AN INVESTIGATION OF SUPPORT VECTOR MACHINE CLASSIFIER IN DETECTING NOCTURNAL AIRWAY OBSTRUCTION FROM SPONTANEOUS HEART RATE COMBINED WITH ECG MORPHOLOGY CHANGES

by

HARSHAN RAVI

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE IN BIOMEDICAL ENGINEERING

THE UNIVERSITY OF TEXAS AT ARLINGTON

AUGUST 2011

ACKNOWLEDGEMENTS

I would like to give a special thanks to Dr.Khosrow Behbehani for his gracious and extended support for my masters and research, for his motivation, patience and immense knowledge. His guidance helped me in all the research and writing of this thesis. Without his valuable suggestions and guidance, this thesis would not have been possible.

I would also like to thank Dr.George Alexandrakis for being my Committee member and a teacher, who is always helpful to the students in need and teaching two interesting optic courses.

Dr. Baohong Yuan deserves special thanks as my thesis committee member and for his encouragement and insightful comments.

My sincere thanks also goes to Dr. Hanzhang Lu for offering me the research assistantship position in his lab, funding me for my last two semesters and leading me working on diverse exciting projects.

I would like to thank my parents Ravi Chandra Sekhar Rao and Ravi JayaLakhsmi for giving birth to me at the first place and supporting me spiritually throughout my life and my brother Ravi vineel for his encouragement and motivation.

Last but not least, I wish to thank my lab mates and roommates for their valuable advice, friendly help and encouraging words.

July 25, 2011

ABSTRACT

AN INVESTIGATION OF SUPPORT VECTOR MACHINE CLASSIFIER IN DETECTING NOCTURNAL AIRWAY OBSTRUCTION FROM SPONTANEOUS HEART RATE COMBINED WITH ECG MORPHOLOGY CHANGES

Harshan Ravi, M.S

The University of Texas at Arlington, 2011

Supervising Professor: Dr. Khosrow Behbehani

Sleep apnea is a sleep disordered breathing resulting from limitation or cessation to breathing for 10 or more seconds. The prevalence of sleep apnea has increased exponentially in the past decade. The increase in epidemic in sleep apnea is associated with increase in the cases of obesity. It is estimated that 12-18 million of American adults suffer from sleep apnea, which is a sizable sector of the adult population. Sleep apnea is risk factor for hypertension, type II diabetes, and congestive heart failure. Sleep apnea may go undiagnosed for a long period of time after its onset due to complexity of diagnosing it. Nocturnal polysomnography is the standard method for diagnosing sleep apnea. It is often inaccessible and costly, hence making sleep apnea an under diagnosed disease. Further, widespread screen of the vulnerable sector of the population (ages 35 and above) is currently not feasible. To overcome these limitations many physiological markers are investigated as alternatives means of detecting sleep apnea. A number of studies during the past decade have investigated the possibility of detecting sleep apnea using features of the electrocardiogram (ECG).

In this study, a support vector machine (SVM) based classifier was developed to detect obstructive sleep apnea (OSA) and normal breathing using features extracted from nocturnal ECG. NPSG was performed on 16 normal patients and 14 OSA patients. This approach combines both RPE and R-R interval to form a cluster. An optimum centroid is extracted from the cluster, and is used as an input to the SVM. The performance of the proposed algorithm in detecting respiratory event was tested by determining its ability to detect normal breathing, and OSA events in 15 minutes data epochs obtained from volunteer normal 16 subjects and 14 apnea patients. The SVM algorithm was designed and optimized using two heuristic and three numerical optimization techniques. For Manual optimization, a highest learning performance of, accuracy of 91.16%, sensitivity of 95.20% and specificity of 86.20% is achieved for training set and a highest testing performance of, accuracy of 75.98%, sensitivity of 81.20 %, and specificity of 69.87% is achieved for testing set. The computerized optimization resulted in slightly higher performance than the Manual optimization. The highest learning performance achieved for training set is, accuracy of 92.78%, sensitivity of 96.33% and specificity of 88.43% and a highest testing performance of, accuracy of 76.66%, sensitivity of 81.84 %, and specificity of 70.41% is achieved for testing set. The detection rates achieved using SVM is comparable to the results achieved with previous study using other form of classifiers [10].

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

ABSTRACT iv

LIST OF ILLUSTRATIONS ix

LIST OF TABLES xi

Chapter Page

1. INTRODUCTION 1

1.1 Sleep Apnea 1

1.1.1 Definition 1

1.1.2 Types of Sleep Apnea 1

1.1.3 Symptoms and Risk factors 2

1.1.4 Prevalence of Sleep Apnea 3

1.1.5 Diagnosis for Sleep Apnea and Treatment 3

1.1.6 Alternate Means of Diagnosing Sleep Apnea 4

1.2 Hypothesis 7

1.3 Organization of Thesis 7

2. METHODS AND MATERIALS 8

2.1 Physiological effects of Sleep Apnea 8

2.2 Brief Introduction to ECG 9

2.2.1 Effect of Sleep Apnea on QRS complex 10

2.2.2 ECG Lead System 11

2.2.3 Data Preprocessing 12

2.3 K-Means Clustering 15

2.3.1 Metrics to Calculate Distance 18
2.3.2 Flow chart for K-Means Algorithm 21
2.4 Support Vector Machine 22
2.4.1 Binary Classification 22
2.4.2 Linearly Separable Case 23
2.4.3 Non Linearly Separable Case 29
2.4.4 Regularization Parameter C 34
2.4.5 Karush-Kuhn-Tucker Conditions 36
2.4.6 Calculation of Support Vectors and Bias Term b 37
2.4.7 Flow Chart for Support Vector Machines 40
2.5 Optimization of Hyperplane Parameters 41
2.5.1 Manual Optimization 41
2.5.2 Computerized Optimization 42
3. EXPERIMENTAL METHOD 44
3.1 Standard NPSG Data Acquisition 44
3.2 Subject Demographics 45
3.3 Processing of Experimental Data 47
3.4 Performance Evaluation for the Proposed Method 49
3.4.1 Cross Validation 49
3.4.2 Monte Carlo Simulation 49
4. RESULTS 51
4.1 Manual Optimization 51
4.1.1 Optimizing of Hyperplane Parameters using Maximum of accuracy Training Set 51
4.1.2 Optimizing Hyperplane Parameters using Maximum of accuracy Validation Set 54
4.2 Computerized Optimization 56

4.2.1 Optimization of Hyperplane Parameters using Fminbnd 56
4.2.2 Optimization of Hyperplane Parameters using Fminsearch 61
4.3 Comparison between different optimization algorithms 63
5. DISCUSSION 64
5.1 Performance of SVM Detection Algorithm 64
5.2 Comparison of Optimization Results 65
5.2.1 Intra optimization results comparison 65
5.2.1 Inter optimization results comparison 66
5.3 Comparison with Previous Studies 67
6. CONCLUSION AND FUTURE DIRECTIONS 69
6.1 Conclusion 69
6.2 Future Directions 69
APPENDIX
A. EUCLIDEAN DISTANCE 72
B. GENERALIZATION ABILITY OF SVM AND CONCEPT OF SOFT MARGIN 74
C. HOLD OUT AND LEAVE ONE OUT CROSS VALIDATION 79
D. SAMPLE PLOTS 81
REFERENCES 87
BIOGRAPHICAL INFORMATION 93

LIST OF ILLUSTRATIONS

Figure Page

1.1 An illustration of the site of obstruction to airflow 2

2.1 Illustration of Einthoven triangle 10

2.2 Graphical representations of important segments of ECG 10

2.3 Illustrate the change QRS amplitude with respiration 11

2.4 ECG before filtering (Top) and after filtering (bottom) 13

2.5 Graphical illustration of R-R interval after resampling 15

2.6 Trace of centroid calculation for two clusters 17

2.7 Trace of centroid calculation for three clusters 18

2.8 K-Means Clustering with Manhattan distance 20

2.9 Flow chart for the K-means algorithm 21

2.10 Hyperplanes that can be drawn for the classification 23

2.11 Optimum hyperplane with maximum margin 24

2.12 linear separating hyperplane for separable case 24

2.13 Non linearly separable data 30

2.14 Transformation of Non linearly separable data into inner product space 30

2.15 Hyperplane is wider for this case of C=1 35

2.16 Hyperplane is broader for this case of C=10000 35

2.17 Flow chart for the support vector machines 40

3.1 ECG electrode placements 45

4.1 Graphical representation accuracy, sensitivity and specificity mean results for training and testing sets after 500 Monte Carlo runs for Manual 1 52

4.2 Illustration of histogram of training accuracy for
500 Monte Carlo runs for Manual1 53

4.3 Illustration of histogram of testing accuracy for
500 Monte Carlo runs for Manual1 53

4.4 Graphical representation accuracy, sensitivity and specificity
mean results for training and testing sets after
500 Monte Carlo runs for Manual 2 54

4.5 Illustration of histogram of training accuracy for
500 Monte Carlo runs for Manual2 55

4.6 Illustration of histogram of testing accuracy for
500 Monte Carlo runs for Manual2 55

4.7 Graphical representation accuracy, sensitivity and specificity
mean results for training and testing sets after
500 Monte Carlo runs for Fminbnd1 57

4.8 Illustration of histogram of training accuracy for
500 Monte Carlo runs for Fminbnd1 58

4.9 Illustration of histogram of training accuracy for
500 Monte Carlo runs for Fminbnd1 58

4.10 Graphical representation accuracy, sensitivity and specificity
mean results for training and testing sets after
500 Monte Carlo runs for Fminbnd2 59

4.11 Illustration of histogram of training accuracy for
500 Monte Carlo runs for Fminbnd 2 60

4.12 Illustration of histogram of training accuracy for
500 Monte Carlo runs for Fminbnd 2 61

4.13 Graphical representation accuracy, sensitivity and specificity
mean results for training and testing sets after
500 Monte Carlo runs for Fminsearch 62

4.14 Illustration of histogram of training accuracy for
500 Monte Carlo runs for Fminsearch 63

4.15 Illustration of histogram of training accuracy for
500 Monte Carlo runs for Fminsearch 64

LIST OF TABLES

Table Page

3.1 Leads and Combination of their Electrode Placements 45

3.2 Subject Demographics for NOR Group. 46

3.3 Table 3.3 Subject Demographics for OSA Group. 47

3.4 List of the Clips Selected from Subjects 48

4.1 Summary of the Sensitivity, specificity and accuracy mean and standard deviation (std) for training and testing sets after 500-run Monte-Carlo simulation for Manual1 52

4.2 Summary of the Sensitivity, specificity and accuracy mean and standard deviation (std) for training and testing sets after 500-run Monte-Carlo simulation for Manual2 54

4.3 Summary of the Sensitivity, specificity and accuracy mean and standard deviation (std) for training and testing sets after 500-run Monte-Carlo simulation for Fminbnd1 57

4.4 Summary of the Sensitivity, specificity and accuracy mean and standard deviation (std) for training and testing sets after 500-run Monte-Carlo simulation for Fminbnd2 59

4.5 Summary of the Sensitivity, specificity and accuracy mean and standard deviation (std) for training and testing sets after 500-run Monte-Carlo simulation for Fminsearch 61

4.6 Summary of the Sensitivity, specificity and accuracy mean and standard deviation (std) for training and testing sets after 500-run Monte-Carlo simulation for the five discussed optimizations 63

CHAPTER 1

INTRODUCTION

1.1 Sleep Apnea

1.1.1 Definition

Sleep apnea is a sleep disordered breathing (SDB) resulting from limitation of respiratory airflow or complete cessation of breathing during sleep. If the obstruction to breathing is complete and for ten or more seconds, then the episode is called apnea. During sleep period, apneic events may ensue once or multiple times per hour where the duration of the event may be ten or more seconds. During sleep apnea, patients do not get sufficient oxygen during sleep. The result is patient being completely out of sleep or a transition from deep sleep to light sleep. This results in poor sleep hygiene that leads to fatigue and daytime sleepiness. Sleep hypopnea is a sleep disordered breathing, in which the patient takes abnormally shallow breathing during sleep resulting in disruption of sleep. During sleep hypopnea Carbon dioxide (CO2) level increases and oxygen (O2) level decreases in the blood. This change in O2 and CO2 concentration is directly proportional to obstruction to airflow. This atypical breathing pattern leads to disruption of sleep [10].

1.1.2 Types of sleep apnea

Sleep apnea is categorized into three types: obstructive sleep apnea (OSA), central sleep apnea (CSA) and mixed sleep apnea. Obstructive sleep apnea is due to physical obstruction to airflow at upper airway. Figure 1.1 shows a pictorial representation of site of obstruction. This type of sleep apnea is most prevalent in overweight male between 35 and 50 years old, who has large tonsils, small jaw, and a small opening of the airway [19]. When this condition is present, the muscles of the soft palate and the muscles at the base tongue relax and sag, causing obstruction to breathing during sleep. Approximately, 84% of all sleep apnea patients

suffer from OSA [14]. The severity of OSA is graded with the Apnea-Hypopnea Index (AHI). A person who has an AHI of 30, experiences 30 abnormal respiratory events per hour of sleep which is a severe level of disease. An AHI of 5-15 is cogitated as mild OSA, an AHI of 15-30 as moderate OSA and an AHI greater than 30 as severe OSA. Central sleep apnea is due to lack respiratory effort. When this condition is extant, for a brief moment of time the chest and diaphragm muscles which control breathing do not receive signals from brain. This is a rare form of disorder where airway remains open but muscles controlling breathing fails. About 0.4% of all sleep apnea patients suffer from CSA [14]. Mixed sleep apnea is due to transition between long periods of OSA and brief intervals of CSA. It is estimated 15% of all sleep apnea patients suffer from mixed sleep apnea [14].

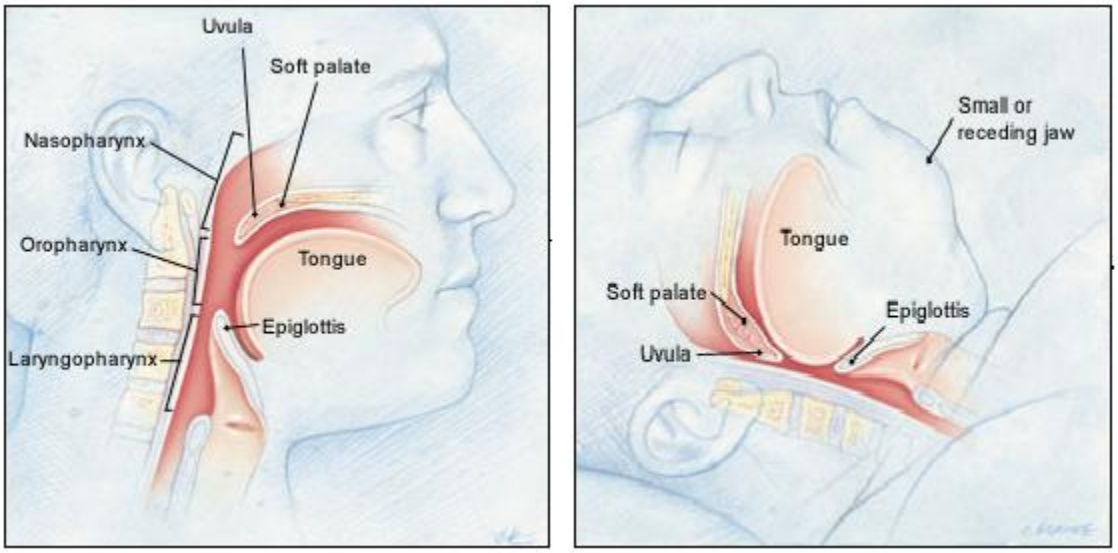


Figure 1.1 An illustration of the site of obstruction to airflow. Adopted form [18]

1.1.3 Symptoms and Risk factors

The most common symptoms of sleep apnea are loud snoring, choking or respiration pause during sleep, and feeling wooziness during day. Some less common symptoms are morning headaches, memory problems, feeling irritable, frequent urination at night, and dry throat [3]. Sleep apnea is also a risk factor for hypertension, angina pectoris, TYPE- || Diabetes and stroke

[15]. Previous studies have also confirmed that OSA causes hypertension and contributes to Essential hypertension [2].

1.1.4 Prevalence of Sleep Apnea

Sleep apnea is considered a common disorder which goes undiagnosed for long period of time [17]. The Wisconsin Sleep Cohort Study estimated in 2008 that roughly one in five adults has mild sleep apnea and one in 25 adults has severe sleep apnea. According to one study about 12-18 million of USA adults suffer from sleep apnea [11]. Sleep apnea cost 15.9 billion dollars as direct cost with an estimated 50 to 100 billion in indirect and related costs a year on American economy [12]. The above discussed studies suggest that quite a sizable population is suffering from sleep apnea and early diagnosis of sleep apnea prevents further advent.

1.1.5 Diagnosis for Sleep Apnea and Treatment

Doctors diagnose sleep apnea based on patient's medical and family history, report of symptoms and nocturnal sleep studies using polysomnograph. Nocturnal polysomnograph (NPSG) is a standard tool to diagnosis sleep apnea. This test is conducted during overnight stay in sleep laboratory where multiple physiological markers are recorded simultaneously. Electroencephalography (electrical activity of brain), Electromyography (skeletal muscles electrical activity), Electrooculography (eye movements), Electrocardiography (heart rate), oxygen saturation and Plethysmography (respiratory flow) are measured during the test. Using NPSG data, a certified sleep specialist blind to aim of the study scores apnea on the recorded results; if apneic episodes are more than five episodes per hour of a sleep and they are concomitant with blood oxygen desaturation. Further, sleep expert considers if apnea episodes are associated with the frequent arousal and irregular heartbeat. The NPSG is very costly procedure. This prevents a large clump of people from the availing the service. But who can afford the service find it tedious and laborious due to long duration of scan. It is also a fact that scarcity of sleep labs and sleep lab experts is detrimental to timely diagnosis and treatment. The treatment of sleep apnea varies with the severity of sleep apnea. For mild OSA, treatment

may consists of refraining from sleeping on ones back, or using decongestant therapy for patients having a problem with nose congestion, or weight reduction for obese patients. For severe OSA, the treatment consists of using nasal continuous positive airway pressure (CPAP), which relives OSA. CPAP uses a small mask held onto the nose by straps, or has soft plastic pillows that are inserted into the nostrils. CPAP is a respiratory ventilator that applies titrated pressure to upper airway sufficient to prevent upper airway collapse. Compressed air generated by the CPAP machine is applied to the patient's nostrils via mask. Typically, CPAP the mask is held on to the face by straps that wrap over head, or has soft plastic pillows that are inserted into the nostrils. Machine can control the pressure of the air entering the nose. The column of pressurized airflow in the patient airway, exert the pressure to keep the upper airway open.

1.1.6 Alternate Means of Diagnosing Sleep Apnea

To overcome the shortfalls of the NPSG, a number of alternate means of detecting sleep apnea have been investigated. Respiration was one of the physiological parameter which was investigated extensively for the detection of sleep apnea.

1.1.6.1 Diagnosing sleep apnea measuring changes in respiration

Sleep apnea is characterized by cyclic repetition of breathing cessation and restoration during sleep. Measuring the respiration during sleep could give qualitative information regarding sleep apnea. Respiration can be measured in a number of ways. One such method is use of thermistor/thermocouple sensor. A thermistor/thermocouple based flow sensor, worn at below the nose and above upper lip was used as means for diagnosing sleep apnea. Change in temperature of air due to breathing pattern changes the resistance of the sensor for thermistor and voltage in the case of thermocouple [20]. The sensor keeps track of decrease or complete cessation of respiration throughout the night. After few hours of recording, they are scored, tallying the number of respiratory events per hour of recording. The position of the sensor should be fixed to get qualitative information. Since, it plays important role in the detection of sleep apnea; it requires continuous monitoring of sensor positioning. Furthermore, a conflict in

results was provided with study of accuracy of the monitor. It was unclear whether performance depends on the signal acquisition system, including the sensor, or on the analytical algorithm [21].

Respiration can also be assessed from the measurements of chest and abdominal movements. Changes in respiration may be measured effectively by using plethysmography. There are number of Plethysmography techniques available for use. However, for diagnostic purpose primarily Elastomeric Plethysmography, Impedance Plethysmography and respiratory inductance Plethysmography are used. In Elastomeric Plethysmography, a belt embedded with piezo-electrical sensor is fastened around chest or abdomen, which experiences a change in tension due chest and abdominal movement. This change is directly converted into voltage. The method is cost effective and easy to use, but suffers from trapping artifact. This results in undermining and/or overrating the actual degree of chest movement. Impedance Plethysmography is based on the principle that the human body presents high impedance to electric current. Two electrodes which carry weak alternating current are connected to the skin to measure impedance. Movement of the cross-section changes the impedance. This result in a non-linear signal, which is useful for qualitative analysis. Since, the method use electric current to detect impedance, one should take into account the frequency range used, to prevent interference with other electric equipment such as defibrillators and pacemaker. The respiratory inductance Plethysmography uses the Faraday's Law and Lenz's Law of electromagnetic to detect respiratory effort. A coiled elastic belt is worn around the chest or abdomen. When an alternating current is passed through the coil, a magnetic field is generated (Faraday's law). The respiratory excursion changes the cross section area of the body, inducing an opposite current that can be measured [22]. This method is proven to be useful in measuring the chest or abdominal movement effectively but suffers from the calibration difficulties where extensive calibration paradigms or two position calibration is required.

To make diagnosis of sleep apnea more viable, cost effective, patient friendly and accurate researchers started looking extensively into other physiological parameters. Electrocardiogram (ECG) is one such physiological marker which is researched extensively as an alternative mean of diagnosing sleep apnea.

1.1.6.2 Diagnosing sleep apnea using ECG

Sleep apnea was found to have a significant impact on the electrical activity of heart [34]. George Moody et al were the first to report the changes in ECG with the changes in respiration. They presented a signal-processing technique which derives respiratory waveforms from ordinary ECGs, permitting reliable detection of respiratory efforts. More than 13 groups came together for an international competition to explore the time and frequency domain analysis of HRV to detect sleep apnea [7]. Sugesseted few possible ways of detecting sleep apnea with reasonable accuracy, but could not reach conclusion at the end. An algorithm was developed for automatic detection of OSA using spectrogram of ECG [4]. For this purpose heart rate (HR), S-pulse amplitude and pulse energy were extracted from ECG. However, due to complexity of patterns and variation among subjects, manual classification of ECG was more accurate than the proposed algorithm. Specific features extracted from ECG which contains frequency information(R-R interval) and morphology(R-wave amplitude) was used to detect sleep disorder breathing (SDB) [8]. This technique uses the frequency domain analysis of the time varying features, where time domain analysis is completely neglected. A bivariate time varying autoregressive model (TVAM) was used to evaluate beat by- beat power spectral densities for both the RR intervals and the QRS complex areas was used to detect sleep apnea with considerable accuracy [24]. However, the algorithm overestimates apnea and is useful only for screening. In other study by Quiceno-Manrique et al, time frequency distribution and dynamic features extracted from ECG were used for the detection Sleep apnea [5]. Most recently a novel method of analyzing time frequency plots generated from heart rate variability was used to detect sleep apnea. This technique used textural features generated from normalized gray-level

co-occurrence matrix (NGLM) of images generated by short time discrete Fourier transform (STDFT) of the HRV [9]. In the above both cases, even though good detection accuracy was achieved, both methods did not considered the amplitude variation in ECG due to sleep apnea. A quadratic discriminant analysis classification system was used to detect sleep apnea in children [6]. The method presented the use of temporal features such as mean and Standard deviation of R-R interval within that segment and first 3 serial correlation coefficients and frequency features for instance power spectral density (PSD) and ECG derived respiration. In other study by AH Khandoker et. al, support vector machines (SVMs) were used for the screening of obstructive sleep apnea Syndrome from ECG recordings [43].

After a thorough literature search it is observed that a limited research has been performed in investigating the combined effect of R-R interval and R peak envelope (envelope of R peaks) for the detection of sleep apnea. Also, it is observed that SVM an attractive machine learning theory is scarcely used in the field of sleep apnea detection. It was deemed to investigate both R-R interval and RPE together for the detection of sleep apnea with SVM as a classifier.

1.2 Hypothesis

In this study, we are investigating the performance of SVM for the detection of sleep apnea using limited features extracted from ECG. For this purpose, we used centroid extracted from the cluster of R-R interval and RPE derived from ECG, as an input to SVM classifier. In this thesis we will discuss a proposed approach to test the above hypothesis.

1.3 Organization of Thesis

The content of this thesis is presented as follows. Chapter 2 presents ECG the algorithm to detect the presence of apnea using ECG. The experimental method in evaluating the performance of the proposed algorithm is presented in Chapter 3. Chapter 4 describes the results obtained from the sleep data. Chapter 5 discusses performance results for the proposed algorithm. Chapter 6 contains the conclusions and future work.

CHAPTER 2

METHODS AND MATERIALS

This chapter discusses in detail about the methods and materials that are used in this project. First section 2.1 of the chapter deals with the physiological effects of sleep apnea and also discusses the brief Introduction to ECG. Section 2.2 describes the data preprocessing and the proposed method to extract ECG features. Section 2.3 details the concept of K-Means clustering. Section 2.4 discusses the mathematical foundations for support vector machines (SVM). Last section discuses the optimization techniques used for the optimization of hyperplane parameters.

2.1 Physiological effects of Sleep Apnea

Sleep is the time for cardiovascular relaxation due to decrease in metabolic rate and sympathetic nervous system activity (SNA) and increase in vagal parasympathetic outflow to the heart. The SNA such as Heart rate, cardiac output and blood pressure decreases. However, OSA counteracts this cardiovascular quiescence during sleep. This mediates blood pressure and heart changes to be intermediary between wakefulness and sleep. Physiological effect of sleep apnea can be summarized as follows.

First, there is initial decrease in the drive to breathe caused by both baroreflex and chemoreflex. As a result of this, upper airway collapses. Also, this leads to reduced activity of the pulmonary stretch receptors which in turn increase the drive to breathe. Moreover, the subsequent hypoxic-hypercapnic state in turn increases the drive to breathe. In addition, the decrease in O_2 concentration in blood due to hypoxic state influences the peripheral chemoreceptors, whereas increase in CO_2 concentration affects central chemoreceptors. Additionally, Chemoreceptor activation also increases sympathetic activity which consecutively increases arterial blood pressure, which then activates the baroreceptors [25, 26]. Increase in blood pressure changes

the intrathoracic pressure there by applying pressure to baroreceptors. This in response induces reflex sympathoexcitation. Finally, these reflexes eventually terminate the apneic event and educe arousal, which is repeated many times during the night.

The ECG is most economical and globally accepted the most accurate method of detecting cardiac arrhythmia. A brief introduction to ECG and the effect of sleep apnea on ECG is provided below.

2.2 Brief Introduction to ECG

Electrocardiography abbreviated as ECG is a non invasive technique to measure electric activity of heart over time. This electrical activity is recorded with the help of ECG electrodes. The electrodes are twelve in number. The connection between two limb (arm or leg) electrodes is called a lead. Einthoven assigned the leads between three limb electrodes standard lead I, lead II and lead III referring to the two arm electrodes and the left leg electrode. The three electrodes when joined form an equilateral triangle where the heart electrically constitutes the null point. The resultant triangle is called Einthoven triangle. The Einthoven's triangle is helpful in determining the electrical axis of the heart. Figure 2.1 illustrates the Einthoven triangle.

A typical ECG waveform recording is graphically represented in Figure 2.2. Each peak and depression in ECG waveform has a particular significance in the heart activity. Electrical signal needed for heart to pump blood to various parts of the body is generated at Sino Arterial (SA) node. The electric impulse generated at SA node travels to both right and left atrium where the impulse trigger both atrium to contract, the P wave of ECG represents the conduction throughout left and right atrium. After, the signal is relayed to Atrioventriclular (AV) node, where the signal is delayed for few moments. This allows both left and right atrium to empty all its content into both ventricles, PR segment in ECG wave represents this delay. The ventricles receive impulse from AV node through purkinje fibers where the fibers stimulate myocardial cells of both ventricles to contract. This is represented by huge QRS complex in ECG wave. The whole phenomenon so far is influenced by sympathetic nervous system. T and U of ECG

represent the ventricular repolarization where the above two events are controlled by parasympathetic nervous system.

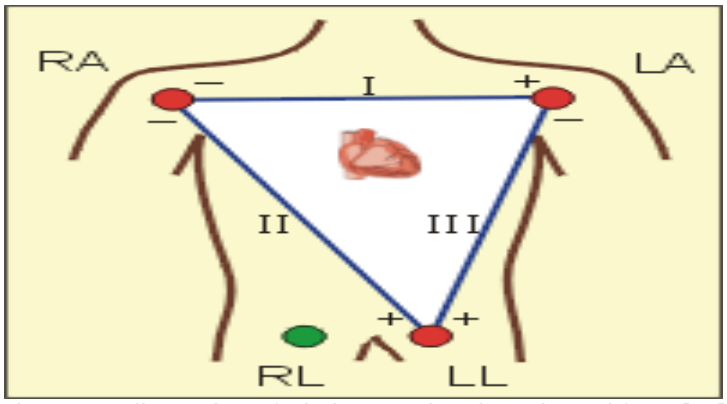


Figure 2.1 Illustration of Einthoven triangle. Adopted from [39]

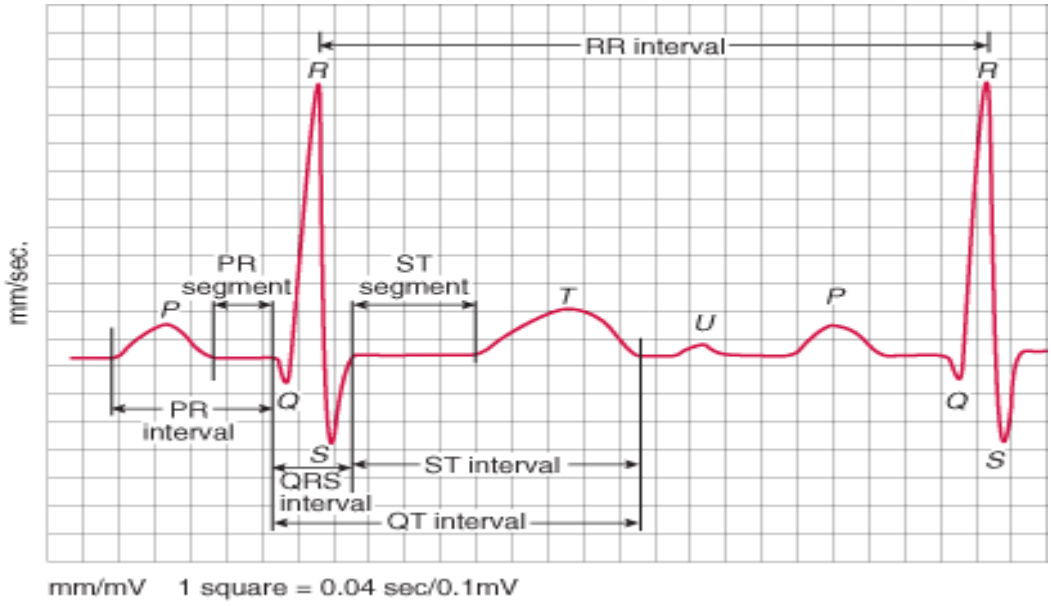


Figure 2.2 Graphical representations of important segments of ECG. Adopted from [27].

2.2.1 Effect of Sleep Apnea on QRS complex

The ECG measured from the chest electrodes are influenced by respiration. The expansion and the contraction of the chest during respiration results in motion of the chest electrodes. Short term thoracic impedance changes are also associated with filling and emptying of the lungs during respiration. These physical influences of respiration result in QRS amplitude variations in the observed ECG [34]. Figure 2.3 depicts a typical change in QRS during sleep apnea.

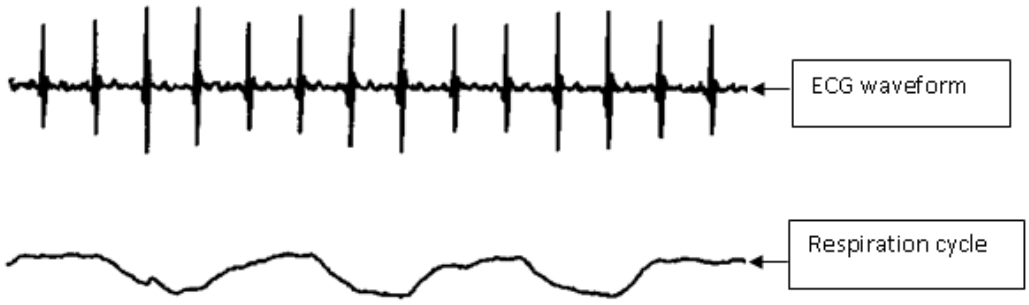


Figure 2.3 Illustrate the change QRS amplitude with respiration. Upper waveform indicates ECG and lower wave form indicates the measure of respiration. Adopted from [34].

When the obstructive sleep apnea cause the respiration to cease, the movement of the ECG electrodes with respect to the heart also ceases. Hence, the modulation of the QRS amplitude due to respiration is altered

2.2.2 ECG Lead System

Twelve lead configuration is the general norm of measuring ECG. A total of 6 chest leads, 3 bipolar limb leads and 3 augmented unipolar leads are used for measuring ECG. Each of these leads provides additional information in comparison with other leads.

Single lead or combination of leads was investigated for detection of sleep apnea using ECG. The use of multi Lead ECG has been shown to detect sleep apnea with acceptable results [35, 36]. But the selection of Lead combinations was not standardized in these studies. Recently, Single lead ECG is investigated extensively as a means of detecting sleep apnea by many researchers [37, 38, 39]. These studies have shown that single-lead ECG is adequate for detection of SDB with reasonable accuracy of detection. In this study, only one lead (lead 1) is used.

R-peak envelope (RPE) extracted from lead I showed highest sensitivity for the detection of OSA [8, 10]. ECG Lead I had a sensitivity of 88.23% of detecting OSA compared to 70.0% of

Lead II and 72.97% of Lead V6 [8]. In the following subsection data processing used for the proposed study is described. The standard NPSG data collection is discussed in next chapter.

2.2.3 Data Preprocessing

Two time series were derived from ECG Lead I. The first time series was envelope of the R waves, referred to as the R-peaks. The second time series, called the R-R interval is derived from the detected R peaks. The entire night's lead I ECG data was divided into epochs of 900 seconds. The baseline wander removal, detection of R- Peak and subsequent detection of R-R interval is summarized in the following subsections.

2.2.3.1 Baseline Wander Removal

Patient movement, dirty electrodes, and loose electrodes may cause baseline wandering, which changes the position of isoelectric line. Respiration is also thought to be a cause of baseline wandering. A high pas, linear-phase FIR filter with a cut-off frequency of 0.8 Hz and length 200 was used to remove baseline wander. Bidirectional filtering was employed to null group delay of the filter. Figure 2.4 shows a sample ECG of 30 seconds before and filtering [8].

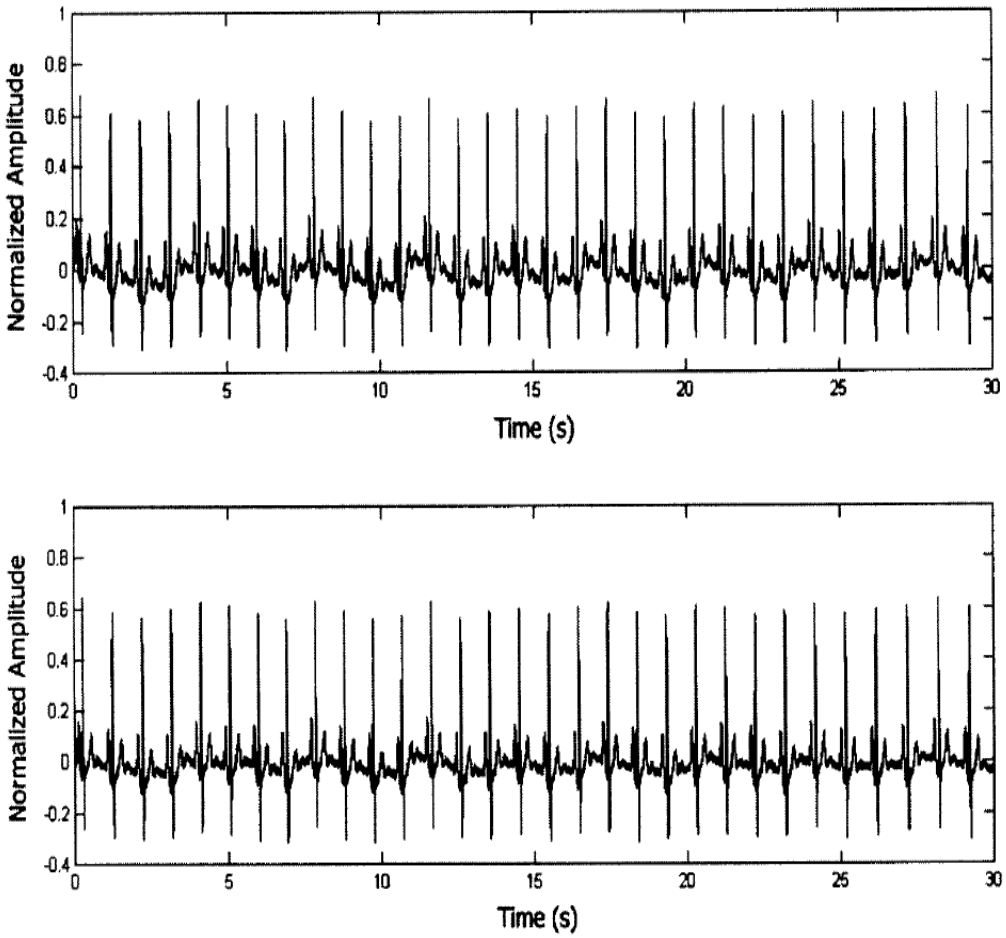


Figure 2.4 ECG before filtering (Top) and after filtering (bottom). Adopted from [8]

2.2.3.2 R Peak Detection

The entire night's ECG lead I data was divided into epochs of 900 seconds duration. A Hilbert transform based algorithm suggested by Benitez et al was used to detect R peaks [8, 10]. In the RPE time series, a single value was obtained for each beat in the epoch. Every value has a time reference of the corresponding beat at which it was extracted. This results in unevenly sampled data. The RPE was evenly re-sampled at 10 Hz using cubic spline interpolation. The 900-second outlier free RPE time series is then evenly resampled at 1 Hz to produce a 900-points time series. A mean error detection of 1% was for a total of 1.78 million beats detected by the algorithm. Testing of the R peak detection algorithm carried out by manual verification of

peaks detected from MIT-BIH and also with the data collected for the study. False detection was corrected. In a clip, if more than 10% of the beats were premature or detection error percentage was more than 15% the clip was rejected.

2.2.3.3 Computation of R-R Interval

R-R interval was derived from R peaks as a discrete event time series. The outliers were removed using a heuristic method based on the mean and standard deviation of R-R interval of entire epoch. The R-R interval greater than three times the standard deviation above mean are considered as outliers. These outliers were removed from the epoch, if they are at least 0.5 seconds greater than prior or subsequent R-R interval [8]. The entire 15 minute outlier free R-R interval was evenly sampled at 10Hz using cubic spline interpolation, using MATLAB function spline. Figure 2.4 gives a graphical representation of R-R interval when plotted against time.

For the proposed study we extracted 191 clips, 86 normal (NOR) and 105 apneic (OSA) clips, which will be discussed in detail in next chapter. Where two features RPE and R-R interval was extracted from each clip. Both the features are combined in a two dimensional space with R-R interval in abscissa and RPE in ordinate to form a cluster. A single centroid was extracted from this cluster using K-means function in matlab. The theoretical approach for calculation of centroid is provided below with examples.

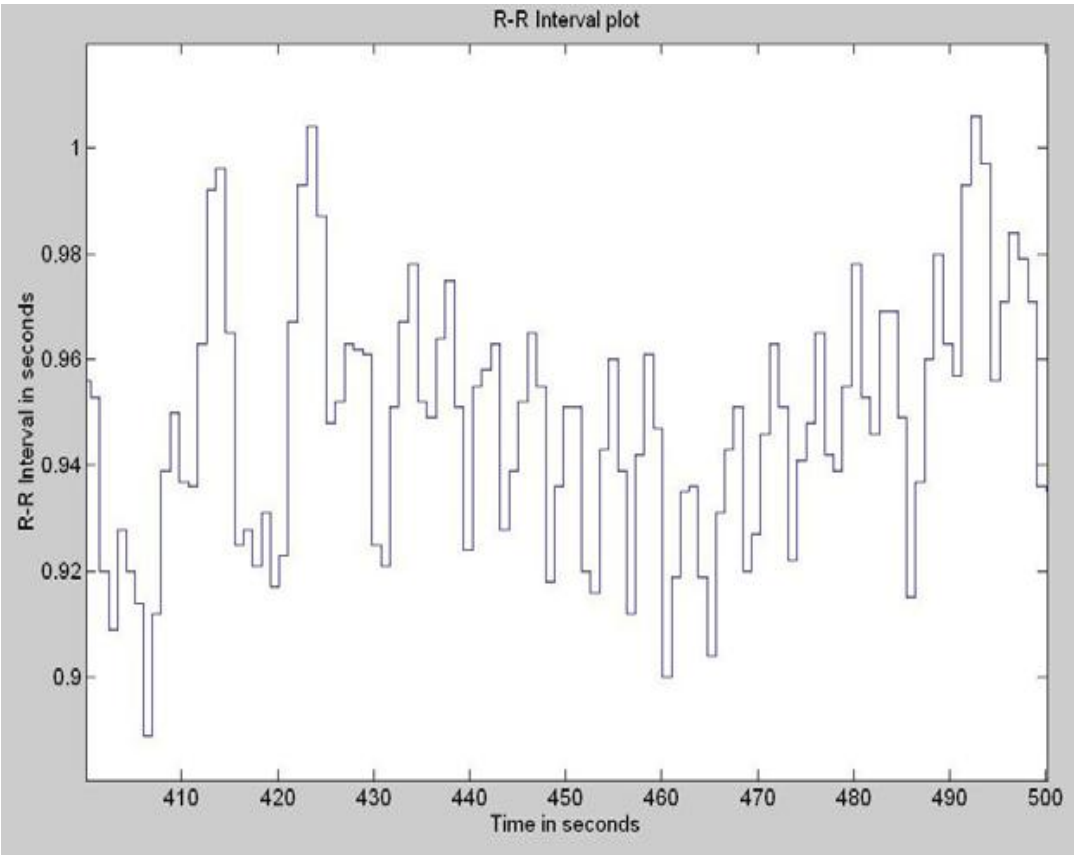


Figure 2.5 graphical illustration of R-R interval after resampling. Adopted from [9]

2.3 K-Means Clustering

K-means is a simple cluster analysis technique, where the data is clustered into K-clusters. The data in a cluster have the data points which are closest to mean or median of the cluster. The data is classified into K clusters and each cluster has its distinct centroid farthest away from other clusters and closest to the cluster it belongs.

K-means algorithm partitions the data set with *n* observations into k-clusters, where each observation is assigned to a cluster with the nearest mean or median. The algorithm follows simple and easy way to classify the observations through K a priori fixed clusters. The idea is to find a centroid for each cluster in a sequential approach, where outcome of the result solely depends on the initial selection of centroids. After initial selection, each observation of the data set should be assigned to a centroid. Where each observation assigned is closest to the centroid. When the assignment is complete all the observations should be grouped into

respective clusters. At the end of the above step, new k-centroids are calculated as barycenters of the clusters generated. After calculating new centroids, the clustering is done on the same data set with nearest new centroid. The centroids change their locations at the end of each clustering step. The above steps are repeated several times, until the centroid stops moving further from its previous location or the change in the location of centroid are within acceptable range. As a result all the observations in the data set are clustered into K-clusters with each cluster having a distinct centroid [28].

The algorithm defines an objective function J for the above mentioned steps and tries to solve it sequentially until the objective function is minimize. The objective function is squared error function. The objective function for K-means clustering is given by the following function.

$$J = \sum_{j=1}^{K} \sum_{i=1}^{n} \left\| x_i^{(j)} - C_j \right\|^2 \quad (2.3.1)$$

Where $\left\| x_i^{(j)} - C_j \right\|^2$ is the distance measure between an observation $x_i^{(j)}$ and centroid C_j. J is an indicator of the distance of the *n* data points from their respective cluster centers. Optimizing function J results in optimized centroid values. The optimization reaches a final solution when the distance between data points in a cluster to its centroid is bare minimum. The optimization is said to be converged when it reaches final solution to the optimization problem. Figure 2.6 and 2.7 illustrates the movement of centroids from their initial guess till the centroids stops moving further.

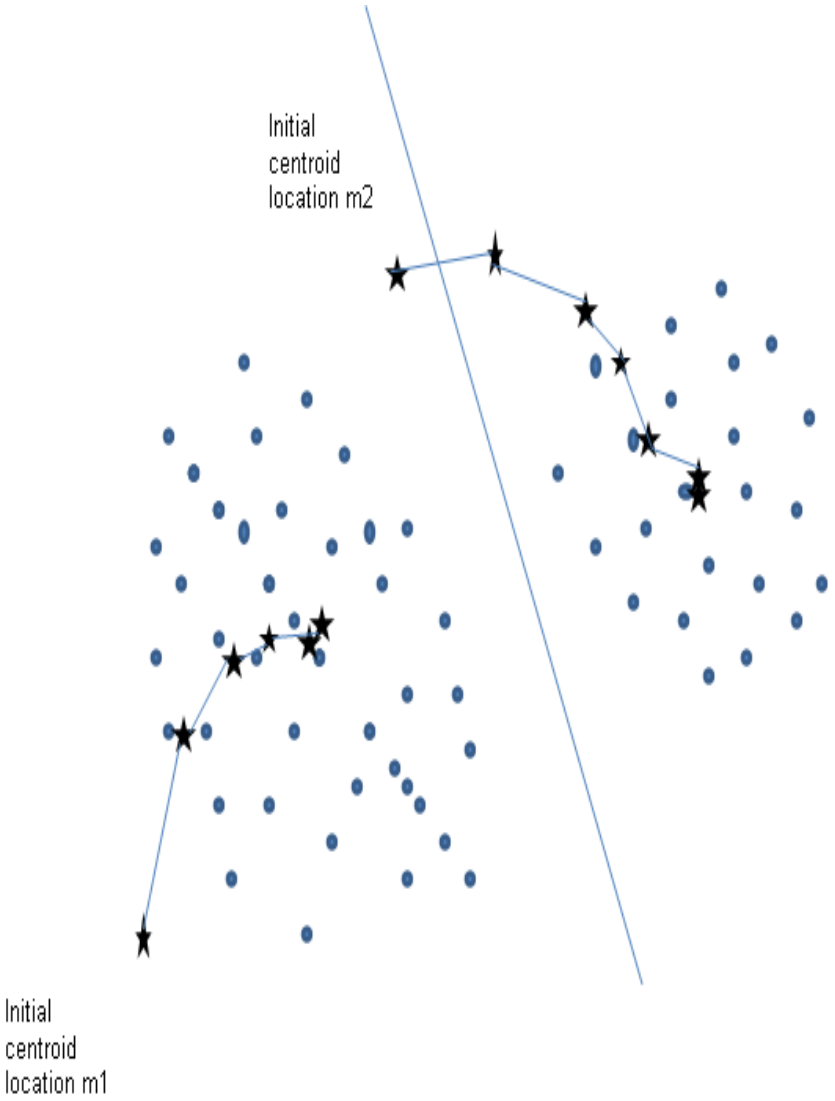


Figure 2.6 Trace of centroid calculation for two clusters

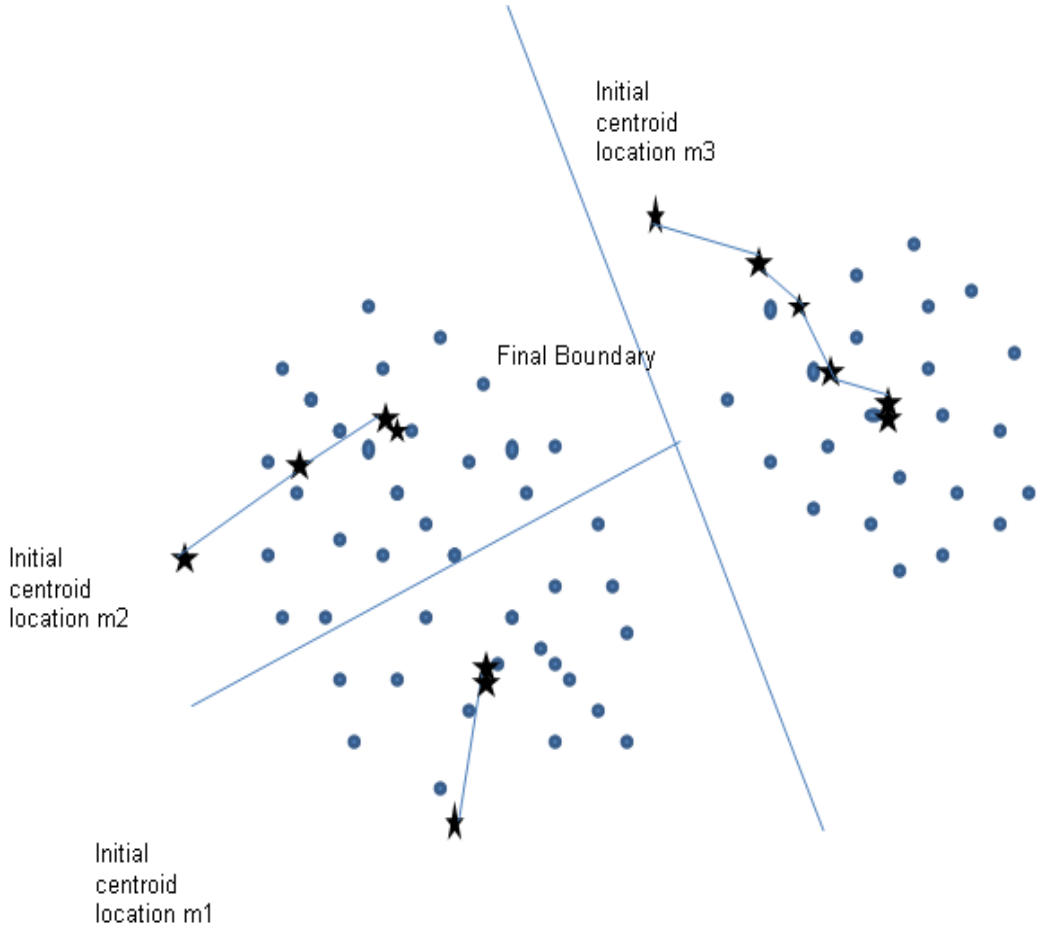


Figure 2.7 Trace of centroid calculation for three clusters

The algorithm of K-Means clustering converges very fast in practice, though sometimes takes exponential time to converge. The K-means algorithm is simple, scalable and efficient. Sometimes the algorithm has a local minima problem. The local minima problem can be avoided with good initial guess for centroid. The centroid selection is carried out randomly, but there are better and efficient heuristic approaches to select initial centroids. The selection of number of clusters also plays an important role in achieving an optimal solution.

2.3.1 Metrics to Calculate Distance

Calculation of distance between centroid and each data point in data set plays an important role in the K-Means clustering technique. There are number of distance metrics available for use. The distance metrics used are Euclidean distance, Manhattan distance, maximum norm,

Mahalanobis distance, cosine angle between vectors, Hamming distance. Minkowski Metric is used for higher dimensional data. The metric is given by

$$d_p\,(x_i, x_j\,) = \left(\sum_{k=1}^{d} \left|x_{i,k} - x_{j,k}\right|^p\right)^{\frac{1}{p}} \qquad (2.3.1.1)$$

Where d is the dimensionality of the data. The Euclidean and Manhattan distance are special cases of Minkowski metric with p=2 and p=1 respectively.

The distance metric gives the similarity of one group with respect to other group. The most popular distance metrics are Euclidean distance or 2-norm distance and Manhattan distance or 1-norm distance. In this study we use Manhattan distance for the calculation of distance.

2.3.1.1 Manhattan distance

For the given two points in space with co-ordinates (x, y) and (a, b), the Manhattan distance d between two points is given by

Manhattan distance: $d = |x - a| + |y - b|$ (2.3.1.2)

When Euclidean distance is used to obtain centroids, the resulting cluster centroids are median of all the data points in a cluster [29]. Where the median of cluster minimizes the Manhattan distance (also known as city block) given by equation

$$d\,(x_i, x_j\,) = \left(\sum_{k=1}^{d} \left|x_{i,k} - x_{j,k}\right|\right) \qquad (2.3.1.3)$$

The centroids extracted using K-means algorithm is fed to a classifier to separate normal (NOR) clips from OSA clips. The selection of NOR and OSA will be discussed later in next chapter. In this study a support vector machine (SVM) based binary classifier is used for the classification purpose. The mathematical foundations for the SVM are described in next section.

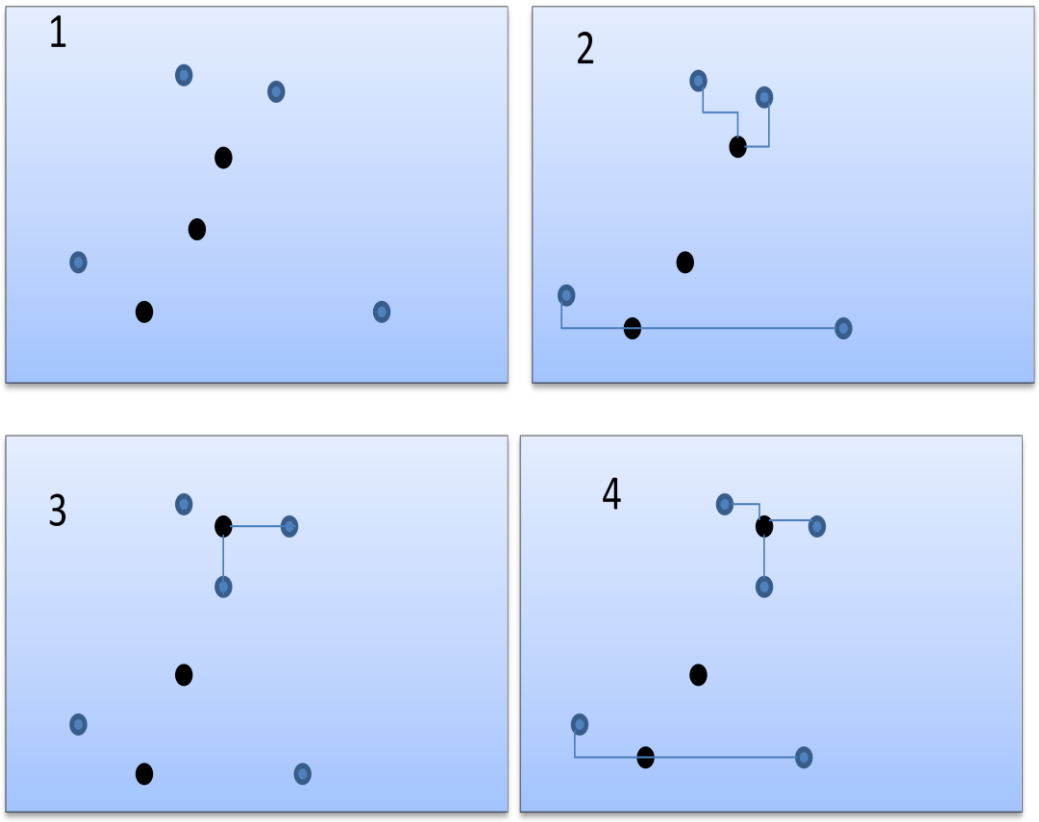


Figure 2.8 K-Means Clustering with Manhattan distance. Black circles represent the centroids.

2.3.2 Flow chart for K- Mean's Algorithm

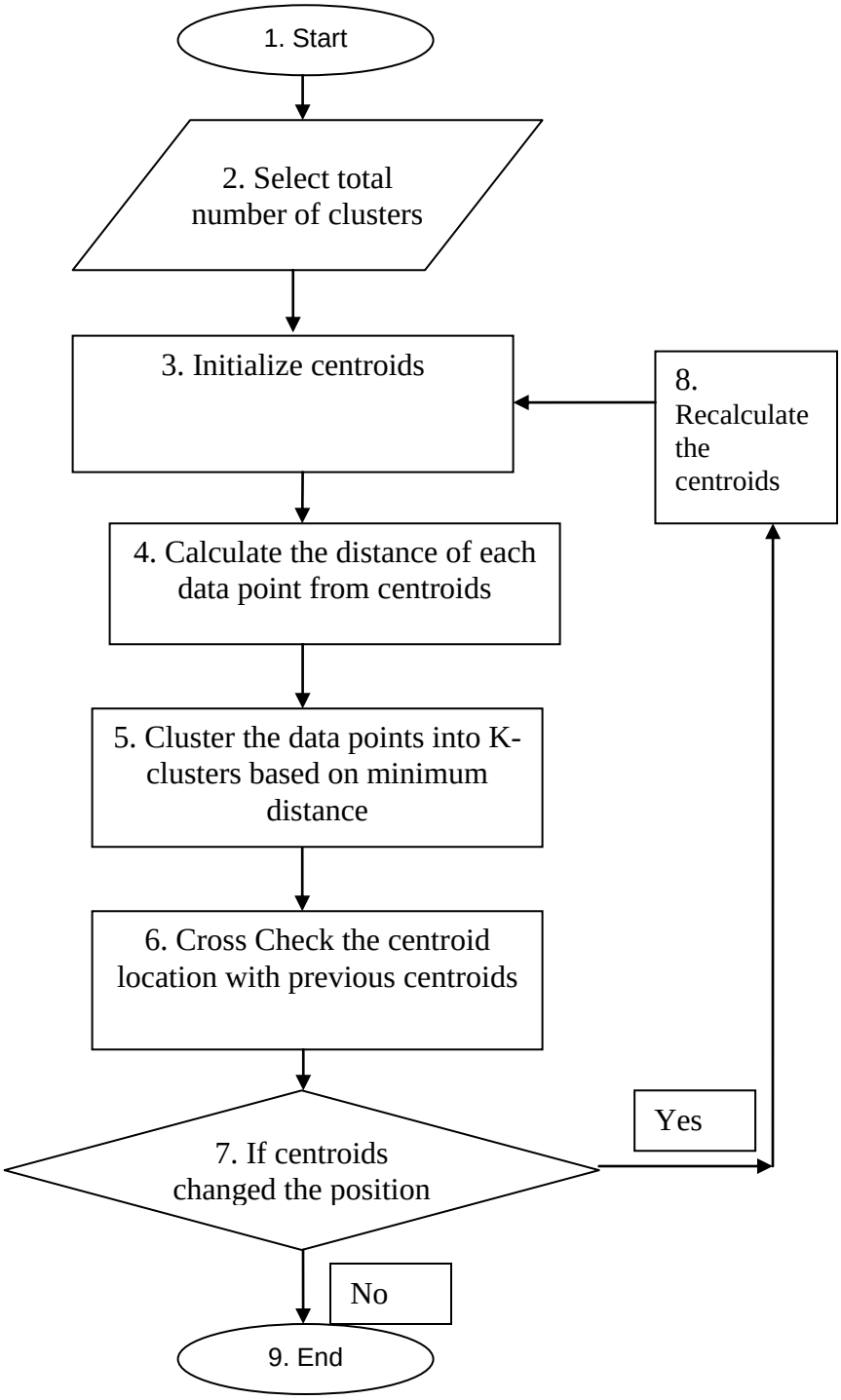


Figure 2.9 Flow chart for the K-means algorithm.

2.4 Support Vector Machine

The concept of support vector machine (SVM) was introduced by Vladimir Vapnik in mid 1990's while he was working at Bell laboratories. Generalization and optimization theory in SVM has firm theoretical and mathematical foundation. SVM finds its applications in wide variety of fields and has successfully been applied in many real world problems such as pattern recognition, text classification, web page classification, weather prediction, and intrusion detection.

Support Vector Machine is considered a very powerful tool due to its high generalization capability, having a rather simple geometrical interpretation, a sparse solution, and the ability to process high dimensional data. SVM generated results are highly stable and reproducible. However, they suffer from speed and size of training, especially with large non linear data [28]. SVM belongs to the class of binary classifier and the idea is extended to multi-class through the use of several binary classifiers.

In the following sections, the mathematical foundations for binary classification using SVM are expounded in detail.

2.4.1 Binary Classification

Binary classification is the task of classifying data into two classes based on the weight associated with certain property of the data. Classical example of Binary classification is simple" Yes" or" No" answer to questions, face detection, and disease detection.

The detailed mathematical foundation of support vector machines and underlying Vapnik-Chervonenkis dimension (VC Dimension) is covered in literature of statistical learning theory (Vapnik 1998) [29]. In the following section mathematical background of SVM for Linearly separable and non linearly separable classes are discussed in brief. The beauty of SVM lies in geometrical interpretation of its mathematical formulation where one can perceive mathematical formulations with geometrical analogies.

2.4.2 Linearly Separable Case

In the linearly separable case, the data of two classes do not overlap and are easily separated by single decision surface. The Figure 2.10 gives an example of two-dimensional linearly separable data. The lines which separate the data are referred as hyperplanes. An infinite number of hyperplanes can be drawn between two classes which can separate the data precisely, but the aim of any classification algorithm is to find an optimal hyper plane which gives good generalization of results to all members of sampled population. Vapnik proposed in his theory that the hyper plane which has a large margin, the separation gap between two classes yields better results. In this case the hyperplane which evenly divides the margin between the two classes is the optimal hyperplane, as shown in Figure 2.11.

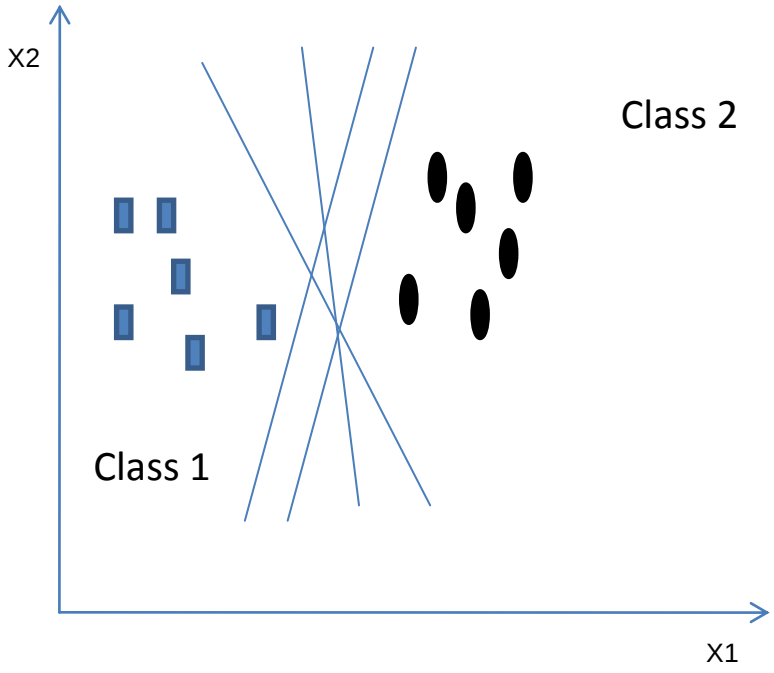


Figure 2.10 hyperplanes that can be drawn for the classification

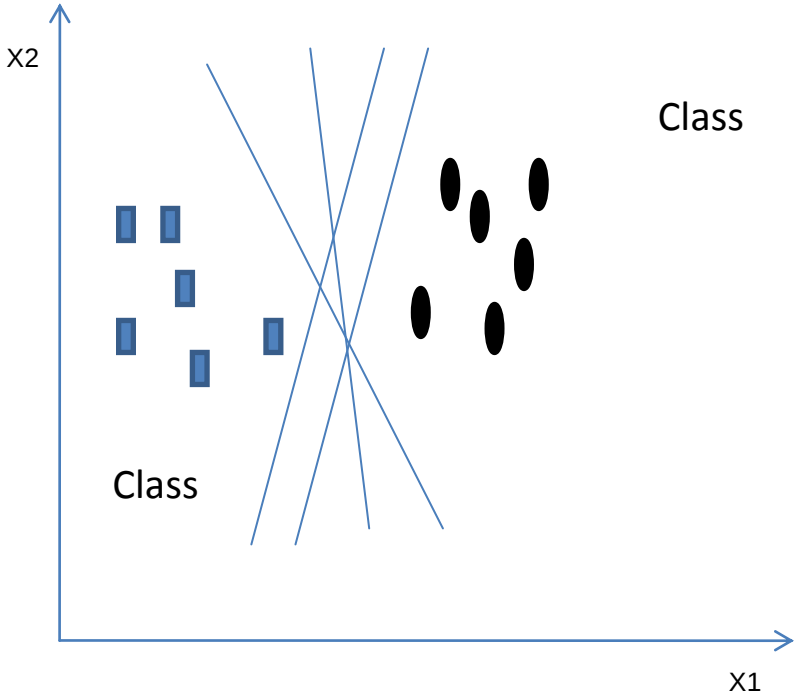


Figure 2.11 Optimum hyperplane with maximum margin.

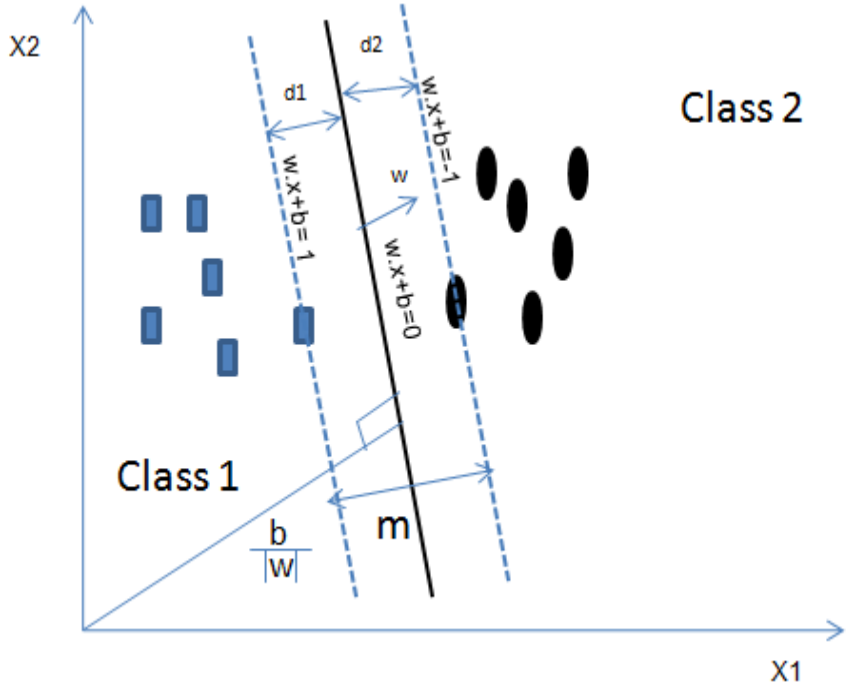


Figure 2.12 linear separating hyperplane for separable case.

The training example in the above Figure 2.12 is of the form: $\{x_i, y_i\}$ *i= 1, 2, 3….p* and $x_i \in R$; $y_i \in \{1, 1\}$, we call $\{x\}$ the input vector which contains two features x_1 and x_2 and $\{y\}$ is the targeted output or labels.

As shown in Figure 2.12, the equation of separating hyperplane is given by

$$w.x + b = 0 \quad (2.4.2.1)$$

where w, b is normal and bias to hyperplane respectively.

We can define two hyperplanes parallel to the separating hyperplanes defined by (2.4.2.1). They represent that planes that cut through the closest data points on either side of the separating hyperplane. The equation of the two parallel hyperplanes is given by

$$w.x + b = 1 \quad (2.4.2.2)$$

$$w.x + b = -1 \quad (2.4.2.3)$$

The hyperplanes given by (2.4.2.2) and (2.4.2.3) are parallel, at a maximum distance from each other and still separate data points. The distance between these two hyperplanes is margin of the hyperplane. Our aim is to maximize this margin.

The perpendicular distance from a point (m, n) to line represented by equation Ax+By+C=0 is given by

$$D = \frac{Am + Bn + C}{\sqrt{A^2 + B^2}}$$

The perpendicular distance from origin (0, 0) to line represented by (2.4.2.1) is given by

$$d = \frac{b}{\|w\|} \quad (2.4.2.4)$$

where $\|w\|$ is defined as $\sqrt{w^T w}$

The distance from origin to (2.4.2.2) and (2.4.2.3) is respectively given by

$$d_- = \frac{\|1-b\|}{\|w\|} \quad (2.4.2.5)$$

$$d_+ = \frac{\|-1-b\|}{\|w\|} \quad (2.4.2.6)$$

The distance d1 between the hyperplane and the parallel hyperplane passing through the data points is given by

$$d_2 = \frac{\|1+b\|-\|b\|}{\|w\|} \quad (2.4.2.7)$$

$$d_2 = \frac{1}{\|w\|} \quad (2.4.2.8)$$

By symmetry $d_2 = d_1 = \frac{1}{\|w\|}$

Margin m of the hyperplane is given by

$$m = d_1 + d_2 \quad (2.4.2.9)$$

$$m = \frac{2}{\|w\|} \quad (2.4.10)$$

From figure 2.12 we can write the following constraints for the optimization problems

$$\mathrm{w}.x_i + \mathrm{b} \geq 1 \text{ if } y_i = +1 \quad (2.4.2.11)$$

$$\mathrm{w}.x_i + \mathrm{b} \leq -1 \text{ if } y_i = -1 \quad (2.4.2.12)$$

Where y_i is t he target output of the input x_i

Combining equation (2.4.2.11) and (2.4.2.12) we can get one equation for above constraints

$$y_i(\mathrm{w}.x_i + \mathrm{b}) \geq 1 \quad (2.4.2.13)$$

We want to maximize equation (2.4.2.10). The numerator in equation (2.4.2.10) is a constant. In order to maximize (2.4.2.10) we have to minimize the denominator. The final optimization problem is given by

Minimize $\frac{1}{2}\ \|w\|^2$ (2.4.2.14)

Subject to $y_i(\mathrm{w}.x_i + \mathrm{b}) \geq 1$, where *i*=1, 2, 3, 4……p (2.4.2.15)

Where p is the number of data points in the cluster.

The above two equations represent the primal form of our convex optimization problem. Our goal is to find w, b of the hyperplane so that the following conditions are satisfied

Minimize $\frac{1}{2}\ \|w\|^2$ (2.4.2.16)

Subject to $y_i(\mathrm{w}.x_i + \mathrm{b}) - 1 \geq 0$, where *i*=1, 2, 3, 4…p (2.4.2.17)

This is a non linear quadratic optimization problem subject to linear inequality constraints. The above two equations should satisfy the Karush-Kuhn-Tucker (KKT) conditions for solution to be optimum. The KKT conditions are necessary and sufficient conditions for optimal solution for the above describe optimization problem. The KKT conditions are described in detail later in the chapter. The KKT conditions that the minimizer has to satisfy are

$$\frac{d}{dw}\mathcal{L}(w, b, \lambda) = 0 \quad (2.4.2.18)$$

$$\frac{d}{db}\mathcal{L}(w, b, \lambda) = 0 \quad (2.4.2.19)$$

$\lambda_i \geq 0$, *i*=1, 2… p (2.4.2.20)

$\lambda_i[(y_i(\mathrm{w}.x_i + \mathrm{b}) - 1)] = 0$ *i*=1, 2… p (2.4.2.21)

Where λ is the vector of the Lagrange multipliers, λ_i and $\mathcal{L}(w, b, \lambda)$ is a Lagrange function defined as

$$\mathcal{L}(w, b, \lambda) = \frac{1}{2}w^T w - \sum_{i=1}^{N}\lambda_i[(y_i(\mathrm{w}.x_i + \mathrm{b}) - 1)] \quad (2.4.2.22)$$

Combining (2.4.2.18) and (2.4.2.22) we get

$$\frac{d}{dw}\left(\frac{1}{2}\ w^T w - \sum_{i=1}^{N}\lambda_i[(y_i(\mathrm{w}.x_i + \mathrm{b}) - 1)]\right) = 0 \quad (2.4.2.23)$$

Taking derivative with respect to w gives,

$$w - \sum_{i=1}^{N}\lambda_i\ y_i x_i = 0 \quad (2.4.2.24)$$

$$w = \sum_{i=1}^{N}\lambda_i\ y_i x_i \quad (2.4.2.25)$$

Taking derivative with respect to b we get,

$$\frac{d}{db}\left(\frac{1}{2}\ w^T w - \sum_{i=1}^{N}\lambda_i[(y_i(\mathrm{w}.x_i + \mathrm{b}) - 1)]\right) = 0 \quad (2.4.2.26)$$

$$\sum_{i=1}^{N}\lambda_i\ y_i = 0 \quad (2.4.2.27)$$

The above minimization problem belongs to convex programming optimization. Since the optimization problem is a convex optimization problem and the constraints associated with it are linear constraints. We can use the concept of Lagrange duality. The primal problem can be transformed into dual form using the Wolfe dual representation given by

Maximize $\mathcal{L}(w, b, \lambda)$ (2.4.2.28)

Subject to $\sum_{i=1}^{N} \lambda_i \ y_i = 0$ (2.4.2.29)

$\lambda \geq 0$ (2.4.2.30)

Substituting (2.4.2.25) and (2.4.2.27) in (2.4.2.22) we get

$\max_{\lambda}(-\sum_{i=1}^{N} \lambda_i\left[\left(y_i\left(\sum_{j=1}^{N} \lambda_j \ y_j x_j . x_i + \mathrm{b}\right) - 1\right)\right])$ (2.4.2.31)

Subject to $\sum_{i=1}^{N} \lambda_i \ y_i = 0$ (2.4.2.32)

$\lambda \geq 0$ (2.4.2.33)

When we solve (2.4.2.31) we get,

$\max_{\lambda} (\sum_{i=1}^{N} \lambda_i - \frac{1}{2}\sum_{i,j}^{N} \lambda_i \lambda_j y_i y_j x_i^T x_j)$ (2.4.2.34)

Subject to $\sum_{i=1}^{N} \lambda_i \ y_i = 0$ (2.4.2.35)

$\lambda \geq 0$ (2.4.2.36)

The above problem can also be written as

$\min_{\lambda} (\frac{1}{2}\sum_{i,j}^{N} \lambda_i \lambda_j y_i y_j x_i^T x_j - \sum_{i=1}^{N} \lambda_i)$ (2.4.2.37)

Subject to $\sum_{i=1}^{N} \lambda_i \ y_i = 0$ (2.4.2.38)

$\lambda \geq 0$ (2.4.2.39)

When we solve the above quadratic programming we get Lagrange multipliers λ_i, the Lagrange multipliers which satisfies the KKT conditions are optimum Lagrange Multipliers. The data points associated with Lagrange multipliers greater than zero are known as support vectors. Later in

the chapter it is shown that these support vectors are enough to define a hyperplane. *b* is calculated indirectly, derivation of bias term *b* will be described in detail later in the chapter.

Now we know *w, b* parameters of the hyperplane. The unseen data can know be classified using the following equation

$$\mathcal{F}(x) = sign(w.x + b) \quad (2.4.2.40)$$

where $sign(w.x + b) = \begin{cases} -1, & (w.x + b) < 0 \\ 1, & (w.x + b) > 0 \end{cases}$

Where x is the data to be classified, based on the output of the $\mathcal{F}(x)$ the data is classified into two groups.

The concept of soft margin was introduced to deal with the cases where classification is not possible without errors using single hyperplane. The concept of soft margin is explained in appendix B. kernel trick and mathematical foundation of non-linearly separable data is discussed in next section.

2.4.3 Non Linearly Separable Case

A typical non-linearly separable data is shown in figure 2.13. A non linearly separable data cannot be precisely classified into two classes using a single separating hyperplane. As shown in Figure 2.13, the data cannot be classified correctly using a single hyperplane. This does not mean that the data cannot be separated exactly into groups; polynomial curves and circles can be used to classify. However, to find the optimal curve to fit the data is difficult. For this purpose, the data is first transferred into a higher dimensional space. This transformation makes the data linearly separable in higher dimensional space. A single hyperplane can be drawn to this data in higher dimension space which precisely classify the data into two groups shown in figure 2.14.

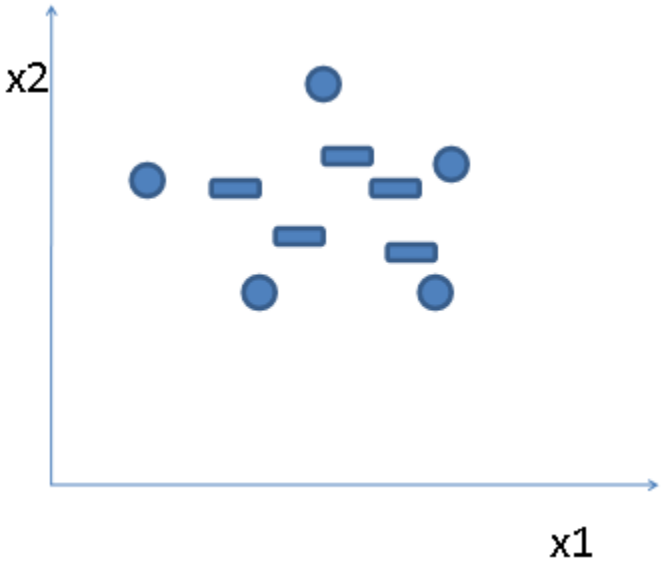


Figure 2.13 Non linearly separable data

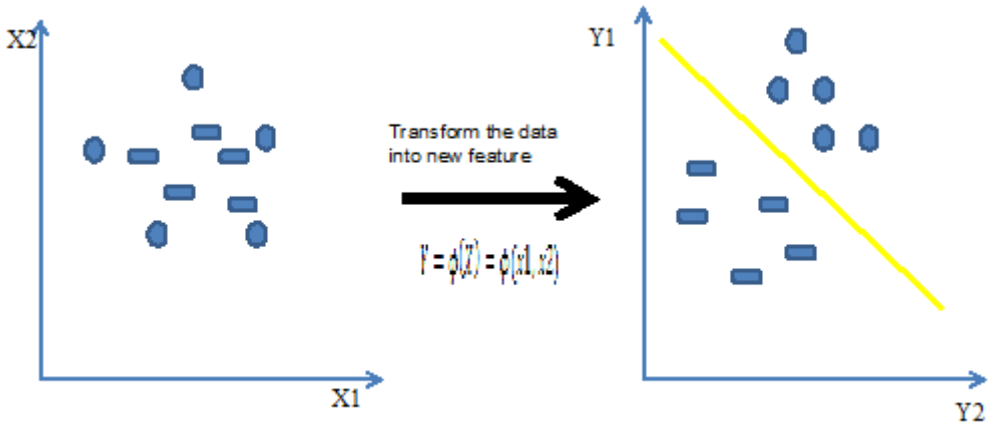


Figure 2.14 Transformation of Non linearly separable data into inner product space.

The transformation into feature space is carried out using a mapping function $Y = \phi(X)$ where the input n dimensional data is mapped into feature space of n' dimensions where n'>n. The mapping function maps the input data into a considerably high dimensional space, this makes data linearly separable eventually. The final solution of quadratic optimization problem depends on the calculation of dot product of the input data points as illustrated in equation 2.4.2.37.The complexity of keeping track of mapping function and calculating the dot product of the data points imposes an additional burden on computation. Storing the high dimensional data requires a lot of memory. To avoid the problems associated with mapping data into high dimensional

space and keeping track of the dot product of data points in high dimensional data, SVM makes use of well established Kernel trick. The use of Kernel trick does not have any impact on the algorithm; it simply replaces the dot product of two vectors with Kernels.

2.4.3.1 Kernel Trick

The Kernel trick pertains to mapping data in input space into much higher dimensional inner product space, without actual knowledge of the mapping function. The data is transferred into higher dimensional space with the hope that the data in the higher dimensional space achieves linear structure. Finding a linear hyperplane in the higher dimensional space to the data is same as finding a polynomial curve which classifies the data in the input space.

The Kernel function used in kernel trick is given by

$$K(x_i, x_j) = < \phi(x_i), \phi(x_j) >$$

Where ϕ is any transformation function which maps the data in input space into higher dimensional space, $< \phi(x_i), \phi(x_j) >$ represents the inner product of the two vectors. The Kernel functions do not require the prior knowledge of the transformation function but should satisfy Mercer's condition. The Mercer's condition is given by

$$\sum_{i=1}^{n} \sum_{j=1}^{n} K(x_i, x_j) a_i a_j \geq 0. \qquad (2.4.3.1.1)$$

Where $a_i \, \epsilon R$ and $a_j \, \epsilon R$

The equation (2.4.3.1.1) represents that, for finite sequence of $x_1, x_2, x_3 \ldots \ldots \ldots x_p$, and of all possible real number values for a_i, a_j, should always result in a positive value. This kernel function which satisfies the above equation is called positive semi definite function. There are number of Kernels which are used in SVM. In this thesis we use polynomial Kernel.

$K = (x_i . x_j + 1)^n$ where i, j=1, 2, 3…….p.

where n is the degree of the polynomial which is characteristic to a hyperplane.

Let us again consider the primal form of objective function defined for soft margin given appendix B. As shown in Figure B.1 the slack variable as assigned to data points which are misclassified or within the margin.

Minimize $\frac{1}{2}\ \|w\|^2 + C\sum_{i=1}^{N}\xi_i$ (*2.4.3.*1)

Subject to $y_i(\mathrm{w}.\phi(x_i) + \mathrm{b}) \geq 1 - \xi_i$, where *i*=1, 2, 3, 4……*N* (*2.4.3.*2)

$\xi_i \geq 0$ where *i*=1, 2, 3, 4……*N* (*2.4.3.*3)

Where ϕ is a mapping functioning transforming the input data into higher dimensional space, C is a hyperplane parameter known as regularization parameter and ξ is a slack variable which is discussed in detail in appendix B.

The Lagrange representation is given by

$\mathcal{L}(w, b, \lambda, \mu) = \frac{1}{2}\ \|w\|^2 + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\mu_i\xi_i - \sum_{i=1}^{N}\lambda_i\,[y_i(w\phi(x_i) + b) - 1 + \xi_i]$ (*2.4.3.*4)

Where μ is the Lagrange multiplier for slack variable ξ_i.

The corresponding KKT conditions are given by

$\frac{d}{dw}\mathcal{L}(w, b, \lambda, \xi_i) = 0$ (*2.4.3.*5)

$\frac{d}{db}\mathcal{L}(w, b, \lambda, \xi_i) = 0$ (*2.4.3.*6)

$\frac{d}{d\xi_i}\mathcal{L}(w, b, \lambda, \xi_i) = 0$ (*2.4.3.*7)

$\lambda_i \geq 0$, *i*=1, 2… N (*2.4.3.*8)

$\mu_i \geq 0$, *i*=1, 2… N (*2.4.3.*9)

$\xi_i\,\mu_i \geq 0$, *i*=1, 2… N (*2.4.3.*10)

$\lambda_i[(y_i(\mathrm{w}.\phi(x_i) + \mathrm{b}) - 1 + \xi_i)] = 0$ (*2.4.3.*11)

Taking derivatives with respect to w, b and ξ_i

$w - \sum_{i=1}^{N}\lambda_i\ y_i\phi(x_i) = 0$ (*2.4.3.*12)

$$w = \sum_{i=1}^{N} \lambda_i \; y_i \phi(x_i) \quad (2.4.3.13)$$

$$\sum_{i=1}^{N} \lambda_i \; y_i = 0 \quad (2.4.3.14)$$

$$C - \lambda_i \; - \mu_i = 0 \quad (2.4.3.15)$$

The associate Wolfe dual representation is given by

Maximize $\mathcal{L}(w, b, \lambda, \mu)$ (2.4.3.16)

Subject to $w = \sum_{i=1}^{N} \lambda_i \; y_i \phi(x_i)$ (2.4.3.17)

$$\sum_{i=1}^{N} \lambda_i \; y_i = 0 \quad (2.4.3.18)$$

$$C - \lambda_i \; - \mu_i = 0 \qquad i=1, 2\ldots N \quad (2.4.3.19)$$

$$\lambda_i \geq 0, \qquad i=1, 2\ldots N \quad (2.4.3.20)$$

$$\mu_i \geq 0, \qquad i=1, 2\ldots N \quad (2.4.3.21)$$

Substituting the above equality constraints into Lagrangian dual we get,

$$\max_{\lambda}\left(\sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i,j}^{N} \lambda_i \lambda_j y_i y_j \; \phi(x_i)^T \phi(x_j)\right) \quad (2.4.3.22)$$

Subject to $0 \leq \lambda_i \leq C$ $i=1, 2\ldots N$ (2.4.3.23)

$$\sum_{i=1}^{N} \lambda_i \; y_i = 0 \quad (2.4.3.24)$$

The dot product $\phi(x_i)^T . \phi(x_j)$ is replaced by kernel function

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_j) \quad (2.4.3.25)$$

Now the dual appears as

$$\max_{\lambda}\left(\sum_{i=1}^{N} \lambda_i - \sum_{i,j}^{N} \lambda_i \lambda_j y_i y_j \; K(x_i, x_j)\right) \quad (2.4.3.26)$$

Subject to $0 < \lambda_i < C$ $i=1, 2\ldots N$ (2.4.3.27)

$$\sum_{i=1}^{N} \lambda_i \; y_i = 0 \quad (2.4.3.28)$$

This is a typical quadratic optimization problem which can be solved using many optimization routine available freely. The optimization routine used for our study is QPAS developed in matlab [48]. The dual representation in this case is same as that of linear separable case with small change in inequality constraints. The parameters C penalize the slack variable which is discussed in appendix B, where C controls the size of the margin. The importance of C is discussed in next section.

2.4.4 Regularization Parameter C

The parameter C of the hyperplane is known as regularization parameter. Selection of C plays an important role in the generalization of results to all members of sampled population. The Lagrange multipliers of the points either within margin or on opposite side of the classifier are equal to maximum allowable value of C. These points have highest significance in the final solution of w. The slack variable ξ_i and their Lagrange multipliers μ_i do not influence the solution, but indirectly reflect through C. The Parameter C has an important role to play in the size of margin. This can be easily expounded through equation (1) of soft margin discussed in appendix B. The second term in the equation (1) in appendix B has a momentous influence on the cost function. The value C controls the trade-off between width of the margin and the number of allowable misclassified data. Hence, one should be careful while selecting a value for C. For example, if a higher C value is selected, this imposes a high penalty on the misclassified data points and subsequently reduces the width of the margin and results in over fitting the data. If the value of C is small, this may lead to under fitting the data. The parameter is also known as the trade off parameter, as it controls the tradeoff between errors of SVM on training data and margin maximization. Figure 2.11 and 2.12 clearly shows the effect of C on width of the margin. The selection of the parameter C is not a mathematical but instead operator selected where the selection is based on the data.

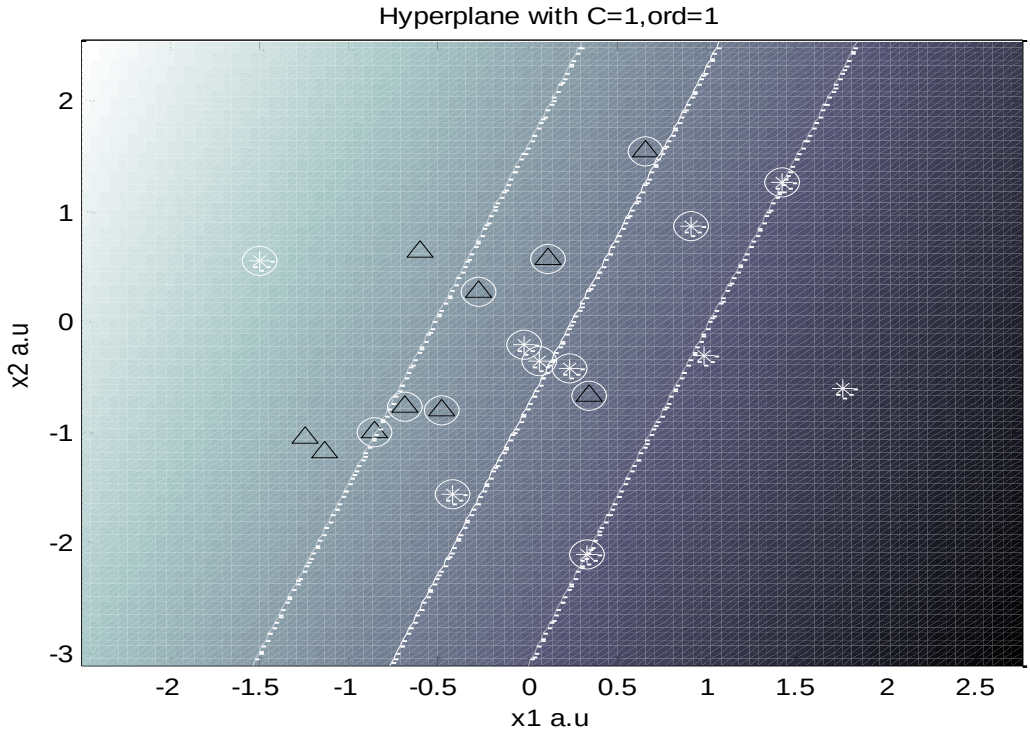


Figure 2.15 Hyperplane is wider for this case of C=1.

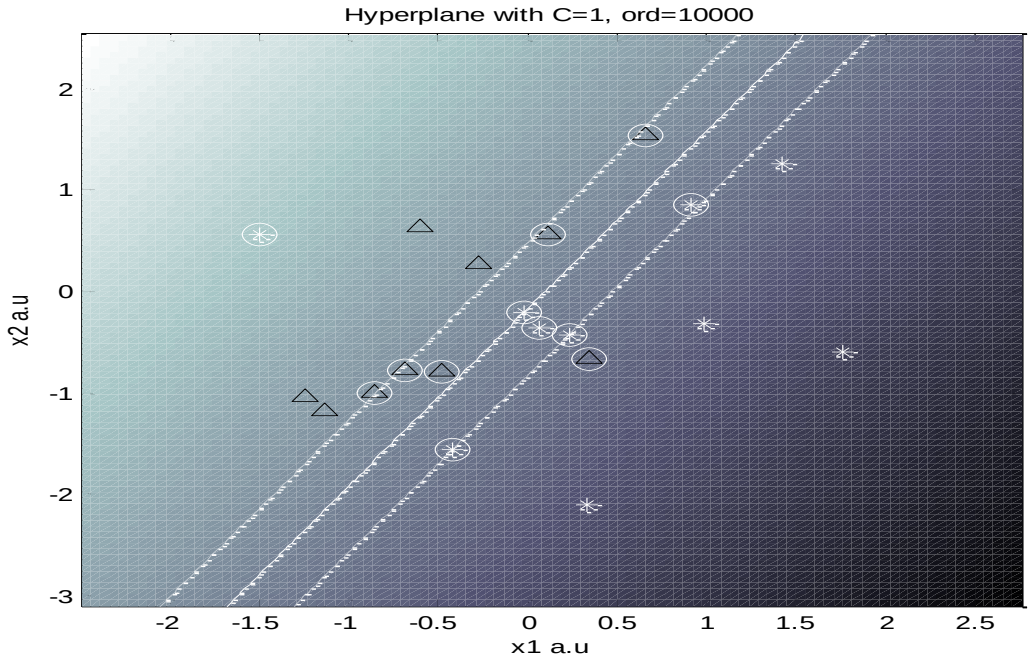


Figure 2.16 Hyperplane is broader for C=10000.

2.4.5 Karush-Kuhn-Tucker Conditions

Karush-Kuhn-Tucker (KKT) conditions are necessary and sufficient conditions for a solution of non–linear programming to be optimal. If a function f where *f* is function of *x* need to be minimized over a set equality and inequality constraints, then $f(x)$ should satisfy KKT conditions for its solution to be optimal. The KKT conditions for an optimizing problem discussed below.

Consider the following optimization problem

Minimize $f(x)$ (*2.4.5.*1)

Subject to $g_i(x) \leq 0$, *for i =1, 2, 3…m* (*2.4.5.*2)

$$h_j(x) = 0 \qquad \textit{for j =1, 2, 3…q} \qquad (2.4.5.3)$$

where : $g: R^n \to R$, $h: R^n \to R$, and m, q represents the number of inequality and equality constraints.

Where $f(x)$ the function to be minimized, subjected to $g_i(x)$ and $h_j(x)$ which are inequality and equality constraints, respectively. Moreover, let us assume that they are continuously differentiable at all points x'. Where x' is the minimizer for the above optimization problem. The Lagrange of the function is given by

$$\mathcal{L}(x', \lambda_i, \mu_i) = f(x') + \sum_{i=1}^{m} \lambda_i \, g_i(x') + \sum_{j=1}^{l} \mu_i \, h_j(x') = 0 \qquad (2.4.5.4)$$

where λ_i, μ_i are Lagrange multipliers of inequality and equality constraints respectively. Then KKT conditions are given by

optimality

$$\frac{d}{dx}\mathcal{L}(x, \lambda, \mu) = 0 \qquad (2.4.5.5)$$

$$\frac{d}{d\lambda}\mathcal{L}(x, \lambda, \mu) = 0 \qquad (2.4.5.6)$$

$$\frac{d}{d\mu}\mathcal{L}(x, \lambda, \mu) = 0 \qquad (2.4.5.7)$$

Primal feasibility

$g_i(x) \leq 0,$ *for I =1, 2, 3…m* (*2.4.5.*8)

$h_j(x) = 0$ *for j=1, 2, 3…q* (*2.4.5.*9)

Non-negativity condition

$\lambda_i \geq 0,$ *for i=1, 2, 3…m* (*2.4.5.*10)

Complimentary slackness,

$\lambda_i g_i(x) = 0$ *for i=1, 2, 3…m* (*2.4.5.*11)

The equations from (2.4.5.5)-(2.4.5.11) are sufficient and necessary conditions for the solution of optimization problem to optimal.

2.4.6 Calculation of Support Vectors and Bias Term b

The calculation of support vectors and bias is carried explicitly using KKT conditions as discussed above.

2.4.6.1 Calculation of Support Vectors

The calculation of support vector machines is based on the dual feasibility of KKT conditions. The condition states that the Lagrange multipliers of data points should always be equal and greater than zero for the solution of the optimization to be optimum. As illustrated in equation *2.4.3.*27, the final quadratic optimization depends only on Lagrange multipliers greater than zero and less than C. Where the data points associated with the above Lagrange multiplier are support vectors. The data points which satisfy the above conditions are collectively known as Support Vectors (SVs). It is widely proved in literature that these SVs are more than enough to represent a separating hyperplane [16]. Hence, SV guide and control the boundary of a separating hyperplane. The number of SV's is always less than the number of data points itself. This makes computations involved with classifying huge data sets less complex, instead of

using all the data points we can use few support vectors for future classification purpose. For example

S= [1, 2, 3, 4, 5, 6]; be data set where 1-3 belongs to class 1 and 4-6 belong class 2 in a data set. Suppose the Lagrange multipliers be $\lambda =$ [0. 1.5, 1, 0, 0, 1.25]. Then the support vectors are SV= [2, 3, 6]. The data set is reduced to half for the above case. In Figure 2.15 the circled ones are the support vector.

2.4.6.2 Calculation of Bias Term b

The bias term b is calculated explicitly using complementary slackness condition of KKT conditions which is discussed in section 2.4.5. The calculation of bias is same for both the linear and non linear separable case. Let us consider the bias calculation for non separable case. The complementary slackness condition is given by

$$\lambda_i[(y_i(\mathrm{w}.\phi(x_i)+\mathrm{b})-1+\xi_i)]=0 \qquad (2.4.6.2.1)$$

As discussed above, the data points on the margin have optimal solution for optimization problem and Lagrange multipliers associated with support vectors are always greater than zero. Where the value of slack variable ξ_i is always zero for support vectors as discussed in appendix B. The details about the slack variables are discussed in soft margin section in appendix B.

Consider a point on the supporting hyperplanes. Then

$$y(\mathrm{w}.\phi(x)+\mathrm{b})-1=0 \qquad (2.4.6.2.2)$$

Taking 1 to other side and multiplying both sides by y

$$y^2(\mathrm{w}.\phi(x)+\mathrm{b})=\mathrm{y} \qquad (2.4.6.2.3)$$

Where $y^2=1$, so (2.4.6.2.3) becomes

$$\mathrm{w}.\phi(x)+\mathrm{b}=y \qquad (2.4.6.2.4)$$

$$w=\lambda_j y_j \phi(x_j) \qquad \text{for j=1, 2, 3....N} \qquad (2.4.6.2.5)$$

where N is number of support vectors.

$$\lambda_j y_j \phi(x_j)\phi(x) + \mathrm{b} = y \qquad (2.4.6.2.6)$$

$$\mathrm{K}\,(x, x_j)(\,\lambda_j y_j) + \mathrm{b} = y \qquad (2.4.6.2.7)$$

$$\mathrm{b} = y_{\,i} - \mathrm{K}\,(x, x_j)(\,\lambda_j y_j) \qquad (2.4.6.2.8)$$

If we consider the linearly separable case the $\mathrm{K}\,(x_i, x_j)$= $x_i . x_j$.

The equation for b changes to

$$\mathrm{b} = y - x_i . x_j . \,(\,\lambda_j y_j) \qquad (2.4.6.2.9)$$

Using any support vector the bias can be determined, but for numerical stability it is better to average over all the support vectors.

$$b = 1/r(\textstyle\sum_{i=1}^{r} y_i - \mathrm{K}\,(x_i, x_j)\,(\,\lambda_j y_j)) \qquad \text{for I, j=1, 2, 3……N} \qquad (2.4.6.2.10)$$

Where is number of support vectors.

2.4.7 Flow Chart for Support Vector Machines

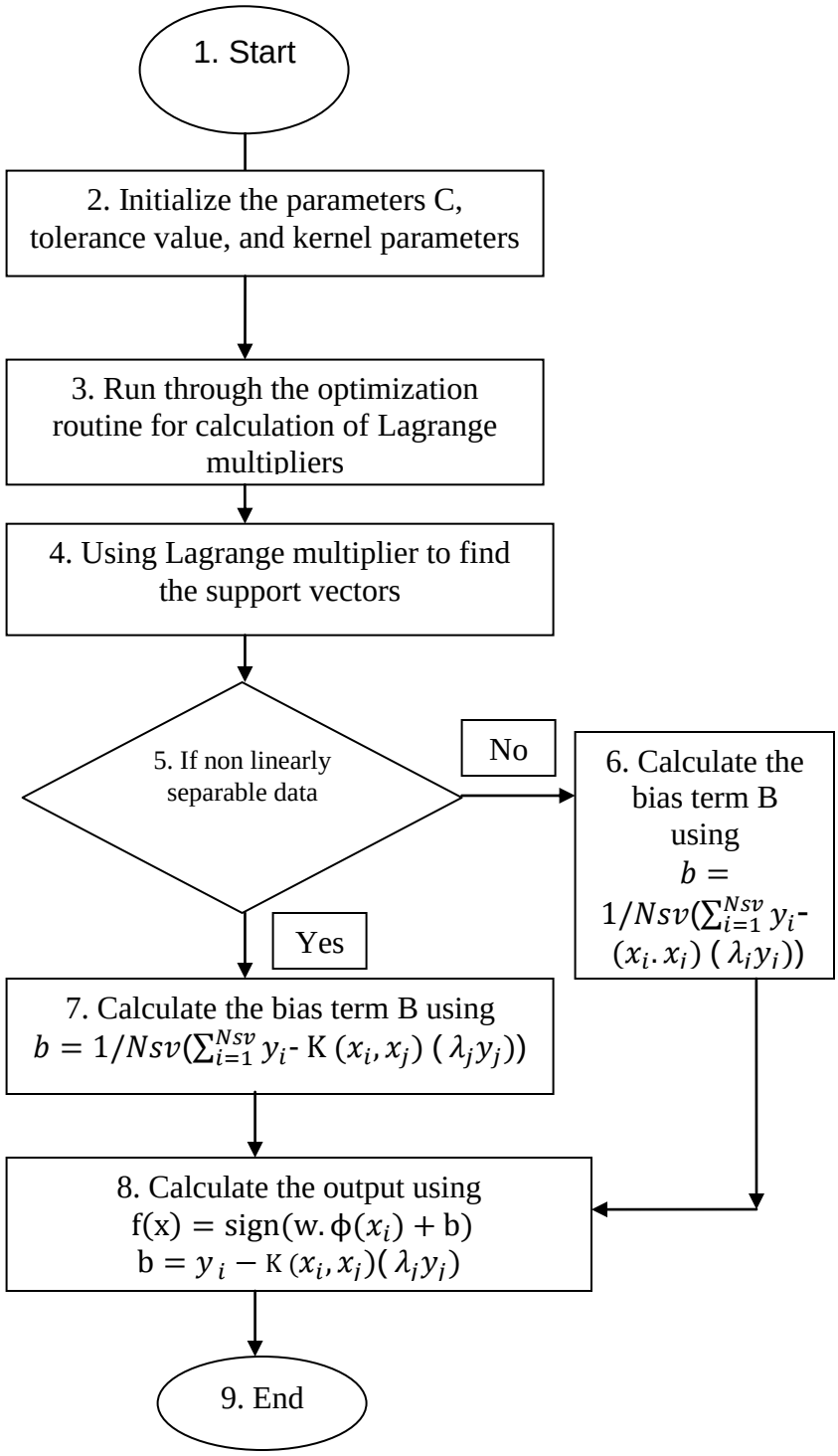


Figure 2.17 Flow chart for the support vector machines.

2.5. Optimization of Hyperplane Parameters

The regularizing parameter (C) and n the order of the hyperplane are optimized to get an optimum separating hyperplane. There are number of heuristic approaches available for the optimization purpose. However, in this thesis we use two optimization techniques for the selection of C and n. we refer them as manual optimization and computerized optimization. The first optimization method is manual optimization where we define a selection criterion. The second method is computerized optimization where we use the derivative free optimization functions available in matlab to generate an optimized value for C and n.

2.5.1 Manual Optimization

The manual optimization method of selection of C and n is based idea that, the combination C and n which provides maximum accuracy value for the training set will also give a good generalization of results for testing set. Since, during the testing step SVM was blind to the data set; the testing accuracy is the accuracy for SVMs. In this method two techniques were used. The maximum of accuracy for training set and maximum of accuracy of validation set. The range over which C and n operate is found using iterative search. The range for C is [1, 16] and for n is [1, 30]. All the optimization techniques use the above discussed range.

2.5.1.1 Maximum of Accuracy for Training Set

In this method, C and n are incremented sequentially within the selected range. Each C and n value is run against the training data to get an accuracy value. The combination of C and n which gives highest accuracy value for the training set is selected as the optimum value. This referred to as manual1. The resulting values of C and n are used for learning the training parameters; the bias term(b), and Lagrange multipliers(λ). These training parameters are fixed, in order to run the test set and calculate the accuracy of the SVM.

2.5.1.2 Maximum of accuracy for validation set

This method of selection of C and n is also same as above described method, but with a small difference. The training data in this method is divided into new training and validation data. In

this study 10-fold cross validation is used. Again, the C and n are incremented sequentially within the selected range. This referred to as manual2. The new training data is trained with a C and n value to obtain the training parameters which are tested against the validation set. The combination of C and n which gives a highest accuracy for validation set is selected. The new training data and the validation data are combined to form the old training data. The optimized combination of C and n are trained using the training set in order to get the λ and b values, which are tested against the test set. The detail description of cross validation is given in next chapter.

2.5.2 Computerized Optimization

Computerized optimization is used to check how well the SVM routine performs with matlab optimization function. In this method, we used optimization toolbox available in matlab for the optimization of C and n. The range for C and n described is also used in this optimization. The derivative free optimization functions Fminbnd and Fminsearch available in the matlab are used for this purpose.

2.5.2.1 Optimization using Fminbnd

Fminbnd finds the minimum of a single variable function on fixed interval. SVM routine is the single variable function and the variables are C in this case. Since the Fminbnd minimizes only single variable function, n is assumed to be constant while optimizing C. This optimization is referred to as fminbnd1. Another optimization is also carried out in this optimization. Where, first C is optimized with n as constant. Followed by optimization of C, n is optimized. This is referred to as fminbnd2. Once the optimization is completed, the training and testing accuracy are calculated for the optimized C and n Value as explained in section 2.5.1.1. The resulting testing accuracy is the accuracy of the classifier.

2.5.2.2 Optimization using Fminsearch

Fminsearch finds the minimum of an unconstrained multivariable function. The optimization of C and n are carried out together. Once the optimization is complete, the training and testing

accuracy is computed for optimized C and n as described in section 2.5.1.1. The training and testing accuracies are stored for the current data. The whole procedure is repeated for new training and testing data set.

CHAPTER 3

EXPERIMENTAL METHOD

In this chapter experimental method used for the purpose of the study is discussed in detail. Section 3.1 discusses the standard Nocturnal polysomnograph (NPSG) data acquisition. Section 3.2 describes the subject demographics. Section 3.3 explains the processing of the experimental data. The final section discusses the Performance evaluation of the proposed method.

The data used in this study was previously collected and prepared by our laboratory researchers in collaboration with Sleep Consultants Inc., Fort Worth, TX. A brief explanation of the acquisition method is provided below.

3.1 Standard NPSG Data Acquisition

Data was acquired overnight from the subjects during sleep. A total of 8 physiological parameters were measured using 18 channels. 18 Channels used to record the parameters are, nine for ECG, three for EEG, one channel each for EOG, chin EMG, chest and abdominal movements, nasal airflow, and percent oxygen saturation. For sleep stage scoring, EEG acquired from position C1-A2 of 10-20 system was used [8]. A frequency of 25-100 Hz was used to sample and acquire EOG, chin EMG, chest and abdominal movements, percent oxygen saturation and EEG. A frequency of 1024 Hz was used to sample and acquire remaining nine ECG channels (Lead I, II, III and V1-V6) and nasal airflow [10]. A sleep expert, blind to the goal of the study, manually scored the entire original data. Table 3.1 gives the Lead configuration, electrode position, their placements and their abbreviations. Figure 3.1 shows the electrode placements. For more detailed information on data collection, refer to [8, 10].

Table 3.1 Leads and Combination of their Electrode Placements

Lead configuration	Electrode Placements	Electrode Abbreviations(with reference to figure 2.3)
Lead I	Right arm-Left arm	RA-LA
lead II	Right arm -left Leg	RA-LL
lead III	Left arm-Left Leg	LA-LL
V1	V1(precordial)-Left leg	V1-LL
V2	V2(precordial)-Left leg	V2-LL
V3	V3(precordial)-Left leg	V3-LL
V4	V4(precordial)-Left leg	V4-LL
V5	V5(precordial)-Left leg	V5-LL
V6	V6(precordial)-Left leg	V6-LL

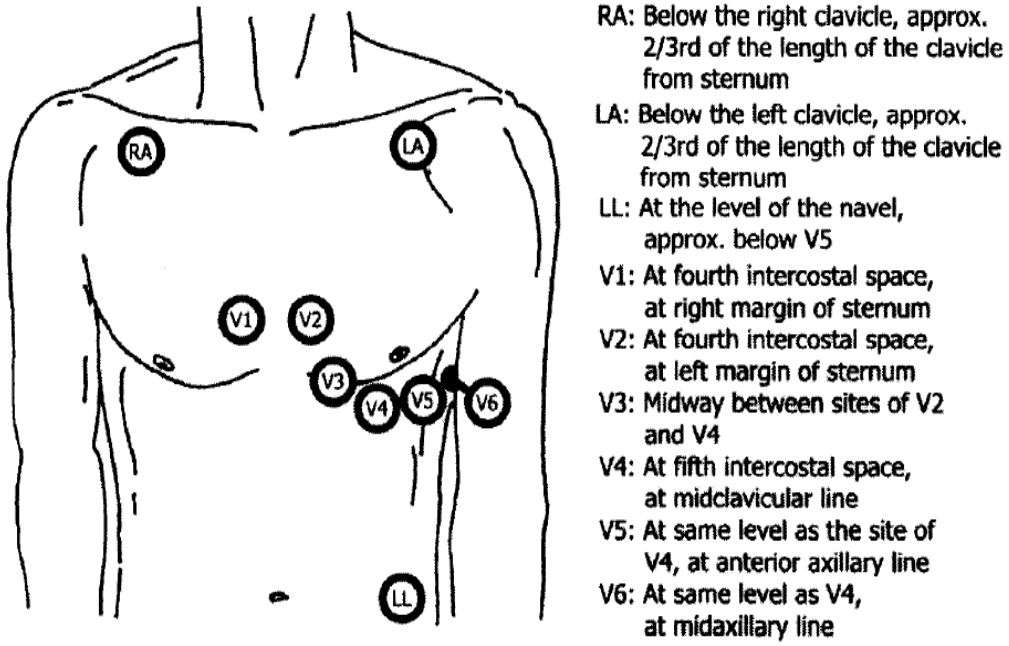


Figure 3.1 ECG electrode placements. Adopted from [8]

3.2 Subject Demographics

Thirty (30) adult volunteers were recruited for the purpose this study as detailed in [9]. Sixteen normal subjects were recruited as a control group, where none of them had prior sleep related problems and had not previously participated in NPSG studies. They were selected from broad public and the selection criterion was based on a questionnaire on health history and sleep

hygiene. Subjects from the group with prior no respiratory problems are referred as 'NOR' group. Fourteen subjects were recruited from volunteer patients who were previously diagnosed as having SDB. These groups of subjects are referred as 'OSA' group [8]. Tables 3.2 shows the summary of NOR subject demographics and the sleep expert score of their NPSG studies. Table 3.3 contains demographics and NPSG test results for OSA subjects.

Table 3.2 Subject Demographics for NOR Group

subject id	Gender	Age (Years)	Weight (kg)	Heigh t(m)	BMI (KG/m^2)	AHI
N01	M	43	87	1.85	25.4	3
N02	M	36	66	1.73	22.1	6
N03	F	58	64	1.6	25	0
N04	M	62	65	1.68	23	2
N05	M	49	95	1.75	31	4
N06	F	42	82	1.7	28.4	6
N07	F	40	61	1.6	23.8	2
N08	F	35	46	1.58	18.4	0
N09	M	38	68	1.65	25	6
N10	M	56	86	1.75	28.12	2
N11	F	54	54	1.6	22.3	3
N12	M	39	100	1.78	31.6	11
N13	F	36	81	1.68	28.7	2
N14	F	43	NA	NA	NA	20
N15	M	59	93	1.88	26.3	1
N16	F	42	78	1.65	28.7	14
Mean ± standard Deviation		46 ±9.38	73.08 ±16.5	1.69 ±0.09	25.34 ±3.86	3.75 ±3.11

Where AHI is the apnea hypopnea index and BMI is body mass index. Subject N12 and N14 were removed from the study set for having high AHI. Subjects N13 and N15 were removed from the study for abnormal ECG Lead I recordings.

Table 3.3 Subject Demographics for OSA Group

subject id	Gender	Age (Years)	Weight (kg)	Height (m)	BMI (KG/m^2)	AHI
N17	M	50	99	1.83	29.6	9
N18	M	38	91	1.88	25.7	4
N19	F	49	67	1.75	21.9	19
N20	M	39	157	1.9	43.5	70
N21	F	47	91	1.65	33.4	57
N22	M	37	64	1.63	24.1	8
N23	M	56	128	1.85	37.4	37
N24	F	44	89	1.7	30.8	20
N25	F	49	59	1.6	23	62
N26	M	49	100	1.8	30.9	14
N27	M	57	105	1.8	32.4	4
N28	F	54	92	1.52	39.8	30
N29	F	69	76	1.52	32.9	38
N30	M	66	95	1.75	33.2	65
Mean ± standard deviation		50.28 ±9.60	93.79 ±25.2	1.73 ±0.13	31.33 ±6.29	31.21 ±23.89

3.3 Processing of Experimental Data

As was described earlier, a sleep specialist, blind to the aim of the study, manually scored the entire original data. Using the scored data two types of 15 minute long data epochs or clips were extracted from the full night recorded data. Two types clips extracted were extracted from

the ECG. ECGs with amplifier saturation, excessive movement artifact, noise bursts, or large baseline wander, and epochs with high-frequency noise were discarded. The epoch are labeled 'NOR' if the breathing over entire 15 minutes clip showed normal breathing and the clip was extracted from normal subject. The epochs are labeled 'OSA' if 4 or more apneic events were scored during the entire 15 minutes clip and the clip was acquired from OSA subjects.

For the purpose of this study, a clip, which is extracted from a normal (NOR) subject was given a diagnostic value of negative one (-1). Any clip, which is extracted from apneic (OSA) subject, was given a diagnostic value of one (1). Table 3.4 summarizes the number of clips extracted, per subject for the purpose of this study.

Table 3.4 List of the Clips Selected from Subjects

NOR		OSA	
Subject ID	No. of clips	Subject ID	No. of clips
N01	7	N17	4
N02	6	N18	4
N03	7	N19	**5**
N04	8	N20	12
N05	7	N21	13
N06	4	N22	6
N07	7	N23	11
N08	11	N24	13
N09	8	N25	5
N10	5	N26	8
N11	6	N27	2
N12	6	N28	5
		N29	8
		N30	9
Total Nor	86	total OSA	105

3.4 Performance Evaluation for the Proposed Method

3.4.1 Cross Validation

Cross validation is a technique used to assess the reliability of a learner for unknown data. This technique is very useful in predicting the output of a learner for a new unseen data. The fundamental design of the cross validation is to divide the data into two groups, one for training and other for testing. Where the learner is trained with training data group and determine the parameters of the learner. These parameters are used for testing the performance of the model on unseen data, the testing data. In this study we use K-fold cross validation.

3.4.1.1 K-Fold Cross Validation

The data set is divided into K-subsets and the hold out method discussed in appendix C is used K-times. Each time, out of the K subsets one subset is set aside for testing and remaining K-1 is used for training. The training parameters obtained are used to test the testing subset. The errors are averaged over all of the K trails.

3.4.2 Monte Carlo Simulation

Monte Carlo simulation is a technique used to test the robustness of a complex system using random samples of parameters or inputs. Monte Carlo simulation is a computational analysis technique, where repeatedly and randomly resampling the data could give a desired result. Monte Carlo simulation uses a set of random numbers as input and iteratively evaluates the deterministic models. This is repeated numerous times. The average of the results given by these simulations gives the reliability of a particular system. Monte Carlo simulations find most of their applications in simulating physical and mathematical systems. Monte Carlo simulation is a variant of K-fold cross validation. The data in Monte Carlo simulation is divided randomly into learning set and validation set.

In order to test the performance of the SVM routine used in this study, the selected feature set, used as input vectors to SVM, are randomly assigned into two groups: training and testing, with a 2:1 ratio. With total of 191 clips, this is translated to 127 clips for training, and the remaining

64 for testing. K-fold validation is used for the purpose of optimizing the parameters for the SVM algorithm. A 500 run Monte Carlo simulations are performed to evaluate the performance of the SVM for randomly generated unseen data.

3.4.2.1 Quantification of Accuracy, Sensitivity, and Specificity

The number of distinct possible arrangements that the example vectors can assume can be calculated using the un-ordered arrangement equation:

Number of arrangements $= \frac{191!}{(191-127)! \times 127!} = 1.3031 \times 10^{53}$

500 is a large sampling of these arrangements which would provide a reasonable estimate of the algorithm performance. The performance of the SVM is evaluated in terms average of the sensitivity, specificity and accuracy of 500 runs. They are defined as follows [9]

$$\text{Sensitivity} = \frac{OSA_p}{Total\ Number\ of\ positive\ clips\ \ tested} \text{ x100\%}$$

$$\text{Specificity} = \frac{NOR_p}{Total\ Number\ of\ negative\ clips\ \ tested} \text{ x100\%}$$

$$\text{Accuracy} = \frac{OSA_p + NOR_p}{Total\ NUmber\ of\ clips\ \ tested} \text{ x100\%}$$

Where OSA_p, NOR_p are correctly classified OSA, NOR clips respectively.

The probability that a diagnostic test is positive, given that subject is apneic is referred to as sensitivity. The probability that a diagnostic test is positive, given that the subject is normal is referred to as specificity. The probability that the diagnostic test is correctly performed is referred as accuracy.

CHAPTER 4

RESULTS

This chapter reports the training and testing performance of the detection algorithm used for the purpose of the study. As explained in Chapter 2, the RPE and R-R interval are the features extracted from ECG waveform. The entroid extracted from the cluster of these two features is fed to the SVM. Also, as was discussed earlier in chapter 2 polynomial kernel is used for the SVM algorithm. Two optimization techniques are used for the optimization of hyperplane parameters as discussed in Chapter 2 section 2.5. The performance of these 2 optimizations is reported in this section.

4.1 Manual Optimization

As was discussed in Chapter 2, the manual optimization refers to optimization where the user provides the selection criterion. In this optimization process two types of selection criterion for regularization parameter (C) and degree of the polynomial kernel (n) are used. In this study, the range of C and n is selected using iterative search. The range for C is selected as [1, 16] and n is selected as [1, 30] using the iterative search. This range for C and n is used for all the optimization techniques employed for the purpose of the thesis. The results for performance are reported below.

4.1.1 Optimizing hyperplane parameters using Manual1

The selection criterion for the optimization technique is based on the maximum accuracy value for training set. For ease of reference, this method is referred to as Manual1. The detail analysis of selection criterion is given section 2.5. Figure 4.1 Illustrates the graphical representation of average of training and testing results obtained for the Manual1 optimization technique for 500 Monte Carlo runs. Where error bars represent one standard deviation of computed accuracy values. Figure 4.2 and Figure 4.3 illustrates the histogram of the training accuracy and testing

accuracy for 500 Monte Carlo runs, respectively. Table 4.1 summarizes the performance results of training and testing for 500 Monte Carlo runs.

Table 4.1 Sensitivity, specificity and accuracy mean and standard deviation (std) for training and testing sets after 500-run Monte-Carlo simulation for the optimization

	Training			Testing		
	Accuracy	sensitivity	specificity	Accuracy	sensitivity	Specificity
Mean±Std	90.28 ±3.01	94.43 ±3.15	85.18 ±4.47	75.98 ±4.58	81.20 ±7.16	69.67 ±8.02
Max	96.85	100	96.49	87.5	97.14	89.65
Min	81.88	82.85714	73.68	60.93	57.14	41.37

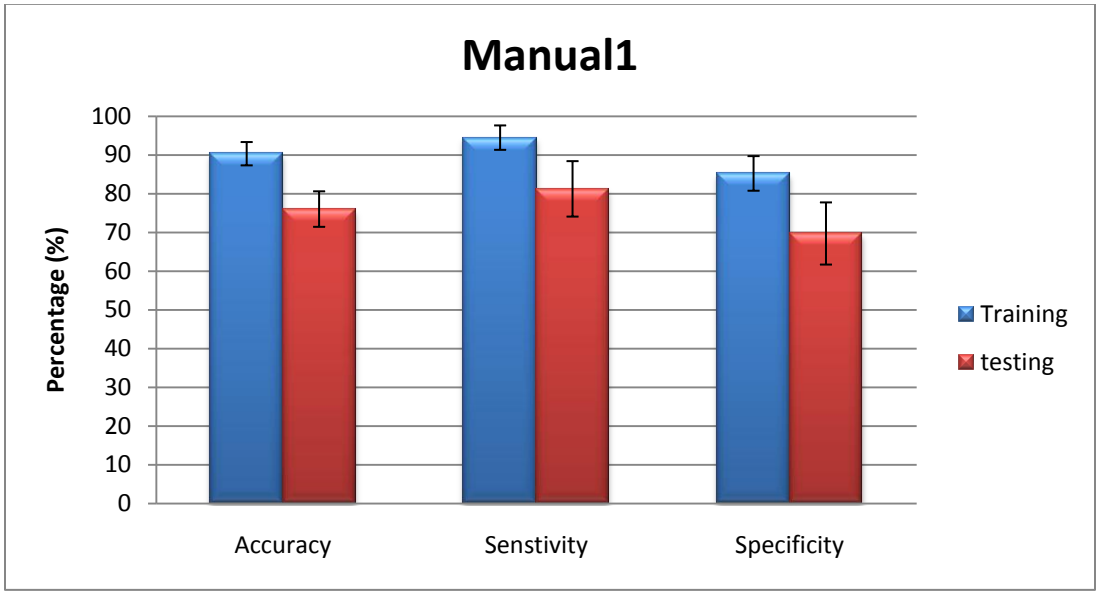


Figure 4.1 Graphical representation accuracy, sensitivity and specificity mean results for training and testing sets after 500 Monte Carlo runs. Error bars represents the standard deviation.

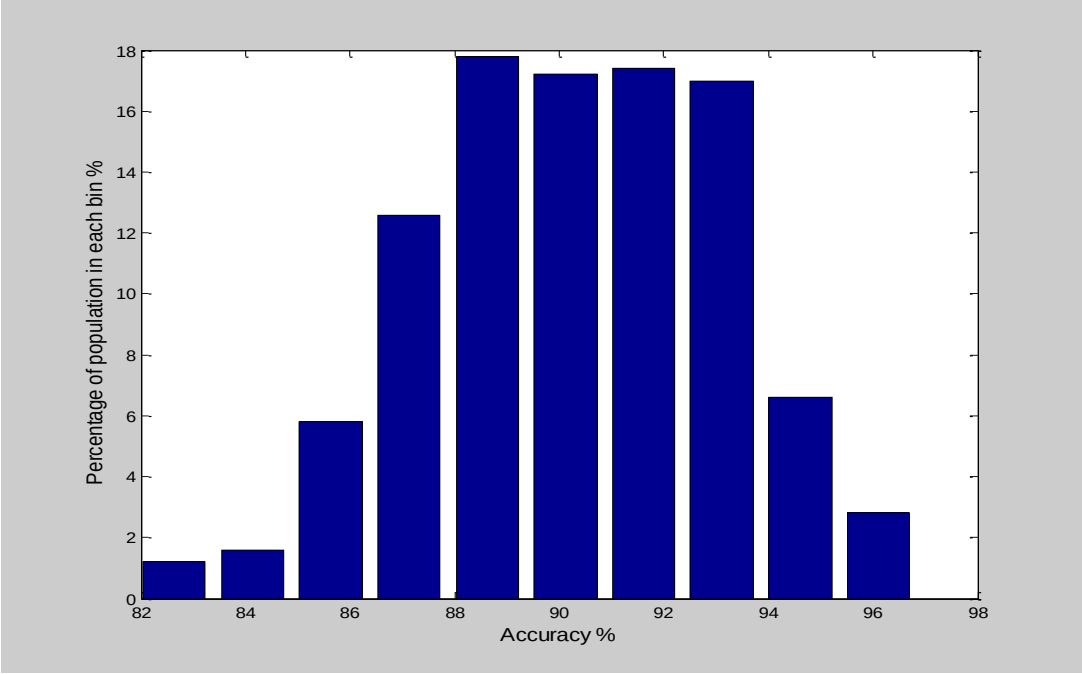


Figure 4.2 Illustration of histogram of training accuracy for 500 Monte Carlo runs. The graph represents percentage of accuracy values in each of the 10 equally divided accuracy bins.

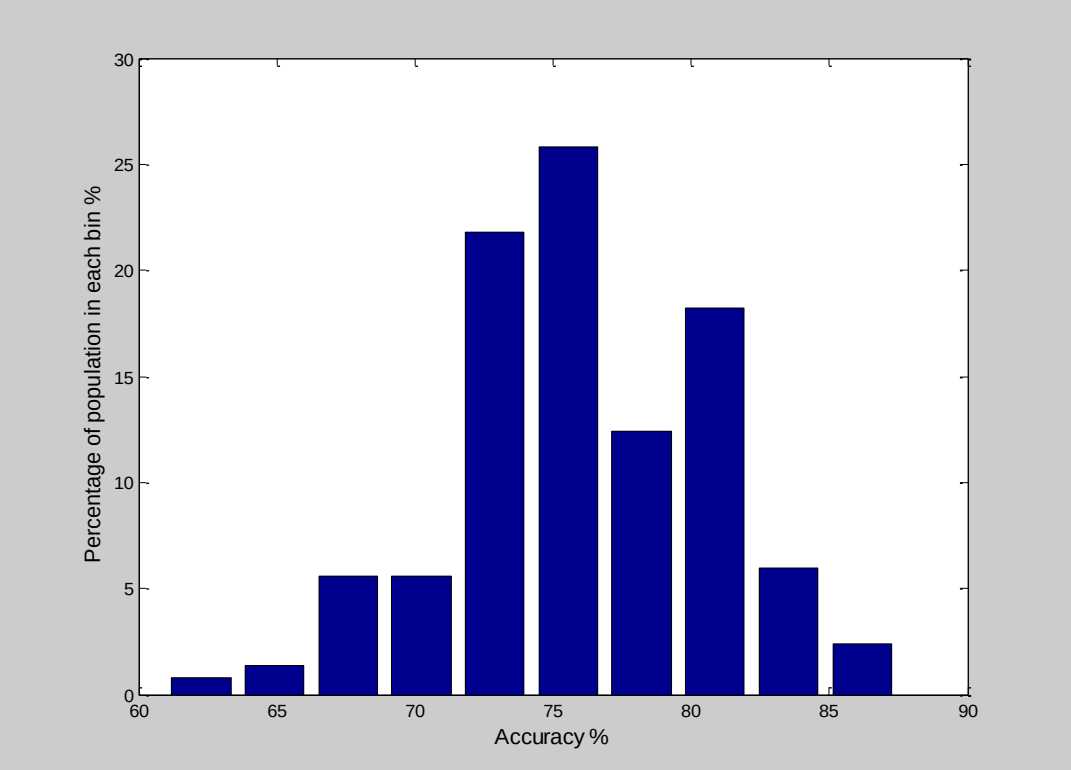


Figure 4.3 Illustration of histogram of testing accuracy for 500 Monte Carlo runs. The graph represents percentage of accuracy values in each of the 10 equally divided accuracy bins.

4.1.2 Optimizing hyperplane parameters using maximum of accuracy Validation set

The selection criterion for hyperplane parameters for the optimization is based on the Maximum accuracy values for validation set. This routine is referred to as Manual2 in this study. The selection criterion is discussed in detail in section 2.5. Figure 4.4 Illustrates the graphical representation of average of training and testing results obtained for the above optimization technique for 500 Monte Carlo runs. Where error bars represent one standard deviation the computed values. Figure 4.5 and Figure 4.6 illustrates the histogram of the training accuracy and testing accuracy for 500 Monte Carlo runs, respectively. Table 4.2 summarizes the performance results of training and testing for 500 Monte Carlo runs.

Table 4.2 Sensitivity, specificity and accuracy mean and standard deviation (std) for training and testing sets after 500-run Monte-Carlo simulation for the optimization

	Training			Testing		
	Accuracy	sensitivity	specificity	Accuracy	sensitivity	Specificity
Mean±Std	91.16 ±2.81	95.2 ±2.93	86.21 ±4.09	75.82 ±5.15	80.02 ±7.45	70.79 ±8.32
Max	97.63	100	96.49	92.18	97.14	96.55
Min	83.46	85.71	71.92	57.81	48.57	44.82

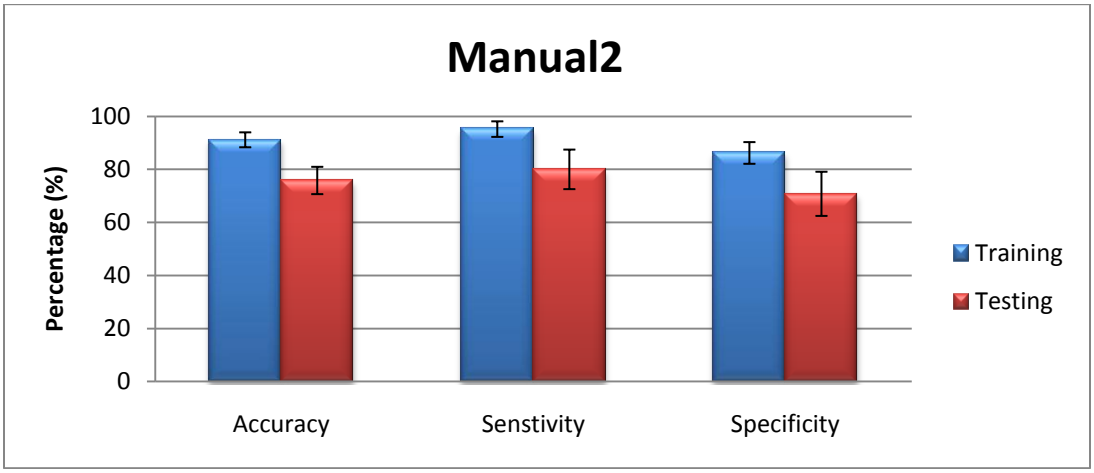


Figure 4.4 Graphical representation accuracy, sensitivity and specificity mean results for training and testing sets after 500 Monte Carlo runs. Error bars represents the standard deviation.

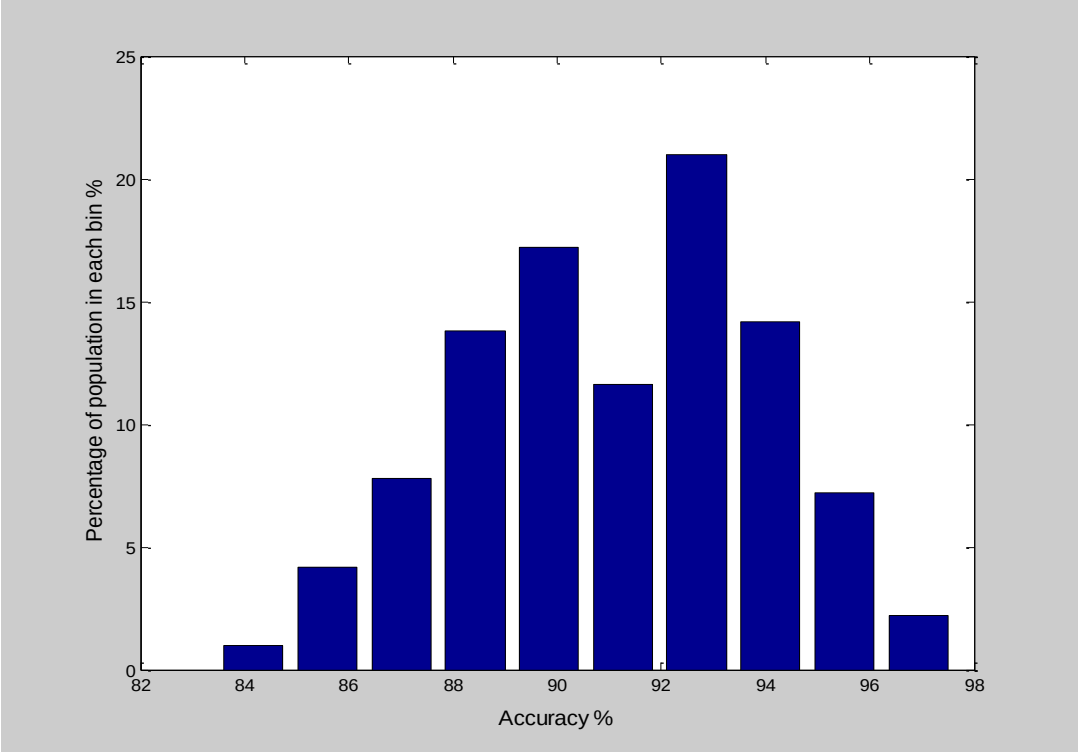


Figure 4.5 Illustration of histogram of training accuracy for 500 Monte Carlo runs. The graph represents percentage of accuracy values in each of the 10 equally divided accuracy bins

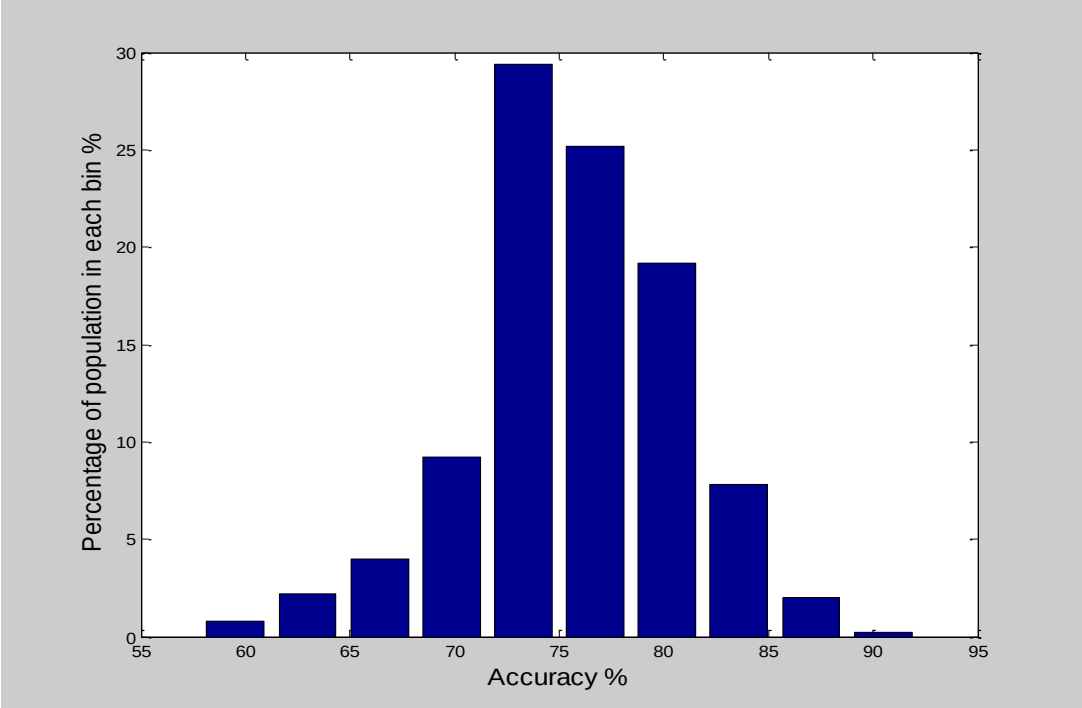


Figure 4.6 Illustration of histogram of testing accuracy for 500 Monte Carlo runs. The graph represents percentage of accuracy values in each of the 10 equally divided accuracy bins.

4.2 Computerized Optimization

As explained in Chapter 2 section 2.5, computer optimization was applied to explore the possibility of improving the performance of the proposed algorithm. In this optimization, the optimization toolbox in $MATLAB^{®}$ is used for the selection of C and n. For the purpose of this thesis, two optimization functions Fminbnd and Fminssearch is used. The Fminbnd optimizes a single parameter at a time within a fixed bound and Fminsearch optimizes multi parameter. The bound for parameter is not fixed in Fminsearch. The performance results are reported below. The range for C and n is same as the range used for Manual optimization.

4.2.1 Optimization of hyperplane parameters using Fminbnd

Two types of optimization are performed using Fminbnd function in matlab. The first optimization technique optimizes C alone. The selection of C as parameter for optimization is explained below. This technique is referred to as fminbnd1 for the purpose of the thesis. The second optimization technique optimizes C followed by n. This is referred as Fminbnd2.

4.2.1.1 Optimization using Fminbnd1

In this optimization we can optimize either C or n. But optimizing C alone gives better results than optimizing n alone. Firstly, the optimization results in decimal number. The parameter C can be a decimal number whereas n must be an integer. Secondly, C controls the margin of the hyperplane. C is a suitable parameter for optimization using this optimization. The optimization of C is performed using Fminbnd1.

The optimization is carried out for each run in 500 Monte Carlo simulations. Figure 4.7 Illustrates the graphical representation of average of training and testing results obtained for the above optimization technique for 500 Monte Carlo runs. Where error bars represent one standard deviation of computed accuracy values. Figure 4.8 and Figure 4.9 illustrates the histogram of the training accuracy and testing accuracy for 500 Monte Carlo runs respectively. Table 4.3 summarizes the performance results of training and testing for 500 Monte Carlo runs.

Table 4.3 Sensitivity, specificity and accuracy mean and standard deviation (std) for training and testing sets after 500-run Monte-Carlo simulation for the optimization

	Training			Testing		
	Accuracy	sensitivity	specificity	Accuracy	sensitivity	specificity
Mean±Std	92.34 ±2.53	96.36 ±2.54	87.4 ±4.12	75.51 ±5.16	80.36 ±7.13	69.66 ±9.15
Max	98.42	100	98.24	89.06	100	93.1
Min	83.46	87.14	77.19	57.81	54.28	37.93

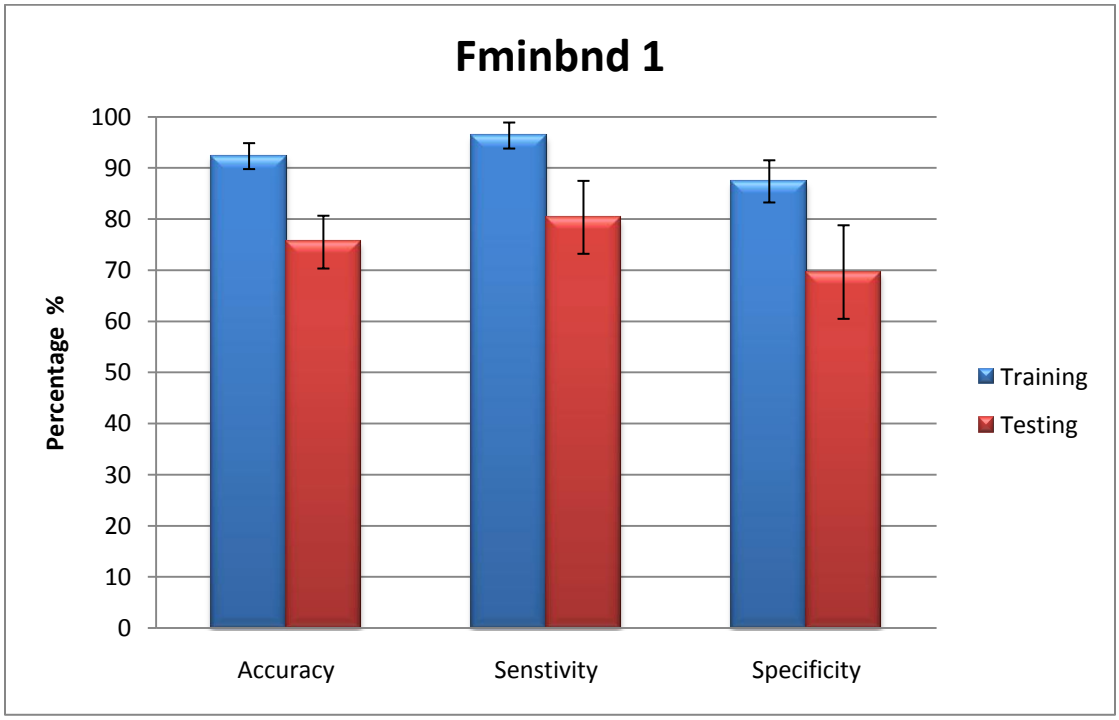


Figure 4.7 Graphical representation accuracy, sensitivity and specificity mean results for training and testing sets after 500 Monte Carlo runs. Error bars represents the standard deviation.

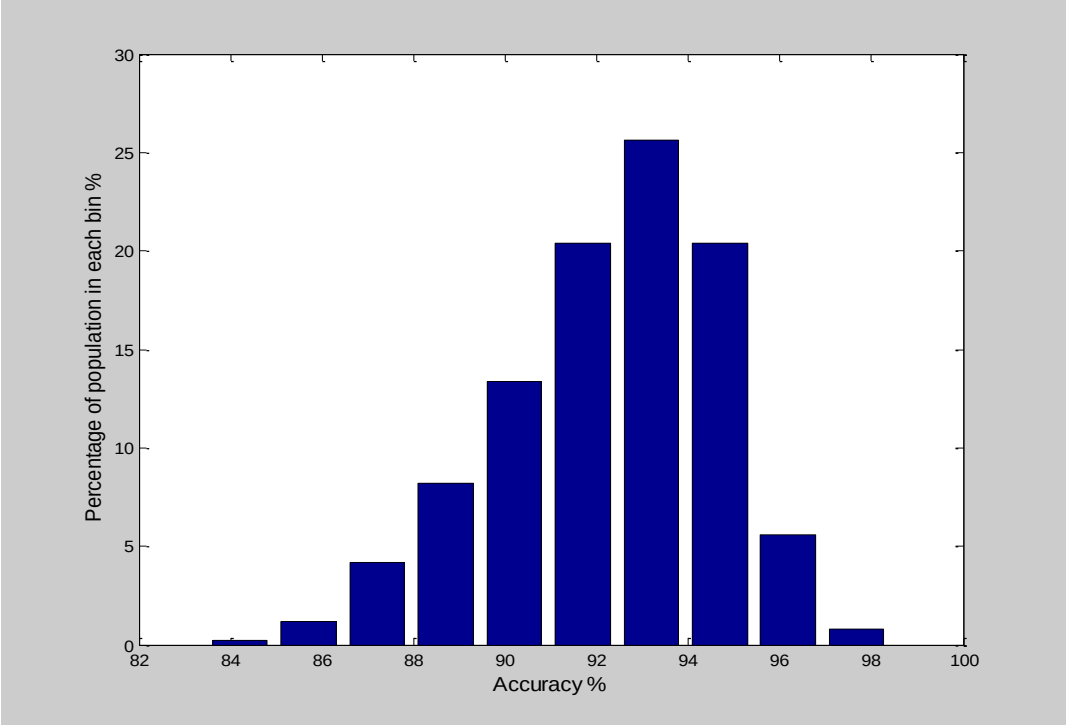


Figure 4.8 Illustration of histogram of training accuracy for 500 Monte Carlo runs. The graph represents percentage of accuracy values in each of the 10 equally divided accuracy bins.

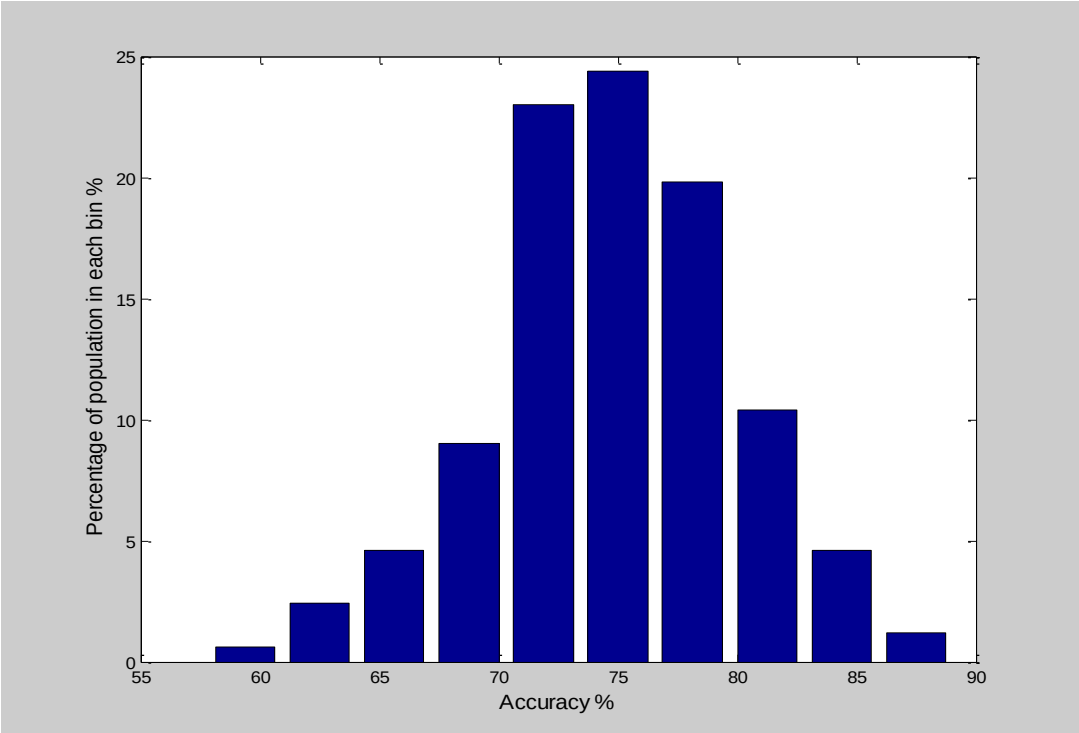


Figure 4.9 Illustration of histogram of testing accuracy for 500 Monte Carlo runs. The graph represents percentage of accuracy values in each of the 10 equally divided accuracy bins.

4.2.1.2 Optimization using Fminbnd2

The optimization of two parameter of hyperplane, namely C and n, is performed sequentially using the same optimization routine as in section 4.2.1.1. Figure 4.10 Illustrates the graphical representation of average of training and testing results obtained for the above optimization technique for 500 Monte Carlo runs. Where error bars represent one standard deviation of computed accuracy values. Figure 4.11 and Figure 4.12 illustrates the histogram of the training accuracy and testing accuracy for 500 Monte Carlo runs respectively. Table 4.4 summarizes the performance results of training and testing for 500 Monte Carlo runs.

Table 4.4 Sensitivity, specificity and accuracy mean and standard deviation (std) for training and testing sets after 500-run Monte-Carlo simulation for the optimization

	Training			Testing		
	Accuracy	sensitivity	specificity	Accuracy	sensitivity	specificity
Mean±Std	87.35 ±3.82	91.41 ±4.23	82.36 ±5.37	76.66 ±4.81	81.84 ±6.75	70.41 ±8.98
Max	97.63	100	98.24	85.9375	97.142	93.1
Min	59.84	77.14	28.07	54.6875	60	31.03

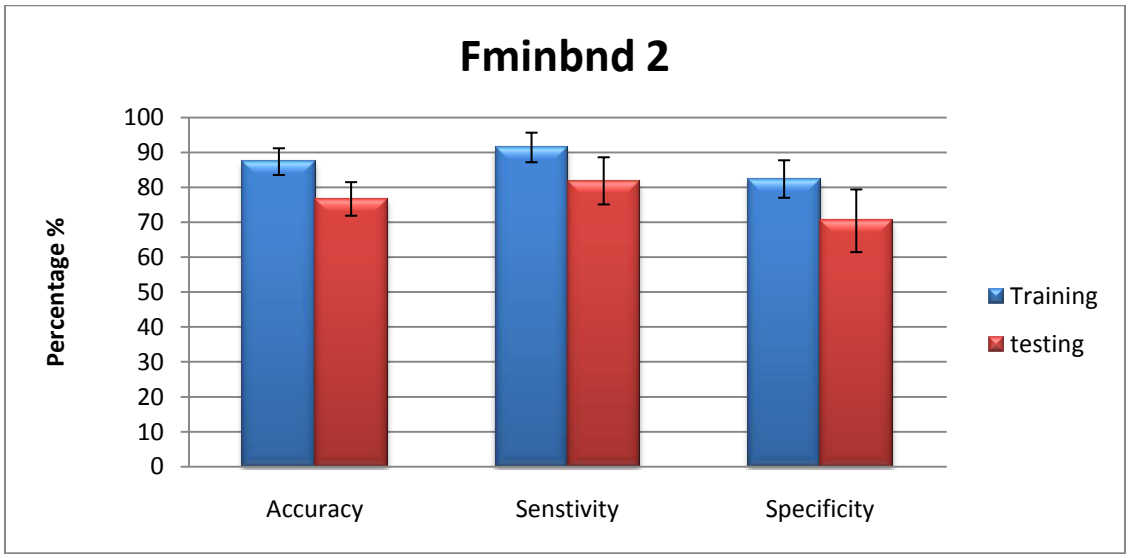


Figure 4.10 Graphical representation accuracy, sensitivity and specificity mean results for training and testing sets after 500 Monte Carlo runs. Error bars represents the standard deviation.

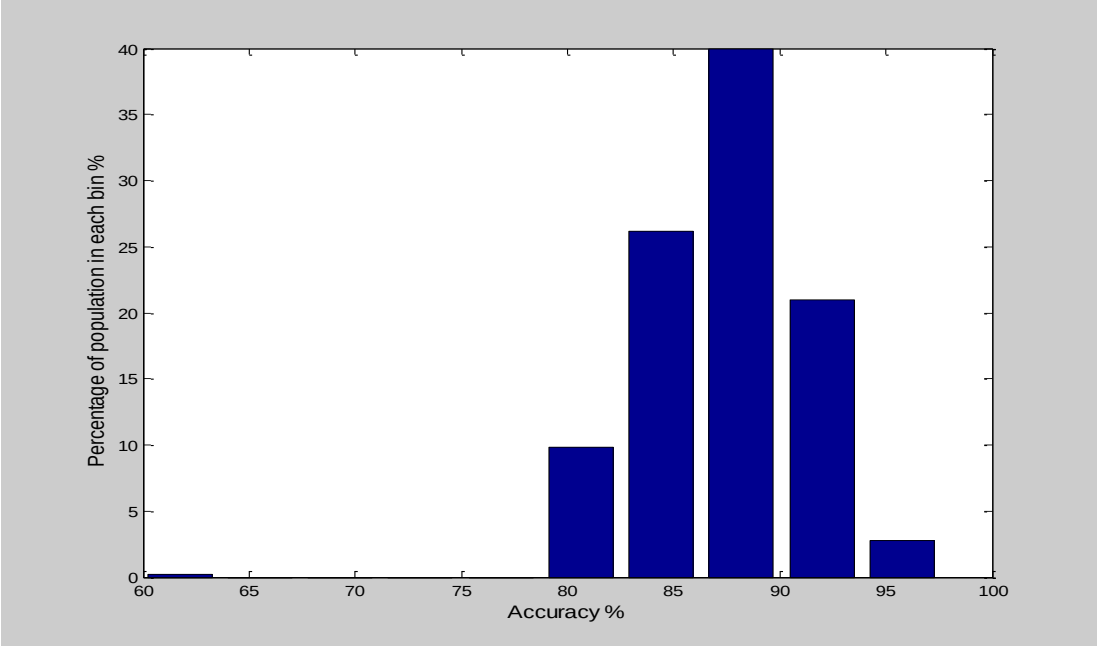


Figure 4.11 Illustration of histogram of training accuracy for 500 Monte Carlo runs. The graph represents percentage of accuracy values in each of the 10 equally divided accuracy bins.

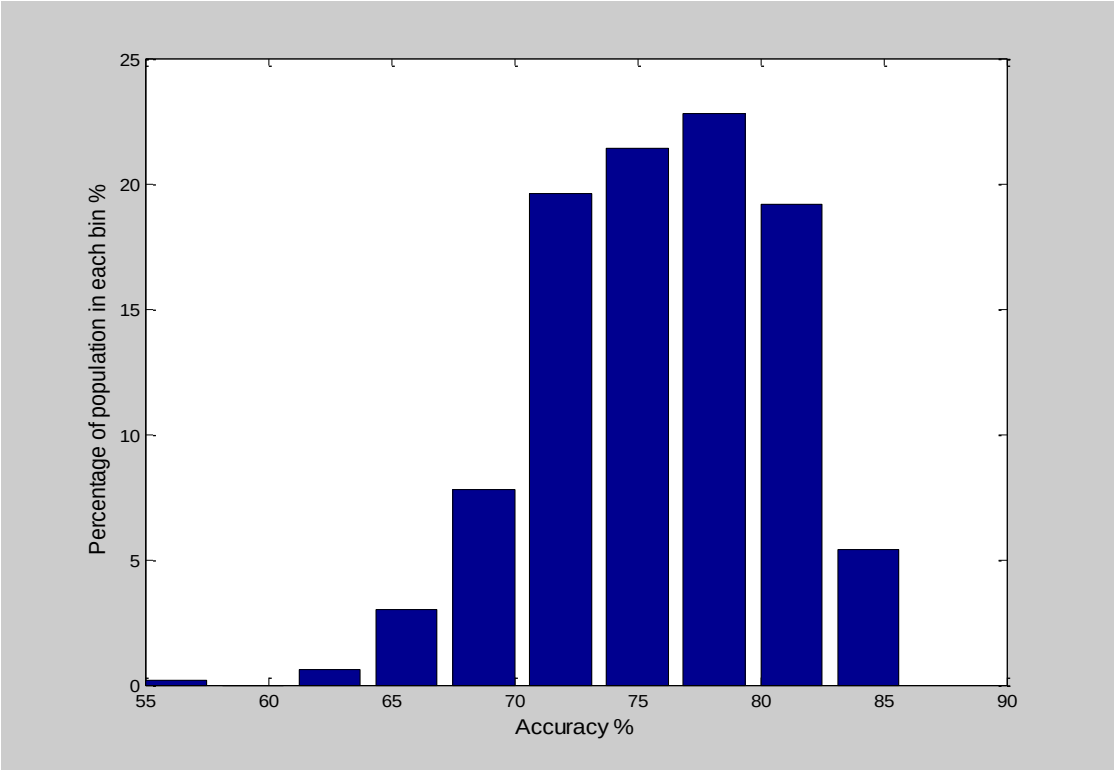


Figure 4.12 Illustration of histogram of testing accuracy for 500 Monte Carlo runs. The graph represents percentage of accuracy values in each of the 10 equally divided accuracy bins.

4.2.2 Optimization of hyperplane parameters using Fminsearch

Both C and n are optimized together in this optimization. Figure 4.13 Illustrates the graphical representation of average of training and testing results obtained for the above optimization technique for 500 Monte Carlo runs. Where error bars represent one standard deviation of computed accuracy values. Figure 4.14 and Figure 4.15 illustrates the histogram of the training accuracy and testing accuracy for 500 Monte Carlo runs respectively. Table 4.5 summarizes the performance results of training and testing for 500 Monte Carlo runs.

Table 4.5 Sensitivity, specificity and accuracy mean and standard deviation (std) for training and testing sets after 500-run Monte-Carlo simulation for the optimization

	Training			Testing		
	Accuracy	sensitivity	specificity	Accuracy	sensitivity	specificity
Mean±Std	92.78±2.53	96.33±2.54	88.43±4.40	74.74±5.6	78.57±8.24	70.12±8.74
Max	99.21	100	98.24	87.5	97.142	93.1
Min	83.46	87.14	77.19	56.25	42.85	44.82

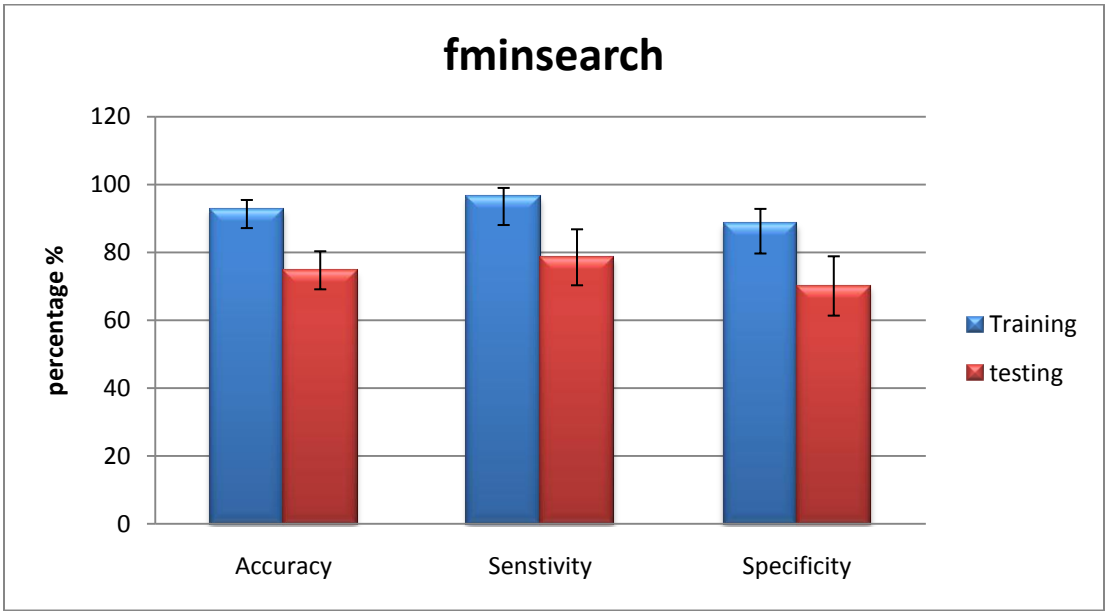


Figure 4.13 Graphical representation accuracy, sensitivity and specificity mean results for training and testing sets after 500 Monte Carlo runs. Error bars represents the standard deviation.

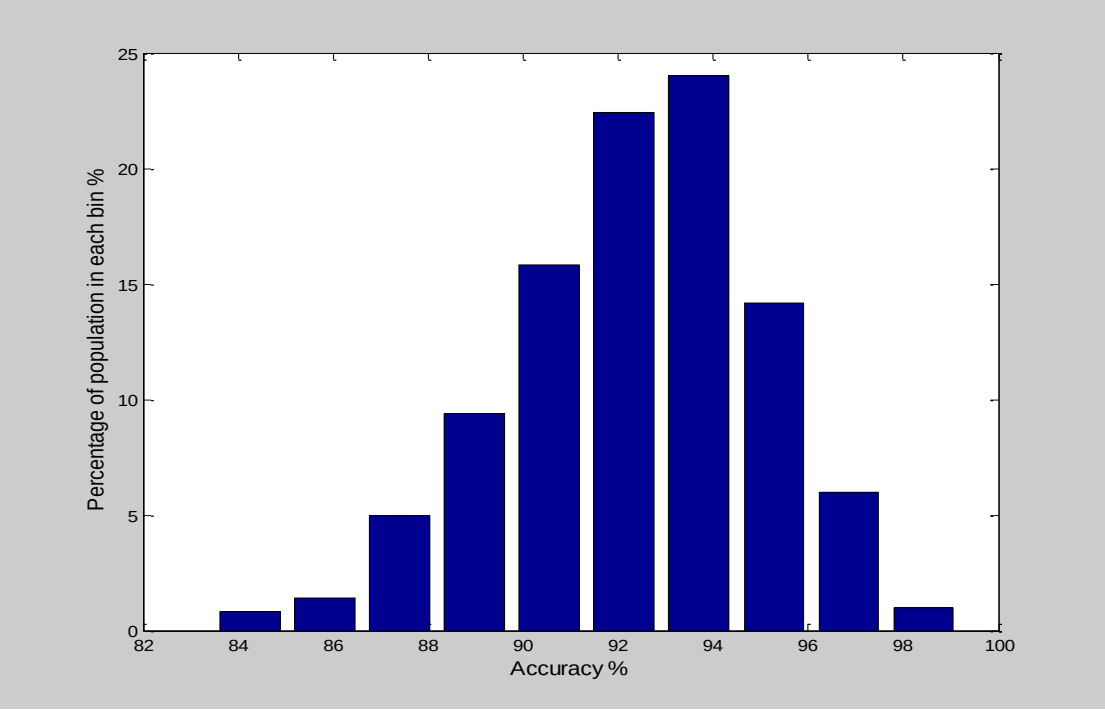


Figure 4.14 Illustration of histogram of training accuracy for 500 Monte Carlo runs. The graph represents percentage of accuracy values in each of the 10 equally divided accuracy bins.

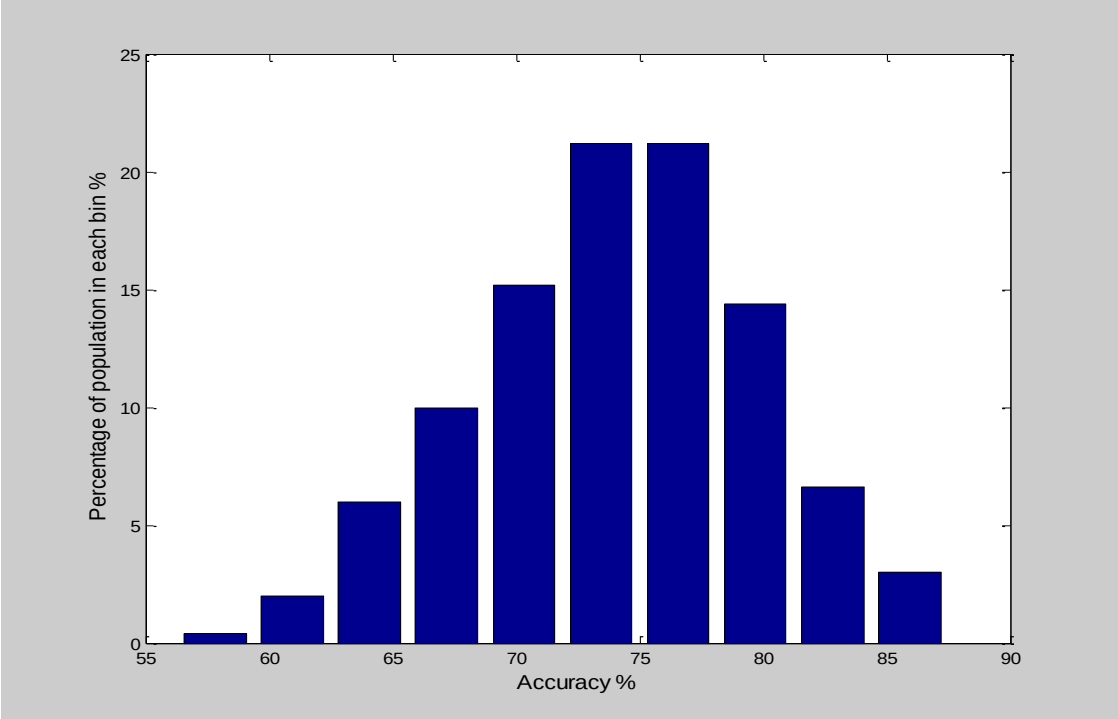


Figure 4.15 Illustration of histogram of testing accuracy for 500 Monte Carlo runs. The graph represents percentage of accuracy values in each of the 10 equally divided accuracy bins.

4.3. Comparison between different optimization algorithms

A summary of the testing performance results of the Monte Carlo simulation for the five optimization techniques is accounted in Table 4.6.

4.6 Summary of the sensitivity, specificity and accuracy mean and standard deviation (std) for testing sets after 500-run Monte-Carlo simulation for a the five discussed optimization techniques.

Optimization routine	Testing (Mean ±std)		
	Accuracy	Sensitivity	specificity
Manual1	75.98±4.59	81.20±7.16	69.67±8.03
Manual2	75.83±5.15	80.02±7.45	70.79±8.32
fminbnd1	75.51±5.16	80.36±7.13	69.66±9.15
fminbnd2	76.66±4.81	81.84±6.75	70.41±8.98
fminsearch	74.74±5.61	78.57±8.25	70.12±8.74

CHAPTER 5

DISCUSSION

This chapter contains the discussion of the different results obtained in this investigation. Section 5.1 discusses the performance of SVM detection algorithm. Section 5.2 discusses the comparison of the results acquired for optimization. Section 5.3 discusses the comparison of the results with previous studies.

5.1 Performance of SVM Detection Algorithm

The performance of a SVM is analyzed in terms of training and testing performance. The training performance measures the ability of a SVM to learn through training, whereas the testing performance measures the capability of SVM to classify the unseen data using the parameters obtained during training phase. Hence, the optimization approaches presented in this study were employed to optimize the learning section so that the testing would improve. In the ultimate application of algorithm, its testing performance is of interest.

The performance of the detection algorithm is measured in terms of accuracy, sensitivity and specificity. As discussed in chapter 3, the sensitivity is referred to as the probability of a diagnostic test is positive, given that subject is apneic and specificity is referred to as the probability of a diagnostic test is negative, given that subject is normal. In this study, we are investigating the ability of SVM for the detection of OSA events. The results reported for different optimization techniques in chapter 4 reveals that the sensitivity of SVM is higher than its specificity. This implies that the algorithm is able to detect more of OSA clips correctly than NOR clips. This is result as we are investigating the detection of OSA clips using features extracted from ECG.

5.2 Comparison of Optimization results

This section compares the results acquired for different optimization routines. The comparison of results within Manual and computerized optimization is referred as intra optimization comparison and the comparison of results between Manual and computerized optimization is referred as inter optimization comparison.

5.2.1 Intra optimization results comparison

For manual optimization, the results reported in chapter 4 indicate that the training and testing performance obtained for both the Manual1 and Manual2 are somewhat similar. But Manual2 is able to achieve highest accuracy value of 97.63 % for training and 92.18 % for testing. Even though the overall performance of Manual1 to some extent better than Manual2 in the testing part, but Manual2 is able to achieve highest detection accuracy for both testing and training. The histograms for Manual1 and Manual2 reveal the distribution of accuracies across 500 Monte Carlo simulation runs. A histogram shows basic information about the data set, such as central location, width of spread, and shape. The testing accuracy histograms for both the optimization reveal that, the accuracy value for 65% of the 500 runs are populated near or above 76%, but the rest of runs have accuracy lower than 75% which drag the mean of the 500 runs to a value lower than 76%. This indicates that the Manual2 optimization is better compared to Manual1.

For computerized optimization, the results obtained suggested that, even though both Fminbnd1 and Fminsearch both performed well with the training data, Fminbnd2 performed exceptionally well with testing data. Histogram for training accuracies also revealed that the width of the distribution for Fminbnd1 and Fminsearch is compact, whereas it is widespread for Fminbnd2 often resulting in accuracy values less than 70%. The histogram of testing accuracies show that the accuracy of Fminbnd2 has about 65% of its population near or above 75% accuracy value, even though it has a single valve at 57% that had less impact on the final overall performance of the SVM. The 57% accuracy value was due abrupt termination of

optimization toolbox. The final performance of SVM is estimated by testing performance, which implies that Fminbnd2 has a higher capability of detecting OSA than the other computerized optimizations.

5.2.2 Inter Optimization Results Comparison

This section discusses the optimization results for Manual and Computerized optimization. In Manual optimization a mean accuracy of 75.98% is achieved for the 500 Monte Carlo runs. The histograms indicate that the maximum accuracy value achieved using Manual optimization is around 90%. Even though the accuracy values of computerized optimization never reached 90 %, the mean accuracy value achieved in Fminbnd2 is higher than that of Manual optimization. The highest average sensitivity, the ability to detect OSA was higher for Fminbnd2.

It is evident from the section 4.3 that optimization of C and n using Fminbnd2 gives an advantage over other optimization techniques used in this study. The average accuracy value of 76.66 % is achieved over the 500 Monte Carlo runs. Where the sensitivity achieved using the optimization is 81.84%. The OSA detection capability is higher to Fminbnd2 compared to other optimization routines.

The results obtained from all the optimization techniques indicate that the detection algorithm is able to learn well with training set, but is not able to replicate the same results with the testing set as expected. There may be several reasons for deterioration of testing accuracies.

In the study [10], Sanjee reported that the changes in ECG are evident few seconds before and after the onset of the apneic events. If the event occurs at the end of the 15-minute epoch, there may not be any change in the trend or if the event onset is just after the 15 minute epoch, there may be some changes reflected in the cluster trend. Therefore, a centroid extracted from normal clip of fifteen minutes duration can be classified as an OSA clip, if there are some events just after the clips end. Also, a centroid extracted from a cluster of OSA clip is classified as normal it has events at the end of the fifteen minute epochs. This may be the reason for false detection. Any missed R-peak detection may also induce false detection [47]. Furthermore, it is

reported in previous study [8], that changes that occur during Hypopnea are not completely reflected in RPE. This may be one of the reasons Hypopnea clip to be classified as NOR clip.

SVM is sensitive to data imbalances. In a data set of two classes, if there is unequal distribution of data points in two classes then the data set is called unbalanced data set. The dominant class tries to push the hyperplane away from it. This leads to the misclassification of the of the minority class. The data used in this is unbalanced, where OSA clips being the dominant class. Therefore, sensitivity of the training and testing is always greater than specificity.

The detection accuracy could be boosted, if the algorithm is trained for more subjects as achieved in [43].

5.3 Comparison with previous studies

A pilot study by Al-Abed M et al [44] used parameters extracted from cluster of RPE and R_R interval to classify OSA and NOR subjects. An accuracy of 97.5%, a sensitivity of 100% and a specificity of 95% was reported for classifying the OSA and NOR subjects in this study. A total of 9 parameters were extracted each cluster of RPE and R-R interval.

Our proposed method is superior in three ways. Firstly, only a single parameter extracted from each cluster of RPE and R-R interval is used for classification of OSA and NOR subjects as opposed to 9 parameters. Secondly, it does not require the use of neural networks, which are computationally complex. Thirdly, the total subjects used for classification purpose are 26 as compared to 14 in their study.

In the study conducted by Sanjee [10], an algorithm was proposed for the detection of normal, Chyene-Stokes Respiration (CSR) and OSA events in fifteen minute epochs. This resulted in detection rate rates of 87.8 %, 88.6% and 85.4% in the training set and 70.7%, 72.7%, and 87.5 % in the test set for normal, CSR and OSA events. It is evident from the results in the previous that our proposed algorithm has shown performance results comparable to the previous study.

In other study by de chazel et al [38], 64 features extracted from ECG were used to detect sleep apnea. They used a complex neural network as a classifier. The study reports sensitivity and specificity of 100% for classifying normal subjects from OSA subjects.

The algorithm used in thesis is superior in two ways. Only one feature is used for the detection purpose opposed to 64 features. The algorithm takes advantage of computationally simple SVM opposed to computational complex neural network used in their study.

This approach showed detection results comparable to the previous study [10]. The limitation of this study is that the detection results are low compared to the previous study [9, 38]. The lower detection results can be associated to, fewer features used for the detection, and the tolerance of the SVM to data imbalances.

CHAPTER 6

CONCLUSION AND FUTURE DIRECTIONS

6.1 Conclusion

The performance of the support vector machines in detecting the presence of OSA in overnight ECG recording is tested in this study. This approach combines both RPE and R-R interval to form a cluster. An optimum centroid is extracted from the cluster, and is used as an input to the SVM. The performance results showed small or no change across the different optimization techniques. For Manual optimization, a highest learning performance of, accuracy of 91.16%, sensitivity of 95.20% and specificity of 86.20% is achieved for training set and a highest testing performance of, accuracy of 75.98%, sensitivity of 81.20 %, and specificity of 69.87% is achieved for testing set. The Computerized optimization resulted in slightly higher performance than the Manual optimization. The highest learning performance achieved for training set is, accuracy of 92.78%, sensitivity of 96.33% and specificity of 88.43% and a highest testing performance of, accuracy of 76.66%, sensitivity of 81.84 %, and specificity of 70.41% is achieved for testing set. The detection rates achieved using SVM is comparable to the results achieved with previous study [10]. However, the detection rates are lower than the results acquired using some other study [9, 38]. The testing sensitivity achieved for both Manual and computerized optimization is more than 80%. This is indicative that the SVM is useful tool for initial screening purpose but may not be as useful for probable diagnostic purpose.

6.2 Future Directions

In this research, we concentrated on feasibility of conventional SVM for the detection of OSA in overnight ECG recoding. Future studies could concentrate on increasing the performance of the classifier using advancement in SVM such as recently developed v-SVM. The support vectors play an important role in the final solution of SVM. The generalization error for a support

vector is directly proportion to the number of support vectors as discussed in Appendix B. The lower the support vectors the higher the generalization ability of the SVM. v-SVM has the advantage of controlling the number of support vectors in advance, thus controlling the performance of the SVM. In general they are two different problems with the same optimal solution set. However, compared to conventional SVM, the dual formulation of v-SVM is more complicated, due to which the probable use of v-SVM for large scale SVM training is not possible [49, 50].

The major limitation of SVM is its sensitivity to the data imbalances. The future study should concentrate on correction techniques for data imbalances. Especially, the future study should be directed toward using an offset based method for the correction of the imbalances. This method is based on calculating offset for unbalanced support vectors resulting from unbalanced data [45].

Additional future directions could include:

- Use of additional features from ECG, such as changes in P wave which shows significant difference between normal breathing segments and OSA contaminated segments in each OSA patient [46].
- Use of information from multiple leads. Even though ECG from lead I showed highest sensitivity for the detection of OSA events, incorporation of ECG from additional leads could boost the detection rates. This information can be easily recorded and stored using Holter monitors at any point of time.
- Inclusion of some easy-to-measure parameters such as SaO2 would increase the performance of the classifier. Even though they show lower detection rates compare to other features, inclusion of them could further increase the detection rate.
- Use of Shorter data clips. Data clips used in this thesis are all 15 minutes long. In the OSA clip the apneic event could be recorded anywhere within the 15 minute clip. The detection rate could be further increased by using shorter interval clips which are

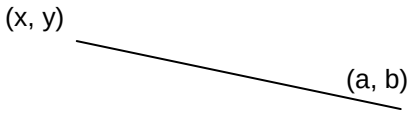
accurate indicative of the apneic events. Given the fact that the shortest apnea is 10 sec, the clip length should be around one minute long.

APPENDIX A

EUCLIDEAN DISTANCE

Euclidean Distance:

For the given two points in space with co-ordinates (x, y) and (a, b), the Euclidean distance d, between the two points is given by.



Euclidean distance: $d = \sqrt{(x-a)^2 + (y-b)^2}$ (1)

When Euclidean distance is used to obtain centroids, the resulting cluster centroids are mean of all the data points in a cluster. Where the mean of cluster minimizes the square distance given by equation

$$d\,(x_i, x_j\,) = \left(\sum_{k=1}^{d} \left|x_{i,k} - x_{j,k}\right|^2\right)^{\frac{1}{2}} \quad (2)$$

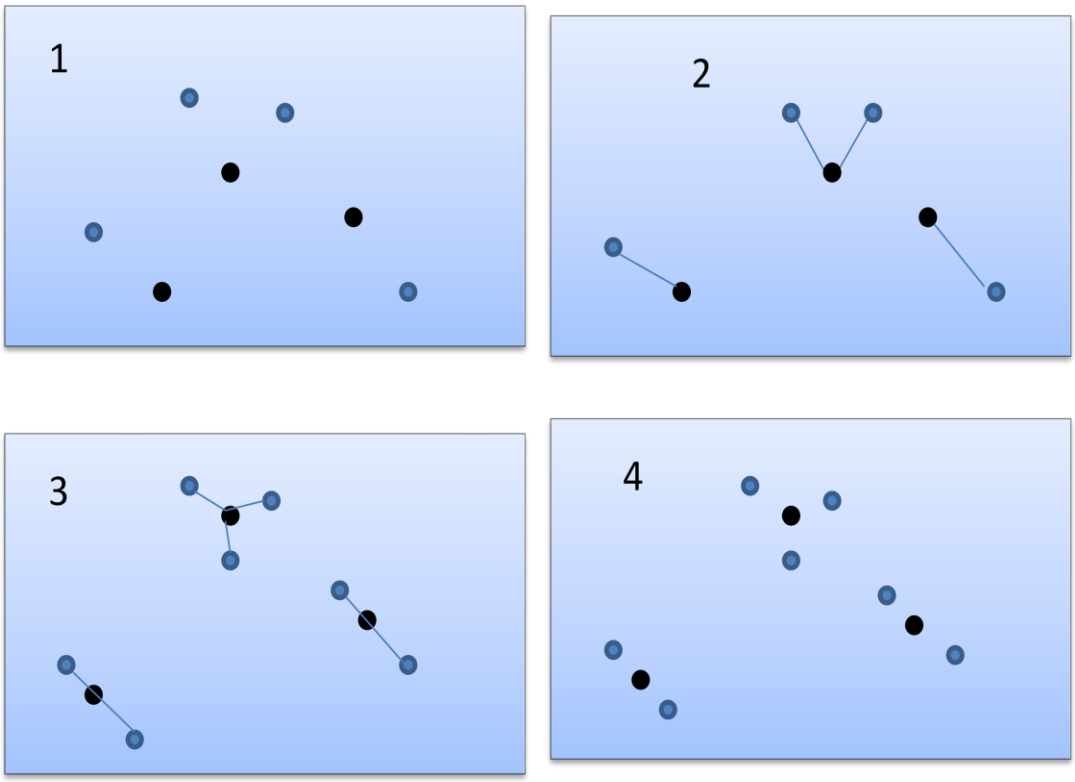


Figure A.1 K-Means Clustering with Euclidean distance. Black circles represent the centroids.

APPENDIX B

GENERALIZATION ABILITY OF SVM AND CONCEPT OF SOFT MARGIN

Generalization ability of SVM.

Lately SVM's attracted lot of interest due to their good generalization of performance. The studies conducted on generalization ability of SVM reported their good generalization performance [32]. Suppose the optimum hyperplane is found for a data, assume that N_s be the number of support vectors and *N* be the total number data points. It is widely proved in the literature that the expected out-of-sample error (the portion of unseen data that will be misclassified),π is bound by

$$\pi \leq \frac{N_S}{N-1} \qquad (1)$$

The following implies that simpler systems are better and for SVM's, fewer support vectors are compact, simpler and enough representation of the hyperplane. Hence, generalization performance should be better for SVM [33].

Soft Margin

To classify the data with minimum number of errors, otherwise cannot be completely separated by linear separating hyperplane the constraints of the primal form should be relaxed to some extent. The concept was first developed by Corinna Cortes and Vladimir Vapnik. For the purpose of relaxing the constraints a slack variable ξ_i was introduced to relax the constraints. As shown in the Figure 1 the training data points fall into three categories.

Case 1: The data points which fall outside the hyperplane and are classified correctly. These are completely in compliance with constraints in the primal problem defined above.

Case 2: The data points which fall inside the margin but on the same side as the respective class. They satisfy the following inequality.

$$0 \leq y_i(\mathrm{w}.x_i + \mathrm{b}) < 1 \qquad (1)$$

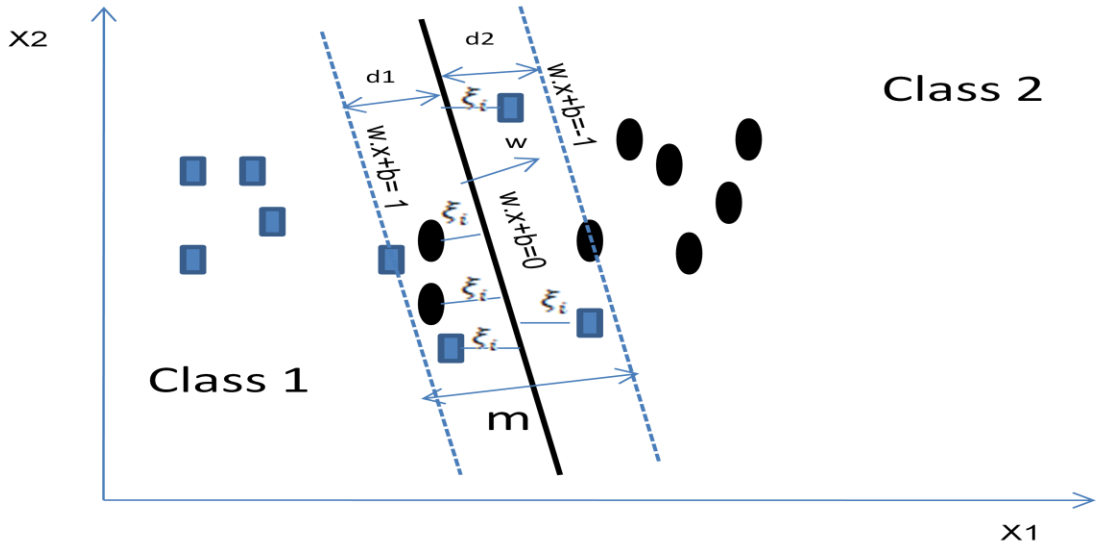


Figure B.1 Illustrates slack variables assigned to unclassified data.

Case 3: The data points which fall on the opposite side of the margin. These data points obey the following inequality

$$y_i(\mathrm{w}.x_i + \mathrm{b}) < 0 \quad (2)$$

All the three cases can be represented by using one equation. Now the constraints of the primal problem become

$$y_i(\mathrm{w}.x_i + \mathrm{b}) \geq 1 - \xi_i \quad (3)$$

Where $\xi_i > 0$ (4)

The value of the slack Variable ξ_i is different for each of the three cases discussed above. For the first case the value of slack variable is zero as the data points are in harmony with the constraints. The value of slack variable for the second case is in between zero and one i.e. $0<\xi_i \leq 1$. The value of slack variable is greater than one for the third case.

The new optimizing problem with additional slack variables to the constraints is more involved, yet it falls under the same rationale as described for the linear separable case. In addition, the underlying principle does not change much with respect to linear separable case with the addition of slack variables. The induction of slack variables into SVM revolutionized the way it

was used and made it possible to see it as a potential tool for pattern classification for almost all practical cases.

The primal problem is given by

Minimize $\frac{1}{2}\ \|w\|^2 + C\sum_{i=1}^{N}\xi_i$ (1)

Subject to $y_i(\mathrm{w}.x_i + \mathrm{b}) \geq 1 - \xi_i$, where *i*=1, 2, 3, 4.....*N* (2)

$\xi_i \geq 0$ where *i*=1, 2, 3, 4.....*N* (3)

Where C is a penalty imposed on the slack variables. C is an arbitrary constant. The above problem is again a convex programming, and the corresponding Lagrangian is given by

$$\mathcal{L}(w, b, \lambda, \mu) = \frac{1}{2}\ \|w\|^2 + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\mu_i\xi_i - \sum_{i=1}^{N}\lambda_i\,[y_i(wx_i + b) - 1 + \xi_i] \quad (4)$$

The corresponding KKT conditions are given by

$$\frac{d}{dw}\mathcal{L}(w, b, \lambda, \mu) = 0 \quad (5)$$

$$\frac{d}{db}\mathcal{L}(w, b, \lambda, \mu) = 0 \quad (6)$$

$$\frac{d}{d\xi_i}\mathcal{L}(w, b, \lambda, \mu) = 0 \quad (7)$$

$\lambda_i \geq 0$, *i*=1, 2… N (8)

$\mu_i \geq 0$, *i*=1, 2… N (9)

$\xi_i\,\mu_i \geq 0$, *i*=1, 2… N (10)

$$\lambda_i[(y_i(\mathrm{w}.x_i + \mathrm{b}) - 1 + \xi_i)] = 0 \quad (11)$$

Taking derivatives with respect to w, b and ξ_i

$$w - \sum_{i=1}^{N}\lambda_i\ y_i x_i = 0 \quad (12)$$

$$w = \sum_{i=1}^{N}\lambda_i\ y_i x_i \quad (13)$$

$$\sum_{i=1}^{N}\lambda_i\ y_i = 0 \quad (14)$$

$$C - \lambda_i - \mu_i = 0 \quad (15)$$

The associate Wolfe dual representation is given by

Maximize $\mathcal{L}(w, b, \lambda, \mu)$ (16)

Subject to $w = \sum_{i=1}^{N} \lambda_i \, y_i x_i$ (17)

$$\sum_{i=1}^{N} \lambda_i \, y_i = 0 \quad (18)$$

$$C - \lambda_i - \mu_i = 0 \qquad i=1, 2\ldots N \quad (19)$$

$$\lambda_i \geq 0, \qquad i=1, 2\ldots N \quad (20)$$

$$\mu_i \geq 0, \qquad i=1, 2\ldots N \quad (21)$$

Substituting the above equality constraints into Lagrangian dual we get,

$$\max_{\lambda}\left(\sum_{i=1}^{N} \lambda_i - \frac{1}{2}\sum_{i,j}^{N} \lambda_i \lambda_j y_i y_j x_i^T x_j\right) \quad (22)$$

Subject to $0 \leq \lambda_i \leq C$ $i=1, 2\ldots N$ (23)

$$\sum_{i=1}^{N} \lambda_i \, y_i = 0$$

APPENDIX C

HOLD OUT AND LEAVE ONE OUT CROSS VALIDATION

Holdout Method

In this method the data is completely divided into two sets one for training and other for testing. A function approximator is used to fit a function to the training data set. Then the function approximator is used to predict the output of the unseen testing data. This is repeated many times and results obtained are averaged to give the reliability of the system.

Leave-one-out cross validation

The data set is divided into n-subsets, where n is the number of data points in the data set. Each time, one subset is set aside for testing and remaining n-1 is used for training. The training parameters obtained are used to test the testing subset. The errors are averaged over all of the n trails.

APPENDIX D

SAMPLE PLOTS

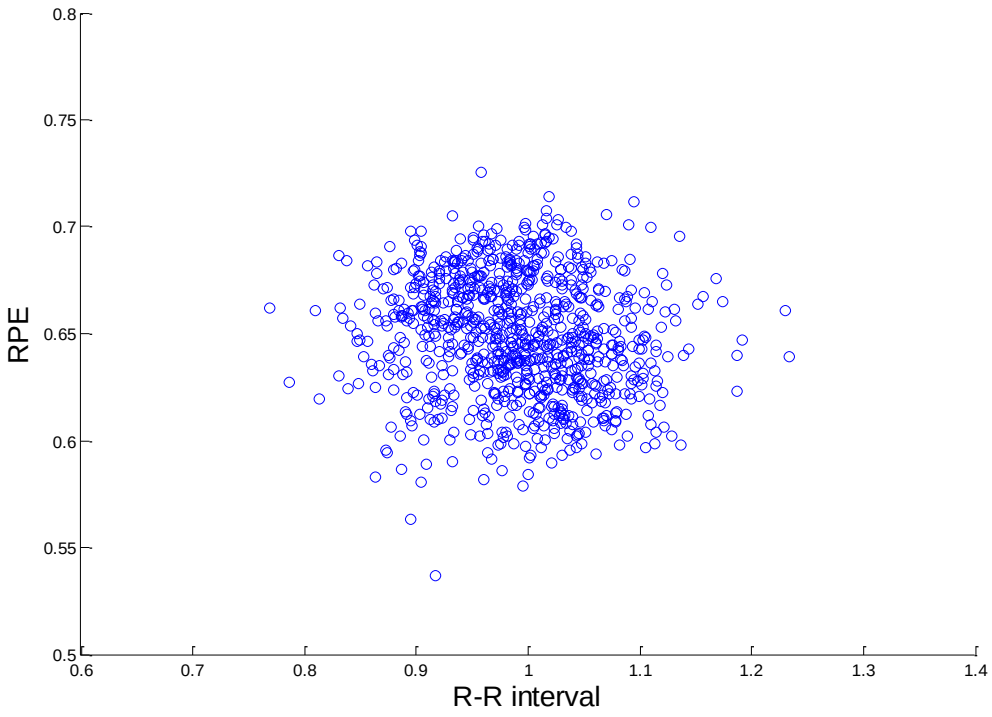


Figure D.1 A scatter plot of RR vs. RPE signals for 899 points representing 15 min of overnight ECG LI recordings for a normal clip.

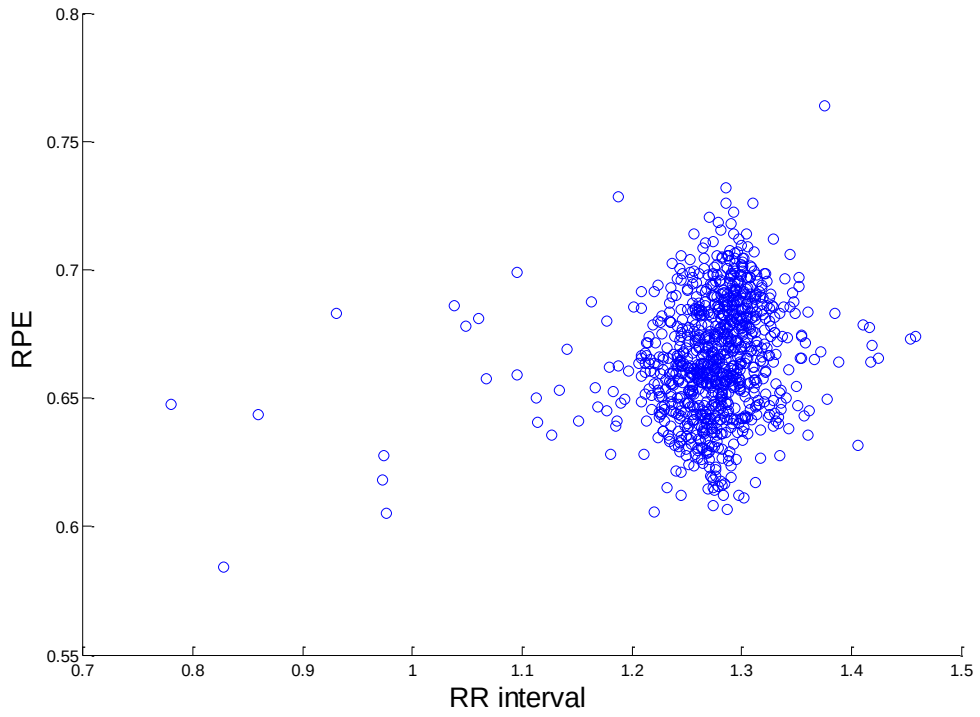


Figure D.2 A scatter plot of RR vs. RPE signals for 899 points representing 15 min of overnight ECG LI recordings for an OSA clip.

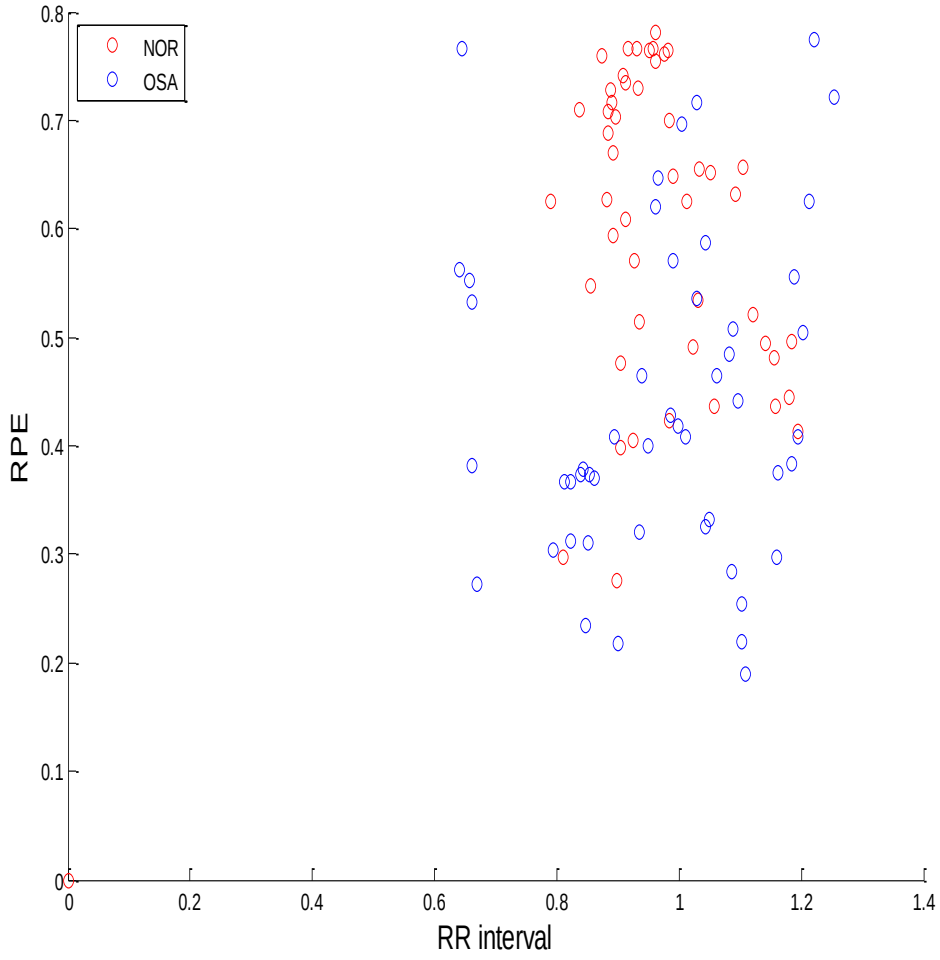


Figure D.3 A sample distribution of centroids extracted from a scatter plot of RR interval vs RPE from both OSA and normal clips.

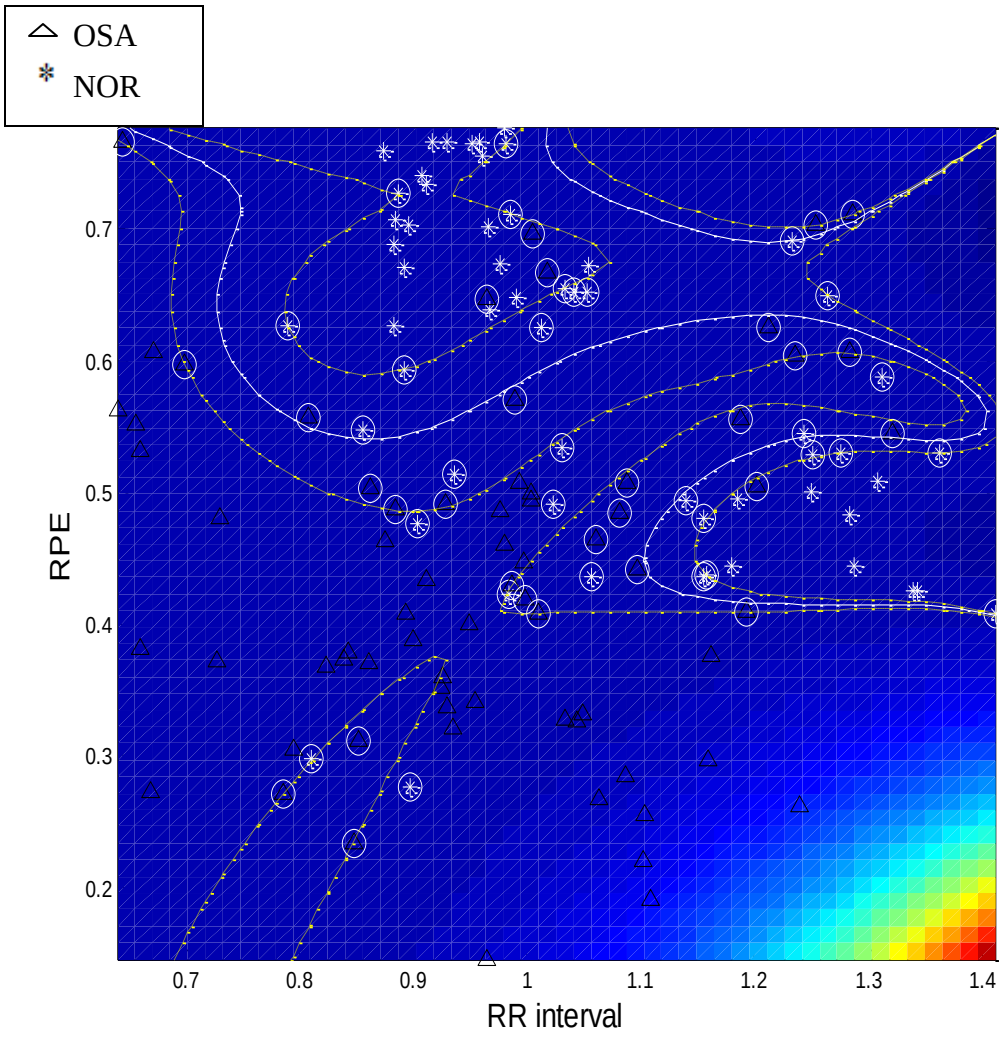


Figure D.4 A support vector classification of centroids extracted from scatter plot of RR interval vs RPE for the purpose of training.

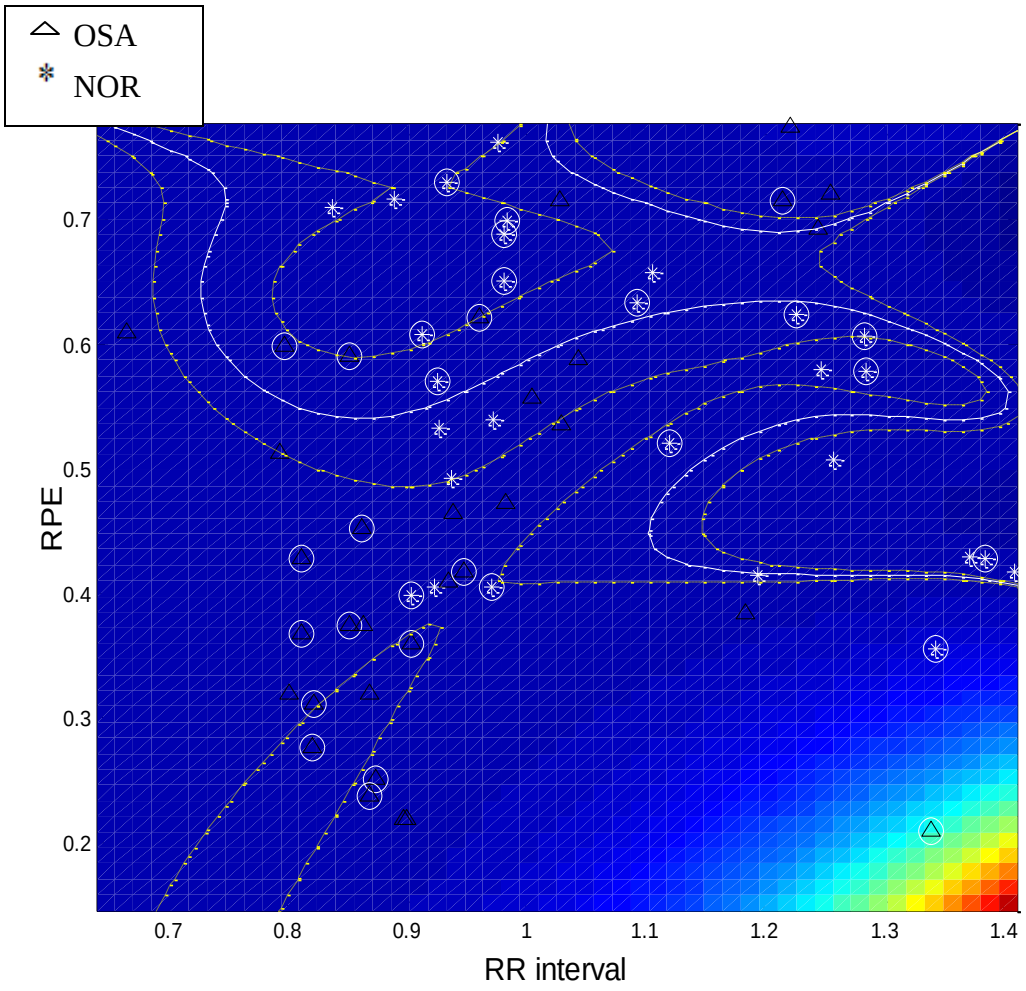


Figure D.5 An optimized SVM classifier hyperplane for the purpose of testing unseen data.

REFERENCES

[1].American Association of Oral and maxillofacial surgeons “Snoring and sleep apnea“ http://www.aaoms.org/sleep_apnea.php. Date visited 2011 May 10.

[2]. Donald S. Silverberg, M.D., Adrian Iaina, M.D., Arie Oksenberg, PhD, “Treating Obstructive Sleep Apnea Improves Essential Hypertension and Quality of Life” http://www.aafp.org/afp/2002/0115/p229.html. Date Visited 02011 May 12.

[3]. “http://sleep-apnea.emedtv.com/sleep-apnea/symptoms-of-sleep-apnea.html”. Date visited 2011 May 15.

[4]. McNames, J.N, Fraser, A.M, "Obstructive sleep apnea classification based on spectrogram patterns in the electrocardiogram," Computers in Cardiology *2000* , vol., no., pp.749-752, 2000

[5]. Quiceno-Manrique, A.F, Alonso-Hernandez, J.B, Travieso-Gonzalez, C.M, Ferrer-Ballester, M.A, Castellanos-Dominguez, G, "Detection of obstructive sleep apnea in ECG recordings using time-frequency distributions and dynamic features," Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE, vol., no., pp.5559-5562, 3-6 Sept. 2009

[6]. Redmond B. Shouldice, Louise M. O'Brien, Ciara O'Brien, Philip de Chazal, David Gozal, and Conor Heneghan “Detection of obstructive sleep apnea in pediatric subjects using surface lead electrocardiogram features”. *Sleep.* 2004 June 15; 27(4)**:** 784–792**.**

[7]. T. Penzel, J. McNames, P. de Chazel, B. Raymond, A. Murray, and G.Moody, “Systematic Comparison of Different Algorithms for Apnea Detection based on Electrocardiogram Recordings”, Medical and Biological Engineering and Computing, 40, pp. 402-407, 2002.

[8]. Vijendra. S, “An Investigation in the Detection of Sleep-Disordered Breathing using the Electrocardiogram”, Master’s Thesis, Dept of Biomedical Engineering, University of Texas at Arlington, Arlington, TX, USA, 2003.

[9]. Al-abed.M, “A Novel approach in the Detection of Obstructive sleep apnea from electrocardiogram signals using neural network classification of textural features extracted from time-frequency plots”, Master’s Thesis, Dept of Biomedical Engineering, University of Texas at Arlington, Arlington, TX, USA, 2006.

[10]. Sanjee. R, “An Technique for the Detection of Cheyne Stokes Breathing and Obstructive Sleep Apnea using Electrocardiogram”, Master’s Thesis, Dept of Biomedical Engineering, University of Texas at Arlington, Arlington, TX, USA,2005.

[11]. Terry Young, “Rationale, design and findings from the Wisconsin Sleep Cohort Study: Toward understanding the total societal burden of sleep disordered breathing” Sleep Med Clinic, 2009 March 1; 4(1): 37–46.

[12]. “Overview of the findings of the national commission on sleep disorders research (1992)”. http://www.stanford.edu/~dement/overview-ncsdr.html. Date visited 2011 june 27.

[13]. Quiet sleep, “Sleep apnea” Sleep apnea http://www.quietsleep.com/. 2011 May 15.

[14]. Timothy I. Morgenthaler, Vadim Kagramanov, Viktor Hanak, Paul A. Decker, “Complex Sleep Apnea Syndrome: Is It a Unique Clinical Syndrome?” Pub Med center 2006 September 4: Sleep Volume 29, issue 09

[15]. PatrickJ.Burns, “Sleep Apnea and the Cardiovascular system”. http://www.acponline.org/about_acp/chapters/mt/mtg08_burns.pdf. 2011 May 15.

[16].Dustin Boswell: Papers & project I have written “Introduction to Support Vector Machines” http://dustwell.com/PastWork/IntroToSVM.pdf. 2011 May 16.

[17]. National Heart Lung and Blood Institute, “Sleep apnea” http://www.nhlbi.nih.gov/health/dci/Diseases/SleepApnea/SleepApnea_WhatIs.html. Date visited 2011 May 17.

[18]. “Sleep apnea and obesity”: http://www.cimedicalcenter.com/Sleep-Apnea.html. Figure by Christy Kames (1999). Date visited 2011 May 20

[19]. Food and Drug Administration (FDA), “Types of Adult sleep Apnea” http://www.enotalone.com/article/7997.html.Date visited 2011 May 20.

[20]. Shochat T, Hadas N, Kerkhofs M, et al. “The SleepStrip: an apnea screener for the early detection of sleep apnea syndrome”. Eur Respir J 2002; 19: 121–126.

[21]. H. Nakano, T. Tanigawa, T. Furukawa and S. Nishima “Automatic detection of sleep-disordered breathing from a single-channel airflow record”. Eur Respir J 2007; 29: 728–736

[22]. Gandis G. Mazeika, MD, Rick Swanson, RPSGT, CRTT “Respiratory Inductance Plethysmography An Introduction”. www.pro-tech.com; Date Visited 2011 May 29.

[23]. New Choice Health “Sleep Study (Polysomnography) Procedure & Cost Information”. http://www.newchoicehealth.com/Directory/Procedure/51/Sleep%20Study%20(Polysomnography).Date visited 2011 May 29, 2011.

[24]. Martin O. Mendez∗, Anna Maria Bianchi, Thomas Penzel, Sergio Cerutti “Sleep Apnea Screening by Autoregressive Models from a Single ECG Lead” IEEE transactions on BIOMEDICAL ENGINEERING: December 2009, VOL. 56, NO. 12: 2838 - 2850

[25]. Vikram rao, “influence of hypoxia, hypercapnia and inspiratory efforts on sympathoexcitation during

Apnea” Master’s Thesis, Dept of Biomedical Engineering, University of Texas at Arlington, Arlington, TX, USA,2008.

[26] T. Douglas Bradley, “Obstructive Sleep Apnea and Heart Failure: Pathophysiologic and Therapeutic Implications” CardiologyRounds: www.cardiologyrounds.org. Date visited June 1, 2011.

[27]. white coat, ‘ECG’s” http://white-coat.tumblr.com/post/5618063042/ecgs. Date visited June 1, 2011

[28]. “A Tutorial on Clustering Algorithms”, http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/index.html. Date visited June 4 2011.

[29]. Benjamin J. Anderson, Deborah S. Gross, David R. Musicant, Anna M. Ritz, Thomas G. Smith, Leah E.Steinberg, “Generating Normalized Cluster Centers with KMedians” http://www.cs.brown.edu/%7Earitz/files/SDMpresentation.pdf. Date visited June 5 2011.

[30] Vapnik, Vladimir N. Statistical Learning Theory. New York, NY: John Wiley & Sons. 1998.

[31]. Sergios Theodoridis, Konstantinos Koutoumbas, Pattern Recognition Third Edition. USA: Elsevier. 2006

[32]. Marti A. Hearst, “Support Vector Machines”, IEEE intelligent systems vol 13, no. 4 (1998), P: 18-28

[33].Dustin Boswell, “Introduction to Support Vector Machines”, http://dustwell.com/PastWork/IntroToSVM.pdf. August 2006. Date visited 15 Feb 2011.

[34]. George B. Moody, Roger G. Mark, Andrea Zoccola, and Sara Mantero “Derivation of Respiratory Signals from Multi-lead ECGs” IEEE Computer Society Press; Computers in Cardiology *1985*, vol. 12, pp. 113-116.

[35]. C Maier, H Dickhaus, P Laguna, “Amplitude Variability Extraction from Multi-Lead Electrocardiograms for Improvement of Sleep Apnea recognition”, IEEE Computer Society Press; Computers in Cardiology 2005; 32:pp.355-358.

[36] H. Dickhaus and C. Maier, ‘‘Detection of sleep apnea episodes from multilead ECGs considering different physiological influences,’’ Methods Inf. Med., vol. 46, no. 2, pp. 216–221, 2007.

[37]. JE Mietus, C-K Peng, PCh Ivanov, AL Goldberger, “Detection of Obstructive Sleep Apnea from Cardiac Interbeat Interval Time Series”, IEEE Computer Society Press; Computers in Cardiology *2000*, vol. 27, pp. 753-756.

[38]. de Chazal P, Heneghan C, Sheridan E, Reilly R, Nolan P, O'Malley M.''Automatic classification of sleep apnea epochs using the electrocardiogram", IEEE Trans. Biomed. Eng. 2000; Vol. 50, no.6, pp.686-696.

[39]. Richard E. Klabunde, “Electrocardiogram Standard Limb Leads (Bipolar)”, http://www.cvphysiology.com/Arrhythmias/A013a.htm. Date visited 06/25/11.

[40]. Anne M. Gillis, Ricardo Stoohs, and Christian Guilleminault, “Changes in the QT Interval during Obstructive Sleep Apnea”. Sleep. 1991, 14(4): pages 346-350

[41]. Hyunsoo Kim, Peg Howland, Haesun Park, “Dimension Reduction in Text Classification with Support Vector Machines”. Journal of Machine Learning Research 6 (2005) 37–53.

[42]. Hidalgo, Sonia Sosa and E. Gomez Trevino, "Application of the kernel method to the inverse geosounding problem", Hugo Neural Networks, vol. 16, pp. 349-353, 2003.

[43]. Khandoker, A.H, Karmakar, C.K, Palaniswami M, "Screening obstructive sleep apnoea syndrome from electrocardiogram recordings using support vector machines," Computers in Cardiology, 2007, vol., no., pp.485-488, Sept. 30 2007-Oct. 3 2007

[44]. Al-Abed, Mohammad; Behbehani, Khosrow; Burk, John R.; Lucas, Edgar A.; Manry, Michael; "Cross correlation and scatter plots of the heart rate variability and R-peak Envelope as features in the detection of obstructive sleep apnea," Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE, vol., no., pp.3488-3491, 20-25

[45]. Boyang Li, Jinglu Hu, and Kotaro Hirasawa, “Support Vector Machine Classifier with WHM Offset for Unbalanced Data”. Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol.12, No.1 pp. 94-101, 20

[46]. Yachuan Pu, Patterson. R, Bornemann, M.C, "Detecting pediatric obstructive sleep apnea using ECG," Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE, vol.1, no., pp. 342- 345 Vol.1, 17-2

[47]. Suhas.S.R, Behbehani. K, Vijendra. S, Burk. J.R, Lucas.E.A, "Time domain analysis of R-wave attenuation envelope for sleep apnea detection," Engineering in Medicine and Biology Society, 2004. 26th Annual International Conference of the IEEE, vol.2, no., pp.3885-3888

[48]. Adrian Wills, “QPC”, School of Elec. Eng. & Comp. Sci., University of Newcastle, Callaghan, NSW, 2308, AUSTRALIA.

[49]. Chih-Chung Chang, Chih-Jen Lin “Training *v*-Support Vector Classifiers: Theory and Algorithms” Neural Computation, Volume 13, Issue 9, Page 2119-2147, September 1, 2001.

[50]. X. Wu and R. Srihari, “New ν-support vector machines and their sequential minimal optimization,” in ICML, 2003.

BIOGRAPHICAL INFORMATION

Harshan Ravi was born in a Hyderabad, India also known as the city of pearls in the year 1986. He was good at cultural activities and leadership skills. His skills helped him to get a job in Tata consultancies Limited, India, as a software engineer, but his enthusiasm to pursue higher education, motivated him to join Master's Degree in Bioengineering at University of Texas at Arlington, 2008. He was awarded dean fellowship from 2008-2010 for his exceptional GRE score. His research interests include discrete signal processing, instrumentations and machine learning theory. He is planning on doing PhD in bioengineering and joined Dr.Hanzhang Lu lab at UT Southwestern.