STATISTICAL MODELS FOR DETECTING EXISTENCE OF OBSTRUCTIVE SLEEP APNEA, PREDICTING ITS SEVERITY, AND FORECASTING FUTURE EPISODES

By

WORANAT WONGDHAMMA

Bachelor of Engineering in Electrical Engineering
Mahidol University
Nakhon Pathom, Thailand
2003

Master of Science in Industrial Technology
Texas A&M University – Commerce
Commerce, Texas
2006

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December, 2015

STATISTICAL MODELS FOR DETECTING

EXISTENCE OF OBSTRUCTIVE SLEEP APNEA,

PREDICTING ITS SEVERITY, AND FORECASTING

FUTURE EPISODES

Dissertation Approved:

Dr. Sunderesh Heragu

Dissertation Adviser

Dr. Williams Kolarik

Dr. Tieming Liu

Dr. Sundar Madihally

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude and deepest appreciation to my advisor, Dr. Sunderesh Heragu. He has been supportive and has given me the freedom to pursue my research without objection. His guidance has taken me from the point where I thought that I may have to drop out of school to the point where there are only a few steps left for me to finish the degree program. It has been a great honor for me to be one of his Ph.D. students.

I would like to express my heartfelt gratitude and infinite appreciation to my former advisor, Dr. Satish Bukkapatnam. I thank him for introducing me to the wonderful world of scientific research. I thank him for his support, guidance, encouragement, and patience. He was always there to listen and to give advice. He taught me how to be persistent in order to accomplish any goal. It has been a great honor for me to be one of his Ph.D. students.

In addition to my advisors, I would like to express my sincere gratitude to my dissertation committee members, Dr. William Kolarik, Dr. Tieming Liu, and Dr. Sundar Madihally for their support, encouragement, and insightful comments throughout my Ph.D. study. It has been my fortune to be able to receive guidance from these prodigies.

Special thanks are due to Akkarapol Sa-ngasoongsong for his unending friendship and support. I also want to thank all my friends and colleagues, especially Hoang Tran, Trung Le, Changqing Cheng, and Prahalad Rao for their great friendship and the encouragement that I needed.

I am deeply thankful to my wife, Suruetai Paisitthanyaphong, for her understanding and love. Her support and encouragement were in the end what made this dissertation possible. I would like to especially thank my sisters, Aumpira Wongdhamma, and Nivimar Wongdhamma, for taking care of every responsibility I have back home. Without their support, I would not be able to put my all into finishing this dissertation. I will forever be thankful to my aunts, Kalyanee Boulabeiz, and Nanta Ross, for their generous care and support. With their help, my home away from home has become a lot more welcoming.

Finally and most importantly, I would like to express my eternal gratitude and love to my beloved parents, Dr. Thonglaw Wongdhamma, and Dr. Wallapa Pimthong, for their unconditional love, support, and patient. They have selflessly sacrificed their lives for my sisters and for me. Nothing in this world can be more fortunate than being their son.

Dedicated to my beloved parents

Acknowledgements reflect the views of the author and are not endorsed by committee members or Oklahoma State University.

Name: WORANAT WONGDHAMMA

Date of Degree: DECEMBER, 2015

Title of Study:

STATISTICAL MODELS FOR DETECTING EXISTENCE OF OBSTRUCTIVE SLEEP APNEA, PREDICTING ITS SEVERITY, AND FORECASTING FUTURE EPISODES

Major Field: INDUSTRIAL ENGINEERING AND MANAGEMENT

Abstract:

This dissertation presents three statistical models based on data mining and nonlinear time-series analysis techniques as an alternative method for the diagnosis and treatment of obstructive sleep apnea disease (OSA). From a diagnosis perspective, our method reduces the time and cost associated with the conventional method by first screening a non-OSA subject from the population, then individually determining the OSA's severity by utilizing the data from a single-lead electrocardiogram (ECG) device that is worn overnight at the subject's location. Our OSA forecasting model can be used to activate an OSA therapy device such as a continuous positive airway pressure (CPAP) machine or a hypoglossal nerve stimulator (HNS) as needed or before an OSA episode so that the latter can be averted in real time.

In particular, our contributions are:

1) Detect the existence of OSA in an individual based on the pattern of biological physiology and simple clinical data with a low false negative rate and reasonable accuracy (FNR: 5.3%, Accuracy: 84.47%). People with some degree of probability of having OSA will be confirmed by the next model.

2) Determine the OSA severity by classifying the OSA episode (event) from one-lead ECG data collected overnight (accuracy: 92.26% with 10,052 equally sampled events from 24 subjects). The advantage of our model is that the variations (i.e., different body build, age, gender, activity, health conditions, and race) have very little effect on the prediction because the neighboring patterns in the reconstructed phase spaces have very little or no correlation to those variations. This benefit can be seen from our model's performance compared to two other models that exist in the literature.

3) Forecast an incoming OSA episode in real time using the one-lead ECG data (accuracy: 92%, 88%, and 87% for 1, 5, and 10 minutes ahead). This forecasting model with any appropriate OSA episode prevention device (i.e., HNS, and just-in-time CPAP) will allow for an effective OSA treatment method for CPAP nonadherence OSA sufferers.

4) Develop a wearable device that can collect the biological data via a single-lead ECG as a home sleep test (HST) device.

TABLE OF CONTENTS

Chapter Page

Chapter Page

Chapter Page

LIST OF TABLES

Table Page

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

CHAPTER I

INTRODUCTION

1.1 Research motivation

Sleep is a crucial part of life. The human body becomes fatigued during the day because of numerous activities and rejuvenates itself during the night while sleeping, creating a daily life cycle of degradation and renewal. In general, during the resting or sleep state, the body constantly executes many involuntary actions dealing with recovery, regeneration, and realignment functions. Except for the degradation from the normal aging process, sleeping is like a machine in a maintenance state. Good sleep fosters a good working state for the body. Conversely, a disturbed sleep will not restore the body to its normal working state. Because activities need to be performed during the day, lack of sleep induces more fatigue in the body. Sleep disorders prevent the body from rejuvenating. One form of sleep disorders is sleep apnea, a common disorder marked by frequent pauses in breathing or shallow breaths during sleep. The most prevalent form of apnea, called obstructive sleep apnea (OSA), is due to a partly or completely obstructed airway. Clinically, OSA is identified as a major risk factor for hypertension, arrhythmias, stroke, myocardial infarction, congestive heart failure, and death [3-8]. Approximately 1 in 15 adults, or about 18 million Americans, have moderate or severe OSA [9] and more than half of them remain undiagnosed. An estimated 50 to 70 million Americans suffer from chronic sleep apnea [10], and hundreds of billions of dollars are spent each year in direct medical costs for screening and treatment [11].

OSA has subtle observable symptoms during the day and, more importantly, it is almost impossible for the person with OSA to realize that he or she has the disease because it occurs when the person is asleep. A vast majority of OSA patients seek treatment and/or receive the diagnosis after their condition becomes moderate or severe. Unlike standard hypertension or diabetes, in which a test is performed routinely during an annual physical exam, a sleep apnea diagnosis is made only after a patient expresses sleep discomfort or upon a doctor's recommendation. The gold standard for OSA screening and diagnosis involves administering polysomnography (PSG) or a sleep study. PSG involves the patient spending the entire night in a sleep clinic with many sensors attached to several parts of the body for recording several biological signals. To complete the study, sleep apnea episodes (i.e., impeded or difficult breathing events) are marked manually by a sleep specialist or a sleep doctor by looking for specific patterns in the multiple bio-signal time-series collected overnight. The purpose of a PSG is to determine the severity of the sleep apnea condition by noting how often a subject stops breathing (apnea) on average in each hour of sleep. Sleep clinics are known to collectively perform 1.17 million screening tests per year in the U.S. [11], a very low number compared to the number of individuals with OSA, estimated at 70 million. The waiting time for diagnosis and screening of suspected patients ranges from 2 to 10 months [11]. Baseline estimates of 5-year diagnosis and treatment charges for a patient with OSA are about $4,210 [12]. The OSA diagnosis process mentioned above has two key problems. 1) Because of the limited equipment and number of facilities for PSG, a relatively small percentage of the OSA population can be tested. 2) Each apnea episode must be manually and individually marked by a sleep technician or sleep doctor, a laborious task that requires a significant amount of time to complete. There is a clear need to expedite the diagnosis process to reduce the overall medical costs and the adverse impact on an individual's health.

Currently, with the advancement of the microelectromechanical systems (MEMs) and information technology, acquisition of the signals used in a sleep study no longer need to be confined to the sleep laboratory or hospital. We have designed a wearable sensor with just a one-lead electrocardiogram (ECG) that patients can wear during the night in their normal environment to collect the signals needed for a PSG.

Also, the availability of the OSA's clinical diagnosis data is more and more accessible. This opens up the opportunity to develop a precise offline and online diagnosis tool.

1.2 Research objective

The objective of this research is to develop predictive models to detect OSA, determine its severity, and forecast obstructive sleep apnea episodes in real time.

- Detecting OSA: develop a simple OSA screening rule with considerable accuracy by uncovering the pattern of biological physiology and simple clinical data.
- Determine OSA severity: develop an automatic OSA episode prediction method by tracking nonlinear dynamic cardio-respiratory coupling using mainly a one-lead ECG signal for a home sleep test (HST) device.
- Forecast OSA episodes in real time: develop an effective forecasting model using nonlinear dynamic cardio-respiratory coupling features to predict a forthcoming OSA episode for improving OSA treatment efficiency.

Another objective of this research is to develop a wireless wearable sensor to collect the data required for the objective above.

CHAPTER II

BACKGROUND

This chapter, describes the basic background concerning the sleep process, a pathology of the upper airways as it relates to OSA, symptoms and effects of OSA, and a standard for OSA diagnosis. The purpose of this chapter is to provide a basic background about OSA for those who are not familiar with the disease.

2.1 Sleep process

Sleep is an essential part of life. The fatigue due to any active activity that has accumulated during the day is eliminated during the night by a sleep process. In general, throughout this resting stage, in order to maintain good health, many important involuntary mechanisms of the body automatically deal with the recovery, regeneration, and realignment functions unhampered by interference from any voluntary mechanism [13].

In the past, most people believed that sleep was a passive process. Actually, it was found in the 1950s that sleep is a dynamic activity in which our brains are very active during the process [14]. Sleep can be divided into two major states, rapid eye movement (REM) sleep and non-rapid eye movement (NREM) sleep [15-17]. As the name suggests, one of the main differences between these states is the movement activity of the eyes. The NREM sleep can be further divided into four stages depth of sleep increases [15]. In some literature, stages 3 and 4 are grouped together and called slow wave sleep or deep sleep. The characteristics of each sleep stage are described by different combinations of Electroenchalograph (EEG), Electrooculograph (EOG), and Electromyelograph (EMG) as shown in Table 1 [16, 18].

In normal adults, NREM sleep stages predominate during the first third of the night, accounting for 75-80% of normal sleep time (1-5% in stage 1, 45-50% in stage 2, 12% in stage 3, and 13-15% in stage 4); then REM sleep predominates in the last half of the night, accounting for 20-25% of normal sleep time [16]. The length of a normal cycle of sleep starts with NREM sleep stages 1 to 4. REM sleep is about 90-110 minutes [14]. However, this pattern can vary with factors such as age, medical history, physiology, and sleep disorders [15]. The plot of sleep stages versus sleep time is called a hypnogram. The hypnogram of a typical healthy sleep in adults is shown in Figure 2.1.

Table 2.1: EEG, EOG, and EMG characteristics during awake and different sleep stages [16]

Stage	EEG	EOG	EMG
Awake	Predominant alpha (8-13 Hz) activity (more than 50% of epoch) mixed with EEG beta (>13 Hz)	Slow and rapid	High
NREM 1	Theta (4-7 Hz) activity, sometimes with vertex sharp waves	Slow	Decreased from awake
NREM 2	Theta (4-7 Hz) activity and sleep spindles or K complexes	None	Decreased from awake
NREM 3	High-voltage (>75 uV) delta (<3.5 Hz) activity accounting 20-50% of the EEG activity	None	Decreased from awake
NREM 4	High-voltage (>75 uV) delta (<3.5 Hz) activity accounting more than 50% of the EEG activity	None	Decreased from awake
REM	Low-voltage, mixed-frequency activity	Rapid	Almost absent

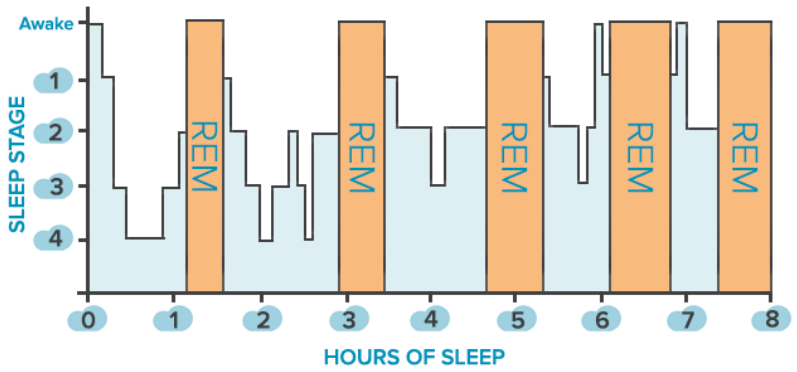


Figure 2.1: Stages of healthy sleep
(Reproduced from [19])

Sleep does not mandatorily follow each stage in sequence. However, in normal cases, after the transition from wakefulness, sleep begins in NREM 1, developing to deeper stages later, NREM stages 2, 3, and 4. After NREM 4, the sleep stage returns to repeat NREM 2 before reverting to REM sleep. Once REM sleep is complete, the sleep stage normally goes back to NREM 2. It is believed that the first two

cycles of NREM sleep in the beginning of the night and the REM sleep cycles at the end of the night are the most important sleep stages. As seen from Figure 2.1, deep sleep occurs mostly in the first two cycles of sleep. REM sleep dominates in the later parts of the night by prolonging REM sleep and shortening NREM sleep. During NREM sleep, cell division, protein synthesis, and astrocyte maturation are maximal and many of the anabolic drives including the secretion of growth hormone peak during deep NREM sleep (stages 3 and 4) [15]. These activities contribute to the overall physical rest and reinforcement of the immune system. Additionally, the metabolic rate, sympathetic nervous system activity, heart rate, cardiac output, and systemic vascular resistance fall during NREM sleep; that is, NREM sleep is a state of cardiovascular relaxation [20]. It has been suggested by [16, 21] that the most negative impact on daytime wakefulness and performance is due to sleep disturbance during the NREM sleep stages. However, this does not mean that the REM sleep is not important. It was found that rats whose lifespans are normally 2-3 years survive only about 5 weeks on average if they are deprived of REM sleep [14]. In contrast to NREM sleep, the characteristics of REM sleep are closer to the wakefulness state in terms of the variability of the control systems' outputs [15]. As in deep NREM sleep, protein production increases in REM sleep. The REM sleep process relates to several regions of the brain such as cerebral cortex, which is responsible for learning, thinking, and organizing information, and the pons, which disables the neurons in the spinal cord resulting in limb muscle paralysis [14]. REM sleep contributes more to psychological rest and long-term emotional well-being and may also maintain memory function [18].

2.2 Obstructive sleep apnea (OSA)

Sleep disturbances are very common and occur in several forms such as insomnia (difficulty of sleep initiation or maintenance), circadian rhythm disorders, and sleep-related disorders. However, one disorder that differs from the others, in that patients are not be able to see the symptoms themselves, is sleep apnea. Sleep apnea is a common and chronic disorder marked by frequent pauses in breathing or shallow breaths during sleep. These pauses in breathing cause arousals which alter the changes in the normal sleep stage in such a way that the sleep stage reverts to a shallower stage instead of moving to a deeper stage. In

severe cases, sleep apnea creates arousals that hinder the normal sleep process, on average, more than 50 times per hour of sleep. The most prevalent form of sleep apnea is called obstructive sleep apnea (OSA), because it is caused by a partly or completely obstructed airway. An apnea episode is characterized by at least a 10-second cessation of respiratory air flow (apnea) or a severely reduced flow with a 4% drop in blood oxygen level (hypopnea) [22]. An OSA episode may last a few seconds or minutes and episodes can recur up to one hundred times in an hour. Technically, OSA severity is commonly defined by the number of apnea and hypopnea episodes per hour of sleep (apnea-hypopnea index, AHI) [23]. Clinically, OSA is identified as a major risk factor for hypertension, arrhythmias, stroke, myocardial infarction, congestive heart failure, and death [3-8].

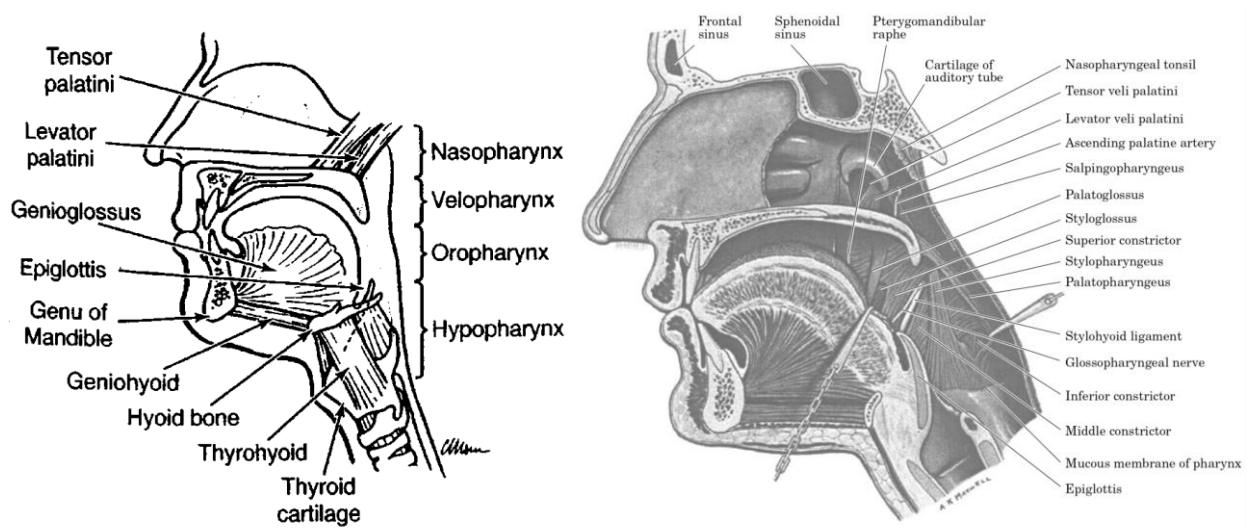
2.2.1 Pathophysiology of the upper airway (UA) in OSA patients

Two main factors predispose a person to OSA episodes, the anatomy of the upper airway (UA) and the control mechanism of related muscles from the brain. In this study, we focus only on OSA disorders in the adult population, which is the largest group of OSA patients. Unlike other animals, the human body has a very unique upper airway, which is collapsible to accommodate respiration, speech, and deglutition functions [24].

Shown in Figure 2.2, the pharyngeal airway has very complex structures containing many muscles and bones which range from the nasal choanae to the epiglottis. The air enters the airway mainly from the nasal passages through four pharyngeal sub-sections: the Nasopharynx (NP), the Velopharynx (Retropalatal oropharynx), the Oropharynx (Retroglossal oropharynx), and the Hypopharynx [24-26]. Four muscle groups interactively control the patency of the upper airway lumen: 1. the muscles regulating the position of the soft palate (alai nasi, tensor palatini, levator palatine), 2. the tongue (genioglossus, geniohyoid, hyoglossus, styloglossus) 3. the hyoid apparatus (hyoglossus, ginioglossus, digastric, geniohyoid, sternohyoid) and 4. the posterolateral pharyngeal walls (palatoglossus, pharyngeal constrictors) (see Figure 2.2 A and B) [27].

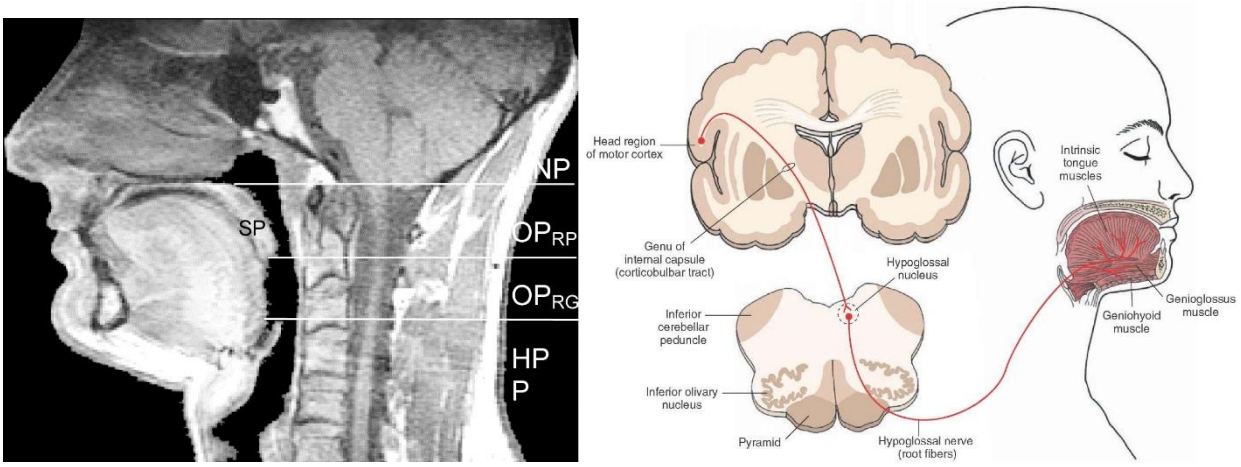
The most confined region that potentially contributes to the collapse of the airway is the retropalatal in the Oropharynx region (OP_{RP}) as shown in Figure 2.3 A. The main points of interest in this area are the extrinsic tongue muscle (genioglossus muscle), soft palate, and lingual tonsils where there is no bony structure for these muscles to attach to. During the wakefulness state, these muscles receive activation signals from several parts of the brain in order to stabilize their contractions and stiffness to be ready to serve their functions properly. The largest dilator muscle here is the genioglossus (GG) muscle, which is the back part of the tongue and contributes to the collapse of the airway.

The GG contraction is controlled by action potential or nerve impulses from the widening of the hypoglossal nerve shown in Figure 2.3 B. This contraction causes the anterior movement of the tongue and pharyngeal airway [28]. The control system's inputs in this case are from the respiratory pattern generator, negative pressure receptors within the upper airway, chemical receptors (O2 and CO2), and the neurons that regulate the wakeful or sleep state [25]. During inspiration, the intraluminal pressure in the upper airway becomes relatively negative, which induces the collapsible muscles to loosen, resulting in a narrower airway. The human respiratory control system counters that problem by stiffening and contracting most upper airway muscles to enlarge the airway. Several studies show near-maximal action potential burst activity (muscle tone) from the hypoglossal nerve to the GG muscle during inspiration, which moves the back part of the tongue forward to open up the airway [24-26, 28]. This behavior also applies to most upper airway muscles such as the alae nasi (the expanded outer wall of cartilage on each side of the nose), the tensor palatini, the stylopharyngeus, and the styloglossus (see Figure 2.2) but with different timing, preceding the beginning of inspiration [29]. However, this ability to react to the threat of a collapsible airway tends to be absent after the body state has shifted from wakefulness to sleep onset, NREM and REM sleep [24, 25, 30].



(A) (B)

Figure 2.2: (A) Anatomy of the upper airway showing the main pharyngeal segments (reproduced from [27]) (B) Anatomy of the upper airway muscles (reproduced from [26])



(A) (B)

Figure 2.3: (A) Midsagittal magnetic resonance image (MRI) of the head of an adult (reproduced from [24])
(B) Origin and distribution of the hypoglossal nerve (cranial nerve XII) with its innervation of the muscle of the tongue (reproduced from [31])

In OSA, the sleep stages affect the upper airway's ability to maintain the patency in various ways due to mechanical and chemical stimuli. Several studies [24, 30, 32] found that OSA typically occurs in NREM sleep stages 1 and 2 and in REM sleep. The results of the investigations also suggest that during NREM sleep, the GG muscle activation is modulated mainly by a combination of hypercapnia (a condition of abnormally elevated carbon dioxide (CO2) levels in the blood) and inspiratory resistive loading. In other words, unlike in the wakefulness state, the GG muscle is relatively insensitive to either chemical (O2, CO2

levels in the blood) or mechanical (UA resistive loading) stimuli alone during NREM sleep [32]. However, mild OSA patients or heavy snorers seem to be able to regain airway patency when they are in NREM sleep stages 3 and 4 [30]. Many researchers have suggested that OSA occurring during NREM stages 1 and 2 is most likely from the instability of the respiratory control mechanisms due to the transition from wakefulness to sleep [33, 34]. When the sleep stage shifts into REM sleep, with the change in the body control mechanism, the brain shuts off the body's muscles, including the GG muscle and other UA muscles [30]. The lack of the muscle contraction mechanism caused by the paralysis, also leads to the airway obstruction in the case of people whose UA is small. The review above covers the main pathophysiology of OSA in the adult population; however, other, less common factors, could cause a collapse of the airway containing the pharyngeal anatomy, the ventilation control stability, the respiratory arousal threshold, and the lung volume [25].

2.2.2 Obstructive sleep apnea cycle

Obstructive sleep apnea (OSA) episodes can recur throughout the sleep cycle. The severity depends on how well the body maintains the patency of the airway with the changes in body control mechanisms during each sleep stage. Summarized from [35], the OSA cycle is explained by the diagram in Figure 2.4.

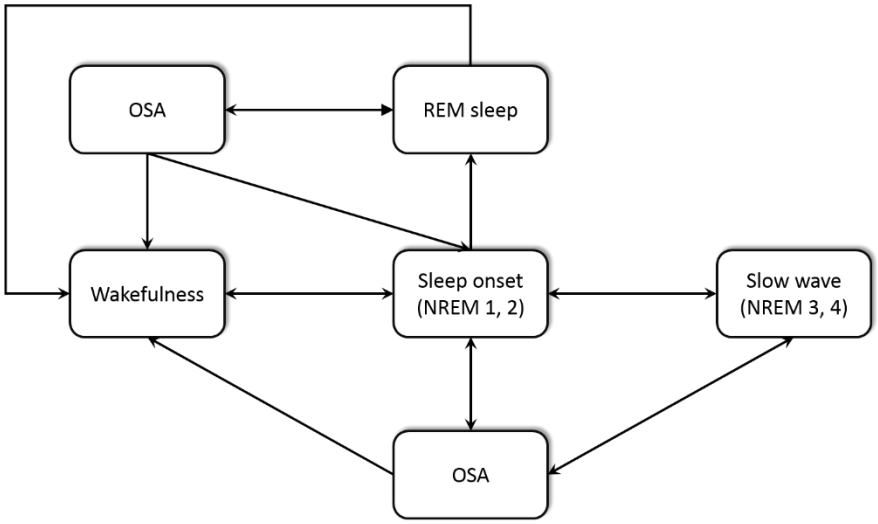


Figure 2.4: State transitions between sleep cycles and OSA

The sleep process starts from a wakefulness state and moves to sleep onset which is a group of NREM stages 1 and 2. As mentioned earlier, these two stages have shown instability in the respiratory control system, especially the contraction control of GG muscles, which becomes less sensitive to resistive loading and chemical stimuli (CO2 and O2). In other words, to activate the GG muscles to maintain patency, the airway resistance and chemical stimuli (CO2 and O2) detected must be higher than the normal threshold during the wakefulness state. This results in an increase in the airway resistance compared to a wakefulness state, which predisposes the airway to collapse if the breathing effort pressure is less than the critical pressure, the cut-off pressure before OSA takes place.

Two scenarios are possible when the airway is partially or fully collapsed: 1) If the person gains control of the UA muscle before reaching the body's arousal thresholds (too low O2 or too high CO2 levels in the blood), there will not be any impact on the sleep stage, 2) but if the person cannot regain the control before the arousal thresholds are reached, the brain will try to terminate the apnea by creating arousal. Arousal is a sleep interruption that lasts from 3 – 15 seconds [36]. It is believed to be a protective mechanism against asphyxia (a condition arising when the body is deprived of oxygen) during sleep. However, the result of the arousal is the lightening of sleep, shifting from a deeper sleep stage to a lighter sleep stage [37]. These repeated arousals in OSA patients lead to sleep fragmentation, poor sleep quality, and importantly, diminished amounts of slow wave (NREM 3, 4) and REM sleep.

2.2.3 Obstructive sleep apnea symptoms and consequences

It is commonly known that most cardiovascular diseases are the product of prolonged hypertension [38]. The underlying mechanisms inducing the hypertension in OSA patients are not well understood, and many theories are found in the literature. First, a prolonged sympathetic activity as a consequence of the recurrence of oxygen desaturation during apneas alters vascular function and structure, resulting in hypertension development over time [38]. This theory is supported by evidence that elevated sympathetic nerve activity is also found in OSA patients during the daytime and especially during the onset of apnea

[39]. Another well-known theory is that the elevated blood pressure during apneas causes endothelial dysfunctions, resulting in increased peripheral vascular resistance. In other words, the heart needs to pump harder to overcome the increased resistance in the blood vessels, resulting in permanent hypertension [40]. The most substantial medical consequences of prolonged hypertension are cardiovascular diseases which link directly to increased cardiovascular mortality [38-40]. Moreover, sleep impairment due to apneas causes a metabolic syndrome. That results in obesity and diabetes, which also induce cardio vascular complications [40]. The daytime effects of untreated OSA are excessive daytime sleepiness, cognitive impairment, decreased work productivity, and increased risk of accidents [41].

2.2.4 Classification of obstructive sleep apnea severity

One of the challenges associated with the diagnosis of OSA is that the OSA suspect is not able to see or investigate the direct symptoms himself; someone else must share the same room to observe the symptoms. In most cases, a spouse is the one who identifies the symptoms, such as snoring accompanied by breathing stoppage, then a gasp when breathing starts. The by-symptoms such as excessive daytime sleepiness and fatigue may be noticeable but they show up only when the OSA has already progressed to moderate or severe phases. To diagnose the disease, patients are normally screened with a set of questionnaires. If the results of the questionnaires indicate that the patient is in an OSA high risk group, he or she is referred for a polysomnography (PSG) or sleep study. PSG is the gold standard for classifying the severity of the sleep apnea, but the cost is quite high, estimated at $1,000 - $3,000 nationwide. The study is so expensive because a sleep technician needs to cooperate with an OSA suspect overnight to monitor and collect the required bio-rhythm signals: cardiac activity (via electrocardiogram (ECG)), brain activity (via electroencephalography (EEG)), visual (eye) movement (via electrooculogram (EOG)), muscle activity (via electromyography (EMG)), airflow rate, oxygenation, respiratory movement, and body position [42]. After collecting all needed data, a sleep technician or sleep doctor manually marks the OSA episodes (impeded or difficulty in breathing; by definition, either obstructive apnea or hypopnea) by looking for specific patterns in the multivariate signals, i.e., complete loss of respiratory activity for at least 10 seconds

followed by a decrease in blood oxidation or brain and muscle arousal activity. To complete the study, the number of apnea and hypopnea events per hour of sleep or an apnea-hypopnea index (AHI) is calculated to quantify the severity of OSA. The steps for an AHI calculation [17] are as follows: (1) add the total number of apnea and hypopnea events, (2) divide that number by the total number of minutes of actual sleep time, and (3) multiply the result by 60. The severity of sleep apnea is classified as follows: AHI < 5 - no sleep apnea, 5 ≤ AHI ≤15 – mild sleep apnea, 15 ≤ AHI ≤ 30 – moderate sleep apnea, and AHI > 30 - severe sleep apnea. With labor-intensive procedures, the PSG can be administrated to only about 5 cases a night in a normal-to-big sleep laboratory. The waiting time for diagnosis and screening of suspected patients ranges from 2 to 10 months [11]. Baseline estimates of 5-year diagnosis and treatment charges for a patient with OSA are $4,210 [12]. With its high cost and complexity, PSG is not accessible and affordable to everyone. This wide gap between current methods for sleep studies, and the need means that balancing the complexity and cost with a practical, rational alternative will offer more opportunities to everyone to be able to have a diagnosis in periodically basis. There is a clear need to expedite the diagnosis process to reduce the overall medical costs and the adverse impact on an individual's health. Following are feasible several feasible solutions: 1) using other sleep-monitoring devices for out-of-center (OOC) testing, 2) implementing an automatic OSA episode classification method instead of manual classification by a sleep technician in both a conventional sleep study (PSG) and an out-of-center (OOC) or home sleep test, or 3) establishing a protocol to periodically diagnose people who are at risk of having OSA as a preventive measure.

CHAPTER III

REVIEW OF LITERATURE

The history of the sleep apnea disorder started around the 19th century, when the term "pickwickian" was used to describe an obese somnolent patient [43]. At that time, the research concentrated only on the patient's obesity rather than on the breathing disorder during sleep. Gastaut et al. [44] were the first scientific investigators who, in 1966, showed evidence of recurrent sleep apnea episodes in pickwickian patients using a polygraphic or what we call polysomnography. Today, although we know more about the sleep apnea disorder, a vast number of people with the disease remain undiagnosed. One of the main reasons is that the diagnosis process is quite complex. Traditionally, sleep apnea suspects not only need to go through a normal clinical examination during the day but also need to spend an entire night at a sleep clinic to be completely diagnosed. This chapter provides a review of the sleep apnea literature as follows: sleep apnea screening, sleep apnea severity diagnosis processes, forecasting of sleep apnea episodes, and the wearable devices for detecting a sleep apnea episode.

3.1 Sleep apnea screening processes

Screening is the first step in the sleep apnea diagnosis process. The purpose is to filter high probability sleep apnea suspects from the normal population and recommend them for further diagnosis, usually, a PSG. The screening is most often initiated in one of two ways, with the direct complaint of a subject who demonstrates sleep apnea's symptoms such as an excessive daytime sleepiness and hypertension, or with a subjective impression from a doctor. In this section, the literature is divided into three categories: 3.1.1 screening questionnaires, 3.1.2 clinical diagnostic models, and 3.1.3 screening devices. All study results were compared to PSG at the end to validate the accuracy. Each type of screening has advantages and disadvantages. Screening devices are the most accurate but require a data collection (mostly overnight) and interpretation process. Clinical diagnostic models tend to give reasonable accuracy but the complications of real-life implementation have made them not as popular. Questionnaires give the least overall accuracy and the results are quite inconsistent. However, because of its simplicity and low cost, this screening method is always the first choice for physicians [42].

3.1.1 Screening questionnaires

In the screening questionnaire category, in general, the questionnaires were developed with regard to symptoms that patients with sleep apnea are most likely to have. Then, a scoring method was derived from the results of modeling techniques such as logistic regression and multivariate regression used for investigating risk factors. Finally, a simple set of rules to best screen people with a high risk of having sleep apnea was developed from the screening method. The first popular set of questionnaires was the Wisconsin sleep questionnaires. Originally, in 1993, Young et al. [45] used a logistic regression method to analyze data from the Wisconsin Sleep Cohort to estimate the prevalence of undiagnosed sleep-disordered breathing among adults. A random sample from a general population (no history of sleep disorders) of 602 subjects was studied by an overnight PSG. The study came up with a simple set of questions asking about snoring habits and evidence of sleep apnea. This set of questionnaires is popularly called the Wisconsin sleep questionnaires. The performance of the questionnaires in the general population for screening of OSA

(AHI≥5) is 79% sensitivity and 46% specificity (28% positive predictive value (PPV) and 89% negative predictive value (NPV)) and for screening of moderate OSA (AHI≥15) is 87% sensitivity and 40% specificity (11% PPV and 97% NPV) [46].

In 2006, Sharma et al. [47] studied the prevalence and risk factors of OSA in an Indian population in Delhi, India. They used a modified version of the Wisconsin sleep questionnaires which also included questions about demographics, sleep symptoms, medical history, and medications. Of 2,150 respondents, 554 were classified as habitual snorers with a high risk of OSA and 1,596 were classified as non-habitual snorers with a low risk of OSA. To validate the performance of the questionnaires, 77 OSA suspects underwent a PSG and 36 subjects were identified as having OSA. Among the non-OSA suspected cases, 74 subjects underwent a PSG and 2 subjects were identified as having OSA. The performance of the modified Wisconsin sleep questionnaire in the general population for screening of OSA (AHI≥5) is 95% sensitivity and 64% specificity (46% PPV and 97% NPV).

The Berlin questionnaire, an outcome of the Conference on Sleep in Primary Care held in April 1996 in Berlin, Germany, became another popular tool to screen the general population to identify patients with sleep apnea. Besides height, weight, age, and gender, the questions asked about snoring history, daytime sleepiness, sleep apnea, and high blood pressure history. In 1999, Netzer et al. [48] recruited 100 subjects of 744 who responded, to undergo a PSG in order to validate the accuracy of the questionnaire. A logistic regression model was used to examine the relative effects of age, gender, BMI, blood pressure (high/low), and neck circumference. The performance of the questionnaires in the general population for screening for OSA (RDI≥5) is 86% sensitivity and 77% specificity (88% positive predictive value (PPV) and 72% negative predictive value (NPV)) and for screening of moderate OSA (RDI≥15) is 54% sensitivity and 97% specificity (97% PPV and 48% NPV).

From the reported results, the Berlin sleep questionnaire seem to perform better in classifying the sleep apnea suspect group than the Wisconsin questionnaire. However, in 2006 and 2008, studies were performed by Chung et al. [49] (211 subjects) and Sharma et al. [50] (104 subjects) to validate the performance of the Berlin questionnaire in subjects without a history of sleep disorders (general

population). The results from both studies are as follows: in Chung's study, screening of OSA (AHI≥5) is 69% sensitivity and 56% specificity (77% positive predictive value (PPV) and 44% negative predictive value (NPV)) and for screening for moderate OSA (AHI≥15) is 79% sensitivity and 50% specificity (50% PPV and 78% NPV) [46, 51] and in Sharma's study screening of OSA (AHI≥5) is 85% sensitivity and 95% specificity (96% positive predictive value (PPV) and 81% negative predictive value (NPV)). Comparing the three studies, the results are quite inconsistent.

In a perioperative period, patients with a sleep apnea disorder have different characteristics such as sensitivity to anesthetic agents, and adverse postoperative events. For these reasons, in 2006, Gross et al. [52], in the name of the American Society of Anesthesiologists (ASA), developed a checklist composed of a set of questionnaires for anesthesiologists to evaluate patients with OSA. The checklist's question categories were not so different from those in the Berlin questionnaire but the questions were more detailed. Validated by Chung et al. [51], using AHI≥15 as a cut-off value for OSA, the sensitivity and specificity were 79% and 37% respectively (45% PPV and 73% NPV).

In 2008, Chung et al. developed two sets of questionnaires [49] called the STOP and the STOP-Bang questionnaires (S-snore, T-tired, O-apnea observed, P-high blood pressure, B-body mass index (BMI), A-age, N-neck circumference, and G-gender). The STOP questionnaire was developed to be self-administered by patients but the STOP-Bang was clinician-administered. Using AHI≥15 as a cut-off value for OSA, the sensitivity and specificity of the STOP were 74% and 53% respectively (51% PPV and 76% NPV), and the sensitivity and specificity of the STOP-Bang were 93% and 43% respectively (52% PPV and 90% NPV).

Shown in Tables 3.1 and 3.2, the average performance of each questionnaire popularly used for OSA screening is depicted in terms of sensitivity, specificity, PPV, and NPV. All questionnaires except the Berlin, tend to focus on capturing true negative cases more than true positive case (specificity > sensitivity). This makes sense in clinical testing because in reality, we do not want to have high false negative rates (people who are diagnosed as not having OSA but actually have it). However, economically, high false positive rates (people who are diagnosed as having OSA but do not have it) generate more cost and a longer

waiting time for patients. Also, some studies [53-55] have found that clinical symptoms such as snoring, a history of sleep apnea, and daytime sleepiness are not good predictors of OSA in children. In [56], in particular, the results suggested that there is no statistical difference in age, gender, and OSA symptoms between children with OSA and normal children. Thus, by fixing the model performance to have low false negative rates which already makes sense, there is still a need to improve the accuracy in term of true positive cases in order to reduce the cost associated with doing a PSG on people who actually do not have OSA. Also, OSA in children seems to have different characteristics than OSA in adults so there is still a need to create a suitable prediction rule to be used exclusively for children.

Table 3.1: Questionnaire performance for screening OSA patients using ≥5 as a cut-off value

Questionnaires	Total N.	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Wisconsin (average from [45], [47])	753	87	55	37	93
Berlin (average from [48], [50], [51])	415*	80	76	87	65.67
ASA [51]	211*	72	38	72	38
STOP [49]	177*	66	60	78	44
STOP-Bang [49]	177*	84	56	81	61
Total	1345**				
All average		77.8	57	71	60.33

*Some studies ([51], [49]) validated several questionnaires performance at the same time.
** include only non-intersected subjects.

Table 3.2: Questionnaire performance for screening OSA patients using ≥15 as a cut-off value

Questionnaires	Total N.	Sensitivity (%avg.)	Specificity (%avg.)	PPV (%avg.)	NPV (%avg.)
Wisconsin [45]	602	87	40	11	97
Berlin [48], [51]	311*	66.5	73.5	73.5	63
ASA [51]	211*	79	50	50	78
STOP [49]	177*	74	53	51	76
STOP-Bang [49]	177*	93	43	51	90
Total	1090**				
All average		79.9	51.9	47.3	80.8

*Some studies ([51], [49]) validated several questionnaires performance at the same time.
** include only non-intersected subjects.

3.1.2 OSA screening models

Because the performance of the screening questionnaires is not satisfactory, many prediction models for OSA screening have been developed by researchers and engineers. Most of the research has been in develop OSA screening models using popular clinical predictors and demographic data such as weight, height, wrist circumference, age, and gender to predict the prevalence of obstructive sleep apnea in each individual. With the same idea as screening questionnaires, if the results indicate that the subjects are likely to have OSA, they are referred to be diagnosed using PSG. In addition, some groups have developed methods to classify OSA patients using a signal collected directly from the individual. This approach definitely yields better results because the individual's dynamics are directly quantified. However, the overnight data collection process is still needed. This make no difference to the OSA severity measure process from PSG.

Various types of data have been used in attempting to develop simple prediction rules to quickly screen people with OSA. For medical research groups, clinical data such as the body mass index (BMI), age, gender, waist and neck circumference, coexisting medical conditions, and life styles are important variables [5-13]. Linear multivariate methods are commonly used for modeling. Kwiatkowska et al. [57] generated at obstructive sleep apnea prediction rule by combining the knowledge of a physician and the data driven analysis. The BMI, gender, age, and hypertension are fed into decision trees for prediction. The best model performances in term of overall accuracy reported were 69.81% accurate, 82.5% sensitivity and 50% specificity. In their next research [58], they integrated a Semio-fuzzy framework to automatically develop OSA prediction rules by creating a self-learning algorithm. However, the performance of the developed model was not reported. In addition to the basic demography and clinical data, Dixon et al. [59] used two biomedical factors, fasting insulin, and glycosylate to their multivariate linear and logistic analysis. They concluded that neck circumference could replace BMI and gender in the analysis. Ramachandran et al. [60] derived their prediction rules from 10 clinical variables using a logistic regression. Then, from the weight of each variable based on its importance in the model, a scale called the perioperative

sleep apnea prediction (P-SAP) was established for its ease of use. The reported performances showed a very good sensitivity, 93.9 %, but very poor specificity, 32.3%.

Some supervised machine learning methods for OSA classification were used in this task. Chen et al. [61] applied a support vector machine (SVM) for classification of normal and different stages of obstructive sleep apnea severity. From the PSG data collected overnight, they found that the oxygen desaturation index (ODI) alone was the best predictor of severe obstructive sleep apnea. Their method and model of classifying people with OSA yielded 42.86% sensitivity, 94% specificity, and 87.72% accuracy (high accuracy but very low sensitivity might come from unbalanced targets, OSA < non-OSA). Some researchers have used artificial neural networks for obstructive sleep apnea prediction. Kirby et al. [62] came up with a generalized regression neural network (GRNN) to classify and predict patients with obstructive sleep apnea using 23 variables from demographic data, clinical features, medical history, daytime sleepiness history, and bed partner observations. The performance of their model was quite remarkable, with an accuracy of 91.3%, sensitivity of 89.9%, and specificity of 80% (the number of subjects with OSA was more than twice the number of normal subjects (281 vs 124)).

Although, these studies showed good results in predicting obstructive sleep apnea from basic clinical data, some of them did not consider prior probabilities in a model training process. Using an unbalanced target dataset to train a binary classifier, the model will tend to fit itself to the target with a majority number of observations. As is well known in the data mining community, such models are impracticable. Also, these methods still lack practical accuracy and the ability for early detection of obstructive sleep apnea. The current approaches for prediction and prognosis of sleep apnea are based on the population's statistics and not on individual dynamics.

3.2 OSA episode classification models

OSA episode monitoring and detection research has also received considerable attention in the literature. Several OSA detection approaches based on correlating the statistical patterns of heart rate, respiration rate, and oxygen saturation (SpO2) signals during OSA episodes have been reported [63-65].

As mentioned in the previous chapter, normally sleep apnea episodes are detected by means of capturing specific patterns, loss of respiratory air flow which results in loss of blood oxygen desaturation, with arousals at the apnea episode termination. To do so, the multichannel bio-signals recording such as PSG is required. However, because of the limited equipment and number of facilities for PSG, a relatively small percentage of the OSA population can be tested. To solve this problem, an OSA episode testing device should be very simple and easy to use (i.e., fewer recording channels needed and mobile) to enable the subject to self-administrator the test at home. Methods of OSA episode prediction along with those devices are also needed. Thus, in this section, the review of literature focuses on the literatures that proposes sleep apnea episode prediction methods using a single channel bio-signal recording. In most of the literature, a single ECG signal was used because the ECG gives so much information, not only from the cardiovascular system but also the respiratory and autonomic nervous control system. In brief, the workflow found in the OSA episode prediction methods utilizing an ECG signal is as follows:

- One-lead ECG (time-series)
    - Transformation applied
        - R-R interval time-series.
            - Detection of ECG's QRS complex.
        - ECG derived respiration (EDR) time-series.
            - R peak magnitude cubic spline interpolation
            - Wavelet transformation
    - Features extraction (quantification of segmented time-series)
        - Heart rate variability (HRV)
        - Wavelet transformation
        - Basic statistics of the signal of interest
    - Prediction (classification) model development
        - Regression models (multiple regression, and logistic regression)

        - Artificial neural network
        - Support vector machine
        - K-nearest neighborhood
        - Manual thresholding of the developed predictor(s)
    - Validation method
        - K-fold cross validation
        - Leave one out
    - Performance measurements
        - Percent event classification accuracy
        - Misclassification rate
        - Area under receiver characteristics curve (AUC)
        - True positive, and true negative rate (sensitivity and specificity)

The rise in the popularity of minute-by-minute sleep apnea episode prediction was due to the release of an apnea dataset from the Physionet database in 2000. This dataset was released by a collaboration of Penzel et al. [66, 67] in a part of a challenge to develop a method to detect OSA episodes from the given dataset. One approach to using the hidden information is to decompose a signal to different levels based on the instantaneous frequency and choose only the pieces that directly relate to an OSA episode.

One popular choice found in the literature is wavelet transformation. Based on the one-lead ECG data provided, one of the participant groups, Raymond et al. [68] combined information from an ECG-derived respiration (EDR) signal using a wavelet decomposition method and R-R interval time-series. The EDR signal was then decomposed using discrete harmonic wavelet decomposition (DHWT). The authors captured the arousal after the termination of the apnea episode from a tachy/bradycardia cycle in the heart rate using median powers of EDR DHWT levels 11 and 12. To classify an apnea episode, the sum of these two powers was calculated, then compared with the median signal power during minute k. If these powers are larger than twenty percent of the median, the minute k and the first half of minute k+1 will be considered as having evidence of arousal from an apnea episode. Using this criterion, the accuracy of the OSA episode

classification was 81 %. This research showed that the decomposition of EDR signal to several levels based on the instantaneous frequency could be used as a predictor to differentiate OSA episodes from normal breathing.

Using a single lead ECG for deriving an instantaneous heart rate (IHR), then decomposing the signal using another type of wavelet transformation, Delibasoglu et al. [69] discovered that the variance of the 5^{th}, 6^{th}, and 7^{th} detail components decomposed by a wavelet decomposition method were good predictors for OSA episode detection. Eight levels of Daubechies order 3 were used in the wavelet transformation. The decomposition was applied to each data segment with a length of 6 minutes. They found that the 4^{th} minute was the deciding minute for an OSA episode. A nonlinear autoregressive (NARX) type artificial neural network (ANN) was then trained for classification of OSA episodes. The reported ANN architecture is 6, 3, and 1 neurons for input, hidden, and output layers respectively. Levenberg-Matquardt error back propagation is used as a training optimization algorithm. The best accuracy achieved in their method was 82.58 % for minute-based OSA episode classification without reporting the true positive and false positive rates. With the same idea, Roche et al. [70] extracted features from heart rate fluctuations using wavelet decomposition. Then, they used classification and regression trees for obstructive sleep apnea prediction. The results show that the sensitivity and specificity of this method are higher than 90%.

Using another type of transformation method, not a wavelet transformation, to extract features that correlate with the OSA episode, Mietus et al. [71] applied a Hilbert transform to the filtered normal sinus to normal sinus (NN) time-series that derived from the R-R interval. The signal results from a transformation data window of 5 minutes were analyzed. The average, standard deviation, and time fraction of the amplitudes and frequencies were extracted. The thresholds for detection of sleep apnea episodes were developed from those features, yielding 82.1% accuracy of OSA episode classification.

Rather than using transformation techniques to extract useful information from an ECG signal as described above, some groups have found that statistical quantifications calculated from the R-R interval are also good predictors of OSA episodes. Laiali et al. [72] developed a prediction model to detect the OSA episode from one-lead ECG using standard heart rate variability (HRV) features such as mean, standard

deviation, and inter quartile range of partitioned R-R time-series as inputs for the classification technique. A support vector machine (SVM) classifier, a supervised machine learning method that finds the optimal separating plane to separate each target class in the feature space, was trained and tested on sleep apnea recordings. Their best results were reported as 96.5% accuracy for detection of OSA episodes from 15-second data partition.

However, this reported performance is questionable. The authors did not clearly report the number of data points and the ratio of the two targets used in the training and testing processes. Generally, the performance of the model is heavily affected by these two factors. If not enough data is selected for training and testing the model, the trained model may not be generalized enough to be used in other data. Because it does not capture all possible patterns of the OSA episodes. Also, if the ratio of target variables (apnea and non-apnea episodes) is not equal, without any weight correction, the classifier will focus on capturing the events with the majority ratio because this will lead to better results (i.e., event classification accuracy) which are not practically true. For example, if there are 4000 non-apnea and 1000 apnea events, the model training process will focus on classifying a true positive non-apnea event rate with a weight of four times more than capturing a true positive apnea event rate, leaving a higher false positive apnea event rate because it gives a better overall classification performance which is actually deceptive.

Mendez et al. [73] selected three sets of features used for OSA episode prediction based on physiological evidence. R-R interval time-series and the area under QRS were used as starting signals. A bivariate time-varying autoregressive model was applied to each R-R interval and QRS area time-series to get a power spectral density (PSD) and coherence in three frequency bands: very low frequency (VLF: 0 – 0.04 Hz), low frequency (LF: 0.04 – 0.15 Hz), and high frequency (HF: 0.15 – 0.4 Hz). Module and phase were defined by relationships between the R-R interval and QRS area time-series. The other features described by the authors, mean, variance, kurtosis, and skewness, were calculated from the R-R interval and QRS area time-series minute by minute. Two methods for feature selection were used in this case: a wrapper approach through a leave-one-out (LOO) and k-nearest neighbor (KNN)-Model selection. Overall, ten features, mostly power spectrum density (PSD) in different frequency ranges, were selected based on

their contribution in the OSA classification model. The most important feature is the ratio of a very low frequency (VLF) power to the total power of the R-R interval time-series. Two modeling methods, an artificial neural network (ANN) with 3-30 neurons and KNN, were trained by splitting the data into three groups, training, validating, and testing, with percentages of 60, 20, and 20 of the total data respectively. The performance of each model was evaluated by its accuracy, sensitivity, and specificity in the testing data group. The KNN model obtained its best results by using all ten feature inputs: accuracy, sensitivity, and specificity are 88, 85, and 90 percent respectively. The performance of the ANN model was comparable with accuracy, sensitivity, and specificity of 88, 85, and 90 percent correspondingly.

One of the challenges in applications for detection and prediction of OSA episode is that the dynamics of human physiology control systems change in different sleep stages. Based on that assumption, Noviyanto et al. [74] not only developed a method to predict OSA episodes but also to determine sleep quality by predicting sleep stages using information embedded in ECG signals. Overall, of thirty-nine sleep stages, correlated features were extracted from R-R intervals, ECG-derived respiration (EDR), and raw ECG signals using 18 records from the MIT-BIH polysomnographic database, Physionet. Four classifiers were used in this study: a Bayesian network, a multilayer perceptron artificial neural network, a k-nearest neighbor, and a random forest. The thirty nine features proposed, four feature sets were selected based on the degree of dependence or predictability of one variable on another. Based on Noviyanto et al.'s [74] experiment, the features derived from the raw ECG signal were better predictors than those from the R-R interval or the EDR. The best combination of features and classifier to predict sleep stages in this study were the full set of features extracted from the raw ECG (without using feature selection) and a random forest method. This combination yielded an accuracy of 80.95 % on average percent correct classification.

Because of the high correlation between cardio-respiratory control and neural control systems, Yilmaz et al. [75] developed a new feature set extracted from R-R interval time-series called a mean absolute deviation (MAD) to predict sleep stages and sleep apnea episodes. Three classification methods, k-nearest neighbor (kNN), support vector machine (SVM), and quadratic discriminant analysis (QDA), were trained using the MAD features based on PSG data collected from seventeen subjects. The authors

concluded that these features could be used in sleep stage prediction in both healthy and OSA groups. However, the performance for classification of some sleep stages is still questionable.

Rather than using single lead ECG as the main signal to detect sleep apnea episodes, Tagluk and Sezgin [76] used the electroencephalogram (EEG) in this application. A bispectrum analysis was used to identify phase relationships between each EEG subband, namely, delta, theta, alpha, beta, and gamma, resulting in quantifications of the quadratic phase coupling (QPC). For OSA episode classification, an artificial neural network (ANN) model was trained with the architecture of 5x300 input layers, 10 hidden layers, and 2x1 output layers. The accuracy of OSA episode classification reported in the literature was 96.15 %. Although the performance of this method showed significant accuracy in detecting the OSA episode, the implementation of EEG signal acquisition required much more complicated hardware than that used for obtaining the ECG signal. Also, because several sensors must be mounted directly on the user's scalp and temple, the reliability of data collected at home by a user is questionable. For these reasons, the ECG signal is still more suitable in practice.

Because OSA symptoms can be observed directly by a sound produced from the upper airway system, based on breathing sounds, Doukas et al. [77] attempted to develop a preliminary tool for detection of sleep breath disorders. In their experiment, three microphones were installed over the bed in order to capture subjects' breathing and snoring sounds. The breathing and snoring events were then classified by computing the likelihood of their presence based on Laplacian and Gaussian distributions of snore and breathing signals. To classify an OSA event, an appropriate threshold is applied to the probability of OSA the calculated by a state diagram of sleep breath disorder detection from each data segment. However, the authors reported only the detection of snore events from the recorded sounds, and concluded that this system could be used as a preliminary remote assessment method in case patients present both OSA and snoring because the snoring events are closely related to OSA episodes.

With the same assumption, Mikami [78] analyzed the nonlinear dynamic of snoring sounds. Although the results cannot predict obstructive sleep apnea, he concluded that the sounds from a simple snorer have higher nonlinearities than those of obstructive sleep apnea patients. Focusing on the direct

breathing evidence for apnea and hypopnea prediction using empirical mode decomposition and Hamming distance on the airflow signal, Robinson and Conway [79] achieved 94.53% sensitivity and 100% specificity for their prediction model. However, the airflow signal is one of the hardest signals to obtain because in order to directly measure the in and out airflow from the nostrils, which is one direct evidence of OSA, the sensor must be clipped directly onto the user's nose. Thus, this sensor type is not a popular choice in the literature.

Most of the reported studies were done solely on apnea episode detection in adults. However, a sleep apnea disorder can develop in humans of every age. Williamson et al. [80] focused on developing an on-line method to predict OSA in infants. In their study, three physiology signals, namely a respiratory signal from an abdominal effort belt, a one-lead ECG, and the oxygen saturation percentage from a pulse oximeter, were collected and used for manually marking OSA episodes by sleep doctors. The authors computed thirteen features from the R-R interval (RRI) from the ECG signal and inter-breath intervals (IBIs) from the respiratory signal. These features were also extracted from RRI and IBIs from several time delay steps to be used for detection of OSA episodes in current time (i.e., features of IBIs and RRI in time k-5, k-4, …, k were used for prediction of an OSA episode in time k). Because the total number of extracted features was high and many features were highly correlated, a principle component analysis (PCA) was applied to all features mainly in order to reduce the feature dimensionality. Six principle components (PC) were used for training an equal-prior quadratic classifier (QC) to classify the apnea state. By varying the detection cutoff probability of each subject, the area under the receiver operating characteristic curve (AUC) was generated and used as a prediction performance indicator. The best model trained by all features mentioned earlier performed with an overall AUC of 73%. However, in this study, the prediction time window resolution is 5.5 minutes, whereas other studies always used a one-minute time window. Another limitation of this method is that the OSA episode could be predicted only after recording data for three minutes and ten and a half minutes after the previous apnea detection. Focusing on the online detection of abnormal breathing patterns in adults, Varady et al. [81] developed an online pattern recognition method capable of classifying three different breathing patterns: normal breathing, hypopnea, and apnea. In their

study, two forms of respiration signals, a nasal airflow (NAF) and a thoracic or abdominal respiratory effort (RIP), captured by a thermistor and inductance plethysmography were used for deriving the instantaneous respiration amplitude (IRA) and instantaneous respiration interval (IRI). Four different feed-forward artificial neural network (ANN) models were developed to coordinate four feature sets, 25 features from NAF, 50 features from NAF and RIP, 50 features from IRA and IRI, and 100 features from IRA and IRI, respectively. The reported results showed that the features extracted from IRA and IRI were good predictors for OSA episode classification.

Rather than detection of OSA episodes, Maier and Dickhaus [82] developed a method to detect apnea episodes specifically resulting from the compromised regulation of the central respiratory drive, called central sleep apnea (CSA). The EDR signal was extracted from a one-lead ECG signal using the QRS area method explained in [83]. With the assumption that the modulation of the respiratory signal in ECG is absent during CSA episodes, the recurrence analysis was then applied to the EDR signal to analyze the respiratory dynamics. A phase space reconstruction of the respiratory dynamics was carried out with both dimension (D) and time delay (Tau) of 3. Then, to obtain the recurrence plot, the phase space was projected onto a 2-dimensional plot by thresholding the distance between each point to every point in the phase space. The only points within a threshold that remain in the plot are regarded as true neighbors. The literature reports that the pattern of points influenced by a CSA episode was best detected when fixing thresholds that gave a recurrence rate of 20 percent. To classify the CSA episodes, a vertical line pattern in the recurrence plot called the trapping time (TT), was constantly calculated for each minute epoch. Each CSA episode was detected when the value of the TT exceeded the threshold set by the author. The best detection performance reported in this case using a one-lead ECG signal had a sensitivity of 82.8 % and a positive predictive value (PPV) of 41%. Babaeizadeh et al. [84] derived the respiratory signal from the ECG signal and used it to classify obstructive sleep apnea based on the length of time that the individual stops breathing.

Although there have been many significant advancements in this field of research as seen from the results reported above, the credibility of some of the results is questionable because the reported validation

process is not clear. For example, many studies did not pay attention to the stratification of the proportion of a target used for training their models. In an unbalanced sample of the target situation (the sample number of target 0, non-apnea event, is not equal to that of target 1, apnea event), without any decision cost (weight) associated with each target to balance the model decision, the trained model will automatically be biased toward the target that has a larger percentage in the total sample.

Also, the accuracy of some OSA episode detection methods is hampered by the challenges associated with capturing and quantifying the nonlinear dynamics of the physiological process underlying the measured signals. These detection methods use information from each signal in isolation (i.e., features are extracted from one signal); they do not attempt to capture the dynamic coupling and interrelationships of the signals that can be used to improve OSA detection accuracy. There is a need for further research to study deeply the changes in individual dynamics that cause obstructive sleep apnea. Then, with a good understanding of the interactions and the ways in which the system evolves to develop obstructive sleep apnea, based on that knowledge, it may be possible to individually predict the risk of before the person's dynamics advance into matured obstructive sleep apnea. Finally, to enable preventive measures to intervene in the recurrence of OSA episodes, a real-time OSA episode detection method is needed. In the literature, only one OSA online prediction method [80] was found. There is still a clear need to develop a method that optimizes the information collected in real time to detect and forecast OSA episodes.

3.3 OSA episode forecasting models

To prepare a counter process to prevent an OSA episode from happening, it is crucial to be able to systematically forecast the future OSA state in order to minimize the intervention measures that could also wake the person up during the OSA prevention process. The very first group found in the literature who attempted to forecast sleep apnea was Bock and Gough in 1998 [85]. However, their perspective was not to forecast the sleep apnea episode but the time-series of respiratory signals that may show sleep apnea episode patterns. In their study, the time-series of heart rate, respiration, and oxygen saturation were used as predictor signals. An Elman recurrent neural network was trained with the followed parameters:

momentum, 0.001; learning rate, 0.40; and one hidden layer with eighteen neurons. The maximum prediction step reported was 32 data samples of the respiration signal, which was equal to 16 seconds. The authors quantified the performance of their forecasting model by comparing the largest Lyapunov exponent and correlation dimension of the actual respiratory signal and the predicted respiratory signal during apnea events. The reported Lyapunov exponent error and correlation dimension were thirteen and nine percent respectively. However, in our opinion, the performance error quantification used in this research would be very hard to use in a real application.

In 2010, Waxman et al. [86] developed a forecasting method to predict OSA episodes for thirty seconds to two minutes into the future by utilizing EEG, ECG (by means of the heart rate variability (HRV)), nasal airflow pressure, oronasal temperature, submental EMG, and EOG from PSG. The assumption in this study was that the pattern of the PSG before sleep apnea onset could be used to predict sleep apnea episodes. After segmenting the data into several durations, 30, 60, 90, and 120 seconds, all signal segments were demeaned. A discrete wavelet transform was then applied, resulting in new sets of signals. The authors did not report which wavelet transformation method was used to transform the data and which levels of signals were used for training the model. The features used for training the artificial neural network were the amplitude and timing of each of the three minima and three maxima, the ratio between the mean of the three maximal amplitudes and the mean of the three minimal amplitudes, the root-mean-square (RMS) value, and the RMS value relative to that of the original signal. Large-scale memory storage and retrieval (LAMSTAR) was used in training the classifier. The best performance reported was as follows: 30-seconds apnea forecasting achieving sensitivity and specificity up to 80.6 ± 5.6 and 72.8 ± 6.6 percent during NREM sleep, and 74.4 ± 5.9 and 68.8 ± 7.0 percent during REM sleep. However, the performance of forecasting OSA episodes further into the future was less accurate, with sensitivity and specificity around 50 percent in two-minute forecasting of an episode during REM sleep.

To the best of our knowledge, little research has thoroughly investigated the area of forecasting sleep apnea episodes. Also, the forecasting of sleep apnea episodes utilizing a single signal was not found in the literature.

3.4 Wireless wearable sensory systems for monitoring of sleep apnea episodes

In this section, prior research on wearable wireless systems for monitoring respiratory and/or heart related signals that can be used for out-of-center (OOC) OSA testing are reviewed. The first, is a tight, full-body suit referred to as WEALTHY that uses a conductive fabric to collect five ECG signals and a piezoresistive fabric to capture respiration and movements [87]. The second device, from the MyHeart project, is a wearable sensor device that emphasizes heart condition monitoring [88]. This device embeds conductive fabric as an electrode to sense the ECG signals in five positions. Similar to the WEALTHY system, piezoresistive fabric is implemented to detect respiration signals. The MagIC [89] combines conductive fabric and piezoresistive fabric into a vest. Textile electrodes and strain-gauge fabric are used for obtaining two-lead ECG and respiratory signals. The research in [87-89] used the new technology of conductive fabric as ECG electrodes. However, the quality of the signals obtained by the conductive fabric is far inferior to the quality from conventional electrodes because of a very high sensitivity to noise from movements and the pressure needed at measurement points for a quality signal. For more serious applications such as classification of OSA episodes, detection of myocardial infarction, and ambulatory heart monitoring, the reliable electrodes used in a clinical procedure are more suitable.

Another group of researchers has developed a miniaturized wearable device to monitor both heart and respiratory activities. A wrist device called AMON is a monitoring and alert system designed for cardiac/respiratory patients [90]. The device can monitor a one-lead ECG, SPO_2, blood pressure, and one body position. Another wireless wrist device, HealthGear [91], is an SPO_2 sensor developed for sleep apnea event detection and heart rate monitoring. In this case, the performance of AMON is not satisfactory because the sensing points are only around the wrist, not at appropriate points (i.e. chest ECG signals). The HealthGear [91] simultaneously measures oxygen saturation level and heart rate. These signals, while useful, are not adequate to monitor the entire cardiovascular system. Another group of devices is those equipped with a multi-sensor array. A wearable device, LifeShirt system from Vivometrics [92], is composed of a fit-type jacket, two inductive respiratory bands, standard electrodes for a one-lead ECG, pulse oximeter, and recorder unit. Another multi-sensor device, Bioharness [93], is a chest belt for real-

time monitoring of ECG, respiration, heart rate, skin temperature, posture, and activity. These systems show high quality signals in the literature, but both are limited to a one-lead ECG, which is restricted to extracting only heart rate variability. To this point, the devices outlined in are the only wearable systems that have the ability to collect bio-signals during sleep. Still, to the best of our knowledge, the ability to process the collected signals in real time to predict the sleep apnea episode has yet to be found in the literature.

CHAPTER IV

RESEARCH METHODOLOGIES

This chapter provides a background and reviews the methodologies used in all the analysis reported in this study.

4.1 Decision trees

A decision tree (DT) is a non-parametric supervised machine learning technique that is simple, yet very powerful because the concept is easy to understand and it is effectiveness in modeling. A decision tree is a set of rules that is optimized for directed data mining [94]. The goal of the decision tree is to assign a class to the target field based on the majority values of the input variables. The tree is built by splitting the records at each node according to the best cut-off value of an input variable. Starting from the first node, the tree determines which of the input fields make the best split. The algorithm is as follows [95]:

Given: A training sample of N observations on class variable C that takes values [1, 2, …, k] and p predictor variables, $X_1, \ldots$, Xp

Goal: Find a model for predicting values of Y from new X values.

4.1.1 Algorithm:

1. Start at the root node
2. Do a split search. If X is an interval variable, each unique value of X could serve as a potential split point. If X is a categorical variable, the average value of the target is taken within each categorical input level. This average value serves as a potential split point.
3. For each split point, two groups are generated. Cases with input values less than the split point are put in the left branch. The ones with input values equal to or greater than the split point are classified into the right branch. Compute the purity measure of the split branches (left and right) from each split point. Choose X and the corresponding split point that maximizes the purity measure.
4. Compute the purity measure of the split branches (left and right) from each split point. Choose X and the corresponding split point that maximize the purity measure.
5. If the stopping criterion is reached, go to step 6. Otherwise, apply step 2 to each branch.
6. Prune the tree by comparing the average square error in the training and validation datasets.

For binary dependent (target) variable ($Y \epsilon [0, 1]$) , in this study, the probability Chi-square is used as a purity measure. At the splitting point, the number of cases in the left and right branches is counted, forming a 2x2 contingency table with the column specifying the branch direction (left or right) and the row specifying the target value (0 or 1). A Pearson Chi-square statistic is used to quantify the independence of the counts in each table's column.

$$\chi^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{4.1}$$

Where O_{ij}= Observed frequency of row i and column j in the contingency table

E_{ij}= Expected frequency of row i and column j in the contingency table

$= \frac{(R_i x C_j)}{N}$

N = Total number of observations

R_I = Row i^{th} marginal total

C_j = Column j^{th} marginal total

The chi-square (χ^2) value indicates a difference in the proportion of counts in the left and right branches. A large χ^2 suggests that the proportions of 0 and 1 in the left and right branches are different, which is a good split. It is also used as a criterion for stopping the splitting. By converting to a p-value (df=1), assuming identical target (0, 1) proportions in each branch, the p-value indicates the likelihood of obtaining the observed value of the statistic [94]. Suggested by [94], for the split to occur, this p-value must exceed a threshold of 0.2. However, for a large dataset, these p-values can be very small so that the quality of the split will be reported by:

$$\text{Logworth} = -\log(\text{p-value}) \tag{4.2}$$

(Corresponding to the p-value of 0.2, a logworth is approximated 0.7)

For a binary target, as in our case (non-apnea = 0, apnea = 1), some other impurity indexes are widely used, such as [96]:

$$\text{Gini Impurity index; } i(p) = 1-p_1^2-p_0^2 = 2p_1(1-p_1) \tag{4.3}$$

Where p_0 is the proportion of subjects without apnea disorder

p_1 is the proportion of subjects with apnea disorder

The theory behind the Gini impurity index is that when two cases are chosen at random with replacement from a node, the probability that both cases are subjects without apnea disorder is p_0^2 and that both are subjects with apnea disorder is p_1^2. Thus, $1-p_1^2-p_0^2$ reflects to the probability that the two cases chosen at random with replacement are different. The Gini index can take a minimum value of zero when the node is pure and maximum value of 0.5 when both targets are equal in the node. Another impurity index is:

$$\text{Entropy; } i(p) = -\sum_{i=0}^{1} p_i log_2(p_i) \tag{4.4}$$

The rarity of an event is measured as $-\log_2(p_i)$. A large rarity measure means the event is rare. To interpret this equation, suppose the probability of subjects with apnea in the node is 0.01 so that conversely, the probability of subjects without apnea in the node is then 0.99. Then, the entropy is $i(p) = -[0.01*\log_2(0.01) + 0.09*\log_2(0.09)] = 0.0808$. Compared to a case where the probabilities of subjects with

and without apnea are equal, the entropy of this case is $i(p) = -[0.5*\log_2(0.5) + 0.5*\log_2(0.5)] = 1$. These results show that, when the proportion of the targets is most impure (equal number of subject with and without apnea in the same node), the entropy will take the value of 1 but when the node purity approaches to maximum (only one case in the node) the entropy will approach the value of zero.

4.1.2 Stopping rules and the pruning process

As suggested in the Chi-square splitting rule, the DT will stop the splitting process when the logworth of the split is less than the threshold. However, a significant level from the logworth is not the only option for limiting the tree's size. Some other stopping criteria are the number of observations needed to split the nodes, the maximum depth size that the lead can grow, and the maximum number of splitting nodes.

These stopping criteria are just the initial mechanisms to prevent an over-fitting problem. The last process for the DT is called "pruning". After the last splitting node, the tree is called a "maximal tree" (see Figure 4.1). The pruning process removes the splits in the tree that are not relatively significant. The pruning algorithm is as follows [96]:

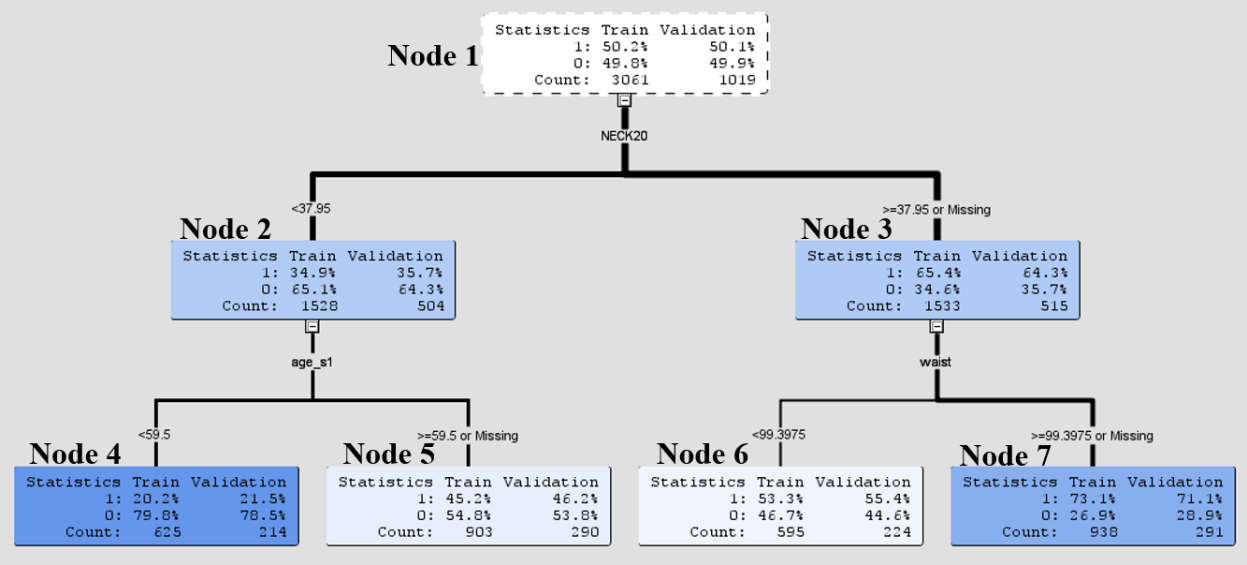


Figure 4.1: Maximal tree from an apnea/non-apnea classification

1. From the maximal tree with M leaves shown in Figure 4.1 (M=4), if we remove one split at the end of the tree, we will get a sub-tree size M-1. For example, removing splitting point age_s1 leaves a sub-tree size 4-1 = 3 which are nodes 2, 6, and 7. This sub-tree is called sub-tree_3.1 (the first 3-node sub-tree). Next, removing the splitting point waist also leaves a sub-tree size 3 called sub-tree_3.2 (the second 3-node sub-tree) with nodes 4, 5, and 3. Thus, from the maximal tree in this case, we have two cases of sub-tree sized 3.
2. We continue the process until reaching a sub-tree size 1. Next we look for the nodes that leave sub-trees size 2. By removing, the splitting point age_s1 and waist, there is a sub-tree size 2, called sub-tree_2.1 with nodes 2 and 3 which is the only sub-tree size 2.
3. The process corresponding to Figure 4.1 reaches a sub-tree size 1 by removing the splitting point NECK20. What is left is only node 1, or the root node.
4. We calculate the assessment criteria from each sub-tree and choose the sub-tree that gives the best validation assessment criteria with the smallest sub-tree.

From the algorithm explained above, we will have 2 three-node sub-trees, 1 two-node sub-tree, and 1 one-node sub-tree. The assessment criterion used in this application is the misclassification rate. The validation misclassification rate is calculated for each sub-tree as shown in Table 4.1 below.

Table 4.1: Summary of validation misclassification of each sub-tree from Figure 4.1

Tree	**Number of leaves**	**Validation Misclassification**
Root node	1	1-0.501 = 49.90%
Sub-tree_2.1 (node 2,3)	2	1-[0.643x504+0.643x515]/1019 = 35.7%
Sub-tree_3.1 (node 2,6,7)	3	1-[0.643x504+0.554x224+0.711x291]/1019 = 35.71%
Sub-tree_3.2 (node 3,4,5)	3	1-[0.785x214+0.538x290+0.643x515]/1019 = 35.71%
Sub-tree_4.1 (node 4,5,6,7)	4	

From the validation misclassification rate calculated in Table 4.1, sub-tree_4.1, sub-tree_3.1, and sub-tree_3.2 should be pruned because there is no gain in accuracy by going beyond subtree_2.1. Thus, the optimal tree in this case is the tree with nodes 1, 2, and 3.

4.1.3 Variable importance

One advantage of a DT is that the relative importance of input variables is also calculated after the maximal tree is constructed. The relative importance of input variable v in sub-tree T is computed by the square root of the summation of a product between the agreement of node τ in sub-tree T and a reduction of the sum square error from the predicted values as shown below [97]:

$$I(v;T) = \sqrt{\sum_{\tau \epsilon T} a(s_v, \tau) \, \Delta \, SSE(\tau)} \tag{4.5}$$

Where s_v indicates the primary or surrogate splitting rule using variable v.

$a(s_v, \tau)$ is an agreement measurement for the rule using variable v in node τ.

$a(s_v, \tau)$ = 1 if $s_v$is the primary splitting rule

= agreement measurement if s_v is a surrogate rule

= 0 otherwise

The agreement measurement is the proportion of the training samples assigned to the same branch after the split results from the primary splitting variable and the surrogate splitting variable at a particular τ node.

$\Delta \, SSE(\tau)$ is the reduction in the sum of square errors from the predicted values.

$\Delta \, SSE(\tau)$ = $SSE(\tau) - \sum_{b \in B(\tau)} SSE(\tau_b)$

In our case where the target variable is binary, SSE is computed by

$$SSE(\tau) = N\left(1 - \sum_{j=1}^{J} \left(\hat{p}_j{}^2\right)\right) \quad for\ training\ data \tag{4.6}$$

$$SSE(\tau) = N\left(1 - \sum_{j=1}^{J} \left(2p_j - \hat{p}_j\right)\hat{p}_j\right) \quad for\ validation\ data \tag{4.7}$$

Where p_j is the proportion of the validation data with target value j and N ($p_j \ and \ \hat{p}_j$are evaluated at node τ).

$B(\tau)$ = set of branches from τ

τ_b = child node of τ in branch b

$N(\tau)$ = number of observations in τ

δ_{ij} = 1 if Y_i = j, 0 otherwise

$\widehat{p_j}(\tau)$ = average δ_{ij} in the training data in τ

The intuitive idea for this formula is to evaluate the effects of a variable that reduces the prediction's sum square error in the tree. The variable that contributes greatly in correctly predicting the target variable will have a relatively high importance in the model compared to other variables, as reflected by the decreased sum square error. Another popular method of evaluating the variable importance is the Gini variable importance (GVI) [98]. The idea is almost similar to the previous case, but the GVI evaluates the variable importance by looking at the Gini impurity measurement reduced by the particular splitting variable over the tree. The average of the Gini variable importance in tree T when the target is a categorical variable can be computed as follows [98]:

$$\widehat{GVI}(X_i) = \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{J} d_{ij}\, I_{ij}\right) \tag{4.8}$$

$$d_{ij} = \hat{G} - \left(\frac{N_L}{N}\hat{G}_L + \frac{N_R}{N}\hat{G}_R\right) \tag{4.9}$$

$$\hat{G} = 2\hat{p}(1-\hat{p}) \tag{4.10}$$

Where d_{ij} = the decrease in impurity produced by variable X_i at j[th] node of the t[th] tree

N = the number of observations at node j

N_R, N_L = the number of observations of the right and left nodes after splitting

$\hat{G}_L, \hat{G}_R$ = the Gini indexes of the left and right node

I_{ij} = 1 if $s_v$is the primary splitting rule

= agreement measurement if s_v is a surrogate rule

= 0 otherwise

4.1.4 Surrogate splitting rule and imputation

One of the advantages of a decision tree is that, unlike a regression model that needs a complete case (no missing data) to do an analysis, a DT can handle missing values in the dataset by using a surrogate splitting rule when records in the primary splitting process are missing. To choose which variables could be used as surrogate variables, the rank of agreement is evaluated at each splitting point. The first surrogate variable is the variable that has the highest agreement with the primary variable. The agreement measure between a main splitting variable and a surrogate variable is the proportion of observations in the training samples that are assigned to the same branches [97]. Cases with missing or unseen values that were not used in the primary splitting rule but are used in the surrogate rule at the same node are also counted as observations not assigned to the same branch. If there are also missing cases in the best or first surrogate variable, the next or second surrogate variable is then used instead, and so on. If all surrogate variables assigned to the nodes cannot be used because of missing cases, these cases will be assigned to the node that has a bigger proportion (the majority of the cases).

Imputation is a process for statistically filling in the incomplete cases in the dataset. Although this process is not needed for a DT modeling technique as mentioned earlier, to use other techniques such as a regression, artificial neural network, and k-nearest neighbor, complete case data with no missing values in the dataset is needed. With listwise deletion, where all cases with missing data are deleted, if there are many missing cases in the dataset, much information will be eliminated uselessly. In this case, a DT can be used for statistically predicting the missing values.

The process is described as follows [99, 100]:

1. Suppose we have n x p data matrix Y. Arrange Y such that $Y=(Y_P,Y_C)$, where the columns in Y_P are the cases where one or some observations are missing and the columns in Y_C are the complete cases observed from Y.
2. Given i = 1, 2, …, p and $Y_i = i^{th}$ column of Y_P, build a DT using Y_C as a training sample. With the DT constructed, apply each Y_i to the model. Draw a sample from the node to which each missing

value of Y_i is assigned, with predictive distribution conditioned on Y_C. Use each sample to impute each missing value in Y_i.

3. Given i = 1, 2, …, p and Y_{-i} = matrix Y that was imputed in the previous step with its i^{th} column removed, build a DT using Y_{-i} (previously imputed). With the DT constructed, apply Y_i to the model. Draw a sample from the node to which each missing value of Y_i is assigned, with selection probabilities generated by the Bayesian bootstrap [101] within each leaf of Y_{-i}.
4. Repeat step 3 for I times to check for a convergence. Recommended by [99] to use l = 10.
5. Repeat steps 1-4 m times, yielding m imputed sets.

Because this approach assumes a large sample size, it should be applied with caution. In our case, the student t test is used to statistically test whether there is a difference between the mean along with the F test for the equality of variances. A key advantage of this method is that it does not depend on the missing-data mechanism because of the non-parametric process [102]. However, the main disadvantage is a computational cost from many repetitions of the process.

4.2 Logistics regression

One limitation of a linear regression is that it cannot deal effectively with a categorical dependent variable. Because in our case, the dependent variable is binary or dichotomous (i.e., 0 = no apnea, 1 = apnea), a logistics regression which is designed especially for a model with a binary dependent variable is one of the most suitable modeling methods. Although logistic regression is one of the generalized regression models, some assumptions are quite different from those in a simple linear regression. The assumptions of a logistic regression are as follows [103]:

- A linear relationship is not required between the dependent and independent variables.
- The dependent variable must be binary.
- The independent variables can be interval and categorical. They do not need to be normally distributed, linearly related, nor of equal variance within each group.

- For categorical, independent variables, the groups must be mutually exclusive and exhaustive. A case can only be a member of only one group.

The intuitive background of a logistic regression is a simple regression with a non-linear transformation. In other words, the logistics regression is a linear regression having a dependent variable as an S-curve or non-linear transformation of a probability ranging from 0 to 1. A logistics regression model is shown below [103-106]:

Logistic regression equation:

$$logit[p(x)] = \ln\left[\frac{p(x)}{1-p(x)}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \quad (4.11)$$

Predicted probability:

$$p(x) = \frac{exp^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k)}}{1+exp^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k)}} = \frac{1}{1+exp^{-logit[p(x)]}} \quad (4.12)$$

Where $p(x)$ = the probability that a case is in a particular category (i.e., success vs not success)

β_0 = the constant of the equation

$\beta_1, \beta_2, \ldots, \beta_k x_k$ = the coefficients of the predictor variables

In order to transform a probability, p(x), having a range of 0 to 1 to $-\infty\ to\ \infty$, the logit or logistic function $logit[p(x)] = \ln\left[\frac{p(x)}{1-p(x)}\right]$ is used. As seen from Figure 4.2, a logit function stretches a bounded probability, p(x), to an S curve that can take any number. Note that at a probability of 0.5, the logit of p(x) is zero. After the regression model is obtained, to be able to do a discriminant analysis, the logit is then transformed back to a probability by using equation 4.12. Then, a cut-off probability value needs to be assigned to differentiate between the range of probability that is 0 or 1 (i.e., non-apnea (0) or apnea (1)). The underlying distribution of the model is a binomial distribution so that a parameter estimation using the least squares technique cannot be used in this case. A maximum likelihood estimation which is described in the next section is used for that task.

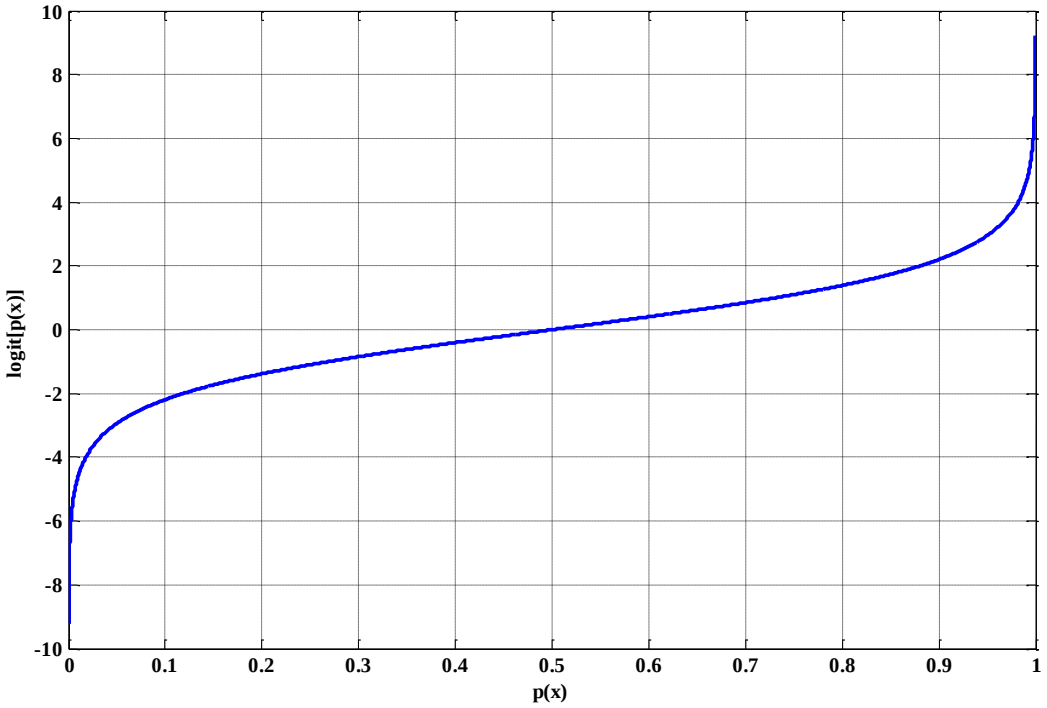


Figure 4.2: Logistic or logit transformation function

4.2.1 Parameter estimation

From equation 4.11, we can rewrite the equation in matrix form as [106]:

$$\log\left[\frac{p_i}{1-p_i}\right] = \sum_{k=0}^{K} x_{ik}\beta_k \quad i = 1,2,\ldots\ldots,N \tag{4.13}$$

Where p = a column vector with length N elements and p_i is the probability that a case is in a particular category for any given observation in the i^{th} population. Assuming p is the probability of success (1) in this case,

x = the independent variable(s) with N rows and K+1 columns. The first element, $x_{i0} = 1$.

K = the number of independent variables specified in the model

β = the parameter vector as a column vector of length K+1

The goal in this section is to estimate K+1 unknown parameters β such that the model best describes all observations. Let Y be a length N column vector as a random variable representing the number of

successes in population i, where i = 1,2,…,N and y is a column vector that contains elements y_i representing the observed counts of the number of successes for each population. Because each y_i is binary, the joint probability density function of Y is [106]:

$$f(y|\beta) = \prod_{i=1}^{N} \frac{n_i!}{y_i!\,(n_i - y_i)!} {\pi_i}^{y_i} (1 - \pi_i)^{n_i - y_i} \tag{4.14}$$

To estimate the parameters, the maximum likelihood estimation is used. Having the same form as the joint probability density function above but switching known and fixed variables y and β, the likelihood function is:

$$f(\beta|y) = \prod_{i=1}^{N} \frac{n_i!}{y_i!\,(n_i - y_i)!} {\pi_i}^{y_i} (1 - \pi_i)^{n_i - y_i} \tag{4.15}$$

From the equation above, the maxima and minima can be calculated by differentiating the equation with respect to each β_i and evaluating the first derivative that is set equal to zero. Then, using the values obtained from solving the first derivative previously to evaluate the second derivative, the maxima point is found when the second derivative at that point is negative. Note that the variance-covariance matrix is also formed by the second derivative equation. Simplified first and second derivatives of the log likelihood function are shown below [106]:

$$\text{First derivative: } \frac{\partial l(\beta)}{\partial \beta_k} = \sum_{i=1}^{N} y_i x_{ik} - n_i \pi_i x_{ik} \tag{4.16}$$

$$\text{Second derivative: } \quad \frac{\partial l(\beta)}{\partial \beta_k \partial \beta_{k'}} = -\sum_{i=1}^{N} n_i x_{ik} \pi_i (1 - \pi_i) x_{ik'} \tag{4.17}$$

4.2.2 Test of significance for a model and the model's parameters

Wald χ^2 statistics are normally used to test the significance of each parameter in the model. The Wald calculation for the β_k parameter is [105]:

$$Wald = \frac{\beta_k}{SE_{\beta_k}} \tag{4.18}$$

Each Wald statistic is then compared with a χ^2 distribution with a degree of freedom = 1 against the null hypothesis H_0: $\beta_k = 0$. Note that for a dataset with a small sample size, there may be a bias such that it produces a large estimation of the coefficient and an inflated standard error resulting in a lower Wald statistic, inaccurately underestimating the significance of some parameter in the model.

For a model test of significance, a log likelihood (LL) ratio test can be used to compare the likelihood of the data under the full model (e.g., with all parameters) and another model, normally an intercept-only model. The test statistic is:

$$G = -2[(Log\ likelihood\ of\ intercept - only\ model) - (\text{Log}\ likelihood\ of\ a\ full\ model)]$$

G has a χ^2 with degrees of freedom equal to the number of parameters used in the full model. The significance is assessed by a χ^2test with the null hypothesis H_0: $\beta_1 = \beta_2 = \ldots = \beta_k$. The significance level of 0.05 is normally used in both Wald and LL tests.

4.2.3 Model performance testing

In our case, where the modeling focus is on creating a mathematical tool to distinguish between people with sleep apnea (target = 1) and people without sleep apnea (target = 0), the end result of the logistic regression, which is the probability of having a sleep apnea, is not a sufficient criterion to differentiate between these two groups. The probability cut-off to divide the people into two group (i.e., classified as non-apnea if $p < 0.5$ and as apnea if $p \geq 0.5$) must be chosen carefully based on the user's objective (e.g., minimizing the false negative rate). Thus, the method for assessing the performance of the logistic model is discussed in the receiver operating curve (ROC) section (4.4) below.

4.3 Artificial neural network

An artificial neural network (ANN) is a parametric machine learning technique developed with inspiration from the structure and function of biological neurons. An ANN model basically maps inputs

into outputs using mathematical functions. To introduce the basic idea of ANNs, Figure 4.3 is a diagram of a single-input with one neuron ANN.

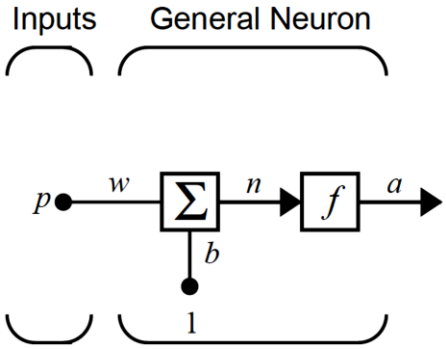


Figure 4.3: Artificial neural network architecture of a single-input with one neuron (reproduced from [107])

Scalar input p is multiplied by scalar weight w. Then, the product of w x p is summed at a summer with bias b. The summation output is called net input n, which is then passed on to transfer function f. The final product from the transfer function is output a. In mathematical notation, the neuron output is:

$$a = f(wp + b) \tag{4.19}$$

The neuron output depends on which transfer function is used in the equation. The choice of criteria for choosing a transfer function depends on several factors. One of those factors is the task for the ANN. In our case, because the probability of an event is a desirable measurement for the ANN output, the log-sigmoid transfer function is chosen because it will bound the output into the range 0 to 1 which corresponds to the objective. Another reason to use the log-sigmoid transfer function that it is differentiable, which is a requirement for using a backpropagation algorithm for training a multilayer network [107].The log-sigmoid function used is:

$$a = \frac{1}{1 + e^{-n}} \tag{4.20}$$

The basic idea in using the ANN is to find a combination of parameters w, p, and b, and a transfer function that gives the desired output given the inputs. As commonly known, a single layer ANN can only solve a problem that is linearly separable. However, the architecture of an ANN can be adjusted to

accommodate the complexity of the problem. Solving more complex problem will require more inputs, neurons, and layers. The architecture in an ANN shown in Figure 4.4 is much more complex.

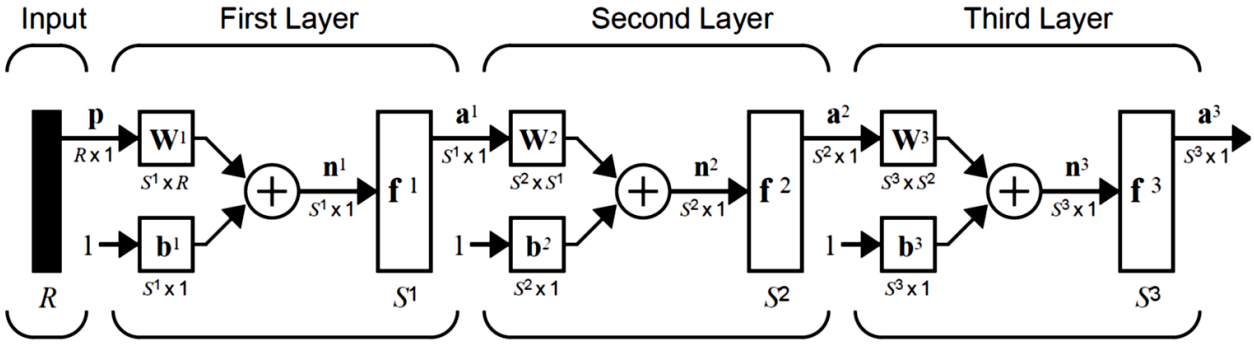


Figure 4.4: Three-layer multilayer perceptron network (reproduced from [107])

This ANN consists of three layers with neurons as needed. This particular network architecture is called a multi-layer perceptron network (MLP). Each layer can have any number of neurons (S), which is a set of weight, bias, summer, and transfer functions. The inputs of this ANN are composed of R inputs expressed in terms of a matrix as **p**. Each layer includes the weight matrix $\mathbf{W^i}$ with the dimension of S^i x R, the summers, the bias vector $\mathbf{b^i}$ with the dimension of S^i, a net input vector $\mathbf{n^i}$ with the dimension of S^i, and an output vector $\mathbf{a^i}$ with the dimension of S^i, where S^i indicates the number of neurons used in the layer, and a super subscript i which indicates which layer those parameters belong to. The last layer, whose output is the output of the entire network, is called the output layer and the other layers are called hidden layers. Also, the output of the previous layer becomes an input in the next layer. Thus, the equation for the output of each layer is:

$$a^{m+1} = f(W^{m+1}a^m + b^{m+1}) \quad for\ m = 0,1,\dots,M \tag{4.21}$$

Where m = the number of layers in the network

4.3.1 Training the multilayer perceptron (MLP) network with the backpropagation algorithm

This section describes how to estimate the parameters, weights, and biases that will best estimate the proper output (i.e., the probability of having apnea given the inputs). In this section, the reasoning and mathematical background for training the MLP network are provided, in particular, the backpropagation algorithm, which is a supervised learning technique used in such a task. The intuition in this process is that

we provide the inputs to the network, including independent and dependent variables. In each case, the value of the independent variables is passed through the network. Then, after the output generated by the network is determined, the performance measurement is computed and sent back into the network to adjust the parameters. This process continues until some criteria are met (i.e., after the mean square error is less than the threshold value, or after the number of iterations is reached). Assume we have a set of data with known output:

$$\{\boldsymbol{p}_1, \boldsymbol{t}_1\}, \{\boldsymbol{p}_2, \boldsymbol{t}_2\}, \dots, \{\boldsymbol{p}_N, \boldsymbol{t}_N\} \tag{4.22}$$

Where **p** = an input vector to the network

t = the corresponding output vector

N = the number of cases (data points)

To train, each input set is fed into the network with randomized parameters as an initial condition (only in the first iteration). The outputs are then compared to the desired targets. To assess the performance of the network, the mean square error is calculated using the following equation [107]:

$$F(\boldsymbol{x}) = \frac{1}{N}\sum_{n=1}^{N}(\boldsymbol{t}_n - \boldsymbol{a}_n)^T(\boldsymbol{t}_n - \boldsymbol{a}_n) = E[(\boldsymbol{t}_n - \boldsymbol{a}_n)^T(\boldsymbol{t}_n - \boldsymbol{a}_n)] \tag{4.23}$$

Where **x** = a vector containing all network weights and biases

a = an output vector of the network

Intuitively, a small mean square error is desirable because it means that there is a small difference between the network's generated outputs and the actual targets. Thus, in general, the training algorithm should minimize the mean square error, called a performance index. If we consider the mean square error as a performance surface where the global minima point is the combination of parameters that gives the minimal mean square error, by this means, the training process actually is a searching algorithm that iteratively looks for the steepest descent direction to the minima point. One such method is the least mean square (LMS) [107]. In each iteration, the mean square error function is approximated using a Taylor series expansion with a limited number of expansion terms. The algorithm will look for the steepest path to the local minima of the approximation. The searching direction is determined by the negative gradient of the

direction vector of the parameters. With a sufficient number of iterations, the algorithm will eventually reach the global minima, where the parameters give the lowest mean square error. Assume that the current searching point is at iteration k, and the estimated mean square error is:

$$\hat{F}(\boldsymbol{x}) = [\boldsymbol{t}(k) - \boldsymbol{a}(k)]^T[\boldsymbol{t}(k) - \boldsymbol{a}(k)] = \boldsymbol{e}^T(k)\boldsymbol{e}(k) \tag{4.24}$$

Using the steepest descent algorithm [107], we can approximate new weights and biases for the next iteration k+1:

$$w_{i,j}^m(k+1) = w_{i,j}^m(k) - \alpha \frac{\partial \hat{F}}{\partial w_{i,j}^m} \tag{4.25}$$

$$b_i^m(k+1) = b_i^m(k) - \alpha \frac{\partial \hat{F}}{\partial b_i^m} \tag{4.26}$$

Where α = learning rate; $\alpha < \frac{2}{\lambda_{max}}$ for a stable search

λ_{max} = maximum eigenvalue of the Hessian matrix of the performance index

$w_{i,j}^m$ = weight at i[th] neuron, j[th] input, and m[th] layer

b_i^m = bias at i[th] neuron and m[th] layer

To implement these equations, in our case, the output of the MLP network depends not only on the weights and biases of the output layer but also on the outputs of previous layers, so the derivative terms cannot be directly computed. The chain rule is then applied to those equations:

$$\frac{\partial \hat{F}}{\partial w_{i,j}^m} = \frac{\partial \hat{F}}{\partial n_i^m} \times \frac{\partial n_i^m}{\partial w_{i,j}^m} \tag{4.27}$$

$$\frac{\partial \hat{F}}{\partial b_i^m} = \frac{\partial \hat{F}}{\partial n_i^m} \times \frac{\partial n_i^m}{\partial b_i^m} \tag{4.28}$$

Where n_i^m = the net input of layer m in i[th] neuron.

$= \sum_{j=1}^{S^{m-1}} w_{i,j}^m a_j^{m-1} + b_i^m$

So that $\frac{\partial n_i^m}{\partial w_{i,j}^m} = a_j^{m-1}$ and $\frac{\partial n_i^m}{\partial b_i^m} = 1$. Hagan et al. [107] defined $s^m = \frac{\partial \hat{F}}{\partial n_i^m}$ as the sensitivity of $\hat{F}$ to changes in the i[th] element of the net input at layer m, and used a Jacobian matrix, $\frac{\partial n^{m+1}}{\partial n^m}$, to derive the recurrence relationship for the sensitivities. The final form of s^m is as follows:

$$\boldsymbol{s}^M = -2\dot{\boldsymbol{F}}^M(\boldsymbol{n}^M)(\boldsymbol{t} - \boldsymbol{a})^T \tag{4.29}$$

$$\boldsymbol{s}^m = \dot{\boldsymbol{F}}^m(\boldsymbol{n}^m)(\boldsymbol{W}^{m+1})^T s^{m+1} \tag{4.30}$$

Where m = $M-1, M-2, \ldots, 1$

M = the last layer of the network

$$\dot{F}^m(n^m) = \begin{bmatrix} \dot{f}^m(n_1^m) & 0 & \cdots & 0 \\ 0 & \dot{f}^m(n_2^m) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & \dot{f}^m\left(n_{S^m}^m\right) \end{bmatrix}$$

Because s^m can only be computed backwards (from last layer back to the first layer), this algorithm is commonly known as backpropagation. Finally, with the steepest descent and back propagation algorithms, the equations used to minimize the square error iteratively are:

$$\boldsymbol{W}^m(k+1) = \boldsymbol{W}^m(k) - \alpha \boldsymbol{s}^m(\boldsymbol{a}^{m-1})^T \tag{4.31}$$

$$\boldsymbol{b}^m(k+1) = \boldsymbol{b}^m(k) - \alpha \boldsymbol{s}^m \tag{4.32}$$

4.3.2 Considerations for training MLP networks

Some limitations for training the ANN must be considered. First, the architecture of the network needs to be adjusted to suit the complexity of the problem. To the best of our knowledge, there is still no scientific method for estimating the number of layers, and neurons, including the weights, or biases that best fit the problem. As recommended by [108], these constraints cannot be known before training the initial network. If too many parameters are used, the network will most likely introduce the over-fitting problem. If there are too few parameters in the network, they will not be enough to fit the underlying process of the problem.

Second, depending on the transfer function used, data preprocessing may be needed before training the network. The data points and associated target should range between the boundaries of the transfer functions used in the network to prevent a saturation problem. For example, if the tangent-sigmoid is used as a transfer function, the input data should be normalized to the range of [-1, 1]. Furthermore, some authors [108] recommend setting the target to be less than the range of the transfer function (i.e., [-0.76,0.76] in this case) to avoid the saturation problem.

Third, the stopping criteria for training the network should be carefully chosen. An ANN is very prone to an overfitting problem. One recommended way to avoid this problem, is to at least partition the dataset into two sets, training and validation, and stop the training when the performance index on the validation dataset increases. Another stopping criterion is the norm of the gradient of the performance index. It might be possible for a single layer network for a perfectly linear separable problem to have a zero gradient at the minimum. However, in MLP networks with more complex problems, it is not likely to be the case. Normally, a mean square error (as a performance index) is set at 10^{-6} as a training stopping criterion [108].

4.4 Receiver operating characteristic (ROC) curves

In classification applications, many models such as artificial neuron networks, and logistics regression, mostly predict a success percentage rate as their output. However, real-world applications such as medical diagnosis, direct marketing, and pattern recognition, focus on the discrimination power of the model (i.e., potential buyer or not, having a disease or not, and having the same pattern or not). The last step before a model can do such task is to define the probability cut-off value that best separates the two target groups (binary targets in our case). Some models use 50 percent as a default threshold (i.e., $y < 50$: does not have the disease, $y \geq 50$: have the disease). However, this value does not guarantee good discrimination between two target groups (in cases where the target is binary). In practical situations, when considering two populations, one with disease, the other without a disease, we rarely find a perfect separation between the two groups. For example, a laboratory test of an enzyme level in the blood could be

used to diagnose a disease. A high or low enzyme level can indicate whether the person has the disease or not. However, practically, it is almost impossible to have a clear-cut border line to distinguish between the groups. As shown in Figure 4.5, the criterion or cut-off value must be selected to discriminate between the two populations.

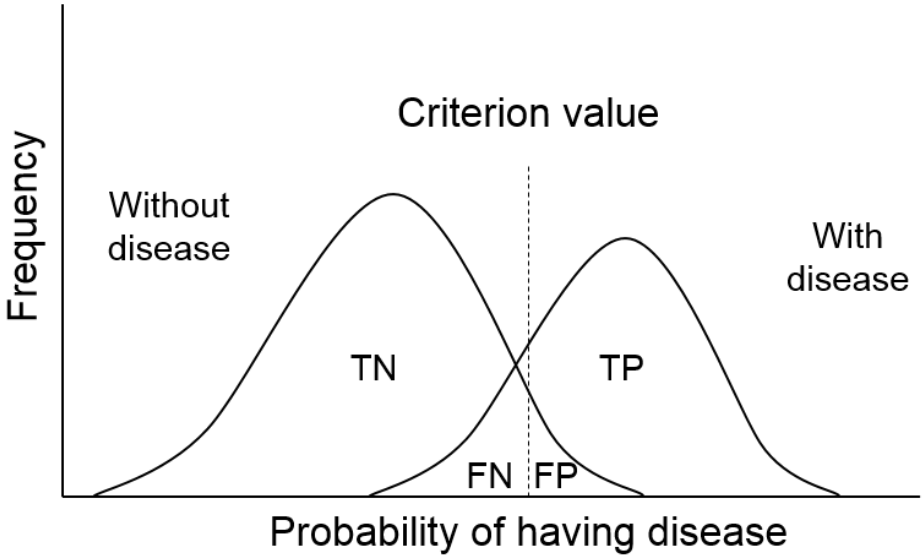


Figure 4.5: Binomial distribution of populations with and without disease

Corresponding to each cut-off value, will be some cases correctly and incorrectly classified by the tool (i.e., a laboratory test result, classification models). There are four possibilities; cases with a disease and correctly classified as having the disease (TP = true positive fraction), cases without a disease and correctly classified as not having the disease (TN = true negative fraction), cases without a disease but incorrectly classified as having the disease (FP = false positive fraction), and cases with a disease but incorrectly classified as not having the disease (FN = false negative fraction). In these four cases, tool correctly classified two, TP and TN, and incorrectly classified two cases, FP and FN. The performance indexes widely used in the literature with their equations are included in Table 4.2:

In an ideal situation, we would definitely want our tool to discriminate among the cases such that sensitivity or TPR and specificity or TNR are 100%. However, that is usually not the case because of the variations between people (in cases of a diagnostic test). There is always a trade-off between TPR and TNR. Referring to Figure 4.5, when the cut-off criterion is set to a lower probability (the cut-off point is shifted

to the left of the graph), the tool will likely be sensitive to classifying the cases into the group with the disease, which most likely also increases the FP fraction and decreases the FN fraction. In this case, we will have a tool that has high sensitivity or TPR and low specificity or TNR. Inversely, if the cut-off point is set to the higher threshold (to the right of the graph), the tool will tend to be less sensitive but more specific to the disease, resulting in lower sensitivity or TPR and higher specificity or TNR.

Table 4.2: Derivation of performance indexes [109]

Performance index	Derivations
Sensitivity or true positive rate (TPR)	TP/(TP+FN)
Specificity or true negative rate (TNR)	TN/(FP+TN)
Positive predictive value (PPV)	TP/(TP+FP)
Negative predictive value (NPV)	TN/(TN+FN)
False positive rate (FPR)	FP/(FP+TN) = 1 – TNR
False negative rate (FNR)	FN/(FN+TP) = 1 – TPR

The tool that helps visualize these trade-offs between sensitivity and specificity when adjusting the cut-off threshold is called a receiver operating characteristic (ROC) [110-113]. (To clarify, the ROC curve idea and background are totally different from the operating characteristic (OC) curve which is widely used in the quality control discipline. The ROC curve (see Figure 4.6) is empirically derived from the confusion matrix.). To construct the ROC in this study, the procedure is as follows [112]:

1. Decide a discrete period for a cut-off value for evaluating the confusion matrix. A finer period will give a smooth ROC curve but require more computational cost. In this study, the period of 0.01 or 1% is used.
2. Vary the cut-off threshold with the period set in 1). For each threshold, feed each classifier with the test dataset. From the result of each classifier in each cut-off threshold, build the confusion matrix.
3. Calculate sensitivity or TPR and specificity or TNR from each confusion matrix.

4. Plot each coordinate obtained in 3) on the graph. The sensitivity or TPR is plotted against the Y-axis and the FPR or (1-specificity) is plot against the X-axis.

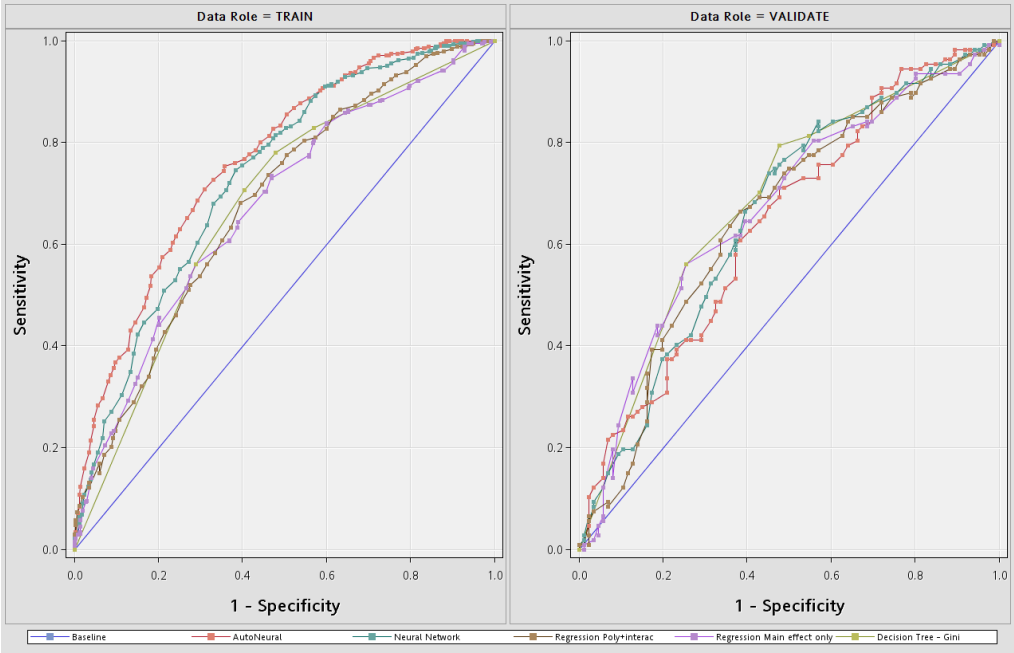


Figure 4.6: Example of ROC curves constructed from confusion matrixes of several modeling methods in training and validation datasets

In a decision tree (DT) classifier which is designed to produce only a class decision at each end point, the algorithm above is not applicable for such model type [112], but, we can approximately construct the ROC curve by evaluating the confusion matrixes obtained from each final leaf of the DT [94]. However, some authors [111, 112] have concluded that discrete classifiers such as DTs produces only a single point in the ROC space derived from the final confusion matrix.

Again, the operating cut-off value depends on the user's objective. Popular choices of objectives are: minimizing misclassification rate, minimizing the FPR, minimizing the FPR, maximum TPR, maximum TPR, maximizing profit, minimizing cost, and so on. Furthermore, the ROC curve is not only used for visualization and evaluation of the model's performance. Once projected on the same scale (same cut-off unit. e.g., probability), it can be used for comparing the performances of several models. By plotting the ROC curves from different models on the same graph, for example, if accuracy is the objective, the

most northwest point closest to the (0,1) coordinate on the ROC graph gives the most accurate model and cut-off threshold point.

Another criterion commonly used for choosing the best model is to calculate the area under the ROC curve (AUC). The AUC is equal to the probability that the classifier will correctly identify positive cases when randomly presented with a pair of positive and negative cases [114]. In other words, the AUC is the average sensitivity over the entire range of specificity or vice versa [115]. Hanley and McNeil [114] showed that the AUC is equivalent to the Wilcoxon test of ranks. To use the AUC as a ranking tool for choosing the best model is simple. After plotting the ROC curve, the AUC is the area under the ROC curve with the maximum value of 1. The tool or model that gives a maximum AUC will have a maximum average performance over every cut-off threshold. However, because the AUC reflects the average performance of the model, the model that produces the maximum AUC may not produce the minimum classification error.

An example of ROC curves is shown in Figure 4.6. Generally, the ROC shows the trade-off between TPR and FPR when varying the cut-off threshold (each point on the ROC plot corresponds to each cut-off probability). Referring to Figure 4.5, when the cut-off probability moves to the left, not only does the TPR increase but the FPR also increases. This behavior is also shown in the ROC, so we can choose which cut-off value corresponds to the desired performance. In Figure 4.6, the plots in the left or TRAIN column, are results from the confusion matrixes built from the training dataset. Likewise, the plots in the right or VALIDATE column, are the results from the validation dataset. The ROC chart always includes a diagonal line that refers to the confusion matrix derived from the probability that the target was randomly chosen with equal probability. The AUC may be the best choice visually, in order to tell which model is the best on average. Starting from the ROC curves of the training dataset, it is clear that the neural network (ANN) model is the best. Because it gives the largest area under the ROC curve (spotted by the line that is the furthest northwest at almost every point). However, for the ROC curves on the VALIDATE dataset, the ANN model does not provide the largest AUC, most likely because of the overfitting problem, so the ANN model may not be the best choice globally. Thus, we may say, virtually, that the decision tree (DT) gives the average best performance in both training and validation datasets. However, as mentioned earlier, it

solely depends on the objective function of the modeler. The largest AUC does not guarantee the best model at every cut-off point. It is the modeler's task to carefully choose the cut-off value that agrees with the modeler's objective function.

CHAPTER V

OSA SCREENING RULES BASED ON THE PATTERNS OF PHYSIOLOGICAL DATA

This chapter depicts the process of developing an OSA screening model. All procedures for processing the raw data for analysis are also explained here. At the end of this chapter, the result (screening rule) and the performance of the data analysis based on the data mining method are thoroughly described.

The objective for developing an OSA screening rule in this chapter is to confidently identify people who do not have OSA disease from the population using easily obtained variables such as weight, height, sleep habits, and other anthropological variables. Then, the people that the model cannot confidently classify as not having OSA will use the diagnosis method in the next chapter for a more precise identification. Technically, we want to build a classification model such that it can classify non-OSA people with a low false negative rate (using a 5% threshold in this chapter).

5.1 Database information

In this chapter, we used demographic, anthropological, and sleep-related data of patients from the Sleep Heart Health Study (SHHS) to build the OSA screening rules for differentiating between people with and without OSA. The main objective for developing these rules is to confidently separate people without OSA from the population. In the other words, statistically, the model is developed with the objective of minimizing the false negative fraction (positive = with OSA, negative = without OSA). Then, people who do not show strong evidences of not having OSA will be conveniently diagnosed by their own biorhythms collected overnight using a mobile and lightweight device explained in chapter 6. The rules developed in this chapter help more efficiently and economically prioritize people who need to go through an OSA diagnosis process.

5.1.1 Sleep Heart Health Study (SHHS)

The SHHS is a multi-center cohort study implemented by the National Heart Lung & Blood Institute [116, 117]. The purpose of the SHHS is to determine whether sleep-related breathing problems are associated with an increased risk of hypertension and other cardiovascular diseases. Because this set of data was collected from human subjects, the data were obtained by formal registration with the related approval from the Oklahoma State University Institutional Review Board (IRB). The data collected in the SHHS study focused on early middle-aged men and women who had not reported or experienced cardiovascular disease events (heart attack, surgery, or medications). After assessment of the responses from the pre-screening questionnaire sent to about 20,000 people, 6,441 men and women aged 40 years and older were enrolled in this study. These subjects were selectively recruited from many ongoing cohort studies including Atherosclerosis Risk in Communities (ARIC) Study sites in Washington County, Maryland, and Minneapolis, Minnesota; the Cardiovascular Health Study (CHS) sites in Sacramento, California, Washington County, Maryland, and Pittsburgh, Pennsylvania; the Framingham Offspring and Omni cohorts in Tucson, Arizona; the Strong Heart Study sites in Arizona, Oklahoma, and South Dakota; and New York City Populations assessed in studies of hypertension.

Participants younger than 65 years old were selected into the study only if they had a history of snoring or any indicators of sleep-disordered breathing. For patients aged 65 years or more, snoring is no longer a good predictor of the presence of sleep-disorder breathing; thus, they were selected for the study even without a snoring history. According to SHHS protocol [116, 117], three types of data were collected; sleep data, covariate data, and outcomes data.

Sleep data collection was composed of two procedures, a sleep habits questionnaire and a PSG study. The sleep habits questionnaire was designed to assess baseline sleep habits and any problems, history or treatment of sleep-disordered breathing. This data was also used in our analysis to build the OSA screening rules. A PSG study was done by EEG-based polysomnography (PSG) in the subject's home. Twelve channels of data were recorded as follows: Oximetry, Heart Rate, Chest wall and abdomen movement, Nasal/oral airflow, Body position, Electroencephalogram (EEG) (2 central; one for redundancy in case of failure/loss), Electrooculogram (EOG) (bilateral), Electromyogram – chine (EMG), and Electrocardiogram (ECG). The PSG data length ranges from 4-8 hours for each subject. Along with the PSG data, the target variable (0=no apnea, and 1=apnea) evaluated by a certified sleep technician or doctor is also provided. In this study, we use the ECG as a primary signal for classification of a sleep apnea episode, as explained later in the chapter.

Covariate data is the data collected regarding demography, health history, cardiovascular risk factors, and cardiovascular events. However, the variables in this data were ranked by three priorities; critical, important, and useful data. Only critical variables were present in all data records. Because the SHHS was carried out on top of other studies which may have collected totally different data variables, variables were considered important or useful and that had already been collected from those cohorts was also added to the SHHS. Thus, some data variables may exist in only some records. For examples of the variables in this data collection are the prevalence of cardiovascular disease (CVD) in the last 3 months, blood pressure, neck circumference, weight, medications taken in the past two weeks, self-reported hypertension, gender, age, usual caffeine and alcohol intake, and cholesterol level. Next, the outcome data collection focused on events relating to cardiovascular diseases, stroke, hypertension (higher than the

threshold level), and mortality. Finally, overall, 1,242 variables were collected in the SHHS study. However, not all variables were related to OSA disease. Before these data could be used in an analysis, only those that relate to OSA disease must be statistically selected. This process is explained in the next section.

5.2 Data preprocessing process

The raw data from most data sources usually cannot be analyzed directly for several reasons: incomplete or missing data, spurious extreme data or outliers, and non-conformance with a normality assumption used in linear modeling techniques. In this section, details of the data preparation are explained.

5.2.1 Data cleansing

From the raw demographic, anthropology, and sleep-related data obtained from the SHHS, 5,804 records with 1,242 variables are contained in a dataset. A binary variable named "apnea" was created from thresholding an apnea-hypopnea index (AHI) (AHI<5, then apnea = 0; AHI≥5, then apnea=1). This variable was used to identify a person as having or not having an apnea disorder. This variable is the target variable in this section. Next, before statistical analysis can be done to create OSA screening rules, all the data must be clean. In the other words, extreme values or outliers, missing values, and the skewness in distribution of each variable should be taken care of. The overall workflow used in this section is shown in Figure 5.1.

The process started by manually rejecting variables not correlated with OSA disease. All administrative variables such as a flag of the availability of some data (e.g., has a PSG record or not, has a sleep-habit form or not), the data collection date, hospital visiting id, and all PSG measurement variables were excluded because they are not related to the screening process. After these two data categories were excluded, there are 297 variables left in the dataset. These variables are classified as CVD outcomes, demographics, anthropometry, clinical measurements, medical history, medications, self-administrated questionnaire, health interview on life quality, and sleep habits.

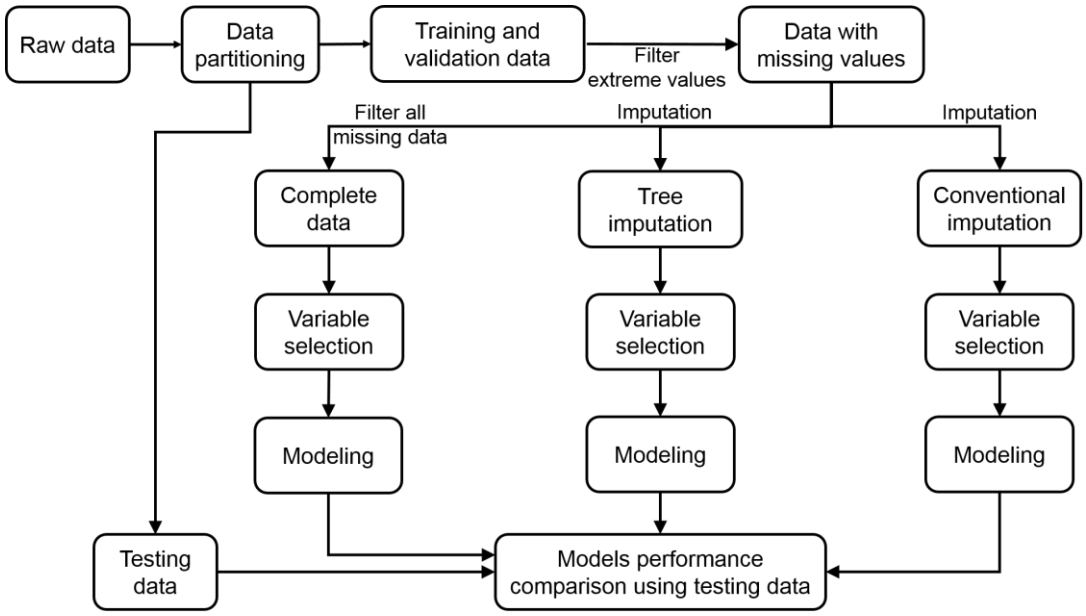


Figure 5.1: OSA screening model development workflow

Because most of the data was obtained from questionnaires with Likert scale answers and non-ranking answers, all "don't know" answers were treated as missing values. Descriptive statistics were used to examine the data characteristics of each variable. Of the 297 variables, only 3 did not contain any missing values. The percentage of missing values in this dataset ranges from less than 1% to 98%. The variables with missing value percentages more than 20% were dropped from the analysis because the results from analyzing these data will likely be bias, as suggested by experts in [118]. Suspicious values (e.g., too high or too low values such as triglycerides of more than 1000 mg/dL, or age equal to 0) were also manually filtered out. In binary variables, the variables with a majority response (either "yes" or "no") of more than 90% with no differentiation of OSA group found, were also dropped from the analysis because they gave very little information on identifying OSA.

Rather than dropping the unrelated variables, we combined some redundant variables to safely impute missing values. Some variables could be used as substitutes for each other. For example, the variables "weight" and "weight20" both refer to the weight of the subjects but at different times. The variable "weight" was measured at the body check-up in the hospital while "weight20" was measured on the same day that the blood pressure was measured, which was the day before the PSG. These two dates were 2-4 weeks apart so that the change in weight should not be significantly different. Thus, the new

variable “weightim” (weight imputed) was created by averaging those two variables if they were both present, or using only one if the other was missing.

We also created new meaningful variables from variables that were hard to interpret. For example, sleep time and wake up time were used to calculate the total sleep duration. Those who smoke pipe, cigar, or cigarette before bed were grouped into one variable indicating whether the subject smokes before going to bed or not. The ordinal variables with too many levels were converted to a binary level, low or high only. At this point, 201 variables are left for analysis.

5.2.2 Treatment of missing values

As shown in Figure 5.1, we managed the missing values in two ways, listwise deletion or complete case analysis, and imputation. In complete case analysis, all missing values were filtered. Only the observations with complete responses remain in the dataset. The advantages of this method are simplicity and comparability across analyses [119]. This method is a default method used for dealing with missing values in many statistical analysis tools. The obvious drawback of this method is that, with a significantly lower number of data point N, the statistical power is reduced because we did not use all available information. Also, if the missing data were not missing completely at random (i.e., the respondents tended not to respond to the sensitive questions), this method will likely introduce bias into the dataset. After implementing this method, there are 952 complete cases for analysis.

For the imputation of missing values, we used a decision tree with surrogate method (see topic 4.1.4 in Chapter 4). After the imputation, to make sure that the imputed values do not introduce bias into the dataset, in the other words, that there is no statistical difference between the original and the imputed datasets, we used a student’s t-test for testing whether the mean of each interval variable is statistically different in the case of the interval variables. For most of the interval variables, because the sample size is large (N=5104), from the central limit theorem, the data distribution could be approximated as normal. However, for the variables that are heavily skewed, the Wilcoxon rank sum test is used for testing whether two distributions (imputed vs non-imputed) are statistically different. The null hypothesis is that the two

distributions are identical and the alternative hypothesis is that the two distributions differ only with respect to the median [120]. The test showed that most variables were not significantly different. Some of the variables did have a sign for a significant change in mean, median, or mode between non-imputed and imputed datasets. However, we use these variables in the dataset because they may reflect the real underlying process of the meaningfulness of the data. The statistical differences between the imputed and non-imputed groups may occur because an underlying process of the missingness is not completely random (i.e., the respondents intended not to answer some sensitive questions) which makes some sense in our case. As mentioned earlier, this dataset was the product of data collected from several cohorts. Some cohorts may not have collected some data variables that others did.

To make sure that the tree imputation did not introduce bias into the dataset, we introduced one more imputation dataset obtained by using the conventional imputation method for comparison. For interval variables, if the distribution of a variable was normal or close to normal, the overall mean of the variable was used to replace the missing value. If the distribution of the variable was not normal or skewed, we replaced the missing value with the median of the variable. For the categorical variables, the mode for each variable was used for the imputation. The advantage of this method is that no bias is introduced into the distribution of the imputed data. The other, non-missing value fields for the observation are not wasted as in the listwise method. However, the disadvantages are that this method inflates the significance of any statistical tests and deflates the variance [121]. However, for supervised classification applications, this method has been confirmed to perform well [121-123]. We use these three datasets for training the screening models and then compare their performances in the end. The assumption is that if the tree imputation really captures the underlying process for the missingness, the modeling performance from the dataset should perform better than that of the other groups. Otherwise, its performance should be worse.

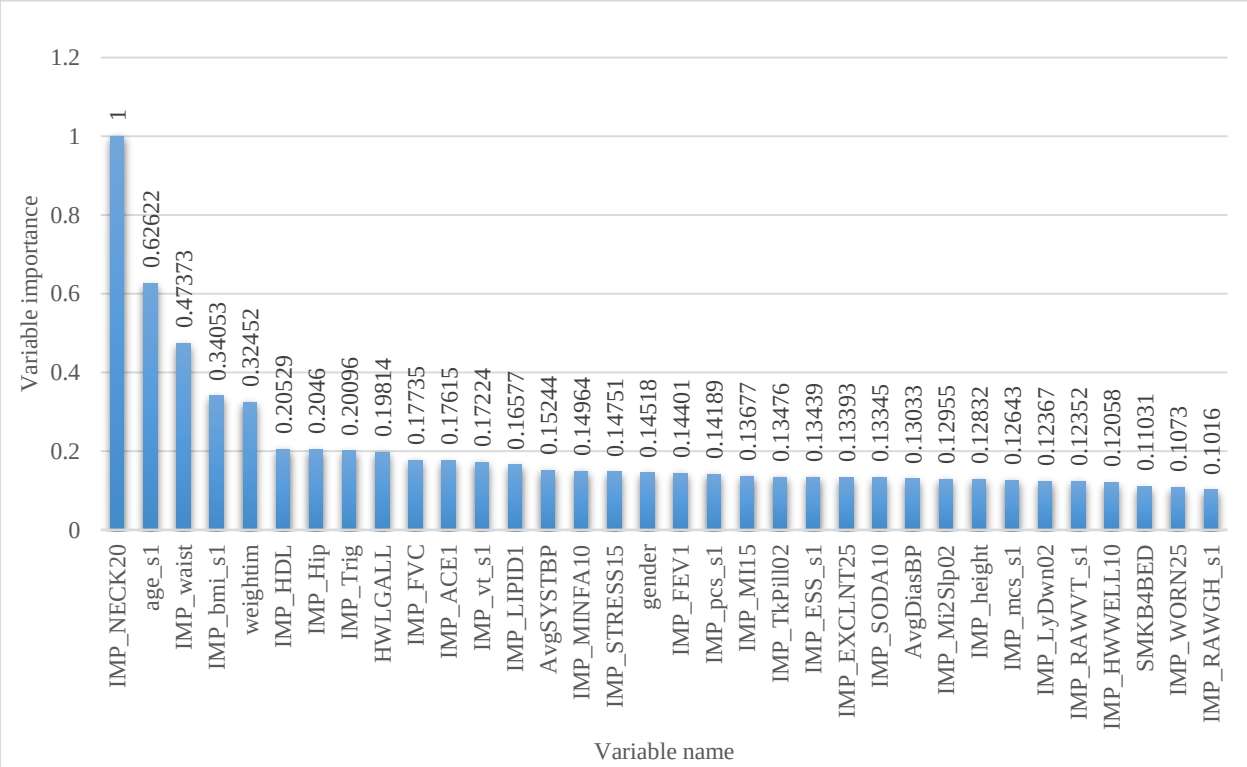


Figure 5.2: Decision tree's variable importance obtained from tree imputed data

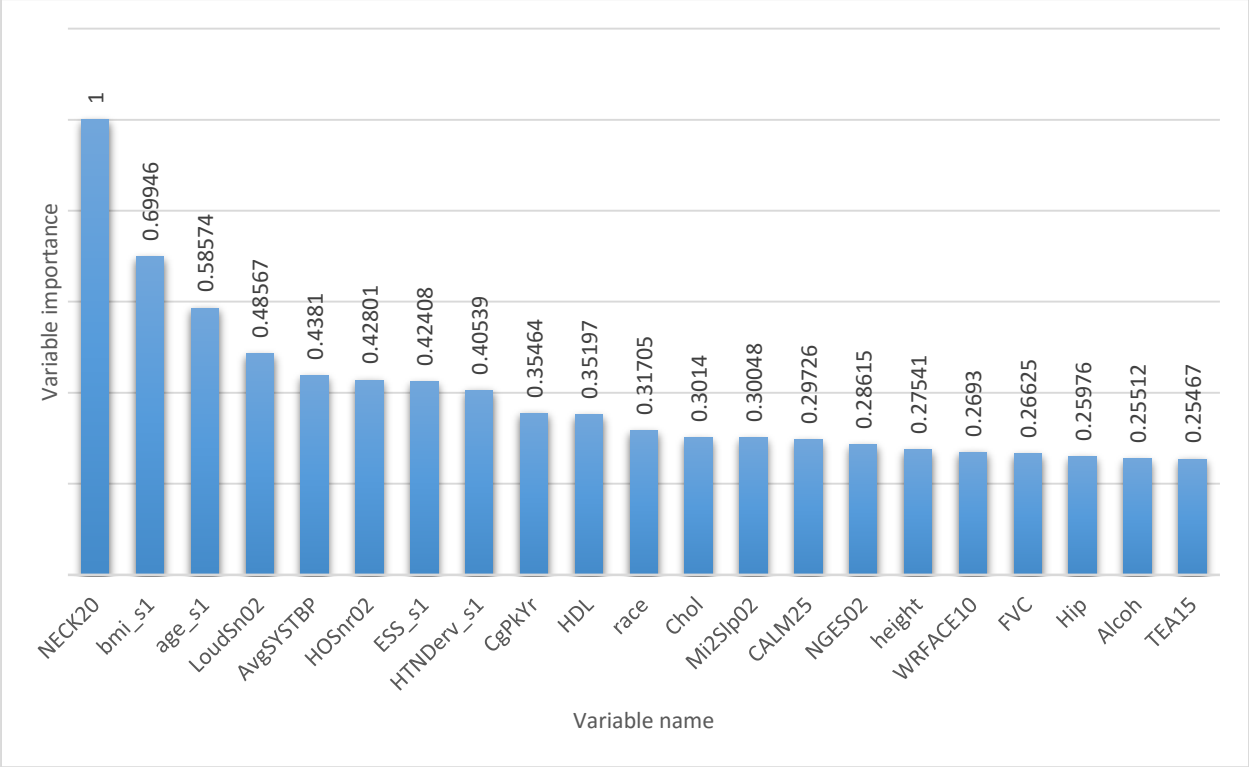


Figure 5.3: Decision tree's variable importance obtained from complete case data

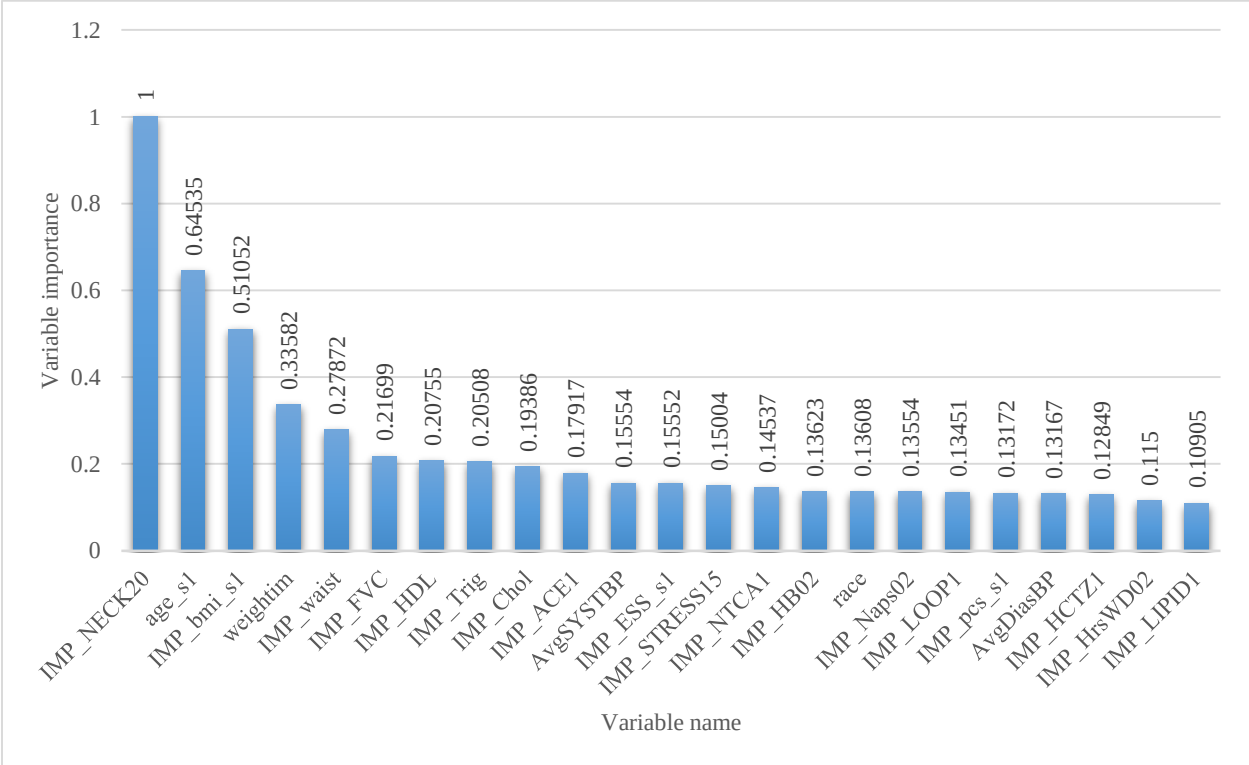


Figure 5.4: Decision tree's variable importance obtained from conventional imputed data

Table 5.1: Selected variables from DT's variable Importance in different datasets

Importance ranking	Tree imputation	Conventional imputation	Listwise
1	NECK20	NECK20	NECK20
2	age_s1	age_s1	bmi_s1
3	waist	bmi_s1	age_s1
4	bmi_s1	weightim	LoudSn02
5	weightim	waist	AvgSYSTBP
6	HDL	FVC	HOSnr02
7	Hip	HDL	ESS_s1
8	Trig	Trig	HTNDerv_s1
9	HWLGALL	Chol	CgPkYr
10	FVC	ACE1	HDL
11	ACE1	AvgSYSTBP	race
12	vt_s1	ESS_s1	Chol
13	LIPID1	STRESS15	Mi2Slp02
14	AvgSYSTBP	NTCA1	CALM25
15	MINFA10	HB02	NGES02
16	STRESS15	race	height
17	gender	Naps02	WRFACE10
18	FEV1	LOOP1	FVC
19	pcs_s1	pcs_s1	Hip
20	MI15	AvgDiasBP	Alcoh
21	TkPill02	HCTZ1	TEA15
22	ESS_s1	HrsWD02	
23	EXCLNT25	LIPID1	
24	SODA10		
25	AvgDiasBP		
26	Mi2Slp02		
27	height		
28	mcs_s1		
29	LyDwn02		
30	RAWVT_s1		
31	HWWELL10		
32	SMKB4BED		
33	WORN25		
34	RAWGH_s1		

5.2.3 Variable selection based on a variable importance derived from decision tree

Now, we have three datasets: listwise, tree imputation, and conventional imputation. It is impractical to input all 201 variables into an OSA screening rule. Only the variables that give the most information or are closely relate to OSA disease should be used. In our case, we use the variable importance from a decision tree (DT) to pre-select the potential variables to be included in the screening model. Because the DT algorithm does not rely on the assumption of linearity, the selected variables can have either a linear or non-linear relationship with the target variable. The DT algorithm determines the importance of the variable by selecting a variable that contributes to correctly predicting the target variable. In other words, if using a Gini impurity measurement as a criterion (see section 4.1.1 and 4.1.3 in Chapter 4), the most important variable is the one that decreases the overall Gini impurity throughout the tree the most. The number of selected variables depends on the variables used in creating splitting rules (primary and surrogate variables). The criteria for variables to be selected is to have a p-value at the splitting point exceeding 0.2. The plots of variable importance obtained from a DT algorithm in different datasets are shown in Figures 5.2, 5.3, and 5.4.

In Figures 5.2, 5.3, and 5.4, the variable importance is ranked by the relative importance ranging from 1 to 0. The most important variable has an importance value of 1. The least important variable has an importance value close to 0. Table 5.1 shows all the variables selected from the DT's variable importance in the different datasets: tree imputation, conventional imputation, and listwise datasets. The variables that were chosen from all three datasets were highlighted in red text and those from two datasets are highlighted in green text. Interestingly, three variables, NECK20, age_s1, and bmi_s1, were at the very top of the three datasets. These variables are neck circumference in centimeters, age, and body mass index (BMI). In the literature, these variables are some of the important predictors for determining OSA. Some other interesting variables that were chosen and that are also mentioned in the literature [124] are triglycerides (Trig), HDL cholesterol (HDL), lipid (LIPID), hip and waist size (Hip, waist), blood pressure (AvgDiasBP, AvgSYSTBP), and Epworth sleep score (ESS_s1).

The model methodologies used are 1) logistics regression, 2) decision tree (DT), and 3) artificial neural network (ANN). These three modeling methodologies do not require the assumption that the data are in a specific distribution pattern. Also, a linear relationship is not a requirement in these cases. However, for a logistic regression, the independent variables must be statistically independent to prevent multicollinearity [104], the situation in which two or more independent variables in a regression model are highly and linearly correlated. The problem arises because we estimate the partial coefficient by holding other variables at a constant and then measure the change in the dependent variable while varying the focused independent variable. If two or more variables are highly correlated, it is difficult to vary one while fixing the others at a constant, so the obtained partial coefficient does not truly exhibit the underlying nature of the data [104]. Some common causes for multicollinearity are 1) using variables that are derived from other variables (e.g., bmi = weight (kg) / [height (m)]^2), 2) using almost the same variables in the model (e.g., weights measured in the same week) and 3) coincidental correlated variables. Thus, it is best to perform a diagnostic test for multicollinearity before using all the variables in any regression model.

5.2.4 Variance-inflation factors

Although it may be possible to use pairwise correlations between the independent variables for detecting multicollinearity, the small correlation values do not guarantee the absence of multicollinearity, so the pairwise correlation is not widely used in this task [104]. The more suitable statistics in this case are variance-inflation factors (VIFs). For each independent variable, the VIF indicates how much larger the variance of the estimated coefficient would be if it did not correlate with other independent variables. The VIF for any independent variable (x_j) can be computed as in [98]:

$$VIF(x_j) = \frac{1}{(1 - R_j^2)} \tag{5.1}$$

Where R_j^2= the coefficient of determination of the regression of $x_j$on all other independent variables

In other words, practically, R_j^2 is the coefficient of determination when fitting a multiple regression model using x_j as a dependent variable and the other independent variables in the original model as independent variables. Thus, if R_j^2 is close to zero, it means that the other independent variables in the original model cannot be used to explain x_j. Therefore, x_j in this case is not involved in any multicollinearity. However, if R_j^2 is non-zero, the closer the value of R_j^2 to 1, the closer the VIF to infinity. This means that x_j can be well explained by other independent variables in the original model, which indicates serious multicollinearity. It has been suggested that a VIF value higher than 10 suggests serious multicollinearity [104].

To check for multicollinearity in our datasets, because the VIF is calculated based on a linear multiple regression, the normality assumption is needed. Each variable is checked for normality and transformed such that its distribution is as close as possible to a normal distribution using a power transformation by Box and Cox [125]. The variance-inflation factor for each variable was then calculated as shown in Table 5.2.

Table 5.2: Variance-inflation factor calculated from the tree imputed dataset

Parameter Estimates						
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Variance Inflation**
Intercept	1	-0.7263	0.2401	-3.0200	0.0025	0.00
SQRT_IMP_NECK20	1	1.0486	0.1381	7.6000	<.0001	3.68
age_s1	1	0.0088	0.0008	10.6300	<.0001	2.11
IMP_waist	1	0.0038	0.0009	4.0500	<.0001	3.95
SQRT_IMP_bmi_s1	1	0.2011	0.0845	2.3800	0.0173	3.64
LOG_IMP_HDL	1	0.0958	0.1027	0.9300	0.3513	1.60
LOG_IMP_Hip	1	-0.0058	0.3609	-0.0200	0.9872	2.05
LOG_IMP_Trig	1	0.3709	0.1748	2.1200	0.0339	1.30
SQR_HWLGALL	1	0.0359	0.0510	0.7000	0.4825	1.23
SQRT_IMP_FVC	1	-1.0120	0.1931	-5.2400	<.0001	11.32
IMP_ACED1	1	0.0249	0.0761	0.3300	0.7433	1.01
EXP_IMP_vt_s1	1	-0.0932	0.0930	-1.0000	0.3162	27.69
EXP_IMP_gh_s1	1	0.1216	0.1131	1.0700	0.2825	44.19
IMP_LIPID1	1	-0.0166	0.0202	-0.8200	0.4128	1.08
AvgSYSTBP	1	-0.0001	0.0004	-0.1700	0.8648	1.84
LOG_IMP_MINFA10	1	0.1549	0.1015	1.5300	0.127	1.36
IMP_STRESS15	1	0.0043	0.0089	0.4900	0.6266	1.05
gender	1	-0.1177	0.0244	-4.8200	<.0001	3.70
SQRT_IMP_FEV1	1	1.0780	0.1699	6.3500	<.0001	9.92
PWR_IMP_pcs_s1	1	-0.0640	0.0702	-0.9100	0.3626	2.97
IMP_MI15	1	-0.0145	0.0275	-0.5300	0.5986	1.11
IMP_TkPill02	1	-0.0139	0.0062	-2.2300	0.0257	1.11
SQRT_IMP_ESS_s1	1	0.1615	0.0468	3.4500	0.0006	1.61
IMP_EXCLNT25	1	0.0078	0.0113	0.6900	0.4908	3.38
SQRT_IMP_SODA10	1	0.0597	0.0475	1.2600	0.2088	1.04
AvgDiasBP	1	0.0007	0.0007	0.9800	0.3254	1.81
LOG_IMP_Mi2Slp02	1	0.1808	0.1740	1.0400	0.2986	1.24
IMP_height	1	-0.0029	0.0012	-2.3800	0.0171	3.42
PWR_IMP_mcs_s1	1	0.0721	0.0689	1.0500	0.2958	2.37
IMP_LyDwn02	1	-0.0053	0.0078	-0.6900	0.493	1.57
EXP_IMP_RAWVT_s1	1	0.0642	0.0915	0.7000	0.4833	32.79
IMP_HWWELL10	1	-0.0025	0.0083	-0.3000	0.7664	1.15
LOG_SMKB4BED	1	-0.2875	0.1149	-2.5000	0.0124	1.05
IMP_WORN25	1	-0.0097	0.0102	-0.9500	0.3437	2.83
EXP_IMP_RAWGH_s1	1	-0.0832	0.1101	-0.7600	0.45	55.18

The variables SQRT_IMP_FVC, EXP_IMP_vt_s1, EXP_IMP_gh_s1, EXP_IMP_RAWVT_s1, and EXP_IMP_RAWGH_s1 have a VIF > 10 and the VIF of variable SQRT_IMP_FEV1 is close to 10. This indicates serious multicollinearity. In this case, before blindly removing all variables, we should make sure which of the variables are most likely to have high correlations. We found that FVC (forced vital

capacity) and FEV1 (forced expiratory volume in one second) are closely related. Also, the variables gh_s1 and vt_s1 are the standardized versions of the variables RAWGH_s1 and RAWVT_s1 respectively. We start removing the variable with the highest VIF and then recalculate the VIFs to see whether the other variable VIFs decrease. After removing the variable EXP_IMP_RAWGH_s1 and recalculating the new VIFs, the highest VIF is from the variable SQRT_IMP_FVC. Then, SQRT_IMP_FVC is removed and the VIFs are reevaluated. Now, although there is no variable with VIF more than 10, we know that the variables EXP_IMP_vt_s1, EXP_IMP_gh_s1, EXP_IMP_RAWVT_s1, and EXP_IMP_RAWGH_s1 are redundant. After each pair is removed, the R^2 values of the two multiple regression models are compared. Removing the unstandardized variables, EXP_IMP_RAWVT_s1 and EXP_IMP_RAWGH_s1, gave a slightly better R^2. Thus, we choose to remove the unstandardized variables and recalculate the VIFs. The new VIF is calculated as shown in Table 5.3. This procedure was also carried out in the other two datasets. The final results are shown in Tables 5.4, and 5.5.

The datasets obtained from this procedure are used for the three modeling techniques, decision tree (DT), logistic regression, and artificial neural network (ANN). Although, multicollinearity might be good for a DT technique because the redundant or highly correlated variables will help in the case of missing data as a surrogate split, in this application, we would like to create a screening rule that will be easy to use with the fewest variables possible (a parsimonious rule). We also assume that the person who uses this rule or model has all the data variables required.

Table 5.3: Variance-inflation factors recalculated from the tree imputed dataset after removing high VIF (>10) and redundant variables

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	-0.7010	0.2404	-2.9200	0.0036	0.00
SQRT_IMP_NECK20	1	1.0843	0.1382	7.8500	<.0001	3.67
age_s1	1	0.0089	0.0008	10.7600	<.0001	2.11
IMP_waist	1	0.0038	0.0009	4.0800	<.0001	3.95
SQRT_IMP_bmi_s1	1	0.2004	0.0846	2.3700	0.0179	3.64
LOG_IMP_HDL	1	0.0816	0.1029	0.7900	0.4276	1.59
LOG_IMP_Hip	1	0.0967	0.3612	0.2700	0.789	2.04
LOG_IMP_Trig	1	0.3866	0.1752	2.2100	0.0274	1.30
SQR_HWLGALL	1	0.0400	0.0512	0.7800	0.4344	1.23
IMP_ACED1	1	0.0251	0.0763	0.3300	0.7419	1.01
EXP_IMP_vt_s1	1	-0.0374	0.0378	-0.9900	0.3225	4.54
EXP_IMP_gh_s1	1	0.0418	0.0352	1.1900	0.2342	4.25
IMP_LIPID1	1	-0.0185	0.0202	-0.9100	0.3617	1.08
AvgSYSTBP	1	0.0000	0.0004	0.0700	0.9439	1.83
LOG_IMP_MINFA10	1	0.1572	0.1017	1.5500	0.1221	1.36
IMP_STRESS15	1	0.0039	0.0089	0.4300	0.6656	1.05
gender	1	-0.1000	0.0242	-4.1200	<.0001	3.63
SQRT_IMP_FEV1	1	0.3119	0.0868	3.5900	0.0003	2.58
PWR_IMP_pcs_s1	1	-0.0581	0.0679	-0.8600	0.3916	2.75
IMP_MI15	1	-0.0120	0.0276	-0.4400	0.6623	1.10
IMP_TkPill02	1	-0.0138	0.0062	-2.2200	0.0265	1.10
SQRT_IMP_ESS_s1	1	0.1660	0.0469	3.5400	0.0004	1.60
IMP_EXCLNT25	1	0.0080	0.0113	0.7100	0.4803	3.38
SQRT_IMP_SODA10	1	0.0570	0.0475	1.2000	0.2306	1.04
AvgDiasBP	1	0.0007	0.0007	0.9800	0.3271	1.81
LOG_IMP_Mi2Slp02	1	0.1880	0.1744	1.0800	0.281	1.24
IMP_height	1	-0.0047	0.0012	-3.9400	<.0001	3.17
PWR_IMP_mcs_s1	1	0.0700	0.0678	1.0300	0.3015	2.27
IMP_LyDwn02	1	-0.0066	0.0078	-0.8400	0.3982	1.57
IMP_HWWELL10	1	-0.0015	0.0083	-0.1800	0.8549	1.15
LOG_SMKB4BED	1	-0.3338	0.1149	-2.9100	0.0037	1.05
IMP_WORN25	1	-0.0084	0.0101	-0.8300	0.405	2.75

Table 5.4: Variance-inflation factors recalculated from the conventional imputed dataset after removing high VIF (>10) and redundant variables

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	-1.6933	0.1374	-12.3300	<.0001	0.00
SQRT_IMP_NECK20	1	1.1835	0.1183	10.0000	<.0001	2.67
age_s1	1	0.0092	0.0008	11.2600	<.0001	2.02
SQRT_IMP_bmi_s1	1	0.1132	0.0997	1.1400	0.2563	5.00
Weightism	1	0.0018	0.0011	1.6800	0.0923	7.71
IMP_waist	1	0.0028	0.0009	3.1900	0.0014	3.44
SQRT_IMP_FVC	1	0.0746	0.0901	0.8300	0.408	2.43
LOG_IMP_HDL	1	-0.0666	0.1034	-0.6400	0.5194	1.60
LOG_IMP_Trig	1	0.1824	0.1867	0.9800	0.3286	1.47
SQRT_IMP_Chol	1	0.2778	0.1031	2.6900	0.0071	1.24
IMP_ACE1	1	-0.0023	0.0201	-0.1200	0.9082	1.08
AvgSYSTBP	1	-0.0001	0.0004	-0.2900	0.7714	1.84
SQRT_IMP_ESS_s1	1	0.1396	0.0406	3.4300	0.0006	1.20
IMP_NTCA1	1	-0.0307	0.0301	-1.0200	0.3082	1.02
IMP_HB02	1	0.0136	0.0067	2.0300	0.042	1.07
Race	1	0.0194	0.0126	1.5500	0.1215	1.18
LOG_IMP_Naps02	1	0.0389	0.1141	0.3400	0.7331	1.23
IMP_LOOP1	1	0.0393	0.0328	1.2000	0.2313	1.11
PWR_IMP_pcs_s1	1	-0.0287	0.0461	-0.6200	0.533	1.26
AvgDiasBP	1	0.0007	0.0007	0.9600	0.3393	1.82
IMP_HCTZ1	1	-0.0203	0.0282	-0.7200	0.4715	1.04
IMP_HrsWD02	1	0.0009	0.0055	0.1600	0.8698	1.03
IMP_LIPID1	1	-0.0189	0.0202	-0.9300	0.3508	1.07

Table 5.5: Variance-inflation factors recalculated from the listwise dataset after removing high VIF (>10) and redundant variables

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	-1.7556	0.4472	-3.9300	<.0001	0.00
NECK20	1	0.0237	0.0066	3.6100	0.0003	3.26
bmi_s1	1	0.0128	0.0071	1.7900	0.0731	5.15
age_s1	1	0.0097	0.0021	4.6000	<.0001	1.66
LoudSn02	1	0.0413	0.0198	2.0800	0.0374	1.60
AvgSYSTBP	1	-0.0008	0.0009	-0.8800	0.3779	1.25
HOSnr02	1	0.0482	0.0169	2.8600	0.0043	1.59
ESS_s1	1	-0.0070	0.0037	-1.8800	0.0599	1.14
HTNDerv_s1	1	0.0699	0.0333	2.1000	0.0358	1.20
CgPkYr	1	-0.0026	0.0009	-3.0200	0.0026	1.22
HDL	1	-0.0004	0.0011	-0.3800	0.7039	1.31
race	1	0.1223	0.0507	2.4200	0.0159	1.07
Chol	1	0.0006	0.0004	1.4500	0.1461	1.07
Mi2Slp02	1	0.0000	0.0009	0.0100	0.995	1.08
CALM25	1	0.0160	0.0153	1.0400	0.2981	1.13
NGES02	1	0.0223	0.0167	1.3400	0.1811	1.24
height	1	0.0009	0.0030	0.3000	0.7619	3.62
WRFACE10	1	0.0330	0.0188	1.7500	0.0799	1.05
FVC	1	-0.0126	0.0249	-0.5100	0.6125	2.90
Hip	1	-0.0014	0.0030	-0.4500	0.6529	3.60
Alcoh	1	0.0049	0.0024	2.0500	0.0405	1.18
TEA15	1	-0.0129	0.0149	-0.8600	0.3882	1.03

5.3 OSA screening rule construction

In this section, the three datasets prepared in the previous sections are used for constructing the classification models. The target or independent variable in our case is the indicator that the person currently has OSA disease (AHI≥5, then apnea=1) or does not have OSA disease (AHI<5, then apnea=0).

As shown in Figure 5.1, at the beginning of the workflow, the raw dataset is partitioned into thirds, for training, validation, and testing. The model training process will initially involve the training and validation data partitions. The testing data partition is left out of the modeling process and will be use in the end for evaluating the performance of the developed models. The training partition, the main part of the dataset, is used for finding patterns and creating initial models. The validation partition is used for

evaluating the performance of the models developed from the training data partition. One of the main reasons for using the validation dataset is to prevent overfitting. As is commonly known, adding more complexity to the model improves the classification performance in the training dataset. However, an overly complex model will fit itself to the noise in the data. This over-complexity can be seen if the model is applied to the validation dataset and each training iteration is checked and the model performances used in both training and validation datasets are comparable. The training is stopped if the validation MSE increases which reflects the point where the overfitting occurs. Finally, the model is revalidated with the testing dataset, which is not involved in the training process, to see how the model performs in the real situation.

5.3.1 OSA screening rule derived from the logistic regression method

As the normality assumption is not required for the logistic regression, we use the variables in the datasets directly in modeling process without any transformation. Doing so more convenient for the end user, as the implementation will not require any calculation in variable transformation. As a standard procedure, the logistic regression starts by determining which variables to be included in the regression. In this research, we use the stepwise method, which is explained as the process develops.

Intuitively, the stepwise variable selection method comprises of two steps in one iteration: determining which variable is significant enough to enter the model and determining the significance level of each parameter in the model. In this study, a significance level of 0.05 is used for both entering and staying in the model. Before the selection process starts, the algorithm builds the baseline model with only the intercept term. This model predicts the overall average target values for all cases. Then, the variable selection starts. First, every variable is used for fitting the one-variable model. The first variable selected into the model is the one that most improves it over the baseline model (the highest Chi-square). Then, the algorithm reevaluates the statistical significance of all included inputs. If any included input is not significant enough, it is removed. The process terminates when there is no variable significant enough to be included in the model and the inputs already in the model are all significant. Then, the performance of the model in each iteration is evaluated and the one with the best performance is chosen as the final model.

Figure 5.6 summarizes the steps of the stepwise algorithm. In the first step, the interaction of variables age_s1 and NECK20 enters the model. After the parameter is estimated by a maximum likelihood algorithm, the overall model's significance is tested as shown in Figure 5.7. Because the overall model is significant at level 0.05, we conclude that this variable is significant enough to enter the model. Also, the significant of each variable in the model is reevaluated by the Wald Chi-square statistics as shown in Figure 5.7. All variables are significant at the 0.05 level. This process progresses iteratively as shown in Figure 5.6. At step 8 with the entry of the interaction between age_s1 and bmi_s1, after every variable is reevaluated, the variables age_s1 and weightim became insignificant so they were removed from the model. The stepwise algorithm terminated at step 14. The performance (misclassification rate in this case) of the models obtained in every step is evaluated using the validation dataset. In this case, the model in step 3 with the variables Intercept, IMP_NECK20, IMP_waist*age_s1, age_s1*weightim, was chosen to be the final model because it has the lowest misclassification rate. For the other two datasets, listwise and DT imputation, the results for fitting the logistic regression are shown in Figures 5.8 and 5.9.

Summary of Stepwise Selection

Step	Effect Entered	Effect Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq	Validation Misclassification Rate
1	age_s1*weightim		1	1	446.9992		<.0001	0.3288
2	IMP_NECK20		1	2	92.1480		<.0001	0.3281
3	IMP_waist*age_s1		1	3	30.2999		<.0001	0.3216
4	IMP_Chol*IMP_ESS_s1		1	4	19.6143		<.0001	0.3242
5	IMP_HB02*race		8	5	27.3326		0.0006	0.3301
6	IMP_LIPID1*IMP_NTCA1		1	6	9.6933		0.0018	0.3320
7	IMP_HB02*IMP_LIPID1		4	7	14.0753		0.0071	0.3353
8	IMP_bmi_s1*age_s1		1	8	6.6592		0.0099	0.3333
9		age_s1*weightim	1	7		0.2349	0.6279	0.3320
10	IMP_FVC*IMP_Trig		1	8	5.1954		0.0226	0.3359
11	age_s1*age_s1		1	9	5.4051		0.0201	0.3301
12		IMP_waist*age_s1	1	8		2.5491	0.1104	0.3327
13	IMP_bmi_s1*IMP_waist		1	9	18.4207		<.0001	0.3275
14		IMP_bmi_s1*age_s1	1	8		1.6678	0.1966	0.3275

The selected model, based on the misclassification rate for the validation data, is the model trained in Step 3. It consists of the following effects:

Intercept IMP_NECK20 IMP_waist*age_s1 age_s1*weightim

Figure 5.6: Summary of the stepwise selection algorithm used in the logistic regression from the conventionally imputed dataset

Likelihood Ratio Test for Global Null Hypothesis: BETA=0

-2 Log Likelihood Intercept Only	-2 Log Likelihood Intercept & Covariates	Likelihood Ratio Chi-Square	DF	Pr > ChiSq
4949.016	4471.421	477.5947	1	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Standard Estimate	Wald Error	Chi-Square	Pr > ChiSq	Standardized Estimate	Exp(Est)
Intercept	1	-3.2548	0.1670	379.91	<.0001		0.039
age_s1*weightim	1	0.000660	0.000033	397.52	<.0001		1.001

Figure 5.7: Global null hypothesis test in the first iteration in the logistic regression from the conventionally imputed dataset

Summary of Stepwise Selection

Step	Effect Entered	Effect Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq	Validation Misclassification Rate
1	IMP_NECK20*IMP_waist		1	1	456.1934		<.0001	0.3503
2	IMP_NECK20*age_s1		1	2	126.4418		<.0001	0.3222
3	IMP_ESS_s1*IMP_bmi_s1		1	3	21.5478		<.0001	0.3209
4	IMP_FEV1*IMP_Trig		1	4	10.7390		0.0010	0.3183
5	SMKB4BED*age_s1		1	5	8.9351		0.0028	0.3183
6	IMP_ESS_s1*IMP_Trig		1	6	5.7279		0.0167	0.3229
7	IMP_height*IMP_pcs_s1		1	7	6.3515		0.0117	0.3177
8	IMP_NECK20*IMP_pcs_s1		1	8	6.3444		0.0118	0.3151
9	IMP_height*age_s1		1	9	17.8056		<.0001	0.3203
10	IMP_LIPID1*IMP_TkPill02		4	10	11.3483		0.0229	0.3216
11	IMP_HWWELL10*SMKB4BED		1	11	4.8922		0.0270	0.3216
12	IMP_LyDwn02*IMP_WORN25		15	12	25.4235		0.0445	0.3229
13		IMP_LyDwn02*IMP_WORN25	15	11		24.7727	0.0531	0.3216

The selected model, based on the misclassification rate for the validation data, is the model trained in Step 8. It consists of the following effects:

Intercept IMP_ESS_s1*IMP_Trig IMP_ESS_s1*IMP_bmi_s1 IMP_FEV1*IMP_Trig IMP_NECK20*IMP_pcs_s1 IMP_NECK20*IMP_waist IMP_NECK20*age_s1 IMP_height*IMP_pcs_s1 SMKB4BED*age_s1

Figure 5.8: Summary of the stepwise selection algorithm used in the logistic regression from the Tree imputed dataset

Summary of Stepwise Selection

Step	Effect Entered	Effect Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq	Validation Misclassification Rate
1	NECK20*age_s1		1	1	70.4029		<.0001	0.3679
2	HOSnr02		3	2	33.6186		<.0001	0.3886
3	NECK20*bmi_s1		1	3	14.5143		0.0001	0.4093
4	race		2	4	10.9433		0.0042	0.4041
5	Alcoh*Chol		1	5	5.0840		0.0241	0.3990
6	Alcoh		1	6	7.2802		0.0070	0.3938
7	AvgSYSTBP*CgPkYr		1	7	6.7319		0.0095	0.4041
8	Alcoh*bmi_s1		1	8	4.5188		0.0335	0.4041
9	CALM25*NGES02		20	9	31.4673		0.0493	0.3679
10		CALM25*NGES02	20	8		28.0027	0.1093	0.4041

The selected model, based on the misclassification rate for the validation data, is the model trained in Step 1. It consists of the following effects:

Intercept NECK20*age_s1

Figure 5.9: Summary of the stepwise selection algorithm used in the logistic regression from the listwise dataset

Shown in Table 5.6, the important OSA predictors appear that in all three models are the interactions between NECK20 (neck circumference in centimeters), Epworth sleep scale (ESS), age, waist size in inches, forced expiratory volume in one second (FEV1), triglycerides, and weight. The other interesting OSA predictors are interactions between age and the indicator for smoking before going to bed (SMKB4BED), and the Physical Component Scale Standardized Score (pcs_s1). The initial performance evaluated from the validation misclassification rates are 32.16% in model I, 31.51% in model II, and 36.79% in model III. These initial performances suggest that the listwise (complete) dataset is not a very good choice, as seen in the higher misclassification rates compared to the other two models. Again, to find the best model here, the final performance must be evaluated by selecting the appropriate cut-off value for each model such that the false negative rate (FNR) is below the threshold (5% in this study). That topic is discussed later in this chapter.

Table 5.6: Effects selected in the final logistic regression models from different datasets

Effects in the final model		
Model I: Conventional imputation	**Model II: Tree imputation**	**Model III: Listwise**
NECK20	ESS_s1*Trig	NECK20*age_s1
waist*age_s1	ESS_s1*bmi_s1	
age_s1*weightim	FEV1*Trig	
	NECK20*pcs_s1	
	NECK20*waist	
	NECK20*age_s1	
	height*pcs_s1	
	SMKB4BED*age_s1	

5.3.2 OSA screening rule derived from the decision tree method

The decision tree (DT) is a non-parametric method that does not have any equation or coefficient. The tree is built based on the patterns found from the training data. As explained in the methodology section, tree construction starts with the root node, which is where the best split occurs. The objective function of each split is to minimize the impurity measurement computed from the ratio of the target group (apnea vs. non-apnea). The algorithm will search for the variable and its value where the impurity measurement is

minimized. In this study, a Gini impurity index is used. The other criteria (splitting and stoppage) used in this study are shown in Table 5.7.

We applied the DT algorithm to our three datasets: tree imputation, conventional imputation, and listwise datasets. The same data partitions used in the logistic regression model are also used in this modeling section. The tree construction from the tree imputation dataset is used to demonstrate. The final tree obtained from this dataset is shown in Figure 5.10.

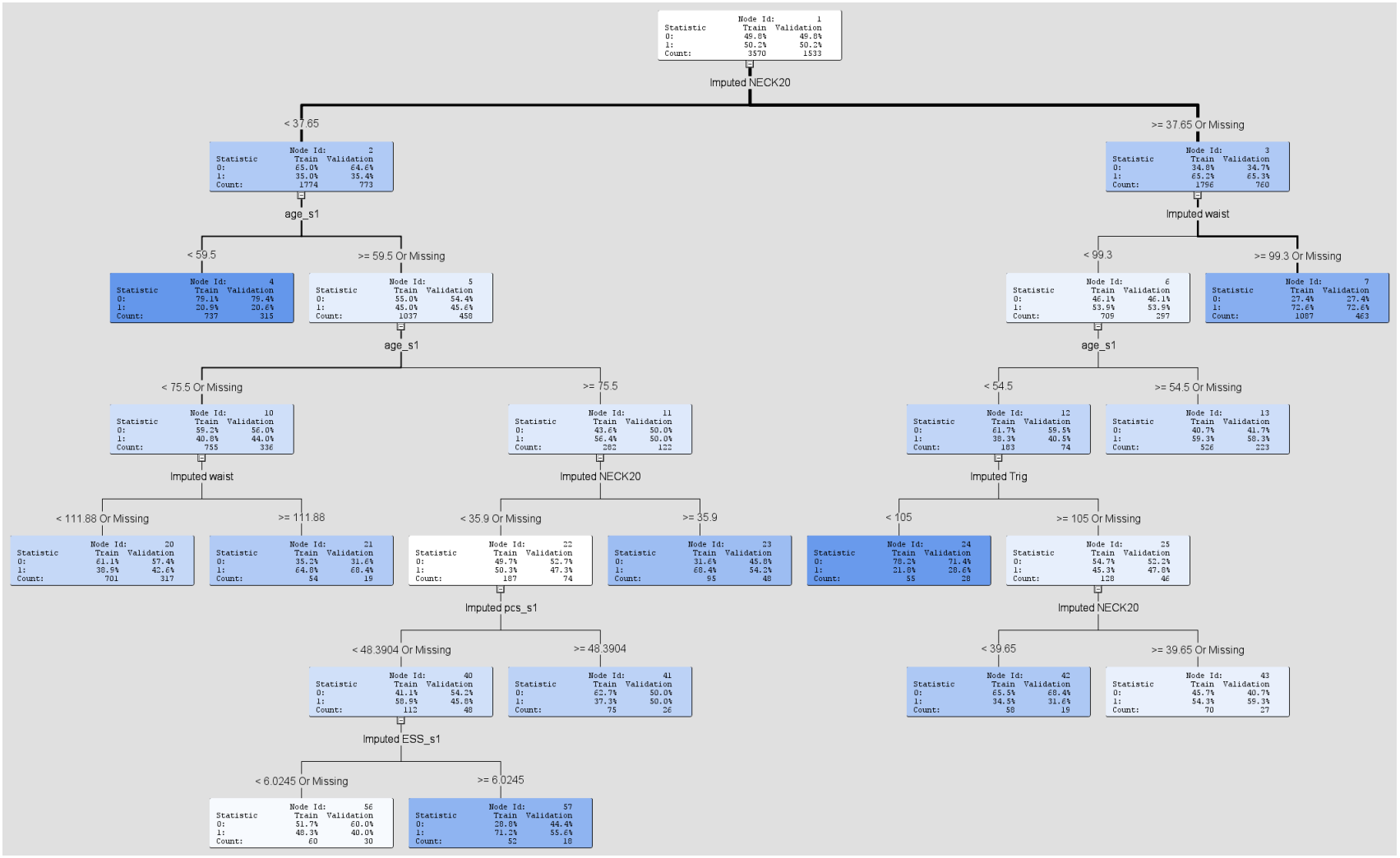


Figure 5.10: Decision tree constructed from a tree imputed dataset

Table 5.7: Criteria used for constructing decision trees in this study

Impurity measurement	Gini impurity index
Splitting significance level	0.2
Maximum split branch at each split	2 (binary tree)
Maximum tree depth	10
Number of cases required at each leaf	50
Initial performance measurement	Validation misclassification rate

Table 5.8: Variable importance calculated from every splitting in the DT constructed from the tree imputed dataset

Obs	NAME	NRULES	NSURROGATES	IMPORTANCE
1	IMP_NECK20	3	4	1
2	IMP_waist	2	4	0.91696
5	IMP_Trig	1	5	0.84861
12	age_s1	3	2	0.63831
13	IMP_pcs_s1	1	6	0.57035
18	IMP_ESS_s1	1	2	0.24228

Shown in Figure 5.10, the root node starts from variable NECK20, which is the neck size, then it spits into two leaves. This process kept going until one of the stoppage criteria shown in Table 5.7 has been reached. The relative variable importance is the reflection of how well the variables perform in the classification. In this model, the neck circumference is the most important variable. However, there is no guarantee that the construction of the tree stopped at the criteria will be optimal. This is where the pruning process and validation dataset are used. The pruning process removes the splits that are not relatively significant in the tree (see the methodology chapter for more details). Furthermore, the plot of the initial performance indexes (in our case, training and validation misclassification rates) obtained from both training and validation data partitions is used for comparison (see Figure 5.11). The plot suggests that the optimal number of leaves is twelve. As the tree grows larger, it is expected that the classification performance will be better, just as when we add more variables to the regression model. However, growing too many leaves also introduces the overfitting problem, where the tree starts to fit the noise in the data rather than capturing the true patterns for distinguishing between the target groups. The best point to stop

the tree's growth is the first point before the performance index of the validation partition starts to get worse (an increased validation misclassification rate in this case).

Shown in Figure 5.11, the misclassification rates in both training and validating partitions decrease quickly after the very first leaf then steadily decrease until saturate at 12 leaves. The validation

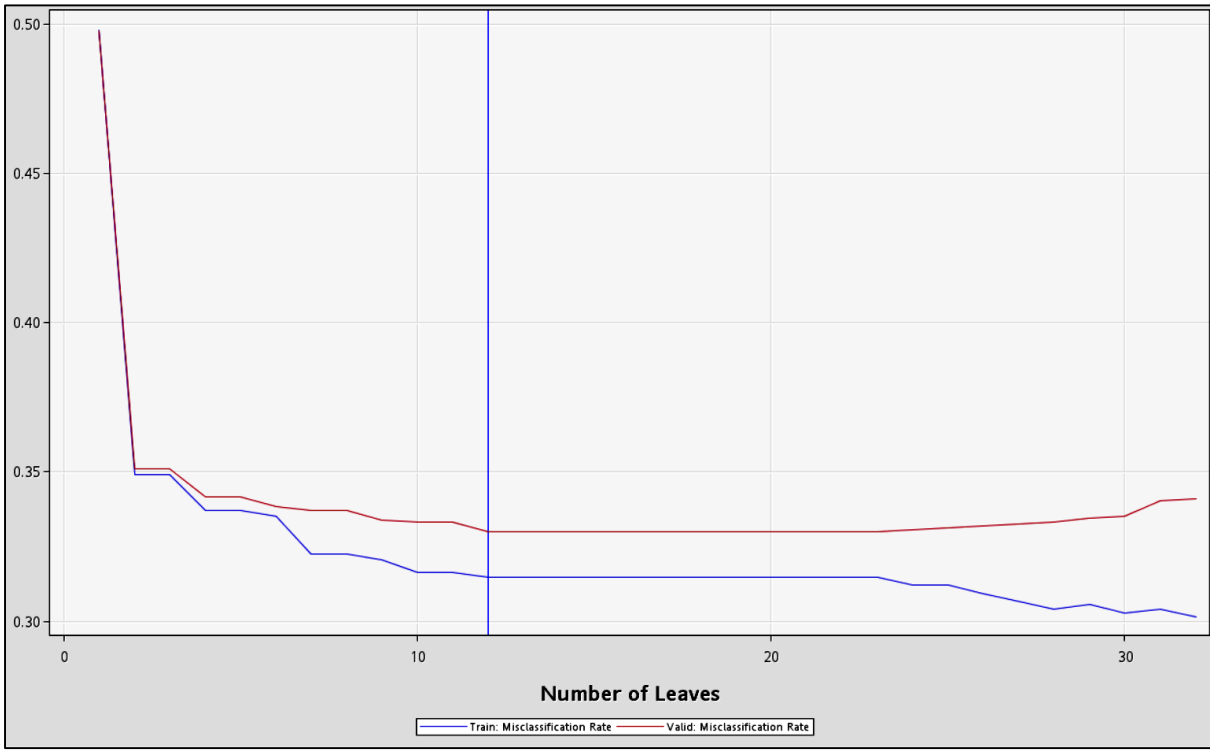


Figure 5.11: Misclassification rates plotted from performances of decision tree models built from training and validation data partitions of the tree imputed dataset

misclassification rate started to increase when the tree had around 23 leaves, so it is better to keep the tree simple and avoid the overfitting problem. Only 12 leaves are used in the final classification tree. The procedure explained above is also used in constructing trees from the other two datasets. The tree models are shown in Figures 5.11 and 5.12. The importance of the variables in each model are included in Table 5.9.

Table 5.9: Important variables in the final tree models from different datasets

Variables used for building tree models		
Model I: Conventional imputation	**Model II: Tree imputation**	**Model III: Listwise**
IMP_NECK20	IMP_NECK20	bmi_s1
IMP_waist	IMP_waist	age_s1
IMP_Trig	IMP_Trig	
IMP_bmi_s1	age_s1	
IMP_ESS_s1	IMP_pcs_s1	
age_s1	IMP_ESS_s1	
IMP_pcs_s1		
IMP_HB02		

The variables used for building the classification trees listed in Table 5.9 are mostly in accordance not only for the tree models but also the logistic regression models developed in the previous section. The neck circumference is still the best predictor for differentiating people with and without OSA. The other important predictors are waist size, triglyceride level, body mass index (BMI), age, Epworth sleep score (ESS), Physical Component Scale Standardized Score (pcs_s1), and frequency of being awakened with heartburn or indigestion (HB02). The initial performance of the tree models evaluated by the validation misclassification rate are 33.65% for model I, 33.00% for model II, and 39.37% for model III. These initial performance indexes agree with the results obtained from the logistic regression models. This confirms that the modeling using the listwise (complete) dataset does not give a very good result compared to the imputed one because a lot of useful information is meaninglessly discarded. One of the advantages of using a tree model is that it is very easy to understand and implement. For example, from the tree model in Figure 5.10, we can write out all the splitting rules as English rules as follows:

Table 5.10: English rules created from the tree constructed from the tree imputed data

----

Node = 4

----

if age_s1 < 59.5

AND Imputed NECK20 < 37.65

then

Tree Node Identifier = 4

Number of Observations = 737

Predicted: apnea=1 = 0.21

Predicted: apnea=0 = 0.79

Node = 7

----

if Imputed waist >= 99.3 or MISSING

AND Imputed NECK20 >= 37.65 or MISSING

then

Tree Node Identifier = 7

Number of Observations = 1087

Predicted: apnea=1 = 0.73

Predicted: apnea=0 = 0.27

Node = 13

----

if age_s1 >= 54.5 or MISSING

AND Imputed waist < 99.3

AND Imputed NECK20 >= 37.65 or MISSING

then

Tree Node Identifier = 13

Number of Observations = 526

Predicted: apnea=1 = 0.59

Predicted: apnea=0 = 0.41

Node = 20

----

if age_s1 < 75.5 AND age_s1 >= 59.5 or MISSING

AND Imputed waist < 111.88 or MISSING

AND Imputed NECK20 < 37.65

then

Tree Node Identifier = 20

Number of Observations = 701

Predicted: apnea=1 = 0.39

Predicted: apnea=0 = 0.61

Node = 21

----

if age_s1 < 75.5 AND age_s1 >= 59.5 or MISSING

AND Imputed waist >= 111.88

AND Imputed NECK20 < 37.65

then

Tree Node Identifier = 21

Number of Observations = 54

Predicted: apnea=1 = 0.65

Predicted: apnea=0 = 0.35

Node = 23

----

if age_s1 >= 75.5

AND Imputed NECK20 < 37.65 AND Imputed NECK20 >= 35.9

then

Tree Node Identifier = 23

Number of Observations = 95

Predicted: apnea=1 = 0.68

Predicted: apnea=0 = 0.32

Node = 24

----

if age_s1 < 54.5

AND Imputed waist < 99.3

AND Imputed Trig < 105

AND Imputed NECK20 >= 37.65 or MISSING

then

Tree Node Identifier = 24

Number of Observations = 55

Predicted: apnea=1 = 0.22

Predicted: apnea=0 = 0.78

Node = 41

----

if age_s1 >= 75.5

AND Imputed pcs_s1 >= 48.3904

AND Imputed NECK20 < 35.9 or MISSING

then

Tree Node Identifier = 41

Number of Observations = 75

Predicted: apnea=1 = 0.37

Predicted: apnea=0 = 0.63

Node = 42

----

if age_s1 < 54.5

AND Imputed waist < 99.3

AND Imputed Trig >= 105 or MISSING

AND Imputed NECK20 < 39.65 AND Imputed NECK20 >= 37.65

then

Tree Node Identifier = 42

Number of Observations = 58

Predicted: apnea=1 = 0.34

Predicted: apnea=0 = 0.66

Node = 43

----

if age_s1 < 54.5

AND Imputed waist < 99.3

AND Imputed Trig >= 105 or MISSING

AND Imputed NECK20 >= 39.65 or MISSING

then

Tree Node Identifier = 43

Number of Observations = 70

Predicted: apnea=1 = 0.54

Predicted: apnea=0 = 0.46

Node = 56

----

if age_s1 >= 75.5

AND Imputed pcs_s1 < 48.3904 or MISSING

AND Imputed NECK20 < 35.9 or MISSING

AND Imputed ESS_s1 < 6.02451 or MISSING

then

Tree Node Identifier = 56

Number of Observations = 60

Predicted: apnea=1 = 0.48

Predicted: apnea=0 = 0.52

Node = 57

----

if age_s1 >= 75.5

AND Imputed pcs_s1 < 48.3904 or MISSING

AND Imputed NECK20 < 35.9 or MISSING

AND Imputed ESS_s1 >= 6.02451

then

Tree Node Identifier = 57

Number of Observations = 52

Predicted: apnea=1 = 0.71

Predicted: apnea=0 = 0.29

To use these rules for classification of OSA, the user will have to match the data to the case. For example, for the model developed by the tree imputed data, 6 variables are required as inputs: NECK20, waist, Trig, age_s1, pcs_s1, and ESS_s1. We pick one data record from the dataset as follows:

Table 5.11: Attributes required for using the tree model derived from tree imputed data from an observation number 29

IMP_NECK20	43
IMP_waist	108
IMP_Trig	145
age_s1	57
IMP_pcs_s1	57.26
IMP_ESS_s1	10

Applying these data to the tree in Figure 5.10 and English rules in Table 5.10, this case will end at node 7. At node 7, the prediction of the case uses a default cut-off value at 50%, or the majority rule so because 72.6% of the people have OSA (target = 1), we will group the person in this case as having OSA. Again, before using any of the tree models developed in this section, the cut-off value must be appropriately defined so that the model behaves according to our objective, to identify people without OSA (negative case) with the lowest the false negative rate (<5%). The method to select the cut-off value is is discussed later in this chapter.

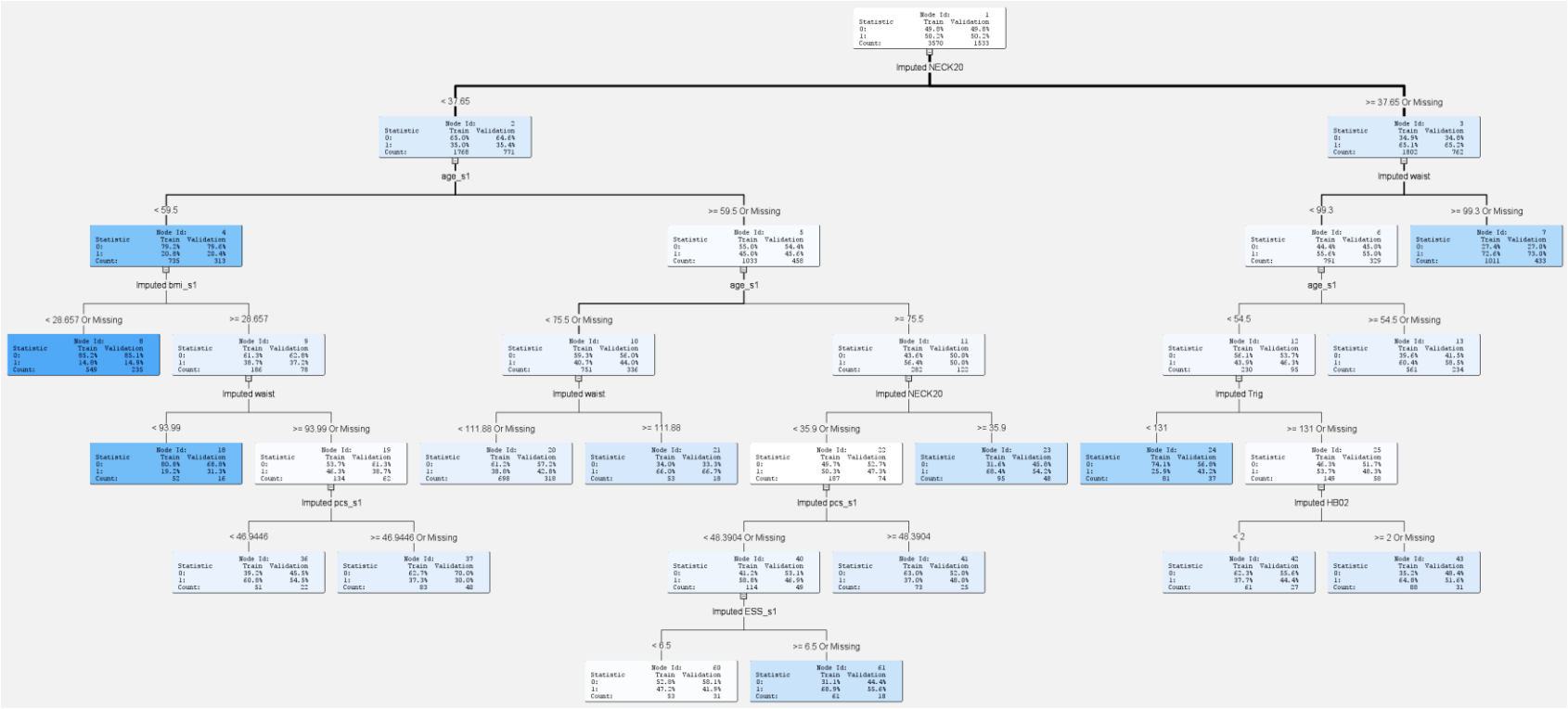


Figure 5.12: Decision tree constructed from a conventionally imputed dataset

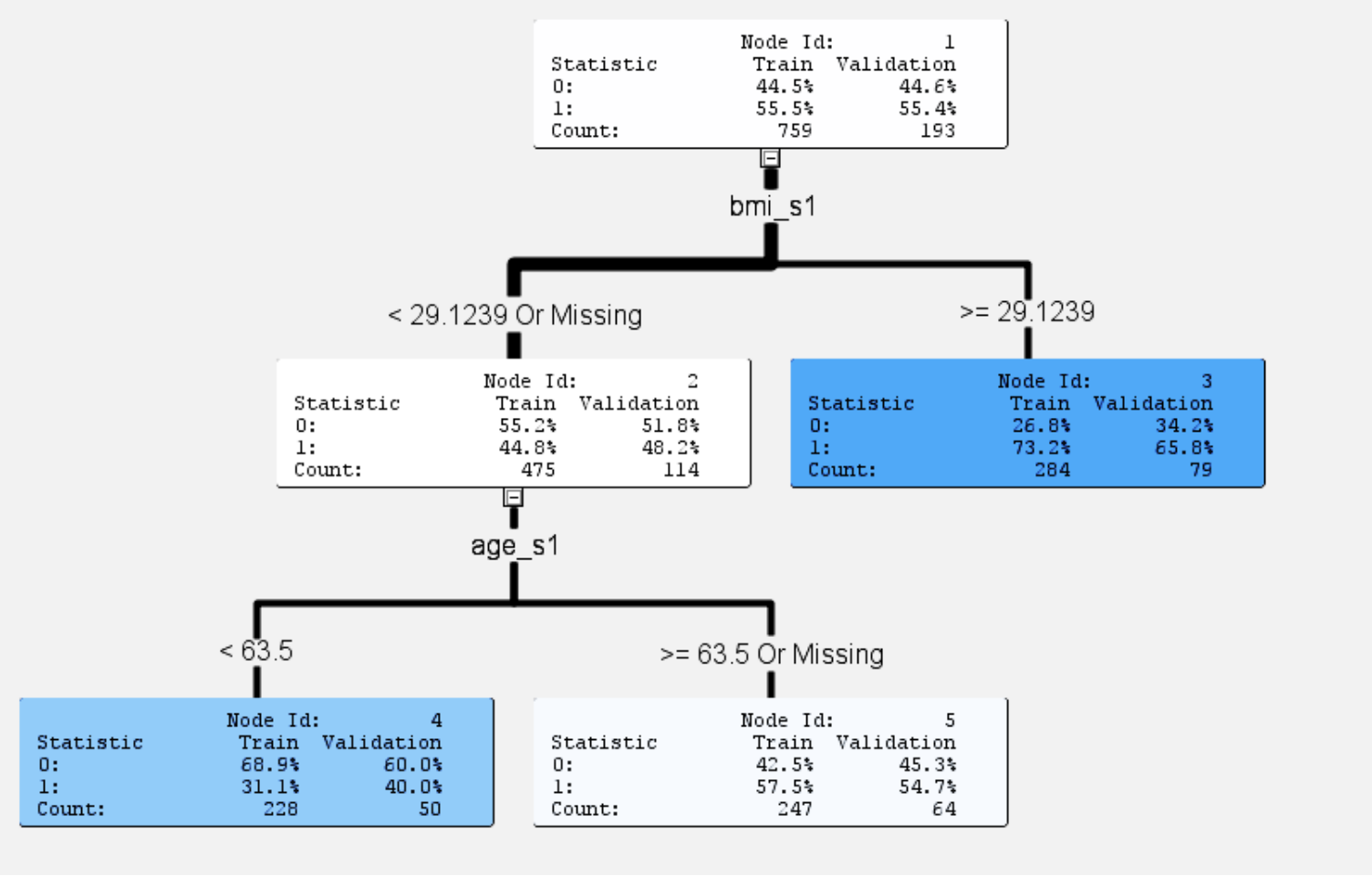


Figure 5.13: Decision tree constructed from a listwise dataset

5.3.3 OSA screening rule derived from an artificial neural network method

An artificial neural network (ANN) is a mathematical network that has many interconnections, imitating the structure of biological neural networks. As introduced in the methodology section, the main idea of the ANN is to map the input to the corresponding output via the pattern recognized from the inputs. The architecture of the ANN depends mainly on the complexity of the problem. In this section, we chose to use one hidden-layer perception network. The log-sigmoid function is used as a transfer function, which is appropriated to be used with a binary target, such as our case. The architecture of the ANN is shown in Figure 5.14 below:

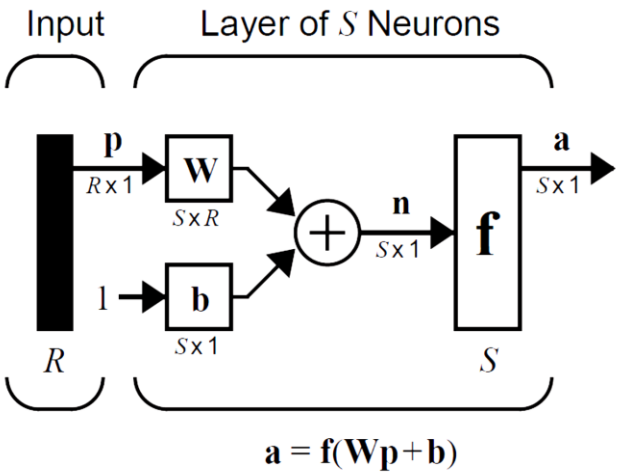


Figure 5.14: Multilayer perceptron (MLP) network (reproduced from [108])

The training process for an ANN is to estimate parameters, bias, and weights that are associated with each input and which will be summed and translated to the output by the log-sigmoid function as a transfer function. As mentioned in the methodology section, the output of the network is the probability that the person will have OSA. In this study, the training algorithm used is a backpropagation algorithm. Intuitively, backpropagation is actually a searching algorithm that searches for the minimum of the mean square error as a performance surface. Iteratively, the search direction is determined by the negative gradient of the direction vector of the parameters. For detailed information, see the artificial neural network section in the methodology chapter.

In practical training, as is commonly known, the ANN does not have any algorithm to select the independent variables that truly correlate to the dependent variable. It is computational-intensive and using all the inputs to train the ANN model may introduce the risk of overfitting. In this case, we use variables selected from the decision trees in the previous sections as inputs for training the ANN models. The variables used are shown in Table 5.12 below:

Table 5.12: Selected variables from DT and logistic regression models

Variables used for training ANN using dataset from:		
Conventional imputation: Variable selected by DT	**Tree imputation: Variable selected by DT**	**Listwise: Variable selected by DT**
IMP_NECK20	IMP_NECK20	NECK20
IMP_waist	age_s1	age_s1
IMP_Trig	IMP_waist	bmi_s1
IMP_bmi_s1	IMP_FVC	
IMP_ESS_s1	IMP_HDL	
age_s1	AvgSYSTBP	
IMP_pcs_s1	IMP_pcs_s1	
IMP_HB02	IMP_RAWMH_s1	
	IMP_RAWPF_s1	

Because most of the variables in the same group are selected by both tree and logistic regression models, all variables are used as inputs for the ANN training. Regarding the ANN architecture used (one hidden layer), we can write the equation for the hidden units in our case as:

$$H_1 = f_1(w_{01} + w_{11}x_1 + w_{21}x_2 + \cdots + w_{k1}x_k) \quad (5.2)$$

$$H_2 = f_2(w_{02} + w_{12}x_1 + w_{22}x_2 + \cdots + w_{k2}x_k)$$

$$H_i = f_i(w_{0i} + w_{1i}x_1 + w_{2i}x_2 + \cdots + w_{ki}x_k)$$

Where k = the number of input variables used in ANN model

i = the number of hidden units used in ANN model

f = any chosen transfer function

Thus, the final model of the general ANN in our case is:

$$y = (w_0 + w_1H_1 + w_2H_2 + \cdots + w_kH_k) \quad (5.3)$$

$$or\ log\frac{p}{(1-p)} = (w_0 + w_1H_1 + w_2H_2 + \cdots + w_kH_k)$$

Where y = a summation of the output from every hidden unit

p = posterior probability that the person will have OSA

As suggested by [108], for the pattern recognition application, a hyperbolic tangent function, $h = (e^{x} - e^{-x})/(e^{x} + e^{-x})$, is one of the most appropriate transfer functions in the hidden layer. To the best of our knowledge, there is no scientific method for estimating the optimal number of parameters or hidden units used in an ANN model. We first tried the number of hidden units equal to the number of input variables. Then, we increased the number of hidden units. If the performance of the model increases (without overfitting), we keep increasing the number of hidden units until the performance does not improve. However, if the performance of the model does not increase, we decrease the number of hidden units until the performance decreases. Then, use the last number of hidden units before the performance decreased. This way, we will not use too many hidden units, which will help prevent the overfitting problem and also save the computational cost of so many terms in the model.

Using a back propagation as a training algorithm to train the ANN model from the tree imputation dataset, we started with 9 hidden units, which equals the number of input variables. Then, we increased the number to 10, 20, and 30 hidden units to see whether the model improves in terms of the average square error. Finally, we iteratively decreased the number of hidden units to one. The results in terms of validation average square error is shown in Table 5.13.

Table 5.13: Model selection based on an average square error of validation data partition in the tree imputation dataset

Model	Number of hidden units	Validation: average square error	Validation: misclassification rate
1	7	0.20733	0.31181
2	4	0.20771	0.31181
3	30	0.2078	0.30724
4	1	0.20827	0.31442
5	20	0.20885	0.31507

Table 5.13 shows only the top 5 models based on the average square error (ASE) of the validation data partition in the tree imputation dataset (the lower the better). It is common to see that increasing the number of parameters improves the performance of a model (lower ASE). However, the tradeoff of having many parameters is the model complexity and the risk of model overfitting. We found that in our case, the optimal number of hidden units is one, because there is no significant improvement in the model with more hidden units, as seen from the ASE and misclassification rates (MISC). It is better to keep a model simple if the performance is comparable. The estimated parameters for the model are shown in Table 5.14. From these estimated parameters, we can implement the ANN model according to the previously mentioned architecture by using a pseudo code shown in Table 5.15. The prefix "S_" in front of each parameter name refers to a standardized version of that parameter. This procedure transforms the data to a zero mean and a unit variance in order to avoid saturation at the transfer function nodes (tanh, and logistic function) which transform the output to (-1, 1) and (0, 1).

Table 5.14: Parameters estimated for an ANN model from the tree imputed dataset (Model 1)

LABEL	FROM	TO	WEIGHT
AvgSYSTBP -> H11	AvgSYSTBP	H11	0.0284
BIAS -> H11	BIAS	H11	0.0859
IMP_FVC -> H11	IMP_FVC	H11	-0.0774
IMP_HDL -> H11	IMP_HDL	H11	0.0133
IMP_NECK20 -> H11	IMP_NECK20	H11	0.6979
IMP_RAWMH_s1 -> H11	IMP_RAWMH_s1	H11	0.0245
IMP_RAWPF_s1 -> H11	IMP_RAWPF_s1	H11	-0.0277
IMP_pcs_s1 -> H11	IMP_pcs_s1	H11	-0.0187
IMP_waist -> H11	IMP_waist	H11	0.3576
age_s1 -> H11	age_s1	H11	0.3705
BIAS -> apnea1	BIAS	apnea1	-0.0759
H11 -> apnea1	H11	apnea1	1.8355

Table 5.15: Pseudo-code for implementing the ANN model from the estimated parameters in Table 5.14

```
H11 =    0.0284029085087 * S_AvgSYSTBP +   -0.07736196891919 *
         S_IMP_FVC +  0.01326096899979 * S_IMP_HDL
         +    0.69788460732452 * S_IMP_NECK20 +    0.02452487611675 *
         S_IMP_RAWMH_s1 +    -0.0277362803885 * S_IMP_RAWPF_s1
         +    -0.0186558726798 * S_IMP_pcs_s1 +    0.35755807366539 *
         S_IMP_waist +    0.37051793731458 * S_age_s1 ;

H11 =    0.08592866364664 + H11 ;
H11 =    TANH(H11 );

P_apnea1 =    1.83548737907284 * H11 ;
P_apnea1 =   -0.07588404449354 + P_apnea1 ;
P_apnea1 =    1.0 / (1.0 + EXP(- P_apnea1));
```

Table 5.16: Estimated ANN parameters from the conventional imputed dataset (Model 2)

LABEL	FROM	TO	WEIGHT
AvgSYSTBP -> H11	AvgSYSTBP	H11	-0.0116
BIAS -> H11	BIAS	H11	0.2649
HWLGALL -> H11	HWLGALL	H11	-0.0395
IMP_FVC -> H11	IMP_FVC	H11	0.0424
IMP_HDL -> H11	IMP_HDL	H11	0.0695
IMP_HrsWD02 -> H11	IMP_HrsWD02	H11	-0.0316
IMP_HrsWE02 -> H11	IMP_HrsWE02	H11	0.0256
IMP_NECK20 -> H11	IMP_NECK20	H11	0.5358
IMP_RAWMH_s1 -> H11	IMP_RAWMH_s1	H11	0.0902
IMP_RAWPF_s1 -> H11	IMP_RAWPF_s1	H11	-0.0612
IMP_SODA15 -> H11	IMP_SODA15	H11	0.1157
IMP_STRESS151 -> H11	IMP_STRESS151	H11	-0.0309
IMP_STRESS152 -> H11	IMP_STRESS152	H11	-0.0283
IMP_Trig -> H11	IMP_Trig	H11	0.0998
IMP_bmi_s1 -> H11	IMP_bmi_s1	H11	0.2570
IMP_height -> H11	IMP_height	H11	0.0126
IMP_mcs_s1 -> H11	IMP_mcs_s1	H11	-0.0632
IMP_waist -> H11	IMP_waist	H11	0.0845
age_s1 -> H11	age_s1	H11	0.4493
race1 -> H11	race1	H11	-0.0803
race2 -> H11	race2	H11	-0.1681
BIAS -> apnea1	BIAS	apnea1	-0.2141
H11 -> apnea1	H11	apnea1	2.0299

The result of the pseudo-code in Table 5.15 is variable P_apnea1, which is the probability of the person is having sleep apnea. By default, the person will have a positive result when having P_apnea1 more than or equal to 0.5. However, the cut-off probability at 0.5 is not always optimal. The method used for choosing a cut-off probability value is discussed in the next section. Using the same process mentioned previously, we also constructed the ANN models from the other two datasets as shown in Table 5.16 and Table 5.17.

Table 5.17: Estimated ANN parameters from the listwise dataset (Model3)

LABEL	FROM	TO	WEIGHT
BIAS -> H11	BIAS	H11	-0.0681
BIAS -> H12	BIAS	H12	4.0603
BIAS -> H13	BIAS	H13	-2.7317
BIAS -> H14	BIAS	H14	-2.4503
BIAS -> apnea1	BIAS	apnea1	0.7937
H11 -> apnea1	H11	apnea1	1.3904
H12 -> apnea1	H12	apnea1	4.6940
H13 -> apnea1	H13	apnea1	4.8610
H14 -> apnea1	H14	apnea1	0.9326
NECK20 -> H11	NECK20	H11	0.0394
NECK20 -> H12	NECK20	H12	1.9683
NECK20 -> H13	NECK20	H13	-1.0289
NECK20 -> H14	NECK20	H14	2.8934
age_s1 -> H11	age_s1	H11	-0.0776
age_s1 -> H12	age_s1	H12	4.2579
age_s1 -> H13	age_s1	H13	-2.8372
age_s1 -> H14	age_s1	H14	3.2981
bmi_s1 -> H11	bmi_s1	H11	1.6384
bmi_s1 -> H12	bmi_s1	H12	-2.1128
bmi_s1 -> H13	bmi_s1	H13	1.3514
bmi_s1 -> H14	bmi_s1	H14	-5.8023

Again, to prevent the overfitting problem, we used the same process by monitoring the initial performance of the models from the validation dataset. The final model performances (validation misclassification rates) are 31.44% for model I, 31.05% for model II, and 35.23% for model III. Comparing these initial performances, we see that the ANN models constructed from the tree imputed dataset and

conventionally imputed dataset have almost the same validation misclassification rates. However, the number of parameters from the tree imputed dataset that are required in the ANN model is significantly smaller (12 vs. 23 parameters). Also, as expected, the ANN model derived from the listwise dataset performed poorly compared to the other two ANN models. Still, we need to fine tune every model developed so far to achieve the objective of minimizing the false negative rates of the classification of people with OSA.

5.4 Performance comparison of OSA screening rules

To choose which model is the best for an OSA screening task, we select the model that gives the highest predictive accuracy in the testing data partition while keeping the false negative rate within the threshold. In this study, we choose two thresholds, 5% and 10% false negative rates. To maintain these thresholds, the ROC curve is used as a tool to monitor each model's performance while varying the model's percentage cut-off threshold (i.e., $y < 50$: does not have the disease, $y \geq 50$: have the disease). The procedure is as follows:

1. From each final model, vary the cut-off percentage with the resolution of 1%.
2. Search for the cut-off percentages that give false negative rates lower than the defined thresholds in both training and validation datasets.
3. Apply the cut-off percentages to the testing dataset.

Because the testing dataset was separated out before the training procedure started, we can use the results obtained from this dataset to directly compare the performance of each model trained from all three data preparation scenarios. The results are shown in Tables 5.18 – 5.23 below:

Table 5.18: Performance of models trained from the tree imputed dataset with 5% FNR threshold

Model	Cut-off	TPR	TNR	FPR	FNR	Accuracy
REGM	0.26	0.9455	0.2314	0.7686	0.0545	**0.5898**
REGMI	0.23	0.9339	0.2078	0.7922	0.0661	0.5723
DT	N/A	N/A	N/A	N/A	N/A	N/A
ANN	0.23	0.9494	0.2000	0.8000	0.0506	0.5762

Table 5.19: Performance of models trained from the conventional imputed dataset with 5% FNR threshold

Model	Cut-off	TPR	TNR	FPR	FNR	Accuracy
REGM	0.28	0.9518	0.4007	0.5993	0.0482	0.7068
REGMI	0.20	0.9518	0.3615	0.6385	0.0482	0.6894
DT	0.17	0.9662	0.3344	0.6656	0.0338	0.6854
ANN	0.01	0.9542	0.6597	0.3403	0.0458	**0.8233**

Table 5.20: Performance of models trained from the listwise dataset with 5% FNR threshold

Model	Cut-off	TPR	TNR	FPR	FNR	Accuracy
REGM	0.27	0.9689	0.1294	0.8706	0.0311	0.5508
REGMI	0.26	0.9728	0.1843	0.8157	0.0272	**0.5801**
DT	N/A	N/A	N/A	N/A	N/A	N/A
ANN	0.14	0.9922	0.0314	0.9686	0.0078	0.5137

Table 5.21: Performance of models trained from the tree imputed dataset with 10% FNR threshold

Model	Cut-off	TPR	TNR	FPR	FNR	Accuracy
REGM	0.32	0.9105	0.3451	0.6549	0.0895	**0.6289**
REGMI	0.31	0.8872	0.3255	0.6745	0.1128	0.6074
DT	0.33	0.8833	0.3176	0.6824	0.1167	0.6016
ANN	0.23	0.9144	0.3333	0.6667	0.0856	0.6250

Table 5.22: Performance of models trained from the conventional imputed dataset with 10% FNR threshold

Model	Cut-off	TPR	TNR	FPR	FNR	Accuracy
REGM	0.38	0.9012	0.5241	0.4759	0.0988	0.7336
REGMI	0.34	0.9012	0.5000	0.5000	0.0988	0.7229
DT	0.33	0.9036	0.5031	0.4969	0.0964	0.7256
ANN	0.79	0.9470	0.7169	0.2831	0.0530	**0.8447**

Table 5.23: Performance of models trained from the listwise dataset with 10% FNR threshold

Model	Cut-off	TPR	TNR	FPR	FNR	Accuracy
REGM	0.36	0.9222	0.3373	0.6627	0.0778	**0.6309**
REGMI	0.36	0.8949	0.3176	0.6824	0.1051	0.6074
DT	0.33	0.7899	0.4431	0.5569	0.2101	0.6172
ANN	0.22	0.9066	0.1255	0.8745	0.0934	0.5176

Tables 5.18 – 5.20 show the performance of the models from the three datasets with the threshold of 5% FNR. The models trained from the conventional imputed dataset perform, on average, better than the

models trained from the other two datasets in terms of predictive accuracy. Some points needed to be considered in choosing which model is best. First, the cut-off criteria should limit the FNR to under the predefined thresholds. Also, it is expected that the performance of the model may change slightly when it is applied to the different datasets. However, the performance variations obtained when applying the models to the three data partitions (training, validation, and testing) are low. Moreover, the performances of the models in the testing dataset should already reflect the real-world performances (leave-one-out method). Second, if a model satisfies all the point mentioned above, then the higher prediction accuracy is preferable. Third, if two or more models satisfy the first point and have the same prediction accuracy, the least complex model should be chosen as determined by the number of variables in the model. In other words, the model should be easy to use because of the fewer variables and parameters.

With the considerations explained, and a 5% FNR threshold, the artificial neural network model with 2-hidden layers and 80 neurons, trained from the conventional imputed dataset fits the predefined criteria the best. At the cutoff 0.01, the model performs with 4.58% FNR, 95.42% TPR, and prediction accuracy of 82.33%. The same criteria are used for selecting the best model for limiting the FNR to 10%. The best model is the same model selected in the previous case with a different cut-off value. At a cutoff of 0.79, the model performs with 5.31% FNR, 94.70% TPR, and prediction accuracy of 84.47%.

Now, we have a model that can confidently screen people without OSA out of the population with a FNR of less than 5% and 10% (more than 95% and 90% TPR respectively). This means that people whom the model classifies as non-OSA can be assured that they do not have OSA (with less than 5% or 10% chance of incorrect rejection). At the same time, we have a high false positive rate, meaning that if the person has some degree of probability of having OSA, the model will sensitively identify that person as having OSA. This is exactly our intention. For people identified as having OSA, we will use the model developed in Chapter VII and the ECG of a subject collected overnight using our wireless-wearable device introduced in Chapter VIII to confirm the existence of OSA and to determine its severity.

CHAPTER VI

OSA EPISODE PREDICTION BASED ON THE PATTERNS OF RECURRENCE QUANTIFICATION ANALYSIS FEATURES

Chapter 5 described the model to predict the existence of OSA disorder in an individual. In standard clinical procedure, a polysomnography (PSG), or a sleep study is required for the sleep physician not only confirm the OSA symptoms but also determine the feasible treatment options for OSA patients. The PSG is basically the recording of a patient's multiple biosignals while sleeping. The number of signals required depends on the diagnostic purpose according to the standard set by the American Academy of Sleep Medicine [126, 127]. In our case, the PSG data obtained from the SHHS study consists of twelve channels of biosignals as follows: oximetry (blood oxygen percentage), heart rate, chest wall and abdomen movement, nasal/oral airflow, body position, electroencephalogram (EEG) (2 central; one for redundancy in case of failure/loss), electrooculogram (EOG) (bilateral), electromyogram – chine (EMG), and electrocardiogram (ECG). Figure 6.1 shows a patient wearing most of the sensors and electrodes during the sleep study. In addition to needing to wear so many sensors and electrodes during PSG, patients also need to sleep in a sleep laboratory or hospital. An attended PSG requires at least one sleep technician on site to constantly monitor the patient. At the end of the sleep study, a certified sleep technician or sleep doctor manually looks at the time-series of the biosignals recorded overnight for evidences of the symptoms of OSA. The evidence for the OSA episode reflected from biosignals can be any or a combination of abrupt

changes in heart rate, lower or no amplitude of respiratory signals (airflow, and abdomen and ribcage respiratory efforts), and lower blood oxygen level. The OSA episode is then confirmed by the arousals that can be noticed by any or a combination of muscle movement (EMG), lighter sleep stage, changes in heart rate, and ventilatory overshoots.

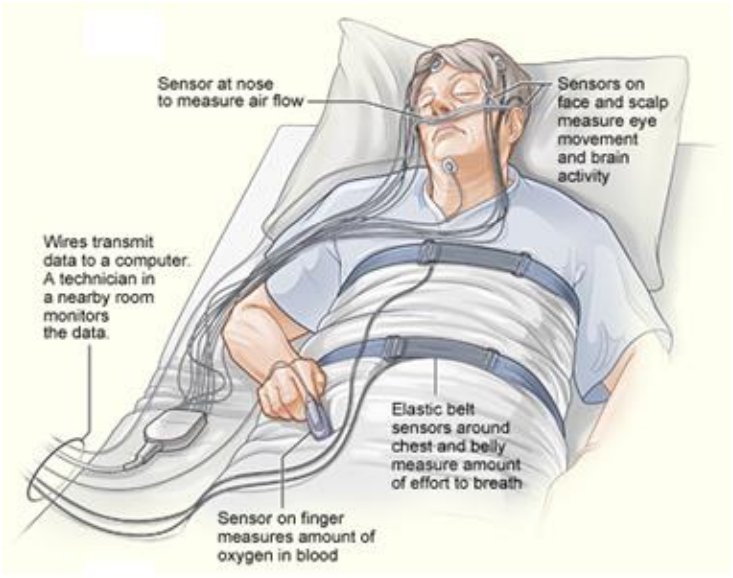


Figure 6.1: Patient wearing sensors and electrodes during a sleep study (reproduced from [1])

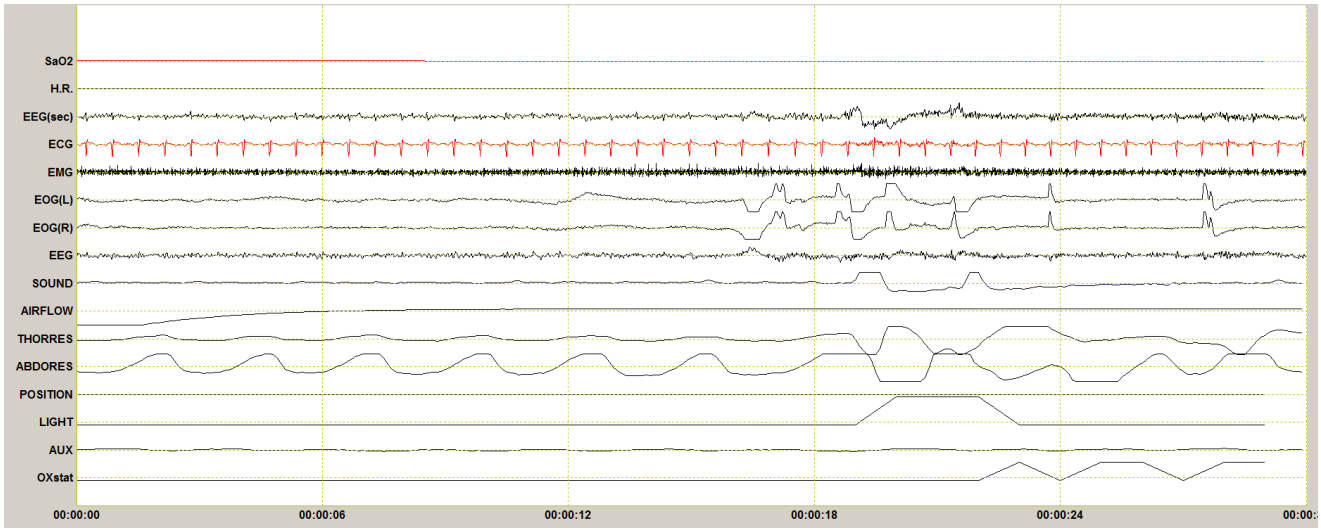


Figure 6.2: Multiple time-series of biosignals from PSG

After manually identifying such evidence, the certified sleep technician or sleep physician calculates the number of sleep apnea episodes per hour, which is called the apnea-hypopnea index (AHI). The AHI number is a sleep apnea severity indicator. The calculation can be done using the following equation:

$$AHI = \frac{(Apnea\ events + Hypopnea\ events) \times 60}{Actual\ sleep\ time\ (mins)} \tag{6.1}$$

The severity of sleep apnea is classified as follows: AHI < 5 - no sleep apnea, 5 ≤ AHI ≤15 – mild sleep apnea, 15 ≤ AHI ≤ 30 – moderate sleep apnea, and AHI > 30 - severe sleep apnea.

As mentioned in Chapter 2, because of the labor-intensive process and the limited number of sleep laboratories, a high number of suspected OSA patients remain undiagnosed. Although a smaller, multisensory PSG device such that the sleep study could be done at a patient's home has been developed, a sleep technician is still required to place the sensors and electrodes on the patient due to the complexity of the sensor and electrode placement. Later in Chapter 8, we report the development of a wireless, wearable device that can be used for a home sleep study. Even though the wearable device was designed with ease of use in mind, when implemented with its full function, a proneness for a bad quality signal still exists because of the multiple sensors and electrodes. For these reasons, in this chapter, we introduce several aspects of digital signal processing, non-linear time-series analysis, and data mining techniques that can be used to transform only one-lead ECG data into other meaningful signals that more correlate to OSA symptoms and then extract meaningful quantifications using a nonlinear time-series analysis technique called a recurrence quantification analysis (RQA). By studying the patterns of these quantifications, the status of the OSA episode can be classified. The AHI is calculated from the status of the OSA episodes overnight. It is used to determine OSA severity, which is important for determining the treatment options.

The organization of this chapter is as follows. First, we explained about the digital signal processing (DSP) methods to obtain the signals that are significant predictors of an OSA episode from a one-lead ECG signal. Then, we describe the non-linear time-series analysis methods that are used for unfolding the nonlinearity and nonstationarity of the signals previously extracted from the ECG signal. Next, the

recurrence quantification analysis (RQA) is portrayed as it is used for extracting the meaningful quantifications that will be used as patterns for the data mining methods for learning to classify an OSA state.

6.1 Extracting OSA predictors from one-lead electrocardiogram (ECG)

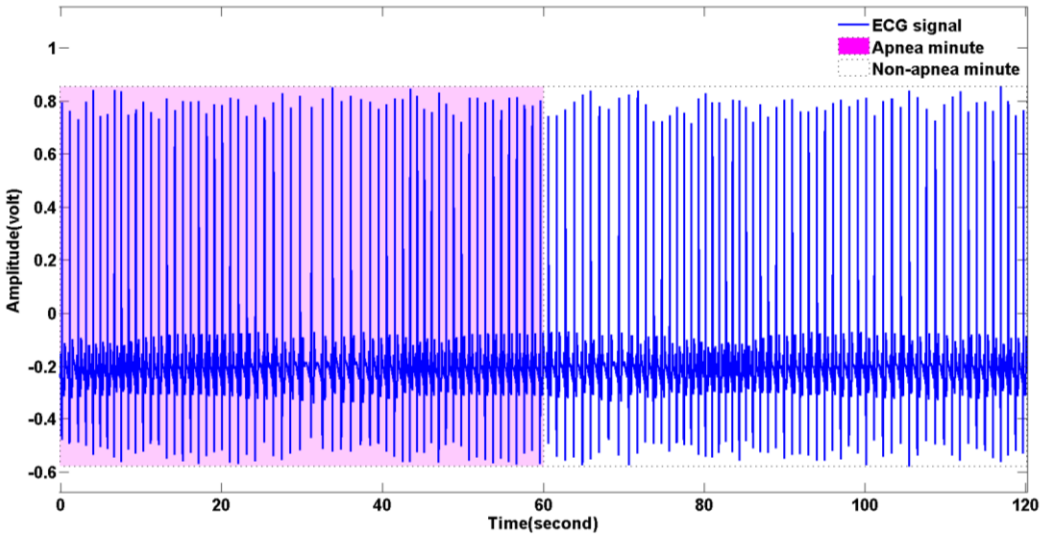


Figure 6.3: ECG plot with classified regions of apnea (shaded) vs non-apnea minute

Although we know from the literature that there is much information embedded in the ECG signal, visually, regarding the classification of OSA episodes, the ECG itself does not give much information. Shown in Figure 6.3 is the ECG signal with a length of 120 seconds or 2 minutes. Visually, there is no significant difference between the first minute, marked by the sleep physician as an OSA episode minute, and the second minute, marked as a non-OSA episode minute. However, with the right pieces of information (features) and using appropriate extraction methods, we can obtain the signal or information that will be more useful or related to the problem. Feature extraction is a process for extracting a new set of features or information from the original feature through some functional mapping [128]. The feature extraction techniques used in this study are explained in a subsection of this section.

6.1.1 ECG and R-R interval time-series

In this section, the basic idea of the electrical signal from the heart is intuitively explained. A knowledge of several markers from the ECG signal, such as P, Q, R, S, and T waves, is required in the digital analysis step to extract the useful information related to the classification of OSA. Furthermore, the methods to derive each signal required in further statistical analysis are also explained within each subsection. The derivation of time-series or signals that are covered in this section are the R-R interval (HRV), ECG-derived respiration, and the slow waves that represent a respiratory sinus arrhythmia (RSA).

6.1.1.1 ECG origin

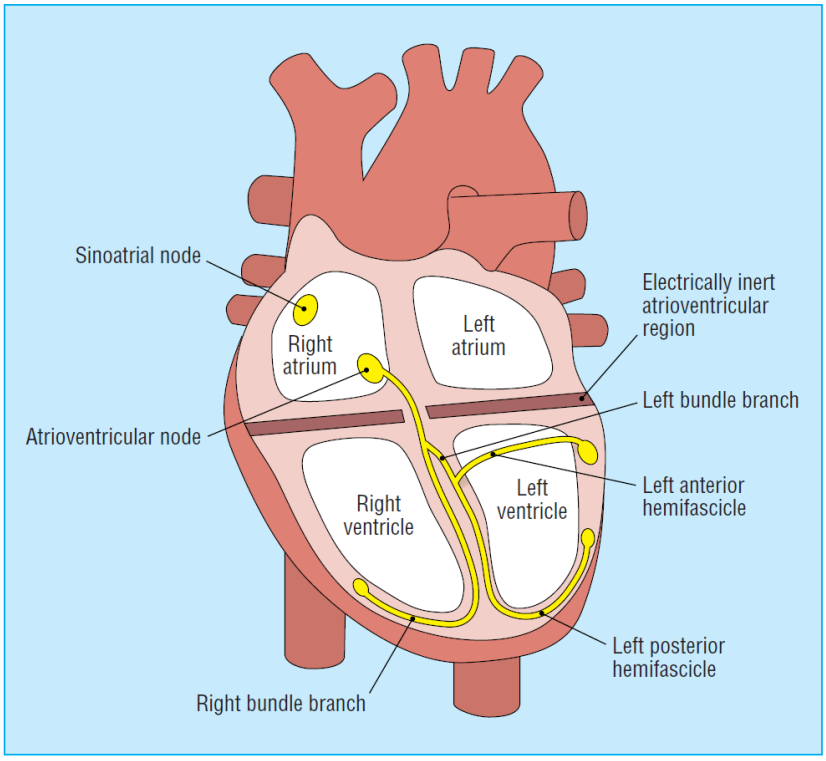


Figure 6.4: Heart chambers and pacemaker points (reproduced from [129])

The heart is a vital organ that pumps oxygenated blood from the lungs throughout the body and also drives the deoxygenated blood from the body back to the lungs. As shown in Figure 6.4, there are four chambers in the heart and each one has its own responsibility. In brief, the deoxygenated blood from a body

enters the right atrium, which acts as a buffer for the deoxygenated blood; then it is pumped into the right ventricle. The right ventricle pumps this deoxygenated blood to the lungs where the blood receives oxygen from the alveoli in exchange for the carbon dioxide. Then, the oxygenated blood comes back to the heart via the left atrium. The left atrium pumps this blood to the left ventricle, where the oxygenated blood is then pumped throughout the body.

The coordination of the pumping process in the heart is controlled by pacemaker cells [130]. The dominant group of pacemaker cells in the heart is called the sinoatrial (SA) node or sinus node, located in the right atrium as shown in Figure 6.4. The SA node initiates atrial depolarization. This depolarization is propagated to an atrioventricular (AV) node located in the lower part of right atrium. From the AV node, the depolarization impulse is propagated to the ventricles via the special conductor tissue connected to the AV node. The pumping processes (contraction and relaxation) result from the depolarization (contraction) and repolarization (relaxation) of the myocardial cells [129]. An electrocardiogram (ECG) is the result of those electrical changes picked up via electrodes placed on the chest and limb surface [129].

6.1.1.2 PQRST Signal

In general, a PQRST signal is typically a result of cardiac contraction and relaxation cycle as explained in the last section. Specifically, starting from the SA node in the right atrium, the depolarization generated here results in the P wave of the ECG. The repolarization period of the atrium (after the depolarization) occurs during the PR interval as shown in Figure 6.5. Next, after the depolarization impulse travels to the AV node, the impulse is transmitted to the ventricles via the AV left and right bundle branches (see Figure 6.4), resulting in the contraction of the ventricles. The electrical signal from the depolarization of the ventricles that can be picked up at the chest surface is the QRS complex in the ECG. Finally, the repolarization of the ventricles results in the T wave.

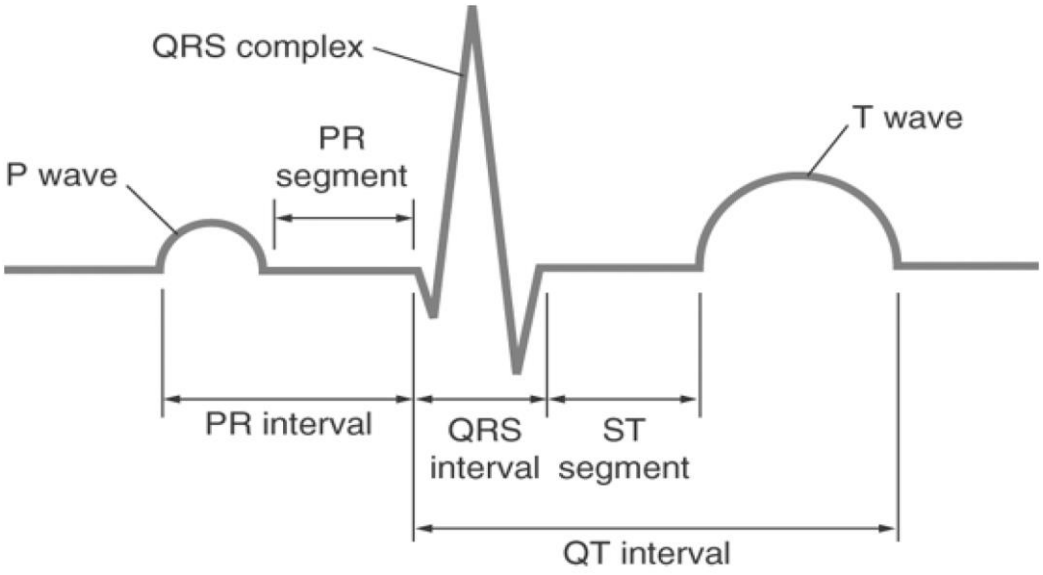


Figure 6.5: PQRST signal in one cycle of a cardiac contraction and relaxation (reproduced from [130])

6.1.1.3 One-lead ECG system

As described in the previous section, the ECG is the recording of the cardiac electrical activity from the chest and limb surfaces. The different locations of the recordings results in the different angles of view for looking at the heart's electrical activities. An ECG lead is composed of two electrodes, positive and negative, called a bipolar lead [131]. Since the heart is a three-dimensional organ, the electrical activities from one point to another are traveling in three dimensions as well. The electrode placement of each lead defines the recording direction from the negative to the positive electrode. As a result, the ECG signal corresponding to the electrode placement is shown as the difference in magnitude between positive and negative electrodes with time in 2-dimensional space.

The one-lead ECG is the non-standard definition used in the ECG handheld recorders or other exercise equipment referred to lead I or II in a standard 12-lead ECG [132]. This one-lead ECG focuses on detecting the R-R interval to obtain a heart rate. With a simple bipolar limb lead which uses only two or three electrodes, the left arm and right arm if using lead I, and the right arm and left leg if using lead II, and one can use either lead I or II to obtain a heart rate easily, because the R wave is dominant in lead I and II.

By tracking the R-R interval, we can see the sinus arrhythmia, which is the variation in the heart rate that occurs during inspiration and expiration, a higher heart rate during inspiration and a lower heart rate during expiration [129]. Moreover, we can also extract the respiratory signal using a technique called ECG-derived respiration (EDR) from the ECG signal collected from a one-lead ECG. To clarify, in our case, the wearable device reported in Chapter VIII is modified to record only a one-lead ECG, lead I.

6.1.1.4 R-R interval time-series (HRV)

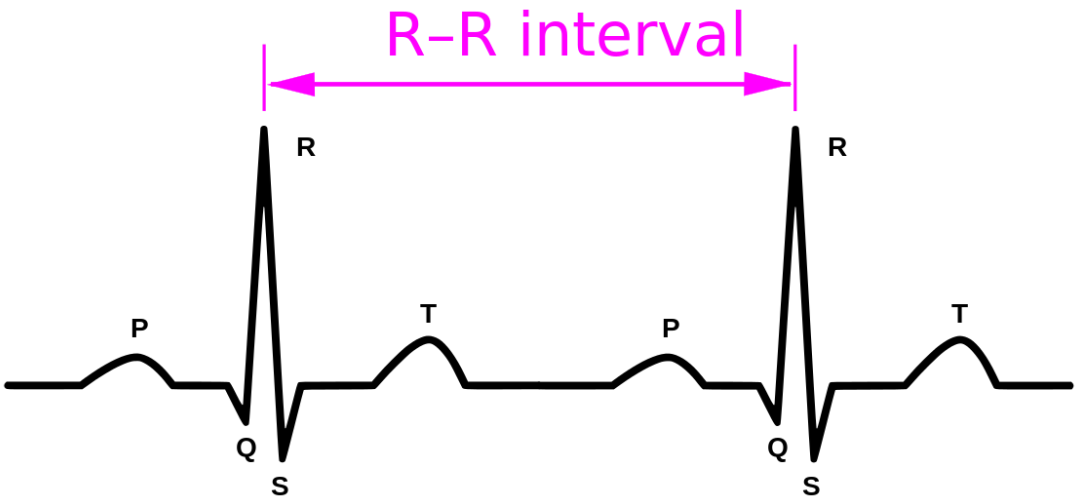


Figure 6.6 R-R interval (reproduced from [133])

The R-R interval is the time between two adjacent R peaks in the ECG signal. The physical meaning of the R-R interval is the beat-to-beat time of the heart's left ventricle. This quantity tells how fast the heart pumps the blood to the body via blood vessels. The heart rate, which is, the more familiar term, is the average of the R-R intervals within one period of time. The popular unit for the heart rate is beats per minute (bpm), which tells how many R-R intervals in a one-minute period.

The heart rate is controlled by the brain with the perturbation from a respiratory sinus arrhythmia (RSA), baroreflex, and thermoregulation [134]. In our study, because the heart rate related to sleep environments such as air pressure and temperature does not significantly change during the night, the main factor that significantly affects the fluctuation in heart rate is the RSA. The RSA is the phenomenon that the heart rate increases during inspiration and decreases during expiration. This normal fluctuation should

be steady in all sleep stages at a steady state. Because of the effects of an OSA episode, it is expected that after the closure of the airway occurs, the blood oxygen level is constantly decreased. This causes the heart to pump more blood through the lungs trying to increase the amount of blood going through the oxygen and carbon dioxide exchange process to compensate for the decreased blood oxygen, which abruptly increases the heart rate. After the blood oxygen decreases below a threshold, the brain shifts the sleep stage back to a lighter stage to gain back control over the muscles around the constricted airway. This behavior is called an arousal. It opens up the airways, which terminates the OSA episode. Sometimes, this arousal process induces ventilatory overshoots, which may cause an instability in the respiratory control system which eventually induces the OSA episode to recur [135, 136]. We believe that these processes explained earlier are captured by the patterns of the R-R interval time-series that change over time. For this reason, the R-R interval time-series is extracted from the ECG signal for further analysis.

The process for obtaining the R-R interval time-series, is explained as follows:

1. Filter the ECG signal with a band pass (0.04 - 150 Hz) to remove the baseline wandering and other artifacts. The baseline wandering is the phenomenon where the ECG signal is highly modulated with the low frequency signal. Figure 6.7 shows a baseline-wandering affected ECG in the solid line and the band pass filtered signal in the dashed line.
2. Normalize the filtered ECG signal to range (-1, 1). Search for the R-peaks where the derivative of the signal is zero (maximum point) with the threshold of the amplitude higher than the T-wave peaks. However, in this step, in many cases, the amplitudes of R-peaks and T-peaks are not different. To solve this problem, we selectively filter the T-wave using a coiflet wavelet decomposition. From our experiment, with the ECG sampled at 100 Hz, to filter the T-wave, the ECG signal is decomposed using the coiflet5 wavelet at level 8. Then remove all levels from levels 5 to 8, which correspond to the T-wave and synthesize the signal back. The result is shown in Figure 6.8. The T-wave amplitude of the wavelet-denoising-ECG signal in the dashed line is substantially lower compared to the original signal, in the solid line. After the T-wave parts are

smaller, we can search for all R-peak amplitudes, and their corresponding locations as shown with the ‘o’.

3. Calculate the distance between the consecutive R-R peaks and convert into time units. In our study, we use seconds as the time unit for the R-R intervals. The time-series is then formed from these distance time points. An example of an R-R interval time-series is shown in Figure 6.9.

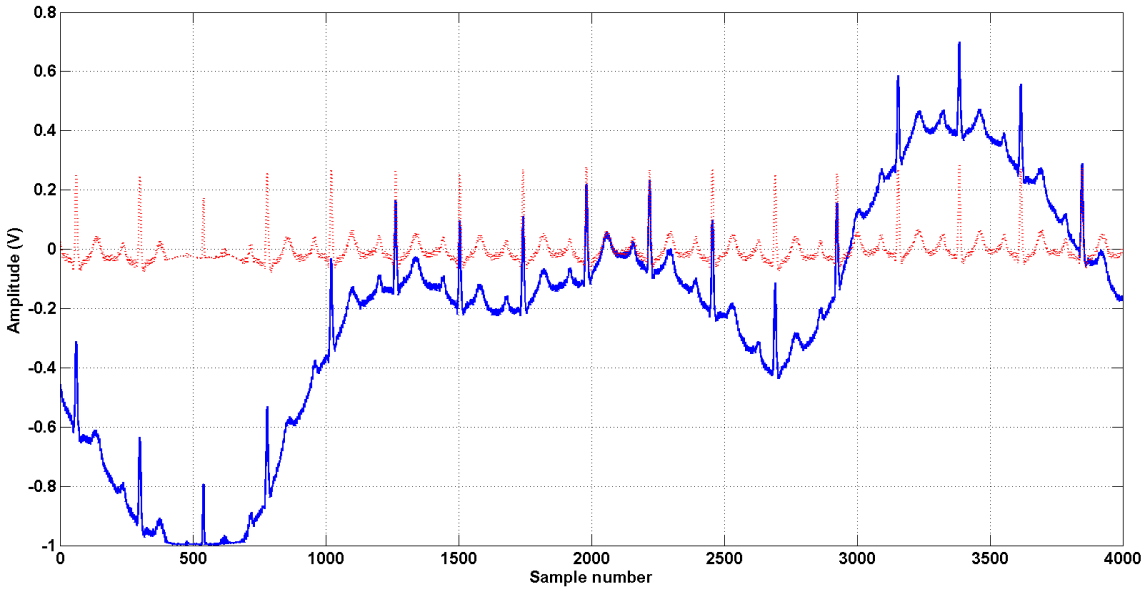


Figure 6.7: ECG plot showing baseline wandering effect in solid plot and band-pass filtered signal in dash line

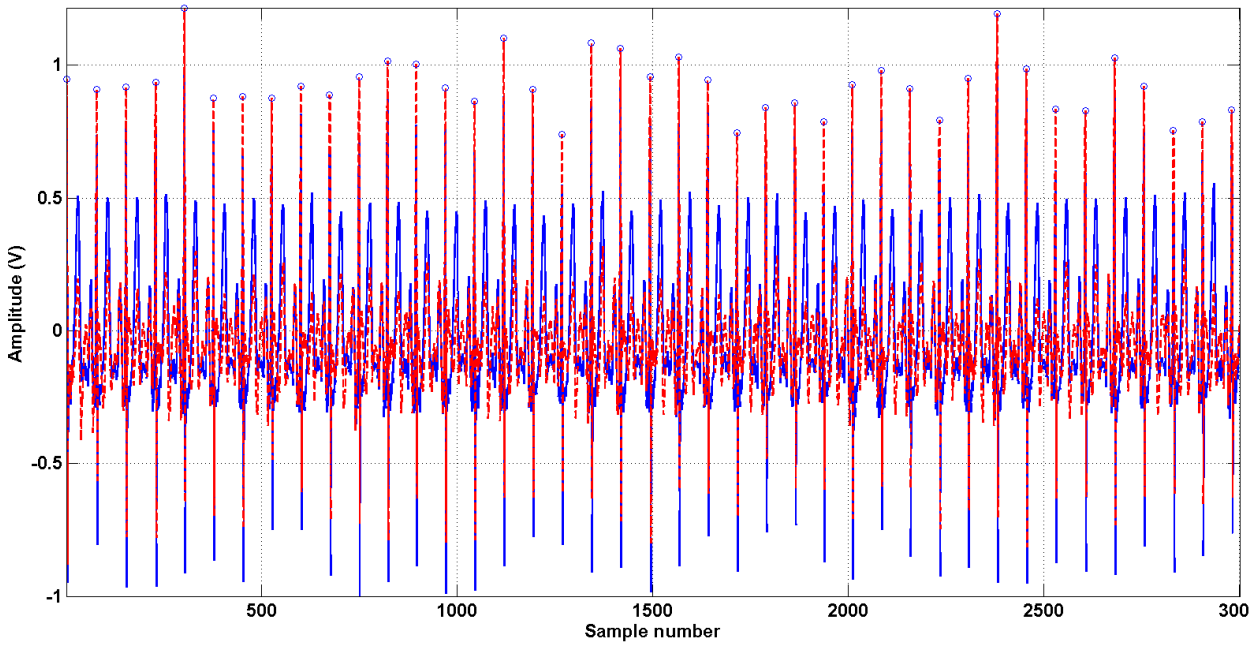


Figure 6.8: The plots of the ECG signal with high amplitude T-waves in solid plot and the ECG with T-wave selectively removed by a denoising technique using a coiflet wavelet

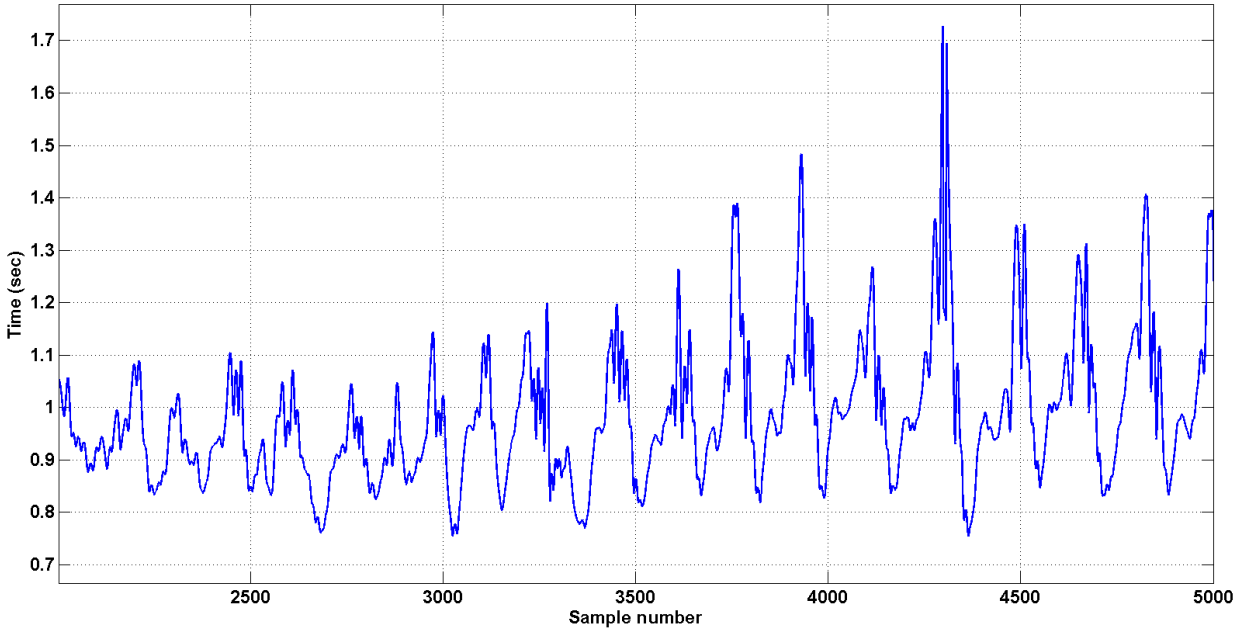


Figure 6.9: An example of an R-R interval time-series

The R-R interval time-series shown in Figure 6.9 has a duration of 500 seconds (the signal was resampled to a sampling rate of 10 Hz). From the mean wandering and non-uniform variation, we can roughly see the nonstationary and nonlinear pattern of the data which hinder the ability to predict or forecast

using some traditional methods such as linear regression. We will discuss how we manage the nonstationary and nonlinearity of these time-series later in this chapter.

6.1.2 ECG derived respiration (EDR)

Some of the most important information for detecting sleep-disordered breathing (SDB) such as OSA is the respiratory signal. Two main types of respiratory signals are used in a sleep study, respiratory effort and nasal respiration. The respiratory effort captures the effort from the lung to inhale or exhale air. This reflects that the control signal from the brain may be absent, as in the central sleep apnea (CSA) which is excluded from this study. The nasal respiration is used as evidence of whether the person stops breathing or not. These two signals are used in PSG to help classify the type of sleep apnea, whether it is OSA or CSA. In our study, which focuses mainly on OSA, the respiratory effort is sufficient to represent the evidence of breathing activity.

To physically capture the respiratory effort, two respiratory effort belts measure the changes in thoracic or abdominal circumference during the respiration periods, inhaling and exhaling, as shown in Figure 6.1. The drawback of this method is that the belt itself is an elastic band. The user must properly adjust the belt such that it fits snugly to the point of measurement, which tends to introduce an uneasy feeling leading to less efficient sleep. The nasal respiratory signal is detected by a nasal thermistor (Figure 6.1). The problem with this sensor type is that the thermistor may fail to detect minor changes in the airflow. Moreover, due to the sensitivity problem, the recorded signal tends to be delayed (lag) compared to other signals (120 to 720 msec) [137]. The air tube that hangs from the nostrils to the ears with some parts of the sensor inside the nose causes irritation and may fall off during the night when the user changes sleep positions. In order to use the most information from the ECG signal so that the additional sensor can be omitted, we derive the respiratory signal from the amplitude variability of the ECG's R-peaks, using a method modified from [138-140]. This method based on the fact that the impedance across the thoracic cavity changes during inspiration and expiration (the chest surface moves closer or further away relative to the heart position) causing the variation in the transthoracic impedance. The impedance is higher during

inspiration (more distance between the heart and chest surface), and lower during expiration (less distance between the heart and chest surface). If we neglect the very small variation in the amplitude of the ECG itself, the amplitude of the ECG's QRS complex, especially the R wave amplitude, changes [138-140] accordingly to the respiratory activity. Specifically, the EDR can be derived from the changes in R-wave amplitude from the suitable lead ECG (normally, leads I and II). We modified the methods explained in [138-140] because their methods do not work with an ECG signal in which the T-wave amplitude is too large compared to the amplitude of the R-wave. Our process to derive the respiratory signal from the single lead ECG is briefly explained as follows:

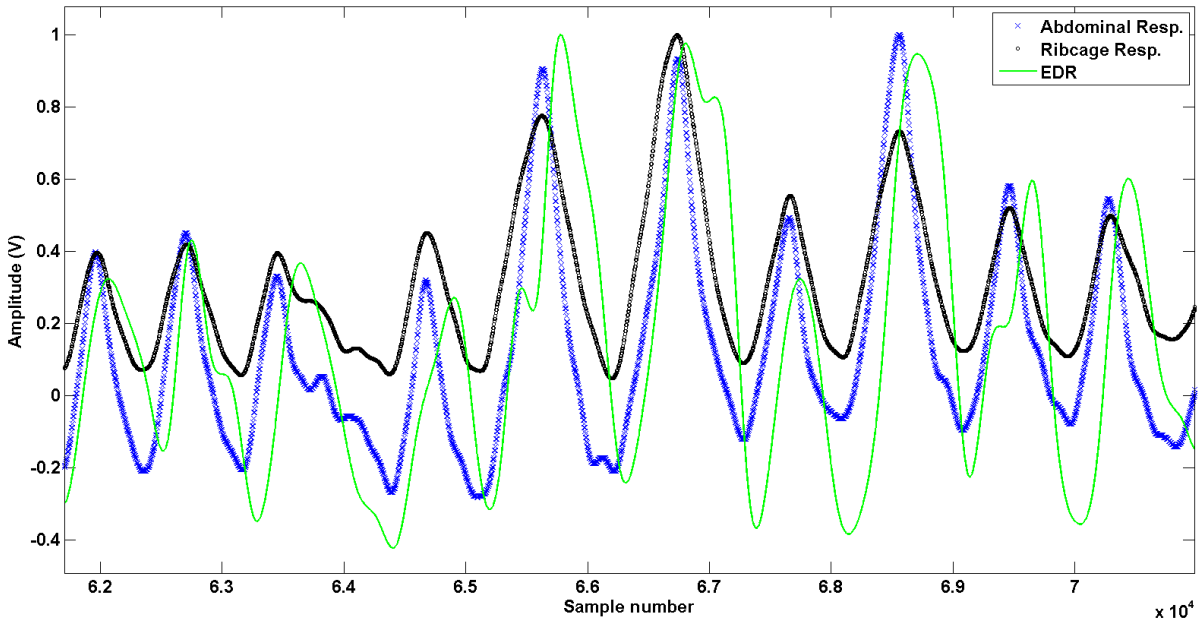


Figure 6.10: The plots of abdominal respiratory in marker 'x', ribcage respiratory in marker 'o', and ECG-derived respiration (EDR) in solid line

1. Use steps 1-2 in deriving the R-R interval time-series section to get the amplitudes and locations of every R-peak from the ECG signal.
2. Interpolate the R-peak amplitudes and locations obtained from the previous step with a cubic spline method. The resulting signal is the ECG-derived respiration (EDR) signal. An example of the result is shown in Figure 6.10.

Figure 6.10 shows the plots of the respiratory efforts measured from respiratory bands at abdominal (x) and ribcage (o) positions and the EDR. Although the EDR seems to be a little slower (lag) than the other two signals, when we quantify the features in the resolution of one minute, this minor difference is not significant in the analysis. Also, the respiratory rate calculated from the zero crossing rate of both signals is not statistically different. More detail is provided in the analysis sections.

6.1.3 Slow wave signal derived from wavelet decomposed ECG (SWS)

In the previous section, we derived an EDR signal that very well resembles the real respiratory signal (in terms of waveform and zero crossing rate). It is used as a representative to reflect the state of the respiratory system. As mentioned in the background section, there is much information embedded in the heart rate variability (HRV). One more piece of information that is as important as the respiratory activity is the coupling of the cardiovascular and the respiratory systems. This coupling can be captured by the behavior of the respiratory sinus arrhythmia (RSA), which is mostly embedded as a high frequency band (HF, 0.15-0.4 Hz) of the HRV or R-R interval time-series [141, 142]. In this section, we explain how we extract this signal from the R-R interval time-series using wavelet decomposition.

Wavelet decomposition is a modified short-time Fourier transform that represents the decomposed signals in both time and frequency domains through a time windowing function or a mother wavelet function [143]. Traditionally, the Fourier transform is used for analyzing the signal in the frequency domain. However, when transforming a nonlinear time-series that contains short transients in the time domain to the frequency domain, the result is damped and long-duration vibrations [141]. In other words, they cannot be detected in the frequency domain after the Fourier transformation. In contrast to the Fourier transform, which assumes the signal to be stationary, wavelet analysis does not have such a limitation so it works well with a nonstationary time-series. This time-frequency localization advantage is a well-known characteristic of a wavelet decomposition.

The wavelet transformation process comprises two main phases, analysis or decomposition and synthesis or reconstruction phases. If certain conditions are met, the signal can be perfectly reconstructed

from the coefficients obtained from the analysis or decomposition phase. For this reason, wavelet decomposition is very popular in signal denoising applications. The user can selectively delete the decomposed coefficients corresponding to the noise and reconstruct the denoised signal. However, in our research, the main idea is the opposite. Using an ECG signal as a base signal, we selectively choose the coefficients that most correlate to the occurrence of sleep apnea events, delete other unrelated coefficients, and reconstruct the signal.

The basis of the mother wavelet or the type of windowing technique used depends on the window length. Since this study does not focus on how to develop a wavelet function, we will briefly explain how the wavelet decomposition works intuitively. The main method for wavelet decomposition in this study is the multiresolution analysis (MRA) developed by Mallat in 1989 [144]. In general discrete wavelet transformation (DWT), the signal is passed through a series of high pass filters and low pass filters as shown in Figure 6.11:

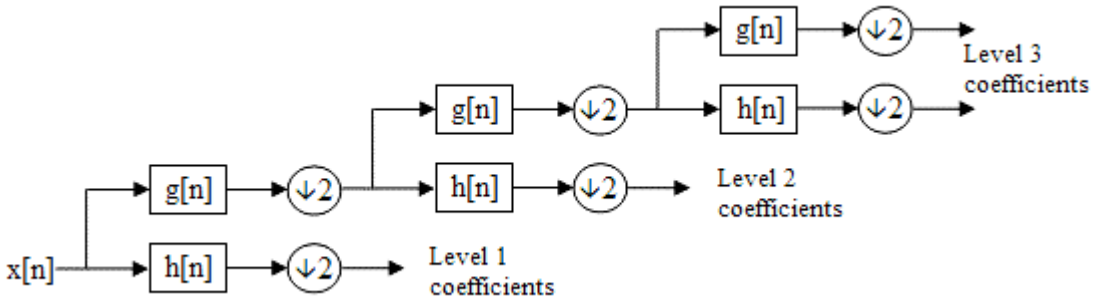


Figure 6.11: Discrete wavelet transform using multiresolution analysis (MRA) with 3 level filter banks (reproduced from [2])

The DWT procedure starts with feeding the time-series x[n] to the half band low pass filter with an impulse response g[n] and a high pass filter with an impulse response h[n]. In mathematical expression, the filtering process is the convolution of the signal with the impulse response of the filter:

$$x[n] * h[n] = \sum_{k=-\infty}^{\infty} x[k] \cdot h[n-k] \tag{6.2}$$

Regarding the Nyquist theory, after passing the signal through either a half band low pass filter or a half band high pass filter, half of the samples can be eliminated. This is denoted by the symbol ↓2 in

Figure 6.11. The result of the first high pass filter is level 1 detail coefficients. Likewise, the result of the first low pass filter is level 1 approximation coefficients. To perform further analysis, the level 1 approximation coefficients are used as a signal to be passed through another set of half band low pass and high pass filters. In theory, the decomposition can be done for n levels. However, in practice, the analysis levels depend on the number of samples in the original signal. It should be noted that because the decomposition process involves downsampling by a factor of two, the number of samples required in wavelet analysis must be to the power of two.

In the synthesis phase, to be able to perfectly reconstruct the signal back from the wavelet coefficients in every decomposed level, the pair of low pass and high pass filters must form orthonormal bases. To satisfy that constraint, the relationship between them is [145]:

$$h[L-1-n] = (-1)^n \cdot g[n] \tag{6.3}$$

Where h[n] is the impulse response of a high pass filter

g[n] is the impulse response of a low pass filter

L is the filter length in number of samples

When a filter pair that satisfies equation 6.3 is used, the reconstruction process is exactly the reverse of the analysis process. The coefficients at every level are upsampling by a factor of two, then passed through the synthesis filter pairs. The relationship between the analysis and synthesis filters is that they are identical to each other but time reversed. There are many choices of low pass and high pass filter pairs in wavelet analysis. In our case, we choose to use a coiflet wavelet family because of its nearly linear phase filter property [146]. This results in minimal phase distortion in the outcome signal, which is desirable in most digital signal analysis applications. Moreover, the coiflets have a maximum number of vanishing moments meaning that complex functions can be represented with a sparser set of wavelet coefficients. For further information, reference [145] provides very good information on the theory and application of wavelet decomposition.

To derive the slow wave signal using wavelet analysis, we looked for the combination of coefficients in the higher levels which corresponds to the low frequency signals. At the beginning of the

analysis, the frequency band that we looked for from the extracted signal was between 0.1 – 0.6 Hz which corresponds to the dominant frequency of the human respiratory signal [142]. In particular, after we narrowed down the analysis levels to the frequency band mentioned above (approximately level 7 and above), we tried several combinations of coefficients that give the synthesis signal with the frequency response closest to the HF component. We found that the synthesized wavelet signal from the ECG that is significantly correlated with the state of an OSA episode is a result of the coefficients obtained from levels 8 and 9 of DWT decompositions using the coiflet5 wavelet. The impulse responses of the analysis and synthesis filter pairs for the coiflet5 wavelet are shown in Figure 6.12.

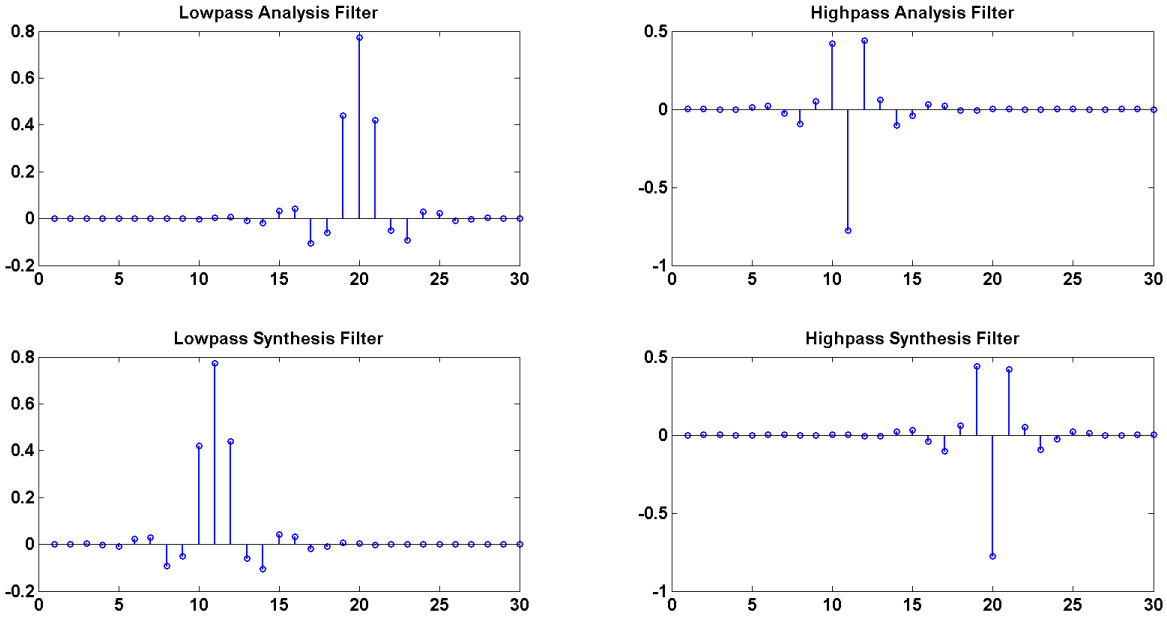


Figure 6.12: Impulse responses of the analysis and synthesis filter pairs for coiflet5 wavelet

The process to derive the slow wave signal is summarized as follows:

1. Filter the ECG signal with a band pass (0.04 - 150 Hz) to remove the baseline wandering and other artifacts. The baseline wandering is the phenomenon where the ECG signal is highly modulated with the low frequency signal.
2. Decompose the ECG signal from step 1 into 9 levels using the coiflet5 wavelet. To clarify, the ECG signal is passed through the analysis filter pairs (see Figure 6.12) and downsampled by a factor of two 9 times (see Figure 6.11).

3. Reconstruct the signal with the coefficients from levels 8 and 9 by setting the coefficients in every decomposed level (except levels 8 and 9) to zero. Exactly reverse what was done in step 2, but use synthesis filter pairs (see Figure 6.12) and upsample by a factor of two.

6.2 Unfolding nonlinearity of the signals using phase space reconstruction analysis

In section 6.1, we explained how three important signals, R-R intervals, EDR, and slow waves, were extracted from the one-lead ECG signal. However, as commonly known, nonlinear or irregular behaviors embedded in the data always impede the predictability in both time-series and pattern recognition applications. This problem happens in many aspects. One of them is that most of the time, we can only observe, at most, one of the process activities, not all the influences that compose to the process. If we assume that all influences composing to the process are independent and we can directly observe their activities, most likely, the process can be approximately represented by a linear mathematical model which is built by a linear combination of the effects from all influences. However, most of the time, we can only observe, at most, one effect which may not directly relate to the problem (i.e., ECG and the OSA state). Still, the observed data is most likely contain a lot of information including what is directly related to the problem because of its nonlinearity (i.e., coupling and interaction between the effects).

Introduced by Takens in 1981 [147], the higher dimensional phase (state) space reconstruction can be used to rebuild the topology of all the states represented in the system. This theorem allows us to map the observed data back to estimated numbers of variables that compose to the dynamics of the system (i.e., the observed one dimensional ECG signal to a three-dimensional ECG signal). Then, we can use all the unfolded information to infer the solution to our original problem (i.e., classification of the OSA episode state) using statistical modeling methods explained in Chapter 4.

The organization of this section is as follows: First, the state space representation (roughly interchangeable with the phase space) is introduced. Then, we depict the method to derive the important parameters, embedding dimension and time delay, which are required for reconstructing our time-series to their suitable state space dimension. Finally, we explain the method to quantify the data patterns in the

higher state space in order to better distinguish the data patterns between apnea and non-apnea states. The idea of unfolding the nonlinearity of the signals using phase space reconstruction in this section is partially from the works of our research group [148-151].

6.2.1 State space representation

Anything that evolves over time can be thought of as a dynamic system. Normally, the variables used in explaining an instantaneous dynamic system are called state variables. The set of all possible values of state variables is called the state space. Another analogy used to refer to the continuous and finite-dimensional state space is called the phase space [152-154]. To create a mathematical dynamic system, two parts needed to be considered:

1) The part of the system that evolves over time or a state vector. The state vectors give a complete description of the system at any particular time. Thus, in d-dimensional space or phase space at time t, we can describe the state of a system by a state vector $\vec{y}(t)$ formed by its state variables [147, 155]:

$$\vec{y}(t) = y^1(t), y^2(t), \dots, y^D(t) \tag{6.4}$$

The superscript is used for referring to the index of the state vectors.

2) A function or rule that describes how the state vector evolves over time. Basically, when all state vectors with all collected state variables are represented in successively for enough time, the movement of the state will reveal a phase space trajectory or orbit. This movement behavior explains the dynamics of the system. The movement of the state vector $\vec{y}(t)$ can be represented by its velocity vector, $\dot{y}(t)$ [155]:

$$\dot{y}(t) = \partial_t \vec{y}(t) \tag{6.5}$$

With these two quantities, the dynamical system is actually a model that explains the temporal evolution of the system [152-154]. For example, considering an object moving in 3-dimensional space, the set of variables that can completely explain the system includes 3 positions in each dimensional space (P_x, P_y, P_z in a Cartesian coordinate system) and 3 velocities in each dimensional space (V_x, V_y, V_z) over time. If their velocity vector is known, the state of the system at any time can be determined.

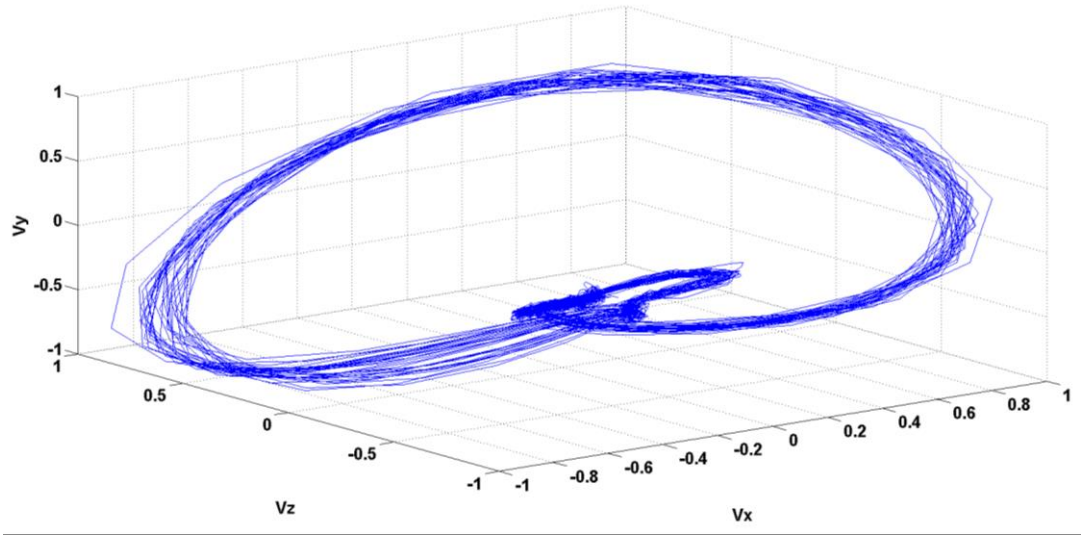


Figure 6.13: State space representation of vectorcardiogram (VCG)

Figure 6.13 shows a state space representation of a vectorcardiogram (VCG). This representation is obtained by using a 3-lead vectorcardiogram to capture the heart electrical activity. It clearly shows the path in which the heart electrical activity travels around the heart in 3-dimensional space. Not only do we gain advantages in term of the ability to visually inspect the system's time evolution but also the ability to statistically track the evolution in order to predict and forecast the state of the system. This comes back to our original problem statement that in most cases, we cannot collect all the state variables in the state space which are required to represent the dynamic system. However, we can use only one collected state variable to reconstruct the topology of the state space representation. The related theories and methods are described in the next section.

6.2.2 Phase space reconstruction

In 1981, Floris Taken introduced the method to reconstruct the phase space or state variables of dynamic systems with one observable using a time delay method [147]:

$$\vec{y}_i = x_i, x_{i+\tau}, x_{i+2\tau}, \dots, x_{i+(m-1)\tau} \tag{6.6}$$

Where $\vec{y}_i$ is the reconstructed phase space vector

x_i is the observable at time step i

τ is the time delay (in time step)

m is the embedding dimension

It should be noted that because the observable is usually collected in equally-spaced, discrete time-series, the index in the equation above refers to the time step index not the actual time. Intuitively, the main idea is to look for the time delay at which the specific number of state variables (m dimensions) becomes the most independent. Doing so requires that the observable have a nonlinear relationship or coupling with the other state variables in the dynamic system [155-157]. Also, equation 6.6 below guarantees that the topological structures of the original phase space will be reserved if [147]:

$$m \geq 2d + 1 \tag{6.7}$$

Where m is the embedding dimension

d is the real dimension of the original phase space

It should be noted that [158] and [159] suggested that a value of m that satisfies $m > d$ can be sufficient. This idea has been extended by several researchers but the topic is still open for research because of some limitations. For example, autocorrelation is one of the methods that can be used to determine a suitable time delay but only if the underlying dynamics of the system is linear.

Before this equation can be effectively implemented to reconstruct the phase space of any system, two parameters need to be estimated: embedding dimension, m, and time delay, τ. After thoroughly reviewing several studies in the domain [147, 152, 155-157, 159-161], the suitable methods for our application are mutual information [162, 163] and false nearest neighbors [161, 164].

6.2.2.1 Mutual information

The mutual information procedure requires that the time delay be determined first because it is a parameter needed to correctly determine the embedding dimension. The mutual information is derived from

information theory which, given that x has been observed, tells us how much information is gained about y. In addition to linear dependency, it also measures the generalized correlation, which is sensitive to any relationship, including nonlinear relationships [163]. The formula below is used to compute the mutual information of the time-series in order to determine a suitable time lag for reconstruction of the system's dynamic topology [162, 163, 165]:

$$I(\tau) = -\sum_{x,y} p_{x,y}(\tau)\log\frac{p_{x,y}(\tau)}{p_x p_y} \tag{6.8}$$

Where $I(\tau)$ is the mutual information of the original time-series, x_i, with its time lag version for τ time step time-series, y_i.

x_i is the observable at time step i

τ is the time delay (in time step)

m is the embedding dimension

The average information gained in equation 6.8 is based on the entropy calculation, which measures the disorder in the system. A straightforward way to calculate the mutual information is as follows: The time-series is partitioned into equal size bins used to construct a histogram, then the probability of the particular point in the time-series falling into a particular bin of the histogram is calculated. Then, the entropies of 1) the binned version of the original time-series, (x), 2) the binned version of the τ time lag version of the original time-series, y, and 3) the joint distribution of x and y, are calculated. Finally, the mutual information is the summation of the x entropy and y entropy subtracted from the entropy of the joint probability distribution of x and y. Although the entropy is sensitive to the number of bins used, the mutual information is not heavily affected by it [163]. However, to be on the safe side, because using too small a bin number could decrease the sensitivity and too large a bin number could cause a loss of generality, we use the Freedman-Diaconis rule [166] to select the optimal size of the bins to be used in the histogram. This rule is related to the interquartile range and number of data points of the time-series regardless of the distribution of the data. In our application, the mutual information of the original time-series with its varying time lag is calculated. We vary the time lag (τ) from 1 to 100 and calculate the mutual information in every

time lag. The ideal case would be that there is a time lag that provides zero mutual information meaning that the two variables are independent. However, in real applications, several time lags could yield approximately the same minimum mutual information. It has been suggested that the first local minimum of the mutual information is the suitable choice [155, 156]. An example of the plot of the mutual information against the time lag of the heart rate variability is shown in Figure 6.14.

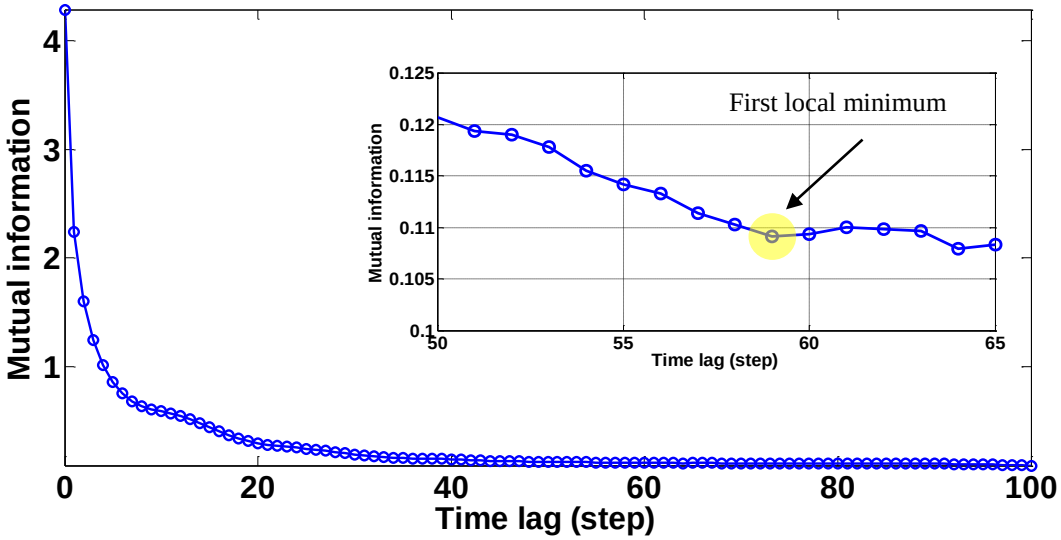


Figure 6.14: The plot of mutual information vs. time lag (step) of the subject a01's heart rate variability

Figure 6.14 shows the typical behavior of the plot of the mutual information of the heart rate variability (subject a01) against a one-hundred-time lag (τ). As mentioned above, we look for the first local minimum of the mutual information, which is at time lag 59 with the mutual information of 0.1092 in this case. This time lag indicates the time that makes the additional state variables the least dependent and will be used in the next procedure to find the minimal embedding dimension (m).

6.2.2.2 False nearest neighbors (FNNs)

The main idea for the false nearest neighbor method is that in the phase space representation, the states or points that are close to each other are considered to be true neighbors. In particular, in the same evolution, the states of the closest neighbors to the focused state should not be too different [159, 161]. Now, if we take one state variable off the phase space, from d to (d-1) state variables, and represent the same data in the phase space, the projection of the points in the (d-1) dimension will never be the same as

in the original dimension and the points that are closest to each other in this representation may not be true neighbors because of the lack of one axis in the state space (folding). The points that are not true neighbors in the phase space representation but are neighbors in the lower dimensions are false neighbors [159, 161]. As we may imagine, if we further take more state variables off the state space representation, the number of false nearest neighbors will significantly increase.

Using this idea in reverse, when we reconstruct the phase space from only one time-series by increasing the number of embedding dimensions by one and examine whether the closest neighbors in the lower dimension are still the neighbors in the higher dimension, the minimum embedding dimension required to unfold the attractors is the dimension that gives the first minimum false nearest neighbors [159, 161]. We use equations 6.9 and 6.10 below from [161] and [159] to implement the FNNs idea. The square of the Euclidian distance between point $x(n)$ and its neighbor is [161]:

$$R_d^2(n,r) = \sum_{k=0}^{d-1} [x(n+k\tau) - x^{(r)}(n+k\tau)]^2 \tag{6.9}$$

The criterion for the neighbor point to be a false neighbor is [161]:

$$\left[\frac{R_{d+1}^2(n,r) - R_d^2(n,r)}{R_d^2(n,r)}\right]^{1/2} = \frac{|x(n+\tau d) - x^{(r)}(n+\tau d)|}{R_d(n,r)} > R_{Threshold} \tag{6.10}$$

Where $R_d(n,r)$ is the Euclidian distance between point $x(n)$ and its r[th] neighbor in dimension d

d is the starting dimension or state variable

n is the index number of a point in time-series x

τ is the time delay (in time step)

As seen in equations 6.9 and 6.10, the process for calculating the FNNs starts from choosing a number of nearest neighbors (r), the starting dimension (d), and the distance threshold ($R_{Threshold}$) to be used in the equations. It is suggested from [161] that only one neighbor (r=1) is sufficient when considering all the points in the dataset. Kennel et al. also suggested that the $R_{Threshold} \geq 10$ is enough to clearly identify the false neighbors. In our case, the starting dimension is d=1. Then, after all required parameters

are determined, starting at d=1, the Euclidian distance between each point and its neighbor is calculated. Then, the phase space is reconstructed with the dimension d+1 and the Euclidian distance between the point and its previously assigned neighbor. If the ratio between the increase in distance from dimension d to d+1 to the distance in the d dimension is larger than the threshold ($R_{Threshold} = 10$ in this case), we flag that index as a false neighbor. The percentage between false and true neighbors in this d dimension is calculated. Next, we move to the next starting dimension, which is d=2, and repeat the same process. We repeat the procedure until the number of d dimensions is very high (10 dimensions were used in our case). The percentages of false neighbors in every embedded dimension are then plotted against the corresponding dimension. The minimal embedding dimension required to unfold the phase space is the dimension that gives the first minimum in the plot (zero false neighbors or close to zero). An example of the results from our heart rate variability data is shown in Figure 6.15 below:

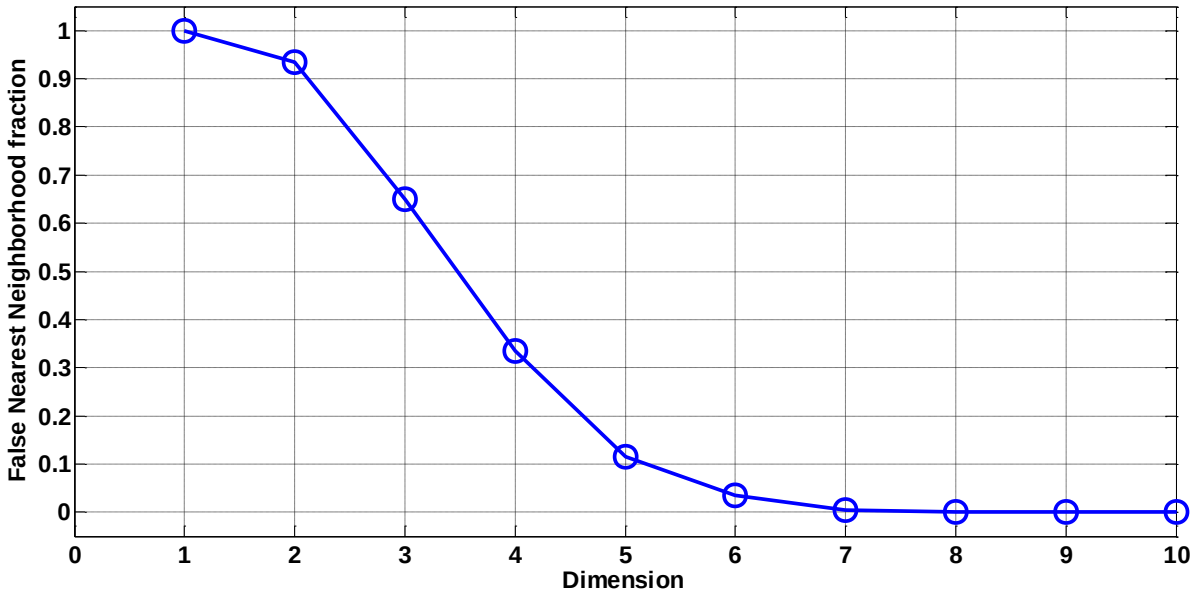


Figure 6.15: The plot of false nearest neighborhood fraction vs. dimension of the subject a01's heart rate variability

As seen from the plot above, the FNNs percentage declines with the increasing number of embedding dimensions and approaches to zero after the 7th dimension. It should be noted that sometimes, the algorithm overestimates the FNNs percentage, resulting in too many dimensions used for reconstructing a phase space. Also, depending on the noise level in the data, in our case, a very low FNN rate (≤0.05) is

acceptable in the phase space reconstructing process. We discuss the limitations when using too many and too few dimensions in the process in the results section.

6.2.2.3 Phase space reconstruction on simulation data

In order to fully demonstrate the effectiveness of the reported algorithms, we choose to reconstruct the phase space of the Lorenz system [167], which is one of the most well-known studies on chaos dynamics. Lorenz system is famous for the simplicity of the three coupling nonlinear differential equations that display chaos behavior. It was originally developed by Ed Lorenz in 1963 to be a simplified version of the equations used for describing the atmospheric convection system. It basically depicts the movement of two different air temperatures in the earth's atmosphere (hot air rises then cool air sinks) that creates a convection that never crosses the same path. Lorenz is able to simplify an original 12 state variable system [168] to the equations below:

$$\begin{aligned} \dot{x} &= \sigma(y - x) \\ \dot{y} &= -xz + rx - y \\ \dot{z} &= xy - bz \end{aligned} \tag{6.11}$$

There are three parameters in the system: the Prandtl number (σ), the Rayleigh number (r), and the aspect ratio of the convection cylinders (b) (for more information please see [167, 169]). The original value of these parameters set by Lorenz is σ=10, r=28, and b=8/3. We simulated this equation for 5,000 points with 0.01 distance in each time step. The plot of these simulated equations is shown in Figure 6.16.

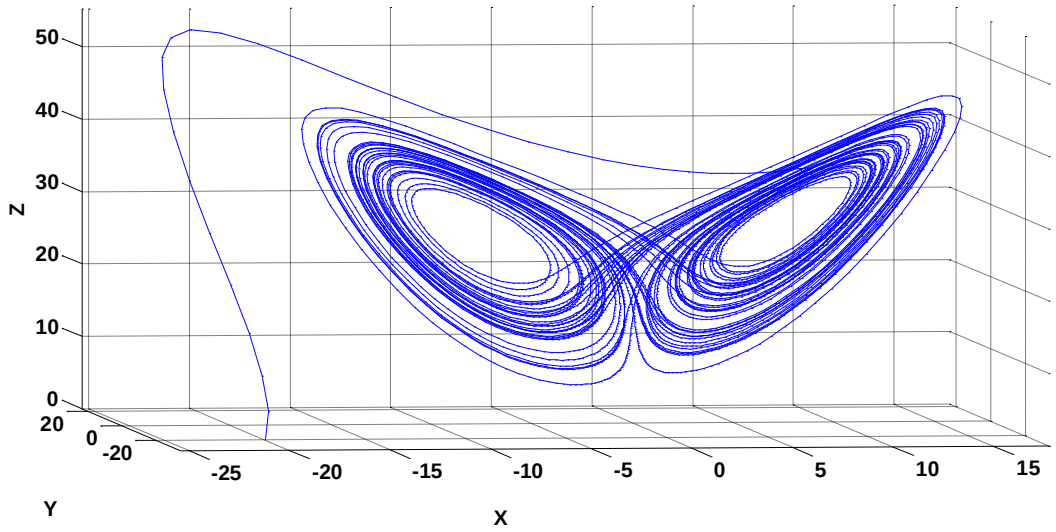


Figure 6.16: The Lorenz attractor plot in 3 dimensional space with parameters σ=10, r=28, and b=8/3

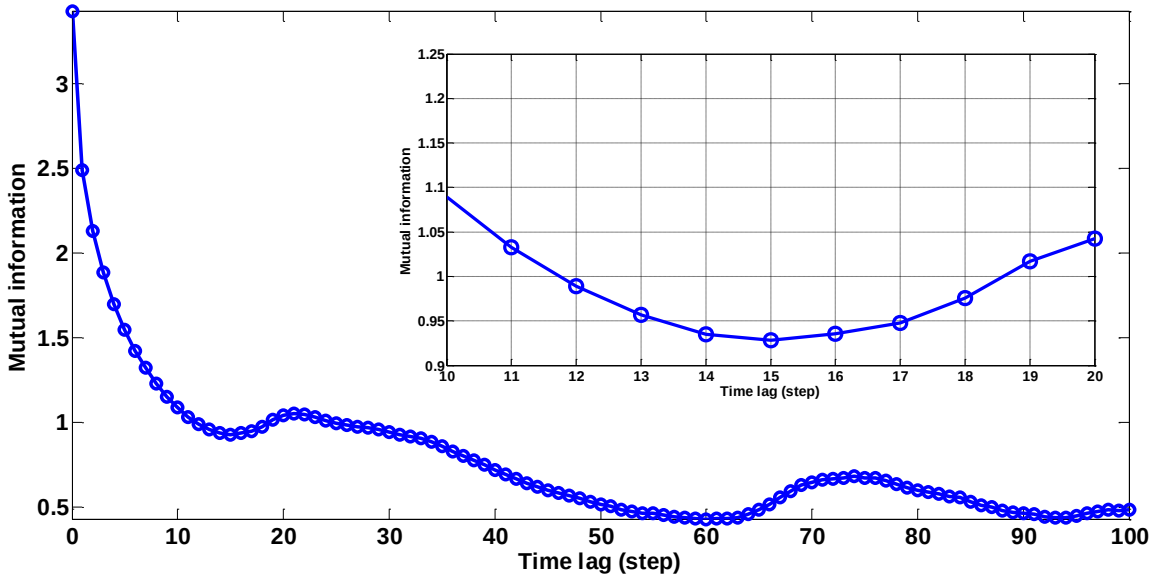


Figure 6.17: The plot of mutual information vs. time lag (step) of the simulated Lorenz system

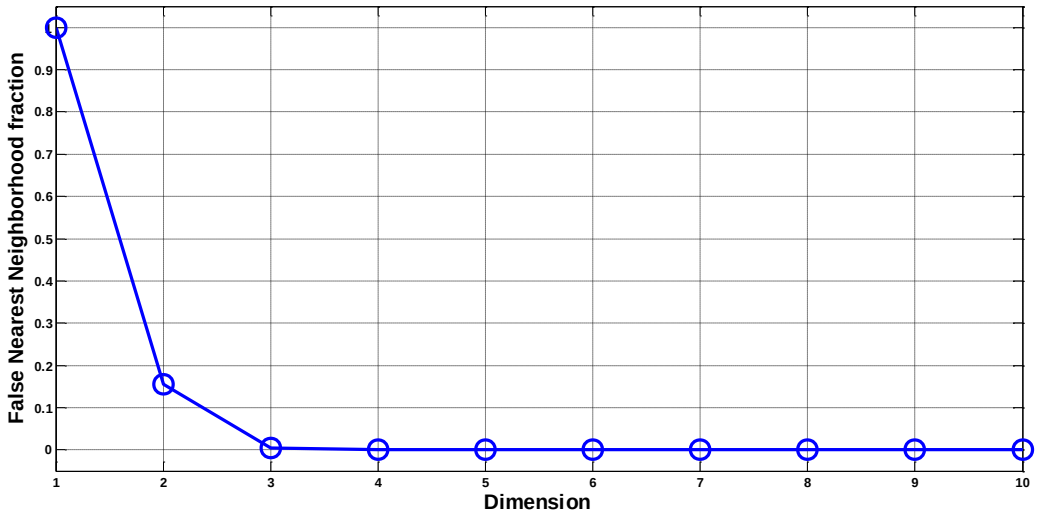


Figure 6.18: The plot of FNNs vs. dimension of the simulated Lorenz system

We reconstruct the phase space of the Lorenz attractor shown in Figure 6.16 from the data from only one time-series generated from one of the equations in equation 6.11. We choose the data point from the $\dot{y}$ or dy equation. The first step is to determine the time delay by using mutual information. Suggested from the mutual information plot in Figure 6.17, a suitable time delay determined from the first local minimum is at time lag 15.

Using the time lag of 15, the calculated FNN percentage of the corresponding embedding dimension is shown in Figure 6.18. As seen from the plot, the FNN approaches to zero at dimension 3 (FNN<0.5%). This result suggests the use of embedding dimension 3, which is the same as the true number of the true state variables. Finally, using equation 6.6, a time-series dy, time delay (τ) = 15, and embedding dimension (m) = 3, the reconstructed Lorenz attractor phase space is shown below:

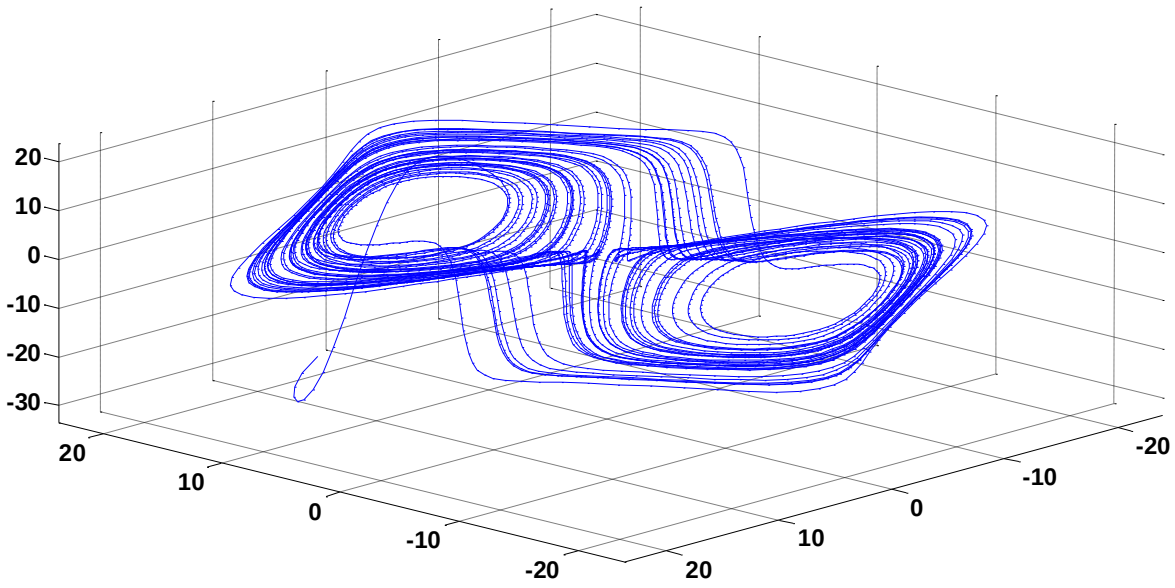


Figure 6.19: The reconstructed Lorenz attractor from only $\dot{y}$ or dy time-series in equation 6.11

The reconstructed phase space of the Lorenz system shown in Figure 6.19 above preserves the same dynamic structure or topology of the original phase space as seen from the same number of attractors and the patterns of the dynamics that evolve over time. We will use this method to reconstruct the phase space from the three time-series data that we extracted from the ECG signal to better observe the different neighboring patterns which can be used to differentiate the state of a sleep apnea episode. The implementation details and statistical results are described in the next chapter.

6.3 Recurrence plot (RP)

Although the reconstructed phase space can unfold the dynamic structure of the focused system, we still need to quantify the different patterns of the states in the phase space so that we can use them in our statistical modeling methods in order to differentiate between the apnea and non-apnea states. Moreover, one of the limitations in phase space reconstruction is that in higher dimensions (>3), we lose the visualization ability which is crucial for tracking the evolution of the dynamics. Realizing this problem, Eckmann et al. [170] introduced a method to statistically project a multidimensional system into a two-dimensional representation such that we can distinguish many state patterns regardless of the number of

state variables. By considering every point in the phase space and its neighbors, we can choose to represent only the states that are close together or within distance threshold ε_i using the following equation [156]:

$$\boldsymbol{R}_{i,j}^{m,\varepsilon_i} = \Theta(\varepsilon_i - \|\boldsymbol{x_i} - \boldsymbol{x_j}\|, \boldsymbol{x_i} \in \mathbb{R}^m, i, j = 1,2, \dots, N \quad (6.12)$$

Where $\boldsymbol{R}$ is the two dimensional squared matrix

$\boldsymbol{x}$ is the state vector in the phase space

m is the embedding dimension used for constructing the phase space

N is the number of states

ε is the distance threshold between state i and j

Equation 6.12 results in a two-dimensional square matrix showing all the states that appear within the distance of ε. In particular, when considering state $\boldsymbol{x_i}$, all the states $\boldsymbol{x_j}$ that occur within a close distance of ε are considered to be neighbors in the phase space. Using the Heaviside function, the close states (neighbors) are then denoted by one, otherwise zero (black dot and white dot on the plot respectively). After considering all states in the phase space, we can plot this information using time or the consecutive state number (which is actually the time with the equally-spaced distance in the sampling period) and this plot is called a recurrence plot. An example of recurrence plots using the information from the Lorenz system is shown in Figures 6.20, and 6.21.

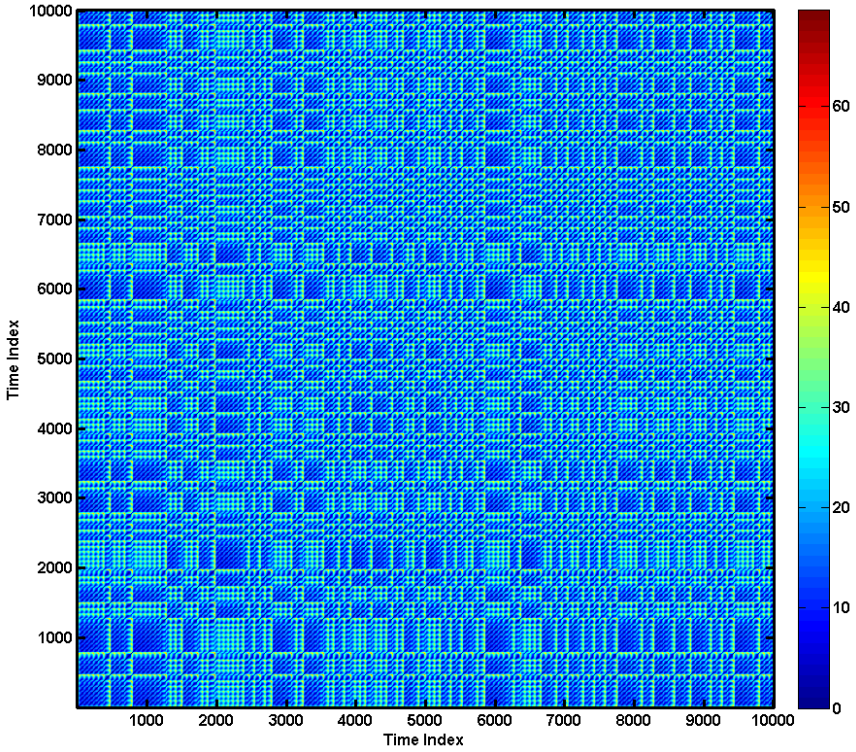


Figure 6.20: The unthresholded recurrence plot of Lorenz attractor using information from Figure 6.15

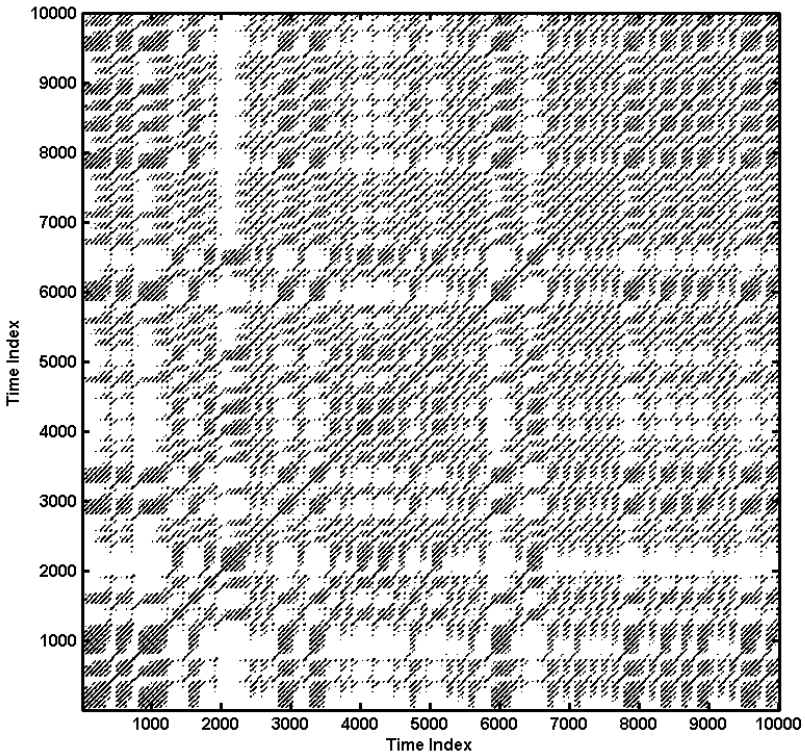


Figure 6.21: The threshold recurrence plot of Lorenz attractor using information from Figure 6.18

In the recurrence plot constructing process, all the Euclidian distances between states $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ where $i, j = 1,2, \dots, N$ have to be calculated. The recurrence plot using these distances as color-coded points (without any threshold) are called a non-threshold or global recurrence plot (see Figure 6.20). A cool color tone (cyan to dark blue) represents a closer distance between the two states and a hot color tone (cyan to red) represents a further distance between the two states. It should be noted that without thresholding the distance between the two states, the recurrence patterns can be visually spotted from this global recurrence plot very well. Now, after applying equation 6.12 to the distance between all the states in the phase space, we can plot all the states that are close together as a black dot, shown in Figure 6.21. However, we need to consider choosing a value for the threshold before the recurrence plot can be built. Many studies [159, 160, 165, 171, 172], provide recommendations for selecting such a threshold: 1) few to ten percent of the maximum phase space diameter (distance between the farthest two states in the phase space), 2) 20-40 percent of the signal's standard deviation, and 3) choose the threshold such that the recurrence rate (discussed in the next section) is fixed at 10%. The main idea is that if the threshold value is too large, all states in the phase space will be neighbors, so we cannot distinguish the patterns in the recurrence plot, and if the threshold value is too small, there will be too few points showing up on the recurrence plot so the recurrence patterns will be hard to spot. However, no underlying dynamics are alike. Several experiments may be required until the results fit the user's requirements.

6.4 Recurrence quantification analysis (RQA)

The recurrence plot was developed mainly for visual analysis of the recurrence patterns of the dynamic system. To avoid subjective justification from users and to enable statistical analysis of the discovered patterns, Zbilut and Webber [173, 174] originally developed statistical measures to quantify the recurrence patterns especially on the threshold recurrence plot (features 1-5). Later, several researchers further extended the measures from the recurrence plot based on the development of the complex network concept on reconstructed phase space (features 6-12) [155, 156, 175-178]. The statistical measures (features) used in this research are as follows:

1) Recurrence rate or percent recurrence (REC or RR). This measurement is simply the density of the recurrence points (black dots in the recurrence plot) in the specified threshold (ε).

$$RR = \frac{1}{N^2}\sum_{i,j=1}^{N} \boldsymbol{R}_{i,j}^{m,\varepsilon} \tag{6.13}$$

2) Determinism (DET) is a percentage of the points in the diagonal lines (excluding where i=j) and total the recurrence points within the threshold (ε). This feature reflects the determinism structure of the system. In the recurrence plot, long diagonal lines are found in periodic systems, whereas short diagonal lines are found in chaotic systems and there will be no diagonal line in a stochastic or random system. The DET can be calculated by:

$$DET = \frac{100(number\ of\ points\ in\ diagonal\ lines)}{(number\ of\ recurrence\ points)} \tag{6.14}$$

Webber and Marwan [165] suggested the calculation method by first grouping all the diagonal lines that have the same length into a histogram:

$$H_D(l) = \sum_{i,j=1}^{N} (1 - R_{i-1,j-1})(1 - R_{i+l,j+l}) \prod_{k=0}^{l-1} R_{i+k,j+k} \tag{6.15}$$

Then, we can calculate the DET by [165]:

$$DET = \frac{\sum_{l=d_{min}}^{N} lH_D(l)}{\sum_{i,j=1}^{N} R_{i,j}} \tag{6.16}$$

In equation 6.16, d_{min}specifies the minimum shortest length for the diagonal lines. It is suggested by [165] that $d_{min} = 2$ for typical systems.

3) Average diagonal line length.

$$\langle D \rangle = \frac{\sum_{l=d_{min}}^{N} lH_D(l)}{\sum_{l=d_{min}}^{N} H_D(l)} \tag{6.17}$$

In the recurrence plot, this measurement can be interpreted as the average time that two segments of the trajectory are close to each other, so it can be used for predicting the occurrence of the next segment of the same trajectory.

4) Maximal diagonal line length (D_{max}) is the longest length of the diagonal line in the considered recurrence plot. This measurement has an inverse correlation with the highest positive Lyapunov exponent [157]. Thus, the system will be less stable (chaotic) if it has a short D_{max} (higher possibility that the trajectories will diverge).

$$D_{max} = \arg\max H_D(l) \tag{6.18}$$

5) Shannon information entropy (ENT) of all diagonal line lengths. The entropy measures the signal complexity from the probability of various diagonal line lengths, which are binned into a histogram. Equation 6.15 is also used to calculate this measurement. :

$$ENT = -\sum_{l=d_{min}}^{N} p(l) \ln p(l) \tag{6.19}$$

$$where\ p(l) = \frac{H_D(l)}{\sum_{l=d_{min}}^{N} H_D(l)}$$

Entropy has a unit of bit/bin. Very low entropy indicates low complexity in the system, which happens when most or all of the diagonal lines are the same length (periodic system) or there are no diagonal lines (a stochastic or random system).

Feature 6.8 are based on the structure of the vertical lines in the recurrence plot (as opposed to the diagonal lines in the previous features). With the same procedure for calculating the diagonal features, all the vertical lines with the same length in the recurrence plot are grouped into the histogram [165]:

$$H_V(l) = \sum_{i,j=1}^{N} (1 - R_{i,j-1})(1 - R_{i,j+l}) \prod_{k=0}^{l-1} R_{i,j+k} \tag{6.20}$$

6) Laminarity (LAM) is a percentage of the points in the vertical lines which reflects the laminar states within the system (the state does not change or changes very slowly).

$$LAM = \frac{\sum_{l=v_{min}}^{N} lH_V(l)}{\sum_{i,j=1}^{N} R_{i,j}} \tag{6.21}$$

The same concept as with the diagonal features is applied in the vertical features. The minimum vertical line length (v_{min}) must be chosen. Marwan and Webber [165] suggested a typical value of $v_{min} = 2$.

7) Trapping time (TT) is the average vertical line length in the recurrence plot. As the vertical line structure of the recurrence reflects the time before one state changes to another state, the average vertical line length or TT is the average time that the system stays in one state (trapping).

$$TT = \frac{\sum_{l=v_{min}}^{N} lH_V(l)}{\sum_{l=v_{min}}^{N} H_V^{\varepsilon}(l)} \tag{6.22}$$

8) Maximal vertical line length (V_{max}) is the longest vertical line that appears in the recurrence plot. Although it is still not clear how the V_{max} influences the system, it can tell what state the system stays in the longest in the recurrence plot.

$$V_{max} = \arg\max H_V(l) \tag{6.23}$$

9) Mean of the recurrence times of the first type ($\bar{T}_1$) [179] is the time average between the consecutive states in the state space with neighbor threshold ε. This time is actually the length of the vertical line in the recurrence plot from one state to the next consecutive state within the considered threshold ε. First, consider the reference point (state) $x_0$within threshold ε.

$$B_\varepsilon(x_0) = \{x: \|x - x_0\| \leq \varepsilon\} \tag{6.24}$$

From the phase space trajectory of length N, denote the consecutive states that belong to $B_\varepsilon(x_0)$.

$$S = \{x_{t_1}, x_{t_2}, \dots, x_{t_i}\} \tag{6.25}$$

Then, consider all the time differences in all consecutive states of the trajectory in S.

$$T_1(k) = \{t_{k+1} - t_k\}, k = 1,2, \dots, (i-1) \tag{6.26}$$

Finally, the mean of the recurrence times of the first type ($\bar{T}_1$) is an average of all time differences.

$$\bar{T}_1 = \frac{1}{(i-1)} \sum_{k=1}^{(i-1)} T_1(k) \tag{6.27}$$

10) Mean of the recurrence times of the second type ($\bar{T}_2$) [179]. When the considered neighborhood, ε, is big or dense, there might be several revolutions of states in the same trajectory that pass through the considered neighborhood, ε, $(S_1, S_2, \dots, S_p)$. After removing the sojourn points (the points in each revolution except the first point that entered the neighborhood), the remaining points are called the recurrence points of the second type (see Figure 6.22).

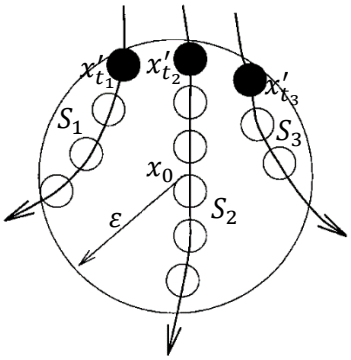


Figure 6.22: Recurrence points of the second type (solid circles) and the sojourn points (open circles) in $B_\varepsilon(x_0)$ (reproduced from [179])

$$S' = \{x'_{t_1}, x'_{t_2}, \dots, x'_{t_i}\} \tag{6.28}$$

Then, we calculate the average time between the consecutive evolutions of the recurrence points of the second type in S'.

$$T_2(k) = \{t'_{k+1} - t'_k\}, k = 1,2, \dots, (i-1) \tag{6.29}$$

Finally, the mean of the recurrence times of the second type ($\bar{T}_2$) is an average of all time differences without the sojourn points.

$$\bar{T}_2 = \frac{1}{(i-1)} \sum_{k=1}^{(i-1)} T_2(k) \tag{6.30}$$

11) Recurrence time entropy (RTE) is an analogous entropy measurement of the diagonal line lengths in feature 5 but referring to the entropy of the vertical line instead. This measurement quantifies the complexity of the vertical line lengths in the RP and the range of the recurrences [176].

$$RTE = -\frac{1}{\ln V_{max}} \sum_{l=1}^{V_{max}} H_V(l) \ln H_V(l) \tag{6.31}$$

6.5 Power spectrum density (PSD)

The power spectrum density is the method for explaining how the variance of the data or time-series is distributed in the frequency domain. To be specific, it quantifies the variance of the data at all frequencies (for more information please see [180] and [181]). In our application, this quantification can be used for classification of the sleep apnea states (apnea vs. non-apnea). The two main assumptions are the following: 1) the variation of the heart rate variability is high shortly after the occurrence of an OSA episode because of the burst in the heart rate. This happens as the brain tries to force the heart to supply more oxygen to the body (the concentration of oxygen is lower during the apneic state) to compensate for the decreased oxygen in the blood. And 2) the variation of the respiratory signal is also high shortly after the termination of the OSA episode, as the body tries to get the blood oxygenated as fast as possible so that the amplitude of the respiratory signal will be higher with a slower breathing rate. There are several ways to calculate the PSD, such as by calculating the summation of the covariance function of the signal, using a band-pass filter to filter only each frequency, then calculating the magnitude of the variation in each frequency range, and so on. The method used in this study is a short-time Fourier transform, which is a discrete Fourier transform with a windowing technique. First, the segmented time-series is transformed to the Fourier domain using a discrete-time Fourier transform:

$$X(k) = \sum_{n=0}^{N-1} x[n] e^{-j\frac{2\pi}{N}kn} \tag{6.32}$$

Where k is the number of sinusoid cycles per N sample

N is the number of data samples

The equation 6.32 transforms any sequence of N numbers in a time domain (time-series $x[n]$) into $X(k)$, which is a combination of the sinusoidal components (amplitude and phase) of function $x[n]$. The amplitude and phase are:

$$|X(k)| = \frac{\sqrt{Re(X(k))^2 + Im(X(k))^2}}{N} \tag{6.33}$$

$$\arg(X(k)) = atan2(Im(X(k), Re(X(k))) \tag{6.34}$$

Because the Fourier transformation assumes signal $x[n]$ to be periodic and uses the combination of different amplitudes and phases of sinusoidal signals as fundamental signals in representing $x[n]$ in the frequency domain, the error introduced when this technique is applied to the non-periodic signal is high between the discontinuity between the last sample and the repeated first sample. This type of error is known as a spectral leakage, where the actual narrow energy band was represented as a wide energy range. One effective remedy is to use a window technique to convert signal $x[n]$ to fit the periodic requirement, amplitude zero at the beginning, rising to the maximum, and decaying to zero again in the end (for more information, please see [180]). The windowing technique used in this study is a Hanning window known as one of the versatile windowing techniques that gives good frequency resolution and amplitude accuracy, and less spectral leakage [180]:

$$w(n) = 0.5\left[1 - cos\left(2\pi\frac{n}{N}\right)\right], 0 \le n \le N \tag{6.35}$$

$$v(n) = x[n] \cdot w(n) \tag{6.36}$$

Where $w(n)$ is the Hanning window function with the length of $L = N + 1$.

N is the number of data samples to be transformed, and

$v(n)$ is the windowed signal

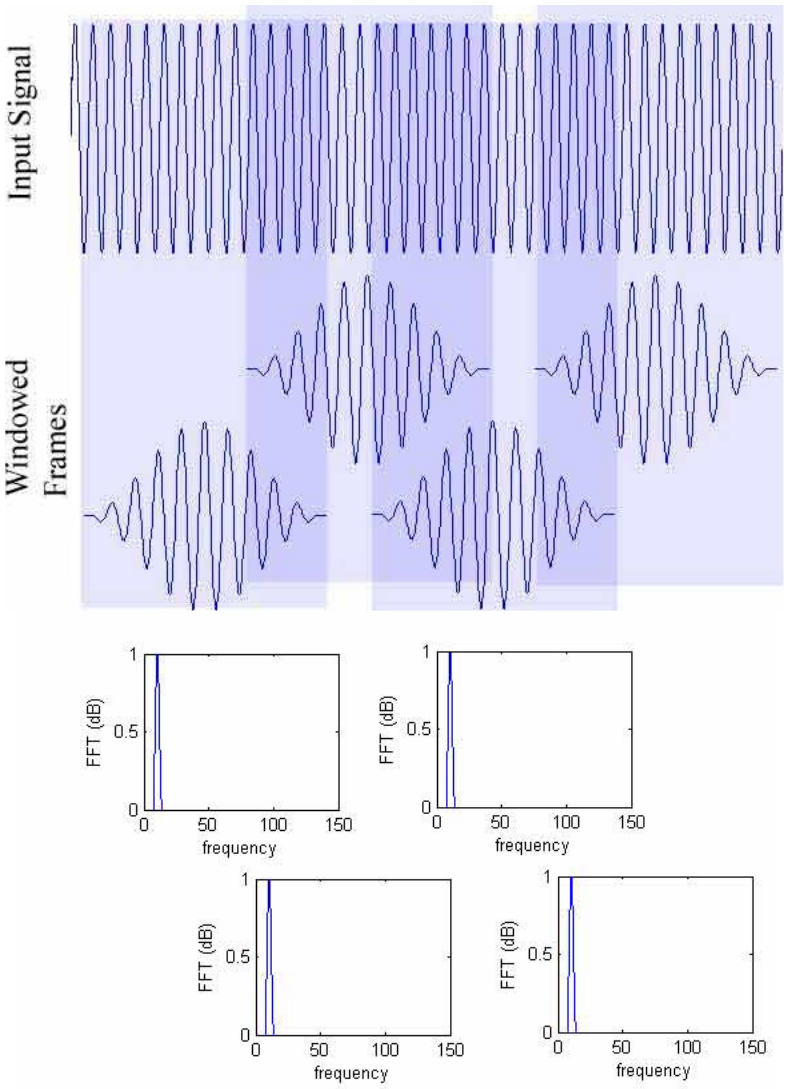


Figure 6.23: Overlap and window processing before DFT process for better frequency resolution (reproduced from [182])

This windowing method is often used with the overlapping technique to gain better resolution of the transformation to compensate for the loss of data at the beginning and the end of the actual signal. An example of this application is shown in Figure 6.23. The overlapping parts are where the amplitudes were attenuated by the window effect. So far, to recapitulate the process, in order to calculate the power spectrum of random signal $x[n]$, we first apply the window $w(n)$ to each segmented part of $x[n]$. The length of the window and the overlap part depends on the user's need. If the larger window size (many data points) is used, the corresponding DFT will have good resolution in the frequency domain but poor localization in

the time domain. Vice versa, if the smaller window size (fewer data points) is used, the corresponding DFT will have poor resolution in the frequency domain but good localization in the time domain. Finally, the power spectral density (PSD) can be calculated by:

$$PSD(k) = \frac{1}{N}|V(k)| \tag{6.37}$$

Where $V(k)$ is the DFT of $v(n)$

$PSD(k)$ is the power spectral density (V^2/Hz)

To conclude this chapter, we explained how three important time-series can be extracted from only a one-lead ECG: the heart rate variability (HRV), the ECG respiratory derived (EDR), and the slow wave signal derived from the wavelet decomposed ECG. Then, we reconstructed the dynamic topologies from each of the time-series, in order to unfold the evolutions of the state's trajectory. This is done based on nonlinear dynamics and chaos theories by embedding the points in the time-series to the higher dimension. Finally, recurrence quantification analysis is used to quantify the neighbor patterns of the state in the recurrence plot such that we can use the data mining methods explained in Chapter 4 to explore the significant patterns that can be used to classify the OSA state. The implementation process and results are thoroughly explained in the next chapter.

CHAPTER VII

OSA EPISODE CLASSIFICATION AND FORECASTING MODELS BASED ON RECURRENCE QUANTIFICATION ANALYSIS (RQA)

In Chapter 6, we explained all methods to process the one-lead ECG time-series to the quantifications that can be further used in any statistical modeling method. In this chapter, we will implement the modeling methods explained in Chapter 4 to the quantification outcomes explained in Chapter 6. The objective are 1) to build OSA episode classification models and 2) to build OSA episode forecasting (look ahead to the future) models. The goal is to improve the OSA diagnosis process by reducing the number of sensors required to only a one-lead ECG signal and reduce the labor of the registered sleep technician and sleep doctor in manually marking the OSA episode by the automatic OSA episode classification model. The ideas and contributions of this chapter are partially from the works of our research group [148-151]. The organization of this chapter is as follows. First, the description of the database is clarified. The criteria for choosing the data records used in the analysis are also explained. Next, we will explain the data preprocessing procedure to be done before the analysis can be carried out. Then, the analysis results are reported and discussed. Finally, we conclude this chapter with the findings from the analysis and the future work that may further improve this application.

7.1 CINC data Polysomnography database information

The data used for the analysis in this chapter is made available by Dr. Thomas Penzel of Philipps-University, Marburg, Germany [63]. It is available for a public download at www.physionet.org in the section that has a challenge from Physionet and computers in cardiology 2000. The database consists of 35 recordings of a single lead ECG signal with a sampling rate of 100 Hz and the non-overlapping-minute-by-minute sleep apnea episode condition annotations (apnea and non-apnea states). These annotations were made by the human experts from a standard procedure to indicate the absence or presence of apnea. All apnea events are obstructive and mixed. The hypopnea events are considered apnea events. The criteria for the hypopnea are defined as a combination of a drop in respiratory flow below 50%, a drop in oxygen saturation of at least 4%, and hyperventilation at the apnea termination. Pure central sleep apnea (CSA) and Cheyne-Stokes respiration are excluded in this analysis. Each record contains data with a length of six to ten hours. There are three groups of subjects: apnea, borderline, and control groups. The criteria for differentiating among the groups is the apnea-hypopnea index (AHI). The subjects with AHI equal to or greater than 10 are classified in the apnea group (group a), between 5 and 10 are classified in the borderline group (group b), and less than 5 are classified in the control group (group c). The subjects are men and women between ages 27 and 63. Their weight and BMI are between 53 and 135 kg, and 20.3 and 42.1 respectively. The AHI ranges from 0 to 93.5. Some records are excluded from the analysis due to the inferior quality of the signals (undetectable R-peaks, loss of ECG signal in some areas, and so on). Finally, 24 subjects are included in the analysis. We will discuss the preprocessing procedure in the next section.

7.2 Data preprocessing and feature extraction

In this section, we explain the workflow of the process to construct the OSA episode classification and forecasting models from the overnight one-lead ECG. The detailed explanation of every algorithm used is in Chapter 6. The overall procedure beginning with the data preprocessing and analysis is shown in Figure 7.1. Each individual's ECG time-series was prepared separately. The ECG signal was visually inspected to see if there is any degradation or loss in any part of the signal. If the quality was bad (most part of the ECG

is too noisy or the R-peaks are absent), we excluded the entire ECG signal from the analysis is because we compare the actual AHI and the calculated AHI from the model, for which we need the entire ECG signal as the primary signal. To extract the ECG-derived respiration (EDR) and heart rate variability (HRV) signals from the ECG signal, we detected all R-peaks from the ECG signal (see Chapter 6 for all algorithm details). For the slow wave signal (SWS), wavelet analysis and synthesis with the thresholding process were performed. These processes were carried out on every chosen ECG signal, so for each individual, we have the overnight three time-series (EDR, HRV, and SWS) needed for further analysis.

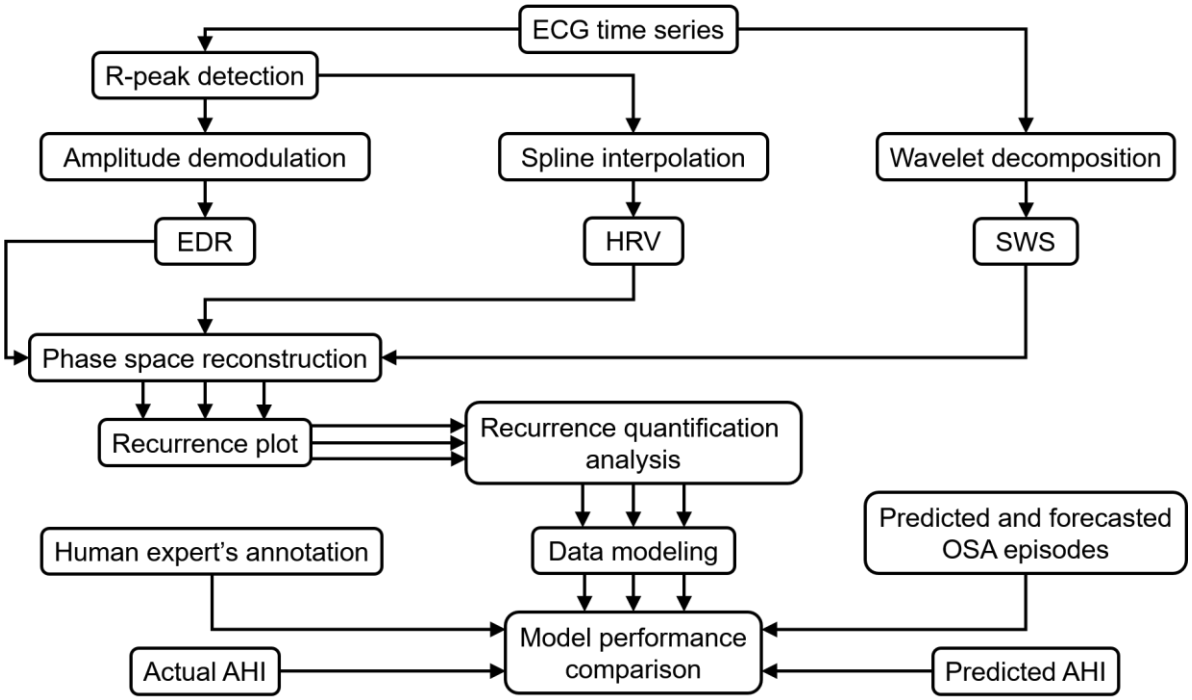


Figure 7.1: OSA episode classification and forecasting model development workflow

7.2.1 Phase space reconstruction parameters

Because of the variations in the underlying dynamics of each person, the phase space reconstruction parameters are individually calculated. Each time-series is then used for the calculation of the embedding time delay (mutual information) and embedding dimension (false nearest neighbor (FNN) rate). The embedding parameters for a phase space reconstruction of the three time-series (HRV, EDR, SWS) extracted from each subject's overnight one-lead ECG are shown in Table 7.1.

Table 7.1: The embedding parameters for the phase space reconstruction of the three time-series (HRV, EDR, and SWS) extracted from one-lead ECG in each subject

Subject	HRV		EDR		SWS	
	Time delay	Dimension	Time delay	Dimension	Time delay	Dimension
a1	49	6	12	7	12	7
a2	54	8	8	2	10	2
a3	66	6	11	7	20	6
a4	89	6	11	7	14	7
a5	91	6	15	7	24	8
a7	46	7	10	7	20	7
a8	72	6	15	7	23	7
a9	51	6	11	7	22	7
a12	80	6	18	7	19	6
a13	62	5	9	7	21	7
a15	72	5	15	7	21	7
a18	52	6	12	7	21	7
a19	51	6	16	7	30	8
a20	39	6	15	7	21	7
Group a average	62.43	6.07	12.71	6.64	12.71	6.64
b1	27	6	16	7	19	7
b2	67	6	16	7	8	6
b3	66	5	15	7	9	6
b4	28	6	17	6	20	7
Group b average	47	5.75	16	6.75	16	6.75
c1	14	6	14	6	22	6
c3	26	6	18	7	20	7
c4	10	7	13	7	20	7
c7	31	6	14	7	23	7
c8	26	7	16	7	8	6
c10	12	6	17	7	21	7
Group c average	19.83	6.33	15.33	6.83	15.33	6.83

Noticeably, in the HRV, and EDR time-series, the embedding delays of the subjects in group a (apnea) are quite different from those of the subjects in group c (normal). A quick student T-test also indicated the statistical difference between them (HRV: p = 9.83E-7, and EDR: p=3.51E-2). These results suggest that the underlying dynamics of the HRV and EDR time-series between apnea and normal groups are different. Although we expect the same result for the SWS, it is not the case.

The embedding dimensions suggested from the false nearest neighbor (FNN) rate for all time-series are approximately the same in all groups. This suggests that the number of variables that influence the

dynamics of each time-series is the same. The embedding dimensions shown in Table 7.1 are the dimensions where the FNN rate is less than 5% or equal to approximately zero.

7.3 Phase space reconstruction and recurrence plot (RP) of OSA episode predictors

After the required parameters are known, we can use equation 6.6 to reconstruct the phase space. However, in order to visualize the reconstructed phase space, the maximum dimension is limited to 3 (instead of 6 or 7). We choose to use ECG data from subject a05 because of the almost equal ratio of overall apnea and non-apnea minutes so we can choose only the consecutive apnea and non-apnea minutes to reconstruct the phase space. Also, we use the data from the same subject in order to eliminate the inter-personal variation and to maximize the different features influenced exclusively by apnea and non-apnea dynamics. Figures 7.2, 7.3, and 7.4 show the reconstructed phase spaces of 10 consecutive non-apnea and apnea minutes from HRV, EDR, and SWS time-series respectively.

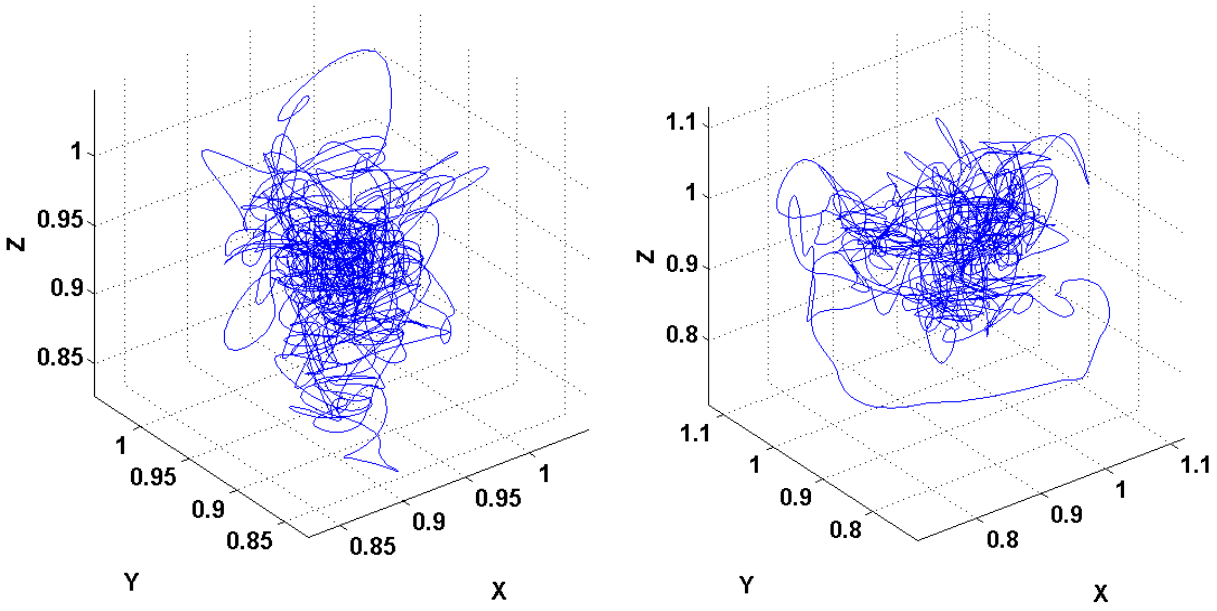


Figure 7.2: The phase space reconstruction of the HRV time-series (non-apnea: left, apnea: right)

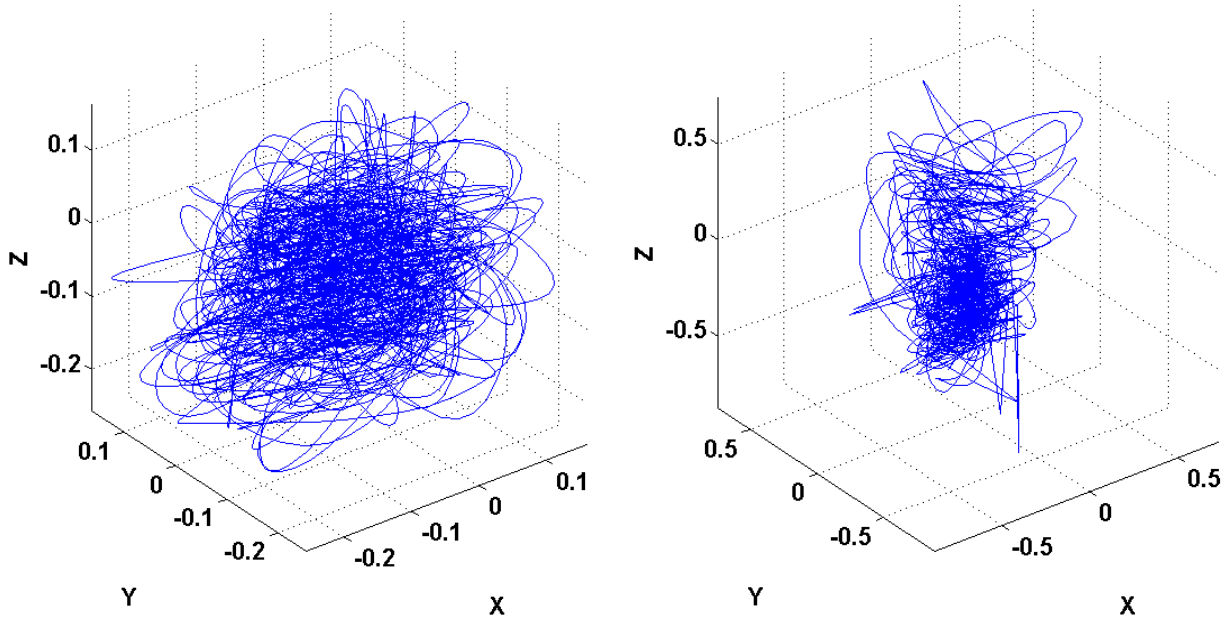


Figure 7.3: The phase space reconstruction of the EDR time-series (non-apnea: left, apnea: right)

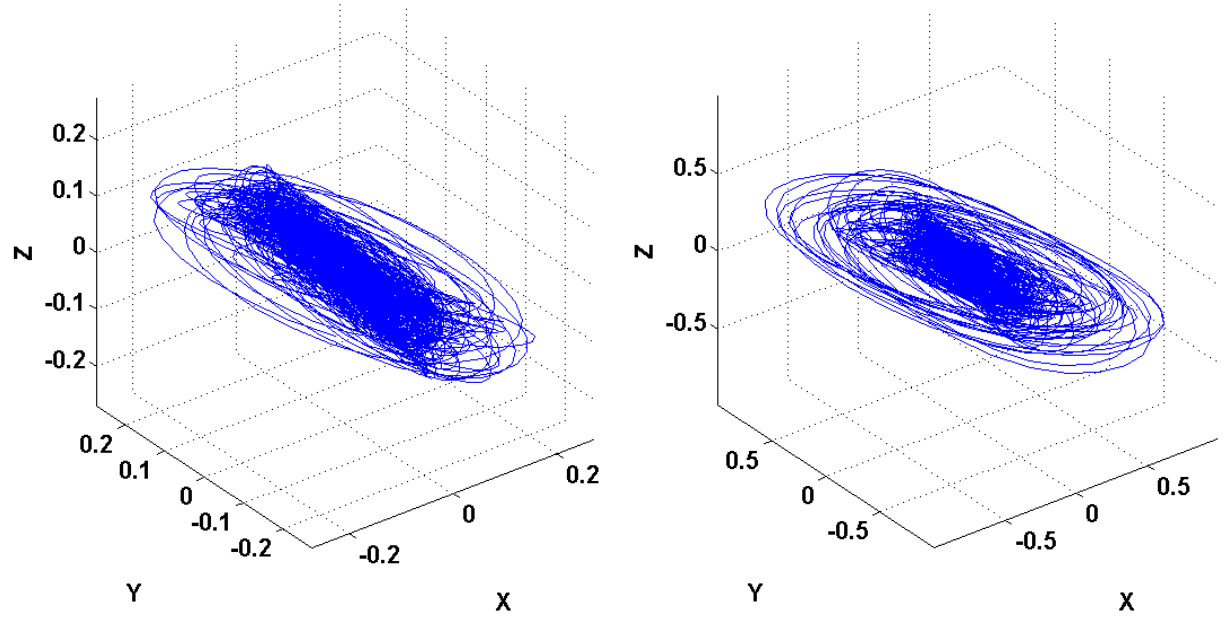


Figure 7.4: The phase space reconstruction of the SWS time-series (non-apnea: left, apnea: right)

Comparing the two phase spaces constructed from 10 consecutive non-apnea and apnea minutes, we can see some differences between them. In HRV phase spaces, it seems that the one from the non-apnea

minutes expresses more variation of the trajectory in the Z direction but less variation in the X direction compared to the one from the apnea minutes. In EDR phase spaces, the one constructed from non-apnea minutes seems to have approximately equal variations in every direction (almost sphere like), whereas the variations of the one from the apnea minutes are more constricted in the X and Y directions. Lastly, the phase spaces reconstructed from the SWS time-series express themselves almost as an X-Y plane with almost no variation in the Z direction. However, it seems that the one reconstructed with non-apnea minutes has a more confined trajectory (very dense) in the center whereas the trajectory of the one from the apnea minutes evolves more outside.

Although there are some differences between the two subjects (apnea vs. non-apnea), it might be too subjective to use the phase space plots to directly distinguish between apnea and non-apnea state subjects because there might be some person-to-person variations. The quantitative approach is more suitable and will be discussed later. Also, using half of the suggested dimension ($3 < 7$) to reconstruct the phase spaces may not fully reveal all the folded patterns of the underlying dynamics of all three time-series.

Once the phase spaces are reconstructed, we now can project these points on the phase spaces onto the two-dimensional plots, which is a recurrence plot. This time we reconstruct all phase spaces using the recommended dimensions and then create the recurrence plots using equation 7.12.

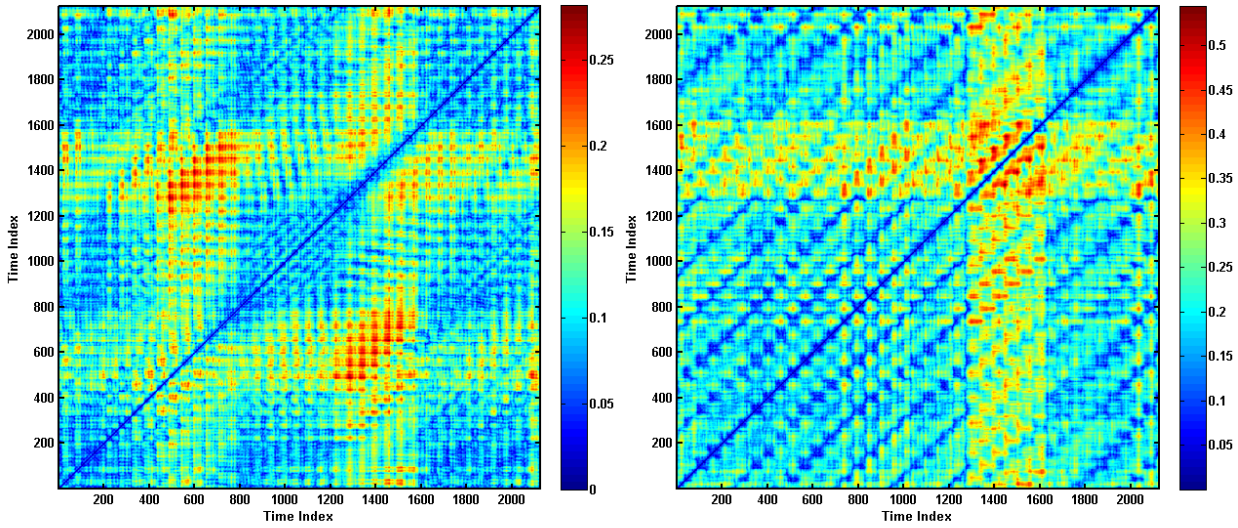


Figure 7.5: The unthresholded recurrence plots of the HRV time-series (non-apnea-left, apnea-right)

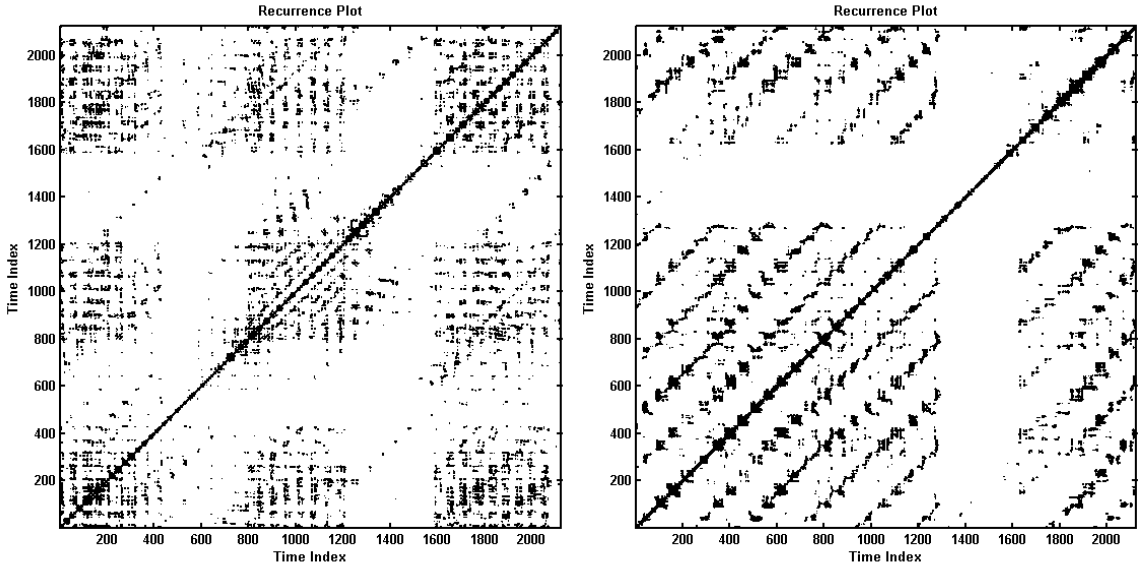


Figure 7.6: The thresholded recurrence plots of the HRV time-series (non-apnea-left, apnea-right)

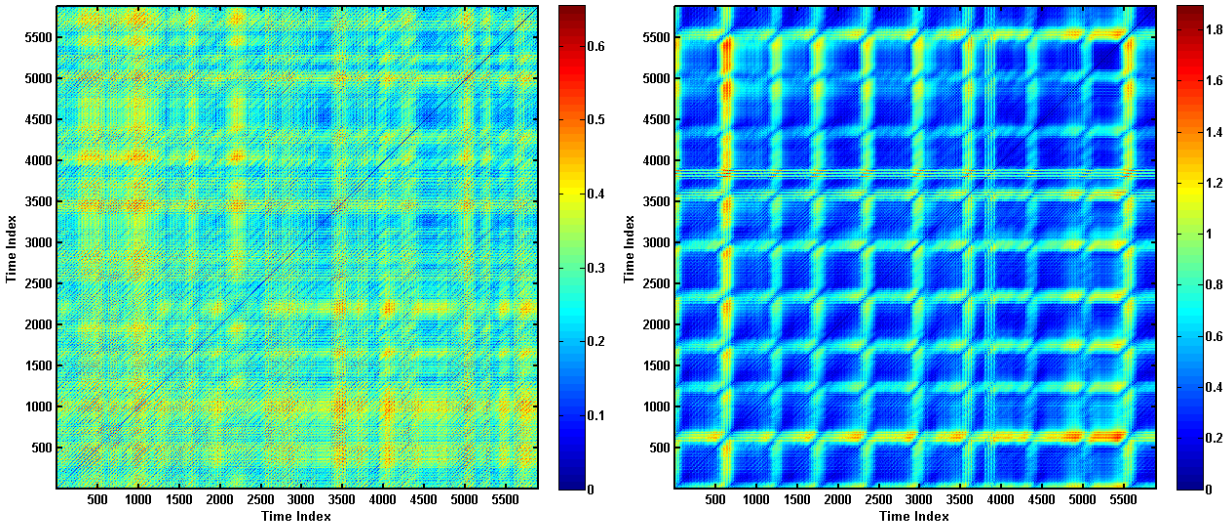


Figure 7.7: The unthresholded recurrence plots of the EDR time-series (non-apnea-left, apnea-right)

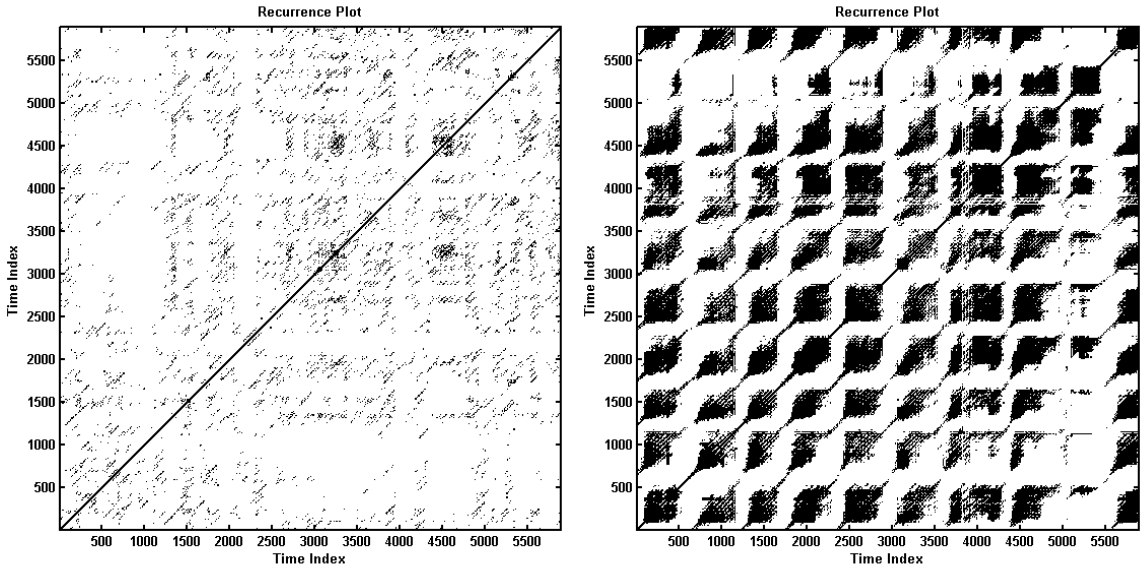


Figure 7.8: The thresholded recurrence plots of the EDR time-series (non-apnea-left, apnea-right)

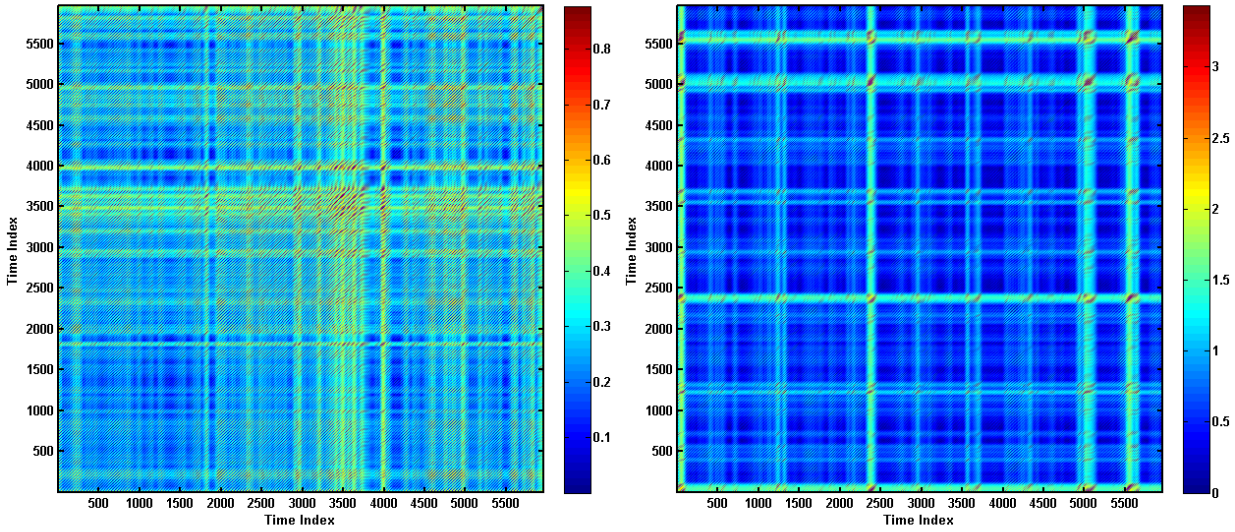


Figure 7.9: The unthresholded recurrence plots of the SWS time-series (non-apnea-left, apnea-right)

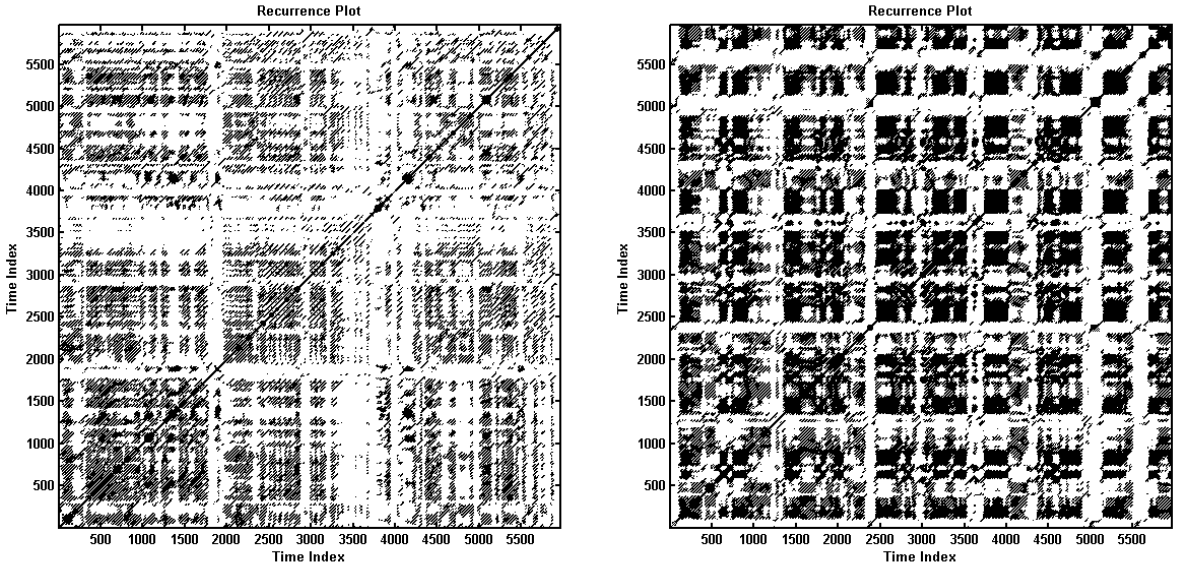


Figure 7.10: The thresholded recurrence plots of the SWS time-series (non-apnea-left, apnea-right)

Figures 7.5, 7.7, and 7.9 are the unthresholded recurrence plots (RP) of HRV, EDR, and SWS respectively. The color coding represents the distance between one state and all other states in the phase space. The cooler colors (blue) represent shorter distances between the pair of states and the hotter colors (red) represent the farther distances. We defined a true neighbor threshold using a recommended value, $\varepsilon =$ 10 percent of the maximum phase space diameter. Then, if the distance between the two states is less than ε (meaning that they are true neighbors), we represent the relationship by a black point (recurrence point) on the RP, otherwise a white point (not a true neighbor). This type of plot seen in Figures 7.6, 7.8, and 7.10 is called a thresholded recurrence plot (RP).

Focusing on the threshold RPs, the patterns between the apnea and non-apnea states are obvious. For the HRV thresholded RP from the apnea minutes, we can see more diagonal structures with some of the vertical and horizontal lines. In contrast, the one from the non-apnea minutes expresses more vertical and horizontal structures creating box-like structures all over the plot. For the RPs from the EDR time-series, the obviously different patterns make it easier to differentiate between apnea and non-apnea states. The EDR RP from the non-apnea minutes shows a very sparse diagonal structure all over, whereas the one from the apnea minutes expresses very dense diagonal structures which form an obvious box-like structure.

Note that the sampling rate for the EDR time-series is 10 Hz, so the 600 time-indexes corresponds to one minute. Thus, we can see that the box-like structure in the EDR RP corresponds to the OSA episode and a very short time period of the white band that corresponds to the apnea termination of the episode. These behaviors can also be seen in the SWS RPs with slight differences. The SWS RP from the non-apnea minutes forms dense diagonal structures all over. However, when compared to the one from the apnea minutes, the vertical structures are denser and form more obviously box-like structures. The RP structures mentioned above are not only used in a qualitative classification of the OSA states but they are also used in the quantitative measure by converting those structures into the measurement. The implementation and results will be discussed in the next section.

7.4 Implementation of recurrence quantification analysis (RQA) on the recurrence plot (RP) of OSA episode predictors

So far, we have explained how the RP can be used to discriminate visually between apnea and non-apnea states. In this section, we explain in detail how we implement the methods explained in section 6.4. However, we also encountered some limitations when trying to use them. Coming back to our original problem, how we can maximize the information embedded in the one-lead ECG to predict the OSA state, we are bound by the one-minute resolution annotation given by human experts. In the other words, we need to be able to interpret all information embedded in a one-minute ECG for the OSA state. After the HRV, EDR, and SWS time-series were extracted from the ECG time-series, each signal was resampled to a suitable sampling rate. In our study, the HRV was sampled at 4 Hz which is 4 samples per second. The EDR and SWS, were resampled at 10 Hz (10 samples per second). Recall that the time delay used for reconstructing the HRV phase space for subject a05 is 91 samples (see Table 7.1). Referring to the equation 6.6, for the subject a05, the time delay used for reconstructing the phase space is 91 samples with the dimension of 6. This means that to reconstruct one point in the phase space, we need at least $x_{i+(m+1)\tau}$ data points, which is equal to 547 points in this case. Thus, if we used the suggested time delay and dimension, we cannot reconstruct even one point in the HRV phase space for subject a05. Recall that equation 7.7,

$m \geq 2d + 1$, originally guarantees that the topology of the reconstructed phase space will resemble the real topology if $m \geq 2d + 1$. However, since we cannot reconstruct the phase space using the suggested dimension (m), we decrease the number of m until we can see the neighbor patterns which seem to work well in our case. Also, because the OSA symptoms are present not only in the apneic minute but some significant symptoms also appear before and after the apneic minute (i.e., increased heart rate and overshot respiratoration), we also included the three time-series of minutes before and after the apneic minute. With the limitations previously mentioned, the different data partition schemes and the smaller numbers of dimensions are used for RQA calculations.

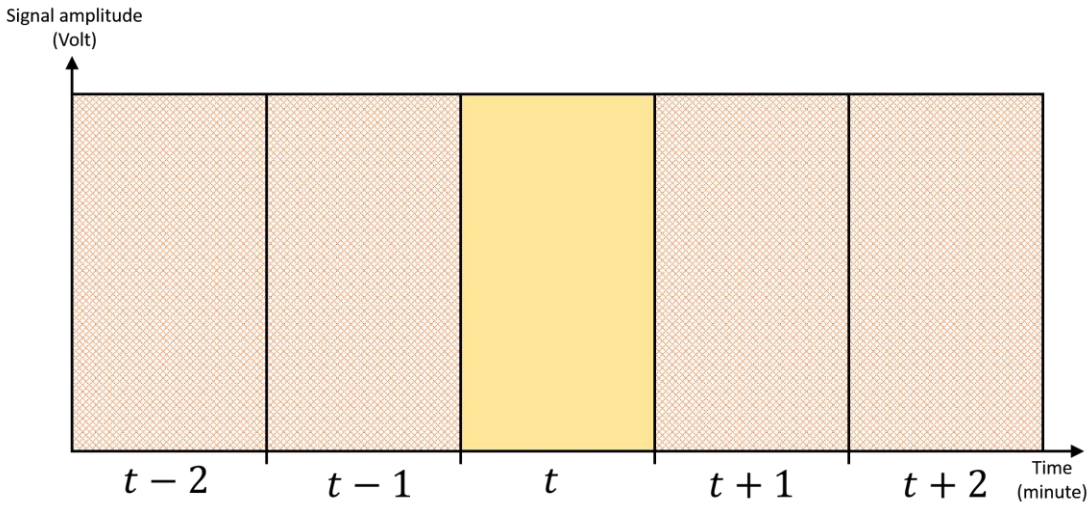


Figure 7.11: The data segmentation scheme

In Figure 7.11, if we focus on predicting the OSA state at t minute, rather than considering the information only in t minute, we can extend the boundary to the consecutive minutes (i.e., from $(t-1)$ to $(t+1)$ or from $(t-2)$ to $(t+1)$). For convenience, we will call the minutes after the focus minute, or t, advance minutes (a#) and the ones before the focus minutes as lagged minutes (l#). There are seven data partition schemes, l0a0, l1a0, l1a1, l1a2, l2a0, l2a1, l2a2. As mentioned above, the reason behind these data partitioning schemes work in this case is that there are significant apnea symptoms before apnea happens (i.e., increased heart rate or lower respiratory amplitude) and after apnea termination (i.e., respiration

overshooting or lower heart rate). This way we not only alleviate the data shortage problem but also improve the prediction accuracy.

To prepare each time-series for the RQA process, after all time-series are clean (apply band-pass filter, and remove outliers and spurious data points), we deleted the parts of the signals that do not have apnea annotations. Then, we partitioned the data corresponding to each partition scheme. After that, we reconstructed the phase space from each data partition using the time delay information in Table 7.1. For the dimension number, the maximum we can use is 4 dimensions, so we reconstructed the phase space using 3 and 4 dimensions because more data points may more clearly reveal apnea and non-apnea patterns. After the phase space of each data partition was reconstructed, we projected all the points in each phase space to the RP. All points in the RP are considered the neighbors because the distance threshold is automatically applied when we constrict the signal to only 1 – 5 minutes depending on which data partition scheme is used. Finally, all eleven RQA features were calculated from each RP using a Matlab toolbox available from [183]. In addition to the RQA features, the PSD and normalized PSD were also used for building classification and forecasting models. The results of the different data segmentation schemes are discussed in the next sections.

7.5 OSA episode classification model based on RQA features

From the total of 10,052 data points (equally sampled), we partitioned data into three portions, training, validation, and testing, with the data ratios of 50, 35, and 15 percent respectively. With the total of 39 features (RQAs and PSDs) extracted from each data partition of HRV, EDR, and SWS signals, we applied the three modeling techniques explained in Chapter 4 namely, a decision tree (DT), logistic regression (LR), and an artificial neural network (ANN), to capture the different patterns of OSA states. We also varied factors such as the number of dimensions and the data partition schemes used for reconstructing the phase spaces of HRV, EDR, and SWS. Because the dominant frequencies of EDR and SWS are not significantly different, we chose to use the same dimension when reconstructing their phase

spaces. The results for the best model of each data partition scheme and the different choices of dimensions are shown in Table 7.2.

Table 7.2: The model performance of the best modeling method used in each data partition and embedding dimension scheme

Model	Embedding dimension			Period used		Best modeling method	Misclassification rate (%)		
	HRV	EDR	SWS	$t - x_1$	$t - x_2$		Training	Validation	Testing
d3d3l0a0	3	3	3	0	0	ANN-MLP40	8.03	11.74	10.92
d3d3l1a0	3	3	3	1	0	ANN-MLP45	6.4	10.03	9.47
d3d3l1a1	3	3	3	1	1	ANN-MLP64	4.21	8.64	9.33
d3d3l1a2	3	3	3	1	2	ANN-MLP35	5.11	8.5	9
d3d3l2a0	3	3	3	2	0	ANN-MLP60	3.58	8.98	9.2
d3d3l2a1	3	3	3	2	1	ANN-MLP61	4.55	8.01	8.54
d3d3l2a2	3	3	3	2	2	ANN-MLP64	4.99	7.33	7.94
d3d4l0a0	3	4	4	0	0	ANN-MLP30	9.67	11.88	11.05
d3d4l1a0	3	4	4	1	0	ANN-MLP3	9.63	10.86	9.33
d3d4l1a1	3	4	4	1	1	ANN-MLP64	5.31	8.81	8.87
d3d4l1a2	3	4	4	1	2	ANN-MLP60	5.79	8.3	8.8
d3d4l2a0	3	4	4	2	0	ANN-MLP25	6.5	9.69	9.4
d3d4l2a1	3	4	4	2	1	ANN-MLP30	3.52	8.55	7.88
d3d4l2a2	3	4	4	2	2	ANN-MLP60	5.27	8.01	7.74
d4d4l0a0	4	4	4	0	0	Cannot calculate – Not enough points			
d4d4l1a0	4	4	4	1	0	ANN-MLP64	7.06	10.63	9.27
d4d4l1a1	4	4	4	1	1	ANN-MLP64	5.79	9.15	9.8
d4d4l1a2	4	4	4	1	2	ANN-MLP60	5.45	8.55	8.87
d4d4l2a0	4	4	4	2	0	ANN-MLP50	6.52	9.61	9.66
d4d4l2a1	4	4	4	2	1	ANN-MLP60	4.11	8.33	9
d4d4l2a2	4	4	4	2	2	ANN-MLP55	2.2	7.76	8.01

Shown in Table 7.2, using the results from the testing data partition as a benchmark, the best model, an ANN with a data partitioning scheme that includes 2 minutes before and after the analyzed minute, has a 7.74% misclassification rate or 92.26% accuracy. This result suggests that expanding the phase space data points to the adjacent minutes helps improve the classification accuracy. In the best case, extending the time to 2 minutes before and after the focus minute improves the classification accuracy almost 4 percent. As to the influence of the number of dimensions, although we see the overall best model when embedding

the HRV, EDR, and SWS with 3, 4, and 4 dimensions respectively, there seems to be no significant improvement between 3 and 4 embedding dimensions, as seen from their misclassification rates. Again, of the three modeling methods, the multilayer perceptron ANN provides the best results in every variation. In addition, we cannot calculate the RQA in is one case because there are not enough points in the reconstructed phase space because of the high dimensional embedding parameters (4-4-4) and no adjacency minutes considered. Table 7.3 reports the summary of the best parameters for model construction and the model performances based on the testing (left out) data partition. Furthermore, we also apply this model to each subject's ECG data to see how it performs individually. The results are reported in Table 7.4.

Table 7.3: Summary of the best model parameters and performances based on the results from the testing data partition

Parameter	Value
Data partitioning scheme	2 lag and 2 advance minutes (5 minutes total)
Embedding time delay	Individually calculated (based on the subject)
Embedding dimension	HRV: 3, EDR: 3, and SWS: 4
Modeling method	ANN-MLP60
Features	39 features from RQAs and PSDs
Accuracy	92.26%
True positive rate (TPR)	93.72%
True negative rate (TNR)	90.87%
False positive rate (FPR)	9.13%
False negative rate (FNR)	6.28%

Table 7.4: Summary of the best classification model performance on each individual ECG data

Apnea group	Accuracy	Total sleep min	Actual Apnea min	Predicted Apnea min	Actual non-apnea min	Predicted non-apnea min	Actual AHI	Predicted AHI
a01	98.76	484	467	471	17	13	57.89	58.39
a02	95.03	523	417	419	106	104	47.84	48.07
a03	91.25	514	243	262	271	252	28.37	30.58
a04	99.18	487	450	452	37	35	55.44	55.69
a05	91.31	449	274	301	175	148	36.61	40.22
a07	88.93	506	319	343	187	163	37.83	40.67
a08	87.5	496	189	195	307	301	22.86	23.59
a09	91.83	490	378	374	112	116	46.29	45.80
a12	96.15	572	531	545	41	27	55.70	57.17
a13	92.04	490	241	242	249	248	29.51	29.63
a15	91.09	505	368	379	137	126	43.72	45.03
a18	98.14	484	435	438	49	46	53.93	54.30
a19	96.38	497	201	201	296	296	24.27	24.27
a20	94.85	505	315	315	190	190	37.43	37.43
Borderline group	Accuracy	Total sleep min	Actual Apnea min	Predicted Apnea min	Actual non-apnea min	Predicted non-apnea min	Actual AHI	Predicted AHI
b01	96.05	482	19	26	463	456	2.37	3.24
b02	96.29	512	93	94	419	418	10.90	11.02
b03	97.94	436	71	76	365	360	9.77	10.46
b04	98.52	406	10	12	396	394	1.48	1.77
Control group	Accuracy	Total sleep min	Actual Apnea min	Predicted Apnea min	Actual non-apnea min	Predicted non-apnea min	Actual AHI	Predicted AHI
c01	99.13	457	0	4	457	453	0.00	0.53
c03	99.56	449	0	2	449	447	0.00	0.27
c04	99.77	435	0	1	435	434	0.00	0.14
c07	97.17	424	4	8	420	416	0.57	1.13
c08	100	508	0	0	508	508	0.00	0.00
c10	99.76	402	1	0	401	402	0.15	0.00

Considering the classification accuracy based on OSA severity, the accuracy range in the Apnea group (a##) is the widest compared to the ones in the borderline (b##) and control (c##) groups. The best prediction accuracy is 99.18% with subject a04 and the worst prediction accuracy is 87.5% with subject a08. The reason for the difference seems to be related to the complexity of the OSA state, which can be determined by the number of OSA state changed for each subject's entire night's sleep. To clarify, we counted the number of events when the state changed from non-apnea to apnea and from apnea to a non-

apnea state. A higher number of state changes indicates more complexity in the OSA pattern, so it is harder to accurately predict. Conversely, it is easier to predict if the apnea pattern does not change much. For subject a04, there are only 4 state changes, but for subject a08, the OSA state changes 64 times the most of all the subjects. For the borderline group, the prediction accuracy is good, averaging at 97.2%. This result is also because the OSA pattern is not very complex, with the maximum number of the OSA state changes at 26 times. Finally, for the subjects in the control group, our model does not have any problem predicting the OSA events in this group with an average accuracy of 99.23%. Again, the results in Table 7.3 confirm those in Table 7.2, that our best model tends to make more false positive events as seen from more predicted apnea minutes than the actual apnea minutes in almost every case.

Now, the most important output needed from the OSA classification process is the apnea-hypopnea index (AHI). This number classifies the subjects into categories based on the severity of OSA. The calculation is as follows [184]:

$$AHI = \frac{total\ number\ of\ apnea\ and\ hypopnea\ events}{total\ number\ of\ sleep\ time\ in\ minute} \cdot 60 \qquad (7.1)$$

Shown in Table 7.3, the differences between the actual AHI and the predicted AHI in every case is very small. The biggest difference is only 3 AHI. However, it should be noted that the AHI calculation is based on averaging OSA events; since the predicted outcome is a binary (0: non-apnea or 1: apnea), wrong predictions may contribute to better or worse AHI results. In our case, because our model tends to create more false positive events, the predicted AHIs are slightly overestimated in almost every case.

7.6 Importance predictors for OSA episode classification models based on RQA features

Although the DT algorithm is always not the best modeling method based on the accuracy or misclassification rate, its straightforward algorithm makes the results very interpretable, unlike the ANN algorithm. To fully understand each predictor's role in the apnea state, based on the overall Gini reduction from the DT model, we can determine how each variable contributes in the model (see Chapter 4 for more

information). The plot of the average Gini impurity reduction of the top 20 variables from the best classification model is shown in Figure 7.12.

Table 7.5: Full description of each variable name in Figure 7.12 (see Chapter 6 for more information)

Variable name	Full description
Entropy_EDR	Shannon information entropy of all diagonal line lengths of EDR signal
longdiag_EDR	Maximal diagonal line length of EDR signal
RTE_EDR	Recurrence time entropy of EDR signal
PSD_SWS	Power spectrum density of SWS signal
PSD_RR	Power spectrum density of RR signal
Rect1st_EDR	Mean of the recurrence times of the first type of EDR signal
Rect2nd_EDR	Mean of the recurrence times of the second type of EDR signal
NPSD_EDR	Normalized power spectrum density of EDR signal
Entropy_RR	Shannon information entropy of all diagonal line lengths of RR signal
PSD_EDR	Power spectrum density of EDR signal
longvert_EDR	Maximal vertical line length of EDR signal
RR_EDR	Recurrence rate or percent recurrence of EDR signal
Avediag_EDR	Average diagonal line length of EDR signal
RR_SWS	Recurrence rate or percent recurrence of SWS signal
NPSD_RR	Normalized power spectrum density of RR signal
Rect2nd_RR	Mean of the recurrence times of the second type of RR signal
Trap_EDR	Trapping time of EDR signal
longdiag_SWS	Maximal diagonal line length of SWS signal
NPSD_SWS	Normalized power spectrum density of SWS signal
RTE_SWS	Recurrence time entropy of RTE signal

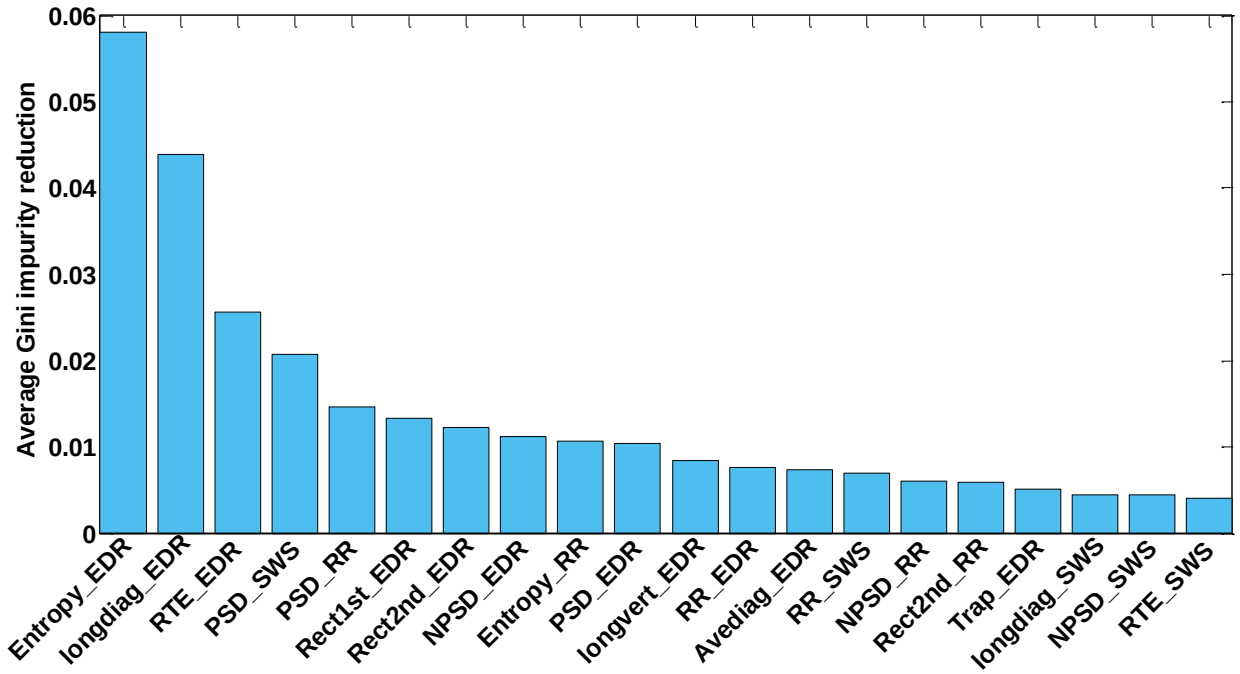


Figure 7.12: Variable importance based on the average Gini impurity reduction for the DT OSA episode classification model (see full description of each variable name in Table 7.5)

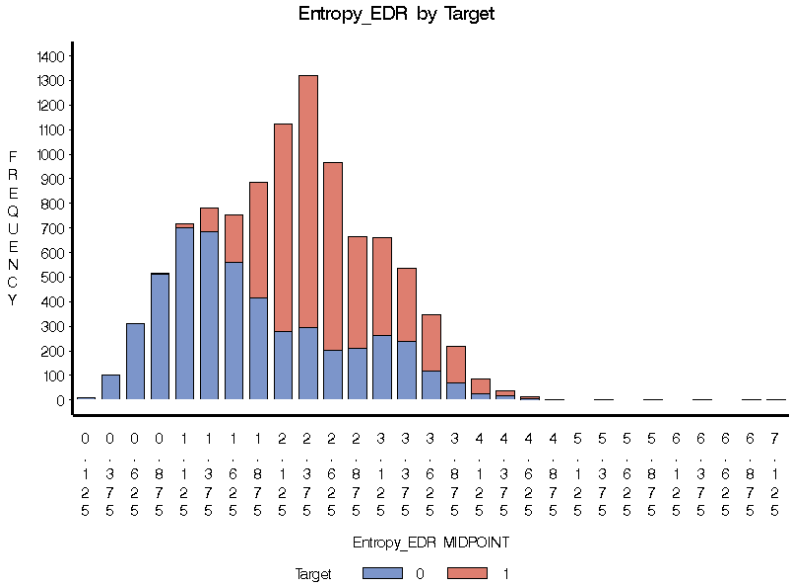


Figure 7.13: The histogram of EDR's Shannon entropy grouped by OSA episode state

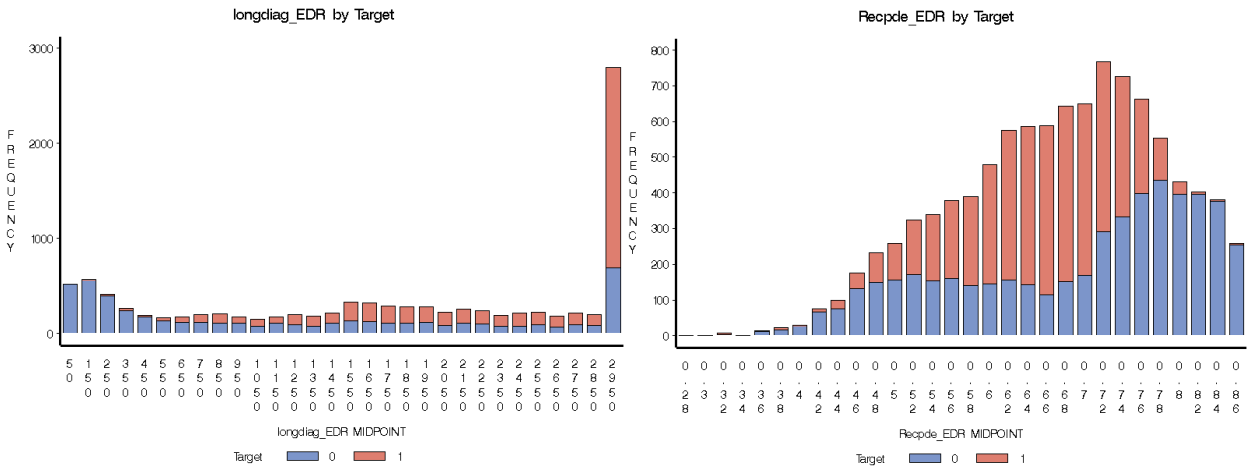


Figure 7.14: The EDR's longest diagonal line grouped by OSA episode state (left) and the EDR's recurrence time entropy grouped by OSA episode state (right)

The three most important features are from the EDR signal. The Shannon information entropy and the longest diagonal line contributes to about one third of the overall Gini impurity reduction (32.44%). This result is also supported by Figure 7.13. Clearly, the Shannon information entropy for the non-apnea states tends to have a lower value than for the apnea states. Recall that lower entropy indicates lower complexity in the system, represented by the sparse diagonal lines shown in Figure 7.8 for the non-apnea state RP. On the other hand, many diagonal lines with various lengths, which contributes to higher entropy, can be spotted easily in the RP plot for apnea states in Figure 7.8. Evidently, many continuous diagonal lines of EDR are most likely to show up in the RP of the apnea states (see Figures 7.8 and 7.13). Because the diagonal line structure related to the time between segments of the trajectory is close, the longest diagonal line suggests the stability of the system. With shorter longest diagonal lines in the RP, the system will tend to change its state (the next closest trajectory to the current state will tend to diverge).

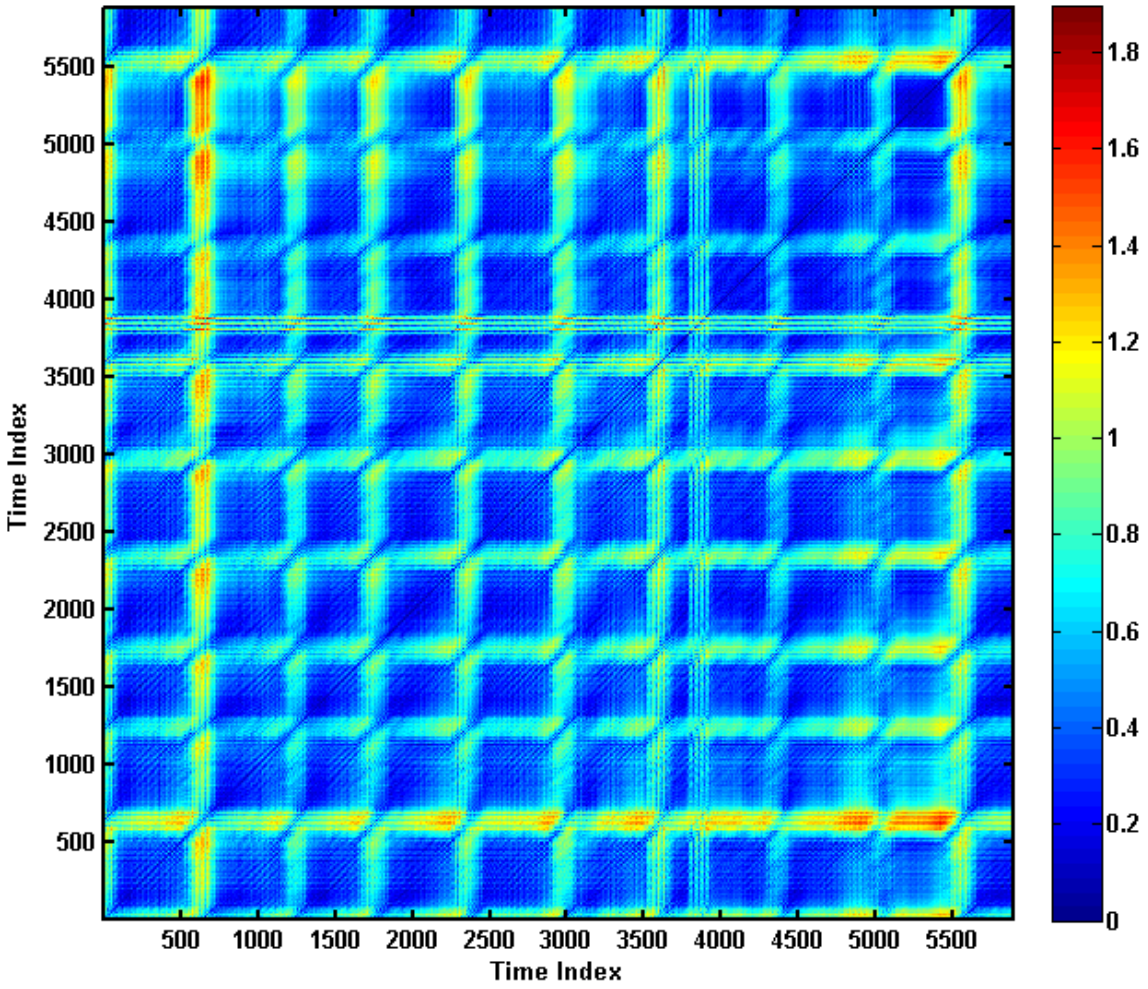


Figure 7.15: The unthresholded recurrence plots of the EDR time-series (10 consecutive apnea minutes)

The next important variable is the recurrence time entropy of the EDR time-series. This feature is actually the entropy of the vertical lines in the RP. It might be quite hard to see those vertical lines in Figure 7.8 (right) because of the high density of the RP points. However, it is quite easy to see many vertical lines in the unthresholded RP in Figure 7.15 that are potentially neighbors, indicated by the dark blue color (short distances to each other). Moreover, we can distinguish most apnea states from the mid-range and the non-apnea states from the low and high range of the recurrence time entropy as shown in Figure 7.14. According to [177], this measurement is used to determine the repetitiveness of the trajectory in the phase space. To be specific, the purely periodic signals will have RTE value equal to one, and as the complexity (chaos)

increases, the RTE also increases. The maximum RTE value is equal to one in a purely random white-noise case.

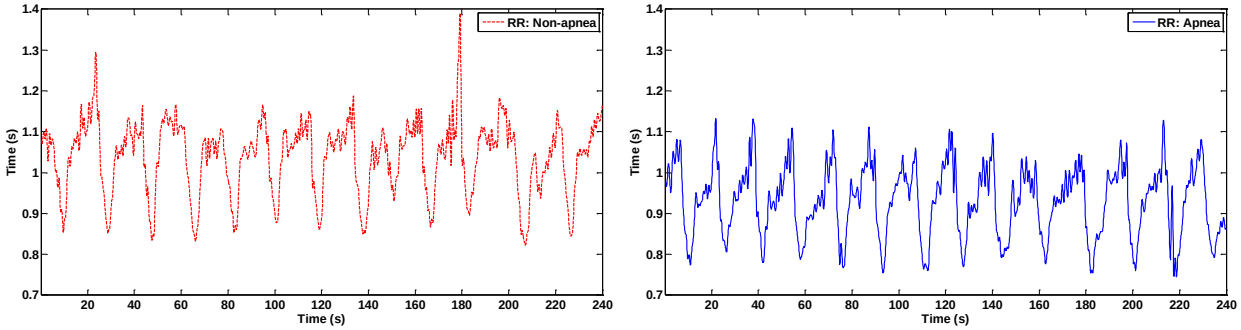


Figure 7.16: The RR's time-series plots: Non-apnea minutes (left), Apnea minutes (right)

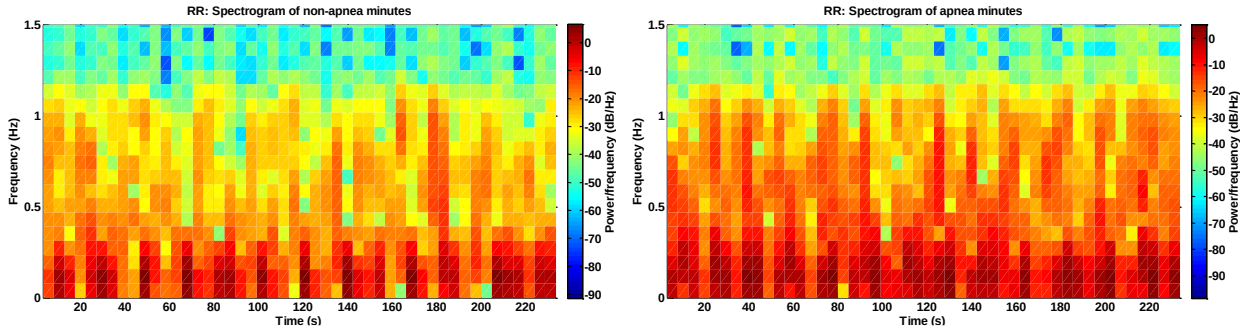


Figure 7.17: The RR's spectrogram: Non-apnea minutes (left), Apnea minutes (right)

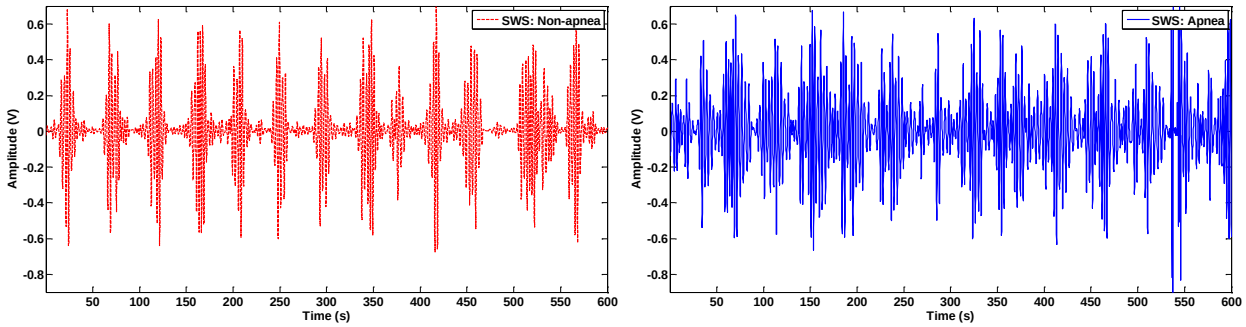


Figure 7.18: The SWS's time-series plots: Non-apnea minutes (left), Apnea minutes (right)

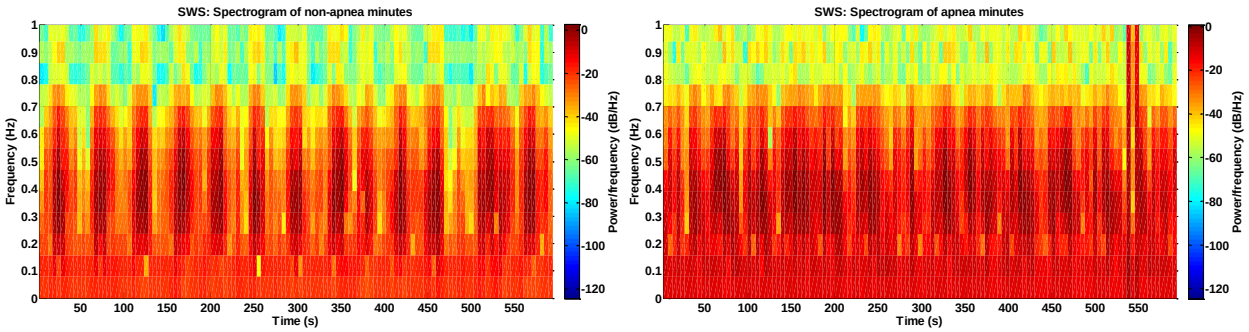


Figure 7.19: The SWS's spectrogram: Non-apnea minutes (left), Apnea minutes (right)

The next two important variables are the power spectrum densities (PSDs) of the SWS and RR time-series. Again, the PSD explains how the variation in the time-series is distributed over the frequency domain. In the feature extraction process, we calculated the PSD value for each minute. The results tell us all the power in every frequency up to half of the sampling frequency. Then, we can combine the total power of the frequency range of interest as used in one of our features. Considering the RR time-series from the same person between the apnea and non-apnea minutes, shown in Figure 7.16, the difference between the RR time-series of the apnea and non-apnea minutes can be spotted from the overall mean shift. The mean value of the apnea RR time-series is much smaller than the value from the non-apnea RR time-series. Moreover, they are also different in shapes and peak sizes. One of the reasons is that the heart rate becomes higher during a sleep apnea episode, so the HRV interval becomes shorter. Although it is quite clear that the mean values of the RR time-series are different between the two states, when it comes to the comparisons across persons, these mean values cannot be used as a features to differentiate between the two states due to heart rate variations between persons. However, if we consider this signal in the frequency domain, the active frequency ranges between the two states are quite different as shown in the spectrogram in Figure 7.17. We can see that in the non-apnea state, the HRV's variation is most active in the frequency range of about 0 – 0.35 Hz, indicated by the shades of red (see Figure 7.17). For the apnea minutes, the active frequency range is from 0 – 1 Hz.

For the SWS time-series (see Figure 7.18), we can see that the signal variation is sharply active for short periods divided by inactive periods in the non-apnea minutes. However, in the apnea minutes, the

active periods become longer, leaking into what is supposed to be a non-active period. These patterns are captured very well in the PSD shown in the spectrograms in Figure 7.19. Also, note that in the apnea minutes, the power in the frequency range 0-0.1 is highly active compared to the non-apnea minutes. Also, in the apnea minutes, the increase in active frequency regions can be clearly seen in Figure 7.19.

These features show very prominent patterns for us to easily differentiate differences between apnea and non-apnea states. However, for the remaining features shown in Figure 7.12, we could not find explicit patterns that can be explained visually. However, because of their Gini impurity reduction values, we believe that these features can be useful when used altogether with other reported features.

7.7 Performance comparison of OSA episode classification models

As reported in Chapter 3, we are not the first research group to come up with the idea to use information from an ECG signal to classify the OSA states. Several groups who use other methods and achieved very good OSA classification results. To compare our classification method to the others, we implement two methods from the two other respectable groups, [73] and [69], with the same data used in our study.

7.7.1 Brief explanation of comparison methods

In short, Mendez et al. [73] extracted two signals from the ECG signal, HRV (RR intervals) and QRS area time-series. The QRS area is estimated by a summation of the amplitude before and after the R-peak for 100 milliseconds. From there, the best 10 features based on these two signals are further extracted and quantified to each corresponding one-minute window. Roughly, all of Mendez et al.'s [73] features are power spectrum density (PSD) and magnitude squared coherence in different frequency bands, namely very low frequency (VLF: 0 – 0.04 Hz), low frequency (LF: 0.04 – 0.15 Hz), and high frequency (HF: 0.15 – 0.4 Hz). In digital signal processing, the coherence indicates the causal correlation between input x and output y at each frequency with a value between 0 and 1. The magnitude squared coherence can be calculated from:

$$C_{xy} = \frac{\left|P_{xy}\right|^2}{P_{xx}P_{yy}} \tag{7.2}$$

Where P_{xy} is the cross power spectral density (PSD) of x and y

P_{xx} is the power spectral density (PSD) of x

P_{yy} is the power spectral density (PSD) of y

They define two additional quantities, module (Mod) and phase (Pha) [73]:

$$Mod\ (n) = \sqrt{QRS_{area}(n)^2 + RR(n)^2} \tag{7.3}$$

$$Pha(n) = tan^{-1}\left(\frac{QRS_{area}(n)}{RR(n)}\right) \tag{7.4}$$

Where n is the heart beat number

Thus, all 10 features using Mendez et al.'s [73] method are summarized in Table 7.6.

Table 7.6: Ten most important variable in order of importance extracted using Mendez et al. [73] method

Feature number	Feature name
1	VLF power of RR intervals normalized to total power
2	Coherence at VLF between RR intervals and QRS area
3	LF/HF Ratio power of QRS area
4	VLF power of module normalized to total power
5	VLF power of QRS normalized to total power
6	HF Power of Phase norm to total power
7	LF Power of Phase norm to total power
8	Coherence at HF between RR intervals and QRS area
9	LF power of module normalized to total power
10	LF power of QRS normalized to total power

In the modeling process, Mendez et al. used a multilayer perceptron ANN with 3-30 neurons in the hidden layer to train their classifier. We replicated all these processes and discuss their model performance in detail at the end of this section.

Another method used for comparing the performance of our OSA classification model is from Delibasoglu et al. [69]. In brief, Delibasoglu and the group extracted the instantaneous heart rate (IHR) from the ECG signal. The IHR is almost analogous to the RR interval. The RR interval tells how long (in

seconds) from one R peak to another, whereas the IHR tells how fast the heart is beating on average (beats per minute) between one R peak and another. They also consider the data partitioning scheme as the data from three minutes before the analysis minute, the analysis minute, and two minutes after the analysis minute, which is 6 minutes in total. Then, the wavelet decomposition using Daubechies order 3 with 8 levels is performed on each data partition so that there are 16 detail and approximation components equally. Further, they quantify these components by calculating the mean variance of each detail component and use these eight quantities as deciding values for OSA state classification. Delibasoglu et al. train the classification model using a three-layer nonlinear autoregressive network with exogenous variables (NARX). It has an input layer with 6 neurons, a hidden layer with 3 neurons, and an output layer with one neuron. We also replicate their processes to build an OSA state classifier.

7.7.2 Performance comparison

To the best of our knowledge, we have replicated and implemented the other two OSA classification methods. Then, we scored all the performance indexes (accuracy, TPR, TNR, FPR, and FNR) using the same testing data partition used with our method. The OSA classification results from the testing data partition are shown in Figure 7.20 as follows:

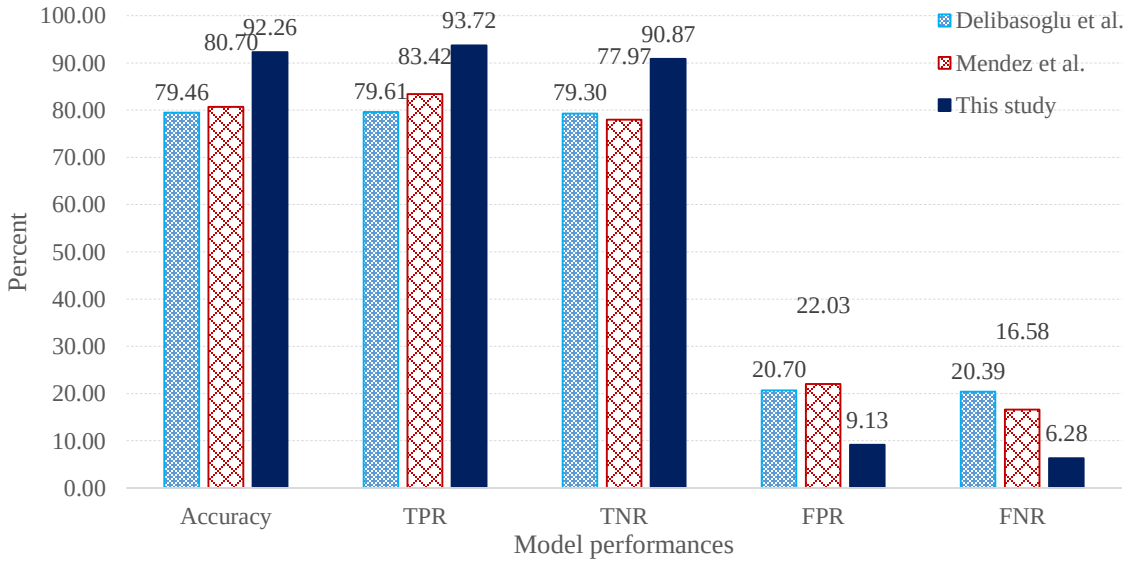
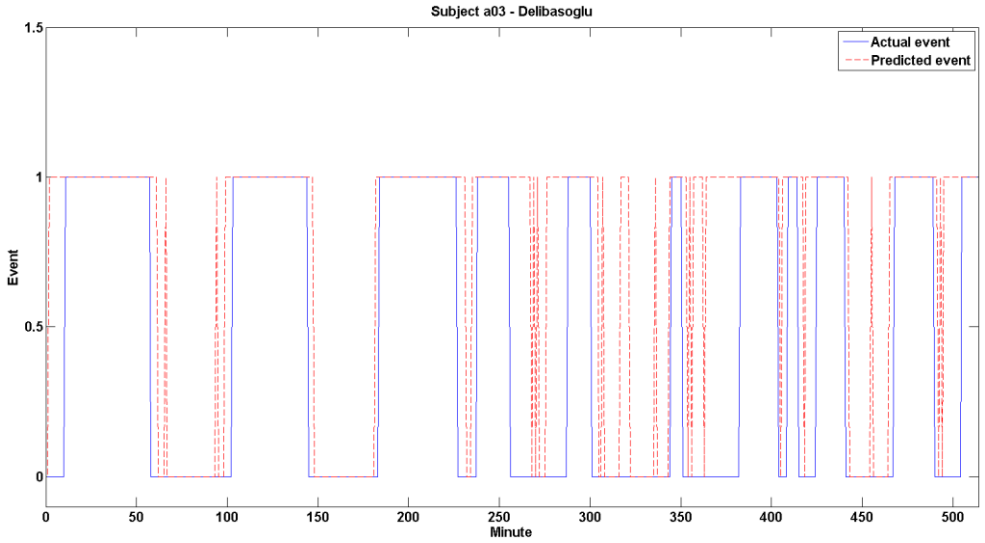
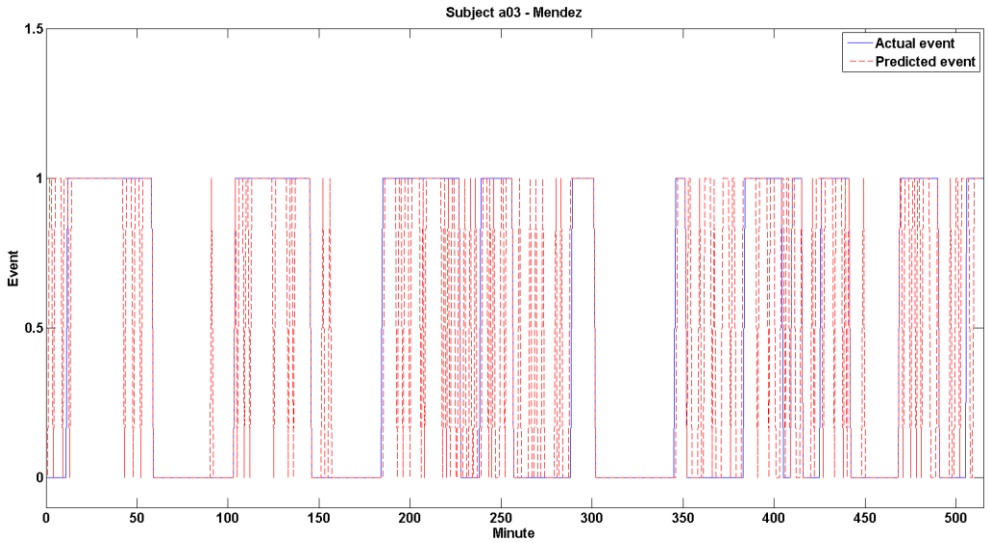


Figure 7.20: Performances comparison of OSA episode classification models

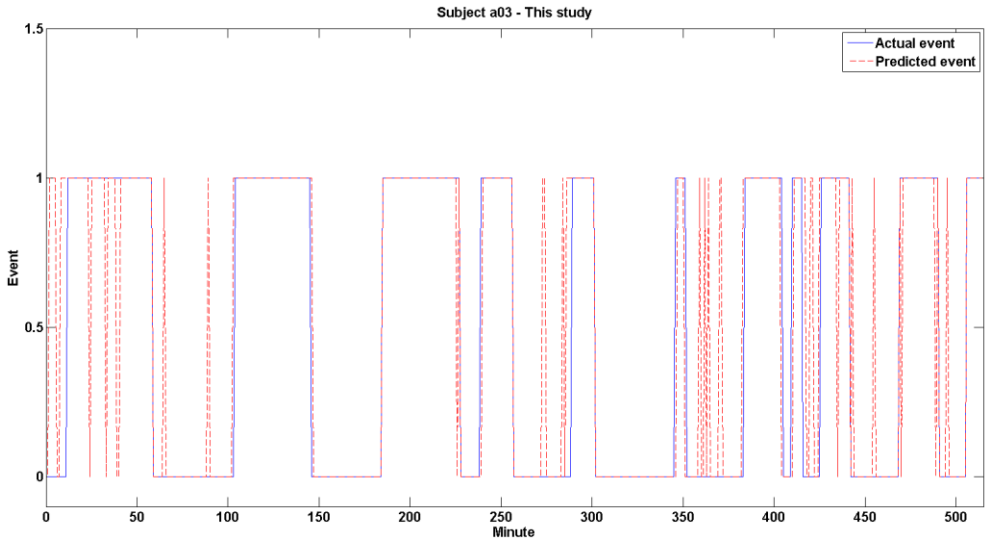
As shown in the Figure 7.20, our classification model outperforms the other two models in every model performance index. Again, these performances are from the pool of apnea and non-apnea events that only tell the model performances on average. Now, to further see if the models are suitable for use in real-life applications, we apply the three models to the subjects across the groups: apnea, borderline, and control groups. The subjects chosen from apnea groups are subject a03, a08, and a13 based on the complexity of their OSA patterns. The number of OSA state changes is 21, 64, and 39 times respectively. For the borderline group, we chose the two subjects with the most complex OSA patterns, b02, and b03, with the number of OSA state changes of 26 and 21 times respectively. For the control group, we chose two subjects with OSA events, c07, and c10. The results are shown in the figures as follows:



Subject	a03
Method	Delibasoglu et al.
Accuracy	74.46%
TPR	100.00%
TNR	51.48%
FPR	48.52%
FNR	0.00%
Actual apnea minute	243
Predicted apnea minute	374
Actual AHI	28.36
Predicted AHI	43.74

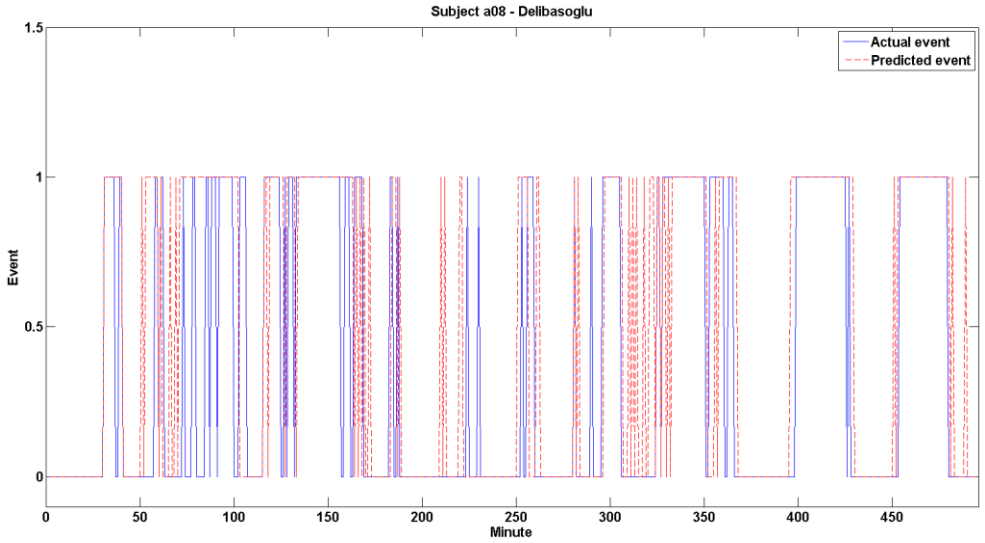


Subject	a03
Method	Mendez et al.
Accuracy	80.35%
TPR	79.01%
TNR	81.55%
FPR	18.45%
FNR	20.99%
Actual apnea minute	243
Predicted apnea minute	242
Actual AHI	28.36
Predicted AHI	28.24

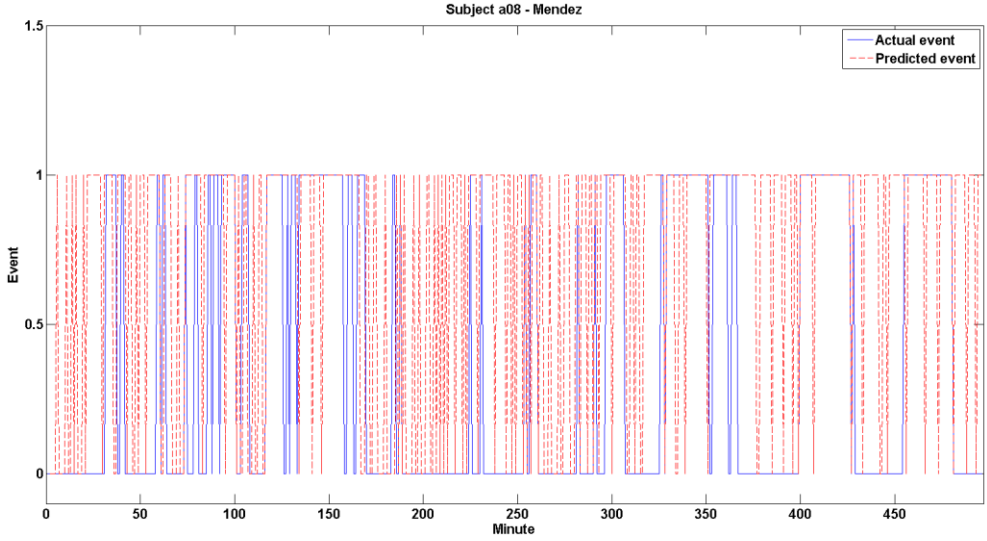


Subject	a03
Method	This study
Accuracy	91.25%
TPR	94.65%
TNR	88.19%
FPR	11.81%
FNR	5.35%
Actual apnea minute	243
Predicted apnea minute	262
Actual AHI	28.36
Predicted AHI	30.58

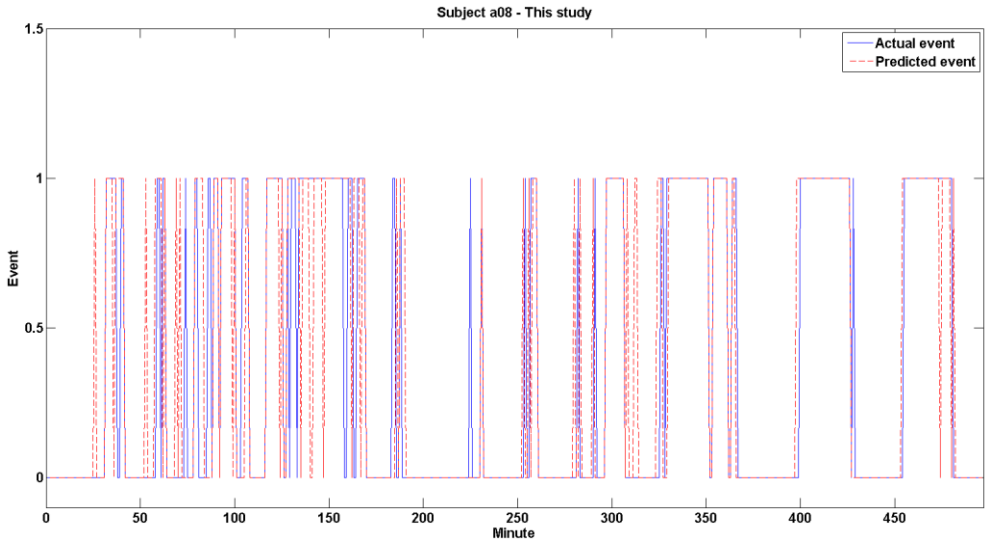
Figure 7.21: Comparison of OSA classification results for subject a03



Subject	a08
Method	Delibasoglu et al.
Accuracy	79.80%
TPR	86.24%
TNR	75.82%
FPR	24.18%
FNR	13.76%
Actual apnea minute	189
Predicted apnea minute	237
Actual AHI	22.86
Predicted AHI	28.73

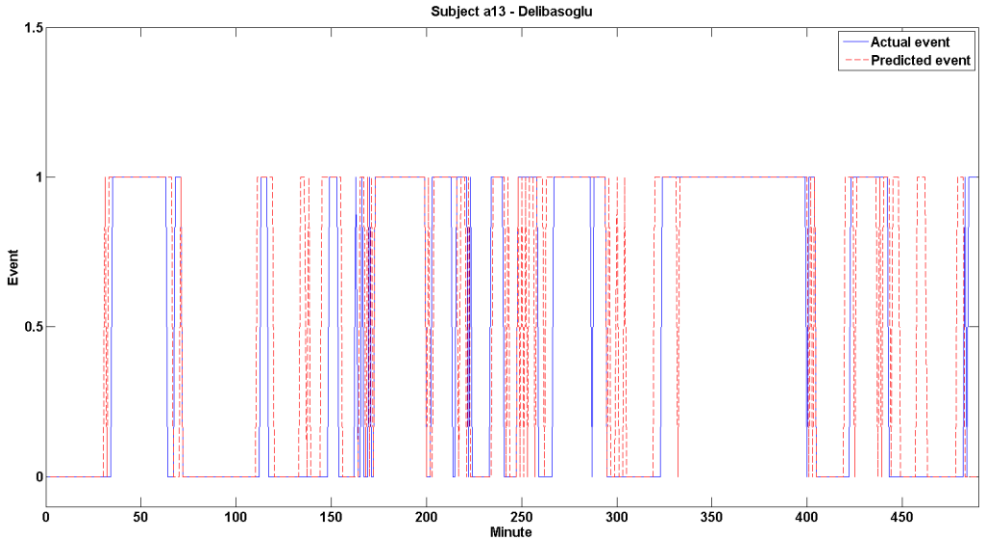


Subject	a08
Method	Mendez et al.
Accuracy	55.65%
TPR	87.30%
TNR	36.16%
FPR	63.84%
FNR	12.70%
Actual apnea minute	189
Predicted apnea minute	361
Actual AHI	22.86
Predicted AHI	43.67

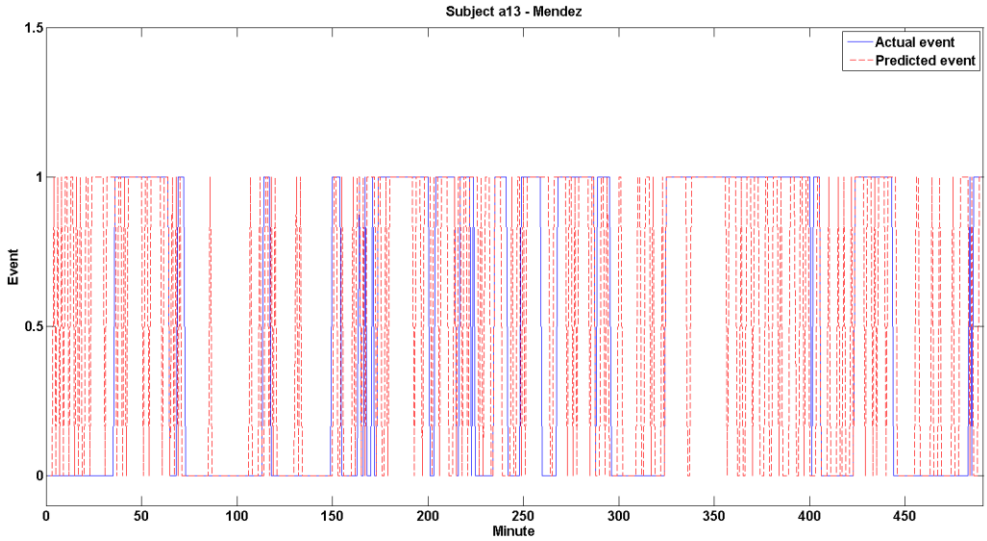


Subject	a08
Method	This study
Accuracy	87.50%
TPR	85.19%
TNR	88.93%
FPR	11.07%
FNR	14.81%
Actual apnea minute	189
Predicted apnea minute	195
Actual AHI	22.86
Predicted AHI	23.58

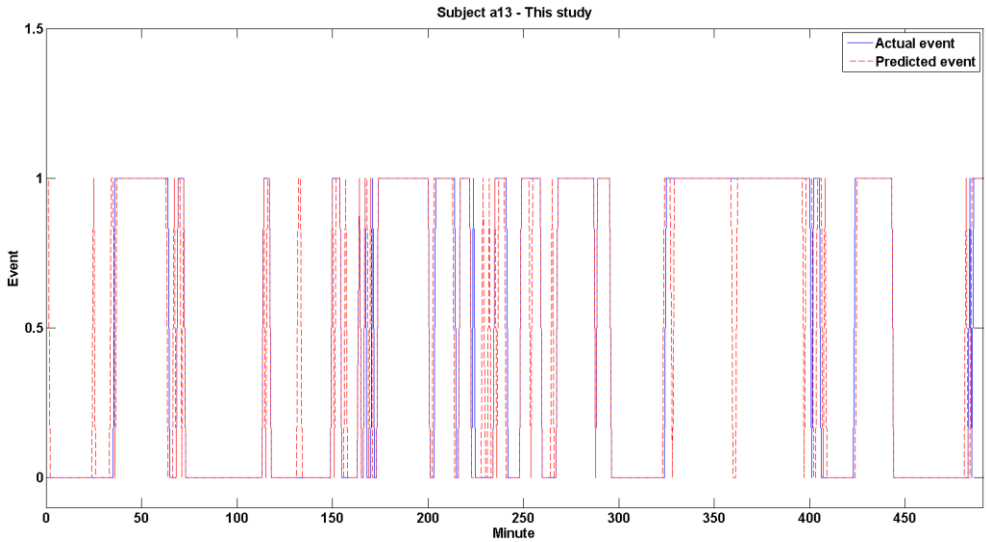
Figure 7.22: Comparison of OSA classification results for subject a08



Subject	a13
Method	Delibasoglu et al.
Accuracy	81.60%
TPR	89.21%
TNR	74.19%
FPR	25.81%
FNR	10.79%
Actual apnea minute	241
Predicted apnea minute	279
Actual AHI	29.63
Predicted AHI	34.23

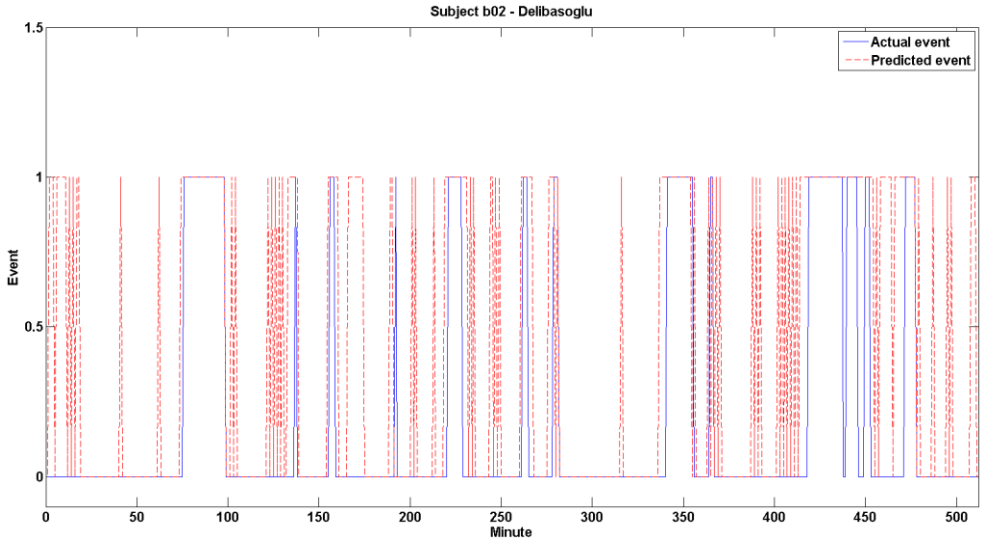


Subject	a13
Method	Mendez et al.
Accuracy	68.37%
TPR	70.54%
TNR	66.27%
FPR	33.73%
FNR	29.46%
Actual apnea minute	241
Predicted apnea minute	254
Actual AHI	29.63
Predicted AHI	31.1

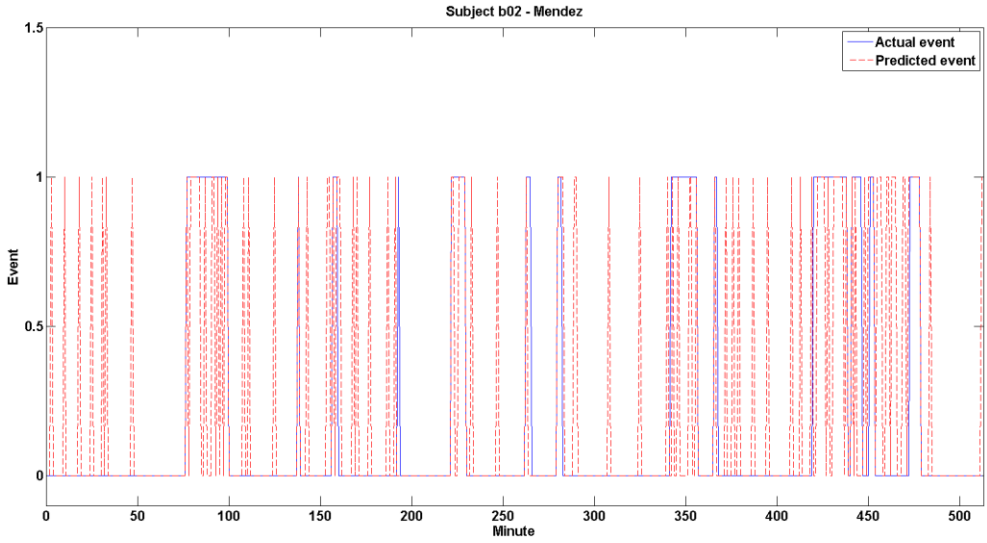


Subject	a13
Method	This study
Accuracy	92.04%
TPR	92.12%
TNR	91.97%
FPR	8.03%
FNR	7.88%
Actual apnea minute	241
Predicted apnea minute	242
Actual AHI	29.63
Predicted AHI	29.51

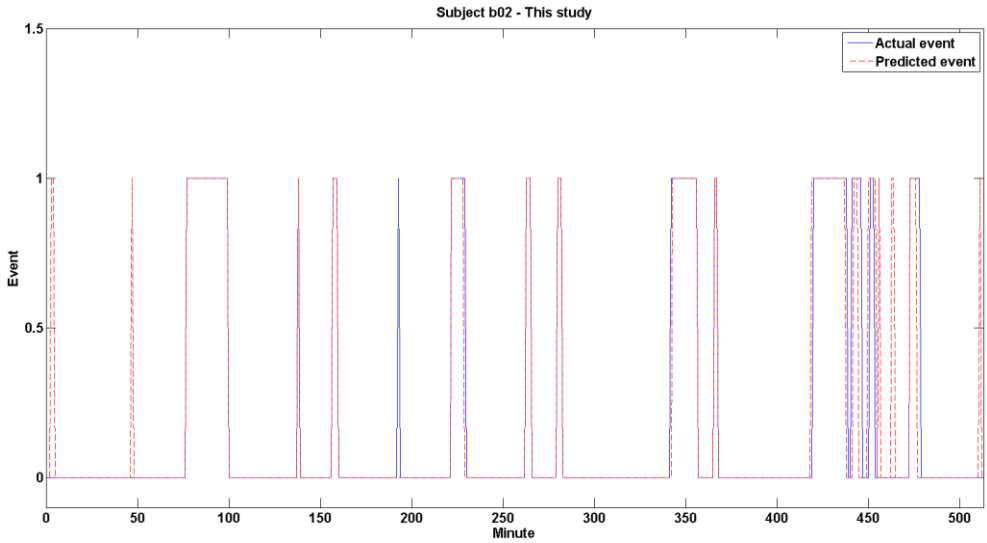
Figure 7.23: Comparison of OSA classification results for subject a13



Subject	b02
Method	Delibasoglu et al.
Accuracy	76.71%
TPR	95.70%
TNR	72.32%
FPR	27.68%
FNR	4.30%
Actual apnea minute	93
Predicted apnea minute	205
Actual AHI	11.01
Predicted AHI	24.07

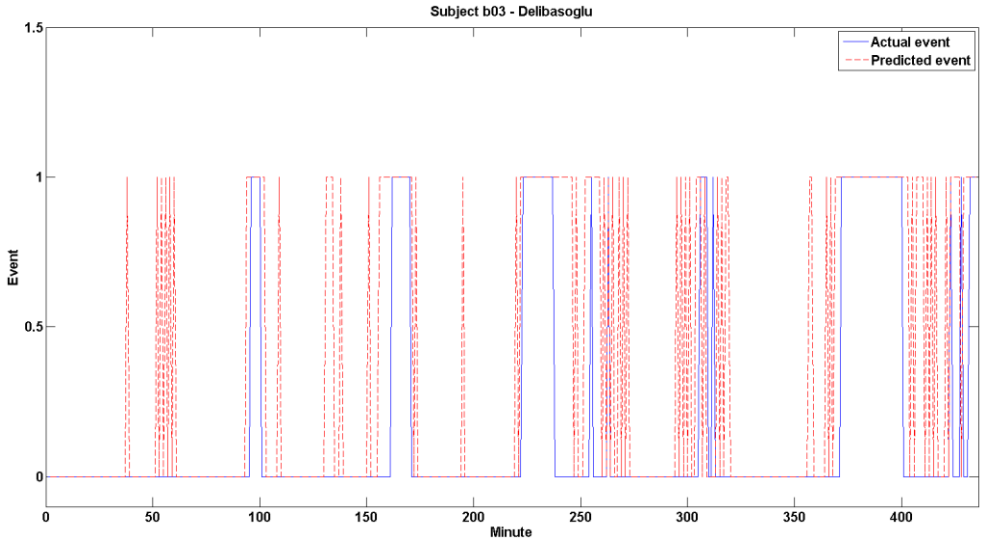


Subject	b02
Method	Mendez et al.
Accuracy	83.01%
TPR	61.29%
TNR	87.83%
FPR	12.17%
FNR	38.71%
Actual apnea minute	93
Predicted apnea minute	108
Actual AHI	11.01
Predicted AHI	12.66

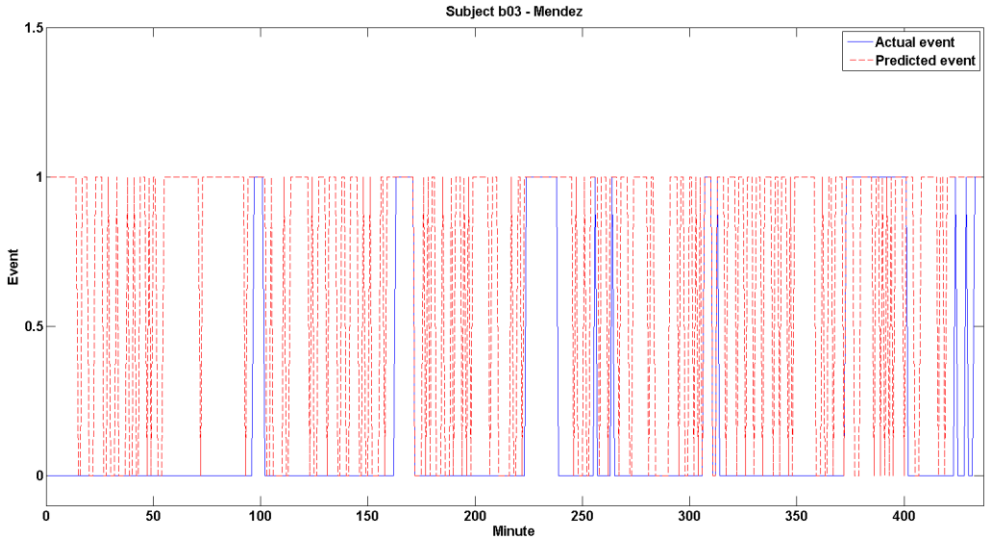


Subject	b02
Method	This study
Accuracy	96.29%
TPR	90.32%
TNR	97.61%
FPR	2.39%
FNR	9.68%
Actual apnea minute	93
Predicted apnea minute	94
Actual AHI	11.01
Predicted AHI	10.89

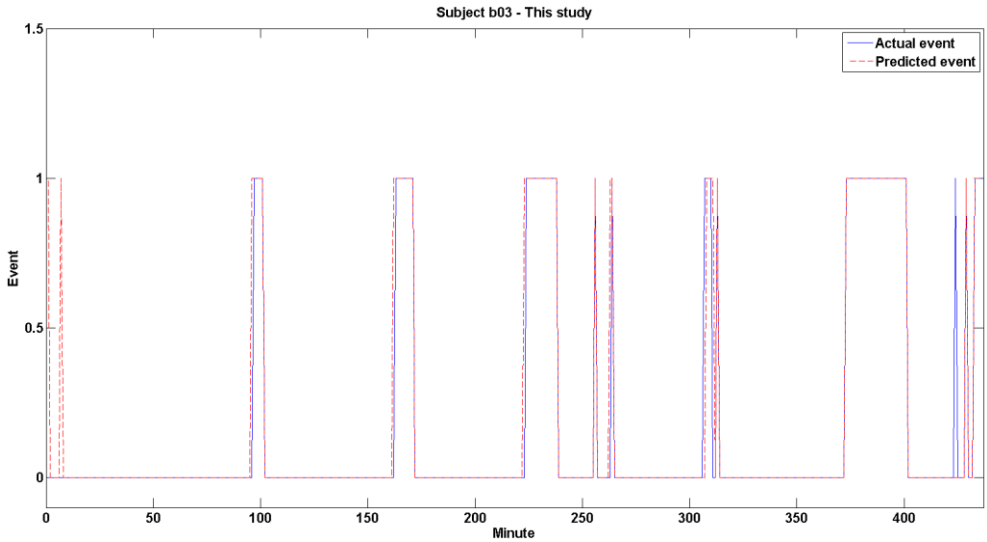
Figure 7.24: Comparison of OSA classification results for subject b02



Subject	b03
Method	Delibasoglu et al.
Accuracy	79.31%
TPR	94.37%
TNR	76.37%
FPR	23.63%
FNR	5.63%
Actual apnea minute	71
Predicted apnea minute	153
Actual AHI	9.77
Predicted AHI	21.1

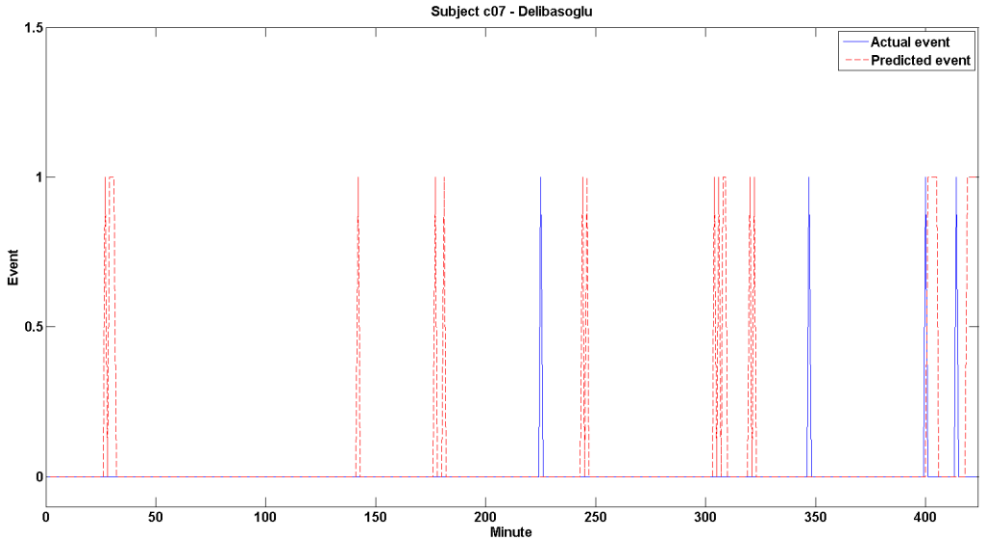


Subject	b03
Method	Mendez et al.
Accuracy	42.89%
TPR	87.32%
TNR	34.25%
FPR	65.75%
FNR	12.68%
Actual apnea minute	71
Predicted apnea minute	302
Actual AHI	9.77
Predicted AHI	41.56

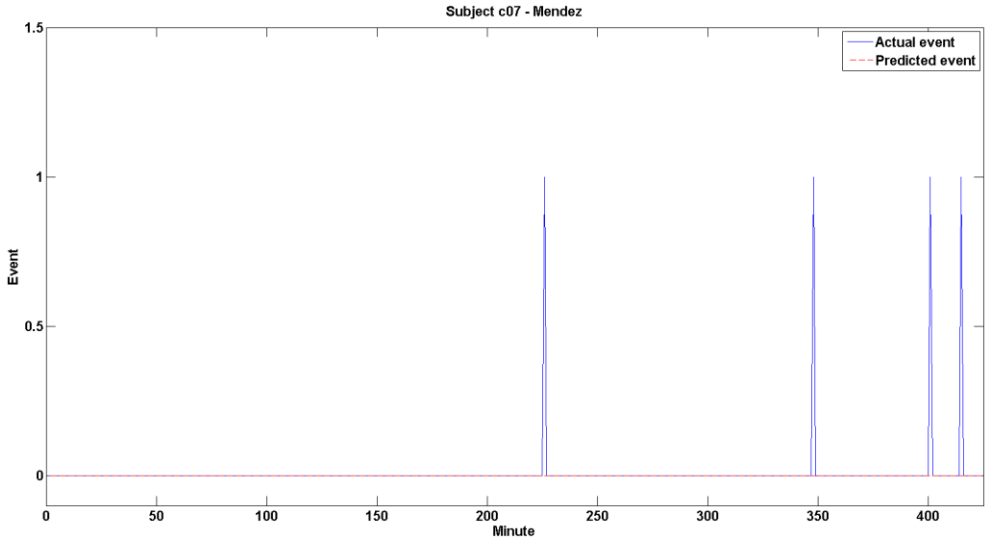


Subject	b03
Method	This study
Accuracy	97.94%
TPR	97.18%
TNR	98.08%
FPR	1.92%
FNR	2.82%
Actual apnea minute	71
Predicted apnea minute	76
Actual AHI	9.77
Predicted AHI	10.45

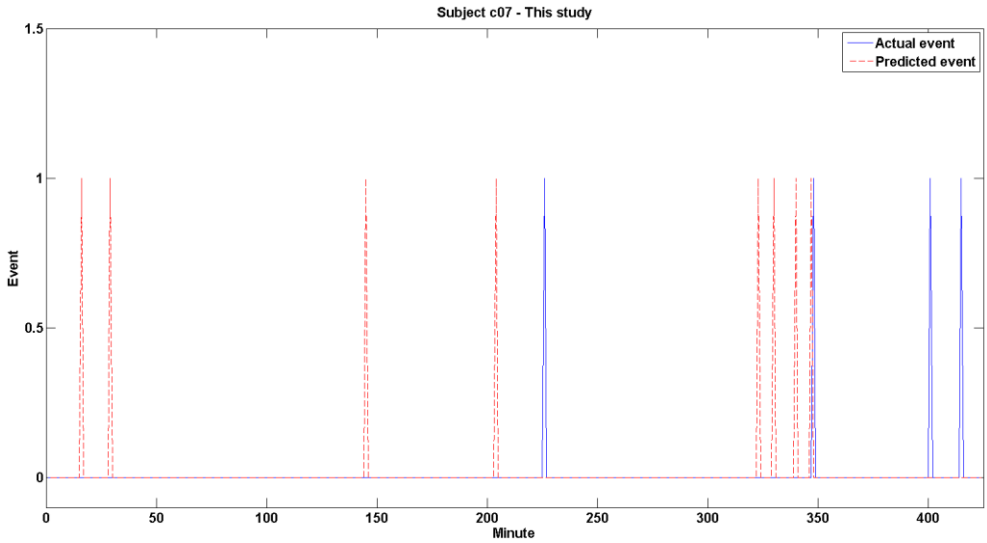
Figure 7.25: Comparison of OSA classification results for subject b03



Subject	c07
Method	Delibasoglu et al.
Accuracy	93.38%
TPR	0.00%
TNR	94.05%
FPR	5.95%
FNR	100.00%
Actual apnea minute	4
Predicted apnea minute	25
Actual AHI	0.57
Predicted AHI	3.55

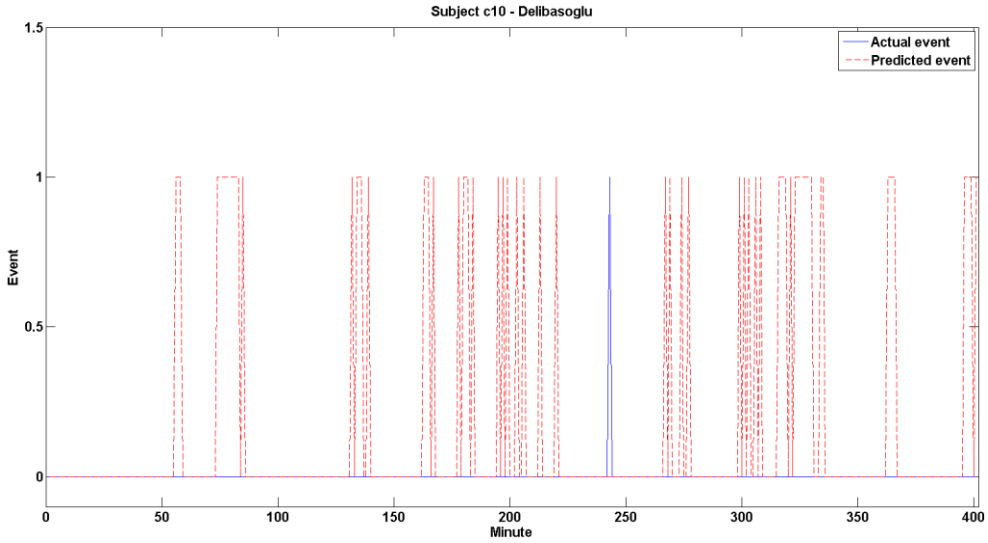


Subject	c07
Method	Mendez et al.
Accuracy	99.06%
TPR	0.00%
TNR	100.00%
FPR	0.00%
FNR	100.00%
Actual apnea minute	4
Predicted apnea minute	0
Actual AHI	0.57
Predicted AHI	0

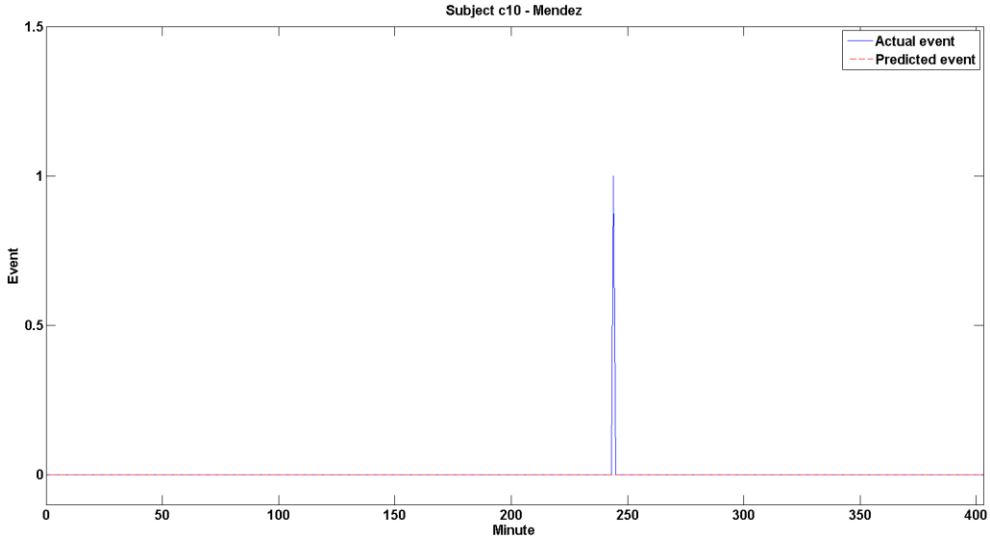


Subject	c07
Method	This study
Accuracy	97.17%
TPR	0.00%
TNR	98.10%
FPR	1.90%
FNR	100.00%
Actual apnea minute	4
Predicted apnea minute	8
Actual AHI	0.57
Predicted AHI	1.13

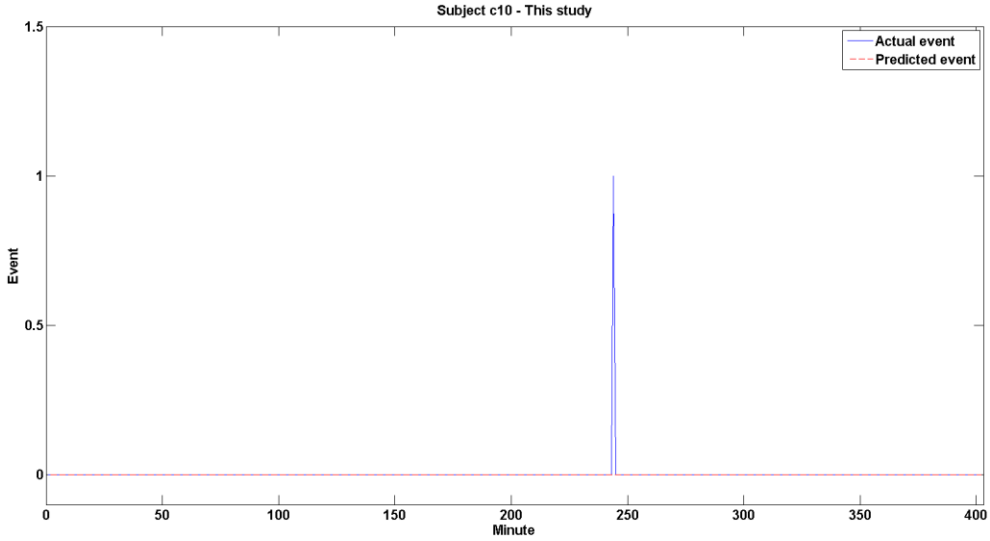
Figure 7.26: Comparison of OSA classification results for subject c07



Subject	c10
Method	Delibasoglu et al.
Accuracy	83.04%
TPR	0.00%
TNR	83.04%
FPR	16.96%
FNR	100.00%
Actual apnea minute	1
Predicted apnea minute	68
Actual AHI	0.14
Predicted AHI	10.17



Subject	c10
Method	Mendez et al.
Accuracy	99.75%
TPR	0.00%
TNR	100.00%
FPR	0.00%
FNR	100.00%
Actual apnea minute	1
Predicted apnea minute	0
Actual AHI	0.14
Predicted AHI	0



Subject	c10
Method	This study
Accuracy	99.75%
TPR	0.00%
TNR	100.00%
FPR	0.00%
FNR	100.00%
Actual apnea minute	1
Predicted apnea minute	0
Actual AHI	0.14
Predicted AHI	0

Figure 7.27: Comparison of OSA classification results for subject c10

Figures 7.21 – 7.27 show the OSA classification results from all three models for each subject individually. The subjects in the apnea group, have a larger number of OSA episodes than the subjects in the other groups. What we should look for is the consistency of the prediction not only in the model's accuracy but also in high TPR and TNR, which means lower FPR and FNR.

For subject a03, although it is quite hard to tell visually which model performs the best, it is quite easy to see which model performs the worst. In this case, we can see that Mendez's model performs the worst from high predictions of both FP and FN events. Surprisingly, for Delibasoglu's model, although it seems the model accuracy is acceptable from Figure 7.21, the FPR is shockingly high, at 48.52%. For overall classification performance (accuracy, TPR, TNR, FPR, and FNR) in this case, our model performs the best in every performance index. In terms of AHI, our model prediction is very close to the AHI calculated from the actual events. However, with lower accuracy, high FPR and FNR, their prediction of AHI is good. Again, because the AHI is calculated from the average number of OSA episodes overnight, wrong predictions can contribute to better or worse results. To tell which model is suitable for the OSA classification application, we should focus on the consistency of the model performances and AHI.

For subject a08 with the most complex OSA episode patterns, our model still performs the best. In addition, our model can follow the complex state-change-patterns very well. The other two models seem to have a problem with the difficult OSA episode patterns, especially, Mendez's model scores with a very high FPR at 63.84%. Finally, our model performs the best with less than one AHI different from the actual AHI. Again, we still can see how consistently each model performs on the data from the apnea group from the results of subject a13.

Moving to the subjects in the borderline group, who have less OSA episodes per night compared to the apnea group, starting from the subject b02, we can see that the models from Delibasoglu and Mendez do not distinctively differentiate the OSA states very well visually. These low performances are also reflected in the very high FPR of Delibasoglu's model (27.68%) and the very high FNR of Mendez's model at (38.71%). Visually and statistically, our model does not have a problem tracking the OSA episode pattern from this subject. Again, for subject b03, the high FPRs in both Delibasoglu and Mendez's models confirm

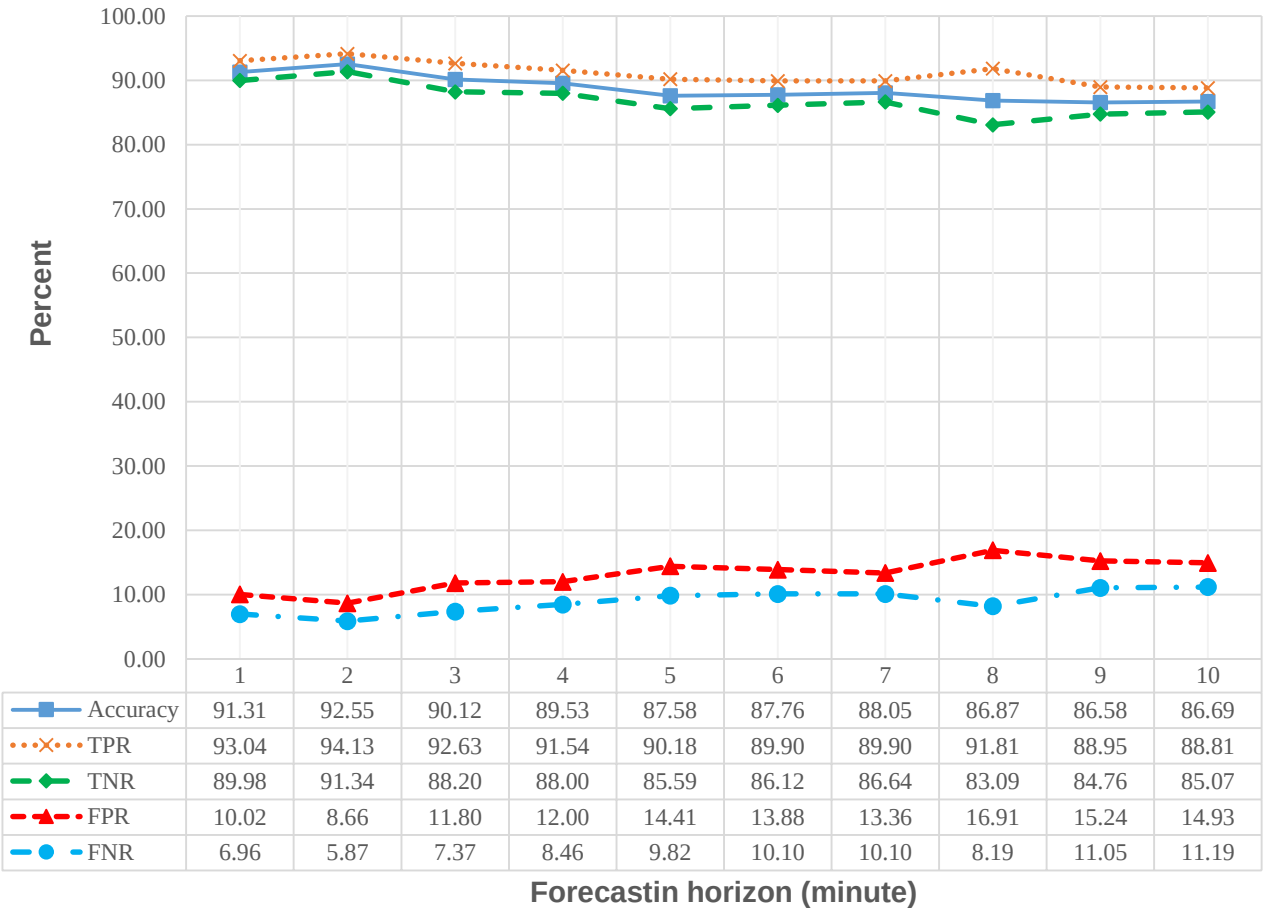
how these models perform in the borderline group. Next, for the control group, Delibasoglu's model suffers from a high FNR, especially when predicting OSA events in subject c10. However, Mendez's and our models can differentiate the non-apnea group very well visually and statistically.

To conclude, in our opinion, Delibasoglu's model is acceptable for classifying OSA events in the apnea group. However, their model suffers when predicting subjects with less sever OSA (few OSA events) as seen from very high FPRs in the borderline and control groups. Mendez's model, seems to perform well in the control group but not in the apnea and borderline groups. Their high FPRs and FNRs make the model performances very inconsistent. In our opinion, human inter-variations cause this inconsistency. For example, an exact baseline heart rate cannot be used to identify a heart's health because each person's build, activity, and other factors are different. We believe that if the normalization process to compensate for these differences is done correctly, these two models will perform better. Finally, the results in all three groups reflect the consistency of our model's performances. This also confirms that the neighboring patterns in the reconstructed phase space of HRV, EDR, and SWS signals can be used to detect OSA events very well and consistently.

7.8 OSA episode forecasting model based on RQA features

In this section, we explain the process for building a model to forecast apnea states in the future time (i.e., at current time t, predict the apnea state at time $(t+1)$ and so on). Our assumption is that there should be significant signs that can be observed at times before t (i.e., $(t-2), (t-1)$) and at time t that tell whether an apnea episode will happen. Thus, our approach is almost the same as for the classification model in the previous sections. The difference is that we shift the target variable (non-apnea = 0, and apnea = 1) of the $(t+i)$ minute (where $i = 1,2,\ldots,10$) to time t. Then, we use the modeling methods explained in Chapter 4 to look for importance features that can be used for classifying the apnea state in that $(t+i)$ minute. Again, out of the three modeling methods, all the best models are from the artificial neural network modeling methods. We also test the effect of the number of past minutes $(t-i)$ that should be included in the phase space reconstruction for the feature extraction process. From the experiments, we found that the

best data partitioning scenario is to include the data from $(t-2), (t-1), \text{and } t$. Including more past minutes (i.e., $(t-3), (t-4)$) or using less data does not improve and also hinders the forecasting accuracy. To emphasize the real-world performance of each forecasting model, in the model training process, we left out a testing partition and used only the training and validation data partitions to train the model (partitioning percentage: 50:35:15). The performances of 10 models for forecasting apnea states at time $(t+i)$ minute(s) (where $i = 1,2,\ldots,10$) in the testing data partition are shown in Figure 7.28 below:



	1	2	3	4	5	6	7	8	9	10
Accuracy	91.31	92.55	90.12	89.53	87.58	87.76	88.05	86.87	86.58	86.69
TPR	93.04	94.13	92.63	91.54	90.18	89.90	89.90	91.81	88.95	88.81
TNR	89.98	91.34	88.20	88.00	85.59	86.12	86.64	83.09	84.76	85.07
FPR	10.02	8.66	11.80	12.00	14.41	13.88	13.36	16.91	15.24	14.93
FNR	6.96	5.87	7.37	8.46	9.82	10.10	10.10	8.19	11.05	11.19

Figure 7.28: The performances of 10 models for forecasting apnea states at time $(t+i)$ minute(s) (where $i = 1, 2, \ldots, 10$)

As expected, the forecasting performances decay with longer forecasting horizons. However, what we did not expect is that the model performance in forecasting the OSA states for time $(t+2)$ is better than the performance for time $(t+1)$ (92.55% > 91.31%). The difference is about 1% which is still very

low. The forecasting accuracy decays slowly with longer forecasting horizons, starting at about 92% at 1 minute ahead and converges to about 86% at 8, 9, and 10 minutes ahead. The true positive rates produced by the model are higher than the true negative rates. This means that the model makes more correct predictions when the targets are equal to one, or the apnea state than when the targets are equal to zero, or the non-apnea state. This result can also be interpreted as the model tending to create more false positive predictions (predicts one but is actually zero) than false negative predictions (predicts zero but is actually one), as seen from about 3% difference, on average. Still, this behavior can be adjusted by changing the model cutoff values as explained in Chapter 4 to fit the application's requirements.

7.9 Important predictors for the OSA episode forecasting model based on RQA features

The top 11 important variables based on the Gini impurity reduction obtained from fitting the DT models with RQA features are shown in Figure 7.29. In this plot, the first column (X-axis) corresponds to the most important variables (left) and less important variables (right) respectively for predicting a one-minute ahead OSA episode (first row of the Y-axis). The next rows (Y-axis) show the trend of the changing patterns of the Gini impurity reduction. The first two variables, which are Shannon entropy and the longest diagonal line of the EDR signal, remain the most important variables (compared to the classification model in the previous section) for deciding whether there will be an OSA episode one to ten minutes ahead. However, what makes these models different from the previously discussed classification model is the patterns of the rest of the features. The recurrence time entropy (RTE) of the EDR signal is not the third important variable; this feature does not have as much influence in forecasting as in the OSA episode classification application. Interestingly, the RR (HRV) entropy has less power to detect an OSA episode with longer forecasting horizons (2 – 10 minutes) and its importance drops to last at the 10-minute forecasting horizon.

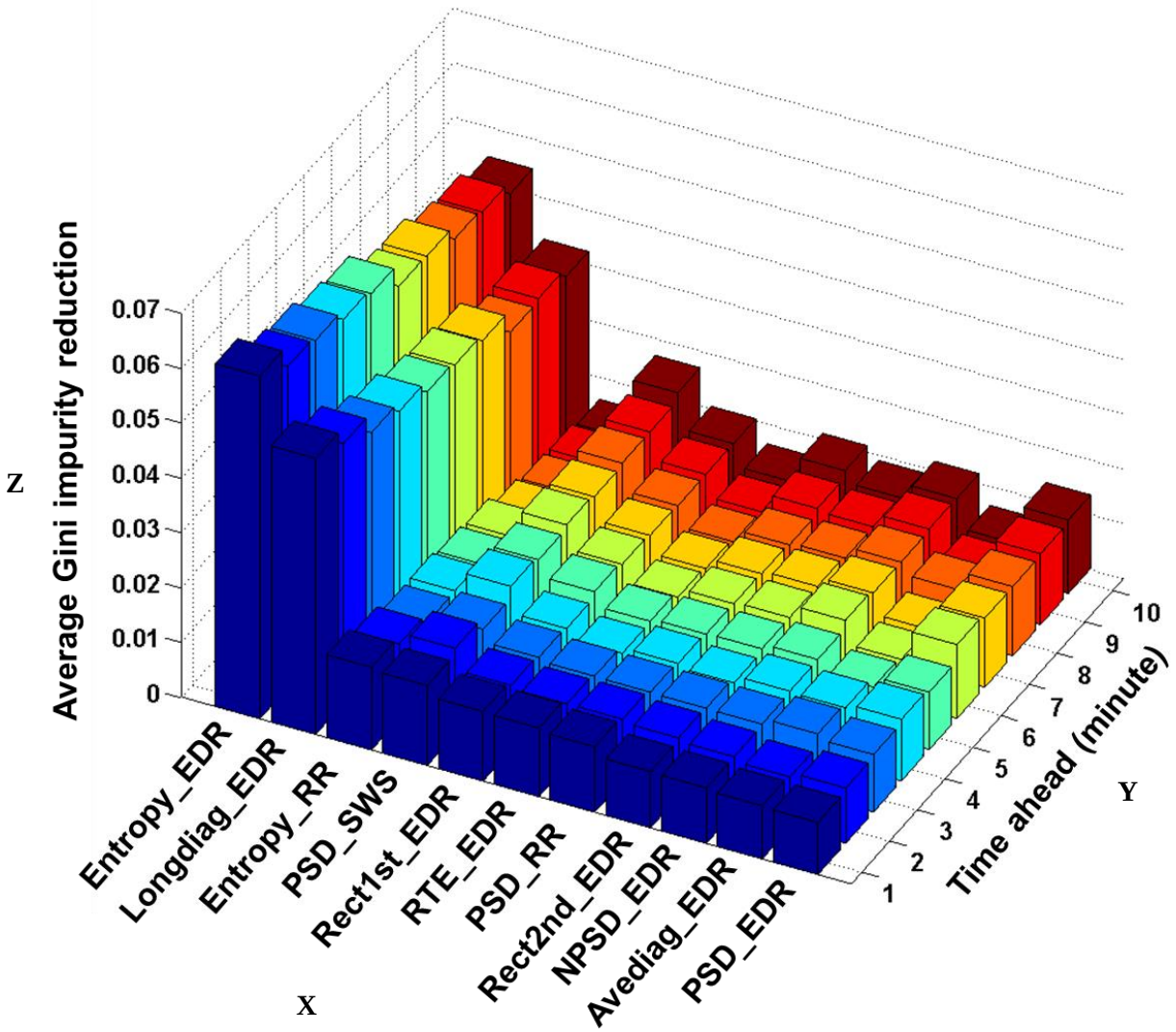


Figure 7.29: Variable importance based on the average Gini impurity reduction for DT OSA episode forecasting models (see full description of each variable name in Table 7.7)

Table 7.7: Full description of each variable name in Figure 7.29 (see Chapter 6 for more information)

Variable name	Full description
Entropy_EDR	Shannon information entropy of all diagonal line lengths of EDR signal
longdiag_EDR	Maximal diagonal line length of EDR signal
Entropy_RR	Shannon information entropy of all diagonal line lengths of RR signal
PSD_SWS	Power spectrum density of SWS signal
Rect1st_EDR	Mean of the recurrence times of the first type of EDR signal
RTE_EDR	Recurrence time entropy of EDR signal
PSD_RR	Power spectrum density of RR signal
Rect2nd_EDR	Mean of the recurrence times of the second type of EDR signal
NPSD_EDR	Normalized power spectrum density of EDR signal
Avediag_EDR	Average diagonal line length of EDR signal
PSD_EDR	Power spectrum density of EDR signal

Another interesting pattern is from the power spectrum density (PSD) of the SWS signal. With the approximately the same Gini impurity reduction as the HRV entropy, this feature's detection power has noticeably increased with the longer forecasting time horizons. The other features that have the same increasing pattern are the EDR signal's normalized power spectrum density (NPSD), and PSD. The remaining features which are the HRV entropy, EDR mean of the recurrence time of the first type (Rec1st), EDR RTE, HRV PSD, EDR mean of the recurrence time of the second type Rec2nd, and EDR average diagonal line length, behave differently with the decreasing trend in the Gini impurity reduction.

CHAPTER VIII

WIRELESS WEARABLE MULTI-SENSORY SYSTEM FOR MONITORING OF SLEEP APNEA AND OTHER CARDIORESPIRATORY DISORDERS

This chapter reports the development of a wireless, wearable, multi-sensor suite for diagnosis of point-of-care cardiorespiratory disorders with a particular focus on sleep apnea screening [150, 151]. The original suite is designed to synchronously gather 3-lead vectorcardiogram signals sampled at 250 Hz from a custom-designed system, heart sound signals gathered at 250 Hz from a MEMs sound sensor (model MT-201 from AD-instrument), digitized oxygen saturation (SPO2) signals at 250 Hz (model PC-60B from St. Alban Medical), and respiratory signals. The signals are transmitted to a near-by processing and visual graphics unit via a Bluetooth module (model RN-800S-CB from Roving Network). Extensive tests suggest that the quality of the signals is comparable to that of commercial and standard recording systems. While this unit offers enhanced functionality in terms of simultaneous echo and phono-cardiorespiratory analysis, the power consumption rates are comparable to those of commercial point-of-care wireless medical devices for cardiorespiratory diagnosis. The system can serve as an out-of-center Type III sleep apnea screening device. This device is highly customizable such that the unneeded sensors could be removed to satisfy the desired application. The ideas and contributions of this chapter are partially from the works of our research group [150, 151]. In this study, we only use the one-lead ECG sensor to corporate with our statistical models reported in Chapter 6 and 7.

Figure 8.1 shows an overview of the hardware module. The arrows capture both power and signal transmissions. All components are powered using two 9 V batteries. The appropriate voltage levels for each component are prepared by the power supply circuit and a voltage regulator. The signals from MEMs sensors are sensed in the form of voltage differences. Normally these signals are confounded with noise such as electrical signals from muscles or electromyography (EMG), movement noise, and even the electromagnetic field from a cell phone. In most cases, the natural frequencies of such noises are not in the same range as those of the target signals. For example, for monitoring purposes, it is acceptable to have ECG's in the frequency range of 0.05 – 40 Hz. In this case, if power line noise at 60 Hz is contaminating our ECG signal, we can use a low pass filter to clean the ECG signal at 40 Hz. Therefore, the signal conditioning circuits are required for eliminating the noises by filtering the signals to their bandwidths and amplifying the signal to achieve a full resolution of the quantization range of an analog-to-digital converter (ADC). Finally, the digitized vital signals are transmitted via a Bluetooth module to a designated receiver that extracts, displays, and stores meaningful data and features. The details of each working block in Figure 8.1 are explained in the following sub-sections.

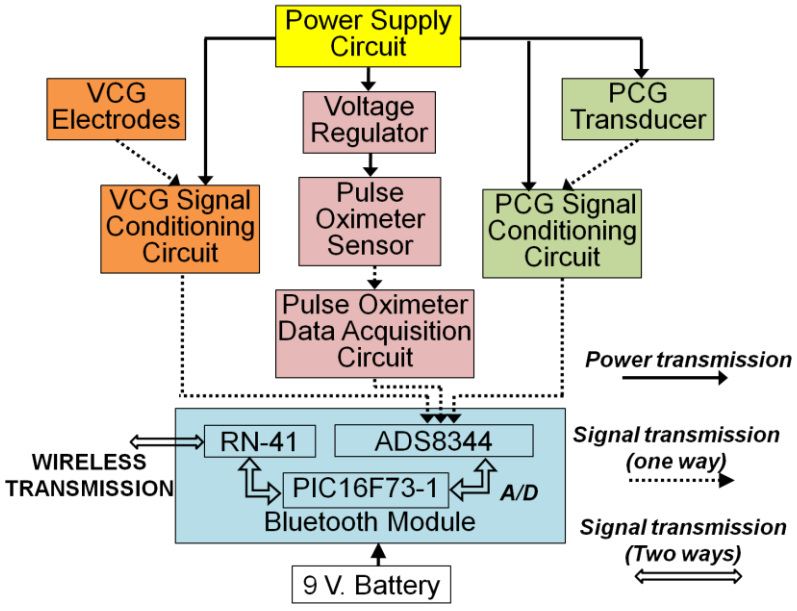


Figure 8.1: Hardware design overview

8.1 Hardware design

8.1.1 Power supply circuits

To determine the power needed for the complete system, the power specification of each component must be evaluated. Commonly, the highest voltage rate needed among the components is determined first in order to specify the power rating of a power supply or a battery, and then the voltage regulator can be used to supply components in the system that require smaller voltages. In this system, the component that requires the highest voltage is the Bluetooth module, requiring 6-12 V. A widely used 9 V battery was chosen. Also, as the specification of an analog-to-digital converter (ADC) in the Bluetooth module requires the input signal to be varied from 0 – 5 V, the signals from the sensors must be amplified to the required range to get the full resolution of the digitizing process. Since the main amplification component is a dual-supply operational amplifier, a virtual ground circuit is created by a power supply circuit (see Figure 1) using an op-amp for dividing the supply voltage from 0 to 9 V to -4.5 to 4.5 V, having a virtual ground at zero as a reference. For a pulse oximeter (SPO2) and its data acquisition circuit, the standard 3.3 V is required. This task is accomplished by implementing a voltage regulator, LM317T, adjusted to supply at 3.3 V.

8.1.2 MEMs-based sensors and conditioning circuits

Primarily, we use a 3-lead VCG sensor, a phonocardiogram transducer, and a pulse oximeter. The VCG sensor is composed of disposable electrodes, leads, and conditional circuits. The type of electrode used depends on the length of time required for each monitoring epoch; the disposable electrode from BIOPAC model EL503 is suitable for short-term general signal monitoring; and for long term VCG data acquisition, a long-term electrode, EL502, and a foam electrode, EL501, are appropriate. These electrodes act as conductors to measure the potential differences according to Frank XYZ leads vectorcardiogram (VCG) at the different positions of the body surface. Compared to the standard 12-lead electrocardiogram (ECG), this system requires only 7 leads for a 3-lead VCG to get the same set of signals by transforming them using a novel method described in [185]. The signals from these leads are fed to the conditional

circuits composed of the filter and amplifier circuits. The frequency range of the VCG signal is about 0.05-100 Hz. However, for monitoring purposes, it is acceptable to use the frequency range of 0.05-40 Hz. An active band pass filter is implemented at 0.05-40 Hz. The signals are also amplified before being fed to the ADC and transmitter circuit.

A second sensor is a phonocardiography (PCG) transducer. This sensor acts as a microphone for converting cardiac-related sounds to an electrical signal. It also captures respiratory, and more specifically snoring sounds, which appear to be one of the major indicators of obstructive sleep apnea. We use an MLT-201 Cardio microphone from ADinstrument. This transducer is designed exclusively to capture heart sounds at frequency 10 – 600 Hz. However, the transducer is very sensitive to low frequency noise (less than 20 Hz) such as that from body movements and breathing. Also, to capture the first (S1) and second (S2) components of the heart sounds, it is sufficient to capture the signal in the range of 20 – 120 Hz. Thus, a low pass filter with a cut-off frequency at 20 Hz is employed to eliminate the low frequency noise captured by the transducer. Since we sample the data at 250 Hz, a high pass filter with a cut off frequency at 120 Hz is used to prevent the aliasing in the signal due to under-sampling. Then, this signal is amplified to an appropriate level to feed to the next module. The pulse oximeter used in this system is a commercial unit, PC-60B. We designed the circuit to acquire the percentage of arterial oxyhemoglobin saturation (SPO_2) directly. The digital signal from the sensor (blood oxygen level) is converted back to an analog signal and conditioned to the fixed ranges of interpretable levels. All signals are shifted above ground level to comply with the Bluetooth module's specifications. Then, we feed the conditioned VCG, heart sound, and SPO_2 signals to the ASD83344, which is an analog-to-digital converter (ADC), which converts them into a serial format and transmits them using a Bluetooth protocol.

8.1.3 Analog-to-digital converter and Bluetooth transmitter/receiver

We chose a Bluetooth protocol because of its high throughput (3 Mbps Maximum), security features, and universal standard. Most people are familiar with Bluetooth protocol and possess devices that are equipped with the Bluetooth technology. For this unit, the commercial platform BlueSentry Bluetooth

wireless transmission module RN-800S-CB fits our requirements for converting the analog signals to digital signals and wirelessly transmitting the data using a protocol with programmable features. The core component in the module that deals with this task is Texas Instruments ADS8344. With its configurable 8 channels single-ended or 4 channels differential input, this module is capable of digitizing the analog signals at 16 bit resolution with overall sampling rate of 3 kHz. The input voltage range is set to be 0-5 V as an industrial standard. After the input signals have been digitized, they are transmitted to the Bluetooth transceiver, RN-41 Class 1 Bluetooth module, using a universal asynchronous receiver/transmitter (UART) protocol.

The controller for the whole module, PIC16F73-1, stores firmware and a command list for a user to be able to program features such as transmission power, transmission speed, sampling rate, and a power consumption scheme. This module supports Bluetooth specifications up to version 2.1 + Enhanced Data Rate (EDR) with full backward compatibility with previous versions, 2.0, 1.2, and 1.1. We can program this module for suitable baud rate speeds varying from 1,200 bps to 921 Kbps. With a class 1 radio, ideally, the communication range is up to 100 m, using maximum output power over the air at 15 dBm with -80 dBm receiver sensitivity. However, the maximum output is rarely used since the communication range is limited by the application. For monitoring purposes, the range is only within a 10-meter radius at most. This range should be initially determined for the optimal configuration of the output power for two main reasons: to conserve the battery and to limit the communication range for security reasons. Other power management techniques such as Sniff mode and Deep sleep are also available. Basically, they are methods to adjust the optimal active time and shut-down time of the radio to be suitable to the applications.

8.1.4 Default performance of the multi-sensor platform

Power consumption: A power supply and a known value resistor are used to form a circuit to measure the power consumption of the entire system. Since the circuit is connected in series, the current (I) passes through a known value resistor (R_t) and the system is the same regarding Ohm's law. The magnitude of the current depends on the maximum resistance of the components in a circuit. In this case, the

measurement resistor (R_t) used is 0.7 ohm, which is smaller than the resistance of the system, so the measured current is contributed solely from the system. A high-precision Fluke multimeter and oscilloscope is used to measure the voltage drop across R_t (V_t); then the current consumption of the system in different working stages is calculated by V_t and R_t (see Table 8.1). As indicated in Table 8.1, with all sensors powered (system on) and continuous data transmission, the current used by the overall system is around 169 mA. In the current setup, we use two 9 V lithium batteries rated at 1200 mAh each. This system can work continuously for approximately 14 hours (more than adequate for a normal 8-9 hours of sleep).

Table 8.1: Power consumption of the system in different working stages

System	Bluetooth status	Current drawn (mA)
On	Discovery mode (not connected)	100
On	Connected to a computer	149
On	Connected and continuous transmission	169

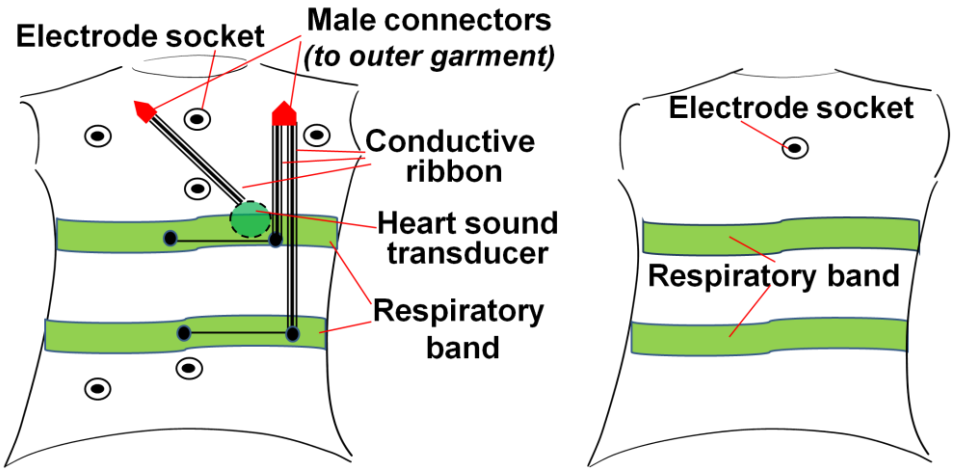
Communication interfacing and transmission test: The module is ready for a connection about ten seconds after it is turned on. In our case, we use Matlab both for initiating communication and real time analysis. Theoretically, the communication range for a class 1 Bluetooth module is up to 100 m. However, we found that the range also depends on the stage of communication. For the discovery mode, the module could be discovered within a range of 10 m, but that range will not work for the pairing mode. The pairing mode requires a maximum range of 2 m. Also, after the pairing procedure is done, 1 m is the maximum range for opening the communication port to initiate a connection. Once the port is opened, this range can be increased up to about 10 m for effective data transmitting/receiving. The additional characteristics of our system are summarized in Table 8.2.

Table 8.2: Default hardware performance

Performance	Maximum
Overall sampling rate	3000 Hz
Communication range	10 m
Initiation time	10 – 15 seconds
Operation time (continuously)	14 hours

8.2 Hardware integration to a garment

From an engineering standpoint, the design of our wearable unit consists of two pieces: first, an inner garment provides housings for MEMs sensors that will be in contact with the user's skin, and second, an outer garment contains all circuit board housings and the electrical connections from all sensors to the circuit boards (see Figures 8.2 and 8.3). The detailed design and tailoring of the garment was carried out by a team of researchers [22] from human sciences and health sciences departments in addition to the authors.



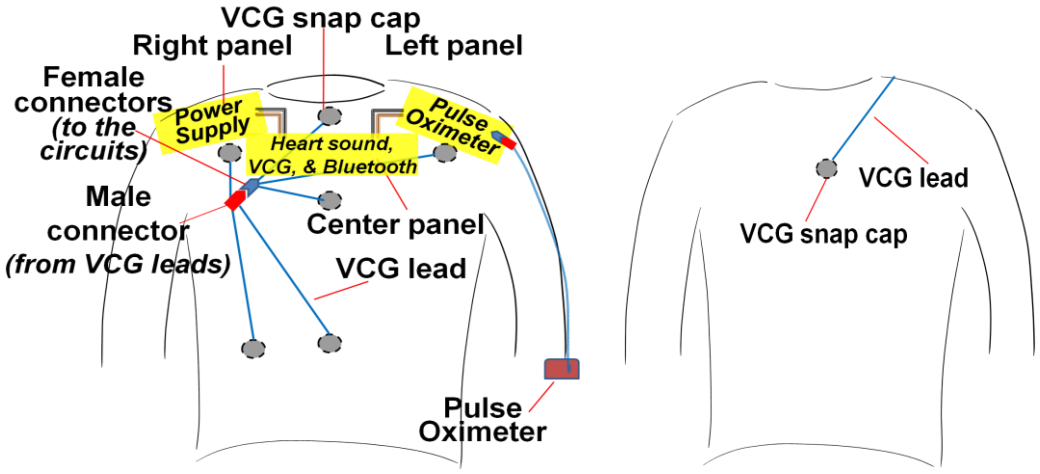
(a) Front side **(b) Back side**

Figure 8.2: Inner garment configuration

8.2.1 Inner garment electronics

Since the VCG electrodes need to be attached to the user's skin, we created sockets at the positions corresponding to the orthogonal 3-lead VCG as shown in Figure 8.2. Users who do not have a background in VCG systems can put electrodes in the correct positions every time. Next, two stripes, shown in Figure

8.2, are housings for respiratory bands. Reusable respiratory inductance plethysmography (RIP) bands from Embla are inserted into these housings. The electronic circuits for this sensor are currently under development and should be ready for the next generation of the garment to precisely collect the respirations from the abdomen and rib-cage. However, the respiratory signal is available in the current system using the electrocardiogram derived respiratory (EDR) method. The upper band also serves another purpose: immediately under the band, around the tricuspid region (near the left lower sternum and opposite the fourth and fifth costal cartilages), is a small pocket to hold the PCG transducer. Pressure from the band will keep the PCG transducer attached to the skin to ensure sufficient signal quality from the transducer.



(a) Front side **(b) Back side**

Figure 8.3: Outer garment configuration

8.2.2 Outer garment electronics and conductive ribbons

Signals are transmitted mostly through multiple three-channel conductive ribbons. They are made from a conventional fabric strip sewed with conductive thread, creating an electrical conducting channel from one end to the other. In our case, the ribbon contains three conductive channels. These conductive ribbons are sewed onto the inner layer (inside) of an outer garment and lead to the circuit board area at the user's chest. We use only the middle channel to transmit the signal and ground the other two to suppress the noise that these conductive threads might pick up from the air. Both ends of each ribbon are fused to the connectors. Twelve-pin and three-pin Molex connectors are used at the circuit board end. In some areas, the ribbons are covered by fabric to prevent them from touching each other. At the sensor end, snap buttons

(male) are sewed onto the inner garment and the female buttons are fused with the middle channel of the conductive ribbons. The positions of the snap button pairs (male and female) of both inner and outer garments are aligned exactly to ensure the correct connections when the garment is put on. Thus, each line leads from each sensor to the corresponding connector at the main circuit board.

A twelve-line ribbon cable (multi-wide planar cable) is used for transmitting each signal to its matching circuit board for an SPO_2 sensor. The line runs from a circuit board located at the chest area through a fabric channel on the left sleeve to an SPO_2 sensor at a user's left index finger.

8.2.3 Wireless-wearable one-lead ECG collecting device

One important feature of our device is that it is highly customizable. Depending on the user's objective, he/she can choose to use only the sensors that fit the applications appropriately. For this study, because the signal of interest is the ECG, we modify our device to collect only one-lead ECG to fit the objective. This further reduces the power consumption and the device's footprint such that it can be encased in any soft-shell material that is suitable to wear, as shown in Figure 8.4. The device is worn using stretchable material for the arm-band so that the user can adjust the tightness as needed. We have also tested this device on ourselves and external subjects, and found that it was an appropriate location for the sensor to be worn overnight as it has little or no effect on the normal sleep process.

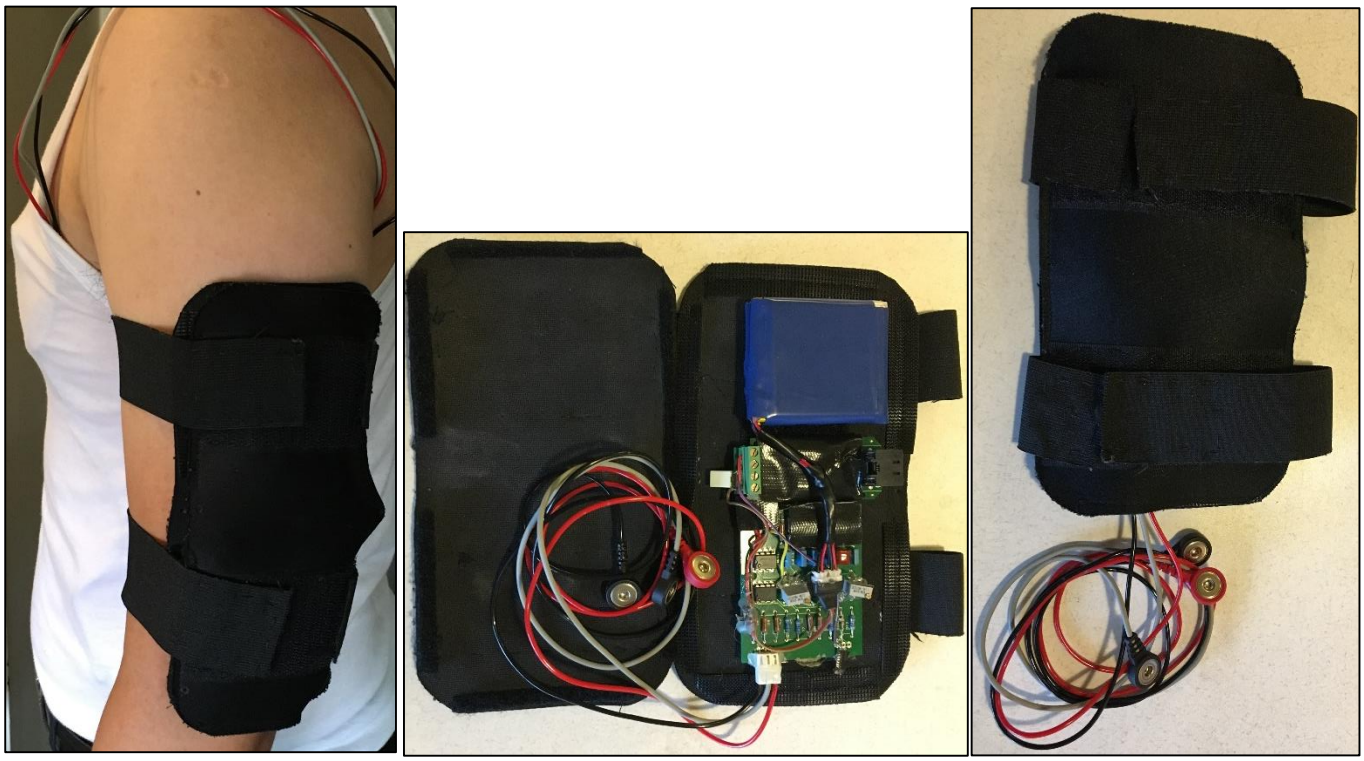


Figure 8.4: Wireless-wearable one-lead ECG collecting device

8.3 Overall performance

To validate the results of our system, we used a LifeShirt system from Vivometrics [18] to obtain VCG signals and respiratory signals as benchmarks for evaluation from 20 subjects. As for the Lifeshirt system, it is equipped with only a 1-lead ECG, and hence does not comprehensively capture the underlying cardiorespiratory dynamics. Also, in order to validate our system, the Lifeshirt system needs to collect each lead individually, corresponding to the X, Y, and Z leads of our system. For heart sounds, the commercial device ds32a+ stethoscope from Thinklabs was used. Comparisons of results are shown in Figures 8.5 through 8.9.

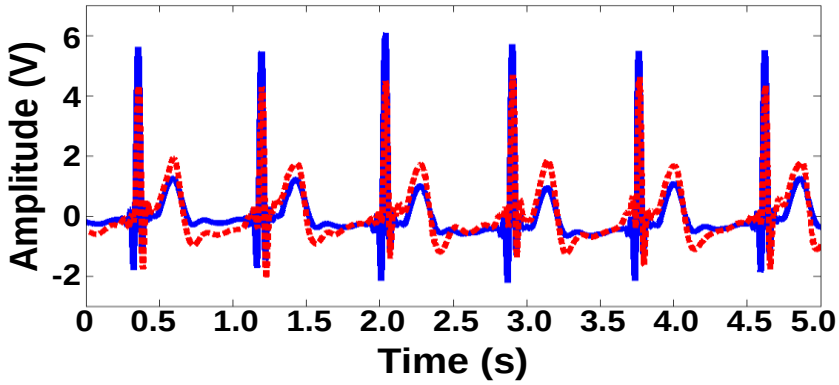


Figure 8.5: Verification of VCG lead X

VCG lead X from Life shirt system

VCG lead X from our system

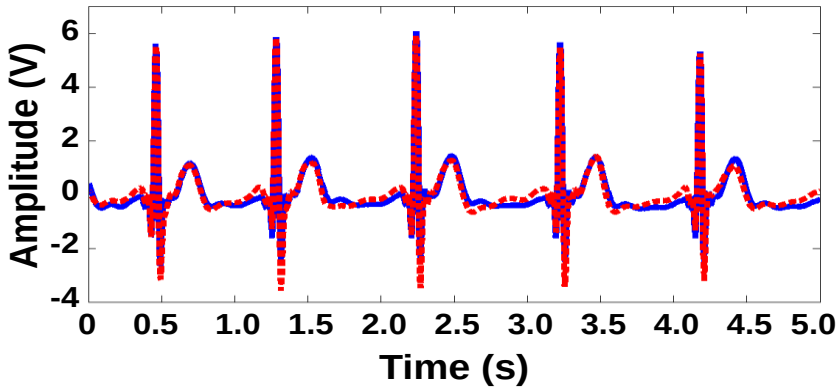


Figure 8.6: Verification of VCG lead Y

VCG lead X from Life shirt system

VCG lead X from our system

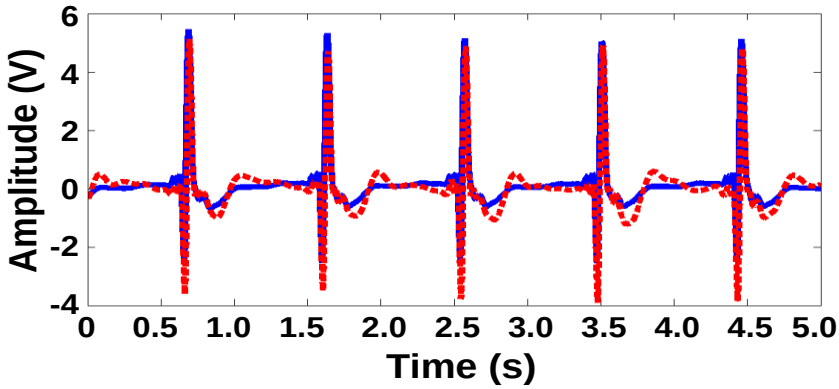


Figure 8.7: Verification of VCG lead Z

VCG lead X from Life shirt system

VCG lead X from our system

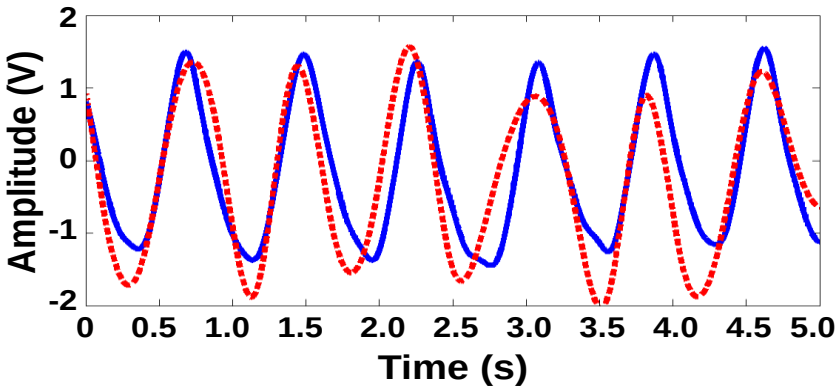


Figure 8.8: Verification of respiratory signal

Respiratory signal from LifeShirt system

EDR signal from our system

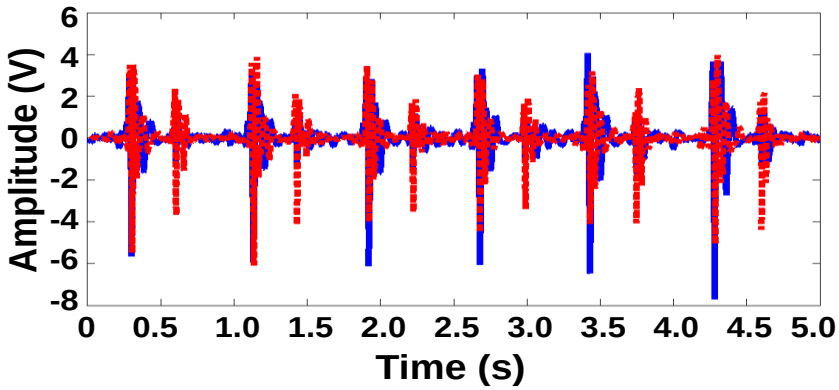


Figure 8.9: Verification of heart sound (HS) signal

HS from ds32a+ Stethoscope

HS from our system

As shown in Figures 8.5, 8.6, and 8.7, the VCG (ECG) signals from both systems closely resemble each other. The slight differences in the signals may come from the different designs in the analog front end circuits of the systems. Also, the electrodes' positions are slightly different since we have to use two sets at the same measuring positions. The important reference points in the time domain, the Q, R, and S complex, are perfectly aligned. The average Pearson correlation coefficient for VCG lead X, Y, and Z signals (acquired for a duration of 5 minutes) between the two systems were 0.7684, 0.743, and 0.6969, respectively. More prominently, the signals from the wearable multi-sensor suite portray the various waveform features more vividly than those from the commercial device.

For the validation of respiratory signals, we compared the signals obtained from on RIP band from a LifeShirt system at the abdomen position with the ECG-derived respiratory signal from our system. As shown in Figure 8.8, the two respiratory signals are almost the same. The inhale-exhale patterns are clear. The Pearson correlation coefficient between them averaged 0.6915 (length of 5 minutes).

Next, we used a commercial digital stethoscope model ds32a+ from Thinklabs as a benchmark medium for the heart sound signal. As seen in Figure 8.9, the amplitudes of the two signals are slightly different. The precise position of the transducer is critical for measuring the heart sounds. With the large surface area of both the stethoscope and transducer, positioning them at the same measuring spot could cause a difference in signal quality. However, S_1 and S_2 heart sounds from both sensors are perfectly aligned. The average Pearson correlation coefficient between them is 0.4031. For the pulse oximeter used in our system, as reported in the previous section, we used a "reverse engineering" approach based on “tapping” the blood oxygen level shown on the pulse oximeter’s screen and feeding the signal directly to the Bluetooth module. The blood oxygen level from our system monitor and the one on the device’s screen are similar.

This chapter presents the design and testing of a novel wireless, multi-sensor, wearable, device that can serve as a type III OSA screening platform to be used for OOC testing. The device has the ability to measure, display, and store 3-lead vectorcardiogram (VCG), 12-lead electrocardiogram (ECG), heart rate (HR), respiration signal, heart sound, and blood oxygen saturation level (SPO_2) simultaneously and synchronously. All the salient patterns of signals received from the system are comparable to those from a commercial product, Lifeshirt [92]. Experimental investigations also suggest substantial correlations (Pearson correlation coefficient > 0.7) between the waveforms from the device and those from the Lifeshirt system. The standout feature of the device is that it combines (a) the capability to wirelessly acquire good quality multi-channel signals vital for analysis of sleep disorders and cardiovascular diseases at a high sampling rates, (b) ease of integration into a garment, and (c) long battery life and data capacity that allow the device to operate through an entire night. These features make this wireless wearable system highly suitable to be used as an OSA device for OOC testing and cardiovascular system monitoring applications.

To conclude this chapter, we have thoroughly explained the detail of developing a wearable wireless sensor which is designed to collect important biology signals overnight. Our device is also highly customizable. This means that we can pick and choose the sensors to be used in different applications appropriately. For this particular study, we customized our device to collect only the one-lead ECG data (see Figure 8.9) which we show that it can be used for diagnosis of an OSA. With the low cost of development, wearability, and self-enabling operation, if used with an automatic OSA episode prediction tool such as the statistical models that we proposed in this study, it can be a self-OSA diagnosis tool that might be widely be available on the shelf for anyone to easily obtain.

CHAPTER IX

CONCLUSION AND FUTURE WORK

9.1 Contributions

This dissertation presents three statistical models based on data mining and nonlinear time-series analysis techniques as an alternative method for the diagnosis and treatment of obstructive sleep apnea (OSA) disease. In general, our contribution from an application perspective is that our new method substantially reduces the time and cost associated with the conventional method by first screening non-OSA subjects out of the population with a low false negative rate and reasonable accuracy (FNR: 5.3%, and accuracy 84.47%), then individually determining the OSA severity by utilizing the data from a single-lead electrocardiogram (ECG) device that is worn overnight at the subject's location with an accuracy comparable to the sleep study or PSG (accuracy: 92.26%). Our method for forecasting the OSA episode can be used to activate an OSA therapy device such as a continuous positive airway pressure (CPAP) machine or a hypoglossal nerve stimulator as needed or before an OSA episode so that the latter can be averted in real time. In particular, we list the contributions of this research as follows:

1) Detect the existence of OSA in an individual based on the pattern of biological, physiological, and simple clinical data with a low false negative rate (FNR: 5.31%, and accuracy: 84.47%). People with some degree of probability of having OSA are confirmed by the next model.

2) Determine the OSA severity by classifying the OSA episode (event) from only one-lead ECG data collected overnight (accuracy: 92.26% with total of 10,052 equally sampled events from 24 subjects).

The advantage of our model is that the variations (i.e., different body build, age, gender, activity, health conditions, and race) have very little effect on the model because the neighboring patterns in the reconstructed phase spaces have very little or no correlation to those variations. This benefit can be seen from our model's performance compared to two other models that exist in the literature.

3) Forecast the incoming OSA episode in real time using the one-lead ECG data (accuracy: 92%, 88%, and 87% for 1, 5, and 10 minutes ahead). This forecasting model with any appropriate OSA episode prevention device (i.e., HNS, and just-in-time CPAP) will allow for an effective OSA treatment method for CPAP nonadherence OSA sufferers.

4) Develop a wearable device that can collect the biological data via a single-lead ECG as a home sleep test (HST) device. This device is intended to be used with the models mentioned in contributions 2 and 3.

The research methodologies include machine learning techniques (decision tree, logistics regression, and artificial neural network), wavelet decomposition, nonlinear time-series analysis based on the recurrence plot (RP) of the phase space reconstruction, and recurrence quantification analysis (RQA). The development of the first model (screening model) especially emphasizes the solutions for dealing with missing values and their effects in the classifier accuracy, the method for selecting the significant variables, the performance and advantage of the machine learning techniques, and the method for fine-tuning the classifier for the desired model performance (i.e., low FNR or FPR). The advantage of this model is the power to confidently reject (screen) subjects without OSA disease out of the focus group with a low false rejection rate (FNR<5%). People with some degree of probability of having OSA will be confirmed by the next model, which requires the collection of the ECG signal overnight.

In the development of the second model (OSA episode prediction), we highlight the nonlinear decomposition technique (wavelet transformation) and other digital processing processes to maximize the information of the OSA symptoms from the one-lead ECG (comparing to more than 20 signals from PSG), resulting in three predictor signals; heart rate variability (HRV), ECG derived respiration (EDR), and slow wave signal from wavelet decomposition (SWS). Then, to unfold the nonlinear patterns from those signals,

we use phase space reconstruction and RP. The advance statistical quantification technique called recurrence quantification analysis (RQA) is used to translate the important patterns of the RP to numbers such that the machine learning techniques can differentiate the OSA states. The advantage of using this combination method is already mentioned in the contribution 2 above. Moreover, aside from the contribution in the statistical perspective, the combination of our OSA screening model, OSA episode prediction model, and one-lead wearable sensor, enable the out-of-center OSA diagnosis that is economical, and convenient (easy to self-operate, and at home), with a comparable performance with the sleep study (PSG).

In the third (OSA forecasting) model, we track the changing patterns of RQAs from HRV, EDR, and SWS over time and show that they are good predictors of an incoming OSA episode for i minute(s) ahead ($i = 1,2, \dots ,10$). This forecasting model with any appropriate OSA episode prevention device (i.e., hypoglossal nerve stimulator, just-in-time CPAP) will allow for an effective OSA treatment method for CPAP nonadherence OSA sufferers.

9.2 Conclusion

In this research, we developed an automated OSA diagnosis system that can help develop new treatment systems. This diagnosis system consists of three elements, an OSA screening model, an OSA's episode classification model, and a wireless wearable sensor for collecting one-lead ECG data. We applied several data mining techniques introduced in Chapter 4 to develop a screening model in Chapter 5 that eliminates people without moderate to severe OSA from the population so that we can focus exclusively on people who have moderate or severe OSA. This is different from the traditional approach where a sleep doctor will use a set of questions to screen people who may have OSA out of the population for a PSG. The reasons behind our approach are as follows. 1) The traditional screening questionnaires were developed from a very small sample in a specific demographic region. It may not be appropriate to use them as the main tool to screen people with OSA from the entire population because of human physiological differences. Unless the person shows strong evidence of having OSA, he or she will not be referred to a

sleep doctor to do a PSG for further diagnosis. This traditional approach is not suitable for early OSA detection. 2) The individual and direct evidence, which is the number of OSA episodes occurring overnight, should be the main piece of information to be used as a tool to diagnose OSA. In the traditional approach, because the PSG requires extensive time and labor, the ability to diagnose a sufficiently large number of people who may have OSA for early detection of the disease is difficult. With the use of the device explained in Chapter 8 and the model to predict OSA episodes from the signal collected from it, the ability to self-diagnose the existence of OSA for a much larger population can be achieved.

In Chapter 7, we have thoroughly explained all methods used for 1) constructing an OSA episode classification model, and 2) constructing an OSA episode forecasting model (predicting an incoming episode). The theory and methods required to achieve such tasks are clearly reported in Chapter 7. We first explain how we extract three more time-series that correlated to OSA symptoms from the ECG signal, namely heart rate variability (HRV), ECG-derived respiration (EDR), and slow wave signal (SWS). These three extracted signals correspond to the cardiovascular system, respiratory system, and autonomic nervous system (ANS), which are directly and indirectly influenced by OSA state variations or breathing and non-breathing patterns during sleep.

The basic idea for building these two models is based on the reconstructed phase space of these signals. Because these signals are highly nonlinear in nature, the observed patterns from the original time-series are influenced by not only the OSA state but also many other systems. Reconstructing the phase spaces of these signals onto the higher (sufficient) dimensions reveals the trajectories that correspond to the changing patterns in the state space. In the other words, having enough state variables to explain how the state of each signal changes over time reveals the true and less complex trajectories compared to each original time-series. Once each signal pattern is completely unfolded and revealed, we can use the machine learning technique to distinguish the patterns which relate to the apnea or non-apnea state.

However, because of variations in age, gender, body build, health conditions, and race, the quantities calculated directly from each time-series such as heart rate and respiratory rate may not be appropriate for use in the general population. The quantities calculated from the neighboring patterns from

the states or points in the reconstructed state space of each time-series prove to be firm measures that are not influenced by the previously mentioned variations. To test our assumption and the OSA episode classification model performance, we compared our model results with two those of other methods in the literature. The results suggest that our method is accurate and suitable to be used in people with different apnea conditions.

Finally, our classification model, along with the wearable wireless sensor device introduced in Chapter 6, can be further refined to develop an economical and reasonably accurate device for a home sleep test (HST) which can help undiagnosed and suspected OSA patients and health enthusiasts. In addition, pairing our forecasting model with the device allows for automatic intervention so that an impending OSA episode can be averted without waking the patient up. The potential use of this forecasting model is not limited to this application. For example, it may also be used to improve the continuous positive airway pressure (CPAP) therapy such that the pressure build-up in the system can be maintained on a low level while there is no apnea. Then, if the incoming apnea episode is forecasted, the device activates itself to build up the pressure to prevent the apnea episode from happening. This will improve the user's comfort leading to more adherence to the therapy.

9.3 Future work

In our approach to uncover the underlying dynamics of the time-series, we have currently reconstructed the phase space, then quantified each windowed signal using recurrence quantification analysis (RQA). Then, we apply the machine learning techniques to learn the patterns to differentiate between the two states, apnea or non-apnea. However, using all 39 features from a combination of the RQA and PSD features from three time-series requires a substantial calculation power to compute these features and train the model. For the RQA calculation, with 16 threads of the OSU HPCC Cowboy supercomputing node (dual Intel Xeon E5-2620 "Sandy Bridge" hex core 2.0 GHz CPUs, with 32 GB of 1333 MHz RAM), it took about an hour to process each person nightlong HRV, EDR, and SWS time-series to the corresponding RQAs. Furthermore, this processing took a normal computer (AMD Phenom II X6 1045T

hex core 2.7 GHz CPU, with 8 GB of 1333 MHz RAM) more than 12 hours to do the same task. One way to address this problem is to use a dimension reduction technique to reduce the number of variables used in the model training process. We have already tried to use some techniques such as a principle component analysis (PCA) and forward and backward variable selection, but the results are not satisfactory. This can be one of the directions for future study.

Furthermore, rather than calculating the RQA from each signal's recurrence plot (RP) individually, a cross recurrence plot (CRP), which simultaneously shows the states in the trajectory of one dynamic system that visits roughly the same space of another dynamic system, could reveal more interactions between the two systems (e.g., HRV and EDR). This may help improve not only the classification and forecasting accuracy but also reduce the computation time (calculation of only one CRP rather than two RPs).

Next, although we have successfully shown the ability to classify and forecast the OSA episodes using only one-lead ECG, there are still gaps that can be further improved, especially in the introduced methods and models. Because of current capability and limitations, we are currently building models that can classify and forecast the OSA episode in one minute. By definition, an OSA event is counted when the symptoms last at least 10 seconds. This is one of the reasons why our classification model tends to overestimate the AHI. Thus, a shorter predicting resolution may be the next goal for us to improve our models to be more accurate and realistic.

Finally, with the real-time data collection and processing capability of our device and using the OSA episode forecasting model, intervention measures can be deployed so that an OSA episode does not occur. However, these intervention measures will need to be closely investigated because they will have to be implemented and tested on real subjects. A partnership with the sleep doctors is required.

REFERENCES

1. National Heart Lung and Blood Institute. *What to Expect During a Sleep Study*. 2012 [cited 2015; Available from: http://www.nhlbi.nih.gov/health/health-topics/topics/slpst/during.
2. Johnteslade. *Wavelet Filter Bank*. 2005; Available from: https://en.wikipedia.org/wiki/File:Wavelets_-_Filter_Bank.png.
3. Malhotra, A. and Loscalzo, J., *Sleep and Cardiovascular Disease: An Overview.* Progress in Cardiovascular Diseases, 2009. **51**(4): p. 279.
4. Leung, R.S.T. and Bradley, T.D., *Sleep Apnea and Cardiovascular Disease.* American Journal of Respiratory and Critical Care Medicine, 2001. **164**(12): p. 2147-2165.
5. Ip, M.S.M., Lam, B., Ng, M.M.T., Lam, W., Tsang, K.W.T., and Lam, K.S.L., *Obstructive Sleep Apnea Is Independently Associated with Insulin Resistance.* American Journal of Respiratory and Critical Care Medicine, 2002. **165**(5): p. 670-676.
6. Marin, J.M., Carrizo, S.J., Vicente, E., and Agusti, A.G., *Long-Term Cardiovascular Outcomes in Men with Obstructive Sleep Apnoea—Hypopnoea with or without Treatment with Continuous Positive Airway Pressure: An Observational Study*. The Lancet, 2005. **365**(9464): p. 1046-1053.
7. Shahar, E., Whitney, C., Redline, S., Lee, E., Newman, A., Javiernieto, F., O'connor, G., Boland, L., Schwartz, J., and Samet, J., *Sleep-Disordered Breathing and Cardiovascular Disease: Cross-Sectional Results of the Sleep Heart Health Study.* American Journal of Respiratory and Critical Care Medicine, 2001. **163**(1): p. 19-25.
8. Go, A.S., Mozaffarian, D., Roger, V.L., Benjamin, E.J., Berry, J.D., Borden, W.B., Bravata, D.M., Dai, S., Ford, E.S., Fox, C.S., Franco, S., Fullerton, H.J., Gillespie, C., Hailpern, S.M., Heit, J.A., Howard, V.J., Huffman, M.D., Kissela, B.M., Kittner, S.J., Lackland, D.T., Lichtman, J.H., Lisabeth, L.D., Magid, D., Marcus, G.M., Marelli, A., Matchar, D.B., Mcguire, D.K., Mohler, E.R., Moy, C.S., Mussolino, M.E., Nichol, G., Paynter, N.P., Schreiner, P.J., Sorlie, P.D., Stein, J., Turan, T.N., Virani, S.S., Wong, N.D., Woo, D., and Turner, M.B., *Heart Disease and Stroke Statistics—2013 Update: A Report from the American Heart Association.* Circulation, 2013. **127**(1): p. e6-e245.
9. Young, T., Skatrud, J., and Peppard, P.E., *Risk Factors for Obstructive Sleep Apnea in Adults.* Journal of the American Medical Association, 2004. **291**(16): p. 2013-2016.
10. Altevogt, B.M. and Colten, H.R., *Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem*. 2006, Washington, DC: National Academies Press.
11. Flemons, W.W., Douglas, N.J., Kuna, S.T., Rodenstein, D.O., and Wheatley, J., *Access to Diagnosis and Treatment of Patients with Suspected Sleep Apnea.* American Journal of Respiratory and Critical Care Medicine, 2004. **169**(6): p. 668-672.
12. Chervin, R.D., Murman, D.L., Malow, B.A., and Totten, V., *Cost-Utility of Three Approaches to the Diagnosis of Sleep Apnea: Polysomnography, Home Testing, and Empirical Therapy.* Annals of Internal Medicine, 1999. **130**(6): p. 496-505.

13. Eidelman, D., *What Is the Purpose of Sleep?* Medical Hypotheses, 2002. **58**(2): p. 120-122.
14. Aittokallio, T., Gyllenberg, M., and Polo, O., *A Model of a Snorer's Upper Airway.* Mathematical Biosciences, 2001. **170**(1): p. 79-90.
15. Dodds, C., *The Physiology of Sleep.* Current Anaesthesia & Critical Care, 2002. **13**(1): p. 2-5.
16. Stanley, N., *The Physiology of Sleep and the Impact of Ageing.* European Urology Supplements, 2005. **3**(6): p. 17-23.
17. Sejnowski, T.J. and Destexhe, A., *Why Do We Sleep?* Brain Research, 2000. **886**(1): p. 208-223.
18. Carskadon, M.A. and Dement, W.C., *Normal Human Sleep: An Overview.* Principles and Practice of Sleep Medicine, 2000. **4**: p. 13-23.
19. Heo, J., Yeo, J., and Kim, Y., *Novel Focus Monitoring Using Diffraction Image of Forbidden Pitch Patterns.* Microelectronic Engineering, 2012. **98**(0): p. 595-598.
20. Bradley, T.D. and Floras, J.S., *Sleep Apnea and Heart Failure Part I: Obstructive Sleep Apnea.* Circulation, 2003. **107**(12): p. 1671-1678.
21. Åkerstedt, T., Billiard, M., Bonnet, M., Ficca, G., Garma, L., Mariotti, M., Salzarulo, P., and Schulz, H., *Awakening from Sleep.* Sleep Medicine Reviews, 2002. **6**(4): p. 267-286.
22. Collop, N.A., Tracy, S.L., Kapur, V., Mehra, R., Kuhlmann, D., Fleishman, S.A., and Ojile, J.M., *Obstructive Sleep Apnea Devices for out-of-Center (Ooc) Testing: Technology Evaluation.* Journal of clinical sleep medicine: JCSM: official publication of the American Academy of Sleep Medicine, 2011. **7**(5): p. 531.
23. Young, T., Peppard, P.E., and Gottlieb, D.J., *Epidemiology of Obstructive Sleep Apnea: A Population Health Perspective.* American Journal of Respiratory and Critical Care Medicine, 2002. **165**(9): p. 1217-1239.
24. Arens, R. and Marcus, C.L., *Pathophysiology of Upper Airway Obstruction: A Developmental Perspective.* Sleep, 2004. **27**(5): p. 997-1019.
25. Jordan, A.S. and White, D.P., *Pharyngeal Motor Control and the Pathogenesis of Obstructive Sleep Apnea.* Respiratory Physiology & Neurobiology, 2008. **160**(1): p. 1-7.
26. Sériès, F., *Upper Airway Muscles Awake and Asleep.* Sleep Medicine Reviews, 2002. **6**(3): p. 229-242.
27. Ayappa, I. and Rapoport, D.M., *The Upper Airway in Sleep: Physiology of the Pharynx.* Sleep Medicine Reviews, 2003. **7**(1): p. 9-33.
28. Kobayashi, I., Perry, A., Rhymer, J., Wuyam, B., Hughes, P., Murphy, K., Innes, J., Mcivor, J., Cheesman, A., and Guz, A., *Inspiratory Coactivation of the Genioglossus Enlarges Retroglossal Space in Laryngectomized Humans.* Journal of Applied Physiology, 1996. **80**(5): p. 1595-1604.
29. Strohl, K.P., Hensley, M.J., Hallett, M., Saunders, N.A., and Ingram, R., *Activation of Upper Airway Muscles before Onset of Inspiration in Normal Humans.* Journal of Applied Physiology, 1980. **49**(4): p. 638-642.
30. Issa, F.G. and Sullivan, C.E., *Upper Airway Closing Pressures in Obstructive Sleep Apnea.* J Appl Physiol, 1984. **57**(2): p. 520-527.
31. *Figure 14-4 Diagram Illustrates the Origin and Distribution of the Hypoglossal Nerve (Cranial Nerve Xii [General Somatic Efferent]) with Its Innervation of the Muscles of the Tongue.* 2014; Available from: http://what-when-how.com/wp-content/uploads/2012/04/tmp15F31.jpg.
32. Stanchina, M.L., Malhotra, A., Fogel, R.B., Ayas, N., Edwards, J.K., Schory, K., and White, D.P., *Genioglossus Muscle Responsiveness to Chemical and Mechanical Stimuli During Non–*

Rapid Eye Movement Sleep. American Journal of Respiratory and Critical Care Medicine, 2002. **165**(7): p. 945-949.
33. Onal, E. and Lopata, M., *Periodic Breathing and the Pathogenesis of Occlusive Sleep Apneas.* The American review of respiratory disease, 1982. **126**(4): p. 676-680.
34. Weiner, D., Mitra, J., Salamone, J., and Cherniack, N.S., *Effect of Chemical Stimuli on Nerves Supplying Upper Airway Muscles.* Journal of Applied Physiology, 1982. **52**(3): p. 530-536.
35. Dempsey, J.A., Veasey, S.C., Morgan, B.J., and O'donnell, C.P., *Pathophysiology of Sleep Apnea.* Physiological Reviews, 2010. **90**(1): p. 47-112.
36. American Sleep Apnea Association. *Arousals and Awakenings*. 2014 [cited 2014 August 28]; Available from: http://www.sleepapnea.org/treat/diagnosis/sleep-study-details.html.
37. Mcnicholas, W., *Arousal in the Sleep Apnoea Syndrome: A Mixed Blessing?* European Respiratory Journal, 1998. **12**(6): p. 1239-1241.
38. Dopp, J.M., Reichmuth, K.J., and Morgan, B.J., *Obstructive Sleep Apnea and Hypertension: Mechanisms, Evaluation, and Management.* Current Hypertension Reports, 2007. **9**(6): p. 529-34.
39. Parish, J.M. and Somers, V.K., *Obstructive Sleep Apnea and Cardiovascular Disease.* Mayo Clinic Proceedings, 2004. **79**(8): p. 1036-1046.
40. Butt, M., Dwivedi, G., Khair, O., and Lip, G.Y.H., *Obstructive Sleep Apnea and Cardiovascular Disease*. International Journal of Cardiology, 2010. **139**(1): p. 7-16.
41. Hiestand, D.M., Britz, P., Goldman, M., and Phillips, B., *Prevalence of Symptoms and Risk of Sleep Apnea in the Us Population: Results from the National Sleep Foundation Sleep in America 2005 Poll.* CHEST Journal, 2006. **130**(3): p. 780-786.
42. The Institute for Clinical and Economic Review *An Action Guide for Obstructive Sleep Apnea: Next Step for Patients, Clinicians, and Insurers*. 2013.http://www.icer-review.org/wp-content/uploads/2014/01/Action-Guide_OSA_Final_Jan2013.pdf
43. Lavie, P., *Nothing New under the Moon. Historical Accounts of Sleep Apnea Syndrome.* Archives of Internal Medicine, 1984. **144**(10): p. 2025-2028.
44. Gastaut, H. and Villeneuve, A., *The Startle Disease or Hyperekplexia: Pathological Surprise Reaction.* Journal of the Neurological Sciences, 1967. **5**(3): p. 523-542.
45. Young, T., Palta, M., Dempsey, J., Skatrud, J., Weber, S., and Badr, S., *The Occurrence of Sleep-Disordered Breathing among Middle-Aged Adults.* New England Journal of Medicine, 1993. **328**(17): p. 1230-1235.
46. Abrishami, A., Khajehdehi, A., and Chung, F., *A Systematic Review of Screening Questionnaires for Obstructive Sleep Apnea.* Canadian Journal of Anesthesia, 2010. **57**(5): p. 423-438.
47. Sharma, S.K., Kumpawat, S., Banga, A., and Goel, A., *Prevalence and Risk Factors of Obstructive Sleep Apnea Syndrome in a Population of Delhi, India.* CHEST Journal, 2006. **130**(1): p. 149-156.
48. Netzer, N.C., Stoohs, R.A., Netzer, C.M., Clark, K., and Strohl, K.P., *Using the Berlin Questionnaire to Identify Patients at Risk for the Sleep Apnea Syndrome.* Annals of Internal Medicine, 1999. **131**(7): p. 485-491.
49. Chung, F., Yegneswaran, B., Liao, P., Chung, S.A., Vairavanathan, S., Islam, S., Khajehdehi, A., and Shapiro, C.M., *Validation of the Berlin Questionnaire and American Society of Anesthesiologists Checklist as Screening Tools for Obstructive Sleep Apnea in Surgical Patients.* Anesthesiology, 2008. **108**(5): p. 822-830.

50. Sharma, S., Vasudev, C., Sinha, S., Banga, A., Pandey, R., and Hande, K., *Validation of the Modified Berlin Questionnaire to Identify Patients at Risk for the Obstructive Sleep Apnoea Syndrome.* Indian Journal of Medical Research, 2006. **124**(3): p. 281-290.
51. Chung, F., Yegneswaran, B., Liao, P., Chung, S.A., Vairavanathan, S., Islam, S., Khajehdehi, A., and Shapiro, C.M., *Validation of the Berlin Questionnaire and American Society of Anesthesiologists Checklist as Screening Tools for Obstructive Sleep Apnea in Surgical Patients.* Anesthesiology, 2008. **108**(5): p. 822-830 10.1097/ALN.0b013e31816d91b5.
52. Gross, J., Bachenberg, K., Benumof, J., Caplan, R., Connis, R., Cote, C., Nickinovich, D., Prachand, V., Ward, D., and Weaver, E., *Practice Guidelines for the Perioperative Management of Patients with Obstructive Sleep Apnea: An Updated Report by the American Society of Anesthesiologists Task Force on Perioperative Management of Patients with Obstructive Sleep Apnea.* Anesthesiology, 2006. **104**(5): p. 1081.
53. Wang, R.C., Elkins, T.P., Keech, D., Wauquier, A., and Hubbard, D., *Accuracy of Clinical Evaluation in Pediatric Obstructive Sleep Apnea.* Otolaryngology-Head and Neck Surgery, 1998. **118**(1): p. 69-73.
54. Suen, J.S., Arnold, J.E., and Brooks, L.J., *Adenotonsillectomy for Treatment of Obstructive Sleep Apnea in Children.* Archives of Otolaryngology–Head & Neck Surgery, 1995. **121**(5): p. 525-530.
55. Goldstein, N.A., Sculerati, N., Walsleben, J.A., Bhatia, N., Friedman, D.M., and Rapoport, D.M., *Clinical Diagnosis of Pediatric Obstructive Sleep Apnea Validated by Polysomnography.* Otolaryngology-Head and Neck Surgery, 1994. **111**(5): p. 611-617.
56. Leach, J., Olson, J., Hermann, J., and Manning, S., *Polysomnographic and Clinical Findings in Children with Obstructive Sleep Apnea.* Archives of Otolaryngology–Head & Neck Surgery, 1992. **118**(7): p. 741-744.
57. Kwiatkowska, M. and Atkins, A.S. *Integrating Knowledge-Driven and Data-Driven Approaches for the Derivation of Clinical Prediction Rules*. in *Machine Learning and Applications, 2005. Proceedings. Fourth International Conference on*. 2005. Los Angeles, CA.
58. Kwiatkowska, M., Atkins, M.S., Ayas, N.T., and Ryan, C.F., *Knowledge-Based Data Analysis: First Step toward the Creation of Clinical Prediction Rules Using a New Typicality Measure.* Information Technology in Biomedicine, IEEE Transactions on, 2007. **11**(6): p. 651-660.
59. Dixon, J.B., Schachter, L.M., and O'brien, P.E., *Predicting Sleep Apnea and Excessive Day Sleepiness in the Severely Obese.* CHEST Journal, 2003. **123**(4): p. 1134-1141.
60. Ramachandran, S.K., Kheterpal, S., Consens, F., Shanks, A., Doherty, T.M., Morris, M., and Tremper, K.K., *Derivation and Validation of a Simple Perioperative Sleep Apnea Prediction Score.* Anesthesia & Analgesia, 2010. **110**(4): p. 1007-1015.
61. Yung-Fu, C., Jen-Ho, C., Liang-Wen, H., Yen-Ju, L., and Chih-Jaan, T. *Diagnosis and Prediction of Patients with Severe Obstructive Apneas Using Support Vector Machine*. in *Machine Learning and Cybernetics, 2008 International Conference on*. 2008. San Diego, CA.
62. Kirby, S.D., Eng, P., Danter, W., George, C.F.P., Francovic, T., Ruby, R.R.F., and Ferguson, K.A., *Neural Network Prediction of Obstructive Sleep Apnea from Clinical Criteria.* CHEST Journal, 1999. **116**(2): p. 409-415.
63. Penzel, T., Mcnames, J., Chazal, P., Raymond, B., Murray, A., and Moody, G., *Systematic Comparison of Different Algorithms for Apnoea Detection Based on Electrocardiogram Recordings.* Medical and Biological Engineering and Computing, 2002. **40**(4): p. 402-407.

64. Lavie, L., *Obstructive Sleep Apnoea Syndrome – an Oxidative Stress Disorder.* Sleep Medicine Reviews, 2003. **7**(1): p. 35-51.
65. M. O. Mendez, J. Corthout, S. Van Huffel, M. Matteucci, T Penzel, S. Cerutti, and Bianchi, A.M., *Automatic Screening of Obstructive Sleep Apnea from the Ecg Based on Empirical Mode Decomposition and Wavelet Analysis.* Physiological Measurement, 2010. **31**: p. 273-289.
66. Goldberger, A.L., Amaral, L.a.N., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.-K., and Stanley, H.E., *Physiobank, Physiotoolkit, and Physionet: Components of a New Research Resource for Complex Physiologic Signals.* Circulation, 2000. **101**(23): p. e215-e220.
67. Penzel, T., Moody, G., Mark, R., Goldberger, A., and Peter, J. *The Apnea-Ecg Database*. in *Computers in Cardiology 2000*. 2000. Cambridge, MA: IEEE.
68. Raymond, B., Cayton, R., Bates, R., and Chappell, M. *Screening for Obstructive Sleep Apnoea Based on the Electrocardiogram - the Computers in Cardiology Challenge*. in *Computers in Cardiology 2000*. 2000. Cambridge, MA: IEEE.
69. Delibaşoğlu, I., Avci, C., and Akbaş, A. *Ecg Based Sleep Apnea Detection Using Wavelet Analysis of Instantaneous Heart Rates*. in *Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies*. 2011. Barcelona, Spain: ACM.
70. Roche, F., Pichot, V., Sforza, E., Court-Fortune, I., Duverney, D., Costes, F., Garet, M., and Barthélémy, J.-C., *Predicting Sleep Apnoea Syndrome from Heart Period: A Time-Frequency Wavelet Analysis.* European Respiratory Journal, 2003. **22**(6): p. 937-942.
71. Mietus, J., Peng, C., Ivanov, P.C., and Goldberger, A. *Detection of Obstructive Sleep Apnea from Cardiac Interbeat Interval Time Series*. in *Computers in Cardiology 2000*. 2000. Cambridge, MA: IEEE.
72. Almazaydeh, L., Elleithy, K., and Faezipour, M. *Detection of Obstructive Sleep Apnea through Ecg Signal Features*. in *Electro/Information Technology (EIT), 2012 IEEE International Conference on*. 2012. IEEE.
73. Mendez, M.O., Bianchi, A.M., Matteucci, M., Cerutti, S., and Penzel, T., *Sleep Apnea Screening by Autoregressive Models from a Single Ecg Lead.* Biomedical Engineering, IEEE Transactions on, 2009. **56**(12): p. 2838-2850.
74. Noviyanto, A., Isa, S.M., Wasito, I., and Arymurthy, A.M., *Selecting Features of Single Lead Ecg Signal for Automatic Sleep Stages Classification Using Correlation-Based Feature Subset Selection.* International Journal of Computer Science Issues (IJCSI), 2011. **8**(5).
75. Yılmaz, B., Asyalı, M.H., Arıkan, E., Yetkin, S., and Özgen, F., *Sleep Stage and Obstructive Apneaic Epoch Classification Using Single-Lead Ecg.* BioMedical Engineering OnLine, 2010. **9**: p. 39.
76. Emin Tagluk, M. and Sezgin, N., *A New Approach for Estimation of Obstructive Sleep Apnea Syndrome.* Expert Systems with Applications, 2011. **38**(5): p. 5346-5351.
77. Doukas, C., Petsatodis, T., Boukis, C., and Maglogiannis, I., *Automated Sleep Breath Disorders Detection Utilizing Patient Sound Analysis.* Biomedical Signal Processing and Control, 2012. **7**(3): p. 256-264.
78. Mikami, T. *Detecting Nonlinearity in Prediction Residuals of Snoring Sounds*. in *2009 ICROS-SICE International Joint Conference*. 2009. Fukuoka, Japan.

79. Robertson, H.J., Soraghan, J.J., Idzikowski, C., and Conway, B.A. *Emd and Pca for the Prediction of Sleep Apnoea: A Comparative Study*. in *Signal Processing and Information Technology, 2007 IEEE International Symposium on*. 2007. Alcala de Henares, Spain.
80. Williamson, J.R., Bliss, D.W., Browne, D.W., Indic, P., Bloch-Salisbury, E., and Paydarfar, D. *Using Physiological Signals to Predict Apnea in Preterm Infants*. in *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on*. 2011. Pacific Grove, CA: IEEE.
81. Várady, P., Micsik, T., Benedek, S., and Benyó, Z., *A Novel Method for the Detection of Apnea and Hypopnea Events in Respiration Signals.* Biomedical Engineering, IEEE Transactions on, 2002. **49**(9): p. 936-942.
82. Maier, C. and Dickhaus, H., *Central Sleep Apnea Detection from Ecg-Derived Respiratory Signals. Application of Multivariate Recurrence Plot Analysis.* Methods of Information in Medicine, 2010. **49**(5): p. 462-6.
83. Moody, G.B., Mark, R.G., Zoccola, A., and Mantero, S., *Derivation of Respiratory Signals from Multi-Lead Ecgs.* Computers in Cardiology, 1985. **12**: p. 113-116.
84. Babaeizadeh, S., Zhou, S.H., Pittman, S.D., and White, D.P., *Electrocardiogram-Derived Respiration in Screening of Sleep-Disordered Breathing.* Journal of Electrocardiology, 2011. **44**(6): p. 700-706.
85. Bock, J. and Gough, D.A., *Toward Prediction of Physiological State Signals in Sleep Apnea.* Biomedical Engineering, IEEE Transactions on, 1998. **45**(11): p. 1332-1341.
86. Waxman, J.A., Graupe, D., and Carley, D.W., *Automated Prediction of Apnea and Hypopnea, Using a Lamstar Artificial Neural Network.* American Journal of Respiratory and Critical Care Medicine, 2010. **181**(7): p. 727-733.
87. Paradiso, R., Loriga, G., and Taccini, N., *A Wearable Health Care System Based on Knitted Integrated Sensors.* Information Technology in Biomedicine, IEEE Transactions on, 2005. **9**(3): p. 337-344.
88. Lymberis, A. and De Rossi, D., *Myheart: Fighting Cardiovascular Disease by Preventive Lifestyle and Early Diagnosis.* Wearable eHealth Systems for Personalised Health Management: State of the Art and Future Challenges, 2004. **108**: p. 36.
89. Di Rienzo, M., Rizzo, F., Parati, G., Brambilla, G., Ferratini, M., and Castiglioni, P. *Magic System: A New Textile-Based Wearable Device for Biological Signal Monitoring. Applicability in Daily Life and Clinical Setting*. in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*. 2005. Shanghai, China,: IEEE.
90. Anliker, U., Ward, J.A., Lukowicz, P., Troster, G., Dolveck, F., Baer, M., Keita, F., Schenker, E.B., Catarsi, F., and Coluccini, L., *Amon: A Wearable Multiparameter Medical Monitoring and Alert System.* Information Technology in Biomedicine, IEEE Transactions on, 2004. **8**(4): p. 415-427.
91. Oliver, N. and Flores-Mangas, F. *Healthgear: A Real-Time Wearable System for Monitoring and Analyzing Physiological Signals*. in *Wearable and Implantable Body Sensor Networks, 2006. BSN 2006. International Workshop on*. 2006. Boston, MA: IEEE.
92. Wilhelm, F.H., Roth, W.T., and Sackner, M.A., *The Lifeshirt: An Advanced System for Ambulatory Measurement of Respiratory and Cardiac Function.* Behavior Modification, 2003. **27**(5): p. 671-691.

93. Hailstone, J. and Kilding, A.E., *Reliability and Validity of the Zephyr™ Bioharness™ to Measure Respiratory Responses to Exercise.* Measurement in Physical Education and Exercise Science, 2011. **15**(4): p. 293-300.
94. Michael Berry, G.S.L., *Data Mining Techniques: Theory and Practice Course Notes*. 2009: Data Miners, Inc.
95. Loh, W.Y., *Classification and Regression Trees.* Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2011. **1**(1): p. 14-23.
96. Sarma, K.S., *Predictive Modeling with Sas Enterprise Miner: Practical Solutions for Business Applications*. 2007, Cary, NC: SAS Institute.
97. Sas Institute Inc. *The Arboretum Procedure*. 2001.http://support.sas.com/documentation/onlinedoc/miner/em43/allproc.pdf
98. Kitchen, C.M.R., *Nonparametric Variable Selection Using Machine Learning Algorithms in High Dimensional (Large P, Small N) Biomedical Applications*, in *Biomedical Engineering, Trends in Electronics, Communications and Software*, Laskovski, A., Editor. 2011, InTech.
99. Burgette, L.F. and Reiter, J.P., *Multiple Imputation for Missing Data Via Sequential Regression Trees.* American Journal of Epidemiology, 2010: p. kwq260.
100. Lee, J.H., Kim, I.Y., and O'keefe, C.M., *On Regression-Tree-Based Synthetic Data Methods for Business Data.* Journal of Privacy and Confidentiality, 2013. **5**(1): p. 5.
101. Rubin, D.B., *The Bayesian Bootstrap.* The Annals of Statistics, 1981. **9**(1): p. 130-134.
102. He, Y., *Missing Data Imputation for Tree-Based Models*. 2006, University of California Los Angeles.
103. Burns, R.P. and Burns, R., *Business Research Methods and Statistics Using Spss*. 2008: SAGE Publications.
104. Freund, R.J. and Wilson, W.J., *Statistical Methods*. 2003, Burlington, MA: Academic Press.
105. Bewick, V., Cheek, L., and Ball, J., *Statistics Review 14: Logistic Regression.* Critical Care, 2005. **9**(1): p. 112-118.
106. Czepiel, S.A. *Maximum Likelihood Estimation of Logistic Regression Models: Theory and Implementation*. 2002.http://www.czep.net/stat/mlelr.pdf
107. Hagan, M.T., Demuth, H.B., and Beale, M.H., *Neural Network Design*. 2002: University of Colorado Bookstore.
108. Hagan, M., Demuth, H., Beale, M., and De Jesus, O., *Neural Network Design (2nd Edition)*. 2014: Martin Hagan.
109. Fawcett, T., *An Introduction to Roc Analysis.* Pattern Recognition Letters, 2006. **27**(8): p. 861-874.
110. Swaving, M., Van Houwelingen, H., Ottes, F.P., and Steerneman, T., *Statistical Comparison of Roc Curves from Multiple Readers.* Medical Decision Making, 1996. **16**(2): p. 143-152.
111. Eng, J., *Receiver Operating Characteristic Analysis: A Primer.* Academic Radiology, 2005. **12**(7): p. 909-916.
112. Fawcett, T., *Roc Graphs: Notes and Practical Considerations for Researchers.* Machine Learning, 2004. **31**: p. 1-38.
113. Mcneil, B.J. and Hanley, J.A., *Statistical Approaches to the Analysis of Receiver Operating Characteristic (Roc) Curves.* Medical Decision Making, 1983. **4**(2): p. 137-150.
114. Hanley, J.A. and Mcneil, B.J., *The Meaning and Use of the Area under a Receiver Operating Characteristic (Roc) Curve.* Radiology, 1982. **143**(1): p. 29-36.

115. Metz, C.E., *Roc Methodology in Radiologic Imaging.* Investigative Radiology, 1986. **21**(9): p. 720-733.
116. Redline, S., Sanders, M.H., Lind, B.K., Quan, S.F., Iber, C., Gottlieb, D.J., Bonekat, W.H., Rapoport, D., Smith, P.L., and Kiley, J.P., *Methods for Obtaining and Analyzing Unattended Polysomnography Data for a Multicenter Study. Sleep Heart Health Research Group.* Sleep, 1998. **21**(7): p. 759-767.
117. Redline, S., Et Al. *Sleep Heart Health Study*. [cited 2014; Available from: http://sleepdata.org/datasets/shhs.
118. Schlomer, G.L., Bauman, S., and Card, N.A., *Best Practices for Missing Data Management in Counseling Psychology.* Journal of Counseling Psychology, 2010. **57**(1): p. 1.
119. Humphries, M. *Missing Data & How to Deal: An Overview of Missing Data*. 2012.http://www.texaslonghornsl.com/cola/centers/prc/_files/cs/Missing-Data.pdf
120. Pappas, P.A. and Depuy, V. *An Overview of Non-Parametric Tests in Sas: When, Why, and How*. in *SouthEast SAS Users Group Conference*. 2004. Nashville, TN.
121. Acuna, E. and Rodriguez, C., *The Treatment of Missing Values and Its Effect on Classifier Accuracy*, in *Classification, Clustering, and Data Mining Applications*. 2004, Springer. p. 639-647.
122. Mundfrom, D.J. and Whitcomb, A., *Imputing Missing Values: The Effect on the Accuracy of Classification.* Multiple Linear Regression Viewpoint, 1998. **25**.
123. Chan, L.S. and Dunn, O.J., *The Treatment of Missing Values in Discriminant. Analysis-I. The Sampling Experiment.* Journal of the American Statistical Association, 1972. **67**(338): p. 473-477.
124. Kostoglou-Athanassiou, I. and Athanassiou, P., *Metabolic Syndrome and Sleep Apnea.* Hippokratia Medical Journal, 2008. **12**(2): p. 81-86.
125. Box, G.E. and Cox, D.R., *An Analysis of Transformations.* Journal of the Royal Statistical Society, 1964: p. 211-252.
126. Force, A.O.S.a.T. and Medicine, A.a.O.S., *Clinical Guideline for the Evaluation, Management and Long-Term Care of Obstructive Sleep Apnea in Adults.* Journal of Clinical Sleep Medicine, 2009. **5**(3): p. 263.
127. Collop, N.A., Anderson, W.M., Boehlecke, B., Claman, D., Goldberg, R., Gottlieb, D.J., Hudgel, D., Sateia, M., and Schwab, R., *Clinical Guidelines for the Use of Unattended Portable Monitors in the Diagnosis of Obstructive Sleep Apnea in Adult Patients.* Journal of Clinical Sleep Medicine, 2007. **3**(7): p. 737-747.
128. Verveer, P.J. and Duin, R.P.W., *An Evaluation of Intrinsic Dimensionality Estimators.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1995. **17**(1): p. 81-86.
129. Meek, S. and Morris, F., *Abc of Clinical Electrocardiography: Introduction. I—Leads, Rate, Rhythm, and Cardiac Axis.* British Medical Journal, 2002. **324**(7334): p. 415.
130. Thaler, M.S., *The Only Ekg Book You'll Ever Need*. Vol. 365. 2010: Lippincott Williams & Wilkins.
131. Richard Pflanzer, W.M., *Physiology Lessons for Use with the Biopac Student Lab*. 2008.
132. Grier, J.W. *How to Use 1-Lead Ecg Recorders to Obtain 12-Lead Resting Ecgs and Exercise ("Stress") Ecgs*. 2008 [cited 2014 03/26]; Available from: http://www.ndsu.edu/pubweb/~grier/1to12-lead-ECG-EKG.html.

133. Atkielski, A. *Schematic Diagram of Normal Sinus Rhythm for a Human Heart as Seen on Ecg, Two Periods Forming a Rr-Interval*. 2009; Available from: http://commons.wikimedia.org/wiki/File:ECG-RRinterval.svg#.
134. Van Ravenswaaij, C., Kollee, L.A., Hopman, J.C., Stoelinga, G.B., and Van Geijn, H.P., *Heart Rate Variability.* Annals of Internal Medicine, 1993. **118**(6): p. 436-447.
135. Xie, A., Teodorescu, M., Pegelow, D.F., Teodorescu, M.C., Gong, Y., Fedie, J.E., and Dempsey, J.A., *Effects of Stabilizing or Increasing Respiratory Motor Outputs on Obstructive Sleep Apnea.* Journal of Applied Physiology, 2013. **115**(1): p. 22-33.
136. Kheirandish-Gozal, L. and Gozal, D., *Sleep Disordered Breathing in Children: A Comprehensive Clinical Guide to Evaluation and Treatment*. 2012: Humana Press.
137. Xiong, C., Sjöberg, B.J., Sveider, P., Ask, P., Loyd, D., and Wranne, B., *Problems in Timing of Respiration with the Nasal Thermistor Technique.* Journal of the American Society of Echocardiography, 1993. **6**(2): p. 210-216.
138. O'brien, C. and Heneghan, C., *A Comparison of Algorithms for Estimation of a Respiratory Signal from the Surface Electrocardiogram.* Computers in Biology and Medicine, 2007. **37**(3): p. 305-314.
139. Lipsitz, L.A., Hashimoto, F., Lubowsky, L.P., Mietus, J., Moody, G.B., Appenzeller, O., and Goldberger, A.L., *Heart Rate and Respiratory Rhythm Dynamics on Ascent to High Altitude.* British Heart Journal, 1995. **74**(4): p. 390-6.
140. Ding, S., Zhu, X., Chen, W., and Wei, D., *Derivation of Respiratory Signal from Single-Channel Ecgs Based on Source Statistics.* International Journal of Bioelectromagnetism, 2004. **6**(1).
141. Chow, S.-M., Ferrer, E., and Hsieh, F., *Statistical Methods for Modeling Human Dynamics: An Interdisciplinary Dialogue*. 2012: Taylor & Francis.
142. Cordes, D., Haughton, V.M., Arfanakis, K., Carew, J.D., Turski, P.A., Moritz, C.H., Quigley, M.A., and Meyerand, M.E., *Frequencies Contributing to Functional Connectivity in the Cerebral Cortex in "Resting-State" Data.* American Journal of Neuroradiology, 2001. **22**(7): p. 1326-1333.
143. Shim, I., Soraghan, J., and Siew, W., *Detection of Pd Utilizing Digital Signal Processing Methods. Part 3: Open-Loop Noise Reduction.* Electrical Insulation Magazine, IEEE, 2001. **17**(1): p. 6-13.
144. Mallat, S.G., *A Theory for Multiresolution Signal Decomposition: The Wavelet Representation.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1989. **11**(7): p. 674-693.
145. Gao, R.X. and Yan, R., *Wavelets: Theory and Applications for Manufacturing*. 2010, Springer.
146. Monzón, L., Beylkin, G., and Hereman, W., *Compactly Supported Wavelets Based on Almost Interpolating and Nearly Linear Phase Filters (Coiflets).* Applied and Computational Harmonic Analysis, 1999. **7**(2): p. 184-210.
147. Takens, F., *Detecting Strange Attractors in Turbulence*, in *Dynamical Systems and Turbulence, Warwick 1980*. 1981, Springer Berlin Heidelberg. p. 366-381.
148. Le, T.Q., Cheng, C., Sangasoongsong, A., Wongdhamma, W., and Bukkapatnam, S.T., *Wireless Wearable Multisensory Suite and Real-Time Prediction of Obstructive Sleep Apnea Episodes.* Translational Engineering in Health and Medicine, IEEE Journal of, 2013. **1**: p. 2700109-2700109.
149. Karandikar, K., Le, T.Q., Sa-Ngasoongsong, A., Wongdhamma, W., and Bukkapatnam, S.T. *Detection of Sleep Apnea Events Via Tracking Nonlinear Dynamic Cardio-Respiratory Coupling*

from Electrocardiogram Signals. in *Neural Engineering (NER), 2013 6th International IEEE/EMBS Conference on*. 2013. IEEE.
150. Wongdhamma, W., Le, T.Q., and Bukkapatnam, S.T. *Wireless Wearable Multi-Sensory System for Monitoring of Sleep Apnea and Other Cardiorespiratory Disorders*. in *Automation Science and Engineering (CASE), 2013 IEEE International Conference on*. 2013. Madison, WI: IEEE.
151. Bukkapatnam, S.T., Le, T., and Wongdhamma, W., *Device and Method for Predicting and Preventing Obstructive Sleep Apnea (Osa) Episodes*. 2013, The Board Of Regents For Oklahoma State University.
152. Baker, G.L. and Gollub, J.P., *Chaotic Dynamics: An Introduction*. 1996: Cambridge University Press.
153. Scheinerman, E.R., *Invitation to Dynamical Systems*. 2012: Courier Corporation.
154. Dq, N. *The Idea of a Dynamical System*. [cited 2015 July 9]; Available from: http://mathinsight.org/dynamical_system_idea.
155. Marwan, N., *Encounters with Neighbours: Current Developments of Concepts Based on Recurrence Plots and Their Applications*. 2003: Norbert Marwan.
156. Marwan, N., Romano, M.C., Thiel, M., and Kurths, J., *Recurrence Plots for the Analysis of Complex Systems.* Physics Reports, 2007. **438**(5): p. 237-329.
157. Webber Jr, C.L. and Zbilut, J.P., *Recurrence Quantification Analysis of Nonlinear Dynamical Systems*, in *Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences*. 2005, National Science Foundation: Arlington, VA. p. 26-94.
158. Ding, M., Grebogi, C., Ott, E., Sauer, T., and Yorke, J.A., *Plateau Onset for Correlation Dimension: When Does It Occur?* Physical Review Letters, 1993. **70**(25): p. 3872.
159. Kantz, H. and Schreiber, T., *Nonlinear Time Series Analysis*. Vol. 7. 2004: Cambridge University Press.
160. Zbilut, J.P. and Webber, C.L., *Recurrence Quantification Analysis.* Wiley Encyclopedia of Biomedical Engineering, 2006.
161. Kennel, M.B., Brown, R., and Abarbanel, H.D., *Determining Embedding Dimension for Phase-Space Reconstruction Using a Geometrical Construction.* Physical Review A, 1992. **45**(6): p. 3403.
162. Fraser, A.M. and Swinney, H.L., *Independent Coordinates for Strange Attractors from Mutual Information.* Physical Review A, 1986. **33**(2): p. 1134.
163. Roulston, M.S., *Estimating the Errors on Measured Entropy and Mutual Information.* Physica D: Nonlinear Phenomena, 1999. **125**(3): p. 285-294.
164. Hegger, R. and Kantz, H., *Improved False Nearest Neighbor Method to Detect Determinism in Time Series Data.* Physical Review E, 1999. **60**(4): p. 4970.
165. Webber Jr, C.L. and Marwan, N., *Recurrence Quantification Analysis: Theory and Best Practices*. 2014: Springer International Publishing.
166. Freedman, D. and Diaconis, P., *On the Histogram as a Density Estimator:L 2 Theory.* Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 1981. **57**(4): p. 453-476.
167. Lorenz, E.N., *Deterministic Nonperiodic Flow*. Journal of the atmospheric sciences, 1963. **20**(2): p. 130-141.
168. Saltzman, B., *Finite Amplitude Free Convection as an Initial Value Problem-I.* Journal of the Atmospheric Sciences, 1962. **19**(4): p. 329-341.

169. Sprott, J., *Simplifications of the Lorenz Attractor.* Nonlinear Dynamics, Psychology, and Life Sciences, 2009. **13**(3): p. 271.
170. Eckmann, J.-P., Kamphorst, S.O., and Ruelle, D., *Recurrence Plots of Dynamical Systems.* Europhys. Lett, 1987. **4**(9): p. 973-977.
171. Thiel, M., Romano, M.C., Kurths, J., Meucci, R., Allaria, E., and Arecchi, F.T., *Influence of Observational Noise on the Recurrence Quantification Analysis.* Physica D: Nonlinear Phenomena, 2002. **171**(3): p. 138-152.
172. Schinkel, S., Dimigen, O., and Marwan, N., *Selection of Recurrence Threshold for Signal Detection.* The European Physical Journal Special Topics, 2008. **164**(1): p. 45-53.
173. Zbilut, J.P. and Webber, C.L., *Embeddings and Delays as Derived from Quantification of Recurrence Plots.* Physics Letters A, 1992. **171**(3): p. 199-203.
174. Webber, C.L. and Zbilut, J.P., *Dynamical Assessment of Physiological Systems and States Using Recurrence Plot Strategies*. Vol. 76. 1994. 965-973.
175. Marwan, N., Donges, J.F., Zou, Y., Donner, R.V., and Kurths, J., *Complex Network Approach for Recurrence Analysis of Time Series.* Physics Letters A, 2009. **373**(46): p. 4246-4254.
176. Marwan, N., Kurths, J., and Foerster, S., *Analysing Spatially Extended High-Dimensional Dynamics by Recurrence Plots.* Physics Letters A, 2015. **379**(10): p. 894-900.
177. Little, M.A., Mcsharry, P.E., Roberts, S.J., Costello, D.A., and Moroz, I.M., *Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection.* BioMedical Engineering OnLine, 2007. **6**(1): p. 23.
178. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D.-U., *Complex Networks: Structure and Dynamics.* Physics Reports, 2006. **424**(4): p. 175-308.
179. Gao, J., *Recurrence Time Statistics for Chaotic Systems and Their Applications.* Physical Review Letters, 1999. **83**(16): p. 3178.
180. Oppenheim, A.V. and Schafer, R.W., *Discrete-Time Signal Processing*. 3rd ed. 2010, Upper Saddle River, NJ: Prentice Hall.
181. Ljung, L., *System Identification*. 1998, Englewood Cliffs, NJ: Springer.
182. Lds-Group *Understanding Fft Windows. Application Note Ano14*. 2003.http://www.physik.uni-wuerzburg.de/~praktiku/Anleitung/Fremde/ANO14.pdf
183. Marwan, N. *Cross Recurrence Plot Toolbox 5.18 (R29.3)*. 2015.http://tocsy.pik-potsdam.de/CRPtoolbox/
184. Benz, R.L., Masood, I., and Pressman, M.R., *Sleep Disorders Associated with Chronic Kidney Disease*, in *Chronic Kidney Disease*. 2012, InTech.
185. Dawson, D., Yang, H., Malshe, M., Bukkapatnam, S.T.S., Benjamin, B., and Komanduri, R., *Linear Affine Transformations between 3-Lead (Frank Xyz Leads) Vectorcardiogram and 12-Lead Electrocardiogram Signals.* Journal of Electrocardiology, 2009. **42**(6): p. 622-630.

VITA

Woranat Wongdhamma

Candidate for the Degree of

Doctor of Philosophy

Thesis: STATISTICAL MODELS FOR DETECTING EXISTENCE OF OBSTRUCTIVE SLEEP APNEA, PREDICTING ITS SEVERITY, AND FORECASTING FUTURE EPISODES

Major Field: Industrial Engineering and Management

Biographical:

Education:

Received the B.Eng. degree in Electrical Engineering from Mahidol University, Nakhon Pathom, Thailand in May 2003

Received the M.S. degree in Industrial Technology from Texas A&M University – Commerce in December 2006

Completed the requirements for the Doctor of Philosophy in Industrial Engineering and Management at Oklahoma State University in December 2015

Experience:

Woranat has research interests in statistical analysis, new product development, data mining, analytics, and sensor networks for real-time monitoring and decision making. After the completion of his M.S. study, he had worked at Media Standard Company Limited as a General Manager. His responsibility included product release planning, copyrights management, and post-production management. He was also a visiting lecturer at the Civil Aviation Training Center (CATC), Thailand. During his Ph.D. study, he had worked as a Graduate Research Associate for the School of Industrial Engineering and Management at Oklahoma State University. He was responsible for the development of the statistical and data mining models (classification and forecasting), and the sensor networks for real-time monitoring and decision making device.

Professional Memberships:

- Institute for Operations Research and the Management Sciences (INFORMS)
- The Industrial Engineering Honor Society (Alpha Pi Mu)
- Institute of Electrical and Electronics Engineers (IEEE)
- Institute for Industrial Engineers (IIE)