UTS UNIVERSITY OF TECHNOLOGY SYDNEY

[Sleep Apnea Detection Using Multi-Bio Signals and Machine Learning]

by [Xilin Li]

Thesis submitted in fulfilment of the requirements for the degree of

[Doctor of Philosophy]

under the supervision of [Dr Ling, Steve Sai Ho]

University of Technology Sydney
Faculty of [Engineering and Information Technology]

[August 2020]

Declaration

I, Xilin Li declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

*If applicable, the above statement must be replaced with the collaborative doctoral degree statement (see below).

*If applicable, the Indigenous Cultural and Intellectual Property (ICIP) statement must be added (see below).

This research is supported by the Australian Government Research Training Program.

Signature: Production Note: Signature removed prior to publication.

Date: 02/08/2021

Acknowledgements

I want to express my sincere gratitude to my Principal Supervisor Dr Ling, Steve Sai Ho, for his continual support, guidance, help, and encouragement during my PhD period. Dr Ling gives me solid support in the topic of Sleep Apnea Detection Using Multi-Bio Signals and Machine Learning, and always provides brilliant insights into my research works. It is my honour to have such a supervisor who always inspires me to achieve higher targets. His conscientious and meticulous attitude on research has a significant influence on my work.

I want to express my sincere gratitude to my Co-supervisor Associate Professor Su, (Steven) Steven Weidong, for their precious comments and suggestions on my research. I also would like to demonstrate my sincere gratitude to my External Supervisor Al-Ani, (Ahmed) Ahmed. He has brought me into the topic of Sleep Apnea Detection and given me his kind support in the first year of my PhD degree.

Special thanks my friends in the CHT research group, in particular, Lin Ye, Wenhui Chen, Haimin Zhang, Zhiyuan Shi, Zhichao Sheng, Kairui Guo, Hanjie Wu, Lingmeng Li for their selfless help. Working together with them will be a good memory.

My deepest gratitude goes to my parents and my girlfriend for their immeasurable support and encouragement throughout my graduate studies.

Abstract

One of the most common types of breathing diseases is obstructive sleep apnea (OSA). OSA affects about 2% - 5% of the total human population. Polysomnography (PSG) is considered the gold standard and records multichannel bio-signals throughout one entire night, including electrocardiogram (ECG), electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG), airflow (AF), abdominal and thoracic efforts (AB and TH), body position, snore, oxygen saturation (SaO_2). Patients may not be able to sleep well because an OSA test is always performed in a sleep laboratory or a hospital, and electrodes are placed on the patient's skin. On the other hand, manual detection of OSA is time-consuming and costly because sleep experts need to monitor and review the overnight PSG signals. To solve these problems, an automatic OSA detection system with high accuracy is desirable.

In this study, the main objective is to construct an OSA monitoring system with high accuracy that is based on multi-domain features from multiple bio-signals using machine learning methods. To achieve this main objective, there are four main contributions presented as follows: 1) feature extraction, 2) feature selection, 3) classification for OSA, and 4) the performance of this monitoring system with the time-window method. Firstly, time-domain, frequency-domain, and non-linear analysis algorithms are used to extract features from multiple bio-signals, such as SaO_2, AF, AB, TH, and ECG recordings. The bio-signals have different bio-patterns between apnea and

normal events, and the features are also related to the apneic bio-physiological patterns. Secondly, a hybrid two-stage feature selection is proposed to select the significant features. Statistical analyses (the first stage) are used to determine independent and significant features, and selected features are put into the feature subset. In the second stage, machine learning methods are utilised to confirm the final feature subset. Thirdly, a multi-errors-reduction (MER) classification system is constructed to classify apnea or normal episodes. Considering the different machine learning algorithms, the structure of the MER system is the stacking method, and it consists of weak learners (boosting methods) and a meta-learner (an artificial neural network (ANN)). Fourthly, the proposed monitoring system is able to provide good performance in the 60-second time-window segmentation method, which means that this monitoring system has the potential for implementation using real data.

In this thesis, the frequently used database was the Sleep Heart Health Study (SHHS) database. This dataset provides over 1,500 patients' PSG signals and each respiratory event and its duration is labelled by a human expert, which is able to evaluate the stability of the proposed monitoring system. According to the duration of each event, multi-domain feature extraction methods are used to provide 66 kinds of features from ECG, SaO_2, AF, TH, and AB signals. The two-stage feature selection procedure then selects 19 kinds of significant features and determines the final feature subset. These 19 features are extracted from ECG, SaO_2, and AB signals. Compared with different machine learning methods, the support vector machine (SVM) method has better performance than other classifiers (66.54% specificity, 97.05% sensitivity, and 81.68% accuracy). It has been found that the SVM method show poor performance when it predicts normal events in less than 60 seconds. Thus, in order to enhance classification results, the 60-second time-window segmentation method is used to improve classification performance and evaluate the potential for implementation

with real data. Using this time-window method, 48 selected kinds of features, which were extracted from five kinds of bio-signals, obtained better performance (sensitivity = 91.94%, specificity = 89.00%, accuracy = 90.71%). The MER classification system is developed to enhance performance. The structure of the MER system is the stacking method. To construct the MER system, four boosting methods (Gradient Boosting, CatBoost, Light GBM, and XGBoost) are base learners, and the meta-learner is the ANN with one input layer, one hidden layer with 4 knots, and one output layer. This MER classification system is able to hold better performance (sensitivity = 96.37%, specificity = 90.83%, accuracy = 94.66%) compared to existing studies.

Publications

The contents of this thesis are based on the following papers that have been published, accepted, or submitted to peer-reviewed journals and conferences.

Journal Papers:

1. X. Li, S.H. Ling, and S. W. Su, "A Hybrid Feature Selection and Extraction Methods for Sleep Apnea Detection Using Bio-Signals," Sensors, vol. 20, no. 20, p. 4323, 2020. [IF:3.275, Q1 in Instruments & Instrumentation, ranked as 15th out of 64].

2. X. Li, S.H. Ling, and S. W. Su, "Sleep apnea classification using ensemble algorithms and neural network with multi-bio signals from polysomnography ," Artificial Intelligence in Medicine. [IF:4.383, ISI: Q1 in Medical Informatics, ranked as 5th out of 26] (Under Review).

Conference Papers:

1. X. Li, A. Al-Ani and S.H. Ling, "Feature selection for the detection of sleep apnea using multi-bio signals from overnight polysomnography," in Proc. 40th Annual International conference of the IEEE Engineering in Medicine and Biology Society, July, Honolulu, US, 2018, pp.1444-1447.

Table of Contents

List of Figures

List of Tables

Nomenclature

Symbols

λ_{feature}	the importance of a feature
d	the result of statistical analyses
p-value	the result of statistical analyses
α	alpha wave
β	beta wave
δ	delta wave
N	the importance of a feature
θ	theta wave

Acronyms / Abbreviations

σ	Width of the Radial basis function kernel
d	Degree of the Polynomial kernel
FN	False Negative
FP	False Positive

N Condition Negative

P Condition Positive

R Regularization parameter of the support vector machine method

TN True Negative

TP True Positive

AB Abdominal

AF Airflow

AHI Apnea-Hypopnea Index

ANN Artificial Neural Network

ANOVA Analysis of Variance

AUC Area Under The ROC Curve

ECG Electrocardiogram

EDR ECG-Derived Respiration

EEG Electroencephalogram

EMG Electromyogram

EOG Electrooculogram

HRV Heart Rate Variability

kNN k-nearest neighbour algorithm

Light GBM Light Gradient Boosting

MER Multi-Errors-Reduction

OSA Obstructive Sleep Apnea

PCA Principal Component Analysis

PSD Power Spectral Density

PSG Polysomnography

RBF Radial Basis Function kernel

SaO_2 Oxygen Saturation

Sen Sensitivity

SHHS Sleep Heart Health Study

Spe specificity

SVM Support Vector Machine

TH Thoracic

WSD Wavelet Spectral Density

Chapter 1

Introduction

Obstructive sleep apnea (OSA) is the most common type of breathing disease that affects sleep. OSA leads to many kinds of symptoms. Daytime sleepiness is the most common presenting complaint [2]. Untreated OSA is correlated to long-term health consequences, including snoring, sleep talking, asthma, fatigue, falling asleep, morning headache, memory loss, weight gain, and limited attention span. These symptoms affect the quality of life and significantly increases health problems, such as cardiovascular disease, sudden death, irritability, hypertension, depression, and learning difficulties [3]. Thus, detection becomes the number one treatment goal.

It is a time-consuming and costly task to diagnose OSA manually. Hence, new and simplified methods and techniques for automatic detection are desirable. This thesis introduces novel methodologies for the OSA detection and constructs an automatic novel OSA detection system with good performance. This chapter begins with the background study of this research in Section 1.1. Thesis motivation, objectives, contributions, and structure of this thesis are discussed in Section 1.2, 1.3, and 1.4 respectively.

1.1 Background

OSA is the repeated temporary closure of the upper airways during sleep (Fig. 1.1). The duration of these apnea events is variable; however, to be clinically significant, the duration must last more than 10 seconds, although they typically do not exceed 120 seconds. The common duration of apnea events is between 20 to 40 seconds [4], which usually decreases blood oxyhemoglobin saturation, which causes arousal since the closed upper airways suddenly open. Due to the frequent arousals during sleep, the sleep architecture of people with OSA is fragmented, which severely affects the refreshing effects of sleep [5]. OSA affects between 3% and 7% of the adult population [6]. The prevalence of OSA is about 3% in children [7], 17% in men and 9% in women aged 50-70 years old [8]. In fact, many patients with OSA are undiagnosed, which may be attributed to a lack of awareness. In addition, primary physicians may not diagnose early OSA if people do not show classic OSA characteristics, such as subjective sleepiness and a high body mass index. There may be substantial sleep-disordered breathing in patients without subjective sleepiness and up to half of OSA patients are not obese [8], [9]. Given the health consequences of OSA, the shortcomings of existing diagnostic approaches, and the economic burden of sleep-disordered breathing, there is a pressing need to develop new diagnosis methods and it has been recommended that OSA should be objectively evaluated. The severity of apnea and hypopnea is measured by the number of events per hour, called the apnea-hypopnea index (AHI). The AHI presents the number of breathing cessations of the upper airway (apnea) and narrowing (hypopnea) during sleep lasting over 10 seconds. Regarding the severity cut-offs, the AHI of healthy people should be lower than 5 episodes per hour, mild, moderate, and severe apnea cases tend to have 5-15, 15-30, and 30 or more events per hour, respectively.

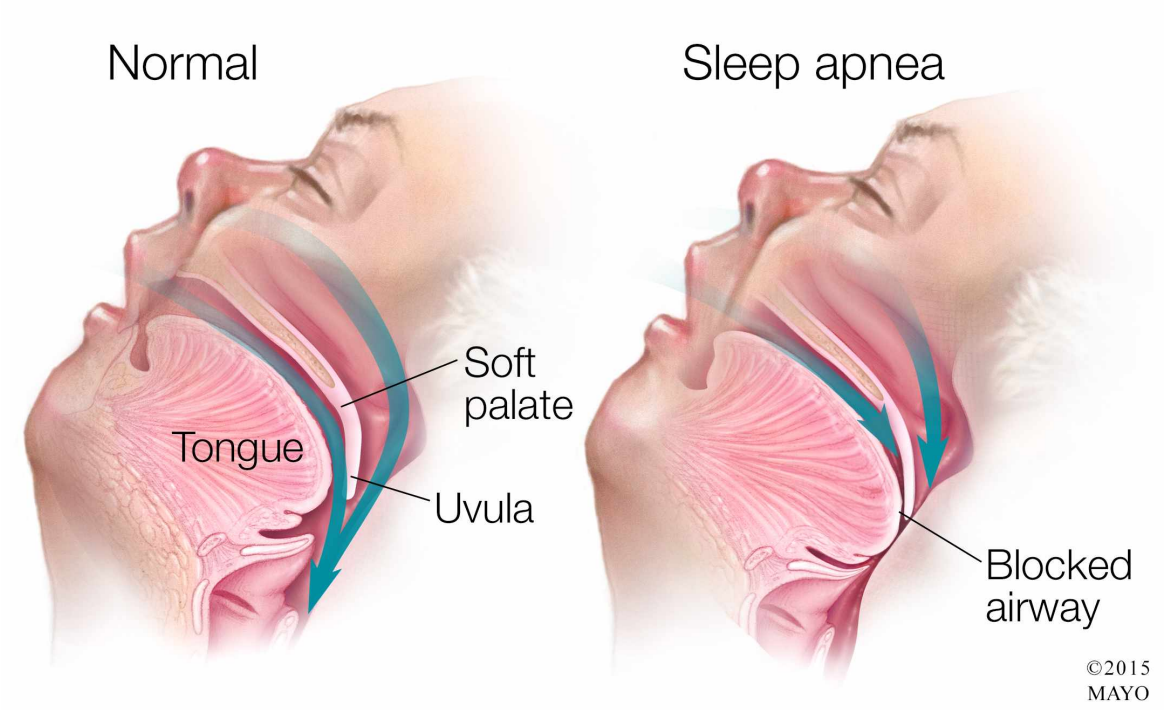

Fig. 1.1 The patient's airway gets blocked because of obstructive sleep apnea (Source: mayoclinic.org)

Polysomnography (PSG) is considered the gold standard for OSA diagnosis (Fig. 1.2). Multichannel bio-signals monitor many body functions and perform a comprehensive recording of bio-physiological changes throughout one night, including electrocardiogram (ECG), electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG), airflow (AF), abdominal and thoracic efforts (AB and TH), body position, snore, oxygen saturation (SaO_2) signals. Among PSG, there are three kinds of bio-signals collected:

1. Bioelectrical potentials caused by the body's own tissues, including ECG (heart rate), EOG (eye movements), EEG (brain waves), and EMG (muscle activity).
2. Waves received from different sensors that translate non-electrical physiological activity into electrical signals:
   (a) Thermistors measure temperature changes.

(b) Inductance respiratory bands record thorax/abdomen effort.

(c) Leg sensors record leg kicks and jerks.

(d) Position sensor measures the physical positioning of the subject.

3. Information from specialised devices:

(a) Oximetry measures oxyhemoglobin saturation.

(b) Nasal pressure (or cannula flow) measures pressure changes during inhalation and exhalation.

OSA is characterised by recurrent collapse of the upper airway during sleep, and the closure results in substantially reduced (hypopnea events) or complete cessation (apnea events) of airflow. It leads to changes in different kinds of muscles, including neck and pharyngeal muscles, respiratory muscles, and muscle changes contribute to the responses in bioelectrical potentials. These bioelectrical potentials of multi-bio signals are always used to diagnose OSA by sleep experts. The remainder of this chapter will introduce some bio-signals with significant bioelectrical potentials.

In general, sleep respiratory disturbances are related to desaturation. SaO_2 signals uses electrodes (e.g. pulse oximetry (see in Fig. 1.3) on the body to record the percentage of oxygen bound to haemoglobin in the blood). Pulse oximetry is a non-invasive and painless test and can measure the changes of a subject's oxygen saturation level or the oxygen levels in the subject's blood. When oxygen is carried efficiently to the extremities furthest from the heart (i.e. legs or arms), there are some changes in the subject's blood and pulse oximetry can detect these changes rapidly. The pulse oximeter is a small, clip-like device that most commonly attaches to a finger. It also can be placed on another body part, such as a toe or an earlobe. To measure the level of oxygen in the blood, during a pulse oximetry reading, small beams of light pass

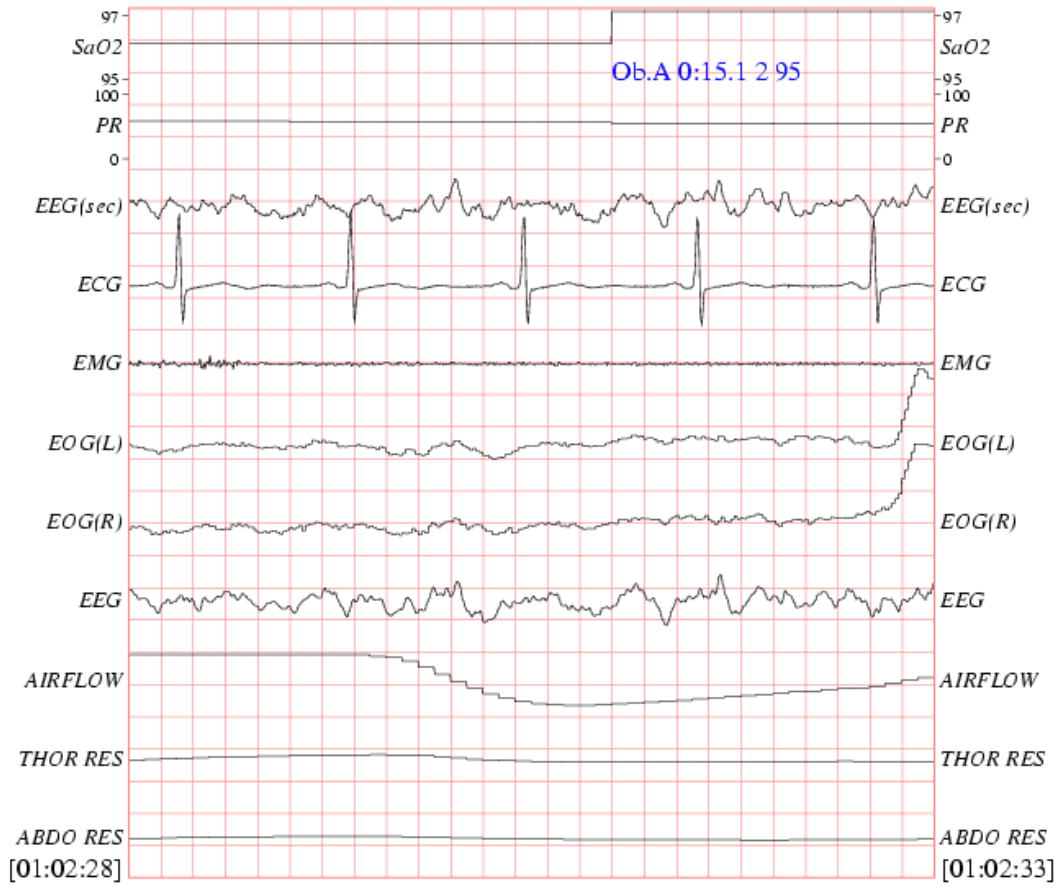

Fig. 1.2 Example of Polysomnography, including EEG, EOG, EMG, abdominal, thoracic, airflow, SaO_2, heart rate, ECG signals (Source: physionet.org)

through the toe, earlobe, or finger. Then, the pulse oximetry obtains the oxygen levels in the subject's blood by the changes of light absorption in oxygenated or deoxygenated blood. An apnea is defined as a complete cessation of breathing lasting more than 10 seconds in adults and 3 to 4 seconds in children and this hypoventilation leads to a drop in the SaO_2. The standard polysomnographic criteria considers the minimum oxygen saturation in adults to be less than 85% and less than 92% in children. The drop usually starts from 10 to 30 seconds after the start of apnea detected from other signals. After the apnea ends, the SaO_2 begins to recover. Fig. 1.4 shows an example of a 300-second session in the SaO_2 signal and indicates the apnea events that happen

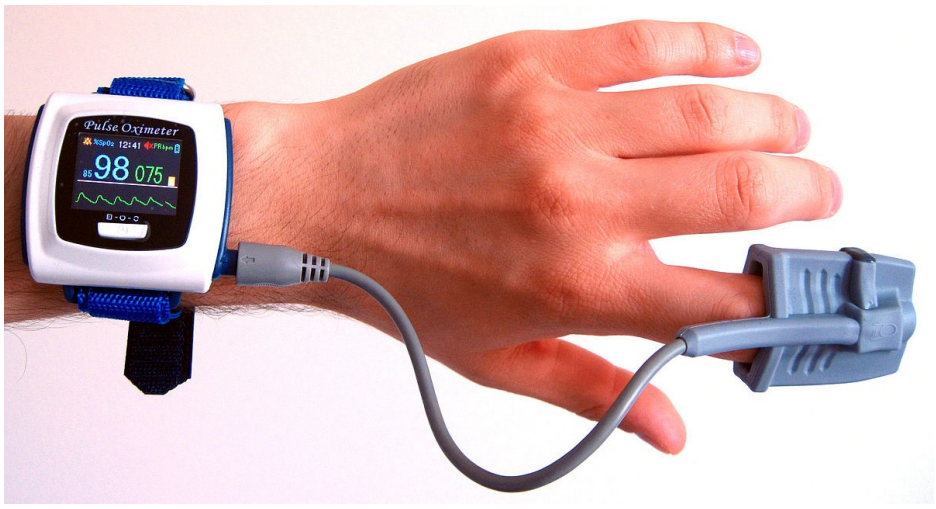

Fig. 1.3 Example of a pulse oximetry attaches to a finger (Source: theprint.in)

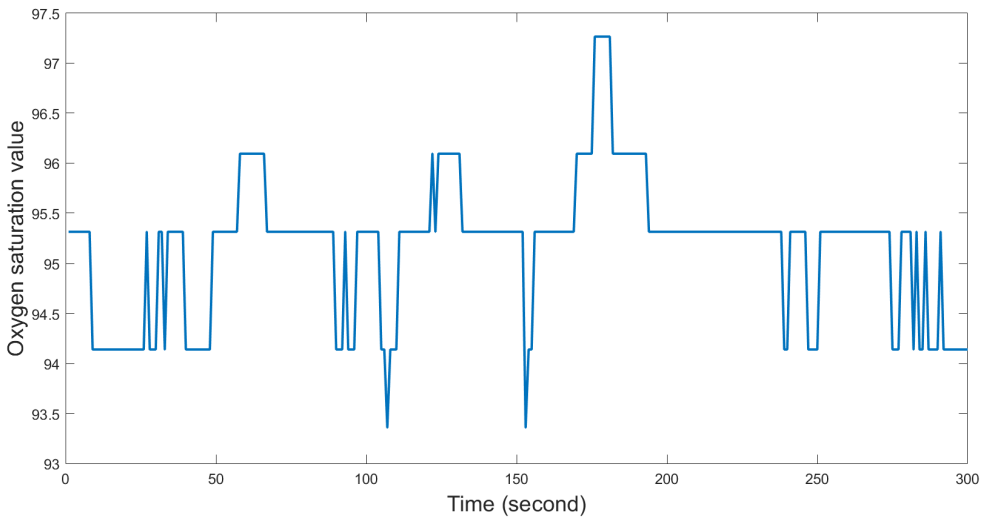

Fig. 1.4 Example of 300-second session shows sudden drops, fast recoveries, and typical saw-tooth morphology of the oximetry curve

in a pseudo periodic pattern. These apnea events construct typical SaO_2 signals and also lead to a modulation frequency peak at the very-low-frequency band. On the other hand, sudden drops in oximetry values followed by fast recoveries are characteristic of apnea, and patients with OSA show an oximetry curve with a typical saw-tooth morphology.

Airflow signal is also considered as a desaturation bio-signal correlated with OSA. The respiratory airflow signal is collected by an electrode (Fig. 1.5) placed under a patient's nose. It can record pressure changes during inhalation and exhalation. Nasal pressure is often placed between the subject's nose and upper lip and the prongs fit into the nares. As for all individuals, there is increased upper airway resistance when upper airway patency decreases during sleep. During an inspiratory phase, the upper airway is most likely to collapse, which leads to relative hypoventilation. Negative upper airway pressure also leads to limited inspiratory flow at sleep onset, which is a dynamic obstruction event. During the period of limited inspiratory flow, although individuals can increase inspiratory efforts, the amount of airflow is not increased. During negative upper airway pressure, healthy subjects can increase the efforts of the respiratory muscles and their nocturnal ventilation may be improved. However, although subjects with OSA also increase the efforts of respiratory muscles, these efforts will lead to a complete closure of the upper airway. Finally, the complete closure causes the consequent cessation of airflow and even hypoxia. When the OSA subjects arouse from sleep, the upper airway patency is restored and effective ventilation is also improved, which leads to the cessation of apnea [10]. Currently, the morphological criterium that is most widely used in apnea reduction is the reduction of respiratory airflow signal (Fig. 1.6) to at least 10% of the base line. In the clinic, at least two missed breaths are scored and classified as apnea. Two missed breaths mean the recurrence of these apnea events is every two normal breaths at most. The low-frequency and high-frequency bands are changed because of these missed and recurrence events. In addition, respiratory efforts, which include AB and TH signals, and the respiratory bands (Fig. 1.7), are able to measure changes in movement because of breathing. During OSA events, people increase respiratory efforts to improve their nocturnal ventilation, and AB and TH signals are correlated with the patient's AF signal and

SaO_2 (Fig. 1.6). During apnea events, there are complete cessations of the upper airway, which lead to the movement of the patient's abdomen and thorax muscles, presenting potential power in apnea classification.

Fig. 1.5 Example of the airflow sensor placed under a patient's nose (Source: jcsm.aasm.org)

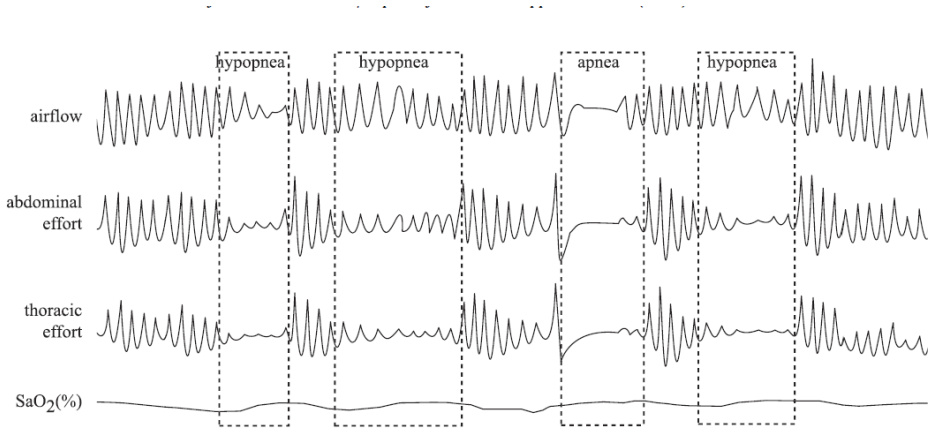

Fig. 1.6 Example of airflow, abdominal, thoracic, SaO_2 signals, and two respiratory efforts are related with the desaturation recordings (airflow and SaO_2 signals) [1]

Sleep respiratory disturbances are accompanied by changes in ECG signals. ECG signals uses electrodes placed on the skin to record the tiny electrical activity of a human's heart, which will change with the heart muscle's electrophysiology of depolarisation and repolarisation during each beat. One of the two snap ECG electrodes

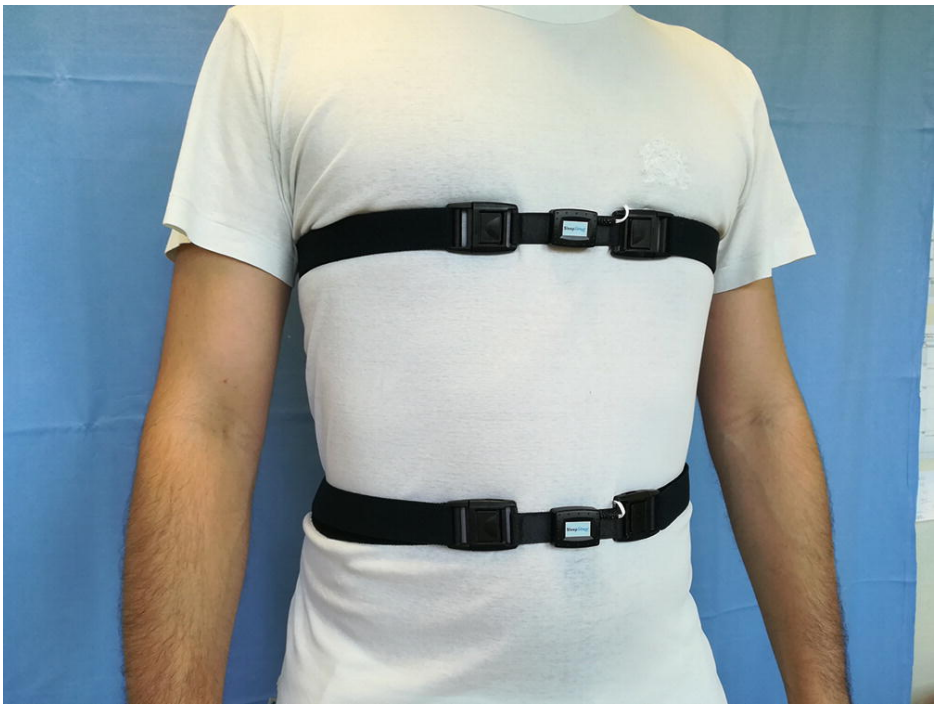

Fig. 1.7 Example of thoracic and abdominal bands (Source: link.springer.com)

is always put below the midpoint of the right clavicle, while the other is placed below the armpit and placed on the mid-rib on the left side. An example of the ECG signal is shown in Fig. 1.8. In each heartbeat, the theoretical waveform is assigned by the letters P, Q, R, S, and T. HRV (R-R intervals of ECG, the beat-by-beat series of heart rates) and EDR (ECG-derived respiration, referred to as the surrogate respiratory signal) has been shown to relate to sleep apnea [11]. Out of all the components, the QRS complex was the most essential and characteristic wave. Several studies have focused on the QRS detection algorithm and feature extraction from the QRS complex. People with OSA show different patterns in the RR interval (Fig. 1.9) with enhanced beat-to-beat variation at lower heart rates and reduced variation at faster rates relative to healthy people. In addition, ECG electrodes are placed on the surface of the chest, where signals can be influenced by the motion of the electrodes caused by the heart and by changes in the impedance of the thoracic cavity.

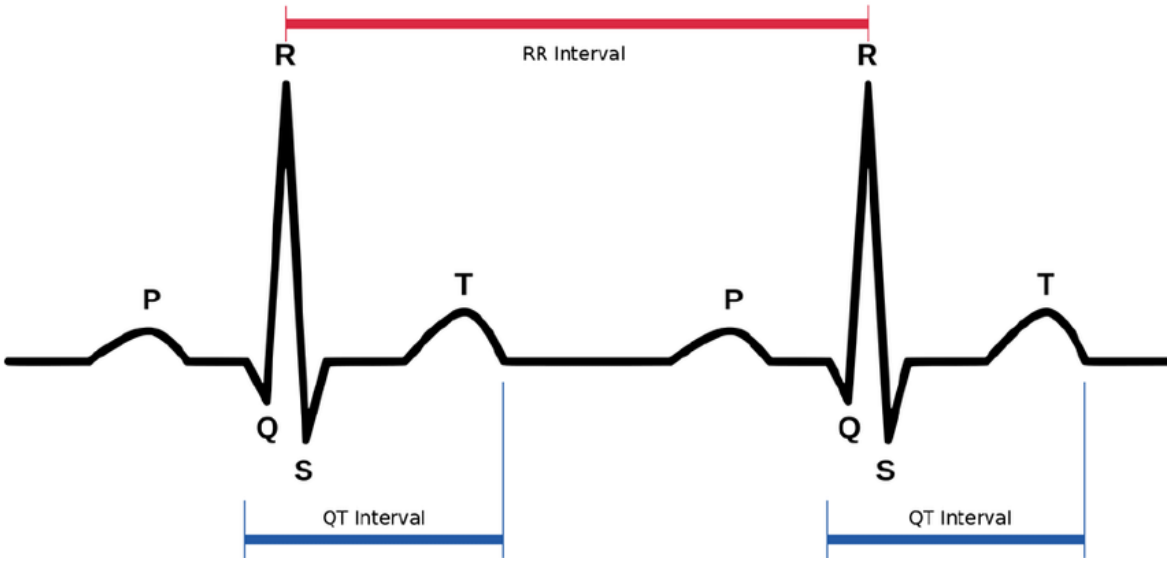

Fig. 1.8 Example of ECG signal shows P, Q, R, S, T waves, QRS complex and RR interval (Source: medictests.com)

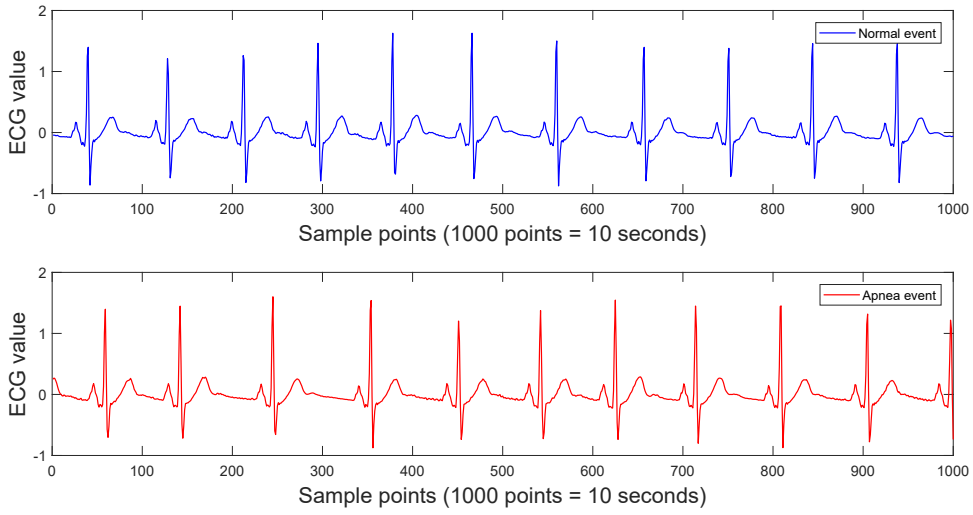

Fig. 1.9 Difference between apnea and normal windows in ECG signal (the top wave is the normal ECG window and the bottom wave is the abnormal ECG window)

EEG, EOG, chin EMG, and leg movement sensors, which detect bio-signals related to OSA, can provide information about sleep quality and quantity. EEG is widely used to record the electrical activity of the brain. Some EEG electrodes are placed on these sites and record the brainwaves from pre-frontal (Fp), frontal (F), temporal (T), parietal (P), occipital (O), and central (C) lobes. Four electrodes are placed on four sites of the midline sagittal plane of the skull, which are called FpZ, Fz, Cz, Oz, respectively. A "Z" refers to zero and these four electrodes are always considered as reference or ground points in PSG sleep studies. Fig. 1.10 shows the letters Fp, F, T, P, O, and C, which identify the lobe or the area of the brain. In apnea studies, the most commonly used electrodes are C3 and C4. To determine C3, first, the zero line of the tape measure is placed and firmly held at Cz mark point. Second, the tape measure is draped over the left side of the subject's head to reach the subject's left pre-auricular point. The length between the left pre-auricular mark and the right pre-auricular mark is measured and 20% of the length is computed. Finally, a mark is made at the 20% location from the Cz mark to the left pre-auricular point, which indicates the site of the C3 electrode. On the other hand, to determine the C4 electrode site, the zero line of the tape measure is again placed and held firmly on the Cz mark point. Then, the tape measure is draped over the right side of the subject's head to reach the subject's right pre-auricular point. Twenty percent of the length from the left pre-auricular mark to the right pre-auricular mark is also computed. Finally, a mark is made at the 20% location from the Cz mark to the right pre-auricular point to determine the C4 electrode site.

A brainwave is utilised when diagnosing various sleep disorders. Basically, the major spectrums of a sleep EEG are delta (δ, 1-4 Hz), theta (θ, 4-8 Hz), alpha (α, 8-12 Hz), and beta (β, 12-30 Hz) signals (Fig. 1.11). There are phenomena that occur at the end and beginning of an apnea event. At the end of an apnea event, the δ wave

often shows some approximate decreases, which often continue to the early or middle phase of the next apnea event, where it then begins to increase. At the end of the next apnea event, the δ wave begins to decrease again. On the other hand, α and θ waves demonstrate the exact opposite of the θ wave phenomenon. There are often approximate increases at the end of each apnea event in α and θ waves, which often continue to the early or middle phase of the next apnea event. At the end of the next apnea event, α and θ waves begin to increase again.

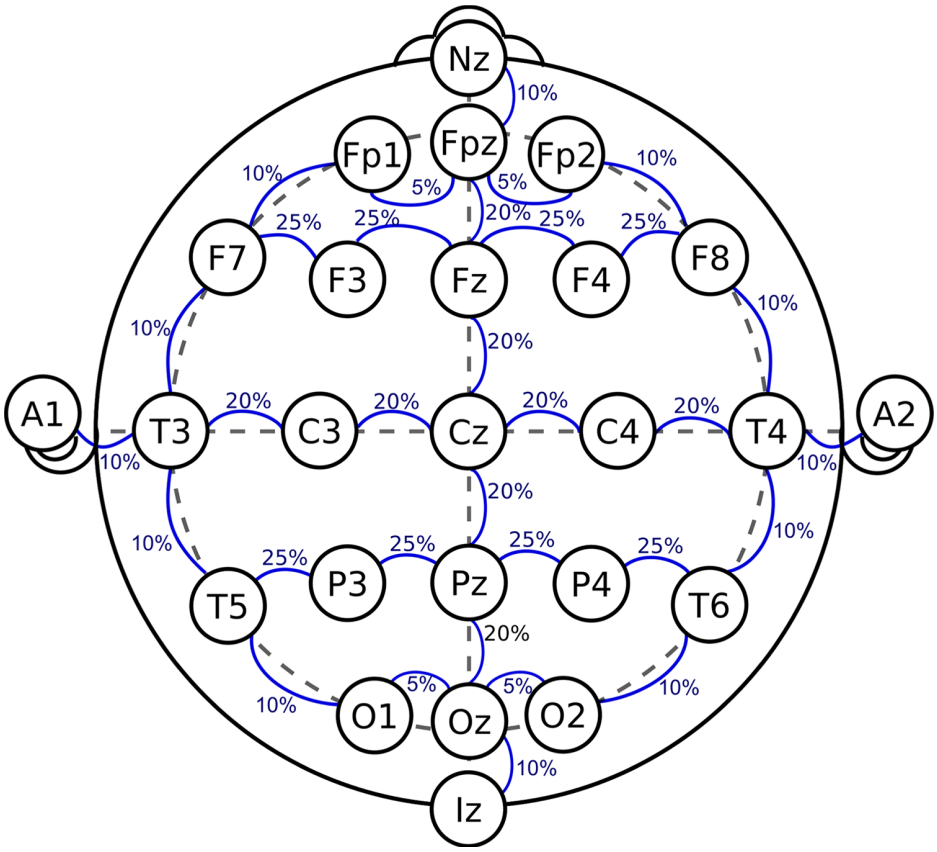

Fig. 1.10 10-20 system EEG placement and electrode distances (Source: ieeexplore.ieee.org)

EOG signals are used to record voltage changes caused by eye movements. One electrode is placed above the right eye, while the other electrode is placed above the left eye. These two electrodes can record the movements of the subject's eyes, and EOG signals are used to help determine different sleep stages (Fig. 1.12). In

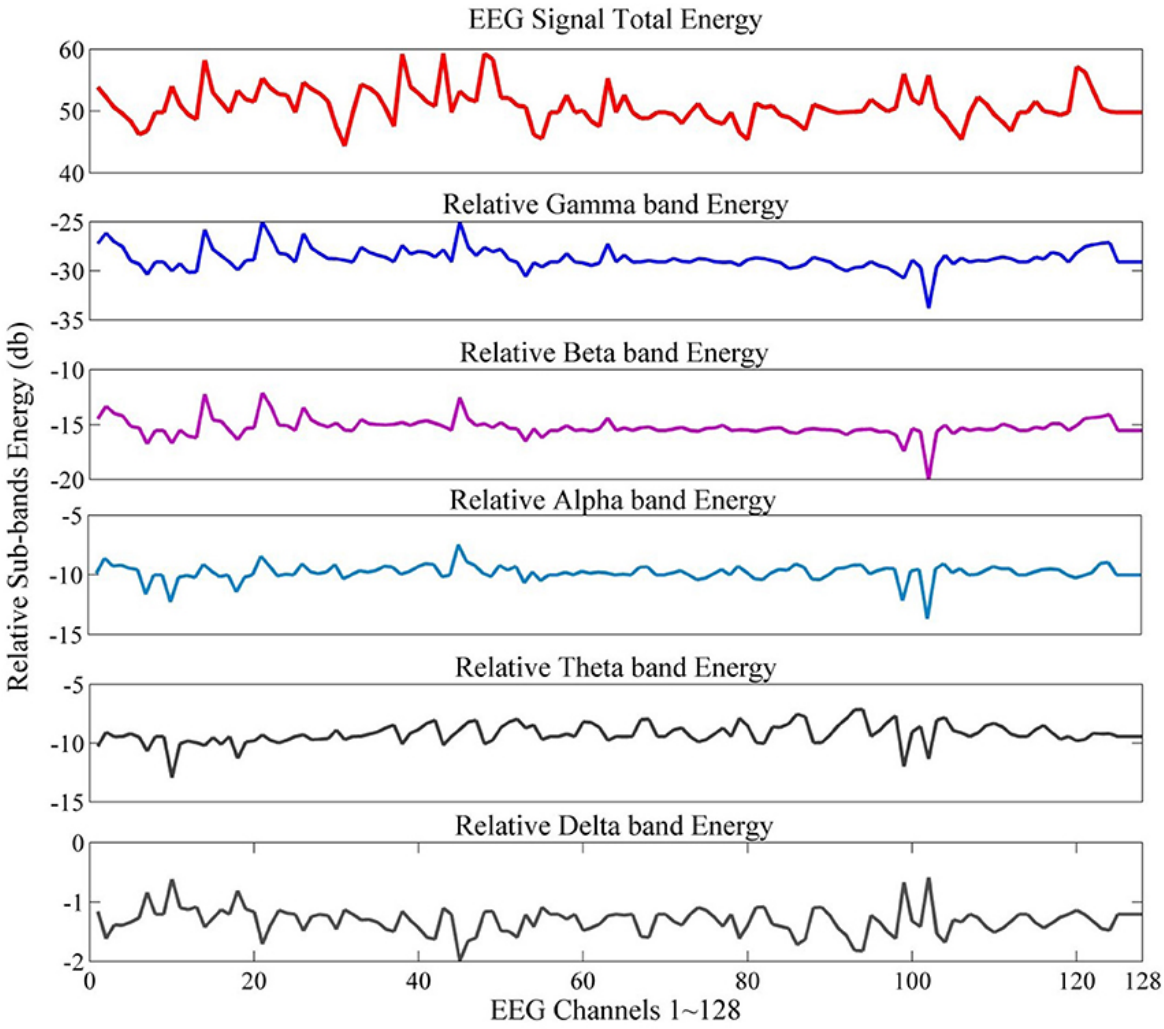

Fig. 1.11 the EEG signal and δ, θ, α, and β component waveforms (Source: frontiersin.org)

the wakefulness stage, EOG signals are random and high amplitude, however,during non-rapid eye movement, there is a slow-rolling EOG wave, while during rapid eye movement, the waves are very flat. In addition, to measure chin muscle waveforms, EMG electrodes are used to measure changes in the geniohyoid muscle since it is a large muscle under the chin (Fig. 1.12). Thus, one of the two electrodes are placed on each side of this muscle and the distance between these two electrodes is at least 3 cm. In general, one electrode is placed on the left ledge of the chin, while the other is placed on the right. Changes in electrical waveforms provide information about whether the muscle tone changes normally during the different sleep stages. The total or partial airflow obstruction leads to OSA and the obstruction also leads to increases in the baseline of muscle tone. During the rapid eye movement stage, chin EMG signals also record abnormally frequent bursts of discrete muscle activation.

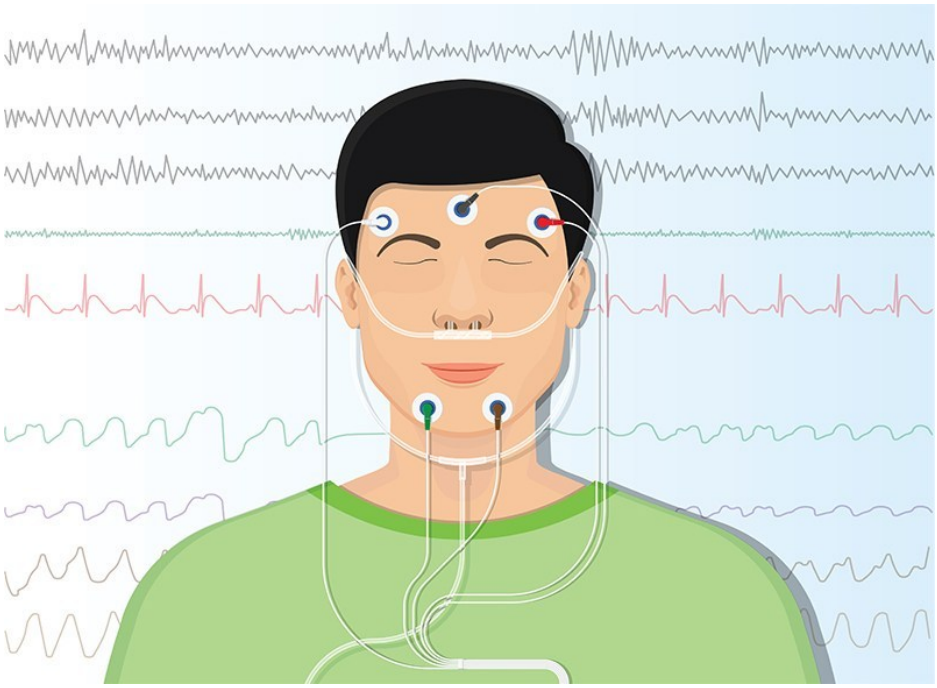

Fig. 1.12 EOG electrode placement and the two EOG electrodes are placed below the outer canthus of each eye, and EMG electrode placement is used to measure the changes of the geniohyoid muscle (Source: cvgclinical.co.za)

1.2 Motivation of Thesis

Individuals with OSA often face many serious diseases, including strokes, heart attacks, coronary failure, and even sudden death during sleep. Traditionally, sleep specialists perform OSA diagnosis according to visual observation of PSG signals. In each PSG test, it is inconvenient for a patient because sensors are placed on the patient's skin overnight and it is always conducted in a hospital or sleep laboratory under the supervision of a sleep technician. In addition, OSA diagnosis is based on the visual observation of sleep specialists, which has many disadvantages. Firstly, the identification of OSA by trained specialists requires expensive human resources and relies on the experience level of the specialist. Secondly, it is an onerous task for specialists to analyse the large volume of data that is collected overnight and specialists are prone to making errors due to fatigue. Thus, the manual detection of OSA is time-consuming and costly. In addition, sleep health experts need to monitor and review the overnight PSG signals, and the sleep apnea research of large-scale populations have been hindered by visual observation of sleep specialists.

Therefore, in order to detect OSA, automatic systems based on machine learning algorithms are desirable to help doctors during the diagnostic process. Furthermore, the development of OSA auto-detection device is desirable for in-home or mobile care, and this kind of device should be more comfortable, portable, and low-cost. Such a device can also be utilised in a preliminary OSA in-home test without sleep doctors and can also help sleep physicians in the early diagnosis and subsequent treatment of OSA. In order to ensure portability and wearability, automatic detection should only focus on a few core bio-signals (e.g., SaO_2 and airflow signals). To ensure the high performance of automatic detection, better classifiers will be selected. In the automatic detection approaches, general steps will include the feature extraction stage, the feature selection

stage, and the classification stage as shown in Fig. 1.13. To achieve the aims of constructing automatic systems, first, the feature extraction methods will obtain multi-domain features from the PSG data of OSA patients and healthy individuals, as well as features related to the bio-physiological criteria of OSA that hold good performance. Sleep experts will also diagnose OSA using these criteria from different kinds of bio-signals. Then, feature selection will be used to remove the redundant features before the classification stage to prevent over-fitting and reduce the computational load so that classifiers will only process useful features. Finally, OSA will be identified using machine learning algorithms, which will use these features as their input.

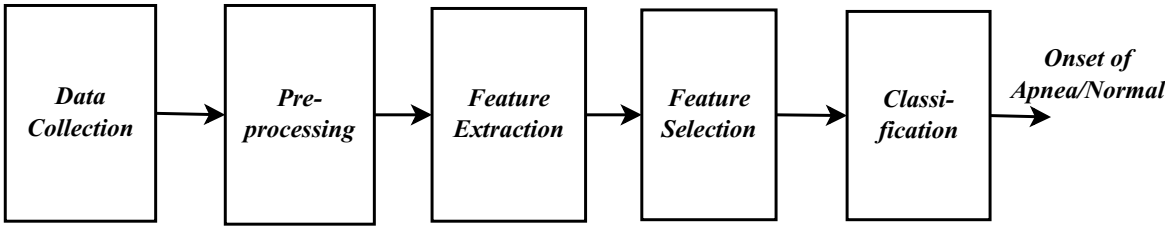

Fig. 1.13 Basic components of the OSA automatic detection progress

1.3 Objectives and Contributions

In this study, the automatic OSA monitoring system (shown in Fig. 1.13) is constructed. The main objective of this research is to develop a higher accuracy OSA monitoring system using multiple bio-signals. To achieve this main objective, four main contributions are presented as follows: feature extraction (in Chapter 3), feature selection (in Chapter 4), evaluation of a time-window segmentation method (in Chapter 5), and classification for OSA with the time-window method (in Chapter 6). Firstly, time-domain, frequency-domain, and non-linear analysis algorithms are proposed to extract features. Secondly, feature selection methods based on statistical analyses and machine learning methods are used to determine the features with good

discrimination. Thirdly, different machine learning algorithms are introduced, such as support vector machine(SVM), linear discrimination, decision tree. According to the performance of classifiers, the time-window segmentation method is used to evaluate the potential implementation of our monitoring system using real data. Fourthly, a multi-errors-reduction (MER) classification system is constructed, and boosting methods, the stacking method, and an artificial neural network (ANN) are used. More details about the contributions as follows:

1.3.1 Multi-Domain Feature Extraction Algorithms

Multi-domain feature extraction methods are proposed to extract the important features from multi-bio sensors. Methods based on time-domain, frequency-domain, and non-linear analyses are utilised to process multiple bio-signals, such as oxygen saturation (SaO_2), airflow (AF), two respiratory efforts (abdominal (AB) and thoracic (TH)), electrocardiogram (ECG), electroencephalogram (EEG), and electromyogram (EMG) recordings. The bio-signals have different bio-patterns between apnea episodes and normal episodes. Different apneic bio-physiological changes from multi-bio signals and extracted features based on the changes as follows:

1. *Oxygen Saturation (SaO_2) Signals*: The drop in SaO_2 usually starts from 10 to 30 seconds after the start of apnea detected from other signals. Once the apnea ends, the SaO_2 begins to recover. Sudden drops in oximetry values followed by fast recoveries are characteristic of apnea and patients with OSA show an oximetry curve with a typical saw-tooth morphology. In addition, apnea events happen in a pseudo periodic pattern and also lead to a modulation frequency change at the frequency band. Time-domain features stand for sudden drops

and fast recoveries, and very-low-frequency features reflect the changes caused by these drops and recoveries.

2. *Airflow Signals*: An apnea event is recorded when the respiratory airflow signal decreases at least 10% from its baseline and lasts at least 10 seconds. Clinically, at least two missed breaths are scored and classified as apnea. Two missed breaths mean that these apnea events occur every two normal breaths at most. The low-frequency, high-frequency, and very-high-frequency bands are changed because of these missed and recurrence events. Time-domain features are obtained from airflow signals and reflect decreases in apnea events. In addition, to extract frequency-domain features, wavelet transformation is used. Wavelet transformation uses a multi-scale basis and a varying window size to analyse non-stationary signals; the size is narrow at high frequencies and broad at low frequencies.

3. *Respiratory Efforts Signals*: AB and TH signals (Fig. 1.6) are used to measure apnea events. Abdomen and thorax muscles are active since the subject restores the closed upper airway by increasing respiratory efforts. Time-domain features and features from the high-frequency band are obtained from abdominal signals, while the time-domain and frequency-domain features are extracted from thoracic signals.

4. *ECG Signals*: ECG records the electrical activity of the subject's heart, and RR interval variation (Fig. 1.9) shows different patterns between apnea events and normal events with enhanced beat-to-beat variation at lower heart rates and reduced variation at faster rates. In addition, ECG electrodes are placed on the subject's chest. ECG-derived respiratory signals (EDR series) are obtained from ECG signals and are influenced by changes in the impedance of the thoracic cav-

ity. Many statistical analyses are utilised to extract the features. Power spectral density (PSD), Shannon's entropy, and wavelet spectral density (WSD) are capable of extracting the frequency-domain features. PSD describes the distribution of frequency power components. While normal events tend to generate a more random signal, apnea events tend to have a more deterministic signal. Shannon's entropy shows the difference of the uncertainty. Recurrence and principal component analysis (PCA) approaches are used to extract non-linear features. Recurrence is a fundamental characteristic of dynamic systems and recurrence quantification analysis describes the structures of recurrence. PCA maximises the variance of each dimension and projects high-dimension signals into lower-dimension data.

5. *EEG Signals*: EEG consists of four major spectrums (δ wave, θ wave, α wave, β wave). During apnea episodes, the δ wave often shows some approximate decreases, while the α wave and θ wave demonstrate approximate increases. Statistical algorithms and wavelet transformation are proposed to extract multiple features based on these criteria.

6. *Chin EMG Signals*: Chin EMG records abnormally frequent bursts of discrete muscle activation during apnea events since apnea leads to the closure of the upper airway and activates the geniohyoid muscle. Statistical analyses are used to extract features from chin EMG signals. Wavelet transformation and PSD are also utilised to obtain frequency-band features.

1.3.2 Two-Stage Feature Selection Method

In this thesis, a hybrid feature selection is proposed to select the significant features from the full feature set, which is an important step in an automatic process. The full

feature set increases the processing time and affects the performance because of the redundant features. Hence, a feature selection plays an important role to remove the redundant features before undergoing a classification stage, which should be able to prevent over-fitting and reduce computational load. The selected features should have good discrimination between apnea and normal events, which means they are more distinct between the two groups.

A two-stage feature selection procedure (Stage 1: the statistical analysis stage and Stage 2: the final feature subset selection stage using machine learning methods) is proposed in this thesis. In the first stage, if a feature is the objective of this stage, it is considered as being more independent and significant, and it is put into the feature subset. In the second stage, machine learning methods are proposed to determine the final feature subset since statistical methodologies only analyse differences between group values and associated labels, which means they are not able to compute differences in a higher dimension.

Firstly, in Stage 1 of the feature selection, the analysis of variance (ANOVA) and the rank-sum test are used. The ANOVA can compute the p-value for each kind of feature by analysing differences between group means and associated labels. The rank-sum test is also able to compute the p-value for each kind of feature by comparing randomly selected values from the first group and the second group, respectively. An ANOVA p-value ≤ 0.05 and a p-value $= 1$ of the rank-sum test means that the feature is independent between the apnea and normal groups. Secondly, in Stage 2 of the feature selection, the selected features are sorted into different classes to select the final significant feature subset according to the results from Stage 1. Each class is fed into machine learning methods using the hill-climbing algorithm, which is able to record the performance (accuracy, sensitivity, and specificity) in each iteration. The final feature subset is then determined by comparing the performance of all iterations.

1.3.3 Automatic OSA Monitoring System Based on Time-Window Method

The proposed automatic OSA monitoring system based on the time-window segmentation method has the potential to be implemented with real data. All PSG datasets provide the duration of each episode labelled by human experts. However, there is no duration parameter in real PSG data. A time-window segmentation method is used to evaluate the proposed monitoring system.

Each bio-signal is segmented into 60-second windows. Our monitoring system consists of three main parts, including the multi-domain feature extraction methods, the two-stage feature selection, and the classification stage. Firstly, in each 60-second window, multi-domain feature extraction methods are used to provide different kinds of features. Secondly, the two-stage feature selection evaluates the importance of each feature and determine the final feature subset. Finally, in the classification stage, the parameters are confirmed by the parameter tuning process. classifiers with the time-window segmentation method performs well compared with the non-time-window method.

1.3.4 Multi-Errors-Reduction Classification System

In this thesis, a multi-errors-reduction (MER) classification system is proposed. Selected significant features are extracted from multiple bio-signals and multiple feature extraction algorithms are used. Multi-domain features are known to have multi-domain errors. A single classifier is known to reduce one kind of error and show good classification performance. However, it is not able to solve multi-domain errors. In order to obtain better classification performance, different classifiers are utilised to reduce er-

rors in different domains and an ANN is used to combine a set of imperfect estimators and reduce multi-domain errors.

This multi-errors-reduction classifier consists of a stacking method, and the structure of the stacking consists of base learners (boosting systems) and a meta-learner (an ANN). In the process of stacking, the k-fold cross-validation procedure is used to train the meta-level classifier. To form a meta-instance, the decisions from the base classifiers combine with the gold standard in each training fold. Then, a meta-level classifier is trained on the meta-instances. In order to generate a set of classifiers, the boosting system iteratively changes the training sets and fits a simple classifier in the updated training sets. Then, it is able to arbitrarily minimise small error probability and predict labels. ANNs have the topology of a directed graph and is inspired by the structure of biological neural networks in human brains.

If the MER parameters are able to provide good performance with larger input data, this classifier is considered to be stable. In this MER classifier, the parameters are determined by the input data after the parameter tuning process. For applications with a large database, a MER classifier is able to perform classification tasks more efficiently in terms of accuracy, sensitivity and specificity.

1.4 Structure of Thesis

The thesis is organised into seven chapters:

Chapter 1

This chapter introduces the Background, the Motivation of Thesis, Objectives and Contributions, and Structure of Thesis.

Chapter 2

This chapter first introduces the current technologies behind the OSA monitoring system. Feature extraction methods, including time-domain, frequency-domain, and non-linear analysis from different kinds of bio-signals are also provided in Chapter 2. It is finalised by introducing simple machine learning algorithms, such as the support vector machine (SVM), the decision tree, the random forest, and complex ensemble methods, including the boosting method and the stacking method.

Chapter 3

This chapter initially introduces the proposed time-domain, frequency-domain, and non-linear analyses for feature extraction methodologies. Oxygen saturation (SaO_2), airflow (AF), two respiratory efforts (abdominal (AB) and thoracic (TH)), electrocardiogram (ECG), electroencephalogram (EEG), and electromyogram (EMG) recordings are utilised to provide multi-domain features.

Chapter 4

This chapter first presents a hybrid feature selection for sleep apnea detection using bio-signals. The proposed two-stage feature selection is utilised to confirm the significant features mostly related to OSA. Stage 1 is the statistical analysis stage, including analysis of variance (ANOVA) and the rank-sum test. According to the results of these two analyses, the importance of all features is obtained. In Stage 2, machine learning methods are used to determine a final feature subset.

Chapter 5

This chapter presents first the classification performance using different classifiers based on the determined feature subset. To improve the classification performance, a 60-second time-window segmentation is used and based on this segmentation, the proposed feature extraction, two-stage feature selection, and classifiers are used to

provide experimental results (accuracy, sensitivity and, specificity). Based on the experimental results, some conclusions are drawn.

Chapter 6

This chapter discusses the structure of a multi-errors-reduction (MER) classification based on ensemble systems and artificial neural networks (ANNs). A stacking system provides the main structure, including weak learners and a meta-learner. Firstly, many different machine learning methods provide their performance (accuracy, sensitivity, and specificity), and the selected significant features are fed into these machine learning methods. Based on a comparison of the performance, some classifiers are considered as weak learners since they hold a better performance. Furthermore, applications on tuning the parameters of weak learners are given. Secondly, constructing an ANN is introduced. Tuning the ANN parameters and determining the number of layers is discussed. Application examples on a larger OSA dataset are given to illustrate the applicability, stability, and merits of the proposed MER approach

Chapter 7

This chapter provides a conclusion of the findings and results in this thesis, and it indicates the directions for future research in this field.

Chapter 2

Literature Review

In this chapter, firstly, a review on automatic obstructive sleep apnea (OSA) detection with one single bio-signals or multiple bio-signals is given. Secondly, a review of machine learning algorithms is also illustrated in this chapter. The advantages and disadvantages of them are discussed.

2.1 An Introduction to Automatic OSA Detection Using Bio-Signals

OSA has been linked to diseases, and overnight polysomnography (PSG) as the gold standard is a high-cost test. The identification of OSA by trained specialists needs expensive human resources and relies on the experience level of the specialist, and the manual detection of OSA is time-consuming. As a result, there is a clear need for an automatic, low-cost detection for the doctors' diagnosis of apnea events, using easily accessible bio-signals.

2.1.1 Automatic Detection with Oxygen Saturation Signals

Oxygen saturation (SaO_2) signals are collected using non-invasive nocturnal pulse oximetry (see in Fig. 1.3), and the diagnosis of OSA using SaO_2 signals is widely used and accepted. Signal processing techniques in the time-domain, frequency-domain, and non-linear methods have been utilized for the OSA detection purposed using SaO_2 signals. The approaches were based on the oxygen desaturation during apnea events.

Time-domain, frequency-domain, and non-linear analyses were utilized to provide 28 features. Sequential forward selection and sequential backward selection had to be implemented. A novel hierarchical classification scheme was presented to classify the levels of severity of OSA using SaO_2 signals. A decision tree classifier and a probabilistic neural network were used to classify non-OSA and mild-OSA cases, and time-domain characteristics and non-linear parameters were the input of this algorithm. A decision tree classifier and a support vector machine (SVM) classifier classified moderate and severe subjects, and the feature subset included the time-domain and non-leaner parameters [12]. Non-linear analysis from overnight SaO_2 signals was utilized to extract information related to OSA, such as approximate entropy, central tendency measure, and Lempel-Ziv complexity. Marcos, J. Victor, et al. [13] evaluated a multilayer perceptron neural network classifier with one hidden layer of knots to assist OSA detection, and the neural network-based classification method was able to illustrate the excellent accuracy of 85.5%, the 89.8% sensitivity, the 79.4% specificity, and the 90.0% area under the ROC Curve (AUC). Otero, Abraham, et al. [14] used a fuzzy set in the detection of apnea episodes using the SaO_2 and the respiratory airflow signals. In this case, the desaturations were obtained from these two signals since an apnea is defined as a decrease in airflow to at least 10% of its baseline, and there is also the drop in SaO_2 signals. The performance was based on a multivariable fuzzy tempo-

ral profile model, and the accuracy, in this case, was 90% with the 3.2% false positive. Gunes S et al. [15] also used four clinical features, and a multi-layer perceptron artificial neural network (ANN) classifier to identify OSA and classify different levels (mild OSA, moderate OSA, severe OSA, and non-OSA). In this case, they designed a novel feature selection algorithm called multi-class f-score feature selection. Each feature was related with each degree of OSA, and the method was based on the feature selection, and the neural network classifier achieved 84.14% classification rate. Identifying apneic events used the expert criteria and the fuzzy-logic-based automated system [16]. From SaO_2 signals, fall duration and fall reduction were extracted. Three features reflecting a desaturation of an apneic event in the SaO_2 signal were obtained from airflow, thoracic, and abdominal respiratory signals, and these features included the degree of membership, the associated reduction, and the duration of each existing apneic interval. An optimized convolution neural network structure was used to develop the apnea event and SaO_2 signals were used. Greedy based optimization was proposed to construct the structure, and three different variants, including the topology transfer, the weighted-topology transfer with rough estimation, and the weighted-topology transfer with fine-tuning were used in the greedy based optimization [17].

2.1.2 Automatic Detection with Airflow Signals

The respiratory airflow signal is collected by using an electrode (Fig. 1.5) placed under a patient's nose. Currently, the morphological criteria that are used the most widely in apnea reduction are the reduction in the respiratory airflow signal to at least 10% of the basal line because the airflow signal shows the respiration and changes.

In a recent work by Otero A et al., sudden changes of respiratory airflow signals were found in apneic polysomnogram fragments, and the amount of air that the subject

is inhaling and exhaling also illustrated changes. The fuzzy set was used to identify OSA according to these changes [5]. The airflow signal provided the absolute value of the five detail coefficients using the wavelet transformation, and these coefficients were the inputs of a feedforward neural network using the Bayesian framework and a regularized cross-entropy function. The goal of the new approach presented in this case was to detect each apnea event into one of three basic sleep apnea types: OSA, central sleep apnea (CSA), and mixed sleep apnea (MSA). The results showed the mean classification accuracy was 83.78% [18]. In [19], a detection method for OSA in children was developed using the airflow and SaO_2 signals. The minimum amplitude and the kurtosis from 0.119 - 0.192 Hz, as well as the skewness from 0.784 - 0.890 Hz, were extracted by the power spectral density (PSD) of airflow, and 3% oxygen desaturation index was obtained from SaO_2 signals. A logistic regression model built these five features achieved: 85.90% sensitivity; 87.40% specificity; 86.30% accuracy. Lee, Hyoki, et al. [20] proposed a rule-based algorithm using airflow signals to detect apnea or hypopnea events automatically. The algorithm achieved good performance with the sensitivity of 86.40% and the accuracy of 84.50%. The paper [21] focused on the apnea/hypopnea detection using the single-channel airflow signal, and the respiratory rate variability signal was also derived from the airflow signal. Three features including the skewness from the respiratory rate variability signal, the peak amplitude, and the band power from the airflow signal were selected by the forward stepwise logistic regression procedure, and this procedure also provided the highest accuracy (82.43%) and AUC (90.30%). The study [22] extracted some frequency-domain and non-linear features from the airflow signal. Spectral features included the mean, the standard deviation, the minimum amplitude, the maximum amplitude, and spectral entropy, and three non-linear features were obtained by central tendency measure, Lempel-Ziv complexity, and sample entropy. The AdaBoost binary model based on regression

trees provided the best performance (the 89.00% sensitivity, the 80.00% specificity, and the 86.50% accuracy). In Study [23], the authors proposed a new probabilistic algorithm. A Gaussian mixture probability model was used to detect apnea events based on the posterior probabilities of the respective events.

2.1.3 Automatic Detection with Respiratory Signals

Some studies are focusing on how to classify sleep apnea using abdominal and thoracic signals since there is a correlation between two respiratory effort signals with patient's airflow signal.

Sezgin, Necmettin, and M. Emin Tagluk. [24] proposed a classification to detect OSA using the abdominal and thoracic respiration signals. Signals were decomposed into seven levels, and the mean of each detail coefficient were the inputs of an ANN. The multilayer neural network had an input layer with seven knots, two hidden layers with fifteen knots, and one output layer with three knots. The accuracies included 85.35% for OSA, 94.74% for CSA, and 80.42% for MSA, and the total accuracy was 86.84%. The study [25] used the wavelet transform method and an ANN. The wavelet transform of depth seven was used to process each abdominal signal, and the detail coefficients were the input of the neural network with an input layer, a hidden layer, and an output layer. Finally, the accuracy for each class was 73.42% for OSA, 94.23% for CSA, and 66.16% for MSA. Thoracic and abdominal signals provided features to classify various types of apnea [26]. Two features from the thoracic signals and two features from the abdominal signals were obtained and then fed into the SVM based on the standard radial kernel. The overall accuracy was 81.80%. Bianchi, M. T., et al. [27] developed an OSA detection algorithm using two respiratory efforts. This algorithm was able to search the time event where there was a decreased respiration

amplitude lasting at least 10 seconds. For classification, the AUC was over 92.00% using the apnea-hypopnea index (AHI) cutoff value of 5, while the AUC was over 85.00% using the 15 cutoff value. A novel approach [1] was presented for the detection of OSA into each basic type: OSA, CSA, and MSA. The features were the coefficients, which were obtained by a wavelet transformation and extracted from the thoracic signal. The classification model was based on the error-correcting output code formed by an ANN. Using this approach, the mean accuracy was 90.27%, and the accuracies for each type were 94.62% (OSA), 95.47% (CSA), and 90.45% (MSA).

2.1.4 Automatic Detection with ECG Signals

Sleep respiratory disturbances are accompanied by changes in the electrocardiogram (ECG). Therefore, there is an increased interest by researchers to utilize these signals. ECG uses electrodes placed on the skin, and these electrodes can record tiny electrical changes of the subject's heart, which will be changed with the heart muscle's electrophysiology of depolarization and repolarization during each beat. ECG electrodes can also record the changes of the respiratory cavity.

This method was used to find the RR intervals and QRS area series [28], and then the time-varying autoregressive model was used for the feature extraction. Ten spectral features were obtained from the high-frequency band, the low-frequency band, and the very-low-frequency band of the RR intervals and QRS area. They found that the very-low-frequency component was the best feature between apneic and no apneic conditions. The k-nearest neighbour and an ANN were used as classifiers to identify sleep apnea, and both algorithms reached the accuracy of over 88.00%, the sensitivity of over 85.00%, and the specificity of over 86.00%. A methodology for OSA detection was presented based on frequency-domain features using non-overlapping 1-minute

heart rate variability segments [29]. Heuristic splitting and spectral splitting based on stochastic variability were used to extract features. Using k-nearest neighbour classifier, the feature set achieved the over 80% accuracy. Hassan A R et al. [30] performed statistical and spectral features of ECG, which were different between apneic and no apneic conditions. They segmented ECG signals into 1 minute and then employed statistical and spectral features, such as mean, variance, kurtosis, spectral centroid, spectral decrease, spectral slope, spectral spread, and spectral flatness. Bootstrap aggregating was used to classify apneic events from normal events dependence on these features, and a decision tree was considered as the weak learners of the bootstrap aggregating. The method yielded the 85.97% accuracy, the 84.14% sensitivity, and the 86.83% specificity. In a previous work [31], it was demonstrated sleep apnea classification automatically. In the feature extraction stage, power spectrum density was analysed by the frequency analysis, and the heart rate variability and QRS complex were used to provide the frequency features. In the classification stage, hidden Markov models were used to confirm apnea patterns from ECG signals. In Study [32], an SA detection was proposed. Frequential stacked sparse auto-encoder method extracted features and time-dependent cost-sensitive model based on the hidden Markov model and the MetaCost algorithm was used to obtain classification results. Chang et al. [33] proposed a deep convolutional neural network model using ECG signals. Features are extracted from ECG signals by using time-domain analysis, frequency-domain analysis of intrinsic mode functions, and the SVM and the random forest classifiers are used to detect sleep apnea [34]. In Study [35], automatic detection was proposed with fuzzy, approximate, and sample entropy as feature extraction methods. The sequential feature selection method was used to pick the most effective features and then the GentleBoost classifier achieved high accuracy levels. Study [36] extracted time-domain features from ECG signals and the SVM model achieved an 82.12% accuracy,

an 88.41% sensitivity, and a 72.29% specificity. In Study [37], ECG signals provided frequency features and the features were fed into an SVM classifier with the Gaussian kernel. The paper [38] proposed an approach to detect OSA using ECG signals. Frequency features were extracted and the stacked autoencoder based deep neural network and the SVM method were considered to classify apnea events.

2.1.5 Automatic Detection with EEG Signals

Electroencephalogram (EEG) uses electrodes placed on the brain to record the electrical activity of the subject's brain. These electrodes measure voltage fluctuations, which result from ionic current within the neurons of the brain. EEG can provide sleep respiratory disturbance information. Therefore, it becomes one of the most significant signals in order to diagnose sleep respiratory disturbance [39].

The order-200 finite impulse response filter and the Hamming window method were utilized to filter out the sleep-related bands (0.5 - 32 Hz) from EEG signals. The study utilized the Hilbert-Huang transformation to obtain five ratio signals of the Hilbert spectrum, and the detector was used to classify OSA based on the duration. The mean accuracy was about 88.00% [40]. Ubeyli E D et al. [41] presented an automatic method to detect OSA from EEG signals, and the EEG signals were recorded from three electrodes (C3, C4, and O2). The level-6 discrete wavelet transformation was used to provide features, such as maximum, minimum, mean and standard deviation of each coefficient. These features were the input of three classifiers in the adaptive neuro-fuzzy inference system, and three classifiers combined the backpropagation gradient descent method and the least-squares methods. The fourth classifier was able to predict by combining the outputs of the three algorithms as input data, and the total classification accuracy was 96.25%, and the specificity was 97.50%, and the sensitivity

was 95.00%. The bispectral analysis was utilized to obtain bispectral characteristics called quadratic phase coupling, and the bispectrum was segmented into each subband of EEG (δ, θ, α, β and γ waves). An ANN was considered as a classifier to detect sleep apnea, and total accuracy of 96.15% was illustrated [42]. This paper [43] calculated the θ energy ratios by the Fourier transform from the original EEG signals, and the regression made the change of the theta ratio more obvious. An ANN of modified adaptive resonance theory was used to classify apnea and normal groups, and it was able to achieve 91.00% accuracy. The result of [44] showed that features extracted by autoregressive. In this study, the autoregressive PSD estimations of the EEG signals were obtained, and the EEG signals were recorded from C3, C4, and O2 electrodes. The least-squares SVM used the radial basis function and classified OSA, and the statistical parameters included the 92.50% specificity, the 97.50% sensitivity, and the 95.00% accuracy. This Study [45] introduced the evolutionary techniques optimized Hermite functions represented EEG signals, and the artificial bee colony algorithm was considered for the feature extraction. The extreme learning machine and least-squares SVM classifiers were used to detect OSA.

2.1.6 Automatic Detection with Multiple Bio-Signals

Currently, many detection methods for sleep apnea using multi-bio signals have been proposed frequently. Al-Angari H M et al. [46] showed that feature extraction from respiratory efforts, ECG, and SaO_2 signals, and an SVM were utilized to classify apneic parts. Two effort signals were computed the phase-locking value. The standard deviation and mean obtained from the RR interval series, the absolute and normalized powers in the very-low-frequency band, the low-frequency band, and the high-frequency band, and the ratio of the low-frequency power over the high-frequency power were computed. The mean and standard deviation of SaO_2 signals were also

computed. Using these features, the SVM implementing a second-order polynomial kernel with weight = 5 achieved the maximum performance (accuracy: 82.40%, sensitivity: 69.90%, specificity: 91.40%). Features were extracted from SaO_2 and ECG signals and used to train linear discriminant model [47]. SaO_2 and ECG signals were time-aligned to 30-second epochs. The features were computed from each epoch of SaO_2 signals, including the minimum SaO_2 value in each epoch, the mean SaO_2 value over each epoch, the average absolute rate of change per second in each epoch, the number of points below 92% saturation, the duration time that the SaO_2 value exceed the baseline value by at least 3%, the duration time that the SaO_2 value was less than the baseline SaO_2 by at least 3%, and the 3rd and the 57th point values in ordered SaO_2 values. In addition, the features of ECG have included the mean of RR interval series, the standard deviation of the RR interval series, the power spectral density of the RR interval series, and the square root of the average of the sum of the differences between two continuous RR intervals. The combined features were fed into a linear discriminant model and achieved an accuracy of 66.7%, a specificity of 67.00%, and sensitivity of 58.10%. The study investigated three types of ANNs [48], and these types were Elman, Radial basis function and feed-forward backpropagation. Airflow, abdominal, and thoracic signals were used to provided features. Three-level Haar wavelet transformation was utilized to decompose these signals, and extracted features included the mean, the variance, the standard deviation, the kurtosis, the geomean, the skewness, and the mad. As a result, the feed-forward neural network achieved the best performance (the 86.60% AUC). This study [49] proposed an automatic process based on two EEG recordings (C3/A2 and C4/A1) and one submental electromyogram (EMG). Considering the EEG derivations, the information about the power of the different subbands was included. Regarding the submental EMG signals, the information of the amplitude of the signal was included. Six classification methods

were utilized and evaluated for the automatic detection: a naive Bayes, an SVM, a k-nearest neighbour, an ANN, a classification tree, and a fisher's linear discriminant. The two ensemble classification methods, the k-nearest neighbour and the random forest, showed the best performance. The time-domain and frequency-domain features were extracted from ECG, pulse photoplethysmography, airflow, and abdominal signals [50]. The combination showed good results (accuracy of 75.00%, sensitivity of 81.80%, and specificity of 73.90%). This study proposed a real-time OSA detection based on ECG and SaO_2 signals [51]. Statistical and non-linear features were extracted and selected, and the final feature subset consisted of 150 features from ECG and SaO_2 signals. Bagging with REPTree obtained the highest accuracy (83.26%) and specificity (84.62%), while AdaBoost with Decision Stump achieved the best sensitivity (86.81%). Respiration signals, SaO_2 signals, and ECG signals were devised in Study [52]. Time-domain, frequency-domain, and non-linear features were extracted from these kinds of bio-signals and then a long short-term memory recurrent neural network model was proposed to detect different types of sleep breathing patterns.

The data extracted from the existing studies include the types of signals used and the metrics (sensitivity, specificity, and accuracy) shown in Table 2.1. From Table 2.1, it can be found that many studies based on single one bio-signal have achieved good results. The study (Übeyli, Elif Derya, et al. [44]) constructs the classification system based on EEG signals and holds high performance (sensitivity = 95.00%, specificity = 97.50%, accuracy = 96.25%). It means that SaO_2, airflow, abdominal, thoracic, ECG, and EEG signals have obviously different bio-patterns between apnea events and normal events. In addition, from Table 2.1, the studies based on multiple bio-signals may be not able to provide good performance if they extracted some irrelevant features. However, as mentioned in Chapter 1, sleep experts review different kinds of bio-signals to diagnose apnea episodes. An automatic OSA detection system based on

multiple bio-signals should be able to provide high performance. This system based on multi-bio signals can also obtain results more stably. These studies extracted some irrelevant features from different bio-signals and the features affected classification performance. It is important to remove irrelevant features and select relevant features.

Table 2.1 Studies to automatically detect OSA based on bio-signals

Study	Bio-signal	Features	Classification	Metrics(%)		
				Sen	Spe	Acc
Marcos, J. Victor, et al. [13]	SaO_2	Non-linear	Neural networks	89.80	79.40	85.50
Otero, Abraham, et al. [14]	SaO_2	Threshold-based	Fuzzy set	N/A	N/A	90.00
Gunes S et al. [15]	SaO_2	Multi-class f-score	Neural networks	N/A	N/A	84.14
Fontenla-Romero, Oscar, et al. [18]	Airflow	Frequency	Neural networks	N/A	N/A	83.78
Gutiérrez-Tobal, Gonzalo C., et al. [19]	Airflow	Frequency	Regression	85.90	87.40	86.30
Lee, Hyoki, et al. [20]	Airflow	Threshold-based	N/A	86.40	N/A	84.50
Gutiérrez-Tobal, G. C., et al. [21]	Airflow	Time, Frequency, non-linear	Regression	N/A	N/A	82.43
Gutiérrez-Tobal, Gonzalo C., et al. [22]	Airflow	Frequency, non-linear	Adaboost	89.00	80.00	86.50
Sezgin, Necmettin, and M. Emin Tagluk. [24]	AB + TH	Frequency	Neural networks	N/A	N/A	86.84
Lin, Yin-Yan, et al. [26]	AB + TH	Time, Frequency	Support vector machine	N/A	N/A	81.80

Continued on next page

Table 2.1 – continued from previous page

Study	Bio-signal	Features	Classification	Metrics(%)		
				Sen	Spe	Acc
Guijarro-Berdiñas, et al. [1]	AB + TH	Frequency	Neural networks	N/A	N/A	90.27
Mendez, Martin O., et al. [28]	ECG	Time, Frequency	Neural networks, K-nearest neighbor	85.00	86.00	88.00
Martínez-Vargas, Juan David, et al. [29]	ECG	Frequency	K-nearest neighbor	N/A	N/A	80.00
Hassan A R et al. [30]	ECG	Time, Frequency	Bootstrap aggregating	84.14	86.83	85.97
Hsu, Chien-Chang, and Ping-Ta Shih. [40]	EEG	Frequency	Threshold-based	N/A	N/A	88.00
Ubeyli E D et al. [41]	EEG	Frequency	Neural networks	95.00	97.50	96.25
Tagluk, M. Emin, and Necmettin Sezgin. [42]	EEG	Frequency	Neural networks	N/A	N/A	96.15
Liu, Derong, Zhongyu Pang, and Stephen R. Lloyd. [43]	EEG	Frequency	Neural networks	N/A	N/A	91.00
Übeyli, Elif Derya, et al. [44]	EEG	Frequency	Support vector machine	97.50	92.50	95.00
Al-Angari H M et al. [46]	AB + TH + ECG + SaO_2	Non-linear	Support vector machine	69.90	91.40	82.40
Cohen, Gregory, and Philip De Chazal. [47]	SaO_2 + ECG	Time, Frequency	Linear discriminant	58.10	67.00	66.70

Continued on next page

Table 2.1 – continued from previous page

Study	Bio-signal	Features	Classification	Metrics(%)		
				Sen	Spe	Acc
Maali, Yashar, and Adel Al-Jumaily. [48]	Airflow + AB + TH	Frequency	Neural networks	N/A	N/A	86.60
Gil, Eduardo, et al. [50]	ECG + SaO_2 + airflow + AB	Time, Frequency	Linear discriminant	81.80	73.90	75.00
Xie, Baile, and Hlaing Minn. [51]	ECG + SaO_2	Time, Frequency	AdaBoost	86.81	84.62	83.26

Table 2.2 and Table 2.3 summarise a literature review on the OSA detection studies using single one bio-signal or multi-bio signals. ECG, SaO_2, airflow, EEG, chin EMG, thoracic, and abdominal signals are widely used to diagnose OSA automatically. We can see that the most used bio-signal is ECG signals.

Table 2.2 Summary of apneic event detection approaches using single one bio-signal

Source	Number of studies	Sample references
SaO_2	19	[12], [13], [14], [15], [16]
Airflow	13	[5], [18], [19], [20], [21], [22]
ECG	52	[28], [29], [30], [31]
EEG	13	[40], [41], [42], [43], [44]
EMG	6	[53], [54], [55], [56], [57], [58]
EOG	1	[59]
Ab+Th	7	[24], [24], [26], [27], [1]

Table 2.3 Summary of apneic event detection approaches using multi-bio signals

Category	Source
Multi-channel respiratory signals	SaO_2+ECG, SaO_2+ECG+Ab+Th, Airflow+SaO_2+ECG, Airflow+SaO_2+ECG+Ab, Airflow+SaO_2, EEG+ECG, EEG+EMG, Airflow+SaO_2+Ab+Th, ECG+Airflow+SaO_2, EEG+EMG+EOG+ECG

2.2 An Introduction to Statistical Analyses

To select the features with good discrimination, we use two statistical tests: the analysis of variance (ANOVA) and the rank-sum test.

2.2.1 An Introduction to Analysis of Variance

ANOVA can analyse differences between group means and associated labels. The error (residual) sum of squares (SSE) is computed while the total sum of squares (SST) is calculated. We obtain the group sum of squares (SSG) then

$$SSG = SST - SSE \tag{2.1}$$

We next obtain group means and sum of squares errors by dividing by their respective degrees of freedom, df_{group} and df_{error}. df_{group} = I-1 and we take the number of groups to be I. df_{error} = n-I, such that n is the number of observations. Resulting means are defined as the group mean squared and mean squared errors, and are defined by

$$MSG = \frac{SSG}{df_{group}} \tag{2.2}$$

$$MSE = \frac{SSE}{df_{error}} \tag{2.3}$$

Finally, we obtain from these the F-ratio:

$$F = \frac{MSG}{MSE} \tag{2.4}$$

This ratio follows an F_{v_1,v_2} distribution, where $v_1 = df_{group}$ and $v_2 = df_{error}$. After the F-ratio and its paired value in the F_{v_1,v_2} distribution are obtained, the p-value is computed.

2.2.2 An Introduction to Rank-Sum Test

As for the rank-sum test statistical analysis, the aim is to confirm that the two groups are independent. Two values are randomly-selected from the first group and the second group, respectively, and these two values are compared. A random sample is selected from each populations. Let n_1 and n_2 be the numbers of the two samples, respectively. We rank the combined $n_1 + n_2$ observations in ascending order and substitute a range of 1,2...to the $n_1 + n_2$ observations. We assign conflicting observations with their mean ranks.

$$u_1 = w_1 - \frac{n_1(n_1+1)}{2} \tag{2.5}$$

$$u_2 = w_2 - \frac{n_2(n_2+1)}{2} \tag{2.6}$$

where w_i (i=1,2) indicates the sum of the ranks corresponding to n_i observations. Let u=min(u_1,u_2) and u will be compared with the desired critical value. In this

step, the exact probability distribution is determined, and then the Rank-sum test also obtains the p-value by using the distribution and a statistical table.

2.3 A Brief Introduction to Machine Learning Algorithms

2.3.1 Support Vector Machine

SVM is introduced by Vapnik [60] and is a classification method and regression analysis. SVM is a binary classifier and in an N-dimensional space it constructs an optimal separating hyperplane (OSH). The input vector is transformed via nonlinear mapping into a K-dimensional space from the N-dimensional space.

A training data $D=\{(x_i,y_i)\}_i^L$ constitutes of every training vector $x_i \in \mathbb{R}^n$ and the matched label $y_i \in (+1, -1)$. Every training vector x_i is transformed via a nonlinear mapping function $\mathbf{z}=\varnothing(\mathbf{x})$ and into a higher-dimension space $\mathbb{F}$. Every vector $\mathbf{w} \in \mathbb{R}^n$ is normal vector to the hyperplane, and the points $\mathbf{z}$ lie on the hyperplane, and margin bias b is in $\mathbb{R}$, and these parameters satisfy $\mathbf{w}\cdot\mathbf{z}+b=0$, while the input data satisfy the equation $y_i(\mathbf{w}\cdot\mathbf{z}+b)\text{-}1 \geq 1 \ \forall i$. The bias is computed by $2/\|\mathbf{w}\|$. When the OSH maximizes the bias to minimizes $\mathbf{w}^2/2$, the SVM constructs the optimal one. The quadratic programming with Lagrange multipliers α_i can solve a convex optimization problem, and transformation into the following dual problem

$$W(\alpha) = \sum_{i=1}^{L} \alpha_i - \frac{1}{2} \sum_{i=1}^{L} \sum_{j=1}^{L} \alpha_i \alpha_j y_i y_j \mathbf{z}_i \mathbf{z}_j \tag{2.7}$$

such that $\sum_{i=1}^{L} \alpha_i \ y_i = 0$ and $\alpha_i \geq 0$.

Table 2.4 List of kernel functions used in support vector machine models

Kernel parameter	Mathematical formula
Linear	$K(x_i,x_j)=x_i,x_j$
Polynomial	$K(x_i,x_j)=(x_i,x_j+1)^d$ where d is the degree of polynomial
Radial basis function (RBF)	$K(x_i,x_j)=e^{-\left\|x_i-x_j\right\|^2/2\sigma^2}$

A kernel function $K(.,.)$ can calculate the term $\mathbf{z}_i \cdot \mathbf{z}_j$, such that $\mathbf{z}_i \cdot \mathbf{z}_j = \varnothing(x_i) \cdot \varnothing(x_j) = K(x_i,x_j)$. After confirming the optimal α, for the vector $\mathbf{w}$, the optimum solution is solved by

$$\mathbf{w} = \sum_{i=SVs} \alpha_i y_i \varnothing(x_i) \tag{2.8}$$

where SVs are the support vectors.

The output is computed using the following equation when the input vector is $x_i \in \mathbb{R}^n$:

$$y = f(x) = sgn(\mathbf{w} \cdot \mathbf{z} + b) = sgn(\sum_{i=SVs} \alpha_i y_i K(x_i, x_j) + b) \tag{2.9}$$

Selecting a kernel function $K(x_i,x_j)$ affects the performance of the SVM in the construction of the SVM. In this study, three kernels are evaluated. Table 2.4 shows three commonly selected kernels and the required parameters.

2.3.2 Decision Tree Algorithm

Decision tree algorithm [61] is a method widely used in the classification area, and it is a tree-like structure which represents the classification rules. The goal is to train a tree based on input variables, and this tree is able to predict the targets of testing variables.

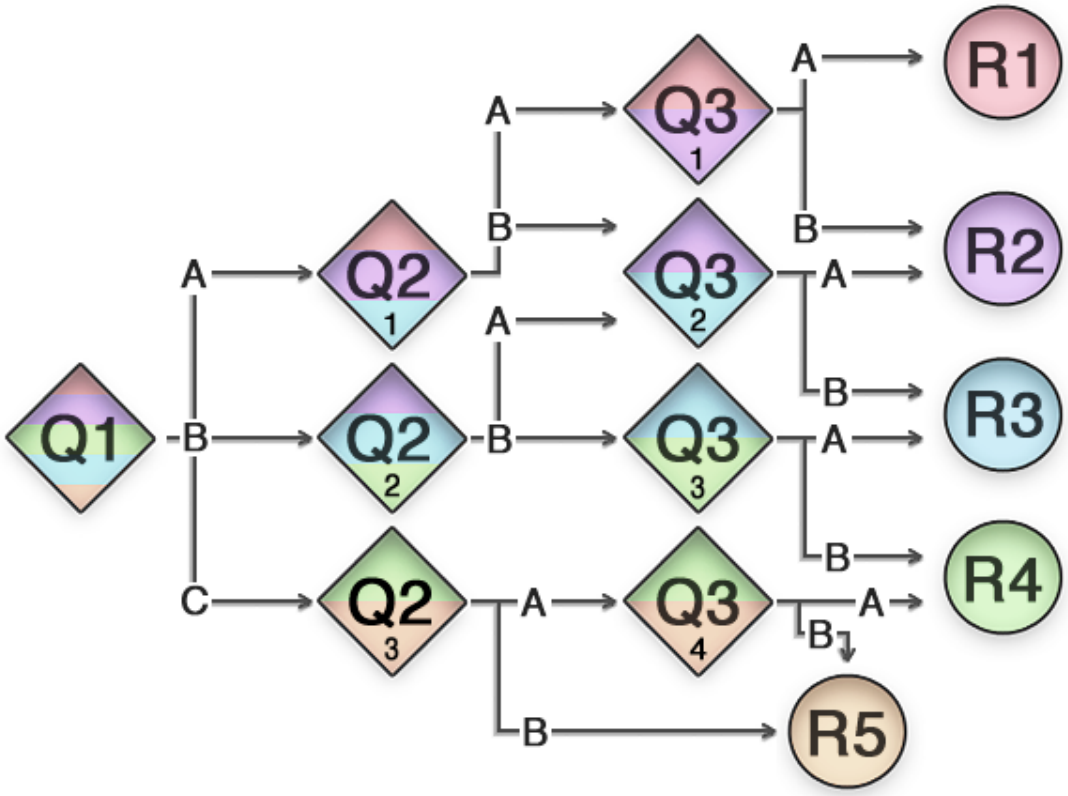

Fig. 2.1 An example of decision tree (Source: matrix.squiz.net)

Decision tree contains a root as the starting node and leaves as the terminal nodes. Leaves denote the labels or targets of variables. Internal (non-leaf) nodes represent the attributes correlated to the objects being classified. Each branch represents each possible value of the internal node, which it originates.

The basic idea of the decision tree algorithm is very simple. Fig. 2.1 shows an example of a decision tree, and this decision tree is to classify one input vector into one of the different results (*R1, R2, R3, R4, R5*). In this tree, all leaves take A, B or C labels. At the root layer, the *Q1* attribute is evaluated. Depending on the value (A, B or C) of the input observation for this attribute, one of the branches in this node is determined. This process is repeated for all internodes (shown in the *Q2* layer and the *Q3* layer) until a label (*R1*, *R2*, *R3*, *R4* or *R5*) is obtained to the

input observation. Decision tree algorithm has several advantages. Firstly, a tree-like structure is simple and easy to understand. Secondly, it does not need to know the distribution of training data.

2.3.3 K-Nearest Neighbors Algorithm

The k-nearest neighbour algorithm (kNN) is a nonparametric, simple and intuitive method [62], where identifies the classes of testing data based on the nearest training data. The kNN algorithm needs to store all training data in the memory, compute the similarity or the distance between the testing data and all training data, and then assign the testing data into the class. In general, the kNN algorithm is followed these step:

1. Compute the similarity or the distance between the input testing set and each training set.

2. Sort the similarity or the distance and determine the k closest classes to the input testing set.

3. Determine the class of the input testing set by using majority voting.

Many studies apply many ways to measure the similarity or the distance between testing sets and training sets. The most widely used measures are the Euclidean Distance and the Manhattan Distance methods. The Euclidean Distance is computed:

$$D(x, y) = (\sum_{i=1}^{m} |x_{\mathrm{i}} - y_{\mathrm{i}}|^2)^{1/2} \tag{2.10}$$

The Manhattan Distance is calculated:

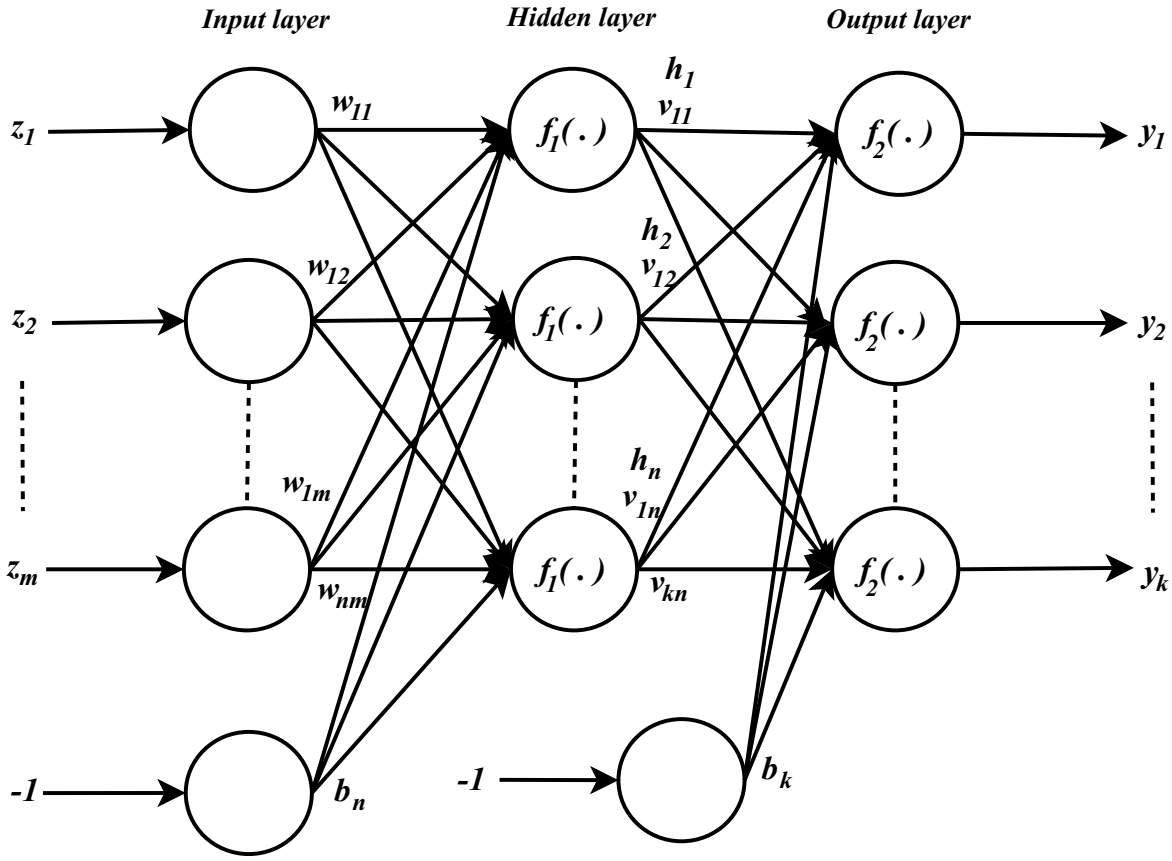

Fig. 2.2 The basic components of an artificial neural network

$$D(x, y) = \sum_{i=1}^{m} |x_{\mathrm{i}} - y_{\mathrm{i}}| \tag{2.11}$$

In the progress of the kNN classification, it is one of the most crucial factors to choose the appropriate k parameter. If the value of k is too small, there is the overfitting problem. If the value of k is too larger, there is the underfitting problem. In general, the value of k is between 3 and 10, which is able to overcome the overfitting and underfitting problems. Calculating each classification error of each value of k, and then confirming the value of k by selecting the least classification error.

2.3.4 Artificial Neural Networks

An ANN is a generic nonlinear function classification method, and it is considered a powerful method and used in the computer technology area. The fundamental characteristic is a parallel-distributed memory, and it is a combination of basic knots, called neurons. It is inspired by the biological neural system, and some methods resemble the characteristics of the biological neural systems (learning or reacting) to construct networks. An ANN schematic representation is shown in Fig. 2.2, and it is a combination of one input layer, some hidden layers, and one output layer. Each input vector is put into the input layer and then distributed to each neuron in the first hidden level. Each neuron has its weight vector, and this weight vector is multiplied by the input vector. Each knot sums the inputs, and a non-linear activation function is used to transform the sum value. Finally, the output layer multiplies all output vectors from the final hidden layer by the weights of the final output layer, and the final predictions are obtained from the output layer by the summations and activations.

The j node (j = 1, 2, 3 ... n) in the hidden layer receives many input vectors represented by p_i (i = 1, 2, 3 ... m), shown in Fig. 2.2. Each input vector is multiplied by a weight factor (w_{ji}), where i is the number of input vectors. b_j is the single bias of the j node. The result of each node is given by Equation 2.12 to perform further calculations. Furthermore, in order to obtain outputs of this ANN, the o node (o = 1, 2, 3 ... k) in the output layer also receives many input vectors computed by the hidden layer and represented by h_j. Each input vector is multiplied by a weight factor (v_{oj}). b_o is the single bias of the o node in the output layer. Each output of the ANN is computed by Equation 2.13.

$$h_j = f_1(\sum_{i=1}^{m}(w_{ji}p_i) - b_j) \tag{2.12}$$

$$y_o = f_2(\sum_{i=1}^{k}(v_{oj}h_j) - b_o) \tag{2.13}$$

Weight factors and bias values are adjustable variables of knots, and they are also able to activate or inhibit knots. As for a weight factor, if it is positive, it will activate the knot. If it is negative, it will inhibit the knot. As for a bias value, when it is positive or larger, it will inhibit the knot. If it is negative or zero, it will activate the knot. The transfer function is used to compute the output of the j knot, and the output from the j knot is calculated by Equation 2.12. The most widely used transfer functions are the step, linear, and sigmoid transformation functions. The step function is simple (Equation 2.14). The output value is zero when the input value is less than 0, and the output value is one when the input value is greater than or equal to 0. The step function is commonly used in the classification area. The linear function is able to return any value since the output is computed by linear Equation 2.15. The sigmoid transfer function is commonly used in multi-layers and backpropagation networks. It outputs zero when the input value is very low, and the output value is one when the input value is very high. The equation of the sigmoid function is shown in Equation 2.16.

$$f(x) = \begin{cases} 0, & x < 0, \\ 1, & x \geq 0. \end{cases} \tag{2.14}$$

$$f(x) = \alpha x + \beta \tag{2.15}$$

$$f(x) = \frac{1}{1 + e^{-x}} \tag{2.16}$$

The number of hidden layers affects the performance of an ANN. Thus, in order to design an appropriate network, one of the most critical tasks is to determine an appropriate number of hidden layers. An ANN with too few hidden nodes only leads to a linear estimate of the actual trend, and it would be incapable of differentiating between complex tasks. On the other hand, the network leads to poor generalization for untrained data, if this network has too many hidden layers since the noise in the data make the network overfit the training data. In addition, as for the network with too many hidden layers, the training period becomes time-consuming. Training algorithms are also an essential part of designing a neural network. An appropriate topology is able to hold better results, if a suitable training algorithm is used in an ANN to help train data, and the better algorithm also helps shorten the training time [63] [64].

2.3.5 Ensemble Algorithm

More research used ensemble systems as classifiers in several areas during the past decade. The basic idea of ensemble learning is to train a set of individual classifiers and then combine the predictions from these individual classifiers, instead of obtaining a final prediction from a single trained classifier [65]. In general, ensemble systems consist of two stages. In the first stage, a set of weak learners fit training data to provide their predictions. In the second stage, the predictions obtained from weak learners in the first stage are combined to provide a final label. Thus, a combination of multiple classifiers can reduce variance and bias, and the results obtained are less dependent on a single classifier. This integration of all good base learners contributes to higher performance levels.

The main motivation for combining individual classifiers is to improve their generalization ability. When each classifier is trained on a limited dataset, it is known to make errors with the assumption. Different classifiers make errors in different patterns. Thus, a combination of multiple classifiers can obtain a better overall recognition capability than the individual classifiers with limitations.

Three main reasons are illustrated to explain why ensemble systems are successful in the machine learning area [66]. Firstly, to predict the training dataset, classifiers usually find the best hypothesis by searching a hypothesis space. However, since the training dataset is limited, a single classifier always is able to confirm many different hypotheses in the space, and all of these hypotheses can fit the training dataset reasonably well. The classifier can not confirm which hypothesis from these hypotheses is better. Ensemble methods can help to obtain a good approximation of the unknown true hypothesis. Secondly, classifiers minimize error functions by searching the minimum, but they always get stuck in local optima. Ensemble algorithms can solve this local optima problem since they minimize error functions by starting the local search from multiple base learners. Thirdly, the unknown true hypothesis may not be in the hypothesis space we are looking for. Ensemble systems can provide a representational enlarged space which includes the unknown true function because of a combination of several hypotheses from base learners.

2.3.5.1 Random Forest Algorithm

Random forest is a popular tree-like ensemble machine learning method [67]. It is a highly adaptive method for datasets with a very larger number of attributes. Random forest is also a robust method and less prone to generalization error and overfitting. The random forest has the potential of real-time since its running time is short.

Random forest is also a tree-bagging algorithm which is the combination of random decision trees. Considering a training data (x_i,y_i) for i = 1, 2, ..., N, where x_i is a vector and y_i is the matched label. The tree-bagging algorithm repeatedly selects a bootstrap training set and then fits trees to these selected training sets. For a testing set, the final classification label is obtained by the majority vote of these fitted trees. Based on the basic idea of the tree-bagging algorithm, the random forest selects a random subset of the features to split the node in the process of fitting trees. The random forest selects this searching method instead of searching in the whole feature set in order to reduce the correlation of the trees. In the progress of building trees, it is one of the crucial tasks to determine the number of trees and the maximum number of nodes in a tree.

2.3.5.2 Boosting Algorithm

To classify apnea events, boosting methods and the stacking method were used in this thesis. The boosting method is based on an idea, and a combination of simple classifiers can hold better performance than any of the simple classifiers alone [68]. With the same training data, a simple classifier (weak learner) is able to produce labels with a probability of error strictly and slightly. Then, a strong learner is able to arbitrarily minimize small error probability and predict labels more accurately by combining weak learners (as illustrated in Fig. 2.3).

In the study, some boosting methods were used, including Gradient Boosting, CatBoost from Yandex, Light Gradient Boosting (Light GBM) and XGBoost [69]. Classic boosting methods minimize the errors by updating the weights of a training set, but Gradient Boosting uses the mistake-residual error directly. CatBoost is a kind of gradient boosting, and it is based on decision trees. “Cat” comes from “Categories” which means CatBoost quickly handles categorical features for a large data set. Light GBM

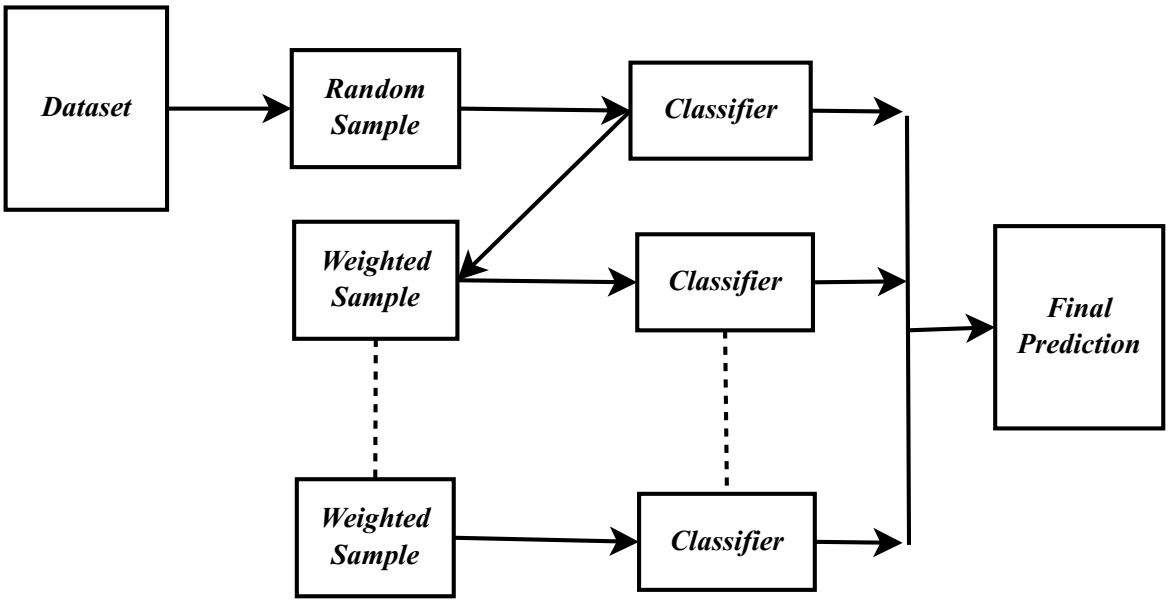

Fig. 2.3 The basic components of the boosting method

is also an algorithm based on a decision tree, and it fits data to split the tree leaf wise. It can reduce the loss and improve accuracy, and this progress is based on the leaf-wise method when growing on the same leaf. XGBoost is a decision-tree-based algorithm, and it uses some methods to avoid overfitting, optimize the classic boosting algorithm and enhance performance, and these methods often include tree-pruning, parallel processing, handling missing values and regularization.

2.3.5.3 Stacking Algorithm

Stacking is one of most widely used ensemble approach [70]. It combines predictions of base classifiers (level-0) in a meta-level (level-1) classifier and a meta-level classifier can correct the decisions of the base level and predict final labels as shown in Fig. 2.4. For training the meta-level classifier, the k-fold cross-validation is used in training data [71]. For forming a meta-instance, in each training fold, the decisions from base classifiers combine with the gold standard. Then, a meta-level classifier is trained on the meta-instances. When a new instance is waiting for classification, first, outputs of

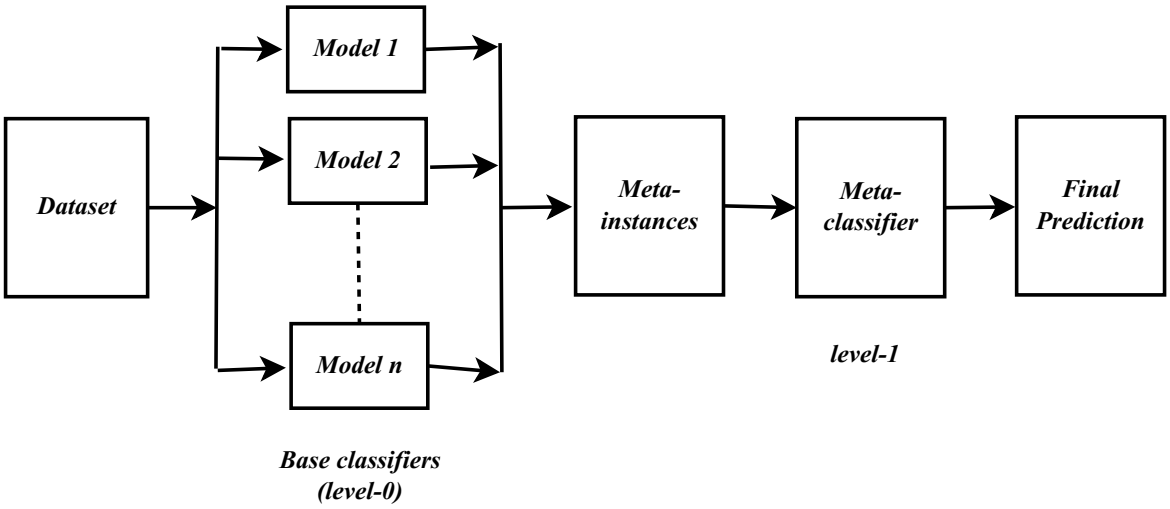

Fig. 2.4 The basic components of the stacking method

base learners are calculated, and then these outputs are put into the meta-classifier, and final results are obtained.

2.4 Summary

In this chapter, previously accepted studies on automatic OSA detection with one single bio-signals or multiple bio-signals are shown. It can be found that many studies based on single one bio-signal have achieved good results (accuracies are more than 85%). The results mean that SaO_2, airflow, abdominal, thoracic, ECG, and EEG signals have obviously different bio-patterns between apnea events and normal events, and they can be used for apnea detection. In addition, sleep experts review different kinds of bio-signals to diagnose apnea episodes, not only rely on one kind of bio-signal. An automatic OSA detection system based on multiple bio-signals is considered in this work because multi-bio signals are more stable.

Secondly, two statistical analysis methods are given, and they are used in the feature selection stage. Two methods do not require the numbers of two groups are

the same. Even in patients with apnea, they breathe normally in most of the night. Therefore, the situation in the real world is that most of the sleep minutes correspond to normal breathing. Thus, the number of apnea events is much less than the number of normal events, and both two analyses are considered to fit this study.

Finally, some single classifiers are shown, and their advantages and disadvantages are discussed in this part. However, multiple bio-signals are used to provide multi-domain features, and single classifiers are not able to solve multi-domain errors. To improve performance, classifier combination methods are also shown in this part.

Chapter 3

Multi-Domain Feature Extraction Methods

3.1 Introduction

During sleep, three basic respiratory disturbances are found. In these types, the most common disorder is obstructive sleep apnea (OSA). The closure of the upper airway is repeated and temporary, and it is defined as OSA in adults. OSA can cause a complete breathing cessation and the cessation lasts greater than 10 seconds [72]. OSA should be objectively assessed to treat it. The apnea/hypopnea index (AHI) can evaluate OSA. To diagnose OSA, overnight polysomnography (PSG) is considered as the gold standard. PSG monitors multiple bodily signals and records the comprehensive biophysiological changes during sleep. These biophysiological recordings include oxygen saturation (SaO_2), midsagittal jaw movement, breath airflow, respiratory events, body position, snoring, electromyography (EMG), electroencephalography (EEG), electrocardiography (ECG), and electrooculography (EOG) [73]. A clinician must characterize pathological events and identify different parameters that are shown on the PSG

to perform the diagnosis. The amount of data obtained is huge since the PSG records the changes in the whole night. As a result, for physicians, it is time-consuming to identify OSA. Therefore, there is a clear need for an automatic, low-cost detection for the detection of apnea events, using highly accurate and easily accessible sensors.

In the progress of automatic detection shown in Fig. 1.13, feature extraction is one of the most crucial steps. Feature extraction is applied to obtain significant features that reflect bio-physiological criteria, and the features can provide good classification performance. A literature review on these feature extraction methods has been in Chapter 2. In order to obtain features with good discrimination, a few core signals should be used. These core signals tend to be the most useful for the detection of OSA. The sensors of these signals tend to be low-cost, easy to collect bio-physiological changes and make the subject comfortable during sleep. Table 2.2 and Table 2.3 summarise a literature review on the OSA detection studies using single one bio-signal or multi-bio signals in Chapter 2. In this thesis, to provide good performance, ECG, SaO_2, airflow, EEG, chin EMG, thoracic, and abdominal signals are considered as the fundamental signals employed to diagnose OSA automatically.

This chapter is organized as follows. The apneic bio-physiological patterns from ECG, SaO_2, airflow, abdominal, and thoracic signals are shown in Section 3.2. The detail methods of multi-domain feature extraction methods using EEG and chin EMG signals are discussed in Section 3.3. A chapter summary is drawn in Section 3.4.

3.2 Apneic Bio-Physiological Patterns and Feature Extraction Methods

We extract features from ECG, SaO_2, airflow, abdominal, and thoracic signals, and these bio-signals have obviously different bio-patterns between apnea episodes and normal episodes. Time-domain, frequency-domain, and non-linear methods are used in the thesis to obtain features based on different bio-patterns.

3.2.1 Apneic Patterns and Feature Extraction from SaO_2 Signals

There are many quantitative indices used to diagnose OSA. The most widely used in SaO_2 include the cumulative time less than a value, a 90% decline from baseline or the number of SaO_2 less than a value, a 3% decline from baseline. Sudden drops in oximetry values continued with fast recoveries are characteristics of apnea. These sudden drops and fast recoveries are shown in Fig. 3.1. It also can be found that there is a typical saw-tooth morphology of the oximetry curve. Time-domain features are correlated with these changes. In addition, apnea events happen in a pseudo periodic pattern, and apnea events also lead to a modulation frequency change at the frequency band, shown in Fig. 3.2. There is a dramatic decrease in the very-low-frequency band in Fig. 3.2.

We obtained 12 kinds of features from SaO_2 signal. *Preprocessing:* Non-physiological artifacts were considered as artifacts when the point-to-point difference was more than 8%, and the median value was calculated in the beginning 10 seconds to replace artifacts.

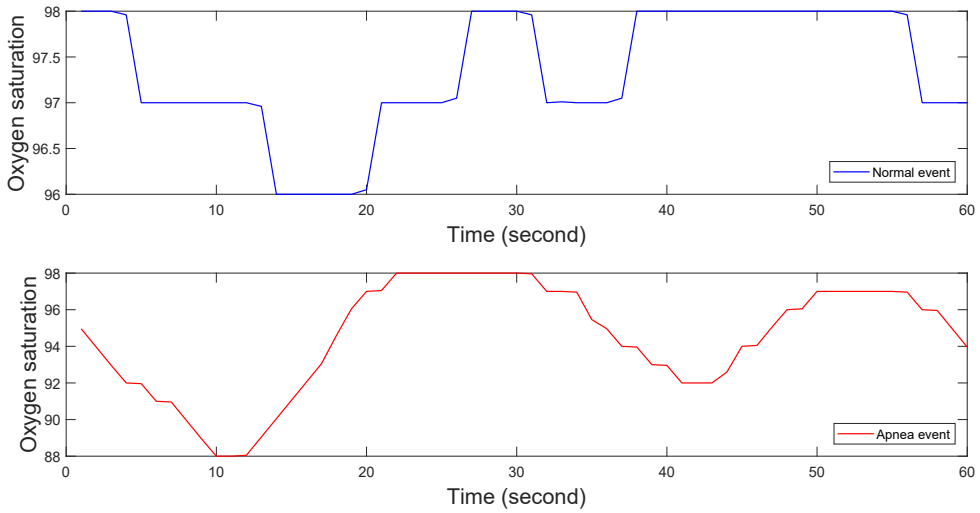

Fig. 3.1 The oxygen saturation difference in time domain between one apnea event and one normal event (the top wave is the oxygen saturation in the normal event and the bottom wave is the oxygen saturation in the abnormal event)

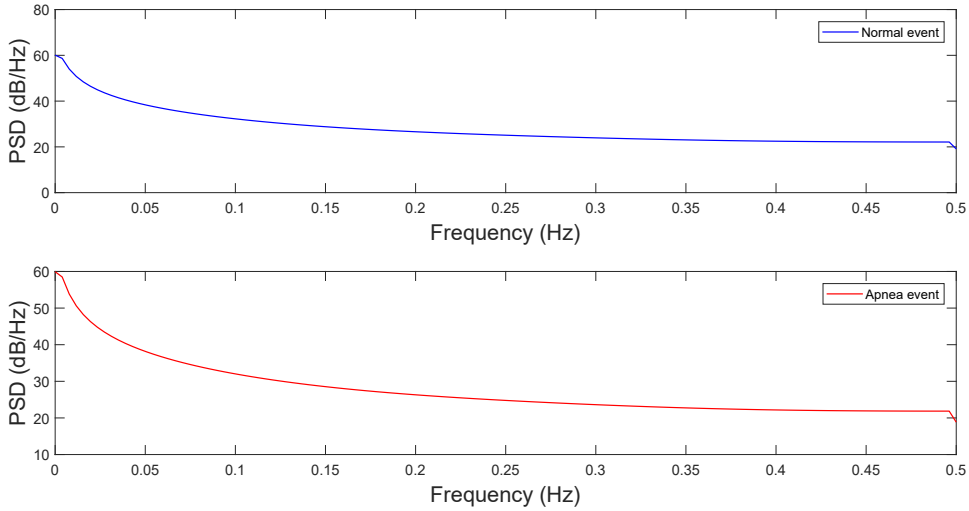

Fig. 3.2 The oxygen saturation difference in frequency domain between one apnea event and one normal event (the top wave is the frequency wave of the normal event and the bottom wave is the wave of the abnormal event)

Multi-domain features: In this study, the feature set includes the median of a window (med), the maximum-to-minimum changes which are more than 2% (MM2), kurtosis (kur), variance (var), minimum (min), mean (mean), the number of zero-cross

Table 3.1 Multi-domain features from SaO_2 signals (The abbreviations are shown and detail methods are shown in Section 3.2.1)

Feature	Short description
med	median of window
var	variance of window
NumZC	number of zero-cross in window
Bel98	time spent below the 98% maximum in window
Abo98	time spent above the 98% maximum in window
MM2	maximum-to-minimum changes more than 2% of window
min	minimum of window
comp	complexity of window
kur	kurtosis of window
mean	mean of window
SD_1	short-term variability of each window
mean_$PSD_{0.016/0.05}$	mean in 0.016 - 0.05Hz frequency band in window

in the window (NumZC), and complexity of a window (comp). Kurtosis can describe the shape (tailedness) of a probability distribution of an SaO_2 window. Complexity can describe the length of the shortest description in an SaO_2 window. The Poincare SD_1 is computed, which shows the short-term variability of an SaO_2 window (SD_1) [74]. Two features which were the time spent below and above the 98% maximum were obtained (Bel98 and Abo98) in an SaO_2 window.

Finally, the power spectral density (PSD) method was used to show the intensity of desaturation events. The distribution of frequency power components is described by PSD. The PSD method includes parametric methods and non-parametric methods. The advantage of parametric methods is higher accuracy, while the advantage of non-parametric is less computational complexity. A 5th-order Yule-Walker autoregressive estimate was used to obtain the mean of PSD in 0.016 - 0.05Hz frequency band (mean_$PSD_{0.016/0.05}$) an SaO_2 window.

In summary, the features extracted from SaO_2 signals are shown in Table 3.1.

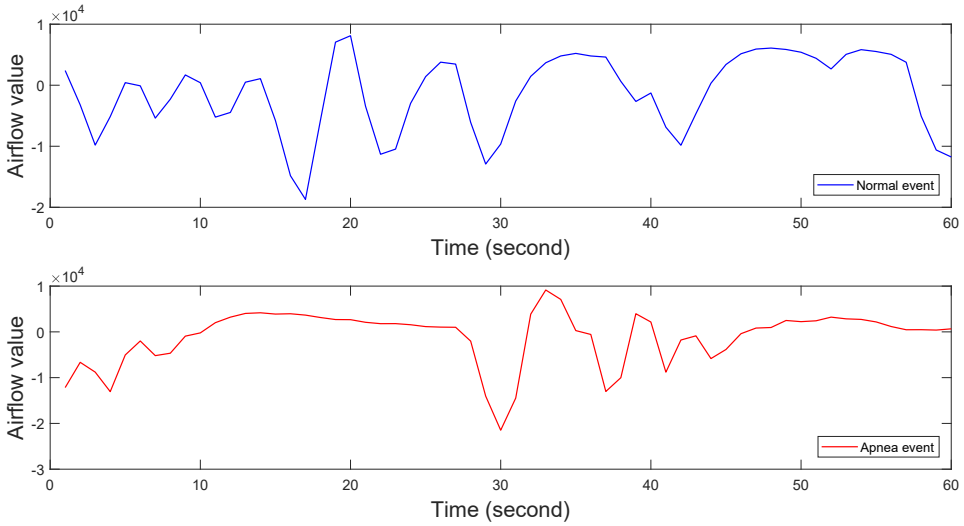

Fig. 3.3 The airflow difference in time domain between one apnea event and one normal event (the top wave is the airflow signal in the normal event and the bottom wave is the airflow signal in the abnormal event)

3.2.2 Apneic Patterns and Feature Extraction from Airflow Signals

The apnea is defined as the closure of the upper airway, and the closure lasts at least 10 seconds. The airflow signal swings up during exhalation and down during inhalation. Fig. 3.3 demonstrates the time-domain difference of an airflow window between the top normal event and the bottom apnea event. We can see in the top normal event the airflow signal swings up and down, but the values of airflow signals in the bottom apnea event are almost zero, which also lasts at least 10 seconds. In addition, apneas are recorded when there are at least two missed breaths of length [75], and this phenomenon is also shown in Fig. 3.3. As for the bio-physiological criteria in frequency-domain, time-domain changes lead to the differences in the very-low-frequency band and the high-frequency band, shown in Fig. 3.4. It can be found that the curve in the bottom apnea event is higher than the curve in the top normal event.

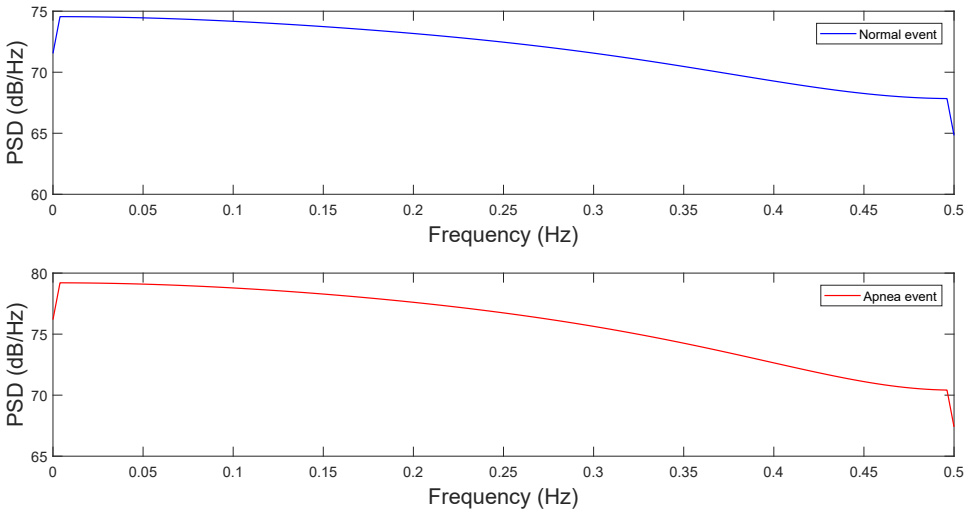

Fig. 3.4 The airflow difference in frequency domain between one apnea event and one normal event (the top wave is the frequency wave of the normal event and the bottom wave is the wave of the abnormal event)

Airflow recordings provided eight features. *Preprocessing:* We obtained the median of a 10-second window, and it was used to correct baseline. For noise removal, a 3rd-order Butterworth low-pass filter was used, and the cut-off frequency was 3 Hz. Airflow recording was re-sampled to 1 Hz.

Multi-domain features: Some statistical features included the mean (mean), the median (med), the standard deviation (std) in each airflow window. On the other hand, decreases and increases of airflow signals throughout the night are related to the frequency domain. PSD and wavelet are used to explore differences of the spectral information between the sleep apnea positive and negative groups in this study.

Wavelet transformation has advantages over classical Fourier transformation. Wavelet transformation uses a multi-scale basis, and varying window size is used to analyst non-stationary signals, and the size is narrow at high frequencies and broad at low frequencies. The functions of scaling and wavelet are applied by wavelet transformation

Table 3.2 Multi-domain features from airflow signals (The abbreviations are shown and detail methods are shown in Section 3.2.2)

Feature	Short description
mean	mean in window
med	median in window
std	standard deviation in window
comp	complexity in window
mean_$PSD_{0/0.1}$	mean in 0 - 0.1Hz frequency band in window
mean_D1	mean of 1st detail coefficient level
mean_D2	mean of 2nd detail coefficient level
mean_D3	mean of 3rd detail coefficient level
mean_A3	mean of approximation coefficient level
mean_$PSD_{0.4/0.5}$	mean in 0.4 - 0.5Hz frequency band in window

to decompose signals at approximate and detailed levels, and there is a correlation between high-pass and low-pass filters and these functions.

In an airflow window, the Welch method used a segment length of 5 samples with 2.5 overlapped samples, and the mean within the frequency ranges of 0 - 0.1 Hz and 0.4 - 0.5 Hz was calculated (mean_$PSD_{0/0.1}$ and mean_$PSD_{0.4/0.5}$). The depth of wavelet transformation was three, and Daubechies wavelets three was also used as the mother wavelet. The wavelet decomposed an airflow window signal to obtain means of one approximation level and three detail coefficient levels (mean_D1 to mean_A3). Finally, there is a variable called complexity (comp). The measure of complexity is related to drops and lengths of the drops in an airflow window, which is related with the closure of the upper airway in apnea events.

In summary, the features extracted from airflow signals are shown in Table 3.2.

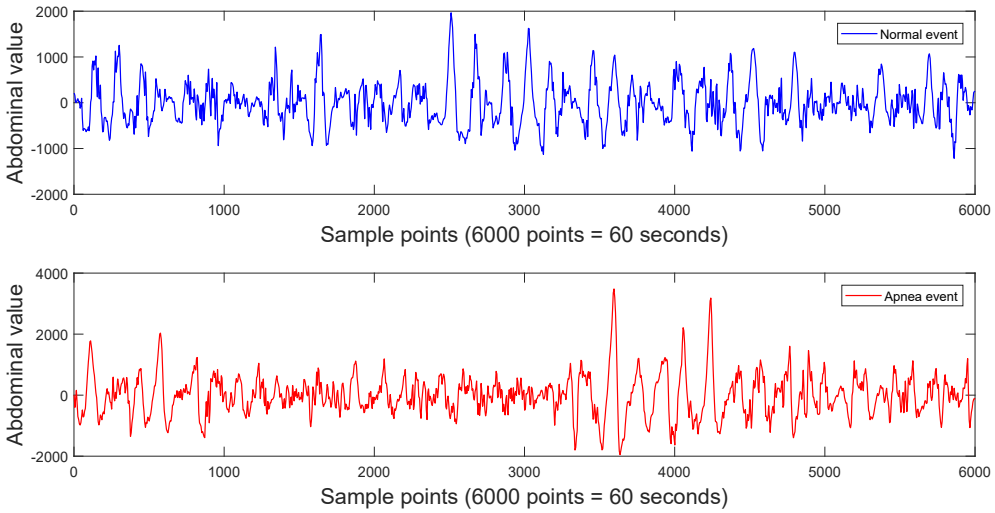

Fig. 3.5 The abdominal difference in time domain between one apnea event and one normal event (the top wave is the abdominal signal in the normal event and the bottom wave is the abdominal signal in the abnormal event)

3.2.3 Apneic Patterns and Feature Extraction from Respiratory Signals

We used time-domain and frequency-domain methods to process abdominal and thoracic signals. For the work, we estimated the performance of abdominal and thoracic signals. Fig. 3.5 shows the time-domain difference of an abdominal window between the top normal event and the bottom apnea event and Fig. 3.6 shows the time-domain difference of a thoracic window between the top normal event and the bottom apnea event. A total closure of the upper airway is scored as the apnea, and it causes respiratory reductions or cessations. We can see the respiratory values in the bottom apnea event are lower than the values in the top normal event in Fig. 3.5 and 3.6. Then, to restore breathing, people increase respiratory efforts to improve their nocturnal ventilation. Fig. 3.5 shows different fluctuation in the middle of the bottom apnea event.

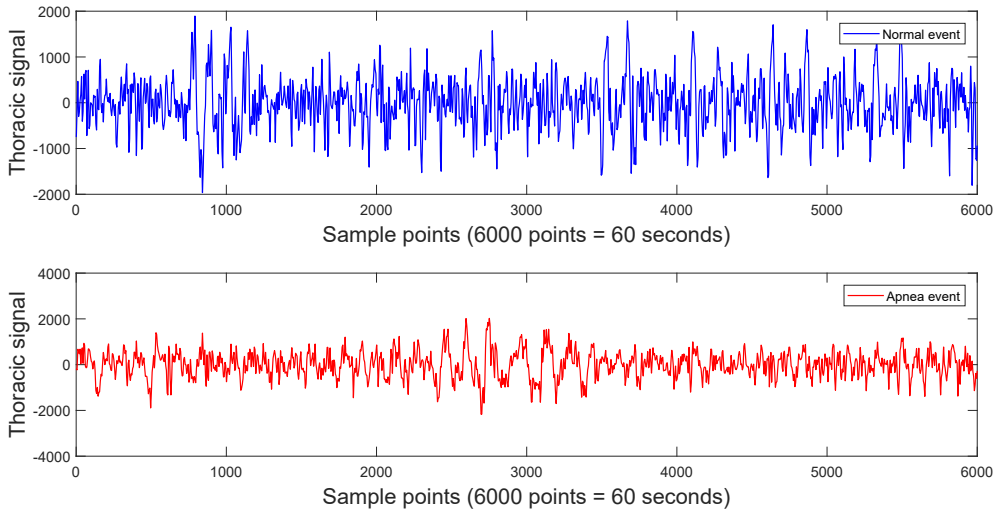

Fig. 3.6 The thoracic difference in time domain between one apnea event and one normal event (the top wave is the thoracic signal in the normal event and the bottom wave is the thoracic signal in the abnormal event)

We used time-domain and frequency-domain methods in abdominal and thoracic recordings. Six features are extracted from abdominal recordings, and six features are extracted from thoracic recordings. *Preprocessing:* Firstly, for abdominal and thoracic recordings, we used the median of a 60-second window to start baseline correction. Then, the Butterworth filter was used to remove noise, and this third filter was a pass-band with a range of 0.05 - 4 Hz.

Multi-domain features: The feature set included the summation and the standard deviation of an absolute abdominal window (sum_abs and std_abs) and the mean of an abdominal window (mean). According to time-domain changes, there is a difference in the very-high-frequency band. Yule-Walker method and wavelet transformation are used to extract frequency-domain features in an abdominal window. The segmentation length of Yule-Walker method is 40 samples, and the mean of the 80 - 100Hz frequency range (mean_$PSD_{80/100}$) is obtained. A wavelet is Daubechies 2 with the depth of 2, and the mean of 1st and 2nd detail levels are computed (mean_D1 and mean_D2).

Table 3.3 Multi-domain features from abdominal signals (The abbreviations are shown and detail methods are shown in Section 3.2.3)

Feature	Short description
sum_abs	summation of each absolute window
mean	mean of each window
std_abs	standard deviation of each absolute each window
mean_$PSD_{80/100}$	mean of the 80 - 100Hz frequency range in each window
mean_D1	mean of 1st detail level
mean_D2	mean of 2nd detail level

Features in frequency-domain and time-domain are extracted from each thoracic window. The feature set consists of the median, summation, mean, variance, and standard deviation of each thoracic window (sum, med, std, mean, and var), and the summation of the 80 - 100 Hz frequency band (sum_$PSD_{80/100}$) is computed by Yule-Walker method in a thoracic window.

In summary, the features extracted from abdominal signals are shown in Table 3.3, and the features extracted from thoracic signals are shown in Table 3.4.

Table 3.4 Multi-domain features from thoracic signals (The abbreviations are shown and detail methods are shown in Section 3.2.3)

Feature	Short description
sum	summation of each window
med	median of each window
var	variance of each window
std	standard deviation of each window
mean	mean of each window
mean_$PSD_{80/100}$	summation of the 80 - 100Hz frequency band in each window

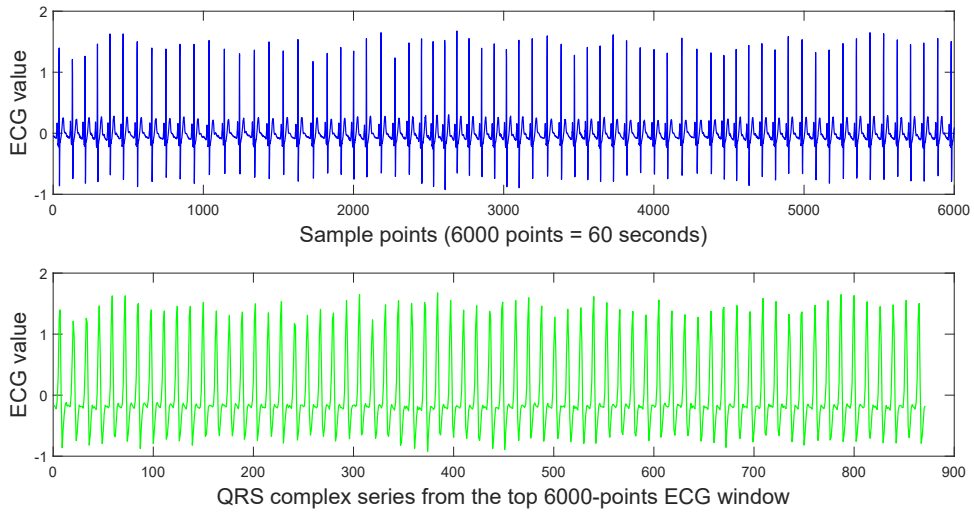

Fig. 3.7 The one-minute ECG signal (the top) and its QRS complex (the bottom)

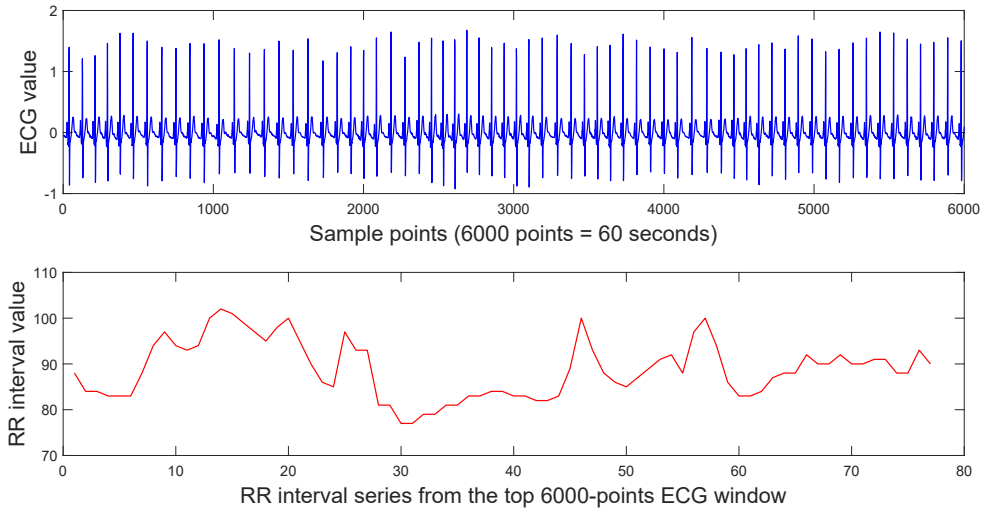

Fig. 3.8 The one-minute ECG signal (the top) and its RR interval series (the bottom)

3.2.4 Apneic Patterns and Feature Extraction from ECG Signals

The R-peaks are identified using the modified Pan-Tompkin algorithm. Here, a symmetric window of 120 ms is used to obtain all QRS (Q wave, R wave, S wave) around the R-peaks, and Fig. 1.8 illustrates the definition of Q wave, R wave, and S wave.

Fig. 3.7 illustrates an one-minute ECG signal and its QRS complex. The heart rate, known as the RR interval, is considered as the time difference of consecutive R peaks. Fig. 3.8 illustrates an one-minute ECG signal and its RR interval series. Besides RR interval series and the QRS complex, ECG-derived respiration (EDR) signals are also derived from ECG signals, and the EDR method of Physionet is used in this thesis. EDR signals are able to reflect respiratory activities between the electrodes and the heart since the electrodes are placed on the patient's thorax, and the electrical impedance changed during breathing is recorded.

Previous studies have stated that the heart rates of OSA patients are lower than the heart rates of healthy people [76]. Based on the bio-physiological criteria, we extract useful time-based features. Fig. 1.9 in Chapter 1 shows the ECG difference between the apnea window and the normal window, and it can be found there are difference in the R amplitude series and the RR interval series. Thus, from Fig. 3.9, the R amplitude series of the apnea event (the bottom wave) has more fluctuations, and the range of the values is also larger. As for the RR interval series shown in Fig. 3.10, the RR intervals in the normal event (the top wave) illustrates pseudo periodic patterns and provides larger values, while the RR intervals in the apnea event (the bottom wave) show an increase. In addition, the motion of the thoracic cavity influences the ECG signals, and respiratory muscles are activated during apnea [77]. Thus, in Fig. 3.11, the EDR signal in the bottom apnea window shows more fluctuations according to active respiratory muscles.

The changes in the time-domain lead to the change in the frequency-domain. According to the changes of ECG signals, EDR signals, and RR interval series in time-domain, it also can be found there are differences of ECG signals (Fig. 3.12), EDR signals (Fig. 3.13), and RR interval series (Fig. 3.14) in frequency-domain. Fig. 3.12 shows ECG difference in the frequency domain between the top normal event and the

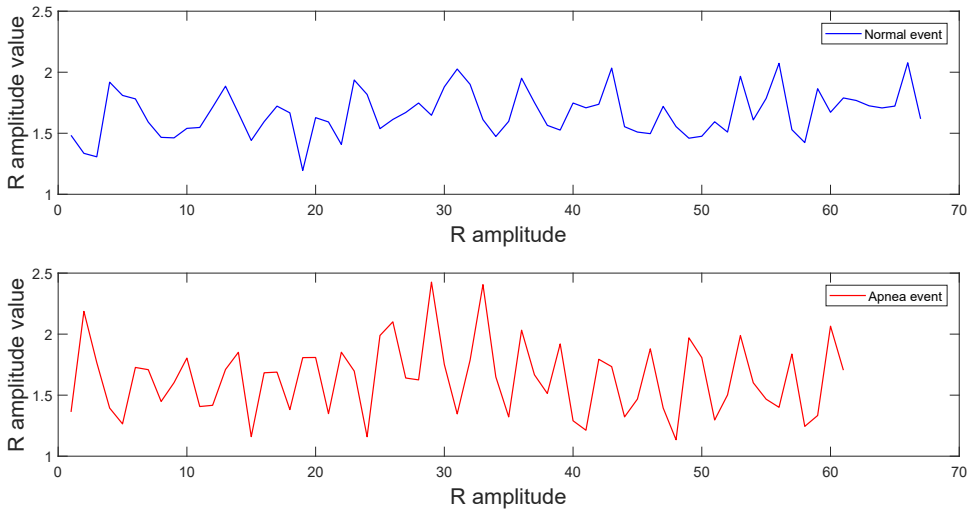

Fig. 3.9 The R amplitude difference between one apnea event and one normal event (the top wave is the normal R amplitude series and the bottom wave is the abnormal R amplitude series)

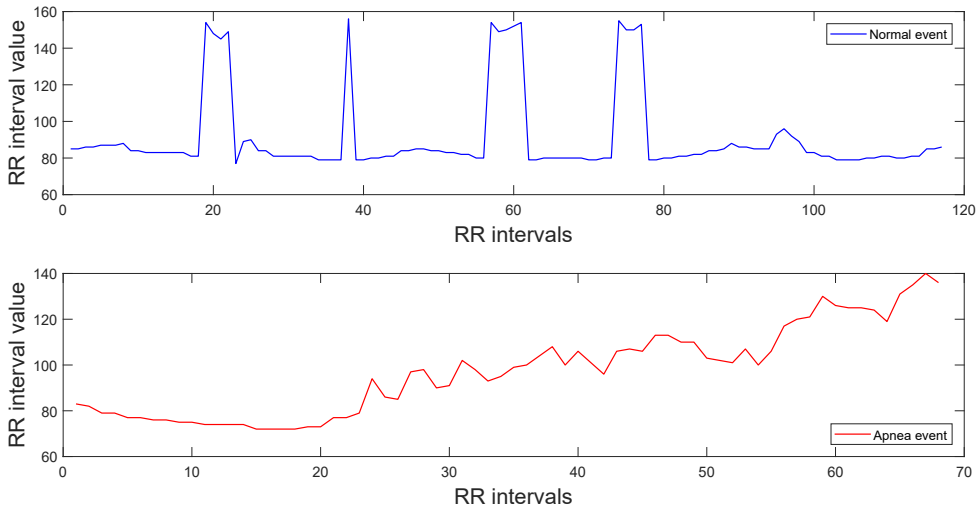

Fig. 3.10 The RR interval difference between one apnea event and one normal event (the top wave is the normal RR interval series and the bottom wave is the abnormal RR interval series)

bottom apnea event. There are larger fluctuations in the low-frequency band of the top normal event, and the frequency values in the high-frequency band in the bottom apnea event are lower than the values in the top normal event. Fig. 3.13 also illus-

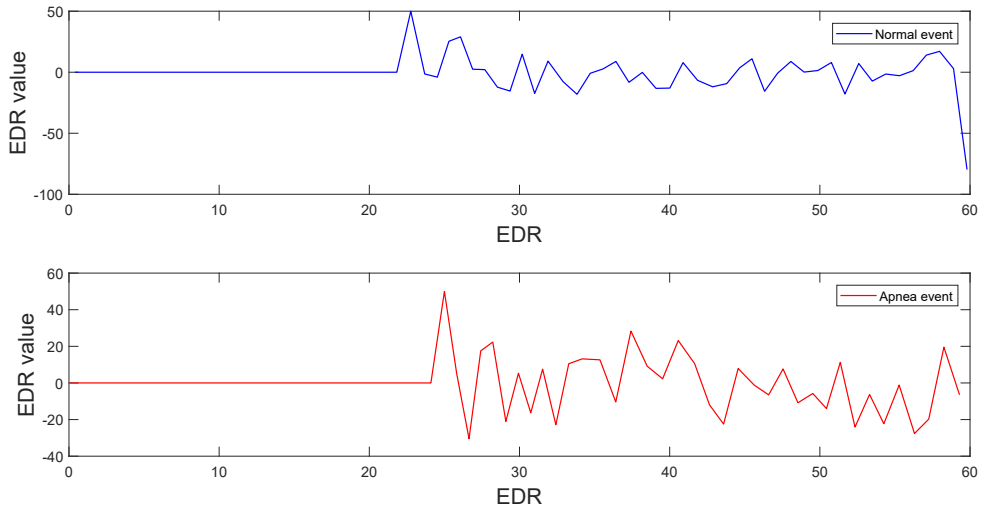

Fig. 3.11 The ECG derivation respiratory (EDR) difference between one apnea event and one normal event (the top wave is the normal EDR signal and the bottom wave is the abnormal EDR signal)

trates EDR difference in the frequency domain between the top normal event and the bottom apnea event the low-frequency band. The different very-low-frequency band are able to use in apnea classification. It also can be found that the difference in the low-frequency band in Fig. 3.14 by comparing the frequency-domain difference in the top normal event and the bottom apnea event. Frequency-domain features of the work are obtained by wavelet transformation and PSD.

In this study, three kinds of feature extraction methods were used to obtain 32 features from the ECG signal. *Preprocessing:* Firstly, the 0.05 - 40 Hz band-pass filter was applied to remove noise and take base-line correction, and the filter was 3rd Butterworth. Then, the R-peaks were found by the modified Pan-Tompkins algorithm. QRS signal was extracted by a symmetric window of 120 ms around the R-peaks. In this work, we used the heart rate correction method from [78]. Finally, the EDR signal was obtained from ECG recordings by the Physionet EDR method.

Table 3.5 Multi-domain features from ECG signals (The abbreviations are shown and detail methods are shown in Section 3.2.4)

Feature	Short description
NN50_RR	the later RR interval is more than the previous one by more than 50 ms
SDSD_RR	standard deviation of standard deviation of the RR interval
tSD_RR	standard deviation between the standard deviation at adjacent 30-second windows
std_RR	the standard deviation of RR interval
mean_RR	mean of each RR interval window
var	variance in each window
kur	kurtosis in each window
CV_EDR	a ratio of the standard deviation to the mean of each EDR window
SS	spectral spread
SD	spectral decrease
entropy_D1	Shannon's entropy in 1st detail coefficient level
entropy_D2	Shannon's entropy in 2nd detail coefficient level
entropy_D3	Shannon's entropy in 3rd detail coefficient level
entropy_D4	Shannon's entropy in 4th detail coefficient level
entropy_D5	Shannon's entropy in 5th detail coefficient level
entropy_D6	Shannon's entropy in 6th detail coefficient level
entropy_D7	Shannon's entropy in 7th detail coefficient level
var_D1	variance in 1st detail coefficient level
var_D2	variance in 2nd detail coefficient level
var_D3	variance in 3rd detail coefficient level
var_D4	variance in 4th detail coefficient level
var_D5	variance in 5th detail coefficient level
var_D6	variance in 6th detail coefficient level
var_D7	variance in 7th detail coefficient level
WSD_RR	Wavelet spectral density in RR series
max_$PSD_{0.03/0.5}$	dominant frequency in 0.03 - 0.5Hz in each RR window
mean_$PSD_{10/20}$	mean in 10 - 20Hz in each ECG window
mean_$PSD_{80/100}$	mean in 80 - 100Hz in each ECG window
SCrC_3_RR	3rd serial correlation coefficient in each RR interval window
SCrC_4_RR	4th serial correlation coefficient in each RR interval window
max_dia_kPCA	maximum of the diagonal matrix of KPCA
RP_2_PC	relative power of the second principal component

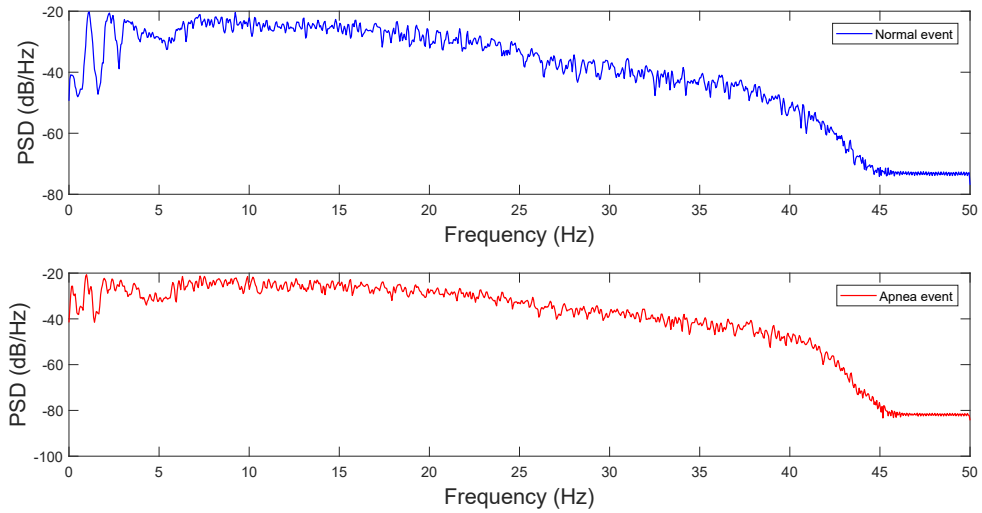

Fig. 3.12 The ECG difference in frequency domain between one apnea event and one normal event (the top wave is the ECG frequency wave of the normal event and the bottom wave is the wave of the abnormal event)

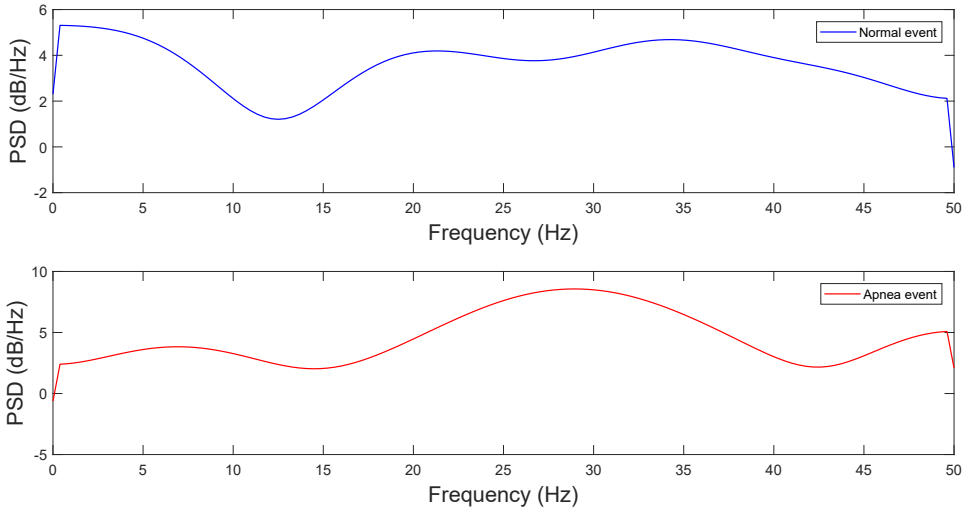

Fig. 3.13 The EDR difference in frequency domain between one apnea event and one normal event (the top wave is the EDR frequency wave of the normal event and the bottom wave is the wave of the abnormal event)

Multi-domain features: First of all, we obtained features using time-domain methods in an ECG window. These features included the number of pairs of adjacent RR intervals that the later RR interval was more than the previous one by more than 50

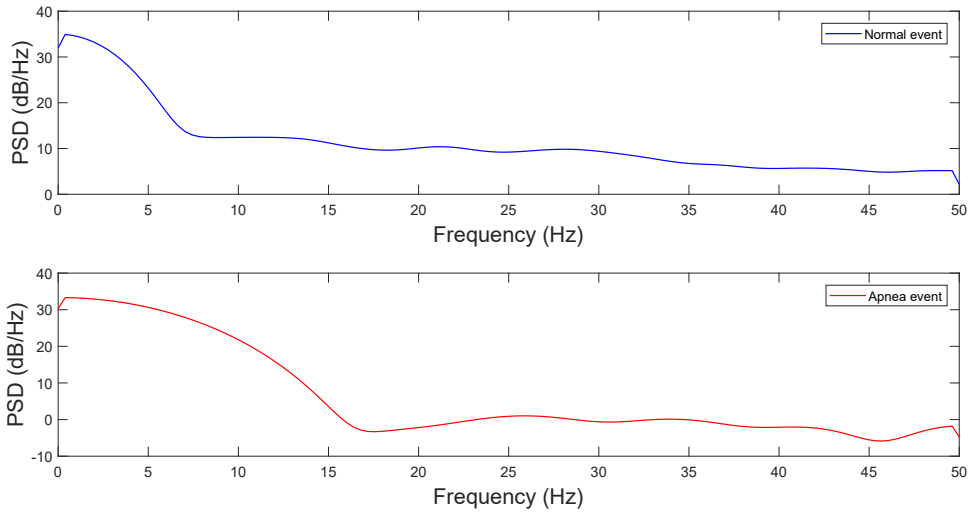

Fig. 3.14 The RR interval difference in frequency domain between one apnea event and one normal event (the top wave is the RR interval frequency wave of the normal event and the bottom wave is the wave of the abnormal event)

ms (NN50_RR), the standard deviation of the standard deviation of the RR interval (SDSD_RR), the standard deviation of the RR interval between the standard deviation at the first 30 seconds and the one at the second 30 seconds (tSD_RR), the standard deviation of RR interval (std_RR), variance and kurtosis of ECG windows (var and kur), mean of heart rate (mean_RR), and a ratio of the standard deviation to the mean of EDR signal (CV_EDR).

Second, in an ECG window, we use PSD and wavelet to extract frequency-domain features. We extracted spectral features such as spectral spread (SS) and spectral decrease (SD). Wavelet transformation was used with a Symlet wavelet with order three, and the level number was 7. Shannon's entropy (entropy_D1 to entropy_D7) and variance (var_D1 to var_D7) were computed using seven detail coefficient levels. Wavelet spectral density (WSD) was used to analyze the heart rate (WSD_RR). PSD was used to process RR interval and ECG signal. From RR interval, the dominant frequency was found in the 0.03 - 0.5 Hz frequency range ($\max\_PSD_{0.03/0.5}$). From

ECG signal, we extracted the mean of PSD in 10 - 20Hz and 80 - 100Hz respectively (mean_$PSD_{10/20}$ and mean_$PSD_{80/100}$).

Finally, two serial correlation coefficients were extracted from RR interval (SCrC_3_RR to SCrC_4_RR) in an ECG window. The serial correlation method can measure the correlation of an ECG window signal with a delayed copy of itself as a function of delay. Here, the kernel principal component analysis (KPCA) was employed in QRS recordings and the maximum of the diagonal matrix of KPCA (max_dia_kPCA) and the relative power of the second principal component (RP_2_PC) were calculated in an ECG window. In summary, the features extracted from ECG signals are shown in Table 3.5.

In the thesis, time-domain, frequency-domain, and non-linear analyses feature extraction methods are applied to extract 66 features from SaO_2, airflow, abdominal, thoracic, and ECG signals. The features are shown in Table 3.6. No.1 - No.12 features are extracted from SaO_2 signals, No.13 - No.22 features are extracted from airflow signals, No.23 - No.28 features are extracted from abdominal signals, No.29 - No.34 features are extracted from thoracic signals, and features No.35 - No.66 are extracted from ECG signals.

Table 3.6 Features extracted from SaO_2, airflow, abdominal signal, thoracic signal, and ECG signals (No.1 - No.12 from SaO_2 signals, No.13 - No.22 from airflow signals, No.23 - No.28 from abdominal signals, No.29 - No.34 from thoracic signals, and feature No.35 - No.66 from ECG signals, four dashed lines divided the table into five parts according to different kinds of signals)

Feature	Short description
1.med	median of window
2.MM2	maximum-to-minimum changes more than 2% of window

Continued on next page

Table 3.6 – continued from previous page

Feature	Short description
3.kur	kurtosis of window
4.var	variance of window
5.min	minimum of window
6.mean	mean of window
7.NumZC	number of zero-cross in window
8.comp	complexity of window
9.SD_1	short-term variability of each window
10.Bel98	time spent below the 98% maximum in window
11.Abo98	time spent above the 98% maximum in window
12.mean_$PSD_{0.016/0.05}$	mean in 0.016 - 0.05Hz frequency band in window
13.mean	mean in window
14.med	median in window
15.std	standard deviation in window
16.mean_$PSD_{0/0.1}$	mean in 0 - 0.1Hz frequency band in window
17.mean_$PSD_{0.4/0.5}$	mean in 0.4 - 0.5Hz frequency band in window
18.mean_D1	mean of 1st detail coefficient level
19.mean_D2	mean of 2nd detail coefficient level
20.mean_D3	mean of 3rd detail coefficient level
21.mean_A3	mean of approximation coefficient level
22.comp	complexity in window
23.sum_abs	summation of each absolute window

Continued on next page

Table 3.6 – continued from previous page

Feature	Short description
24.std_abs	standard deviation of each absolute each window
25.mean	mean of each window
26.mean_$PSD_{80/100}$	mean of the 80 - 100Hz frequency range in each window
27.mean_D2	mean of 2nd detail level
28.mean_D1	mean of 1st detail level
29.sum	summation of each window
30.med	median of each window
31.std	standard deviation of each window
32.mean	mean of each window
33.var	variance of each window
34.sum_$PSD_{80/100}$	summation of the 80 - 100Hz frequency band in each window
35.NN50_RR	the later RR interval is more than the previous one by more than 50 ms
36.SDSD_RR	standard deviation of standard deviation of the RR interval
37.tSD_RR	standard deviation between the standard deviation at adjacent 30-second windows
38.std_RR	the standard deviation of RR interval
39.var	variance in each window

Continued on next page

Table 3.6 – continued from previous page

Feature	Short description
40.kur	kurtosis in each window
41.mean_RR	mean of each RR interval window
42.CV_EDR	a ratio of the standard deviation to the mean of each EDR window
43.SS	spectral spread
44.SD	spectral decrease
45.entropy_D1	Shannon's entropy in 1st detail coefficient level
46.entropy_D2	Shannon's entropy in 2nd detail coefficient level
47.entropy_D3	Shannon's entropy in 3rd detail coefficient level
48.entropy_D4	Shannon's entropy in 4th detail coefficient level
49.entropy_D5	Shannon's entropy in 5th detail coefficient level
50.entropy_D6	Shannon's entropy in 6th detail coefficient level
51.entropy_D7	Shannon's entropy in 7th detail coefficient level
52.var_D1	variance in 1st detail coefficient level
53.var_D2	variance in 2nd detail coefficient level
54.var_D3	variance in 3rd detail coefficient level
55.var_D4	variance in 4th detail coefficient level
56.var_D5	variance in 5th detail coefficient level
57.var_D6	variance in 6th detail coefficient level
58.var_D7	variance in 7th detail coefficient level
59.WSD_RR	Wavelet spectral density in RR series

Continued on next page

Table 3.6 – continued from previous page

Feature	Short description
60.max_PSD$_{0.03/0.5}$	dominant frequency in 0.03 - 0.5Hz in each RR window
61.mean_PSD$_{10/20}$	mean in 10 - 20Hz in each ECG window
62.mean_PSD$_{80/100}$	mean in 80 - 100Hz in each ECG window
63.SCrC_3_RR	3rd serial correlation coefficient in each RR interval window
64.SCrC_4_RR	4th serial correlation coefficient in each RR interval window
65.max_dia_kPCA	maximum of the diagonal matrix of KPCA
66.RP_2_PC	relative power of the second principal component

3.3 Feature Extraction Methods on Another Bio-Signals

SaO_2, airflow, abdominal, thoracic, and ECG signals provide 66 features. Table 2.2 and Table 2.3 demonstrates ECG, SaO_2, airflow, EEG, chin EMG, thoracic, and abdominal signals are widely used to automatically detect OSA. Thus, EEG and chin EMG signals are used to provide other features. The aim of this thesis is to develop an automatic, low-cost detection for the doctors' diagnosis of apnea events, using highly accurate and easily accessible sensors. Although the literature review in Section 2.1.5

illustrates EEG signals are able to hold a good performance, EEG signals are always bad quality, and there is noise in EEG signals. The EEG Headset also makes the subject uncomfortable. In addition, Table 2.2 and Table 2.3 show a literature review on the OSA detection studies using single one bio-signal or multi-bio signals, and only a few studies use the chin EMG signal since chin EMG values are small, which leads to the changes in apnea events are not obvious. Thus, in this thesis, SaO_2, airflow, abdominal, thoracic, and ECG signals are used to provide multi-domain features.

In section 3.3.1 and 3.3.2, a brief explanation of feature extraction from EEG and chin EMG signals are introduced which gives some ideas when researchers adapt these signals for OSA detection in future.

3.3.1 Multi-Domain Features from EEG Signals

From EEG signals, we obtain features from the time-domain and frequency-domain. It has been shown that a decrease in δ wave (1-4Hz) is accompanied by an increase in θ wave (4-8Hz) and α wave (8-13Hz) [79]. We utilized wavelet transformation to decompose each EEG window and obtain approximated δ wave, θ wave and α wave. In an EEG window, we used level-8 wavelet transformation and Daubechies wavelets 2. Since the sample rate is 200 Hz, D8 and D7 (0.78125 - 3.125Hz) were considered as the δ wave, D6 (3.125 - 6.25Hz) was considered as θ wave and D5 (6.25 - 12.5Hz) was considered as the α wave. The A9 wavelet coefficient was removed, and a band-pass filter with cut-off frequencies of 0.5 - 45 Hz was applied to remove noise. The EEG feature set included: mean of absolute value, maximum, minimum, standard deviations and variance of EEG windows of δ wave, θ wave and α wave. In summary, the features extracted from EEG signals are shown in Table 3.7.

Table 3.7 Features extracted from EEG signals (The abbreviations are shown and detail methods are shown in Section 3.3.1)

Feature	Short description
mean	mean of each absolute EEG window
std	standard deviation of each EEG window
max	maximum of each EEG window
var	variance of each EEG window
min	minimum of each EEG window
std_δ	standard deviation of each δ wave
max_δ	maximum of each δ wave
std_α	standard deviation of each α wave
max_α	maximum of each α wave
std_θ	standard deviation of each θ wave
max_θ	maximum of each θ wave
var_δ	variance of each δ wave
min_δ	minimum of each δ wave
var_α	variance of each α wave
min_α	minimum of each α wave
var_θ	variance of each θ wave
min_θ	minimum of each θ wave

3.3.2 Multi-Domain Features from Chin EMG Signals

In the part, we obtain time-domain and frequency-domain features from chin EMG signals. There are bursts of EMG activity in apnea periods [58], which cause changes in the frequency domain. In addition, bursts result in saw-tooth morphology, which causes higher values in the low-frequency band. More chin activity causes an increase in the high-frequency band. In a chin EMG window, power spectrum density of the EMG signal was first obtained, and then the maximum, mean, summation and median for each window were calculated. More specifically, the frequency associated with the maximum value (freq - max of PSD) and the summation within 80 - 100Hz (sum of $PSD_{80/100}$) were obtained. The decomposition of a window was performed using the wavelet transform, and for this work, it was determined to include the variance and

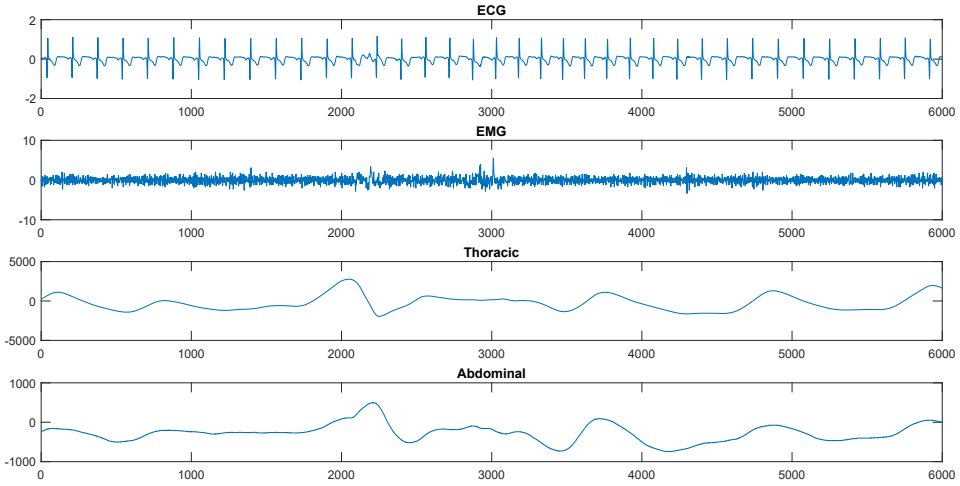

Fig. 3.15 Example of ECG, EMG, thoracic, and abdominal signals from the DREAMS dataset

standard deviation of D1 in the feature set. In summary, the features extracted from chin EMG signals are shown in Table 3.8.

Table 3.8 Features extracted from EMG signals (The abbreviations are shown and detail methods are shown in Section 3.3.2)

Feature	Short description
max_PSD	maximum of each EMG PSD window
freq_max_PSD	the frequency associated with the maximum value of each EMG PSD window
mean_PSD	mean of each EMG PSD window
sum_PSD	summation of each EMG PSD window
med_PSD	median of each EMG PSD window
$\text{sum\_PSD}_{80/100}$	summation within 80 - 100Hz band
var_D1	variance in 1st detail coefficient level
std_D1	standard deviation in 1st detail coefficient level

3.3.3 Experiments and Results

The DREAMS dataset is used to evaluate time-domain and frequency-domain features from EEG and chin EMG signals. The DREAMS Apnea Database consists of 12 PSG signals. Each recording consists of ECG, airflow, SaO_2, EEG, chin EMG, abdominal, and thoracic signals (shown in Fig. 3.15 and 3.16), and lasts for approximately 9 hours. The PSG data are recorded at 200 Hz sampling frequency. These data recordings are annotated based on the respiratory events by an expert. Each signal is divided into 10-second windows since this duration is the minimum obstruction duration to detect an OSA. Finally, we obtain 17 kinds of EEG features and eight kinds of chin EMG features in the DREAMS dataset.

3.4 Summary

In this chapter, time-domain, frequency-domain, and non-linear analyses feature extraction methods are proposed, and ECG, SaO_2, airflow, EEG, chin EMG, thoracic,

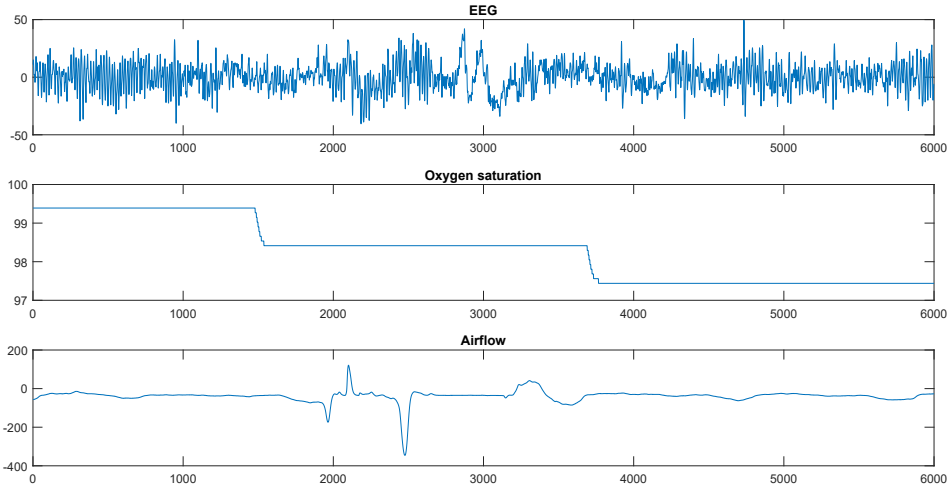

Fig. 3.16 Example of EEG, SaO_2, and airflow signals from the DREAMS dataset

and abdominal signals are used to provide multi-domain features. Features are related to apneic bio-physiological criteria, and they are known to be able to provide good classification performance. ECG signals of OSA patients show distinct patterns in RR intervals, and ECG signals are influenced by the motion of the thoracic cavity. In EEG signals, a decrease in δ wave is accompanied by an increase in θ and α waves. There are bursts of chin EMG activity in apnea events. Sudden downturns and relative sudden recoveries in SaO_2 signals are linked to apnea. There is a decrease in airflow signals, and two missed breaths mean the recurrence of these apnea events is every two normal breaths at most. During apnea periods, the respiratory muscles become active to restore breathing.

The main aim of this thesis is to construct a clear need for an automatic, low-cost detection for the doctors' diagnosis of apnea events, using highly accurate and easily accessible sensors. In order to achieve this aim, the first step is to select highly accurate, easily accessible, and more stable bio-signals. ECG, SaO_2, airflow, thoracic, and abdominal signals are obtained, and 66 kinds of time-domain, frequency-domain, and non-linear analyses features are obtained (shown in Table 3.6). 17 kinds of EEG features (shown in Table 3.7) and 8 kinds of chin EMG features (shown in Table 3.8). However, according to the requirements of highly accurate and easily accessible sensors, EEG and chin EMG signals are not selected to provide multi-domain features.

Chapter 4

Feature Selection Method

4.1 Introduction

A total of features increases the training time of classifier, as time is wasted in processing redundant or potentially detrimental features, which limits real-time detection applications. Also, the total number of features influence the high classification accuracy of the classification algorithm. Thus, before the classification stage, feature selection plays an important role. Feature selection can also prevent overfitting of training and reducing computational load as it requires less signal processing. A small subset of data is obtained by feature selection, and the subset has high-discriminatory power, which maintains performance and shortens the training process time. For the latter feature subset, which provides the most information for discriminating between two classes, we propose a two-stage procedure (Stage 1: the statistical analysis stage and Stage 2: the final feature subset selection stage using machine learning methods). In Stage 1, the two statistical analysis is used to filter out redundant features and select the features with good information. Then, in Stage 2, machine learning methods are used to further validate the features from Stage 1 and show only the most im-

portant features. The Sleep Heart Health Study (SHHS) is used to provide different bio-signals. The dataset was composed of 6441 people over than 40 years old. Each person has two events, one for normal and the other for apnea. Even in patients with apnea, they breathe normally in most of the night.

The novelty of the proposed feature selection method is as follows: 1. Two different statistical methods are used to remove relevant features. One kind of feature needs to meet the requirements of both methods and then is put into the feature subset. 2. Only relying on statistical methods is not adequate to select the features with the most powerful information. To achieve this aim, we use machine learning methods and different kernels and parameters are included. One of the feature subsets which has the most stable performance is considered as the final feature subset.

This chapter is organized as follows. The two-stage feature selection is discussed in Section 4.2. The detail methods of experiments and results are discussed in Section 4.3. A chapter summary is drawn in Section 4.4.

4.2 Two-Stage Feature Selection Method

The total features increase the processing time since classifiers waste time to process redundant features. Classifiers are not unable to hold better performance because of these redundant features. Hence, a feature selection is used to remove redundant features before a classification stage, and it prevents over-fitting and reduces computational load.

For the feature selection, we use a two-stage procedure: Stage 1: the statistical analysis stage and Stage 2: the final feature subset selection stage using machine learning methods, shown in Fig. 4.1. In the first stage, according to the statistical performance of different features, redundant features are removed from the feature set

and other features are put into different classes [21] [80] [50] [81]. Then, we evaluate the performance of different feature classes with a trained support vector machine (SVM) model. In this case, the hill-climbing method, and k-fold cross-validation are used. In this step, k is equal to 5. The feature set is reduced and the best feature class with good performance is selected via compared of sensitivity, specificity and accuracy.

4.2.1 Statistics Analysis

To determine the significance of each feature, ANOVA and rank-sum test are used. ANOVA [80] [82] can analyse differences between group means and associated labels by calculating *SSE*, *SST*, *SSG*, *MSG*, *MSE*, and F-ratio. Then p-value is computed. In the rank-sum test [81] [83], the aim is to confirm that the two groups are independent. Two values are randomly selected from the first group and the second group, respectively and these two values are compared. The rank-sum test also obtains the p-value.

The flowchart of calculating the importance of each feature using statistical analysis methods is shown in Fig. 4.2. At the beginning of the statistical analysis stage, we set a parameter, N, for each feature, which means the importance level of each feature and N is equal to zero. In Fig. 4.2, the first derived feature from the first patient is analysed by the ANOVA, and the rank-sum test method and a pair of p-values is obtained. Here, we use a simple threshold (p-value of ANOVA < 0.05) to determine the feature is positive. This threshold indicated this kind of feature is differently significant between apnea and normal groups in one patient. Similarly, in the rank-sum test, there is a simple value (p-value = 1) to determine the positive or negative feature in this patient. If the feature from the first patient is determined as the positive feature by

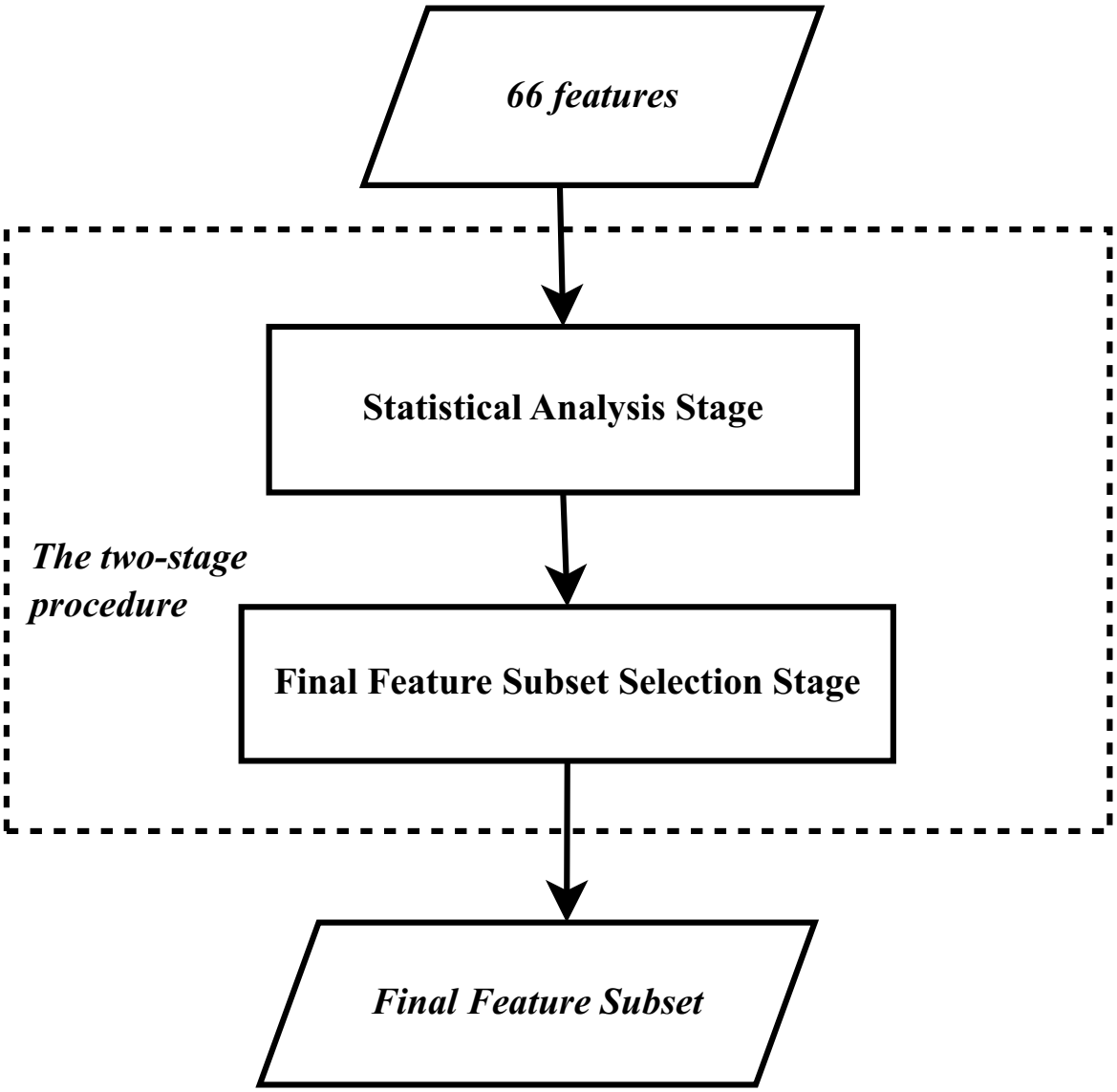

Fig. 4.1 Two-stage feature selection procedure, including Stage 1: the statistical analysis stage and Stage 2: the final feature subset selection stage using machine learning methods

both the ANOVA and the rank-sum test, this feature is considered to be significant between apnea events and normal events in the first patient, and N plus one in this iteration and the result will be considered as the N of the next iteration, which means N is changed with the iteration progress. On the other hand, if one of the p-values from the feature from the first patient is negative, the feature is not considered to be significant between apnea events and normal events in the first patient. Then, the

feature from the next patient is taken by the ANOVA and the rank-sum test method and a pair of p-values is obtained. This process is repeated until all patients' PSG signals are used to provide all pairs of p-values for the feature. Finally, as for the first derived feature, N is obtained and considered as the importance of the feature. In the next iteration, the next derived feature from the first patient is analysed by the ANOVA and the rank-sum test method and a pair of p-values is obtained. Finally, the importance (N) of the next derived feature is obtained. After all features are analyses by the ANOVA and the rank-sum test, the importance list of all features is obtained.

Using the importance list, features is found to be with good discrimination between two groups and selected into the feature subset. We set the λ_{feature} as the number of the importance for each feature, and the maximum value of λ_{feature} is equal to the number of processed patients' PSG signals. In this study, for selecting useful features, we set a threshold, which is equal to the half of the processed PSG signals. One feature is put into the selected subset if its λ_{feature} is more than the threshold. The feature subset is put into the second stage to determine the final feature subset using machine learning methods.

4.2.2 Support Vector Machine Selection

To confirm the best feature subset, SVM is used to select the best classes with the most relevant information in separating apnea and normal and the hill-climbing algorithm is also used. SVM is introduced by Vapnik [60] and a classification method and regression analysis. It is a binary classifier, and in an N-dimensional space, it constructs an optimal separating hyperplane (OSH). The input vector is transformed via nonlinear mapping into a K-dimensional space from the N-dimensional space.

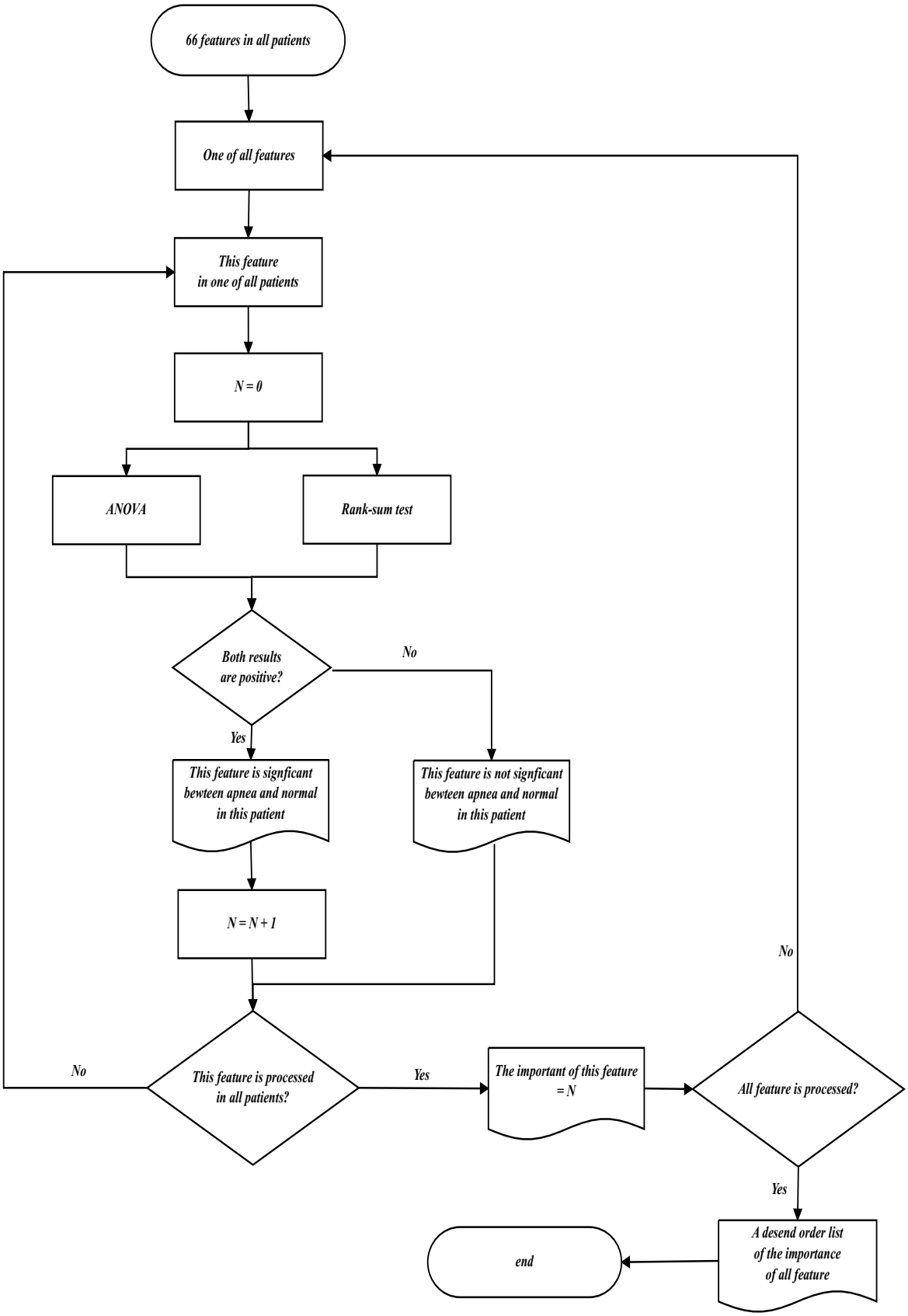

Fig. 4.2 The flowchart of calculating the importance level of each feature

Selecting a kernel function $K(x_i,x_j)$ affects the performance of the SVM in the construction stage. In this study, we evaluate three kernels and select the best one. Three kernels include Linear, Polynomial and Radial basis function (RBF), shown in Table 2.4. To select the best performance of three kernel functions, we use three indexes, accuracy, sensitivity and specificity and three measures to calculate the results of the SVM. The final feature subset is determined by comparing three indexes.

$$sensitivity = \frac{TP}{TP + FN} \tag{4.1}$$

$$specificity = \frac{TN}{TN + FP} \tag{4.2}$$

$$accuracy = \frac{TP + TN}{P + N} \tag{4.3}$$

where FN is the false negative, and TN is the true negative, and N is the condition negative, and FP is the false positive, and TP is the true positive, and P is the condition positive.

The hill-climbing algorithm is used, shown in Fig. 4.3. Initially, the features in the top class are fed into SVM models, and performance is recorded in this iteration. Then the features in the next class are added to put into SVM models, and performance in the second iteration is recorded. The process is repeated until all classes are added, and all performance in iterations is recorded.

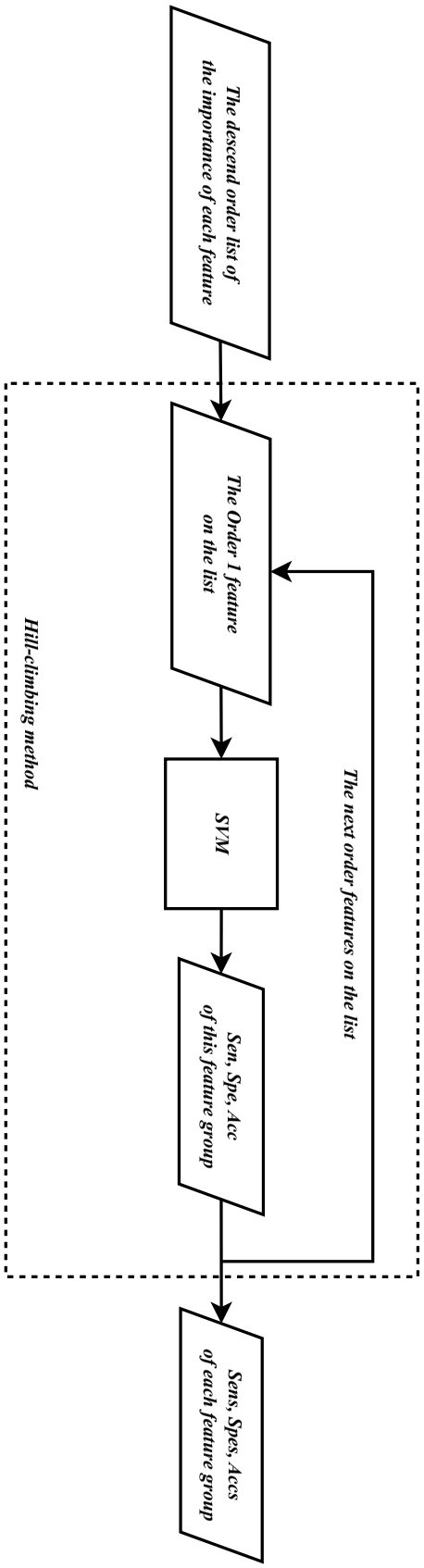

Fig. 4.3 The hill-climbing method based on the descending order list

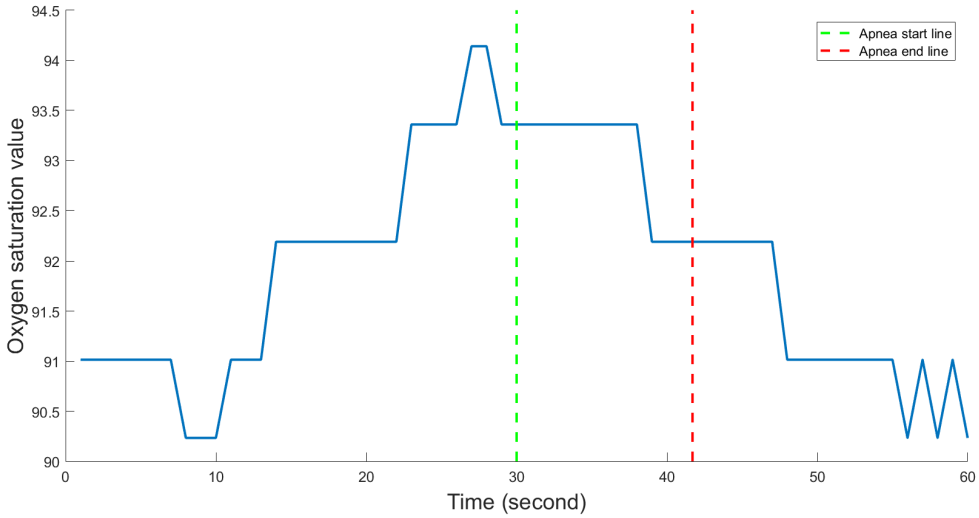

Fig. 4.4 One apnea event duration in SaO_2 signal, and the duration is between the apnea start line and the apnea end line

4.3 Experiments, Results, and Discussion

4.3.1 Collection of Sleep Studies

To construct our apnea classification system, we used the Sleep Heart Health Study (SHHS) provided by the National Heart Lung & Blood Institute. The dataset was composed of 6441 people over than 40 years old. The dataset consisted of respiration events, airflow, finger-tip pulse oximetry, ambient light, heart rate, ECG and body position. Two respiration events were abdominal and thoracic signals with 10 Hz. Pulse oximetry was at 1 Hz, while airflow was sampled at 10 Hz. ECG was at 125 Hz, and heart rate (RR intervals) with 1 Hz was derived from the ECG. Respiratory events were labelled by a human expert, and the sleep physiologist also labelled the event duration for each event, as shown in Fig. 4.4. In this study, ECG, abdominal, thoracic, airflow signals, and SaO_2 signals were used. One example of a PSG signal is shown in Fig. 4.5. In the next section, feature extraction methods are shown.

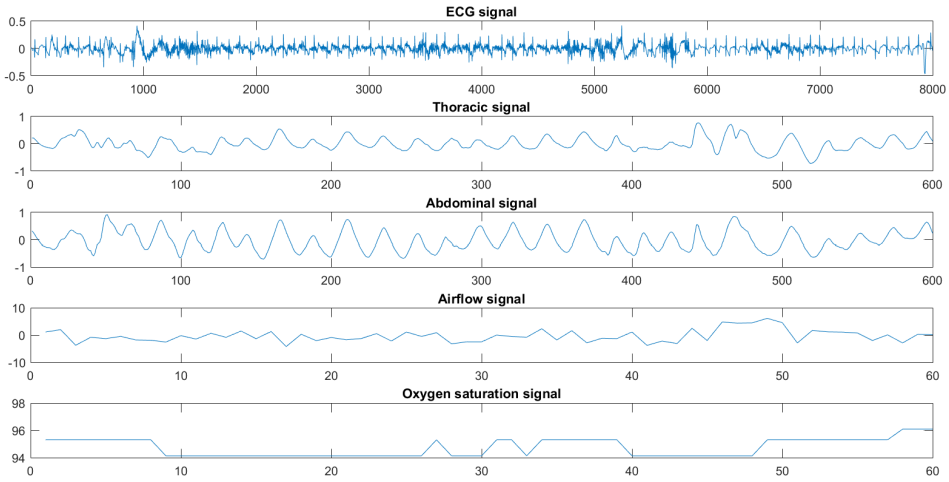

Fig. 4.5 Example of PSG including ECG, abdominal and thoracic signals, airflow and oxygen saturation signals

4.3.2 Feature Extraction Experiments and Results

We obtain 66 kinds of multi-domain features in the SHHS dataset using different feature extraction methods shown in Chapter 3. Nos. 1 - 12 feature are extracted from SaO_2 signals, and Nos. 13 - 22 features are extracted from airflow signals and Nos. 23 - 28 features are extracted from abdominal signals, and Nos. 29 - 34 features are extracted from thoracic signals and Nos. 35 - 66 features are extracted from ECG signals (Table 3.6).

4.3.3 Feature Selection Experiments and Results

Table 4.1 shows the λ_{feature} of each feature. At the beginning of the statistical analysis stage, we set a parameter, N, for each feature, which means the importance level of each feature and N is equal to zero. The first feature is median (med) from SaO_2 signals. To obtain the λ_{feature} of the med feature, first, the med results from the first patient is analysed by the ANOVA and the rank-sum test, and a pair of p-values

is obtained. If the med feature from the first patient is confirmed as the positive feature by both the ANOVA and the rank-sum test, the med feature is considered to be significant between apnea events and normal events in the first patient, and the N of this feature plus one in this iteration and the result will be considered as the N of this feature in the next iteration, which means the N of this feature is changed with the iteration progress. Then, the med feature from all patients is analysed, and all pairs of p-values are obtained. Finally, as for the med feature from SaO_2 signals, N is obtained and equal to 581, which means 581 pairs of No. 1 feature (med) are positive. The λ_{feature} of the med feature from SaO_2 signals is equal to 581.

In Table 4.1, λ_{feature} of feature No.10 is 1,509, which means that 1,509 pairs of the feature No.10 are positive. Similarly, λ_{feature} of feature No. $2 = 748$, which means that this feature has been identified as a positive feature for 748 patients. In this study, 1,574 patients' PSG signals are employed to derive features. To select useful features, we set a threshold, which is equal to 787 (the half of the processed PSG signals). If the λ_{feature} of a feature is ≥ 787, this feature is put into the selected feature set and highlighted with an asterisk (*) in Table 4.1. Finally, 19 features are selected based on λ_{feature} being ≥ 787. The aim of this stage is to discard noisy and irrelevant features and reduce the computational load of the algorithm.

To determine the final feature subset, the SVM algorithm was used. We put selected features into 4 classes {Class A, B, C, D} by comparing λ_{feature} of 19 selected features. The classes were based on the distribution of these features. Firstly, we easily detected Class D, which included five features, since there was a large distance between feature No.8 ($\lambda_{\text{feature}} = 1215$) and feature No.6 ($\lambda_{\text{feature}} = 1414$). Other classes also need to include five or four features since the number of the selected features is 19. Based on this criteria, Class A had five features, and Class B had five features, and Class C had four features (as shown in Table 4.2), and Class {Class A, B, C, D}

Table 4.1 Feature selection results in the experiment with duration using statistics analysis in the SHHS dataset (No.1 - No.12 from oxygen saturation signals, No.13 - No.22 from airflow signals, No.23 - No.28 from abdominal signals, No.29 - No.34 from thoracic signals, feature No.35 - No.66 from ECG signals, four dashed lines divided the table into five parts according to different kinds of signals)

Feature	λ_{feature}	Feature	λ_{feature}
1.med	581	34.sum_$PSD_{80/100}$	341
2.MM2	748	35*.NN50_RR	1432
3.kur	651	36.SDSD_RR	351
4.var	582	37.tSD_RR	563
5.min	661	38.std_RR	330
6*.mean	1414	39.var	678
7*.NumZC	997	40.kur	671
8*.comp	1215	41.mean_RR	637
9.SD_1	774	42.CV_EDR	168
10*.Bel98	1509	43*.SS	1477
11*.Abo98	831	44*.SD	1178
12*.mean_$PSD_{0.016/0.05}$	1565	45*.entropy_D1	1497
13.mean	136	46*.entropy_D2	1505
14.med	149	47*.entropy_D3	1522
15.std	167	48*.entropy_D4	1529
16.mean_$PSD_{0/0.1}$	427	49*.entropy_D5	1527
17.mean_$PSD_{0.4/0.5}$	417	50*.entropy_D6	1529
18.mean_D1	12	51*.entropy_D7	1497
19.mean_D2	17	52.var_D1	606
20.mean_D3	18	53.var_D2	645
21.mean_A3	131	54.var_D3	668
22.comp	635	55.var_D4	700
23*.sum_abs	1475	56.var_D5	660
24.std_abs	474	57.var_D6	626
25.mean	590	58.var_D7	610
26*.mean_$PSD_{80/100}$	947	59.WSD_RR	350
27.mean_D2	31	60.max_$PSD_{0.03/0.5}$	352
28.mean_D1	26	61.mean_$PSD_{10/20}$	711
29.sum	23	62.mean_$PSD_{80/100}$	586
30.med	446	63.SCrC_3_RR	392
31.std	428	64.SCrC_4_RR	336
32.mean	16	65*.max_dia_kPCA	1526
33.var	363	66.RP_2_PC	558

Table 4.2 Feature classes via the distribution of λ_{feature}

Feature No.	Class A 1526-1574	Class B 1491-1525	Class C 1301-1490	Class D 787-1300
SaO_2	12	10	6	7 8 11
Airflow				
Abdominal			23	26
Thoracic				
ECG	48 49 50 65	45 46 47 51	35 43	44

were employed in the hill-climbing method. Features in Class A had larger a λ_{feature} values, and they were highly correlated with the difference between apnea and normal events. The hill-climbing algorithm was used to confirm the best feature subset among these classes. We considered different kernels with different parameters since kernels affected the performance of the SVM algorithm. In this study, we evaluated linear (R), polynomial (R and d) and RBF (R and σ).

In this experiment, the data was obtained after the under-sampled balance method and removed outliers. Even in patients with apnea, most of the night, they are breathing normally. Therefore, the situation is that most of the sleep minutes correspond to normal breathing. The aim of this study is to improve the identification of apnea events (the rare minority group). Machine learning methods tend to produce unsatisfactory classifiers when facing this imbalanced database because they have a bias to the normal group (the majority group). They tend only to predict normal events. The features of the apnea events (the minority class) are treated as noise and often ignored. Thus, in this study, the aim of using the under-sampling method is to increase the frequency of minority and decrease the frequency of the majority class. Both apnea and normal events are considered as useful events, not noise.

The total 66 kinds of features were put into SVM models with different kernels, and parameters and the performance of total 66 features was considered as the standard. The 5-fold cross-validation method was used in this step. For example, in the first

fold, the features from No. 1 - No. 1560 (1/5 * 1574) patients were considered the training set, while the features No. 1561 - No. 1574 patients were considered as the testing set. Then, the performance of the first fold was recorded. Finally, we compared sensitivity, specificity, and accuracy to confirm the best feature subset, and all performance was shown in Table 4.3. The performance was bold if its accuracy was more than 70%. From Table 4.3, under all features, the performance of the SVM (RBF, $\sigma = 25$, $R = 0.2$) with all kinds of features consisted of sensitivity (92.26%) and specificity (63.75%), and this performance was considered as the standard and compared with the performance of other kernels with parameters. For the stability purpose, we saw that Class ABC and Class ABCD were better than Class A, Class AB, and all features. From Table 4.3, Class ABCD was more stable than Class ABC, but Class ABC held the best accuracy (79.06%). Thus, the AUC was used to evaluate Class ABC and Class ABCD. A comparison of AUC under different feature subsets was appreciated in Table 4.4. The AUC of Class ABCD was better than the AUC of Class ABC, and it can be found that Class ABCD had better effectiveness and robustness, and Class ABCD was considered as the feature subset. In the final feature subset, Feature No. 6, 7, 8, 10, 11, 12 were from SaO_2 signals, and Feature No. 23, 26 were from abdominal signals, and Feature No. 35, 43, 44, 45, 46, 47, 48, 49, 50, 52, 65 were from ECG signals.

4.3.4 Discussion

We attained the Class ABCD was the feature subset with high discrimination, including 19 features. The most widely used index to diagnose OSA is that oximetry value is less than a certain threshold or the cumulative time spent below a threshold. In oximetry values, sudden downturns and recoveries affect the frequency band. No. 12 is the mean of PSD within the 0.016 - 0.05Hz frequency range (mean_$PSD_{0.016/0.05}$).

Table 4.3 Sensitivity (%), specificity (%) and accuracy (%) based on the hill-climbing method using SVM models with different kernel functions and parameters in the experiment with duration

Kernels	R	Class A			Class AB			Class ABC			Class A-D			All features		
		Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe	Acc
	0.2	**99.31**	**40.84**	**70.07**	**98.94**	**50.94**	**74.94**	**98.85**	**54.06**	**76.46**	**98.15**	**56.91**	**77.53**	93.99	07.25	50.62
RBF σ=1	1	**99.12**	**43.10**	**71.11**	**98.90**	**51.87**	**75.39**	**98.68**	**54.92**	**76.80**	**98.24**	**57.88**	**78.06**	93.70	11.66	52.68
	10	**98.82**	**47.96**	**73.39**	**98.72**	**54.12**	**76.42**	**98.49**	**56.21**	**77.35**	**97.61**	**59.98**	**78.80**	93.14	14.15	53.64
	0.2	**99.06**	**44.87**	**71.96**	**99.09**	**46.42**	**72.75**	**99.15**	**45.85**	**72.50**	**97.58**	**59.99**	**78.78**	**91.76**	**62.07**	**76.92**
RBF σ=5	1	**99.12**	**45.08**	**72.10**	**99.01**	**46.28**	**72.65**	**99.34**	**46.54**	**72.94**	**97.21**	**60.36**	**78.79**	**91.39**	**63.25**	**77.32**
	10	99.27	40.08	69.68	**99.08**	**48.99**	**74.04**	**98.63**	**54.41**	**76.52**	**97.82**	**58.69**	**78.26**	**89.41**	**65.43**	**77.42**
	0.2	99.96	19.78	59.87	99.84	31.44	65.64	99.80	33.55	66.67	**99.63**	**43.77**	**71.70**	**92.26**	**63.75**	**78.00**
RBF σ=25	1	99.96	24.15	62.06	99.75	35.53	67.64	99.74	37.24	68.49	**98.71**	**49.18**	**73.94**	**91.11**	**64.49**	**77.80**
	10	99.41	39.48	69.45	**99.26**	**42.35**	**70.81**	**99.34**	**42.87**	**71.10**	**96.19**	**60.95**	**78.57**	**91.49**	**64.44**	**77.96**
	0.2	**98.68**	**45.45**	**72.07**	**98.55**	**52.64**	**75.60**	**98.62**	**54.84**	**76.73**	**97.58**	**59.63**	**78.60**	14.01	91.35	52.68
Poly d=2	1	**98.79**	**45.43**	**72.11**	**98.47**	**52.70**	**75.58**	**98.48**	**55.95**	**77.21**	**97.75**	**59.07**	**78.41**	11.99	91.43	51.71
	10	**98.09**	**49.56**	**73.83**	**94.10**	**56.01**	**75.06**	**98.13**	**57.59**	**77.86**	47.75	64.68	56.21	1.41	98.98	50.20
	0.2	99.70	33.62	66.66	3.73	96.74	50.24	0	99.42	49.71	0	99.48	49.74	0	1	50.00
Poly d=3	1	95.91	38.44	67.18	1.02	99.65	50.34	0	99.42	49.71	0	99.84	49.97	0	1	50.00
	10	**99.35**	**40.68**	**70.01**	21.01	89.92	55.47	0	99.99	49.99	0	1	50.00	0	1	50.00
	0.2	0	1	50.00	3.67	95.92	49.80	0	99.26	49.63	0	99.98	49.99	0	1	50.00
Poly d=4	1	0	99.99	49.99	0	99.92	49.96	0	99.97	49.98	0	99.95	49.98	94.24	7.21	50.72
	10	0	1	50.00	1	00.73	50.36	0	99.93	49.96	0	99.94	49.97	0	1	50.00
	0.2	99.92	20.30	60.11	99.67	35.00	67.34	**99.46**	**43.94**	**71.70**	**96.01**	**61.00**	**78.50**	**90.41**	**64.69**	**77.55**
Linear	1	99.91	20.31	60.11	99.65	35.05	67.35	**99.01**	**46.86**	**72.94**	**95.80**	**61.00**	**78.40**	**90.62**	**64.67**	**77.65**
	10	99.92	20.29	60.10	99.76	34.75	67.26	**97.60**	**60.52**	**79.06**	**96.30**	**60.96**	**78.63**	33.31	82.37	57.84

Table 4.4 AUC (%) computed on SVM models with kernels and parameters using Class ABC and ABCD in the experiment with duration

Kernels	R	Class ABC	Class ABCD
RBF $\sigma = 1$	0.2	76.45	77.53
	1	76.80	**78.06**
	10	77.35	**78.79**
RBF $\sigma = 5$	0.2	72.50	**78.78**
	1	72.94	**78.78**
	10	76.52	**78.25**
RBF $\sigma = 25$	0.2	66.67	71.70
	1	68.49	73.94
	10	71.10	**78.57**
Poly d=2	0.2	76.73	**78.60**
	1	77.21	**78.41**
	10	77.86	56.21
Poly d=3	0.2	49.71	49.74
	1	49.71	49.92
	10	49.99	50.00
Poly d=4	0.2	49.63	49.99
	1	49.98	49.97
	10	49.96	49.97
Linear	0.2	71.70	**78.50**
	1	72.93	**78.40**
	10	79.06	**78.63**

No. 6 is the mean of the window (mean), and No. 7 is the number of zero-cross in the window (NumZC), and No. 8 is the complexity (comp). No.10 and No.11 are the time spent above and blew the 98% maximum (Bel98 and Abo98). No. 12 reflects the change of the frequency band, and No. 6, 7, 8, 10, and 11 reflect the change of sudden downturns and recoveries in time-domain. In abdominal signal, No. 23 is the summation of absolute window (sum_abs), and No. 26 is the mean in the 80 - 100Hz frequency range (mean_$PSD_{80/100}$), and these features are related to active abdominal muscles during sleep apnea events. During apnea events, patients with OSA show beat-to-beat variation at lower heart rates relative to healthy subjects. The lower heart rate and the multi-domain change lead to these selected ECG features. No. 35 is the

number if pairs of adjacent RR intervals where the later RR interval is more than 50 ms than the previous one (NN50_RR). Nos. 43 and 44 are spectral spread (SS) and spectral decrease (SD) respectively. Nos. 45-51 are Shannon's entropy (entropy_D1 to entropy_D7) using seven detail coefficient levels. No. 65 is the maximum of the diagonal matrix of KPCA (max_dia_kPCA).

4.4 Summary

In the chapter, to select the features with good discrimination, we develop the two-stage procedure (Stage 1: the statistical analysis stage and Stage 2: the final feature subset selection stage using machine learning methods). There are 66 kinds of features from ECG, SaO_2, airflow, thoracic and abdominal signals provided by the SHHS dataset.

In the first stage, two statistics analysis, the ANOVA and the rank-sum test are applied to remove redundant features for improving the performance and reducing the processing time. Each feature from each patient has a pair of p-values. There are two statistical analysis criteria (the p-value of ANOVA < 0.05 and the p-value of the rank-sum test $= 1$), which are used to determine significant features. λ_{feature} is set as the number of positive pairs for each feature. The threshold is equal to half of the number of processed signals. If the λ_{feature} of a feature is more than the threshold, this feature is put into the feature subset. Finally, 19 features are considered as the significant features in the first stage, and these selected features are asterisk in Table 4.1.

In the second stage, the feature subset from the former phase is divided into four classes {A, B, C, D} by comparing the distribution of the λ_{feature} set of selected features. Features in different classes are put into an SVM classifier using the hill-climbing

algorithm, and different kernels (linear, RBF and polynomial) of SVM models are conducted with different parameters (σ, d and R). The sensitivity, specificity, accuracy, and AUC are used to evaluate the performance of different SVM models. Finally, Class ABCD is the best feature subset since it is more stable than other classes. Class ABCD consists of 19 features from ECG, SaO_2, and abdominal signals. These features are able to reflect the apneic bio-physiological changes, such as sudden drops and fast recoveries in SaO_2 signals, active abdominal muscles during sleep apnea events, and lower heart rate in ECG signals.

Chapter 5

Classification Based on Duration Parameter and Evaluation of The Time-Window Method

5.1 Introduction

Obstructive sleep apnea (OSA) is the most common breathing disorder during sleeping, which is caused by repeated partial or total obstruction of the upper airway [6]. The collapse of the upper airway is a 10-second or more standstill. OSA is a risk for several complications, such as hypertension and cardiac disease. In order to diagnose OSA, the polysomnography (PSG) is considered as the gold standard. PSG records overnight electrocardiogram (ECG), electromyogram (EMG), electrooculogram (EOG), electroencephalogram (EEG), airflow, respiratory effort, oxygen saturation (SaO_2), body position and snore. It is a time-consuming and difficult task for expert physicians to diagnose OSA. Thus, the automatic detection of OSA is of interest.

The automatic detection generally consists of the feature extraction stage, the feature selection stage and the classification stage shown in Fig. 1.13. In Chapter 3, ECG, SaO_2, airflow, thoracic, and abdominal signals are obtained since these bio-signals are collected by highly accurate and easily accessible sensors. Sixty-six kinds of time-domain, frequency-domain, and non-linear analyses features are obtained from multiple bio-signals. All features are correlated with the physiological criteria of apnea. In Chapter 4, the two-stage feature selection (Stage 1: the statistical analysis stage and Stage 2: the final feature subset selection stage using machine learning methods) is used to select a smaller subset with relevant features. For comparison purposes, different kernels and parameters of the support vector machine (SVM) method are included. Finally, 19 features from ECG, SaO_2, and abdominal signals are considered as the final feature subset and the input of classifiers to achieve the automatic detection with good accuracy.

In this chapter, the structure is organised as follows. Section 5.2 covers the performance of classification algorithms with the feature subset, and the reason to not a good performance. Section 5.3 reports the solution that is used to improve performance, and results of this solution. Section 5.4 is the summary.

5.2 Classification Based on Duration Parameter

5.2.1 Experiments and Results

The used machine learning algorithms are shown in Chapter 2, such as the SVM algorithm (shown in Section 2.3.1), the decision tree algorithm (shown in Section 2.3.2), the k-nearest neighbors algorithm (shown in Section 2.3.3), the random forest algorithm (shown in Section 2.3.5.1), the extra trees algorithm, the linear discriminant

analysis algorithm, and the logistic regression algorithm. Extremely randomized trees classifier(extra trees classifier) is one of ensemble learning algorithms, and it is able to combine the results of multiple de-correlated decision trees in the same "forest" and output its final prediction. In concept, it is very similar to the random forest algorithm, but extra trees classifier has a different manner of construction of the decision trees in the forest compared with the random forest algorithm. In some of the used machine learning algorithms, there are different hyper-parameters to be set. For example, in the k-nearest neighbors algorithm, k is needed to be set. In the decision tree algorithm, the number of trees and depth are also needed to be set. Considering hyper-parameters, we employed the Grid Search method to determine each hyper-parameter. In the work, we used the GridSearchCV function from sklearn library in Python, and the 3-fold cross-validation method was also used in the Grid Search method.

As shown in Table 4.1, the nineteen time-domain, frequency-domain, and non-linear features from ECG, SaO_2, and abdominal signals are the input of different classifiers. In the feature subset, No. 6 (the mean of the window (mean)), No. 7 (the number of zero-cross in the window (NumZC)), No. 8 (the complexity (comp)), No.10 (the time spent blew the 98% maximum (Bel98)), No.11 (the time spent above the 98% maximum (Abo98)), and No. 12 (the mean of PSD within the 0.016 - 0.05Hz frequency range (mean_$PSD_{0.016/0.05}$)) are extracted from SaO_2 signals. No. 23 (the summation of absolute window (sum_abs)) and No. 26 (the mean in the 80 - 100Hz frequency range (mean_$PSD_{80/100}$)) are extracted from abdominal signals. No. 35 (the number if pairs of adjacent RR intervals where the later RR interval is more than 50 ms than the previous one (NN50_RR)), No. 43 (spectral spread (SS)), No. 44 (spectral decrease (SD)), Nos. 45 - 51 (Shannon's entropy (entropy_D1 to entropy_D7) using seven detail coefficient levels), and No. 65 (the maximum of the diagonal matrix of KPCA (max_dia_kPCA)) are extracted from ECG signals. Classification methods

Table 5.1 Performance of each classifier with the 19 selected features from ECG, SaO_2 and abdominal signals

Classifiers	Acc (%)	Sen (%)	Spe(%)	AUC (%)
SVM	81.68	97.05	66.54	81.79
Random Forest	81.60	85.27	77.98	81.62
Decision Tree	76.78	75.41	78.13	76.77
Extra Trees	81.35	85.25	77.50	81.38
K-Neighbors	78.28	84.03	72.61	78.32
Logistic Regression	81.28	96.18	66.60	81.39
Linear Discriminant	73.80	88.69	59.13	73.91

include the SVM method, the random forest method, the decision tree method, the extra trees method, the K-neighbors method, and the linear discriminant analysis. The performance of each classifier is evaluated by sensitivity, specificity, accuracy, and area under the ROC curve (AUC), as shown in Table 5.1. The SVM method holds the highest performance (accuracy = 81.68%, sensitivity = 97.05%, specificity = 66.54%, and AUC = 81.79%).

5.2.2 Discussion

In Table 5.1, the SVM method is able to hold the highest accuracy (81.68%) and the highest sensitivity (97.05%), but its specificity is low (66.54%). It means that the SVM method can just classify apnea events perfectly. The random forest algorithm can also hold better accuracy (81.60%). Although its sensitivity (85.27%) is lower than the sensitivity of the SVM method, its specificity (77.98%) is better than the specificity of the SVM method, which means the random forest algorithm can classify both apnea and normal episodes at the same time.

The sensitivity of different classification methods is around 85.00% shown in Table 5.1, which means these methods are able to classify apnea events. However, the specificity is just about 68.00%, and it means that normal events are not detected by

Table 5.2 The duration and its count of each normal event in Patient No. 1537

Normal duration (s)	Count	Normal duration (s)	Count
60	7	630	1
90	2	720	1
120	1	810	1
180	2	1140	1
210	1	1170	1
390	1	1260	3
510	1	1320	1
600	1	1650	2

the classification methods. In order to improve specificity and obtain better results, we check the performance of each testing set and compare the difference between the sets with good results and bad results. It can be found that No. 1537 patient has the 98.18% accuracy, the 100% specificity, and the 96.42% sensitivity, while No. 1490 patient only has the 80.52% accuracy, the 61.45% specificity, and the 100% sensitivity. Table 5.2 shows the Patient No. 1537's duration of the normal events and its count, while Table 5.3 shows the Patient No. 1490's duration of the normal events and its count. From Table 5.2, we can see the normal events of the No. 1537 patient and the duration of all normal events is more than 60 seconds, and these normal events are classified. On the other hand, in Table 5.3, the normal events of the No. 1490 patient

Table 5.3 The duration and its count of each normal event in Patient No. 1490

Normal duration (s)	Count	Normal duration (s)	Count
30	52	390	2
60	18	450	1
90	13	480	1
120	8	540	1
150	5	600	1
180	3	660	1
210	1	780	1
300	2	1050	1
330	4	1110	1
360	2	1260	1

that have the 30 seconds duration are not labelled by normal episodes. By compared Table 5.2 and Table 5.3, it is found that classifiers show bad performance when they predict normal events whose duration was less than 60 seconds.

5.3 Evaluation of The 60-Second Time-Window Algorithm

Based on the result that normal events whose duration is less than 60 seconds are not able to be classified, the time-window method is used, and the length of the time-window is equal to 60 seconds. Firstly, the multi-domain feature extraction methods are used to obtain features from bio-signals, based on the 60-second time-window events. Secondly, the two-stage feature selection is utilized to obtain the final feature subset. Finally, different classifiers are able to provide classification results. The time-window method is evaluated by comparing the results.

5.3.1 Feature Extraction

We extract features from ECG, abdominal, thoracic, airflow, and SaO_2 signals in this work. Frequency-domain, time-domain, and non-linear methods take a total of 66 kinds of features. We obtain 12 kinds of features from SaO_2 signals, and features are Nos.1 - 12 in Table 5.4. Airflow recordings provide features from No.13 to No.22 (shown in Table 5.4). We use time-domain and frequency-domain methods in abdominal and thoracic recordings. Nos.23 - 28 features are extracted from abdominal recordings, and Nos.29 - 34 are extracted from thoracic recordings shown in Table 5.4. In this study, three kinds of feature extraction methods are used to obtain features from ECG

signals. Table 5.4 shows these features from Nos. 35 - 66. The details of multiple feature extraction methods are illustrated in Chapter 3.

5.3.2 Feature Selection

The two-stage procedure (Stage 1: the statistical analysis stage and Stage 2: the final feature subset selection stage using machine learning methods) is used, and the details of the two-stage feature selection is shown in Chapter 4. In the first stage, the statistics analysis results of λ_{feature} are also shown in Table 5.5. It can be seen that λ_{feature} of Feature No.1 is 542, which means that 542 pairs of Feature No.1 are positive. To select better features, the threshold is equal to 787 (the half of the processed PSG signals), and one feature is considered as significant one if its λ_{feature} is more than the threshold. Finally, only six features are considered as significant features, and these selected features are asterisk in Table 5.5. All six features are extracted from SaO_2 signals.

In the final feature subset selection stage using machine learning methods, the hill-climbing algorithm is also used to identify features with the most discrimination. Table 5.6 demonstrates six selected features are classified into three classes Class A, B, C according to the distribution of λ_{feature}. No. 8 feature is put into Class A, and Nos. 2, 4, 5, and 6 features are put into Class B, and No. 9 feature is put into Class C. According to the results in Section 4.3.3, the SVM (RBF, $\sigma = 25$, $R = 0.2$) is selected as the classifier model. Initially, the feature in Class A is picked according to the largest λ_{feature} overall features. In Table 5.6, the Class A feature is found to be No.8 feature comp ($\lambda_{\text{feature}} = 1325$) from SaO_2 signals and put into the SVM model. The performance of this iteration is obtained. Then, the next best features are iteratively added into the SVM model and the performance is obtained. This process is repeated

Table 5.4 Features extracted from SaO_2, airflow, abdominal, thoracic, and ECG signals based on the 60-second time-window segmentation method (No.1 - No.12 from SaO_2 signals, No.13 - No.22 from airflow signals, No.23 - No.28 from abdominal signals, No.29 - No.34 from thoracic signals, and feature No.35 - No.66 from ECG signals, four dashed lines divided the table into five parts according to different kinds of signals)

Feature	Feature
1.med	34.sum_$PSD_{80/100}$
2.MM2	35.NN50_RR
3.kur	36.SDSD_RR
4.var	37.tSD_RR
5.min	38.std_RR
6.mean	39.var
7.NumZC	40.kur
8.comp	41.mean_RR
9.SD_1	42.CV_EDR
10.Bel98	43.SS
11.Abo98	44.SD
12.mean_$PSD_{0.016/0.05}$	45.entropy_D1
13.mean	46.entropy_D2
14.med	47.entropy_D3
15.std	48.entropy_D4
16.mean_$PSD_{0/0.1}$	49.entropy_D5
17.mean_$PSD_{0.4/0.5}$	50.entropy_D6
18.mean_D1	51.entropy_D7
19.mean_D2	52.var_D1
20.mean_D3	53.var_D2
21.mean_A3	54.var_D3
22.comp	55.var_D4
23.sum_abs	56.var_D5
24.std_abs	57.var_D6
25.mean	58.var_D7
26.mean_$PSD_{80/100}$	59.WSD_RR
27.mean_D2	60.max_$PSD_{0.03/0.5}$
28.mean_D1	61.mean_$PSD_{10/20}$
29.sum	62.mean_$PSD_{80/100}$
30.med	63.SCrC_3_RR
31.std	64.SCrC_4_RR
32.mean	65.max_dia_kPCA
33.var	66.RP_2_PC

Table 5.5 Feature selection results in the experiment with the 60-second time-window method using statistics analysis (No.1 - No.12 from SaO_2 signals, No.13 - No.22 from airflow signals, No.23 - No.28 from abdominal signals, No.29 - No.34 from thoracic signals, and feature No.35 - No.66 from ECG signals, four dashed lines divided the table into five parts according to different kinds of signals)

Feature	λ_{feature}	Feature	λ_{feature}
1.med	542	34.sum_$PSD_{80/100}$	603
2*.MM2	1036	35.NN50_RR	497
3.kur	508	36.SDSD_RR	520
4*.var	1032	37.tSD_RR	553
5*.min	1079	38.std_RR	551
6*.mean	1079	39.var	564
7.NumZC	572	40.kur	596
8*.comp	1325	41.mean_RR	583
9*.SD_1	837	42.CV_EDR	534
10.Bel98	480	43.SS	570
11.Abo98	693	44.SD	561
12.mean_$PSD_{0.016/0.05}$	662	45.entropy_D1	554
13.mean	54	46.entropy_D2	580
14.med	253	47.entropy_D3	600
15.std	422	48.entropy_D4	588
16.mean_$PSD_{0/0.1}$	427	49.entropy_D5	535
17.mean_$PSD_{0.4/0.5}$	54	50.entropy_D6	587
18.mean_D1	378	51.entropy_D7	138
19.mean_D2	151	52.var_D1	334
20.mean_D3	345	53.var_D2	329
21.mean_A3	324	54.var_D3	410
22.comp	394	55.var_D4	265
23.sum_abs	332	56.var_D5	293
24.std_abs	198	57.var_D6	311
25.mean	213	58.var_D7	333
26.mean_$PSD_{80/100}$	331	59.WSD_RR	119
27.mean_D2	282	60.max_$PSD_{0.03/0.5}$	399
28.mean_D1	516	61.mean_$PSD_{10/20}$	470
29.sum	516	62.mean_$PSD_{80/100}$	615
30.med	602	63.SCrC_3_RR	78
31.std	618	64.SCrC_4_RR	70
32.mean	671	65.max_dia_kPCA	578
33.var	574	66.RP_2_PC	615

Table 5.6 Feature classes via the distribution of λ_{feature} in the experiment with the 60-second time-window method

Feature No.	Class A 1300-1574	Class B 1000-1299	Class C 787-999
SaO_2	8	2 4 5 6	9

Table 5.7 Sensitivity (%), specificity (%), accuracy (%), and AUC (%) based on the hill-climbing method using the SVM model (RBF, $\sigma = 25$, $R = 0.2$) with six selected features from SaO_2 signals

Class	Accuracy	Sensitivity	Specificity	AUC
Class A	82.44	84.85	77.10	80.98
Class AB	82.51	85.26	76.41	80.83
Class ABC	83.62	86.32	73.63	81.98
66 features	88.76	92.22	81.03	86.61

until all the features have been added, and the best feature subset is confirmed by compared all performance. Considering 66 features as the input of the SVM model, the sensitivity is found to be 92.22%, and the specificity is 81.03%, and the accuracy is 88.76%, and the AUC is 86.61%. These results suggest that the performance of a total of 66 kinds of features is considered as the gold standard. The performance of each iteration and the results of 66 features are shown in Table 5.7. From Table 5.7, the six features are fed into the SVM model, and the obtained performance is the 83.62% accuracy, the 83.32% sensitivity, and the 73.63% specificity. However, all features are able to provide better performance. Sixty-six features are considered as the final feature subset.

5.3.3 Classification

The SVM model provides the 88.76% accuracy, the 92.22% sensitivity, and the 81.03% specificity with 66 multi-domain features. In addition, to confirm the best classification performance, 66 features are fed into other different classifiers, including the random

Table 5.8 Performance of each classifier with the 66 selected features in the experiment with the 60-second time-window method

Classifiers	Acc (%)	Sen (%)	Spe(%)	AUC (%)
SVM	88.76	92.22	81.03	86.61
Random Forest	87.62	90.32	83.85	87.08
Decision Tree	83.69	84.42	82.68	84.55
Extra Trees	86.62	88.88	83.48	86.18
K-Neighbors	88.02	91.80	82.77	87.28
Logistic Regression	87.50	94.23	78.13	86.18
Linear Discriminant	87.12	94.83	76.39	85.61

forest method, the decision tree method, the extra trees method, the k-neighbors method, and the linear discriminant analysis. The Grid Search method was also used to determine each hyper-parameter with the 3-fold cross-validation method. The performance of each classifier is obtained and shown in Table 5.8. The SVM method is able to hold the highest accuracy (88.76%) and the highest sensitivity (92.22%), and its specificity is improved to be 81.03%.

5.3.4 Discussion

In the experiment with duration, 19 selected features are fed into different classifiers, shown in Table 5.8. The logistic regression and the linear discriminant have the highest level of sensitivity (great than 94%), which means they are sensitive for apnea events. However, they have the lowest specificity (less than 79%) and they are not good when they determine normal events. The rest of classifiers, including SVM, the random forest, the decision tree, the extra trees, and k-neighbors, balance both sensitivity performance and specificity performance. It means these classifiers can detect both apnea and normal events well. Among these classifiers, the best classifier is the SVM model because it has a higher sensitivity (92.22%) and acceptable specificity (81.03%).

Table 5.9 Comparison the SVM performance of the duration method and the SVM performance of the 60-second time-window method

Segmentation Method	Acc (%)	Sen (%)	Spe(%)	AUC (%)
duration segmentation method	81.68	97.05	66.54	81.79
60-second time-window method	88.76	92.22	81.03	86.61

In the experiment with duration, 19 selected features are fed into different classifiers, shown in Table 5.9. The SVM model has accuracy = 81.68%, sensitivity = 97.05%, specificity = 66.54%, and AUC = 81.79%. The normal events that last less than the 60s are not detected perfectly in these results. In the experiment with the 60-second time-window segmentation method, ECG, SaO_2, airflow, thoracic, and abdominal signals provide 66 kinds of features using feature extraction methods. After the two-stage feature selection, 66 features are considered as the final feature subset and fed into different classifiers, shown in Table 5.8.

In the experiment with the duration, only 19 features from ECG, SaO_2, and abdominal signals are put into the feature subset, and classifiers hold the good performance. However, in the experiment with the time-window method, more kinds of features are put into this feature subset, and airflow and thoracic signals also provide selected features. To achieve the better performance in this experiment, classifiers have to fit this larger feature subset, because there is no help from a sleep expert to detect the duration of each event. Table 5.9 shows the comparison of the performance of the duration segmentation method and the performance of the 60-second time-window method. It can be seen that although the sensitivity of the 60-second time-window method decreases from 97.05% to 92.22%, it still holds the high level. There is an increase in the specificity of the 60-second time-window method, from 66.54% to 81.03%. The accuracy of the 60-second time-window method is high (88.76%), and the AUC is also high (86.61%).

5.4 Summary

In this chapter, 19 features from ECG, SaO_2, and abdominal signals are considered as the final feature subset in the SHHS dataset, and different classification methods are used to classify apnea and normal episodes with the final feature subset. These classifiers include the SVM method, the random forest method, the decision tree method, the extra trees method, the k-neighbors method, the linear discriminant analysis, and the logistic regression. The SVM method is able to hold highest performance (accuracy = 81.68%, sensitivity = 97.05%, specificity = 66.54%, and AUC = 81.79%). Although the sensitivity is good, the low specificity leads to not good accuracy. In order to improve the specificity, we check the performance of each testing set and compare the difference between the sets with good results and bad results. It is found that classifiers show bad performance when they predict normal events whose duration is less than 60 seconds.

The 60-second time-window method is considered to improve classification performance. Time-domain, frequency-domain, and non-linear feature extraction methods mentioned in Chapter 3 are used to obtain 66 kinds of features. After the two-stage feature selection method, 66 features illustrate the best performance, from five kinds of signals. The results showed that the SVM model outperforms the other classifiers (accuracy = 88.76%, sensitivity = 92.22%, specificity = 81.03%, and AUC = 86.61%). The high performance shows the 60-second time-window method can reflect the differences between apnea and normal events more obviously.

Chapter 6

Obstructive Sleep Apnea Classification Using The Multi-Errors-Reduction (MER) System Based on The 60-Second Time-Window Method

6.1 Introduction

During sleep, three basic respiratory disturbances are found. In these types, the most common disorder is obstructive sleep apnea (OSA). The closure of the upper airway is repeated and temporary, and it is defined as OSA in adults. OSA should be objectively assessed to treat it. To diagnose OSA, overnight polysomnography (PSG) is considered as the gold standard. A clinician must characterize pathological events

and identify different parameters that are shown on the PSG to perform the diagnosis. The amount of data obtained is huge since the PSG records the changes the whole night. As a result, for physicians, it is time-consuming to identify OSA. Therefore, the aims of automatic OSA detection are to analyse the information from the PSG and help doctors to diagnose OSA.

In the automatic detection progress, the feature extraction stage, the feature selection stage, and the classification stage are crucial tasks, shown in Fig. 1.13. In Chapter 4, ECG, SaO_2, airflow, thoracic, and abdominal signals are obtained from the SHHS dataset, and 66 kinds of time-domain, frequency-domain, and non-linear analyses features are obtained from multiple bio-signals. In Chapter 4, the two-stage feature selection (Stage 1: the statistical analysis stage and Stage 2: the final feature subset selection stage using machine learning methods) is used to select a smaller subset with relevant features. Finally, 19 features from ECG, SaO_2, and abdominal signals are considered as the final feature subset. In Chapter 5, the SVM method can provide better performance (accuracy = 81.68%) than other classifiers with these 19 selected features. The specificity is only 66.54%, although the sensitivity is 97.05%. In order to improve the specificity, it is found that classifiers are not able to have good performance when they predict normal events whose duration is less than 60 seconds. The 60-second time-window method is considered to improve classification performance.

The 60-second time-window method is used to segment bio-signals, and time-domain, frequency-domain, and non-linear feature extraction methods mentioned in Chapter 3 are used to obtain 66 kinds of features in each segmentation. After the two-stage feature selection method, 66 features illustrate the best performance from five kinds of signals. The results showed that the SVM model outperforms the other classifiers (accuracy = 88.76%, sensitivity = 92.22%, specificity = 81.03%, and area

under ROC curve (AUC) = 86.61%). Compared with the former performance using the duration parameter, the 60-second time-window method can reflect the differences between apnea and normal events more obviously and improve classification performance.

In this chapter, Section 6.2 covers the feature extraction methods, the two-stage feature selection method, and classification based on the 60-second time-window method. The detail methods of classifiers and the multi-errors-reduction (MER) classification system are discussed in Section 6.3. Section 6.4 is the summary.

6.2 Classification by Machine Learning Methods

6.2.1 Feature Extraction and Feature Selection

In the feature extraction phase, 66 kinds of multi-domain features are extracted from ECG, SaO_2, airflow, thoracic, and abdominal signals, and the detail information is described in Chapter 4. In the feature selection phase, each feature and its λ_{feature} obtained in the statistical analysis stage are illustrated in Table 5.5. The performance of different feature subsets (Table 5.7) are compared in the final feature subset selection stage using machine learning methods. Based on the two-stage feature selection procedure, 66 features are considered as the final feature subset, and the SVM model outperforms the other classifiers (accuracy = 88.76%, sensitivity = 92.22%, specificity = 81.03%, and AUC = 86.61%).

Although the SVM model can give high classification performance, the two-stage feature selection procedure does not determine the significant features. From Table 5.7, 66 features are able to provide the 88.76% accuracy, while the features in Class ABC can hold the 83.62% accuracy, and the six features in Class ABC are only extracted

from SaO_2 signals. It can be found that some significant features from other bio-signals are not put into the feature subset to evaluate classification performance. Thus, the ranking feature algorithm is used to determine the feature subset, instead of the different class method.

Table 5.5 shows each feature and its value of λ_{feature}. λ_{feature} is obtained by the ANOVA algorithm and the rank-sum test method, which means the significant level of each feature. The larger λ_{feature} value illustrates the more important level. The order of all features is obtained according to λ_{feature} in descending order, and the result is shown in Table 6.1, and the table also shows the kind of bio-signal that provides each feature. No.8 feature is comp from SaO_2 signals and the value of No.8 feature is the largest ($\lambda_{\text{feature}} = 1325$). Thus, No.8 feature is considered as Order.1 shown in Table 6.1. On the other hand, No.13 feature (mean_$PSD_{0.4/0.5}$) and No.17 feature (mean) from airflow signals have the smallest λ_{feature} value (λ_{feature} is equal to 54). The two features are put at the end of the order list (Order.65 and Order.66, respectively).

The hill-climbing algorithm is used to identify features with the most discrimination according to the descending order list shown in Table 6.1. Fig. 6.1 illustrates the progress of the hill-climbing algorithm based on the descending order list. After obtained the order list, initially, the single-best feature is picked. In Table 6.1, the Order.1 feature is the comp from SaO_2 signals ($\lambda_{\text{feature}} = 1325$) and put into the SVM model. The performance of this iteration is obtained, including accuracy, sensitivity, specificity, and AUC. The performance is also recorded as the first position in Fig. 6.2. Then, the next best feature (Order.2 min from SaO_2 signals) is picked and added into the feature subset. The subset with two features in this iteration is fed into the SVM model. The performance is recorded and put in the second position in Fig. 6.2. This process is repeated until all features have been added, and the best feature subset is confirmed by compared all performance.

Table 6.1 The descending order of 66 features according to λ_{feature} in the 60-second time-window experiment (the bio-signal provides the feature)

Order	Order
1*.comp (SaO_2)	34*.sum (Thoracic)
2*.min (SaO_2)	35*.mean_D1 (Abdominal)
3*.mean (SaO_2)	36*.kur (SaO_2)
4*.MM2 (SaO_2)	37*.NN50_RR (ECG)
5*.var (SaO_2)	38*.Bel98 (SaO_2)
6*.SD_1 (SaO_2)	39*.mean_$PSD_{10/20}$ (ECG)
7*.Abo98 (SaO_2)	40*.mean_$PSD_{0/0.1}$ (Airflow)
8*.mean (Thoracic)	41*.std (Airflow)
9*.mean_$PSD_{0.016/0.05}$ (SaO_2)	42*.var_D3 (ECG)
10*.std (Thoracic)	43*.max_$PSD_{0.03/0.5}$
11*.RP_2_PC (ECG)	44*.comp (Airflow)
12*.mean_$PSD_{80/100}$ (ECG)	45*.mean_D1 (Airflow)
13*.sum_$PSD_{80/100}$ (Thoracic)	46*.mean_D3 (Airflow)
14*.med (Thoracic)	47*.var_D1 (ECG)
15*.entropy_D3 (ECG)	48*.var_D7 (ECG)
16*.kur (ECG)	49.sum_abs (Abdominal)
17*.entropy_D4 (ECG)	50.mean_$PSD_{80/100}$ (Abdominal)
18*.entropy_D6 (ECG)	51.var_D2 (ECG)
19*.mean_RR (ECG)	52.mean_A3 (Airflow)
20*.entropy_D2 (ECG)	53.var_D6 (ECG)
21*.max_dia_kPCA (ECG)	54.var_D5 (ECG)
22*.var (Thoracic)	55.mean_D2 (Abdominal)
23*.NumZC (SaO_2)	56.var_D4 (ECG)
24*.SS (ECG)	57.med (Airflow)
25*.var (ECG)	58.mean (Abdominal)
26*.SD (ECG)	59.std_abs (Abdominal)
27*.entropy_D1 (ECG)	60.mean_D2 (Airflow)
28*.tSD_RR (ECG)	61.entropy_D7 (ECG)
29*.std_RR (ECG)	62.WSD_RR (ECG)
30*.med (SaO_2)	63.SCrC_3_RR (ECG)
31*.entropy_D5 (ECG)	64.SCrC_4_RR (ECG)
32*.CV_EDR (ECG)	65.mean_$PSD_{0.4/0.5}$ (Airflow)
33*.SDSD_RR (ECG)	66.mean (Airflow)

Fig. 6.1 The hill-climbing method based on the descending order list in the 60-second time-window experiment

Considering 66 features as input, the classification performance is presented in Fig. 6.2. The sensitivity is found to be 94.22%, and the specificity is 81.01%, and the accuracy is 88.76%, and the AUC is 86.61%. These results suggest that the performance of a total of 66 kinds of features is considered as the gold standard. The balanced data was used in this step, which means accuracy and AUC have the same trend, and accuracy is used to show the comparison results. Based on the results of the hill-climbing algorithm, the final feature subset contains Order. 1-48 features (the asterisk (*) in Fig 6.2), and these features achieve the maximum accuracy (88.80%), with good sensitivity (91.95%), specificity (81.82%) and AUC(86.89%). These features are also asterisk (*) in Table 6.1, and they are extracted from ECG, SaO_2, airflow, thoracic, and abdominal signals.

6.2.2 Classification Based on Ranking Order

Classification methods include the support vector machine (SVM) algorithm (shown in Section 2.3.1), the decision tree algorithm (shown in Section 2.3.2), k-nearest neighbors algorithm (shown in Section 2.3.3), the random forest algorithm (shown in Section 2.3.5.1), the extra trees algorithm, the linear discriminant analysis algorithm, and the logistic regression algorithm, some boosting methods (shown in 2.3.5.3). The results of each classifier is shown in Table 6.2 with 48 selected features.

In Table 6.2, it can be found that boosting methods have better performance than other classifiers. The best performance (accuracy = 90.71%) is obtained from CatBoost, and the specificity is 89.00% with the 91.94% sensitivity. The reason is that classic machine learning methods always only use the training data once time, but boosting uses repeatedly the training data with differently weighted versions and obtain some weak classifiers. In each iteration of the boosting algorithm, the weight

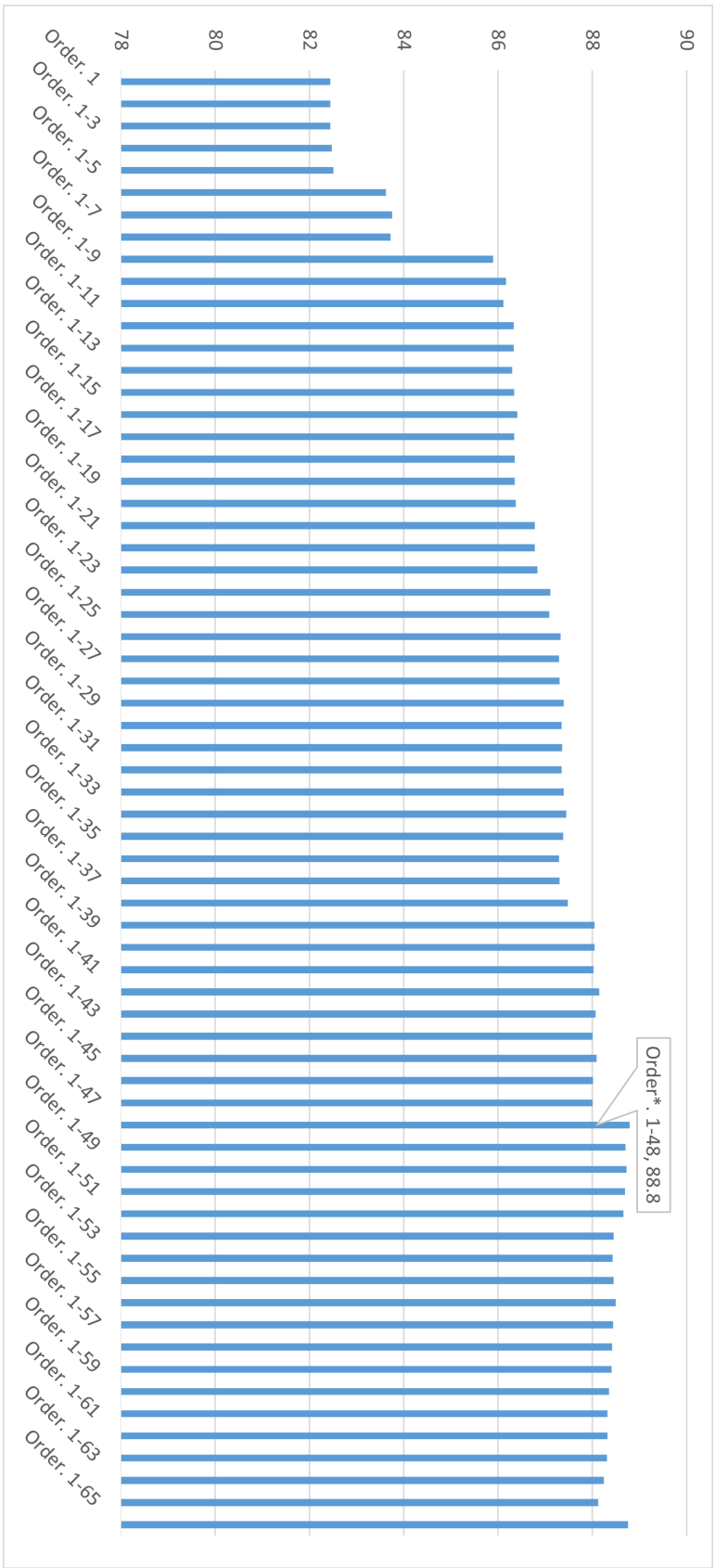

Fig. 6.2 The iteration performance of the hill-climbing algorithm based on the descending order list using the SVM model in the 60-second time-window segmentation method

Table 6.2 Performance of each classifier with the 48 selected features in the 60-second experiment

Classifiers	Acc(%) ± std(%)	Sen(%) ± std(%)	Spe(%) ± std(%)	AUC (%) ± std(%)
SVM	88.80 ± 0.47	91.95 ± 0.43	81.82 ± 0.93	86.89 ± 0.53
Random Forest	87.66 ± 0.57	90.46 ± 0.43	83.15 ± 0.91	87.30 ± 0.62
Decision Tree	83.76 ± 1.20	84.43 ± 1.39	82.83 ± 1.02	83.63 ± 1.17
Extra Trees	86.11 ± 0.48	88.76 ± 0.43	83.41 ± 0.91	86.59 ± 0.54
K-Neighbors	88.99 ± 0.50	91.61 ± 0.58	82.95 ± 0.79	87.28 ± 0.53
Logistic Regression	87.10 ± 0.36	90.87 ± 0.59	81.86 ± 1.00	86.36 ± 0.42
Linear Discriminant	87.65 ± 0.38	93.69 ± 0.44	79.23 ± 1.18	86.46 ± 0.53
Gradient Boosting	90.60 ± 0.42	93.23 ± 0.15	86.94 ± 1.05	90.08 ± 0.51
CatBoost	90.71 ± 0.31	91.94 ± 0.38	89.00 ± 0.41	90.47 ± 0.32
Light GBM	90.34 ± 0.32	91.88 ± 0.28	88.20 ± 0.53	90.04 ± 0.35
XGBoost	90.55 ± 0.30	91.73 ± 0.39	88.90 ± 0.36	90.32 ± 0.30

of each instance of the training data is estimated by the accuracy of the previous classifiers. Thus, this algorithm is able to focus on instances that are always incorrectly detected. In this study, repeatedly the training data with differently weighted versions fitted with complicated processing and boosting methods are able to process complex multi-domain features from different bio-signals.

6.3 Multi-Errors-Reduction Classification System

The 48 selected features (shown in Table 6.1) are extracted from ECG, SaO_2, airflow, thoracic, and abdominal signals. Time-domain, frequency-domain, and non-linear feature extraction methods are used to obtain the selected features. It is a difficult task for a single one machine learning method to classify apnea with high performance

in limited hypotheses when the machine learning method takes multi-domain features from multiple bio-signals.

6.3.1 Methodology

In order to improve the performance of apnea classification, we propose a multi-errors-reduction (MER) classification shown in Fig. 6.3, and it consists of five phases. Firstly, the selected features are fed into the MER classification system. Secondly, some classifiers with good performance are considered as weak learners in level-0. The third phase is the classifier combination method, which is considered as a potential solution to improve classification performance. The basic idea of classifier combination is the misclassified instances of individual classifiers do not overlap, and different individual classifiers can provide different perspectives in classification. Classifier combination can use complementary information to improve performance. Some basic classifier combination schemes include max voting, average voting, and weighted average voting. In this thesis, we use the stacking ensemble method as the classifier combination. In the last phase, a meta-learner is used to provide final predictions.

The general information of stacking ensemble method has been shown in Section. 2.3.5.3. It combines predictions of weak learner (level-0) in a meta-level (level-1) classifier, and a meta-level classifier can correct the decisions of weak learners and predict final labels. Figs. 6.4 show an example of stacking implemented with the first weak learner and 3-fold cross-validation. The set is divided into the train set and

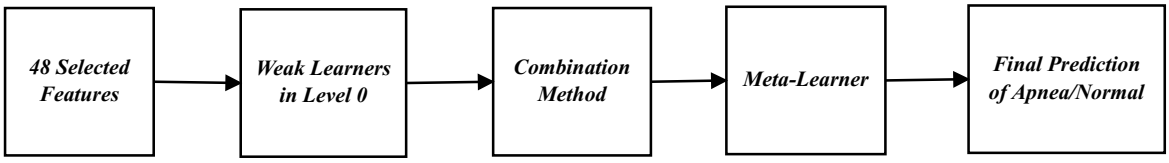

Fig. 6.3 Basic components of the multi-errors-reduction classification system

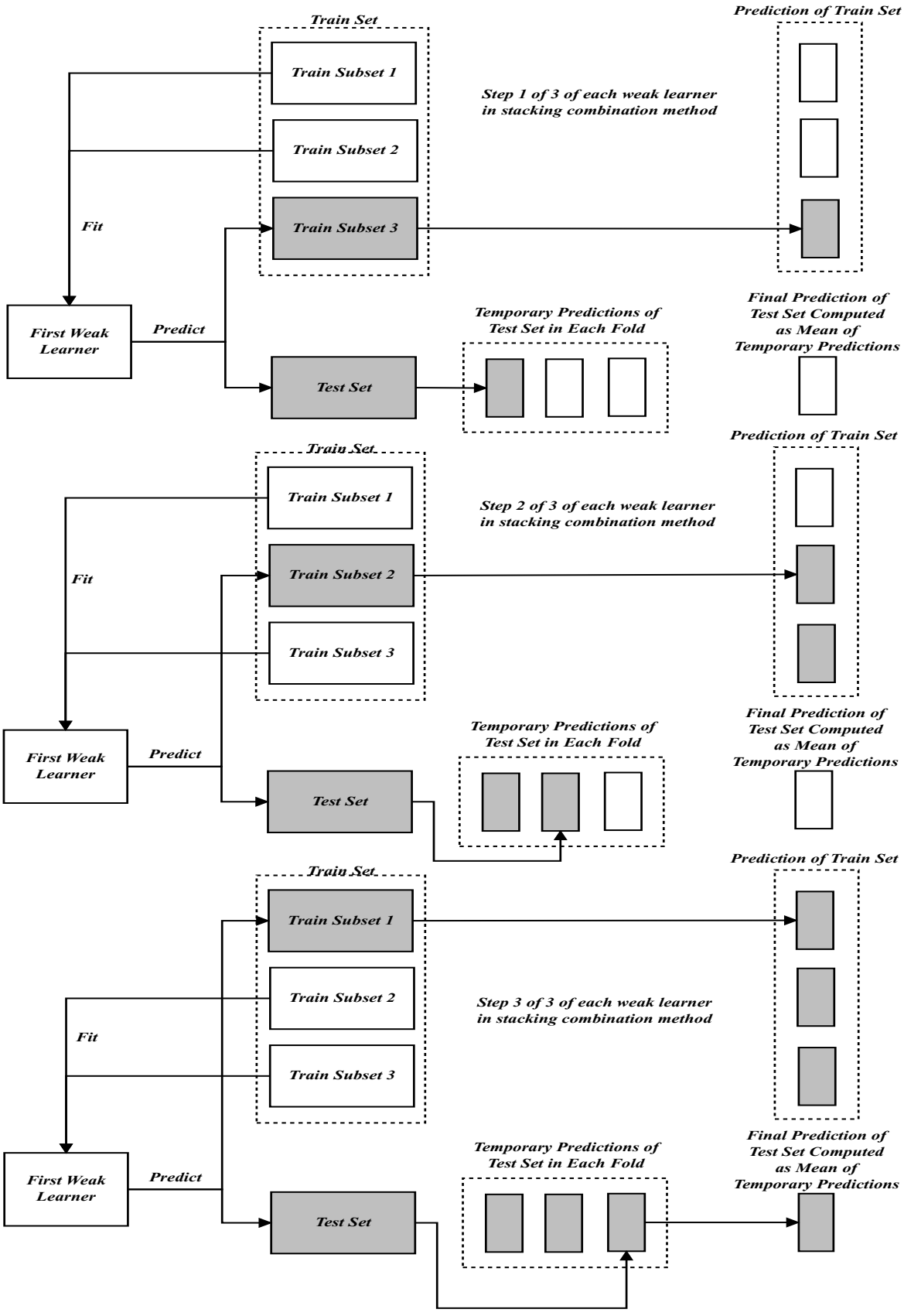

Fig. 6.4 All steps of each weak learner in stacking combination method

the test set. As for the train set, in order to avoid over-fitting, we use 3-fold cross-validation technique. In the first fold, as shown in Fig. 6.4, Train Subset 1 and Train Subset 2 are used to fit the first weak learner, and the trained model is utilized to predict Train Subset 3, and the prediction of Train Subset 3 is recorded. The trained model is also used to predict Test Set, and the prediction of Test Set in the first fold is recorded. In the second fold, as shown in Fig. 6.4, Train Subset 1 and Train Subset 3 are used to fit the first weak learner, and the trained model is utilized to predict Train Subset 2, and the prediction of Train Subset 2 is recorded. The trained model is also used to predict Test Set, and the prediction of Test Set in the second fold is recorded. In the third fold, as shown in Fig. 6.4, Train Subset 2 and Train Subset 3 are used to fit the first weak learner, and the trained model is utilized to predict Train Subset 1, and the prediction of Train Subset 1 is recorded. The trained model is also used to predict Test Set, and the prediction of Test Set in the third fold is recorded. After the 3-fold cross-validation technique, there are three temporary test set predictions, and the mean of all temporary test set prediction sets is the final prediction of Test Set using the first weak learner. This process is repeated until weak learners predict Train Set and Test Set. After the stacking combination method, each vector in Train Set and Test Set has predictions from weak learners, respectively.

In the fourth phase of the MER classification system is the meta-learner. In this thesis, the meta-classifier (level-1) is an artificial neural network (ANN) since it is extensively used for binary classification in sleep apnea studies. ANN is a generic nonlinear function classification method, and it is a combination of basic knots, called neurons (the general information in Section 2.3.4). Fig. 6.5 illustrates the training process of the ANN. Each kind of prediction from one of the weak learners is fed into one neuron in the input layer of an ANN. The ANN obtains the outputs by fitting weights and bias. The outputs are compared with the targets of the input predictions

(the apnea label or the normal label), and the loss parameter is the most widely used as the evaluation index. The evaluation results return to the ANN, and the ANN tweak network weights and bias to reduce the loss parameter. After the training process, the trained ANN is obtained, which has one input layer with neurons, one hidden layer having nodes, and one output layer. Fig. 6.6 illustrates the testing process of the trained ANN. Each kind of prediction from one of the weak learners is fed into one neuron in the input layer of the trained ANN. The determined weights, bias, and activation functions are used to obtain the final predictions.

6.3.2 Experiments, Results and Discussion

The MER classification is shown in Fig. 6.3. Firstly, the 48 selected features are fed into the MER classification system, which are extracted from ECG, SaO_2, airflow, thoracic, and abdominal signals. Secondly, in Table 6.2, four boosting methods (Gradient Boosting, CatBoost, Light GBM, and XGBoost) have better performance than other classifiers, and the accuracy is above 90%. The four boosting methods are considered as weak learners in level-0. The third phase is the classifier combination method. In this thesis, we use the stacking ensemble method as the classifier combination. In the fourth phase, to combine the predictions from four weak learners, an ANN is considered as the meta-learner. The Grid Search method was used to tune the parameters of an ANN. After the training process of the ANN, the trained ANN with one input layer, one hidden layer having four nodes, and one output layer is used to provide the final prediction.

In the 60-second time-window experiment, Table 6.3 shows the performance of the MER classification system and the results of four weak learners (Gradient Boosting, CatBoost, Light GBM, and XGBoost). The average accuracy of four weak learners is

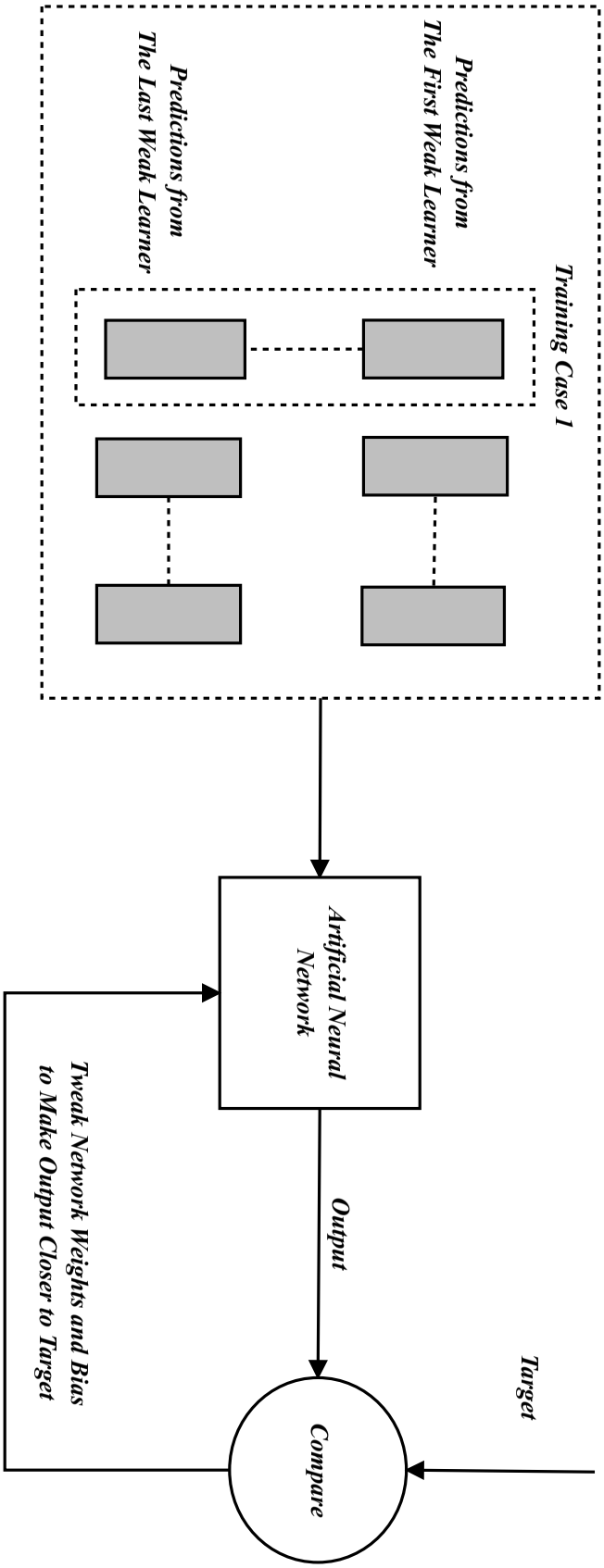

Fig. 6.5 Training process of an ANN in the multi-errors-reduction classification system

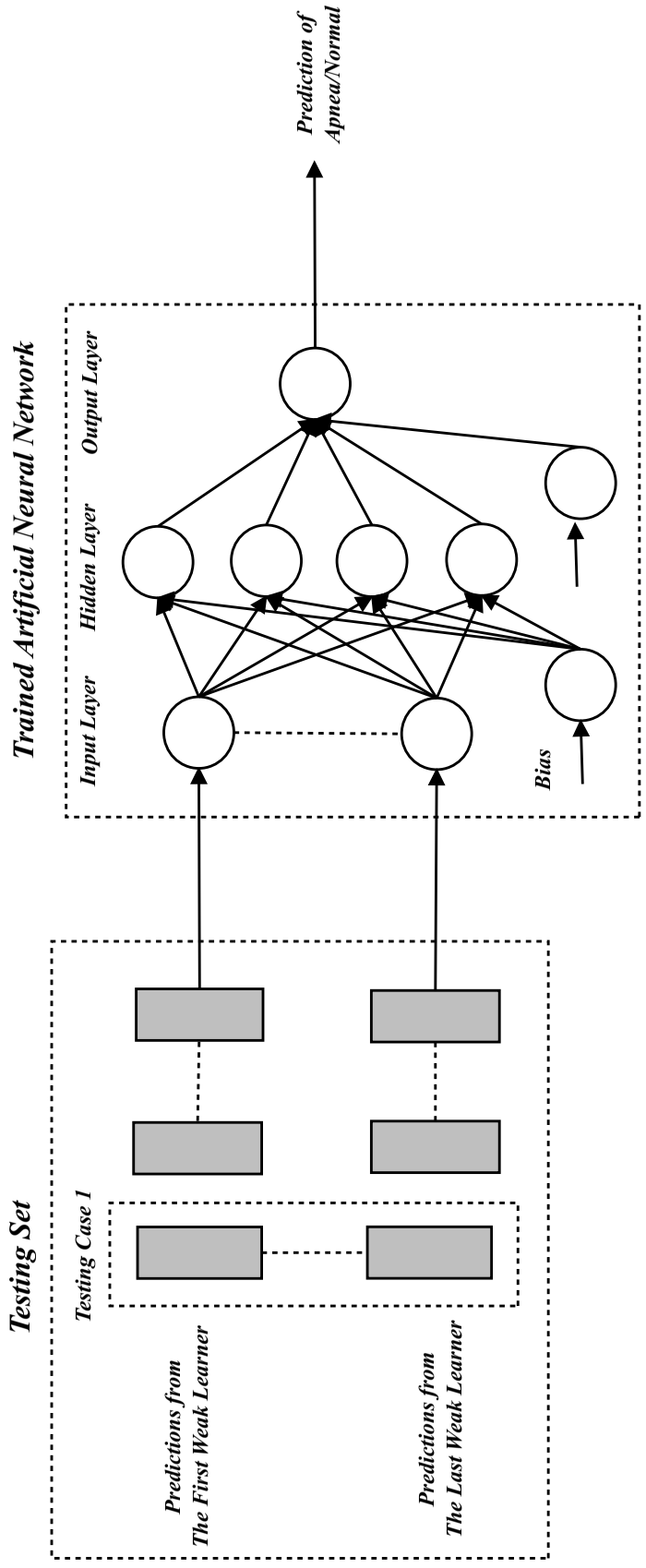

Fig. 6.6 Prediction process of the trained ANN in the multi-errors-reduction classification system

Table 6.3 Performance of four weak classifiers and the multi-errors-reduction classifier with the 48 selected features based on the 60-second time-window segmentation method

Classifiers	Acc (%)	Sen (%)	Spe(%)	AUC (%)
Gradient Boosting	90.60	93.23	86.94	90.08
CatBoost	90.71	91.94	89.00	90.47
Light GBM	90.34	91.88	88.20	90.04
XGBoost	90.55	91.73	88.90	90.32
MER classifier	94.66	96.37	90.83	93.60

Table 6.4 Performance of different classifiers and the multi-errors-reduction classifier with the 48 selected features based on the 60-second time-window segmentation method

Classifiers	Acc (%)	Sen (%)	Spe(%)	AUC (%)
SVM	88.80	91.95	81.82	86.89
Random Forest	87.66	90.46	83.15	87.30
Decision Tree	83.76	84.43	82.83	83.63
Extra Trees	86.11	88.76	83.41	86.59
K-Neighbors	88.99	91.61	82.95	87.28
Logistic Regression	87.10	90.87	81.86	86.36
Linear Discriminant	87.65	93.69	79.23	86.46
MER classifier	94.66	96.37	90.83	93.60

about 90%, and the average sensitivity of four weak learners is about 92%, and the average specificity of four weak learners is about 88%, and the average AUC of four weak learners is about 90%. However, the results (accuracy = 94.66%, sensitivity = 96.37%, specificity = 90.83%, and AUC = 93.60%) of the MER classification system have higher performance than the results from four weak learners.

In addition, the performance of the MER classification system is also compared with the other classifiers (Table 6.4). The average accuracy of different classifiers is about to 86%, and the average sensitivity of different classifiers is about 85%, and the average specificity of different classifiers is about 81%, and the average AUC of different classifiers is about 85%. Using the MER classification system, the accuracy is

Table 6.5 Performance of each classifier and the multi-errors-reduction classifier with the 19 selected features from ECG, SaO_2, and abdominal signals, based on the duration parameter

Classifiers	Acc (%)	Sen (%)	Spe(%)	AUC (%)
SVM	81.68	97.05	66.54	81.79
Random Forest	81.60	85.27	77.98	81.62
Decision Tree	76.78	75.41	78.13	76.77
Extra Trees	81.35	85.25	77.50	81.38
K-Neighbors	78.28	84.03	72.61	78.32
Logistic Regression	81.28	96.18	66.60	81.39
Linear Discriminant	73.80	88.69	59.13	73.91
MER classifier	85.14	94.96	75.45	85.21

improved from about 89% to 94.66%, and the sensitivity is 96.37%, and the specificity is 90.83%, and the AUC is 93.60.

The MER classification system is also used in the duration experiment. The results of the MER system consists of the 85.14% accuracy, the 94.96% sensitivity, the 75.45% specificity, and the 85.21% AUC (Table 6.5). Table 6.5 also demonstrates the results of other classifiers. The MER classification system outperforms the other classifiers.

6.3.3 Discussion

The 60-second time-window segmentation method is used since there is no duration parameter in actual data, and the 60-second time-window method has potential power in actual data. In Section 5.3, 66 kinds of multi-domain features are extracted from ECG, SaO_2, airflow, thoracic, and abdominal signals. In order to reduce the feature set, the ranking feature algorithm is used to determine the final feature subset. Forty-eight selected features are put into the final feature subset from five bio-signals. These 48 kinds of features are related to the bio-physiological criteria. The performance (the 88.80% accuracy) of 48 kinds of features is higher than the performance (the 88.76%

accuracy) of the 66 features, which means some features from abdominal, ECG, and airflow signals are irrelevant.

Four boosting methods (Gradient Boosting, CatBoost, Light GBM, and XGBoost) are used in the single one classification stage, and boosting methods have better classic machine learning methods. The best classifier is CatBoost with accuracy = 90.71%, sensitivity = 91.94%, specificity = 89.00%, and AUC = 90.47%. Although boosting methods have a high level with a time-window segmentation method, we also try to boost the classification to a higher level. The MER classification system is developed. Gradient Boosting, CatBoost, Light GBM, and XGBoost provide four predictions of each event, respectively. After the stacking classification combination method, the ANN with one input layer, one hidden layer having four nodes, and one output layer is used to provide the final prediction. By this classification system, in the 60-second time-window experiment, the accuracy is improved to 94.66% from 90.71%, and in the duration experiment, the accuracy is improved to 85.14% from 81.68%. In this work, we use multi-domain methods to define features from multi-bio signals and reflect apnea events in different spaces. To the extent complex features, the MER classification system provides a more effective approach through comparison with empirical data. The MER classification system generates a more general application, and it combines the advantages from four boosting methods and overcomes the disadvantages.

6.4 Summary

The aim of this chapter is to construct the multi-errors-reduction classification system to detect apnea events. To achieve this aim, firstly, multi-domain feature extractions and the two-stage feature selection are used to select the final feature subset. Multi-bio signals from PSG recordings are used to generate features, and the PSG signals are

collected in the Sleep Heart Health Study (SHHS) database. From the database, we adopt 1,574 patients' PSG recordings and use feature extraction algorithms include time-domain, frequency-domain, and non-linear analysis. In the experiment with the 60-second time-window segmentation method, 66 kinds of features are extracted from ECG, SaO_2, airflow, thoracic, and abdominal signals. In the two-stage feature selection stage, in order to improve performance, the descending order of 66 features is obtained according to λ_{feature}. With the hill-climbing algorithm, 48 features achieve the maximum accuracy (88.80%), with good sensitivity (91.95%), specificity (81.82%), and AUC(86.89%). After the feature selection stage, 48 features are selected from five kinds of bio-signals. In the classification of single one machine learning method, four boosting methods (Gradient Boosting, CatBoost, Light GBM, and XGBoost) are used. The best classifier is CatBoost with accuracy = 90.71%, sensitivity = 91.94%, specificity = 89.00%, and AUC = 90.47%.

The multi-errors-reduction classification system is constructed to improve performance. Firstly, Gradient Boosting, CatBoost, Light GBM, and XGBoost are considered as base classifiers (level-0). After the stacking classification combination method, there is an ANN with one input layer, one hidden layer having four nodes, and one output layer, and four inputs are provided by the results from four boosting methods respectively. After multi-errors-reduction classification system, results (accuracy = 94.66%, sensitivity = 96.37%, specificity = 90.83% and AUC = 93.60%) have high performance in the 60-second time-window experiment. In the duration experiment, results (accuracy = 85.14%, sensitivity = 94.96%, specificity = 75,45%, and AUC = 85.21%). The multi-errors-reduction classification system is able to improve the classification performance.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

Individuals with OSA often face many serious diseases. Traditionally, sleep specialists perform OSA diagnosis according to visual observation of PSG signals. OSA diagnosis is based on the visual observation of sleep specialists, which has many disadvantages, such as the requirement of trained specialists, an onerous task for specialists. Therefore, in order to detect OSA, automatic systems based on machine learning algorithms are desirable to help doctors during the diagnostic process.

The thesis proposes an automatic obstructive sleep apnea (OSA) detection system using multi-domain features from different bio-signals. To achieve this, the automatic OSA detection system is divided into three parts, including the multi-domain feature extraction algorithms, the two-stage feature selection method, and the multi-errors-reduction (MER) classification system.

First of all, since the detection of OSA manually is time-consuming, costly and uncomfortable, the automatic OSA detection system should rely on the core bio-signals,

which tend to be low-cost, easy to collect, and good quality. Thus, it is an important step to select significant bio-signals. ECG, airflow, SaO_2, abdominal, and thoracic signals are considered as the core bio-signals. The bio-signals have different bio-patterns between apnea events and normal events. According to the different apneic bio-physiological changes from multi-bio signals, time-domain, frequency-domain, and non-linear features are extracted from ECG, airflow, SaO_2, abdominal, and thoracic signals. There are 66 features from the five kinds of bio-signals. Twelve features are extracted from SaO_2 signals, and nine feature are extracted from airflow signals, and six features are extracted from abdominal signals, and six features are extracted from thoracic signals, and 32 features are extracted from ECG signals.

All features then are evaluated by the two-stage feature selection (Stage 1: the statistical analysis stage and Stage 2: the final feature subset selection stage using machine learning methods). In the first stage, two statistics analyses (the analysis of variance (ANOVA) and the rank-sum test) are used. Each feature from each patient has a pair of p-values. There are two statistical analysis criteria (the p-value of ANOVA < 0.05 and the p-value $= 1$), which are used to determine significant features. λ_{feature} is set as the number of positive pairs for each feature, and a threshold is equal to half of the number of processed signals. If the λ_{feature} of a feature is more than the threshold, this feature is put into the feature subset. In the second stage, the feature subset from the former phase is divided into different classes by comparing the distribution of the λ_{feature} set of selected features. Features in different classes are put into an SVM classifier using the hill-climbing algorithm. The sensitivity, specificity, accuracy, and area under ROC curve (AUC) are used to evaluate the performance of different SVM models. After the two-stage feature selection procedure, the final feature subset with good discrimination is determined.

In the thesis, the Sleep Heart Health Study (SHHS) database is used since it is able to provide over 1,500 patients' PSG signals, which help to evaluate the stability of our automatic system. 66 time-domain, frequency-domain, and non-linear analysis features are extracted from ECG, airflow, SaO_2, abdominal, and thoracic signals. After the two-stage feature selection, 19 features are considered as the final feature subset, which is extracted from ECG, SaO_2, and abdominal signals. These features are able to reflect the apneic bio-physiological changes, such as sudden drops and fast recoveries in SaO_2 signals, active abdominal muscles during sleep apnea events, and lower heart rate in ECG signals. Compared different classifiers, the SVM method is able to hold highest performance (accuracy = 81.68%, sensitivity = 97.05%, specificity = 66.54%, and AUC = 81.79%). To improve the specificity, it is found that classifiers show bad performance when they predict normal events whose duration is less than 60 seconds.

In the experiment with the 60-second time-window segmentation method, 66 kinds of features are extracted from ECG, SaO_2, airflow, thoracic, and abdominal signals. After the two-stage feature selection method, 48 features achieve the maximum accuracy (88.80%), with good sensitivity (91.95%), specificity (81.82%), and AUC(86.89%), which are extracted from five kinds of bio-signals. In the classification using a single one machine learning method, four boosting methods (Gradient Boosting, CatBoost, Light GBM, and XGBoost) outperforms other classifiers. The MER classification system is constructed to improve performance. Firstly, four boosting methods are considered as base classifiers (level-0). After the stacking classification combination method, there is the ANN with one input layer, one hidden layer having four nodes, and one output layer, and four inputs are provided by the results from four boosting methods respectively. After multi-errors-reduction classification system, results (accuracy = 94.66%, sensitivity = 96.37%, specificity = 90.83% and AUC = 93.60%) have high performance in the 60-second time-window experiment. In the duration exper-

iment, results (accuracy = 85.14%, sensitivity = 94.96%, specificity = 75,45%, and AUC = 85.21%). The MER classification system is able to improve the classification performance.

ECG, SaO_2, airflow, thoracic, and abdominal signals are highly accurate, easily accessible, and more stable bio-signals, and time-domain, frequency-domain, and non-linear analyses can provide features with good information. The two-stage procedure is a combination of the statistical analysis stage and machine learning methods to select the features with good discrimination. The multi-errors-reduction classification system uses Gradient Boosting, CatBoost, Light GBM, and XGBoost as base classifiers (level-0) and an ANN as level-1 and the system can hold the complex multi-domain features from different bio-signals and offer good performance.

7.2 Future Work

Some possible future research directions are shown in this section.

1. Cooperate with hospitals or sleep research institutions. In this thesis, all used datasets are open from the Internet. It is necessary to evaluate the performance of our automatic OSA detection system in real data. Hospitals or sleep research institutions are able to help to collect real OSA PSG recordings since they have PSG machines and sleep experts. Also, in collected PSG signals, different sleep stages and sleep diseases should be labelled by sleep doctors.

2. Use novel sensors. The aim of this study is to obtain useful features from multi-bio signals. The quality of bio-signals affects the performance of our method because there are artifacts and baseline problems in bio-signals with the bad-quality. Although removing artifacts and the baseline correction are used in this

paper, to further improve the auto-detection, it is necessary to upgrade the signal acquisition process and the sensor. Wearable sensors that can be attached directly on human skin can monitor valuable physical and biological information, such as airflow, pressure, respiration, and blood, even if the surrounding conditions always change [84] [85]. Classic sensors monitor a single signal, and it is an urgent need for novel sensors that can monitor changes in a variety of bio-signals. For this purpose, the surface of polyimide nanowires with silver nanoparticles is used to fabricate a novel sensor [86]. With the development of novel sensors, bio-signals are monitored and recorded efficiently, and these bio-signals with good quality improve the performance of feature extraction methods and classifications.

3. Develop in-home or mobile monitoring. More technology companies focus on wearable electric equipment, such as Apple Watch and HUAWEI Watch. Our automatic method will be used in this field. Watches are used to collect limited and useful signals, such as oxygen saturation and airflow signals, and data are transformed to the paired phone by Bluetooth. In the phone, our detection algorithm is running and monitoring apnea events. If the algorithm finds the person monitored is during apnea, the phone alarm the subject by ringing or flashing.

References

[1] B. Guijarro-Berdiñas, E. Hernández-Pereira, and D. Peteiro-Barral, "A mixture of experts for classifying sleep apneas," Expert Systems with Applications, vol. 39, no. 8, pp. 7084–7092, 2012.

[2] G. Slater and J. Steier, "Excessive daytime sleepiness in sleep disorders," Journal of thoracic disease, vol. 4, no. 6, p. 608, 2012.

[3] N. Ben-Israel, A. Tarasiuk, and Y. Zigel, "Nocturnal sound analysis for the diagnosis of obstructive sleep apnea," in 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, pp. 6146–6149, IEEE, 2010.

[4] D. Alvarez-Estevez and V. Moret-Bonillo, "Computer-assisted diagnosis of the sleep apnea-hypopnea syndrome: a review," Sleep disorders, vol. 2015, 2015.

[5] A. Otero, P. Félix, and M. R. Álvarez, "Algorithms for the analysis of polysomnographic recordings with customizable criteria," Expert Systems with Applications, vol. 38, no. 8, pp. 10133–10146, 2011.

[6] N. M. Punjabi, "The epidemiology of adult obstructive sleep apnea," Proceedings of the American Thoracic Society, vol. 5, no. 2, pp. 136–143, 2008.

[7] S. J. Chang and K. Y. Chae, "Obstructive sleep apnea syndrome in children: Epidemiology, pathophysiology, diagnosis and sequelae," Korean journal of pediatrics, vol. 53, no. 10, p. 863, 2010.

[8] P. E. Peppard, T. Young, J. H. Barnet, M. Palta, E. W. Hagen, and K. M. Hla, "Increased prevalence of sleep-disordered breathing in adults," American journal of epidemiology, vol. 177, no. 9, pp. 1006–1014, 2013.

[9] E. L. Gray, D. K. McKenzie, and D. J. Eckert, "Obstructive sleep apnea without obesity is common and difficult to treat: evidence for a distinct pathophysiological phenotype," Journal of Clinical Sleep Medicine, vol. 13, no. 1, pp. 81–88, 2017.

[10] T. Aittokallio, J. S. Malminen, T. Pahikkala, O. Polo, and O. S. Nevalainen, "Inspiratory flow shape clustering: An automated method to monitor upper airway performance during sleep," Computer methods and programs in biomedicine, vol. 85, no. 1, pp. 8–18, 2007.

[11] C. Guilleminault, R. Winkle, S. Connolly, K. Melvin, and A. Tilkian, "Cyclical variation of the heart rate in sleep apnoea syndrome: mechanisms, and usefulness of 24 h electrocardiography as a screening technique," The Lancet, vol. 323, no. 8369, pp. 126–131, 1984.

[12] D. Sánchez-Morillo, M. López-Gordo, and A. León, "Novel multiclass classification for home-based diagnosis of sleep apnea hypopnea syndrome," Expert Systems with Applications, vol. 41, no. 4, pp. 1654–1662, 2014.

[13] J. V. Marcos, R. Hornero, D. Álvarez, F. del Campo, C. Zamarrón, and M. López, "Utility of multilayer perceptron neural network classifiers in the diagnosis of the obstructive sleep apnoea syndrome from nocturnal oximetry," computer methods and programs in biomedicine, vol. 92, no. 1, pp. 79–89, 2008.

[14] A. Otero, P. Félix, S. Barro, and C. Zamarrón, "A structural knowledge-based proposal for the identification and characterization of apnoea episodes," Applied Soft Computing, vol. 12, no. 1, pp. 516–526, 2012.

[15] S. Güneş, K. Polat, and Ş. Yosunkaya, "Multi-class f-score feature selection approach to classification of obstructive sleep apnea syndrome," Expert systems with applications, vol. 37, no. 2, pp. 998–1004, 2010.

[16] D. Álvarez-Estévez and V. Moret-Bonillo, "Fuzzy reasoning used to detect apneic events in the sleep apnea-hypopnea syndrome," Expert Systems with Applications, vol. 36, no. 4, pp. 7778–7785, 2009.

[17] S. S. Mostafa, D. Baptista, A. G. Ravelo-García, G. Juliá-Serdá, and F. Morgado-Dias, "Greedy based convolutional neural network optimization for detecting apnea," Computer Methods and Programs in Biomedicine, vol. 197, p. 105640, 2020.

[18] O. Fontenla-Romero, B. Guijarro-Berdiñas, A. Alonso-Betanzos, and V. Moret-Bonillo, "A new method for sleep apnea classification using wavelets and feedforward neural networks," Artificial Intelligence in Medicine, vol. 34, no. 1, pp. 65–76, 2005.

[19] G. C. Gutiérrez-Tobal, M. L. Alonso-Álvarez, D. Álvarez, F. del Campo, J. Terán-Santos, and R. Hornero, "Diagnosis of pediatric obstructive sleep apnea: Preliminary findings using automatic analysis of airflow and oximetry recordings obtained at patients' home," Biomedical Signal Processing and Control, vol. 18, pp. 401–407, 2015.

[20] H. Lee, J. Park, H. Kim, and K.-J. Lee, "New rule-based algorithm for real-time detecting sleep apnea and hypopnea events using a nasal pressure signal," Journal of medical systems, vol. 40, no. 12, p. 282, 2016.

[21] G. Gutiérrez-Tobal, R. Hornero, D. Álvarez, J. Marcos, and F. Del Campo, "Linear and nonlinear analysis of airflow recordings to help in sleep apnoea–hypopnoea syndrome diagnosis," Physiological measurement, vol. 33, no. 7, p. 1261, 2012.

[22] G. C. Gutiérrez-Tobal, D. Álvarez, F. del Campo, and R. Hornero, "Utility of adaboost to detect sleep apnea-hypopnea syndrome from single-channel airflow," IEEE Transactions on Biomedical Engineering, vol. 63, no. 3, pp. 636–646, 2015.

[23] H. ElMoaqet, J. Kim, D. Tilbury, S. K. Ramachandran, M. Ryalat, and C.-H. Chu, "Gaussian mixture models for detecting sleep apnea events using single oronasal airflow record," Applied Sciences, vol. 10, no. 21, p. 7889, 2020.

[24] N. Sezgin and M. E. Tagluk, "Energy based feature extraction for classification of sleep apnea syndrome," Computers in biology and medicine, vol. 39, no. 11, pp. 1043–1050, 2009.

[25] M. E. Tagluk, M. Akin, and N. Sezgin, "Classıfication of sleep apnea by using wavelet transform and artificial neural networks," Expert Systems with Applications, vol. 37, no. 2, pp. 1600–1607, 2010.

[26] Y.-Y. Lin, H.-T. Wu, C.-A. Hsu, P.-C. Huang, Y.-H. Huang, and Y.-L. Lo, "Sleep apnea detection based on thoracic and abdominal movement signals of wearable piezoelectric bands," IEEE journal of biomedical and health informatics, vol. 21, no. 6, pp. 1533–1545, 2016.

[27] M. Bianchi, T. Lipoma, C. Darling, Y. Alameddine, and M. Westover, "Automated sleep apnea quantification based on respiratory movement," International journal of medical sciences, vol. 11, no. 8, p. 796, 2014.

[28] M. O. Mendez, 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, vol. 56, no. 12, pp. 2838–2850, 2009.

[29] J. D. Martínez-Vargas, L. M. Sepúlveda-Cano, C. Travieso-Gonzalez, and G. Castellanos-Dominguez, "Detection of obstructive sleep apnoea using dynamic filter-banked features," Expert Systems with Applications, vol. 39, no. 10, pp. 9118–9128, 2012.

[30] A. R. Hassan and M. A. Haque, "Computer-aided obstructive sleep apnea screening from single-lead electrocardiogram using statistical and spectral features and bootstrap aggregating," Biocybernetics and Biomedical Engineering, vol. 36, no. 1, pp. 256–266, 2016.

[31] T. Al-ani, C. K. Karmakar, A. H. Khandoker, and M. Palaniswami, "Automatic recognition of obstructive sleep apnoea syndrome using power spectral analysis of electrocardiogram and hidden markov models," in Intelligent Sensors, Sensor Networks and Information Processing, 2008. ISSNIP 2008. International Conference on, pp. 285–290, IEEE, 2008.

[32] K. Feng, H. Qin, S. Wu, W. Pan, and G. Liu, "A sleep apnea detection method based on unsupervised feature learning and single-lead electrocardiogram," IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1–12, 2020.

[33] H.-Y. Chang, C.-Y. Yeh, C.-T. Lee, and C.-C. Lin, "A sleep apnea detection system based on a one-dimensional deep convolution neural network model using single-lead electrocardiogram," Sensors, vol. 20, no. 15, p. 4157, 2020.

[34] R. Tripathy, P. Gajbhiye, and U. R. Acharya, "Automated sleep apnea detection from cardio-pulmonary signal using bivariate fast and adaptive emd coupled with cross time–frequency analysis," Computers in Biology and Medicine, vol. 120, p. 103769, 2020.

[35] A. Zarei and B. M. Asl, "Automatic classification of apnea and normal subjects using new features extracted from hrv and ecg-derived respiration signals," Biomedical Signal Processing and Control, vol. 59, p. 101927, 2020.

[36] N. Pombo, B. M. Silva, A. M. Pinho, and N. Garcia, "Classifier precision analysis for sleep apnea detection using ecg signals," IEEE Access, vol. 8, pp. 200477–200485, 2020.

[37] B. Fatimah, P. Singh, A. Singhal, and R. B. Pachori, "Detection of apnea events from ecg segments using fourier decomposition method," Biomedical Signal Processing and Control, vol. 61, p. 102005, 2020.

[38] H. Singh, R. K. Tripathy, and R. B. Pachori, "Detection of sleep apnea from heart beat interval and ecg derived respiration signals using sliding mode singular spectrum analysis," Digital Signal Processing, vol. 104, p. 102796, 2020.

[39] N. Acır and C. Güzeliş, "Automatic recognition of sleep spindles in eeg by using artificial neural networks," Expert Systems with Applications, vol. 27, no. 3, pp. 451–458, 2004.

[40] C.-C. Hsu and P.-T. Shih, "A novel sleep apnea detection system in electroencephalogram using frequency variation," Expert Systems with Applications, vol. 38, no. 5, pp. 6014–6024, 2011.

[41] E. D. Übeyli, D. Cvetkovic, G. Holland, and I. Cosic, "Adaptive neuro-fuzzy inference system employing wavelet coefficients for detection of alterations in

sleep eeg activity during hypopnoea episodes," Digital Signal Processing, vol. 20, no. 3, pp. 678–691, 2010.

[42] M. E. Tagluk and N. Sezgin, "A new approach for estimation of obstructive sleep apnea syndrome," Expert Systems with Applications, vol. 38, no. 5, pp. 5346–5351, 2011.

[43] D. Liu, Z. Pang, and S. R. Lloyd, "A neural network method for detection of obstructive sleep apnea and narcolepsy based on pupil size and eeg," IEEE Transactions on Neural Networks, vol. 19, no. 2, pp. 308–318, 2008.

[44] E. D. Übeyli, 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, vol. 37, no. 6, pp. 4463–4467, 2010.

[45] S. Taran and V. Bajaj, "Sleep apnea detection using artificial bee colony optimize hermite basis functions for eeg signals," IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 2, pp. 608–616, 2019.

[46] H. M. Al-Angari and A. V. Sahakian, "Automated recognition of obstructive sleep apnea syndrome using support vector machine classifier," IEEE Transactions on Information Technology in Biomedicine, vol. 16, no. 3, pp. 463–468, 2012.

[47] G. Cohen and P. De Chazal, "Automated detection of sleep apnea in infants: A multi-modal approach," Computers in biology and medicine, vol. 63, pp. 118–123, 2015.

[48] Y. Maali and A. Al-Jumaily, "Multi neural networks investigation based sleep apnea prediction," Procedia Computer Science, vol. 24, pp. 97–102, 2013.

[49] I. Fernández-Varela, E. Hernández-Pereira, D. Álvarez-Estévez, and V. Moret-Bonillo, "Combining machine learning models for the automatic detection of eeg arousals," Neurocomputing, vol. 268, pp. 100–108, 2017.

[50] E. Gil, R. Bailón, 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, vol. 57, no. 5, pp. 1079–1088, 2010.

[51] B. Xie and H. Minn, "Real-time sleep apnea detection by classifier combination," IEEE Transactions on information technology in biomedicine, vol. 16, no. 3, pp. 469–477, 2012.

[52] H.-C. Chang, H.-T. Wu, P.-C. Huang, H.-P. Ma, Y.-L. Lo, and Y.-H. Huang, "Portable sleep apnea syndrome screening and event detection using long short-term memory recurrent neural network," Sensors, vol. 20, no. 21, p. 6067, 2020.

[53] J. Huang, J. Zhang, S. P. Lam, S. X. Li, C. K. W. Ho, V. Lam, M. W. M. Yu, and Y.-K. Wing, "Amelioration of obstructive sleep apnea in rem sleep behavior

disorder: implications for the neuromuscular control of osa," Sleep, vol. 34, no. 7, pp. 909–915, 2011.

[54] M. Shokrollahi, S. Krishnan, D. Kumar, and S. Arjunan, "Chin emg analysis for rem sleep behavior disorders," in 2012 ISSNIP Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living (BRC), pp. 1–4, IEEE, 2012.

[55] R. Ferri, J.-F. Gagnon, R. B. Postuma, F. Rundo, and J. Y. Montplaisir, "Comparison between an automatic and a visual scoring method of the chin muscle tone during rapid eye movement sleep," Sleep medicine, vol. 15, no. 6, pp. 661–665, 2014.

[56] Y.-M. Luo, J. Tang, C. Jolley, J. Steier, N.-S. Zhong, J. Moxham, and M. I. Polkey, "Distinguishing obstructive from central sleep apnea events: diaphragm electromyogram and esophageal pressure compared," Chest, vol. 135, no. 5, pp. 1133–1141, 2009.

[57] H. Al-Angari, "Evaluation of chin emg activity at sleep onset and termination in obstructive sleep apnea syndrome," in 2008 Computers in Cardiology, pp. 677–679, IEEE, 2008.

[58] J. W. Burns, F. B. Consens, R. J. Little, K. J. Angell, S. Gilman, and R. D. Chervin, "Emg variance during polysomnography as an assessment for rem sleep behavior disorder," Sleep, vol. 30, no. 12, pp. 1771–1778, 2007.

[59] H. Escola, P. Gaillard, M. Jobert, and C. Tismer, "Automatic detection of ocular movements during sleep with an application to a sleep-disorders study," in 1992 14th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 6, pp. 2738–2739, IEEE, 1992.

[60] C. Cortes and V. Vapnik, "Support-vector networks," Machine learning, vol. 20, no. 3, pp. 273–297, 1995.

[61] J. R. Quinlan, "Simplifying decision trees," International journal of man-machine studies, vol. 27, no. 3, pp. 221–234, 1987.

[62] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," The American Statistician, vol. 46, no. 3, pp. 175–185, 1992.

[63] I. A. Basheer and M. Hajmeer, "Artificial neural networks: fundamentals, computing, design, and application," Journal of microbiological methods, vol. 43, no. 1, pp. 3–31, 2000.

[64] S. Haykin, Neural networks: a comprehensive foundation. Prentice Hall PTR, 1994.

[65] R. Polikar, "Ensemble based systems in decision making," IEEE Circuits and systems magazine, vol. 6, no. 3, pp. 21–45, 2006.

[66] T. G. Dietterich, "Ensemble methods in machine learning," in International workshop on multiple classifier systems, pp. 1–15, Springer, 2000.

[67] T. K. Ho, "Random decision forests," in Proceedings of 3rd international conference on document analysis and recognition, vol. 1, pp. 278–282, IEEE, 1995.

[68] A. J. Ferreira and M. A. Figueiredo, "Boosting algorithms: A review of methods, theory, and applications," in Ensemble machine learning, pp. 35–85, Springer, 2012.

[69] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785–794, 2016.

[70] D. H. Wolpert, "Stacked generalization," Neural networks, vol. 5, no. 2, pp. 241–259, 1992.

[71] S. C. Bagui, "Combining pattern classifiers: methods and algorithms," 2005.

[72] W. Flemons, D. Buysse, S. Redline, A. Oack, K. Strohl, J. Wheatley, T. Young, N. Douglas, P. Levy, W. McNicolas, et al., "Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research," Sleep, vol. 22, no. 5, pp. 667–689, 1999.

[73] N. Pombo, N. Garcia, and K. Bousson, "Classification techniques on computerized systems to predict and/or to detect apnea: A systematic review," Computer Methods and Programs in Biomedicine, 2017.

[74] D. S. Morillo, J. L. Rojas, L. F. Crespo, A. León, and N. Gross, "Poincaré analysis of an overnight arterial oxygen saturation signal applied to the diagnosis of sleep apnea hypopnea syndrome," Physiological measurement, vol. 30, no. 4, p. 405, 2009.

[75] G. C. Gutiérrez-Tobal, M. L. Alonso-Álvarez, D. Álvarez, F. del Campo, J. Terán-Santos, and R. Hornero, "Diagnosis of pediatric obstructive sleep apnea: Preliminary findings using automatic analysis of airflow and oximetry recordings obtained at patients' home," Biomedical Signal Processing and Control, vol. 18, pp. 401–407, 2015.

[76] G. Aljadeff, D. Gozal, V. L. Schechtman, B. Burrell, R. M. Harper, and S. L. Davidson Ward, "Heart rate variability in children with obstructive sleep apnea," Sleep, vol. 20, no. 2, pp. 151–157, 1997.

[77] A. Travaglini, C. Lamberti, J. DeBie, and M. Ferri, "Respiratory signal derived from eight-lead ecg," in Computers in Cardiology 1998, pp. 65–68, IEEE, 1998.

[78] P. De Chazal, 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," IEEE Transactions on Biomedical Engineering, vol. 50, no. 6, pp. 686–696, 2003.

[79] P. Xavier, K. Behbehani, D. Watenpaugh, and J. R. Burk, "Detecting electroencephalography variations due to sleep disordered breathing events," in

Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE, pp. 6097–6100, IEEE, 2007.

[80] L.-W. Hang, H.-L. Wang, J.-H. Chen, J.-C. Hsu, H.-H. Lin, W.-S. Chung, and Y.-F. Chen, "Validation of overnight oximetry to diagnose patients with moderate to severe obstructive sleep apnea," BMC pulmonary medicine, vol. 15, no. 1, p. 24, 2015.

[81] H. N. Yoon, S. H. Hwang, J. W. Choi, Y. J. Lee, D. U. Jeong, and K. S. Park, "Slow-wave sleep estimation for healthy subjects and osa patients using rr intervals," IEEE journal of biomedical and health informatics, 2017.

[82] A. Jafari, "Sleep apnoea detection from ecg using features extracted from reconstructed phase space and frequency domain," Biomedical Signal Processing and Control, vol. 8, no. 6, pp. 551–558, 2013.

[83] D. Alvarez, R. Hornero, J. V. Marcos, and F. del Campo, "Multivariate analysis of blood oxygen saturation recordings in obstructive sleep apnea diagnosis," IEEE Transactions on Biomedical Engineering, vol. 57, no. 12, pp. 2816–2824, 2010.

[84] K. Xu, Y. Lu, T. Yamaguchi, T. Arie, S. Akita, and K. Takei, "Highly precise multifunctional thermal management-based flexible sensing sheets," ACS nano, vol. 13, no. 12, pp. 14348–14356, 2019.

[85] C. Wang, C. Pan, and Z. Wang, "Electronic skin for closed-loop systems," ACS nano, vol. 13, no. 11, pp. 12287–12293, 2019.

[86] P. Bi, X. Liu, Y. Yang, Z. Wang, J. Shi, G. Liu, F. Kong, B. Zhu, and R. Xiong, "Silver-nanoparticle-modified polyimide for multiple artificial skin-sensing applications," Advanced Materials Technologies, vol. 4, no. 10, p. 1900426, 2019.