An Automatic Sleep Apnea Analysis with Soft Computing Approaches

By

Yashar Maali

Submitted in partial fulfilment of the requirements for the
Doctor of Philosophy

Faculty of Engineering and Information Technology
UNIVERSITY OF TECHNOLOGY, SYDNEY

September, 2014

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Candidate

Acknowledgment

Over the last four years I have had the privilege of working with a variety of people who have made my time at University of technology, Sydney an enjoyable and intellectually stimulating experience.

I would like to thank all the people who have helped me along the way and contributed to this dissertation. I am especially grateful to my Supervisor Associate Professor **Dr. Adel Al-Jumaily**. In working with Adel, I have learned how to pursue research problems with intellectual rigor and how to critically evaluate my work.

I would like to present my appreciations to the **Dr. Leon Laks** and Concord Repatriation General Hospital for providing us the data and useful advises during my study.

I would like to thank my friends and colleges who have been encouraged and support me to accomplish my degrees.

Finally, I would like to acknowledge the endless love of my family; who have been constant source of support also who have provided guidance, love and encouragement throughout my life.

Table of Contents

List of Figures

Chapter 2

Chapter 3

Chapter 4

Chapter 6

List of Tables

List of Abbreviations

Adaptive Boosting (AdaBoost)

Adaptive Resonance Theory (ART)

Apnea-hyponea Index (AHI)

Approximate Entropy (ApEn)

Artificial Neural Networks (ANNs)

Automatic CPAP (A-CPAP)

Autonomic Nervous System (ANS)

Backpropagation (BP)

Body Mass Index (BMI)

Bootstrap Aggregation Learning (Bagging)

Central Sleep Apnea (CSA)

Central tendency measure (CTM)

Chest volumes (CV)

Continuous Positive Airway Pressure (CPAP)

Discrete Wavelet Transform (DWT)

Dissolved oxygen (DO)

Electrocardiogram (ECG)

Electroencephalogram (EEG)

Electromyogram (EMG)

Electrooculogram (EOG)

Fast Fourier Transform (FFT)

Fixed CPAP (F-CPAP)

Flow-based auto-CPAP (f-APAP)

Heart rate variability (HRV)

Heart rates (HR)

Hierarchical Multi Master PSO (HMM-PSO)

Intra Class Correlation (ICC)

K-Nearest-Neighbor (KNN)

LArge Memory STorage and Retrieval (LAMSTAR)

Least Squares Support Vector Machine (LS-SVM)

Mixed Sleep Apnea (MSA)

Nocturnal airway-patency appliance (NAPA)

Non Rapid Eye Movement (NREM)

Obstructive Sleep Apnea (OSA)

Oral Appliances (OAs)

Oxygen Saturation (SO2)

Particle Swarm Optimization (PSO)

Photopletismography (PPG)

Polysomnograms (PSG)

Power spectral density (PSD)

Pressure relief is a continuous positive airway pressure (PR-APAP)

Principal Component Regression (PCR)

Probabilistic ANN (PNN)

Quadratic Discriminate Analysis (QDA)

Rapid Eye Movement (REM)

Reasoning Unit (RU)

Receiver operating characteristic (ROC)

Recursive Feature Elimination (RFE)

Respiratory disturbance index (RDI)

Sleep Apnea (SA)

Support Vector Machine (SVM)

United State Dollar (USD)

Abstract

Sleep Apnea (SA) is a common disorder without “age-specific” that affects approximately 2% of women and 4% of men; sleep apnea is characterized by repetitive cessation of breathing during sleep. The consequences of the sleep apnea include daytime sleepiness, impaired cognitive function, impaired memory, neurocognitive dysfunction, and development of cardiovascular disorders, metabolic dysfunction, and impaired quality of life. This thesis investigates the automated detection and prediction of sleep apnea. Many researchers have concentrated on automated detection of sleep apnea, but not much comprehensive or well-ordered work has been done on signal and feature selection or on predicting of the sleep apnea.

The objective is to find the best set of signals as input and the best set of features from selected signals that can be used by a machine learning approaches to study sleep apnea. The best set here is not only refers to a smallest set of signals with a good performance in sleep apnea analysis but also consideration for a set of signals that can be easily acquired from patients.

During the course of this thesis, several algorithms were developed. These algorithms can be used in sleep apnea studies or in wider machine learning areas. The most important contributions of this thesis can be summarized as below:

-Developing a new signal segmentation algorithm designed specifically for sleep apnea by attention to its properties. This algorithm chose times windows with a greater probability of containing at least one sleep apnea event. After that these segmentations are generated, they should be reviewed by the machine learning approaches to be classified as sleep apnea or normal.

-Developing a novel Support Vector Machine (SVM)-based approach named Self-Advising Support Vector Machine (SA-SVM) that transfers more knowledge from the training phase of SVM to the test phase. This idea helps SVM to learn from misclassified data in training phase and use this gained knowledge, in the testing phase. This approach can be used in any binary classification problems and it shows also high impact in sleep apnea detection.

-Developing a new parallel structure for Particle Swarm Optimisation (PSO). Finding the best set of input signals or the best set of features required a huge amount of computation power which a single PSO – or other optimisation approaches- cannot deal with, so a new hierarchical multi-master structure for parallel PSO was developed in this thesis, which quickly revealed its advantages over previous parallel PSO structures.

In this thesis real data has been used from Concord Repatriation General Hospital in Sydney. Obtained result shows a good performance in detection and classification of sleep apnea. Together with detection and classification, a prediction of sleep apnea was also considered. The prediction stage examines some famous neural networks structures and demonstrated how to improve the final result by taking advantage of multi neural network approach.

Chapter 1

Introduction

In this chapter motivations for this thesis are presented, including an outlines of the problem to be addressed that will be expanded upon later in the literature review chapter. After that the thesis questions will be stated, including the reasons for their importance, followed by a chapter-by-chapter synopsis of the thesis contents.

1.1. Why Study Sleep Apnea?

This thesis is about studying sleep apnea detection and prediction. Sleep can be defined on the basis of both the behavior of the person while asleep and related physiological changes that occur to the waking brain's electrical rhythms in sleep. By using bio signals and measuring the physiological behaviour sleep can be classified into two states: non rapid eye movement (NREM) and rapid eye movement (REM) sleep [1]. Physiological changes that occur during wakefulness and NREM and REM can also be classified in to many stages [1]. There are many interesting aspects in the classification of sleep stages and several studies have already studied this area [2-4].

As with studying normal sleep, the study of sleep disorders is important because they are common problem in a general population. As an example, a survey in 1987 [5] reported that at least one symptom of disturbed sleep was present in 41% of all subjects, and sleep disorder is still common now [6]; for instance, Young reported that 1 daytime sleepiness in 5 adults in 2004[7].

Sleep disorders are important not only because they are so common but also because they have several short term and long term side effects [8]. The short term effect leads

to impaired attention and concentration, impaired quality of life, increased rates of absenteeism with reduced productivity and accidents at work, at home, or on the road. The long term consequences of sleep deprivation include increased morbidity and mortality from increasing automobile accidents, coronary artery disease, heart failure, high blood pressure, obesity, type 2 diabetes mellitus, stroke and memory impairment, as well as depression. The long term consequences, however, remain open [1].

Sleep apnea is one of the most common of sleep disorders. Sleep apnea (SA) is characterised by a repeated, temporary cessation of breathing during sleep [9]. Clinically, apnea is defined as the complete or near-total absence of airflow for more than 10 seconds in adults.

The prevalence of SA, is approximately 2% in women and 4% in men whose ages are between 30 and 60 years [10]. It has been reported that in individuals with SA, throughout the night there can be 5–15 episodes per hour for mild cases, and more than 30 episodes per hour for severe cases [11].

1.2. Thesis Motivations?

In this section objectives and goals together with contribution of this thesis are reviewed.

1.2.1. Thesis Questions

The sleep apnea affects many people and the main motivation behind this thesis, is to answer the question " by using minimum recording signals, is it possible to automatically detect sleep apnea accurately as it happen and to predict it before it happen".

This thesis studies the automated detection and prediction of sleep apnea with a machine learning approaches. A machine learning approaches, such as artificial neural networks or support vector machines have had several empirical successes [12-16], of-

ten outperforming other learning methods in a variety of tasks. Moreover, several machine learning methods with different inputs were also investigated in the area of sleep apnea [17-20].

As such, the primary thesis question is “what is the best set of input signals and features for studying sleep apnea?” this question will be examined to discover the best input signals for detecting sleep apnea. This question cannot be separated from another primary question, “What is the best machine learning algorithm for studying sleep apnea?” While these questions are important, all the inputs for sleep studies or all the machine learning methods cannot be considered. Therefore for detecting sleep apnea, common input signals will be examined to select the most important ones by using support vector machine as the classifier.

A third question is, “Can the selected signals predict sleep apnea?". Just few studies investigated prediction of sleep apnea. In this thesis the same selected input signals from the detection phase will be used.

1.2.2. Contributions of the Thesis

The main contribution of the thesis is a better understanding of the detection and prediction of sleep apnea; furthermore some parts of this thesis have contributed to a wider range of machine learning studies. All of these will be discussed in more detail in the related chapters.

Contributions to the Understanding and Detection of Sleep Apnea

- Proposing novel signal segmentation for sleep apnea detection and demonstrating the importance of signal segmentation (chapter 4).
- Finding best signals for sleep apnea detection (chapter 4).
- Study on predicting sleep apnea (chapter 5)

Contributions in a Wider Scope

- Presenting a review on sleep apnea and literature on the machine learning approaches used in the sleep apnea studies (chapter 2 and 3).
- Proposing two dimensionality reduction approaches (Chapter 4)
- Proposing self- advising support vector machine (chapter 6)

Proposing new structure for parallel Particle Swarm Optimization (PSO) (chapter 6)

1.2.3. Structure of the Thesis

In **chapter 2** some general aspects of sleep apnea are introduced. Of course, every detail of sleep apnea cannot be considered, but this chapter gives some essential knowledge of sleep apnea studies for engineers. In section 2.2 different types of sleep apnea are introduced followed by risk factors in section 2.3. Symptoms of sleep apnea also considered in section 2.4 and the cost of this disorder is mentioned in section 2.5. Different methods for diagnosing sleep apnea are reviewed in section 2.6 and the available treatments are reviewed in section 2.7.

In **chapter 3** the literature relating to the application of machine learning methods in sleep studies are reviewed, and because there is a wide range of these methods three most common approaches are selected. In section 3.2 the basis of support vector machines and its applications in sleep apnea are reviewed, and in section 3.3 the basic and most important applications of artificial neural networks are introduced. Particle Swarm Optimization (PSO) is presented in section 3.4. Finally some performance measure indexes are reviewed in section 3.5.

In the first part of the **Chapter 4**, the proposed algorithms related to detecting and classifying of sleep apnea are presented. Signal segmentation is considered in section 4.2 and features generation is introduced in sections 4.3. Two approaches for dimen-

sional reduction are presented in 4.4. Finally an algorithm for signal selection is presented in 4.5. In the second part of this chapter, different experiments about proposed algorithms for detection of sleep apnea are presented followed by conclusion in section 4.7.

Prediction of sleep apnea is studied in the next chapter, **Chapter 5**. This chapter starts by a short introduction about previous works in section 5.1. Then prediction based on the multi Artificial Neural Networks (ANNs) included the proposed linear and non-linear multi ANNs are considered in sections 5.2. Different experiments are examined in section 5.3 for sleep apnea prediction. Finally, conclusions and summarised are given in section 5.4.

Chapter 6 generalized the two proposed algorithms to be used in other areas. In this chapter more information and experiments about the proposed algorithm to improve the support vector machine is introduced in section 6.2. This approach attempts to transfer more knowledge from the training phase to the testing phase. Furthermore, details of the proposed parallel structure for a particle swarm optimization approach together with general benchmarks are presented in section 6.3.

The last chapter, **chapter 7**, is related to the outcomes of this thesis followed by topics for future works.

1.2.4. Publications Resulting from the Thesis

Results of these investigations have been published in a number of papers, these peer reviewed publications including 3 Journals, 2 Lecture Notes in Computer Science (LNCS) chapters, and 6 international conferences.

Journal

1-Yashar Maali, Adel Al-Jumaily, Self-advising Support Vector Machines, International Journal of Knowledge-Based Systems, Volume 52, November 2013, Pages 214–222.

2-Yashar Maali, Adel Al-Jumaily, Automated Detecting and Classifying of Sleep Apnea Syndrome Based on Genetic- SVM , International Journal of Hybrid Intelligent Systems, Vol. 9, No. 4, pp. 203-210, 2012

3-Yashar Maali, Adel Al-Jumaily, Genetic Fuzzy Approach based Sleep Apnea/Hypopnea Detection, International Journal of Machine Learning and Computing, Vol. 2, No. 5, pp. 685-688, 2012

Book chapter

4-Yashar Maali, Adel Al-Jumaily , Hierarchical Parallel PSO-SVM Based Subject-Independent Sleep Apnea Classification, Lecture Notes in Computer Science (LNCS) on Neural Information Processing (ICONIP) Vol. 7666, Editors Huang, Tingwen and Zeng, Zhigang and Li, Chuandong and Leung, ChiSing, Springer Berlin Heidelberg, pp. 500-507, 2012.

5-Yashar Maali, Adel Al-Jumaily, Signal Selection for Sleep Apnea Classification, Lecture Notes in Computer Science (LNCS) on Neural Information Processing (AI 2012), Chapter 56, Editors H. Tingwen, Z. Zhigang, L. Chuandong and L. Chi Sing, Springer Berlin Heidelberg, pp. 661-671, 2012.

Conferences

6-Yashar Maali, Adel Al-Jumaily, Comparison of Neural Networks for Prediction of Sleep Apnea, International congress on Neurotechnology, Electronics and informatics, Algarve, Portugal, 2013.

7-Yashar Maali, Adel Al-Jumaily, Multi Neural Networks Investigation based Sleep Apnea Prediction, The 17th Asia Pacific Symposium of Intelligent and Evolutionary Systems (IES13), Seoul, Korea, 2013.

8-Yashar Maali, Adel Al-Jumaily and Leon Laks, Self-Advising SVM for Sleep Apnea Classification, Proceedings of the Workshop on New Trends of Computational Intelligence in Health Applications (CIHealth 2012), Sydney, Australia, Dec 4, 2012, pp24-33, CEUR-WS.org /Vol-944/cihealth3.pdf.

9-Yashar Maali, Adel Al-Jumaily, A Novel Partially Connected Cooperative Parallel PSO-SVM Algorithm: Study Based on Sleep Apnea Detection, IEEE World Congress on Computational Intelligence (IEEE WCCI 2012), Australia, 2012.

10-Yashar Maali, Adel Al- Jumaily, Automated Detecting Sleep Apnea Syndrome: A Novel System Based on Genetic SVM, 11th International Conference on Hybrid Intelligent Systems, Malaysia, 2011.

11-Yashar Maali, Adel Al-Jumaily, Genetic Fuzzy Approach for detecting Sleep Apnea/Hypopnea Syndrome, 3rd International Conference on Machine Learning and Computing (ICMLC 2011), Singapore, 2011.

Chapter 2

An overview on sleep apnea

2.1. Introduction

This chapter reviews different aspects of sleep apnea. It also presents general information about this syndrome as well as some information on the costs of sleep apnea, including diagnosing and methods of treatment. This review does not aim at listing what have already been published, however it aims at providing general details and concepts about the definitions and related issues of the sleep apnea syndrome.

2.2. Sleep Apnea

This section introduced one of the most important aspects of the sleep syndrome that is called sleep apnea (or sleep apnoea in British English). This is not a new topic, because it was first mentioned in 1837 [21]. Sleep apnea (SA) is characterised by a repeated, temporary cessation of breathing to the lungs during sleep [9]. Clinically, apnea is defined as the complete or near-total absence of airflow for more than 10 seconds in adults. Any decline in breathing signals becomes significant once the amplitude of these signals is reduced by at least around 75% with respect to normal respiration, and occurs for a period of 10 seconds or longer [22]. A hypopnea is an event of less intensity; it is defined as a reduction in baseline of the breathing signal amplitude around 30–50%, also lasting 10 seconds in adults [22].

In general, when a person becomes awake, except for temporary closures during swallowing and speaking, the upper airway remains open, permitting airflow to the lungs [23]. During sleep, the throat lumen may be physically obstructed several times [24] and lead to SA.

The prevalence of SA, is approximately 2% in women and 4% in men whose ages are between 30 and 60 years [10]. It has been reported that in individuals with SA, throughout the night there can be 5–15 episodes per hour for mild cases, and more than 30 episodes per hour for severe cases [11]. SA can be categorised as obstructive, central, and mixed [25].

2.2.1. Obstructive Sleep Apnea (OSA)

Obstructive apnea is the most frequent class of apnea. OSA is recognised by the presence of thoracic and abdominal efforts for continuing breathing while air flow completely stops. In purely obstructive apnea the upper airway closes naturally during inspiration, while subsequent efforts to breathe with the airway closed become larger and larger until either the effort or abnormal blood gases cause the person to wake up. When the airway opens, breathing resumes and blood gases are restored to normal, and the person falls asleep again, setting off another cycle. In obstructive apnea, movement of the chest wall can be observed but flow or nasal pressure tracing has flat tops in inspiration, Figure 2.1 shows a sample of OSA. The oxygen saturation curve is asymmetrical, with a slow decline and quick recovery, while the period of the apnea cycle is variable with the existence of snoring [26].

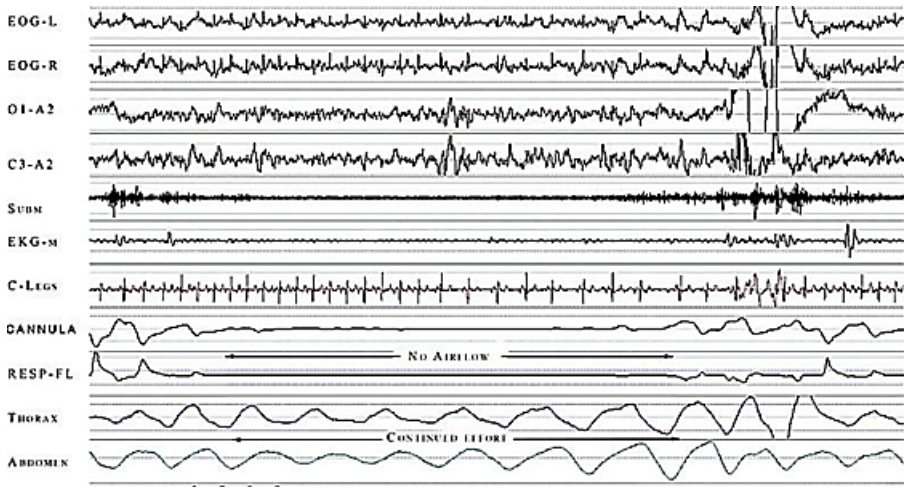

Figure 2.1: polysomnography of an obstructive apnea [27]

2.2.2. Central Sleep Apnea (CSA)

Patients with central apnea have wide open airways even when relaxed in sleep. In these cases periodic breathing results from an unstable negative feedback control system with a combination of high loop gain and a long delay between sensing a blood gas abnormality and compensating for it by adjusting ventilation. In central apnea, breathing effort can not be seen, oxygen saturation has a sinusoidal curve, and the periods of apnea cycles are constant and snoring is often absent [26], Figure 2.2 shows a sample of CSA.

2.2.3. Mixed Sleep Apnea (MSA)

This class of SA is a combination of the two previous ones and is defined by a central respiratory pause followed by an obstructive ventilator effort in a relatively short period of time. In this case the breathing control system is more sensitive to changes in oxygen or carbon dioxide so that obstructed efforts to breathe are greater and when the airway opens, ventilation is higher. Therefore, arterial carbon dioxide falls below normal before the person falls asleep. If it falls below the apneic threshold (the level at which breath-

ing stops in a sleeping person), respiratory efforts will be absent in the first part of the apnea until carbon dioxide rises above the threshold. Mixed apnea thus shows that periodic breathing in sleep is governed by an interaction between the behaviour of the upper airway and the characteristics of the chemoreceptor negative feedback control system [26].

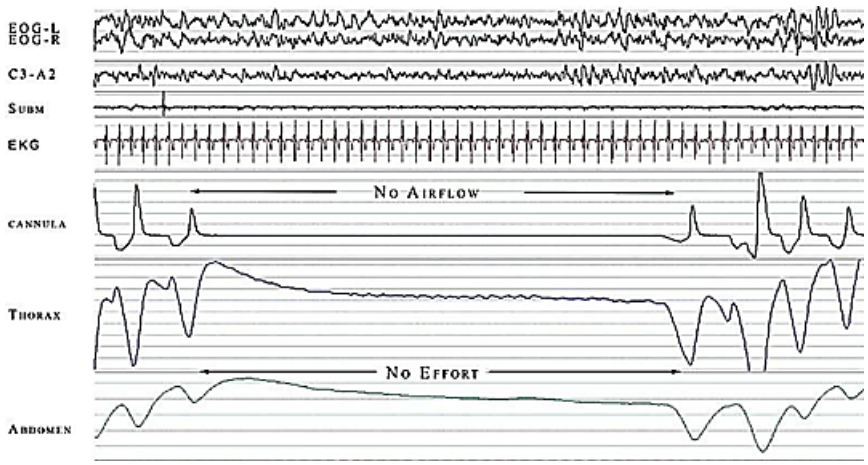

Figure 2.2: polysomnography of a central apnea [27]

The pathophysiologic mechanism which causes respiratory disturbance is the main difference between these categories. **CSA** involves dysfunction of the ventilator control in the central nervous system (loss of ventilator effort); in **OSA** upper airway obstruction is most frequently related to abnormal anatomy and/or abnormal control of the muscles that maintain the patency of the upper airway [28]. Although they are considered physiologically distinct, there is some overlap between the central and obstructive events. For example, some apneas may initially be central, with no evidence of inspiratory effort and after a variable period, respiratory effort commences, but the apnea continues because the upper airway collapses during the central component of the apnea. These events are referred to as mixed apneas. Furthermore, some patients with a central sleep apnea also have clinical features like snoring that are more typical of obstructive sleep apnea [29].

2.3. Risk Factors

Sin et.al [30] showed in their study that the risk factors of OSA and CSA are different and these risk factors are also different between men and women. Based on this study, atrial fibrillation is a risk factor for CSA but not for OSA, whereas hypocapnia is a risk factor for CSA in both sexes. This study also indicated that the single most important risk factor for OSA in men was increasing Body Mass Index (BMI), and the most important risk factor for OSA in females was increasing age. Conversely, increasing age was not a risk factor for OSA in men, nor was increasing BMI a risk factor for OSA in women.

The majority of OSA cases were obese and obesity is the main known risk factor for OSA [31-34]. Beside genetics, the sleep apnea syndrome has a strong familial tendency [35-37]. Various anatomical factors can result in physical obstruction of the airways [38] such as enlarged tonsils [39], enlarged uvula [40], increased tongue size [41] and abnormal craniofacial morphology [42]. These respiratory disorders may lead to hypoxia and hypercapnia, which can trigger arousal from sleep by increasing ventilator drive [43, 44]. Ageing is a factor that leads to conflicting opinions. Some works show a higher prevalence of OSA in older people [45, 46] in general, whereas Young et.al in their study showed that the prevalence of SA increases with age, with a 2- to 3-fold higher prevalence in patients aged over 65 [47]. However, some studies also showed that the respiratory disturbance index (RDI), the total number of apneas divided by the hours of sleep, depended on the BMI and was independent of age [48]. The effect of gender is another topic that has been analysed in several works [49-54]. These studies showed that OSA occurs more in males, and males with OSA were more likely to have symp-

toms of snoring [55, 56], while females with OSA had more symptoms of depression or morning headache [55, 57]. The effect of alcohol and smoking on sleep apnea was also considered in several works [58-64]. It was also noticed that men with a neck circumference of more than 17 inches, or 16 inches in women, can be a potential factor [65, 66].

2.4. Symptoms

Excessive daytime sleepiness is the most common complaint resulting from sleep apnea, with clinical features being a strong feeling of abnormal tiredness during the day, and reduced wake fullness and vigilance [67]. Anyhow the severity of sleepiness does not necessarily depend on the severity of sleep apnea, for example Engleman et al. [68] showed that a patient with an apnea-hyponea index (AHI) of 5-15 per hours of sleep may complain of severe daytime sleepiness but patients with more severe OSA may under report their sleepiness. Another symptom of sleep apnea is snoring [69], but snoring as a sole symptom is not a good predictor of OSA [70], although the absence of snoring makes the probability of OSA less likely [71]. The sleep Heart Health Study showed that snoring is associated with daytime sleepiness and can be independent of the AHI in middle aged and older adults [72]. Another symptom is heart failure, although compounding factors such as obesity, hypertension, and coronary heart disease make this relationship uncertain and an independent correlation remains unproven[73]. Also some other symptoms can be mentioned such as morning headaches, a limited attention span, memory loss, poor judgment, personality changes, and lethargy [74]. These symptoms can significantly decrease the quality of life and increase the risk of accidents [75, 76]. Finally, it should be noted that women and men generally have the same symptoms [55].

2.5. Costs of Sleep Apnea

It is difficult to speak about costs in the context of health care because the direct and indirect costs for individual patients and for society should be considered, and it is also more difficult to calculate economic parameters such as the cost-effectiveness ratios, cost-benefit ratios, and so on, because of the inherent difficulty of computing the utility of healthy or unhealthy people. In spite of these difficulties, several works analysing general health economics have been published, and also for specific diseases such as sleep apnea.

To compute the direct and indirect costs sleep apnea for patients, their family and relatives, and also for society, both in the undiagnosed phase as well as the diagnosing and treating phase should be considered.

Direct costs consist of payment to the physician, drugs and hospital admission. Indirect costs include absence from work, reduction in earning capacity, and accidents related to illness, etc. Kapur et al. [77] calculated the indirect cost of undiagnosed sleep apnea in the USA as 3.4 billion USD per year in 1992. Findley and Suratt in 2001 calculated that treating 500 sleep apnea patients for 3 years would prevent 180 serious crashes (105 with the driver at fault) and 36 injuries. This would save about 369, 000 USD in direct property damage and medical expenses, and 648 000 USD in lost wages, legal expenses, and the administrative costs of insurance companies and government. And the total savings for treating 500 patients for 3 years would exceed 1, 000, 000 USD [78]. It should be noted that treating 500 patients for 3 years would cost roughly less than 600,000 USD [79].

If direct medical costs is considered, a common Continuous Positive Airway Pressure (CPAP) device with a 5 year life span costs around 865 USD for 5 years (173 USD per year)[79]. Total direct medical costs in 2003 were in the range of 350 USD [80] in the USA per year. It should be noted that the cost of treating sleep apnea, unlike many other diseases, is independent of its level of severity [79].

Gamez et al. [81]considered home monitoring with hospital monitoring. This work showed that a home monitoring diagnostic test is 101.34 EURO less than polysomnography at a hospital, and patient satisfaction is significantly higher. They used the same sensors in home or hospital monitoring to ensure that the difference in result is only because of the geographical places. Moreover, another work by the American academy on sleep medicine compares polysomnography with home polygraphy and found that home polygraphy was around 32.30 EURO cheaper than polysomnography [82].

Finally, because untreated sleep apnea can increase the risk of morbidity and mortality, its treatment can result in significant short term and lifetime cost savings. For example, the chance of an SA patient staying in hospital is 1.6-fold [83] more than normal people, and OSA patients used approximately twice as many health care services as non-OSA patients [84].

2.6. Diagnostic

Unfortunately, because of a person's lack of awareness, sleep apnea may go undiagnosed for years [85, 86]. Indeed, a patient is often recognised their spouse, roommate, or family member who has witnessed the periods of apnea alternating with arousals, accompanied by loud snoring [69, 87]. Therefore, patients reporting symptoms of SA should be referred to a sleep centre for an overnight study where a polysomnograph is used. This is an integrated device comprising EEG, EMG, EOG, ECG, oxygen satura-

tion [88], airflow through the mouth and nose, thoracic and abdominal respiration measurement units, thoracic breathing movements, and the position of the body during sleep [11]. The respiratory disturbance index (RDI) and apnea-hyponea index (AHI), which holds the sum of apneas, hypopneas and respiratory arousals per hour during sleep have been standardised from overnight sleep studies. The RDI value is used to diagnose and grade the severity of sleep apnea, and the AHI is used to assess the severity of apnea according to the Chicago criteria. Based on the Chicago criteria, an AHI<5 is referred to as normal, an AHI =5–15 means mild, an AHI=15–30 corresponds to moderate, and an AHI >30 is referred to as severe (The Report of an American Academy of Sleep Medicine Task Force, 1999 [22]).

Sleep experts generally make their decisions based on the degree of AHI, such that [26]:

- If the AHI is less than 10 and the patient lacks important sleepiness or another problem is attributable to sleep apnea, then no further investigation and treatment for sleep apnea is needed. It should be noted that recommendations for weight loss or treatments for snoring may be desirable.
- If the AHI is less than 10 but the patient has important daytime symptoms, then their history of other causes or symptoms (e.g., sleep deprivation, insomnia, medication, and narcolepsy) should be examined and home monitoring should be considered together with polysomnography, and a test of daytime alertness or sleepiness, if available.
- If the AHI is above 10 and below 30 and the patient has no important daytime symptoms or comorbidity, then no further investigation and no treatment for OSA are needed. But recommendations for weight loss or treatment for snoring may be desirable.

- If the AHI is above 10 and below 30 and the patient has important daytime symptoms, then a trial with an auto-adjusting CPAP should be offered. If that succeeds, recommend CPAP and if it fails technically, consider titration during polysomnograms (PSG) in the laboratory. If the patient cannot tolerate CPAP, consider a conservative treatment such as a dental appliance, or surgery. If it succeeds technically but the patient does not find it valuable and has relatively mild symptoms, recommend conservative treatment or a dental appliance. If it succeeds technically but the patient has an important degree of sleepiness that does not improve, investigate for other causes of sleepiness with PSG.
- If the AHI is above 30, a trial with CPAP is recommended even if the patient does not report symptoms, and otherwise follows the same plan as for AHI between 10 and 30. If the patient is very overweight or has severe hypoxemia on the portable monitor, arterial blood gases should be checked and titration of CPAP during polysomnography is preferable. If the patient has severe symptoms or complications with sleep apnea and all else fails, consider a tracheostomy.
- It should be noted that the diagnostic methods are not limited to use of bio signals, because some studies use chemical biomarkers [89] or [90, 91]. Also, some non-electric signals were used in some papers such as the sound of breathing and/or snoring [92-104]; blood pressure [105-107]; airflow signal [108, 109]; pupil size [110]; videos from physical activities [111-114]; tracheal sound [115] and breathing signals (nasal flow, thorax movement, and abdomen movement) [116, 117].

Previously we mentioned to relation between thoracic and abdominal movements, and air flow with sleep apnea. In the rest of this section a study of various significant bio signals which were used in different studies as the main input in diagnosing sleep apnea is presented.

2.6.1. Oxygen Saturation

Oxygen saturation (SO2) or dissolved oxygen (DO) is one of the relative measures of the amount of oxygen that is dissolved or carried in a given medium. Oxygen saturation measures the percentage of hemoglobin binding sites in the bloodstream occupied by oxygen. There are several indexes related to SO2, such as SaO2 which is characterised by arterial oxyhemoglobin saturation measured by an arterial blood gas, while SpO2 which is characterised by arterial oxyhemoglobin saturation that measures non-invasively by pulse oximetry. Generally, there is a decrease in the oxygen saturation level [118, 119].

SaO2: The lack of airflow during apneic periods can lead to recurrent episodes of hypoxemia that can be detected on oximetry as fluctuations in the SaO2 records [120]. There are several works that use SaO2 to diagnose sleep apnea [121-128].

SpO2: Sleep apnea produces a drop in SpO2 which begins approximately 10 to 30 seconds after the apnea has begun. Shortly after hypoventilation ceases the SpO2 should begin to recover. Several papers highlight the rule of SpO2 in diagnosing sleep apnea [129-134].

2.6.2. Electrooculogram (EOG)

A measurement of the electrical activity of eye movements recorded using small metal discs called electrodes applied to the skin near the eyes is useful for monitoring the movement of eyeballs in REM and non-REM sleep. The two kinds of eye movements related to sleep classification are: 1) slow eye movements (SEM), rapid eye movements (REM), which occur during the wakefulness stage (although voluntarily) and the REM phase. The distinction between these two kinds of eye movements is made on the basis of the properties of synchrony, amplitude, and slope of the EOG signals. An analysis of

eye movements for each epoch was used in diagnosing and classify sleep disorders [135]. Some studies proposed that the apnea is longer during rapid eye movement (REM)[136-139] but another study reported more AHI in NREM than REM [140], while yet another study reported that respiratory distribution is not greatly affected by the sleep stage [141]. Using EOG signals to directly diagnose sleep apnea could be a good subject for further research; some existing studies in this area are [142-145].

2.6.3. Photopletismography (PPG)

PPG is an easily acquired measurement and provides a measure of the volume of tissue blood where the pulsatile component of the heartbeat is measured and the peripheral circulation is evaluated. This measurement is tie-related to arterial vasoconstriction or vasodilatation generated by the autonomic nervous system (ANS) and modulated by the heart cycle. When an apnea occurs, sympathetic activity increases as a response to the obstructive event in order to reestablish respiration. This increase in sympathetic activity is associated with vasoconstriction and is possibly related to transient arousal. Vasoconstriction is reflected in PPG by a decrease in the amplitude fluctuation signal [146, 147]. Amplitude reduction in PPG occurs when an apnea event takes place due to changes in the sympathovagal balance [148, 149]. However, other physiological events such as movement and deep inspiratory gasp produces a sympathetic activation and decrements in the PPG envelope amplitude which are unrelated to apnea [150]. There are several works related to the application of PPG signals in diagnosing sleep apnea [151-153].

2.6.4. Electrocardiogram (ECG)

An electrocardiogram measures the electrical activity of the heart and has a close relationship with the activity of the Autonomic Nervous System (ANS). An ECG has many

advantages, in that it can be easily measured in a non-invasive way and with a high signal to noise power ratio [154].

Sleep apnea is a respiratory event so its effects can be clearly observable within other peripheral systems such as the cardiovascular system. Due to this relationship, the electrocardiogram (ECG) can provide very valuable information about apnea events and has been broadly studied for the detection of apnea. One of the most important signals which can be obtained from an ECG is the beat-by-beat series of the heart rate. This signal contains fluctuations which are commonly named the heart rate variability (HRV), which present frequency components between 0 and 0.5 Hz and are linked to the Autonomic Nervous System (ANS) function. Frequency components between 0.15 and 0.5 Hz are generally associated with the vagal tone and are known as high frequency components (HF). Frequencies from 0.02 to 0.15 Hz are a manifestation of the activation of both parasympathetic and sympathetic systems, and are labelled low frequency components (LF). An increase in the LF power is generally associated with orthosympatic activation. Finally, frequencies between 0.0033 and 0.02 Hz contain information regarding slow processes such as thermoregulation [155]. The ratio between HF and LF spectral powers is defined as a measure of the sympatho-vagal balance [156]. During sleep, HRV presents specific dynamics [157] and a complexity that are characteristic of non-rapid eye movement sleep (NREM) and rapid eye movement (REM) sleep. The power spectral distribution of HRV signals shows the powers are concentrated around 0.3 Hz during NREM sleep, in contrast to REM sleep where the high frequency components are less peaked and the low frequency components are prevalent. However, when sleep apnea occurs, there is a reduction in the HRV complexity and the frequency components appear around 0.02 Hz as a result of repeating apnea [158]. There are several

papers that used the ECG signal as their main input in the detection of sleep apnea [20, 159-163].

2.6.5. Electromyogram (EMG)

The Electromyogram EMG signal (also known as the myoelectric signal) is a biomedical signal that measures electric currents generated in muscles during contraction, and represent neuromuscular activities [164]. The nervous system always controls the muscle activity (contraction/relaxation). Therefore, an EMG signal is a complicated signal which is controlled by the nervous system and depends on the anatomical and physiological properties of muscles.

The application of EMG signals in predicting sleep apnea appeared in several works, and in these studies the EMG signals were mainly extracted from the chin [142, 165] or tongue[166, 167].

2.6.6. Electroencephalogram (EEG)

An EEG signal is the record of electrical potentials generated by the cortex and deeper brain structures. During sleep apnea the arousals are characterised by abrupt changes in the EEG frequency (which is suggestive of an awakened state). The American Sleep Disorders Association (ASDA) has defined this arousal as "*An abrupt shift in EEG frequency, which may include theta, alpha, and/or frequencies greater than 16 Hz but not spindles*"[168]. Many significant works have studied the EEG signals in sleep apnea [169-171].

2.7. Treatment

Treatment of SA can range from conservative methods such as oral appliances [172], continuous positive airway pressure (CPAP)[173], to more radical approaches such as the surgical removal of anatomic obstructions [174-177].

2.7.1. Continuous Positive Airway Pressure (CPAP)

Continuous positive airway pressure (CPAP), was first used to treat obstructive sleep apnea patients by Professor Colin Sullivan of Sydney, Australia in 1981[178], and remains the main method for treating the obstructive sleep apnea syndrome. CPAP is a portable electronic device attached to a nasal mask via plastic tubing. CPAP prevents the upper airway from collapsing by putting a positive pressure in the pharynx during sleep. CPAP is a highly effective therapy, but it is not curative and the patients should use the CPAP mask regularly to significantly decrease the sleep fragmentation.

CPAP devices can be classified in different ways. For example, they can be classified based on their mechanism as: (1) automatic CPAP (A-CPAP) devices which automatically adjust pressure. (2) Fixed continuous positive airway pressure (F-CPAP) which requires an in-laboratory titration procedure to determine the effective pressure level (Peff) [179]. It should be noted that some studies argue there is no significant difference between these methods [180], while others argue that A-CPAP has a better compliance, better satisfaction, or increased patient preference [181]. Automatic CPAP also can be classified as flow-based (f-APAP) and vibration-based (v-APAP) machines. Flow-based auto-CPAP (f-APAP), works primarily by measuring the instant flow limitation at the mask with the aid of a pneumotachograph and vibration-based auto-CPAP (v-APAP), uses a pressure transducer to monitor the airway by vibration pattern [182]. Also, by at-

tention to different pressure sources CPAPs can be classified as Bubble CPAP and Ventilator CPAP [183]. The pressure generator for Bubble CPAP is a water bottle and a gas source is the generator for Ventilator CPAP. In some papers it appears that the pressure relief is a continuous positive airway pressure (PR-APAP) which is a recent innovation that attempts to overcome some disadvantages of CPAP by incorporating a reduction in airway pressure at the end of inspiration [184-187].

2.7.2. Oral Appliances

Oral appliances (OAs) are used to correct upper airway obstruction. OAs are now widely prescribed for the treatment of snoring and mild to moderate obstructive sleep apnea, both as primary therapy and as an alternative for patients who are unwilling or unable to tolerate CPAP. There are variety of synonyms for OAs and rather than oral, they may be called intra-oral, dental, or mandibular; and rather than being called an appliance, they may be called a device, splint, or prosthesis [188].

Dental appliances can be classified into three classes. One type of device is designed to reposition the tongue in a more forward position (tongue retaining device)[189]. This type of appliance increases the posterior airway space by holding the tongue away from the posterior pharyngeal wall. A second type of device positions the mandible forward, these are the nocturnal airway-patency appliance (NAPA)[190], the Snore Guard (Dental Sleep Disorder Prevention, Inc.)[191], Herbst [192, 193] and mandibular repositioner [194-196]. The basis for this second OA was based on the fact that the tongue is attached to the genial tubercles of the mandible, so positioning the mandible forward moves the tongue forward. Finally, the third type of dental device is designed to lift the soft palate or reposition the uvula [197]. These devices are used to reduce the vibration of the soft palate that causes snoring.

2.7.3. Invasive Methods

Generally there are two classes of invasive methods; the first is based on operation, and the second uses invasive micro stimulator.

Kuhlo et al. described the first surgical treatment for OSA in 1969 [198]. This treatment effectively eliminated OSA but it was poorly tolerated by patients. Fujita et al. [199] introduced the uvulopalatopharyngoplasty (UPPP) for OSA in 1981. After that, many modifications and variations of this technique were used, most of which are based on an operation on the tongue [200].

The first attempts to electrically stimulate the upper airway muscles were made in 1978 by Guilleminault et al. [142], but their efforts were considered to be a failure. Miki et al. reported their experience with genioglossus intramuscular stimulation [201] after which several attempts were made to activate the muscles of the tongue electrically to maintain airway patency. Transcutaneous stimulation of the tongue is technically awkward [202]. Intra-lingual stimulation of the tongue via percutaneous wires inserted deep and inferior to the frenulum produced only modest reductions in OSA episodes [203]. Similar wires inserted into the anterolateral tongue via a sub-mental approach produced inconsistent changes in the diameter of the airways of normal, awake volunteers [204]. A surgically implanted system developed in collaboration with Medtronic targeted selected branches of the hypoglossal nerve and triggered stimulation from an implanted sub-sternal pressure sensor [205]. Few works also used electrical stimulation of the hypoglossal nerve [206, 207].

2.8. Summary

In this chapter the definition and basic concepts related to sleep apnea have been reviewed, with the aim being to provide a general knowledge of sleep apnea for engineering researchers. It is impossible to mention and review all the literature available in an area as wide as sleep apnea, but the most important areas were reviewed and studies in this field. As well as a definition of sleep apnea, some information about diagnosing, treatment methods, and economical aspect of this area were also given.

Chapter 3

Methodology Review

3.1. Introduction

In previous sections sleep apnea and related topics such as the symptoms, cost, treatment and etc., were introduced, and the scope of the project was reviewed. In this section some popular approaches and methods used in the machine learning phase of the project with their previous applications in sleep apnea studies will be reviewed.

It should be stated that parallel computing has not been used to study sleep apnea, but parallel PSO will be introduced in this section.

3.2. Support Vector Machine

In this section most important aspects about support vector machines are reviewed.

3.2.1. Introduction

The support vector machine (SVM) is a powerful classification method proposed by Vapnik in 1995[208] since then different types of SVM have been proposed. While this method basically discriminated between two classes, it can still be used for multi class problems. SVM can find a good decision boundary between two classes, where the margin between the decision boundary and both classes have been maximised. Consider a binary classification, using a training set of N samples $(\boldsymbol{x_1}, y_1), \dots, (\boldsymbol{x_i}, y_i), \dots, (\boldsymbol{x_N}, y_N) \in \mathcal{R}^n \times \{\pm 1\}$, where $\boldsymbol{x_i}$ is the input vector corresponding to the ith sample that is labelled by y_i depending on its class. SVM aims at separat-

ing the binary labeled training data with a hyper-plane that is at maximum distance from them. This is known as the maximum margin hyperplane. Figure 3.1 shows the basic idea of the SVM, graphically. The pair $(\boldsymbol{w}, b)$ defines the hyperplane with the tion $< \boldsymbol{w}, \boldsymbol{x} > + b = 0$. This hyperplane can separate the train data linearly if

$$y_i(\boldsymbol{w}.\boldsymbol{x_i} + b) \geq 1\,, \qquad i = 1, \dots, N \qquad (3.1)$$

The distance of each training data $\boldsymbol{x_i}$ from the hyperplane is given by

$$d_i = \frac{\boldsymbol{w}.\boldsymbol{x_i} + b}{\|\boldsymbol{w}\|}, \qquad (3.2)$$

and combining inequality (3.1) and (3.2), for all $\boldsymbol{x_i}$ will result in

$$y_i d_i \geq \frac{1}{\|\boldsymbol{w}\|}. \qquad (3.3)$$

Therefore, $\frac{1}{\|\mathbf{w}\|}$ is the lower bound of the distance between the training data $\boldsymbol{x_i}$ and the separating hyperplane. The maximum margin of the hyperplane can be considered as the solution to the problem of maximising the $\frac{1}{\|\mathbf{w}\|}$ subject to the constraint (3.1), or by solving the following problem

$$Minimize \quad z = \frac{1}{2}\boldsymbol{w}.\boldsymbol{w} \qquad (3.4)$$

$$s.t. \qquad y_i(\boldsymbol{w}.\boldsymbol{x_i} + b) \geq 1\,, \qquad i = 1, \dots, N.$$

Consider $(\alpha_1, \alpha_2, \dots, \alpha_N)$ as the N non-negative Lagrange multipliers associated with the constraints (3.1), and without considering a few steps, the resulting decision function is given by [209],

$$f(\boldsymbol{x}) = sign\left(\sum_{\alpha_i > 0} y_i \alpha_i < \boldsymbol{x}, \boldsymbol{x_i} > + b\right), \qquad (3.5)$$

Note that the non-zero α_i is those for which the constraints (3.1) are satisfied by the equality sign. This has an important consequence. Since most of α_i is usually zero the

vector $\boldsymbol{w}$ is a linear combination of a relatively small percentage of the training data $\boldsymbol{x_i}$. These points are called Support Vectors (SV) because they are the closest points to the separating hyperplane and the only points needed to determine the hyperplane. Support Vectors are the training patterns that lie on the boundaries of the margin. In reality, SVM only uses a small subset of the training samples SVs for the classification.

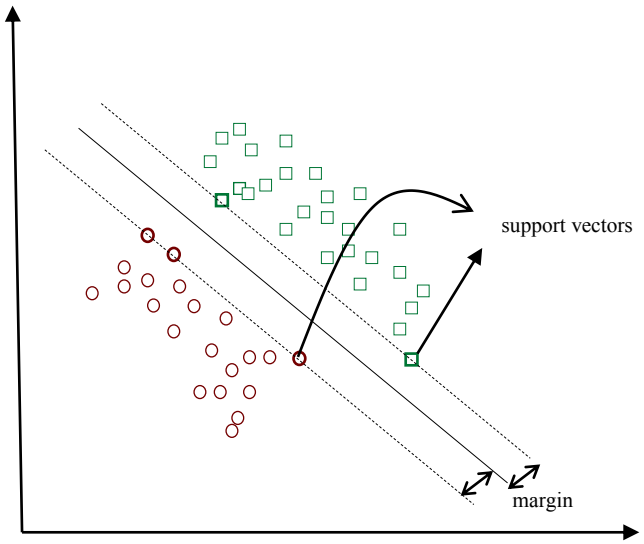

Figure 3.1: Basic ideas of support vector machines

There is also another type of support vectors that consists of the training data that are beyond their corresponding margins. These support vectors are regarded as misclassified data [210].

If the training data are not linearly separable, the problem of searching for a separating hyperplane is meaningless (there may be no separating hyperplane to start with). Fortunately, the previous analysis can be generalised by introducing N non-negative variables $(\boldsymbol{\xi_1}, \boldsymbol{\xi_2}, \dots, \boldsymbol{\xi_N})$ such that,

$$y_i(\boldsymbol{w}.\boldsymbol{x_i} + b) \geq 1 - \boldsymbol{\xi_i}\,, \qquad i = 1, \dots, N. \qquad (3.6)$$

The purpose of the variables $\boldsymbol{\xi_i}$ is to enable a small number of misclassified points. If the data $\boldsymbol{x_i}$ satisfies inequality (3.1), then, $\boldsymbol{\xi_i}$ is zero and (3.6) reduces to (3.1). Instead, if the data $\boldsymbol{x_i}$ does not satisfy inequality (3.1), the extra term -$\boldsymbol{\xi_i}$ is added to the right hand side of (3.1) to obtain inequality (3.6).

It should be noted that by introducing this tolerance parameter actually some training data were ignored in order to have a linearly separating hyperplane. The generalised separating hyperplane is then regarded as the solution to,

$$Minimize \quad z = \frac{1}{2}\boldsymbol{w}.\boldsymbol{w} + C\sum_{i=1}^{N}\boldsymbol{\xi_i} \tag{3.7}$$

$$s.t. \qquad y_i(\boldsymbol{w}.\boldsymbol{x_i} + b) \geq 1 - \boldsymbol{\xi_i}, \qquad i = 1, \dots, N.$$

The purpose of the $C\sum_{i=1}^{N}\boldsymbol{\xi_i}$, is to keep the number of misclassified points under control. Note that this term leads to a more robust solution. The penalty parameter C can be regarded as a regularisation parameter. The above problem tends to maximise the minimum distance 1/w for small C, and minimise the number of misclassified points for large C. For intermediate values of C the solution of the problem (3.7) trades errors for a larger margin. In this case, the decision function is given by,

$$f(\boldsymbol{x}) = sign\left(\sum_{\alpha_i>0} y_i\alpha_i < \boldsymbol{x}, \boldsymbol{x_i} > + b\right), \tag{3.8}$$

$$0 \leq \alpha_i \leq C, \qquad i = 1, \dots, N$$

In order to use the SVM to produce non-linear decision functions, the training data is projected to a higher dimensional inner product space F, called feature space, using a non-linear map $\phi(\boldsymbol{x})$: $\mathcal{R}^n \rightarrow \mathcal{R}^d$. The optimal linear hyperplane is computed in the feature space. Nevertheless, by using kernels it is possible to make all the necessary operations in the input space by using $k(\boldsymbol{x_i}, \boldsymbol{x_j}) =< \phi(x_i), \phi(x_j) >$ as $k(\boldsymbol{x_i}, \boldsymbol{x_j})$ is an inner

product in the feature space. The decision function can be written in terms of these kernels as follows:

$$f(\boldsymbol{x}) = sign\left(\sum_{\alpha_i>0} y_i\alpha_i k(\boldsymbol{x}, \boldsymbol{x_i}) + b\right). \qquad (3.9)$$

Also, the decision value for each $\boldsymbol{x}$ of the test set which can get a negative or positive value depends on the position of the $\boldsymbol{x}$ and the hyperplane, which is defined as equation (3.10).

$$\mathrm{h}(\boldsymbol{x}) = \sum_{\alpha_i>0} y_i\alpha_i k(\boldsymbol{x}, \boldsymbol{x_i}) + b \qquad (3.10).$$

There are 3 common kernel functions in SVM:

Polynomial kernel : $K(x_i x_j) = (x_i x_j + 1)^q$

RBF kernel : $K(x_i x_j) = e^{-\gamma|x_i - x_j|^2}$

Sigmoid kernel : $K(x_i x_j) = tanh(\gamma x_i^T x_j + c)$

Here q, γ, c are kernel parameters.

3.2.2. Application of SVM based Systems in Sleep Apnea Studies

For the first application of SVM in sleep apnea work of Cho et al. in 2005 should be considered [211]. In this study a single channel EEG was used with SVM to detect sleep apnea events. The mean value from the signal used to make the zero-mean distribution of the signal that was above or below a specified range (-150 ~ +150uV), and a band-pass filter, filtered the signal from 0.5 to 50 Hz. To estimate the changes of power spectrum in time, a spectrogram with 257 points (1.285 seconds) and a Hanning window was calculated every 60 seconds. The result of the time-frequency analysis was then used to evaluate the six frequency bands by adding all the values of each band: 0-0.5Hz

(gamma), 0.5-4Hz (delta), 4-8Hz (theta), 8-12Hz (alpha), 12-16Hz (sigma), 16- 30Hz (beta). Next, the mean values computed for each band per one second and the median filter, were applied to the mean values to obtain a smoother signal. The alpha and beta power, and the ratio between the current alpha or beta power and their average of them during the previous ten seconds, were including the ratio between the sigma and alpha plus beta power, which could suggest the presence of sleep spindles, and the mean frequency of the signal at every second, were selected as features. For classification RBF kernel and c=10 are selected. The results showed sensitivity equal to 75.26% and specificity equal to 93.08%. Übeyli et. al. also used single EEG with a least squares support vector machine LS-SVM [212]. The EEG signals (pre and during hypopnoea) from three electrodes (C3, C4 and O2) were considered as a classification problem with the Auto regression (AR) coefficients. The features defining the EEG signals were computed by the Burg AR method. The extracted features were used as inputs of the LS-SVM. This approach could discriminate the EEG signals with very high accuracy (the total classification accuracy was 95.00%).

Single ECG signal is also used with SVM for sleep apnea studies. Khandoker et al. in 2007 [213], and also another works of Khandoker in 2009 [214]. In this study the decomposition of wavelets at 14 levels were applied to HRV and EDR signals. In the classification phase, a leave-one-out cross-validation scheme was adopted to evaluate the ability of the classifier to generalize, and the hill climbing feature of the selection algorithm was used to identify features that contributed the most in separating the two classes. By using a polynomial kernel the accuracy for testing a set was equal to 92.85%.

Also Yildi et.al. in 2011 used nocturnal ECG recordings [215]. In the first stage an algorithm based on Discrete Wavelet Transform (DWT) was used to analysis the ECG re-

cordings for the detection of heart rate variability (HRV) and ECG-derived respiration (EDR) changes. In the second stage a Fast Fourier Transform (FFT) based power spectral density (PSD) method was used for to extract features from the HRV and EDR changes, and 64 features were extracted. In the feature selection step, the hill climbing feature selection algorithm was used and 4-8 features were selected for the final stage. The system correctly recognised 20 out of 20 subjects with OSA and 10 out of 10 normal subjects with the least squares support vector machine (LS-SVM). By comparing the classification performance of the RBF and the poly and linear kernels, it can be noticed that the RBF kernel performed better than the other two. Overall, it emphasized that the classifier could access the best classification performance using a subset consisting of a few good features. Yilmaz et al. in 2010, used a single lead ECG with SVM to detect sleep apnea [216]. The sampling frequency used for ECG acquisition was 200 Hz and the band-pass filter cut-off frequency values were set at 0.5 Hz and 40 Hz. Then the median, inter-quartile range and the mean absolute deviation values were extracted as features from the RR-interval values for each epoch. In the classification k-Nearest-Neighbor (KNN), Quadratic Discriminate Analysis (QDA) and SVM were tested. SVM with a mean accuracy of 87.3% performed the best. Furthermore, in 2011 Isa et.al [217], used an ECG to compare SVM with NN and Naïve-Bayes. This study used PCA for feature selection and concluded that SVM with an RBF kernel had the best classification accuracy.

Nasal air flow was also used as a single signal in studies by Koley and Dey [218, 219]. They used an ensemble of three binary SVM arranged in a one-against-all strategy to classify the feature vector among three categories as normal, apnea, and hypopnea. They used The Recursive Feature Elimination (RFE) technique for feature selection and 8

features from 36 were selected. To detect the optimum values of SVM they used a grid search.

Patangay et al. in 2007 [211] who also used ECG with heart sound and nasal flow to detect sleep apnea events in patients with heart failure. In preprocessing phase R waves detected from ECG, low pass filtering was applied on heart sounds to eliminate ambient noise and for the measurement of S1 heart sound. The S1 heart sound was detected and measured using a simple window technique. Also, the RR intervals were computed for each two minute interval. A spline interpolation was used for the S1 amplitude and RR interval to convert them into uniformly sampled event based series. All the data were then separated into two training sets and tested randomly, and then an SVM with a linear kernel and C=10 was applied to the training sets to determining best features, and those features with absolute weights > 0.1 were selected. By using the selected feature to classify the testing data, 85.5 % the sensitivity and 92.2% specificity were achieved.

SVM integrated with genetic algorithms were used in study by Chen et al. in 2009 [220]. In this study the EEG, EOG and sub-mental EMG, tibia EMG, airflow, inductance plethysmorgraphy, ECG, arterial oxygen saturation and questionnaires included the demographic (age, gender, etc.), anthropometric (weight, height, BMI, waist and neck circumferences, etc.), and symptomatic (diabetes, hypertension, asthma, smoking, alcohol consumption, etc.). In the classification phase genetic was used to select the best feature and the parameter of SVM, and accuracy equal to 89%-92% was reported. A genetic algorithm was also used by Hang et.al.[19] in 2012. In their study the genetic algorithm was used to tune the SVM parameters, including the combined weight of three SVMs in the proposed ensemble system.

3.3. Artificial Neural Networks

In this section different characteristics of Artificial Neural networks together with a review on their application on sleep apnea study are considered.

3.3.1. Introduction

Artificial Neural networks (ANNs) were originally inspired by the brain's nervous system and first introduced by Walter and Warren McCulloch in 1943. The basic idea behind the nervous system is a simple unit, neuron, with a complex connection between too many other neurons, that can perform very complex tasks in a limited time [221].

An artificial neural network is simply shown by Figure 3.2. It contains a single cell which takes p inputs, and each input is multiplied by a specific weight. The output is found according to the sum of the weighted inputs minus a certain threshold (θ). The output of cell (y) follows a step function φ, and therefore the output is either one or zero depending on the sum of the weighted input. For simplicity the threshold value can be expressed as extra input with a value of 1 and a weight of θ[221].

All other models of neural networks are basically on the same as the previous model but with some extension to suit more complex problems. These extensions are based on two factors:

1) Architecture (topology): the main categories of ANNs are feed-forward and feed-backward networks. Neurons in feed-forward networks are organised into layers with a uni-directional connection between them, while neurons in feed-backward nets are connected in both directions. As a result of these connections feed-forward ANNs are static, which means that they produce one set of outputs while the output of feed-backward ANNs depends on the previous state of the net. ANN may produce many outputs, one for each state.

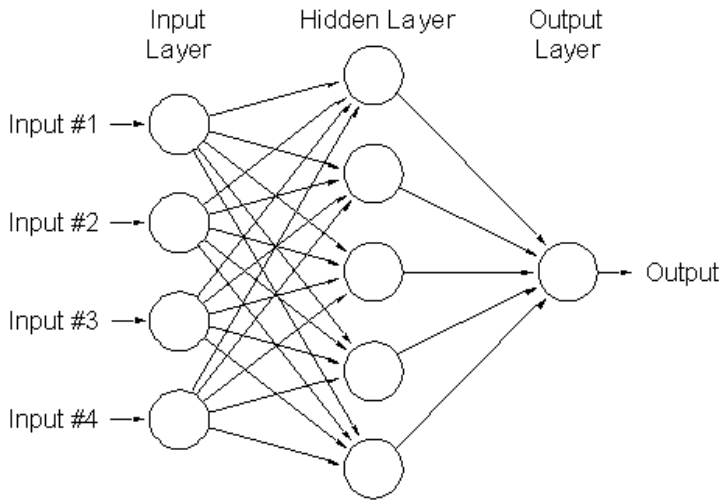

Figure 3.2: Simple artificial neural network structure

2) Learning: the connecting weights in all types of neural networks should be tuned to ensure that the separation between available outputs is adequate. This is why a learning algorithm is needed to modify the weights according to a given training set. The learning procedure can be supervised learning, reinforced or graded learning, and unsupervised learning. In supervised learning the patterns of the training set are given with their actual outputs to train the net by modifying its weights, while in unsupervised learning the weights are updated according to correlated measures between data, without knowing the actual output.

For more information about the learning process and different types of ANNs please see [221]. Learning in neural networks is achieved by adjusting the modifiable connecting weights between the units. In other words, learning in neural networks is a problem of finding a set of connecting weights which would enhance the ability of neural networks to store experiential knowledge; hence, learned knowledge can be used to achieve the desired response in the future. At present there are a variety of learning algorithms available, which can be classified into three main classes: supervised learning,

reinforced or graded learning, and unsupervised learning. The Backpropagation (BP) algorithm is one of the notable supervised learning algorithms. This approach basically refers to an external agent such as a computer program that totally monitors the input and output vector pairs and adjusts the weights such that each network output matches its target value. Other commonly used supervised learning algorithms are Levenberg–Marguardt and gradient descent. The training objectives can be defined as follows [222]:

$$J = \frac{1}{N}\sum_{i=1}^{N}[\hat{y}\,(i) - y(i)]^2 \qquad (3.11)$$

where N is the number of data points, $\hat{y}$ is the network prediction, and y is the target value. In the Levenberg–Marguardt optimisation algorithm, the network weight adjustment can be shown as follows:

$$\Delta W(k+1) = -\eta[\frac{1}{N}\sum_{i=1}^{N}\frac{\partial \hat{y}(t)}{\partial W(k)}\,(\frac{\partial \hat{y}(t)}{\partial W(k)})^T + \delta I]^{-1}.\frac{\partial J}{\partial W(k)} \qquad (3.12)$$

$$W(k+1) = W(k) + \Delta W(k+1) \qquad (3.13)$$

where $W(k)$ and $\Delta W(k)$ are the vectors of weights and weight adaptations at step k respectively, η is the learning rate, δ is a parameter to control the size of the searching step, and I is the identical matrix. Gradient descent is another supervised learning algorithm that can be presented by the following equations:

$$\Delta W(k+1) = -\eta\frac{\partial J}{\partial W(k)}$$

$$W(k+1) = W(k) + \Delta W(k+1) \qquad (3.14)$$

This method follows the gradient rule that the weight vector W is literately updated in the direction of the greatest rate of decrease of the error. Other available methods are the

modified gradient descent and the Gauss–Newton which can be represented respectively as below:

$$\Delta W(k+1) = \alpha \Delta W(k) - \eta \frac{\partial J}{\partial W(k)}$$

$$\Delta W(k+1) = -\eta [\frac{1}{N} \sum_{i=1}^{N} \frac{\partial \hat{y}(t)}{\partial W(k)} \, (\frac{\partial \hat{y}(t)}{\partial W(k)})^T + \delta I]^{-1} . \frac{\partial J}{\partial W(k)}$$

where α is the coefficient of momentum. The term momentum term basically leads to a faster convergence toward minimum without causing a divergent oscillation, but be aware that the Gauss–Newton method can encounter a numerical problem if the term $\sum_{i=1}^{N} \frac{\partial \hat{y}(t)}{\partial W(k)} \, (\frac{\partial \hat{y}(t)}{\partial W(k)})^T$ is close to singular.

The reinforcement learning algorithm is also similar to supervised learning except that the desired output is not provided. Unsupervised learning however, only uses the input vector for network training while the network regulates its own weights without the benefit of knowing what particular output to assign to a given input. The Kohonen Rule for training on the Kohonen network is one example of unsupervised learning.

3.3.2. Elman Neural Networks

The Elman neural network is one kind of globally feed-forward locally recurrent network model proposed by Elman [26]. It occupies a set of context nodes to store the internal states. Thus, it has certain dynamic characteristics over static neural networks, such as the Back-Propagation (BP) neural network and radial-basis function networks.

Elman network consists of four layers: input layer, hidden layer, context layer, and output layer. There are adjustable weights connecting every two adjacent layers. Generally, the Elman neural network can be considered as a special type of feed-forward neural network with additional memory neurons and local feedback. The distinct “local connec-

tions" of context nodes in the Elman neural network make its output sensitive not only to the current input data, but also to historical input data, which is useful in time series prediction. The training algorithm for the Elman neural network is similar to the back-propagation learning algorithm, as both based on the gradient descent principle. However, the role that the context weights as well as initial context node outputs play in the error back-propagation procedure must be taken into consideration in the derivation of this learning algorithm. Due to its dynamical properties, the Elman neural network has found numerous applications in such areas as time series prediction, system identification and adaptive control [27].

3.3.3. Cascade-Forward Neural Network Models

Cascade-forward models are similar to feed-forward networks, but include a weight connection from the input to each layer and from each layer to the successive layers. While two-layer feed-forward networks can potentially learn virtually any input-output relationship, feed-forward networks with more layers might learn complex relationships more quickly. For example, a three layer network has connections from layer 1 to layer 2, layer 2 to layer 3, and layer 1 to layer 3. The three-layer network also has connections from the input to all three layers. The additional connections might improve the speed at which the network learns the desired relationship. Cascade-forward artificial intelligence model is similar to feed-forward back propagation neural network in using the back propagation algorithm for weights updating, but the main symptom of this network is that each layer of neurons related to all previous layer of neurons. This network is a Feed-Forward network with more than one hidden layer. Multiple layers of neurons with nonlinear transfer functions allow the network to learn more complex nonlinear relationships between input and output vectors.

3.3.4. Multi Artificial Neural Networks

The artificial neural network is a massive parallel-distributed processor with the capability of storing experiential knowledge and then making it available for use [223]. Artificial neural networks have been increasingly used in non-linear modelling in industry because the main advantage of neural network models is that they are easy to build. This feature is very useful when modelling complicated non-linear processes where detailed mechanistic models are difficult to develop [222].

Robustness is one of the main criteria used to judge the performance of the neural network, where robustness refers to an inherent insensitivity to the presence of outliers [224]. Low robustness is basically due to an inappropriate learning of the training data or over fitting the training data [225] to improve learning in neural network algorithms, different techniques such as training with regularisation and early stopping have been developed [225]. Among those approaches for a general improvement of the neural network, the combination of multiple neural networks seems to be very effective [225]. Many real world problems are too large and too complex for a single monolithic system to solve alone and also fit the data distribution well enough. In reality there is no perfect model presently available to truly and consistently define the process, therefore the multi-model (consensus/aggregation) approach offers a real prospect of better simulation than the best model included in the combination [226]. The idea behind a combination is that the process inputs contain insufficient information about the outputs, while the neural network model is sub-optimal because it does not completely utilise the information in the inputs. In other words, combining a set of imperfect models (networks) can be thought of as a way to manage the recognised limitations of individual models, where each is known to have errors, but then they are combined in a certain way to min-

imise the effect of these errors [227]. Because each network can waste some of the information in the inputs because of insufficient training or approximation failures [228], by using the information that other models have to offer, a better prediction of the output can be expected [228].

The sufficiency of a combined model can also be proved by Shannon's information theory [228]. What can also be improved [229] is combining three independent ANNs using majority voting so that each one can give the correct classification with a probability ρ while the combined network can give the correct classification with a probability of $3\rho^2 - 2\rho^3$. Therefore, the combined network performs better than the single network when $0.5 < \rho < 1$. For successful applications of combined ANN please refer to [222].

A. Combining of Artificial Neural Networks

Combining classifiers to achieve higher accuracy is an important research topic that has the following different names in the literature: a combination of multiple classifiers [216, 221, 230], classifier fusion [231, 232], mixture of experts [233, 234], classifier ensembles [235, 236], committees of classifiers [237], consensus aggregation [238], voting pool of classifiers [239], composite classifier system [240], divide-and-conquer classifiers [241], etc. The paradigms of these models differ on the assumptions about classifier dependencies, the type of classifier outputs, the aggregation strategy (global or local), and the aggregation procedure, and etc. [220].

In general, information can be classified as pre-classification fusion and post-classification fusion [224]. Pre-classification fusion refers to combining information prior to the application of any classifier or matching algorithm, and post-classification

fusion refers to methods that combine the information after decisions about the classifiers have been obtained [224].

Furthermore, post-classification fusion in ANN methods can be classified as modular and ensemble [222]. Modular methods refer to a neural network model devised on the basis of component networks that are trained using different data and are then combined to create a neural network model. Ensemble on the other hand, applies to a method where the component networks are redundant in the sense that each one provides a solution to the same task or task component. Each of the component networks in an ensemble might reach a solution by different means but any one of them could be used in isolation to provide a solution to the task. During testing, the implication is that all input patterns will be presented to all the component networks and the output of all the networks will be combined to form an ensemble output.

Ensemble combination methods can be separated into several distinctive groups, such as linear, non-linear, Supra Bayesian, and stacked generalisation [222].

B. Linear and non-linear combination

A linear combination is the most used combination method for neural networks. In weighted averaging, individual network outputs are multiplied by appropriate weights and then combined to give the final model prediction [222]. Let us consider L classifier as $D = \{D_1,\ D_2, \dots, D_L\}$ for a model by k class labels as $\Omega = \{\omega_1, \omega_2, \dots, \omega_k\}$. For each arbitrary $z \in R$ each classifier assigns a class label from Ω, and the output of each classifier D_i can be defined as $D_i(z) = [d_i^1(z),\ d_i^2(z), \dots, d_i^k(z)]^T$, where $d_i^j(z)$ is the measure that classifier D_i assigns to the hypothesis that z comes from class ω_j. Similarly the outputs of the L classifier for vector z can be represented in a decision profile $DP(z)$, as follows;

$$DP(z) = \begin{bmatrix} d_1^1(z), \dots, d_1^j(z), \dots, d_1^k(z) \\ \dots \quad \dots \quad \dots \\ d_i^1(z), \dots, d_i^j(z), \dots, d_i^k(z) \\ \dots \quad \dots \quad \dots \\ d_L^1(z), \dots, d_L^j(z), \dots, d_L^k(z) \end{bmatrix}$$

Different types of combinations have been proposed to use this matrix to find the overall output for each class [242]. Among the class conscious methods, the weighted linear combination is the most popular. Various weighted average combiners have been proposed and they can be classified into three categories [243].

In the first category, L weights are assigned, one per classifier. In this case the output for class ω_j is calculated as

$$\mu_j(z) = \sum_{i=1}^{L} w_i . d_i^j(z)$$

where w_i is the weight assigned to the ith classifier. In the simplest case, with simple averaging, L weights are equal to $w_i = 1/L$. This is the most common method, but it is not optimal. In a more reasonable way the weight assigned to each classifier can be based on its estimated error rate obtained using the behaviour of the classifier [244], the correlation between classifiers [245] and the minimum classification error, MCE, criterion [246].

In the second category the output for class ω_j is calculated as follows, where w_{ij} is the weight assigned to the ith classifier for class ω_j,

$$\mu_j(z) = \sum_{i=1}^{L} w_{ij} . d_i^j(z)$$

In the third category of linear combinations, the output for class ω_j is obtained by a linear combination of all elements of the decision profile DP(z) as

$$\mu_j(z) = \sum_{k=1}^{K} \sum_{i=1}^{L} w_{ikj} \, . \, d_i^k(z)$$

Also Principal Component Regression (PCR) approach should be mentioned to. In the PCR Instead of combining the system properties on the original measured variables, the properties are combined on the principal component scores of the measured variables (which are orthogonal and thus well-conditioned) [222].

A linear combination is simple and easy to use but only those models with a high linear correlation to the output variable can be modelled by linear approaches. It is therefore important to develop non-linear aggregations that can combine useful ANN models regardless of the nature of their relationship to the actual output [228].

In non-linear approaches the Bayesian selective combination method [227] can be noticed where, instead of using fixed combination weights, the probability of a particular network being the true model is used as the combined weight for combining that network. The prior probability is calculated using the sum of squared errors of individual networks on a sliding window that covers the most recent sampling times. A nearest neighbour method is used to estimate the network error for a given input data point, which is then used to calculate the combined weights for individuals [227]. Among the non-linear methods the Dempster–Shafer method seems to be the most renowned because it should deal with the uncertainty and ignorance of the classifiers [222]. The Dempster-Shafer theory of evidence is a tool for representing and combining measures of evidence. This theory is a generalisation of Bayesian reasoning and it is more flexible than Bayesian when our knowledge is incomplete and uncertainty and ignorance is exist [221].

Majority voting is another popular choice among methods in the non-linear classification. Zhang [222] implemented majority voting to improve online fault diagnosis using multiple neural networks. Majority voting is applied in the natural way, i.e. if there are more individual classifiers giving an output of 1 rather than 0, then the aggregated classifier takes the output 1 and vice versa. Since the neural network outputs take continuous values between 0 and 1, majority voting cannot be implemented in its natural form. Here it is proposed that majority voting takes the following form:

$$\mu_j(z) = median\,\{d_i(z)\}, i = 1, \dots, l$$

The modified majority voting combination scheme proposed is of the following form:

$$\mu_j(z) = \begin{cases} median\,\{d_{ij}[z]\}, & if\ median\{d_{ij}[z]\} \leq 0.6 \\ max\,\{d_{ij}[z]\}, & if\ median\{d_{ij}[z]\} > 0.6 \end{cases}$$

where $\mathbf{d_{ij}}$ is the **jth** element of the **ith** NNs. The rationale behind this combination scheme is when the majority of individual networks give outputs that are much larger than $\mathbf{0}$ ($> \mathbf{0.6}$), then a higher output for the combination is very likely. On the other hand, when most of the individual networks give outputs which are not much larger than $\mathbf{0}$ ($\leq \mathbf{0.6}$), then the median of the outputs is taken as the final output. The threshold 0.6 is set based on heuristics and it is possible to fine tune this parameter based on a set of training and testing data [229].

For more information about other approaches such as Supra Bayesian, Recursive Least Squares (RLS), Combination Using Fuzzy Logic and etc. please refer to [247] or [222].

C. Training Multiple Artificial Neural Networks

As mentioned before, there are several types of multiple neural networks but the main ideas are basically similar. They can be categorised based on the training methods, they can be classified into two major types [222]. The first category is multiple model neural

networks where the training data used to build the individual networks are completely different, so they can only be built using different inputs in different regions of operation. The idea behind this approach is to adapt different information using different inputs and then combine this information to obtain a better prediction. The learning algorithm in each network can also be different and can be either the supervised or unsupervised method. The second category is to create multiple models using the same training data which has been resampled or partitioned using particular algorithms. Different studies show that combining multiple neural networks can only be effective if the individual networks are accurate and diverse [242]. Therefore, various techniques have been used to create a diverse ensemble. This section will review the two most popular algorithms being used to resample or partition the training data, and they are bagging or bootstrap resampling, and adaboost [222].

Bagging, (short for Bootstrap Aggregation Learning), [242] is the most commonly implicit method for creating an ensemble. This approach refers to the replication of a training data set by resampling the original training data set. Therefore some of the data samples may occur several times, and others may not occur in the sample at all. The individual training sets are independent and the neural networks can be trained in parallel [222]. A training set of size N generates L new training sets, each of size N, by randomly drawing samples of the original training set, where the same sample may be drawn multiple times. Each of the new training sets is used to train exactly one neural network classifier. Hence, an ensemble of L individual networks is obtained from L new training sets, several researchers used this algorithm [248-251].

AdaBoost or ‘adaptive boosting’ [252] proceeds in iterations, with a new network being trained at each iteration. A network is trained initially with equal emphasis on all

training data, and at the end of each iteration the misclassified patterns are identified and have their emphasis increased in a new training set, and then a new network is trained. After the desired numbers of networks have been trained, they are combined by a weighted vote, based on their training error [253].

Another popular algorithm is Negative Correlation Learning, for more information about this approach please refer to [254].

3.3.5. Application of Neural Network based Systems in Sleep Apnea Studies

There are several works related to the application of ANNs in sleep apnea studies, but for the most important ones following works can be mentioned.

In 2005 Fontenla-Romero et al. proposed feed-forward neural networks to classify sleep apnea events as obstructive, central, or mixed. The inputs of the neural network are the 5 detail coefficients of the first level obtained from the transformation of a discrete wavelet of the thoracic effort signal. The mean classification accuracy, obtained over the test set was 83.78 $\pm$1.90% [230].

Another work that used the neural network to detect sleep apnea was Liu et al. in 2008, who used Pupil Size and EEG as the input [231]. They showed that modified adaptive resonance theory (ART) NNs [233], performs better than other structures, and they achieved 91% accuracy classifying subjects with OSA and controls. It should be noted that they published their early results by 2007 [234]. Tagluk and Sezgin also used EEG signals and ANN in 2011[255], in their work. The energy remaining under the delta, theta, alpha, beta, and gamma frequency bands of EEG were then given as inputs to the classification module. For the classifier they used multi-layer perceptron neural networks with a bipolar sigmoid activation function and they achieved 96.15% accuracy for detecting apnea or normal. Tagluk was also involved in two other works in 2010

[256, 257] where abdominal respiration was used as the input and the features were generated by coefficients of a wavelet with a depth of 7 (i.e. 7 levels). To classify sleep apnea as central obstructive and mixed [235] the abdominal respiration signals were separated into spectral components using multi-resolution wavelet transforms and then these spectral components were applied to the inputs of the artificial neural network. The multi-layer neural networks model used in this study has one input layer with 16 nodes, one hidden layer with 15 nodes, and an output layer of 3 nodes. The activation function used for all nodes in the hidden layer and output layer is the bipolar sigmoid function which is robust and highly efficient. Inputs of this ANNS are coefficients of the 7^{th} of wavelet transfer. They reported the mean accuracy and confidence interval for each one of the classes as 73.42% (OSA), 94.23% (CSA) and 66.16% (MSA). Another study by Tagluk in 2011 [236], used EEG as the input and a neural network consisting of one input layer, two hidden layers, and one output layer.

Oxygen saturation was also used as input in some studies, such as that done by Marcos and A´ lvarez [258] in 2012 or another study by Morillo and Gross in 2013 [259].

Marcos and A´ lvarez [258] used Fourteen time-domain and frequency-domain features to quantify the effect of SAHS on SaO2 recordings. The distribution of SaO2 values tends to reflect different properties, depending on the AHI, the mean (μ), variance (σ), skewness (γ), and kurtosis (δ). On the other hand, a non-linear analysis of SaO2 signals using the approximate entropy (ApEn), central tendency measure (CTM), and Lempel-Ziv complexity (LZC) was performed to measure irregularity, variability, and complexity, respectively. Also, as frequency features they used three additional features derived from the PSD function: the total power of the SaO2 signal (ST), the power in the band between 0.010, and 0.033 Hz (SB), and the most significant frequency compo-

nent in that band (PA). They compared multiple linear regression (MLR) and multi-layer perceptron (MLP) neural networks and concluded that the MLP algorithm performed best with an intra-class correlation coefficient (ICC) of 0.91.

Morillo and Gross [259], used a set of 17 time domain, stochastic, frequency domain, and non-linear features initially computed from SpO2 recordings. They used the sequential forward selection (SFS) method and a Probabilistic ANN (PNN) classifier, and using the area under the receiver operating characteristic (ROC) curve (AUC) as a criterion value. SFS consists of two main steps. First, a PNN is designed and the AUC associated to the LOOCV set is computed for each of the variables. The feature with the best value of AUC is selected. Second, all possible two dimensional vectors that contain the winner from the previous step are formed. In each case a new PNN is trained and validated, and its AUC is calculated. Again, the best one is selected. This procedure continues by forming all three dimensional vectors springing from the two-dimensional winner and by selecting the best one. The algorithm finishes when the nth dimensional vector computed from the nth step does not improve the AUC of the PNN that was trained and validated from the winner from the previous step. This system resulted in 92.4 % sensitivity and 95.9 % specificity.

The ECG signal was also used as The ECG signals were preprocessed and segmented to extract the P-waves, and then three P-wave features were extracted: the P-wave duration (Tp), the P-wave dispersion (Pd), and the time interval (Tpr) from the peak of the P-wave to the R-wave. Combinations of the three features were used as features for classification using ANN. In this work, sensitivity, specificity, and accuracy values for detecting OSA were obtained when the ANN classifier was based on the combinations

of the features. The overall accuracy for the combinations Tp and Pd, Pd and Tpr, Tp and Tpr, and Tp, Pd, and Tpr was 89.9%, 76.3%, 89.9%, and 92.3%, respectively.

Three works done by Günes et.al. [260], Chen et.al. [261] and Guijarro-Berdiñas et.al. [18]. Günes et.al. [260], can be considered as feature selection methods in NNs applications in the area of sleep apnea. In their study in 2010 used clinical variables such as the Arousals Index (ARI), the Apnea and Hypoapnea Index (AHI), SaO2 minimum value in stage of REM, and Percent Sleep Time (PST) in the stage of SaO2 intervals larger than 90%, that were obtained from patient polygraph recording, and then they applied the f-score feature selection. As a classifier they used Multi-layer perceptron artificial neural network and the experiments with the MLP back propagation algorithm were done in a MATLAB environment. While MLPANN obtained 63.41% classification accuracy on the diagnosis of OSAS, the combination of MLPANN and multi-class f-score feature selection achieved 84.14% classification accuracy. Chen et.al.[261] in 2011 used anthropomorphic measurements (e.g., age, gender, height, weight, and BMI), systolic blood pressure, diastolic blood pressure, frequency of desaturation (DI3, DI4), frequency of paroxysmal leg movement within an hour, and questionnaire measurements (ESS, SOS), as inputs. They investigated the performance of PSO and GA for feature selection integration with 1-NN method. The experimental results showed that the PSO predicted better than GA. In 2012 Guijarro-Berdiñas et.al.[18] used the nasal airflow signal, the thoracic effort signal, and abdominal signals and features that are generated using the symlet wavelet family (symlet of order: O = 7); specifically, the absolute value of the first 16 coefficients of the level-5-detail. In the feature selection phase they used the SVM recursive feature elimination (SVM RFE) approach and a mixture of classifiers in the classification phase. The classifier trained by different sets learned to discriminate be-

tween obstructive and central characteristics that were applied at the beginning of the input pattern. Thus, the classifier learned to look for common characteristics between the central and mixed patterns in order to distinguish them from obstructive ones. Also, two classifiers trained to discriminate between the obstructive and central characteristics applied at the ending of the input pattern. In this case the classifier will look for common properties between obstructive and mixed patterns in order to distinguish them from central ones. By paying attention to the experimental results, the mean test accuracy obtained by the selected model in 10 different 10-fold cross validations was 90.27% ± 0.79, and the mean test accuracy obtained for each class apnea was 94.62% (obstructive), 95.47% (central) and 90.45% (mixed).

There are very few studies on predicting sleep apnea and of them, few used NNs. The pioneer work on sleep apnea prediction can be found in the paper by Bock and the paper by Gough in 1998 [262]. This study used heart rate, respiration force, and blood oxygen saturation (SaO2.), and also the recurrent networks proposed by Elman [263]. Each of the three time series variables (heart rate, breathing, and blood oxygenation) were used as inputs for network training and testing operations. Each variable was introduced to a unique network node at the input layer; this network had 18 nodes in the hidden layer.

The latest paper in this area is the work of Waxaman, Graupe and Carley in 2010 [264]. They predicted apnea from 30 to 120 seconds in advance. They use LArge Memory STorage and Retrieval (LAMSTAR) neural network. LAMSTAR is a supervised neural network that can process a large amount of data and also provide detailed information about its decision making process. Input signals for this algorithm are EEG, heart rate variability (HRV), nasal pressure, nasal temperature, sub-mental EMG, and electrooculography (EOG). They trained separate LAMSTARs for each 30, 60, 90, and

120 second segment. Their results showed that the best prediction occurred to the next 30 seconds and the performance was less with a longer lead time. However, most predictions up to 60 seconds in the future were correct.

3.4. Particle Swarm Optimisation (PSO)

In this section basic of PSO and some variety of this algorithm are presented.

3.4.1. Single PSO

In 1995, Kennedy and Eberhart introduced particle swarm optimization (PSO)[265]. PSO is a population-based stochastic optimisation technique based on the movement of swarms and inspired by the social behaviour of insects, animals herding, birds flocking, and fish schooling, where a collaborative search for food had the potential for a computational model. In comparison with other metaheurestic algorithms such as genetic algorithms (GAs), PSO has less complicated operations and less defining parameters and it can be coded in just few lines and it is highly dependent on stochastic processes. Because of the simplicity and performance of the PSO, this algorithm has received increasing attention in recent years [266, 267].

In PSO the potential solutions of the optimisation problem are called particles, which is a point in the search space. Compared to genetic algorithms, PSO updates the particles by considering their internal velocity and the position obtained by the experience of all the particles.

Consider a D-dimensional feature space and a PSO with M particles, where $x_i = (x_{i1}, x_{i2}, \dots, x_{iD})^T$ and $v_i = (v_{i1}, v_{i2}, \dots, v_{iD})^T$are the position and velocity of the ith particle respectively $i = 1,2, \dots, M$. The performance of x_i is evaluated by a user defined fitness function. At each iteration the particle updates its own position and velocity

by tracking its best solution and the global best solution discovered by all the particles in the swarm.

Let p_{best} denote the best previous position encountered by the ith particle, and g_{best} denote the global best position so far. By paying attention to p_{best} and g_{best} the ith particle updates its position and velocity according to formula (3.15) and (3.16).

$$v_{id}^{k+1} = v_{id}^{k} + c_1 r_1^k \left(p_{best}^k - x_{id}^k\right) + c_2 r_2^k \left(g_{best}^k - x_{id}^k\right) \qquad (3.15)$$

$$x_{id}^{k+1} = x_{id}^{k} + v_{id}^{k+1} \qquad (3.16)$$

Where r_1^k and r_2^k are random numbers in $[0, 1]$, k is the iteration counter and positive constant c_1 and c_2 are personal and social learning factors. x_{id}^k and v_{id}^k are the current position and velocity of the dth $(d = 1,2, \dots, D)$ dimension in the kth iteration of the ith particle, respectively.

Later, Shi and Eberhart [268] introduced an inertia weight w, which controls the impact of the previous velocity on the current velocity, by modifying Eq.(3.15) to

$$v_{id}^{k+1} = w v_{id}^{k} + c_1 r_1^k \left(p_{best}^k - x_{id}^k\right) + c_2 r_2^k \left(g_{best}^k - x_{id}^k\right) \qquad (3.17)$$

A suitable inertia weight ω provides a balance between global and local exploration and exploitation, and on average results in less iteration for finding a sufficiently optimal solution. Previous studies have shown that a time-dependent weight factor often outperforms a fixed factor [269]. The most common functional form for this weight factor is linear, and it changes with the time step as follows:

$$w(k) = w_{max} - \frac{w_{max} - w_{min}}{N_{iter}} k \qquad (3.18),$$

where N_{iter} is the maximum number of iterations and w_{max} and w_{min} are often selected to be 0.9 and 0.4, respectively [269].

The velocity v_{id} is restricted to the $[v_{min}, v_{max}]$ range. This range determines the resolution of the search regions between the present and target position. If v_{max} is too high, the particle may fly over the good solutions. If v_{max} is too small the particle may not explore beyond local solutions enough, and become trapped in a local optimum.

The constants c_1 and c_2 represent the weights of the stochastic acceleration terms that pull each particle towards p_{best} and g_{best}. Low values allow particles to roam far from target regions before being tugged back, while high values result in abrupt movement towards, or past the target regions. Hence, the acceleration constants c_1 and c_2 are often set to be 2.0[270]. Pseudo code 3.1 shows a sample of the PSO algorithm.

Pseudo code 3.1: *The original Particle Swarm Optimization (PSO)*

1: Initialize randomly the position x and the velocity v of each particle.

2: ***for*** *i = 1 to s* ***do***

3: $Pbest_i = X_i$

4: ***end for***

5: ***while*** *the stopping criterion is not satisfied* ***do***

6: ***for*** *i = 1 to s* ***do***

7: update position X_i *using (3.15) and (3.16)*

8: calculate particle fitness $f(X_i)$

9: update $Pbest_i$*, Gbest, and particles using (3.17) and (3.18)*

10: ***end for***

11: set iteration = iteration + 1

12: ***end while***

3.4.2. Parallel PSO

The recent availability of cheap and fast parallel hardware has encouraged this researcher towards the possibility of implementing a parallel type of meta-heuristic algorithms for large scale problems. Population based algorithms such as genetic algorithms are good candidate for parallelisation [271]. In meta-heuristic algorithms, parallelisation can reduce computation time and also improve the quality of the solutions [272].

PSO can be parallelised by different points of view. In this section the most common parallel structures of PSO are introduced. Because the processes between PSO's are independent, this algorithm is quite suitable to being parallelised. The only dependency existing between the PSO processes is updating the swarm's velocities and positions, which must be shared between particles. Therefore, PSO can be implemented in parallel very efficiently [273].

Parallel processing can generally be classified as pipeline processing and data parallelism. Pipeline processing separates the problem into a cascade of tasks where each task is executed by an individual processor. Data are transmitted through each processor which executes a different part of the process on each of the data elements. Since the program is distributed over the processors in the pipeline and the data moves from one processor to the next, no processor can proceed until the previous processor has finished its task, although more than one processor might be work at same time. The data parallelism method is an alternative approach which involves distributing the data to be processed among all the processors, which then executes its procedure on each subset of the data [274].

Data parallelism is widely used for implementing meta-heuristic algorithms such as PSO. In this case, parallel PSO can be classified into three categories:

(a) Global or Master–slave PSO: This model uses a single swarm and only the fitness evaluations of the particles are done in parallel (in slave processors). Therefore, the nature of PSO has not changed because the algorithm works with the whole population. In this model the parallelisation only tries to speed up the fitness computation [275]. Figure 3.3 represents a model of the master-slave structure. By paying attention to this figure, at each iteration of PSO the master sends the position of particles to the slaves, and then the slaves send back the fitness of the particle to the master. Therefore, only the fitness computations of particles are computed in slaves and other parts of the PSO algorithm are performed in the master.

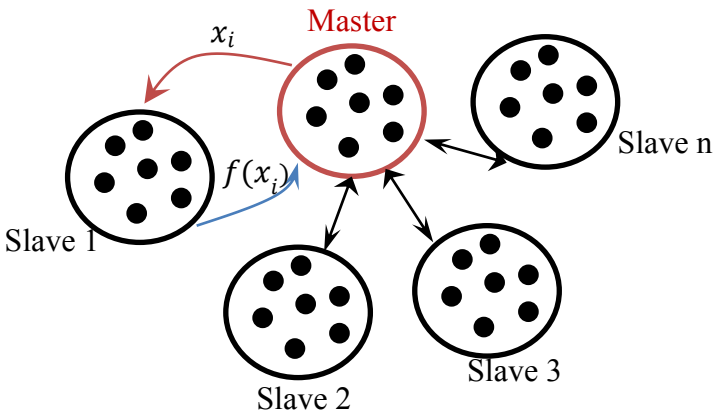

Figure 3.3: Master- Slave parallel PSO structure.

(b) Migration PSO or island model (Coarse-grained): In this model several swarms are available, each of which is maintained by a different processor. Then, according to some ‘‘migration-strategy’’, commonly at a given number of iterations, the particles between these swarms are exchanged [275]. Island PSO models are also referred in literature as multi PSO [276]. Island models are able to control the global information exchange by means of a migration strategy, but parameters such as the ‘‘migration population’’ must be chosen well [277].

Figure 3.4 shows a sample of parallel PSO with an island model structure. It should be noted that because of physical limitations, interaction between the swarms is not simultaneous.

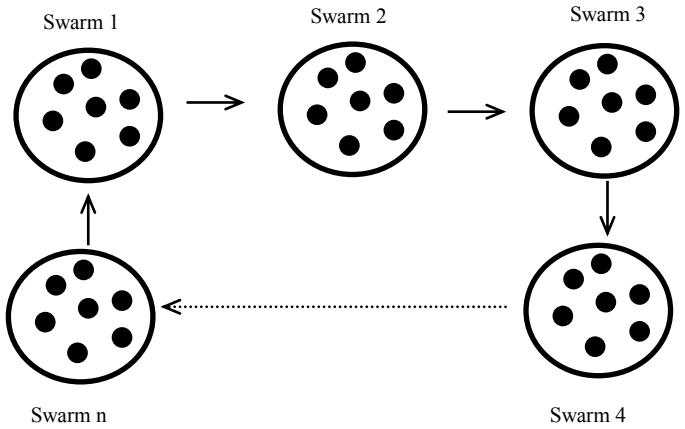

Figure 3.4: Island model PSO structure

In this model the swarms are partitioned into sub-groups, and only those swarms in the same sub-group exchange their information. Interaction between swarms can be performed by different communication structures such as; broadcast, star, migration and diffusion model. For more information about these communication models please refer to [278].

(c) Fine-grained PSO or cellular PSO: In this structure a swarm is separated into a large number of very small sub-populations which are maintained by different processors. The sub-population may only be a particle. Such neighborhood restriction delays the exchange of information between the non-neighbour processes and increases the diversity in the search [275]. The use of local selection and the reproduction rules leads to a continuous diffusion of individuals over the population. Therefore this model is also called the diffusion mode [272]. As this model uses a single population which is distributed among available processors, it allows smaller population sizes than the island model. On the other hand, information is permanently exchanged through neighborhood

connections, which imposes more communication than the island model. It also has the advantage of not having to define a migration strategy [277].

3.5. Performance Indications

There are various measures that can be used to evaluate the performance and assess the different characteristics of machine learning algorithms, although most measures focus on the ability of the classifier to correctly identify classes [279].

Several methods have been used by supervised machine learning approaches to evaluate the performance of learning algorithms such as measuring the quality of classification that are based on the "confusion matrix" which records correct and incorrectly recognized examples for each class. Table 3.1 presents a confusion matrix for binary classification where *tp* is true positive, *fp* is false positive, *fn* is false negative, and *tn* is the true negative amount(s) [279].

Evaluating the classification performance without special interest on a class is the most common and the most used empirical measure where accuracy does not distinguish between the numbers of correct labels of different classes:

$$accuracy = \frac{tp + tn}{tp + fp + tn + fn}$$

Table 3.1: Confusion matrix for a binary classification

label / result	As positive	As negative
Positive	*tp*	*fn*
Negative	*fp*	*tn*

On the other hand, two measures that separately estimate a classifier's performance on different classes are sensitivity and specificity:

$$sensitivity = \frac{tp}{tp + fn}, \quad specificity = \frac{tn}{tn + fp}$$

The another paradigm is focus on one class, especially in bioinformatics, text classification, information extraction, natural language processing, where the number of examples belonging to one class is often considerably lower than the overall number of examples [280]. The experimental setting is as follows: within a set of classes there is a class of special interest (usually positive). Other classes are either left as is – multi-class classification – or combined into one – binary classification. The measures of choice calculated on the positive class are [279]:

$$precision = \frac{tp}{tp + fp}, \quad recall = \frac{tp}{tp + fn}$$

$$F - score = \frac{(\beta^2 + 1) * precision * recall}{\beta^2 * (precision + recall)}$$

All three measures distinguish the correct classification of labels within different classes because they concentrate on one class (positive examples). Recall is a function of correctly classified examples (true positives) and its misclassified examples (false negatives). Precision is a function of true positives and examples that are misclassified as positives (false positives). The F-score is evenly balanced when β = 1 because it favors precision when β > 1, and recall otherwise [2]. A broad evaluation of classifier performance can be obtained by the ROC:

$$ROC = \frac{P(x|positive)}{P(x|negative)}$$

Where $P(x|C)$ represents the conditional probability that a data entry has the class label C and an ROC curve plots the classification results from the most positive to the most negative classification. ROC and the Area under the Curve (AUC) apply to learning with asymmetric cost functions and imbalanced data sets [280]. To obtain the full range of true positives and false negatives, which is why ROC is used in experimental sciences where it is feasible to generate a lot of data. The study of radio signals and biomedical and medical science are a steady source of learning problems. Another possibility of building the ROC is to change the decision threshold of an algorithm. The AUC defined by one run is widely known as balanced accuracy:

$$AUC_b = (sensitivity + specificity)/2.$$

3.6. Summary

In this chapter the SVM, Neural networks, and PSO were reviewed as the main algorithms used in automated sleep apnea studies. Also most of the important researches in the area of sleep apnea area by these methods were reviewed.

In the studies related to sleep apnea based on SVM or ANNs, many different signals and features were used. One of the main disadvantages of previous work can be the lack of study on signal selection and feature selection in sleep apnea studies. Another disadvantage is the lack of strong feature selection in most previous researches, and many of these studies did not provide enough information about their experimental settings and parameters.

Chapter 4

Sleep Apnea Detection and Classification

4.1. Introduction

This chapter consists of two main parts; in the first part, sections 4.2 till 4.5, proposed algorithms for sleep apnea detection and classification will be introduced. The second part, section 4.6, comprises of different experiments related to sleep apnea detection and classification employed by the proposed algorithms in the first part. Figure 4.1 shows the general process for detecting or classifying sleep apnea.

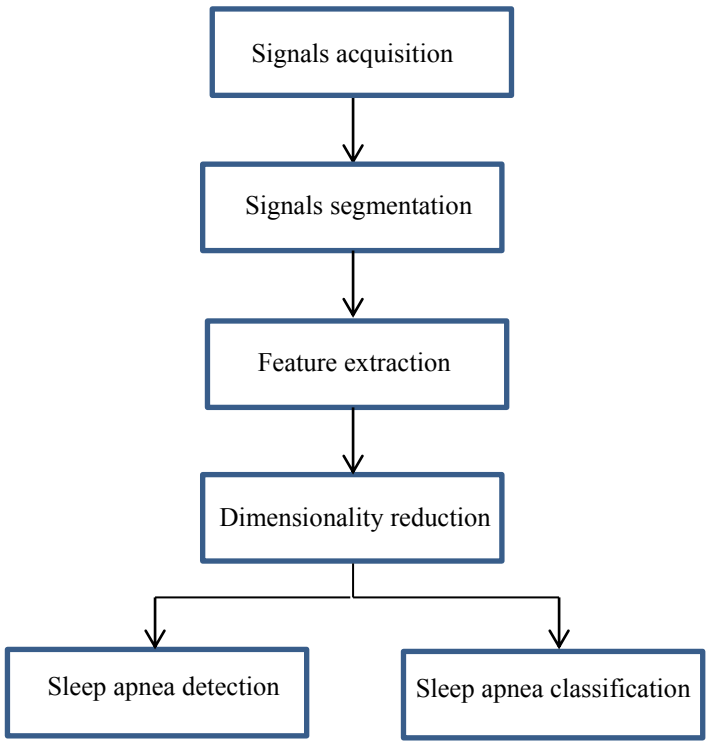

Figure 4.1: Process of the sleep apnea detection or classification

Signal acquisition: Signals from polysomnograms records which obtained from Concord Repatriation General Hospital in Sydney, Australia, were used in this thesis. More information about the database and these signals are presented in the second part of this chapter, in section 4.6.1.

Signal segmentation: Here, in this thesis, signal segmentation means dividing one-dimensional signals into several time windows, epochs. Segmentation is so important since just these specified time windows will be considered in the next steps. The proposed algorithm for signal segmentation is introduced in section 4.2 and related experiment about the performance of this algorithm is presented in section 4.6.2.

Feature extraction: For each segment features are extracted, or in another word generated, from signals belong to the corresponding segment. The proposed pre-processing and feature extraction approaches are presented in section 4.3.

Dimensionality Reduction: In this thesis two approaches are proposed for dimensionality reduction as; 1- feature selection, 2- feature and training data reduction.

For the first approach, feature selection, a PSO-SVM algorithm is proposed in section 4.4.1, this feature selection is employed in different experiments of section 4.6.

In the second approach, feature and training data reduction, together with reducing the features a subset of the training data is selected. More details of this approach are presented in section 4.4.2.

Sleep apnea detection and classification: In this thesis sleep apnea detection means specifying the generated segments to sleep apnea or normal sleep by attention to the extracted features. The detected segments as sleep apnea can be classified to hypopnea, obstructive, central or mixed apnea in the classification phase.

For sleep apnea detection two paradigms are examined as; (1) normal sleep apnea detection or subject dependent and (2) subject independent sleep apnea detection.

In the first paradigm the machine learning approach trained and validated on each participant individually. However the second paradigm detect sleep apnea event in new participant whose signals have never been used to train the machine learning algorithm. More details of these paradigms are presented in section 4.6.8.

4.2. Signal Segmentation

Sleep apnea events last at least 10 seconds and consequently the detection of SA events means looking at continuous time windows and detecting them as either SA or normal sleep. Therefore, in the first step the input signals should be segmented to time windows, and then these segments can be reviewed to be considered as sleep apnea or normal sleep. There are at least two strategies for segmenting input signals;

The first approach is blind segmentation where the signals are segmented from the first second without any further consideration. In this approach each segment is started exactly from the place that previous segment is finished, therefore segments are determined independent to the properties and characterise of the related signals and segments cover whole length of signals.

The second strategy for segmentation selects time windows that have more chance of containing SA events. The resulting segments in this approach are determined based on the properties and characterise of the related signals and some parts of the signals are not covered by the resulting segments.

It should be noted that blind segmentation can reduce the final accuracy, as shown in Figure 4.2, where in Figure 4.2(a) a whole sleep apnea event is correctly located in one segment, but in 4.2 (b) the length of the sleep apnea event in each segment is less than

10 seconds. This means that this sleep apnea event cannot be detected because the length in each segment is less than 10 seconds. The blind segmentation may leads to the situations similar to Figure 4.2 (b) more than another approaches.

In this thesis a segmentation algorithm is proposed to follow the second strategy. The selected length of each segment is considered as 30 seconds based on a consideration of state of the art in sleep disorders, especially in sleep apnea literature. These segments named as Reasoning Units (RUs) [281] and each RU represents an interval that has more chance to contain a sleep apnea event.

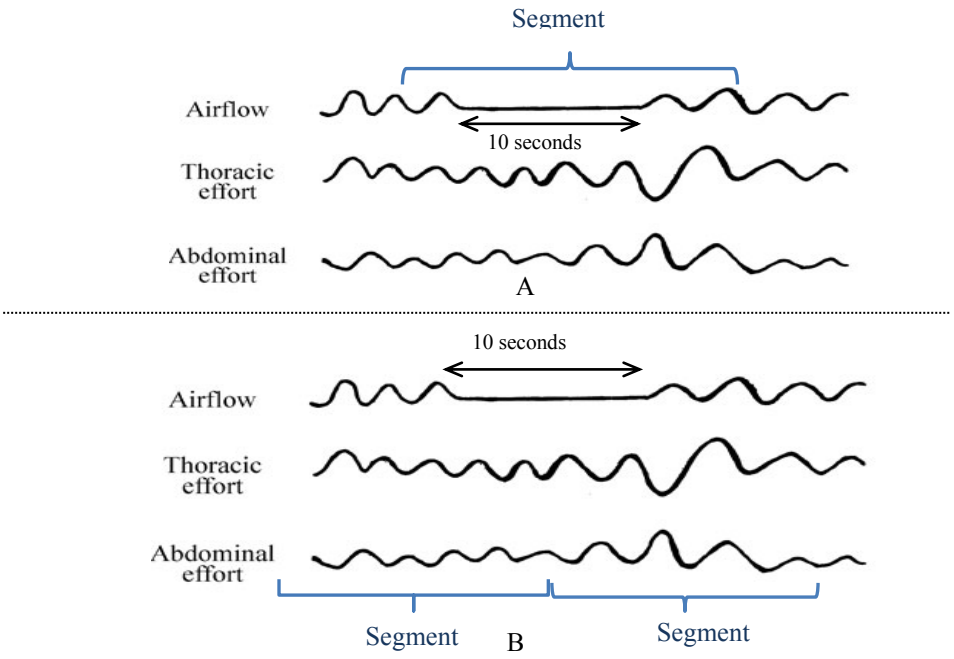

Figure 4.2: Two samples of signal segmentation

The proposed signal segmentation algorithm is based on the [282], with some modification. The original algorithm [282] only used airflow for detecting sleep apnea events but this modified version considered three signals as input to generate RUs. The proposed segmentation algorithm generates some RUs that are not overlapped. The Pseudo code 4.1 presented the proposed segmentation algorithm.

Pseudo code 4.1: *The proposed segmentation algorithm*

1: ***Pre-processing:*** *Consider three input signals and set one of them one of them as the main signal. Each signal is normalised to -1 and 1, and if any of the input signals are missed for more than 60 seconds, those time windows are labelled as not usable and will not considered by the algorithm anymore.*

2: ***Amplitude calculator:*** *Difference between local maximum and local minimum of 10 consequence data, one second, are computed as amplitude value (v_i) for each second.*

3: ***Amplitude classifier****: The amplitude reference value (AVREF) is computed as the last maximum amplitude value in the previous two minutes, if any amplitude value (v_i) is greater than the 0.25*AVREF, it is classified as normal. Otherwise it is classified as a feasible event.*

4: ***Amplitude reviewer:*** *The classified amplitude vectors are reviewed in this step by completing following steps a and b:*

(a) Because a complete respiratory cycle lasts at least 3 seconds, those normal amplitude values located between feasible events that are not separated by more than 3 seconds are re-classified as feasible event.

(b) The amplitude values that primarily classified as a feasible event will be classified as normal if their duration is less than 10 seconds.

5: ***RU generation:*** *In the first step the amplitudes of the main input signal are considered and for each amplitude value, which is labelled as an event, one RU with a length equal to 30 seconds is created. After generating RUs related to the first signal, feasible events of second and third signals are reviewed. In this step a new RU will be created between previous RUs only if a feasible event can be found in the amplitudes of the se-*

cond or third signals, and there is a room for the new RU between the previously made RUs.

Therefore each RU is a time window which indicates a segment of time that we are more interested to have deeper analyses on them, instead of considering the whole length of signals.

4.3. Feature Generation

After signal segmentation the next step is feature generation. In this stage features are generated for each RU from a section of signals that belongs to the corresponding RU.

For feature generation we cannot use the original signals since bio signals, especially signals related to sleep study, vary in amplitude over time. For these signals depending on the sleep stage, when the amplitude decreased, several false apneas could be detected, and on the other hand when the averaged amplitude of the signal increased, some apneas could be not detected [291]. To overcome this problem, the original signal in each RU was normalised by considering the average and variance of a sliding window that started from 15 seconds before and ended after 15 seconds of each 30-second RU; this means that 60-second sliding window is considered.

After normalisation, wavelet packet coefficients were used to generate features for each of these signals. Although Fourier-based methods are standard methods for frequency analysis, they are not well suited for the analysis of non-stationary signals, such as bio signals [283]. Therefore wavelet packet is selected in this thesis. A wavelet packet is a generalization of a Discrete Wavelet Transform (DWT) such that each octave frequency band of the wavelet spectrum is further subdivided into finer frequency bands by using the two-scale relations repeatedly [283].

Different wavelet family were selected for each signal based on state of the arts. The details of feature generation for each signal are presented as follows:

- Airflow, Abdominal movement, Thoracic movement, SPO2 and Snore

Daubechies wavelet packet with order 3 and 4 levels were applied to these signals and features were generated by applying the statistical measures from Table 4.1 to each coefficient at the terminal nodes of the wavelet tree [97, 284, 285], in this Table x represents coefficients of the wavelet packet.

Table 4.1: List of statistical features

$\log(\text{mean}(x^2))$	$\text{kurtosis}(x^2)$	$\text{geomean}(\lvert x\rvert)$
$\text{std}(x^2)$	$\text{var}(x^2)$	$\text{mad}(x)$
$\text{skewness}(x^2)$	$\text{mean}(\lvert x\rvert)$	$\text{mean}(x^2)$
$\text{skewness}(x)$	$\text{kurtosis}(x)$	$\text{var}(x)$
$\text{geomean}(x^2)$	$\text{mad}(x^2)$	$\text{std}(x)$

- EEG

Daubechies wavelet packet with order 2 and depth 7 (7 levels) was applied. After that the frequency ranges of the EEG signal were broken down into Delta (below 3.5 Hz), Theta (4-7 Hz), Alpha (8-13 Hz), and Beta (14-30 Hz) bands [286], and finally the following features were used to represent the time–frequency distribution of the EEG signals [286]:

1. Average quadratic value or Energy of wavelet packet (WP) coefficients for each of the sub bands,

2. Total Energy,

3. Ratio of different Energy values,

4. Average of the absolute values of the coefficients in each sub band, and

5. Standard deviation of the coefficients in each sub band.

Furthermore, more features were generated by applying the statistical measure of Table 4.1 to each coefficient at the terminal nodes of the wavelet tree.

- EOG

To generate the features from the EOG signal, an 8-level Daubechies wavelet packet with an order 3 was used. After that features were generated by applying measures of Table 4.1 to each coefficient of the last layer. Furthermore, the normalised correlation coefficient between the left and right EOG signals was considered as another feature [287].

- EMG

For the EMG signal features were generated by applying statistical measures from Table 4.1 to the coefficients of last layer of 2-level Daubechies with the order 2 [284].

- ECG

Feature for ECG signals were generated by applying statistical measures from Table 4.1 to each coefficient of the last layer of an order 3 Daubechies wavelet packet with 8 levels [288]

By considering this features generation procedure, 208 features were generated for each RU.

4.4. Dimensionality Reduction

In this thesis two approaches are proposed for dimension reduction, the first one is feature selection and the second one is feature and training data selection. Output of the first approach is the whole training data with subset of their features and output of the second approach is subset of training data with subset of their features.

Both of these two approaches are employed through PSO-SVM implementation. For the PSO part either a single weighted PSO proposed by Shi and Eberhart [268] or a new proposed parallel PSO, named Hierarchical Multi Master PSO (HMM-PSO) can be used. Details of the proposed parallel PSO, HMM-PSO, can be found in section 6.3. For the SVM part, traditional SVM or the new proposed SVM algorithm, named Self-Advising SVM, SA-SVM, can be used. Details of the proposed SA-SVM approach can be found in section 6.2, this algorithm is published in International Journal of Knowledge-Based Systems [289].

Presented algorithms in this section have been published in different papers such as a book chapter entitled “Hierarchical Parallel PSO-SVM Based Subject-Independent Sleep Apnea Classification” published in Lecture Notes in Computer Science (LNCS) on Neural Information Processing (ICONIP) [290], and a paper entitled " A Novel Partially Connected Cooperative Parallel PSO-SVM Algorithm: Study Based on Sleep Apnea Detection” which is presented at IEEE World Congress on Computational Intelligence [291], or another paper entitled “Self-Advising SVM for Sleep Apnea Classification” which is presented in Workshop on New Trends of Computational Intelligence in Health Applications [292].

Details of these two approaches for data reduction are as follows;

4.4.1. Feature Selection

In this thesis a PSO-SVM approach is proposed for feature selection. The main task of this PSO-SVM algorithm is to select a subset of features and tuning the parameters of SVM.

For implementing this algorithm, samples should be divided into train, test and validation. The PSO-SVM algorithm used train and test sets in order to optimize the feature subset. Finally a completely unseen validation set is used to measure the generalization capability of the proposed system. As mentioned before, the PSO-SVM is used to select the best combination of features from the original features. In this approach each particle is an array with two parts. The first part is an array with two cells related to the SVM parameters gamma and cost and each of them can get a value between 2^{-5} to 2^{5}. The second part is an array with 208 cells, related to 208 features, that contains weights, numbers between 0 and 1. These weights shows the importance of coresponding features and in each iteration features with the weight higher than a specified threshold, are selected to perform the detection or classification problem.

To compute the fitness of each particle; SVM algorithm is trained by the training set and the selected features from the coresponding particle and the performance of the SVM on the test set with the corresponding feature set is considered as the fitness of that particle. In this thesis the F-score is used as the classification performance indexes. Figure 4.3 shows the flowchart of the proposed PSO-SVM for feature selection.

The stopping condition is set to 100 iterations for the PSO and 0.5 is selected as the threshold to select weight corresponding to features. After finishing of this algorithm, the best subset of features and setting for the SVM is reported for the next step which is sleep apnea detection or classification.

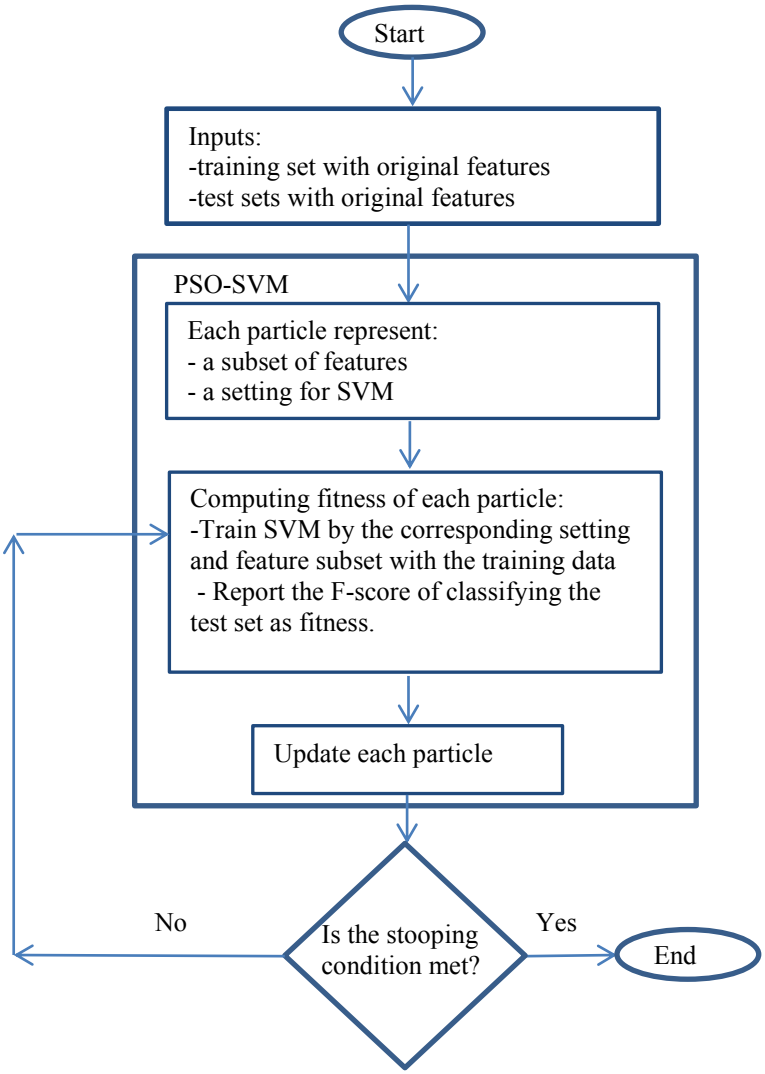

Figure 4.3: Flowchart of the PSO-SVM algorithm for feature selection

Note: Particles are generated randomly, except the first particle. For the first particle cells related to the features are set as 1, which means using all of the features. Therefore using all of the features is also considered by the algorithm.

4.4.2. Feature and Training Data Selection

In this thesis together with feature selection, selection of a subset of training data from the available training set is also proposed. This algorithm checks if as well as feature selection, reducing training data can result in better performance or not. Training pattern reduction or instance selection has been considered in different areas previously [293-295].

The general process of this algorithm is same as the PSO-SVM algorithm for feature selection that described in section 4.4.1. But here each particle consisted of three arrays, where two arrays related to subset of features selection and SVM settings are the same as the 4.4.1. The third array is related to selecting a subset of training data, length of this array is equal to the number of the train data. Each cell in the third array can get a number between zero and one, and if the value of a cell is higher than 0.5 then the corresponding training data is selected. Pseudo code 4.2 shows the proposed algorithm for feature and training data reduction.

Pseudo code 4.2: The proposed algorithm for feature and training data reduction

---

1- Specify the train, test and validation data sets

2- Initialize the PSO.

3- Do until the maximum number of iterations is reached {

4. For each particle

i. Specify the corresponding features subset, training data subset and SVM parameters according to the particle.

ii. Use the SVM with the selected training data to classify the test set.

iii. Set the SVM's F-score as the fitness of the particle.

5. Update particles

6. Go to 3.

7- End Do}

8- Report the best features subset, training subset and the SVM parameters

---

Note: Particles are generated randomly, except for two particles. For the first particle, cells related to the features are set as 1, which means using all of the features. For the second particle, cells related to the training data are set as 1, which means using all of the training data. Therefore using all of the training data is also considered by the algorithm.

4.5. Signal Selection

In this thesis an algorithm utilized to rank the signals based on their importance in detecting apnea events. In this algorithm different combinations between the PSG signals were examined. The results of this work can be useful for selecting the best subset of input signals for detecting or classifying sleep apneic events. This work has been published entitles "Signal Selection for Sleep Apnea Classification" as a book chapter in Lecture Notes in Computer Science (LNCS) on Neural Information Processing (AI 2012) [296].

In the first step, the data from all the patients were separated into the train, test, and validation. Train and test sets are used by algorithm to find a subset of *k* signals with best performance and unseen validation set is used to measure the generalization capability of the selected signals.

Three main tasks were done by PSO-SVM algorithm;

-Searching for the best set of *k* signals, for *k=1, 2,… ,12*.

-Searching the best subset of features for selected signals

-Tuning of SVM's parameters.

In this approach each particle is consists of two part. The first part is an array with two cells related to the SVM parameters gamma and cost and each of them can get a value between 2^{-5} to 2^{5}. The second part an array with a length equal to 208 that contains

weights related to the features of each signal. These features are ordered from the first signal to the last one, as shown in Figure 4.4.

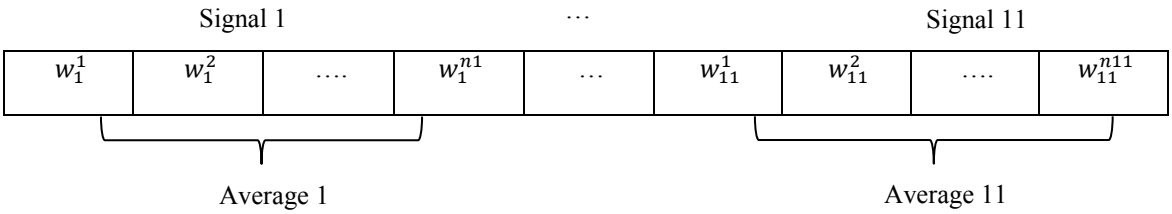

Figure 4.4: Signal selection by the proposed particle structure

Where w_i^j is the weight of the jth feature for the signal number i, and n_i is number of features for signal i. These weights, Figure 4.4, shows the importance of the corresponding features, and where in each iteration the k signals, $k=1,2,\ldots,12$ with the highest average of weights are selected. After selecting the k signals with highest average, features of the selected signals are chosen from features with weights higher than a specified threshold, 0.5.

Each particle is evaluated based on its ability to perform on the test set; F-score of this detection was then considered as the fitness of that particle. After 100 iterations for the PSO-SVM algorithm, the selected signals and their features are reported to the next step. In the next step the selected signals and selected feature and tuned SVM parameters are used by an SVM to detect the unseen validation data, and where the performance of this detection was reported as the performance of the selected signals.

4.6. Experiments and Results

In this section first of all the used database is introduced, and then different experiments related to the detection and classification of sleep apnea that have been employed by the proposed methods in this chapter, are presented.

4.6.1. Database Materials

The polysomnograms (PSG) of 30 subjects (12 women and 18 men) with ages ranging from 20 to 77 years and an average age of 49.4 years is used in this thesis. These data were provided through collaboration with Dr Leon Laks, sleep disorder specialist that acquired in by the Concord Repatriation General Hospital in Sydney, Australia. Each of these data contained 12 signals which are described in Table 4.2.

Table 4.2: Description of database signals

	Signal	**Signal description**
1	EEG C3-A2	Electroencephalography
2	EEG C4-A1	Electroencephalography
3	EEG A1-A2	Electroencephalography
4	EOG left	Electrooculography from left eye
5	EOG right	Electrooculography from right eye
6	EMG chin	Electromyography from chin muscle
7	ECG	Electrocardiography
8	Patient flow	Patient air flow
9	Snore	Snore sound
10	Thoracic effort	Thoracic effort
11	Abdominal effort	Abdominal effort
12	SpO2	Blood oxygen saturation

Figure 4.5 [297, 298], shows a basic setup for PSG recording. As it is clear from this figure, collecting all of these signals is inconvenience for the patient. Therefore one of the main objectives of this thesis is to select fewer signals and select those signals that are more convenience for the patient.

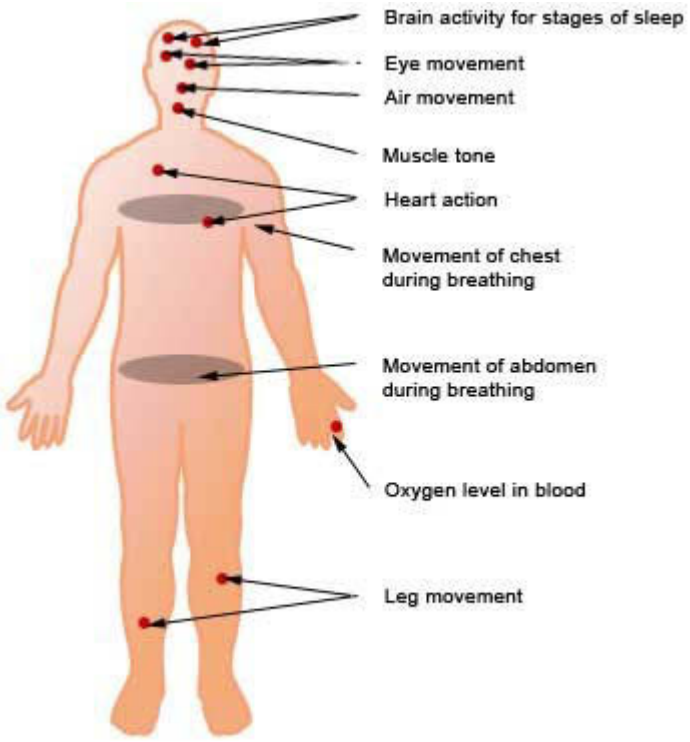

Figure 4.5: Basic polysomnograms setup [297, 298]

The PSG record was scored manually and evaluated by experts with extensive experience in interpreting sleep data, and who rated the recordings used for this thesis. A total of 4527 apneic were found, which means an average of 150.9 events per sample. Figure 4.6, shows distribution of apneic events in each sample.

Figure 4.7 shows that there are a wide variety of apnea events; 3 samples with around 300 events, and there are 13 samples with less than 100 events. From these total events, 2042 were obstructive, 378 were central, 306 were mixed, and 1801 were hypopnea.

By attention to these records, apneic events last between 10 and 88 seconds, with an average of 20.77 seconds, and Figure 4.2 shows their distribution. This figure shows the percentages of events with duration between 10 to 20 seconds; 20 to 40 seconds; 40 to 60 seconds, and 60 to 90 seconds for each of the obstructive, central, mixed, and hypopnea events.

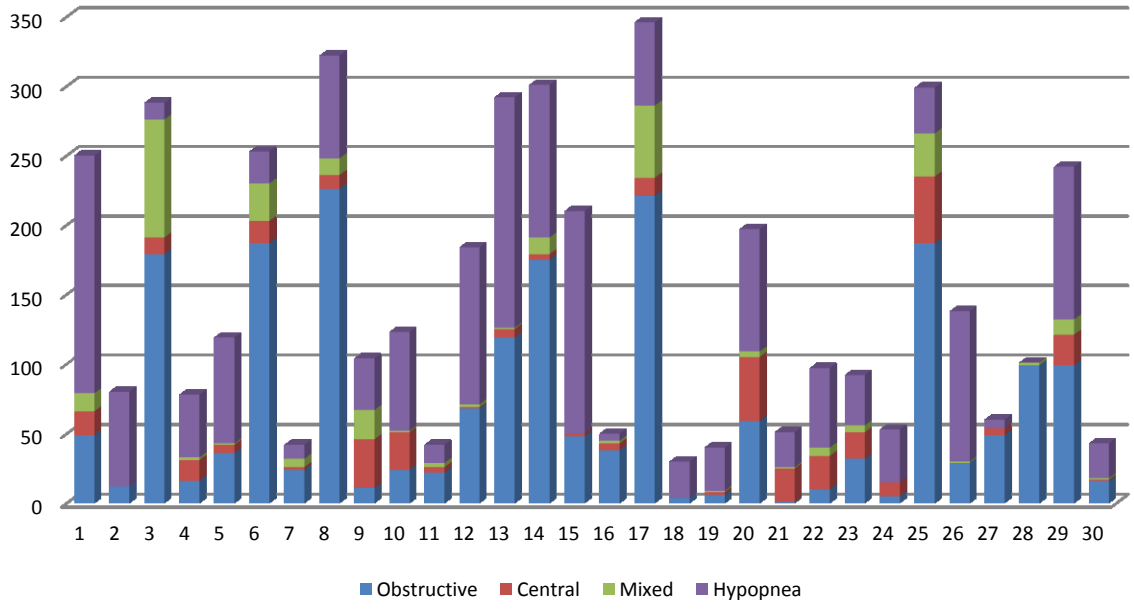

Figure 4.6: Distribution of apnoeic events

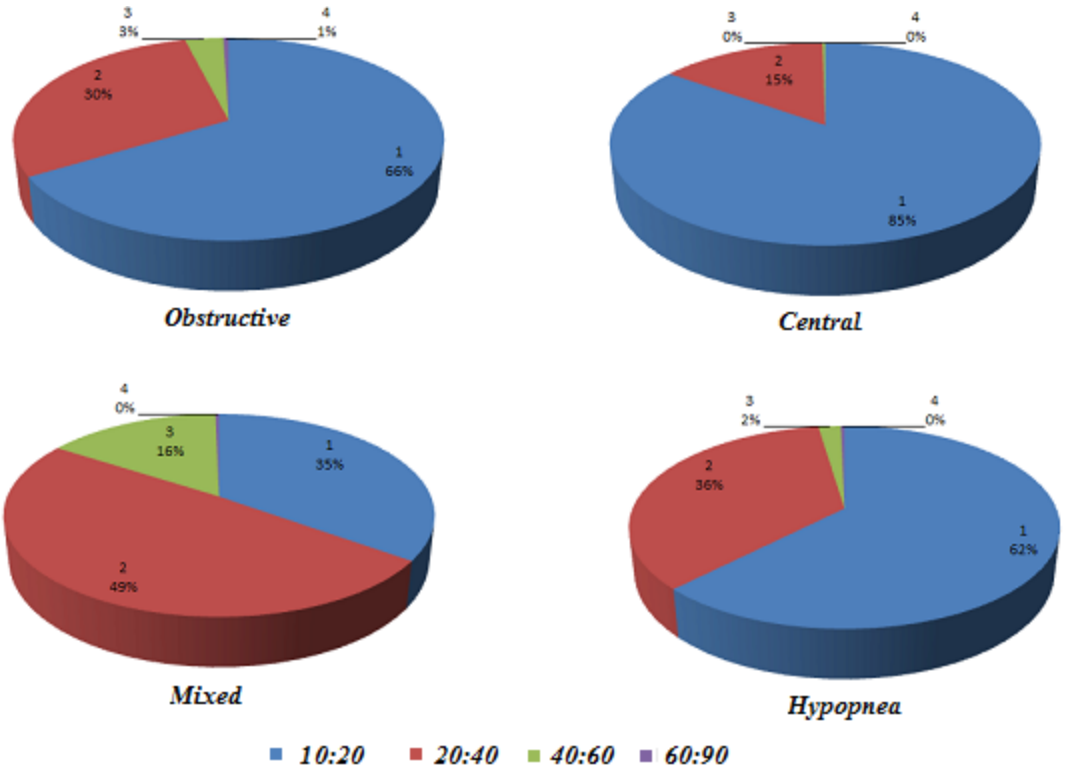

Figure 4.7: Distribution of the duration of apneic events

Obstructive events usually last between 10 and 88 seconds, with an average of 19.80 seconds, central events have duration between 10 and 51.5 seconds, with an average of 15.65 seconds, and mixed events have duration between 10.5 and 68 seconds, with an average of 27.53 seconds. Hypopnea events also have durations between 10 to 79 seconds, with an average of 20.12 seconds.

4.6.2. Signal Selection for Detecting Sleep Apnea

The first step of this thesis wanted to rank the available signals by attention to their impact on sleep apnea detection, the proposed algorithm in this section only considered the performance of signals and other issues such as those signals that easily can be obtained from patient, the cost of acquisition and etc. had not been considered.

The PSO-SVM introduced in section 4.5, with 40 particles and LIBSVM are used here as a common library for support vector machines. This library is available at http://www.csie.ntu.edu.tw/~cjlin/libsvm [299] RBF kernel is also selected for LIBSVM.

First of all RUs are generated by using airflow, abdominal and thoracic movements signals, it is followed by generation of the features for all of 12 signals for each RUs. Then each of the 30 samples is divided into 3 equal sets as, train, test, and validation, randomly. The first two sets were used by PSO-SVM, described in section 4.5 and the final accuracy and F-score are obtained by detecting the validation sets.

Table 4.3 shows the rank of signals and the average performance of each set of signals for detecting the apnea events in the validation sets. It can be noticed that features of the right and left EOG are combined as one signal.

Table 4.3 shows that using more signals did not necessarily lead to better accuracy. One reason for having less accuracy with more signals can be over training by more fea-

tures and therefore lower accuracy for detecting the unseen validation set. It should be noted that this ranking is used by the proposed feature generation method, whereas using other features or even another classifier can lead to a different ranking.

Table 4.3: Signal selection for sleep apnea detection

	Selected signals	**Accuracy**	**F-score**
1 signal	Snore	79.05	0.7436
2 signals	Snore – ECG	79.26	0.7512
3 signals	Snore – ECG – Thoracal	82.93	0.7521
4 signals	Snore – ECG –Thoracal – EMG	84.35	0.8015
5 signals	Snore – ECG –Thoracal – EMG – Abdominal	85.92	0.8062
6 signals	Snore – ECG –Thoracal – EMG – Abdominal – Air flow	86.54	0.8137
7 signals	Snore – ECG –Thoracal – EMG – Abdominal – Air flow – Oxygen saturation	85.15	0.8215
8 signals	Snore – ECG –Thoracal – EMG – Abdominal – Air flow – Oxygen saturation – EEG A1-A2	88.58	0.8452
9 signals	Snore – ECG –Thoracal – EMG – Abdominal – Air flow – Oxygen saturation – EEG A1-A2 – EEG C3-A2	86.77	0.8215
10 signals	Snore – ECG –Thoracal – EMG – Abdominal – Air flow – Oxygen saturation – EEG A1-A2 – EEG C3-A2 – EOG	85.84	0.8119
11 signals	Snore – ECG –Thoracal – EMG – Abdominal – Air flow – Oxygen saturation – EEG A1-A2 – EEG C3-A2 – EOG – EEG C4-A1	81.96	0.8011

The rest of this work will use thoracic and abdominal movements with oxygen saturation as input signals. The main reasons for selecting these signals was not just considering their performance in detecting sleep apnea, but also those signals were considered that can be obtained easily and not get influence by the environment. This study targeted to improve the performance of the selected signals for studying sleep apnea by feature selection and using more powerful machine learning techniques in the rest of this chapter.

4.6.3. A Comparison on Signal Segmentation

After choosing input signals such as thoracic movement, abdominal movement, and oxygen saturation, the proposed segmentation approach is compared with blind segmentation. As described before in blind segmentation, signals were segmented from the beginning point by 30-second windows.

It should be noted again that airflow is not considered anymore in our final signals therefore in the proposed segmentation thoracic movement was used as the main signal, instead of airflow, and abdominal and oxygen saturation were used in the second level. Here the PSO-SVM algorithm described in section 4.4 with single PSO and traditional SVM is used for feature selection. Table 4.4 shows the average results of these two segmentation methods with LIBSVM as the classifier for 5 independent runs. In each run data of each patient is divided into three parts as train, test and validation, randomly. Train and test sets are used by the PSO-SVM algorithm for feature selection. The reported accuracy or F-score in table 4.4 are based on detection of sleep apnea events in the validation set.

Table 4.4: performance of the proposed segmentation approach and Blind segmentation

Samples	Proposed Segmentation			Blind Segmentation		
	Number of RUs	Accuracy	F-score	Number of RUs	Accuracy	F-score
#1	738	94.71±1.7	0.91±0.02	825	94.78±1.1	0.90±0.03
#2	598	94.47±2.6	0.75±0.11	881	95.16±0.5	0.71±0.08
#3	720	89.95±2.3	0.87±0.07	882	87.82±2.6	0.84±0.13
#4	630	97.61±0.5	0.89±0.05	812	90.38±1.7	0.86±0.06
#5	726	96.68±1.2	0.88±0.04	836	84.99±1.1	0.80±0.10
#6	778	90.73±3.1	0.84±0.10	842	88.63±1.3	0.78±0.04
#7	738	97.77±0.4	0.85±0.05	886	90.43±2.2	0.72±0.13
#8	687	92.13±2.4	0.90±0.06	860	82.39±3.5	0.81±0.04
#9	438	95.17±1.0	0.89±0.04	492	84.63±1.4	0.83±0.02
#10	651	97.7±1.2	0.92±0.02	725	90.72±1.2	0.81±0.06
#11	832	97.91±1.1	0.88±0.06	930	83.38±1.7	0.79±0.01
#12	683	95.61±1.3	0.91±0.03	850	85.23±3.1	0.80±0.06
#13	708	92.76±2.2	0.90±0.03	798	87.47±0.2	0.81±0.04
#14	739	91.49±1.5	0.88±0.09	826	84.49±1.0	0.81±0.06
#15	596	86.93±1.3	0.80±0.06	869	86.77±1.1	0.77±0.05
#16	645	95.79±0.8	0.66±0.07	895	90.57±1.3	0.72±0.04
#17	766	92.96±2.4	0.91±0.02	888	86.78±1.2	0.88±0.07
#18	715	97.16±0.5	0.90±0.03	822	81.05±3.4	0.83±0.03
#19	538	96.11±1.2	0.58±0.26	823	87.00±1.7	0.58±0.08
#20	739	92.27±2.4	0.84±0.10	865	87.89±0.03	0.83±0.05
#21	488	97.53±0.5	0.77±0.07	842	86.94±1.0	0.73±0.03
#22	682	96.03±1.7	0.81±0.05	832	86.69±3.6	0.79±0.02
#23	739	96.35±1.5	0.83±0.04	860	86.51±1.3	0.80±0.10
#24	732	96.73±0.8	0.73±0.17	920	81.58±4.2	0.72±0.04
#25	763	91.76±2.3	0.87±0.02	840	84.14±1.8	0.84±0.09
#26	694	95.68±0.7	0.88±0.08	852	86.77±1.8	0.81±0.12
#27	612	95.58±1.3	0.60±0.09	777	89.99±2.1	0.73±0.02
#28	752	96.01±1.1	0.80±0.04	813	88.24±1.4	0.75±0.05
#29	762	96.45±1.1	0.93±0.01	857	85.50±1.1	0.82±0.01
#30	620	97.08±1.0	0.75±0.11	788	98.24±0.7	0.80±0.15

Based on these results, average F-score of the proposed approach and the blind segmentation are 0.831 and 0.789, respectively. The t–test with a level of significance of $\alpha = 0.05$ shows that differences of F-score obtained by these two approaches were statistically significant, by a p-value of 0.0001. If accuracy is considered, the average accuracy of the proposed approach and the blind segmentation are 94.73 and 87.50, respectively, the

p-value for the t-test of average of accuracies is equal to 4.5×10^{-10}, which also shows that difference between accuracies of these two approach is statistically significant.

These results and statistical analysis show that the proposed segmentation algorithm works significantly better than the blind segmentation.

4.6.4. Dimensionality Reduction Approaches for Sleep Apnea Detection

After selecting signals and the segmentation phase, features are generated by attention to the selected signals for each RU. In this section, different approaches for dimension reduction are examined.

In the first step two approaches that proposed in section 4.4 are compared for sleep apnea detection. SVM were considered as the classifier with RBF kernel and a single PSO with 20 particles is choosed with c_1 and c_2 set to be 2.0, the threshold for selecting a feature or training instance is set to be 0.5. Table 4.5 shows results of 5 independent runs of feature selection algorithm implemented by the PSO-SVM algorithm described in 4.4.1, and results of feature and instance selection described in section 4.4.2 together with results obtained with original data set.

The average F-score for these approaches were as 0.7808, 0.85 and 0.8354 respectively. It is very clear that using dimensionality reduction approaches can improve the result in compare with using the original data. To evaluate between these two dimension-reduction approaches, the pair t-test was used, where the p value of t-test was 0.096. These statistical tests indicated that the results obtained by these approaches are not statistically different. Therefore we can use feature selection since it has simpler structure.

Table 4.5: Dimensionality Reduction Approaches

	Original Data set		Feature selection		Feature and training instance selection	
	Accuracy	**F-score**	**Accuracy**	**F-score**	**Accuracy**	**F-score**
#1	78.45±2.2	0.7469±0.04	94.37±1.4	0.8324±0.03	93.14±2.5	0.8215±0.05
#2	80.37±2.3	0.7934±0.12	95.41±0.9	0.8628±0.07	96.39±1.6	0.8524±0.03
#3	79.13±3.4	0.8012±0.09	93.84±1.7	0.8514±0.03	93.48±2.7	0.8558±0.07
#4	81.12±1.6	0.7883±0.13	95.42±2.1	0.8496±0.03	93.28±3.1	0.8374±0.07
#5	80.73±3.1	0.7746±0.06	93.45±1.2	0.8538±0.06	94.15±1.6	0.8554±0.10

4.6.5. Comparing Traditional SVM with SA-SVM

While SVM is used in this thesis, some patterns were misclassified in the training phase. This means that even the classifier cannot classify the training data or the seen data, correctly. During this thesis it has been found that SVM cannot classify patterns that were close to the misclassified data in the training. Although the labels are provided to the SVM, the SVM could not use all of the knowledge provided from the train phase. Therefore more knowledge should be used from the training phase in compare with the traditional SVM; this resulted in the method named Self-Advising SVM (SA-SVM). SA-SVM was better at classifying general binary classification [289], and here the performance of the SA-SVM was examined in comparison to the traditional SVM. For more details of SA-SVM please refer to section 6.2.

Table 4.6: Classification performance of SVM and SA-SVM

samples	SVM		SA-SVM	
	Accuracy	f-score	Accuracy	f-score
#1	93.27±1.6	0.8975±0.03	93.27±1.6	0.8975±0.03
#2	91.85±3.6	0.7839±0.11	91.85±3.6	0.7839±0.11
#3	89.91±2.7	0.8839±0.06	92.53±1.8	0.9428±0.02
#4	95.28±1.7	0.8896±0.04	95.28±1.7	0.8896±0.04
#5	96.58±1.1	0.8749±0.06	96.58±1.1	0.8749±0.06
#6	90.18±2.4	0.8562±0.08	91.79±2.2	0.8893±0.07
#7	98.18±0.6	0.8372±0.11	98.18±0.6	0.8372±0.11
#8	93.64±3.3	0.9037±0.03	93.64±3.3	0.9037±0.03
#9	95.38±1.4	0.8971±0.07	95.38±1.4	0.8971±0.07
#10	97.82±0.8	0.9428±0.03	98.13±0.5	0.9514±0.02
#11	98.86±0.4	0.8863±0.05	98.86±0.4	0.8863±0.05
#12	95.62±2.1	0.9038±0.03	94.83±1.5	0.8974±0.07
#13	92.63±2.6	0.8957±0.03	95.29±0.9	0.9136±0.04
#14	91.17±3.2	0.8639±0.11	91.17±3.2	0.8639±0.11
#15	85.39±5.1	0.8283±0.12	85.39±5.1	0.8283±0.12
#16	94.29±1.3	0.6482±0.04	95.73±1.8	0.8754±0.07
#17	92.48±2.4	0.9027±0.02	92.48±2.4	0.9027±0.02
#18	98.37±0.8	0.8563±0.05	98.37±0.8	0.8563±0.05
#19	97.38±1.1	0.6892±0.09	97.38±1.1	0.6892±0.09
#20	90.48±2.5	0.8739±0.07	94.63±0.9	0.8968±0.06
#21	95.38±2.6	0.8138±0.12	94.73±1.3	0.8759±0.07
#22	95.73±0.9	0.8047±0.11	96.83±1.2	0.8529±0.03
#23	97.31±1.7	0.8569±0.09	97.31±1.7	0.8569±0.09
#24	95.38±1.6	0.7527±0.11	96.83±0.8	0.8439±0.10
#25	92.63±2.4	0.8839±0.07	92.63±2.4	0.8839±0.07
#26	95.73±1.4	0.8874±0.02	97.49±0.9	0.8952±0.04
#27	96.21±2.3	0.7382±0.05	96.21±2.3	0.7382±0.05
#28	97.48±0.7	0.8372±0.12	97.84±0.7	0.8372±0.12
#29	96.17±1.2	0.9472±0.02	96.17±1.2	0.9472±0.02
#30	96.49±1.5	0.8351±0.11	96.49±1.5	0.8351±0.11

Table 4.6 shows the average results of 5 independent runs of these two classifiers on 30 samples, SVM and SA-SVM, where both of these algorithms are implemented by LIBSVM. Here again in each experiment each sample is divided into three parts as train, test and validation, randomly. Train and test sets are used by the PSO-SVM algorithm for feature selection described in section 4.4.1. The reported accuracy and F-score in Table 4.6 are based on detection of sleep apnea events in the validation based on using traditional SVM or the proposed SA-SVM. It should be noted that proposed segmentation was used with the same parameters and kernel for both approaches.

By attention to the F-score, SVM reached an average of 0.8490 while the average F-score of SA-SVM was 0.8706, the p value related to the t-test by a level of significance of $\alpha = 0.05$ was 0.008, which shows the statistical significance of this difference between these two approaches.

If the accuracies of these two classifiers are considered; SVM reached an average accuracy of 94.57 while the average accuracy of SA-SVM was 95.10, the p value related to the t-test by a level of significance of $\alpha = 0.05$ was 0.006, which shows the statistical significance of this difference.

As a conclusion on this experiment, obtained result and statistical analyses show that SA-SVM outperforms the traditional SVM for detection of sleep apnea.

4.6.6. Comparing SA-SVM with Different Machine Learning Algorithms

SVM was used as the classifier in this thesis because of its performance in comparison with other classifiers. As mentioned before, in this thesis input signals are selected to have minimum contact with body, in this section we will compare different algorithms that used same inputs as this thesis, comparing the result with result of other algorithms

with different input signals are out of our scope. Results of the previous section 4.6.5 show that SA-SVM is better than tradition SVM in sleep apnea detection. Here SA-SVM is compared with other classifiers,

A total of 13 models are examined here using:

-Genetic-Fuzzy approach [300]

- Classic Fisher's linear discriminant (LDA)

- Quadratic discriminant (QDA)

- Self Advising Support Vector Machines (SA-SVM)

- 9 configurations of an Artificial Neural Networks (ANN) [301].

9 different configurations of a feed-forward neural network with one hidden layer, trained with a scaled conjugate gradient back propagation algorithm [302] were tested. A different number of neurons ranging from 20 to 100 were used in the hidden layer for each configuration.

PSO-SVM algorithm from section 4.4.1 is considered for feature selection. To select the best model and achieve a good estimation of the real F-score, these algorithms were run 5 independent times. The averages of accuracies and F-scores for 5 independent runs are presented in Table 4.7 which shows that the results of a discriminant analysis and genetic-fuzzy are not good compared to SA-SVM or ANNs.

Table 4.7: Average apnea classification by different classifier

	Accuracy	F-score
SA-SVM	95.29±1.1	0.8724±0.02
LDA	83.30±2.8	0.6380±0.07
QDA	85.63±1.2	0.6957±0.14
Genetic-Fuzzy	88.12±2.4	0.7157±0.09
ANN (20 neurons in hidden layer)	93.72±2.4	0.8131±0.11
ANN (30 neurons in hidden layer)	94.20±1.2	0.8092±0.03
ANN (40 neurons in hidden layer)	94.56±1.1	0.8032±0.05
ANN (50 neurons in hidden layer)	93.83±2.8	0.8056±0.07
ANN (60 neurons in hidden layer)	93.94±3.2	0.8026±0.01
ANN (70 neurons in hidden layer)	94.10±1.3	0.8298±0.06
ANN (80 neurons in hidden layer)	93.52±2.5	0.8138±0.08
ANN (90 neurons in hidden layer)	93.25±3.3	0.7808±0.13
ANN (100 neurons in hidden layer)	92.96±1.9	0.8157±0.1

By attention to results of these experiments, SA-SVM obtained an average F-score of 0.8724 and ANN obtained its best result as an average of 0.8298 with 70 neurons in hidden layer. Average F-score of Genetic-Fuzzy, LDA and QLDA were 0.7157, 0.6380 and 0.6957, respectively. The p value related to the ANOVA test by a level of significance of $\alpha = 0.05$ was 0.0001, which shows the statistically significant difference between these groups. To have a more comprehensive comparison t-test was performed to compare results of SA-SVM and the ANN with 70 neurons, those models with best results, p-value of the t-test significance of $\alpha = 0.05$ was 0.0008 which shows that f-scores of SA-SVM is statistically higher than f-scores achieved by the ANN with 70 neurons. Therefore it can be concluded that SA-SVM performed better than these algorithms.

4.6.7. Classification of Sleep Apnea Events

After detecting the RUs containing a sleep apnea event, these events are classified as hypopnea, central apnea, obstructive apnea, or mixed apnea. Here the feature selection is done by the proposed PSO-SVM approach that described in section 4.4.1. The algorithm was run 5 different times. RBF kernel was selected for the SVM, and single PSO structure was considered with 20 particles.

Table 4.8 tabulates the number of central, obstructive, and mixed events in each of the validation set and training for these 5 runs. This table shows that different numbers of each class were in these runs. Therefore average of performance on these 5 runs can be considered as a good indicator for the performance of the proposed classification approach in the general form.

Table 4.8: Diversity of each class in different runs

	Training				Validation			
	Obstructive	**Central**	**mixed**	**Hypopnea**	**Obstructive**	**Central**	**mixed**	**Hypopnea**
#1	635	112	102	523	655	114	121	591
#2	672	128	114	572	650	153	92	531
#3	638	139	98	518	602	121	128	542
#4	674	132	132	562	668	119	131	493
#5	648	173	104	502	636	125	72	514

The average accuracies for these two approaches were as 81.19 and 88.71 respectively. To evaluate between these results more reliably, the pair t-test was used, where the p value of t-test was 0.007. These statistical tests indicated that the results obtained by the feature selection were much better than the results from the all data.

Table 4.9: Sleep apnea classification accuracies

	Full Data	**With feature selection**
#1	82.24±2.9	89.14±1.3
#2	79.81±1.1	85.72±1.2
#3	84.38±1.7	90.74±2.1
#4	79.29±1.2	90.83±3.3
#5	80.27±1.2	87.12±3.1
Average	81.20	88.71

4.6.8. Subject Independent Sleep Apnea Detection

This section examines subject independent sleep apnea detection. The total data from all of the samples were first integrated as Meta data then two experiments are considered. In the first experiment, the feature selection is performed by proposed PSO-SVM from section 4.4.1. In the second experiment, the training data and feature selection is performed by the proposed PSO-SVM from section 4.4.2. Figure 4.8, represents these two experiments. Figure 4.8.A represents the first experiment and figure 4.8.B shows the second experiment. Therefore in the second experiment instead of the whole training data just a subset of it is used by SA-SVM. To overcome the impact of validation set on the final result 5 independent tests were examined. Same as before F-score was considered to measure the performance of these methods, in these experiments RBF kernel is selected for the SA-SVM, with single PSO. Table 4.9, shows the F-score obtained for these two experiments.

The average F-score for these methods were, 0.8116, and 0.8447, respectively, to obtain a reliable evaluation between the results of these methods; a pair t-test was used between experiment 1 and experiment 2. The p of the t-test with a level of significance of $\alpha = 0.05$ was 0.002 which shows the results of experiment 2, feature and data selection, were better than the results obtained by experiment 1, which is just feature selection. It should be mentioned that in experiment 2, average of 12.7% of instances were removed by the proposed algorithm from the training set.

Therefore, feature and training data selection outperformed just feature selection. In the next experiment single PSO is compared with the proposed parallel PSO, Hierarchical Multi Master PSO (HMM-PSO) for feature and data selection. Details of the HMM-PSO can be found in section 6.3. Table 4.11 mentions the F-score obtained for subject inde-

pendent sleep apnea detection with single or parallel PSO and using the data reduction algorithm introduced in section 4.4.2.

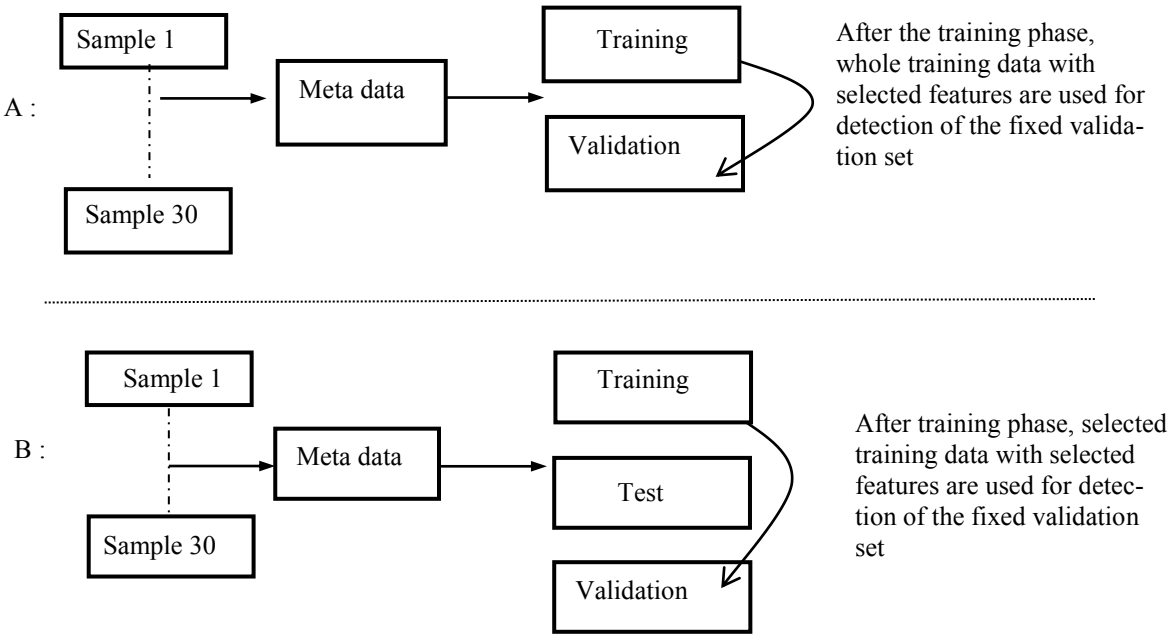

Figure 4.8: Subject dependent and independent sleep apnea detection

The average F-score for single PSO and the proposed parallel PSO are, 0.8381, and 0.8673, respectively. A pair t-test was used with a level of significance of $\alpha = 0.05$, p value of this test was 0.008 which shows the results of the proposed parallel PSO were statistically better than the results obtained by the single PSO for feature and data reduction.

4.7. Summary

This chapter reviewed algorithms proposed for detecting and classifying sleep apnea. The detection process consisted of segmentation, feature generation, dimensionality re-

duction and detection. After detection the apneic events can be classified to hypopnea, obstructive, mixed or central apnea. The main outcomes of this chapter are as follows:

Table 4.10: F-score of subject independent sleep apnea detection

	Experiment 1	**Experiment 2**
#1	0.8025±0.12	0.8292±0.07
#2	0.8172±0.12	0.8527±0.06
#3	0.8342±0.10	0.8473±0.10
#4	0.7975±0.09	0.8353±0.04
#5	0.8067±0.11	0.8553±0.04

Table 4.11: F-score of subject independent detection with single and parallel PSO

	Single PSO	HMM-PSO
#1	0.8178±0.11	0.8326±0.08
#2	0.8317±0.03	0.8243±0.14
#3	0.8026±0.14	0.8317±0.07
#4	0.8542±0.06	0.8617±0.10
#5	0.8274±0.12	0.8463±0.07

- **Segmentation:** A new segmentation algorithm was proposed in this thesis. Results showed that the proposed segmentation method can perform better than the blind segmentation. Segmentation is important in studying sleep apnea because an apneic event is defined by its duration so if the segmentation method cannot locate at least few seconds of an apneic event in a segment then the classifier cannot recognise that segmentation as an apneic in future steps.

- **Feature generation:** In this study generating features from input signals by using different families of wavelet packets were considered. In the first step as many features as possible, 208 features, were generated.

- **Dimensional reduction:** Two approaches were presented in this thesis for dimensional reduction. The first approach was just based on feature subset selection and the second one was feature and training data subset selection. Both of these algorithms were implemented by PSO-SVM approach.

- **Detection:** In this thesis thoracic movement, abdominal movement, and oxygen saturation were used as input signals because these signals are relatively easier to acquire and also these signals get less influence by surrounding factors. In the dimensional reduction step, the proposed PSO-SVM approach for feature selection was selected, based on its performance over the feature and training instance selection algorithm. SVM was selected as the classifier and its performance was improved by proposing SA-SVM approach. The average F-score of SA-SVM was 0.8706 and average accuracy was 95.10% for sleep apnea detection while SVM reached F-score of 0.8490 and accuracy of 94.57%.

- **Subject independence detection:** To have a fully automated sleep apnea detection system, a subject independent approach was designed to detect new samples without training the system on them. For this reason the proposed PSO-SVM approach was used for feature selection and proper training sample selection, implemented by a new proposed parallel PSO algorithm. Results of this section proved that feature and instance subset selection achieved the better result by F-score of 0.8447 and average accuracy of 89.13% while just feature selection resulted in F-score of 0.8116. Also

single PSO is compared with the proposed parallel PSO. The parallel PSO achieved F-score of 0.8673 versus F-score of 0.8381 for the single PSO. By attention to these results it is clear that using parallel structure for feature and training data reduction can lead to a better subject independent detection.

Chapter 5

Predicting Sleep Apnea

5.1. Introduction

Predicting means providing statements without an observed outcome; such as estimating the expected value of some variable of interest at some specified future date. Predicting is sometimes known as forecasting but predicting has a more general meaning. Predicting sleep apnea events before they happen can be useful to prevent side effect of sleep apnea by using automated CPAP or even by artificially stimulating the related muscles.

Prediction of individual sleep apnea events is considered in this chapter, some parts of this chapter have been published in peer reviewed conferences such as a work entitled "Comparison of Neural Networks for Prediction of Sleep Apnea" which is presented in "International congress on Neurotechnology, Electronics and informatics, Algarve, Portugal, 2013" and another work entitled "Multi Neural Networks Investigation based Sleep Apnea Prediction" which is presented in "The 17th Asia Pacific Symposium of Intelligent and Evolutionary Systems (IES13), Seoul, Korea 2013".

There are several works on how predicting has been applied in different areas but there are few studies about predicting sleep apnea. One of the pioneer work is a paper by Dagum and Galper 1995 [303] who developed a time series prediction using a belief networks model and then used it in sleep apnea. They used a multivariate data set containing 34000 recordings, sampled at 2 Hz, of heart rates (HR), chest volumes (CV), blood

oxygen concentrations (SaO_2), and sleep states (REM) from the time series competition of the Santa Fe Institute in 1991. Their results showed they could predict complex non-linear multivariate data, although the CV prediction had more bias than HR and SaO2 because of the rapid and erratic oscillations of the CV time series.

For the second work in this area a paper by Bock and Gough, 1998 [304] should be referred, where they used 4.75 h of heart rates, respiration force, and blood oxygen saturation (SaO_2) collected from a chronic apnea patient and simple recurrent networks (SRN) proposed by Elman [305]. Each of these three time series variables (heart rate, breathing, and blood oxygenation) were used as inputs for network training and testing operations. Each variable was introduced to a unique network node at the input layer; this network also had 18 nodes in the hidden layer. Time-displaced predictions of respiration signals were produced at each of the three network output layer nodes which were taken as representative of the physiologic state of the sleeping patient. They also selected a back propagation [306] to learn this network. The results of this system were evaluated by two accepted statistics of dynamical invariance, (largest Lyapunov exponent [307] and correlation dimension[308]) and were found to be reasonably good, although the λ_l prediction error was 13% and the error was within 9% of the true time series value.

Another published paper in this area was by Waxaman, Graupe and Carley in 2010 [309] who tried to predict apnea 30 to 120 seconds in advance. They used the LArge Memory STorage And Retrieval (LAMSTAR) neural network [310], where LAMSTAR is a supervised neural network that can process large amounts of data and also provide detailed information about its decision making process. The input signals for this algorithm were EEG, the heart rate variability (HRV), nasal pressure, nasal temperature,

sub-mental EMG, and electrooculography (EOG). It should be noted that LAMSTAR can determine the most important input signal in the predicting process. In the pre-processing phase, data that was segmented by 30, 60, 90, and 120 seconds was normalised by dividing by its mean value, after which a discrete wavelet transform was applied. After that, for each level obtained from the wavelet transformer, the amplitude, the timing for each of the three minima and three maxima, the ratio between the mean of the three maximal amplitudes and three minimal amplitudes, the root-mean-square (RMS) value, and the RMS value related to the original signal were computed.

They trained separate LAMSTAR for each 30, 60, 90, and 120 seconds segments, the results showing that the best prediction belonged to next 30 seconds. They also obtained a lower performance for the longer lead time, although most of the predictions up to 60 seconds in the future were correct. Also, the prediction of non-REM events was generally better than REM events.

As mentioned before, there are limited works on predicting sleep apnea, but also some papers can be mentioned that predict biosignals related to sleep apnea which can actually be used in predicting sleep apnea indirectly. First of all, predicting nasal signals was considered in Hoffstein and Mateika's, 1994 [311]. One of the first papers that studied the prediction of EEG signals was Demarco and Tassinari in 1977 [312], predicting EMG signals was also mentioned in works such as [313, 314].

5.2. Prediction of Sleep Apnea with Multi ANNs

In prediction, same signals (the abdominal, thoracic movement and oxygen saturation signals) were used based on the decision in the previous chapter.

The prediction with lead times such as 30, 60, 90, and 120 seconds were considered and for each lead time, features were generated from time segments and different time

segments such as 30, 60, 90, and 120 seconds were investigated, which means 16 different experiments. The F-score was selected again as the performance measure in this study.

The same features as the detection process were generated for these three signals by applying the statistical measures from Table 5.1 to each coefficient of the last layer of the 4 level Daubechies wavelet packet with an order of 3. In this Table *x* is coefficients of wavelet packet.

Table 5.1: List of statistical features

$\log(\text{mean}(x^2))$	$\text{kurtosis}(x^2)$	$\text{geomean}(\lvert x\rvert)$
$\text{std}(x^2)$	$\text{var}(x^2)$	$\text{mad}(x)$
$\text{skewness}(x^2)$	$\text{mean}(\lvert x\rvert)$	$\text{mean}(x^2)$
$\text{skewness}(x)$	$\text{kurtosis}(x)$	$\text{var}(x)$
$\text{geomean}(x^2)$	$\text{mad}(x^2)$	$\text{std}(x)$

Three different ANNs; Elman, cascade-forward back propagation, and feed-forward back propagation networks, were used to predict sleep apnea, for more information about these ANNs please see section 3.3. To improve the predictive performance, the multi neural network was investigated. For this reason post classification multi ANNs were examined in this study. To implement multi ANNs, both linear and nonlinear approaches were investigated.

5.2.1. Linear multi ANNs

The linear weighting approach was used in this study; a weight is assigned per classifier and the PSO-ANNs algorithm is used to find the optimal weights for each of the three neural networks.

The output of each neural network is an array with the same length as the test set, and it is based on being either positive or negative; it shows that if the corresponding test data was predicted as normal or sleep apnea by that specified neural network. In the proposed PSO-ANNs algorithm each particle contains three cells, and each cell represents the weight of one classifier. Pseudo code 5.1, shows the proposed algorithm, first of all samples were divided into test, train and validation. The PSO-ANNs used the train and test set to find the optimal weights. In each iteration of the PSO algorithm, the final output for each particle is the weighted average of three individual outputs obtained, by considering the weights corresponding to that particle and the fitness of each particle is computed as the F-score of the weighted output to predict the test set. Finally, the algorithm is evaluated by predicting the not seen validation set.

Pseudo code 5.1: The proposed algorithm to find the optimal weights for ANNs

---

1- Specify the train, test and validation data sets and train ANNs with the train set and specify the output of each ANNs for the test set.

2- Initialize the PSO.

3- Do until the maximum number of iterations is reached {

4. For each particle

i. compute the weighted average of output of ANNs by attention to the corresponding weight presented in particle.

ii. Compute the f-score of predicting the test set by the resulted average output.

iii. Set the F-score as the fitness of the particle.

5. Update particles

6. Go to 3.

7- End Do}

8- Predict the validation set by the output of the multi ANNs resulted by the optimal weights.

5.2.2. Non-Linear multi ANNs

For the non-linear multi ANNs majority voting was considered, in the majority voting if more individual classifiers give an output of 1 rather than an output of 0, then the aggregated classifier takes the output 1 and vice versa. Since the neural network outputs take continuous values between 0 and 1, median operator was used to implement majority voting.

5.3. Experiments and Results

In this chapter different experiments regarding to the prediction of sleep apnea are discussed.

5.3.1. Artificial Neural Networks Architectures

Three different ANNs; Elman, cascade-forward back propagation, and feed-forward back propagation networks, were used to predict sleep apnea. For each of these ANNs the 208 generated features are set as the inputs, therefore the first layer of these ANNs have 208 nodes.

For the feed-forward ANNs, the radial basis was used as the transfer functions of the hidden layer and the output layer, and the Log-sigmoid and pure line were used as trans-

fer functions of the hidden layer and the output layer of the cascade-forward network. For the Elman networks tangent sigmoid transfer function was used for the hidden and output layers, scaled conjugate gradient back propagation was used as the training function for these three ANNs. The number of nodes in the hidden layer of each ANN was also determined by trial and error.

As mentioned before, this study wanted to predict sleep apnea in the next 30, 60, 90, and 120 seconds. The prediction was based on investigating various lengths of segments from 30 seconds to 120 seconds. For each of these experiments a unique ANN was trained.

5.3.2. Early Stopping

Standard neural network architectures such as the fully connected multi-layer perceptron are prone to over fitting. While the network seemed to get better and better (the error on the training set decreased), but at some point during training it actually began to get worse again (the error on unseen examples increased). There are two basic ways to overcome the over fitting problem: reduce the number of dimensions of the parameter space or reduce the effective size of each dimension. The corresponding ANN techniques for reducing the size of each parameter dimension such as weight decay or early stopping [315] were regularised. Early stopping is widely applied because it is simple to understand and implement and has been reported in many cases as being superior to the regularisation methods. Early stopping can be used either interactively, i.e. based on human judgment, or automatically, i.e. based on some formal stopping criterion. In this work, automatic stopping criteria based on the cross validation error was used, while the validation error was used to estimate the generalisation error [315].

5.3.3. Designing Structure of Neural Networks

As mentioned before, 16 different experiments were considered in the prediction of sleep apnea, including 4 different lead times, where each of them can have 4 different segments. To find the number of nodes in the hidden layer different numbers of nodes were examined, with numbers from 5 to 200 in increments of 10 nodes. Table 5.2 shows the best number of nodes in the hidden layer, for different settings.

Table 5.2: Optimal number of nodes in the hidden layer

Lead time (seconds)	**30**				**60**				**90**				**120**			
Segments (Seconds)	**30**	**60**	**90**	**120**	**30**	**60**	**90**	**120**	**30**	**60**	**90**	**120**	**30**	**60**	**90**	**120**
Feed Forward	30	50	50	70	20	60	70	70	50	70	50	50	60	60	70	60
Cascade forward	10	20	50	30	30	50	50	70	10	50	30	60	40	20	40	10
Elman	30	10	20	20	20	40	10	60	40	70	30	40	30	10	20	10

To find the optimal node numbers, the F-score was used as the first criteria, after that accuracy was considered, and in cases of different options, the structure with lowest node number was selected.

5.3.4. Sleep Apnea Prediction

After choosing the optimal structures for the three selected neural networks for all of the settings, they were then used to detect sleep apnea. Here the samples were divided into train, test and validation by random, and 5 different approaches were examined: three single neural networks, the linear multi ANNs of these networks, and the non-

linear multi ANNs. The weights of the linear multi ANNs were determined by the PSO, as described before, and the median of classifiers was the non-linear multi ANNs.

Tables 5.3-5.6 tabulated the average F-scores and the accuracies for 5 independent experiments.

Table 5.3: Prediction of sleep apnea with 30 seconds lead time

	F-score				Accuracy			
Segments (Seconds)	**30**	**60**	**90**	**120**	**30**	**60**	**90**	**120**
Feed Forward	0.6997	0.7509	0.6914	0.7186	77.83	77.41	79.57	82.31
Cascade forward	0.6507	0.6759	0.6925	0.6254	73.29	80.25	79.77	76.72
Elman	0.5465	0.5786	0.6461	0.5908	81.27	84.57	85.08	84.06
Linear multi ANNs	0.8082	0.8139	**0.8296**	0.7903	83.29	80.25	**85.53**	79.72
Non-linear multi ANNs	0.7146	0.6853	0.7170	0.6990	70.45	75.40	72.55	78.33

Table 5.3 indicates that the linear weight multi ANNs of neural networks achieved the highest F-score and accuracy, with 90-second segments for 30 seconds lead time. The average running time for the validation set was 1.14 seconds.

When predicting with a lead time of 60 seconds, the linear weighted multi ANNs achieved the highest F-score and accuracy using 90-second segment, Table 5.4. The average running time for the validation set was 2.92 seconds.

Table 5.4: Prediction of sleep apnea with 60 seconds lead time

	F-score				Accuracy			
Segments (Seconds)	**30**	**60**	**90**	**120**	**30**	**60**	**90**	**120**
Feed Forward	0.6600	0.5320	0.6177	0.5472	80.04	81.66	77.15	84.15
Cascade forward	0.6392	0.6553	0.6802	0.5374	72.77	70.56	84.27	77.03
Elman	0.5380	0.5961	0.5640	0.5491	83.36	84.25	84.82	76.75
Linear multi ANNs	0.7630	0.8070	**0.8172**	0.8026	82.77	85.06	**85.13**	79.03
Non-linear multi ANNs	0.7697	0.7501	0.7418	0.7394	79.57	81.23	82.26	80.88

Table 5.5: Prediction of sleep apnea with 90 seconds lead time

	F-score				Accuracy			
Segment (Seconds)	**30**	**60**	**90**	**120**	**30**	**60**	**90**	**120**
Feed Forward	0.5466	0.5824	0.6298	0.6093	74.09	80.87	81.07	66.61
Cascade forward	0.5973	0.6320	0.6907	0.5241	79.06	72.55	71.15	76.38
Elman	0.5715	0.5115	0.5426	0.5161	81.99	**83.95**	81.02	80.55
Linear multi ANNs	0.8105	**0.8114**	0.7918	0.7158	79.06	82.00	82.15	82.10
Non-linear multi ANNs	0.7653	0.7406	0.7800	0.7656	80.77	81.91	76.32	81.49

Table 5.5 shows that when predicting with a lead time of 90 seconds, the linear multi ANNs achieved the best F-score by using 60-second segment, but the highest accuracy was achieved by the Elman neural network and with the same data segment. The aver-

age of running time for predicting of the validation set with a lead time of 90 seconds was 3.22 seconds.

Table 5.6: Prediction of sleep apnea with 120 seconds lead time

	F-score				Accuracy			
Segment (Seconds)	**30**	**60**	**90**	**120**	**30**	**60**	**90**	**120**
Feed Forward	0.5211	0.6162	0.5464	0.5372	75.29	66.64	64.66	70.01
Cascade forward	0.6173	0.5900	0.6112	0.5265	71.07	75.90	61.56	57.79
Elman	0.5480	0.5564	0.6272	0.5940	76.47	78.05	76.36	69.55
Linear multi ANNs	0.7261	0.7451	**0.7519**	0.7457	71.07	75.90	**79.07**	77.79
Non-linear multi ANNs	0.7431	0.7150	0.7422	0.6959	76.88	77.63	66.93	69.91

Finally, for predicting with a lead time of 120 seconds, the highest F-score and accuracy were achieved with a linear multi ANNs with 90-second segment. The average running time for this setting was 3.66 seconds.

As a conclusion on this experiment, it is very clear that linear multi ANNs outperforms other algorithm either for F-score or accuracy of prediction.

5.4. Summary

Predicting of individual sleep apnea events was studied in this chapter using three different artificial neural networks together with linear and non-linear multi ANNs. The

weights of the linear multi ANNs were optimised by PSO, and the median of these three networks was the non-linear multi ANNs.

Four lead times, 30, 60, 90 and 120 seconds were studied, and instead of the fixed data segment, different segments from 30 seconds to 120 seconds were examined. The results showed that the linear multi ANNs achieved the highest F-score in all settings, and the highest accuracy, in most cases. Based on this study, the best result is achieved for predicting sleep apnea with 30 seconds lead time and time windows of 90-seconds with F-score of 0.8296**.** But also, the prediction of sleep apnea with a lead time of 90 seconds by using the linear multi ANNs and a 60-second segment can be reasonable setting by attention to the obtained F-score of 0.8114 and accuracy of 82.00% with running time of 3.22 seconds, which can be reasonable for predicting with 90-second lead time.

Chapter 6

Thesis Developed Techniques Generalization

6.1. Introduction

During this study it was necessary to develop several algorithms based on specific application, such as segmentation or feature generation, and feature selection. Moreover, a new idea for SVM had also developed and accordingly proposed a new parallel structure for multi-PSO algorithm. Because these two algorithms can be used in many area of machine learning, they are introduced separately in this chapter.

The proposed SA-SVM algorithm has been published in International Journal of Knowledge-Based Systems [289]and the proposed parallel PSO structure has been used in different papers [290, 291].

6.2. Self-advising SVM

There has been a progressive increase in the use of advice sets using the prior knowledge obtained from experts. However, there are some difficulties in how this knowledge can be applied and expressed in terms of the constraints. Moreover, these approaches include new parameters which increase the computational cost of SVM. However, the ensemble algorithms are iterative procedures that increase the computation cost and do not always improve the performance of the SVMs, in fact they make it worse [316].

This section presents a developed non-iterative method that extracts subsequent knowledge from the training phase. In the traditional SVM method the only information that is used in the test phase from the training phase is the hyperplane position or the first type of support vectors. Subsequent knowledge can be in the form of any further information derived about the first type of support vectors, such as their distribution, and or the knowledge extracted from the second type of support vectors.

As a part of these thesis contributions, an aim is for generating subsequent knowledge from the second type of support vectors. Even with the optimised hyperplane a lot of data can be misclassified, second type of support vectors [210], were selected. This misclassified data can come from 2 potential sources such as outliers, or as data that has not been linearly separated using any of the types of kernels. Classic SVM ignores the training data that has not been separated linearly by kernels during the training phase. This occurs by the introduction of tolerance parameters in the objective function and constraints. However, if data that is similar or identical to this misclassified data appears in the test set, it will be classified wrongly again because the data which is close to the misclassified data is uncertain. This misclassification is not reasonable since it occurred because the available data and information in the training phase was ignored by the SVM algorithm. It should be noted that any method that wants to benefit from misclassified data must have some control on the impact of the outlier data. Actually, if more samples in the proposed method that look like misclassified data were found, this approach may be able to improve the performance of the classifier.

This proposed method focused on the ignorance of SVM from the knowledge that can be acquired from misclassified data by generating advice weights based on the use of misclassified training data, if possible, and by using these weights together with the de-

cision values of the SVM in the test phase. These weights help the algorithm to eliminate the outlier data.

The misclassified data sets MD, in the training phase, can be defined as follows:

$$MD = \bigcup_{i=1}^{N} \boldsymbol{x}_i | y_i \neq sign\left(\sum_{\alpha_j > 0} y_j \alpha_j k(\boldsymbol{x_i}, \boldsymbol{x_j}) + b \right) \qquad (6.1)$$

It must be considered that on the right hand side of the equation (6.1), any SVM decision function and kernel can be used. The MD set can be null, but experimental results have revealed that the occurrence of misclassified data in the training phase is a common occurrence.

For each $\boldsymbol{x}_i$ of MD the neighborhood length (NL) is defined as:

$$NL(\boldsymbol{x}_i) = minimum_{\boldsymbol{x}_j} \left(||\boldsymbol{x}_i - \boldsymbol{x}_j|| \middle| y_i \neq y_j \right)/2. \qquad (6.2)$$

where $\boldsymbol{x_j}, j = 1, \dots, N$, are the training data that do not belong to the MD set.

Note: if the training data is mapped to a higher dimension using a mapping function then the distance between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ can be computed according to the following equation (14), with reference to the related kernel k,

$$||\boldsymbol{\theta}(\boldsymbol{x}_i) - \boldsymbol{\theta}(\boldsymbol{x}_j)|| = \left(k(\boldsymbol{x}_i, \boldsymbol{x}_i) + k(\boldsymbol{x}_j, \boldsymbol{x}_j) - 2k(\boldsymbol{x}_i, \boldsymbol{x}_j) \right)^{0.5}. \qquad (6.3)$$

Based on the findings of NL, for each $\boldsymbol{x}_k$ from the test set, the advised weight $AW(\boldsymbol{x}_k, j)$ for $J = +1$ or -1 is computed as follows,

$$AW(\boldsymbol{x}_k,\ J) =$$

$$\begin{cases} 0, & \forall \boldsymbol{x}_i \in MD,\ \left|\left|\boldsymbol{x}_k - \boldsymbol{x}_i\right|\right| > NL(\boldsymbol{x}_i)\ or\ \ MD = NUL \\ \sum \frac{|h(\boldsymbol{x}_i)|}{1 + ||\boldsymbol{x}_k - \boldsymbol{x}_i||^2} & \exists \boldsymbol{x}_i \in MD,\ \left|\left|\boldsymbol{x}_k - \boldsymbol{x}_i\right|\right| \leq NL(\boldsymbol{x}_i)\ and\ y_i = J \end{cases} \qquad (6.4)$$

These AWs represent how close the test data are to the misclassified data. To conclude the above, the self-advising SVM (SA-SVM) [317] is as follows:

Training phase:

1- Finding the hyperplane by solving problem (3.7) or the related problem, and this means the normal SVM training.

2- Finding the MD set using equation (6.1).

3- If the MD is null, go to the testing phase or else compute NL for each member of MD using equation (6.2).

Testing phase:

1- Compute $AW(\boldsymbol{x}_k, +1)$ and $AW(\boldsymbol{x}_k, -1)$for each $\boldsymbol{x}_{k,}$ from the test set

2- Compute $h(x_k)$ as the absolute value of the SVM decision values for each $\boldsymbol{x}_k$ from the test set by using equation (10), and scale it to [0,1].

3- For each $\boldsymbol{x}_k$ from the test set,

If $\max\big(AW(x_k, \pm 1), h(x_k)\big) = h(x_k)$ then $\boldsymbol{y}_k = sign\left(\sum_{\alpha_j>0} y_j \alpha_j k\big(\boldsymbol{x_k}, \boldsymbol{x_j}\big) + b\right)$, this means normal SVM labeling, otherwise $\boldsymbol{y}_k = +1$ or -1 based on $\max\big(AW(x_k, +1), AW(x_k, -1)\big)$.

Note: If the testing and training data are mapped to a higher dimension, then $||\boldsymbol{x}_k - \boldsymbol{x}_i||$ in step 3 of the test phase should be computed by equation (6.3), and as mentioned previously, any SVM methods and kernels can be used in this algorithm.

6.2.1. Experimental Results

To evaluate the proposed SA-SVM, the experiment adopts 11 datasets from the UCI machine learning repository [318]. These databases were selected from the most common benchmarks for classification and the variety of these databases supports the validation in this study. The number of instances and the attributes of each database are shown in the Table 6.1. It should be noted that for a dataset with multi-class, only data

from two classes were selected and the data with missing values were deleted. Of course, applying the proposed approach for multi-class classification with extension approaches [319, 320] can be considered in future studies.

The platform used in this experimental study is Intel Pentium V, 3.16 GHz CPU, 3.25 GB RAM and Windows 7 which is the operating system. The proposed algorithm was implemented on MATLAB 7.14, and LIBSVM [321] was used for the SVM.

For each dataset C-SVM and $\nu-$SVM were used and then compared them with SA$-C$- SVM and $SA-\nu-$SVM methods to see if the SA algorithm could improve these two types of SVM. It should be noted that because this approach is not an iterative method, it cannot be compared to methods such as AdaBoost or other multi ANNs approaches. In these experiments the RBF kernel was selected and the parameters were tuned by a grid search.

RBF kernel were used for each dataset and the optimal parameters c and gamma were computed by a grid search. A full list of these parameters are presented in Table A.1. In Appendix A.

For a reliable validation 10-fold cross validation was performed for five times. Average$\pm$estandard deviations of the training accuracy with C-SVM and $\nu-$SVM , are shown in Table 6.2. It should be noticed that the training accuracy was computed as the accuracy of SVM in classifying the seen training data set. As Table 6.2 indicates, the training accuracy was less than 100% in most cases. This showed that SVM could not classify the entire pattern correctly, even for the seen part. This becomes more important when a similar pattern to the misclassified training data appears in the test set because these patterns in the test set will also be classified wrongly in classic SVM. However

SA-SVM tries to use the knowledge extracted from misclassified training data to prevent a similar misclassification of test data

Table 6.1: Datasets from the UCI repository

	Dataset	Number of instances	Number of attributes
1	**Australian Credit Approval (Statlog)**	690	14
2	**Breast Cancer Wisconsin (Original)**	683	9
3	**Contraceptive Method Choice**	642	9
4	**German credit Data (Statlog)**	1000	24
5	**Hayes-Roth**	102	5
6	**Hepatitis**	129	17
7	**Ionosphere**	351	33
8	**iris**	100	4
9	**Liver Disorders**	345	6
10	**Spambase**	4601	57
11	**Teaching Assistant Evaluation**	101	5

C-SVM and ν-SVM were performed with and without an SA algorithm on these datasets. Table 6.3 tabulated the testing accuracy and also the percentage of improvement by SA-SVM compared to classic SVM.

From Table 6.3 it is clear that in 9 datasets out of 11, SA-SVM performed better than traditional SVMs, while the SA algorithm improved the performance of SVM by around 10% in iris(**C-SVM**) and Teaching assistant (***ν*-SVM**) databases, by around 2% in Hayes and Spambase (***ν*-SVM**) databases, and the improvements in other databases were around 1%. A full list of these 5 ×ten-fold validations are presented in Tables A.2 and A.3 In Appendix A.

Table 6.4 summarises the means and standard of deviations of accuracies for C-SVM, ν-SVM, and in total with and without self-advising for 10-fold cross- validation and 5-fold cross-validations.

Table 6.2: Accuracy of the training phase of classic SVM

Dataset	C-SVM	$\nu - $ SVM
Australian Credit Approval (Statlog)	82.63±0.68	90.86±0.54
Breast Cancer Wisconsin (Original)	97.35±0.23	97.08±0.27
Contraceptive Method Choice	77.29±0.44	85.94±0.38
German credit Data (Statlog)	78.23±0.63	82.55±0.42
Hayes-Roth	94.55±0.87	81.94±3.06
Hepatitis	76.89±1.74	90.07±0.72
Ionosphere	99.28±0.10	95.22±0.28
iris	90.15±8.99	99.09±0.89
Liver Disorders	77.16±0.75	83.62±0.83
Spambase	97.23±0.07	90.63±0.17
Teaching Assistant Evaluation	98.48±0.61	82.12±1.04

Tables A.2 and A.3, in Appendix A, show that the proposed algorithm improved the accuracy of SVM in most datasets. Table 6.4 shows that the average improvement of the proposed method was higher in the 5-fold cross-validation. This could be result of the size of the training and test set in each run, where these sets were bigger in the 5-fold cross validation than the10-fold cross-validation. A bigger training and test set can in-crease the chance of finding a similarity between misclassified patterns in the training set and patterns in the test set.

Table 6.3: Results of classification

	SVM		Self-Advising SVM	
Data Set	**C-SVM**	**ν-SVM**	**C-SVM**	**ν -SVM**
Australian Credit Approval (Statlog)	83.12 (1.06)	**85.91 (0.50)**	83.85 (1.06)	85.83 (0.60)
Breast Cancer Wisconsin (Original)	96.89 (0.38)	96.33 (0.54)	**96.92 (0.38)**	96.36 (0.52)
Contraceptive Method Choice	73.69 (1.83)	72.68 (1.39)	**74.60 (1.94)**	73.71 (1.01)
German credit Data (Statlog)	76.21 (0.88)	75.47 (0.64)	**76.27 (0.82)**	75.57 (0.62)
Hayes-Roth	78.77 (4.27)	76.54 (4.82)	**80.36 (4.21)**	78.91 (5.22)
Hepatitis	**69.99 (2.80)**	66.88 (2.65)	69.67 (2.77)	67.03 (2.76)
Ionosphere	**94.22 (1.15)**	93.32 (0.85)	**94.22 (1.15)**	93.32 (0.85)
iris	90.06 (9.06)	99.79 (0.46)	**100**	99.79 (0.46)
Liver Disorders	73.80 (0.75)	70.16 (2.20)	**74.97 (0.98)**	71.03 (1.91)
Spambase	93.39 (0.38)	90.38 (0.26)	**93.83 (0.44)**	92.37 (0.23)
Teaching Assistant Evaluation	72.92 (6.90)	76.16 (1.53)	74.74 (6.70)	**86.06 (1.65)**
Average	82.10	82.15	83.58	**83.63**

Table 6.4: Summaries of the experimental results

		10-fold validation		5-fold validation	
		Mean	**SD**	**Mean**	**SD**
C-SVM **(55 runs)**	SVM	82.127	10.56	79.4664	11.47
	SA-SVM	83.614	10.66	81.6788	11.72
	Improvement	1.487		2.21	
$\nu-$SVM **(55 runs)**	SVM	83.549	11.60	81.5014	11.79
	SA-SVM	84.912	11.17	82.6646	11.35
	Improvement	1.363		1.16	
Total (110 runs)	SVM	82.151	11.09	80.4839	11.62
	SA-SVM	83.638	10.79	82.1717	11.50
	Improvement	1.488		1.68	

If the 5-fold cross validation is compared to the method proposed with C-SVM, then the improvement is equal to 2.21% in all of the datasets and when compared with ν-SVM it is equal to 1.68% in total. The distribution of the percentage of improvement in different methods and in the total is shown in Figure 6.1, (Figure 6.1.A in Total, Figure 6.1.B: C-SVM, Figure 6.1.C: ν-SVM). This figure shows that the proposed method improved the accuracies in more than 67% of experiments, half of which are greater than 1%. It can also be noticed that 22% of these improvements are greater than 2% and 11% of these improvements are more than 5%. Moreover, the highest improvement was equal to 25%. In 30% of the experiments this algorithm had exactly the same result as classic SVM and in just 3.6% of the experiments, the proposed algorithm lead to a worse result, which was only 0.62% less accurate than traditional SVM.

To have a more reliable and accurate validation, 11 datasets were used in this section. It is obvious that if a subset of these databases is selected and used for validation then a higher overall improvement can be achieved. Since 5 datasets were tested (Hayes-Roth, iris, liver disorder, spambase and teaching assistance) that achieved an overall im-

provement of 3% by the proposed algorithm but using less datasets would not be as reliable.

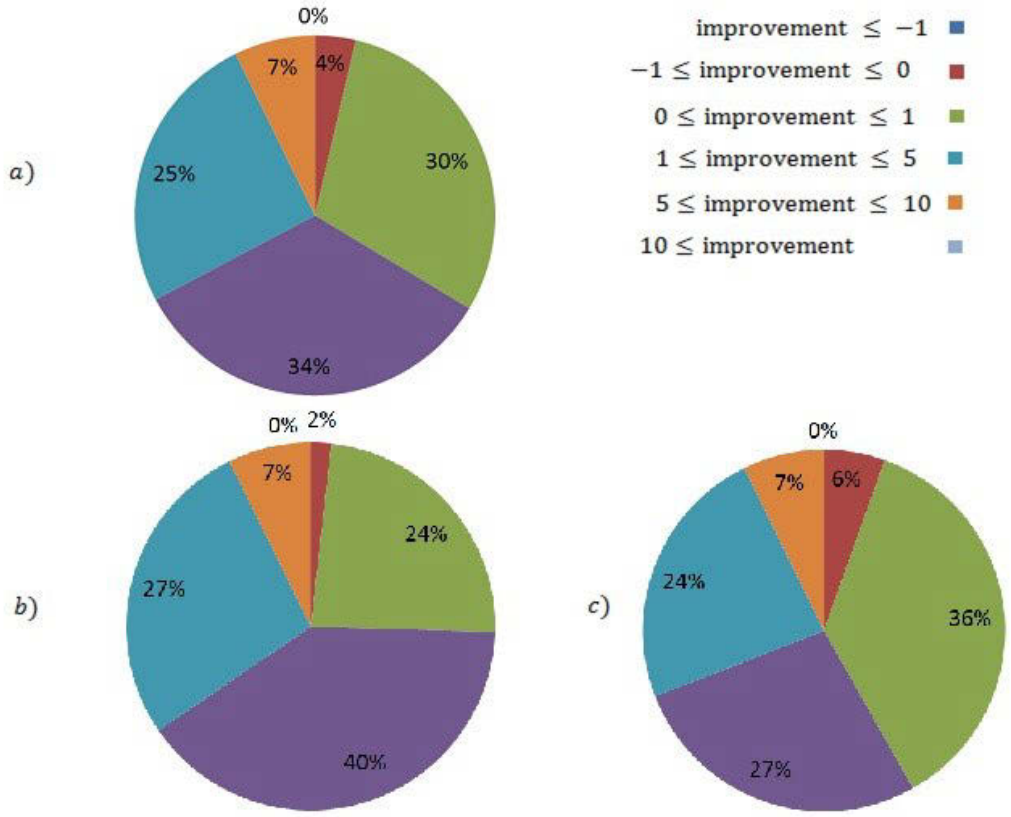

Figure 6.1: Distribution of accuracies improvements by SA-SVM

For a more reliable evaluation of the results of these two methods, a pairwise t-test was performed. This statistical test also proved the significance of improvements to the classifications acquired through the proposed method. The p value for C-SVM and ν-SVM of the t-test are 5.78×10^{-3} and 6.51×10^{-5} respectively, and by attention to all the runs, the p value was 7.20×10^{-5}, which shows statistically significance of difference between results obtained these approaches.

For even more validation the statistical F-score (F-measure) was used as another measure of the performance of classification algorithms. The F-score can be interpreted

as the weighted average of the precision and recall, where the F-score reaches its best value at 1 and its worst score at 0.

Table 6.5 tabulates the averages and standard deviations of f-scores for 10-fold and 5-fold cross validation.

It is also obvious from Table 6.5 that the SA-SVM was better than the SVM methods in the F-score. This claim was strengthened by performing the t-tests. The t-test shows that an average of the f-scores of SA-SVM was also significantly higher than the SVM results since the p-values were equal to 1.55×10^{-3}, 1.21×10^{-4} and 6.01×10^{-4}, respectively for C-SVM and ν-SVM in total. The full results of these 5×5-fold validations are presented in Tables A.4 and A.5 in Appendix A.

Different studies have been proposed to improve the performance of SVM, but the primary disadvantages of these approaches are the inclusion of new parameters, increased complexity of the learner model, and the difficulty of implementation in solving practical problems such as modeling prior knowledge.

The self-advising SVM method proposed here, improved the performance of SVM by transferring more information from the training phase to the testing phase. This information was generated by using misclassified data in the training phase. If the same data or similar patterns appear in the testing data, they will not be misclassified again. Experimental results in this study showed an improvement in accuracy, and the F-score and statistical tests revealed the significance of these improvements. Our results showed that using the misclassified data in the training phase by the proposed method will not increase overtraining. The main advantages of SA-SVM are that it can be applied through any of the SVMs and kernel types, and it does not need the addition of new parameters to the learning phase.

Table 6.5: Averages and standard deviations of F-score

		10-fold validation		5-fold validation	
		Mean	**SD**	**Mean**	**SD**
C-SVM **(55 runs)**	SVM	0.77	0.15	0.75	0.15
	SA-SVM	0.79	0.15	0.78	0.15
	Improvement	0.01		0.02	
$\nu-$ SVM **(55 runs)**	SVM	0.78	0.15	0.77	0.16
	SA-SVM	0.79	0.15	0.79	0.15
	Improvement	0.01		0.01	
Total (110 runs)	SVM	0.77	0.15	0.76	0.15
	SA-SVM	0.79	0.15	0.78	0.15
	Improvement	0.0164		0.0192	

Implementing KNN for computing the advice weights and implementing the proposed approach for multi-class classification may be considered potential areas for future studies.

6.3. Proposed Parallel Structure

In this study, given the enormous size of the search space, a single PSO cannot perform well and may be result in a local optimum with low accuracy. Therefore, a new parallel PSO was used structure to perform better explorations and exploitations in the search space.

In the traditional multi-PSO models all of the swarms are at the same level and exchanging information just happened based on the definition of neighbourhood. But in this structure, swarms were classified as ‘masters’ and ‘slave(s)’. Master swarms have

access to the best particles of other swarms but the slave swarms have no access to information of other swarms, and in reality they just provide information for others. Sending information of the best local particle among the masters and from the slave(s) to the masters can be performed at each iteration by a specified frequency.

In proposed hierarchical model, one of the master swarms is considered to be the central swarm, and all of the swarms, both masters and slaves, send the local best particle to the central swarm. The central swarm computes the global best particle and sends it to the other master swarms. So all of the master swarms update their particles using the global best particle, but the slave swarms only use their own local best particles to update themselves. The Pseudo code 6.1 shows the process of the Hierarchical Multi-Master PSO (HMM PSO).

Pseudo code 6.1: *the Hierarchical Multi-Master PSO (HMM PSO)*

1- Select the number of master and slave swarms, the number of particles for each sub-warm, and also the frequency for sending the information. Select one of the master swarms as the central swarm.

2- Initialize the position and velocity of each particle

3- ***Do*** *in parallel until the maximum number of iterations is reached {*

3-1-Evaluate the fitness value of each particle

3-2-Find out the local best particle in each sub-swarm

*3-3-****If*** *meets the sending condition*

3-3-1- Send the local best particle (lp-best) from each swarm to the central swarm.

3-3-2-Update the global best particle (Gp-best) in the central swarm.

3-3-3-Send the global best particle to the entire master swarms.

*3-4-**End If***

3-5-Update the position of each particle in each swarm

*4- **End Do**}*

5-Returning the best solution (the global best particle) of the algorithm.

6-End.

Figure 6.2 illustrates a sample of the proposed parallel structure with 4 masters and 2 slave swarms. In this implementation, master 1 is selected as the central swarm so all the swarms send their local best particle to this swarm. After computing the global best swarm, the central swarm sends it to all the master swarms. So in this structure, the slave swarm provides information for other swarms but they do not benefit from each other's information.

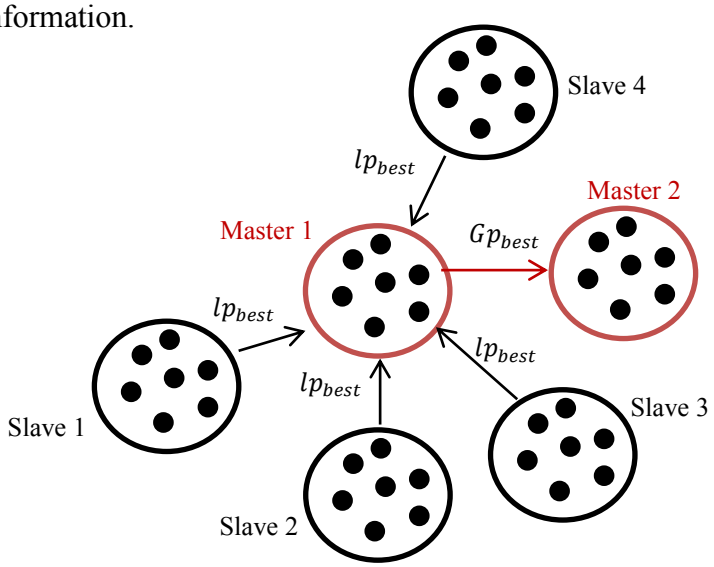

Figure 6.2: Proposed parallel structure with 2 masters and 4 slaves

Indeed, fast convergence is one of the disadvantages of PSO, which is heightened by the parallel structure. This hierarchical model tries to expand both the exploration and exploitation abilities of the parallel PSO, by integrating the isolated swarms and the linked swarms. Slave swarm(s) in this model help prevent premature convergence. PSO parameters for the slaves and masters can be different so these parameters can be adjusted to have greater explorative abilities for the slaves and a greater local search for the masters.

6.3.1. Experimental Studies

In this section, different benchmark functions were used to test the effectiveness and efficiency of the HMM PSO algorithm. For each benchmark function, 30 independent optimisations were performed to collect the associated statistics for evaluation and to have a more robust comparison. In the experiment studies the benchmark function with dimensions of 10 investigations were used to measure the quality of the proposed algorithm's solution.

Six common benchmark problems were used: Ackley, Quartic, Rastrigin, Rosenbrock, Schwefel's problem 2.22 and Sphere. These functions are useful and common for comparing algorithms because they allowed us to explore the behaviour of optimisation algorithms in a simple and standardised way [322]. The benchmark functions are listed below:

Ackley function (A_n)

$$A_n(x) = -20\exp\left(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} {x_i}^2}\right) - exp\left(\frac{1}{n}\sum_{i=1}^{n}\cos(2\pi x_i)\right) + 20 + e.$$

Quartic (Q_n)

$$Q_n(x) = \sum_{i=1}^{n} i{x_i}^4 + rand[0.1).$$

Rastrigin (Ras_n)

$$Ras_n(x) = \sum_{i=1}^{n} ({x_i}^2 - 10\cos(2\pi x_i)) + 10).$$

Rosenbrock (R_n)

$$R_n(x) = \sum_{i=1}^{n} 100({x_i}^2 - x_{i+1})^2 + (x_i - 1)^2.$$

Schwefel's problem 2.22 (Sch_n)

$$Sch_n(x) = \sum_{i=1}^{n} |x_i| + \prod_{i=1}^{n} |x_i|.$$

Sphere (S_n)

$$S_n(x) = \sum_{i=1}^{n} {x_i}^2.$$

Table 6.6 lists the dimension, n, of each function, the ranges of their search space, and their global minimum value. In this study asymmetric initialization is used to compare these algorithms as symmetric initialization can give false impressions of relative performance [323].

Table 6.6: Parameters of the benchmark functions

Function	n	Minimum value	Range of search	Range of initialization
A_n	10	0	[−32.768, 32.768]	[5, 10]
Q_n	10	0	[-1.28, 1.28]	[0.64, 1.28]
Ras_n	10	0	[−5.12, 5.12,]	[2.56, 5.12]
R_n	10	0	[-100, 100]	[15, 30]
Sch_n	10	0	[-10, 10]	[5, 10]
S_n	10	0	[-100, 100]	[50, 100]

A. Experimental Results

The performance of the new method methods was compared with COM-MCPSO, COL-MCPSO and MCPSO-CC [324], The maximum velocity (V_{max}) and minimum velocity (V_{min}) for these algorithms were set at 1 and -1 and a fixed number of maximum generations 1000 was applied to these algorithms with a population size equal to 80. For the parallel methods, a setting of c1 = c2 = 2.05, c3 = 2.0 was considered. The inertia weight was also critical for the convergence behaviour of PSO in these experiments, so a decaying inertia weight starting at 0.9 and ending at 0.4, following Shi and Eberhart [325], was used. The number of slave swarms was set as 4 for COM-MCPSO, COL-MCPSO and MCPSO-CC. All the parameters used in the slave and master swarms were the same as those defined above. Also the sending information frequency was set as 5 iterations each in the HMM PSO algorithm.

B. Architecture Analysis

In this section 6 swarms were used for all the algorithms, but with the HMM PSO algorithm, different numbers of master and slaves swarms can be used. Therefore, for the

HMM PSO algorithm, a different setup with 2 masters 3 masters and 4 masters are examined in this section. For these reason 4 benchmark functions, the Ackley, Rastrigin, Rosenbrock, and Sphere functions were selected randomly. The best, worst, mean, and standard deviation of the function values found in 30 trials are as shown Table 6.7. In this table, the numbers in bold-face type represent the comparatively best values. Based on the results obtained, 2 masters and 4 slaves were selected as the best architecture for the HMM algorithm in the following experiments.

Table 6.7: Analysis of different number of master swarms

	Number of Masters	best	worst	average	std
Ackley	**2**	**19.6**	**19.76**	**19.73**	0.02
	3	20	20	20	5.23E-06
	4	19.71	19.78	19.75	0.01
Rastrigin	**2**	**45.76**	**99.50**	**75.59**	14.36
	3	**45.76**	104.17	80.62	14.29
	4	**45.76**	117.11	83.86	16.63
Rosenbrock	**2**	**11.28**	159.26	49.49	37.60
	3	21.46	**94.57**	50.19	15.64
	4	25.02	105.38	**46.33**	29.99
Sphere	**2**	4.60E-31	**5.97E-28**	**7.08E-29**	1.29E-28
	3	5.98E-31	1.78E-27	2.26E-28	4.26E-28
	4	**5.11E-33**	2.51E-27	4.45E-28	6.31E-28

The experimental results for each algorithm on each test function consisted of the best, worst, mean and standard deviation of the function values found in 30 runs, and they are listed in Table 6.8. In this table the numbers in bold-face type represent the comparative-

ly best values, and the graphs presented in Figures 6.3–6.4 illustrate the evolution of best fitness found by five algorithms, averaged for 30 runs for the benchmark functions. By attention to the Ackley function, Figure 6.3.a, the improved searches by all algorithms seemed to be similar, but the HMM PSO method improved slowly in the beginning but eventually outperformed the other algorithms before 1000 generations. HMM PSO, COM-MCPSO and COL-MCPSO algorithms reached the same best fitness for this function, as shown in Figure 6.3.b Function Q_n is a noisy function and with it, all the algorithms except the MCPSO-CC method, seemed to converge in a similar pattern, as shown in Figure 6.3.b. The HMM PSO method was better in obtaining the best fitness values, while the HMM PSO algorithm reached the best average fitness in 30 trials, but the average difference between these two algorithms was insignificant.

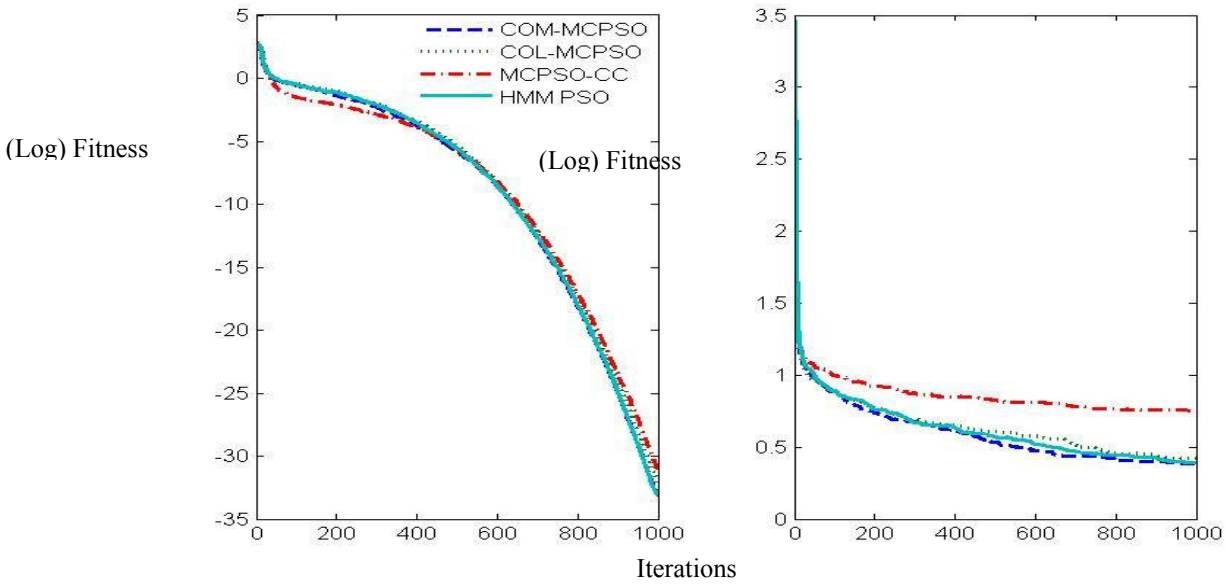

Figure 6.3: Ackley function and Quartic Functions

Table 6.8: Results of different parallel PSO

		COM	COL	CC	MM
A_n	**Best**	**2.66E-15**	**2.66E-15**	6.22E-15	**2.66E-15**
	worst	1.33E-14	3.11E-14	1.13E-13	**6.22E-15**
	mean	**3.85E-15**	1.08E-14	2.71E-14	4.44E-15
	std	2.30E-15	6.62E-15	2.38E-14	**1.80E-15**
Q_n	**Best**	1.01E+00	1.04E+00	1.61E+00	**8.26E-01**
	worst	1.85E+00	1.81E+00	2.54E+00	**1.76E+00**
	mean	**1.47E+00**	1.52E+00	2.12E+00	1.48E+00
	std	0.21	0.20	0.21	**0.17**
Ras_n	**Best**	**9.95E-01**	1.99E+00	3.76E+01	**9.95E-01**
	worst	**8.95E+00**	2.89E+01	6.10E+01	1.19E+01
	mean	**3.73E+00**	7.83E+00	5.03E+01	4.38E+00
	std	**1.82**	5.65	6.11	2.20
R_n	**Best**	1.31E-02	1.56E-01	4.35E+01	**7.51E-03**
	worst	**1.43E+01**	1.52E+01	2.42E+02	1.80E+01
	mean	5.67E+00	5.94E+00	9.90E+01	**5.24E+00**
	std	3.65	**3.38**	35.64	3.58
Sch_n	**Best**	4.59E-27	1.23E-26	9.06E+00	**6.91E-28**
	worst	6.23E-24	3.20E-23	1.90E+01	**1.16E-24**
	mean	6.62E-25	4.62E-24	1.34E+01	**1.60E-25**
	std	1.50E-24	6.59E-24	2.70	**2.90E-25**
S_n	**Best**	**4.36E-32**	2.75E-31	4.80E-29	2.48E-31
	worst	1.85E-28	1.60E-27	4.95E-27	**4.73E-29**
	mean	1.63E-29	1.61E-28	8.45E-28	**5.54E-30**
	std	3.34E-29	3.15E-28	1.13E-27	**9.66E-30**

Function Ras_n , Figure 6.4.a, is a classic optimisation problem so convergence to its global optimum was difficult. In this problem MCPSO-CC and COL-MCPSO clearly performed worse than COM-MCPSO and MM PSO. In this problem the trend for HMM PSO was similar to COM-MCPSO until 600 generations, but after that COM-MCPSO performed better for 30 trials. However, according to the final results, the HMM PSO algorithm obtained the best final results but HMM PSO outperformed it with the average results. For the R_n function, Figure 6.4.b, all the algorithms except the MCPSO-CC method performed the same, while Table 6.13 shows that the HMM PSO algorithm obtained the best and the best average of the final results.

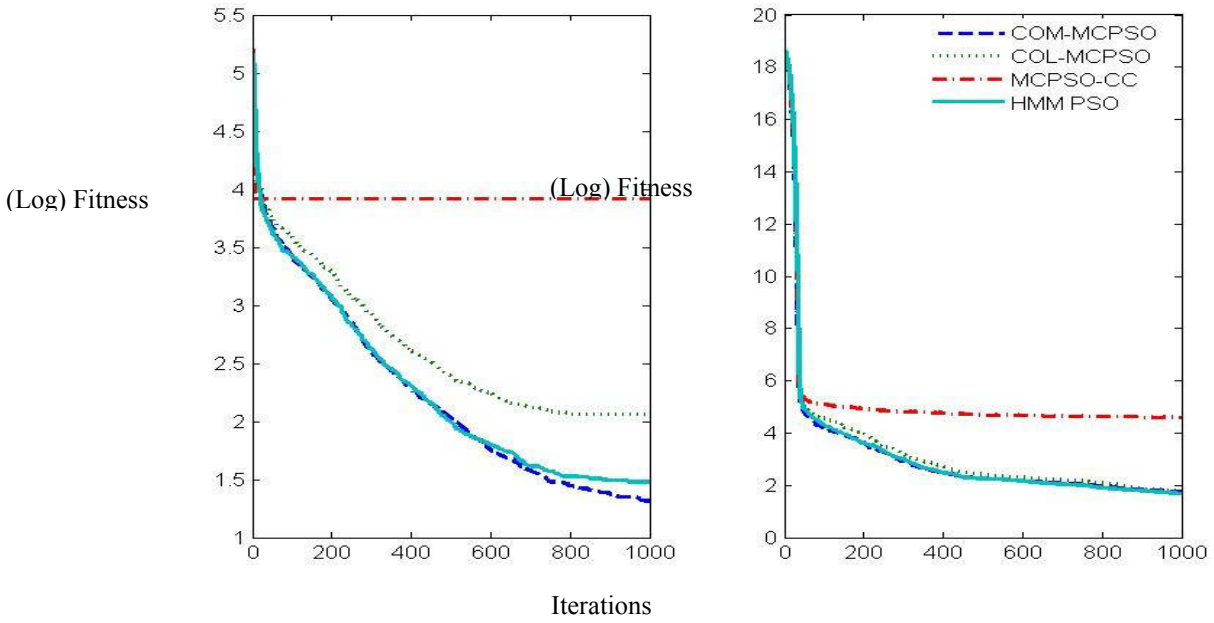

Figure 6.4: Rastrigin and Rosenbrock Functions

With regard to the simpler functions (relative to others) S_n and Sch_n, most algorithms converged extremely fast towards the optimum point, except the MCPSO-CC algorithm for the Sch_n, in Figure 6.5.a. Both the COM-MCPSO and HMM PSO methods had a particularly fast convergence with the same trend on S_n.

With Sch_n , in Figure 6.5.b, the HMM PSO function outperformed the other algorithms in obtaining the best and average final results for Sch_n. Function S_n exhibited a pattern similar to that observed with function Sch_n, except that MCPSO-CC method had performed the same as the other approaches. For this function the HMM PSO algorithm again obtained the best average for the final results but COM-MCPSO obtained the best final fitness in 30 trials.

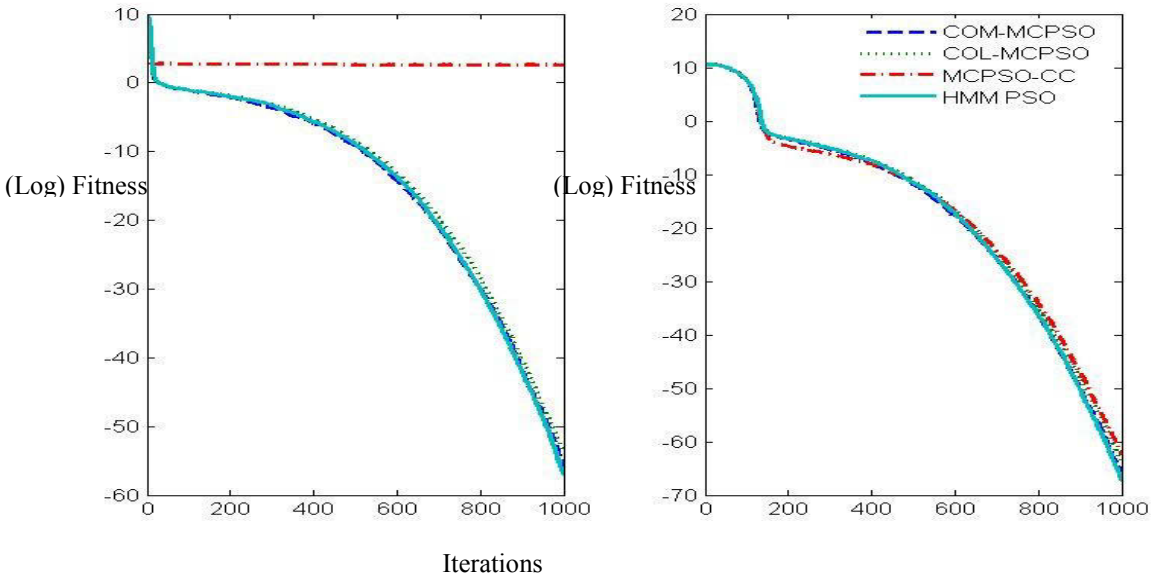

Figure 6.5: S_n and Sch_n Functions

Figures 6.6-6.7 show the multi-swarm process for the functions by the HMM PSO algorithm from a single test run. These figures show that the slave swarms performed well not just in the initial iterations, but also when they made further progress when the master swarm plateaued.

Figure 6.6.a, presents the process of all swarms for the Ackley function for 1000 iterations, and in Figure 6.6.b, by attention to the process between iterations 760 to 880. This figure shows that the master swarms helped each one to improve the fitness function.

Although the master swarms have equal access to the global knowledge, their processes are different and they help each other to improve their fitness. This shows the advantages that multi-master swarms have in relation to the unique master swarm.

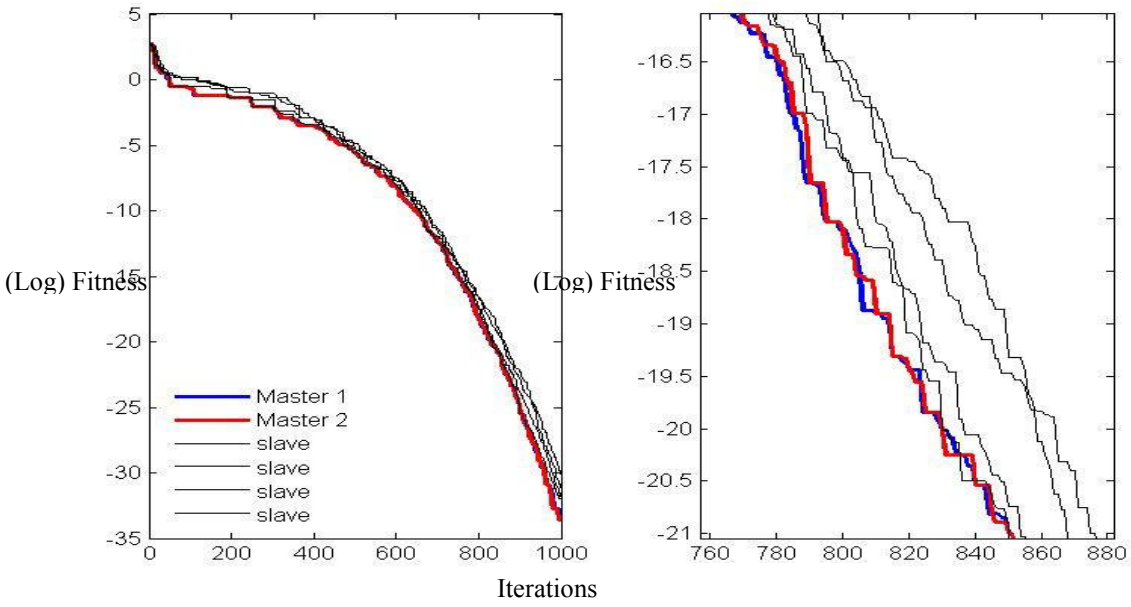

Figure: 6.6: Ackley Functions

Figure 6.7.a, represent the processes of swarms in the HMM PSO algorithm for one try for the Rosenbrock function, while more details of the process of the swarms between iterations 760-880 can be found in Figure 6.7.b. This Figure also shows how coperatioon between all the master swarms and the slave swarms improved the fitness function.

Also, in Figure 6.8 shows the process by which the swarms minised Schwefel's problem 2.22 function, at one try. This figure also shows how important all the slave swarms are, because it is obvious how one of the slave swarms had less impact at improving the fitness function after the first iteration, but improved the process dramatically after iteration 500.

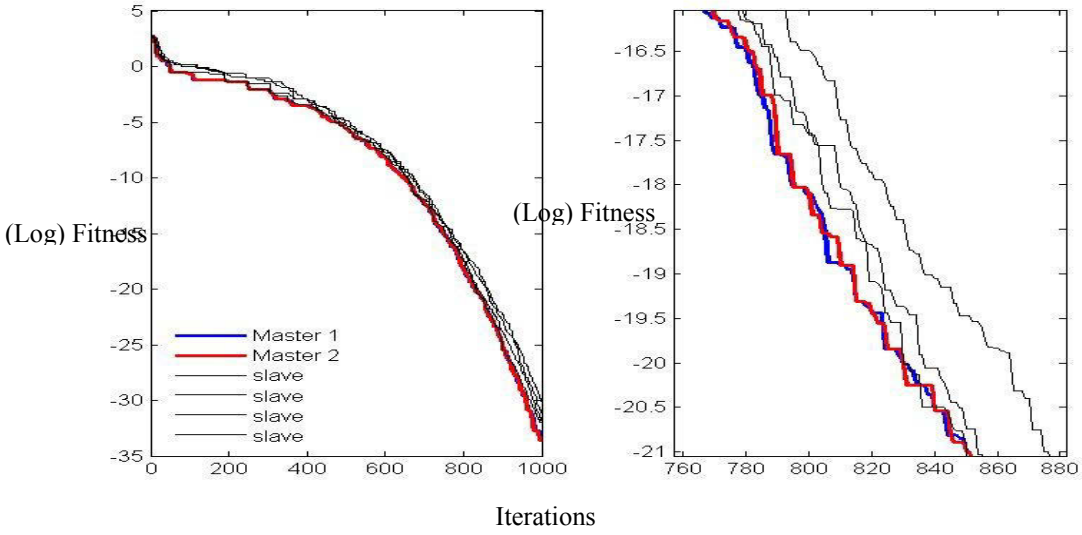

Figure 6.7: Rosenbrock Function

In this section, a multi-master PSO has been proposed to improve the performance of the standard PSO. HMM PSO is a master–slave model that consists of several master swarms and slave swarms. The slave swarms help to amplify the diversity of particles and generate more promising particles for the master swarms. The master swarm uses both its own experience and that of the most successful particles in the slave swarms. The results of six benchmark functions show that the proposed method has superior features, as evidenced by the high quality solution. In a general, an analysis of the results shows that the HMM PSO algorithm was highly competitive with other parallel PSOs on all the problems, usually surpassing its performance. The HMM PSO algorithm obtained the best fitness for most of functions. The only exception was for the Sphere function, for which COM-MCPSO seemed to fine tune its results slightly better than HMM PSO. Moreover, for 4 functions out of 6 the HMM PSO obtained the best average fitness in 30 trials. It can be concluded from this summary that HMM PSO was the best algorithm in this study and the HMM PSO method is a highly consistent strategy in finding the optimum solution compared with other methods. Also, experimental results showed the im-

portance of multi masters, although based on the results, the master swarms have access to the same results, but in different iterations one of them obtained a higher quality result which improved the whole process. When using the single master approaches, all the improvements should be done by just one master.

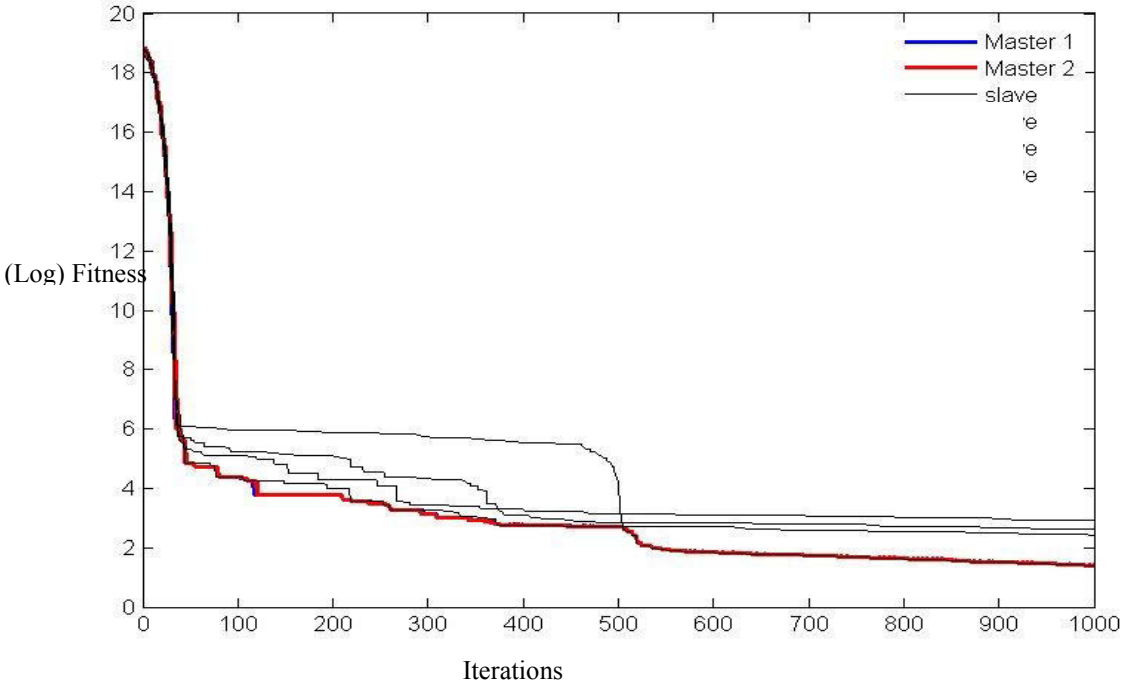

Figure 6.8: Schwefel Function

6.4. Summary

There are two main contributions in this study that can be used in general and not just in sleep apnea studies.

The first algorithm was named self-advising SVM. Different studies have been proposed to improve the performance of SVM, but the main disadvantages of these approaches are the inclusion of new parameters, increased complexity of the learner model, and the difficulty of implementation in solving practical problems such as modelling prior knowledge. The self-advising SVM method proposed here, improved the perfor-

mance of SVM by transferring more information from the training phase to the testing phase. This information was generated by using misclassified data in the training phase, so if the same data or similar patterns appear in the testing data, they will not be misclassified again. The experimental results in this study showed an improvement in accuracy, and the F-score and statistical tests revealed the significance of these improvements. Our results showed that using misclassified data in the training phase by the proposed method is not increasing overtraining. The main advantages of SA-SVM are that it can be applied through any of the SVMs and kernel types, and it does not need the addition of new parameters to the learning phase.

Implementing KNN for computing the advice weights and implementing the proposed approach for multi-class classification may be considered potential areas for future studies.

The second proposed algorithm was a new parallel structure for PSO. HMM PSO is a master–slave model that consists of several master swarm and slave swarms. The slave swarms help to amplify the diversity of particles and then generate more promising particles for the master swarms. The master swarm uses both its own experience and that of the most successful particles in the slave swarms. Experiment results on six benchmark functions showed that the proposed method has superior features, as shown in the high quality of the solution. In a general, an analysis of the results showed that the HMM PSO algorithm to be highly competitive with other parallel PSOs on every problem, indeed, usually surpassing its performance. The HMM PSO algorithm obtained the best fitness for most functions. The only exception being the Sphere function, for which COM-MCPSO seems to fine-tune its results slightly better than HMM PSO. Also for 4 functions out of 6 HMM PSO obtained the best average fitness in 30 trials. From this

summary it can be concluded that HMM PSO was the best performing algorithm in this study and the HMM PSO method is a highly consistent strategy for finding the optimum solution compared to other methods. Moreover, the experimental results showed the importance of multi masters, but based on the results the master swarms have access to the same results, and in different iterations one of them obtained a higher quality result which improved the whole process. When in the single master approaches, all the improvements should be done by just by one master.

Chapter 7

Summary and Future Research

7.1. Introduction

This thesis has established an automated system for sleep apnea study based on soft computing approaches.

The algorithms designed here consists of four main parts; signal segmentation, feature generation followed by dimensionality reduction, and machine learning methods that have been used for detection, classification or prediction. Many experiments have been conducted and several algorithms are developed that can be used in sleep studies or for an even wider range of applications. The conclusions of these experiments and their results are summarised as follows.

7.2. Sleep Apnea Detection

The first part of this thesis commenced with signal segmentation where a new signal segmentation based on nature of sleep apnea was proposed. This segmentation algorithm tries to select time-windows with length of 30 seconds that have a chance of containing a sleep apnea event by attention to the input signals. This segmentation algorithm obtained a higher F-score, 0.831 than blind segmentation, 0.789.

After segmenting the signals, features are generated for each segment by wavelet packet coefficients. Settings for feature generation such as wavelet families, degree and etc. are considered based on state of the arts, in this section 208 feature were generated for each segment.

In this thesis, different subsets of 12 signals from polysomnographic record were ranked based on how well they performed in sleep apnea detection. The best result was obtained by a subset of 8 signals; Snore – ECG –Thoracal – EMG – Abdominal – Air flow – Oxygen saturation – EEG A1-A2, that obtained an accuracy of 88.58%. This ranking was based only on how the signals performed but other criteria were considered in this thesis in order to select signals such as easily be obtained, and signals that are not influenced by the environment. By using these criteria, oxygen saturation, abdominal and thoracic movements signals were used as input signals in this thesis, and improving the performance of these signals by using more powerful machine learning approaches is targeted in this thesis.

In the dimensional reduction phase, two approaches have been considered as (1) feature selection and (2) feature and training instance subset selection, for feature selection a PSO-SVM algorithm was proposed. In this algorithm each particle of PSO represents a subset of features, and SVM evaluates fitness of each particle by attention to its F-score to detect sleep apnea. In this approach PSO was also used to tune the parameters of SVM, this algorithm can also be used for feature selection in other fields. And for the second approach selection of a subset of training data from the available training set was proposed. This algorithm checks if together with feature selection, reducing training data can perform better.

After selecting the signals and features, three main problems were considered in chapter 5: detecting sleep apnea, classifying sleep apnea, and subject independent detection of sleep apnea.

In the detection phase, an improved version of SVM, named Self-Advising SVM was examined. SA-SVM achieved a better result than traditional SVM, and reached an accu-

racy of 95.10% and an F-score of 0.8706 for detecting sleep apnea, while traditional SVM reached accuracy of 94.57% and F-score of 0.8490. It should be stated that the proposed SA-SVM approach can be used in other fields.

Classifying a sleep apnea event to hypopnea, central apnea, obstructive apnea or mixed apnea was also considered in this thesis by using the PSO-SVM algorithm for feature selection. An RBF kernel for the SVM; and one against all strategies was selected, and average accuracy of 88.71% was obtained for 5 independent runs.

Detecting sleep apnea subject independently was also examined in this thesis. For subject independent detection, together with feature selection, instance selection from the training set was also examined. The results showed that reducing the training data with the proposed PSO-SVM algorithm can improve the final performance, such that an average F-score of 0.8447 was obtained and F-score of 0.8116 was obtained for feature selection. In this experiment by attention to the huge size of the search space a new parallel PSO-SVM was proposed, this parallel structure improved the F-score to 0.8673. In the traditional multi-PSO models, all the swarms were at the same level and exchanging information happened based on the definition of neighbourhood. But in this structure the swarms were classified in two different levels as ‘masters’ and ‘slave(s)’. Master swarms have access to the best particle of other swarms but the slave swarms have no access to information of the other swarms, in fact, they only provide information for others. Sending information from the best local particles among the masters and from the slave(s) to the masters can be performed at each iteration by a specified frequency. This parallel PSO structure can also be used in other areas.

7.3. Sleep Apnea Prediction

Predicting individual sleep apnea events was another main part of this thesis, predicting apnea events can be useful to prevent side effect of these events before they happen. In Chapter 6 three different artificial neural networks together with linear and non-linear multi ANNs were used to predict sleep apnea. A PSO algorithm optimised the weights of the linear multi ANNs and the median of these three networks is considered as the non-linear multi ANNs.

Different lead times of 30, 60, 90, and 120 seconds, were used with examining different time windows from 30 seconds to 120 seconds. The results showed that the linear multi ANNs achieved the highest F-score in all settings and choosing a lead time of 90 seconds and using 60 seconds time windows can be reasonable, which obtained an accuracy of 82% and an F-score equal to 0.8114.

7.4. Future Works

In this thesis signals from polysomnographic records were used, but acquiring signals and information by not connected sensors and then using this information for sleep studies can be considered as the future trend in sleep studies. Currently few works have been done such as [326-328] but more works will be done in this area in the near future. Also link of sleep apnea and other diseases such as cardiovascular diseases can be considered more.

Of the methods that used in this thesis, there are a number of directions in which they could be extended, for example:

Signal segmentation can have a huge impact on the final result and while the proposed segmentation algorithm worked properly but it worth further investigation for a more intelligent segmentation algorithm for sleep apnea.

Ranking of signals based on sleep apnea detection was considered in this thesis by attention to their performance in detection. Adding more criteria and using multi criteria decision making algorithms such as TOPSIS [329] can be considered in future.

In this thesis SA-SVM was proposed for binary classification, so extending this approach for multi classification problems such as the classification of sleep apnea to hypopnea, central apnea, obstructive apnea or mixed apnea with SA-SVM can be considered in future works.

Appendix A

In this Appendix some tables from chapter 6 and related to the proposed SA-SVM algorithms are presented. Full information about these tables are presented in section 6.2.

Table A.1: Optimized parameters used for each database

Dataset		C	gamma
Australian Credit Approval (Statlog)	C-SVM	100	0.1250
	V-SVM	32	0.1250
Breast Cancer	C-SVM	8200	0.003
	V-SVM	32	0.5
Contraceptive Method Choice	C-SVM	32	0.002
	V-SVM	32	2
German credit Data (Statlog)	C-SVM	8200	0.003
	V-SVM	32	0.0087
Hayes-Roth	C-SVM	2050	0.1250
	V-SVM	32	0.002
Hepatitis	C-SVM	0.5	0.1250
	V-SVM	32	0.5
Ionosphere	C-SVM	2	0.1250
	V-SVM	32	0.5
iris	C-SVM	0.5	0.003
	V-SVM	0.5	0.003
Liver Disorders	C-SVM	3000	0.03
	V-SVM	32	2
Spambase	C-SVM	3000	0.1250
	V-SVM	32	0.0078
Teaching Assistant Evaluation	C-SVM	3000	2
	V-SVM	32	0.1250

Table A.2: Accuracies of the five-fold validations for C-SVM

Dataset		Run				
		1st	2nd	3rd	4rd	5th
Australian Credit	C-SVM	75.942	63.86	67.39	69.42	69.47
	Self-Advising	77.53	69.52	71.01	74.34	72.96
	Improvement	1.59	5.66	3.62	4.92	3.48
Breast Cancer	C-SVM	96.18	96.92	97.06	97.50	97.06
	Self-Advising	96.33	96.92	97.06	97.65	97.06
	Improvement	0.14	0	0	0.146	0
Contraceptive	C-SVM	72.72	74.45	72.88	71.78	73.98
	Self-Advising	73.82	75.86	74.92	71.94	75.23
	Improvement	1.09	1.41	2.03	0.15	1.25
German credit Data	C-SVM	76.1	75.07	74.7	75.7	77.67
	Self-Advising	75.9	75.27	74.9	75.7	77.77
	Improvement	-0.2	0.20	0.2	0	0.10
Hayes-Roth	C-SVM	76.47	72.27	82.17	80.39	73.26
	Self-Advising	77.45	73.26	82.17	82.35	74.25
	Improvement	0.98	0.99	0	1.96	0.99
Hepatitis	C-SVM	72.65	67.96	67.96	66.40	67.96
	Self-Advising	73.43	69.53	68.75	67.18	67.96
	Improvement	0.78	1.56	0.78	0.78	0
Ionosphere	C-SVM	94.28	95.42	96.57	93.71	94.57
	Self-Advising	94.57	95.42	96.57	93.71	94.57
	Improvement	0.28	0	0	0	0
iris	C-SVM	74.74	74.74	100	100	74.74
	Self-Advising	100	100	100	100	100
	Improvement	25.25	25.25	0	0	25.25
Liver Disorders	C-SVM	72.38	69.18	69.76	69.18	74.70
	Self-Advising	73.25	70.34	70.63	69.47	75.87
	Improvement	0.87	1.16	0.87	0.29	1.16
Spambase	C-SVM	92.64	93.10	93.06	93.29	92.90
	Self-Advising	93.08	93.40	93.34	93.56	93.16
	Improvement	0.43	0.30	0.28	0.26	0.26
Teaching Assistant	C-SVM	72.44	76	63.63	66.32	69.69
	Self-Advising	73.46	77	63.63	67.34	71.71
	Improvement	1.02	1	0	1.02	2.02

Table A.3: Accuracies of the five-fold validations for V-SVM

Dataset						
Australian Credit	V-SVM	86.21	84.90	85.21	84.20	87.08
	Self-Advising	86.06	84.90	85.21	84.20	87.08
	Improvement	-0.14	0	0	0	0
Breast Cancer	V-SVM	96.33	97.36	96.18	96.48	97.50
	Self-Advising	96.48	97.65	96.18	96.48	97.50
	Improvement	0.146	0.29	0	0	0
Contraceptive	V-SVM	71.00	73.82	72.72	72.41	69.43
	Self-Advising	72.57	74.29	72.10	73.35	70.53
	Improvement	1.56	0.47	-0.62	0.94	1.09
German credit	V-SVM	76.57	74.17	74.2	74.97	75.05
	Self-Advising	76.67	74.17	74.7	74.77	75.15
	Improvement	0.10	0	0.5	-0.20	0.10
Hayes-Roth	V-SVM	73.26	76.23	77.45	82	57.84
	Self-Advising	74.25	79.20	79.41	82	61.76
	Improvement	0.99	2.97	1.96	0	3.92
Hepatitis	V-SVM	65.625	67.18	67.96	65.62	65.62
	Self-Advising	66.40	67.96	67.96	65.62	66.40
	Improvement	0.78	0.78	0	0	0.78
Ionosphere	V-SVM	93.71	95.14	95.14	93.71	92.57
	Self-Advising	93.71	95.14	95.14	93.71	92.57
	Improvement	0	0	0	0	0
iris	V-SVM	97.97	100	100	100	99
	Self-Advising	98.98	100	100	100	100
	Improvement	1.01	0	0	0	1
Liver Disorders	V-SVM	70.93	69.18	72.09	67.15	69.47
	Self-Advising	70.93	69.76	72.96	68.02	70.63
	Improvement	0	0.58	0.87	0.87	1.16
Spambase	V-SVM	90.27	89.73	89.99	90.25	89.83
	Self-Advising	91.94	91.36	91.88	91.82	91.77
	Improvement	1.67	1.63	1.89	1.56	1.93
Teaching Assistant	V-SVM	74.74	69.69	76.76	73.46	77
	Self-Advising	80.80	78.78	83.8	79.59	82
	Improvement	6.06	9.09	7.07	6.12	5

Table A.4: F-scores of the five, five-fold validations for C-SVM

Dataset						
Australian Credit	C-SVM	0.81	0.75	0.77	0.78	0.78
	Self-Advising	0.82	0.78	0.78	0.80	0.80
	Improvement	0.01	0.02	0.01	0.03	0.02
Breast Cancer	C-SVM	0.94	0.95	0.95	0.96	0.95
	Self-Advising	0.94	0.95	0.95	0.96	0.95
	Improvement	0.01	0	0	0.01	0
Contraceptive	C-SVM	0.61	0.61	0.60	0.57	0.61
	Self-Advising	0.62	0.63	0.63	0.57	0.62
	Improvement	0.01	0.02	0.03	-0.01	0.01
German credit	C-SVM	0.55	0.55	0.53	0.51	0.57
	Self-Advising	0.55	0.56	0.53	0.51	0.57
	Improvement	-0.01	0.01	0.01	0.01	0.01
Hayes-Roth	C-SVM	0.76	0.74	0.82	0.81	0.73
	Self-Advising	0.76	0.74	0.82	0.82	0.74
	Improvement	0.01	0.01	0	0.01	0.01
Hepatitis	C-SVM	0.61	0.55	0.51	0.51	0.57
	Self-Advising	0.62	0.58	0.52	0.52	0.577
	Improvement	0.0154	0.03	0.01	0.02	0
Ionosphere	C-SVM	0.922	0.93	0.95	0.91	0.92
	Self-Advising	0.92	0.93	0.95	0.91	0.92
	Improvement	0.01	0	0	0	0
iris	C-SVM	0.65	0.65	1	1	0.65
	Self-Advising	1	1	1	1	1
	Improvement	0.34	0.34	0	0	0.34
Liver Disorders	C-SVM	0.77	0.74	0.74	0.75	0.80
	Self-Advising	0.78	0.75	0.75	0.75	0.80
	Improvement	0.01	0.01	0.01	0.01	0.01
Spambase	C-SVM	0.94	0.94	0.94	0.94	0.94
	Self-Advising	0.94	0.94	0.94	0.94	0.94
	Improvement	0.01	0.01	0.01	0.01	0.01
Teaching Assistant	C-SVM	0.74	0.77	0.63	0.64	0.68
	Self-Advising	0.75	0.78	0.63	0.65	0.70
	Improvement	0.01	0.01	0	0.01	0.03

Table A.5: F-scores of the five, five-fold validations for V-SVM

Dataset						
Australian Credit	V-SVM	0.875	0.85	0.86	0.85	0.88
	Self-Advising	0.874016	0.85	0.86	0.85	0.88
	Improvement	-0.01	0	0	0	0
Breast Cancer	V-SVM	0.94	0.96	0.94	0.94	0.96
	Self-Advising	0.94	0.96	0.94	0.94	0.96
	Improvement	0.01	0.01	0	0	0
Contraceptive Method	V-SVM	0.557	0.60	0.56	0.56	0.54
	Self-Advising	0.57	0.61	0.54	0.58	0.55
	Improvement	0.02	0.01	-0.01	0.01	0.01
German credit	V-SVM	0.55	0.50	0.50	0.54	0.55
	Self-Advising	0.55	0.50	0.52	0.54	0.55
	Improvement	0.01	0.01	0.01	-0.01	0.01
Hayes-Roth	V-SVM	0.74	0.80	0.76	0.83	0.55
	Self-Advising	0.75	0.82	0.79	0.83	0.60
	Improvement	0.01	0.02	0.02	-0.01	0.049
Hepatitis	V-SVM	0.57	0.61	0.60	0.59	0.56
	Self-Advising	0.59	0.62	0.60	0.59	0.58
	Improvement	0.01	0.01	0	0	0.01
Ionosphere	V-SVM	0.90	0.92	0.92	0.90	0.88
	Self-Advising	0.90	0.92	0.92	0.90	0.88
	Improvement	0	0	0	0	0
iris	V-SVM	0.97	1	1	1	0.98
	Self-Advising	0.98	1	1	1	1
	Improvement	0.01	0	0	0	0.01
Liver Disorders	V-SVM	0.74	0.73	0.77	0.73	0.74
	Self-Advising	0.74	0.73	0.77	0.73	0.75
	Improvement	0	0.01	0.01	0.01	0.01
Spambase	V-SVM	0.92	0.91	0.92	0.92	0.92
	Self-Advising	0.93	0.93	0.93	0.93	0.93
	Improvement	0.01	0.01	0.01	0.01	0.01
Teaching Assis-tant	V-SVM	0.72	0.67	0.77	0.71	0.77
	Self-Advising	0.79	0.77	0.84	0.78	0.82
	Improvement	0.07	0.10	0.06	0.07	0.05

References

1. Chokroverty, S., *Overview of sleep & sleep disorders.* Indian Journal of Medical Research, 2010. **131**(2): p. 126-140.
2. Gath, I., C. Feuerstein, and A. Geva, *unsupervised classification and adaptive definition of sleep patterns.* Pattern Recognition Letters, 1994. **15**(10): p. 977-984.
3. Cardoso, E., A. Batista, R. Rodrigues, M. Ortigueira, C. Barbara, C. Martinho, and R. Rato, *A Contribution for the Automatic Sleep Classification Based on the Itakura-Saito Spectral Distance*, in *Emerging Trends in Technological Innovation*, L.M. CamarinhaMatos, P. Pereira, and L. Ribeiro, Editors. p. 374-381.
4. Moser, D., P. Anderer, G. Gruber, S. Parapatics, E. Loretz, M. Boeck, G. Kloesch, E. Heller, A. Schmidt, H. Danker-Hopfe, B. Saletu, J. Zeitlhofer, and G. Dorffner, *Sleep Classification According to AASM and Rechtschaffen & Kales: Effects on Sleep Scoring Parameters.* Sleep, 2009. **32**(2): p. 139-149.
5. Klink, M. and S.F. Quan, *Prevalence of reported sleep disturbances in a general adult-population and their relationship to obstractive airways diseases.* Chest, 1987. **91**(4): p. 540-546.
6. Chokroverty, S., C. Sudhansu, Md, Frcp, and Facp, *Approach to the Patient with Sleep Complaints*, in *Sleep Disorders Medicine (Third Edition)*2009, W.B. Saunders: Philadelphia. p. 255-274.
7. Young, T.B., *Epidemiology of daytime sleepiness: Definitions, symptomatology, and prevalence.* Journal of Clinical Psychiatry, 2004. **65**: p. 12-16.
8. Chokroverty, S., C. Sudhansu, Md, Frcp, and Facp, *Sleep deprivation and sleepiness*, in *Sleep Disorders Medicine (Third Edition)*2009, W.B. Saunders: Philadelphia. p. 22-28.
9. Guilleminault, C., J. van den Hoed, and M. Mitler, *Overview of the sleep apnea syndromes. In: C. Guilleminault and Wc Dement, Editors, Sleep apnea syndromes, Alan R Liss, New York.* 1978: p. 1-12.
10. Young, T.B., M. Palta, J. Dempsey, J. Skatrud, S. Weber, and S. Badr, *Occurrence of sleep-disordered breathing among middle-aged adults in the wisconsin sleep cohort study.* American Review of Respiratory Disease, 1993. **147**(4): p. A233-A233.
11. Kryger, M.H., *management of obstractive sleep-apnea.* Clinics in Chest Medicine, 1992. **13**(3): p. 481-492.
12. Pradhan, B., *Use of GIS-based fuzzy logic relations and its cross application to produce landslide susceptibility maps in three test areas in Malaysia.* Environmental Earth Sciences, 2011. **63**(2): p. 329-349.
13. Balabin, R.M. and E.I. Lomakina, *Support vector machine regression (SVR/LS-SVM)-an alternative to neural networks (ANN) for analytical chemistry? Comparison of nonlinear methods on near infrared (NIR) spectroscopy data.* Analyst, 2011. **136**(8): p. 1703-1712.
14. Espejo, P.G., S. Ventura, and F. Herrera, *A Survey on the Application of Genetic Programming to Classification.* Ieee Transactions on Systems Man and Cybernetics Part C-Applications and Reviews, 2010. **40**(2): p. 121-144.
15. Maier, H.R., A. Jain, G.C. Dandy, and K.P. Sudheer, *Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions.* Environmental Modelling & Software, 2010. **25**(8): p. 891-909.
16. Harris, T.J.R. and F. McCormick, *The molecular pathology of cancer.* Nature Reviews Clinical Oncology, 2010. **7**(5): p. 251-265.
17. Gubbi, J., A. Khandoker, and M. Palaniswami. *Classification of obstructive and central sleep apnea using wavelet packet analysis of ECG signals*. in *Computers in Cardiology, 2009*. 2009.
18. Guijarro-Berdiñas, B., E. Hernández-Pereira, and D. Peteiro-Barral, *A mixture of experts for classifying sleep apneas.* Expert Systems with Applications, 2012. **39**(8): p. 7084-7092.
19. Hang, L.-W., H.-H. Lin, J.Y. Chiang, H.-L. Wang, and Y.-F. Chen, *Diagnosis of Severe Obstructive Sleep Apnea with Model Designed Using Genetic Algorithm and Ensemble Support Vector Machine.* Appl. Math, 2013. **7**(1S): p. 227S-336S.
20. Khandoker, A.H., J. Gubbi, and M. Palaniswami, *Automated Scoring of Obstructive Sleep Apnea and Hypopnea Events Using Short-Term Electrocardiogram Recordings.* Ieee Transactions on Information Technology in Biomedicine, 2009. **13**(6): p. 1057-1067.
21. Dickens, C., *The posthumous papers of the pickwick club (1st ed.), Chapman and Hall, London.* 1837.
22. Flemons, W.W., D. Buysse, S. Redline, A. Pack, K. Strohl, J. Wheatley, T. Young, N. Douglas, P. Levy, W. McNicholas, J. Fleetham, D. White, W. Schmidt-Nowarra, D. Carley, J. Romaniuk, and F. Amer Acad Sleep Med Task, *Sleep-related breathing disorders in adults: Recommendations for syndrome definition and measurement techniques in clinical research.* Sleep, 1999. **22**(5): p. 667-689.
23. Shelton, R.L. and J.F. Bosma, *Maintenance of pharyngeal airway.* Journal of Applied Physiology, 1962. **17**(2): p. 209-&.
24. Remmers, J.E., W.J. Degroot, E.K. Sauerland, and A.M. Anch, *Pathogenesis of upper airway occlusion during sleep.* Journal of Applied Physiology, 1978. **44**(6): p. 931-938.
25. Emin Tagluk, M., M. Akin, and N. Sezgin, *ClassIfIcation of sleep apnea by using wavelet transform and artificial neural networks.* Expert Systems with Applications. **37**(2): p. 1600-1607.

26. Whitelaw, W.A. and K.R. Burgess, *Diagnosis of sleep apnoea: some critical issues.* Indian Journal of Medical Research, 2010. **131**(2): p. 217-229.
27. Stearns, J.D. and T.L. Stierer, *Peri-operative identification of patients at risk for obstructive sleep apnea.* Seminars in Anesthesia, Perioperative Medicine and Pain, 2007. **26**(2): p. 73-82.
28. Banno, K. and M.H. Kryger, *Sleep apnea: Clinical investigations in humans.* Sleep Medicine, 2007. **8**(4): p. 400-426.
29. Ward Flemons, W. and W.T. McNicholas, *Clinical prediction of the sleep apnea syndrome.* Sleep Medicine Reviews, 1997. **1**(1): p. 19-32.
30. Sin, D.D., F. Fitzgerald, J.D. Parker, G. Newton, J.S. Floras, and T.D. Bradley, *Risk factors for central and obstructive sleep apnea 450 men and women with congestive heart failure.* American Journal of Respiratory and Critical Care Medicine, 1999. **160**(4): p. 1101-1106.
31. Young, T., P.E. Peppard, and D.J. Gottlieb, *Epidemiology of obstructive sleep apnea - A population health perspective.* American Journal of Respiratory and Critical Care Medicine, 2002. **165**(9): p. 1217-1239.
32. Peppard, P.E., T. Young, M. Palta, J. Dempsey, and J. Skatrud, *Longitudinal study of moderate weight change and sleep-disordered breathing.* Jama-Journal of the American Medical Association, 2000. **284**(23): p. 3015-3021.
33. Tishler, P.V., E.K. Larkin, M.D. Schluchter, and S. Redline, *Incidence of sleep-disordered breathing in an urban adult population - The relative importance of risk factors in the development of sleep-disordered breathing.* Jama-Journal of the American Medical Association, 2003. **289**(17): p. 2230-2237.
34. Romero-Corral, A., S.M. Caples, F. Lopez-Jimenez, and V.K. Somers, *Interactions between obesity and obstructive sleep apnea implications for treatment.* Chest. **137**(3): p. 711-719.
35. Guilleminault, C., M. Partinen, K. Hollman, N. Powell, and R. Stoohs, *Familial aggregates in obstractive sleep-apnea syndrome.* Chest, 1995. **107**(6): p. 1545-1551.
36. Mathur, R. and N.J. Douglas, *Family studies in patients with sleep-apnea hypopnea syndrome.* Annals of Internal Medicine, 1995. **122**(3): p. 174-178.
37. Redline, S., P.V. Tishler, T.D. Tosteson, J. Williamson, K. Kump, I. Browner, V. Ferrette, and P. Krejci, *The familial aggregating of obstractive sleep-apnea.* American Journal of Respiratory and Critical Care Medicine, 1995. **151**(3): p. 682-687.
38. Schwab, R.J., M. Pasirstein, R. Pierson, A. Mackley, R. Hachadoorian, R. Arens, G. Maislin, and A.I. Pack, *Identification of upper airway anatomic risk factors for obstructive sleep apnea with volumetric magnetic resonance imaging.* American Journal of Respiratory and Critical Care Medicine, 2003. **168**(5): p. 522-530.
39. Shintani, T., K. Asakura, and A. Kataura, *Adenotonsillar hypertrophy and skeletal morphology of children with obstructive sleep apnea syndrome.* Acta Oto-Laryngologica, 1996: p. 222-224.
40. Hamans, E., E.A. Van Marck, W.A. De Backer, W. Creten, and P.H. Van de Heyning, *Morphometric analysis of the uvula in patients with sleep-related breathing disorders.* European Archives of Oto-Rhino-Laryngology, 2000. **257**(4): p. 232-236.
41. Do, K.L., H. Ferreyra, J.F. Healy, and T.M. Davidson, *Does tongue size differ between patients with and without sleep-disordered breathing?* Laryngoscope, 2000. **110**(9): p. 1552-1555.
42. Cakirer, B., M.G. Hans, G. Graham, J. Aylor, P.V. Tishler, and S. Redline, *The relationship between craniofacial morphology and obstructive sleep apnea in whites and in African-Americans.* American Journal of Respiratory and Critical Care Medicine, 2001. **163**(4): p. 947-950.
43. Ayas, N.T., R. Brown, and S.A. Shea, *Hypercapnia can induce arousal from sleep in the absence of altered respiratory mechanoreception.* American Journal of Respiratory and Critical Care Medicine, 2000. **162**(3): p. 1004-1008.
44. Benlloch, E., P. Cordero, P. Morales, J.J. Soler, and V. Macian, *Ventilatory pattern at rest and response to hypercapnic in patients with obstractive sleep-apnea syndrome.* Respiration, 1995. **62**(1): p. 4-9.
45. Zamarron, C., F. Gude, Y. Otero, J.M. Alvarez, A. Golpe, and J.R. Rodriguez, *Prevalence of sleep disordered breathing and sleep apnea in 50-to 70-year-old individuals - A survey.* Respiration, 1999. **66**(4): p. 317-322.
46. Young, T., E. Shahar, F.J. Nieto, S. Redline, A.B. Newman, D.J. Gottlieb, J.A. Walsleben, L. Finn, P. Enright, J.M. Samet, and G. Sleep Heart Hlth Study Res, *Predictors of sleep-disordered breathing in community-dwelling adults - The sleep heart health study.* Archives of Internal Medicine, 2002. **162**(8): p. 893-900.
47. Young, T., J. Skatrud, and P.E. Peppard, *Risk factors for obstructive sleep apnea in adults.* Jama-Journal of the American Medical Association, 2004. **291**(16): p. 2013-2016.
48. Ancoli-Israel, S., P. Gehrman, D.F. Kripke, C. Stepnowsky, W. Mason, M. Cohen-Zion, and M. Marler, *Long-term follow-up of sleep disordered breathing in older adults.* Sleep Medicine, 2001. **2**(6): p. 511-516.
49. Young, T., M. Palta, J. Dempsey, J. Skatrud, S. Weber, and S. Badr, *the occurrence of sleep-disordered breathing among middle-aged adults.* New England Journal of Medicine, 1993. **328**(17): p. 1230-1235.
50. Block, A.J., P.G. Boysen, J.W. Wynne, and L.A. Hunt, *Sleep apnea, hypopnea and oxygen desaturation in normal subjects - strong male predominance.* New England Journal of Medicine, 1979. **300**(10): p. 513-517.

51. O'Connor, C., K.S. Thornley, and P.J. Hanly, *Gender differences in the polysomnographic features of obstructive sleep apnea.* American Journal of Respiratory and Critical Care Medicine, 2000. **161**(5): p. 1465-1472.
52. Jordan, A.S. and R.D. McEvoy, *Gender differences in sleep apnea: epidemiology, clinical presentation and pathogenic mechanisms.* Sleep Medicine Reviews, 2003. **7**(5): p. 377-389.
53. Pillar, F., A. Malhotra, R. Fogel, J. Beauregard, R. Schnall, and D.P. White, *Airway mechanics and ventilation in response to resistive loading during sleep - Influence of gender.* American Journal of Respiratory and Critical Care Medicine, 2000. **162**(5): p. 1627-1632.
54. Malhotra, A., Y.Q. Huang, R.B. Fogel, G. Pillar, J.K. Edwards, R. Kikinis, S.H. Loring, and D.P. White, *The male predisposition to pharyngeal collapse - Importance of airway length.* American Journal of Respiratory and Critical Care Medicine, 2002. **166**(10): p. 1388-1395.
55. Young, T., R. Hutton, L. Finn, S. Badr, and M. Palta, *The gender bias in sleep apnea diagnosis - Are women missed because they have different symptoms?* Archives of Internal Medicine, 1996. **156**(21): p. 2445-2451.
56. Redline, S., K. Kump, P.V. Tishler, I. Browner, and V. Ferrette, *Gender differences in sleep-disordered breathing in a community-based sample.* American Journal of Respiratory and Critical Care Medicine, 1994. **149**(3): p. 722-726.
57. Smith, R., J. Ronald, K. Delaive, R. Walld, J. Manfreda, and M.H. Kryger, *What are obstructive sleep apnea patients being treated for prior to this diagnosis?* Chest, 2002. **121**(1): p. 164-172.
58. Teschler, H., M. BerthonJones, T. Wessendorf, H.J. Meyer, and N. Konietzko, *Influence of moderate alcohol consumption on obstructive sleep apnoea with and without AutoSet(TM) nasal CPAP therapy.* European Respiratory Journal, 1996. **9**(11): p. 2371-2377.
59. Sahlin, C., K.A. Franklin, H. Stenlund, and E. Lindberg, *Sleep in women: Normal values for sleep stages and position and the effect of age, obesity, sleep apnea, smoking, alcohol and hypertension.* Sleep Medicine, 2009. **10**(9): p. 1025-1030.
60. Vakulin, A., S.D. Baulk, P.G. Catcheside, N.A. Antic, C.J. van den Heuvel, J. Dorrian, and R.D. McEvoy, *Effects of alcohol and sleep restriction on simulated driving performance in untreated patients with obstructive sleep apnea.* Annals of Internal Medicine, 2009. **151**(7): p. 447-W145.
61. Scanlan, M.F., T. Roebuck, P.J. Little, J.R. Redman, and M.T. Naughton, *Effect of moderate alcohol upon obstructive sleep apnoea.* European Respiratory Journal, 2000. **16**(5): p. 909-913.
62. Balaguer, C., A. Palou, and A. Alonso-Fernandez, *Smoking and Sleep Disorders.* Archivos De Bronconeumologia, 2009. **45**(9): p. 449-458.
63. Scrima, L., M. Broudy, K.N. Nay, and M.A. Cohn, *Increased severity of obstractive sleep-apnea after bedtime alcohol ingestion-diagnostic potential and proposed mechanism of action* Sleep, 1982. **5**(4): p. 318-328.
64. Taasan, V.C., A.J. Block, P.G. Boysen, and J.W. Wynne, *Alcohol increases sleep-apnea and oxygen desaturation in asymptomatic men.* American Journal of Medicine, 1981. **71**(2): p. 240-245.
65. Davies, R.J.O., N.J. Ali, and J.R. Stradling, *Neck circumference and other clinica-features in the diagnosis of the obstractive sleep-apnea syndrome.* Thorax, 1992. **47**(2): p. 101-105.
66. Mortimore, I.L., I. Marshall, P.K. Wraith, R.J. Sellar, and N.J. Douglas, *Neck and total body fat deposition in nonobese and obese patients with sleep apnea compared with that in control subjects.* American Journal of Respiratory and Critical Care Medicine, 1998. **157**(1): p. 280-283.
67. Sauter, Asenbaum, Popovic, Bauer, Lamm, Klösch, and Zeitlhofer, *Excessive daytime sleepiness in patients suffering from different levels of obstructive sleep apnoea syndrome.* Journal of Sleep Research, 2000. **9**(3): p. 293-301.
68. Engleman, H.M., W.S.J. Hirst, and N.J. Douglas, *Under reporting of sleepiness and driving impairment in patients with sleep apnoea hypopnoea syndrome.* Journal of Sleep Research, 1997. **6**(4): p. 272-275.
69. Hoffstein, V., *Snoring*, in *Principles and practice of sleep medicine*, M.H. Kryger, T. Roth, and D.W. C., Editors. 2000, Saunders: Philadelphia, PA: W.B. p. 813-826.
70. Flemons, W.W., W.A. Whitelaw, R. Brant, and J.E. Remmers, *Likelihood ratios for a sleep-apnea clinical-prediction rule.* American Journal of Respiratory and Critical Care Medicine, 1994. **150**(5): p. 1279-1285.
71. Viner, S., J.P. Szalai, and V. Hoffstein, *Are history and physical - examination a good screening- test for sleep-apnea.* Annals of Internal Medicine, 1991. **115**(5): p. 356-359.
72. Gottlieb, D.J., Q. Yao, S. Redline, T. Ali, M.W. Mahowald, and G. Sleep Heart Hlth Study Res, *Does snoring predict sleepiness independently of apnea and hypopnea frequency?* American Journal of Respiratory and Critical Care Medicine, 2000. **162**(4): p. 1512-1517.
73. Ferreira, S., J. Winck, P. Bettencourt, and F. Rocha-Goncalves, *Heart failure and sleep apnoea: To sleep perchance to dream.* European Journal of Heart Failure, 2006. **8**(3): p. 227-236.
74. Kales, A., R.J. Cadieux, E.O. Bixler, C.R. Soldatos, A. Velabueno, C.A. Misoul, and T.W. Locke, *Severe obstractive sleep-apnea .1. onset, clinical course, an characteristics.* Journal of Chronic Diseases, 1985. **38**(5): p. 419-425.

75. Findley, L.J., M.E. Unverzagt, and P.M. Suratt, *Automobile accidents involving patients with obstractive sleep-apnea.* American Review of Respiratory Disease, 1988. **138**(2): p. 337-340.
76. George, C.F., P.W. Nickerson, P.J. Hanly, T.W. Millar, and M.H. Kryger, *Sleep-apnea patients have more automobile accidents.* Lancet, 1987. **2**(8556): p. 447-447.
77. Kapur, V., D.K. Blough, R.E. Sandblom, R. Hert, J.B. de Maine, S.D. Sullivan, and B.M. Psaty, *The medical cost of undiagnosed sleep apnea.* Sleep, 1999. **22**(6): p. 749-755.
78. Findley, L.J. and P.M. Suratt, *Serious motor vehicle crashes: the cost of untreated sleep apnoea.* Thorax, 2001. **56**(7): p. 505-505.
79. Wittmann, V. and D.O. Rodenstein, *Health care costs and the sleep apnea syndrome.* Sleep Medicine Reviews, 2004. **8**(4): p. 269-279.
80. Mar, J., J.R. Rueda, J. Duran-Cantolla, C. Schechter, and J. Chilcott, *The cost-effectiveness of nCPAP treatment in patients with moderate-to-severe obstructive sleep apnoea.* European Respiratory Journal, 2003. **21**(3): p. 515-522.
81. Jurado Gamez, B., J. Redel Montero, L. Munoz Cabrera, M.C. Fernandez Marin, E. Munoz Gomariz, M.A. Martin Perez, and A. Cosano Povedano, *Cost-effectiveness and degree of satisfaction with home sleep monitoring in patients with symptoms of sleep apnea.* Archivos De Bronconeumologia, 2007. **43**(11): p. 605-610.
82. Alvarez, M.D., J.T. Santos, J.C. Guevara, M.G. Martinez, L.R. Pascual, J.L.V. Banuelos, and A.M. Cabello, *Reliability of home respiratory polygraphy for the diagnosis of sleep apnea-hypopnea syndrome: Analysis of costs.* Archivos De Bronconeumologia, 2008. **44**(1): p. 22-28.
83. Amer Acad Sleep, M., *Cost justification for diagnosis and treatment of obstructive sleep apnea.* Sleep, 2000. **23**(8): p. 1017-1018.
84. Ronald, J., K. Delaive, L. Roos, J. Manfreda, A. Bahammam, and M.H. Kryger, *Health care utilization in the 10 years prior to diagnosis in obstructive sleep apnea syndrome patients.* Sleep, 1999. **22**(2): p. 225-229.
85. Ball, E.M., R.D. Simon, A.A. Tall, M.B. Banks, G. NinoMurcia, and W.C. Dement, *Diagnosis and treatment of sleep apnea within the community - The Walla Walla project.* Archives of Internal Medicine, 1997. **157**(4): p. 419-424.
86. Kryger, M.H., L. Roos, K. Delaive, R. Walld, and J. Horrocks, *Utilization of health care services in patients with severe obstructive sleep apnea.* Sleep, 1996. **19**(9): p. S111-S116.
87. Bearpark, H., L. Elliott, R. Grunstein, S. Cullen, H. Schneider, W. Althaus, and C. Sullivan, *Snoring and sleep-apnea - a population study in Australian men.* American Journal of Respiratory and Critical Care Medicine, 1995. **151**(5): p. 1459-1465.
88. Penzel, T., J. McNames, P. de Chazal, B. Raymond, A. Murray, and G. Moody, *Systematic comparison of different algorithms for apnoea detection based on electrocardiogram recordings.* Medical & Biological Engineering & Computing, 2002. **40**(4): p. 402-407.
89. Li, Y.X., V. Chongsuvivatwong, A. Geater, and A. Liu, *Exhaled breath condensate cytokine level as a diagnostic tool for obstructive sleep apnea syndrome.* Sleep Medicine, 2009. **10**(1): p. 95-103.
90. Heruti, R., T. Shochat, D. Tekes-Manova, I. Ashkenazi, and D. Justo, *Association between erectile dysfunction and sleep disorders measured by self-assessment questionnaires in adult men.* Journal of Sexual Medicine, 2005. **2**(4): p. 543-550.
91. Abrishami, A., A. Khajehdehi, and F. Chung, *A systematic review of screening questionnaires for obstructive sleep apnea.* Canadian Journal of Anaesthesia-Journal Canadien D Anesthesie. **57**(5): p. 423-438.
92. Behbehani, K., F.C. Yen, J.R. Burk, E.A. Lucas, and J.R. Axe, *Automatic-control of airway pressure for treatment of obstractive sleep-apnea.* Ieee Transactions on Biomedical Engineering, 1995. **42**(10): p. 1007-1016.
93. Alshaer, H., G.R. Fernie, E. Sejdic, and T.D. Bradley. *Adaptive segmentation and normalization of breathing acoustic data of subjects with obstructive sleep apnea*. in *Science and Technology for Humanity (TIC-STH), 2009 IEEE Toronto International Conference*. 2009.
94. Sola-Soler, J., R. Jane, J.A. Fiz, and J. Morera. *Automatic classification of subjects with and without Sleep Apnea through snoring analysis*. in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*. 2007.
95. Sola-Soler, J., R. Jane, J.A. Fiz, and J. Morera. *Pitch analysis in snoring signals from simple snorers and patients with obstructive sleep apnea*. in *[Engineering in Medicine and Biology, 2002. 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society] EMBS/BMES Conference, 2002. Proceedings of the Second Joint*. 2002.
96. Abeyratne, U.R., C.K.K. Patabandi, and K. Puvanendran. *Pitch-jitter analysis of snoring sounds for the diagnosis of sleep apnea*. in *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*. 2001.
97. Ng, A.K., K.Y. Wong, C.H. Tan, and T.S. Koh. *Bispectral analysis of snore signals for obstructive sleep apnea detection*. in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*. 2007.
98. Ghaemmaghami, H., U.R. Abeyratne, and C. Hukins. *Normal probability testing of snore signals for diagnosis of obstructive sleep apnea*. in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. 2009.

99. Ng, A.K., T.S. Koh, E. Baey, and K. Puvanendran. *Speech-like analysis of snore signals for the detection of obstructive sleep apnea*. in *Biomedical and Pharmaceutical Engineering, 2006. ICBPE 2006. International Conference on*. 2006.

100. Sola-Soler, J., R. Jane, J.A. Fiz, and J. Morera. *Variability of snore parameters in time and frequency domains in snoring subjects with and without Obstructive Sleep Apnea*. in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*. 2005.

101. Mikami, T. *Detecting nonlinear properties of snoring sounds for sleep apnea diagnosis*. in *Bioinformatics and Biomedical Engineering, 2008. ICBBE 2008. The 2nd International Conference on*. 2008.

102. Sola-Soler, J., R. Jane, J.A. Fiz, and J. Morera. *Spectral envelope analysis in snoring signals from simple snorers and patients with Obstructive Sleep Apnea*. in *Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE*. 2003.

103. Ng, A.K. and T.S. Koh. *Using psychoacoustics of snoring sounds to screen for obstructive sleep apnea*. in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*. 2008.

104. Emoto, T., U.R. Abeyratne, M. Akutagawa, H. Nagashino, and Y. Kinouchi. *Feature extraction for snore sound via neural network processing*. in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*. 2007.

105. Penzel, T., R. Fricke, H.F. Becker, R. Conradt, A. Jerrentrup, and J.H. Peter, *Comparison of peripheral arterial tonometry and invasive blood pressure in obstructive sleep apnea.* Sleep, 2001. **24**: p. 450.

106. Penzel, T., K. Kesper, T. Ploch, H.F. Becker, and C. Vogelmeier. *Ambulatory Recording of Sleep Apnea Using Peripheral Arterial Tonometry*. in *Engineering in Medicine and Biology Society, 2004. IEMBS '04. 26th Annual International Conference of the IEEE*. 2004.

107. Castiglioni, P., M.R. Bonsignore, G. Insalaco, G. Parati, and M. Di Rienzo. *Signal processing procedures for the evaluation of the cardiovascular effects in the obstructive sleep apnea syndrome*. in *Computers in Cardiology 2001*. 2001.

108. Nazeran, H., A. Almas, K. Behbehani, J. Burk, and E. Lucas. *A fuzzy inference system for detection of obstructive sleep apnea*. in *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*. 2001.

109. Fu-Chung, Y., K. Behbehani, E.A. Lucas, J.R. Burk, and J.R. Axe, *A noninvasive technique for detecting obstructive and central sleep apnea.* Biomedical Engineering, IEEE Transactions on, 1997. **44**(12): p. 1262-1268.

110. Derong, L., P. Zhongyu, and S.R. Lloyd, *A Neural Network Method for Detection of Obstructive Sleep Apnea and Narcolepsy Based on Pupil Size and EEG.* Neural Networks, IEEE Transactions on, 2008. **19**(2): p. 308-318.

111. Falie, D., L. David, and M. Ichim. *Statistical algorithm for detection and screening sleep apnea*. in *Signals, Circuits and Systems, 2009. ISSCS 2009. International Symposium on*. 2009.

112. Nishida, Y., T. Mori, T. Sato, and S. Hirai. *The surrounding sensor approach - application to sleep apnea syndrome diagnosis based on image processing*. in *Systems, Man, and Cybernetics, 1999. IEEE SMC '99 Conference Proceedings. 1999 IEEE International Conference on*. 1999.

113. Broadway, M., L. Matthews, and M. Kwiatkowska. *A fuzzy logic approach to modeling physical activity levels of obstructive sleep apnea patients*. in *Fuzzy Information Processing Society, 2008. NAFIPS 2008. Annual Meeting of the North American*. 2008.

114. Ching-Wei, W. and A. Hunter. *A robust pose matching algorithm for covered body analysis for sleep apnea*. in *BioInformatics and BioEngineering, 2008. BIBE 2008. 8th IEEE International Conference on*. 2008.

115. Yadollahi, A. and Z. Moussavi. *Apnea detection by acoustical means*. in *Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE*. 2006.

116. Al-Ashmouny, K.M., A.A. Morsy, S.F. Loza, and Ieee, *Sleep apnea detection and classification using fuzzy logic: Clinical evaluation*, in *2005 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vols 1-7*2005. p. 6132-6135.

117. Al-Ashmouny, K.M., H.M. Hamed, A.A. Morsy, and Ieee, *FPGA-based sleep apnea screening device for home monitoring*. 2006 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vols 1-152006. 566-569.

118. Gil, E., J. María Vergara, and P. Laguna, *Detection of decreases in the amplitude fluctuation of pulse photoplethysmography signal as indication of obstructive sleep apnea syndrome in children.* Biomedical Signal Processing and Control, 2008. **3**(3): p. 267-277.

119. Iida, S., M. Kogo, S. Ishii, H. Kohara, and T. Matsuya, *Changes of arterial oxygen saturation (SpO2) following push-back operation.* International Journal of Oral and Maxillofacial Surgery, 1998. **27**(6): p. 425-427.

120. Epstein, L.J.E. and G.R. Dorlac, *Cost-effectiveness analysis of nocturnal oximetry as a method of screening for sleep apnea-hypopnea syndrome.* Chest, 1998. **113**(1): p. 97-103.

121. Yadollahi, A. and Z. Moussavi. *Acoustic obstructive sleep apnea detection*. in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. 2009.

122. Álvarez, D., R. Hornero, M. García, F. del Campo, and C. Zamarrón, *Improving diagnostic ability of blood oxygen saturation from overnight pulse oximetry in obstructive sleep apnea detection by means of central tendency measure.* Artificial Intelligence in Medicine, 2007. **41**(1): p. 13-24.
123. Hornero, R., D. Alvarez, D. Abasolo, C. Gomez, F. del Campo, and C. Zamarron. *Approximate entropy from overnight pulse oximetry for the obstructive sleep apnea syndrome*. in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*. 2005.
124. Alvarez, D., R. Hornero, J.V. Marcos, F. del Campo, and M. Lopez. *Obstructive Sleep Apnea Detection Using Clustering Classification of Nonlinear Features from Nocturnal Oximetry*. in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*. 2007.
125. Victor Marcos, J., R. Hornero, D. Alvarez, F. Del Campo, C. Zamarron, and M. Lopez. *Single layer network classifiers to assist in the detection of obstructive sleep apnea syndrome from oximetry data*. in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*. 2008.
126. Marcos, J.V., R. Hornero, D. Alvarez, F. Del Campo, and C. Zamarron. *A classification algorithm based on spectral features from nocturnal oximetry and support vector machines to assist in the diagnosis of obstructive sleep apnea*. in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. 2009.
127. Alvarez, D., R. Hornero, M. Garcia, F.D. Campo, C. Zamarron, and M. Lopez. *Cross Approximate Entropy Analysis of Nocturnal Oximetry Signals in the Diagnosis of the Obstructive Sleep Apnea Syndrome*. in *Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE*. 2006.
128. Lee, Y.K., M. Bister, P. Blanchfield, and Y.M. Salleh. *Automated detection of obstructive apnea and hypopnea events from oxygen saturation signal*. in *Engineering in Medicine and Biology Society, 2004. IEMBS '04. 26th Annual International Conference of the IEEE*. 2004.
129. Nobuyuki, A., N. Yasuhiro, T. Taiki, Y. Miyae, M. Kiyoko, and H. Terumasa. *Trial of measurement of sleep apnea syndrome with sound monitoring and SpO2 at home*. in *e-Health Networking, Applications and Services, 2009. Healthcom 2009. 11th International Conference on*. 2009.
130. Burgos, A., A. Goni, A. Illarramendi, and J. Bermudez. *SAMON: Sleep apnea monitoring*. in *Bioinformatics and Biomedicine Workshop, 2009. BIBMW 2009. IEEE International Conference on*. 2009.
131. Kaimakamis, E., C. Bratsas, L. Sichletidis, C. Karvounis, and N. Maglaveras. *Screening of patients with obstructive sleep Apnea syndrome using C4.5 algorithm based on non linear analysis of respiratory signals during sleep*. in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. 2009.
132. Otero, A., P. Felix, M.R. Alvarez, and C. Zamarron. *Fuzzy structural algorithms to identify and characterize apnea and hypopnea episodes*. in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*. 2008.
133. Burgos, A., A. Goni, A. Illarramendi, and J. Bemudez, *Real-time detection of apneas on a PDA.* Information Technology in Biomedicine, IEEE Transactions on, 2009. **PP**(99): p. 1-1.
134. Otero, A., C.O.S. Sorzano, P. Felix, M.R. Alvarez, and C. Zamarron. *A fuzzy constraint satisfaction approach to identify and characterize apnea episodes*. in *Bioinformatics and Biomedical Engineering, 2008. ICBBE 2008. The 2nd International Conference on*. 2008.
135. Escola, H., P. Gaillard, M. Jobert, and C. Tismer, *Automatic detection of ocular movements during sleep with application to a sleep-disorders study*, in *Proceedings of the Annual International Conference of the Ieee Engineering in Medicine and Biology Society, Vol 14, Pts 1-7*, J.P. Morucci, R. Plonsey, J.L. Coatrieux, and S. Laxminarayan, Editors. 1992. p. 2738-2739.
136. Findley, L.J., S.C. Wilhoit, and P.M. Suratt, *Apnea duration and hypoxemia during REM-sleep in patients with obstractive sleep-apnea.* Chest, 1985. **87**(4): p. 432-436.
137. Brownell, L.G., P. West, P. Sweatman, J.C. Acres, and M.H. Kryger, *Protriptyline in obstractive sleep-apnea - a double-blind trial.* New England Journal of Medicine, 1982. **307**(17): p. 1037-1042.
138. Smith, P.L., E.F. Haponik, R.P. Allen, and E.R. Bleecker, *The effects of protriptyline in sleep-disordered breathing.* American Review of Respiratory Disease, 1983. **127**(1): p. 8-13.
139. Orr, W.C., R.J. Martin, N.K. Imes, R.M. Rogers, and M.L. Stahl, *Hypersomnolent and non-hypersomnolent patients with upper airway- obstraction during sleep.* Chest, 1979. **75**(4): p. 418-422.
140. Siddiqui, F., A.S. Walters, D. Goldstein, M. Lahey, and H. Desai, *Half of patients with obstructive sleep apnea have a higher NREM AHI than REM AHI.* Sleep Medicine, 2006. **7**(3): p. 281-285.
141. Loadsman, J.A. and I. Wilcox, *Is obstructive sleep apnoea a rapid eye movement-predominant phenomenon?* British Journal of Anaesthesia, 2000. **85**(3): p. 354-358.
142. Guilleminault, C., M.W. Hill, F.B. Simmons, and W.C. Dement, *Obstractive sleep apnea - electromyographic and fiberoptic studies.* Experimental Neurology, 1978. **62**(1): p. 48-67.

143. Fabbri, M., F. Pizza, E. Magosso, M. Ursino, S. Contardi, F. Cirignotta, F. Provini, and P. Montagna, *Automatic slow eye movement (SEM) detection of sleep onset in patients with obstructive sleep apnea syndrome (OSAS): Comparison between multiple sleep latency test (MSLT) and maintenance of wakefulness test (MWT).* Sleep Medicine. **11**(3): p. 253-257.
144. Abdal, H., J.J. Pizzimenti, and C.C. Purvis, *The eye in sleep apnea syndrome.* Sleep Medicine, 2006. **7**(2): p. 107-115.
145. Estrada, E., H. Nazeran, P. Nava, K. Behbehani, J. Burk, E. Lucas, and Ieee, *Itakura Distance: A useful similarity measure between EEG and EOG signals in computer-aided classification of sleep stages*, in *2005 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vols 1-7*2005. p. 1189-1192.
146. Mendelson, Y., *Pulse oximetry-theory and applications for noninvasive monitoring.* Clinical Chemistry, 1992. **38**(9): p. 1601-1607.
147. Nitzan, M., A. Babchenko, B. Khanokh, and D. Landau, *The variability of the photoplethysmographic signal - a potential method for the evaluation of the autonomic nervous system.* Physiological Measurement, 1998. **19**(1): p. 93-102.
148. Somers, V.K., M.E. Dyken, M.P. Clary, and F.M. Abboud, *Sympathetic neural mechanisms in obstractive sleep-apnea.* Journal of Clinical Investigation, 1995. **96**(4): p. 1897-1904.
149. Imadojemu, V.A., K. Gleeson, K.S. Gray, L.I. Sinoway, and U.A. Leuenberger, *Obstructive apnea during sleep is associated with peripheral vasoconstriction.* American Journal of Respiratory and Critical Care Medicine, 2002. **165**(1): p. 61-66.
150. Allen, J., *Photoplethysmography and its application in clinical physiological measurement.* Physiological Measurement, 2007. **28**(3): p. R1-R39.
151. Gil, E., R. Bailon, J.M. Vergara, and P. Laguna, *PTT variability for discrimination of sleep apnea related decreases in the amplitude fluctuations of PPG signal in children.* Ieee Transactions on Biomedical Engineering. **57**(5): p. 1079-1088.
152. Gil, E., M. Mendez, J.M. Vergara, S. Cerutti, A.M. Bianchi, and P. Laguna, *Discrimination of Sleep-Apnea-Related Decreases in the Amplitude Fluctuations of PPG Signal in Children by HRV Analysis.* Ieee Transactions on Biomedical Engineering, 2009. **56**(4): p. 1005-1014.
153. Gil, E., J.M. Vergara, and P. Laguna, *Detection of decreases in the amplitude fluctuation of pulse photoplethysmography signal as indication of obstructive sleep apnea syndrome in children.* Biomedical Signal Processing and Control, 2008. **3**(3): p. 267-277.
154. Corthout, J., S. Van Huffel, M.O. Mendez, A.M. Bianchi, T. Penzel, S. Cerutti, and Ieee, *Automatic screening of Obstructive Sleep Apnea from the ECG based on Empirical Mode Decomposition and Wavelet Analysis*, in *2008 30th Annual International Conference of the Ieee Engineering in Medicine and Biology Society, Vols 1-8*2008. p. 3608-3611.
155. Camm, A.J., M. Malik, J.T. Bigger, G. Breithardt, S. Cerutti, R.J. Cohen, P. Coumel, E.L. Fallen, H.L. Kennedy, R.E. Kleiger, F. Lombardi, A. Malliani, A.J. Moss, J.N. Rottman, G. Schmidt, P.J. Schwartz, and D.H. Singer, *Heart rate variability. Standards of measurement, physiological interpretation, and clinical use.* European Heart Journal, 1996. **17**(3): p. 354-381.
156. Malliani, A., *The pattern of sympathovagal balance explored in the frequency domain.* News in Physiological Sciences, 1999. **14**: p. 111-117.
157. Bonnet, M.H. and D.L. Arand, *Heart rate variability: Sleep stage, time of night, and arousal influences.* Electroencephalography and Clinical Neurophysiology, 1997. **102**(5): p. 390-396.
158. Mendez, M.O., A.M. Bianchi, M. Matteucci, S. Cerutti, and T. Penzel, *Sleep Apnea Screening by Autoregressive Models From a Single ECG Lead.* Ieee Transactions on Biomedical Engineering, 2009. **56**(12): p. 2838-2850.
159. Penzel, T., J.W. Kantelhardt, L. Grote, J.H. Peter, and A. Bunde, *Comparison of detrended fluctuation analysis and spectral analysis for heart rate variability in sleep and sleep apnea.* Ieee Transactions on Biomedical Engineering, 2003. **50**(10): p. 1143-1151.
160. Penzel, T., J.W. Kantelhardt, H.F. Becker, J.H. Peter, and A. Bunde, *Detrended fluctuation analysis and spectral analysis of heart rate variability for sleep stage and sleep apnea identification*, in *Computers in Cardiology 2003, Vol 30*, A. Murray, Editor 2003. p. 307-310.
161. Al-Angari, H.M. and A.V. Sahakian, *Use of sample entropy approach to study heart rate variability in obstructive sleep apnea syndrome.* Ieee Transactions on Biomedical Engineering, 2007. **54**(10): p. 1900-1904.
162. Khandoker, A.H., M. Palaniswami, and C.K. Karmakar, *Support vector machines for automated recognition of obstructive sleep apnea syndrome from ECG recordings.* Ieee Transactions on Information Technology in Biomedicine, 2009. **13**(1): p. 37-48.
163. de Chazal, P., C. Heneghan, E. Sheridan, R. Reilly, P. Nolan, and M. O'Malley, *Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnoea.* Biomedical Engineering, IEEE Transactions on, 2003. **50**(6): p. 686-696.
164. Khushaba, R.N., A. Al-Ani, and A. Al-Jumaily, *Orthogonal fuzzy neighborhood discriminant analysis for multifunction myoelectric hand control.* IEEE Trans Biomed Eng. **57**(6): p. 1410-9.
165. Kurtz, D., J. Krieger, and J.C. Stierle, *EMG activity of cricothyroid and chin muscles during wakefulness and sleeping in sleep apnea syndrome.* Electroencephalography and Clinical Neurophysiology, 1978. **45**(6): p. 777-784.

166. Chua, E.C.P., D.G. McSharry, W.T. McNicholas, and M.M. Lowery. *Towards a genioglossus surface EMG model of obstructive sleep apnea*. in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. 2009.
167. Blumen, M.B., A.P. de La Sota, M.A. Quera-Salva, B. Frachet, F. Chabolle, and F. Lofaso, *Tongue mechanical characteristics and genioglossus muscle EMG in obstructive sleep apnoea patients.* Respiratory Physiology & Neurobiology, 2004. **140**(2): p. 155-164.
168. ASDA, *EEG arousals-scoring rules and examples* Sleep, 1992. **15**(2): p. 174-184.
169. Black, J.E., C. Guilleminault, I.M. Colrain, and O. Carrillo, *Upper airway resistance syndrome - Central electroencephalographic power and changes in breathing effort.* American Journal of Respiratory and Critical Care Medicine, 2000. **162**(2): p. 406-411.
170. Redmond, S.J. and C. Heneghan, *Cardiorespiratory-based sleep staging in subjects with obstructive sleep apnea.* Biomedical Engineering, IEEE Transactions on, 2006. **53**(3): p. 485-496.
171. Sugi, T., F. Kawana, and M. Nakamura, *Automatic EEG arousal detection for sleep apnea syndrome.* Biomedical Signal Processing and Control, 2009. **4**(4): p. 329-337.
172. Schmidtnowara, W., A. Lowe, L. Wiegand, R. Cartwright, F. Perezguerra, and S. Menn, *Oral appliaces for the treatment of snoring and obstractive sleep-apnea - a review.* Sleep, 1995. **18**(6): p. 501-510.
173. Garvey, J.F. and W.T. McNicholas, *Continuous positive airway pressure therapy: New generations.* Indian Journal of Medical Research. **131**(2): p. 259-266.
174. Baptista, P.M., *Surgery for obstructive sleep apnea.* Anales Del Sistema Sanitario De Navarra, 2007. **30**: p. 75-88.
175. Verse, T., A. Baisch, J.T. Maurer, B.A. Stuck, and K. Hormann, *Multilevel surgery for obstructive sleep apnea: Short-term results.* Otolaryngology-Head and Neck Surgery, 2006. **134**(4): p. 571-577.
176. Burstein, F.D., S.R. Cohen, P.H. Scott, G.R. Teague, G.L. Montgomery, and A.V. Kattos, *Surgical therapy for severe refractory sleep-apnea in infants and children- application of the airway zone concept.* Plastic and Reconstructive Surgery, 1995. **96**(1): p. 34-41.
177. Hochban, W., R. Conradt, U. Brandenburg, J. Heitmann, and J.H. Peter, *Surgical maxillofacial treatment of obstructive sleep apnea.* Plastic and Reconstructive Surgery, 1997. **99**(3): p. 619-626.
178. Sullivan, C.E., M. Berthonjones, F.G. Issa, and L. Eves, *Reversal of obstractive sleep-apnea by continuos positive airway pressure applied through the nares.* Lancet, 1981. **1**(8225): p. 862-865.
179. Boisteanu, D., R. Vasiluta, A. Cernomaz, and C. Mucenica, *Home monitoring of sleep apnea treatment: benefits of intelligent CPAP devices*. At-Equal 2009: 2009 Ecsis Symposium on Advanced Technologies for Enhanced Quality of Life: Lab-Rs and Artiped 2009, ed. A. Stoica, T. Arslan, T. Huntsberger, P. Botez, A.T. Erdogan, and A.O. ElRayis2009. 77-80.
180. Hertegonne, K.B., J. Volna, S. Portier, R. De Pauw, G. Van Maele, and D.A. Pevernagie, *Titration procedures for nasal CPAP: Automatic CPAP or prediction formula?* Sleep Medicine, 2008. **9**(7): p. 732-738.
181. Hussain, S.F., L. Love, H. Burt, and J.A. Fleetham, *A randomized trial of auto-titrating CPAP and fixed CPAP in the treatment of obstructive sleep apnea-hypopnea.* Respiratory Medicine, 2004. **98**(4): p. 330-333.
182. Shi, H.B., L. Cheng, M. Nakayama, Y. Kakazu, M. Yin, A. Miyoshi, and S. Komune, *Effective comparison of two auto-CPAP devices for treatment of obstructive sleep apnea based on poly somnographic evaluation.* Auris Nasus Larynx, 2005. **32**(3): p. 237-241.
183. Huang, W.-C., Y.-M. Hua, C.-M. Lee, C.-C. Chang, and Y.-S. Yuh, *Comparison between bubble CPAP and ventilator-derived CPAP in rabbits.* Pediatrics & Neonatology, 2008. **49**(6): p. 223-229.
184. Luo, Y.M., Z.H. Qiu, H.D. Wu, J. Steier, C. Jolley, N.S. Zhong, J. Moxham, and M.I. Polkey, *Neural drive during continuous positive airway pressure (CPAP) and pressure relief CPAP.* Sleep Medicine, 2009. **10**(7): p. 731-738.
185. Mulgrew, A.T., R. Cheema, J. Fleetham, C.F. Ryan, and N.T. Ayas, *Efficacy and patient satisfaction with autoadjusting CPAP with variable expiratory pressure vs standard CPAP: a two-night randomized crossover trial.* Sleep and Breathing, 2007. **11**(1): p. 31-37.
186. Aloia, M.S., M. Stanchina, J.T. Arnedt, A. Malhotra, and R.P. Millman, *Treatment adherence and outcomes in flexible vs standard continuous positive airway pressure therapy.* Chest, 2005. **127**(6): p. 2085-2093.
187. Nilius, G., A. Happel, U. Domanski, and K.H. Ruhle, *Pressure-relief continuous positive airway pressure vs constant continuous positive airway pressure - A comparison of efficacy and compliance.* Chest, 2006. **130**(4): p. 1018-1024.
188. Fleetham, J.A., J.L. Geoffrey, and D.S. Steven, *Sleep apnea |Oral appliances*, in *Encyclopedia of Respiratory Medicine*2006, Academic Press: Oxford. p. 67-70.
189. Cartwright, R., D. Stefoski, D. Caldarelli, H. Kravitz, S. Knight, S. Lloyd, and C. Samelson, *Toward a treatment logic for sleep-apnea- the place of the tongue retaining device.* Behaviour Research and Therapy, 1988. **26**(2): p. 121-126.
190. Soll, B.A. and P.T. George, *Treatment of obstractive sleep-apnea with a nocturnal airway-patency.* New England Journal of Medicine, 1985. **313**(6): p. 386-387.

191. Schmidtnowara, W.W., T.E. Meade, and M.B. Hays, *Treatment of snoring and obstractive sleep-apnea with a dental orthosis.* Chest, 1991. **99**(6): p. 1378-1385.
192. Pancherz, H., *The herbst appliance - its biologic effects and clinical use.* American Journal of Orthodontics and Dentofacial Orthopedics, 1985. **87**(1): p. 1-20.
193. Eveloff, S.E., C.L. Rosenberg, C.C. Carlisle, and R.P. Millman, *Efficacy of a herbst mandibular advancement device in obstractive sleep-apnea.* American Journal of Respiratory and Critical Care Medicine, 1994. **149**(4): p. 905-909.
194. Knudson, R.C., J.B. Meyer, and R. Montalvo, *Sleep-apnea prosthesis for dentate patients.* Journal of Prosthetic Dentistry, 1992. **68**(1): p. 109-111.
195. Knudson, R.C. and J.B. Meyer, *Managing obstractive sleep-apnea.* Journal of the American Dental Association, 1993. **124**(8): p. 75-78.
196. Bonham, P.E., G.F. Currier, W.C. Orr, J. Othman, and R.S. Nanda, *The effect of a modified functional appliance on obstractive sleep-apnea.* American Journal of Orthodontics and Dentofacial Orthopedics, 1988. **94**(5): p. 384-392.
197. Haze, J., *Treatment of obstructive sleep apnea with the equalizer appliance.* J N J Dent Assoc., 1987. **58**(1): p. 34-36.
198. Pang, K.P. and D.J. Terris, *Tongue suspension suture in obstructive sleep apnea.* Operative Techniques in Otolaryngology-Head and Neck Surgery, 2006. **17**(4): p. 252-256.
199. Fujita, S., W. Conway, F. Zorick, and T. Roth, *Surgical-correction of anatomic abnormalities in obstractive sleep-apnea syndrome - UVULOPALATOPHARYNGOPLASTY.* Otolaryngology-Head and Neck Surgery, 1981. **89**(6): p. 923-934.
200. Samir, M., A. Adly, and M. Elshinawy, *Tongue base assessment in obstructive sleep apnea.* International Congress Series, 2003. **1240**: p. 753-758.
201. Miki, H., W. Hida, T. Chonan, Y. Kikuchi, and T. Takishima, *Effects of submental electrical-stimulation during sleep on upper airway patency in patients with obstractive sleep-apnea.* American Review of Respiratory Disease, 1989. **140**(5): p. 1285-1289.
202. Edmonds, L.C., B.K. Daniels, A.W. Stanson, P.F. Sheedy, and J.W. Shepard, *The effect of transcutaneous electrical-stimulation during wakefulness and sleep in patients with obstractive sleep-apnea.* American Review of Respiratory Disease, 1992. **146**(4): p. 1030-1036.
203. Schwartz, A.R., D.W. Eisele, A. Hari, R. Testerman, D. Erickson, and P.L. Smith, *Electrical stimulation of the lingual musculature in obstructive sleep apnea.* Journal of Applied Physiology, 1996. **81**(2): p. 643-652.
204. Mann, E.A., T. Burnett, S. Cornell, and C.L. Ludlow, *The effect of neuromuscular stimulation of the genioglossus on the hypopharyngeal airway.* Laryngoscope, 2002. **112**(2): p. 351-356.
205. Schwartz, A.R., M.L. Bennett, P.L. Smith, W. De Backer, J. Hedner, A. Boudewyns, P. Van de Heyning, H. Ejnell, W. Hochban, L. Knaack, T. Podszus, T. Penzel, J.H. Peter, G.S. Goding, D.J. Erickson, R. Testerman, F. Ottenhoff, and D.W. Eisele, *Therapeutic electrical stimulation of the hypoglossal nerve in obstructive sleep apnea.* Archives of Otolaryngology-Head & Neck Surgery, 2001. **127**(10): p. 1216-1223.
206. Yoo, P.B., D.M. Durand, and Ieee, *A neural prosthesis for obstructive sleep apnea*, in *2005 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vols 1-7*2005. p. 5254-5256.
207. Kezirian, E.J., A. Boudewyns, D.W. Eisele, A.R. Schwartz, P.L. Smith, P.H. Van de Heyning, and W.A. De Backer, *Electrical stimulation of the hypoglossal nerve in the treatment of obstructive sleep apnea.* Sleep Medicine Reviews. **In Press, Corrected Proof**.
208. Vapnik, V.N., *An overview of statistical learning theory.* Ieee Transactions on Neural Networks, 1999. **10**(5): p. 988-999.
209. Lauer, F. and G. Bloch, *Incorporating prior knowledge in support vector machines for classification: A review.* Neurocomputing, 2008. **71**(7-9): p. 1578-1594.
210. Zhan, Y.Q. and D.G. Shen, *An adaptive error penalization method for training an efficient and generalized SVM.* Pattern Recognition, 2006. **39**(3): p. 342-350.
211. Cho, S., J. Lee, H. Park, and K. Lee. *Detection of arousals in patients with respiratory sleep disorders using a single channel EEG*. in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*. 2006. IEEE.
212. Übeyli, E.D., D. Cvetkovic, G. Holland, and I. Cosic, *Analysis of sleep EEG activity during hypopnoea episodes by least squares support vector machine employing AR coefficients.* Expert Systems with Applications, 2010. **37**(6): p. 4463-4467.
213. Khandoker, A.H., C.K. Karmakar, and M. Palaniswami. *Screening obstructive sleep apnoea syndrome from electrocardiogram recordings using support vector machines*. in *Computers in Cardiology, 2007*. 2007.
214. Khandoker, A.H., C.K. Karmakar, and M. Palaniswami, *Automated recognition of patients with obstructive sleep apnoea using wavelet-based features of electrocardiogram recordings.* Computers in Biology and Medicine, 2009. **39**(1): p. 88-96.
215. Yildiz, A., M. Akın, and M. Poyraz, *An expert system for automated recognition of patients with obstructive sleep apnea using electrocardiogram recordings.* Expert Systems with Applications, 2011. **38**(10): p. 12880-12890.
216. Lam, L. and C.Y. Suen, *Optimal combinations of pattern classifiers.* Pattern Recognition Letters, 1995. **16**(9): p. 945-954.

217. Isa, S.M., M.I. Fanany, W. Jatmiko, and A.M. Arymurthy. *Sleep Apnea Detection from ECG Signal: Analysis on Optimal Features, Principal Components, and Nonlinearity*. in *Bioinformatics and Biomedical Engineering, (iCBBE) 2011 5th International Conference on*. 2011.
218. Koley, B.L. and D. Dey. *Selection of features for detection of Obstructive Sleep Apnea events*. in *India Conference (INDICON), 2012 Annual IEEE*. 2012.
219. Koley, B.L. and D. Dey, *Automatic detection of sleep apnea and hypopnea events from single channel measurement of respiration signal employing ensemble binary SVM classifiers.* Measurement, 2013. **46**(7): p. 2082-2092.
220. Kuncheva, L.I., J.C. Bezdek, and R.P. Duin, *Decision templates for multiple classifier fusion: an experimental comparison.* Pattern Recognition, 2001. **34**(2): p. 299-314.
221. Rogova, G., *Combining the results of several neural network classifiers.* Neural Networks, 1994. **7**(5): p. 777-781.
222. Ahmad, Z., R. Mat Noor, and J. Zhang, *Multiple neural networks modeling techniques in process control: a review.* Asia - Pacific Journal of Chemical Engineering, 2009. **4**(4): p. 403-419.
223. Haykin, S., *Neural networks: a comprehensive foundation*1994: Prentice Hall PTR.
224. Jain, A., K. Nandakumar, and A. Ross, *Score normalization in multimodal biometric systems.* Pattern Recognition, 2005. **38**(12): p. 2270-2285.
225. Ahmad, Z. and J. Zhang, *Combination of multiple neural networks using data fusion techniques for enhanced nonlinear process modelling.* Computers & Chemical Engineering, 2005. **30**(2): p. 295-308.
226. Atiya, A.F., S.M. El-Shoura, S.I. Shaheen, and M.S. El-Sherif, *A comparison between neural-network forecasting techniques-case study: river flow forecasting.* Neural Networks, IEEE Transactions on, 1999. **10**(2): p. 402-409.
227. Ahmad, Z. and J. Zhang, *Bayesian selective combination of multiple neural networks for improving long-range predictions in nonlinear process modelling.* Neural Computing & Applications, 2005. **14**(1): p. 78-87.
228. Sridhar, D.V., E.B. Bartlett, and R.C. Seagrave, *An information theoretic approach for combining neural network process models.* Neural Networks, 1999. **12**(6): p. 915-926.
229. Zhang, J., *Improved on-line process fault diagnosis through information fusion in multiple neural networks.* Computers & Chemical Engineering, 2006. **30**(3): p. 558-571.
230. Woods, K., W.P. Kegelmeyer Jr, and K. Bowyer, *Combination of multiple classifiers using local accuracy estimates.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1997. **19**(4): p. 405-410.
231. Cho, S.-B. and J.H. Kim, *Combining multiple neural networks by fuzzy integral for robust classification.* Systems, Man and Cybernetics, IEEE Transactions on, 1995. **25**(2): p. 380-384.
232. Gader, P.D., M.A. Mohamed, and J.M. Keller, *Fusion of handwritten word classifiers.* Pattern Recognition Letters, 1996. **17**(6): p. 577-584.
233. Jacobs, R.A., M.I. Jordan, S.J. Nowlan, and G.E. Hinton, *Adaptive mixtures of local experts.* Neural computation, 1991. **3**(1): p. 79-87.
234. Jacobs, R.A., *Methods for combining experts' probability assessments.* Neural computation, 1995. **7**(5): p. 867-888.
235. Drucker, H., C. Cortes, L.D. Jackel, Y. LeCun, and V. Vapnik, *Boosting and other ensemble methods.* Neural computation, 1994. **6**(6): p. 1289-1301.
236. Filippi, E., M. Costa, and E. Pasero. *Multi-layer perceptron ensembles for increased performance and fault-tolerance in pattern recognition tasks*. in *Neural Networks, 1994. IEEE World Congress on Computational Intelligence., 1994 IEEE International Conference on*. 1994. IEEE.
237. Bishop, C.M., *Neural networks for pattern recognition*1995: Oxford university press.
238. Benediktsson, J.A. and P.H. Swain, *Consensus theoretic classification methods.* Systems, Man and Cybernetics, IEEE Transactions on, 1992. **22**(4): p. 688-704.
239. Battiti, R. and A.M. Colla, *Democracy in neural nets: Voting schemes for classification.* Neural Networks, 1994. **7**(4): p. 691-707.
240. Dasarathy, B.V. and B.V. Sheela, *A composite classifier system design: concepts and methodology.* Proceedings of the IEEE, 1979. **67**(5): p. 708-713.
241. Chiang, C.-C. and H.-C. Fu. *A divide-and-conquer methodology for modular supervised neural network design*. in *Neural Networks, 1994. IEEE World Congress on Computational Intelligence., 1994 IEEE International Conference on*. 1994. IEEE.
242. Nabavi-Kerizi, S.H., M. Abadi, and E. Kabir, *A PSO-based weighting method for linear combination of neural networks.* Computers & Electrical Engineering, 2010. **36**(5): p. 886-894.
243. Kuncheva, L.I., *Combining pattern classifiers: Methods and algorithms (kuncheva, li; 2004)[book review].* Neural Networks, IEEE Transactions on, 2007. **18**(3): p. 964-964.
244. Liu, C.-L., *Classifier combination based on confidence transformation.* Pattern Recognition, 2005. **38**(1): p. 11-28.
245. Verikas, A., A. Lipnickas, K. Malmqvist, M. Bacauskiene, and A. Gelzinis, *Soft combination of neural classifiers: A comparative study.* Pattern Recognition Letters, 1999. **20**(4): p. 429-444.

246. Ueda, N., *Optimal linear combination of neural networks for improving classification performance.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2000. **22**(2): p. 207-215.
247. Khotanzad, A., H. Elragal, and T.-L. Lu, *Combination of artificial neural-network forecasters for prediction of natural gas consumption.* Neural Networks, IEEE Transactions on, 2000. **11**(2): p. 464-473.
248. Chawla, N., T.E. Moore, Jr., K.W. Bowyer, L.O. Hall, C. Springer, and P. Kegelmeyer. *Bagging is a small-data-set phenomenon*. in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. 2001.
249. Domingos, P. *Why Does Bagging Work? A Bayesian Account and its Implications*. in *KDD*. 1997. Citeseer.
250. Friedman, J.H. and P. Hall, *On bagging and nonlinear estimation.* Journal of statistical planning and inference, 2007. **137**(3): p. 669-683.
251. Poggio, T., R. Rifkin, S. Mukherjee, and A. Rakhlin, *Bagging regularizes*, 2002, DTIC Document.
252. Freund, Y. and R.E. Schapire. *Experiments with a new boosting algorithm*. in *ICML*. 1996.
253. Brown, G., *Diversity in neural network ensembles.* 2004.
254. Liu, Y., X. Yao, and T. Higuchi, *Evolutionary ensembles with negative correlation learning.* Evolutionary Computation, IEEE Transactions on, 2000. **4**(4): p. 380-387.
255. Emin Tagluk, M. and N. Sezgin, *A new approach for estimation of obstructive sleep apnea syndrome.* Expert Systems with Applications, 2011. **38**(5): p. 5346-5351.
256. Emin Tagluk, M., M. Akin, and N. Sezgin, *Classification of sleep apnea by using wavelet transform and artificial neural networks.* Expert Systems with Applications, 2010. **37**(2): p. 1600-1607.
257. Tagluk, M.E. and N. Sezgin, *Classification of Sleep Apnea through Sub-band Energy of Abdominal Effort Signal Using Wavelets + Neural Networks.* Journal of Medical Systems, 2010. **34**(6): p. 1111-1119.
258. Marcos, J.V., R. Hornero, D. Alvarez, M. Aboy, and F. Del Campo, *Automated prediction of the apnea-hypopnea index from nocturnal oximetry recordings.* Biomedical Engineering, IEEE Transactions on, 2012. **59**(1): p. 141-149.
259. Morillo, D.S. and N. Gross, *Probabilistic neural network approach for the detection of SAHS from overnight pulse oximetry.* Medical & biological engineering & computing, 2013: p. 1-11.
260. Güneş, S., K. Polat, and Ş. Yosunkaya, *Multi-class f-score feature selection approach to classification of obstructive sleep apnea syndrome.* Expert Systems with Applications, 2010. **37**(2): p. 998-1004.
261. Chen, L.-F., C.-T. Su, K.-H. Chen, and P.-C. Wang, *Particle swarm optimization for feature selection with application in obstructive sleep apnea diagnosis.* Neural Computing and Applications, 2012. **21**(8): p. 2087-2096.
262. Bock, J. and D.A. Gough, *Toward prediction of physiological state signals in sleep apnea.* Biomedical Engineering, IEEE Transactions on, 1998. **45**(11): p. 1332-1341.
263. Elman, J.L., *Distributed representations, simple recurrent networks, and grammatical structure.* Machine learning, 1991. **7**(2-3): p. 195-225.
264. Waxman, J.A., D. Graupe, and D.W. Carley, *Automated prediction of apnea and hypopnea, using a LAMSTAR artificial neural network.* American Journal of Respiratory and Critical Care Medicine, 2010. **181**(7): p. 727-733.
265. Kennedy, J. and R. Eberhart, *Particle swarm optimization*. In Proceedings of the IEEE International Conference on Neural Networks., ed. I.P.o.t.I.I.C.o.N. Networks. Vol. 4. 1995, Perth.
266. Rana, S., S. Jasola, and R. Kumar, *A review on particle swarm optimization algorithms and their applications to data clustering.* Artificial Intelligence Review, 2011. **35**(3): p. 211-222.
267. Thangaraj, R., M. Pant, A. Abraham, and P. Bouvry, *Particle swarm optimization: Hybridization perspectives and experimental illustrations.* Applied Mathematics and Computation, 2011. **217**(12): p. 5208-5226.
268. Shi, Y.H., R. Eberhart, and Ieee, *A modified particle swarm optimizer*. 1998 Ieee International Conference on Evolutionary Computation - Proceedings1998. 69-73.
269. Shi, Y. and R.C. Eberhart, *Parameter Selection in Particle Swarm Optimization*, in *Proceedings of the 7th International Conference on Evolutionary Programming VII*1998, Springer-Verlag.
270. Cui, S.M. and D.S. Weile, *Application of a parallel particle swarm optimization scheme to the design of electomagnetic absorbers.* Ieee Transactions on Antennas and Propagation, 2005. **53**(11): p. 3616-3624.
271. Herrera, F., M. Lozano, and C. Moraga, *Hierarchical distributed genetic algorithms.* International Journal of Intelligent Systems, 1999. **14**(11): p. 1099-1121.
272. Madhuri and K. Deep. *A state-of-the-art review of population-based parallel meta-heuristics*. in *Nature & Biologically Inspired Computing, 2009. NaBIC 2009. World Congress on*. 2009.
273. Kalivarapu, V., J.-L. Foo, and E. Winer, *Synchronous parallelization of Particle Swarm Optimization with digital pheromones.* Adv. Eng. Softw., 2009. **40**(10): p. 975-985.
274. Chang, J.F., S.C. Chu, J.F. Roddick, and J.S. Pan, *A parallel particle swarm optimization algorithm with communication strategies.* Journal of Information Science and Engineering, 2005. **21**(4): p. 809-818.

275. Waintraub, M., R. Schirru, and C. Pereira, *Multiprocessor modeling of parallel Particle Swarm Optimization applied to nuclear engineering problems.* Progress in Nuclear Energy, 2009. **51**(6-7): p. 680-688.
276. Zhang, J. and X. Ding, *A Multi-Swarm Self-Adaptive and Cooperative Particle Swarm Optimization.* Engineering Applications of Artificial Intelligence, 2011. **24**(6): p. 958-967.
277. Waintraub, M., R. Schirru, and C.M.N.A. Pereira, *Multiprocessor modeling of parallel Particle Swarm Optimization applied to nuclear engineering problems.* Progress in Nuclear Energy, 2009. **51**(6–7): p. 680-688.
278. Koh, B.I., A.D. George, R.T. Haftka, and B.J. Fregly, *Parallel asynchronous particle swarm optimization.* International Journal for Numerical Methods in Engineering, 2006. **67**(4): p. 578-595.
279. Sokolova, M., N. Japkowicz, and S. Szpakowicz, *Beyond accuracy, F-Score and ROC: A family of discriminant measures for performance evaluation*, in *AI 2006: Advances in Artificial Intelligence, Proceedings*, A. Sattar and B.H. Kang, Editors. 2006. p. 1015-1021.
280. Goutte, C. and E. Gaussier, *A probabilistic interpretation of precision, recall and F-score, with implication for evaluation*, in *Advances in Information Retrieval*, D.E. Losada and J.M. FernandezLuna, Editors. 2005. p. 345-359.
281. Alvarez-Estevez, D. and V. Moret-Bonillo, *Fuzzy reasoning used to detect apneic events in the sleep apnea-hypopnea syndrome.* Expert Systems with Applications, 2009. **36**(4): p. 7778-7785.
282. Hernandez-Pereira, E., B. Fernandez-Rey, M. Cabrero-Canosa, and V. Moret-Bonillo, *An amplitude signal based technique for hypopneas detection*. Procedings of the 11th Iasted International Conference on Artificial Intelligence and Soft Computing, ed. A.P. DelPobil2007. 127-130.
283. Park, S.-G., H.-J. Sim, H.-J. Lee, and J.-E. Oh, *Application of non-stationary signal characteristics using wavelet packet transformation.* Journal of Mechanical Science and Technology, 2008. **22**(11): p. 2122-2133.
284. Kiatpanichagij, K. and N. Afzulpurkar, *Use of supervised discretization with PCA in wavelet packet transformation-based surface electromyrogram classification.* Biomedical Signal Processing and Control, 2009. **4**(2): p. 127-138.
285. Ng, A.K., K.Y. Wong, C.H. Tan, T.S. Koh, and Ieee, *Bispectral analysis of snore signals for obstructive sleep apnea detection*, in *2007 Annual International Conference of the Ieee Engineering in Medicine and Biology Society, Vols 1-16*2007. p. 6196-6199.
286. Ebrahimi, F., M. Mikaeili, E. Estrada, H. Nazeran, and Ieee, *Automatic Sleep Stage Classification Based on EEG Signals by Using Neural Networks and Wavelet Packet Coefficients*, in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vols 1-8*2008. p. 1151-1154.
287. Kempfner, J., G.L. Sorensen, H.B.D. Sorensen, P. Jennum, and Ieee, *Automatic REM Sleep Detection Associated with Idiopathic REM Sleep Behavior Disorder*. 2011 Annual International Conference of the Ieee Engineering in Medicine and Biology Society2011. 6063-6066.
288. Gubbi, J., A. Khandoker, and M. Palaniswami, *Classification of sleep apnea types using wavelet packet analysis of short-term ECG signals.* Journal of Clinical Monitoring and Computing, 2012. **26**(1): p. 1-11.
289. Maali, Y. and A. Al-Jumaily, *Self-advising support vector machine.* Knowledge-Based Systems, 2013. **52**: p. 214-222.
290. Maali, Y. and A. Al-Jumaily. *Hierarchical parallel PSO-SVM based subject-independent sleep apnea classification*. in *Neural Information Processing*. 2012. Springer.
291. Maali, Y. and A. Al-Jumaily. *A novel partially connected cooperative parallel PSO-SVM algorithm: Study based on sleep apnea detection*. in *Evolutionary Computation (CEC), 2012 IEEE Congress on*. 2012. IEEE.
292. Maali, Y., A. Al-Jumaily, and L. Laks, *Self-Advising SVM for Sleep Apnea Classification.*
293. Javed, I., M.N. Ayyaz, and W. Mehmood. *Efficient Training Data Reduction for SVM based Handwritten Digits Recognition*. in *Electrical Engineering, 2007. ICEE '07. International Conference on*. 2007.
294. Wang, J., P. Neskovic, and L. Cooper, *Training Data Selection for Support Vector Machines*, in *Advances in Natural Computation*, L. Wang, K. Chen, and Y. Ong, Editors. 2005, Springer Berlin Heidelberg. p. 554-564.
295. Cano, J.R., F. Herrera, and M. Lozano, *On the combination of evolutionary algorithms and stratified strategies for training set selection in data mining.* Applied Soft Computing, 2006. **6**(3): p. 323-332.
296. Maali, Y. and A. Al-Jumaily, *Signal selection for sleep apnea classification*, in *AI 2012: Advances in Artificial Intelligence*2012, Springer. p. 661-671.
297. Health, N.I.o. *What To Expect During a Sleep Study*. Available from: https://www.nhlbi.nih.gov/health/health-topics/topics/slpst/during.html#.
298. Osrmedical. *Polysomnography*. Available from: http://www.osrmedical.com/en/division-diagnostic/polysomnographie/.
299. Chang, C.-C. and C.-J. Lin, *LIBSVM: A library for support vector machines.* ACM Trans. Intell. Syst. Technol., 2011. **2**(3): p. 1-27.
300. Maali, Y. and A. Al-Jumaily, *Genetic Fuzzy Approach based Sleep Apnea/Hypopnea Detection.*
301. Fontenla-Romero, O., B. Guijarro-Berdinas, A. Alonso- Betanzos, and V. Moret-Bonillo, *A new method for steep apnea classification using wavelets and feedforward neural networks.* Artificial Intelligence in Medicine, 2005. **34**(1): p. 65-76.

302. Moller, M.F., *A SCALED CONJUGATE-GRADIENT ALGORITHM FOR FAST SUPERVISED LEARNING.* Neural Networks, 1993. **6**(4): p. 525-533.
303. Dagum, P. and A. Galper, *Time-series prediction using belief network models.* International Journal of Human-Computer Studies, 1995. **42**(6): p. 617-632.
304. Bock, J. and D.A. Gough, *Toward prediction of physiological state signals in sleep apnea.* Ieee Transactions on Biomedical Engineering, 1998. **45**(11): p. 1332-1341.
305. Elman, J.L., *Distribute representations, simple recurrent networks, and gramimatical structure.* Machine Learning, 1991. **7**(2-3): p. 195-225.
306. Werbos, P.J., *Backpropagation through time: what it does and how to do it.* Proceedings of the IEEE, 1990. **78**(10): p. 1550-1560.
307. Brown, R., P. Bryant, and H.D.I. Abarbanel, *Computing the lyapunov spectrum of a dynamic system from an observed time-series.* Physical Review A, 1991. **43**(6): p. 2787-2806.
308. Skinner, J.E., C. Carpeggiani, C.E. Landisman, and K.W. Fulton, *Correlation dimension of hertbeat intervals is reduced in conscious pigs by myocardial-ischemia.* Circulation Research, 1991. **68**(4): p. 966-976.
309. Waxman, J.A., D. Graupe, and D.W. Carley, *Automated Prediction of Apnea and Hypopnea, Using a LAMSTAR Artificial Neural Network.* American Journal of Respiratory and Critical Care Medicine. **181**(7): p. 727-733.
310. Graupe, D. and H. Kordylewski, *A large memory storage and retrieval neural network for adaptive retrieval and diagnosis.* International Journal of Software Engineering and Knowledge Engineering, 1998. **8**(1): p. 115-138.
311. Hoffstein, V. and S. Mateika, *Predicting nasal continous positive airway pressure.* American Journal of Respiratory and Critical Care Medicine, 1994. **150**(2): p. 486-488.
312. Demarco, P. and C.A. Tassinari, *Extreme somatosensory evoked-potential(ESEP) - EEG sign forcasting possible occurrence of seizures in children.* Electroencephalography and Clinical Neurophysiology, 1977. **43**(4): p. 560-561.
313. Arslan, Y.Z., M.A. Adli, A. Akan, and M.B. Baslo, *Prediction of externally applied forces to human hands using frequency content of surface EMG signals.* Computer Methods and Programs in Biomedicine. **98**(1): p. 36-44.
314. Song, D., P. Hendrickson, V.Z. Marmarelis, J. Aguayo, J.P. He, G.E. Loeb, T.W. Berger, and Ieee, *Predicting EMG with generalized volterra kernel model*, in *2008 30th Annual International Conference of the Ieee Engineering in Medicine and Biology Society, Vols 1-8*2008. p. 201-204.
315. Prechelt, L., *Automatic early stopping using cross validation: quantifying the criteria.* Neural Networks, 1998. **11**(4): p. 761-767.
316. Wang, R.H., *AdaBoost for Feature Selection, Classification and Its Relation with SVM*, A Review*, in *International Conference on Solid State Devices and Materials Science*, G. Lee, Editor 2012. p. 800-807.
317. Maali, Y. and A. Al-Jumaily, *Self-advising support vector machine.* Knowledge-Based Systems, 2013. **52**(0): p. 214-222.
318. Frank, A. and A. Asuncion, *UCI Machine Learning Repository*.
319. Kre\, U.H.-G., \#223, and el, *Pairwise classification and support vector machines*, in *Advances in kernel methods*1999, MIT Press. p. 255-268.
320. Dietterich, T.G. and G. Bakiri, *Solving multiclass learning problems via error-correcting output codes.* J. Artif. Int. Res., 1995. **2**(1): p. 263-286.
321. Chang, C.C. and C.J. Lin, *LIBSVM: A library for support vector machines* 2001. p. Available from http:||www.csie.ntu.edu.tw/~cjlin/libsvm.
322. Gardner, M., A. McNabb, and K. Seppi, *A speculative approach to parallelization in particle swarm optimization.* Swarm Intelligence, 2012. **6**(2): p. 77-116.
323. Angeline, P., *Evolutionary optimization versus particle swarm optimization: Philosophy and performance differences*, in *Evolutionary Programming VII*, V.W. Porto, N. Saravanan, D. Waagen, and A.E. Eiben, Editors. 1998, Springer Berlin Heidelberg. p. 601-610.
324. Niu, B., Y.L. Zhu, X.X. He, and H. Wu, *MCPSO: A multi-swarm cooperative particle swarm optimizer.* Applied Mathematics and Computation, 2007. **185**(2): p. 1050-1062.
325. Eberhart, R.C. and S. Yuhui. *Particle swarm optimization: developments, applications and resources*. in *Evolutionary Computation, 2001. Proceedings of the 2001 Congress on*. 2001.
326. Zito, D., D. Pepe, M. Mincica, F. Zito, A. Tognetti, A. Lanata, and D. De-Rossi, *SoC CMOS UWB Pulse Radar Sensor for Contactless Respiratory Rate Monitoring.* Biomedical Circuits and Systems, IEEE Transactions on, 2011. **5**(6): p. 503-510.
327. Lai, J.C.Y., X. Ying, E. Gunawan, E.C. Chua, A. Maskooki, G. Yong Liang, L. Kay-Soon, S. Cheong Boon, and P. Chueh-Loo, *Wireless Sensing of Human Respiratory Parameters by Low-Power Ultrawideband Impulse Radio Radar.* Instrumentation and Measurement, IEEE Transactions on, 2011. **60**(3): p. 928-938.
328. Changzhi, L., J. Cummings, J. Lam, E. Graves, and W. Wenhsing, *Radar remote monitoring of vital signs.* Microwave Magazine, IEEE, 2009. **10**(1): p. 47-56.
329. Lai, Y.J., T.Y. Liu, and C.L. Hwang, *TOPSIS FOR MODM.* European Journal of Operational Research, 1994. **76**(3): p. 486-500.