is chosen for SVM. The proposed method is compared to CART (classification and regression trees) and fuzzy ARTMAP. Experiment results show that SVM outperform other methods.

One-against-all SVM classification is performed on MIT-BIH database in [20]. RBF kernel is chosen for SVM in this study too. High accuracy is obtained.

Another study of SVM on MIT-BIH database is [21]. SVM is applied after adaptive Hermite decomposition. One-against-all approach is realized with SVM. Sensitivity and specificity are the performance metrics in this study. Experiments are designed on interpersonal and intrapersonal partitions. Results imply that intrapersonal classification is more reliable and accurate.

In [22], classification of ECG is based on HRV (heart rate variability) signal. Feature reduction with GDA (generalized discriminant analysis) and SVM classifier forms the proposed method. GDA is applied on MIT-BIH database. Experiments are carried out on normal rhythm and 5 different arrhythmia types. Resulting features from GDA is compared to original features. Performance metrics, sensitivity, specificity, positive predictability and accuracy, are shown to be improved with GDA. With a one-against-all and one-against-one comparison, one-against-all outperforms the latter.

Another study of SVM on ECG is applied to UCI database [23]. PCA and ICA are used for dimensionality reduction and then SVM is applied. The results of SVM classifier is compared to k-NN (K nearest neighbor) and DT (decision tree) algorithms. According to experimental results improved SVM results shown to outperform competing classification results.

With the premising results obtained from SVM classification, new studies are performed in order to combine SVM with various approaches. SVM combined with fuzzy approach is called fuzzy-SVM (FSVM). Simply, this technique defines weight values for each input and leads SVM algorithm to detect support vectors based on these weights. Studies on this subject can be partitioned into two classes: methodology and application.

One of the methodology studies combines Fuzzy C-Means (FCM) with SVM [24, 25]. The input data is first clustered by FCM. A fuzzy membership value for each input is obtained as a result of FCM. These values are used as weight values and FSVM is run with the new augmented input set. FSVM is compares with standard SVM and Least Squares Error (LSE). The results show that FSVM is robust in folding and outperforms the compared algorithms.

In [26], Mahalanobis distance is used in FCM. This modified version of FCM is improved to differentiate main training data points and outliers according to their relative importance in the training data. Membership values from FCM are used for FSVM. Datasets from UCI machine learning repository are used in experiments. Besides the original datasets, new benchmark datasets are formed via adding artificial outlier inputs. Test results imply that higher accuracy and robustness to outliers are gained with FSVM.

In [27], an automatic method is proposed for generating fuzzy membership values for data points from a heuristic strategy by using two factors, the confident factor and the trashy factor, and a mapping function. If the heuristic value of an input is bigger than the confident factor, the fuzzy membership value of input is set to 1. If the heuristic value of an input is smaller than the trashy factor, the fuzzy membership value of input is set to a minimum value. Other inputs which lie between the two factors are assigned proportional weights. FSVM is applied to the dataset with weights. 13 benchmark datasets are used in experiments and the results show that FSVM minimizes the train and test error compared to SVM.

In discrete SVM, the count of the number of misclassified examples is used instead of the distance from the separating hyper plane. Different membership functions are

defined to run with fuzzy discrete SVM in [28]. Distance based membership function calculates Euclidean distance to class mean for each input. Score based membership function uses a generic classification algorithm to assign a score value. Frequency based membership function calculates the mean of different discrete SVM classification models. Accuracy and sensitivity are used as performance metrics. Fuzzy discrete SVM accomplishes an improvement compared to FSVM according to test results.

In [29], one against one (OAO) FSVM is compared to OAO-SVM for text categorization. A ratio and exponential form based on distance to class centers are used as fuzzy membership functions. Precision, recall and f-measure are calculated after four fold and three fold test runs. OAO-FSVM with exponential distance weights outperforms the compared methods in both tests.

As a last modification of FSVM found in this research, iterative form is tested in [30]. Initial membership values are calculated by applying a non-fuzzy SVM. Output probabilities from FSVM are used as a feedback to the next step. Two stopping conditions are tested: fixed number of iterations, iterations until convergence. The test results show that FSVM outperforms SVM but convergence method leads to over fitting.

As an application study, in [31], FSVM is applied to electroencephalographic (EEG) signals. EEG is the recording of electrical activity along the scalp produced by the firing of neurons within the brain [32]. Wavelet-based features are selected and RBF kernel is used for FSVM. The weight value of an input is defined as the number of data points in its neighborhood. This study shows that FSVM is more robust to noise compared to SVM and also accuracy is advanced with weight values.

The literature survey concludes that ECG analysis is an important area in biomedical studies. Many techniques are applied to ECG datasets. The general aim is to extract features from the ECG signal and then apply the classification method. In this study, a numerical dataset, which is the UCI Arrhythmia dataset, is preferred. The dataset contains several features for each input. The meaningful features are grouped in a reduced dataset for better performance. As inputs provide different levels of representation for their labeled classes, a fuzzy classification model, FSVM, is preferred as the main proposed

method. In order to present the classification model in a readable format, rule base extraction techniques are also applied in this study. The details of the proposed method are explained in the next section.

1.3. Proposed Method

This thesis study is specialized on ECG data. The aim of this study is to detect ECG anomalies correctly. Diverse techniques are applied to the dataset to compose a comparative study. Figure 1.5 represents the main flow of this study. There are two categories of medical studies: interpersonal and intrapersonal. In interpersonal studies, samples are collected among different patients. Unlike interpersonal approach, in the intrapersonal studies samples are collected from the same patient. This thesis introduces an interpersonal classification study of ECG data. UCI Arrhythmia dataset has 278 valid features. Hence the first module of the flow is dimensionality reduction. As the proposed method is FSVM and weight values are needed, the second module is the calculation of fuzzy membership values. A number of classification techniques are applied to the resulting datasets as the third module. To produce a readable summary of the classification results, a rule base development step is achieved as the forth and last module. Each module is explained in detail in the following paragraphs.

Dimension reduction is not the main objective of this study, but has a great importance as it is the beginning phase. Other modules are applied on the outputs of this module. Five techniques are used in this module: principal component analysis (PCA), factor analysis (FA), recursive feature elimination with support vector machines (RFE-SVM), correlation based feature selection (CFS) and empirical feature selection (EFS). Detailed diagram of this module is given in Figure 1.6.

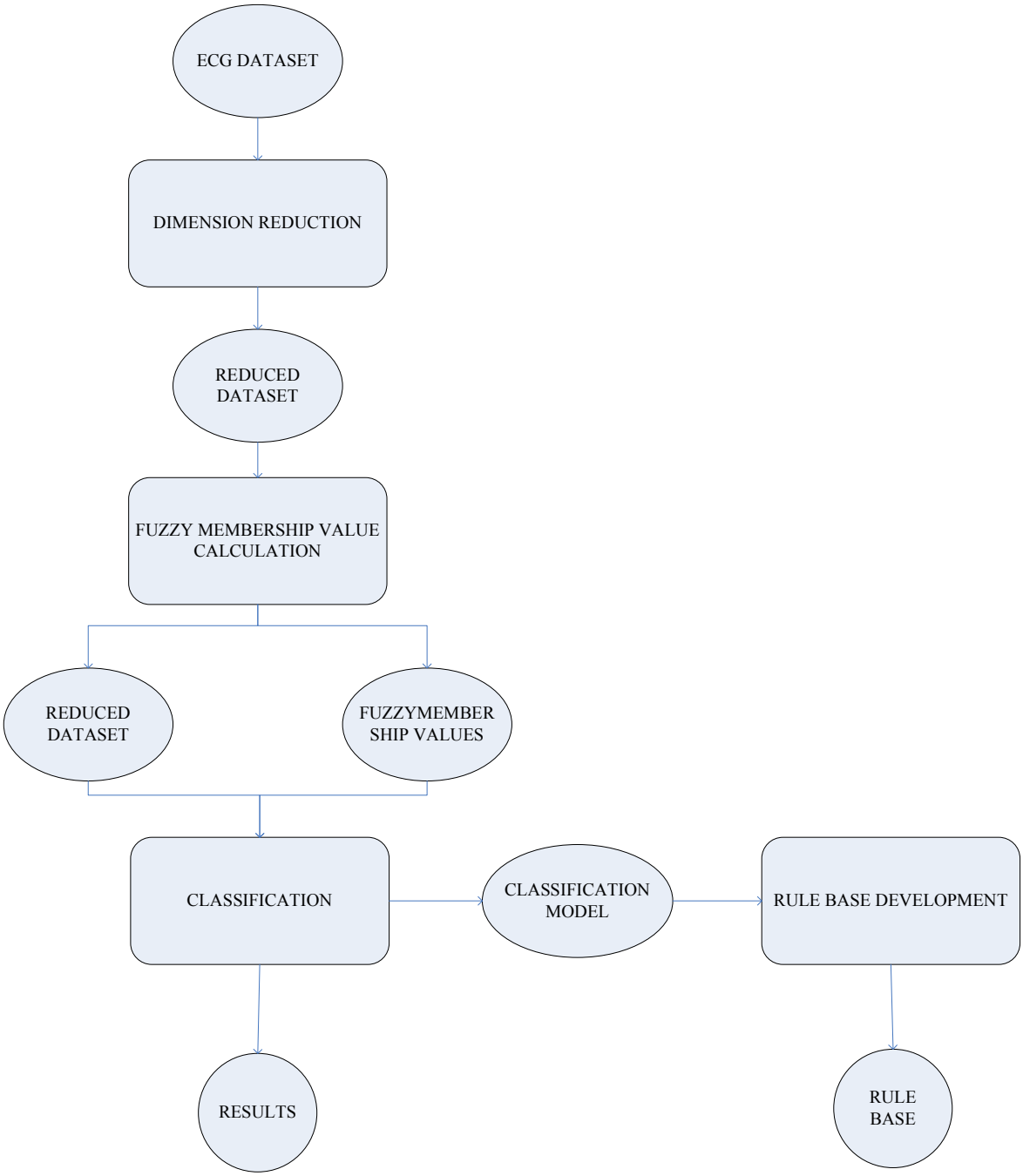


Figure 1.5 Main flow diagram of the proposed method

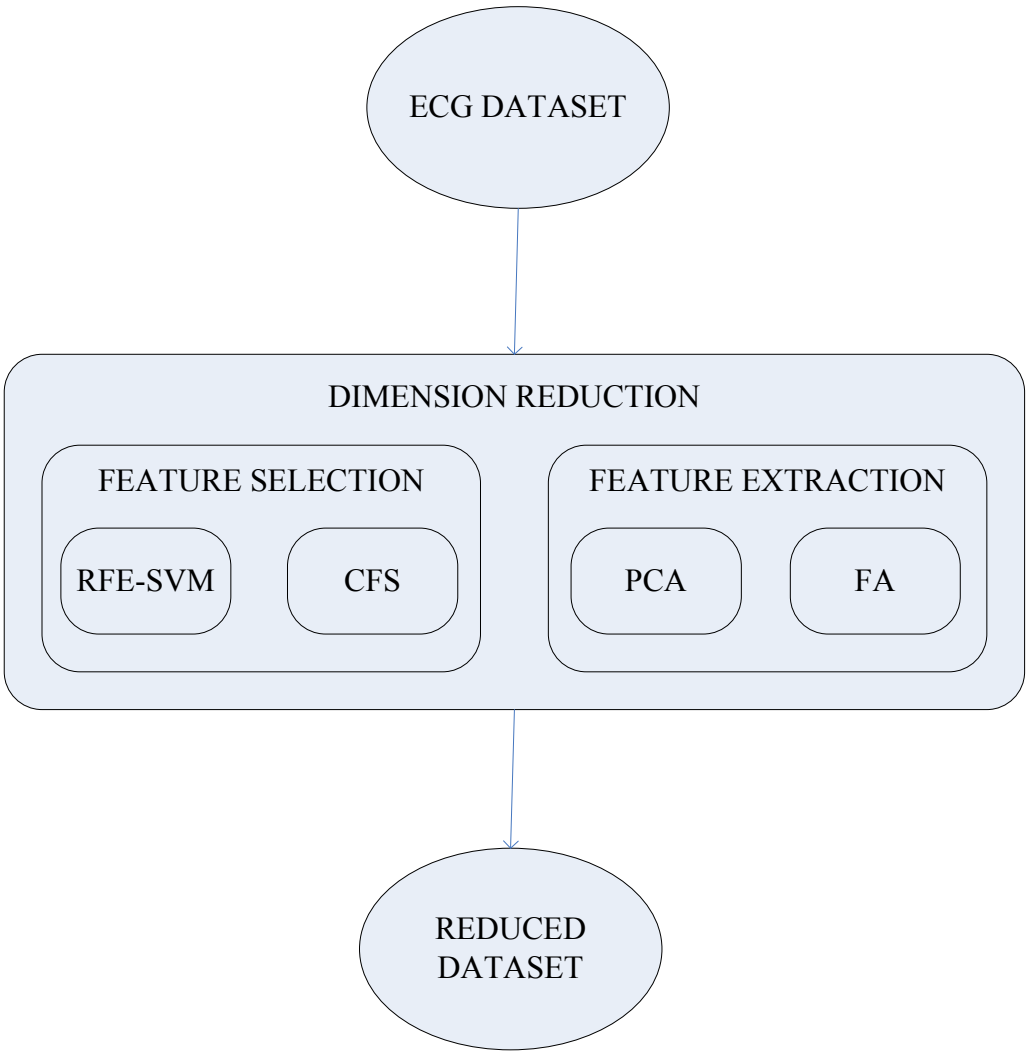


Figure 1.6 Flow diagram of the dimension reduction module

The second module is an intermediate module. The calculations of fuzzy membership values are realized in this module. Each technique assigns a different membership value due to its calculation basis. Five types of fuzzy membership values are applied: one class weighing method (OCW), distance to class mean (DTCM), distance to one class mean (DTOCM), cardinality (CAR) and fuzzy c-means (FCM) clustering. The block diagram is shown in Figure 1.7.

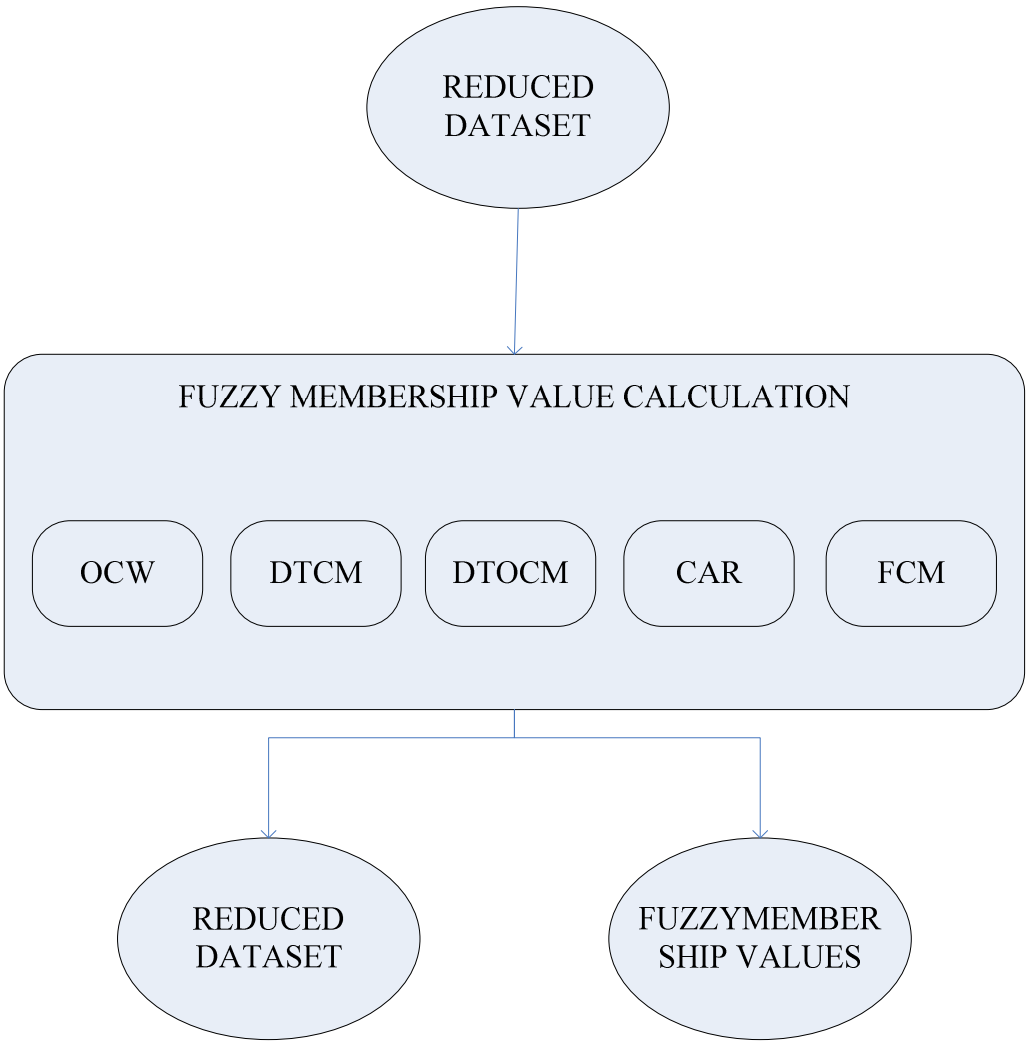


Figure 1.7 Flow diagram of the fuzzy membership value calculation module

The main module is the calculation module. The former modules serve in order to maximize the performance of this module. The latter module arranges the results of this module aiming to reach more readable results. Four techniques are used in this module: k nearest neighbors (k-NN), multilayer perceptron (MLP), support vector machines (SVM) and fuzzy support vector machines (FSVM). Figure 1.8 depicts the block diagram of this module.

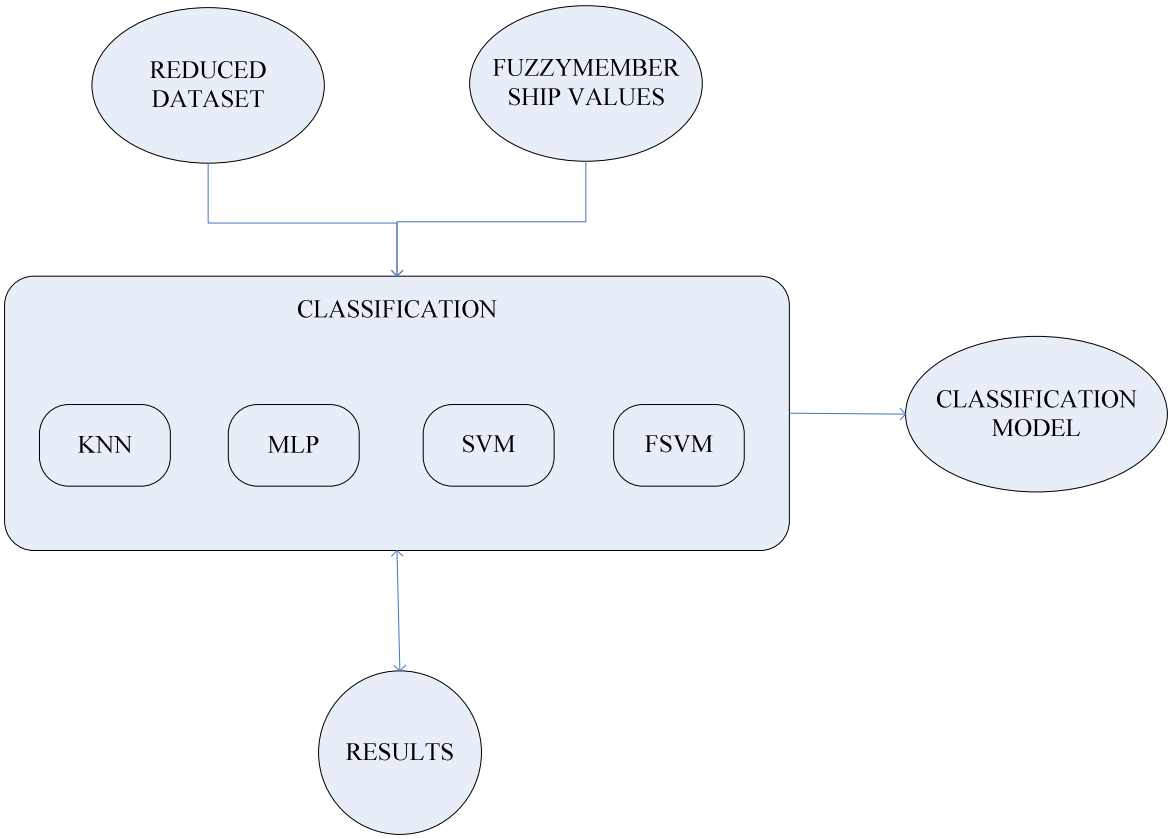


Figure 1.8 Flow diagram of the classification module

The last module is the rule base development module, Figure 1.9. The aim of this module is to create a diverse summary of the classification module output. The produced rule base can be used by cardiac experts as a helping tool in the reading of ECG. Four different techniques are used for comparison: C4.5 decision tree, partial decision tree (PART), Repeated Incremental Pruning to Produce Error Reduction (RIPPER) and adaptive neuro-fuzzy inference system (ANFIS).

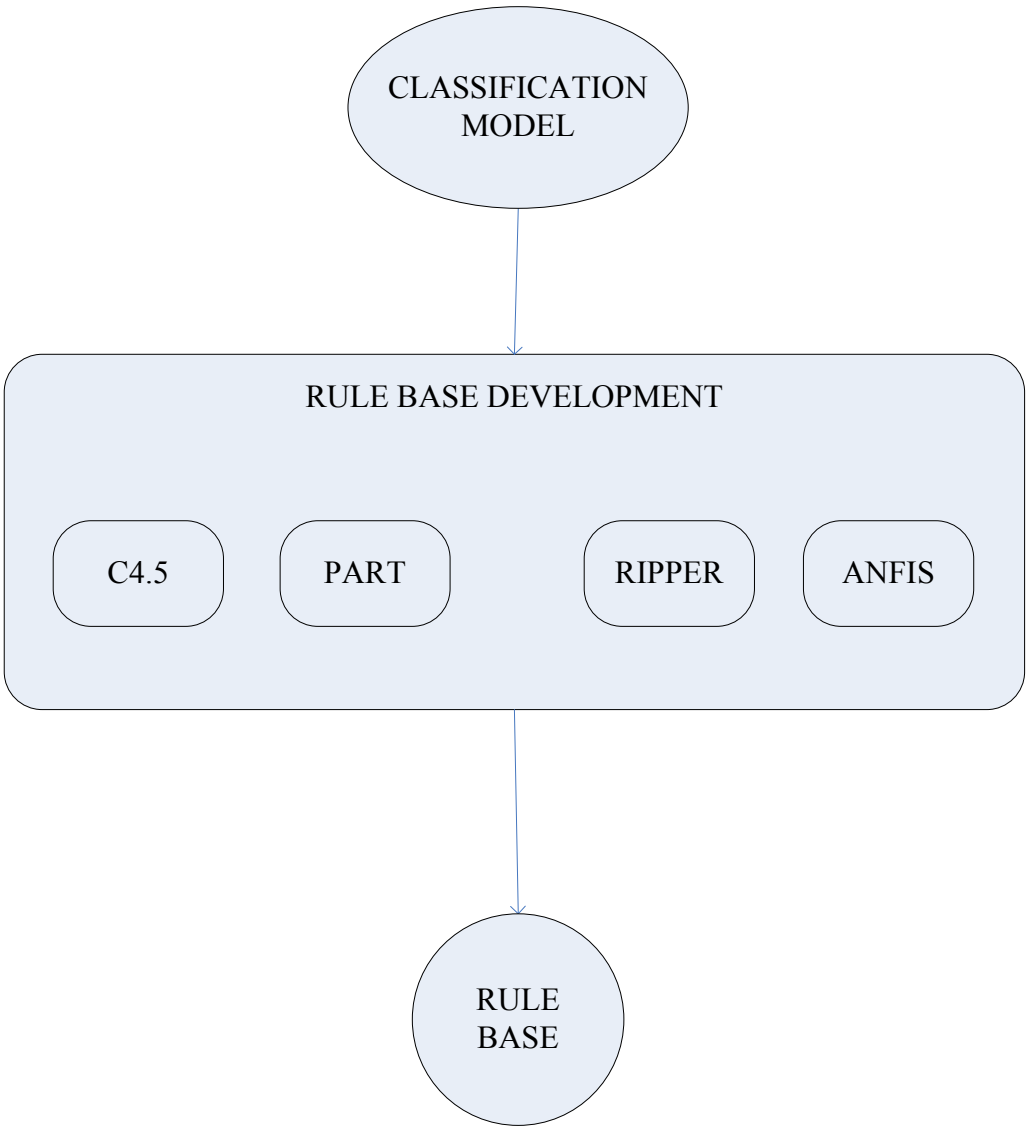


Figure 1.9 Flow diagram of the rule base development module

1.4. Dataset

The techniques discussed so far are specialized in order to analyze the ECG data. The main goal of the study is to detect anomalies with high performance. Therefore selected datasets include normal ECG besides anomalies. Two sources of datasets are used in this study: UCI Arrhythmia Dataset (UCI-ECG) and a dataset composed of new collection of ECGs (OWN-ECG). Both sets are explained in detail in the following paragraphs.

The Arrhythmia dataset from the UCI machine learning repository [6] is used as the main dataset in this study. All the methods are tested and compared via the results from this dataset. This dataset is composed by researchers from Bilkent and Baskent Universities in 1998. The dataset consists of 452 inputs which have 279 features and a class value. Each input corresponds to a different patient. Hence, the dataset enables interpersonal classification of ECG data. 279 features are either numeric or nominal. The first 4 features give personal information: age, sex, weight and height. The rest presents different measurements from the ECG paper. They can be categorized into duration, amplitude or existence of waves. Each measurement is repeated for each lead. That is the reason of the huge number of features. In the original dataset, the inputs are classified into 16 groups. One consists of normal ECG and the rest presents anomalies. In this study, all anomalies are grouped into one class and the resulting dataset contains inputs of two groups. Missing values among the features are removed. This leads a reduction to 278 features and 420 inputs. This resulting dataset is referred as UCI-ECG in the rest of the document. The official dataset description of the UCI Arrhythmia dataset can be found in APPENDIX A.

The second dataset used in this study is a new dataset. It is constructed by collecting ECG outputs from Kardiosis ECG tool with the help of the manufacturing firm TEPA [33]. This is a small collection compared to UCI-ECG. It is used in order to verify this study on real ECG data. There are 28 ECG inputs, 16 of them are normal and the rest 12 of them have cardiac diseases. This dataset is used to verify the best method. Therefore, only the resulting features from the best method are calculated from the inputs. The dataset with the computed features from this new ECG collection is referred as OWN-ECG in the rest of the document.

1.5. Outline

Medical information about human cardiac activities and ECG are presented in the previous parts. Besides, the literature survey on ECG and classification techniques is provided. In the light of the information gained from previous studies, the proposed method is constructed and described in detail. The datasets used in this thesis are also explained. In the following chapters, more details on the proposed method, the applied

tests and test results are provided. Dimension reduction techniques are explained in Chapter 2. Fuzzy membership functions are detailed in Chapter 3. Chapter 4 presents the ECG analysis techniques. Chapter 5 provides tests and test results. Chapter 6 states experiment results for the real life ECG dataset. This document ends with a conclusion part in Chapter 7.

2. DIMENSION REDUCTION

An important pre-processing phase of most data mining applications is the dimension reduction phase. Although high dimensional datasets provide diverse information about the target concept, many features represent the same aspect of the target. The threshold degree of correlation is application specific. On the contrary, there is a concrete reality that more features causes a need for complex algorithms and longer processing time. Therefore, dimension reduction is a major step in data mining.

Two main areas in dimension reduction are: feature selection and feature extraction. The former ranks the available features and chooses a subset beginning from the top rated feature or avoiding from the worst. These correspond to forward and backward selection techniques. Feature extraction aims to compose groupings of features. By this way, the number of resulting features is decreased but the information from most or all of the features are used.

A huge dataset is selected as the main dataset in this thesis. There are 278 features for each input. Hence, the dimension reduction has great importance and impact on the results. Four techniques are selected and applied to the dataset. The resulting subsets are further used in classification and compared with each other. The following subchapters provide detailed information on the dimension reduction techniques used in this study.

2.1. Feature Selection

Feature selection techniques intend to reduce dimension of the dataset by eliminating some features. In order to achieve the appropriate elimination, a ranker is applied before decision. All features are sorted in decreasing importance. Either at one step or iteratively, dimension reduction is applied. Feature selection methods can be formed as forward or backward selection. The former method selects most important features at each iteration. The latter reduces dimension by deleting the least significant feature at each step. The resulting set after feature selection is applied consists of a subset of the original features.

Two feature selection techniques are applied in this study: Recursive Feature Elimination with Support Vector Machine (RFE-SVM) and Correlation Based Feature Selection (CFS). Both of the algorithms are explained in detail in the following sections.

2.1.1. Recursive Feature Elimination with Support Vector Machine

Recursive Feature Elimination is a recursive backward elimination technique. At each iteration, weight values are assigned to all the features in the dataset. A feature with the lowest weight value is eliminated and the algorithm is run with the rest until desired dimension is met. Weight values can be calculated in several ways. One option is to get use of classification techniques. In [34], Support Vector Machine classification is combined with Recursive Feature Elimination and the technique of Recursive Feature Elimination with Support Vector Machine is developed.

The first step in RFE-SVM is to train a linear SVM with all features. The second step is to compute weight vector. The weight vector is calculated with the following function:

$$w = \sum_{i=1}^{l} \alpha_i y_i x_i \tag{2.1}$$

This weight function is taken from the definition of SVM algorithm. l is the input count. x_i is the i^{th} input and y_i is the class label of the i^{th} input. α_i is the dual coefficient which is non-zero only for support vectors. The detailed description of α_i is presented in 4.1.3. Feature ranking is computed by using the weight vector. Ranking score for a feature is the square of the corresponding weight value.

$$rank_i = \left(w_i\right)^2 \tag{2.2}$$

The feature with the smallest ranking score is eliminated from the feature set. This loop is continued until the desired number of resulting features is met. The elimination of features can be done one-by-one or a number of features at an iteration. In case of dimension reduction with huge feature sets, the latter is more preferable.

2.1.2. Correlation Based Feature Selection

Correlation Based Feature Selection is a feature selection algorithm which is proposed in [35]. The ranking is done according to a correlation based heuristic evaluation function. The ranking is done in subset basis rather than feature basis. The aim is to end up with a subset which contains features highly correlated with the class value but uncorrelated with each other.

CFS ranks feature subsets in a search space of all possible subsets. The heuristic function which provides the ranking values is the following:

$$M_S = \frac{k\overline{r_{cf}}}{\sqrt{k + k(k-1)\overline{r_{ff}}}} \tag{2.3}$$

M_S is chosen to represent the heuristic function as it is also referred as the heuristic merit. S denotes the subset of k features. $\overline{r_{cf}}$ is the mean feature-class correlation and $\overline{r_{ff}}$ is the mean feature-feature correlation. The subset with the highest heuristic merit is the output of CFS.

The feature-class and feature-feature correlation computations have the main importance in CFS algorithm. In [35], two alternatives for calculating the inter-correlation between two features are proposed: Relief and MDL. Relief is proposed in [36]. The idea behind the Relief algorithm can be stated as the following:

$$\mathrm{Re}lief_X = P(Xdiffers \mid differentclass) - P(Xdiffers \mid sameclass) \tag{2.4}$$

Relief algorithm states correlation by calculating the difference of probabilities. Relief value of a feature, X, is the probability of class change minus the probability of no class change in case the value of X changes. The mathematical formulation of $\mathrm{Re}lief_X$ can be stated as the following:

$$\mathrm{Re}\,lief_X = \frac{G_{ini}' \times \sum_{x \in X} p(x)^2}{(1 - \sum_{c \in C} p(c)^2) \sum_{c \in C} p(c)^2} \tag{2.5}$$

$$G_{ini}' = \left[\sum_{c \in C} p(c)(1 - p(c)) \right] - \sum_{x \in X} \left(\frac{p(x)^2}{\sum_{x \in X} p(x)^2} \sum_{c \in C} p(c \mid x)(1 - p(c \mid x)) \right) \tag{2.6}$$

In the equations above, c is the class variable. G_{ini}' is an attribute quality measure. As Relief is used as a feature-feature correlation calculation, a symmetrical version that is computed for two features is applied in CFS.

MDL is the minimum description length algorithm proposed in [37]. In [38], MDL is defined as an attribute quality measure in CFS:

$$MDL = \frac{(\mathrm{Pr}\,ior_MDL - Post_MDL)}{n} \tag{2.7}$$

$$\mathrm{Pr}\,ior_MDL = \log_2 \binom{n}{n_1, \ldots, n_C} + \log_2 \binom{n + C - 1}{C - 1} \tag{2.8}$$

$$Post_MDL = \sum_j \log_2 \binom{n_{.j}}{n_{1j}, \ldots, n_{Cj}} + \sum_j \log_2 \binom{n_{.j} + C - 1}{C - 1} \tag{2.9}$$

In the above equations, n is the number of training inputs, C is the class count, $n_{i.}$ is the number of training inputs in class C_i, $n_{.j}$ is the number of training inputs with the j^{th} value of the given feature and n_{ij} is the number of training inputs from class C_i which have the j^{th} value of the given feature.

2.2. Feature Extraction

The objective of feature extraction techniques is to map the original dataset to a dataset with reduced dimensions with minimum loss of information. Different from feature selection, feature extraction algorithms groups the features rather than eliminating one by one. The grouping can be a linear combination or a higher order combination of available features. The idea behind is not to loose different aspects of information but combine them to decrease the processing cost.

Two popular feature extraction techniques are applied in this study: Principal Component Analysis (PCA) and Factor Analysis (FA). Each algorithm is explained in detail.

2.2.1. Principal Component Analysis

Principal Component Analysis is a classic statistical tool for dimension reduction. It depends on the eigenvalues of the covariance matrix of data. It is used as a benchmarking method in all dimension reduction studies.

The first step of preparation for detecting principle components is to subtract the mean of every feature from each dimension. The resulting set is a dataset whose mean is zero. The covariance matrix of the resulting data set is constructed. From the covariance matrix, the eigenvectors and corresponding eigenvalues are computed. The eigenvectors are sorted in descending order. These eigenvectors present the significance of the eigenvector. Hence, most significant eigenvectors are chosen depending on the sorted list of eigenvalues. A principal component is the multiplication of a selected eigenvector and the input data. The PCA output of the dataset is the multiplication of the selected set of eigenvectors and the input data.

The selection of the significant number of eigenvectors is the main problem of PCA. In the case that too many eigenvectors are chosen, PCA doesn’t provide a good dimension reduction. Choosing a small set of eigenvectors leads to loss of information. Therefore

eigenvalues are needed to be correctly investigated before elimination. The best case is that the rejected eigenvalues sum up to a small number presenting a small significance.

2.2.2. Factor Analysis

Factor Analysis is a method to express the variability of the dataset which consists of observed variables by means of fewer unobserved variables. The unobserved variables are called factors. A linear combination of these factors plus error terms is used to model the observed variables.

The decision on the number of factors, m, is the first step of FA. After that, all features are rewritten as a linear combination of these factors and an error term is added.

$$\chi_i = a_1 f_1 + a_2 f_2 + a_3 f_3 + \cdots + a_m f_m + e_i, \forall \mathrm{X}i \in S \tag{2.10}$$

$$\mathrm{X} = AF + \varepsilon \tag{2.11}$$

In Equation 2.10, the multiplication is shown by detail and in Equation 2.11 the matrix form is presented. There are assumptions on these equations. Error terms, e_i, are independent from each other. F and ε are independent from each other. Expected value of factors is 0 and the variance of each factor is 1.

By solving Equation 2.11 for F, the factors are defined. As a smaller number is chosen for factor count, the dimension of the resulting dataset represented by the factors is smaller than the original dataset.

3. FUZZY MEMBERSHIP FUNCTIONS

Second module of the proposed system is a contributory module. This unit is a preparative unit for the following calculations. The results are taken into consideration in the next module while classification via Fuzzy Support Vector Machines. The algorithm of Fuzzy Support Vector Machines is explained in detail in the next chapter. The main advantage in the membership function value assignment can be expressed at this moment for the clarity of the aim of the following calculations. By allocating fuzzy membership function values for each input, a labeling of the input is achieved and the importance of that input while modeling in the algorithm depends on this membership value.

The idea of Fuzzy Support Vector Machines is proposed in [39]. Three fuzzy membership value calculation techniques are suggested in [39]. The first one assigns membership values due to the generation time of the input. This can be applicable only for a restricted set of applications where creation time of the input is meaningful. The second calculation method focuses on the accuracy of classifying one class. Therefore, this method assigns a higher membership value to the target class and lower membership values to other classes. The last method calculates the mean of each class and assigns memberships inversely proportional to the distance of the input point to its class mean. Hence, outliers are assigned smaller membership values and have minimum effect on the classification model.

In this study, five fuzzy membership functions are taken into consideration. The first one is the one class weighing which is mentioned in the above paragraph. The second method is also mentioned in the above paragraph, assigning membership values according to the distance to class mean. The third method is an alternative method which is developed from the second method. This method assigns membership values due to the distance to one class mean. The fourth method depends on the cardinality of same class inputs. Finally, the last method computes a Fuzzy C-Means clustering and uses the membership value output of this algorithm. These techniques are described in the following sections.

3.1. One Class Weighing

This thesis focuses on the correct classification of the ECG samples which contain anomalies. The focus point refers that anomaly class has more importance than normal class for this study. Hence, the application of one class weighing is appropriate for objective of this thesis study. OCW is used in the rest of the document as an abbreviation for one class weighing.

For simplicity, first a two class classification problem is taken into account. Assume that the two classes are a positive class and a negative class as in this study. The membership function depends on two constant values. One is the positive class membership value and the second is the negative class membership value. If the aim is the accurate classification of the positive class, then the weight of the positive class is higher than the other. In case the aim of classifying the negative class has more importance, the corresponding weight of the negative class is higher than the other. In cases which the problem consists of more than two classes, the same manner is applied. The inputs of the target class are assigned higher membership values than other classes.

$$fmf_1(x_i) = \begin{Bmatrix} w_+, x_i \in S_+ \\ w_-, x_i \in S_- \end{Bmatrix} \tag{3.1}$$

The above function, fmf_1, presents the one class weighing membership function for the two class problem. This function assigns w_+ for inputs of the positive class and w_- w_+ for inputs of the negative class. w_+ and w_- can be arranged in order to express the aim of the classification.

3.2. Distance To Class Mean

The outliers are the samples which violates the main class properties. Therefore, they misshape the classification model. In order to prevent the effects of outliers, assignment of low membership values is taken into consideration. The initial requirement is to detect the outliers. In this method, the distance to the class mean is used to detect outliers. DTCM is used in the rest of the document as an abbreviation for distance to class mean.

The idea behind this method is that the core of the class reflects the class properties more accurately. The core of the class is defined as the vicinity of the class mean. Outliers are not in that area and this information proposes the usage of distance to class mean for outlier detection. If an input is at a small distance to class mean, than it is considered as a core class member and has a higher membership value. If an input is at a longer distance to class mean, than it is considered as an outlier and has a smaller membership value.

The class mean is calculated as a point having feature values as the average of each feature among the dataset. The class radii are calculated for normalizing the distance values. The class radii are calculated as:

$$\begin{aligned} r_+ &= \max(\text{dist}(x_+, x_i)), x_+ = meanofS_+, x_i \in S_+ \\ r_- &= \max(\text{dist}(x_-, x_i)), x_- = meanofS_-, x_i \in S_- \end{aligned} \tag{3.2}$$

The dist function is the Euclidean distance function. With the use of class radii, the distance to class mean membership values are calculated with the following function:

$$fmf_2(x_i) = \begin{Bmatrix} 1-[\text{dist}(x_+, x_i)/(r_+ + \delta)], x_i \in S_+ \\ 1-[\text{dist}(x_-, x_i)/(r_- + \delta)], x_i \in S_- \end{Bmatrix}, \delta > 0 \tag{3.3}$$

δ is a very small number greater than zero. The usage of δ prevents division by zero errors. The above function, fmf_2, presents the distance to class mean membership function for the two class problem. This function calculates a membership value which is inversely proportional to the distance of the input to its class mean.

3.3. Distance To One Class Mean

The outliers are defined with respect to a target class in this fuzzy membership value function. If the classification problem focuses on the accurate classification of one class, then the definition of outliers can be refined to outliers of the target class. All the inputs which are members of other classes are treated as outliers and assigned lower membership

values. DTOCM is used in the rest of the document as an abbreviation for distance to one class mean.

This method requires the definition of the target class mean only. The class mean of the target class is calculated as a point having feature values as the average of each feature among the members of this class. The definition and calculation of target class radius is changed in order to take all inputs into consideration. The target class radius is calculated as:

$$r_t = \max(\text{dist}(x_t, x_i)), x_t = t\arg etclassmean, x_i \in S \tag{3.4}$$

The dist function is the Euclidean distance function. The distance to target class mean membership values are calculated with the following function:

$$fmf_3(x_i) = 1 - [\text{dist}(x_t, x_i) / (r_t + \delta)], x_i \in S, \delta > 0 \tag{3.5}$$

δ is a very small number greater than zero. The above function, fmf_3, presents the distance to one class mean membership function for the two class problem. This function calculates a membership value which is inversely proportional to the distance of the input to the target class mean.

3.4. Cardinality

Each class is assumed to be centered at a mean value and most of the class members are assumed to be located in the vicinity of that center. Typical class members are located near the center and the outliers are distant members. The class population is assumed to decrease as the distance to the center increases. Cardinality of a sample input is the number of same class member inputs at a maximum distance of a predefined radius value to that sample. It reflects the class population around the sample point. Therefore, cardinality is a measure of the membership value. CAR is used in the rest of the document as an abbreviation for cardinality method.

$$fmf_4(x_i) = \begin{Bmatrix} \text{count}(x_j \mid \text{dist}(x_i, x_j) \le t), i \ne j, x_i, x_j \in S_+ \\ \text{count}(x_j \mid \text{dist}(x_i, x_j) \le t), i \ne j, x_i, x_j \in S_- \end{Bmatrix}, c > 0 \quad (3.6)$$

The above function, fmf_4, presents the cardinality membership function for the two class problem. This function counts the number of same class members which are located at most a predefined distance, t, to the sample.

3.5. Fuzzy C-Means Clustering

Fuzzy C-Means clustering is a supervised learning technique. FCM is first proposed in [40] and the general case is developed in [41]. The aim of FCM algorithm is to calculate and assign each data point a membership value for each cluster. The cluster count is provided to FCM as a parameter. The membership values are calculated due to the similarity of the sample point and cluster centers. Both cluster centers and membership values are computed in an iterative way. When the stopping condition is met, the iteration ends and the resulting membership values are presented as output.

FCM algorithm is based on minimizing the following objective function:

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^{m} \left\| x_i - c_j \right\|^2, 1 \le m < \infty \quad (3.7)$$

In the objective function, m is the degree of fuzziness. N is the number of inputs and C is the cluster count. x_i is the i^{th} input in the dataset. u_{ij} is the fuzzy membership value of input x_i in cluster j. c_j is the center of cluster j. Finally, || || is any norm representing the similarity between an input and a cluster center.

The minimization of the objective function is an iterative process. As a first step, the membership value matrix is initialized.

$$U^{(0)} = [u_{ij}] \quad (3.8)$$

At each step, the vector of cluster centers is updated using the membership values in that step. The cluster center c_j is updated with the following function:

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} \cdot x_i}{\sum_{i=1}^{N} u_{ij}^{m}} \tag{3.9}$$

After the new cluster centers are calculated, fuzzy membership values are updated due to them. The update function of fuzzy membership values is the following function:

$$u_{ij} = \frac{1}{\sum_{l=1}^{C} \left(\frac{\left\| x_i - c_j \right\|}{\left\| x_i - c_l \right\|} \right)^{\frac{2}{m-1}}} \tag{3.10}$$

The stopping condition checks the change in fuzzy membership values.

$$\left\| U^{(k+1)} - U^{(k)} \right\| < \varepsilon, \varepsilon > 0 \tag{3.11}$$

If the stopping condition is met, the resulting matrix of membership values is $U^{(k+1)}$ and the resulting cluster centers are presented in $C^{(k+1)}$. In $U^{(k+1)}$, each row column represents a cluster and each row represents an input. u_{ij} is the fuzzy membership value of input x_i in cluster j. If there exists a target cluster, the column which represents membership values of that cluster is chosen as fuzzy membership values for the dataset.

4. ECG ANALYSIS

In the previous chapters, pre-processing methods are explained for the main goal of this thesis study which is to correctly detect anomalies in ECG. As a first step in pre-processing, dimension reduction techniques are applied to the dataset. Fuzzy membership value calculation is carried out on the reduced datasets. The results of these phases are the reduced dataset and corresponding fuzzy membership values. The classification module takes these as inputs. Popular classification techniques are applied and their results are compares in order to find the best strategy.

Model construction is the key point in these methods. The reduced dataset is separated into train and test segments. The train segment is analyzed by the classification methods and a model is built. This model is further used to predict the classes of the samples in the test segment. The classification model is also used as a knowledge base and processed to produce a rule base on ECG analysis.

ECG analysis is separated into two modules: classification and rule base development. The techniques used in these modules are explained in the following sections.

4.1. ECG Classification

The ECG classification module is responsible for analyzing the test inputs and labeling them as a member of either the positive class or the negative class. Before the analysis of the test inputs, the classification module is trained with a training set. The product of the training phase is the classification model of ECG. In the test phase, the labeling is accomplished according to the knowledge in the classification model.

Four classification techniques are applied to UCI-ECG dataset: k-nearest neighbor (k-NN), multilayer perceptron (MLP), support vector machine (SVM) and fuzzy support vector machine (FSVM).

4.1.1. k-Nearest Neighbor

k-Nearest Neighbor algorithm is a supervised learning algorithm. It depends on distance calculations. The next sample is labeled by considering k nearest neighboring samples. The selected neighbors vote for the sample in question and the decision is made by choosing the maximum voted label.

The main idea behind k-NN is that class members have common characteristics. The characteristic properties are expressed in terms of features. But some features may be misleading. Hypothetically, in a space of meaningful features, classes occur as populations around class centers. k-NN depends on this distribution. The success of k-NN relies on the selection of the feature space and the distance formula.

The threshold k value is provided by the user. Distance calculations are performed over the train dataset. Various distance formulas can be used within k-NN: Euclidian, Mahalanobis, Minkowski, Manhattan or some content specific distance. Distance values are sorted and k voting samples are chosen in descending distance. The test sample is labeled as the winning label after k voting samples state their votes.

4.1.2. Multilayer Perceptron

Multilayer Perceptron is a feedforward neural network classification algorithm. It is a modification of the standard linear perceptron.

The model of MLP consists of input layer, hidden layer and output layer. The input and output layers are constant in all MLP networks but the number of hidden layers vary. Learning is achieved by changing connection weights after each input is processed, based on comparing the amount of error in the output to the expected result. The comparison is carried out by backpropagation which is defined as a generalization of the least mean squares algorithm.

Because MLP is a standardized neural network algorithm, it is applied to the ECG data in this study for benchmarking.

4.1.3. Support Vector Machine

The technique of support vector machines is proposed and defined in [42, 43, 44]. For a two class classification problem, SVM maps the inputs in the dataset to a high-dimensional feature space and locates the hyperplane to separate the two classes with a maximum margin. The decision on the hyperplane and the calculation of the maximum margin depends on the outliers so called support vectors.

The mathematical formulation of SVM is a quadratic programming problem. Suppose a set S of n labeled inputs is given in the following form:

$$(y_1, x_1), \ldots, (y_n, x_n) \quad \forall x_i \in R^N, \forall y_i \in \{1, -1\} \tag{4.1}$$

Each N dimensional data point x_i is member of either of the two classes which is represented by y_i. Generally, the two classes are referred as the positive and the negative classes and 1 and -1 are used as class labels. In order to make easier the search for the optimal hyperplane, the inputs are mapped to a higher order feature space Z. The mapping is accomplished by a function called a kernel function. The input points x_i are mapped to z_i in Z with the use of the kernel function in the form $z_i = \varphi(x_i)$. In Z, the hyperplane is formulated as:

$$w \cdot z + b = 0 \tag{4.2}$$

$$y_i(w \cdot x_i + b) \geq 1 \tag{4.3}$$

The parameters w and b in Equation 4.2 are selected to enable the inequality defined in Equation 4.3. For a linearly separable set S, a unique optimal hyperplane can be localized with a maximum margin between classes. For the non-separable case, Equation 4.3 is changed into Equation 4.4 by introducing non-negative slack variables, ξ_i.

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i \tag{4.4}$$

In the non-separable case, the optimal hyperplane is found by solving the following optimization problem:

$$\begin{aligned} &minimize\ \frac{1}{2} w \cdot w + C \sum_{i=1}^{n} \xi i \\ &subject\ to \\ &y_i (w \cdot z_i + b) \geq 1 - \xi i \quad i = 1, ..., n \\ &\xi i \geq 0 \quad i = 1, ..., n \end{aligned} \tag{4.5}$$

The optimization problem is a quadratic programming problem and can be solved by transforming into Lagrangian form:

$$\begin{aligned} &maximize\ W(\alpha) = \sum_{i=1}^{n} \alpha i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j z_i \cdot z_j \\ &subject\ to \\ &\sum_{i=1}^{n} y_i \alpha_i = 0 \\ &0 \leq \alpha_i \leq C \quad i = 1, ..., n \end{aligned} \tag{4.6}$$

According to the Kuhn–Tucker theorem, the solution of the Equation 4.6 satisfies the following equations:

$$\overline{\alpha_i} \left(y_i \left(\overline{w} \bullet z_i + \overline{b} \right) - 1 + \overline{\xi i} \right) = 0, i = 1, \ldots, n \tag{4.7}$$

$$\left(C - \overline{\alpha_i} \right) \overline{\xi i} = 0, i = 1, \ldots, n \tag{4.8}$$

In Equation 4.6, $z_i \cdot z_j$ can be expressed by multiplication of kernel functions as $\varphi(x_i) \cdot \varphi(x_j)$ or by a generalized kernel function form as $K(x_i, x_j)$. By updating Equation 4.6 with the generalized Kernel form, the non-linear separating hyperplane can be found with the following formula:

$$
\begin{aligned}
&maximize\, W(\alpha) = \sum_{i=1}^{n} \alpha i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(xi, xj) \\
&subject\ to \\
&\sum_{i=1}^{n} y_i \alpha_i = 0 \\
&0 \le \alpha_i \le C \quad i = 1, \ldots, n
\end{aligned}
\tag{4.9}
$$

and the decision function is:

$$f(x) = sign(w \cdot z + b) = sign(\sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b) \tag{4.10}$$

The decision function is used in order to classify the inputs. If the decision function returns positive, the input is classified as a member of the positive class. If the decision function returns negative, the input is classified as a member of the negative class. The equation in the decision function is equal to 0 for the support vectors.

4.1.4. Fuzzy Support Vector Machine

SVM method considers all the inputs as having equal weight. This assumption leads the outliers to affect the classification model as a representing class member. In order to overcome this situation and provide a solution for assigning weight values to inputs, fuzzy support vector method is introduced [39]. A membership function is defined for the dataset and weight values for the inputs are calculated as the degree of membership. The degree of membership values are attached to the dataset and the new dataset is in the following form where s_i is the degree of membership value for input i:

$$(y_1, x_1, s_1), \ldots, (y_n, x_n, s_n) \quad \forall x_i \in R^N, \forall y_i \in \{1, -1\}, \forall s_i \in \{0, 1\} \tag{4.11}$$

Equation 4.9 is rewritten for FSVM with membership values as:

$$
\begin{aligned}
&maximize\, W(\alpha) = \sum_{i=1}^{n} \alpha i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(xi, xj) \\
&subject\ to \\
&\sum_{i=1}^{n} y_i \alpha_i = 0 \\
&0 \le \alpha_i \le s_i C \quad i = 1, ..., n
\end{aligned}
\tag{4.12}
$$

The Kuhn-Tucker conditions are:

$$\overline{\alpha_i}\left(y_i\left(\overline{w} \bullet z_i + \overline{b}\right) - 1 + \overline{\xi i}\right) = 0, i = 1, \ldots, n \tag{4.13}$$

$$\left(s_i C - \overline{\alpha_i}\right)\overline{\xi i} = 0, i = 1, \ldots, n \tag{4.14}$$

FSVM method introduces the concept of membership values into SVM leading the need of defining membership functions. A membership value provides a degree of the representativeness of the input for its class. Hence, the membership function has to be formed to measure this representativeness. For calculating degree of membership, classical membership functions can be used as well as data specific membership functions. In this study, the functions described in Chapter 3 are used for assigning the degree of membership.

4.2. Rule Base Development

Classification methods construct a model of the dataset in the training phase. This model is used to label the test data in the test phase. The classification model is a data structure in machine language and does not provide meaningful and readable information for human beings. In order to translate the classification model into human language, rule base development techniques are used. These techniques process the dataset and extract the information of the definition of the dataset in a readable form.

In this study, 4 rule base extraction methods are run on ECG classification model: C4.5 Decision Tree (C4.5), Partial Decision Tree (PART), Repeated Incremental Pruning

to Produce Error Reduction (RIPPER) and Adaptive Neuro-Fuzzy Inference System (ANFIS). The methods are described in detail in the following sections.

4.2.1. C4.5 Decision Tree

C4.5 Decision Tree algorithm is proposed by Quinlan in [39]. It is an extension to ID3 algorithm which is also proposed by Quinlan. In addition to ID3, C4.5 is capable of pruning and handling continuous and discrete attributes, missing values.

Features in the dataset are sorted in descending entropy values. The first feature in the list is taken as the root of the decision tree. The root feature is divided into its children via defining a separation value. The next node is selected as the feature with maximum entropy and which is not used in the path from the root. By adding features in this manner, the whole tree is constructed.

In rule base development process, tree construction is accomplished with the training set. Therefore, the constructed tree may be very specific to inputs in the training set. This handicap is solved by pruning the decision tree. Pruning is done by replacing a subtree by a leaf node. The replacement takes place according to the expected error rate in the subtree being greater than in the single leaf.

After tree construction and pruning, the resulting decision tree represents the rule base. Each leaf node corresponds to a rule which can be rewritten by adding up the conditions of features from the root to the leaf node.

4.2.2. Repeated Incremental Pruning to Produce Error Reduction

RIPPER algorithm is proposed by Quinlan [46]. It implements a separate-and-conquer approach for decision tree generation. After each rule generation, the training inputs covered by the rule are removed and the process is continued with the rest of the training set. The RIPPER method has a post-processing phase in which each rule is either replaced or revised. The choice for using the original rule or its replacement or revision is made depending on the minimum description length criteria.

4.2.3. Partial Decision Tree

Partial Decision Tree algorithm is proposed in [47]. This method combines both C4.5 and RIPPER algorithms and avoids global optimization phase. PART algorithm runs as a loop. A rule is built from the dataset as the first step in the loop. A pruned decision tree is constructed for each rule creation and the leaf node with maximum coverage is taken as the rule of that iteration. After the rule is decided, the instances which are covered by this rule are removed from the dataset. This rule creation loop continues until all inputs are processed.

4.2.4. Adaptive Neuro-Fuzzy Inference System

An adaptive network consists of nodes and connections between them. The adaptability comes in account such that the output of each node depends on the node's parameters. These parameters are decided on the learning phase in order to reduce error measure. A fuzzy inference system consists of a knowledge-base representing the data, a decision-making unit, a fuzzification and a defuzzification unit. An adaptive neuro-fuzzy inference system combines the two and produces a feedforward neural network with supervised learning capability.

ANFIS provides a type-3 neuro-fuzzy inference system based on Takagi or Sugeno type if-then-else rules. In a type-3 FIS, the output of each rule is constructed as a linear combination of input variables plus a constant term. The final output of a type-3 FIS is the weighted average of each rule's output.

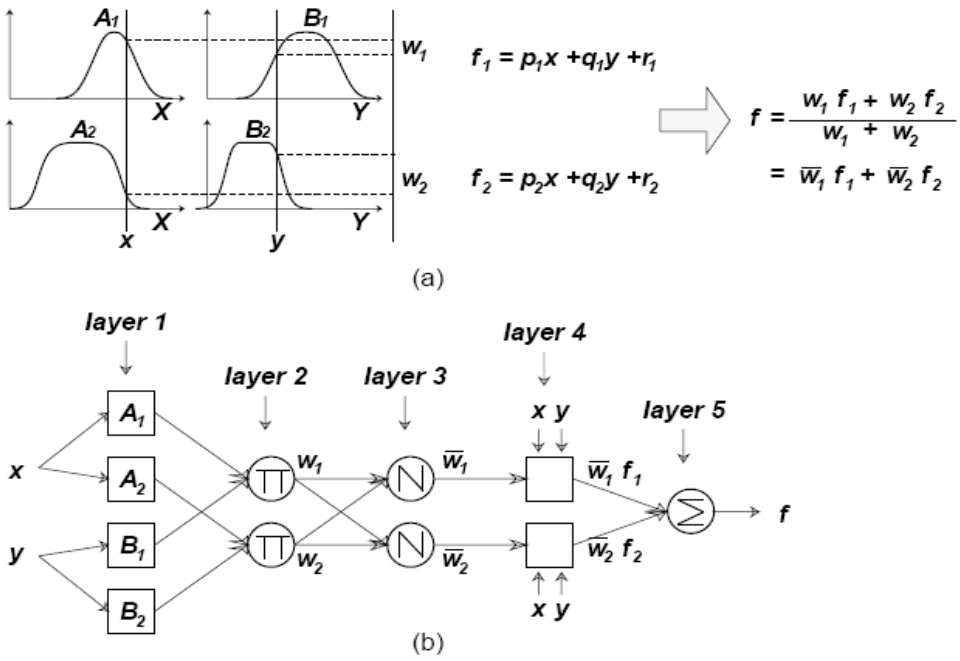


Figure 4.1 ANFIS Structure a) Type-3 Reasoning b) Type-3 ANFIS [48]

In Figure 4.1, the type-3 reasoning and the corresponding ANFIS structure can be examined. There are 5 layers in ANFIS network. The first layer is the input layer and each node in this layer represents a membership function. The second layer provides a combination of membership functions from the previous layer. The third layer calculates the firing strength of rules. The fourth layer combines the output of layer 3 with node parameters for adaptability. The fifth layer computes the output as a summation of incoming values.

In ANFIS, hybrid learning is used. The parameter identification is accomplished in the forward pass of the hybrid learning algorithm. The network is run until the fourth layer and the parameters are identified by the least squares estimate. The parameters are updated and finalized by the gradient descent in a backward pass in which the error rates propagate backward.

5. EXPERIMENTAL RESULTS

This research aims to identify ECG anomalies correctly. The main dataset of the study is the UCI Arrhythmia dataset, UCI-ECG. A classification model is constructed on this dataset. The proposed method is a combination of several techniques in each module.

The first module is the dimension reduction module. The UCI-ECG dataset has 278 features. In order to minimize the cost of processing huge number of features, the number of features is reduced in this module. Both feature selection and feature extraction methods are applied to the dataset. RFE-SVM and CFS are used as feature selection algorithms. PCA and FA are used as feature extraction algorithms. The outputs of these methods produce the experiment datasets.

The second method is the fuzzy membership value calculation module. This module is a preparation module for FSVM algorithm. The degree of membership of each input is expressed as a numeric value by means of membership functions. Five membership functions are applied for comparison: one class weighing, distance to class mean, distance to one class mean, cardinality and FCM.

The third module is the classification module. It is the main module of the proposed method. This module is responsible for learning the class characteristics and predict for test inputs. k-NN, MLP, SVM and FSVM are the classification techniques applied in this module. The results of these techniques are compared with respect to several performance metrics.

The last module is the rule base development module. This module takes the model produced by the classification algorithm and interprets the knowledge in the model in a way close to human language. The interpretation is grouped in a rule base. The resulting rule base is compared with expert opinion.

This chapter presents the experiments performed for the research and compares and interprets the results in order to reach the best ECG analysis method. The selected method

is applied on the OWN-ECG in order to verify the results. OWN-ECG is a small ECG dataset which is constructed for this research study.

5.1. Performance Metrics

The proposed method includes four main classification algorithms. Each algorithm is tested with several sets of parameters. Hence, the research produces a result set which is a combination of all outputs from algorithms. In order to compare the results and the algorithms, performance metrics are used. The basic metrics are the pure classification results. They are expressed as four metrics:

- TP: true positive predictions, input is a member of positive class and predicted as so
- FN: false negative predictions, input is a member of positive class and predicted as negative
- TN: true negative predictions, input is a member of negative class and predicted as so
- FP: false positive predictions, input is a member of negative class and predicted as positive

Other performance metrics are calculations on these four criteria. The main performance measure is accuracy. It is a measure of closeness of the predictions to the actual values. The mathematical formula of accuracy is the following:

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \tag{5.1}$$

The second important performance metric is f-score. But precision and recall have to be explained before f-score for the ease of understanding. Precision is a measurement of consistency, the ability of producing the same prediction under changing conditions. Recall is the proportion of exact detection of the target class to the total targeted inputs. Recall is also called sensitivity or the true positive rate.

$$\Pr ecision = \frac{TP}{TP + FP} \tag{5.2}$$

$$\mathrm{Re}call = \frac{TP}{TP + FN} \tag{5.3}$$

F-score is a summary of both precision and recall. It is calculated as the harmonic mean of both.

$$F - Score = \frac{2 \times \Pr ecision \times \mathrm{Re}call}{\Pr ecision + \mathrm{Re}call} \tag{5.4}$$

The next performance metric is Matthews correlation coefficient. It is considered as a balanced metric as both positive and negative measures are taken into account and it presents consistent values even if class counts are unbalanced.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{5.5}$$

Specificity is also considered in comparing the methods. Specificity is the proportion of negative class members which are correctly predicted.

$$Specificity = \frac{TN}{TN + FP} \tag{5.6}$$

The last performance metric is the receiver operating characteristic analysis, referred as ROC curve analysis. The ROC curve is a plot of sensitivity versus 1-specificity. A classification method is as good as the area under the ROC curve. Therefore, area under ROC curve, AUC, is also a performance metric.

5.2. Experimental Results for Dimensionality Reduction

UCI-ECG dataset has 278 features. In order to minimize the cost and time complexity of the proposed method, the number of features has to be reduced. The

dimension reduction also clears the correlation between features. The balance of correlation removal and information gathering is the main point in dimension reduction techniques.

Both feature selection and feature extraction algorithms are applied on normalized form of UCI-ECG. The details of dimension reduction experiments and result are presented in following sections.

5.2.1. Experimental Results for Feature Selection

In feature selection algorithms, dimension reduction is achieved by removing the unnecessary features from the dataset. Features to be removed are selected by ranking. The resulting dataset consists of the most informative features.

5.2.1.1. Recursive Feature Elimination with Support Vector Machine. WEKA platform is used for this method. SVMAttributeEval algorithm in WEKA corresponds to RFE-SVM. The algorithm takes the size of the resulting feature set and the number of features to be eliminated at each iteration as parameters. 10 features are eliminated at each iteration. 7 sets with different feature counts are constructed and labeled as RFE-SVMi where i represents the number of features. RFE-SVM3, RFE-SVM5, RFE-SVM7, RFE-SVM10, RFE-SVM13, RFE-SVM15 and RFE-SVM20 are the resulting datasets.

5.2.1.2. Correlation Based Feature Selection. CfsSubsetEval method in WEKA platform is used for this algorithm. Backward selection is applied and 10 features are eliminated at each iteration. 7 sets with different feature counts are constructed. The datasets are labeled as CFSi where i is the number of features. CFS3, CFS5, CFS7, CFS10, CFS13, CFS15 and CFS20 are the resulting datasets.

5.2.2. Experimental Results for Feature Extraction

Feature extraction algorithms aim to combine features to form new features and decrease the size of feature set. The important point in feature extraction is to clear correlation but not to loose information.

5.2.2.1. Principal Component Analysis. PCA algorithm is applied by the PrincipalComponents method in WEKA platform. Proportion of variance is selected as 0.95. The resulting datasets are PCA3, PCA5, PCA7, PCA10, PCA13, PCA15 and PCA20 where the number in the label represents the feature count.

5.2.2.2. Factor Analysis. *factoran* function in Matlab is used for FA [49]. The original UCI-ECG dataset is huge for being a parameter for *factoran* function. Hence, a smaller set is used as an input to FA. On account of the main proposed method being FSVM which is built on SVM algorithm, the largest RFE-SVM output is selected as the starting dataset for FA. *factoran* is run on RFE-SVM20. The output of *factoran* is a matrix of factor coefficients for each feature. With the help of the coefficient matrix and the initial feature set, the expressions for factors are formed. The RFE-SVM20 input set enables FA to create 5 different sized datasets: FA3, FA5, FA7, FA10 and FA13.

5.2.3. Empirical Feature Selection

The structure of the ECG signal and the interpretation details from the expert's view is presented in the first chapter. ECG signal is characterized by the P, Q, R, S and T waves. The duration and amplitude of the waves are important. The length of the intervals between P wave and QRS complex and QRS complex and T wave are meaningful for defining the regularity. Heart rate is another indicator of the regularity of the cardiac activity.

On account of the characteristic properties listed in the above paragraph, an empirical feature set is constructed. This dataset is referred as EMP-ECG in the rest of this document. This empirical dataset is used as a reference dataset in order to check the success of the applied methods. The list of the features in the empirical dataset is presented in Table 5.1.

Table 5.1 Feature list of EMP-ECG dataset

Feature Label	Explanation
QRSDuration	Average of QRS duration in msec
P-Rinterval	Average duration between onset of P and Q waves in msec
Q-Tinterval	Average duration between onset of Q and offset of T waves in msec
Tinterval	Average duration of T wave in msec
Pinterval	Average duration of P wave in msec
Heartrate	Number of heart beats per minute
V2_Amplitude_Q_wave	Amplitude of Q wave in V2 Lead in milivolt
V2_Amplitude_R_wave	Amplitude of P wave in V2 Lead in milivolt
V2_Amplitude_S_wave	Amplitude of S wave in V2 Lead in milivolt
V2_Amplitude_P_wave	Amplitude of P wave in V2 Lead in milivolt
V2_Amplitude_T_wave	Amplitude of T wave in V2 Lead in milivolt

The histogram of the features in EMP-ECG is presented in Figure 5.1. The normalized form of the EMP-ECG is used for the histograms. The last feature labeled as CLASS carries the normal-abnormal information of the instance. Abnormal ECG instances are grouped in the positive class and the normal ECG instances fall in the negative class. The positive class is represented with the blue color and red is used to represent the negative class. From the histograms, it can be clearly observed that the anomaly class covers a widespread area in all of the features whereas the normal class is restricted in a small interval. Another noticeable point is that, the mean values of both class distributions are almost the same. This leads to the fact that an anomaly can have partial affects on the ECG signal.

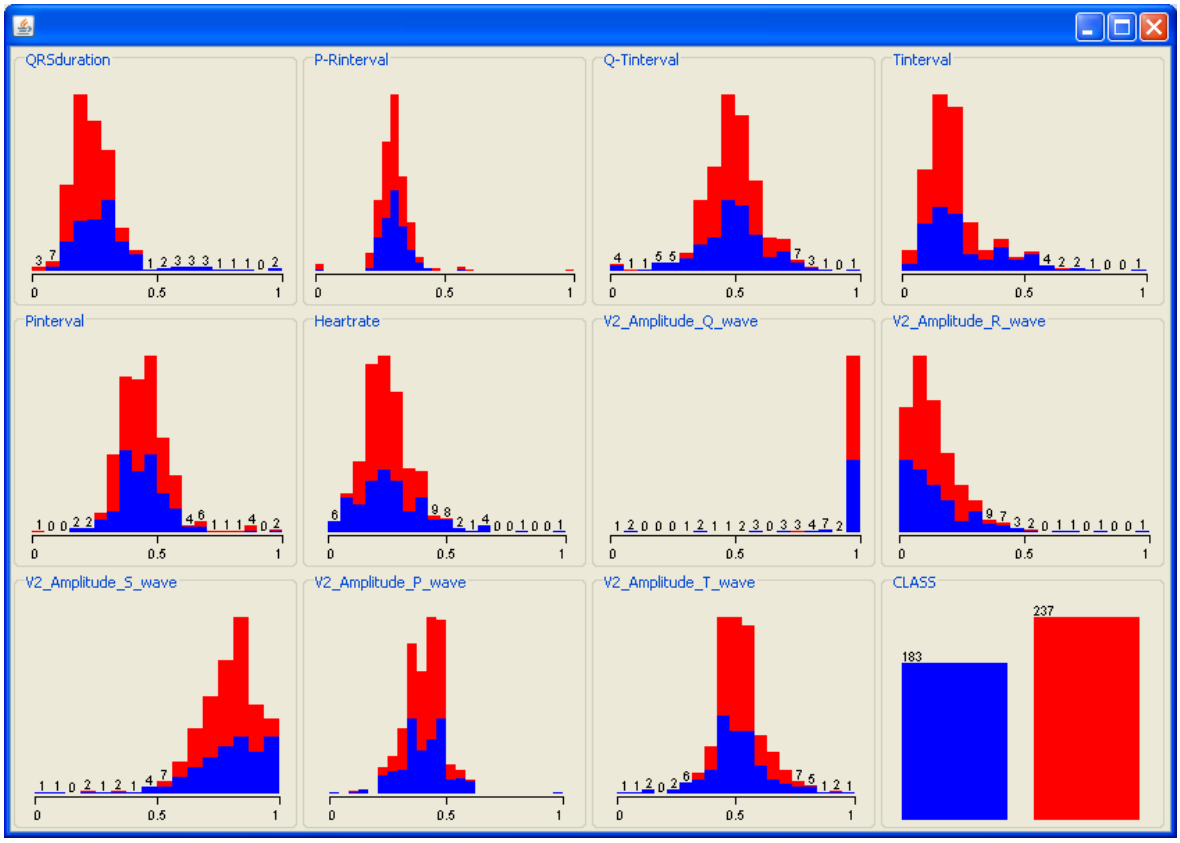


Figure 5.1 EMP-ECG dataset

5.3. Experimental Results for Fuzzy Membership Value Calculation

Fuzzy membership values are calculated in order to be used in FSVM algorithm. Different approaches lead to different membership functions. Five types of membership functions are applied in this research.

One class weighing function directs the focus of the classification method to a specific class. Distance to class mean function calculates distance of each input to its class mean and assigns membership values inversely proportional to these distances. Distance to one class mean function differentiates the former by calculating all distances to the target class mean. Cardinality membership function counts same class inputs in the vicinity of the sample. As a last method, the output of FCM clustering algorithm is used as membership values.

The experiment steps for applying these functions and their results are discussed in the following sections.

5.3.1. One Class Weighing

The purpose of this research is to detect ECG anomalies accurately. According to this purpose, the analysis is constructed in favor of the anomaly class. One class weighing function for fuzzy membership calculation reflects this favor. The general form of the membership function can be found in the third chapter. For this study, the following membership function is applied for one class weighing:

$$fmf_1(x_i) = \begin{Bmatrix} 0.8, x_i \in S_+ \\ 0.2, x_i \in S_- \end{Bmatrix} \tag{5.7}$$

All inputs belonging to the positive class are assigned 0.8 as the membership value. The remaining instances which belong to the negative class are given 0.2 as the membership value. While the instances of the positive class dominate to shape the classification model, the characteristics of the negative class are also taken into account for better performance.

5.3.2. Distance To Class Mean

Distance to class mean function is used to assign membership values. The first step is to locate the class means for all of the reduced datasets. Class means are located by calculating the average of each feature in the dataset and taking them as dimensions of the class mean. The next step is to apply the distance to class mean membership function on all of the inputs. For ease of referencing, the formula is repeated here.

$$fmf_2(x_i) = \begin{Bmatrix} 1-[\text{dist}(x_+, x_i)/(r_+ + \delta)], x_i \in S_+ \\ 1-[\text{dist}(x_-, x_i)/(r_- + \delta)], x_i \in S_- \end{Bmatrix}, \delta > 0 \tag{5.8}$$

x_+ is the positive class mean, x_- is the negative class mean. $r_+ + \delta$ and $r_- + \delta$ are normalizing terms. r_+ is the maximum distance from x_+, similarly r_- is the maximum distance from x_-. δ is a term in order to prevent from division by zero error. Euclidean distance is used in this research as the dist function. By normalizing the distance value and subtracting from 1, a membership value inversely proportional to the distance is calculated for each input.

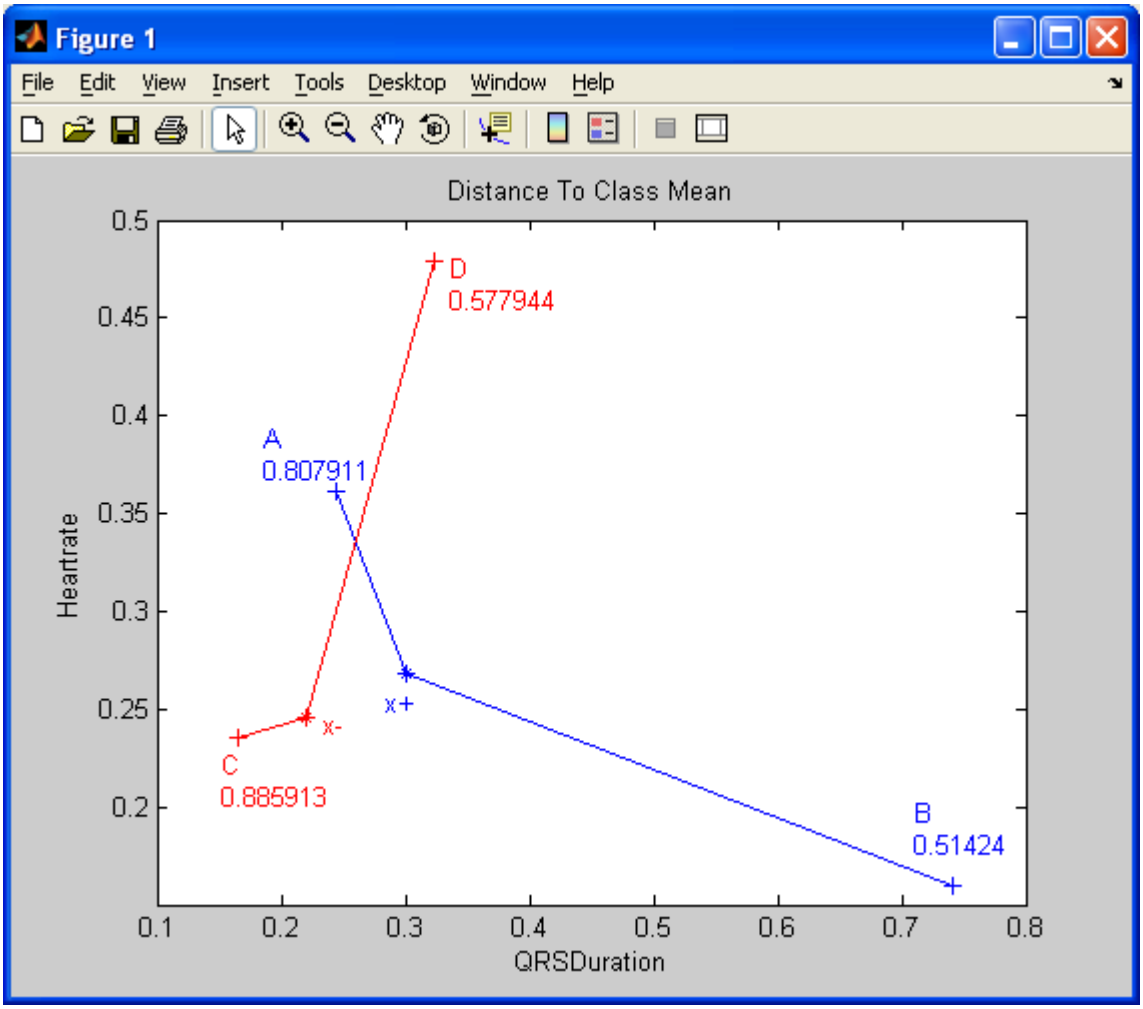


Figure 5.2 Illustration of DTCM method

Examples are provided in Figure 5.2 for clarifying the DTCM method. QRSDuration and Heartrate are chosen as axes. Points A and B are positive class members, C and D are negative class members. Euclidean distance is used in calculations. Hence, distances can be represented by drawing a line between the class centers and members. A is closer to the

positive class center than B. The assigned membership value of A is 0.807911. The assigned membership value of B is smaller than that and is 0.51424. The same situation is observed for the negative class. C is closer to the negative class center than D. The membership value of C, 0.885913, is bigger than the one for D, 0.577944.

5.3.3. Distance To One Class Mean

Distance to one class mean function is an alternative for DTCM. First, the class means are located as DTCM. Then, distance of each input to the positive class mean is calculated. Euclidean distance is applied for distance calculations. For ease of referencing, the formula is repeated here.

$$fmf_3(x_i) = 1 - [\text{dist}(x_t, x_i) / (r_t + \delta)], x_i \in S, \delta > 0 \tag{5.9}$$

The difference between Equation 5.8 and Equation 5.9 is that only the positive class mean is used in distance calculations. Inputs are assigned weight values inversely proportional to their distance to the positive class mean.

5.3.4. Cardinality

In this research, the term cardinality refers to the count of same class members at a maximum distance of a threshold value. For this study, the following membership function is applied for cardinality:

$$fmf_4(x_i) = \begin{Bmatrix} \text{count}(x_j \mid \text{dist}(x_i, x_j) \le 0.1), i \ne j, x_i, x_j \in S_+ \\ \text{count}(x_j \mid \text{dist}(x_i, x_j) \le 0.1), i \ne j, x_i, x_j \in S_- \end{Bmatrix}, c > 0 \tag{5.10}$$

0.1 defines the threshold value, the maximum distance value, in the cardinality calculation. This threshold value is decided after several trials and the one which maximizes the accuracy is chosen. For each input, a spherical area with radius 0.1 is drawn and the number of same class members which lie in the sphere is counted. The resulting number is assigned as the fuzzy membership value.

5.3.5. Fuzzy C-Means Clustering

Fuzzy C-Means clustering is applied to the reduced datasets in order to produce fuzzy membership values. The *fcm* method in Matlab is used for FCM algorithm. The cluster count and the degree of fuzziness are provided as parameters for the *fcm* method. The cluster count is taken as 2. The default degree of fuzziness value in Matlab is 2. CFS7 dataset is used as an example dataset for investigating the most efficient degree of fuzziness value. 1.57 is found to produce better results than other trials. Hence, experiments are run for both 2 and 1.57 as degree of fuzziness.

fcm outputs membership values for both classes. As the main purpose is to detect the anomalies, the membership values for the positive class are used in the further methods.

5.4. Experimental Results for ECG Classification

The experiments which are carried out in the classification module are explained in the following sections. Four classification algorithms are used in experiments: k-NN, MLP, SVM and FSVM. Each algorithm is used with different sets of parameters. Therefore, this study promises to provide intra-method comparison besides the inter-method comparison.

5.4.1. k-Nearest Neighbor

k-NN algorithm predicts the class label of an instance depending on its neighbors. *IBk* method in WEKA platform is used for k-NN algorithm. The input k presents the neighbor count to consider. Euclidean distance is used for distance calculations. 6 different k values are used in the experiments. Odd numbers are chosen for k in order to overcome the equivalence case. 10 fold cross-validation is used in experiments. Accuracy, f-score and MCC are mentioned in the figures in this section. All results of k-NN experiments can be found in 0.

When accuracy results are considered, Figure 5.3, KNN15 is found to be consistent among the trials. The highest accuracy rates are reached by KNN7 which provides 79.28% for RFE-SVM13 and 79.52% for RFE-SVM7 and KNN9 which provides 79.76% for RFE-

SVM7. But when overall performance, which can be calculated by averaging the accuracy results for the datasets, is taken into account, KNN15 is more accurate. KNN15 provides an average of 73.27% accuracy for the ECG datasets whereas the average accuracy is 73% for KNN7 and 73.15% for KNN9.

F-score results are demonstrated in Figure 5.4. KNN7 is shown to outperform other k-NN trials. When RFE-SVM7 dataset is considered, 0.73 is scored by KNN7 whereas 0.732 is scored by KNN9 and 0.72 is scored by KNN15. KNN7 outperforms KNN9 and KNN15 for CFS and PCA datasets. The average f-score value for KNN7 is 0.62 which is higher than 0.618 for KNN9 and 0.614 for KNN15.

When Mathew's Correlation Coefficient is considered, KNN15 produces the best performance. The average of MCC results is 0.462 for KNN15 which is greater than 0.452 for KNN7 and 0.457 for KNN9. When all three performance metrics are taken into account, KNN15 is found to be the leading k-NN trial. In the inter-comparison of the classification methods, KNN15 is used to represent k-NN method.

A consensus on the behavior of k-NN trials on datasets can be drawn when the three graphs are considered together. The highest values are obtained for RFE-SVM datasets. FA results are the second highest and CFS results comes the third. k-NN values for PCA datasets are not as good as the other datasets. k-NN results for EMP-ECG are not satisfying. These results show that RFE-SVM and FA datasets have a more concentrated distribution in their feature space. Therefore, the neighboring instances provide a useful and correct opinion on the label of the sample in question.

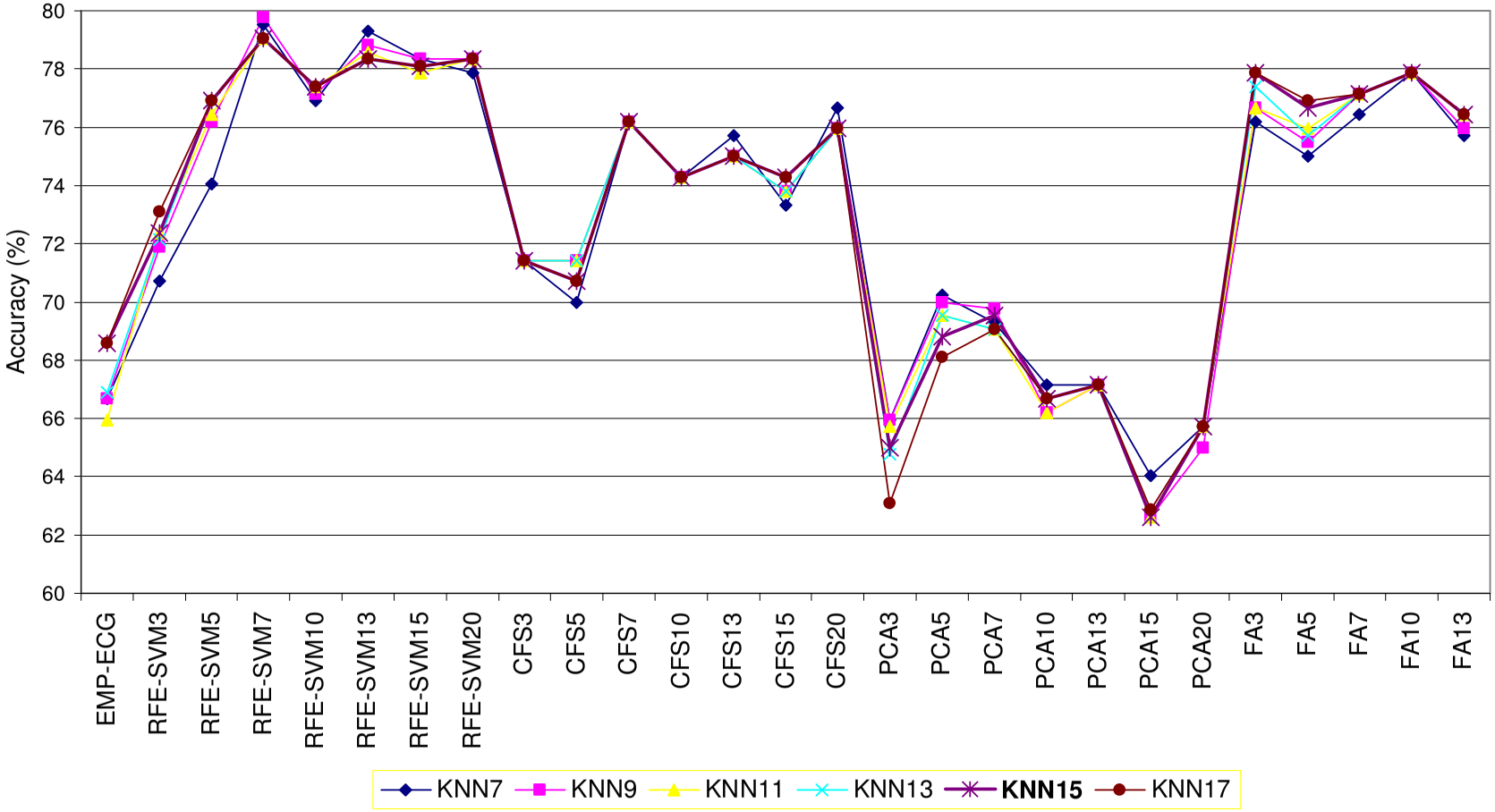


Figure 5.3 Accuracy test results of k-NN

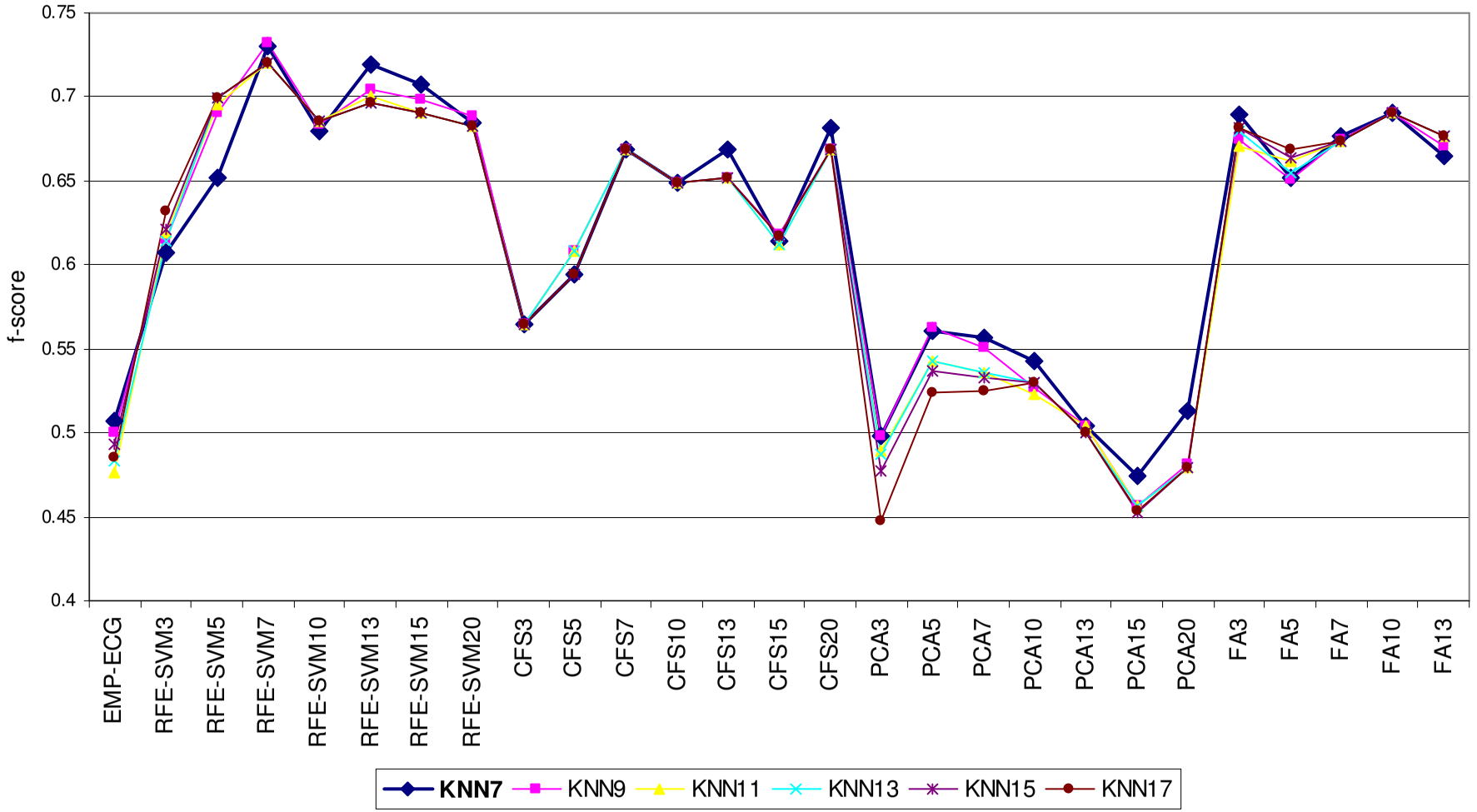


Figure 5.4 f-score test results of k-NN

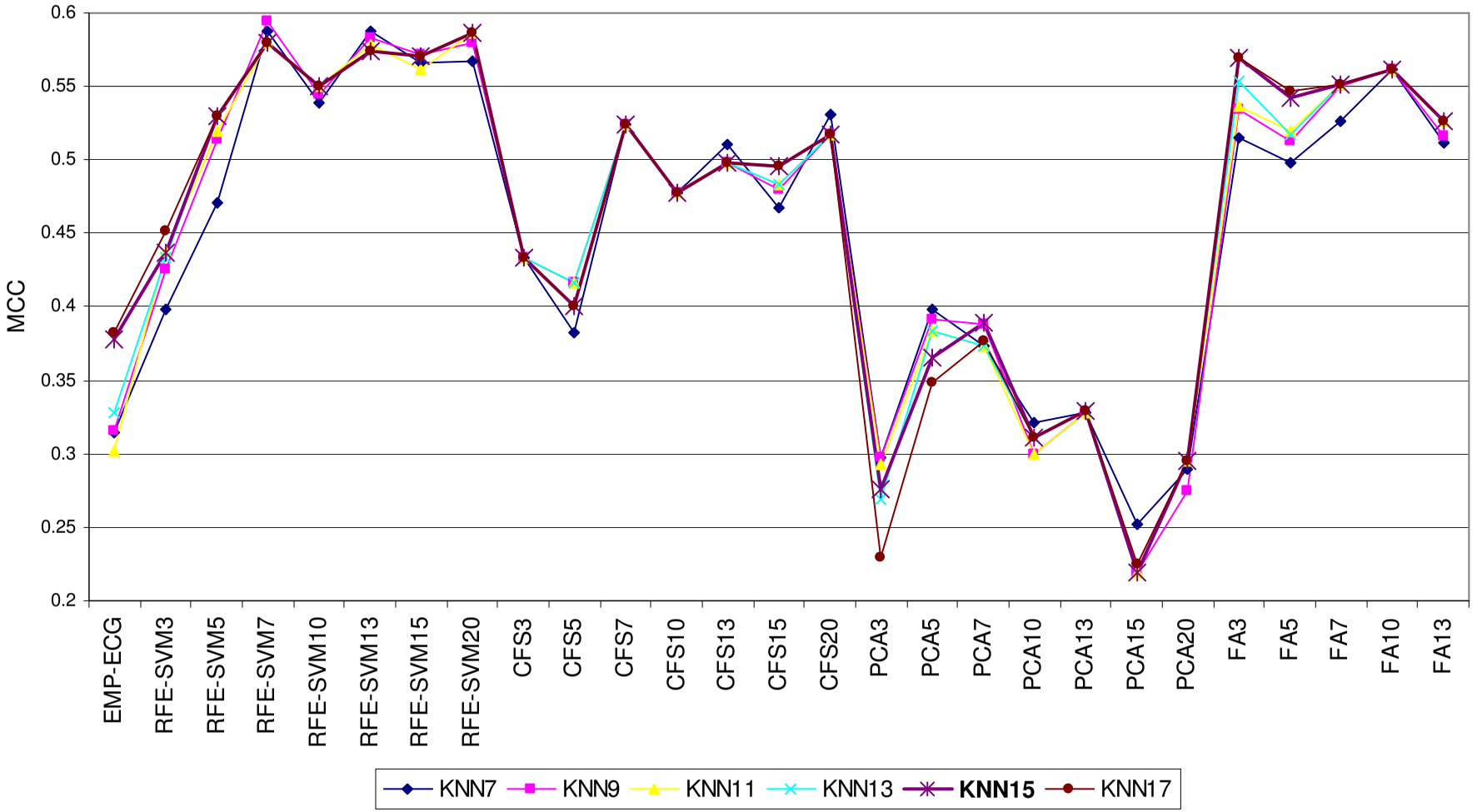


Figure 5.5 Matthews correlation coefficient test results of k-NN

5.4.2. Multilayer Perceptron

MLP is applied as *MultilayerPerceptron* in WEKA platform. *MultilayerPerceptron* uses backpropagation for classification. All nodes that are produced by *MultilayerPerceptron* are sigmoid nodes. The number of hidden nodes is decided based on the number of features and classes.

$$N_{hiddennodes} = (N_{features} + N_{classes}) / 2 \tag{5.11}$$

The reset parameter is set to true. This enables to reset the network and begin training again in case of the network diverges. The number of epochs is set to 500. 10 fold cross-validation is used in experiments. The accuracy, f-score and MCC results are presented here. All results of MLP experiments can be found in APPENDIX C.

Considering the accuracy results, MLP performs the best for CFS datasets, except CFS3. CFS3 dataset provides an accuracy rate of 70.23% while the average accuracy result for other CFS datasets is 79.36%. The three featured sets provide less accurate results for all of the dataset types. RFE-SVM datasets are classified with the second best accuracy rates by MLP with an average of 78.96% except RFE-SVM3. FA and EMP-ECG follow them. The worst results are obtained for PCA datasets. MLP method provides the maximum accuracy for FA10 dataset with 81.66%.

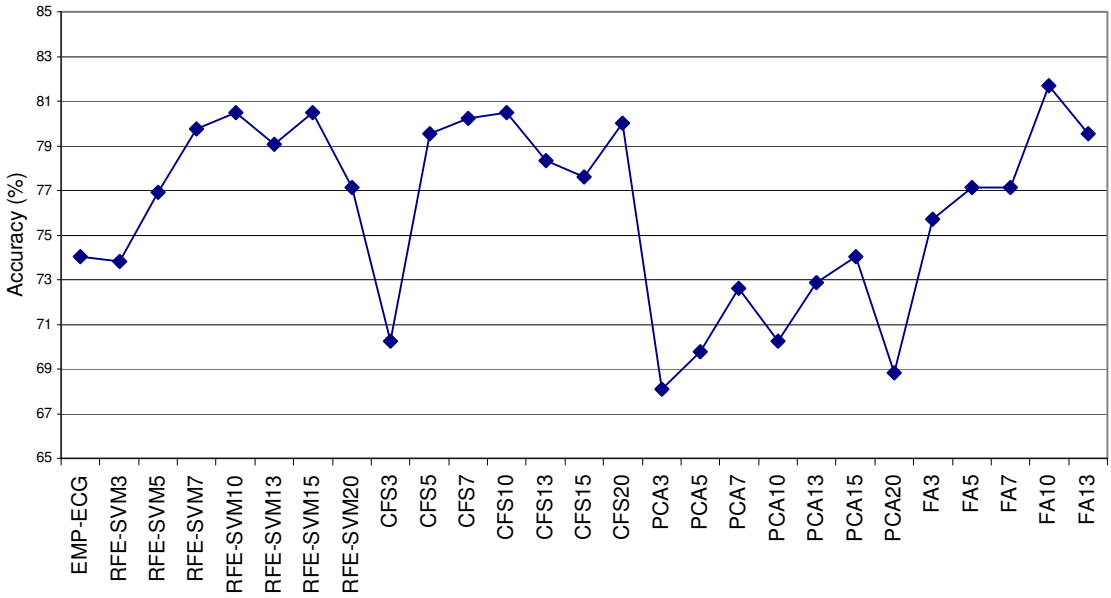


Figure 5.6 Accuracy test results of MLP

MLP method also provides the maximum f-score for FA10 dataset with 0.776. f-score results for CFS and RFE-SVM datasets are leading with an average of 0.747 for CFS and 0.729 for RFE-SVM. The three featured dataset results are not calculated in these averages. FA and EMP-ECG follow them. The last set in the f-score ranking is PCA. The MCC curve presents similar ranking for the performances.

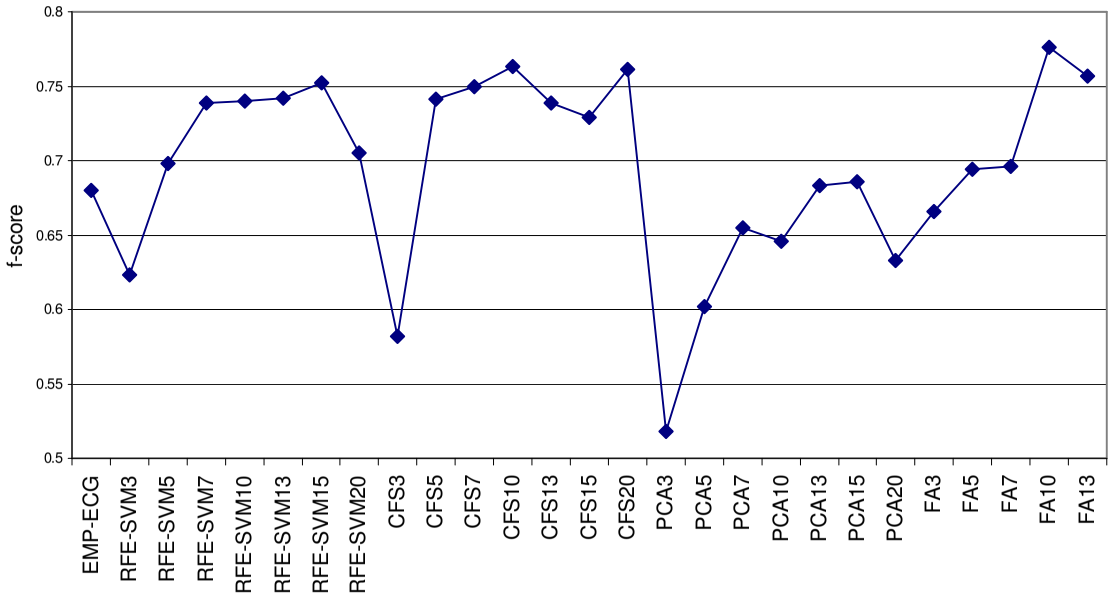


Figure 5.7 f-score test results of MLP

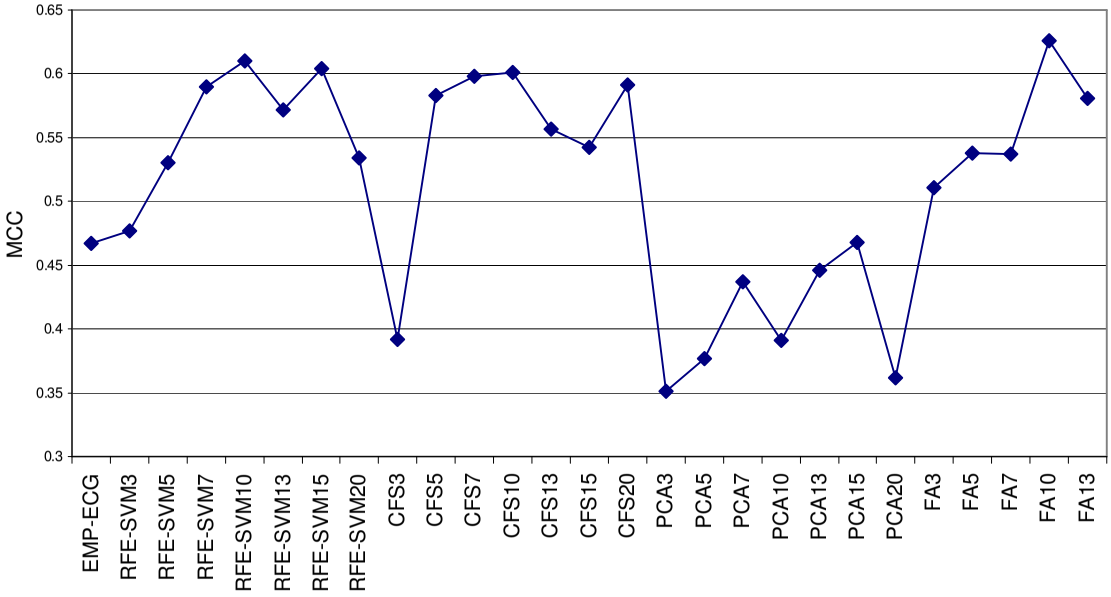


Figure 5.8 Matthews correlation coefficient test results of MLP

5.4.3. Support Vector Machine

SVM algorithm detects the support vectors among the inputs and shapes the class separations according to these support vectors. LIBSVM is a library for SVM [50]. The LIBSVM library is widely used in academic research studies. The feedbacks from the users lead the improvement of the library. The WEKA integration of the LIBSVM is used in this study for SVM algorithm. The cost and gamma values are calculated for each dataset. The calculation of cost and gamma values are achieved by using a helper script developed for LIBSVM. This script code performs trials on both values and measures the accuracy of the SVM algorithm. The optimum parameters are decided according to higher accuracy.

Table 5.2 Optimum cost and gamma values for SVM

Dataset	Cost	Gamma
EMP-ECG	32768	0.03125
RFE-SVM3	8192	0.125
RFE-SVM5	8192	8
RFE-SVM7	512	0.5
RFE-SVM10	0.5	8
RFE-SVM13	0.5	2
RFE-SVM15	0.5	2
RFE-SVM20	512	0.125
CFS3	32768	0.5
CFS5	512	2
CFS7	128	2
CFS10	8	8
CFS13	8	8
CFS15	2	8
CFS20	8	2
PCA3	32768	2
PCA5	8192	0.001953125
PCA7	32768	0.03125

PCA10	8192	0.03125
PCA13	32	2
PCA15	8	2
PCA20	8	2
FA3	2048	0.125
FA5	0.5	8
FA7	32	0.5
FA10	128	0.5
FA13	32768	0.03125

The optimum cost and gamma values are provided as parameters to SVM algorithm. The SVM kernel is chosen as the radial basis function. 10 fold cross-validation is used in experiments. The figures for accuracy, f-score and MCC results are presented here. All results of SVM experiments can be found in APPENDIX D.

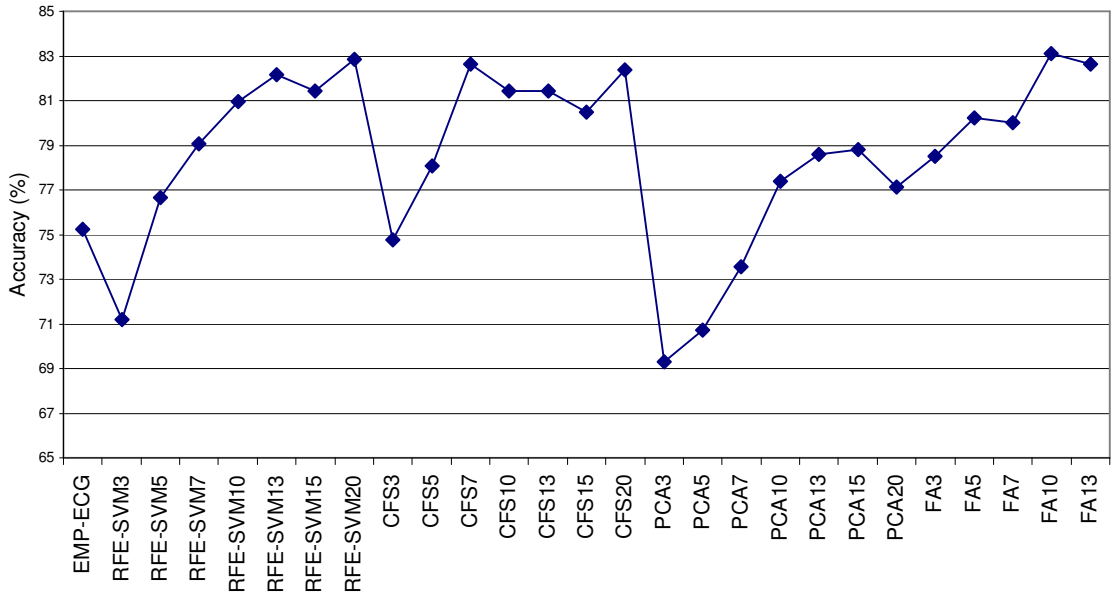


Figure 5.9 Accuracy test results of SVM

The accuracy results show that SVM performance is better with CFS datasets, except CFS3. An average accuracy of 81.07% is achieved with CFS datasets and FA comes the

second with 80.88% accuracy and RFE-SVM follows with an average accuracy of 80.51%. The highest accuracy rate, which is 0.784, is captured for FA10 dataset. The EMP-ECG accuracy is slightly higher than the average of PCA accuracy results.

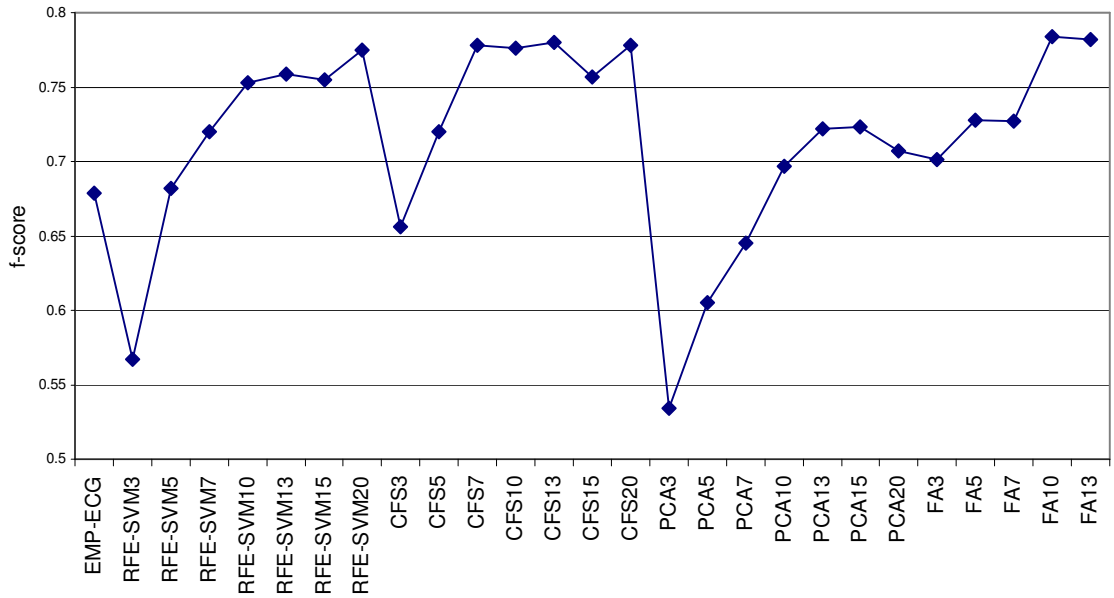


Figure 5.10 f-score test results of SVM

CFS and FA datasets share the highest f-score values with 0.764 and 0.744 average values respectively. RFE-SVM provides slightly lower performance compared to FA with an average of 0.74. EMP-ECG and PCA datasets share the worst f-score values. A similar ranking can be observed from the MCC curve, Figure 5.11.

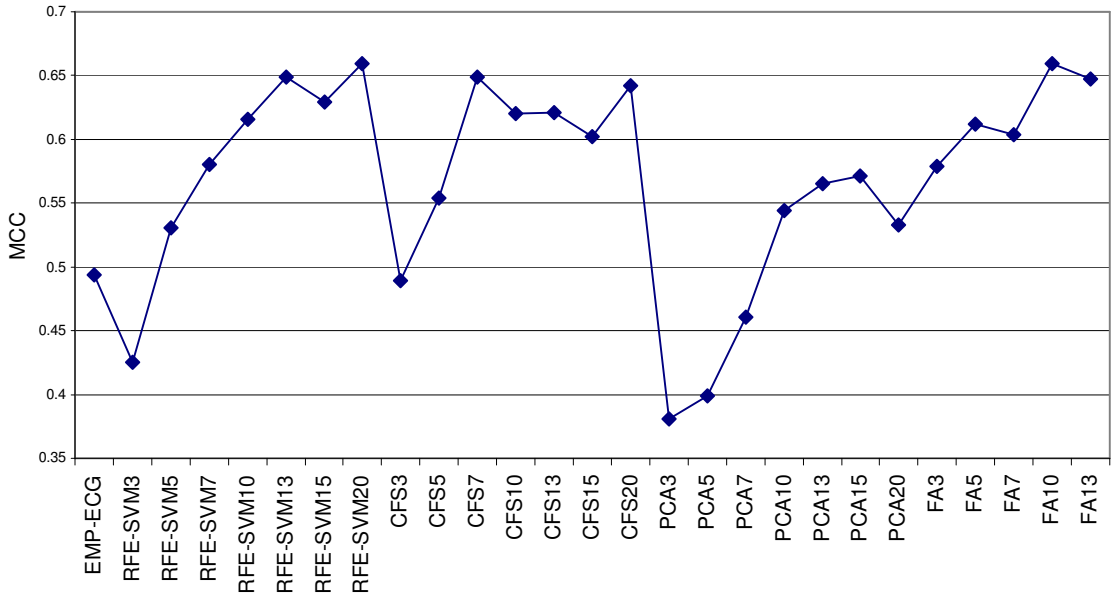


Figure 5.11 Matthews correlation coefficient test results of SVM

5.4.4. Fuzzy Support Vector Machine

FSVM algorithm is an alternative version of SVM algorithm. Introduction of the concept of fuzziness provides better classification performance. The inputs carry class characteristics at different levels. Fuzzy membership values provide measure of representativeness of the inputs to the classes. The version of LIBSVM which allows the use of weights is used for FSVM method. Membership value calculations are described in detail in Chapter 3. FSVM algorithm is applied with 5 different membership value sets. The optimum cost and gamma values for SVM algorithm are also taken into account. 10 fold cross-validation is used in FSVM experiments. The figures for accuracy, f-score and MCC results are presented here. All results of FSVM experiments can be found in APPENDIX E.

CFS, RFE-SVM and FA datasets perform above 80% accuracy with FSVM method. With a closer look, CFS10 dataset gain the maximum accuracy value that is 83.33% with FSVM-DTCM algorithm. When averages are taken into account, FSVM-FCM-2 method outperforms the rest of the FSVM methods with an average accuracy of 78.45%.

FSVM-FCM-2 method provides the best averages for f-score and MCC with 0.71 and 0.566 respectively. When the highest value is taken into account FSVM-DTCM draws the attention for CFS10 dataset. f-score for CFS10 with FSVM-DTCM method is 0.798 and the MCC value is 0.66. The both maximum values are nearly 0.1 higher than the best of averages.

When membership value calculation techniques are compared, membership functions form a similar curve in performance graphs, except OCW. FSVM-OCW and occasionally FSVM-CAR provide worst results for the three performance metrics discussed in this section. FSVM-FCM-2 provides the highest averages for the performance metrics whereas FSVM-DTCM provides the highest values for the three performance graphics.

5.4.5. Classifier Evaluation

Classification and dimension reduction techniques are compared separately in the previous sections. This section provides an inter-classifier comparison. Besides, the discussion on the dimension reduction algorithms is advanced. Each performance metric is taken into consideration. Performance measures are presented in tables. The best value is labeled with bold text.

The KNN15 method is chosen to represent the performance of k-NN algorithm. The reason for this choice is that KNN15 outperforms other k-NN trials in performance metrics. MLP and SVM are taken into account. As the main proposed method is FSVM, this method is represented with all fuzzy membership value functions.

Representative datasets for each dimension reduction method is used in this scope. The representative datasets are selected based on the accuracy metric. The 13-featured sets of PCA, FA and RFE-SVM are chosen besides the 10-featured CFS dataset. The EMP-ECG is also considered.

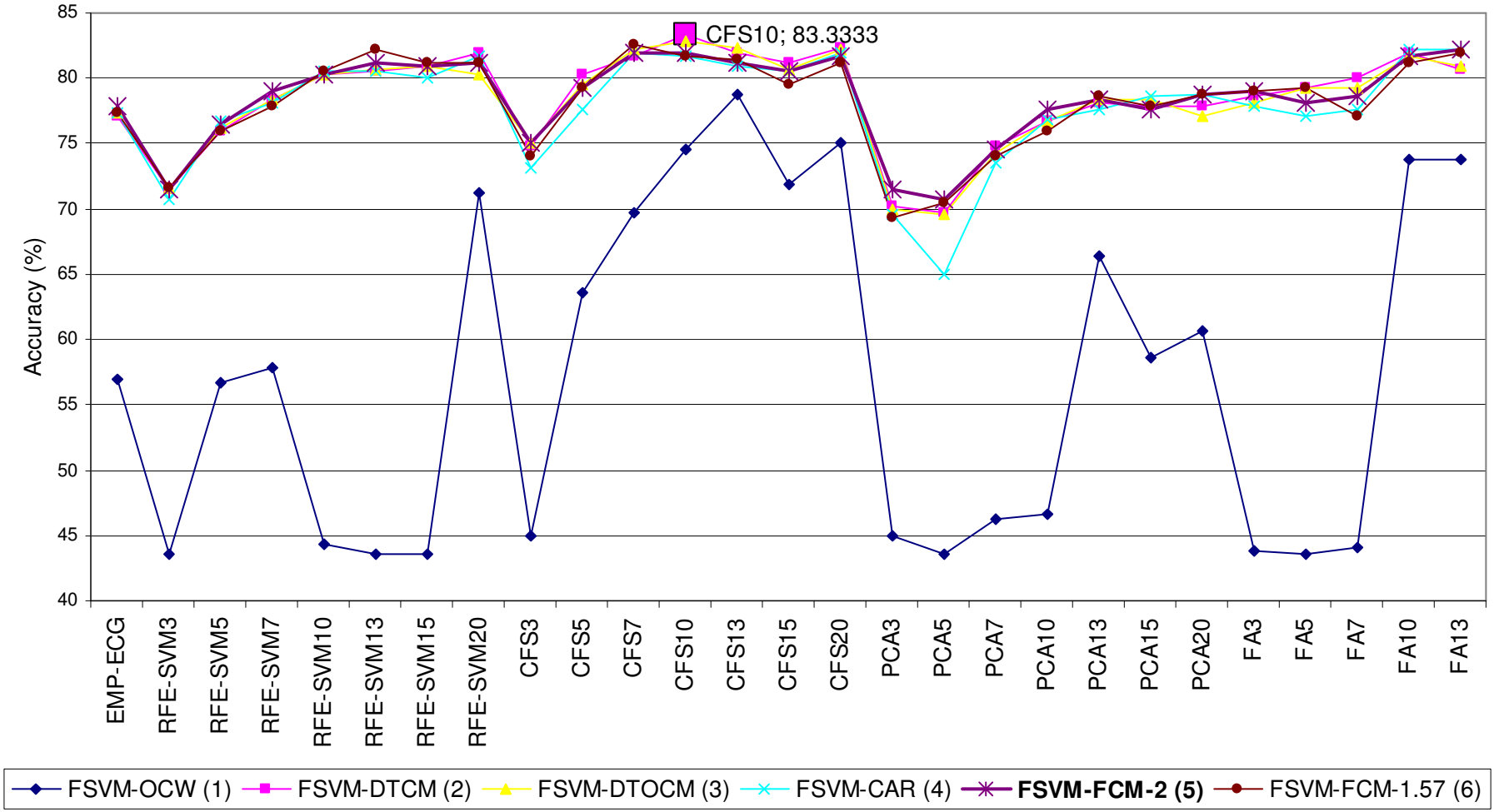


Figure 5.12 Accuracy test results of FSVM

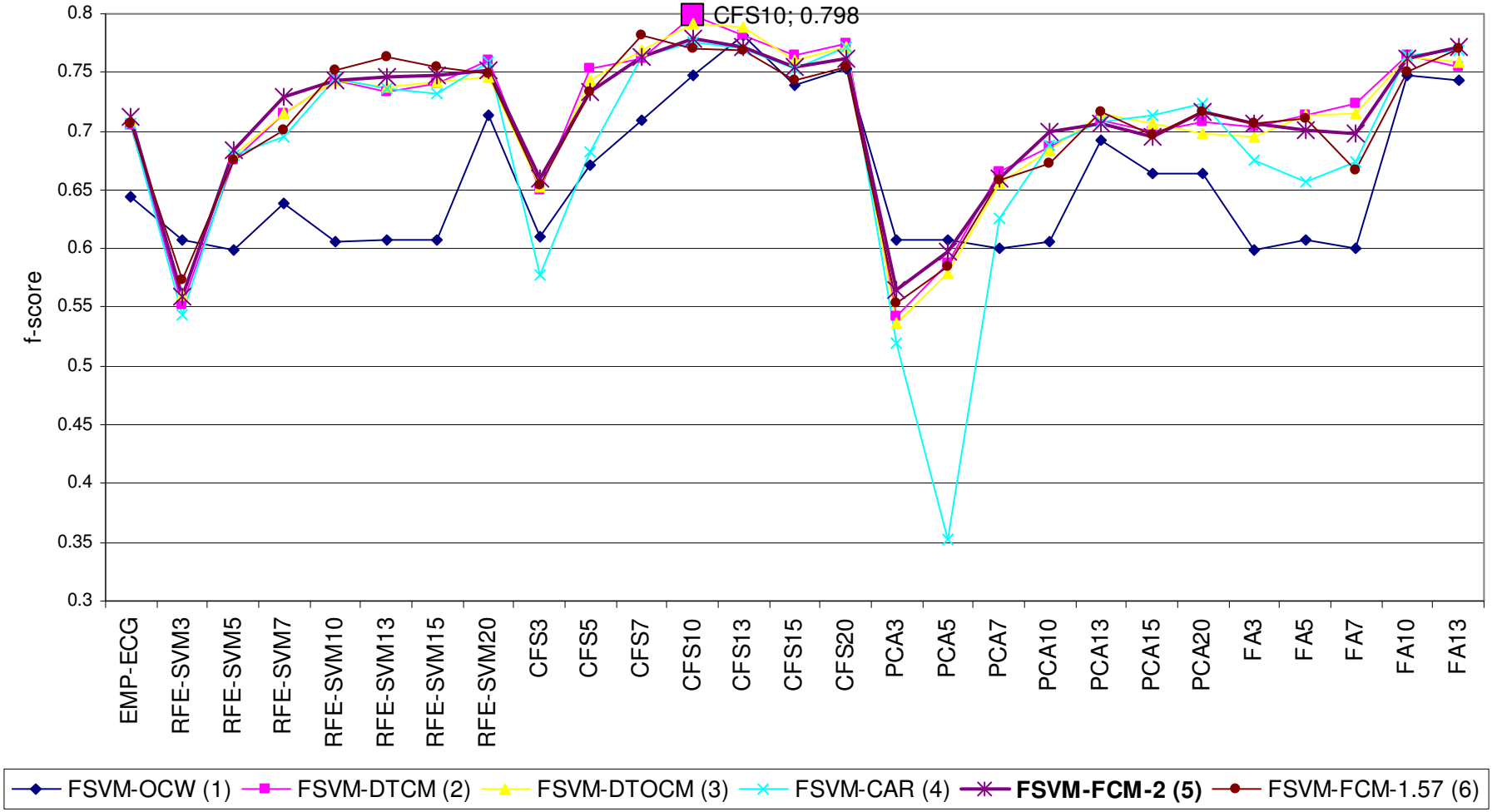


Figure 5.13 f-score test results of FSVM

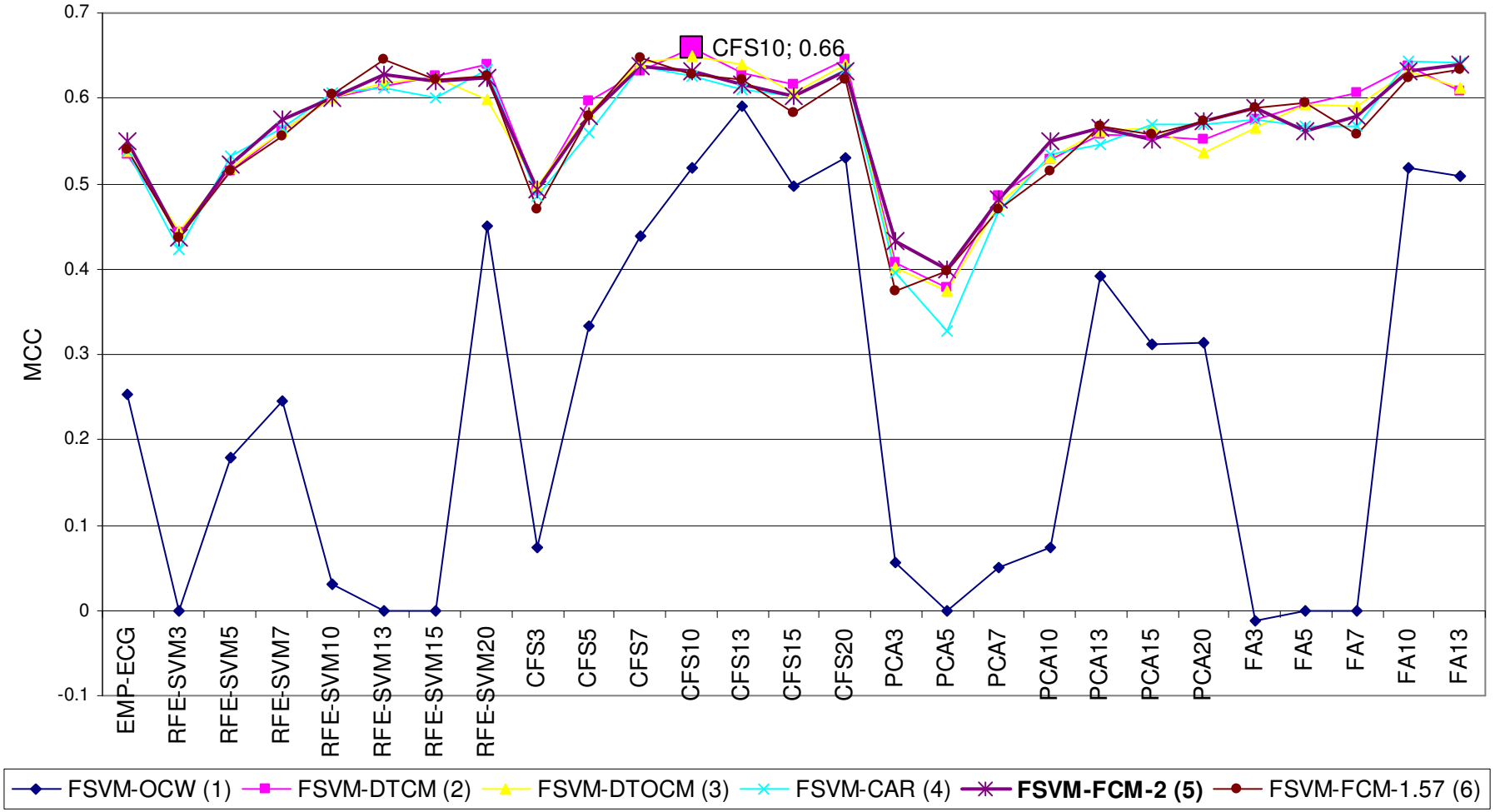


Figure 5.14 Matthews correlation coefficient test results of FSVM

Table 5.3 Classifier accuracy comparison

	PCA13	FA13	RFE-SVM13	CFS10	EMP-ECG
KNN15	67.1429	76.4286	78.3333	74.2857	68.5714
MLP	72.8571	79.5238	79.0476	80.4762	74.0476
SVM	78.5714	82.619	82.1429	81.4286	75.2381
FSVM-OCW	66.4286	73.8095	43.5714	74.5238	56.9048
FSVM-DTCM	78.0952	80.7143	80.4762	**83.3333**	77.1429
FSVM-DTOCM	78.3333	80.9524	80.7143	82.8571	77.381
FSVM-CAR	77.619	82.1429	80.4762	81.6667	77.381
FSVM-FCM-2	78.3333	82.1429	81.1905	81.9048	77.8571
FSVM-FCM-1.57	78.5714	81.9048	82.1429	81.6667	77.381

Accuracy has an increasing trend over all datasets from k-NN, MLP and SVM to FSVM. FSVM accuracy results are changing depending on the membership function. The accuracy results for FSVM-OCW are the worst among other membership calculation functions. FSVM depends on differentiating the inputs by fuzzy membership values. In FSVM-OCW, all members of the positive class are assigned the same value. The distinction between characteristic class members and outliers is not applicable with this method. The best accuracy rate is gained with FSVM-DTCM method.

Table 5.4 Classifier f-score comparison

	PCA13	FA13	RFE-SVM13	CFS10	EMP-ECG
KNN15	0.5	0.677	0.696	0.649	0.493
MLP	0.683	0.757	0.742	0.763	0.68
SVM	0.722	0.782	0.759	0.776	0.679
FSVM-OCW	0.693	0.743	0.607	0.747	0.644
FSVM-DTCM	0.709	0.754	0.733	**0.798**	0.705
FSVM-DTOCM	0.715	0.759	0.738	0.792	0.71
FSVM-CAR	0.708	0.769	0.736	0.776	0.707
FSVM-FCM-2	0.707	0.772	0.746	0.779	0.712
FSVM-FCM-1.57	0.717	0.77	0.763	0.77	0.706

SVM algorithm is more successful in the f-score basis compared to k-NN and MLP. F-score results of SVM are shown to be increased with FSVM methods. Almost all FSVM methods with all datasets provide improvement compared to SVM algorithm. When MCC values are considered, the success of FSVM methods is repeated. The FSVM-DTCM method also provides the highest rates in f-score and MCC as in accuracy. The zero value for RFE-SVM13 with FSVM-OCW method occurs because all the inputs are predicted as positive class members.

Table 5.5 Classifier Matthews correlation coefficient comparison

	PCA13	FA13	RFE-SVM13	CFS10	EMP-ECG
KNN15	0.329	0.526	0.574	0.478	0.378
MLP	0.446	0.581	0.572	0.601	0.467
SVM	0.565	0.647	0.649	0.62	0.494
FSVM-OCW	0.392	0.509	0	0.518	0.253
FSVM-DTCM	0.558	0.609	0.615	**0.66**	0.534
FSVM-DTOCM	0.562	0.613	0.619	0.65	0.539
FSVM-CAR	0.545	0.641	0.613	0.626	0.539
FSVM-FCM-2	0.566	0.639	0.628	0.631	0.55
FSVM-FCM-1.57	0.567	0.634	0.645	0.627	0.54

Precision is the ratio of the correct positive predictions among all the positive predictions. FSVM-CAR provides the highest precision values for PCA, RFE-SVM and CFS datasets. For FA and EMP-ECG, the best precision is gained with FSVM-FCM method.

Table 5.6 Classifier precision comparison

	PCA13	FA13	RFE-SVM13	CFS10	EMP-ECG
KNN15	0.742	0.839	0.897	0.8	0.831
MLP	0.695	0.784	0.803	0.81	0.734
SVM	0.83	0.862	0.922	0.818	0.78
FSVM-OCW	0.576	0.649	0.436	0.658	0.503
FSVM-DTCM	0.842	0.849	0.904	0.847	0.804

FSVM-DTOCM	0.838	0.846	0.905	0.84	0.806
FSVM-CAR	0.895	0.928	**0.945**	0.886	0.81
FSVM-FCM-2	0.859	0.87	0.906	0.832	0.821
FSVM-FCM-1.57	0.844	0.864	0.903	0.849	0.814

Sensitivity or recall is the rate of correct positive class predictions among all positive inputs. On the other hand, specificity presents the rate of correct negative class predictions among all negative inputs. The two performance metrics are complementary. FSVM methods lead to higher values for the two complementary metrics.

Table 5.7 Classifier sensitivity or recall comparison

	PCA13	**FA13**	**RFE-SVM13**	**CFS10**	**EMP-ECG**
KNN15	0.377	0.568	0.568	0.546	0.35
MLP	0.672	0.732	0.689	0.721	0.634
SVM	0.639	0.716	0.645	0.738	0.601
FSVM-OCW	0.869	0.869	**1**	0.863	0.896
FSVM-DTCM	0.612	0.678	0.617	0.754	0.628
FSVM-DTOCM	0.623	0.689	0.623	0.749	0.634
FSVM-CAR	0.623	0.683	0.623	0.727	0.628
FSVM-FCM-2	0.601	0.694	0.634	0.732	0.628
FSVM-FCM-1.57	0.623	0.694	0.661	0.705	0.623

Table 5.8 Classifier specificity comparison

	PCA13	**FA13**	**RFE-SVM13**	**CFS10**	**EMP-ECG**
KNN15	0.899	0.916	0.949	0.895	0.945
MLP	0.772	0.844	0.869	0.869	0.823
SVM	0.899	0.911	0.958	0.873	0.869
FSVM-OCW	0.506	0.869	**1**	0.654	0.316
FSVM-DTCM	0.911	0.907	0.949	0.895	0.882
FSVM-DTOCM	0.907	0.903	0.949	0.89	0.882
FSVM-CAR	0.895	0.928	0.945	0.886	0.886
FSVM-FCM-2	0.924	0.92	0.949	0.886	0.895
FSVM-FCM-1.57	0.911	0.916	0.945	0.903	0.89

When combined metrics like f-score and MCC are considered besides accuracy, CFS10 dataset reaches the top values. When detail metrics like precision, sensitivity and specificity are considered, RFE-SVM accomplishes better results. The main performance metric in this research is accuracy and MCC as the classes have unbalanced number of inputs. F-score presents a balance of sensitivity and specificity. Therefore, CFS10 dataset is chosen to present the best performance and with FSVM-DTCM.

ROC curve analysis is performed for the leading dataset only. The best performance is obtained for CFS10 and using the FSVM-DTCM method. The ROC curve is presented in Figure 5.15.

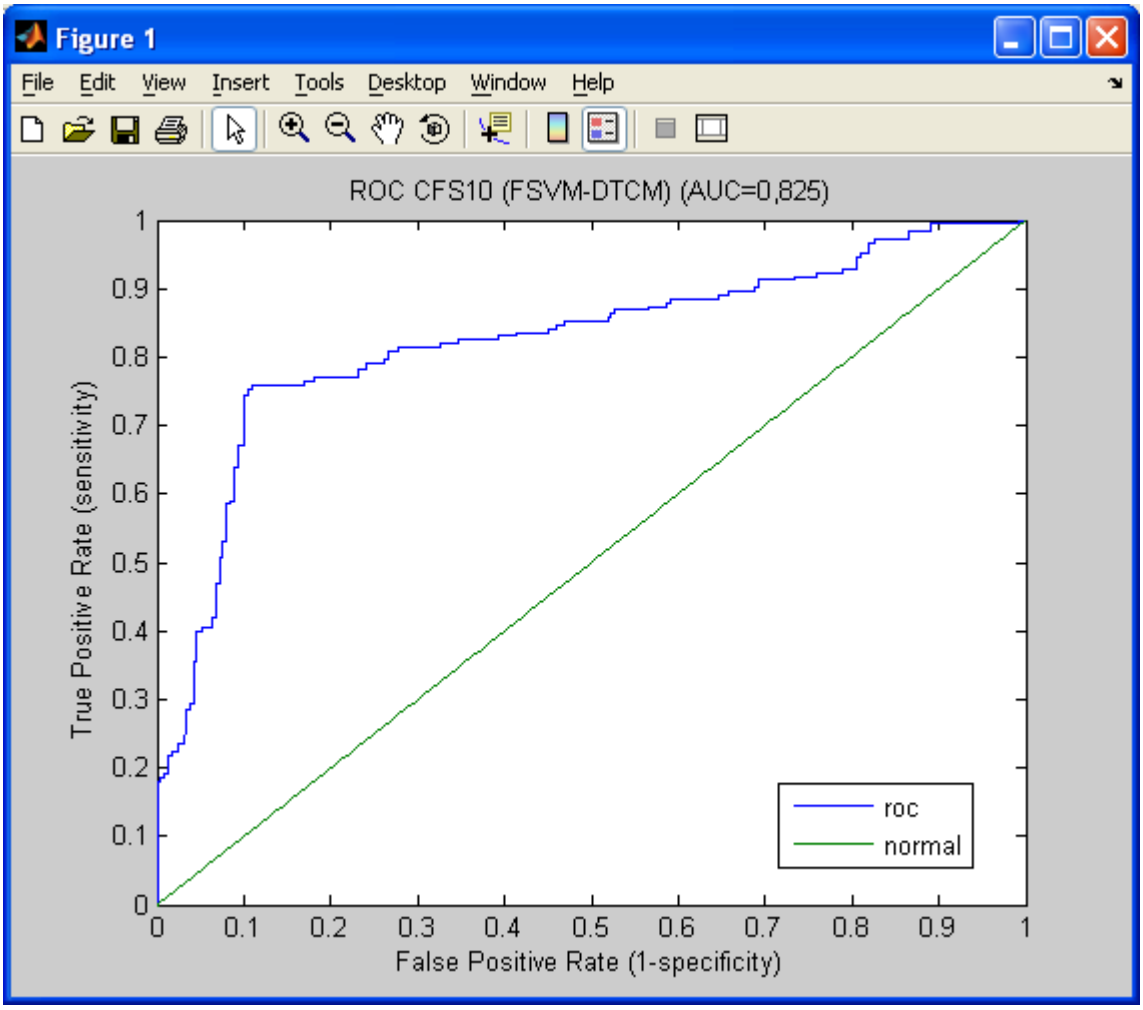


Figure 5.15 ROC curve of CFS10 for FSVM-DTCM

CFS10 dataset is found to provide the best overall performance. The features in CFS10 and their histograms are presented in Figure 5.16. Heartrate and QRSduration features draw attention. Besides, amplitude of the T wave in two leads, AVL and V6, are taken into consideration. Average width of Q wave in AVF lead is chosen. There are features representing the common types of disturbance. The amplitude of R' wave which is a corruption of the R wave is chosen. The amplitude of JJ wave which occurs besides the S wave is used. The number of deflections is taken into account. Area calculations are also used. QRSA is the average area under the QRS complex. QRSTA also depends on the average area under the QRS complex.

CFS10 contains significant features of the ECG signal that can be realized from the ECG paper easily. In addition, common irregular waves are also considered. Two area based features are present in CFS10. The combination of all these features leads the best performance with the proposed method. The EMP-ECG contains the primary wave features. With the better performance gained by CFS10, it can be concluded that complex features can improve the accuracy of the diagnosis. Hence, automated ECG analysis tools can be helpful to cardiologists.

5.5. Experimental Results for Rule Base Development

Rule base development provides a handshake for the classification method and the expert's opinion. Another advantage of the rule base is to present a more readable decision set for the reader. The input of the rule base extraction module is the classification model which is constructed from the correctly classified inputs of CFS10. This model dataset is referred as CFS10-MODEL in the rest of the document.

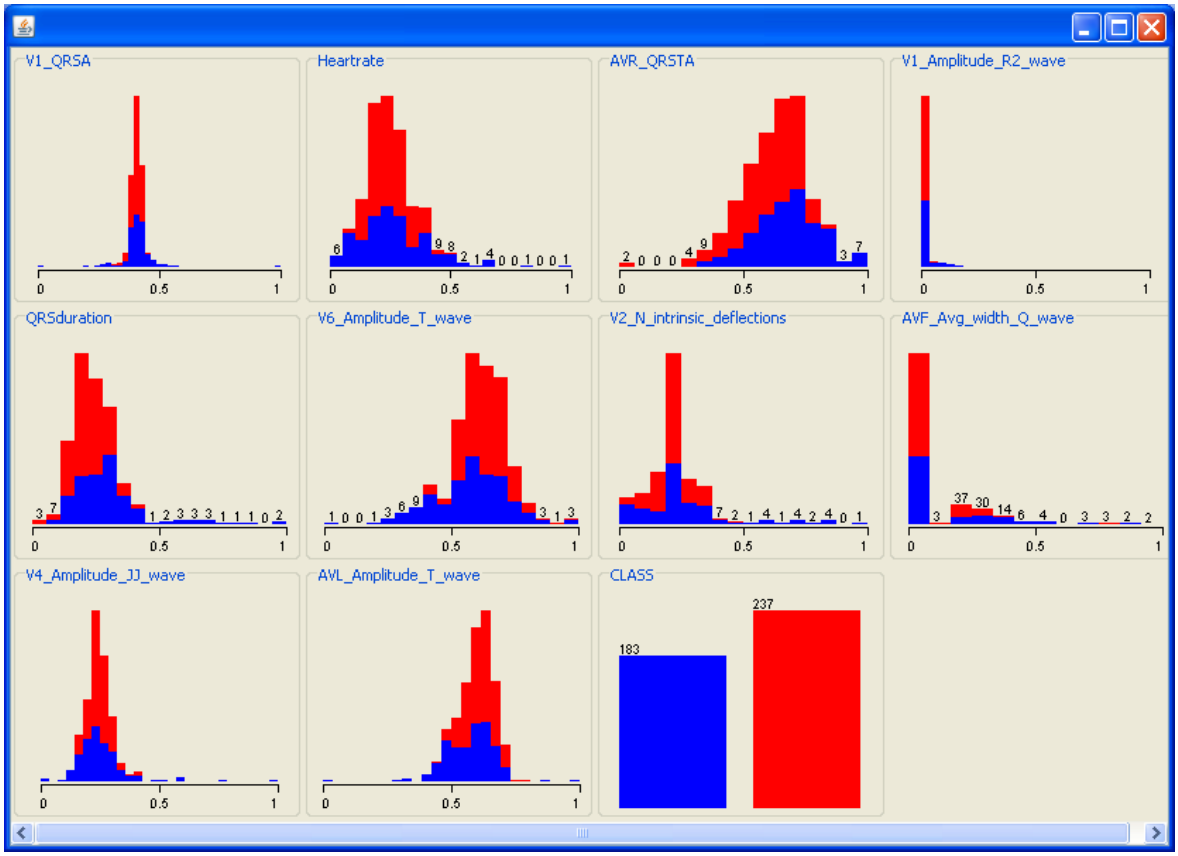


Figure 5.16 CFS10 dataset

Four techniques are considered for rule base development. Decision Tree is used as a base rule set. A Partial Decision Tree algorithm is applied for comparison if it provides any improvement against the decision tree algorithm. Repeated Incremental Pruning to Produce Error Reduction is also applied to the classification model. As a final method, Adaptive Neuro-Fuzzy Inference System is constructed in order to provide a rule base. The experiment details for both methods and the accuracy results for the produced rule basis are provided in the following sections.

5.5.1. C4.5 Decision Tree

Decision tree results are obtained by the *J48* method in WEKA platform. It is an implementation of C4.5 decision tree algorithm. CFS10-MODEL is taken as input and the C4.5 algorithm is applied on 10-fold basis. Reduced error pruning is chosen. The resulting

tree at the end of 10 folds has 9 leaf nodes: 7 for positive class and 2 for negative class, Table 5.9.

Table 5.9 C4.5 initial rule base

1	If V1_Amplitude_R2_wave <= 1 and V6_Amplitude_T_wave <= -0.4, then CLASS 1
2	If V1_Amplitude_R2_wave <= 1 and V6_Amplitude_T_wave > -0.4 and AVF_Avg_width_Q_wave <= 28 and Heartrate <= 58, then CLASS 1
3	If V1_Amplitude_R2_wave <= 1 and V6_Amplitude_T_wave < -0.4 and AVF_Avg_width_Q_wave <= 28 and Heartrate > 58 and Heartrate <= 98 and AVL_Amplitude_T_wave <= -0.8, then CLASS 1
4	If V1_Amplitude_R2_wave <= 1 and V6_Amplitude_T_wave < -0.4 and AVF_Avg_width_Q_wave <= 28 and Heartrate > 58 and Heartrate <= 98 and AVL_Amplitude_T_wave > -0.8 and V2_N_intrinsic_deflections <= 32, then CLASS -1
5	If V1_Amplitude_R2_wave <= 1 and V6_Amplitude_T_wave < -0.4 and AVF_Avg_width_Q_wave <= 28 and Heartrate > 58 and Heartrate <= 98 and AVL_Amplitude_T_wave > -0.8 and V2_N_intrinsic_deflections > 32 and V2_N_intrinsic_deflections <= 52, then CLASS -1
6	If V1_Amplitude_R2_wave <= 1 and V6_Amplitude_T_wave < -0.4 and AVF_Avg_width_Q_wave <= 28 and Heartrate > 58 and Heartrate <= 98 and AVL_Amplitude_T_wave > -0.8 and V2_N_intrinsic_deflections > 32 and V2_N_intrinsic_deflections > 52, then CLASS 1
7	If V1_Amplitude_R2_wave <= 1 and V6_Amplitude_T_wave < -0.4 and AVF_Avg_width_Q_wave <= 28 and Heartrate > 58 and Heartrate > 98, then CLASS 1
8	If V1_Amplitude_R2_wave <= 1 and V6_Amplitude_T_wave < -0.4 and AVF_Avg_width_Q_wave > 28, then CLASS 1
9	If V1_Amplitude_R2_wave > 1, then CLASS 1

Our target class is the positive class; therefore the rules representing the target class is selected to form the rule base. There are minor overlaps in the selected rule set. The

resulting rule base for C4.5 algorithm consists of the 7 non-overlapping rules defining the positive class, Table 5.10.

Table 5.10 C4.5 resulting rule base

RB551.1.	If V1_Amplitude_R2_wave <= 1 and V6_Amplitude_T_wave <= -0.4, then CLASS 1
RB551.2.	If V1_Amplitude_R2_wave <= 1 and V6_Amplitude_T_wave > -0.4 and AVF_Avg_width_Q_wave <= 28 and Heartrate <= 58, then CLASS 1
RB551.3.	If V1_Amplitude_R2_wave <= 1 and V6_Amplitude_T_wave > -0.4 and AVF_Avg_width_Q_wave <= 28 and 58 < Heartrate <= 98 and AVL_Amplitude_T_wave <= -0.8, then CLASS 1
RB551.4.	If V1_Amplitude_R2_wave <= 1 and V6_Amplitude_T_wave > -0.4 and AVF_Avg_width_Q_wave <= 28 and 58 < Heartrate <= 98 and AVL_Amplitude_T_wave > -0.8 and V2_N_intrinsic_deflections > 32, then CLASS 1
RB551.5.	If V1_Amplitude_R2_wave <= 1 and V6_Amplitude_T_wave > -0.4 and AVF_Avg_width_Q_wave <= 28 and Heartrate > 98, then CLASS 1
RB551.6.	If V1_Amplitude_R2_wave <= 1 and V6_Amplitude_T_wave > -0.4 and AVF_Avg_width_Q_wave > 28, then CLASS 1
RB551.7.	If V1_Amplitude_R2_wave > 1, then CLASS 1

In the resulting rule set, the amplitude of the R' and T waves, the duration of Q wave, heart rate and number of intrinsic deflections are chosen as the determinant features. The amplitude of R' wave is the root of the tree and is classified according to being greater or less than 1. The amplitude of T wave, the width of Q wave and the number of intrinsic deflections are also separated into two portions. Heart rate is considered in three slots intersecting at 58 and 98 heart beats in a minute.

5.5.2. Repeated Incremental Pruning to Produce Error Reduction

JRip method in WEKA platform is exercised for Repeated Incremental Pruning to Produce Error Reduction rule base development. Pruning is performed for rule base tree

generation and stopping condition includes the control for the error rate to be greater than or equal to 0.5. The produced rule base consists of 10 rules: 9 for positive class and 1 for negative class, Table 5.11.

Table 5.11 RIPPER initial rule base

1	If QRSduration >= 93 and AVR_QRSTA >= -19.8 and QRSduration >= 100, then CLASS 1
2	If V6_Amplitude_T_wave <= -0.1, then CLASS 1
3	If V1_Amplitude_R2_wave >= 1, then CLASS 1
4	If Heartrate <= 58, then CLASS 1
5	If AVF_Avg_width_Q_wave >= 28, then CLASS 1
6	If Heartrate >= 98, then CLASS 1
7	If AVL_Amplitude_T_wave <= -1, then CLASS 1
8	If V2_N_intrinsic_deflections >= 56, then CLASS 1
9	If QRSduration >= 111, then CLASS 1
10	Else CLASS -1

The positive class is taken as the base class for the rule extraction and the leaving instances are considered as members of the negative class. This situation is represented as the tenth rule. After overlap reduction among rules, the resulting rule base consists of 7 rules for the positive class, Table 5.12.

Table 5.12 RIPPER resulting rule base

RB552.1.	If QRSduration >= 111 or (QRSduration >= 93 and AVR_QRSTA >= -19.8), then CLASS 1
RB552.2.	If Heartrate <= 58 or Heartrate >= 98, then CLASS 1
RB552.3.	If V6_Amplitude_T_wave <= -0.1, then CLASS 1
RB552.4.	If V1_Amplitude_R2_wave >= 1, then CLASS 1
RB552.5.	If AVF_Avg_width_Q_wave >= 28, then CLASS 1
RB552.6.	If AVL_Amplitude_T_wave <= -1, then CLASS 1

RB552.7.	If V2_N_intrinsic_deflections >= 56, then CLASS 1

The selected distinguishing features are the duration of the QRS complex and Q wave, heart rate, the amplitude of R' and T waves, QRSTA of the AVR Lead and the number of intrinsic deflections. The separation points of heart rate, which are 58 and 98, are consistent with the C4.5 rule base.

5.5.3. Partial Decision Tree

PART method in WEKA platform is applied to CFS10-MODEL for Partial Decision Tree rule base development. An unpruned tree with 10 leaf nodes is constructed for the dataset: 7 for positive class and 3 for negative class, Table 5.13.

Table 5.13 PART initial rule base

1	If AVF_Avg_width_Q_wave > 24, then CLASS 1
2	If V1_Amplitude_R2_wave > 1, then CLASS 1
3	If AVL_Amplitude_T_wave <= -1, then CLASS 1
4	If V6_Amplitude_T_wave <= -0.1 and AVR_QRSTA > -19.8, then CLASS 1
5	If Heartrate > 58 and Heartrate <= 96 and QRSduration <= 99 and V2_N_intrinsic_deflections <= 32, then CLASS -1
6	If AVR_QRSTA > -20.5 and AVF_Avg_width_Q_wave <= 0, then CLASS 1
7	If Heartrate > 54 and QRSduration <= 107 and V2_N_intrinsic_deflections <= 52 and V4_Amplitude_JJ_wave > -0.5, then CLASS -1
8	If AVF_Avg_width_Q_wave > 0, then CLASS 1
9	If V4_Amplitude_JJ_wave > -0.5, then CLASS 1
10	Else CLASS -1

The intersections between the rules are revised and the overlapping rules are cleared. The eighth rule covers the first rule and is selected to be in the resulting rule base instead of the two. The resulting rule base consists of 6 rules for the positive class, Table 5.14.

Table 5.14 PART resulting rule base

RB553.1.	If AVF_Avg_width_Q_wave > 0, then CLASS 1
RB553.2.	If V1_Amplitude_R2_wave > 1, then CLASS 1
RB553.3.	If AVL_Amplitude_T_wave <= -1, then CLASS 1
RB553.4.	If V6_Amplitude_T_wave <= -0.1 and AVR_QRSTA > -19.8, then CLASS 1
RB553.5.	If AVR_QRSTA > -20.5 and AVF_Avg_width_Q_wave <= 0, then CLASS 1
RB553.6.	If V4_Amplitude_JJ_wave > -0.5, then CLASS 1

The duration of the Q wave, amplitude of R', T and JJ waves and QRSTA of the AVR Lead are marked as the major features by PART algorithm. The tree structure can be observed from the C4.5 rules. PART algorithm constructs a union of partial decision trees, therefore rules are not in a successive order.

5.5.4. Adaptive Neuro-Fuzzy Inference System

ANFIS which is an adaptive neuro-fuzzy inference system in Matlab is used to construct the neuro-fuzzy rule base from the classification model. Grid partitioning is applied in the production of the membership functions. Trapezoidal membership functions are preferred and the number of membership functions for each feature is entered as input. 3 membership functions for Heartrate feature is selected and for the rest of the features 2. In the training phase, hybrid method is used with 0.05 error tolerance and 30 epochs. The resulting structure of the Sugeno type fuzzy inference system can be seen from Figure 5.17. On the left side the features are listed vertically and they are taken as input to the Sugeno FIS. On the right side, the output box takes place representing if the input ECG has anomaly or not. The Sugeno box in the middle represents the FIS structure which takes features as input and produces the output according to the membership functions.

The membership functions for each feature in the CFS10 dataset is provided in Figure 5.18. The first feature is the duration of the QRS complex. The trapezoid for in1mf1 ends at 140. The medical sources show that a normal QRS duration lies between 4 and 120miliseconds. The membership function in1mf1 covers a wider area but is representative for normal QRS duration. The second feature is the heart rate. The membership function

in2mf2 represents the normal heart rate. The trapezoid for in2mf2 starts at 58 and ends at 96. In the medical literature, the normal heart rate lies in the stated range. The lower and upper heart rates are indicator of anomaly. The sixth feature is the amplitude of T wave in AVL lead and the tenth feature in V6 lead. The normal amplitude of T wave is below 1milivolt. In the two membership function graphs, the intersection point of the two membership functions is below 1 which is consistent with the medical knowledge.

5.5.5. Rule Base Evaluation

The performance measure of the rule set is selected as accuracy. The corresponding accuracy measures for the rule bases are listed in Table 5.15.

Table 5.15 Rule base accuracy performance

	Accuracy
C4.5	90.00%
RIPPER	**93.14%**
PART	92.57%
ANFIS	79.43%

The C4.5 method provides the lowest accuracy rate with 90% for the classification model. PART and RIPPER algorithms reach higher accuracy rates as 92.57% and 93.14% respectively. With the post-processing step, RIPPER is also more accurate on the classification model than PART. The advantage of using ANFIS is that it generates the fuzzy membership functions for modeling the data.

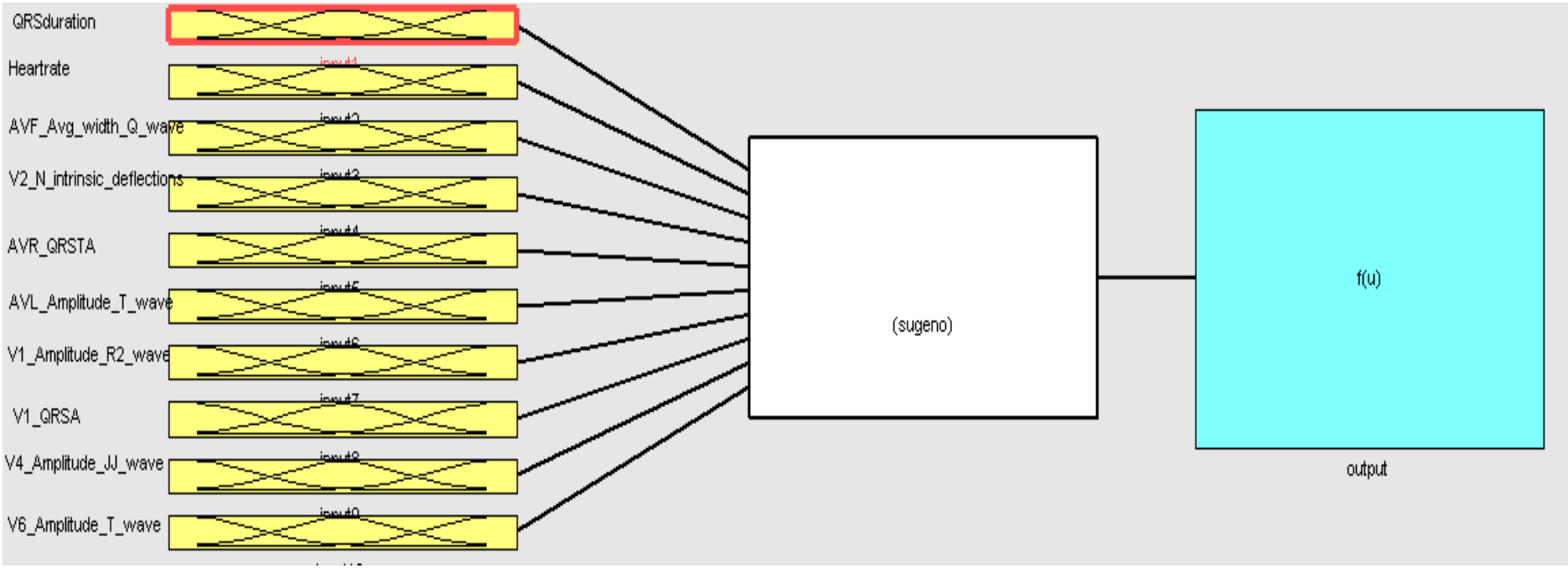
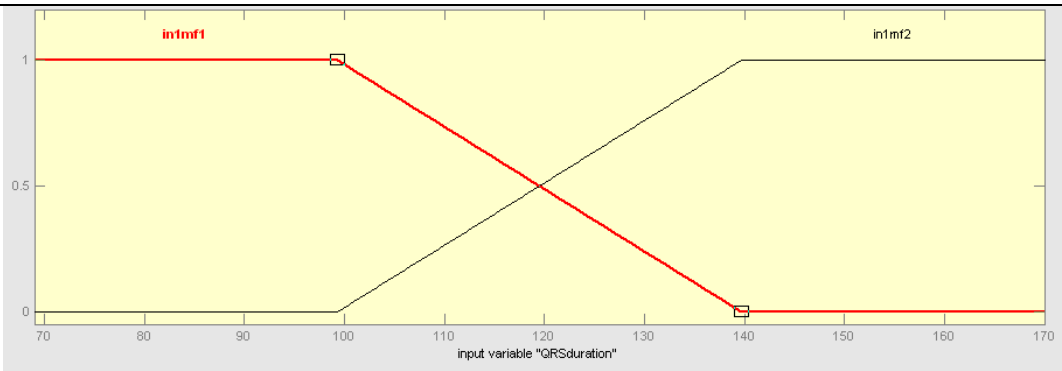
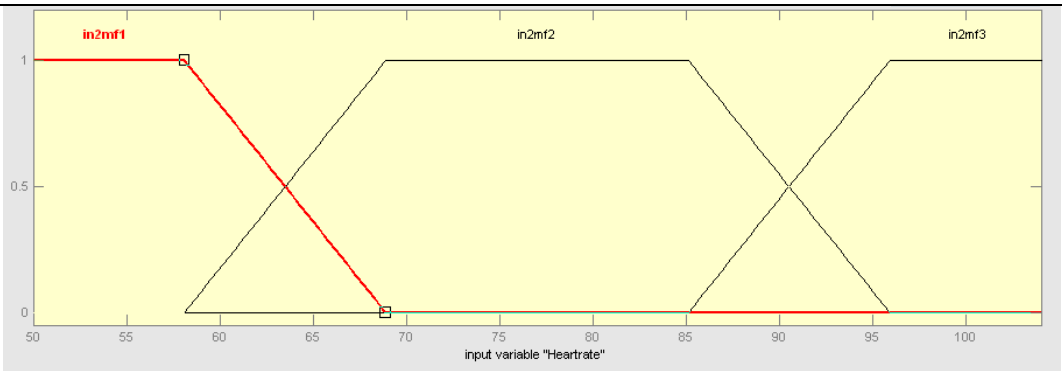
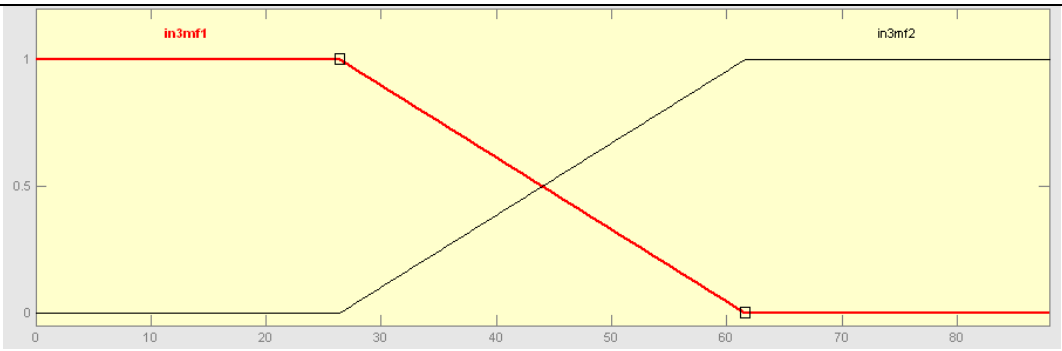
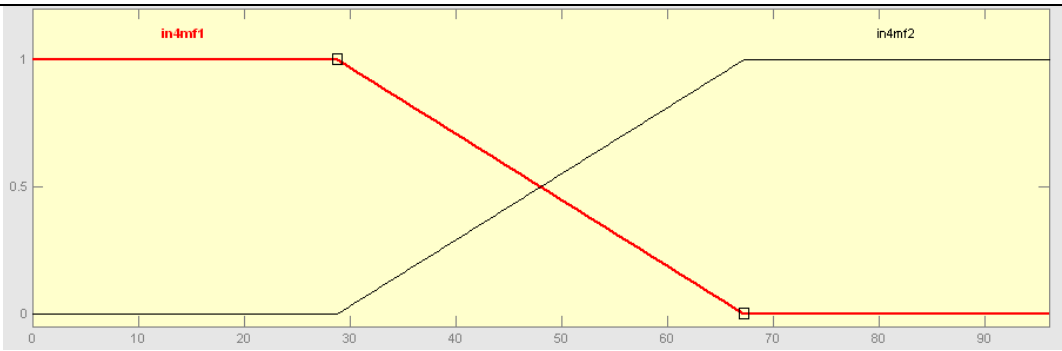
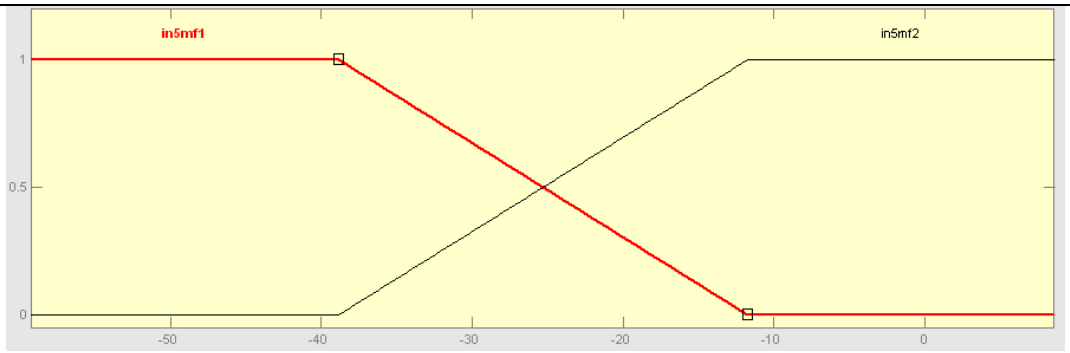
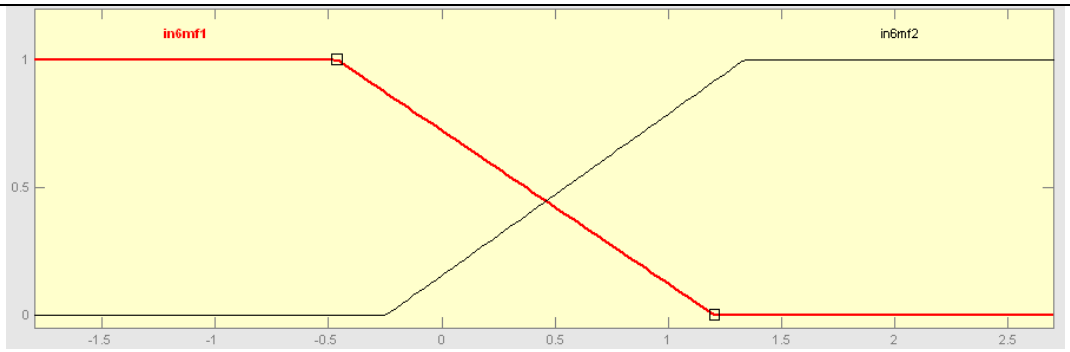
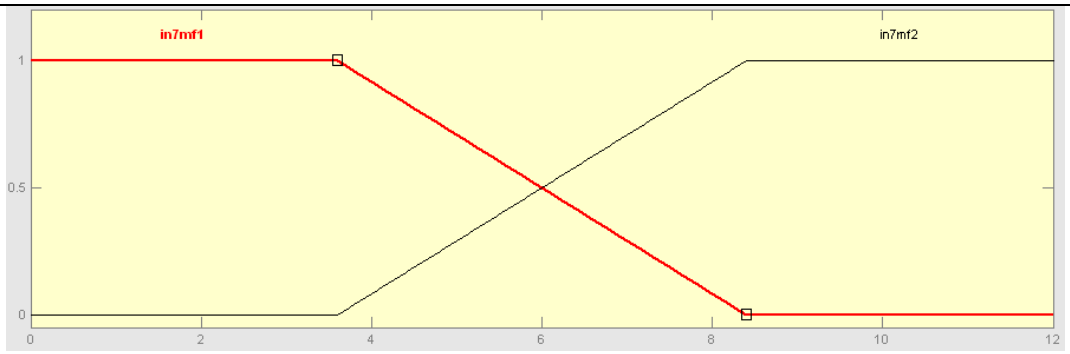
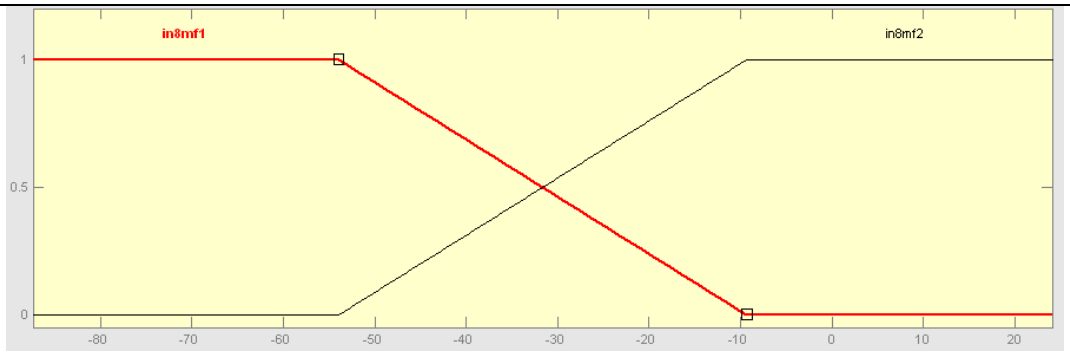
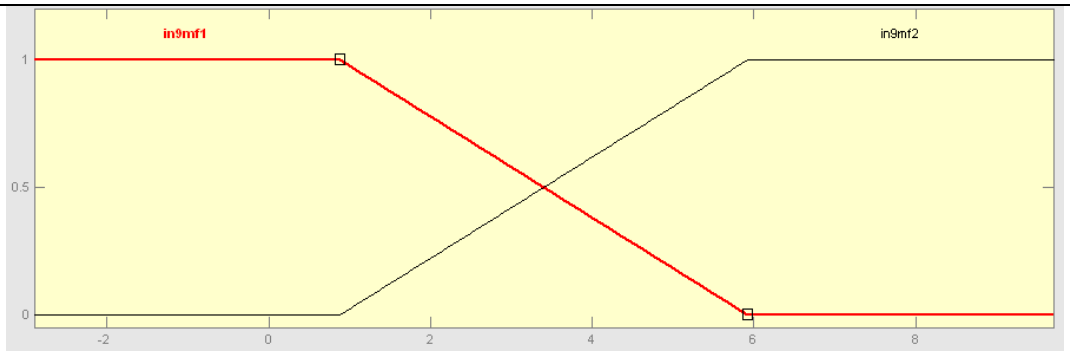


Figure 5.17 FIS Structure of ECG classification model of ANFIS

1	in1mf1 in1mf2 1 0.5 0 70 80 90 100 110 120 130 140 150 160 170 input variable "QRSduration"
2	in2mf1 in2mf2 in2mf3 1 0.5 0 50 55 60 65 70 75 80 85 90 95 100 input variable "Heartrate"
3	in3mf1 in3mf2 1 0.5 0 0 10 20 30 40 50 60 70 80
4	in4mf1 in4mf2 1 0.5 0 0 10 20 30 40 50 60 70 80 90

5	in5mf1 in5mf2 1 0.5 0 -50 -40 -30 -20 -10 0
6	in6mf1 in6mf2 1 0.5 0 -1.5 -1 -0.5 0 0.5 1 1.5 2 2.5
7	in7mf1 in7mf2 1 0.5 0 0 2 4 6 8 10 12
8	in8mf1 in8mf2 1 0.5 0 -80 -70 -60 -50 -40 -30 -20 -10 0 10 20
9	in9mf1 in9mf2 1 0.5 0 -2 0 2 4 6 8

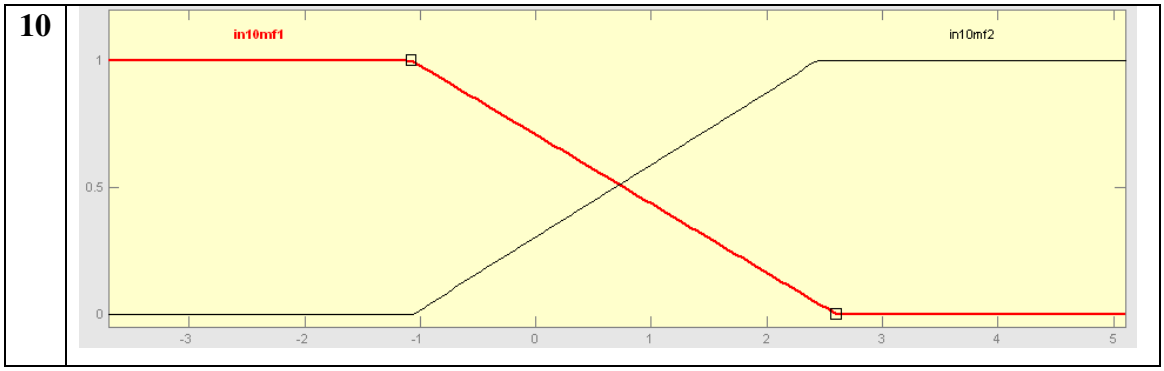


Figure 5.18 Membership functions from ECG FIS

6. EXPERIMENTS WITH REAL ECG DATA

The proposed method is compared with other methods. The advance in classification is presented according to different performance metrics. The presented theoretical success is validated with a new set of ECG. The ECGs are collected the Kardiosis ECG tool by TEPA. The new ECG collection is called OWN-ECG and contains 28 instances. The feature set in CFS10 is chosen to be applied to OWN-ECG because of the overall success of CFS10 in experiments. Kardiosis ECG provides all the features of UCI-ECG as the dataset is also produced using the same tool.

The proposed method, FSVM, is applied to OWN-ECG with all fuzzy membership functions. The results are listed in Table 6.1.

Table 6.1 FSVM results of OWN-ECG

	Accuracy
FSVM-OCW	75.00%
FSVM-DTCM	82.14%
FSVM-DTOCM	82.14%
FSVM-CAR	**85.71%**
FSVM-FCM-2	82.14%
FSVM-FCM-1.57	82.14%

Among the different fuzzy support vector machines, FSVM-CAR produces the highest accuracy with 85.71% and FSVM-OCW produces the lowest accuracy with 75%. The rest of the FSVM methods lead to the same accuracy value which is 82.14%. The average of the accuracy results of FSVM methods is 81.55%. The results of the OWN-ECG dataset imply that the proposed method can be applied to real life ECG data with high accuracy rate. Although the performance of each FSVM method may vary according to the distribution of the dataset, the average results provide premising diagnosis rates.

7. CONCLUSION

ECG is the most frequently used and quick first order medical tool for cardiac treatment. The initial diagnosis is produced and the further tests are planned according to ECG results. The cardiac analysis from the ECG paper is a challenging task which requires experience and focused attention. Hence, the success of the judgment of a cardiologist depends on his experience. This dependency leads to automated ECG analysis tools which provide contributory information for the cardiologist by extracting and interpreting valuable features from the ECG.

This research presents results of an extensive classification study on ECG arrhythmia and also rule base development. First module of the proposed system is the dimension reduction. In the second module fuzzy membership values are calculated for each input. In the third module, the data and the membership values are processed to classify the inputs as containing anomaly or normal ECG. FSVM method is shown to outperform the other classification algorithms. The success of the FSVM method depends on the elasticity of weighing of the inputs to the classification model. The results show that an input may not provide as valuable information as the other for learning the pattern of the dataset. The inputs which mislead the model are called outliers. FSVM overcomes the effects of outliers by assigning low membership values to them.

After classification module, the classification model is used for rule base extraction. Both tree based and fuzzy rule base extraction methods are applied to the ECG classification model. When accuracy metric is taken into account, RIPPER method outperforms C4.5, PART and ANFIS. ANFIS generates fuzzy membership functions besides rules and therefore is separated from the other methods.

In order to verify the classification results, a real life ECG dataset is constructed. The FSVM algorithms are run on the new dataset. The new ECG dataset leads to higher accuracy compared to the UCI dataset.

APPENDIX

APPENDIX A. The official dataset description of the UCI Arrhythmia dataset

1. Title: Cardiac Arrhythmia Database

2. Sources:

(a) Original owners od Database:

-- 1. H. Altay Guvenir, PhD.,
Bilkent University,
Department of Computer Engineering and Information Science,
06533 Ankara, Turkey
Phone: +90 (312) 266 4133
Email: guvenir@cs.bilkent.edu.tr

-- 2. Burak Acar, M.S.,
Bilkent University,
EE Eng. Dept.
06533 Ankara, Turkey
Email: buraka@ee.bilkent.edu.tr

-- 2. Haldun Muderrisoglu, M.D., Ph.D.,
Baskent University,
School of Medicine
Ankara, Turkey

(b) Donor: H. Altay Guvenir

Bilkent University,
Department of Computer Engineering and Information Science,
06533 Ankara, Turkey
Phone: +90 (312) 266 4133
Email: guvenir@cs.bilkent.edu.tr

(c) Date: January, 1998

3. Past Usage:

1. H. Altay Guvenir, Burak Acar, Gulsen Demiroz, Ayhan Cekin
"A Supervised Machine Learning Algorithm for Arrhythmia Analysis"
Proceedings of the Computers in Cardiology Conference,
Lund, Sweden, 1997.

The aim is to determine the type of arrhythmia from
the ECG recordings.

4. Relevant Information:

This database contains 279 attributes, 206 of which are linear valued and the rest are nominal. Concerning the study of H. Altay Guvenir: "The aim is to distinguish between the presence and absence of cardiac arrhythmia and to classify it in one of the 16 groups. Class 01 refers to 'normal' ECG classes 02 to 15 refers to different classes of arrhythmia and class 16 refers to the rest of unclassified ones. For the time being, there exists a computer program that makes such a classification. However there are differences between the cardiolog's and the programs classification. Taking the cardiolog's as a gold standard we aim to minimise this difference by means of machine learning tools."

The names and id numbers of the patients were recently removed from the database.

5. Number of Instances: 452

6. Number of Attributes: 279

7. Attribute Information:

-- Complete attribute documentation:

1 Age: Age in years , linear
2 Sex: Sex (0 = male; 1 = female) , nominal
3 Height: Height in centimeters , linear

4 Weight: Weight in kilograms , linear
5 QRS duration: Average of QRS duration in msec., linear
6 P-R interval: Average duration between onset of P and Q waves in msec., linear
7 Q-T interval: Average duration between onset of Q and offset of T waves in msec., linear
8 T interval: Average duration of T wave in msec., linear
9 P interval: Average duration of P wave in msec., linear
Vector angles in degrees on front plane of:, linear
10 QRS
11 T
12 P
13 QRST
14 J

15 Heart rate: Number of heart beats per minute ,linear

Of channel DI:
Average width, in msec., of: linear
16 Q wave
17 R wave
18 S wave
19 R' wave, small peak just after R
20 S' wave

21 Number of intrinsic deflections, linear

22 Existence of ragged R wave, nominal
23 Existence of diphasic derivation of R wave, nominal
24 Existence of ragged P wave, nominal
25 Existence of diphasic derivation of P wave, nominal
26 Existence of ragged T wave, nominal
27 Existence of diphasic derivation of T wave, nominal

Of channel DII:

28 .. 39 (similar to 16 .. 27 of channel DI)

Of channels DIII:

40 .. 51

Of channel AVR:

52 .. 63

Of channel AVL:

64 .. 75

Of channel AVF:

76 .. 87

Of channel V1:

88 .. 99

Of channel V2:

100 .. 111

Of channel V3:

112 .. 123

Of channel V4:

124 .. 135

Of channel V5:

136 .. 147

Of channel V6:

148 .. 159

Of channel DI:

Amplitude , * 0.1 milivolt, of

160 JJ wave, linear

161 Q wave, linear

162 R wave, linear

163 S wave, linear

164 R' wave, linear

165 S' wave, linear

166 P wave, linear

167 T wave, linear

168 QRSA , Sum of areas of all segments divided by 10,
(Area= width * height / 2), linear
169 QRSTA = QRSA + 0.5 * width of T wave * 0.1 * height of T
wave. (If T is diphasic then the bigger segment is
considered), linear

Of channel DII:
170 .. 179
Of channel DIII:
180 .. 189
Of channel AVR:
190 .. 199
Of channel AVL:
200 .. 209
Of channel AVF:
210 .. 219
Of channel V1:
220 .. 229
Of channel V2:
230 .. 239
Of channel V3:
240 .. 249
Of channel V4:
250 .. 259
Of channel V5:
260 .. 269
Of channel V6:
270 .. 279

8. Missing Attribute Values: Several. Distinguished with '?'.

9. Class Distribution:

Database: Arrhythmia

Class code :	Class :	Number of instances:
01	Normal	245
02	Ischemic changes (Coronary Artery Disease)	44
03	Old Anterior Myocardial Infarction	15
04	Old Inferior Myocardial Infarction	15
05	Sinus tachycardy	13
06	Sinus bradycardy	25
07	Ventricular Premature Contraction (PVC)	3
08	Supraventricular Premature Contraction	2
09	Left bundle branch block	9
10	Right bundle branch block	50
11	1. degree AtrioVentricular block	0
12	2. degree AV block	0
13	3. degree AV block	0
14	Left ventricule hypertrophy	4
15	Atrial Fibrillation or Flutter	5
16	Others	22

APPENDIX B. All k-NN results

KNN7 Results												
Dataset	**TP**	**FN**	**TN**	**FP**	**Accuracy**	**Precision**	**Neg.Precision**	**Recall**	**F-Score**	**Sensitivity**	**Specificity**	**MCC**
CFS3	78	105	222	15	71.4286	0.839	0.679	0.426	0.565	0.426	0.937	0.433
CFS5	92	91	202	35	70	0.724	0.689	0.503	0.594	0.503	0.852	0.383
CFS7	101	82	219	18	76.1905	0.849	0.728	0.552	0.669	0.552	0.924	0.524
CFS10	100	83	212	25	74.2857	0.8	0.719	0.546	0.649	0.546	0.895	0.478
CFS13	103	80	215	22	75.7143	0.824	0.729	0.563	0.669	0.563	0.907	0.51
CFS15	89	94	219	18	73.3333	0.832	0.7	0.486	0.614	0.486	0.924	0.467
CFS20	105	78	217	20	76.6667	0.84	0.736	0.574	0.682	0.574	0.916	0.531
PCA3	71	112	206	31	65.9524	0.696	0.648	0.388	0.498	0.388	0.869	0.297
PCA5	80	103	215	22	70.2381	0.784	0.676	0.437	0.561	0.437	0.907	0.398
PCA7	81	102	210	27	69.2857	0.75	0.673	0.443	0.557	0.443	0.886	0.373
PCA10	82	101	200	37	67.1429	0.689	0.664	0.448	0.543	0.448	0.844	0.321
PCA13	70	113	212	25	67.1429	0.737	0.652	0.383	0.504	0.383	0.895	0.328
PCA15	68	115	201	36	64.0476	0.654	0.636	0.372	0.474	0.372	0.848	0.252
PCA20	76	107	200	37	65.7143	0.673	0.651	0.415	0.513	0.415	0.844	0.29
RFE-SVM3	95	88	202	35	70.7143	0.731	0.697	0.519	0.607	0.519	0.852	0.398
RFE-SVM5	102	81	209	28	74.0476	0.785	0.721	0.557	0.652	0.557	0.882	0.471

RFE-SVM7	116	67	218	19	79.5238	0.859	0.765	0.634	0.73	0.634	0.92	0.588
RFE-SVM10	103	80	220	17	76.9048	0.858	0.733	0.563	0.68	0.563	0.928	0.539
RFE-SVM13	111	72	222	15	79.2857	0.881	0.755	0.607	0.719	0.607	0.937	0.588
RFE-SVM15	110	73	219	18	78.3333	0.859	0.75	0.601	0.707	0.601	0.924	0.566
RFE-SVM20	101	82	226	11	77.8571	0.902	0.734	0.552	0.685	0.552	0.954	0.567
FA3	111	72	209	28	76.1905	0.799	0.744	0.607	0.69	0.607	0.882	0.515
FA5	98	85	217	20	75	0.831	0.719	0.536	0.652	0.536	0.916	0.498
FA7	104	79	217	20	76.4286	0.839	0.733	0.568	0.677	0.568	0.916	0.526
FA10	104	79	223	14	77.8571	0.881	0.738	0.568	0.691	0.568	0.941	0.562
FA13	101	82	217	20	75.7143	0.835	0.726	0.552	0.665	0.552	0.916	0.512
EMP-ECG	72	111	208	29	66.6667	0.713	0.652	0.393	0.507	0.393	0.878	0.315

KNN9 Results												
Dataset	**TP**	**FN**	**TN**	**FP**	**Accuracy**	**Precision**	**Neg.Precision**	**Recall**	**F-Score**	**Sensitivity**	**Specificity**	**MCC**
CFS3	78	105	222	15	71.4286	0.839	0.679	0.426	0.565	0.426	0.937	0.433
CFS5	93	90	207	30	71.4286	0.756	0.697	0.508	0.608	0.508	0.873	0.416
CFS7	101	82	219	18	76.1905	0.849	0.728	0.552	0.669	0.552	0.924	0.524
CFS10	100	83	212	25	74.2857	0.8	0.719	0.546	0.649	0.546	0.895	0.478
CFS13	98	85	217	20	75	0.831	0.719	0.536	0.652	0.536	0.916	0.498

CFS15	89	94	221	16	73.8095	0.848	0.702	0.486	0.618	0.486	0.932	0.48
CFS20	102	81	217	20	75.9524	0.836	0.728	0.557	0.669	0.557	0.916	0.517
PCA3	71	112	206	31	65.9524	0.696	0.648	0.388	0.498	0.388	0.869	0.297
PCA5	81	102	213	24	70	0.771	0.676	0.443	0.563	0.443	0.899	0.391
PCA7	78	105	215	22	69.7619	0.78	0.672	0.426	0.551	0.426	0.907	0.388
PCA10	79	104	199	38	66.1905	0.675	0.657	0.432	0.527	0.432	0.84	0.3
PCA13	70	113	212	25	67.1429	0.737	0.652	0.383	0.504	0.383	0.895	0.328
PCA15	66	117	197	40	62.619	0.623	0.627	0.361	0.457	0.361	0.831	0.219
PCA20	68	115	205	32	65	0.68	0.641	0.372	0.481	0.372	0.865	0.275
RFE-SVM3	94	89	208	29	71.9048	0.764	0.7	0.514	0.615	0.514	0.878	0.426
RFE-SVM5	112	71	208	29	76.1905	0.794	0.746	0.612	0.691	0.612	0.878	0.514
RFE-SVM7	116	67	219	18	79.7619	0.866	0.766	0.634	0.732	0.634	0.924	0.594
RFE-SVM10	104	79	220	17	77.1429	0.86	0.736	0.568	0.684	0.568	0.928	0.544
RFE-SVM13	106	77	225	12	78.8095	0.898	0.745	0.579	0.704	0.579	0.949	0.583
RFE-SVM15	105	78	224	13	78.3333	0.89	0.742	0.574	0.698	0.574	0.945	0.572
RFE-SVM20	101	82	228	9	78.3333	0.918	0.735	0.552	0.689	0.552	0.962	0.58
FA3	102	81	220	17	76.6667	0.857	0.731	0.557	0.675	0.557	0.928	0.534
FA5	96	87	221	16	75.4762	0.857	0.718	0.525	0.651	0.525	0.932	0.513
FA7	100	83	224	13	77.1429	0.885	0.73	0.546	0.675	0.546	0.945	0.55

FA10	104	79	223	14	77.8571	0.881	0.738	0.568	0.691	0.568	0.941	0.562
FA13	103	80	216	21	75.9524	0.831	0.73	0.563	0.671	0.563	0.911	0.516
EMP-ECG	70	113	210	27	66.6667	0.722	0.65	0.383	0.5	0.383	0.886	0.316

KNN11 Results												
Dataset	**TP**	**FN**	**TN**	**FP**	**Accuracy**	**Precision**	**Neg.Precision**	**Recall**	**F-Score**	**Sensitivity**	**Specificity**	**MCC**
CFS3	78	105	222	15	71.4286	0.839	0.679	0.426	0.565	0.426	0.937	0.433
CFS5	93	90	207	30	71.4286	0.756	0.697	0.508	0.608	0.508	0.873	0.416
CFS7	101	82	219	18	76.1905	0.849	0.728	0.552	0.669	0.552	0.924	0.524
CFS10	100	83	212	25	74.2857	0.8	0.719	0.546	0.649	0.546	0.895	0.478
CFS13	98	85	217	20	75	0.831	0.719	0.536	0.652	0.536	0.916	0.498
CFS15	87	96	223	14	73.8095	0.861	0.699	0.475	0.612	0.475	0.941	0.483
CFS20	102	81	217	20	75.9524	0.836	0.728	0.557	0.669	0.557	0.916	0.517
PCA3	69	114	207	30	65.7143	0.697	0.645	0.377	0.489	0.377	0.873	0.293
PCA5	76	107	216	21	69.5238	0.784	0.669	0.415	0.543	0.415	0.911	0.384
PCA7	75	108	215	22	69.0476	0.773	0.666	0.41	0.536	0.41	0.907	0.373
PCA10	78	105	200	37	66.1905	0.678	0.656	0.426	0.523	0.426	0.844	0.3
PCA13	70	113	212	25	67.1429	0.737	0.652	0.383	0.504	0.383	0.895	0.328
PCA15	66	117	197	40	62.619	0.623	0.627	0.361	0.457	0.361	0.831	0.219

PCA20	66	117	210	27	65.7143	0.71	0.642	0.361	0.479	0.361	0.886	0.295
RFE-SVM3	94	89	210	27	72.381	0.777	0.702	0.514	0.619	0.514	0.886	0.438
RFE-SVM5	113	70	208	29	76.4286	0.796	0.748	0.617	0.695	0.617	0.878	0.519
RFE-SVM7	113	70	219	18	79.0476	0.863	0.758	0.617	0.72	0.617	0.924	0.58
RFE-SVM10	104	79	221	16	77.381	0.867	0.737	0.568	0.686	0.568	0.932	0.55
RFE-SVM13	105	78	225	12	78.5714	0.897	0.743	0.574	0.7	0.574	0.949	0.579
RFE-SVM15	104	79	223	14	77.8571	0.881	0.738	0.568	0.691	0.568	0.941	0.562
RFE-SVM20	98	85	231	6	78.3333	0.942	0.731	0.536	0.683	0.536	0.975	0.586
FA3	100	83	222	15	76.6667	0.87	0.728	0.546	0.671	0.546	0.937	0.537
FA5	99	84	220	17	75.9524	0.853	0.724	0.541	0.662	0.541	0.928	0.52
FA7	99	84	225	12	77.1429	0.892	0.728	0.541	0.674	0.541	0.949	0.551
FA10	104	79	223	14	77.8571	0.881	0.738	0.568	0.691	0.568	0.941	0.562
FA13	104	79	217	20	76.4286	0.839	0.733	0.568	0.677	0.568	0.916	0.526
EMP-ECG	65	118	212	25	65.9524	0.722	0.642	0.355	0.476	0.355	0.895	0.302

KNN13 Results												
Dataset	**TP**	**FN**	**TN**	**FP**	**Accuracy**	**Precision**	**Neg.Precision**	**Recall**	**F-Score**	**Sensitivity**	**Specificity**	**MCC**
CFS3	78	105	222	15	71.4286	0.839	0.679	0.426	0.565	0.426	0.937	0.433
CFS5	93	90	207	30	71.4286	0.756	0.697	0.508	0.608	0.508	0.873	0.416

CFS7	101	82	219	18	76.1905	0.849	0.728	0.552	0.669	0.552	0.924	0.524
CFS10	100	83	212	25	74.2857	0.8	0.719	0.546	0.649	0.546	0.895	0.478
CFS13	98	85	217	20	75	0.831	0.719	0.536	0.652	0.536	0.916	0.498
CFS15	87	96	223	14	73.8095	0.861	0.699	0.475	0.612	0.475	0.941	0.483
CFS20	102	81	217	20	75.9524	0.836	0.728	0.557	0.669	0.557	0.916	0.517
PCA3	70	113	202	35	64.7619	0.667	0.641	0.383	0.487	0.383	0.852	0.269
PCA5	76	107	216	21	69.5238	0.784	0.669	0.415	0.543	0.415	0.911	0.384
PCA7	75	108	215	22	69.0476	0.773	0.666	0.41	0.536	0.41	0.907	0.373
PCA10	79	104	201	36	66.6667	0.687	0.659	0.432	0.53	0.432	0.848	0.311
PCA13	69	114	213	24	67.1429	0.742	0.651	0.377	0.5	0.377	0.899	0.329
PCA15	66	117	197	40	62.619	0.623	0.627	0.361	0.457	0.361	0.831	0.219
PCA20	66	117	210	27	65.7143	0.71	0.642	0.361	0.479	0.361	0.886	0.295
RFE-SVM3	93	90	210	27	72.1429	0.775	0.7	0.508	0.614	0.508	0.886	0.433
RFE-SVM5	113	70	210	27	76.9048	0.807	0.75	0.617	0.699	0.617	0.886	0.53
RFE-SVM7	113	70	219	18	79.0476	0.863	0.758	0.617	0.72	0.617	0.924	0.58
RFE-SVM10	104	79	221	16	77.381	0.867	0.737	0.568	0.686	0.568	0.932	0.55
RFE-SVM13	104	79	225	12	78.3333	0.897	0.74	0.568	0.696	0.568	0.949	0.574
RFE-SVM15	103	80	225	12	78.0952	0.896	0.738	0.563	0.691	0.563	0.949	0.57
RFE-SVM20	98	85	231	6	78.3333	0.942	0.731	0.536	0.683	0.536	0.975	0.586

FA3	101	82	224	13	77.381	0.886	0.732	0.552	0.68	0.552	0.945	0.554
FA5	97	86	221	16	75.7143	0.858	0.72	0.53	0.655	0.53	0.932	0.517
FA7	99	84	225	12	77.1429	0.892	0.728	0.541	0.674	0.541	0.949	0.551
FA10	104	79	223	14	77.8571	0.881	0.738	0.568	0.691	0.568	0.941	0.562
FA13	104	79	217	20	76.4286	0.839	0.733	0.568	0.677	0.568	0.916	0.526
EMP-ECG	65	118	216	21	66.9048	0.756	0.647	0.355	0.483	0.355	0.911	0.328

KNN15 Results												
Dataset	**TP**	**FN**	**TN**	**FP**	**Accuracy**	**Precision**	**Neg.Precision**	**Recall**	**F-Score**	**Sensitivity**	**Specificity**	**MCC**
CFS3	78	105	222	15	71.4286	0.839	0.679	0.426	0.565	0.426	0.937	0.433
CFS5	90	93	207	30	70.7143	0.75	0.69	0.492	0.594	0.492	0.873	0.401
CFS7	101	82	219	18	76.1905	0.849	0.728	0.552	0.669	0.552	0.924	0.524
CFS10	100	83	212	25	74.2857	0.8	0.719	0.546	0.649	0.546	0.895	0.478
CFS13	98	85	217	20	75	0.831	0.719	0.536	0.652	0.536	0.916	0.498
CFS15	87	96	225	12	74.2857	0.879	0.701	0.475	0.617	0.475	0.949	0.496
CFS20	102	81	217	20	75.9524	0.836	0.728	0.557	0.669	0.557	0.916	0.517
PCA3	67	116	206	31	65	0.684	0.64	0.366	0.477	0.366	0.869	0.276
PCA5	76	107	213	24	68.8095	0.76	0.666	0.415	0.537	0.415	0.899	0.366
PCA7	73	110	219	18	69.5238	0.802	0.666	0.399	0.533	0.399	0.924	0.389

PCA10	79	104	201	36	66.6667	0.687	0.659	0.432	0.53	0.432	0.848	0.311
PCA13	69	114	213	24	67.1429	0.742	0.651	0.377	0.5	0.377	0.899	0.329
PCA15	65	118	198	39	62.619	0.625	0.627	0.355	0.453	0.355	0.835	0.219
PCA20	66	117	210	27	65.7143	0.71	0.642	0.361	0.479	0.361	0.886	0.295
RFE-SVM3	95	88	209	28	72.381	0.772	0.704	0.519	0.621	0.519	0.882	0.437
RFE-SVM5	113	70	210	27	76.9048	0.807	0.75	0.617	0.699	0.617	0.886	0.53
RFE-SVM7	113	70	219	18	79.0476	0.863	0.758	0.617	0.72	0.617	0.924	0.58
RFE-SVM10	104	79	221	16	77.381	0.867	0.737	0.568	0.686	0.568	0.932	0.55
RFE-SVM13	104	79	225	12	78.3333	0.897	0.74	0.568	0.696	0.568	0.949	0.574
RFE-SVM15	103	80	225	12	78.0952	0.896	0.738	0.563	0.691	0.563	0.949	0.57
RFE-SVM20	98	85	231	6	78.3333	0.942	0.731	0.536	0.683	0.536	0.975	0.586
FA3	100	83	227	10	77.8571	0.909	0.732	0.546	0.682	0.546	0.958	0.569
FA5	97	86	225	12	76.6667	0.89	0.723	0.53	0.664	0.53	0.949	0.542
FA7	99	84	225	12	77.1429	0.892	0.728	0.541	0.674	0.541	0.949	0.551
FA10	104	79	223	14	77.8571	0.881	0.738	0.568	0.691	0.568	0.941	0.562
FA13	104	79	217	20	76.4286	0.839	0.733	0.568	0.677	0.568	0.916	0.526
EMP-ECG	64	119	224	13	68.5714	0.831	0.653	0.35	0.493	0.35	0.945	0.378

KNN17 Results												
Dataset	**TP**	**FN**	**TN**	**FP**	**Accuracy**	**Precision**	**Neg.Precision**	**Recall**	**F-Score**	**Sensitivity**	**Specificity**	**MCC**
CFS3	78	105	222	15	71.4286	0.839	0.679	0.426	0.565	0.426	0.937	0.433
CFS5	90	93	207	30	70.7143	0.75	0.69	0.492	0.594	0.492	0.873	0.401
CFS7	101	82	219	18	76.1905	0.849	0.728	0.552	0.669	0.552	0.924	0.524
CFS10	100	83	212	25	74.2857	0.8	0.719	0.546	0.649	0.546	0.895	0.478
CFS13	98	85	217	20	75	0.831	0.719	0.536	0.652	0.536	0.916	0.498
CFS15	87	96	225	12	74.2857	0.879	0.701	0.475	0.617	0.475	0.949	0.496
CFS20	102	81	217	20	75.9524	0.836	0.728	0.557	0.669	0.557	0.916	0.517
PCA3	63	120	202	35	63.0952	0.643	0.627	0.344	0.448	0.344	0.852	0.23
PCA5	74	109	212	25	68.0952	0.747	0.66	0.404	0.524	0.404	0.895	0.349
PCA7	72	111	218	19	69.0476	0.791	0.663	0.393	0.525	0.393	0.92	0.377
PCA10	79	104	201	36	66.6667	0.687	0.659	0.432	0.53	0.432	0.848	0.311
PCA13	69	114	213	24	67.1429	0.742	0.651	0.377	0.5	0.377	0.899	0.329
PCA15	65	118	199	38	62.8571	0.631	0.628	0.355	0.454	0.355	0.84	0.225
PCA20	66	117	210	27	65.7143	0.71	0.642	0.361	0.479	0.361	0.886	0.295
RFE-SVM3	97	86	210	27	73.0952	0.782	0.709	0.53	0.632	0.53	0.886	0.452
RFE-SVM5	113	70	210	27	76.9048	0.807	0.75	0.617	0.699	0.617	0.886	0.53
RFE-SVM7	113	70	219	18	79.0476	0.863	0.758	0.617	0.72	0.617	0.924	0.58

RFE-SVM10	104	79	221	16	77.381	0.867	0.737	0.568	0.686	0.568	0.932	0.55
RFE-SVM13	104	79	225	12	78.3333	0.897	0.74	0.568	0.696	0.568	0.949	0.574
RFE-SVM15	103	80	225	12	78.0952	0.896	0.738	0.563	0.691	0.563	0.949	0.57
RFE-SVM20	98	85	231	6	78.3333	0.942	0.731	0.536	0.683	0.536	0.975	0.586
FA3	100	83	227	10	77.8571	0.909	0.732	0.546	0.682	0.546	0.958	0.569
FA5	98	85	225	12	76.9048	0.891	0.726	0.536	0.669	0.536	0.949	0.547
FA7	99	84	225	12	77.1429	0.892	0.728	0.541	0.674	0.541	0.949	0.551
FA10	104	79	223	14	77.8571	0.881	0.738	0.568	0.691	0.568	0.941	0.562
FA13	104	79	217	20	76.4286	0.839	0.733	0.568	0.677	0.568	0.916	0.526
EMP-ECG	62	121	226	11	68.5714	0.849	0.651	0.339	0.485	0.339	0.954	0.383

APPENDIX C. All MLP results

Dataset	TP	FN	TN	FP	Accuracy	Precision	Neg.Precision	Recall	F-Score	Sensitivity	Specificity	MCC
CFS3	87	96	208	29	70.2381	0.75	0.684	0.475	0.582	0.475	0.878	0.392
CFS5	123	60	211	26	79.5238	0.826	0.779	0.672	0.741	0.672	0.89	0.583
CFS7	124	59	213	24	80.2381	0.838	0.783	0.678	0.75	0.678	0.899	0.598
CFS10	132	51	206	31	80.4762	0.81	0.802	0.721	0.763	0.721	0.869	0.601
CFS13	129	54	200	37	78.3333	0.777	0.787	0.705	0.739	0.705	0.844	0.557
CFS15	126	57	200	37	77.619	0.773	0.778	0.689	0.729	0.689	0.844	0.542
CFS20	134	49	202	35	80	0.793	0.805	0.732	0.761	0.732	0.852	0.591
PCA3	72	111	214	23	68.0952	0.758	0.658	0.393	0.518	0.393	0.903	0.351
PCA5	96	87	197	40	69.7619	0.706	0.694	0.525	0.602	0.525	0.831	0.377
PCA7	109	74	196	41	72.619	0.727	0.726	0.596	0.655	0.596	0.827	0.437
PCA10	114	69	181	56	70.2381	0.671	0.724	0.623	0.646	0.623	0.764	0.391
PCA13	123	60	183	54	72.8571	0.695	0.753	0.672	0.683	0.672	0.772	0.446
PCA15	119	64	192	45	74.0476	0.726	0.75	0.65	0.686	0.65	0.81	0.468
PCA20	113	70	176	61	68.8095	0.649	0.715	0.617	0.633	0.617	0.743	0.362
RFE-SVM3	91	92	219	18	73.8095	0.835	0.704	0.497	0.623	0.497	0.924	0.477
RFE-SVM5	112	71	211	26	76.9048	0.812	0.748	0.612	0.698	0.612	0.89	0.53
RFE-SVM7	120	63	215	22	79.7619	0.845	0.773	0.656	0.739	0.656	0.907	0.59

RFE-SVM10	117	66	221	16	80.4762	0.88	0.77	0.639	0.74	0.639	0.932	0.61
RFE-SVM13	126	57	206	31	79.0476	0.803	0.783	0.689	0.742	0.689	0.869	0.572
RFE-SVM15	124	59	214	23	80.4762	0.844	0.784	0.678	0.752	0.678	0.903	0.604
RFE-SVM20	115	68	209	28	77.1429	0.804	0.755	0.628	0.705	0.628	0.882	0.534
FA3	102	81	216	21	75.7143	0.829	0.727	0.557	0.666	0.557	0.911	0.511
FA5	109	74	215	22	77.1429	0.832	0.744	0.596	0.694	0.596	0.907	0.538
FA7	110	73	214	23	77.1429	0.827	0.746	0.601	0.696	0.601	0.903	0.537
FA10	133	50	210	27	81.6667	0.831	0.808	0.727	0.776	0.727	0.886	0.626
FA13	134	49	200	37	79.5238	0.784	0.803	0.732	0.757	0.732	0.844	0.581
EMP-ECG	116	67	195	42	74.0476	0.734	0.744	0.634	0.68	0.634	0.823	0.467

APPENDIX D. All SVM results

Dataset	TP	FN	TN	FP	Accuracy	Precision	Neg.Precision	Recall	F-Score	Sensitivity	Specificity	MCC
CFS3	101	82	213	24	74.7619	0.808	0.722	0.552	0.656	0.552	0.899	0.489
CFS5	118	65	210	27	78.0952	0.814	0.764	0.645	0.72	0.645	0.886	0.554
CFS7	128	55	219	18	82.619	0.877	0.799	0.699	0.778	0.699	0.924	0.649
CFS10	135	48	207	30	81.4286	0.818	0.812	0.738	0.776	0.738	0.873	0.62
CFS13	138	45	204	33	81.4286	0.807	0.819	0.754	0.78	0.754	0.861	0.621
CFS15	128	55	210	27	80.4762	0.826	0.792	0.699	0.757	0.699	0.886	0.602
CFS20	130	53	216	21	82.381	0.861	0.803	0.71	0.778	0.71	0.911	0.642
PCA3	74	109	217	20	69.2857	0.787	0.666	0.404	0.534	0.404	0.916	0.381
PCA5	94	89	203	34	70.7143	0.734	0.695	0.514	0.605	0.514	0.857	0.399
PCA7	101	82	208	29	73.5714	0.777	0.717	0.552	0.645	0.552	0.878	0.461
PCA10	109	74	216	21	77.381	0.838	0.745	0.596	0.697	0.596	0.911	0.544
PCA13	117	66	213	24	78.5714	0.83	0.763	0.639	0.722	0.639	0.899	0.565
PCA15	116	67	215	22	78.8095	0.841	0.762	0.634	0.723	0.634	0.907	0.571
PCA20	116	67	208	29	77.1429	0.8	0.756	0.634	0.707	0.634	0.878	0.533
RFE-SVM3	79	104	220	17	71.1905	0.823	0.679	0.432	0.567	0.432	0.928	0.425
RFE-SVM5	105	78	217	20	76.6667	0.84	0.736	0.574	0.682	0.574	0.916	0.531
RFE-SVM7	113	70	219	18	79.0476	0.863	0.758	0.617	0.72	0.617	0.924	0.58

RFE-SVM10	122	61	218	19	80.9524	0.865	0.781	0.667	0.753	0.667	0.92	0.616
RFE-SVM13	118	65	227	10	82.1429	0.922	0.777	0.645	0.759	0.645	0.958	0.649
RFE-SVM15	120	63	222	15	81.4286	0.889	0.779	0.656	0.755	0.656	0.937	0.629
RFE-SVM20	124	59	224	13	82.8571	0.905	0.792	0.678	0.775	0.678	0.945	0.659
FA3	107	79	225	12	78.487	0.899	0.74	0.575	0.701	0.575	0.949	0.579
FA5	111	72	226	11	80.2381	0.91	0.758	0.607	0.728	0.607	0.954	0.612
FA7	112	71	224	13	80	0.896	0.759	0.612	0.727	0.612	0.945	0.604
FA10	129	54	220	17	83.0952	0.884	0.803	0.705	0.784	0.705	0.928	0.659
FA13	131	52	216	21	82.619	0.862	0.806	0.716	0.782	0.716	0.911	0.647
EMP-ECG	110	73	206	31	75.2381	0.78	0.738	0.601	0.679	0.601	0.869	0.494

APPENDIX E. All FSVM results

FSVM-OCW Results												
Dataset	**TP**	**FN**	**TN**	**FP**	**Accuracy**	**Precision**	**Neg.Precision**	**Recall**	**F-Score**	**Sensitivity**	**Specificity**	**MCC**
CFS3	181	2	8	229	45	0.441	0.8	0.989	0.61	0.989	0.034	0.074
CFS5	157	29	112	125	63.5934	0.557	0.794	0.844	0.671	0.844	0.473	0.333
CFS7	156	27	137	100	69.7619	0.609	0.835	0.852	0.71	0.852	0.578	0.438
CFS10	158	25	155	82	74.5238	0.658	0.861	0.863	0.747	0.863	0.654	0.518
CFS13	159	24	172	65	78.8095	0.71	0.878	0.869	0.781	0.869	0.726	0.591
CFS15	167	16	135	102	71.9048	0.621	0.894	0.913	0.739	0.913	0.57	0.498
CFS20	160	23	155	82	75	0.661	0.871	0.874	0.753	0.874	0.654	0.53
PCA3	179	4	10	227	45	0.441	0.714	0.978	0.608	0.978	0.042	0.056
PCA5	183	0	0	237	43.5714	0.436	0	1	0.607	1	0	0
PCA7	169	14	25	212	46.1905	0.444	0.641	0.923	0.6	0.923	0.105	0.05
PCA10	172	11	24	213	46.6667	0.447	0.686	0.94	0.606	0.94	0.101	0.074
PCA13	159	24	120	117	66.4286	0.576	0.833	0.869	0.693	0.869	0.506	0.392
PCA15	172	11	74	163	58.5714	0.513	0.871	0.94	0.664	0.94	0.312	0.311
PCA20	163	20	92	145	60.7143	0.529	0.821	0.891	0.664	0.891	0.388	0.313
RFE-SVM3	183	0	0	237	43.5714	0.436	0	1	0.607	1	0	0
RFE-SVM5	136	47	102	135	56.6667	0.502	0.685	0.743	0.599	0.743	0.43	0.18

RFE-SVM7	156	27	87	150	57.8571	0.51	0.763	0.852	0.638	0.852	0.367	0.245
RFE-SVM10	180	3	6	231	44.2857	0.438	0.667	0.984	0.606	0.984	0.025	0.031
RFE-SVM13	183	0	0	237	43.5714	0.436	0	1	0.607	1	0	0
RFE-SVM15	183	0	0	237	43.5714	0.436	0	1	0.607	1	0	0
RFE-SVM20	151	32	148	89	71.1905	0.629	0.822	0.825	0.714	0.825	0.624	0.45
FA3	176	7	8	229	43.8095	0.435	0.533	0.962	0.599	0.962	0.034	-0.012
FA5	183	0	0	237	43.5714	0.436	0	1	0.607	1	0	0
FA7	176	7	9	228	44.0476	0.436	0.563	0.962	0.6	0.962	0.038	-0.001
FA10	163	20	147	90	73.8095	0.644	0.88	0.891	0.748	0.891	0.62	0.518
FA13	159	24	151	86	73.8095	0.649	0.863	0.869	0.743	0.869	0.637	0.509
EMP-ECG	164	19	75	162	56.9048	0.503	0.798	0.896	0.644	0.896	0.316	0.253

FSVM-DTCM Results												
Dataset	**TP**	**FN**	**TN**	**FP**	**Accuracy**	**Precision**	**Neg.Precision**	**Recall**	**F-Score**	**Sensitivity**	**Specificity**	**MCC**
CFS3	98	85	216	21	74.7619	0.824	0.718	0.536	0.65	0.536	0.911	0.492
CFS5	126	57	211	26	80.2381	0.829	0.787	0.689	0.753	0.689	0.89	0.597
CFS7	123	60	220	17	81.6667	0.879	0.786	0.672	0.762	0.672	0.928	0.632
CFS10	138	45	212	25	83.3333	0.847	0.825	0.754	0.798	0.754	0.895	0.66
CFS13	136	47	208	29	81.9048	0.824	0.816	0.743	0.781	0.743	0.878	0.63

CFS15	128	55	213	24	81.1905	0.842	0.795	0.699	0.764	0.699	0.899	0.617
CFS20	127	56	219	18	82.381	0.876	0.796	0.694	0.774	0.694	0.924	0.645
PCA3	74	109	221	16	70.2381	0.822	0.67	0.404	0.542	0.404	0.932	0.407
PCA5	90	93	203	34	69.7619	0.726	0.686	0.492	0.587	0.492	0.857	0.379
PCA7	106	77	208	29	74.7619	0.785	0.73	0.579	0.666	0.579	0.878	0.485
PCA10	107	76	215	22	76.6667	0.829	0.739	0.585	0.686	0.585	0.907	0.529
PCA13	112	71	216	21	78.0952	0.842	0.753	0.612	0.709	0.612	0.911	0.558
PCA15	108	75	219	18	77.8571	0.857	0.745	0.59	0.699	0.59	0.924	0.556
PCA20	113	70	214	23	77.8571	0.831	0.754	0.617	0.708	0.617	0.903	0.551
RFE-SVM3	74	109	226	11	71.4286	0.871	0.675	0.404	0.552	0.404	0.954	0.442
RFE-SVM5	105	78	214	23	75.9524	0.82	0.733	0.574	0.675	0.574	0.903	0.514
RFE-SVM7	114	69	215	22	78.3333	0.838	0.757	0.623	0.715	0.623	0.907	0.562
RFE-SVM10	121	62	216	21	80.2381	0.852	0.777	0.661	0.744	0.661	0.911	0.6
RFE-SVM13	113	70	225	12	80.4762	0.904	0.763	0.617	0.733	0.617	0.949	0.615
RFE-SVM15	114	69	226	11	80.9524	0.912	0.766	0.623	0.74	0.623	0.954	0.625
RFE-SVM20	121	62	223	14	81.9048	0.896	0.782	0.661	0.761	0.661	0.941	0.639
FA3	107	76	223	14	78.5714	0.884	0.746	0.585	0.704	0.585	0.941	0.576
FA5	108	75	225	12	79.2857	0.9	0.75	0.59	0.713	0.59	0.949	0.592
FA7	110	73	226	11	80	0.909	0.756	0.601	0.724	0.601	0.954	0.607

FA10	123	60	221	16	81.9048	0.885	0.786	0.672	0.764	0.672	0.932	0.637
FA13	124	59	215	22	80.7143	0.849	0.785	0.678	0.754	0.678	0.907	0.609
EMP-ECG	115	68	209	28	77.1429	0.804	0.755	0.628	0.705	0.628	0.882	0.534

FSVM-DTOCM Results												
Dataset	**TP**	**FN**	**TN**	**FP**	**Accuracy**	**Precision**	**Neg.Precision**	**Recall**	**F-Score**	**Sensitivity**	**Specificity**	**MCC**
CFS3	99	84	216	21	75	0.825	0.72	0.541	0.653	0.541	0.911	0.497
CFS5	124	59	210	27	79.5238	0.821	0.781	0.678	0.743	0.678	0.886	0.582
CFS7	125	58	220	17	82.1429	0.88	0.791	0.683	0.769	0.683	0.928	0.641
CFS10	137	46	211	26	82.8571	0.84	0.821	0.749	0.792	0.749	0.89	0.65
CFS13	138	45	208	29	82.381	0.826	0.822	0.754	0.788	0.754	0.878	0.64
CFS15	129	54	210	27	80.7143	0.827	0.795	0.705	0.761	0.705	0.886	0.606
CFS20	127	56	218	19	82.1429	0.87	0.796	0.694	0.772	0.694	0.92	0.639
PCA3	73	110	221	16	70	0.82	0.668	0.399	0.537	0.399	0.932	0.402
PCA5	88	95	204	33	69.5238	0.727	0.682	0.481	0.579	0.481	0.861	0.374
PCA7	103	80	209	28	74.2857	0.786	0.723	0.563	0.656	0.563	0.882	0.476
PCA10	106	77	216	21	76.6667	0.835	0.737	0.579	0.684	0.579	0.911	0.53
PCA13	114	69	215	22	78.3333	0.838	0.757	0.623	0.715	0.623	0.907	0.562
PCA15	110	73	219	18	78.3333	0.859	0.75	0.601	0.707	0.601	0.924	0.566

PCA20	111	72	213	24	77.1429	0.822	0.747	0.607	0.698	0.607	0.899	0.536
RFE-SVM3	76	107	225	12	71.6667	0.864	0.678	0.415	0.561	0.415	0.949	0.444
RFE-SVM5	106	77	214	23	76.1905	0.822	0.735	0.579	0.679	0.579	0.903	0.518
RFE-SVM7	114	69	215	22	78.3333	0.838	0.757	0.623	0.715	0.623	0.907	0.562
RFE-SVM10	121	62	216	21	80.2381	0.852	0.777	0.661	0.744	0.661	0.911	0.6
RFE-SVM13	114	69	225	12	80.7143	0.905	0.765	0.623	0.738	0.623	0.949	0.619
RFE-SVM15	115	68	225	12	80.9524	0.906	0.768	0.628	0.742	0.628	0.949	0.624
RFE-SVM20	122	61	215	22	80.2381	0.847	0.779	0.667	0.746	0.667	0.907	0.599
FA3	105	78	223	14	78.0952	0.882	0.741	0.574	0.695	0.574	0.941	0.566
FA5	108	75	225	12	79.2857	0.9	0.75	0.59	0.713	0.59	0.949	0.592
FA7	109	74	224	13	79.2857	0.893	0.752	0.596	0.715	0.596	0.945	0.591
FA10	124	59	219	18	81.6667	0.873	0.788	0.678	0.763	0.678	0.924	0.631
FA13	126	57	214	23	80.9524	0.846	0.79	0.689	0.759	0.689	0.903	0.613
EMP-ECG	116	67	209	28	77.381	0.806	0.757	0.634	0.71	0.634	0.882	0.539

FSVM-CAR Results												
Dataset	**TP**	**FN**	**TN**	**FP**	**Accuracy**	**Precision**	**Neg.Precision**	**Recall**	**F-Score**	**Sensitivity**	**Specificity**	**MCC**
CFS3	77	106	230	7	73.0952	0.917	0.685	0.421	0.577	0.421	0.97	0.485
CFS5	101	82	225	12	77.619	0.894	0.733	0.552	0.683	0.552	0.949	0.56

CFS7	122	61	222	15	81.9048	0.891	0.784	0.667	0.763	0.667	0.937	0.638
CFS10	133	50	210	27	81.6667	0.831	0.808	0.727	0.776	0.727	0.886	0.626
CFS13	134	49	206	31	80.9524	0.812	0.808	0.732	0.77	0.732	0.869	0.611
CFS15	125	58	213	24	80.4762	0.839	0.786	0.683	0.753	0.683	0.899	0.603
CFS20	128	55	216	21	81.9048	0.859	0.797	0.699	0.771	0.699	0.911	0.633
PCA3	69	114	223	14	69.5238	0.831	0.662	0.377	0.519	0.377	0.941	0.396
PCA5	40	143	233	4	65	0.909	0.62	0.219	0.353	0.219	0.983	0.327
PCA7	93	90	216	21	73.5714	0.816	0.706	0.508	0.626	0.508	0.911	0.468
PCA10	107	76	216	21	76.9048	0.836	0.74	0.585	0.688	0.585	0.911	0.534
PCA13	114	69	212	25	77.619	0.82	0.754	0.623	0.708	0.623	0.895	0.545
PCA15	112	71	218	19	78.5714	0.855	0.754	0.612	0.713	0.612	0.92	0.569
PCA20	117	66	214	23	78.8095	0.836	0.764	0.639	0.724	0.639	0.903	0.57
RFE-SVM3	73	110	224	13	70.7143	0.849	0.671	0.399	0.543	0.399	0.945	0.423
RFE-SVM5	104	79	218	19	76.6667	0.846	0.734	0.568	0.68	0.568	0.92	0.532
RFE-SVM7	105	78	223	14	78.0952	0.882	0.741	0.574	0.695	0.574	0.941	0.566
RFE-SVM10	120	63	218	19	80.4762	0.863	0.776	0.656	0.745	0.656	0.92	0.607
RFE-SVM13	114	69	224	13	80.4762	0.898	0.765	0.623	0.736	0.623	0.945	0.613
RFE-SVM15	115	68	221	16	80	0.878	0.765	0.628	0.732	0.628	0.932	0.6
RFE-SVM20	121	62	222	15	81.6667	0.89	0.782	0.661	0.759	0.661	0.937	0.634

FA3	97	86	230	7	77.8571	0.933	0.728	0.53	0.676	0.53	0.97	0.575
FA5	92	91	232	5	77.1429	0.948	0.718	0.503	0.657	0.503	0.979	0.567
FA7	97	86	229	8	77.619	0.924	0.727	0.53	0.674	0.53	0.966	0.568
FA10	122	61	223	14	82.1429	0.897	0.785	0.667	0.765	0.667	0.941	0.644
FA13	125	58	220	17	82.1429	0.88	0.791	0.683	0.769	0.683	0.928	0.641
EMP-ECG	115	68	210	27	77.381	0.81	0.755	0.628	0.707	0.628	0.886	0.539

FSVM-FCM-2 Results												
Dataset	**TP**	**FN**	**TN**	**FP**	**Accuracy**	**Precision**	**Neg.Precision**	**Recall**	**F-Score**	**Sensitivity**	**Specificity**	**MCC**
CFS3	102	81	213	24	75	0.81	0.724	0.557	0.66	0.557	0.899	0.494
CFS5	120	63	213	24	79.2857	0.833	0.772	0.656	0.734	0.656	0.899	0.579
CFS7	122	61	222	15	81.9048	0.891	0.784	0.667	0.763	0.667	0.937	0.638
CFS10	134	49	210	27	81.9048	0.832	0.811	0.732	0.779	0.732	0.886	0.631
CFS13	133	50	208	29	81.1905	0.821	0.806	0.727	0.771	0.727	0.878	0.616
CFS15	126	57	212	25	80.4762	0.834	0.788	0.689	0.755	0.689	0.895	0.602
CFS20	123	60	220	17	81.6667	0.879	0.786	0.672	0.762	0.672	0.928	0.632
PCA3	78	105	222	15	71.4286	0.839	0.679	0.426	0.565	0.426	0.937	0.433
PCA5	91	92	206	31	70.7143	0.746	0.691	0.497	0.597	0.497	0.869	0.4
PCA7	104	79	209	28	74.5238	0.788	0.726	0.568	0.66	0.568	0.882	0.481

PCA10	109	74	217	20	77.619	0.845	0.746	0.596	0.699	0.596	0.916	0.55
PCA13	110	73	219	18	78.3333	0.859	0.75	0.601	0.707	0.601	0.924	0.566
PCA15	107	76	219	18	77.619	0.856	0.742	0.585	0.695	0.585	0.924	0.552
PCA20	113	70	218	19	78.8095	0.856	0.757	0.617	0.717	0.617	0.92	0.574
RFE-SVM3	76	107	224	13	71.4286	0.854	0.677	0.415	0.559	0.415	0.945	0.437
RFE-SVM5	107	76	214	23	76.4286	0.823	0.738	0.585	0.684	0.585	0.903	0.523
RFE-SVM7	118	65	214	23	79.0476	0.837	0.767	0.645	0.729	0.645	0.903	0.575
RFE-SVM10	121	62	216	21	80.2381	0.852	0.777	0.661	0.744	0.661	0.911	0.6
RFE-SVM13	116	67	225	12	81.1905	0.906	0.771	0.634	0.746	0.634	0.949	0.628
RFE-SVM15	118	65	222	15	80.9524	0.887	0.774	0.645	0.747	0.645	0.937	0.62
RFE-SVM20	120	63	221	16	81.1905	0.882	0.778	0.656	0.752	0.656	0.932	0.623
FA3	106	77	226	11	79.0476	0.906	0.746	0.579	0.706	0.579	0.954	0.589
FA5	108	75	220	17	78.0952	0.864	0.746	0.59	0.701	0.59	0.928	0.562
FA7	104	79	226	11	78.5714	0.904	0.741	0.568	0.698	0.568	0.954	0.58
FA10	123	60	220	17	81.6667	0.879	0.786	0.672	0.762	0.672	0.928	0.632
FA13	127	56	218	19	82.1429	0.87	0.796	0.694	0.772	0.694	0.92	0.639
EMP-ECG	115	68	212	25	77.8571	0.821	0.757	0.628	0.712	0.628	0.895	0.55

FSVM-FCM-1.57 Results												
Dataset	**TP**	**FN**	**TN**	**FP**	**Accuracy**	**Precision**	**Neg.Precision**	**Recall**	**F-Score**	**Sensitivity**	**Specificity**	**MCC**
CFS3	103	80	208	29	74.0476	0.78	0.722	0.563	0.654	0.563	0.878	0.47
CFS5	120	63	213	24	79.2857	0.833	0.772	0.656	0.734	0.656	0.899	0.579
CFS7	131	52	216	21	82.619	0.862	0.806	0.716	0.782	0.716	0.911	0.647
CFS10	129	54	214	23	81.6667	0.849	0.799	0.705	0.77	0.705	0.903	0.627
CFS13	130	53	212	25	81.4286	0.839	0.8	0.71	0.769	0.71	0.895	0.622
CFS15	124	59	210	27	79.5238	0.821	0.781	0.678	0.743	0.678	0.886	0.582
CFS20	122	61	219	18	81.1905	0.871	0.782	0.667	0.755	0.667	0.924	0.621
PCA3	80	103	211	26	69.2857	0.755	0.672	0.437	0.554	0.437	0.89	0.374
PCA5	87	96	209	28	70.4762	0.757	0.685	0.475	0.584	0.475	0.882	0.397
PCA7	105	78	206	31	74.0476	0.772	0.725	0.574	0.658	0.574	0.869	0.469
PCA10	104	79	215	22	75.9524	0.825	0.731	0.568	0.673	0.568	0.907	0.514
PCA13	114	69	216	21	78.5714	0.844	0.758	0.623	0.717	0.623	0.911	0.567
PCA15	107	76	220	17	77.8571	0.863	0.743	0.585	0.697	0.585	0.928	0.558
PCA20	113	70	218	19	78.8095	0.856	0.757	0.617	0.717	0.617	0.92	0.574
RFE-SVM3	80	103	221	16	71.6667	0.833	0.682	0.437	0.573	0.437	0.932	0.436
RFE-SVM5	105	78	214	23	75.9524	0.82	0.733	0.574	0.675	0.574	0.903	0.514
RFE-SVM7	109	74	218	19	77.8571	0.852	0.747	0.596	0.701	0.596	0.92	0.555

RFE-SVM10	124	59	214	23	80.4762	0.844	0.784	0.678	0.752	0.678	0.903	0.604
RFE-SVM13	121	62	224	13	82.1429	0.903	0.783	0.661	0.763	0.661	0.945	0.645
RFE-SVM15	121	62	220	17	81.1905	0.877	0.78	0.661	0.754	0.661	0.928	0.622
RFE-SVM20	118	65	223	14	81.1905	0.894	0.774	0.645	0.749	0.645	0.941	0.626
FA3	106	77	226	11	79.0476	0.906	0.746	0.579	0.706	0.579	0.954	0.589
FA5	107	76	226	11	79.2857	0.907	0.748	0.585	0.711	0.585	0.954	0.594
FA7	96	87	228	9	77.1429	0.914	0.724	0.525	0.667	0.525	0.962	0.557
FA10	119	64	222	15	81.1905	0.888	0.776	0.65	0.751	0.65	0.937	0.624
FA13	127	56	217	20	81.9048	0.864	0.795	0.694	0.77	0.694	0.916	0.634
EMP-ECG	114	69	211	26	77.381	0.814	0.754	0.623	0.706	0.623	0.89	0.54

REFERENCES

1. Prehospital 12-Lead ECG, *http://ems12lead.blogspot.com/2008/10/12-lead-ecg-lead-placement-diagrams.html*, 2008.

2. Basic Heart Anatomy, *http://kpep.org/specific%20images/basic_heart_anatomy.jpg*, 2008.

3. Normal ECG Image, *http://www.swharden.com/blog/images/ecg2.jpg*, 2008.

4. Pediatric Electrocardiography, *http://pediatriccardiology.uchicago.edu/MP/ECG/ECG1.htm*, 2008.

5. MIT-BIH Arrhythmia Database, *http://www.physionet.org/physiobank/database/html/mitdbdir/mitdbdir.htm*, 1997.

6. UCI Machine Learning Repository Arrhythmia Data Set, *http://archive.ics.uci.edu/ml/datasets/Arrhythmia*, 1998.

7. Silipo, R., and C. Marchesi, "Artificial Neural Networks for automatic ECG analysis", *IEEE Transactions on Signal Processing,* Vol. 46, No. 5, pp. 1417-1425, 1998.

8. Stamkopoulos, T., K. Diamantaras, N. Maglaveras, and M. Strintzis, "ECG analysis using nonlinear PCA neural networks for ischemia detection", *IEEE Transactions on Signal Processing,* Vol. 46, No. 11, pp. 3058-3067, 1998.

9. Wikipedia Ischemia Page, *http://en.wikipedia.org/wiki/Ischemia*, 2009.

10. European ST-T Database, *http://www.physionet.org/physiobank/database/edb/*, 2003.

11. Ghongade, R., and A. Ghatol, *An Effective Feature Set for ECG Pattern Classification,* Lecture Notes in Computer Science*,* No. 4901, pp. 25-32, 2007.

12. Soman, T., and P. O. Bobbie, "Classification of Arrhythmia Using Machine Learning Techniques", *WSEAS Transactions on Computers,* Vol. 4, No. 6, pp. 548-552, 2005.

13. Weka Data Mining Software in Java, *http://www.cs.waikato.ac.nz/ml/weka/*, 2000.

14. Llamedo-Soria, M., and J. P. Martínez, "An ECG classification model based on multilead wavelet transform features", *Computers in Cardiology,* No. 35, pp. 105-108, 2007.

15. Neagoe, V. E., I. F. Iatan, and S. Grunwald, "A neuro-fuzzy approach to classification of ECG signals for ischemic heart disease diagnosis", *AMIA Annu Symp Proc*, pp. 494-498, 2003.

16. Anuradha, B., "Cardiac Arrhythmia Classification Using Fuzzy Classifiers", *Journal of Theoretical and Applied Information Technology*, pp. 353-359, 2008.

17. Lei1, W. K., B. N. Li1, M. C. Dong, and M. I. Vai, "AFC-ECG: An Intelligent Fuzzy ECG Classifier", *Soft Computing in Industrial Applications,* No. 39, pp. 189–199, 2007.

18. Jovic, A., and N. Bogunovic, *Analysis of ECG Records Using ECG Chaos Extractor Platform and Weka System*, Information Technology Interfaces, pp. 347-352, 2008.

19. Zhang, H., and L. Q. Zhang, *ECG analysis based on PCA and Support Vector Machines,* Neural Networks and Brain, pp. 743-747, 2005.

20. Tang, X., and M. Zhiwen, *Electrocardiogram Classification Method Based on SVM*, Intelligent Systems and Knowledge Engineering, 2007.

21. Lei, W. K., B. N. Li, M. C. Dong, and B. B. Fu, *An Application of Morphological Feature Extraction and Support Vector Machines in Computerized ECG Interpretation,* Sixth Mexican International Conference on Artificial Intelligence, 2007.

22. Asl, B. M., S. K. Setarehdan, and M. Mohebb, "Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal", *Artificial Intelligence in Medicine,* Vol. 44, No. 1, pp. 51-64, 2008.

23. Uyar A., *Intelligent arrhythmia classification based on support vector machines*, M.S. Thesis, Boğaziçi University, 2006

24. Çelikyilmaz, A., and I. B. Türksen, "Fuzzy Functions with Support Vector Machines", *Information Sciences,* Vol. 177, no. 23, pp. 5163-5177, 2007.

25. Xiong, S. W., H. B. Liu, and X. X. Niu, *Fuzzy Support Vector Machines Based On FCM Clustering*, Machine Learning and Cybernetics, No. 5, pp. 2608-2613, 2005.

26. Zhang, Y., F. Xie, D. Huang, and M. Ji, *Support vector classifier based on fuzzy c-means and Mahalanobis distance*, Journal of Intelligent Information Systems, 2009.

27. Lin, C., and S. Wang, "Training algorithms for fuzzy support vector machines with noisy data", *Pattern Recognition Letters,* Vol. 25, No. 14, pp. 1647 – 1656, 2004.

28. Orsenigo, C., and C. Vercellis, *Evaluating Membership Functions for Fuzzy Discrete SVM*, Lecture Notes in Computer Science, pp. 187-194, 2007.

29. Chiang, H. M., and T. Y. Wang, "OAO-FSVM classifier approaches to text categorization", *Proceedings of the 13th Asia Pacific Management Conference*, pp. 1454-1462, 2007.

30. Shilton, A., and D. Lai, *Iterative Fuzzy Support Vector Machine Classification*, Fuzzy Systems Conference, pp. 1-6, 2007.

31. Xu, Q., H. Zhou, Y. Wang, and J. Huang, "Fuzzy support vector machine for classification of EEG signals using wavelet-based features", *Medical engineering & physics,* Vol. 31, No. 7, pp. 858-865, 2009.

32. Niedermeyer, E., and F. Lopes da Silva, *Electroencephalography*, Williams & Wilkins, 1999.

33. TEPA Inc., *http://www.tepa.com.tr/*, 2006

34. Furlanello, C., M. Serafini, S. Merler, and G. Jurman, "Gene selection and classification by Entropy-based Recursive Feature Elimination", *Proceedings of the International Joint Conference on Neural Networks,* No. 4, pp. 3077-3082, 2003.

35. Hall, M. A., *Correlation-based feature selection for machine learning*, Doctoral Dissertation, Waikato University, 1998.

36. Kira, K., and L. A. Rendell, "A Practical Approach to Feature Selection", *International Conference on Machine Learning*, pp. 249 – 256, 1992.

37. Rissanen, J., "Modeling by Shortest Data Description", *Automatica,* Vol. 14, No. 5, pp. 465-471, 1978.

38. Kononenko, I., "On biases in estimating multi-valued attributes", *International Joint Conference on Artificial Intelligence,* No. 2, pp. 1034–1040, 1995.

39. Lin, C. F., and S. D. Wang, "Fuzzy Support Vector Machines", *IEEE Transactions on Neural Networks,* Vol. 13, No. 2, pp. 464-471, 2002.

40. Dun, J. C., “A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters”, *Journal of Cybernetics,* Vol. 3, No. 3, pp. 32-57, 1973.

41. Bezdek, J. C., *Pattern Recognition with Fuzzy Objective Function Algoritms*, Kluwer Academic Publishers, 1981.

42. Boser, B. E., I. Guyon, and V. Vapnik, *A Training Algorithm for Optimal Margin Classifiers,* Computational Learning Theory, pp. 144-152, 1992.

43. Vapnik, V., *The Nature of Statistical Learning Theory*, Springer-Verlag, 2000.

44. Cortes, C., and V. Vapnik, “Support-Vector Networks”, *Machine Learning,* Vol. 20, No. 3, pp. 273-297, 1995.

45. Quinlan, J. R., *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers, 1993.

46. Cohen, W. W., “Fast Effective Rule Induction,” *In Proc. of the 12th International Conference on Machine Learning*, pp. 115-123, 1995.

47. Frank, E., and I. H. Witten, “Generating Accurate Rule Sets Without Global Optimization”, *Proceedings of the Fifteenth International Conference on Machine Learning*, pp. 144 – 151, 1998.

48. Jang, J. R., “ANFIS: Adaptive Neuro-Fuzzy Inference Systems”, *IEEE Transactions on Systems, Man, and Cybernetics,* Vol. 23, No. 3, pp. 665-685, 1993.

49. The Mathworks, *http://www.mathworks.com/,* 2000.

50. LIBSVM A Library for Support Vector Machines, *http://www.csie.ntu.edu.tw/~cjlin/libsvm/*, 2009.