Pressure Sensitive Mat: an Alternative Sensor to Detect Sleep-Related Breathing Disorders

Thesis submitted

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical and Computer Engineering

Ottawa-Carleton Institute for Electrical and Computer Engineering

School of Electrical Engineering and Computer Science

Faculty of Engineering, University of Ottawa

uOttawa

Abstract

Sleep Apnea (SA) is a common disorder that affects approximately 2% of middle-aged women and 4% of middle-aged men. It is characterized by repetitive cessation of breathing during sleep. SA has significant health and social consequences such as daytime sleepiness, impaired quality of life, and in the worst case, myocardial infarction and sudden cardiac death. It has been estimated that approximately 80% of individuals with moderate to severe SA syndrome have not been diagnosed. The lack of patient sleep histories has caused low identification of SA and referral rates, especially in primary care facilities. Moreover, due to the inadequate prevalence of overnight polysomnography (PSG) as a standard clinical test of SA, patients suspected of having this sleep disorder have to wait several months for diagnosis and treatment.

The costly and time-consuming nature of PSG and the lack of sleep clinics have created a demand for suitable home-based health monitoring devices. Over the years, several devices have been developed to monitor sleep unobtrusively, while an individual is lying in bed. However, most of these devices would either disrupt the sleep of the patient or be disrupted by the patient during routine bed sheet changes. Pressure measurement using a Pressure Sensitive Mat (PSM) enables a non-contact approach for monitoring patient vital signs such as respiration rate. The PSM has the potential to replace obtrusive breathing sensors in the sleep lab and to be used as a pre-screening tool for patients suspected of having sleep apnea.

This thesis proposes multiple algorithms applicable to PSM in order to assess sleep quality. First, fusion techniques are proposed to extract a breathing signal from PSM. Second, a wide range of machine learning approaches including a simple threshold-based algorithm, a linear support vector machine (SVM) and two deep learning methods (i.e., a temporal convolutional network (TCN) and a bidirectional long short-term memory (BiLSTM) network) are compared to find a good-

performing method for automatically detecting central sleep apnea (CSA) events from PSM signals. The results show that the accuracy of the model with the best performance is 95.1% and it is achieved by the BiLSTM network. Finally, by applying SVM, personalized systems are optimized to investigate long-term sleep pattern changes such as central apnea index (CAI), bed occupancy (BO), day-clock, and night-clock from previously recorded data.

Acknowledgments

I have a great appreciation for all the people whom I have had the privilege of working with while carrying out my Ph.D. project.

Firstly, I would like to pass my gratitude to my academic supervisors, Dr. Martin Bouchard, Dr. Rafik Goubran and my medical advisor, Dr. Frank Knoefel, for the enormous support I received throughout my thesis. I am forever grateful for their professional mentoring.

A big thank you to my lab co-habitants for their assistance, encouragement, and friendship. In particular: Qianhui Wan, Hitham Jleed, Hala As'ad, Sareh Soleimani Gilakjani, Bruce Wallace, and Julien Lariviere.

My life leading up to this point has been shaped by unconditional love, unwavering support, and encouragement from my dear friends and beloved family, especially my parents. I would not be where I am without them. This thesis is dedicated to them.

Lastly, I would like to offer my special thanks to Omid Zobeiri for his innumerable acts of support, coaching, and listening to me complain during the more difficult months of my Ph.D. I wouldn’t have been able to accomplish this without him.

Table of Contents

List of Figures

List of Tables

List of Acronyms and Abbreviations

AASM	American Academy of Sleep Medicine
AED	Average Number of Events per day
AHI	Apnea-Hypopnea Index
AI	Apnea Index
ANN	Artificial Neural Network
APAP	Auto-titrating Positive Airway Pressure
ASV	Adaptive Servo-Ventilation
AVE	Average
BCG	Ballistocardiogram
BMI	Body Mass Index
BPAP	Bi-level Positive Airway Pressure
BO	Bed Occupancy
BPM	Breath Per Minute
CAI	Central Apnea Index
cECG	Capacitive Electrocardiography
CNN	Convolutional Neural Network
CPAP	Continuous Positive Airway Pressure
CSA	Central Sleep Apnea
CS-SVM	Cost-Sensitive SVM
DL	Deep Learning
DWT	Discrete Wavelet Transform
ECG	Electrocardiogram

EDS	Excessive Daytime Sleepiness
EEG	Electroencephalogram
EMG	Electromyogram
EOG	Electrooculogram
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FNE	First Night Effect
FN	False Negative
FP	False Positive
HHT	Hilbert-Huang Transform
HGNS	Hypoglossal Nerve Stimulation
HoPSG	Home Polysomnogram
IIR	Infinite Impulse Response
ILSVRC	ImageNet Large-Scale Visual Recognition Challenge
KNN	K-Nearest-Neighbor
LabPSG	Laboratory Polysomnograms
LDA	Linear Discriminant Analysis
LSTM	Long Short-Term Memory
MSA	Mixed Sleep Apnea
MVDR	Minimum Variance Distortionless Response
MAAs	Mandibular Advancement Appliances
NEPAP	Nasal Expiratory Positive Airway Pressure
NM-1	NearMiss-1

NM-2	NearMiss-2
NPV	Negative Predictive Value
NREM	Non-Rapid Eye Movement
OA	Oral Appliances
OSA	Obstructive Sleep Apnea
PAP	Positive Airway Pressure
PCC	Pearson's Correlation Coefficient
PPG	Photoplethysmography
PPGI	Photoplethysmographic Imaging
PPV	Positive Predictive Value
PSG	Polysomnogram
PSD	Power Spectral Density
PSM	Pressure Sensitive Mats
PTT	Pulse Transition Time
PVDF	Polyvinylidene Fluoride
RDI	Respiratory Disturbance Index
REM	Rapid Eye Movement
RFE	Recursive Feature Elimination
RIP	Respiratory Inductance Plethysmography
RMS	Root Mean Square
RNN	Recurrent Neural Network
SA	Sleep Apnea
SCSB	Statistic Charge Sensitive Bed

SDB	Sleep Disorder Breathing
SFM	Spectral Flatness Measure
SNR-MAX	Signal to Noise Ratio Maximum
SPLs	Soft Palate Lifters
SpO2	Saturation of Peripheral Oxygen
SRI	Sleep Regularity Index
STFT	Short-Time Fourier Transform
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
TRDs	Tongue-Retaining Devices

List of Symbols

μ	Regularization factor
r_c	Correlation coefficient
p_v	P-value
α	Lagrange multiplier
ξ	Slack variable
C	Cost constant
γ	Ratio of oversampled apnea class instances to undersampled not-apnea class instances.

Chapter 1: Introduction

1.2 Introduction

This chapter details the motivation for this thesis. The chapter provides an overview of the problem statement and research focus. Moreover, the contributions are stated in section 1.6 and the thesis structure is described in section 1.7.

1.3 Motivation

Through evidence-based reviews by the American Academy of Sleep Medicine (AASM), the current standard clinical practice to diagnose or rule out sleep-related disorders such as sleep apnea (SA) is in-laboratory polysomnography (PSG) [1], [2]. PSG is a comprehensive recording of the physiological changes that occur during sleep. A subject's polysomnography is recorded during nighttime sleep at a sleep center and then visually analyzed offline by sleep specialists. However, the procedure is significantly time-consuming and costly.

Moreover, unfamiliar surroundings in sleep laboratories and the knowledge of a clinical observer might affect the patients' sleep characteristics (i.e., "the first night effect" (FNE)) [3]. Also, a controlled environment like a sleep laboratory might not hold home's sleep-disrupting factors, which can potentially result in an incorrect diagnosis. A simple acquisition system for home monitoring can address both of these concerns in assessing sleep quality.

1.4 Problem Statement

Pressure Sensitive Mat (PSM) is a portable and informative technology for home monitoring of patients who are suspected of having sleep-related disorders. PSM data can be collected along with

traditional PSG data during a patient stay at the sleep lab, or it can be used as a primary standalone home monitoring, leading to a significant reduction in the diagnostics cost and waiting time for a sleep study. Based on the level of severity measured, the results of the home monitoring through PSM could be further used to prioritize patients for PSG or eliminate further comprehensive tests. Home monitoring through PSM can also be used as an alternative system, targeted to those with allergies to adhesives, children, and the elderly who have a difficult time tolerating the currently used obtrusive breathing sensors (e.g., belts placed snuggly around the chest and abdomen, thermistors and nasal pressure cannulas positioned inside the nostrils).

The use of other unobtrusive technologies such as radar, laser, Capacitive Electrocardiography (cECG), Photoplethysmography (PPG), video motion analysis, Photoplethysmographic Imaging (PPGI), and thermography has been studied for home monitoring [4]. However, they impose restrictions on the positioning of the sensor and the distance between the sensor and user. In contrast, a PSM system does not add any limitation to the sleeping environment, and once in place, it has no risk of being affected by the subject either intentionally or inadvertently.

1.5 Objectives

The main objective of this thesis is to investigate the application of pattern recognition and signal processing methods for sleep monitoring on smart PSM-equipped beds and to analyze their performance. The application of the algorithms developed is mainly for day-to-day home monitoring of apneic subjects and to provide statistics of the apnea severity over several days. They provide assessment results as per common clinical standards, i.e., providing the count of apnea events separately.

1.6 Summary of Contributions

The work completed for this thesis resulted in a number of publications, some as the first author and some as coauthor. Contributions related to this thesis proposal are listed as follows:

1. Development of two signal fusion methods to produce a higher quality breathing signal, including a Pearson's Correlation Coefficient (PPC) analysis as a combining method and a Signal to Noise Ratio Maximum (SNR-MAX) method based on Minimum Variance Distortionless Response (MVDR) beam-former theory. Comparisons were made between the proposed methods and previously developed methods of signal fusion. Applied to validation data in laboratory conditions, both methods showed promising results and improvement. The results are presented in Chapter 4. The findings were presented in [5] and published in [6]:

[5] H. Azimi, M. Bouchard, R. Goubran, and F. Knoefel, "Smart Bed Technology For Breathing Monitoring," presented at *Medical Devices Innovation Institute (MDII) and NSERC CREATE-BEST Poster Day,* Ottawa, ON, Canada, Sep.2017.

[6] H. Azimi, S. S. Gilakjani, M. Bouchard, S. Bennett, R. A. Goubran, and F. Knoefel, "Breathing signal combining for respiration rate estimation in smart beds," in *2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2017, pp. 303–307.

Later in [7], the performance of the proposed methods was evaluated for close-to-reality data where the participants were allowed to move around on the bed and have different types of breathing, such as shallow breathing, deep breathing, and periods of apnea. Comparing to other data fusion methods, the SNR-MAX was found to be the best method of signal combining.

[7] S. S. Gilakjani, H. Azimi, M. Bouchard, R. A. Goubran, and F. Knoefel, "Improved sensor selection method during movement for breathing rate estimation with unobtrusive pressure sensor arrays," in *2018 IEEE Sensors Applications Symposium (SAS)*, 2018, pp. 1–6.

2. Development of a simple method for apnea detection by Dual Respiratory Inductance Plethysmography (RIP) based on AASM rules. The study was intended to investigate the effect of all types of SA, including hypopnea, CSA, Obstructive Sleep Apnea (OSA), and Mixed Sleep Apnea (MSA) on thoracic and abdominal signals. The novelty of the research was to use RIP signals instead of the airflow signal, which is commonly used to detect SA. The study resulted in a paper [8], and it is presented in Chapter 5.

[8] H. Azimi, S. S. Gilakjani, M. Bouchard, R. A. Goubran, and F. Knoefel, "Automatic apnea-hypopnea events detection using an alternative sensor," in *2018 IEEE Sensors Applications Symposium (SAS)*, 2018, pp. 1–5.

3. Development and comparison machine learning approaches to find a good-performing method for automatically detecting central sleep apnea (CSA) events from PSM signal. Two deep learning approaches were implemented for automatic detection of CSA events: a temporal convolutional network (TCN) and a bidirectional long short-term memory (BiLSTM) network. The deep learning models designed were compared to a linear support vector machine (SVM) classifier and a simple threshold-based algorithm presented in Chapter 5. The analysis is described in Chapter 6 of the thesis and resulted in conference publications [9], [10] and a journal paper [11].

[9] H. Azimi, M. Bouchard, R. A. Goubran, and F. Knoefel, "Apnea Event Detection Methodology using Pressure Sensors," in *2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2019.

[10] H. Azimi, M. Bouchard, R. Goubran, and F. Knoefel, "Smart Home And Sleep Apnea: Application To Eldercare," presented at *Engineering and Computer Science Graduate Poster Competition*, Ottawa, ON, Canada, 2019.

[11] H. Azimi, P. Xi, M. Bouchard, R. Goubran, and F. Knoefel, "Machine Learning-Based Automatic Detection of Central Sleep Apnea Events From a Pressure Sensitive Mat," *IEEE Access*, vol. 8, pp. 173428–173439, 2020, doi: 10.1109/ACCESS.2020.3025808.

4. Development of a longitudinal design for long-term investigation of intra-subject changes in sleep quality of older adults, with information extracted from pressure sensor array data collected in a period of approximately one year for each patient. With this scheme, the stability of sleep characteristics is determined for each subject. For this purpose, the trained patient-specific system was applied to the whole recorded data of each patient ranging from 8 to 12 months, on a day-by-day basis. Sleep characteristics such as central apnea index (CAI), bed occupancy (BO), the day clock (defined as 10 am to 10 pm) and the night clock (defined as 10 pm to 10 am) were extracted through PSM data that potentially can be used to track a patient's treatment progress in the long-term. The results are presented in Chapter 7, were published in a journal paper [12], and will be presented at a conference [13].

[12] H. Azimi, M. Bouchard, R. Goubran, and F. Knoefel, "Unobtrusive Screening of Central Sleep Apnea from Pressure Sensors Measurements: A Patient-Specific Longitudinal Study," *IEEE Trans. Instrum. Meas.*, pp. 1–1, 2020, doi: 10.1109/TIM.2020.2981111.

[13] F. Knoefel, H. Azimi, M. Bouchard, and R. Goubran, "Detecting Apnea in Community Dwelling Older Adults," presented at the *International Society of Gerontechnology (ISG) 12th World Conference of Gerontechnology,* Trondheim, Norway, Oct. 2020.

5. Comparing the performance of a sleep apnea algorithm in [9] between the Matlab classification environment and IBM Watson cloud classification services, to investigate the potential use of IBM Watson Analytics for the real-time analysis of health sensor data from PSM for large numbers of users. The result were presented in [14] and published in [15]. The full published paper of [15] can be found in Appendix B, as it is not included in the core of the thesis.

[14] H. Azimi et al., "Real Time Data Analytics using IBM Cloud for Sleep Apnea," presented at *the AGE-WELL's 5th Annual Conference*, Moncton, NB, Canada, Oct. 22, 2019.

[15] H. Azimi et al., "Cloud Processing of Bed Pressure Sensor Data to Detect Sleep Apnea Events," in *2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, Bari, Italy, 2020, p. 5.

6. Contributed to the development of an adaptive window length movement detection algorithm to identify movement onsets and offsets. This method was compared favorably to previous work on movement detection, and the results were published in [16].

[16] S. S. Gilakjani, H. Azimi, M. Bouchard, S. Bennett, R. A. Goubran, and F. Knoefel, "Movement detection with adaptive window length for unobtrusive bed-based pressure-sensor array," in *2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2017, pp. 355–360.

1.7 Thesis Structure

This thesis contains eight chapters beginning with some general aspects of SA in Chapter 2. In section 2.2, different types of sleep apnea are introduced, followed by risk factors in section 2.3. Symptoms of sleep apnea are considered in section 2.4, and the cost of this disorder is mentioned in section 2.5. Chapter 3 describes different experimental setups and the experiments that were undertaken for data collection. It includes data acquisition setup, as well as controlled and nocturnal experimental procedures. Chapter 4 studies fusion methods to extract breathing signal from PSM data. Chapter 5 describes the primary idea of detecting SA using RIP signals as a gold standard sensor for PSM. Chapter 6 compares machine learning approaches (including the method developed in Chapter 5) to find a good-performing method for automatically detecting CSA events from PSM signal. Chapter 7 evaluates the use of the algorithm for automatic detection of CSA events discussed in Chapter 6 in a long-term sleep monitoring setup. Finally, in Chapter 8, the major conclusions and future scope of the work are presented.

Chapter 2: Sleep Apnea & Sleep Monitoring Technologies

2.1 Introduction

This chapter details background information needed to fully understand the motives behind this thesis project and the previous work upon which this project is built.

2.2 Sleep Apnea

Sleep apnea (SA) is a potentially serious sleep disorder which is characterized by repeated periods of reduction or complete cessation of airflow and interrupted sleep. Clinically, SA is defined when there is a drop in the peak signal excursion by ≥ 90% of the pre-event baseline using an oronasal thermal sensor (diagnostic study), PAP device flow (titration study), or an alternative apnea sensor such as RIPsum, RIPflow, for ≥ 10 s. The common duration of the apneic event, though, is usually about 20 to 40 s [17], [18]. There are three forms of sleep apnea, which are defined as follows.

2.2.1 Central Sleep Apnea

Central sleep apnea (CSA) is characterized by a complete cessation of both respiratory movements and airflow, accompanied by 3% or more arterial oxygen desaturation. Figure 2.1 shows thorax (Red) and abdomen (Black) signals while CSA is happening, from St Vincent's University Hospital, Dublin Database.

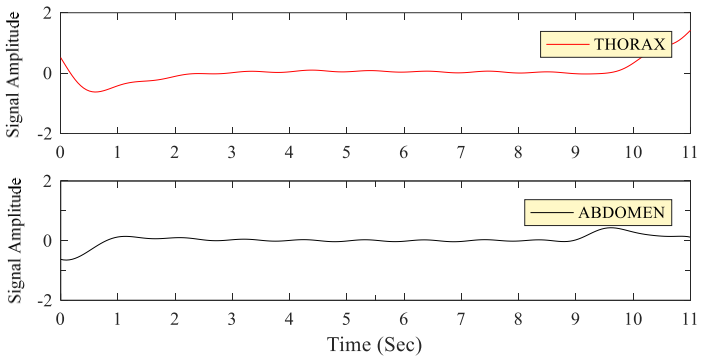


Figure 2.1 Sample of CSA, thorax signal (red), and abdomen signal (black).

2.2.2 Obstructive Sleep Apnea

Obstructive sleep apnea (OSA) is the more frequent pattern, characterized by the presence of abdominal and thoracic efforts for continuing breathing, while airflow completely stops, and accompanied with 3% or more arterial oxygen desaturation. A patient struggling with OSA has an upper airway that naturally falls closed during inspiration unless it is braced open by the contraction of dilator muscles that are activated at the beginning of inspiration. When the person falls asleep, these muscles become less active. The airway falls closed, obstructing breathing [19], resulting in signals as in Figure 2.2.

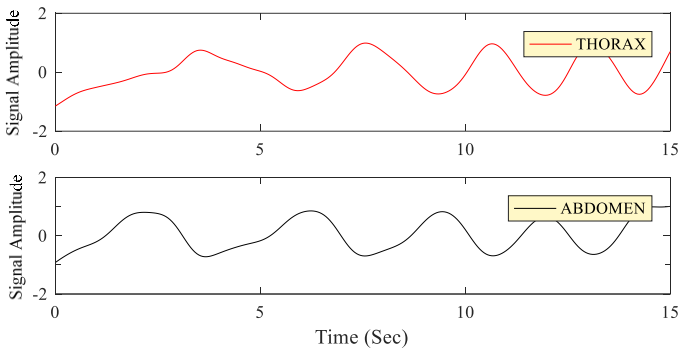


Figure 2.2 Sample of OSA, thorax signal (red), and abdomen signal (black).

Paradox Breathing

During normal breathing, in order to inhale, the diaphragm pushes downwards to decrease pressure in the chest cavity. Therefore, it allows the lungs to expand and fill with air. In most cases, in contrast to normal breathing where the diaphragm descends and the rib cage expands for inhalation, in OSA the rib cage and diaphragm move in opposite directions (i.e. diaphragm down/rib cage in, diaphragm up/rib cage out). [20], [21]. The same concept will happen even if a foreign object causes airway blockage. This type of breathing is known as paradox breathing. It prevents a patient from inhaling enough oxygen, which is important for many bodily functions. Paradox breathing also makes it difficult to exhale carbon dioxide, which is a waste product of the respiratory system. Figure 2.3 describes the performance of the respiratory system in normal and paradox breathing [22]. The main difference between these two types of breathing can be seen in the movement of the chest and abdomen. There is no airflow during paradox breathing since the airway is closed.

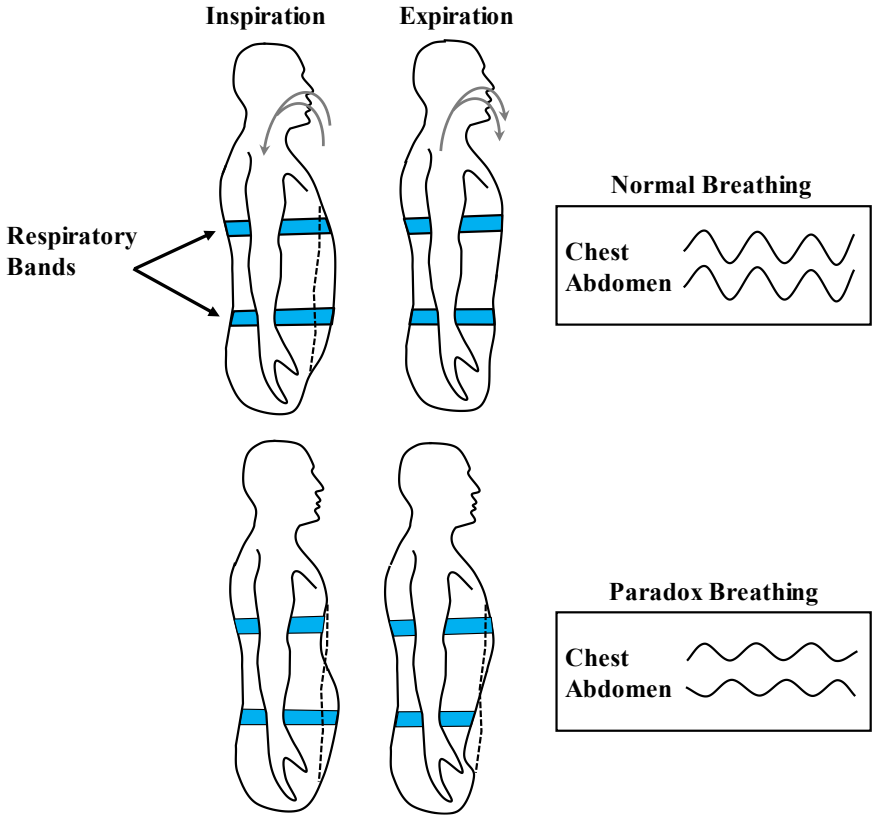


Figure 2.3 Performance of the respiratory system during normal and paradox breathing.

2.2.3 Mixed Sleep Apnea

Mixed sleep apnea (MSA) is a combination of the previous two types of apneas, defined by a CSA followed by an OSA [23], as shown in Figure 2.4.

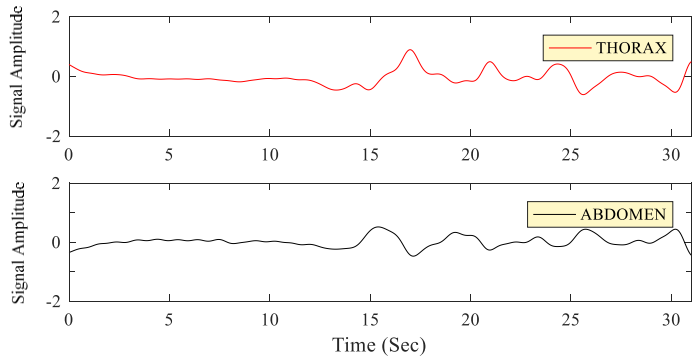


Figure 2.4 Sample of MSA, thorax signal (red), and abdomen signal (black).

2.3 Sleep Hypopnea

Hypopnea is defined when there is a drop in the peak signal excursion by ≥ 30% of pre-event baseline using nasal pressure (diagnostic study), PAP device flow (titration study), or an alternative sensor, for ≥ 10 s in association with either ≥ 3% arterial oxygen desaturation or arousal [17]. Hypopnea simply means abnormally slow or shallow breathing, while apnea means periods of no breathing. This abnormality is not the subject of this thesis and it is defined here for clarity only.

2.4 Risk Factors

Based on the Adult Obstructive Sleep Apnea Task Force of the American Academy of Sleep Medicine (AASM), patients at high risk for OSA include individuals with congestive heart failure, atrial fibrillation, treatment-refractory hypertension, type 2 diabetes, stroke, nocturnal dysrhythmias, pulmonary hypertension, obese people, and those being evaluated for bariatric surgery [24]. Obesity is a well-known risk factor for OSA [25]–[27]. However, the results in [28] showed that risk factors for OSA differ between men and women. According to this study, increasing body mass index (BMI) and age were the most important risk factors for OSA in men and women, respectively. Besides, it indicated that atrial fibrillation is a risk factor for CSA but not for OSA, whereas hypocapnia is a risk factor for CSA in both sexes.

Various anatomical factors can cause physical obstruction of the airway [29], such as enlarged tonsils [30], enlarged uvula [31], increased tongue size [32] and abnormal craniofacial morphology [33]. The effect of certain health behaviors such as cigarette smoking and alcohol on sleep apnea was also considered in several works [34]–[36]. Based on these studies, cigarette smokers are at greater risk for sleep-disordered breathing than those who never smoke. Heavy smokers have the greatest risk, while former smokers are not at increased risk for sleep-disordered breathing.

Alcohol consumption may cause or exacerbate OSA, and this may especially be the case with alcohol consumed shortly before bedtime.

2.5 Symptoms

The most common symptom that SA patients suffer from is excessive daytime sleepiness (EDS). Clinically, EDS's features include abnormal daytime tiredness, reduced wakefulness, and vigilance [37]. However, considering the degree of EDS, it may not correlate with the severity of SA measured by the Apnea-hypopnea index (AHI) (number of apnea-hypopnea events per hour of sleep) [38]. Snoring is part of the spectrum of sleep-disordered breathing that may be a symptom of OSA, but not all patients who snore have clinically significant sleep apnea. Snoring may be present in 30% to 50% of the general adult population, whereas only 2% of women and 4% of men have clinically significant OSA [39]. Morning headache, nocturia, memory loss, and decreased concentration can be mentioned as other symptoms of SA [24], [40], [41].

2.6 Cost of Sleep Apnea

AASM estimated that more than 29 million United States adults suffer from moderate to severe OSA, with an estimated 80% of those individuals living unaware and undiagnosed, contributing to more than $149.6 billion in healthcare and other costs in 2015 [42]. In Canada, an estimated 858,900 (3%) Canadian adults (18 years and older) reported being told by a health professional that they have SA [43]. In addition to those who reported being diagnosed with sleep apnea, over 1 in 4 (26%) adults reported symptoms and risk factors that are associated with a high risk of having or developing obstructive sleep apnea. It is estimated that the associated cost of diagnosing and treating every adult suffering from OSA in the United States would be $50 billion. However, it would result in saving approximately up to $100 billion [42]. Therefore, in the general

population, early diagnosis and screening of people struggling with this disorder not only can improve the efficiency of their day-to-day life, but it can also reduce future health service usage.

2.7 Related Works for Detecting and Diagnosing SA

The diagnosis of SA starts with a sleep history that is typically obtained in one of three settings:

1. As part of routine health maintenance evaluation.
2. As part of an evaluation of symptoms of OSA.
3. As part of a comprehensive evaluation of patients at high risk for SA.

Based on the sleep history of patients, those patients with high risk should have the diagnosis confirmed and severity determined by objective testing [1]. Apnea-Hypopnea Index (AHI) is the key measure used for case identification, quantifying disease severity, and defining disease prevalence in normal and clinical populations [44]. It is the number of apneas or hypopneas recorded during the study per hour of sleep. Based on the Chicago criteria, an AHI<5 is referred to as normal, a 5≤AHI<15 means mild, a 15≤AHI<30 corresponds to moderate, and an AHI ≥30 is referred to as severe [45]. After obtaining the patient's AHI by performing objective testing, sleep experts make a decision based on the degree of AHI. If an AHI is less than 10, and the patient lacks important sleepiness, no further investigation and treatment for sleep apnea are needed. However, the patient is recommended to lose weight or pursue treatment for snoring. If an AHI is less than 10, but the patient suffers from important daytime symptoms, further investigation and repeating objective testing such as PSG and home monitoring should be considered. If an AHI is more than 10 and less than 30, based on the severity of a patient's daytime symptoms, different treatments such as a trial of continuous positive airway pressure (CPAP), a dental appliance, or surgery will be followed. For an AHI of more than 30, if the patient is very overweight, arterial blood gases should be checked, and titration of CPAP during polysomnography is preferable. For

an extreme condition such as severe symptoms or complications of sleep apnea and when all else fails, tracheostomy must be considered [19].

According to the classification proposed by the AASM, sleep diagnostic devices for objective testing could be categorized into four types. Type-1 is the standard PSG system which includes conventional bio-signals, allowing acquisition of physiologic data from different organs, such as brain activity (EEG), muscle activity (EMG), eye movement (EOG), cardiac function (ECG, heart rate variability), as well as respiratory parameters such as airflow, respiratory movement/effort, and oxygen saturation. This method has been proven accurate with a low failure rate because technical staff attend the study. PSG, however, is considered relatively expensive and technically complex. The analysis of the PSG represents a significant cost both in time and in effort for the clinician. This high cost associated with the manual and visual review of the PSG can eventually degenerate into loss of quality in the analysis due to the accumulated tiredness throughout the revision and because of the complexity of the task itself. Type-2, Type-3 and Type-4 devices, known as Portable monitoring (PM), include a minimum of seven signal channels of bio-signals (from EEG, EOG, chin EMG, ECG or heart rate, airflow, respiratory effort, and oxygen saturation), four signal channels (from ventilation or airflow, at least 2 channels of respiratory movement or respiratory movement and airflow, heart rate or ECG, and oxygen saturation), and one signal channel including oxygen saturation or airflow [46]–[48] respectively. The signals alone are effective in detecting SA in the case of binary decision-making, whereas multiple signals are good for multi-class detection [49]. For instance, to distinguish among different types of SA, including OSA, CSA, and MSA, the presence of respiratory movement/effort along with other signals is necessary.

Commonly for all types of sleep diagnostic devices, pattern recognition techniques can be used to extract information from recorded physiological signals and thus help sleep clinicians diagnose SA disorder and its severity. The ultimate goal of the pattern recognition problem is to classify the objects into a number of categories or classes. For this purpose, the role of the classification models is to divide the feature space into disjoint regions assigned to class labels. Depending on the type of classifiers, they may provide different kinds of borders between the regions. Generally, a learning machine is a function which determines a hypothesis from the data:

$$f : x \rightarrow y \tag{2.1}$$

The function transforms objects from the data domain x to the set y of possible target values. The learning model, f usually depends on some adaptive parameters. In this context, learning can be described as a process in which a learning algorithm searches for parameters of the model, to solve a given task.

In the last few decades, by applying signal processing and pattern recognition approaches, several clinically useful sensor modalities and techniques have been developed to diagnose SA disorder. Those modalities and techniques are described in the following sub-sections. The subsections are categorized according to the machine learning techniques used to detect SA events from recorded physiological signals. They are: traditional classification approaches (e.g., Naïve Bayes, Linear Discriminant Analysis (LDA), neural networks, kernel methods, similarity-based approaches, and decision trees) and Deep Learning (DL) methods.

2.7.1 Traditional Classification Approaches

In this subsection, for equitable comparison between a great variety of signals and devices that have adopted traditional classification models to detect SA events, the articles are categorized and summarized based on the type of sleep diagnostic device. Each sub-section (including 2.7.1.1,

2.7.1.2 and 2.7.1.3) contains information such as the methodology and results of articles adopting associated device type and signals.

2.7.1.1 Type-2 & Type-3 Devices, Multichannel Respiratory Activity

Here, an overview of event detection approaches using at least two respiratory/neurophysiological bio-signals is presented. In [50], Large Memory Storage And Retrieval (LAMSTAR) artificial neural networks were used to predict apnea and hypopnea. Signals included in this study were EEG, heart rate variability, nasal pressure, oro-nasal temperature, submental EMG, and right (EOG). Sensitivity and specificity up to 80.6 +/- 5.6 % and 72.8 +/- 6.6 % were respectively achieved for apnea detection. Several studies have used fuzzy interference systems (FIS) to detect SA from the combination of airflow, oxygen saturation, and thoracic and abdominal signals [51]–[53]. A support vector machine (SVM) classifier was used in [54] for the minute-by-minute classification of OSA. An accuracy of 82.4% (sensitivity: 69.9%, specificity: 91.4%) was achieved using polynomial kernels and features extracted from ECG, oxygen saturation, thoracic and abdominal signals. Considering that SA events have different patterns in comparison to normal breathing, in [55] an artificial neural network (ANN) was used to detect apnea events in respiratory signals, including thoracic and abdominal efforts and airflow signals, with an accuracy of 90%. In a few studies, wavelet features from respiratory signals (thoracic and abdominal efforts and airflow signal) were fed to classification approaches such as ANN and ensemble methods for SA detection [56], [57]. The advantage of thoracoabdominal bands combined with airflow is that the signal registered by the bands is not dependent on the patient having to breathe solely through the nose. Moreover, thoracoabdominal bands are already widely used in polysomnography to evaluate whether apneic events are obstructive or central [58].

Many studies have used two types of signals for the goal of SA detection. They are either airflow and oxygen saturation signals or thoracic/abdominal signals with oxygen saturation signals. The respiratory and oximetry signals are the most effective physiological signals on which a reliable and accurate SA detection system is based. In [59], a combination of thoracic and abdominal signals and oxygen saturation were used as inputs of an amplitude-based analyzer to detect SA. They aimed to automate the manual approach of SA detection by designing a decision-making algorithm. The algorithm achieved detection sensitivity and specificity of 73.6, and 90.8% respectively. The combination of airflow and oxygen saturation signals are the most common signals used in the literature to detect SA, using different approaches such as time-delay neural network (TDNN) [60], threshold-based algorithms [61], [62], and fuzzy rules [63].

2.7.1.2 Type-4 Devices, Single Respiratory Activity

In this section, SA detection based on a single respiratory signal of airflow, thoracic, and abdominal efforts, or oxygen saturation is discussed. These works followed a common methodology: detecting the effects caused by each apnea and hypopnea in the signal under study, scoring them as apneic-related events, and deriving the corresponding diagnostic index.

Airflow-Based Detection

The airflow signal is the most important respiratory signal used to detect SA because the impact of airway obstruction is reflected in this signal. Standard sleep studies use the features of air temperature and nasal pressure as estimates of airflow [64].

Most of the airflow-based detection studies distinguished between apneic and normal events using binary classification methods. A real-time adaptive apnea event detection method was proposed in [65] using a two-stage classifier model. An SVM classifier was used to distinguish

normal and abnormal (apneic) episodes. Average accuracies of apnea detection, when compared with polysomnography-based respective indices on unseen subjects during online tests, were found to be 94.9%. A FIS was developed in [66] to detect obstructive sleep apnea (OSA) by analyzing the airflow signal. In this study, a fuzzy logic system processed two parameters, i.e., the normalized area and the standard deviation of consecutive 3-second intervals of baseline, to detect SA events with an accuracy of 83%. In another study [67], data from the airflow signal was first categorized and labeled into three classes by applying the fuzzy engine: normal breathing, abnormal, and not-sure events. The not-sure class contained events for less than 2 s. After eliminating outliers (deep breathing with high amplitude) from normal breathing class, all three classes were fed to a center of gravity clustering engine for final classification. The database used in the study included ten subjects, and the result was illustrated for each patient separately with high specificity and sensitivity over 90%.

Although the airflow signal plays an essential role in diagnosing SA, the type of sensor used to record airflow may cause the signal to be subject to certain limitations such as inaccurate measurement caused by a bad seal between the mask and the patient (if applicable), inaccurate full-range detection of flow values, slow response time [68], patient discomfort, etc. These factors can reduce the reliability of SA detection methods based on airflow signals [46].

Thoracic & Abdominal Efforts-Based Detection

Thoracic & abdominal efforts are part of exhalation and inhalation. According to AASM, these effort signals especially extracted from respiratory inductance plethysmography (RIP) sensors can be used as alternative signals to detect SA [17]. RIP is a method of evaluating pulmonary ventilation by measuring the movement of the chest and abdominal wall. It has been regarded highly by the AASM due to its great accuracy, sensitivity, and high patient safety [69]. Although

RIP accuracy and precision are limited in obese patients and significantly decrease during sleep, this sensor is still clinically useful [70]. SA detection using only thoracic and abdominal signals together or separately has rarely been done. In [71], four clinical features were each extracted from a 1-minute epoch of a thoracic signal, and a KNN classifier was used to detect apneic episodes in eight eight-hour patient records collected from the Physionet database. The classifier achieved a specificity of 88.1% and a sensitivity of 95.7%. Two studies have used an abdominal signal after the initial detection of SA by airflow signal to distinguish OSA, CSA, and MSA [72], [73].

Few studies have used the combination of both thoracic and abdominal signals, especially for detecting OSA, where in most cases there will be a phase difference between these two signals while OSA is happening. By breath-by-breath analysis of the local maximum and minimum points of the signal magnitude in [74], the phase difference between thoracic and abdominal signals was detected. OSA and CSA events were distinguished with an accuracy of 90%. The method was applied to a small database, containing only six PSG recordings. In [23], abdominal and thoracic respiration signals were separated into spectral components by using a multi-resolution discrete wavelet transform (DWT). Then the result of DWT was fed to an ANN to classify OSA, CSA, and MSA. The purpose of this study was classifying different types of apnea, assuming apneic events had already been detected. The accuracies obtained for each one of the three classes were 85.35% (OSA), 94.74% (CSA), and 80.42% (MSA). The global accuracy obtained was 86.84%. An adaptive threshold-based method was adopted in [75] to detect SA. There it was assumed that the amplitudes of thoracic and abdominal signals decreased not only during CSA but also during OSA. The accuracy of the method was evaluated by the correlation between the algorithm scored AHI values, and the technician scored PSG AHI values.

Oxygen Saturation-Based Detection

Oxygen saturation measures the percentage of hemoglobin binding sites in the bloodstream occupied by oxygen [76]. Pulse oximetry monitors the saturation of peripheral oxygen (SpO2) by shining red and infrared light through a fingertip, ear, or toe. The amount of red or infrared light that is absorbed corresponds to the concentration of oxygenated hemoglobin and deoxyhemoglobin in the blood, and therefore, the oxygen concentration in the blood can be determined [77]. Despite being prone to error and false desaturation measurements due to motion artifact, noise, or missed readings, the pulse oximeter has proven to be a valuable tool in detecting and diagnosing sleep apnea. In order to reduce errors, pulse oximeters average readings and report oxygen levels every 3–12 s, which may cause a delay in alarms or monitoring [46].

In [78], it was stated that the mean of the lowest oxygen saturation of healthy subjects for different age groups is 90.4%. During SA, oxygen saturation drops by up to 3-4% from the normal percentage [4], [79]. It is worth mentioning that a person suffering from SA, could easily stay in the normal range of oxygen saturation during an apnea event (e.g. going from normal 99% to 95% during SA) and consequently the disorder could not be diagnosed based on this criteria.

Like the airflow signal, oxygen saturation is one of the main signals to detect SA. As a solo signal, oxygen saturation has been mostly used to characterize OSA positive and OSA negative patients, applying different approaches such as spectral characteristics of heart rate obtained by nocturnal pulse oximetry [80], nonlinear analysis of nocturnal oximetry signal [81], [82], and classification [83]–[85]. However, it should be noted that oxygen desaturation does not always follow after SA [86].

2.7.1.3 Environmental Sleep Monitoring Technologies

In order to record the data at home without any personnel, the data must be easy to record, especially for those without a high degree of technical knowledge. In [87], the comparison between home PSG (HoPSG) and laboratory PSG (LabPSG) for the diagnosis of SA has shown that HoPSG has a higher failure rate (20%) than LabPSG (5%) either due to lost data or poor quality data. There are more difficulties in obtaining technically accurate home night studies. In the study, discordant results (i.e. more than ±10 events per hour of sleep difference) have been observed in patients with poor airflow signal at home (11%) because the patients had to set up the thermistors themselves. Some did not adjust the device as recommended, or others damaged the equipment. Poor quality or lost data lead to repeated night studies and increasing cost, and it may have lead to false or inaccurate results. Therefore, the study suggested that sensors must be designed to be solid, easy to handle, and suited to home studies. It is important to feel comfortable during sleep. Therefore, it must be possible to record the signals without restricting with gauges and cables patient movements during the night [88]. Sleep apnea, breathing rate, heart rate, and movement during sleep can be detected even without attaching sensors directly to the body by installing sensors in the user's native sleep environment. These types of devices have the advantage of being unobtrusive and adequate for longer-term monitoring without the user's intervention [89].

In [90], a non-contact radio-frequency sensor called "SleepMinder" was placed next to a bed and was set-up to continuously measure the bio-motion due to breathing and body-movement during sleep. It uses 5.8 GHz Doppler radar with a modulation system that limits both the maximum and minimum range [91]. The device applies phase demodulation, amplitude, and correlation-based signal-processing methods to detect SA. Comparing to PSG recordings of 129 subjects, AHI estimated by the device had a correlation of 91% and could detect SA (AHI>15)

with a sensitivity of 89% and a specificity of 92%. In [92], a thin foam mattress called Sonomat, with only four polyvinylidene fluoride film-based sensors having dimensions that fit a standard single sized bed was used to diagnose SA. The mattress generated a voltage when deformed due to respiratory and body movements. It was also able to record breathing sound as an alternative measure of airflow. Apneic events captured by the mattress were visually and aurally scored by a technician using breath sound trace and compared to the result of relative PSG recordings. The breathing movement signal was used to visually distinguish different types of apneic events after primary annotation of them using breathing sound signals. Exploring the signal captured by Sonomat, the technician could identify more than 93% of PSG defined respiratory events and the absence of them on 91% of occasions. No computer-based methodology was developed and designed in the paper to automate the process of detecting apneic events and the main goal was to evaluate the ability of device in recording of apneic events.

In [93], a force-coupling pad equipped with a ballistocardiography (BCG) system was installed on top of the mattress of the bed to detect minute forces generated during cardiac contraction and relaxation, and body movement from the respiratory effort and postural changes. The PSG of participants was recorded while the force-coupling pad was installed on the bed. A sensitivity of 89.2% and specificity of 94.6% were attained in the automated detection of SA. However, no details about the apnea detection method were discussed in the paper. In [94], a shirt made of elastic fibers (to monitor respiration activity) and conductive metal-based fibers (to monitor heart electrical activity (ECG)) was used to diagnose OSA. Parts of the respiratory signal containing movement artifacts were considered as outliers and removed from the analysis. A threshold on the amplitude of the respiratory signal extracted from the shirt was then adopted to detect SA. By adjusting the threshold, “a sensibility (sic) of 98.1% and specificity of 84.3% were achieved [94].”

In [95], an air-mattress with a balancing tube was used to noninvasively monitor the signals of heartbeat and respiration, as well as the events of snoring, sleep apnea, and body movement. The proposed system consisted of multiple cylindrical air cells, two sensor cells, and 18 support cells. The physiological signals were measured by the change in the pressure difference between the sensor cells. A simple threshold-based detection algorithm based on a variance analysis with a moving window technique detected simulated apnea events. Sensitivity and positive predictive value (PPV) of the detection algorithm for sleep apnea events were 93% and 88%, respectively. No validation results in terms of comparison with PSG data were presented in the study.

It was shown in [96] that SVM and NN classifiers with depth video and audio signals recorded by a Microsoft Kinect™ camera during sleep could be used to detect and differentiate SA events, including OSA, CSA, and MSA.

Respiratory events were detected in [97] from a series of digital images provided by a digital video camera. The technology was based on the volume of air circulating into the lungs, which is proportional to the patient's movement while breathing. The proposed apnea detection algorithm compared signal variations from the mean respiratory movement used as a reference point for the magnitude detected in the previous minutes. In the analysis of 50 patients, the system was found to have a sensitivity of 100% and specificity of 83%. The PPV and negative predictive value (NPV) were 97% and 100%, respectively.

The use of PSM to monitor vital signs has been an active research area, mainly due to the unobtrusive and non-contact nature of the sensors. The pressure sensor array is well-suited as a home-based device because it does not require any involvement from the patient for installation and data acquisition. The array is placed unobtrusively under the mattress (i.e. without contact to the body, and is too thin to be felt). Therefore, it can be a good alternative to a HoPSG device. The

data can be obtained in a completely natural home environment without the involvement of external factors such as discomfort caused by electrodes, limitation of movements by gauges and cables, potential psychological consequences of being under scrutiny, and a change of environment. The absence of these factors can provide patient sleep characteristics that are closer to real-life sleep, especially for SA diagnosis. In this thesis, PSMs are used to detect respiration and apnea events. The technical characteristic and structure of the pressure sensors are presented in Chapter 3.

2.7.2 Deep Learning

Traditionally, to perform a specific task without using explicit instructions and just by relying on patterns and inference, raw data needs to be transformed before it is fed to a system, and this transformation is performed or decided by a human. In most cases, the system consists of three sequential processing steps: (1) preprocessing, (2) feature extraction, (3) classification [98]. The first two steps extract information from the data establishing the analysis of the system. One can say that manually extracting features requires specific expert knowledge and the process is both time-consuming and domain-specific [99]. Feature extraction algorithms may fail to extract the most relevant information from the data. Furthermore, the amount of information fed to the traditional machine learning algorithms has to be limited since these algorithms would perform worse when dealing with high dimensional inputs, and therefore the decision quality may suffer due to restricted information [98]. Accordingly, automated extraction and selection of task-specific features can be valuable to solve complex real-world problems. Deep learning (DL) has revolutionized the data analytics modeling concept from “expert-driven” feature engineering to “data-driven” feature construction. It has had a great influence on many data analytics applications,

such as speech recognition, image classification, computer vision, and natural language processing [100]. Two of the most popular DL networks are:

1. Convolutional neural network (CNN)

CNN is a feed-forward network, inspired by the neurobiological model of the visual cortex, where the cells are sensitive to small regions of the visual field. A CNN comprises three main parts:

- **Convolution Layer**

A convolutional layer consists of multiple maps of neurons, known as feature maps or filters. Neurons are the basic processing units in a neural network. Each neuron within a feature map is only connected to a local patch of neurons in the previous layer, known as the receptive field. The input data are convolved by different convolutional filters via shifting the receptive fields progressively. The convolution operation acts as a feature extractor by learning from the diverse input signals. The convolutional filters share the same parameters in every small portion of the image, largely reducing the number of hyperparameters in the model.

- **Pooling Layer**

The pooling layer can be placed between convolution layers to take the mean, max, or other statistics of the features at various locations in the feature maps, and therefore reduce the variance and capture essential features. These convolution and pooling operations can be used repeatedly, with each layer learning to detect different features in the input data.

- **Classification Layer**

Finally, in the last layer of a CNN, a fully-connected classifier is used for the classification of the features extracted by the previous convolutional and pooling layers. This layer provides a weighted sum of all the outputs from the previous layer to determine a specific target output. Fully-connected indicates that each neuron in the previous layer is connected to all the neurons in the

current layer and each connection has a specific weight. The total number of fully-connected neurons in the final layer specifies the number of classes [98], [99], [101], [102].

2. Long short-term memory (LSTM)

LSTM is a type of recurrent neural network (RNN) that can learn long-term dependencies between time steps of sequence data. Unlike CNN, an LSTM can remember the state of the network between predictions. The essential components of an LSTM network are:

- **Sequence Input Layer**

A sequence input layer incorporates time-series data into the network.

- **LSTM Layer**

An LSTM layer learns long-term dependencies between time steps of sequence data over time. The layer contains hidden units providing inputs to memory cells and corresponding gate units. All units (except for gate units) have directed connections (serve as inputs) to all units in the layer above.

Transfer Learning

In many machine learning and data mining algorithms, it has been assumed that the training and future data must be in the same feature space and have the same distribution. However, sometimes for developing an optimized system especially in the field of DL, there is only sufficient training data in one domain of interest while a classification task may be required in another domain of interest. In such a case, transfer learning has become a new learning framework to address the problem. DL models have proven that they are highly reusable, especially convolutional architectures and CNN models where inputs are very high dimensional, such as images, text or audio. Normally, transfer learning is done by transferring the weights of some layers from a pre-

trained DL network on a source task to a second network used for another task [101]. Knowledge transfer can significantly improve the performance of learning, bypassing the problem of training data scarcity [103]. The idea behind this approach is based on the fact that the human brain can discover the underlying structure in previously learned knowledge and transfer it to new tasks [104].

There are two ways to use DL models as pre-trained networks in transfer learning, namely, feature extraction and fine-tuning. The first approach consists of using a pre-trained network to extract features from a new dataset. These features are then fed to a new machine learning algorithm that is trained from scratch. The second approach consists of slightly adjusting the pre-trained network together with the new classifier in the top layers [105].

Given models with different architecture for classifying images such as AlexNet, ZFNet, GoogLeNet, ResNet [106] and VGGNet [100],[106] from the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC), signals can be graphically represented as an image by time-frequency methods to be able to use transfer learning with any of these models, in a fashion similar to image classification. Time-frequency methods including Short-Time Fourier Transform (STFT), DWT and Hilbert-Huang transform (HHT) [108] represent a signal in both the time and frequency domains simultaneously.

Application of DL for SA Detection

Recently, DL has been favored in all fields, whether in the medical or non-medical areas, as it does not require engineering knowledge to extract features. SA detection has not been an exception to this. Multi-class classification of sleep apnea/hypopnea events has been proposed based on a three-layer LSTM model and using PPG signals. The PPG signals were measured by the nocturnal polysomnography with 7 h from 82 patients suffering from sleep apnea. The results of the LSTM

model revealed the following performances: a PPV of 94.16% for normal events, 81.38% for apnea events, and 97.92% for hypopnea events; sensitivity of 86.03% for normal events, 91.24% for apnea events, and 99.38% for hypopnea events. Interestingly, the method had a higher performance of hypopnea classification, which had been a drawback of previous studies mentioned in the article [109]. In another study for data recorded by SpO2 sensor, the detection of sleep apnea has been performed by a CNN model built from scratch. The total number of samples presented for the study from SpO2 sensors of 50 patients was 90,000. Applying cross-entropy cost function, the overall accuracy of binary SA classification was 91.3% with a loss rate of 2.3. The CNN model outperformed the other models including LDA, SVM, bagging representation tree, and ANN [110]. In [102], DL approaches were validated with an aim to find the optimal method for automatic detection of SA events from ECG signals. The ECG signals collected from 86 patients were pre-processed, normalized, and segmented into 10 s. intervals. Six deep learning approaches were then designed and implemented for automatic detection of SA events, including deep neural network (DNN), one-dimensional (1D)-CNN, two-dimensional (2D)-CNN, LSTM, and gated-recurrent unit (GRU). The accuracy of the best-performing model was 99.0%.

Unlike the previously mentioned papers, in [107] transfer learning was applied, using pre-trained CNN models including AlexNet and VGG-16 for feature extraction (and not as a classifier) from spectrogram images of pulse transition time (PTT) signals. With the features obtained, SA patients and healthy individuals were classified by SVM and KNN algorithms. Among the different combinations of classifiers and feature extractors, an highest accuracy of 92.64% ± 1.44 was achieved with the SVM algorithm as a classifier and the VGG16 CNN as a feature extractor.

2.8 SA Treatment

Treatment of SA can range from conservative methods to more radical approaches, which are briefly introduced in this section.

2.8.1 Positive Airway Pressure (PAP) Device

A positive airway pressure (PAP) device usually consists of three major parts: a positive pressure generator, a nasal or oral interface such as a mask, and a tube connecting the two parts. The pressure generator is a fan or turbine which is used to apply external pressure to the patient's upper airway [46]. PAP devices are normally divided into four different types. The first developed PAP device was continuous PAP (CPAP) [111], widely accepted as the gold standard for the treatment of OSA. While CPAP is effective at treating OSA, it may not suppress CSA, especially in patients with heart failure [112]. Other types of PAP devices are bi-level positive airway pressure (BPAP) with bi-level pressure by manual titration, auto-titrating positive airway pressure (APAP) with adjustable pressure, adaptive servo-ventilation (ASV) initially used for CSA patients with pressure support. PAP devices are extremely effective when used correctly. However, their major challenge remains the patients' intolerance and non-adherence.

2.8.2 Oral Appliances (OA)

OAs physically alter the mandible, tongue, or soft palate during sleep, to prevent the collapse of upper airway muscles [113]. OAs have become more popular, as they are found to be more comfortable than CPAP and have higher adherence rates. Still, in comparison to CPAP, CPAP has higher efficacy. Therefore, OAs tend to be a secondary form of treatment for OSA and are recommended when a patient becomes noncompliant to CPAP [114]. There are three types of OAs:

soft palate lifters (SPLs), tongue-retaining devices (TRDs), and mandibular advancement appliances (MAAs or OA).

2.8.3 Invasive Methods

Generally, there are two classes of invasive methods; the first is upper airway surgery, and the second uses an implantable hypoglossal nerve stimulation (HGNS) device. The device has an acceptable performance with regards to patient safety, respiration sensors, lead failures, and stimulation failures [115]. In order to undergo HGNS, patients require surgery to implant an electrical stimulator. A stimulation electrode is placed on the hypoglossal nerve, which stimulates tongue-protrusion. Stimulation is delivered just before inspiration, when the upper airway is most prone to narrowing and collapse [115], [116].

Recently, stimulation of the phrenic nerve has also been explored to treat sleep apnea. In this method, the diaphragm is stimulated to restore a normal, physiological breathing pattern during sleep [117]. Although phrenic nerve stimulation is still in development and trials, it has shown promise as an alternative to CPAP [118].

2.8.4 Other

Nasal expiratory positive airway pressure (NEPAP) is a recent alternative to CPAP. In NEPAP, two nasal valves are placed on the nostrils. They provide low resistance during inhalation and high resistance during exhalation. The high resistance causes a positive pressure to build up through the airway and eventually open the soft tissues up [119].

Positional OSA is a subset of OSA that can be defined by a two-fold increase in AHI while in a supine position versus a non-supine position. Based on this definition, about half of all OSA

patients have positional OSA. Techniques and devices such as alarm systems, pillows with straps, tennis ball technique, and vibrating devices are adopted to treat positional OSA [46].

Another potential method for controlling breathing abnormalities is the constant inhalation of CO_2. Theoretically, increasing CO_2 should stabilize central respiratory motor output, as well as improve upper airway caliber and prevent upper airway obstruction [46]. While there are some very promising studies regarding the use of inspired CO_2, there are also several potentially dangerous safety concerns that should be explored in longer-term studies before using this as a possible treatment for SA [120].

2.9 Summary

In this chapter, the definition and basic concepts related to SA have been reviewed, with the aim being to provide a general knowledge of sleep apnea for engineering researchers. It is impossible to mention and review all the literature related to apnea event detection for a topic as wide as sleep apnea, but the most important ones were reviewed and studied.

Although PSG is the most accurate clinical test to diagnose SA, the analysis of the PSG recordings needs manual and visual review conducted by a sleep technician, and it is significantly time-consuming. Several clinically useful sensor modalities and techniques have been developed to be a suitable alternative to PSG. According to the literature, the most common and effective physiological signals for the goal of SA detection are airflow and oxygen saturation signals, which are used either together or solely. Although these signals are considered valuable in detecting and diagnosing SA, depending on the type of sensor for data acquisition, their use comes with limitations and disadvantages such as sensor losses, slow response time, inaccurate measurement, patient discomfort, and being obtrusive [121].

Statistics have revealed that early diagnosis and treatment of SA could significantly reduce healthcare costs. Home monitoring can be an effective approach to the timely diagnosis of SA. Moreover, the idea can be easily accepted by patients since they can remain in the comfort of their home and therefore may sleep better in an environment in which they are familiar with [122]. Consequently, the most discernible impact that familiar-surroundings can have is on minimizing FNE which is one of the well-known disadvantages of LabPSG [121].

In order to monitor individuals who lack technical knowledge at home, the data must be easy to record and reliable [1], [122]. Poor quality or lost data is not cost-effective and can lead to repeated night studies [87], [121]. Home monitoring should be carried out in such a way that along with protecting privacy, the patient feels comfortable during sleep, without restricted movements (caused by gauges and cables) [123].

In this thesis, PSM has been chosen as the main device for home monitoring and SA detection. The extracted signals from PSM can represent the respiratory movements of the upper body. The advantage of PSM over other devices and sensor modalities is that by keeping personal privacy, it can be unobtrusively placed under the mattress of the patient. Once the device is installed, it does not need the user's intervention, and it is adequate for longer-term monitoring.

In previous studies, depending on the device and the corresponding data, different algorithms, and statistical models including traditional classification approaches (e.g., SVM, ANN, FIS and KNN) and DL networks have been applied to detect SA. In the following chapters, a comprehensive comparison is performed between the so-called canonical supervised learning approach and DL methods for sleep assessment and CSA event detection based on the characteristics of the data extracted from PSM. In different chapters, several factors such as the properties and amount of data required, the use of a generative model (to model the actual

distribution of each class) or a discriminative model (to create a decision boundary between the classes) [124] are taken into consideration to choose the right computer-based approach. The objective is to find the best-performing method for automatic detection of CSA from PSM signals.

According to AASM, a device for home monitoring must allow sleep technicians/technologists to look at the raw data and extracted events, and if necessary to edit the automated scoring [1]. Ideally, if a home monitoring device is to replace or assist the standard PSG, it should be able to provide the required information to relevant authorities for visualizing the events (or related changes caused by apneic events in output signals of the device) and therefore understanding patients' sleep characteristics better. In the reviewed studies, there is no discussion of this issue, and the evaluation of the proposed approach has been done in terms of the AHI or segment-by-segment approach through blind segmentation, where one apneic event can be covered by more than one segment. The systems designed in the following chapters allow detection and visualization of the apneic events and comparing them to a ground truth on an event-by-event basis. Therefore, the systems can provide the opportunity for a sleep technician to view events and perhaps edit the scoring if needed.

Along with obesity, age is one of the most important risk factors for SA, and the main source of data used for this project is from older adults ranging from age 65 and beyond, living in affordable seniors housing.

Chapter 3: Experimental Setup & Procedures

3.1 Introduction

The goal of this chapter is to provide information about the technology and general setup for data collection, including PSM and other auxiliary equipment. In this thesis, experiments and data collection were performed in both controlled and uncontrolled manners. A controlled dataset was captured from healthy adults in controlled environments, while an uncontrolled dataset was collected from older adult patients in their homes.

3.2 Ethics

For all human experimentation, the research ethics boards of both the University of Ottawa and Carleton University approved the study. All subjects provided consent to gather data and present results.

3.3 Equipment

3.3.1 Pressure Sensitive Mat

The pressure-sensitive mats (PSM) used in all experiments were the "Bed Occupancy Sensors" (BOS) model from S4 Sensors Inc. (Victoria, B.C.). They are composed of Kinotex^{TM} sensors, which are fiber optic pressure sensors with a high sensitivity to pressure changes [125].

Each optic sensor contains two plastic optical fibers, an emitter, and a receiver. The emitter and receiver strands are beside each other in a cellular foam structure, shown in Figure 3.1. Light-emitting diode (LED) light is emitted through the emitter cable to the foam surface. When pressure is applied to a sensor, changing the pressure on the foam surface results in changes in the intensity of the light. The light returns to a photo-diode through a receiver cable. The photodiode converts the light to an output voltage which varies with the pressure being used on the sensor.

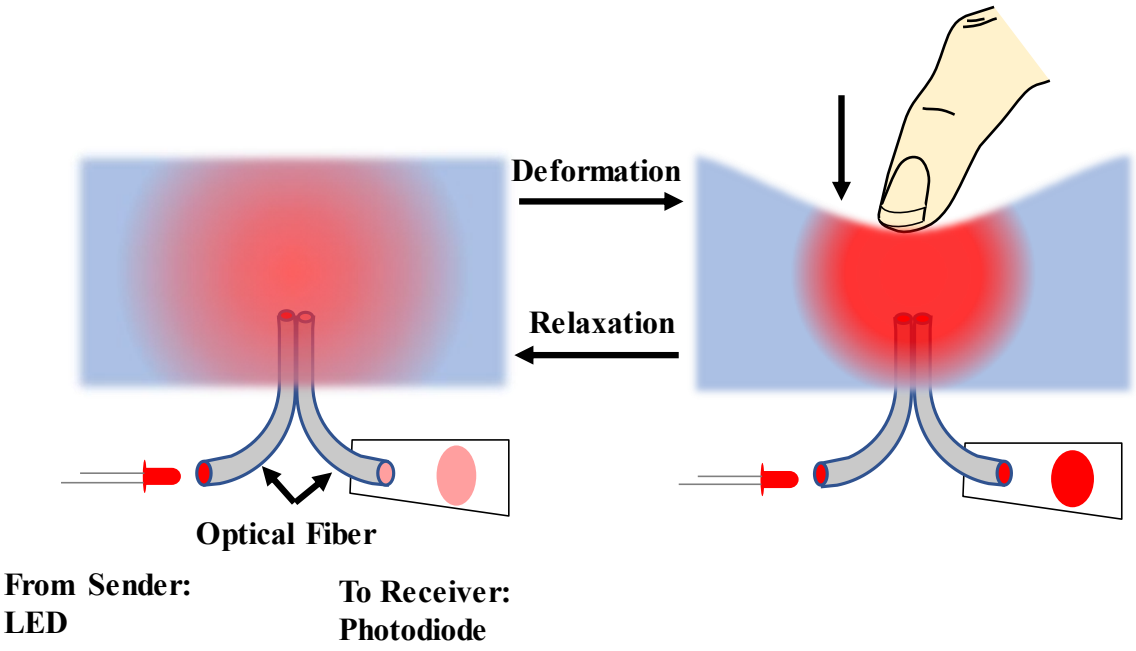


Figure 3.1 Cross-sectional view of an optic-based pressure sensor.

In terms of the number of sensors and configuration, there are two types of PSMs used in this thesis: type A and type B. Both mats have the same functionality.

PSM-Type A

PSM-Type A is composed of three panels as one piece containing 72 sensors in total. The 72 sensors are distributed evenly, as Figure 3.2 shows. The head of each panel is a strip that contains

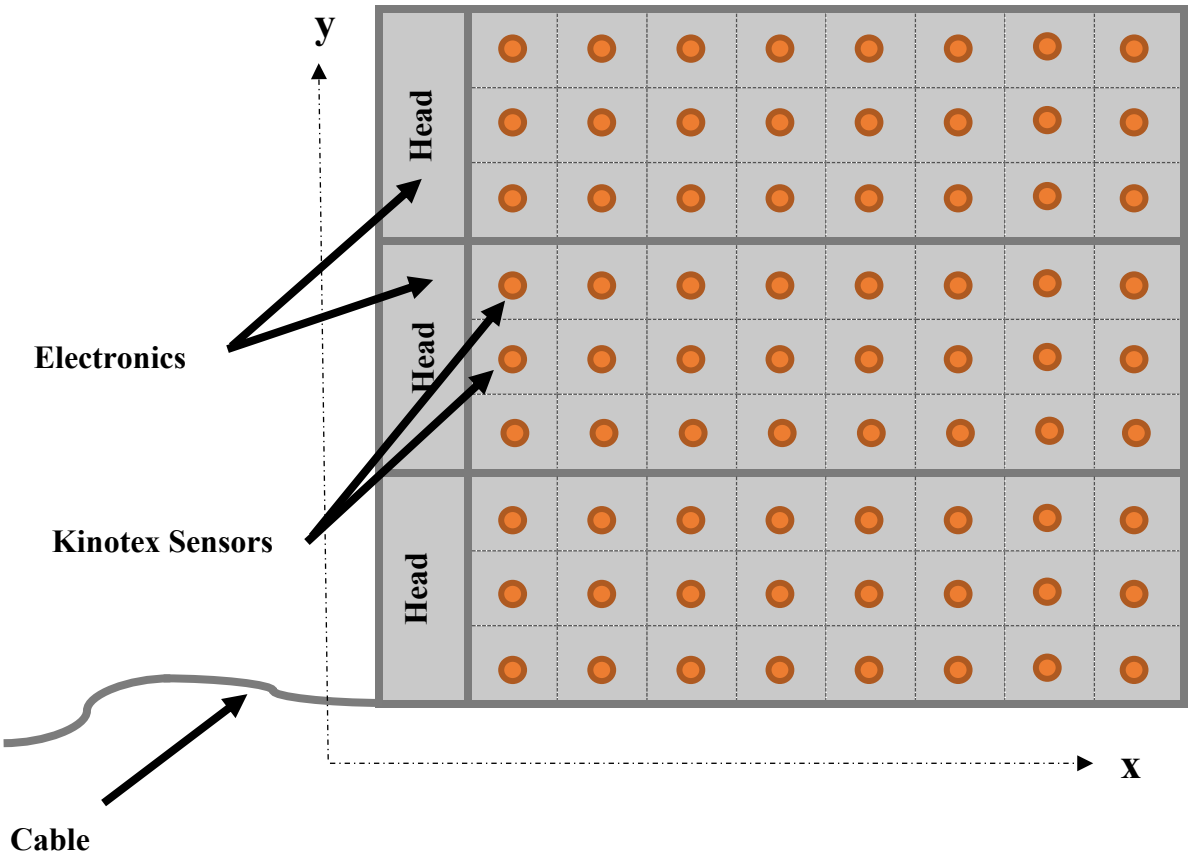


Figure 3.2 PSM-Type A inner sensor distribution diagram.

the electronics for analog to digital conversion, sampling, and transmission of the sensor data. Sensors are embedded in foam as an eight by nine matrix, with eight sensors lining the width and nine sensors lining the length of the mat. Data from 72 sensors are collected simultaneously with a sampling frequency of 20 Hz.

PSM-Type B

PSM-Type B is composed of three separate panels. Each panel consists of an eight by three pressure array. A set of PSM-Type B mat has 72 sensors, which are the same number of sensors

contained in a PSM-Type A mat. Figure 3.3 shows the sensor layout structure of the PSM-Type B mat. PSM-Type B collects data simultaneously with a sampling frequency of 10 Hz.

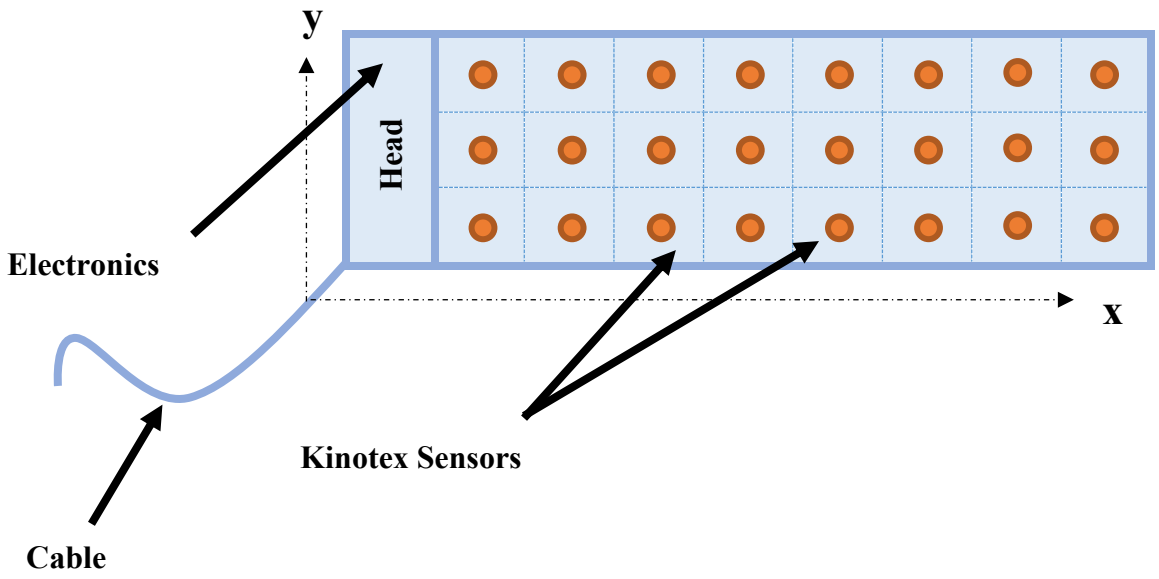


Figure 3.3 PSM-Type B inner sensor distribution diagram.

PSM needs to be connected to a recording box to save data for further processing. The recording box is placed under the bed and saves the data on a secure digital (SD) card, Figure 3.4.

Figure 3.4 Recording box, back view.

3.3.2 BioHarness 3.0

The BioHarness 3.0 is a physiological monitoring telemetry device which consists of a chest strap and an electronics module that attaches to the strap, as shown in Figure 3.5. The device can store and transmit vital sign data, including ECG, heart rate, breathing signal and respiration rate,

body orientation, and activity. For this thesis, breathing signals and respiration rates extracted from the BioHarness chest sensor were used as a gold standard. The algorithms developed to extract breathing signals and respiration rates from PSMs were validated, comparing their outputs with the gold standard information in a controlled experiment.

In [126], the BioHarness 3.0 was compared to two validated systems: a standard laboratory metabolic system with open-circuit spirometry and 12-lead electrocardiogram (Vmax Spectra System, VIASYS, from Yorba Linda Inc., CA) and a portable cardiopulmonary breath-by-breath gas exchange analyzer (Model K4 b2, by COSMED Inc., Rome, Italy) to measure heart and respiration rates. The results revealed that the device is reasonably accurate, and therefore, it can be a good candidate as a reference device for the present study.

Figure 3.5 BioHarness 3.0

3.4 Experimental Setup

3.4.1 Supervised Data

In order to collect supervised data, a PSM-Type A was placed under a bed mattress. It covered the upper part of the body from head to hip. While participants lay on the mattress, they wore 2 BioHarness 3.0 belts, as in Figure 3.6. The belts were positioned around the chest and abdomen to capture the breathing effort. Based on the guideline in [127], the abdominal band was placed just

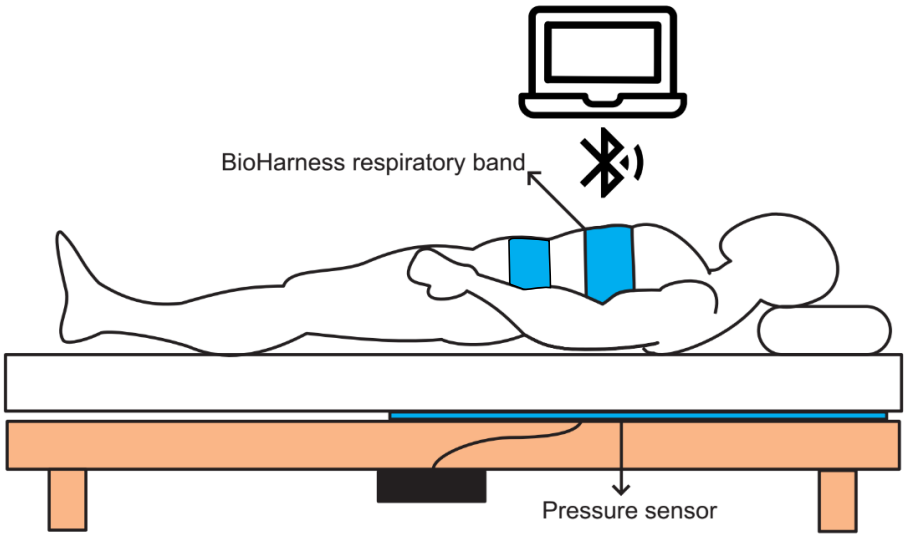


Figure 3.6 Diagram of experimental setup

above the umbilicus, and the thoracic band was placed between the armpit and nipple line. The belts were adjusted, so they were secure, but not tight. Participants were asked to lie on four different sleep positions, respectively the supine, on the left-side, on the right-side, and prone. In each position, they were instructed to do different maneuvers of breathing at certain time intervals, including normal breathing, CSA simulation, and OSA simulation. Participants tried holding their breath for at least 10 s to simulate CSA. For OSA simulation known as Müller's maneuver [128], they were asked to try exerting force with their lungs as if trying to breathe during 10 s of holding their breath. This is because, during OSA, the air volume in the lung remains constant, but the

motion of the diaphragm continues. Therefore, during an obstructed inhalation attempt, the abdomen grows in circumference while the thorax reduces in circumference.

Overall, 13 participants took part in the experiment. Table 3.I indicates the characteristics of the participants.

Table 3.I Participants characteristics, supervised data.

Parameters	**No.**
Gender	
Female	11
Male	2
BMI (Mean ± SD)	22.6 ± 2.5 (kg/m²)
Age (Mean ± SD)	28 ± 3 (years)

3.4.2 Non-Supervised Data

To collect non-supervised data, volunteers from different communities passed some inclusion criteria, which were:

- 65 years of age or more, female/male;
- Being community-dwelling and living in affordable seniors housing, or discharged from the Geriatric Rehabilitation Unit at Elisabeth-Bruyère Hospital.

Sensor system setups were installed in the seniors' homes for long-term monitoring and collecting pressure data continuously. The monitoring time was 8-12 months. Except for one visit every month, the rest of the data collection was not supervised. In the home environment, PSMs were first taped securely to a plywood board to have a stable surface between the box spring and

mattress. If there was no box spring, PSMs were put on the top of the bed frame underneath the mattress. The home set up had the same configuration as Figure 3.6 except no chest sensor was attached to the patient. The primary purpose of obtaining the data described in this section was to monitor patient mobility through the analysis of the bed-exit [129].

A part of the recordings was collected through PSM-Type B mats, composed of three separate panels, and another part using PSM-Type A mats, composed of three panels as one piece. Recorded data was in ASCII format and auto-saved as CSV files hourly. For PSM-Type A, each hour had one large CSV data file, and for PSM-Type B, it had three separate smaller files. Files were concatenated on a day-by-day basis for 24 hours, from noon to noon, to have a daily observation, especially for sleep assessment. Table 3.II illustrates the type of PSM used for each participant and

Table 3.II Participants information, non-supervised data.

Participant No.	No. of Recording Days	Type of PSM	Start Day	Stop Day
1	152	A	2013.Jun.06	2013.Nov.03
2	126	B	2013.Jun.07	2013.Nov.08
3	295	B	2013.Jul.10	2014.Jun.29
4	101	B	2013.Sept.03	2013.Dec.12
5	397	A	2013.Nov.06	2014.Feb.25
		B	2014.Apr.16	2015.Jan.26
6	110	B	2013.Dec.02	2014.Mar.21
7	306	A	2013.Nov.26	2014.Jul.29
		B	2014.Jul.29	2015.Jan.20
8	239	A	2013.Nov.25	2014.May.28
		B	2014.Jun.11	2014.Dec.19
9	204	A	2014.Feb.06	2014.Dec.19

the period for data collection. No other health information is known about the patients because research Ethics approval for this study did not include access to the patient's health information.

3.5 Clinical Data

The database presents small-scale clinical tests. Each subject was monitored for one night in the Royal Ottawa Mental Health Centre (ROMHC) Sleep Disorders Clinic, Ottawa, ON. Two sets of PSM-Type B pressure mats (each set includes 3 PSMs-Type B) were placed under the mattress of the bed so as not to interface with bed cranks (if any existed), and the recorder was placed on the floor under the bed. Figure 3.7 shows the resulting setup. For this project, only the data recorded from the top set of PSMs, which covers the area from head to hip and contains breathing information, is considered for further processing.

Data were collected simultaneously from all PSG sensors and PSMs. Ethics approval was obtained from ROMHC Sleep Disorders Clinic, Carleton University, and the University of Ottawa.

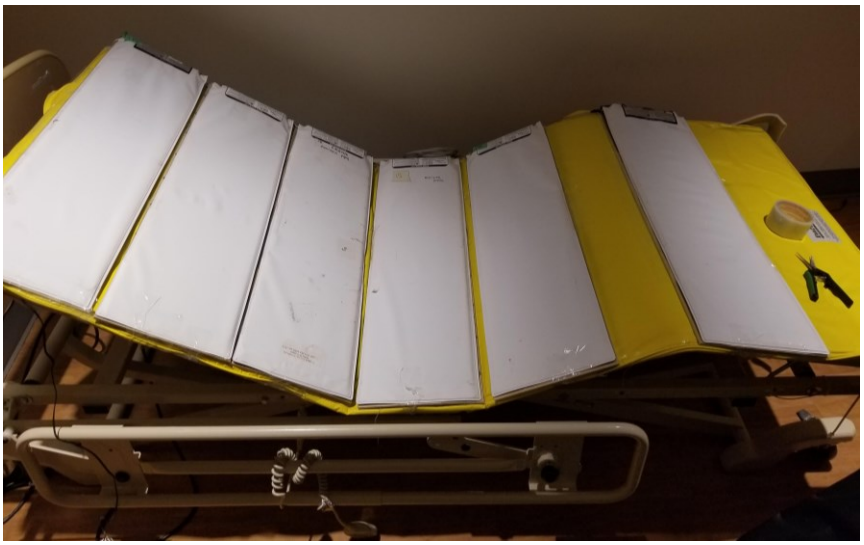


Figure 3.7 Experimental setup at ROMHC Sleep Disorders Clinic.

As the PSG sensors are unobtrusive and had been previously installed under the mattress and calibrated, participation in the study involved no extra effort, test, or signal calibration from the patient's perspective. Two rooms were each equipped with PSMs and data were collected 24 hours a day. The logging program created one new file for every hour. The sensor arrays remained in the same position for the entire data collection period and their placement underneath the mattress did not interfere with daily linen changing. Log files were collected bi-weekly. The incomplete physical connection of the sensor's cable with the data acquisition box and disconnection of the laptop from the power outlet was responsible for most cases of incomplete and missing data logs. 13 PSG files were scored by polysomnographers. Data of two patients were discarded from the database; one was removed because the PSG was split to CPAP in the middle of PSG recording and the other one was removed because PSM data was not available (perhaps due to disconnection of the acquisition box from the power outlet). The characteristics of the patients with acceptable data are illustrated in Table 3.III.

Table 3.III Patient characteristics, ROMHC Sleep Disorders Clinic.

Parameters	**No.**
Gender	
Female	7
Male	4
BMI (Mean ± SD)	31.7 ± 9.5 (kg/m²)
Age (Mean ± SD)	45 ± 13 (years)
Height (Mean ± SD)	66.2 ± 4.1 (inches)
Weight (Mean ±SD)	197.7 ± 64 (lbs)

3.6 Other Data (Public Domain Databases)

Table 3.IV Patient characteristics, St Vincent's University Hospital, Dublin Database (Physionet).

Parameters	No.
Gender	
Female	4
Male	21
BMI (Mean ± SD)	31.6 ± 4.0 (kg/m²)
Age (Mean ± SD)	50 ± 10 (years)
Sleep Efficiency (Mean ± SD)	77.2 ± 10.93 (%)
Study Duration (Mean ±SD)	6.93 ± 0.53 (hour)
AHI (Mean ±SD)	24.1 ± 20.3
Normal: AHI<5	1
Mild: 5≤AHI<15	10
Moderate: 15≤AHI<30	6
Severe: AHI≥30	8

St. Vincent's University Hospital/ University College Dublin sleep apnea database (https://archive.physionet.org/pn3/ucddb/) has been used for this research [130]. It consists of twenty-five full overnight polysomnograms with simultaneous three-channel Holter ECG from adult subjects who had been clinically diagnosed with OSA. The subjects were randomly selected over six months from patients referred to the Sleep Disorders Clinic at St Vincent's University Hospital, Dublin, Ireland for possible diagnosis of OSA, CSA, or primary snoring. They had to be above 18 years of age, with no known cardiac disease, autonomic dysfunction, and not on medication that is known to interfere with heart rate. Table 3.IV contains subject characteristics

and information. Sleep efficiency is defined as the number of minutes of sleep divided by the number of minutes in bed. A normal range is approximately 85 to 90% or higher.

Polysomnograms were obtained using the Jaeger-Toennies system (Erich Jaeger GmbH, Germany). Signals recorded were EEG (C3-A2), EEG (C4-A1), left EOG, right EOG, submental EMG, ECG (modified lead V2), oro-nasal airflow (thermistor), ribcage movements, abdomen movements (uncalibrated strain gauges), oxygen saturation (finger pulse oximeter), snoring (tracheal microphone) and body position. Three-channel Holter ECGs (V5, CC5, V5R) were recorded using a Reynolds Lifecard CF system (Reynolds Medical, UK). An experienced sleep technologist annotated onset time and duration of respiratory events (obstructive, central, mixed apneas and hypopneas, and periodic breathing episodes). The same sleep technologist scored sleep stages according to standard Rechtschaffen and Kales rules [131].

Chapter 4: Respiration Measurements

4.1 Introduction

Bedside measurement of the respiration rate, as one of the primary vital signs, is judged to be of high clinical importance. For measuring the respiration rate, there are clear advantages to use noncontact respiration monitoring methods in comparison to those which need to be attached to the body. Simply, the subject can be more comfortable (especially for long term monitoring) as he is not tied to an instrument.

Moreover, measurement accuracy can be improved as distress caused by a contact device may alter the respiration rate [132]. PSMs have proved to be accurate for breathing rate monitoring of patients in hospitals and smart homes [133]. An extensive investigation of respiration monitoring and movement detection using a PSM was conducted in [134], [135]. In [136], the effect of mattress type and its thickness were investigated in order to validate PSMs for physiological monitoring. The results showed that mattress type, thickness, and the presence of pillows could all affect the strength of acquired signals. Extracted signals from PSM placed under a mattress can be heavily muted and may feature less distinct localized loading. Therefore, more sophisticated signal processing algorithms may be necessary to extract physiological signals. Whether the breathing signal is available from all pressure sensors individually or is weakened due to the mattress and other environmental conditions, signal fusion generally can improve measurement accuracy [137] and produces a single representation of an individual's breathing by fusing data from all sensors. The signals from PSMs can be combined to generate a single output signal with better signal quality. In ambient systems using the linear combining paradigm, sensor signals are conditioned

by pre-processing, aligned, multiplied by a gain factor and then summed. By adopting this procedure, in [138] different offline combining techniques were presented for respiratory signal extraction, including selection combining (SC) and equal gain combining (EGC).

This chapter proposes two simple breathing signal fusion as a linear combining scheme, where individual sensor inputs are weighted by a scalar before summation. These are Pearson's Correlation Coefficient (PCC) based signal fusion and Signal to Noise Ratio (SNR)-Max based signal fusion, which is an SNR-maximizing method based on an MVDR beam-former derivation. The performance of the two proposed methods is evaluated and compared with some previously published methods. These results were published in [6].

4.2 Breathing Signal Combining

4.2.1 Pre-processing

The typical respiratory rate for a healthy adult at rest is 12 to 20 breaths per minute (bpm). However, the complete range of possible breathing rates, including extreme conditions, can be from 4 to 48 bpm, which corresponds to a frequency between 0.07 Hz and 0.8 Hz [4]. For the extraction of the breathing waveform, in this work, a 70th-order Finite Impulse Response (FIR) linear phase bandpass filter with a passband corresponding to the above range (designed by window-based method) is used to filter the signal from each sensor. In comparison to IIR filters, FIR filters are simple to design, stable, and can be forced to have linear phase (by designing the filter taps to be symmetrical about the center tap position) [139]. After filtering, data at the edges (beginning and end) in the output signal is discarded due to "edge effects."

4.2.2 Reference Signal & Polarity Adjustment

For data combining, the signals' gain is assumed to be constant over a short period during which the signal itself is assumed to be relatively stationary. Therefore, after filtering, the signals from each channel are segmented into 30 s segments with a 50% overlap for further processing. These parameters were chosen for conformity to the standard in sleep analysis [140]. After segmentation, the residual mean loading value of the sensors is removed by taking an average. Then, each segment is analyzed to find a reference signal. Among all PSM signals, the strongest signal in terms of signal quality is selected as the reference signal. This selection can be performed based on different criteria. Some have used the signal with the greatest average power in the time domain, whereas others considered the reference signal as the signal exhibiting the highest power in the frequency band of interest. In [141], [142], the sensor with the highest total power in the respiratory frequency band was selected, measured using the power spectral density (PSD). Signals with strong breathing signals will include a peak in the PSD, which will increase the power in that band.

One can say that the reference signal by itself can be used to detect the breathing rate. However, the selection of a single signal is more likely to result in an error, particularly in the presence of interference or noise sources. Therefore, it is good practice to use data fusion in conjunction with a reference signal instead of using only a reference signal. In this thesis, the reference signal is chosen as the sensor for which the output signal has the maximum power in the respiration frequency band.

The PSM captures breathing motion generated along the torso and shoulders of the participant. These signals can have different polarities. The pressure applied to an area of the rigid plastic covering the array can cause the area next to it to bend upward. Figure 4.1 shows the signals from three sensors filtered by the FIR linear phase bandpass filter. It can be seen that not all signals have

the same polarity. Also, the quality of all signals based on their location relative to the body and the breathing movement is not the same.

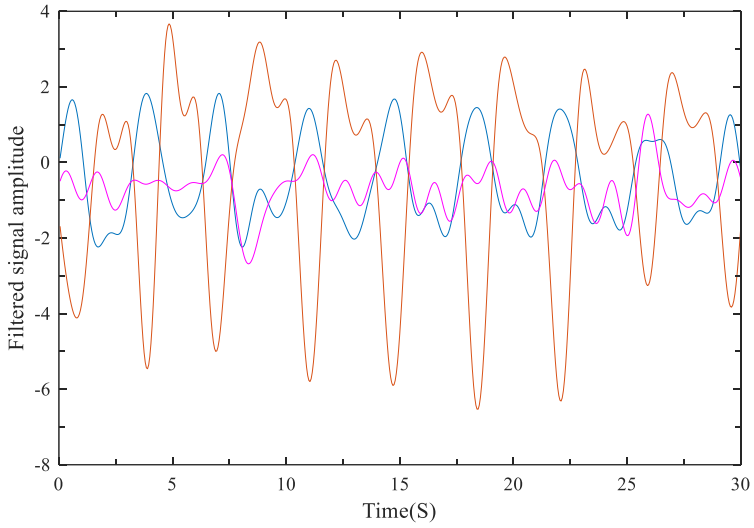


Figure 4.1 Recorded pressure in three adjacent sensors. Inverse polarities can be clearly seen between two of the signals.

Simply adding signals together may result in destructive interference - a loss of information. Therefore, before data fusion, the polarity of each signal in each segment is compared to the polarity of the reference signal, using the sign of the PCC. Although PCC measures linear dependence between two signals, it can also be used to determine direct or opposite proportionality based on its sign [143]. Each sensor signal is negated if the related sign in the corresponding PCC is negative.

4.2.3 Signal Combining

Defining $Z_i[n]=\left[z_{i,1}[n] \quad z_{i,2}[n] \quad \cdots \quad z_{i,N}[n]\right]^T$ as the noisy sensor signals at sampling time n in the i^{th} segment, and $W_i=\left[w_{i,1} \quad w_{i,2} \quad \cdots \quad w_{i,N}\right]^T$ as the weights applied to each sensor of the array in the same segment, then the final output of the signal combination can be written as:

$$y_i[n] = W_i^T Z_i[n] \quad (4.1).$$

The difference between various combining methods is how the weights are computed. Figure 4.2 illustrates the implementation of a linear combiner for each sample within a segment. In each segment, given a reference sensor signal, each sample is manipulated through two steps: first polarity adjustment, next gain adjustment. Apart from polarity adjustment which is implemented due to the characteristics of PSM signals, the rest of the figure represents the standard process of linear combiners [144].

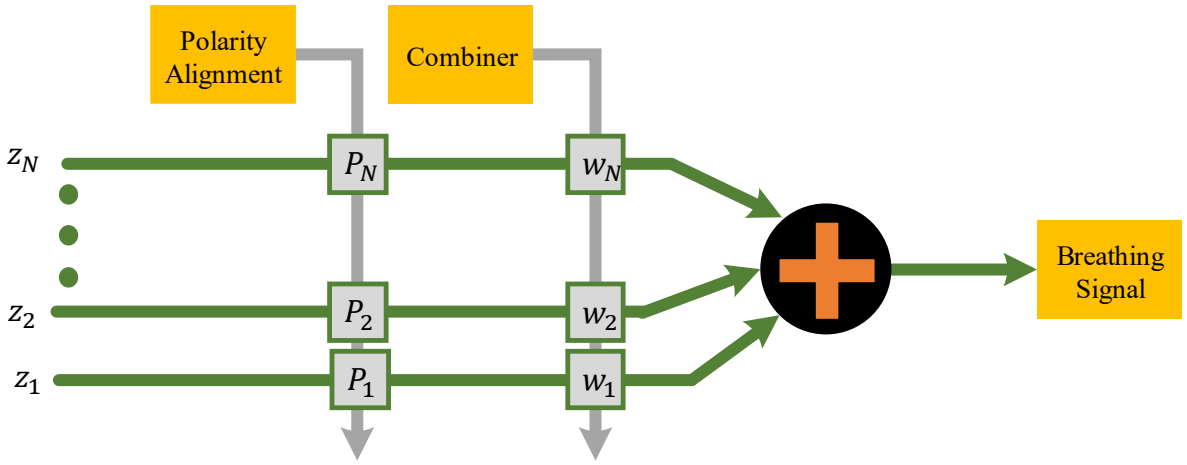


Figure 4.2 Combining samples for each segment

4.2.3.1 Correlation Analysis as a Combining Method

In this method, the combiner uses the PCC between each signal and the reference signal as the weights when adding all the sensor signals together in each segment. In this case, signals which are more correlated to the reference signal have higher weights than the others. It should be noted that there is no need for a separate polarity adjustment step when PCCs are used, as the sign information is part of the PCC. Let $z_{i,r}[n]$ be the reference signal samples in the i^{th} segment and $z_{i,k}[n]$ the k^{th} sensor in each segment. The weights are then computed as:

$$W_i = \begin{bmatrix} \frac{\sum_{n=1}^{M}(z_{i,r}[n]-\bar{z}_{i,r})(z_{i,1}[n]-\bar{z}_{i,1})}{\sqrt{\sum_{n=1}^{M}(z_{i,r}[n]-\bar{z}_{i,r})^2\sum_{n=1}^{M}(z_{i,1}[n]-\bar{z}_{i,1})^2}} \\ \frac{\sum_{n=1}^{M}(z_{i,r}[n]-\bar{z}_{i,r})(z_{i,2}[n]-\bar{z}_{i,2})}{\sqrt{\sum_{n=1}^{M}(z_{i,r}[n]-\bar{z}_{i,r})^2\sum_{n=1}^{M}(z_{i,2}[n]-\bar{z}_{i,2})^2}} \\ \vdots \\ \frac{\sum_{n=1}^{M}(z_{i,r}[n]-\bar{z}_{i,r})(z_{i,N}[n]-\bar{z}_{i,N})}{\sqrt{\sum_{n=1}^{M}(z_{i,r}[n]-\bar{z}_{i,r})^2\sum_{n=1}^{M}(z_{i,N}[n]-\bar{z}_{i,N})^2}} \end{bmatrix} \quad (4.2)$$

where M is the number of samples in each segment.

4.2.3.2 SNR-MAX as a Combining Method

The problem of combining signals from different sensors and maximizing the signal to noise ratio at the output can also be addressed using the classical theory of linear beamforming. Beamforming or spatial filtering is a signal processing technique used in sensor arrays for directional signal transmission or reception [145]. Using this technique, elements in an antenna array are combined in such a way that signals at particular angles experience constructive interference while others experience destructive interference. Up to a scalar term, for a single-tap per sensor beamformer (similar to a frequency domain beamformer), it has been demonstrated that the solutions of a maximum-SNR combining beamformer and an MVDR beamformer are the same [146]. The MVDR beamformer is a data-adaptive beamforming solution whose goal is to minimize the variance of the combined output signal under a distortionless condition on a useful signal component. It can thus mitigate the effect of the noise.

The MVDR beamformer weights are obtained as follows:

$$W_i = \frac{\left(R_{ZZ,i} + \mu I\right)^{-1} D_i}{D_i^H \left(R_{ZZ,i} + \mu I\right)^{-1} D_i} \quad (4.3)$$

where μ is a regularization factor (i.e., diagonal loading) used in practical situations in order to avoid ill-conditioning in matrix inversions, and $R_{ZZ,i}$ is the auto-correlation matrix of the noisy input vector $Z_i[n]$ calculated as:

$$R_{ZZ,i} = \frac{1}{M} \sum_{n=1}^{M} \left(Z_i[n] Z_i[n]^T \right) \tag{4.4}$$

Moreover, D_i can be seen as a "steering vector" which includes the response between a target source and each sensor. The SNR at the output of the MVDR filtering is invariant to a constant scaling factor applied to the D_i column vector. Therefore, the D_i vector can be normalized by the value of the reference signal component, and the resulting D_i vector components are the single-tap impulse responses between the reference signal and the other signals. Such single-tap impulse responses can be obtained from a ratio of the cross-covariance between the reference signal and the other signals, over the auto-covariance of the reference signal, as follows:

$$D_i = \begin{bmatrix} \frac{\sum_{n=1}^{M} (z_{i,1}[n] - \bar{z}_{i,1})(z_{i,r}[n] - \bar{z}_{i,r})}{\sum_{n=1}^{M} (z_{i,r}[n] - \bar{z}_{i,r})^2} \\ \frac{\sum_{n=1}^{M} (z_{i,2}[n] - \bar{z}_{i,2})(z_{i,r}[n] - \bar{z}_{i,r})}{\sum_{n=1}^{M} (z_{i,r}[n] - \bar{z}_{i,r})^2} \\ \vdots \\ \frac{\sum_{n=1}^{M} (z_{i,N}[n] - \bar{z}_{i,N})(z_{i,r}[n] - \bar{z}_{i,r})}{\sum_{n=1}^{M} (z_{i,r}[n] - \bar{z}_{i,r})^2} \end{bmatrix} \tag{4.5}$$

Note that for bandpass filtered data, the subtractions of the means in (4.2) and (4.5) become unnecessary, as the signals have zero mean. When the value of the D_i vector is not accurately known (i.e., mismatch), using the noisy signal correlation matrix $R_{ZZ,i}$ instead of a noise-only correlation matrix can lead to significant degradation of the beamformer performance [146],

including target or useful signal attenuation. However, this is less of a concern if we assume that the noise is spatially uncorrelated between sensors and that it has the same variance in each sensor, such that $R_{ZZ,i}$ becomes a scaled identity matrix. In this case, up to a scalar term equation (4.2) becomes:

$$W_i = D_i \tag{4.6}$$

It can also be noted that in the computation of the elements of D_i, the denominator is common to all components and it can be discarded, as it becomes only a global scaling factor. To calculate the output of each segment, the filter weights are applied to the noisy input samples, again as follows:

$$y_i[n] = W_i^T Z_i[n] \tag{4.7}$$

where $n = 1, 2, \ldots, M$.

4.3 Experiment

A participant lay supine on top of a mattress while she was wearing the BioHarness™ 3 belt system. During the data collection, the participant was asked to breathe normally.

4.3.1 Respiratory Rate Estimation

To evaluate the performance of the respiration rate extraction methods, a comparison with a "gold standard" extracted signal can be a good metric to judge the performance of the method.

Researchers have used different approaches to estimate the respiration rate from an extracted respiration signal. Based on the literature, the use of the signal auto-correlation has proven to be efficient for respiration rate estimation [141], [147], [148]. In this section, the auto-correlation

method is used to estimate the respiration rate from the signals produced by both combining methods, as well as for the gold standard signal from the belt system.

4.4 Results

After data collection, the output signals of the pressure array system were time-aligned to the output of the belt system using cross-correlation.

The result of the two proposed methods are compared to two previous methods from [138]: Selection Combining (SC) and Equal Gain Combining (EGC). The SC method selects the single best channel as the output for each segment, i.e., the reference signal. The EGC combines polarity adjustment with a unit gain for all channels, to perform the channel combining. The outputs of the two proposed combiners for the same segment are shown in Figure 4.3, with the related reference output of the belt system. The level of the signals in Figure 4.3 has been adjusted to have the same root mean square (RMS) value. Both of the proposed methods follow the reference signal quite closely.

To measure the breathing periods, peaks were detected from the auto-correlation function of the gold standard reference signal and the output signals of the different combiners. Linear regression was used in order to model the relationship between the respiration rates estimated from the output of the different combiners and the reference signal, over all the segments. Figure 4.4 illustrates the linear regression data of the extracted respiration rate, while Table 4.I presents the parameters of the linear regressions. The linear regression model is fitted to the data with regression values of

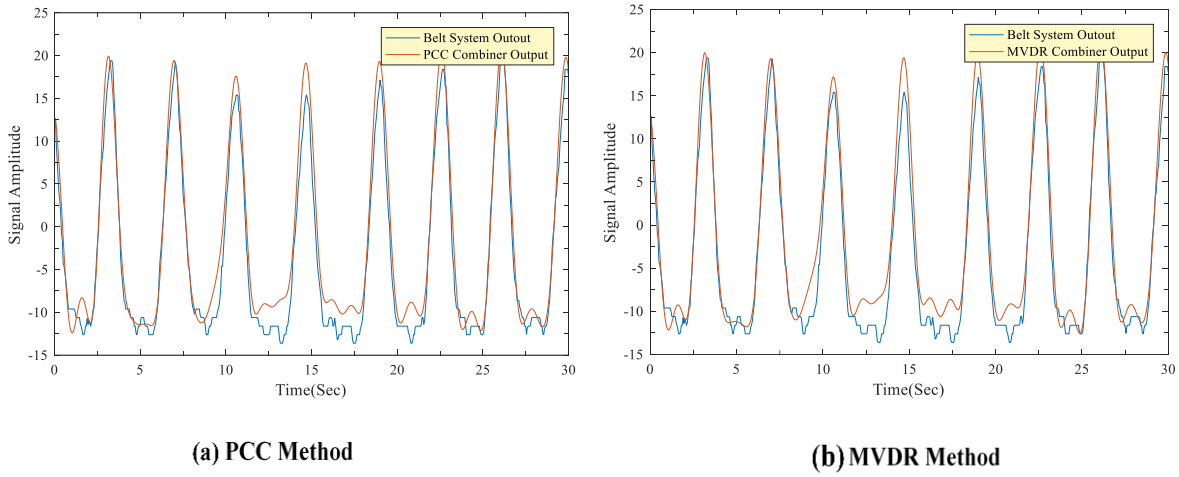
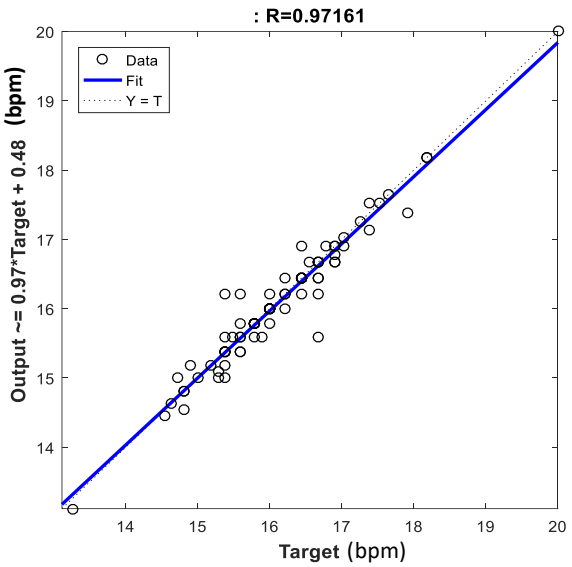


Figure 4.3 Result of the proposed combiners for the same segment.

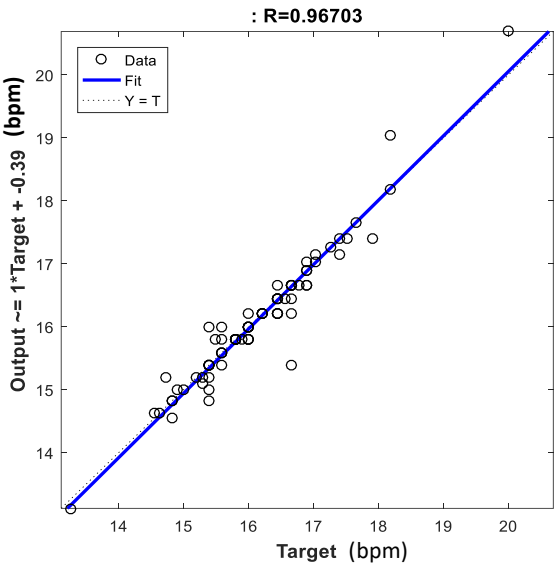
97.16% and 96.70 % for the proposed PCC and MVDR method, respectively. This means that the differences between the respiration rates estimated from the output of the combiners and the reference data are small, and both proposed combining methods could lead to good estimates of the respiratory rate. Moreover, Table I indicates that the two methods proposed in this section produce a better performance than the SC and EGC methods from [138].

Table 4.I Parameters of linear regression for the combiners

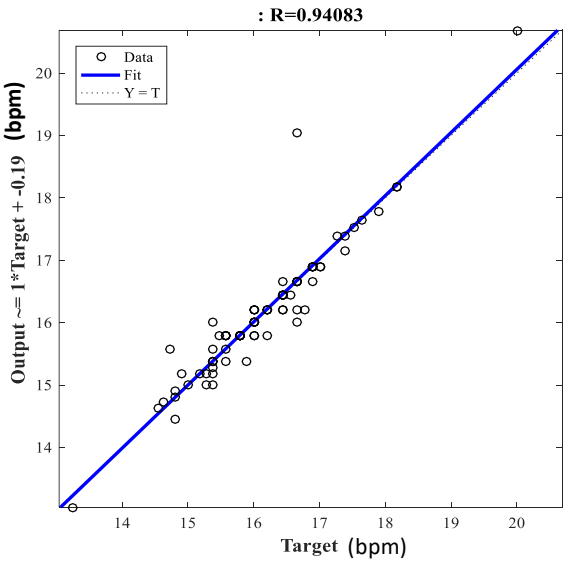
	Regression Value	**Slope of Regression**	**Offset of Regression**
PCC	**0.9716**	0.9680	0.4815
MVDR (SNR-MAX)	**0.9670**	1.0223	-0.3922
SC	**0.9408**	1.0124	-0.1854
EGC	**0.8956**	1.0329	-0.5904



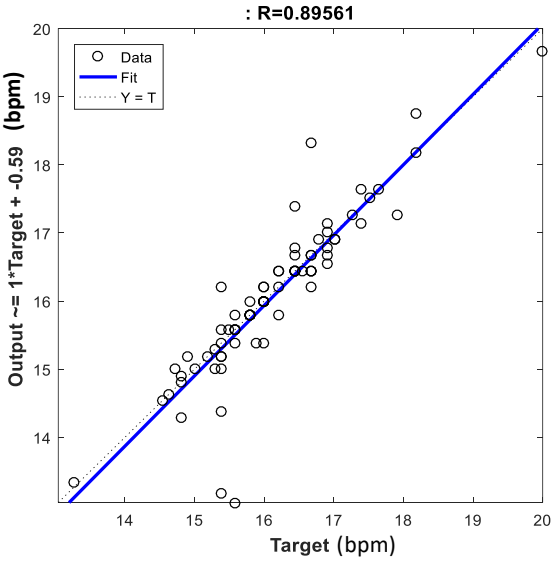
(a) PCC method



(b) MVDR method



(c) SC method



(d) EGC method

Figure 4.4 Linear regression of the extracted respiration rate for the different combining methods versus the gold standard.

4.5 Summary

In this chapter, two simple methods of combining the signals of an in-bed PSM were presented and evaluated: PCC and SNR-MAX. Both methods showed promising results as a signal combiner for respiratory signal extraction. Moreover, given the simplicity of both combining methods, they can be applicable options in many cases.

Body movement can affect the breathing signal, creating artifacts [135]. Therefore, during the data collection, the participant was asked to breathe normally and avoid any type of movement.

To evaluate the performance of the proposed methods for close-to-reality data, participants were allowed to move around on the bed and have different types of breathing, such as shallow breathing, deep breathing, and periods of CSA, in experiments with more noisy data [7]. Compared to other data fusion methods described in this chapter, for more complex data containing movement in bed and different types of breathing, the SNR-MAX was found to be the best method of sensor signal combining, resulting in the highest Pearson Correlation Coefficient with the respiratory band signal. Therefore, for Chapter 6 and Chapter 7, to extract a breathing signal before making a CSA detection, PSM signals will be combined by applying the SNR-MAX method.

Chapter 5: An Alternative Signal for Airflow Signal

5.1 Introduction

As stated in Chapter 2, apnea, without considering its type, is characterized by repeated periods of reduction or complete cessation of airflow. Thoracoabdominal signals are an indirect measurement of airflow with high accuracy, sensitivity, and patient safety [46]. At several stages of this thesis, thoracoabdominal signals were used as a gold standard in comparison to PSMs. To address the question of whether PSMs can detect different types of SA and distinguish them from normal breathing, first it is necessary to consider whether thoracoabdominal signals without the accompaniment of airflow signal can detect and differentiate all types of apnea or not.

According to AASM, the summation of thoracoabdominal signals can be used as an alternative signal to score apneic events [17]. In this chapter, the summation of dual respiratory inductive plethysmography (RIP) signals, RIP-sum, is adopted to detect apnea and hypopnea events. Although calibration of the posture can affect the RIP-sum signal and consequently the accuracy of detecting apneic events using RIP signals, they still allow detection of a relative change in signal volume. Figure 5.1 shows samples of obstructive, central, and mixed apnea events from the University College Dublin Sleep Apnea database [149]. The RIP-sum shows absent or minimal excursions while apneic events happen. In comparison, the thorax and abdomen signals (consequently the RIP-sum signal) show clear breathing with no apneic events. Obstructive sleep apnea (OSA) is the more frequent pattern, characterized by the presence of abdominal and thoracic

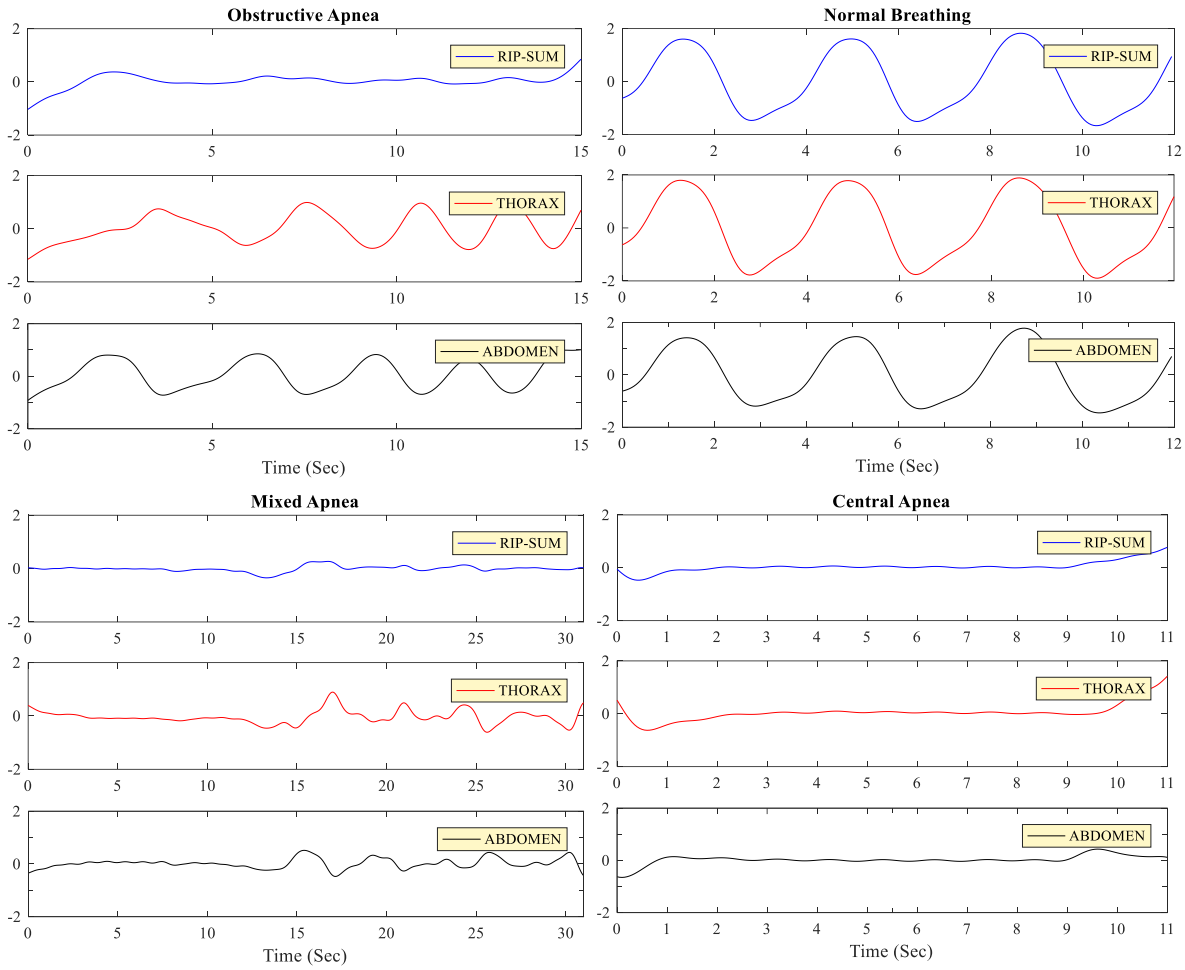


Figure 5.1 Different apnea events and normal breathing illustrated by thorax signal (red), abdomen signal (black) and their sum, i.e., the RIP-sum signal (blue).

efforts for continuing to breathe, while airflow completely stops. In most cases, there is paradox breathing (180 degrees difference in phase) between the abdomen and thorax signals while OSA happens [20]. Therefore, the summation of these as a RIP-sum signal is characterized as no signal or a relatively diminished signal. Central sleep apnea (CSA) is characterized by a complete cessation of both respiratory movements and airflow. Mixed sleep apnea (MSA) is a combination of the previous two, defined by a CSA followed by an OSA [23]. Hypopnea events are defined when the person does shallow breathing, or in other words, abdominal and thoracic efforts are diminished compared to baseline breathing.

In sections 5.3, a simple automated detection method is proposed to investigate the use of RIP-sum signals for detecting apneic events. Section 5.4 contains an evaluation of the result based on event-based metrics. Finally, limitations of the study are discussed in section 5.5.

5.2 Database

The database used in this chapter was the University College Dublin Sleep Apnea Database. It was previously introduced in detail in Chapter 3, section 3.6. The database contains 25 full overnight polysomnograms.

5.3 Method and Material

5.3.1 Pre-processing

In this database, the RIP-sum signal was sampled at 128 Hz. For sleep-disordered breathing detection, respiratory channels are typically recorded with a lower sampling rate (10-50 Hz), because further sampling of the data provides no additional useful information. Therefore, the RIP-sum signal was downsampled to 20 Hz to decrease the computational cost.

The typical respiratory rate for a healthy adult at rest is 12 to 20 breaths per minute (bpm). However, the complete range of possible breathing rates, including extreme conditions, can be from 4 to 48 bpm, which corresponds to a frequency between 0.07 Hz and 0.8 Hz [6]. For the extraction of the breathing waveform, an FIR linear phase bandpass filter with a passband corresponding to the above range was used to filter the RIP-sum signal, as in section 4.2.2. There were some parts of the data where the signal was not available because of a sensor failure (sequence of zeros). If the signal was not available for more than at least 10 s, the algorithm no longer considered those time windows, to avoid false detection later.

5.3.2 Proposed Method

The proposed method included two steps. Each step is defined in detail as follows:

a) Power calculator: According to the AASM in adults, the breathing baseline is defined by the three largest breaths in the 2 minutes preceding the onset of an event in individuals without a stable breathing pattern [17]. The breathing baseline is estimated to determine a threshold that is used to detect an apneic event. For this purpose, the previous 2 minutes of data were divided into *L*-second windows with *M%* overlap, and for each window, the power of the signal was calculated. Let $x_i(n)\ n=1,2,\cdots,N$ be the data samples of the i^{th} sub-segment of length N. Then for each sub-segment, the power was calculated according to the equation:

$$P(i)=\frac{1}{N}\sum_{n=1}^{N}\left|x_i(n)\right|^2 \tag{5.1}$$

b) Classifier: to compute the threshold, the powers of the sub-segments calculated in the last 2 minutes were sorted in ascending order. Then only 20 percent of sub-segments with the largest power were considered for further processing, in order to be robust to cases where there were several events in the previous 2 minutes. This means that the threshold was computed as a fraction of the power of the sub-segment with an 80 percentile position (in terms of power) from the "sorted" sub-segments. This fraction of the power helped to tune the threshold, as the power of the 80-percentile segment was too large to be considered as the threshold by itself. If the power of an event evaluation segment after 2 minutes of data was smaller or equal to the threshold, it was classified as an event, and the flag vector related to this part of the signal was set to one.

5.3.3 Post-processing

This part of the algorithm was used to obtain a representation of the detected events similar to the annotations of the database:

a) A sequence of segments which were all flagged as events were merged as a single event;

b) Because a complete respiratory cycle lasted at least 3 s, parts of a signal located between events and which were not separated by more than 3 s were reclassified as events;

c) Finally, according to AASM, the duration of apnea-hypopnea events should be at least 10 s. Therefore, detected events with a duration of less than 10 s were re-flagged as NOT-events.

Figure 5.2 shows the schematic display of the algorithm while it detects an apneic event. Here a 6-second window size with a 90% overlap was used. The power per segment signal shows a fall during an apneic event. An event is defined as a fall if the power per segment is less than the threshold, adjusted based on the power information of the previous 2 minutes of the signal.

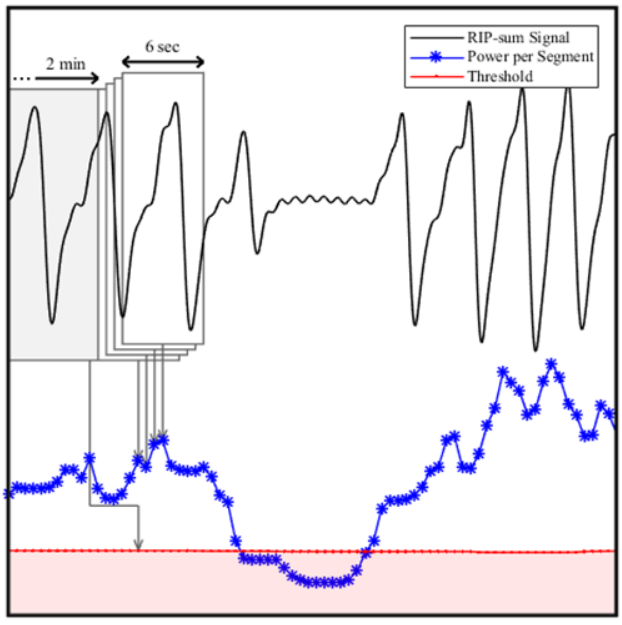


Figure 5.2 Schematic display of the algorithm

5.4 Results

In order to tune and evaluate the algorithm, data from all patients were randomly selected either in the training set (80%) or in the testing set (20%). To avoid overfitting, the training and testing set randomization was done five times, and the average performance was calculated. Each time three parameters, including the size of the windows in the range from 4 to 12 s, the amount of overlap from 50% to 90% and the threshold were tuned based on the highest F-score achieved on the training dataset.

The evaluation of the proposed algorithm was event-based, which compared the proposed method outputs and the corresponding reference event by event. Knowing the start time and duration of the annotated events, it identified the quantity of each possible covering situation as illustrated in Figure 5.3. These various possible configurations were gathered in four categories: correct detection, false detection, missed detection, and multiple coverings implying automatic detection.

Once the numbers for these various categories are known, the number of true positive events (#TP), false-positive events (#FP) and false-negative events (#FN) of a confusion matrix are deduced as:

• TP: an event in the output of the proposed method that has a temporal position overlapping with the temporal position of an event with the same label in the reference. Based on the definition of TP, multiple covering events are counted as correct detection too;

• FP: an event in the proposed method output that has no correspondence to an event with the same label in the reference;

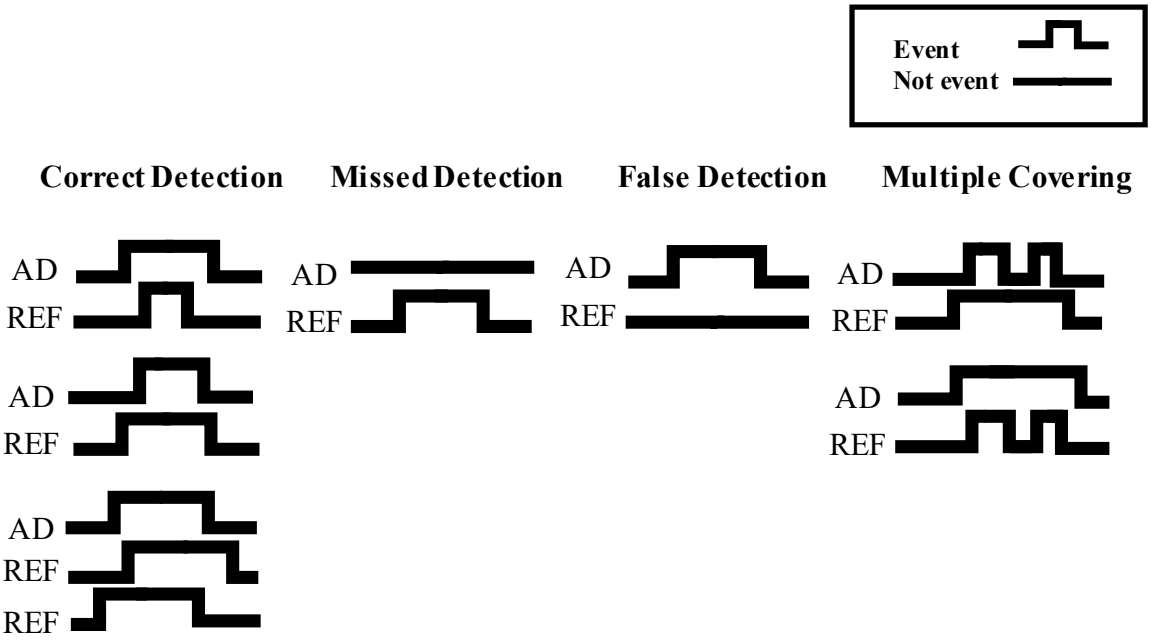


Figure 5.3 Different possible coverings between label in the reference (REF) and automatic detected events (AD).

• FN: an event in the reference that has no correspondence to an event with the same label in the proposed method output.

It should be noted that the event-based metrics have no meaningful true negatives (TN), except for the case where we measure in both the system output and the reference the total length of time segments with no active events [136]. In [151], for the sleep spindles detection, the number of TN events were estimated based on the total duration of the signal, the number of TP, FP, and FN events and the average duration of the events. However, for this study, this approach was not suitable because there is no certainty about the average duration of an apneic event, and the different classes of events in the proposed output and reference signals have different average times. Therefore, the F-score is used as a parameter to evaluate and tune the proposed method.

The F-score is the weighted average of the sensitivity (TP rate) and precision. It reaches its best value at 1 (perfect precision and sensitivity) and worst at 0.

The F-score is calculated as:

$$\text{F-score}=\frac{2\times\text{Precision}\times\text{Sensitivity}}{\text{Precision}+\text{Sensitivity}} \quad (5.2),$$

where precision and sensitivity (or TP rate) are deduced from the confusion matrix as:

$$\text{Sensitivity}=\text{TP}_{\text{rate}}=\frac{\text{TP}}{\text{TP}+\text{FN}} \quad (5.3)$$

$$\text{Precision}=\frac{\text{TP}}{\text{FP}+\text{TP}} \quad (5.4).$$

In adults, apnea and hypopnea events are defined when there is a drop in the peak signal excursion by at least 90% of pre-event level and by at least 30% of pre-event baseline (in association with 3% or more arterial oxygen desaturation or arousal), for a duration of at least 10 s [17]. The database included annotated events labeled as "PB-EVENT" when there was no proof of oxygen level desaturation (although the length of the event was more than 10 s), or "POSSIBLE" when the length of the event was less than 10 second (although there was 3% or more oxygen level desaturation). Since there was no certainty for such cases, none of these two types of annotated events were counted as TP, FP, and FN. Table 5.I indicates the average result of the proposed method on the test dataset. In addition to the average performance of the system, the standard deviation also indicates how much the system is robust when facing different testing and training dataset. The optimized parameters were chosen based on the highest F-score achieved on the testing dataset among all five sets of data. They are the window size of 5 s, the overlap of 90%, and the threshold of 0.2.

5.5 Discussion

In this chapter, a simple algorithm that can automatically detect sleep-disordered breathing from the RIP-sum signal has been developed. The algorithm uses the signal power to detect the variation

Table 5.I Average result of the proposed method on the test dataset.

F-score	Sensitivity	Precision
65.2% (± 6.7%)	72.1% (± 7.9%)	59.6% (± 6.3%)

of RIP-sum signal based on an adaptive threshold, which can decrease the number of false-positive detection events in case of a change in the breathing baseline.

To detect an event, the comparison between the system output and the reference can be done by a segment-based evaluation (fixed-length intervals) or an event-based evaluation (event-instance level). The event-based metrics measure the ability of the system to detect the correct event in the correct temporal position. However, in the segment-based evaluation, the signal is segmented and re-annotated from the first second, and the segments cover the whole length of the signal. After detection, each segment is then evaluated to define the confusion matrix. For sleep apnea detection, segment-based evaluation can reduce the final accuracy, especially in cases where the whole apneic event is not correctly located in just one segment, and part of it is covered by other segments too. In this chapter, the evaluation of the proposed method was event-based. The performances of previous systems developed in the same field are not directly comparable, because their methodology, their databases, and especially their assessment methods were different. In [152], the same database as in this chapter was used to detect apnea. The ECG and SpO2 signals were considered for apnea detection. Sensitivity, specificity, and accuracy values all around 82% were achieved based on the features extracted and the classifiers used. To extract features, first, the signals were re-annotated to have a segment-based evaluation approach. When events were across two adjacent segments, the authors assumed that each segment contained 5 consecutive seconds

of apnea/hypopnea, and the event was labeled as “Apnea” (apneic). This assumption prevents the results of their study to be directly compared with the results in this study. Moreover, they used a different type of sensors and a different evaluation approach.

SA detection is an anomaly detection problem, which deals with skewed classes where one class (in this case normal breathing) is over-represented in the dataset. Accuracy as the traditional performance measure may not be very useful for skewed data because the results obtained may not be truly representative of the actual performance. To overcome this, it is suggested to use precision and sensitivity or equivalently the F-score. In this study, the results showed that the proposed method has a relatively high diagnostic performance (F-score > 60%). It is important to know that there is always a trade-off between precision and sensitivity. An inspection of precision and sensitivity values reveals that the proposed method output contains less undetected events (higher sensitivity) and more events detected wrongly (lower precision). This can be described as one of the limitations of using a single channel signal to detect SA, especially hypopnea events since there is no access to information from other channels of data such as SpO2 (to detect oxygen desaturation), position (to have more accurate breathing baseline) and EEG (to check whether the arousal happens after a hypopnoea/apnea event or not) in order to decrease the number of FP events. The proposed algorithm could be adjusted to increase precision at the expense of sensitivity, or vice versa. However, both were valued equally here. Several previous works on SA detection, such as [153] evaluated their methods by accuracy, which makes these works not directly comparable with our work.

In comparison to the use of SpO2 and airflow signals for monitoring of sleep-disordered breathing, especially for apneic events detection, one of the potential advantages of using the dual

RIP signals as a single type of sensor is the ability to distinguish obstructive and central apnea events.

5.6 Summary

In this chapter, a simple method based on ASSM rules was presented to detect all types of apnea from the RIP-sum signal. The main purpose of this chapter was to learn more about different types of apnea through clinical data and their effect on respiratory signals. Given that at several stages of this thesis, thoracoabdominal signals were used as a gold standard in comparison to PSMs, processing and reviewing these signals helped us to understand the possible effects of SA on PSM data better and in conditions that were more close to real-life.

In the following chapter, the method implemented in this chapter will be compared to machine learning approaches to find the best method for detecting CSA events from PSM data.

Chapter 6: CSA Event Detection

6.1 Introduction

Most often, when it comes to sleep apnea, discussions focus on OSA. However, CSA can be as important as OSA, mainly caused by neurological ailments. The occurrence of CSA results in unstable breathing during the night. It initiates a cascade of events, including the inertia of the ventilatory control system, upper airway narrowing/occlusion, hypoxia, hypercapnia, and transient EEG arousal. The ventilatory control system exhibits marked inertia upon cessation of the ventilatory motor output. Consequently, rhythmic breathing does not resume until arterial PCO_2 is 4 to 6 mmHg above eupneic PCO_2. In addition, central apnea results in narrowing or occlusion of the pharyngeal airway in which additional force is required to overcome gravitational craniofacial forces or mucosal tissue adhesion. Finally, the combination of gas exchange derangements and transient EEG arousals results in ventilatory overshoot, hypocapnia, and recurrent central apnea. This sequence explains the importance of CSA detection and furthermore the identification of its triggers [154].

Automated CSA detection needs techniques applying mathematical tools to extract important diagnostic information from biomedical and biological data. Based on the size and complexity of the data, computers are used for processing, visualizing, and even classifying samples.

Traditional pattern recognition techniques for such applications consist of three main steps: processing of recorded signals, extraction of suitable features, and classification of those features with decision making for diagnosis. After measuring the relevant signals by various sensors, these signals are pre-processed for scaling and de-noising. This is necessary because the measured signals often contain some undesirable noise and artifacts. The sources of noise and artifacts

usually are the activities of other biological systems that interfere with the desired signal and variations due to sensor imperfections, sensor displacement, sensor saturation, etc. Next, the pre-processed signals are presented to suitable signal processing techniques to extract features, often involving the transformation of time-domain signals into other domains. Such features are expected to represent or describe the status and conditions of the subject under study. Although the direct interpretation of such features may not be well understood, these features play an important role in classification. The final step comprises classification and in the case of this study, diagnosis. In this step, the extracted features are fed to a classifier to distinguish among different classes of samples, for example, apnea and normal breathing. The classifiers are trained and optimized using a suitable set of examples.

However, the design or choice of features to be extracted in these conventional classification approaches requires some specific expert knowledge. The process is time-consuming and domain-specific [99]. Manually designed feature extraction algorithms may fail to extract the most relevant information from the data. Further, the amount of information fed to the traditional machine learning algorithms must be limited, since these algorithms would perform worse when facing high dimensional inputs. Therefore, the decision quality may suffer due to restricted information [98]. Consequently, automated design or selection of task-specific features can be valuable to solve complex real-world problems, such as SA detection.

In the current study, by adopting the same PSMs, to choose the right computer-based approach for a classification task, one needs to take into consideration several factors, such as the use of a generative model (to model the actual distribution of each class) or a discriminative model (to create a decision boundary between the classes) [124], and the properties and amount of data required. Upon these criteria, a comprehensive comparison between SVM as the so-called canonical

supervised learning approach [9] and DL methods, designed for the automatic detection of CSA events from PSM signals is performed. The DL approach consists of two different models: the temporal convolutional network (TCN) [155] and the Bidirectional LSTM (BiLSTM) network. The former is a representative CNN model and the latter is a representative recurrent neural network (RNN) model. Both methods are used with time-domain characteristics of the PSM signals. The BiLSTM model is used to construct the RNN-based approach since the CSA events are occurring sequentially and repeatedly during sleep. Just like the BiLSTM, the TCN can be used to model sequence-to-sequence or sequence-to-one tasks. Furthermore, the basic threshold-based method in Chapter 5 is adopted for comparison purposes. The objective is to find the best-performing and generalized method for automatic detection of CSA from PSM signals.

Section 6.2 introduces the proposed method and the information about the database used for this study in detail. Section 6.3 contains the results of different methods, described and compared through training and testing. In section 6.4, strengths and weaknesses are discussed for all the constructed models. Finally, section 6.5 summarizes and concludes the chapter.

6.2 CSA Scoring

Clinically, the validity of PSMs for detecting CSA events as related to gold standard physiological measurements has been proved in [156], where the PSM signals were compared to two other signals: RIP bands alone and RIP bands combined with airflow or oxygen-saturation sensor. PSMs provide similar performance as respiratory bands. Both types of sensors capture thorax and abdomen movements during breathing. The major difference between them is that while PSMs measure pressure caused by breathing, respiratory bands measure volume changes. Figure 6.1 shows samples of CSA from a) University College Dublin Sleep Apnea database introduced in section 3.5, and b) simulated CSA captured while a subject wore respiratory bands in our lab.

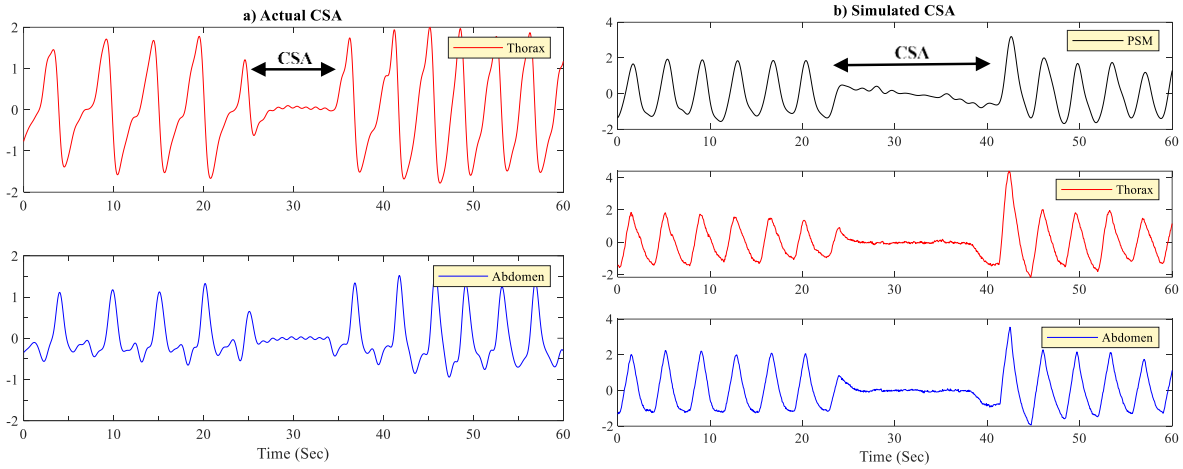


Figure 6.1 CSA, a) Actual event from university College Dublin Sleep Apnea database; b) Simulated event extracted from PSM and respiratory bands.

The next step is to evaluate the performance of PSM in real-life conditions as a home monitoring device. The data adopted for th is chapter were collected in an uncontrolled environment, without wearing respiratory bands as a gold standard sensor. In sleep clinics, SA scoring is associated with a visual review of the PSG performed by a sleep technician. Here, in order to develop a computer-based method for automatic detection of CSA events from non-supervised data described in section 3.4.2 (i.e., 7 days of non-supervised data from each patient), start points and end points of events suspected to be apneic in part of the data are marked manually by a scorer (this thesis's author) who understands the behavior of PSM sensors against the gold standard.

But first, data from section 3.5 (i.e., from Royal Ottawa Mental Health Centre (ROMHC) Sleep Disorders Clinic) is used here to examine the reliability and validity of the scorer regarding the scoring of PSM signals. Matched PSG and PSM data were time-synchronized to ensure that analysis periods were identical. The scoring (i.e., interpretation of raw data signals) started from when "lights out" was noted in the PSG recording. Next, the scorer who was blinded to both the identity and relationship of paired recordings of PSM and PSG manually scored eleven PSM signals in random

order using AASM rules (i.e., a complete cessation of respiratory movements captured by PSM for at least 10 s). From each of the separately scored records, the number of CSA events and their temporal position in PSM signals were obtained.

6.2.1 Statistical Analysis

The scorer agreement is measured to quantify issues of the scorer's reliability arising from the interpretive nature of the scoring. It is most often measured with Cohen's Kappa coefficient [93], [157], [158]. In the sleep study, Cohen's Kappa coefficient is calculated based on epoch-based comparison of scored SA events, where each epoch of 30 s would have a status of apneic or not-apneic. Here for a more detailed investigation, CSA events scored from PSM data were compared to those annotated in paired PSG individually. Events were then classified as:

- TP: PSG and PSM events occurred at the same time;
- FP: Event present within PSM recording but no evidence of an event in the paired PSG recording;
- FN: Event present within a PSG recording but no evidence of an event in the paired PSM recording.

As the absence of a respiratory event is not scored within PSG or PSM recordings, the determination of TN events is not applicable for the event-by-event evaluation.

Overall, as illustrated in Table 6.I, the number of CSA events scored in PSM data were not significantly different from the PSG events. There was only one considerable disagreement for patient 5 with severe respiratory disturbance and an AHI value above 30. The event-by-event evaluation approach is very stringent in dealing with FP and FN events, especially if the total number of events, including TP, FP, and FN, are few. According to Table 6.I, in the worst scenario

Table 6.I Comparison of individual scored CSA events with those annotated in PSG data.

	CSA Events						
		Scorer			Evaluation		
Patient No.	PSG	TP	FN	FP	Precision	Sensitivity	F-Score
1	0	0	0	0	---	---	---
2	4	3	1	0	100	75.0	85.7
3	0	0	0	0	---	---	---
4	0	0	0	0	---	---	---
5	8	6	2	16	27.3	75.0	40.0
6	0	0	0	1	0	---	---
7	2	2	0	1	66.7	100	80
8	17	14	3	5	79.0	83.3	81.1
9	0	0	0	0	---	---	---
10	0	0	0	0	---	---	---
11	1	1	0	0	100	100	100

(i.e., patient 5), the precision of 27.3% means that overall there are only 16 CSA events misclassified, and most of the CSA events in the PSG data were scored in the PSM data. In severe respiratory disturbance (i.e., an AHI value above 30 [92]), such numbers of CSA events wrongly detected (FP) or ignored (FN) would not cause any significant misdiagnosis, and recommended treatment can be followed.

Since the severity of SA disorder is determined based on the number of SA events, the correlation coefficient was used as a form of reliability testing between the number of events scored in PSM (TP and FP events) and the number of events annotated in paired PSG data. A P-

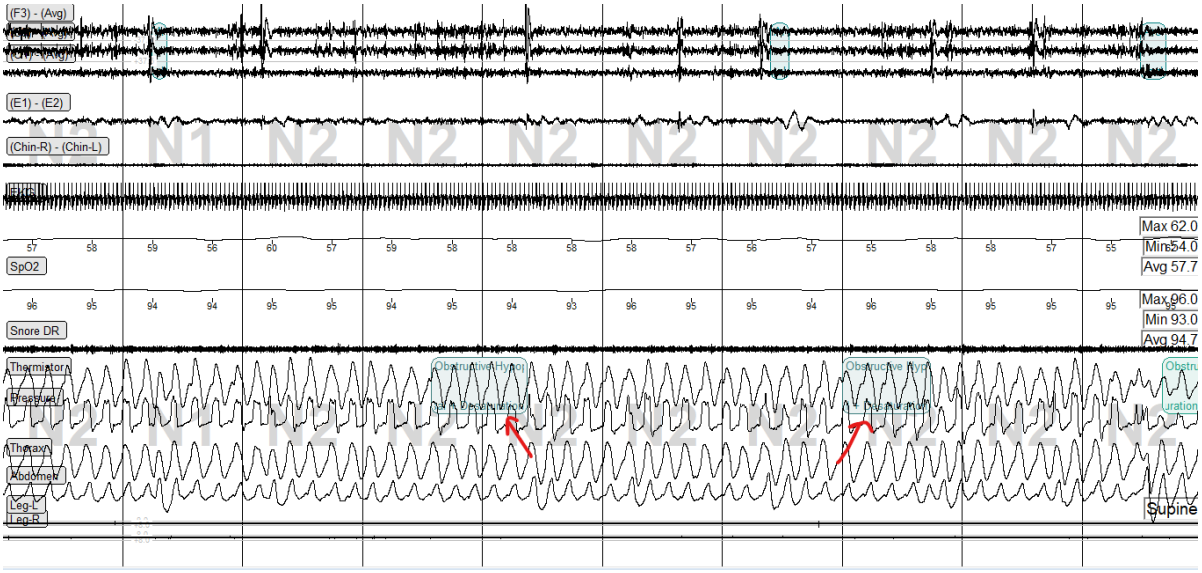


Figure 6.2 Screenshot of PSG data, patient (1). Example of inter-operator variability from PSG data for scoring hypopnea events. The sleep technician noted: *"I do not agree with these hypopneas; no oxygen desaturation or arousals marked, not enough decrease in the in the airflow."*

value (p_v) of < 0.05 was considered to be significant. There was a good correlation (r_c) between the two procedures, and no significant difference was found ($r_c = 0.8809, p_v < 0.01$).

Although the reliability of the scorer with PSM data for this study has been established from the above, the issue of suffering from intra- and inter-operator variability because of manual physiological signal interpretation is not negligible, even with expert knowledge. Figure 6.2 illustrates an example of inter-operator variability for scoring hypopnea events where the presence of oxygen desaturation or arousals are necessary. The sleep technician did not agree with two events scored as hypopneas because she could not confirm oxygen desaturation or arousals, and according to her observation, there was not enough decrease in the amplitude of the airflow signal (thermistor signal).

The scoring of PSG is at high risk of poor reliability due to subjective interpretation and non-standardized definitions of related variables [157]. Additionally, physiological signal interpretation is a tedious process where human errors can be caused by fatigue. It has been

suggested to use computer-supported signal analysis systems, which can avoid intra- and inter-operation variability and errors by fatigue. Additionally, in comparison to human interpretation, most computer-based interpretation can be performed quicker and be more cost-effective [98], [105] [157]. One way to design such a system is to choose criteria based on overall visual inspection of the sleep signal. However, this method still is not completely objective since the criteria are selected by the designer of the system. In the following, to resolve this problem and design an automatic system, a methodological approach is proposed using different methods.

6.3 Method and Material

6.3.1 Database

The database used in this section was part of the long-term nocturnal data presented in section 3.4.2, Chapter 3 (i.e., non-supervised data). A week of data from each of the 9 participants was randomly selected.

6.3.2 Pre-processing

The raw data collected by PSMs in non-laboratory environments are voluminous and noisy, and therefore require preprocessing. We propose a pre-processing pipeline composed of the following steps:

6.3.2.1 Occupancy Extraction

As a subject leaves the bed, a significant drop in PSM signal amplitude occurs. To decrease the computational complexity and mostly to discard irrelevant data, times when the bed was not occupied were detected and removed based on the algorithm in [159]. The basic idea behind the adopted algorithm was first introduced in [160]. However, the threshold selection in the algorithm depended on anthropometric information of participants. Therefore, in [159] for times when no

prior anthropometric information was available (such as in this study), to select a proper threshold, some modifications were applied to the original algorithm.

According to the algorithm, a single output was first obtained from the summation of all 72 raw sensor signals, from noon to noon for each night, denoted as P. For occupancy extraction, none of the steps for signal combining such as polarity adjustment (explained in chapter 4) were applied, because the goal was to obtain the load signal and not the breathing signal. The occupancy then was determined when P exceeded a threshold, T, defined by:

$$T = \beta + \tau \tag{6.1}$$

where β is the base value obtained as the minimum value of P. To make T independent from the person who is lying in the bed, it was established based on using the maximum and minimum value of P, respectively denoted by $Max(P)$ and $Min(P)$, such that:

$$\tau = \frac{Max(P) - Min(P)}{n} \tag{6.2}$$

where *n* can be adjusted depending on the sensitivity required in an application. To avoid false detection of occupancy, a condition was stipulated where occupancy or non-occupancy less than 30 s was disallowed. Also, if the difference between $Max(P)$ and $Min(P)$ was below the threshold, it meant that no occupancy occurred during data collection time. Figure 6.3 shows the evolution of P, defined from noon to noon the next day. As indicated by the red line, there is a significant drop between the loaded and unloaded part of the signal. The subject was only in bed from 1 (am) to 4 (am) approximately.

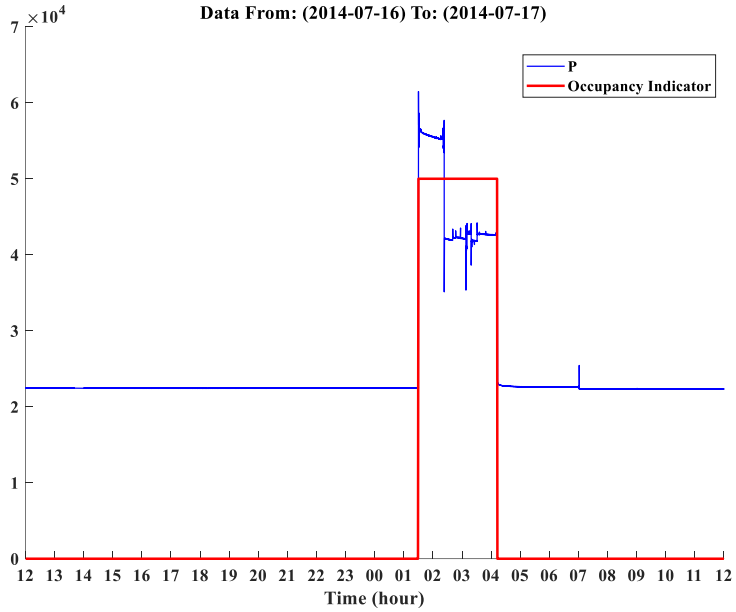


Figure 6.3 Occupancy extraction

The occupation data obtained in this step were kept for further processing, and the rest of the data were discarded.

6.3.2.2 Bandpass Filtering

After discarding the unloaded parts of the signals, an FIR linear phase bandpass filter with a passband corresponding to 0.07-0.8 Hz filtered each of the 72 signals separately, as in section 4.2.1.

6.3.2.3 Combining & Concatenating Signals

The SNR-MAX method explained in section 4.2.3.2 was applied to the sensor array signals to generate a single output signal with better signal quality. Signals were combined every 30 s, with a 50% overlap between the 30-second windows or segments. Different window sizes may be needed to detect CSA, especially since the minimum acceptable length for CSA events is 10 s. Therefore, after signal combining for every 30-second segments, all segments were concatenated

to have one single signal using the following method to achieve a smooth transition across overlapping segments [161], [162], Figure 6.4:

1- As mentioned in section 4.2.2, depending on the polarity of the chosen reference signal in each segment, the resulting combined signal can have different polarity compared to its preceding segment. Therefore, for every two consecutive segments, if the correlation of the overlapping samples of the two segments was negative, the second segment was inverted by multiplying by -1.

2- The last N overlapping samples of the first segment were gradually attenuated by multiplying with a raised-linear function changing from 1 to 0 ($f(n)$; $n = 1,..., N$), while the first N samples of the second segment were gradually amplified by multiplying with $1 - f(n)$.

3- All the modified segments were aligned based on their starting point in the original signal and then added together to construct a concatenated signal.

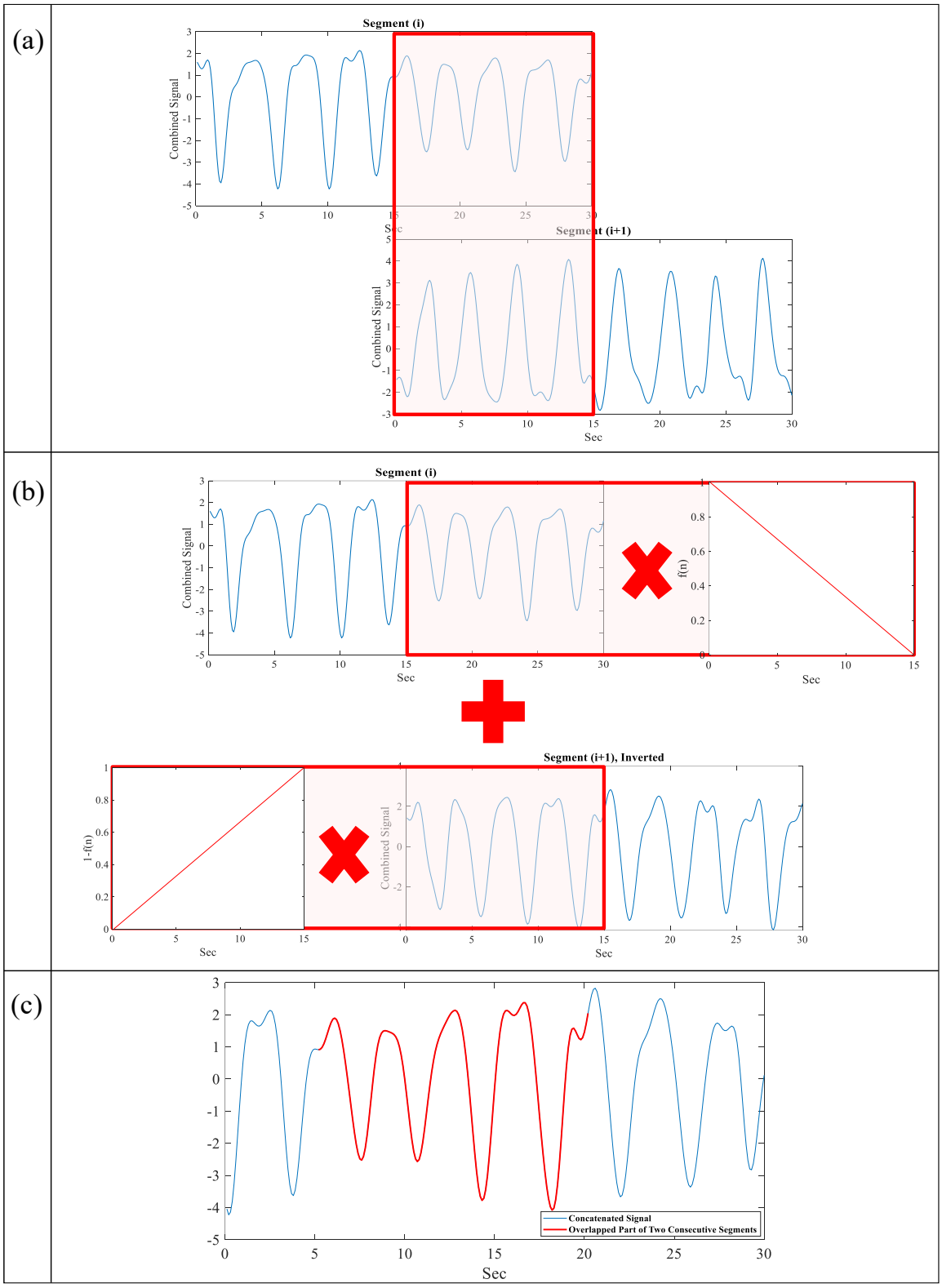


Figure 6.4 Concatenation of signal segments. (a) Sample input segments. (b) First, segment (i+1) is inverted. Then, the output is computed by gradually attenuating the end of the segment (i), gradually incrementing the start of the inverted segment (i+1), and add adding the two parts to yield a smooth transition. (c) Concatenated signal.

6.3.2.4 Signal Normalization

PSM signal amplitude can be affected by body movements and different body postures [137]. In order to provide a more accurate estimate of respiratory depth and volume, the signals must be normalized by the means of the body movements. Body movements as a result of bed entries/exits, rollovers, sleep starts, posture shifts, or small movements such as arm/leg twitches and gasps (deep breathing), cause large fluctuations in the signal. These fluctuations are detected here as a value that is more than three median absolute deviations (MAD) away from the median of the combined signal [9]. After detecting movements, each part of the signal between any two movements is normalized in two consecutive steps: first removing the average of the signal, next dividing the signal by the maximum absolute value of the signal, excluding the 5% largest negative and positive values.

Sorting data is computationally expensive. However, it can increase the performance of the classifier. Dividing the signal by the maximum absolute value of the signal without ignoring outliers might result in a signal with a much smaller breathing range, which can potentially be confused with an apnea event.

The resulting output of this pre-processing pipeline, $q[n]$, is a single normalized signal per day. It is then fed to all the different systems designed in section 6.3.3, section 6.3.4, and section 6.3.5. Depending on the method used, $q[n]$ is segmented into smaller sequences of 9 s with a 50% overlap, or it is injected into a detection system as a single observation for each day. Regardless of different data segmentation, the data of two patients (equivalent to 14 days of data) are randomly selected as test data. The test data are the same for all methods, to allow a meaningful evaluation and comparison.

6.3.3 Power Detector Method

The simple method for detecting SA events from RIP-sum has been developed in Chapter 5. The algorithm is used in this chapter to investigate its ability in detecting CSA events from PSM signals. The summary of the approach is as follows. The approach includes two steps:

6.3.3.1 Power Calculator

According to AASM, the breathing baseline is defined by the 3 largest breaths in the 2 minutes preceding the onset of an event in adults without a stable breathing pattern [17]. The breathing baseline determines a threshold for detecting an apneic event. For this purpose, the previous 2 minutes of data are divided into 9 s sub-segments with a 50% overlap, and for each sub-segment the power of the signal is calculated.

Let $x_i[n], n = 1, 2, \cdots, M_p$ be the data samples of the i^{th} sub-segment taken from $_{q[n]}$, the normalized signal previously described. Then for each sub-segment, the power is calculated according to:

$$P_i = \frac{1}{M_p} \sum_{n=1}^{M_p} | x_i[n] |^2 \tag{6.3}$$

where M_p is the number of samples within a sub-segment (i.e., 90 samples for a PSM with a sampling rate of 10 Hz).

6.3.3.2 Detector

The powers of the sub-segments calculated in the last 2 minutes are sorted in ascending order to compute a threshold. The threshold is calculated as a fraction (FC) of the sub-segment power with an 80-percentile position from the sorted sub-segment powers, in order to be robust to cases where there are several events or movements (with extreme signal power) in the previous 2 minutes. The value of the 80-percentile segment power is too large to be considered as the threshold by itself, and

therefore the fraction FC helps to tune the threshold. If the power of a 9 s segment after 2 minutes of data is smaller or equal to the threshold, it is classified as an event and the flag vector related to this part of the signal is set to one.

6.3.4 Classic SVM

SVM as a kernel-based classifier has been found to often provide good performance, especially for binary classification in the field of biomedical pattern recognition [163]. Initially, SVM was designed to solve binary classification and regression problems. SVM has been successfully applied in many areas [164]–[166]. Finding an optimized hyperplane is the fundamental idea of SVM. Such a goal can be defined as the maximization of the minimum distance, i.e., margin, between the hyperplane and the closest training data points, i.e., support vectors [167]:

$$\max_{w,b} \min\{\| x - x_i \| : w^T x + b = 0, i = 1,...,m\}, \tag{6.4}$$

where w is the weight vector and b is the bias for the model. In linear binary SVM classification, w and b can be rescaled in such a way that the point closest to the hyperplane $w^T x + b = 0$ lies on hyperplanes $H_1 : w^T x + b = 1$ and $H_2 : w^T x + b = -1$ for either class, as illustrated in Figure 6.5. Suppose that all the training data satisfy the following constraints:

$$w^T x_i + b \geq 1 \quad \text{for} \quad y_i = +1 \tag{6.5}$$

$$w^T x_i + b \leq -1 \quad \text{for} \quad y_i = -1 \tag{6.6}.$$

These can be combined into one set of inequalities:

$$x_i : y_i[w^T x_i + b] - 1 \geq 0 \tag{6.7}$$

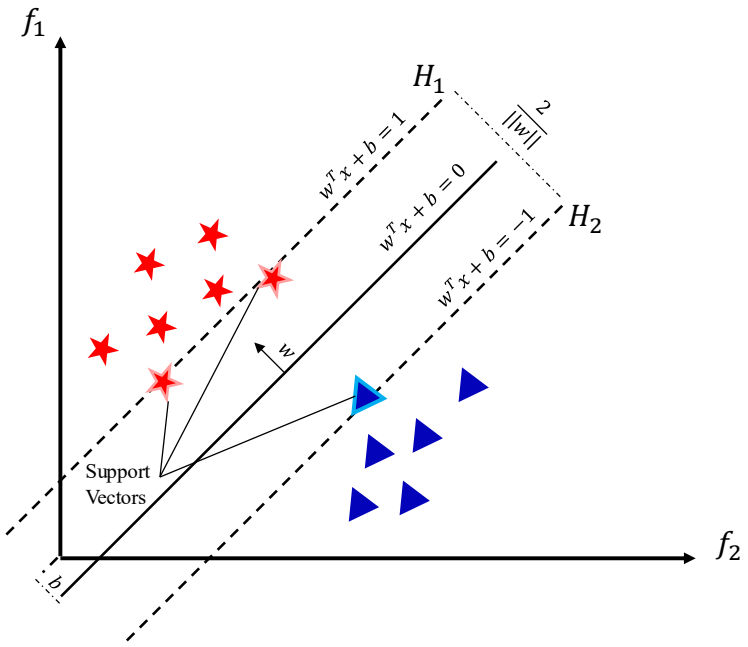


Figure 6.5 Maximum margin hyperplane and margins for a SVM trained with samples from two classes.

As it is shown in Figure 6.5, the width of the margin is simply equal to $2/\|w\|$. Support vectors are the training points for which the equality in equation (6.7) holds. To find the pair of hyperplanes, H_1 and H_2, which give the maximum margin by minimizing $\|w\|$, the optimization problem in (6.4) can be restated as $\tau(w): \min_{w,b} \tau(w) = \frac{1}{2}\|w\|^2$, subject to constraints [168]. To solve it, a Lagrangian is constructed as in (6.8):

$$L(w,b,\alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{m} \alpha_i (y_i[x_i^T w + b] - 1), \tag{6.8}$$

where $\alpha_i > 0$ are Lagrange multipliers. The minimization of (6.8) leads to $\sum_{i=1}^{m} \alpha_i y_i = 0, w = \sum_{i=1}^{m} \alpha_i y_i x_i$.

Based on Karush-Kunh-Thucher (KKT) conditions, $\alpha_i (y_i[x_i^T w + b] - 1) = 0$ for $i = 1, \ldots m.$ The non-zero α_i have a crucial role in the solution of the optimization problem, known as support vectors.

The optimization problem can be transformed into the dual optimization problem:

$$\max_{\alpha} W(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j x_i^T x_j \tag{6.9}$$

with constraints $\alpha_i \geq 0, i = 1,...,m$, and $\sum_{i=1}^{m} \alpha_i y_i = 0.$

As a result of optimization, the decision function can be written as:

$$D(x) = \text{sgn}(\sum_{i=1}^{m} \alpha_i y_i x^T x_i + b) \tag{6.10}.$$

The dot product $x^T x'$ in (6.9) and (6.10) can be replaced by a kernel function $k(x,x') = \phi(x)^T \phi(x')$ to extend the linear discriminant SVM to a nonlinear machine. The generalized decision function is:

$$D(x) = \text{sgn}(\sum_{i=1}^{m} \alpha_i y_i k(x, x_i) + b). \tag{6.11}.$$

Given the decision function in (6.11), the dot product is the simplest kernel. The polynomial kernel is defined by:

$$k_p(x,x') = [\gamma(x^T x') + \theta]^q \tag{6.12}$$

where q is an integer and $\theta = 0$ or $\theta = 1$.

Another powerful kernel is the Gaussian Radial Basis Function (RBF) represented as:

$$k_G(x,x') = \exp[-\gamma \| x - x' \|^2]. \tag{6.13}.$$

In comparison with non-linear kernels, a linear kernel is less prone to overfitting and faster to train. However, depending on data, a different type of kernel may be required. Comparing the results of various types of kernels (by applying them to cross-validation data), including RBF, third-order polynomial, and linear kernel showed that the use of nonlinear kernels did not

significantly improve the accuracy of the model with our dataset. Therefore, a binary linear SVM classifier was considered to classify segments of PSM signals into two categories, i.e., not-apneic "NA" or apneic "A". "A"-labeled segments are those that have at least 50% of their length falling within a start-to-end time of apnea events.

Thirty-four (34) features from the time and frequency domains were extracted from the remaining segments and fed to the SVM [9]. A comprehensive explanation of the features is as follows.

6.3.4.1 Feature Extraction

For automatic detection of CSA, different algorithms are applied to extract suitable features. Some features are the result of time series data, directly fed to the classifier, whereas others first involve a transformation of time series data into the frequency domain. In order to extract features, $_{q[n]}$ is segmented by rectangular 9-second window functions with 50% overlap, whose length is chosen based on the best performance of the system on the training database. It was observed that increasing the window duration beyond 10 s would reduce the performance of the proposed system as the CSA events can be of 10 s by definition. Increasing the window duration affects the performance of the event detector (which will be described later in the chapter) rather than the classifier. At the opposite, if the window length is too small and does not contain the major part of a respiration cycle, the performance of the classifier drops. However, the performance of the event detector to correct the misclassified windows and the detection of events improves. By considering all these factors and the number of calculations required, a window duration of 9 s is chosen. The next parameter which is adjusted is the amount of overlap. It is expected that a small amount of shift (large overlap) will improve the performance by increasing the resolution of the entire event detection process but will also impose a computational burden. The amount of overlap was chosen to be 50%, since on average (at a rate of 15 bpm) it is long enough to provide a complete respiratory cycle to the next window.

1. Time Domain Based Features

Time series analysis includes methods for analyzing time-series signals directly in order to extract meaningful statistics and other characteristics of the signal. Here are the brief descriptions of different time domain-based measures extracted in this work:

Statistical Measures: These first to fourth order statistical parameters, i.e., mean (M1), variance (M2), skewness (M3) and kurtosis (M4), are computed for each segment of data to quantify the central tendency, degree of dispersion, asymmetry, and peakedness, respectively. Other statistical measures extracted from time series segments are: the median as the measure of the "center" of a signal since it is less sensitive to outliers, the maximal (Max) and minimal (Min) value of the signal, the range as the difference between the Min and Max values of a signal, the root-mean-square (RMS) and the standard deviation (SD).

Sub-segment ratio (SSR): Each segment is divided into five sub-segments. Then, the maximum absolute value is obtained in each sub-segment. Afterward, the SSR is calculated for each segment based on dividing the highest maximum absolute level by the lowest maximum absolute level from the five sub-segments [159].

Hjorth Parameters: Hjorth parameters [169] are quite popular for the analysis of biomedical signals, especially EEG signals for sleep stage classification [170], [171]. For a signal x of length N, and x' as its first derivative, the parameters are calculated using the following equations:

$$\text{Mobility}=\sqrt{\frac{\text{Var}(x')}{\text{Var}(x)}} \tag{6.14}$$

$$\text{Complexity}=\frac{\text{Mobility}(x')}{\text{Mobility}(x)} \tag{6.15}$$

where $Var(x)$ and $Var(x')$ are the variance of x and x', respectively. x' is approximated by the difference between successive elements of x.

Non-linear combinations (i.e., multiplication and division) of existing features can be considered as new features and be helpful to optimize the linear classifier. A linear classifier cannot consider these complex features unless they are explicitly provided for the classifier. Non-linear features in table 6.I, are adopted based on this principle. A brief explanation of these features can be found in [172].

2. Frequency Domain Based Features

Time-domain based features are not always the best representation of signals for many signal processing applications. In many cases, the most useful information is hidden in the frequency content of the signal. Here, from the auto-power spectral density (PSD) of each segment, the following features were extracted. In this section, the frequency band of a signal of interest means 0.2-0.33 Hz, corresponding to the normal breathing range (i.e., 12-20 bpm).

Spectral moments: Median and standard-deviation (SD), as well as the first to fourth order spectral moments in the frequency domain, were calculated. These features characterize the shape of the spectral density of the signal in each segment. In [173], the spectral moments were applied to classify sleep stages from PSG recordings.

Peak power frequency (PPF): frequency at which the highest power exists. Also, PPF within the frequency band of interest, $PPF_{FB,}$ and its proportion to PPF (PPF_{FB}/ PPF) are calculated as two extra features.

Spectral flatness measure (SFM): The ratio of the geometric mean of the signal PSD to the average of the signal PSD is a measure of the spectral flatness of the signal (SFM) also known

as tonality coefficient. The meaning of tonal in this context is in the sense of the number of peaks or resonant structure in a power spectrum, as opposed to a flat spectrum of white noise. A high spectral flatness represents a similar amount of power in all spectral bands, as in white noise. Therefore, the graph of the spectrum would appear relatively flat and smooth. A low spectral flatness indicates that the spectral power is concentrated in a relatively small number of bands. Therefore, the spectrum would appear spiky, such as a mixture of sine waves [174].

$$\text{SFM}=\frac{\text{GeometricMean(X)}}{\text{Mean(X)}} \tag{6.16}$$

Relative Spectral Powers (P_R): The percentage of the total area under the curve corresponding to the frequency band of interest.

The non-linear features calculated for time domain-based features are also calculated for frequency domain based features. For time-domain based features, x is a segment of the signal, whereas for frequency-domain based features, X represents the power spectrum of each segment. The sample average removed from the signal and the maximum absolute value of the signal between any two body movements, both used to produce the normalized signal $q[n]$, are also included as two features for each segment. This is in order to avoid information loss, especially when the part of the signal between two consecutive body movements is too short and mostly noisy. The information on the extracted features is summarized in Table 6.II.

6.3.4.2 Dealing with Imbalanced Classes

Apnea detection is an anomaly detection problem, which deals with imbalanced classes where one class (in this case normal breathing) is significantly over-represented in the dataset. A simple way to balance the datasets and consequently prevent the “accuracy paradox”, where one class is completely ignored by the classifier and the accuracy is only reflecting the underlying class

distribution, not the actual performance, is to perform resampling of the classes, i.e., oversampling the minority class or undersampling the majority class. In practice, each of these approaches has its drawbacks. Oversampling the minority class with a high factor introduces duplicated instances from a small pool of observations. Therefore, it can lead to model overfitting. On the other hand, undersampling the majority class can result in eliminating important instances that provide important differences between the two classes. Here, a combination of both strategies is used to reduce the negative impact of each one. Class "NA" are undersampled randomly. Then, the synthetic minority oversampling technique (SMOTE, a well-known method for oversampling) is used to oversample class "A". SMOTE creates synthetic instances of the minority class by forming convex combinations of neighboring instances. It effectively draws lines between minority points in the feature space and samples along these lines. This allows balancing the dataset without too much overfitting, as new synthetic examples are created rather than using duplicates [175].

Table 6.II List of extracted features.

	Formula/Description
Time-domain based features	- Variance, Skewness, Kurtosis, Standard Deviation (SD), Root Mean Square (RMS), Mean, Maximum (Max), Median, Range, Minimum (Min), Hjorth Parameters (Complexity & Mobility) - Sub-Segment-Ratio [25] $\frac{Max(x)}{RMS(x)}$ $\frac{RMS(x)}{Mean\left(Abs\left(x\right)\right)}$ $\frac{Max(x)}{Mean\left(Abs\left(x\right)\right)}$ $\frac{Max(x)}{Mean(Abs(x))^2}$ Note: for time-domain based features, x is a segment of the signal.
Frequency-domain based features	- Spectral Variance, Spectral Skewness, Spectral Kurtosis, Spectral SD, Spectral Mean, Spectral Median, all corresponding to the selected frequency bands*. - Peak Power Frequency corresponding to the selected frequency bands $\left(PPF_{FB}\right)$. - Peak Power Frequency *(PPF)* $\frac{PPF_{FB}}{PPF}$ - Spectral Flatness Measure corresponding to the selected frequency bands. $\frac{Max(X)}{RMS(X)}$ $\frac{RMS(X)}{Mean(X)}$ $\frac{Max(X)}{Mean(X)}$ $\frac{Max(X)}{Mean(X)^2}$ - The percentage of the total area under the curve corresponding to the selected frequency bands. Note: for frequency-domain based features, X represents the power spectrum of each segment.
Normalization Procedure	- Sample average removed from signal. - Maximum absolute value of the signal (excluding the 5% largest negative and positive values).

* Selected frequency bands correspond to normal breathing range (i.e. 12-20 bpm).

In a simple example, consider the depicted distributions in Figure 6.6. In this figure, the stars and triangles represent the minority and majority classes, respectively. Top figures are schematically indicative of randomly removing samples of the majority class. On the other hand, in the bottom, plots represent SMOTE for the minority class. Samples are created along the line between filled stars, highlighted by unfilled stars. These synthetic samples help break the ties introduced by simple oversampling, and furthermore, augment the original dataset in a manner that generally significantly improves learning.

Applying the linear SVM classifier, 5-fold cross-validation is performed on the training dataset to optimize the factors for undersampling the majority and oversampling minority classes, respectively. It should be noted that test data are not balanced to reflect future data (nature prevalence/ratio).

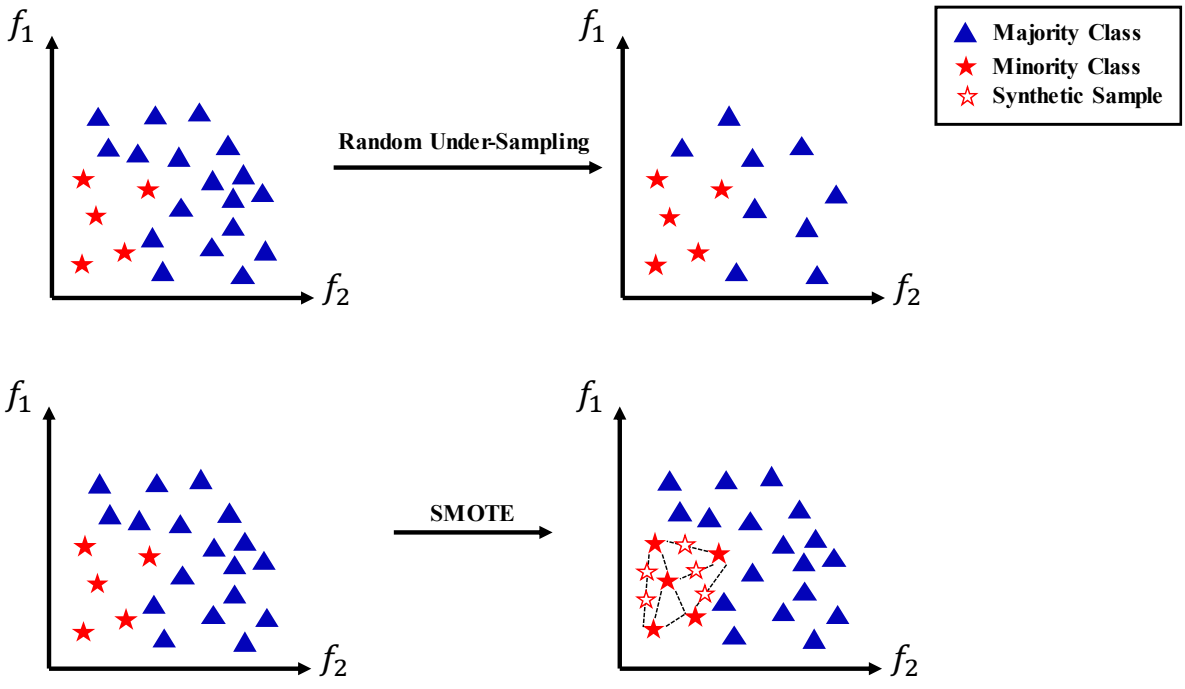


Figure 6.6 A dataset with imbalance problem in two-dimensional feature space.

6.3.4.3 Feature Selection

The extraction of many features is often found to be helpful for classification. However, it can be computationally expensive. Moreover, it can cause the "curse of dimensionality" such that the classifier overfits the training data and loses generality. On the other hand, there can be some redundant features, as well as some irrelevant features, which can complicate the classification and diagnosis system. Therefore, selecting a suitable set of features is also an important process to construct a model. To select features with the "univariate methods", features can be ranked according to their relevance. In this case, the feature that provides a good class separation by itself will rank high and will, therefore, be chosen. However, the underlying feature independence assumptions made by univariate methods can lead to the following:

- Features that are not individually relevant may become relevant in the context of others;
- Features that are individually relevant may not all be useful because of possible redundancies.

In comparison, multivariate methods consider feature dependencies, and therefore potentially, they can achieve better results. There are various multivariate feature selection methods such as wrapper, embedded, and greedy methods. Each of these methods can be adopted in a forward selection or a backward elimination process. In a forward selection, the procedure starts with an empty set and progressively add features, yielding to the improvement of a performance index. On the other hand, in a backward elimination, the procedure starts with all the features and progressively eliminates the least useful ones. Both procedures are reasonably robust against overfitting and provide nested feature subsets. Nevertheless, they may lead to a different subset of features and, depending on the application and the objectives, one approach may be preferred over the other one. In applications involving a selection of the best feature set for improving

classification performance, rather than identifying the best feature, the backward elimination process is often preferred.

In this chapter, after balancing the dataset, the Recursive Feature Elimination (RFE) technique, SVM-REF, is applied to obtain the best set of features. In this multivariate method, features are selected in a sequential backward elimination manner [165]. At each step, a linear SVM learning model is built based on the current feature subset, F. The weight of each feature w was calculated according to the support vectors on the hyperplane of the SVM classifier. The features were then ranked based on the absolute value of the weights as a ranking criterion, and the feature with the worst ranking was eliminated from F. This procedure was repeated until F was empty. The most important features were identified as the last features that were removed from F.

6.3.5 Deep Learning

6.3.5.1 BiLSTM

An LSTM is a type of RNN that can learn long-term dependencies between time steps of sequential data. Unlike CNNs, an LSTM can remember the state of the network between predictions. The essential components of an LSTM network are a sequence input layer and an LSTM layer. A sequence input layer incorporates time-series data into the network. An LSTM layer learns long-term dependencies between time steps of sequence data. The layer contains hidden units providing inputs to memory cells and their corresponding gate units. All units (except for gate units) have connections (serve as inputs) to all units in the higher layers [176].

As an extension of the traditional LSTM network, the bidirectional LSTM (BiLSTM) network can improve model performance on sequence classification problems. While the LSTM layer considers the time sequence in the forward direction, the BiLSTM layer considers the time sequence in both forward and backward directions [177]. In problems where the complete time

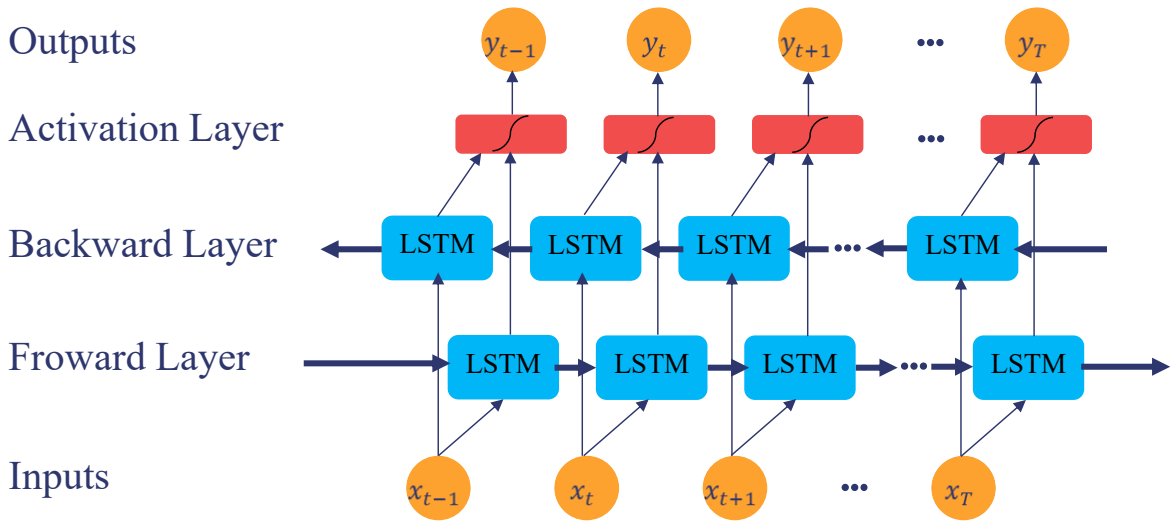


Figure 6.7 The BiLSTM network with two recurrent layers side-by-side two.

steps of the input sequence are available, the BiLSTM network trains two, instead of one LSTM network, on the input sequence. The process involves duplicating the first recurrent layer in the network so that there are now two layers side-by-side as shown in Figure 6.7. It provides the input sequence as an input to the first layer and a reversed copy of the input sequence to the second layer. This can provide additional context to the network and result in faster and better learning of a model.

6.3.5.2 TCN

Sequence modeling for most deep learning practitioners is synonymous with recurrent networks. Yet recent results have shown that a simple convolutional architecture known as a temporal convolutional network (TCN) can outperform recurrent networks such as LSTMs across a wide range of tasks and datasets while demonstrating longer effective memory [155].

A general TCN architecture consists of multiple residual blocks. As illustrated in Figure 6.8, each block comprises two sets of dilated causal convolution layers with the same dilation factor, followed by normalization, rectified linear unit (ReLU) activation, and spatial dropout layers. The input to each block is added to the output of the block (including a 1-by-1 convolution on the input

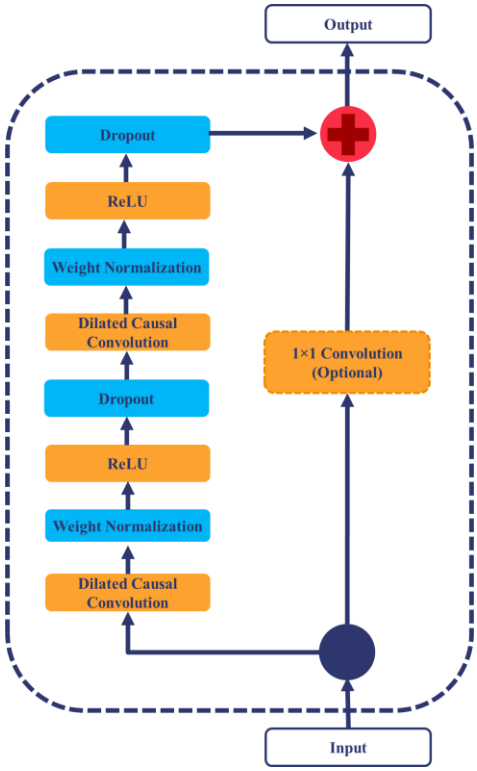


Figure 6.8 TCN residual block.

when the number of channels between the input and output does not match) and a final activation function is applied. TCN combines dilations and residual connections with the causal convolutions needed for autoregressive prediction. Weight normalization is applied to the convolutional filters, and a spatial dropout is added after each dilated convolution for regularization [155].

TCN is based on two principles:

a. The convolutions in the architecture are causal, where an output at time t is convolved only with elements from time t and earlier in the previous layer. Therefore, there is no information "leakage" from the future to the past.

b. The architecture produces an output of the same length as the input, just as with an RNN. It adopts a 1D fully convolutional network architecture, where each hidden layer is the same length as the input layer.

6.3.5.3 Dealing with Imbalanced Classes

Dealing with class imbalance is not just affecting classic machine learning methods, it can also be problematic for DL approaches. Indeed, in the case of imbalanced classes, networks would learn that they can achieve a high accuracy simply by classifying all observations as a member of the majority class. To avoid this bias, for both DL approaches, i.e. BiLSTM and TCN, a weighted classification layer is adopted to compute the weighted cross-entropy loss.

Weighted cross-entropy is an error measure between two continuous random variables. For prediction scores S and training targets T, the weighted cross-entropy loss between S and T is given by:

$$L = -\frac{1}{N_{OB}} \sum_{n=1}^{N_{OB}} \sum_{i=1}^{K} w_i T_{ni} \log(S_{ni}) \tag{6.17}$$

where N_{OB} is the number of observations, K is the number of classes (2 for the case of apnea detection), and w is a vector of weights for each class. The vector of class weights w is inversely proportional to the number of training examples in each class, to give each class equal total weight in the loss [178]. Therefore, in comparison to majority class instances, each instance of the minority class contributes more to the final loss and the majority class is prevented from overexposure to the network representation.

Table 6.III Confusion matrix

		Predicted Class	
		Class Type: "A"	**Class Type: "N"**
Actual Class	**Class Type: "A"**	**Number of class "A" segments correctly identified as class "A" (TP)**	**Number of class "A" segments incorrectly identified as class "N" (FN)**
	Class Type: "N"	**Number of class "N" segments incorrectly identified as class "A" (FP)**	**Number of class "N" segments correctly identified as class "N" (TN)**

6.3.6 Performance Evaluation

Regardless of the method used, the output of the classifier model is a series of class labels for every time step (or separated by an interval equal to the 50% segment shift). A confusion matrix for a binary classification model is presented in Table 6.III. Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class. The classification performances are assessed in terms of:

- Sensitivity: a measure of the ability of the classifier to detect "A" segments

$$\text{Sensitivity}=\frac{\text{TP}}{(\text{TP}+\text{FN})}\times 100\% \quad (6.18),$$

- Specificity: a measure of the ability of the classifier to detect "N" segments.

$$\text{Specificity}=\frac{\text{TN}}{(\text{TN}+\text{FP})}\times 100\% \quad (6.19),$$

- Precision: a measure of the ability of the classifier to reject false detections of "A" segments

$$\text{Precision}=\frac{\text{TP}}{(\text{TP}+\text{FP})}\times 100\% \quad (6.20),$$

- Accuracy: overall performance of the classifier.

$$\text{Accuracy}=\frac{\text{TP}+\text{TN}}{(\text{TP}+\text{TN}+\text{FP}+\text{FN})}\times 100\% \quad (6.21),$$

where TP, TN, FP, and FN are abbreviations for true positives, true negatives, false positives, and false negatives respectively.

Moreover, a rule-based algorithm is adopted from section 5.3.3 of Chapter 5 to obtain an event-by-event evaluation of the presented methods in terms of precision, sensitivity and F-score. For event-based evaluation as opposed to the previous segment-based definitions, TP, FP, and FN represent the conditions in section 5.4 of Chapter 5. Precision is a measure of exactness (i.e., of the events labeled as apnea, how many are labeled correctly), whereas sensitivity is a measure of completeness (i.e., how many events of the class "Apnea" were labeled correctly). These two performance metrics, much like accuracy and error, share an inverse relationship between each other. An attempt to maximize precision usually leads to lower sensitivity values and vice versa. However, unlike accuracy and error, precision and sensitivity are not both sensitive to changes in data distributions. Inspection on the precision and sensitivity formulas yields that precision is sensitive to data distributions, while sensitivity is not. An assertion based solely on sensitivity is misleading since sensitivity provides no insight into how many events are incorrectly labeled as apnea. On the other hand, precision cannot state how many apnea events are labeled incorrectly. The F-score metric combines precision and sensitivity as a measure of the effectiveness of the event detection system, in terms of a ratio of the weighted importance on either sensitivity or precision (here both were weighted equally). As a result, F-score provides more insight into the functionality of the system than the accuracy metric. However, it is still sensitive to data distributions [179], [180].

Figure 6.9 summarizes the process of CSA detection by applying the proposed approaches. After preprocessing, for the first two methods (i.e., power detector method and SVM), $_{q[n]}$ is a sequence of data fed to the feature extractor to drive the most informative features from the data to the classifiers. In this step, while 34 features are extracted for the SVM classifier, the number of features for the power detector method is reduced to one feature only. One of the advantages of

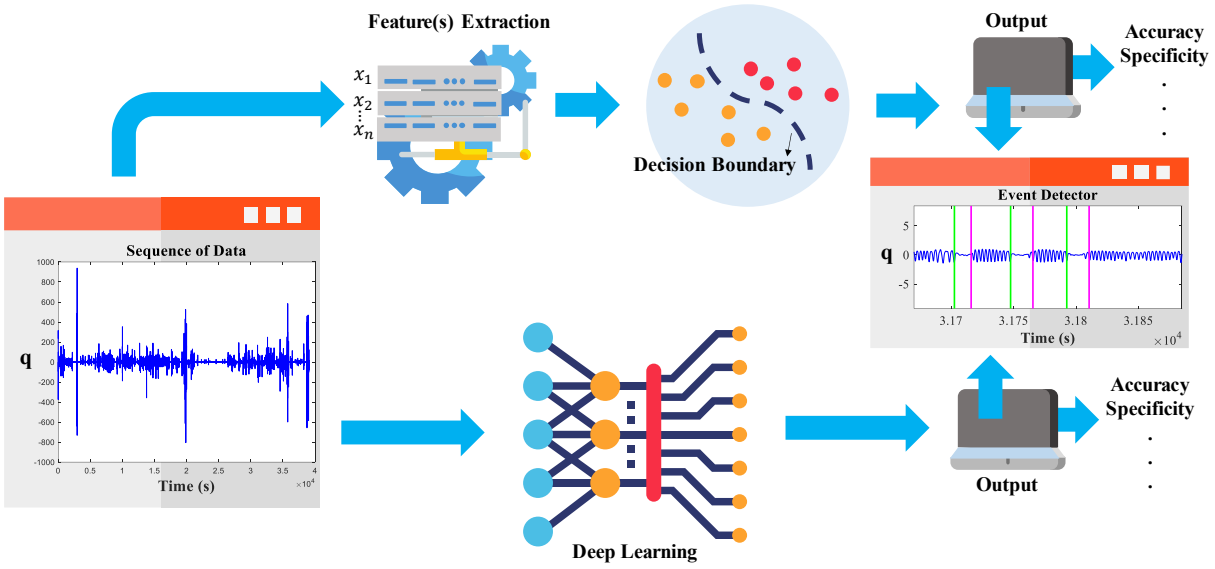


Figure 6.9 Flowchart of detection system. For the event detector, the green and magenta lines are the start and end of the CSA events, respectively.

SVM is that it performs a feature reduction during the training process by assigning more weights to more important features, and therefore reduces the effect of possibly uninformative features on classification. Next, the features are used as inputs to the classifiers for training and optimizing the models. For DL approaches (i.e., BiLSTM and TCN) the processes of feature extraction and network optimization both occur during the training of the networks. All methods are evaluated on the same test data. Finally, a rule-based algorithm analyzes the results of classification as a series of class labels, to extract CSA events individually.

6.4 Results

A total of four approaches are compared in this section to detect CSA events. The model produced by each approach is tuned using a part of the training dataset as validation data. The results of optimization with each method are described next.

6.4.1 Power Detector Method

For the power detector method, 49 days of data are used to tune the FC parameter. FC is the only parameter that needs to be optimized in this method. As described earlier, FC is the fraction of the 80-percentile sub-segment power in the 2 minutes preceding the onset of an event. According to Figure 6.10, the F-score on the training data is gradually improved by increasing FC up to FC=0.1, and further increasing FC reduces the F-score considerably. Therefore, FC=0.1 is applied to the test data (14 days) to evaluate the method.

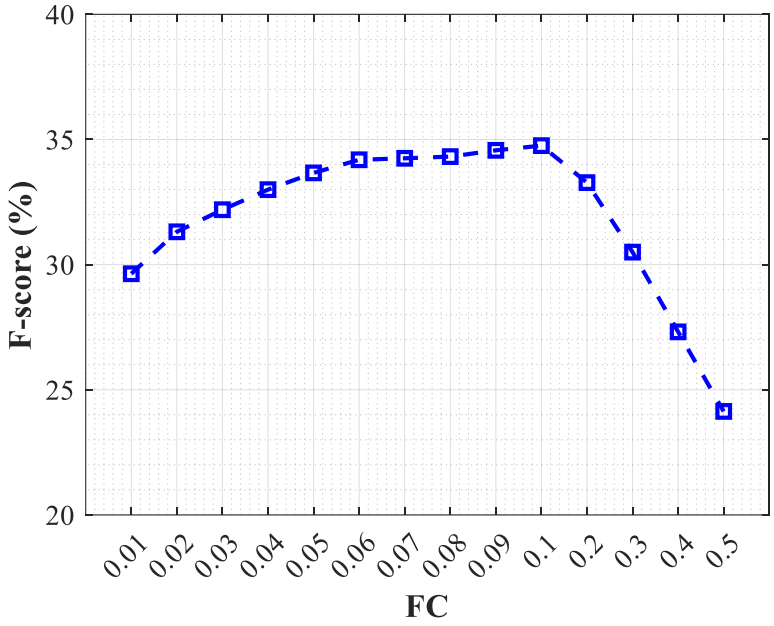


Figure 6.10 F-score of power detector method on training data as a function of FC.

6.4.2 SVM

From each night of each patient's recording, features were calculated from a sliding window and consequently arranged in feature vectors. By leaving out the testing dataset (20%, seven days of 2 patients) commonly used for all presented methods in this chapter, the remaining data (80%, seven days of 7 patients) were adopted to train the classifier model and to check its performance. Data

from all patients were randomly selected either in a training dataset (80%, seven days of 7 patients) or in a testing dataset (20%, seven days of 2 patients). Through the training phase after removing outliers (i.e. the segments that had overlap with gasps and all types of movements detected in section 6.3.2), the training dataset was normalized to zero mean and unit standard deviation. The test dataset was normalized by subtracting the training set mean and dividing with the corresponding training set standard deviation. Normalizing the test dataset using the training parameters is best as sampling errors may negatively bias the predictions. The conceptual idea is that testing data points are supposed to represent real-world data and mimic new, previously unseen data after training the model. In a real application, the new, unseen data could be just 1 data point to be classified. The test dataset is used to get a good estimate of how the trained model performs on any new data. If the test dataset is normalized separately, "information leakage from the future" would happen, and therefore, the result would be an overly optimistic evaluation of the prediction accuracy of the trained model.

The ratio of undersampled "NA" class instances to oversampled "A" class instances is optimized by applying 5-fold cross-validation on the training dataset. Let γ be the ratio of oversampled "A" class instances to undersampled "NA" class instances, which is tested for different values from 0.05 to 0.95, with an increment of 0.1. Therefore, the range of γ is determined by two extremes, where:

- The smallest γ represents an undersampling of "N" class instances to reach the number of "A" class instances, without oversampling the "A" class instances.
- The largest γ represents mostly an oversampling of "A" class instances to reach the number of "NA" class instances, without undersampling the "NA" class instances.

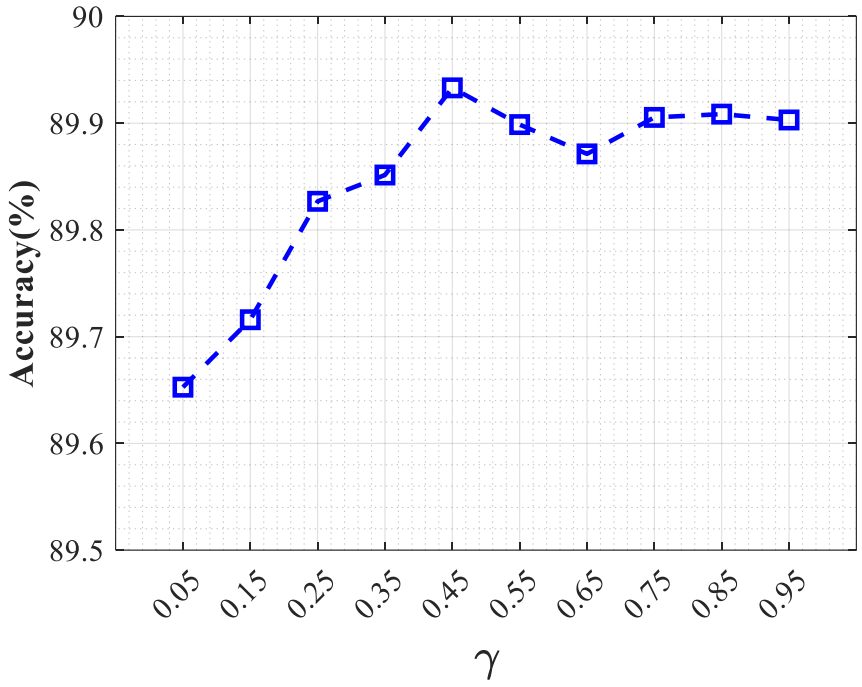


Figure 6.11 Accuracy of the SVM classifier on validation data as a function of the ratio of oversampled "A" class instances to undersampled "NA" class instances.

All other values of γ are between these two extremes. According to Figure 6.11, the classifier achieved the highest F-score on validation data for $\gamma = 0.45$.

After balancing the data for the feature selection, the features were ranked by applying the SVM-RFE technique to find the optimal number of features. 5-fold cross-validation with a limit of 10^6 iterations for convergence was used to score different feature subsets and select the best scoring collection of features. It was observed that the validation accuracy, averaged over all folds, does not improve substantially when using more than 10 features. Table 6.IV shows the list of the selected features along with their respective rank.

6.4.3 Deep Learning

For the BiLSTM and TCN neural models, each of the 63 days of training signals is presented to the network as an individual input sequence. 49 and 14 days of data are used for training and

Table 6.IV List of selected features through SVM-RFE technique.

	Feature Ranking	Formula/Description
Time-domain based features	1	SD
	2	MAX
	3	SSR
	4	$\frac{Max(x)}{RMS(x)}$
	5	$\frac{RMS(x)}{Mean\left(Abs\left(x\right)\right)}$
	6	$\frac{Max(x)}{Mean\left(Abs\left(x\right)\right)}$
	7	$\frac{Max(x)}{Mean(Abs(x))^2}$
Frequency-domain based features	8	$\frac{RMS(X)}{Mean(X)}$
	9	$\frac{Max(X)}{Mean(X)^2}$
	10	Spectral variance

testing, respectively. The BiLSTM and TCN make predictions based on the individual time steps of the sequence data.

6.4.3.1 BILSTM

The architecture of the network contains a BiLSTM layer with 64 hidden units, returning the hidden state output for each input time step, with a fully connected layer of size two, followed by a softmax layer and a weighted classification layer. The number of epochs for training is set to 20, so that the network makes 20 passes through the training data. The training data is shuffled before each training epoch.

The batch size is set to 1 so that the network considers one day of data at a time. The learning rate is set to 10^{-3}. The model parameters are optimized by minimizing weighted cross-entropy loss functions, based on the update rule of the adaptive moment estimation (ADAM) solver [181]. The ADAM solver normally performs better with RNNs than the default stochastic gradient descent with momentum (SGDM) solver [182].

6.4.3.2 TCN

TCN architecture consists of four residual blocks. Each block contains dilated causal convolution layers, each with 175 filters of size 5. The weighted cross-entropy loss is calculated for every batch. The final network is trained via stochastic gradient descent by looping over the sequences in the training dataset computing parameter gradients and updating the network parameters via the ADAM update rule. The number of epochs is set to 10, with a batch size of 1. For each epoch, the training data is shuffled. The learning rate has an initial value of 10^{-5}, which is multiplied by 0.1 every four epochs.

For both DL approaches, i.e., BiLSTM and TCN, the number of epochs is determined by evaluating the performance of the networks on the training dataset and the validation dataset (i.e., data of 2 patients selected randomly). When using more than the final selected number of epochs in each method, the networks start to overfit and their ability to generalize to the unseen validation dataset begins to get worse.

A synopsis of the results is shown in Table 6.V. The table consists of two parts: Model Performance, Event Detector Performance. As mentioned in Section III, in addition to the performance of each classifier, the performance of each method is also assessed based on event-by-event evaluation as detected by a rule-based algorithm.

Table 6.V Performance evaluation of the approaches considered.

	Model Performance (%)			Event Detector Performance (%)		
	Accuracy	Specificity	Sensitivity	Sensitivity	Precision	F-score
Power Detector	---	---	---	86.0	39.2	53.9
SVM	89.3	92.4	65	94.7	56.1	70.5
BiLSTM	**95.1**	**96.0**	**85.7**	**88.0**	**82.0**	**85.0**
TCN	91.8	93.8	70.7	80.7	77.9	79.2

6.5 Discussion

The number of SA events per hour of sleep indicates the severity of SA and therefore it is important for the clinical diagnosis of this disorder. Developing and validating algorithms for SA monitoring only by comparing the number of estimated events with the one reported in the database is not sufficient. It is important to validate any detection method through a temporal event-by-event evaluation. Additionally, previous studies have suggested that respiratory event duration is an important physiological biomarker of SA and can be used for better management of the pathophysiology of this disorder [183].

In this chapter, although methods were implemented only for the detection of CSA events, i.e., a sub-type of SA events, the number of these events per hour of sleep can be useful to determine the severity of SA disorder which is clinically more significant than assessing severity [184], [185]. Therefore, in addition to evaluating the performance of the optimized models as illustrated in Table 6.IV, we evaluated the implemented methods through a rule-based algorithm and event-by-event metrics. The rule-based algorithm can also provide information about individual event durations since it is based on the temporal information of events in signal sequences.

PSM does not directly measure the airflow or respiratory effort. It captures movement information transmitted through a mattress on which a person is lying. The patients are not constrained to certain positions on the mattress, so the positions of limb, spinal cord, torso and shoulder positions can vary with respect to sensor locations. Nevertheless, respiratory signals and therefore CSA events can still be captured by applying signal processing to PSM data. PSMs have provided a good detection of CSA compared to gold-standard data [156]. However, for the method implemented in [156], a patient-specific classification or decision threshold had to be defined, where a value above that threshold indicated normal breathing, and a value below indicated CSA. The proposed algorithm needed to adapt to the different signal strengths seen with different mattress types and to individual body weights, using a relative threshold rather than a fixed threshold on the moving variance. Additionally, the ability of a threshold-tuning algorithm to correctly classify CSA events was affected by different positions especially supine position which tends to have the lowest signal-to-noise ratio [137]. This can also be described as the weakness of the proposed approach in the pre-processing stage, since it failed to reduce the sensitivity of the algorithm to individual factors such as body position, the patient's weight and type of mattress [137], especially when this information is not available.

In this work, in order to mitigate the sensitivity of PSM to different body positions, the 72 measured PSM signals are combined based on their quality of capturing the breathing signal, and then the output is normalized to achieve consistency in the strength of the signal. However, according to the results in Table 6.IV, given the sensitivity of PSM to body movements, the power detector method as a simple method with an adaptive threshold is less reliable for CSA event detection, in comparison to more complex methods. The method is an event-based approach. It

achieves the lowest precision (i.e., many FP events) on test data among all the implemented methods.

In Type-4 devices of sleep monitoring, signals such as oronasal thermal signals, positive airway pressure flow, or alternative signals such as RIPsum (sum of the thorax and abdomen belt signals) are usually used to score apneic events for adults [17]. These signals are less sensitive to movement than PSM. Perhaps applying a simple threshold-based method to these signals could achieve better performance. Here, although the threshold in the power detector method is tuned adaptively, great signal amplitude variations due to body movements can cause the misdetection of normal breathing as an apnea, especially since regular breathing produces only a very low variance in comparison to body movements. Moreover, PSMs are used as unsupervised home-monitoring devices. Unlike supervised sensing with PSG, several unknown environmental parameters can affect the process of data acquisition using PSMs. Optimizing a threshold over all these parameters can be more challenging than for other approaches.

In comparison to the power detector method, SVM has better performance with an F-score of 70.8%. SVM is a discriminative model that attempts to model the training data even if that data is noisy. As mentioned in section 6.3.4, in order to deal with imbalanced classes, the ratio of undersampled "NA" class instances to oversampled "A" class instances is optimized by applying 5-fold cross-validation on the training dataset. We found that for our setup and data, different values of γ did not change the accuracy of the classifier significantly. At $\gamma = 0.45$, in comparison to other values, the classifier has a slightly better performance. Therefore, this value was used to generate the data. It should be noted that although the value of γ, and consequently the scale of resampling the two classes "NA" and "A", does not have much effect on the system optimization and associated metrics, it is a must-do step to prevent the "accuracy paradox" from imbalanced

classes. Using SVM had advantages over simple thresholding algorithms, such as automatically learning the detection criteria based on the recorded data. It can also combine different features of the signal instead of one feature, based on the quality of each feature to detect an event and therefore achieve an automated and more generalized detection system than decision threshold methods. Once the system is optimized, including the classifier model, it can be used to detect apnea events without considering any patient-related modification.

SVM treats the input data as a feature vector and therefore discards the temporal information of the signals. In contrast, BiLSTM is a static model that inherently models temporal relations among time steps of sequence data or incorporates temporal coherence (by regularization and/or temporal pooling) [186]. For BiLSTM, the batch size is set to 1 to prevent the network from interpreting signals incorrectly due to padding (so that all sequences have the same length in the same batch of data).

Unlike RNNs (including BiLSTM) where the predictions for later time steps must wait for their predecessors to complete, in TCN networks convolutions can be done in parallel since the same filter is used in each layer. Therefore, in both training and testing, the input sequence can be processed as a whole in TCN networks, instead of sequentially as in BiLSTM networks.

An epoch is one complete presentation of the training data to a DL network. The number of epochs used depends on a variety of factors, such as network architecture and solver, as well as the complexity of the problem and the data available for the training procedures. Due to the fundamental differences in the structure of TCN and BiLSTM networks, the number of epochs is individually optimized for each method.

For the BiLSTM and TCN approaches, the need for manual feature extraction is eliminated and discriminative features are directly learned from temporal PSM signals. These methods can be more

flexible than SVM since each building block can be modular [187]. The DL models are capable of learning representations of the key factors and interactions from the data itself, through direct feature learning in a supervised manner. The SVM classifier dealt with CSA detection by choosing a limited number of variables to consider, which can be the reason for a noticeable amount of false-positive predictions. For the basic power detector method, this shortcoming is more problematic since the extracted information from a signal is summarized into one variable only. In comparison, the architectures of SMV and power detector method ease the implementation stage [188]. However, DL networks are computationally expensive, and they require high-end GPUs to train in a reasonable amount of time. Here, all approaches were implemented using MATLAB™ 9.6 with a NVIDIA GTX1050 Ti™ GPU on a Windows 10™ environment. Using the GPU, the TCN and BiLSTM networks needed respectively 182 min and 265 min to be trained, whereas with a single CPU the SVM method was trained in 118 min. These execution times are independent of the time taken to optimize the hyperparameters for each method.

According to Table 6.IV, the BiLSTM method outperforms all other methods with an F-score of 85%. Contrary to the results for generic data in [155], we observe that when dealing with PSM sequence data, the BiLSTM network leads to better performance than the TCN network.

It is important to know that there is always a trade-off between precision and sensitivity. An inspection of precision and sensitivity values in Table 6.IV reveals that commonly the results of all four adopted methods contain less undetected events (higher sensitivity) and more events detected wrongly (lower precision). Although for screening purposes, in comparison with the opposite condition (lower sensitivity vs higher precision), having higher sensitivity vs. lower precision is more welcomed, it can be due to the limitations of using single-channel type-4 devices to detect SA. When single-channel type-4 devices are used, there is no access to information from

other channels of data such as SpO2 (to detect oxygen desaturation), position (to have more accurate breathing baseline) and EEG (to check whether the arousal happens after an apnea event or not), in order to decrease the number of FP events [45]. As an example, when using nasal pressure as a type-4 device (one or two signal channels, typically arterial oxygen saturation and airflow), the signal can produce false-positive events if the patient is intermittently mouth breathing. Also, it can be a poor-quality signal if the patient is mouth breathing for long periods of time, which can result in false hypopnea events. For patients with underlying lung disease (e.g., Chronic Obstructive Pulmonary Disease (COPD)) or those who are studied at altitude, the baseline oxygen saturation is lower because of the shape and thresholds of the oxyhemoglobin desaturation curve. Adopting oximetry analysis designed to detect transient drops in oxygen saturation in these patients may show more desaturations, which could improve the sensitivity of a monitor but would likely adversely affect its specificity [189].

It should be noted that in both cases, i.e., using oxygen saturation or nasal pressure alone, if the signals are combined with other types of signals, the false results could decrease significantly. In clinical guidelines for the use of unattended portable monitors, AASM claimed that type-4 home-unattended studies had a wide variance of false positives (ranging from 41% to 73%) in terms of diagnosis of an OSA patient [1]. However, given the easy-to-use in comparison to PSG, pressure sensor arrays for home testing can certainly be useful for screening SA and to improve compliance with physicians' recommendations for sleep studies.

Overall, in comparison with the two other conventional classification methods, the implemented DL models had a greater ability to detect CSA events with the highest accuracy of 95.1% achieved by the BiLSTM. This is most likely due to exploring a broader range of features and finding a more suitable feature set by DL networks than the ones manually defined by the human operator.

6.6 Summary

In comparison with the manual operation of detecting and diagnosing SA by an expert, computer-supported signal analysis systems can avoid errors due to intra- and inter-operation variability and fatigue caused by the tedious process of annotation [177], [178]. In addition, most computer-based analyses can be done quicker and more cost-effectively [98]. For automatic and computer-based detection of CSA from PSM signals, in this work, we implemented two DL approaches including BiLSTM (as a representative of RNNs) and TCN (as a representative of CNNs), and we compared them to approaches used in previous studies: SVM (as representative of manual features combined with conventional classification approaches) [9] and a simple threshold-based method (implemented based on AASM rules for SA detection and diagnosis) [8]. In our experiments, the implemented DL models have shown a greater ability to detect CSA events, because of their better capacity to automatically extract essential features for the tasks.

In supervised learning and especially in DL, a large amount of training data are required but are often not available in the medical domain. Obtaining datasets that are comprehensively labeled and annotated still remains a challenge in the biomedical field [101]. Our study was affected by this limitation to some extent. Even with more than 400 hours of information, its variety was limited to only 9 patients. Nonetheless, we believe that the nocturnal data collected by PSM proved to contain valuable information and our study paves the way for the practical use of PSM as a home monitoring device on a larger scale.

This chapter presented the theoretical background for various signal processing and pattern recognition steps required to detect CSA. It includes pre-processing of measured signals, applications of several signal processing techniques for extraction of features, selection of a suitable set of features, and classification/detection. The ultimate goal of pursuing this study is to

observe older adults' sleep behavior over time, which is a valuable measure from a clinical perspective. Although the results showed that DL approaches had better performance than SVM, due to the small amount of labeled data available for each patient, SVM will be further developed in the following chapter to apply it to longitudinal data of patients and to obtain an estimate of respiratory changes (including apnea), time in bed, total sleep time, and sleep efficiency over time.

Chapter 7: Longitudinal Sleep Study

7.1 Introduction

Sleep is a physiological need; a state where the response of the brain to environmental stimuli has stopped reversibly. Despite developments in the field of sleep medicine, neither the society nor the physicians are adequately informed about sleep and sleep disturbances. This is while diseases associated with sleep are frequent in the population and can have significant consequences on the everyday life of patients. Evaluating a patient with sleep disorder requires multiple steps. First, identifying the main symptom. Second, assessing a detailed history of the sleep and wakefulness cycle. Third, collecting the medical history of the patient, a list of previously used medications, use of tobacco and alcohol, family history, detailed information about the school, work, family and social life, and a physical exam of bodily systems.

Relevant laboratory tests are performed for differential diagnosis, while PSG provides the definitive diagnosis [190]. Even though PSG offers useful information, it has serious limitations since it is collected in a clinical environment (instead of home) and can only assess a single-night sleep. Sleep is a dynamic process that varies from day-to-day. Therefore, it is essential to measure multiple nights of sleep data for medical, research, and wellness reasons. In this regard, the home diary approach can be preferred since this method captures the experience of people in their home environment longitudinally. However, it lacks objectivity. Home sleep monitors can bridge these two extreme approaches through some objective measures over time, ideally in parallel with subjective diary reports. Using home monitoring can allow individuals to identify patterns of interest that are related to their subjective senses [123], [191]. They offer the potential to provide a

more realistic platform, capturing many nights of sleep data. Longitudinal home monitoring allows tracking the sleep dynamics of the same patient at different points in time and reducing the between-subject variation of the measurements. It provides the possibility to link sleep variability to the patient's lifestyle and behaviors such as caffeine, alcohol, medications, stress, and exercise. As these activities vary from day to day, some complex effects and interactions are expected to occur. The longitudinal analysis provides the opportunity to quantify sleep based on these various factors and to implement personalized behavioral modification plans to optimize sleep [123], [192], [193] (specific medical or psychiatric treatment, or simply providing feedback [194]).

One of the most notable example of individual variability in sleep involves sleep apnea symptoms. Sleep apnea is described as the most disruptive sleep disorder in the field of sleep medicine, causing sleep disturbance. It fragments or interrupts rapid eye movement (REM) sleep in many individuals. Significantly, half of the individuals with severe sleep apnea experience daytime sleepiness (as the most noticeable symptom of sleep apnea), whether assessed by subjective reports or by objective measurements.

In this chapter, the concept of the proposed algorithms for automatically detecting CSA events from PSM data in Chapter 6 is expanded for long-term sleep monitoring. The focus of the previous chapter was to provide a generalized model for detecting CSA events, regardless of the sleep characteristics of individuals. While it could be used as a primary diagnosis tool, leading to a significant reduction in the diagnostics cost and waiting time for a sleep study, it can be improved by considering subject-specific factors. PSM has proved to be a promising source of data for health monitoring [6], facilitation of preventive care and aid in the management of ongoing illness [12]. The device can perform similarly to actigraphy, which is one of the most popular means of monitoring sleep quality longitudinally at home. Actigraphy records movements by devices

equipped with an accelerometer [195]. Home monitoring, as the name implies, is conducted in the home, where control on the measuring process is lost. Individuals in the home may spend hours on the couch watching television (very quietly, as during sleep) or they may forget to put the actigraph on again after a shower. In both cases, the time spent in these conditions would be scored as sleep by an actigraph [195]. In contrast, these measurement errors would be avoided using a PSM, as the presence or absence of a person in the bed is easily noticeable from the pressure changes due to body weight and respiratory movements. PSM does not need human intervention for collecting information. Therefore, compared to the actigraph, human errors have less impact on PSM measurement.

In this chapter, the goal is running a longitudinal design for long-term investigation of intra-subject changes in sleep quality from PSM data. With this scheme, the stability of sleep characteristics is to be determined within a subject. This approach has substantial clinical utility given the known night-to-night variability of sleep apnea [196]–[198]. Furthermore, the relation between the severity of CSA and different sleep measures can be investigated concurrently and over time.

7.2 Method and Material

7.2.1 Database

The database contains long-term nocturnal data described in Section 3.4.2 of Chapter 3 (i.e., non-supervised data). A week of data from each of the 9 participants randomly selected for Chapter 6 was also used here to optimize the individual CSA event detection models.

7.2.3 Pre-processing

The pre-processing approach was explained in detail in section 6.2.3. The pre-processing steps are schematically illustrated in Figure 7.1.

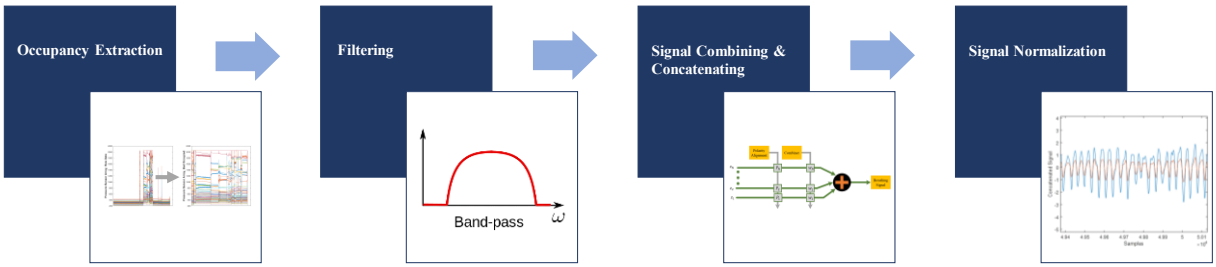


Figure 7.1 Data pre-processing steps.

7.2.4 Feature Extraction

This chapter adopted all the features used in the previous chapter. A comprehensive explanation of these features can be found in section 6.3.4.1.

7.2.5 System Optimization

7.2.5.1 Application of Learning Curve Concerning the Amount of Data

When it comes to building a personalized model, the first important question is how much data is needed to provide an optimized model for long-term monitoring. Assume that a limited amount of data is available, and it is split into a training set and a validation set. One single instance is taken from the training set and used to estimate a model. Then the model's error is measured on the validation set and on that single training instance. The error on the training instance will be zero since it is quite easy to fit a single data point perfectly. The error on the validation set, however, will be very large. That is because the model is built around a single instance, and it almost certainly will not be able to generalize accurately on data that it has not seen before. The

error scores will vary more or less as the training set is changed. The error scores for the validation set (and for the training sets) can be plotted and monitored to investigate the amount of data needed. The resulting curve(s) is called "learning curve(s)", and it should not be confused with other learning curve functions showing the evolution of a cost function over time, as used in machine learning, adaptive signal processing, or optimization. It shows how the error changes as the training set size increases. Here, for each patient, one week of labeled data is available, as shown in Table 7.I.

As a common issue among all patients, the distribution of the two classes Apnea "A" and Normal breathing "N" is imbalanced. For each patient, in each iteration of generating the validation curve, the presence of apneic segments from class "A" in the training dataset is required, as well as balancing the training data. According to Table 7.I, the minimum number of events per day among all patients is 6, and the minimum average time per event is 16 s. 16 segments of apnea events are used as the minimum value M from class "A" to generate the validation curve. Subsequently in each iteration, to overcome the imbalanced data problem, a random selection of segments from class "N", the majority class, has to be performed such that both classes have the same size. The procedure is well known in the literature as the "undersampling" method [179]. As a result of this technique, the minimum training dataset is 32.

The maximum training dataset is two times L, where L is the total number of class "A" instances in the main training data. Using cross-validation, the procedure of generating the learning curve will be repeated k times, where here k is chosen as 7, which is the total number of days of labeled data. It means that in each fold, data of one day will be put aside as a test dataset, and the remaining data of six days will be used to train the classifier. Cross-validation is a very useful

Table 7.I Average time and number of events per day for each patient in a week of data.

Patient No	Average Number of Events per Day (AED)	Average Time for each Event (seconds)
1	6 (±6)	16 (±4)
2	13 (±8)	18 (±6)
3	15 (±6)	17 (±5)
4	59 (±33)	17 (±6)
5	59 (±24)	17 (±5)
6	68 (±15)	17 (±6)
7	82 (±19)	18 (±6)
8	162 (±56)	19 (±5)
9	210 (±40)	18 (±5)

technique for assessing the effectiveness of the model. Moreover, in each fold, to investigate how much the system performance depends on the characteristics of the selected training dataset, the data are randomly shuffled 50 times before creating the learning curve. Once the data is shuffled, no further changes are made in the order of instances, so that one complete learning curve can be generated. Training is performed over iterations $I = 1,2,...(L/M)$. In each iteration, based on the order of the instances in the data matrix, a new unseen patch of data including 16 examples from class "A" and 16 examples from class "N" will be added to the existing training dataset.

Figure 7.2 depicts a summary of the learning curve generation for each patient. SVM training does not depend on the data ordering. Unlike the backpropagation learning algorithm for ANN, a given SVM will always deterministically converge to the same solution for a given data set, regardless of the initial conditions. However, like all other learning models with associated learning

algorithms that analyze data for classification and regression, SVM performance depends greatly on the characteristics of the data to be classified. Here, shuffling the training data allows a different combination of 6 days to be selected and given to SVM. This is more critical since not all "N" instances but just a fraction of them (i.e, same as the number of "A" instances in each iteration) are adopted for training. Figure 7.3 shows the performance of the system on the test dataset for each patient.

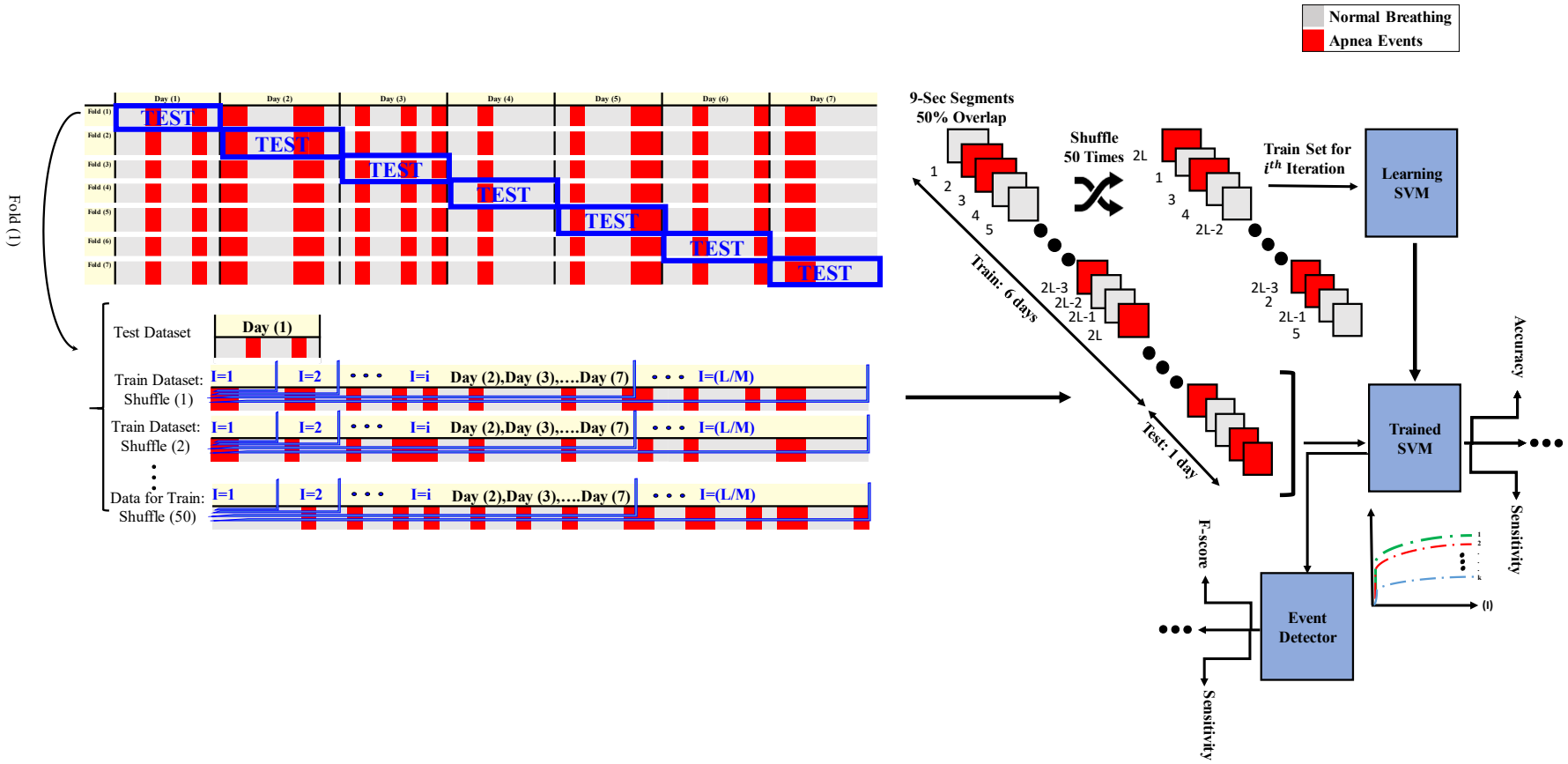


Figure 7.2 Process of generating a learning curve for each patient.

.

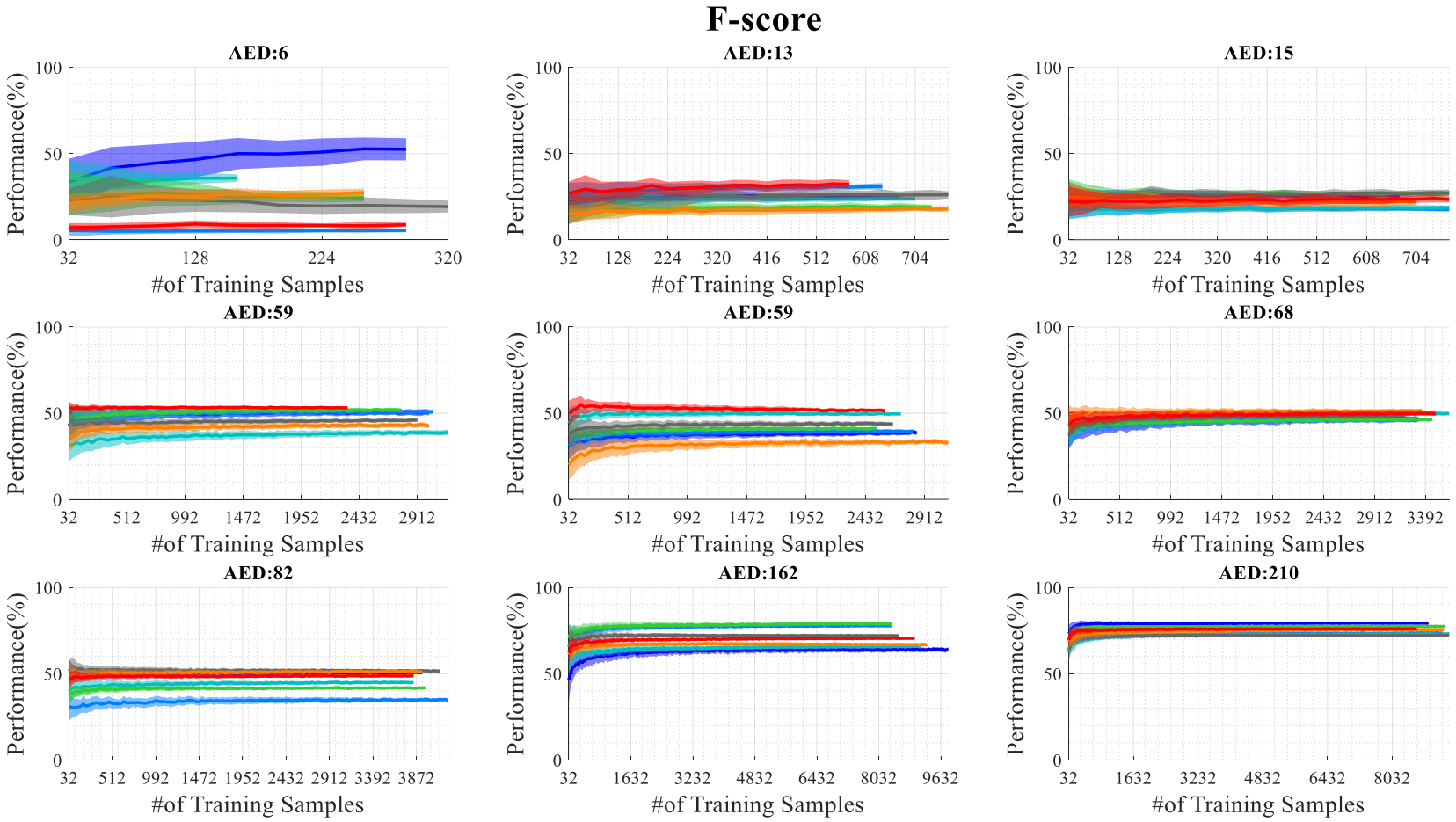


Figure 7.3 The system's performance in terms of F-score on the test dataset for each patient.

The plots are arranged based on the patients' average number of events per day (AED) in ascending order, from left to right and from top to bottom. Each plot contains seven curves, representing the results of different folds with a different day assigned for test data. Each curve denotes the average performance of the system in a 50-time shuffle. The standard deviation of system performance in a 50-time shuffle is calculated to investigate the dependency of the system on the data characteristics, and it is displayed as shade around the average curve: the thinner the shade, the less system dependency there is on the characteristics of selected data. If the system performs properly, apart from individuals' differences and different folds of the dataset, as the training size increases, the average performance of the classifier will increase because the model can generalize better from a higher amount of information. As the curve reaches a saturation point, it means that adding more training data points will not lead to significantly better models. A comparison of the different curves in each plot indicates how the model performs in general when used to make predictions on data not used during the training of the model.

According to Figure 7.3, inter-individual and intra-individual differences show that:

- As the training size increased, the average performance of all systems increased. This is because adding more instances adds diversity and therefore decreases the generalization error of the models.
- As AED increases, the overall performance of the personalized systems increases from an F-score of 24% for the patient with AED of 6 to an F-score of 75% for the patient with AED of 210, which is the largest AED observed. The system built for the patient with the greatest AED had the least dependence on the characteristics of the selected data due to more SA events (and therefore, larger train dataset).

- More importantly, the performance curves reached a saturation point. This means that regardless of different AED a week of labeled data is enough to build a generalized model for long-term monitoring and adding more training data points will not lead to a significantly better system.

Overall, the analysis shows that the more severe the disorder is, the fewer the number of days of data is needed from a patient to build a personalized model.

7.2.5.2 Dealing with Imbalanced Classes

As was discussed in section 6.3.4.2, the CSA detection problem is faced with imbalanced data, where class “N” as a majority class dominates over class “A”, the minority class. The problem causes the model to be more biased towards the majority class and to suffer from the accuracy paradox. There are many approaches to deal with this problem. They can be generally classified into two major categories of 1) sampling-based, and 2) cost function based. Sampling-based methods can be broken into three major categories: a) oversampling b) undersampling c) hybrid of oversampling and undersampling. In the previous chapter, a combination of oversampling of class “A” instances and undersampling of class “N” examples was implemented. The oversampling approach was not by replicating class “A” instances but instead by constructing new minority class data instances via the SMOTE algorithm.

In this chapter, for each patient, a limited amount of data is available, and the diversity of the data is narrowed down to only one patient's information. Given these two issues, sampling-based methods may change the original data distribution. Therefore, here a cost function based method was used to address the class imbalance problem. A cost-sensitive classifier learns more characteristics of the minority class instances by setting a high cost to the misclassification of a

minority class sample. Contrary to the possible disadvantage of sampling-based methods, this method does not modify the data distribution [199].

As explained in section 5.2.7, for linear SVM in binary classification, the objective is to minimize equation (6.1):

$$\frac{\|w\|^2}{2}+(C\sum_{i=1}^{n}\xi_i) \tag{7.1}$$

The algorithm uses slack variables, ξ_i, to penalize the objective function for observations that cross the margin boundary for their class. $\xi_i=0$ for observations that do not cross the margin boundary for their class, otherwise $\xi_i\geq 0$.

w contains the coefficients that define a vector orthogonal to the hyperplane. C and n are the cost (or error penalty) and the number of data instances, respectively. C controls the trade-off between the margin and the misclassification errors. A larger C means that a higher penalty is assigned to misclassification errors

In order to provide different costs associated with the two different kinds of errors, equation (6.1) can be modified to the cost-sensitive SVM (CS-SVM) as follows:

$$\frac{\|w\|^2}{2}+(C^+\sum_{\{i\,|\,y_i=+1\}}^{n_+}\xi_i)+(C^-\sum_{\{j\,|\,y_j=-1\}}^{n_-}\xi_j) \tag{7.2}$$

where C^+ is a cost constant for positive cases and C^- is a cost constant for negative cases. n^+ is the total number of positive cases and n^- is the total number of negative cases [199], [200].

The cost for each patient was optimized based on the average F-score of the event detector, using cross-validation. Six out of seven days of data (six days, six folds) were used for the cross-

validation process. One day of data was put aside as unseen data to evaluate the cost-optimized system. Given the cost structure array as:

$$C = \begin{bmatrix} 0 & a \\ b & 0 \end{bmatrix} \quad (7.3),$$

the cost ratio can then be defined as (b/a).

Figure 7.4 depicts the cross-validation results as a function of the cost ratio. By changing the cost ratio, the F-score changes. The optimized cost ratio is the point on the curve that, on average, produces the best performance in all folds. An overview of the plots for nine patients shows that the optimized cost ratio follows the class imbalance. A lower cost ratio is required when the class imbalance is lower. Table 7.II illustrates the performance of the system on unseen data. The results are arranged based on AED in ascending order.

The rows of the confusion matrix in Table 6.III, section 6.3.6 represented the positive and negative instances. The proportion of the two rows is representative of the class distribution of the data. Any metric such as accuracy that uses values from both rows is highly sensitive to data distributions and imbalances [179]. The cost function-based approach to address imbalanced data does not modify the data distribution. Therefore, although the accuracy of the classifier is presented in Table 7.II, it cannot be a good metric for evaluation.

For more investigation, the results of the optimized system using a cost function based approach were compared with a hybrid of oversampling and undersampling method applied in the previous chapter, i.e., a combination of oversampling of class “A” instances using SMOTE and random undersampling of class “N” instances. Class “N” instances were randomly undersampled by a factor of two. Class “A” instances were oversampled to have the same size as the (undersampled)

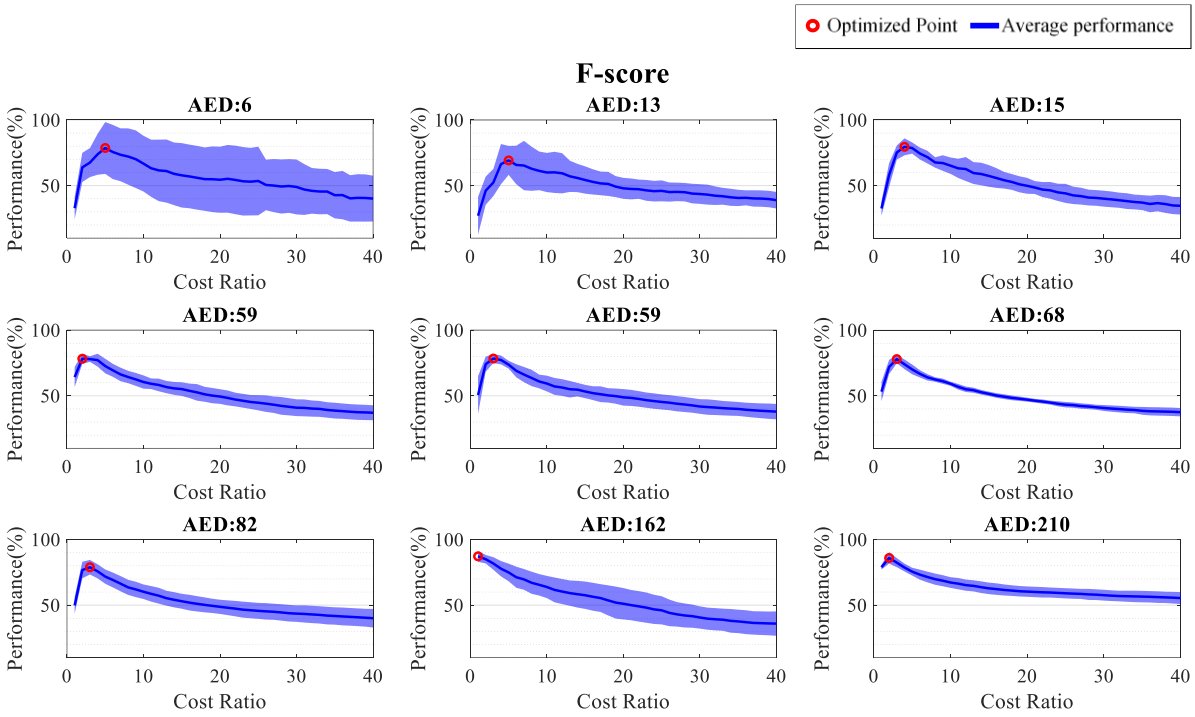


Figure 7.4 The cross-validation results as a function of cost ratio.

instances of the class "N." Additionally, not only random but also two informed undersampling approaches using KNN classifier were considered here [201]. They are:

- NearMiss-1 (NM-1): selecting those class "N" instances whose average distance to the three closest class "A" instances is the smallest.
- NearMiss-2 (NM-2): selecting the class "N" instances whose average distance to the three farthest class "A" instances is the smallest.

After resampling of both classes, the ratio of their size became (1:1). For a fair comparison, these approaches were applied to the same test dataset used to obtain Table 7.II results.

According to Figure 7.5 for all patients, the optimized system had the best performance using the cost function approach. In comparison with other approaches for patients with low AED, the superiority of this approach is more pronounced. Among all undersampling methods combined

with SMOTE, random undersampling and informed undersampling, NM-2, had the best and the worst performance, respectively.

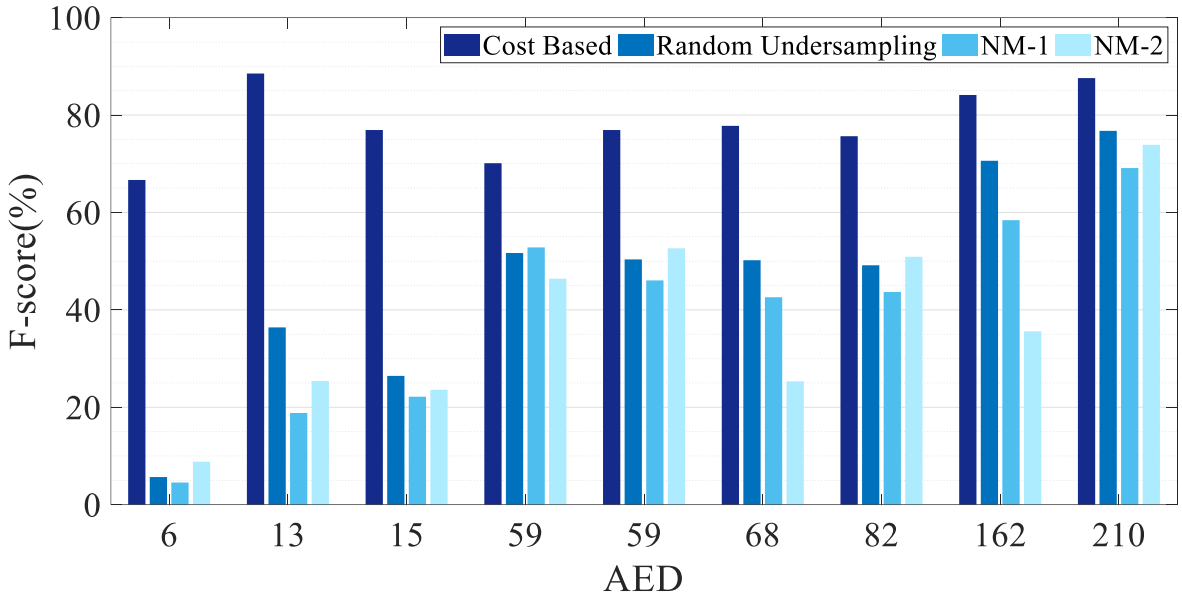


Figure 7.5 System performances using different approaches for addressing imbalanced data: "Cost Function Based' and "Hybrid of SMOTE and Undersampling".

7.2.6 Descriptive Statistics of the Sleep Measures

For each patient, after optimizing the cost matrix, the final classifier was trained on all data from 7 days. Then the data of the whole year was fed to the system day by day to extract sleep-related information. In this section, the main findings of the longitudinal study are described and broken down into the specific sleep measures and their association with habitual night sleep.

7.2.6.1 Central Apnea Index (CAI)

Bed occupancy (BO) was defined as the total time spent in bed during the 24-hour cycle. For every 24-hour cycle from 12 pm to 12 pm the next day, BO was extracted by an occupancy extraction algorithm, as a part of the pre-processing procedure. This algorithm was described in

Table 7.II Performance of optimized cost sensitive model on unseen data.

Patient No	AED	Classifier Performance (%)			Event Detector Performance (%)		
		Specificity	Sensitivity	Accuracy	F-score	Sensitivity	Precision
1	6 (±6)	99.6	55.6	99.5	66.7	100	50
2	13 (±8)	99.5	56.3	98.7	88.5	100	79.4
3	15 (±6)	99.7	51.7	99.2	76.9	83.3	71.4
4	59 (±33)	95.2	48.3	89.3	70.1	73.1	67.4
5	59 (±24)	98.8	46.3	96.5	76.9	75.9	77.9
6	68 (±15)	98.3	51.4	96.5	77.8	81.7	74.2
7	82 (±19)	98.2	52.0	96.4	75.7	84.5	68.5
8	162 (±56)	98.7	52.5	93.8	84.1	80.4	88.2
9	210 (±40)	94.4	59.1	87.5	87.6	96	80.6

detail in section 5.2.3. Bed-occupied time included nighttime sleep and daily naps. CAI was defined by the number of CSA events detected per hour of BO, as shown in equation 6.4. CAI can indicate the severity of the SA, similarly to the AHI defined in chapter two.

$$CAI = \frac{\#CSA_{Events}}{BO} \tag{7.4}.$$

As an example, Figure 7.6 shows the CAI for one of the subjects for his entire data. The light-blue part of the figure represents the week that was selected randomly for labeling.

7.2.6.2 CSA duration per Night

As the output of the system, in addition to the number of extracted events and their temporal position for each night, the duration of each event was investigated. The histogram in Figure 7.7 groups total extracted CSA events into different time intervals from the 212 days of data of the subject in Figure 7.6. The height of each bar shows how many events fall into each range.

According to the bar chart, for this specific patient, most events have a duration of 13 to 14 s. The CAI per night and duration of detected events for the other eight patients are presented in Appendix A.

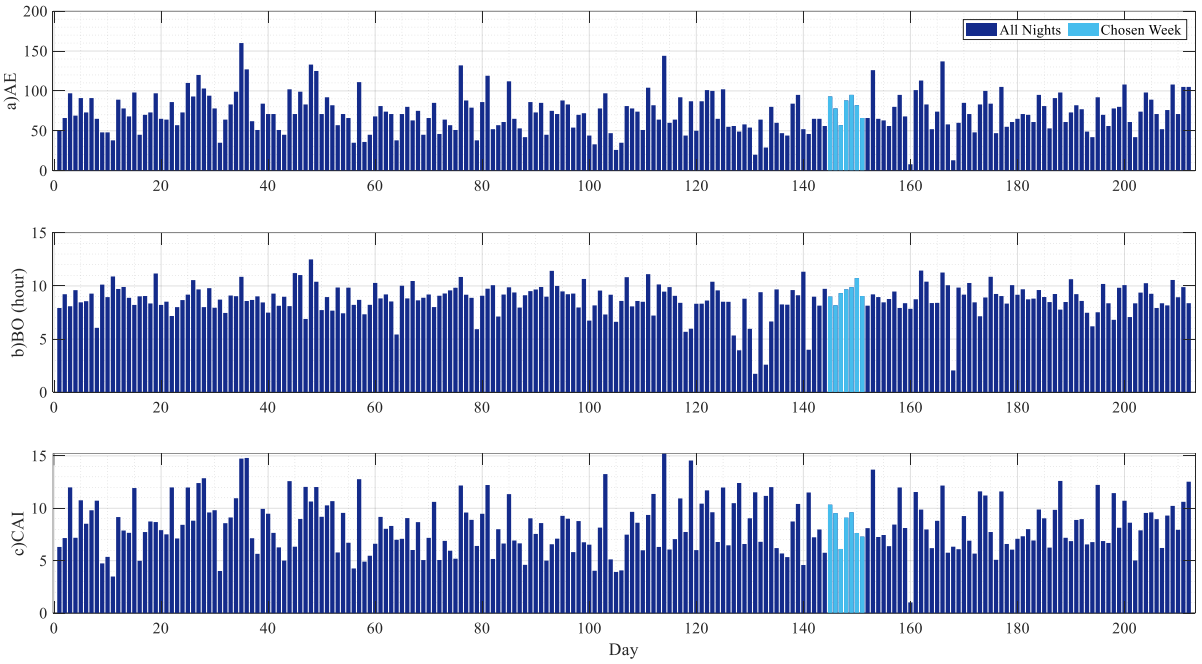


Figure 7.6 CAI per night for one of the patients.

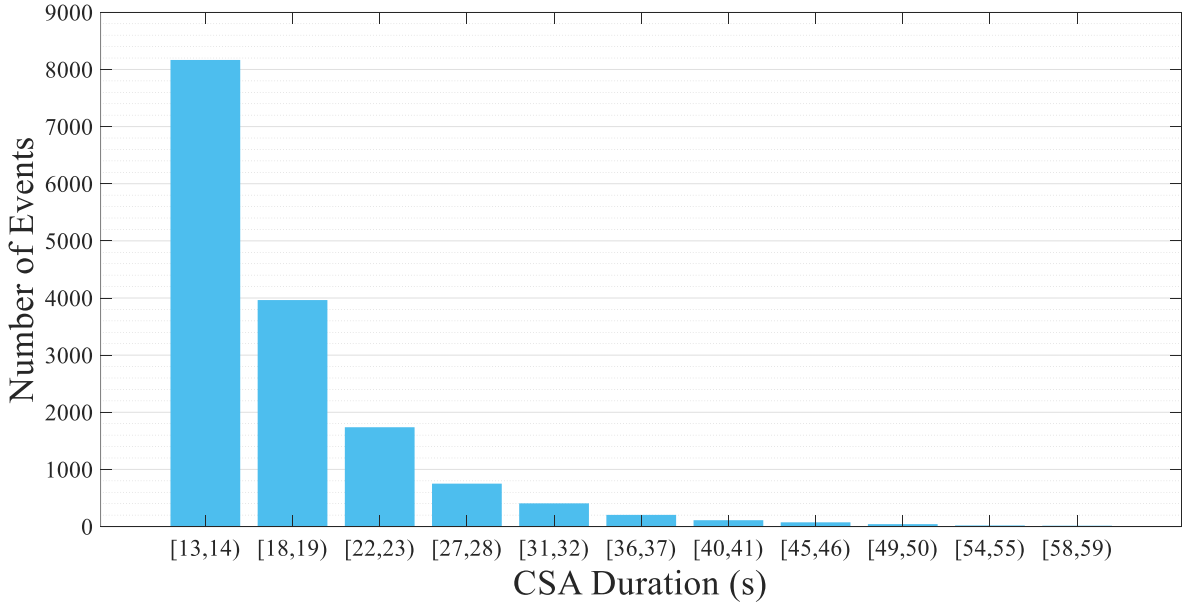


Figure 7.7 Histogram of the CSA events' duration (in seconds) for the same subject of Figure 7.6.

7.2.6.3 Sleep Regularity Index (SRI)

For the intra-individual investigation, SRI calculates the percentage probability of an individual being in the same state (asleep vs. awake) at any two time-points 24 h apart, averaged across the study [202]. The index was scaled so that a patient who slept and woke up at the same time each day scored 100. In contrast, a patient who slept and woke up at random scored zero. For this study, it was assumed that the presence of a patient in the bed was equivalent to being asleep. SRI was originally developed for a minute-by-minute computation. Here one-hour epochs were used, but otherwise, the same description was followed. Given N days of recordings divided into 24 hours, let $s_{i,j}=1$ if the patient was sleeping on the day i in epoch j, and $s_{i,j}=0$ if he was awake. The SRI was calculated as:

$$SRI=-100+\frac{200}{M(N-1)}\sum_{j=1}^{M}\sum_{i=1}^{N-1}\delta(s_{i,j},s_{i+1,j}) \tag{7.5}$$

where $\delta(s_{i,j},s_{i+1,j})=1$ if $s_{i,j}=s_{i+1,j}$ and zero otherwise [203].

7.2.6.4 Clock Night and Clock Day Sleep

To determine whether daily and nightly sleep have any effect on the sleep characteristics of an individual, BO was calculated separately for the day clock (defined as 10 am to 10 pm) and the night clock (defined as 10 pm to 10 am).

7.2.7 Associations between Sleep Variables

As an application of longitudinal study, in addition to calculating sleep measures for each patient, Pearson correlation can be used to quantify associations between these calculated indices. Pearson correlation indicates a direct, mathematical relationship between sleep measures to study inter-individual differences.

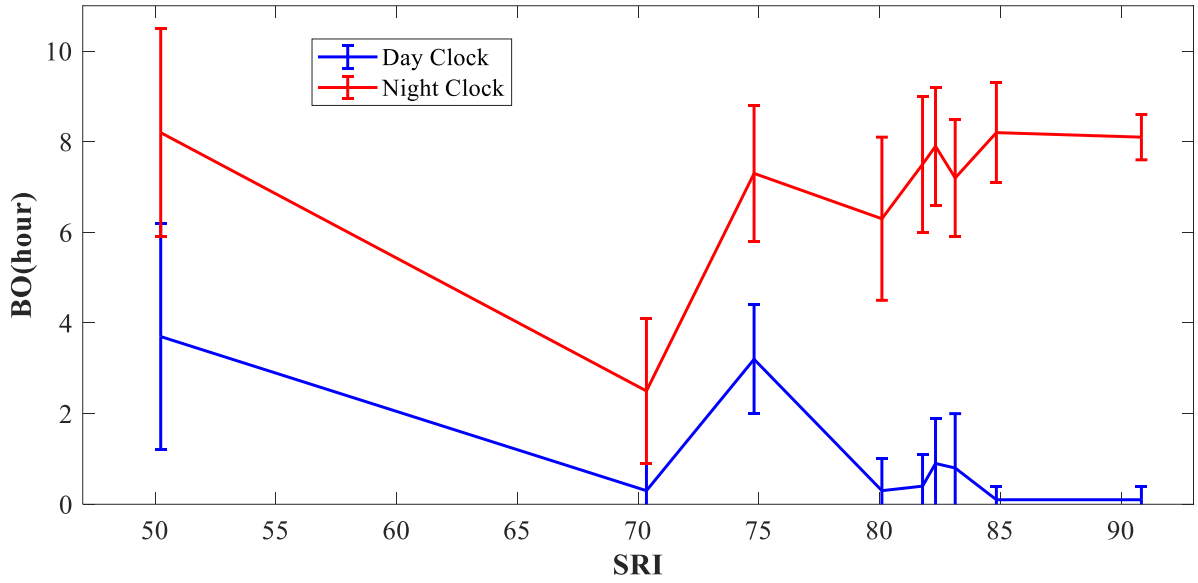


Figure 7.8 BO as a function of SRI.

7.2.7.1 SRI and BO

Individuals' SRI ranged from 50.2 to 90.8. Daily and nightly average sleep duration ranged from (0.1-3.7) and (2.5-8.2) hours, respectively. Figure 6.8 shows that a patient with a larger SRI (meaning a more regular sleeper) obtained more sleep during the night clock and significantly less sleep during the day clock than irregular sleepers with smaller SRI. The results of Pearson correlation analysis between SRI and BO clarified that the sleep irregularity was negatively correlated with greater day-time sleep ($r_c = -0.77, p_v < 0.02$). However, SRI was nearly independent of night-time sleep ($r_c = 0.16, p_v = 0.67$).

7.2.7.2 CAI and BO

There was no association found between CAI and BO, neither in day-time sleep ($r_c = -0.04, p_v = 0.91$) nor night-time sleep ($r_c = 0.4, p_v = 0.28$).

7.2.8 Breathless Events

Clinically, a SA event must be at least 10 s, normally not exceeding 2 minutes [17]. However, the duration or definition of an SA event can be debatable. Perhaps the effect that several events with duration less than ten can have on a patient's health may be equal or even worse than a few clinically accepted SA events. Additionally, because of this restricted definition, in the approach used to extract SA events, many events can be ignored or not detected. Therefore, it can be helpful to have a system for the detection of events with less than 10 s. The approach proposed in this thesis has proved to be significantly capable of detecting SA events. It is also flexible in detecting events with a duration of less than 10 s. For this purpose, the combined signal of PSM was segmented by a window of 2 s with a 1-second overlap instead of 9 s with 50% overlap as done previously. The personalized detection systems were then trained on the information of 2-second segments and applied to all the recordings of each patient. It should be noted that the performance

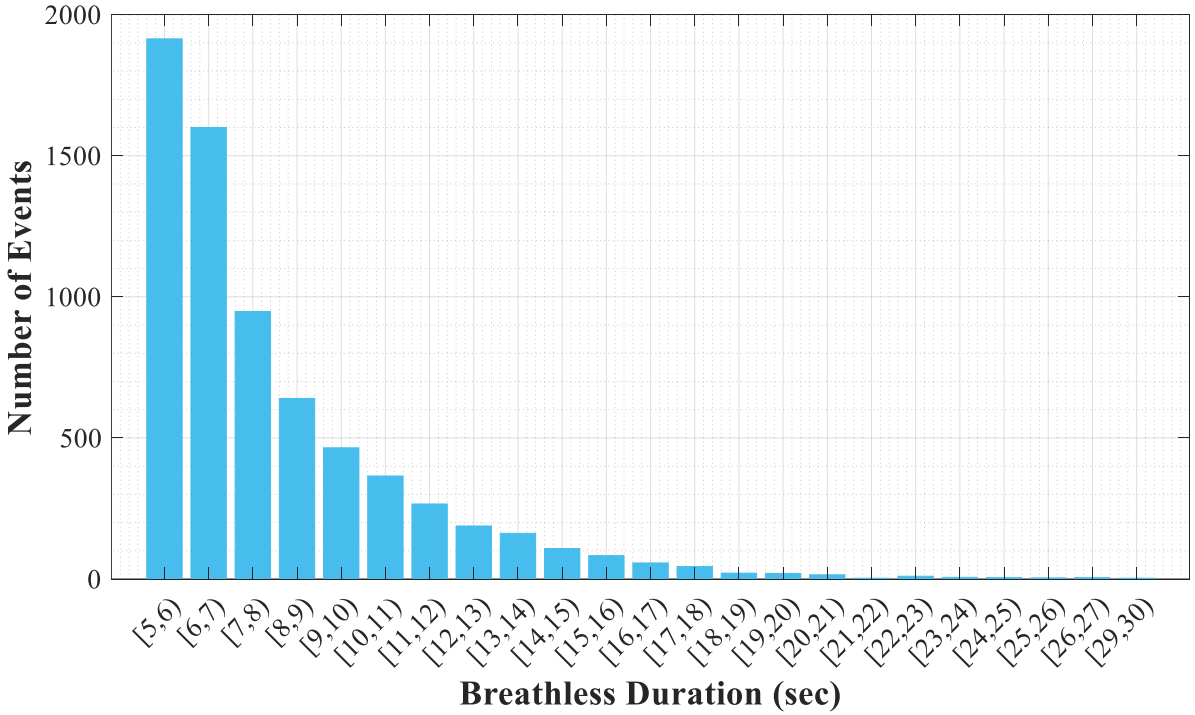


Figure 7.9 Breathless events with different durations.

of the system for segments of 2 s with 1-second overlap was evaluated for CSA events with a duration of 10 s or more. The system can be evaluated more accurately if the events with less than 10 s are manually and visually labeled. Figure 7.9 illustrates the histogram of all the extracted events, with different duration. In 212 occupied days, there were even more events with a duration of [9 10) than SA events with a duration of [10 11) s.

7.3 Summary

In this chapter, a personalized system was built for longitudinal monitoring of sleep characteristics. Table 7.III provides a summary of the sleep characteristics of each patient extracted from longitudinal data. One of the important points to be considered for designing the personalized systems was the amount of data required. It was shown that seven days of data would be enough for all patients with different SA severity. In acute conditions, i.e., for a patient with a great number of CSA events per day, even one day of data could be enough to optimize the model. After extracting the relevant features, the next challenge was to deal with imbalanced data. Different approaches, including cost function based and resampling approaches, were examined to combat imbalanced data. The results showed that the cost function-based method outperformed other methods. Next, a model was trained by using the data from seven consecutive days, and the trained cost-optimized system was then applied to the whole recorded data on a day-by-day basis. The output of the system was the total number of CSA events per day, along with their temporal position in each day and their duration. The CAI was defined as a measure of CSA severity. The application of the CAI is that after an initial diagnosis, it can be monitored in the long-term to track a patient's treatment progress. The proposed approach was also optimized to detect events with less than 10 s, which can provide useful new clinical information. Moreover, in addition to CAI, other sleep characteristics such as BO and SRI were extracted through PSM data. Irregular sleep

Table 7.III Summary of sleep characteristics from longitudinal study for each patient.

Patient No	No of Occupied Days	BO (hour)		No of Detected CSA Events	Average CAI	SRI
		Clock Day	Clock Night			
1	211 out of 239	0.3 (±0.6)	2.5 (±1.6)	7 (±6)	2.6	70.3
2	150 out of 152	0.1 (±0.3)	8.1 (±0.5)	15 (±7)	1.9	90.8
3	296 out of 306	0.4 (±0.7)	7.5 (±1.5)	70 (±44)	8.9	81.8
4	200 out of 204	0.8 (±1.2)	7.2 (±1.3)	57 (±22)	7.3	83.1
5	122 out of 126	3.2 (±1.2)	7.3 (±1.5)	84 (±51)	8	74.8
6	212 out of 295	0.9 (±1)	7.9 (±1.3)	73 (±24)	8.4	82.3
7	100 out of 110	3.7 (±2.5)	8.2 (±2.3)	127 (±59)	10.7	50.2
8	393 out of 397	0.3 (±0.7)	6.3 (±1.8)	80 (±63)	12.1	80.1
9	99 out of 101	0.1 (±0.3)	8.2 (±1.1)	205 (±60)	24.6	84.8

patterns are common in individuals of all ages. Greater sleep irregularity is associated with a higher 10-year risk of heart disease, as well as greater obesity, hypertension, fasting glucose, and diabetes. The analysis in this chapter illustrated that patients with low SRI (high sleep irregularity) tended to sleep more during the day ($r_c = -0.77, p_v < 0.02$), which is consistent with the results of other studies [202], [203]. There was no association between CAI and BO, neither in day-time sleep ($r_c = -0.04, p_v = 0.91$) nor night-time sleep ($r_c = 0.4, p_v = 0.28$).

Chapter 8: Conclusion & Future Work

8.1 Conclusion & Contributions

Millions of Americans suffer from a disorder of sleep and wakefulness [204]. Sleep apnea is a serious sleep disorder that occurs when a person's breathing is interrupted during sleep. It was estimated that 24% of men and 9% of women have some degree of sleep apnea [205]. Nowadays, these figures are probably higher than those based on these older studies, since the obesity rate was not as large as today. Failure to treat sleep apnea can lead to poor health and other chronic diseases, such as cardiovascular disease [206]. The current clinical standard for diagnosing and monitoring sleep apnea, PSG, requires patients to sleep in a foreign environment while wired up to several uncomfortable sensors. The obtrusive nature and sometimes limited access to PSG [207] have led to a keen interest in unobtrusive methods to monitor an individual's sleep. The value in unobtrusive sensors for sleep and apnea monitoring is that they require no user interaction and can collect long-term health information without modifying the subject's behavior in a variety of home-based contexts. The PSMs used in this study lay underneath a mattress to detect movement and respiration. Previously, the ability of PSM to detect breathing was proven in [136], [148]. In this thesis, building upon the earlier related work [138], [208], two simple signal combining methods including SNR-MAX and Correlation-based Analysis (PCC) were first developed to enhance the ability of PSM for respiratory rate detection. An experiment was designed to examine how well the combined signals could detect the breathing rate of an individual lying on the bed.

In comparison with the breathing rate derived from a respiratory band, both signal combining methods showed good results to extract respiratory signal and consequently respiration rate, outperforming some previous combining methods. The results were published in [6]. Later in [7], the performance of the proposed methods was evaluated for close-to-reality data, where the participants were allowed to move around on the bed and have different types of breathing, such as shallow breathing, deep breathing, and periods of apnea. Compared to other data fusion methods, the SNR-MAX was found in [7] to be the best method of sensor signal combining, resulting in the highest Pearson Correlation Coefficient with the respiratory band signal.

Large movements by an individual lying in bed (i.e., rolling over or moving an arm or leg) result in large excursions of the pressure sensor signals, which are orders of magnitude greater than the PSM breathing signal. These large movements completely "washout" the ability of the PSM system to detect breathing during the movement. In [16], a method was developed to detect large movements. It made use of a previously developed method for movement detection based on control levels [209]. The novel contribution of the paper was to use an adaptive window length to calculate a moving average and a moving variance, by measuring the distance between two consecutive peaks in the signal related to consecutive movements. The proposed method was applicable for different postures and breathing patterns of the bed occupant. The results showed that the proposed scheme provided significant improvements compared to a previous method. It led to an average movement detection offset as low as 1.32 s, with no false-positive events and low false-negatives.

PSM data obtained from periods of quiescence can be analyzed for the presence of apneic events. Previously, a small-scale validation of the PSM compared to traditional PSG was presented for use as a CSA screening tool [156]. The results revealed that PSM has a great capacity to be used

to detect CSA. In this thesis, advanced signal processing and pattern recognition techniques have been applied for the detection of CSA using PSM, to improve on existing strategies such as in [137]. To detect CSA two approaches were considered: one for longitudinal monitoring and one for a sleep lab study. Both approaches commonly include a sequence of steps as follows. First, to decrease the computational complexity and mostly to discard irrelevant data, periods during which the bed was not occupied were detected and removed. Then, SNR-MAX signal combining was used on the pressure sensor signals. After detecting movements, depending on the available data different approaches were studied, compared, and optimized to detect CSA events. Finally, the event detection was not solely based on the decision of the classifiers under the study. Instead, a heuristically obtained rule-based algorithm was used to identify the CSA events from the time-sequenced decisions of the classifiers.

8.1.1 Sleep Lab Study

To design a generalized and well-suited model for the application of sleep lab study where only one night of data is available from patients to assess in terms of CSA detection, 4 different machine learning approaches were implemented and compared. These approaches ranged from a simple threshold-based algorithm (developed in Chapter 5) to SVM as one of the most popular traditional machine learning approaches and finally to deep learning methods including TCN and BiLSTM. In the implementation of each of these methods, the necessary steps were considered to optimize them, so that it was still possible to make a reasonable comparison between all methods. A threshold was tuned based on the training data for the simple threshold-based algorithm. The training data for SVM were balanced by the undersampling of the majority class and oversampling of the minority class. The proportion of resampling the data from each class was investigated through 5-fold cross-validation on training data. A feature selection technique based on SVM-RFE

was adopted to find the optimal set of features. Optimization of deep learning approaches including TCN and BiLSTM was more complicated and time-consuming since there where more hyperparameters to consider for tuning. The learning rate and the number of epochs are examples of these hyperparameters to be optimized. Overall, the BiLSTM method outperformed all other methods with an F-score of 85% and had a better performance on the common test data adopted for all methods.

8.1.2 Longitudinal Monitoring

The main intention of this research was to monitor elderly patients using PSM systems deployed into their homes. In order to reach a preliminary diagnosis and eliminate the first night effect (FNE), individuals suspected of having sleep apnea can have a PSM outfitted in their bed before scheduling an overnight PSG at the sleep lab. Moreover, those patients already diagnosed with sleep apnea can be monitored over time while they sleep in their unaltered sleeping environment, to track their treatment progress. However, due to the sheer amount of data, it would be unreasonable to assume that a sleep technologist can visually score the collected PSM data each night. To overcome this matter, in this thesis, seven consecutive days (one week) of each patient's data were visually labeled. The first step taken was to investigate whether the amount of data available for each patient was enough to design an optimized personalized system, or not. In this regard, a 7-fold cross-validation procedure (7 days of data, each day as a separate fold) was applied to each patient's week of data. For each fold, the training dataset was shuffled 50 times to eliminate the sensitivity of the procedure to the order of the data. Finally, the learning curves as a result of each fold in terms of average F-score of 50 processes were provided. The analysis showed that a week of data was enough to build a personalized system for longitudinal monitoring of all patients with different SA severity. After extracting the relevant features, the next challenge was to deal

with imbalanced data. Given the limited amount of labeled data for each patient, different approaches, including cost function based and resampling approaches, were examined to cope with imbalanced data and leave the original data distribution unchanged. A comparison among applied methods revealed that the cost function-based method outperformed other methods. The output of the optimized, personalized system was the total number of CSA events per day, along with their temporal position in each day and their duration. The CAI was defined as a measure of CSA severity. Extracting CAI as a parameter allowed to monitor the severity of the disorder through a year of data for each patient. For the purpose of home monitoring, CAI is valuable because the progression of treatment can be tracked while the disorder is being treated.

Moreover, CAI, BO, and SRI were then extracted for each patient as sleep measures from longitudinal data. The inter-individual analysis between patients showed that irregular sleepers slept during the day clock more than regular sleepers. There was no correlation between CAI and BO, neither during the day nor during the night.

Overall, the results of the thesis suggested that PSM could be beneficially used in both the sleep lab and in patients' homes for longitudinal monitoring purposes. However, unsupervised collecting and analyzing PSM data presents some unique challenges. Since systems were collecting data in a real-life context, there were more noise and uncertainty factors to affect the authenticity and veracity of data than there would be in a laboratory setting. For instance, home environments can be varied when participants move their bedroom furniture or unplug the technology system by mistake. These factors present extra challenges to collecting data in the home environment compared to the laboratory environment. Perhaps a sleep diary for recording an individual's sleeping and waking times with related information could be helpful to reduce these uncertainty factors. While lab controlled data may produce higher accuracy in some ways, data collected in

the home environment reflects the long-term realities of daily life that must be considered in order to deliver better diagnostics and treatments.

8.2 Future Work

Sleep monitoring is a growing and very interesting area of research. There are many possible paths for future work. Some of these are listed below.

8.2.1 Sleep Position Detection

Many patients who suffer from sleep apnea experience worse symptoms when they sleep on their back [210]–[212]. Developing a method for distinguishing sleeping positions would likely enhance the ability of PSM systems to detect the presence of sleep apnea in individuals, especially for home monitoring purposes.

One of the challenges presented by in-home PSM data collection will be accounting for multiple individuals sleeping on the bed/mattress system. Subjects can be identified by image processing of the area occupied on PSM and other individual's characteristics such as weight and height.

8.2.2 Weight Monitoring

In patients with heart failure, fluid retention can weaken their functional capacity. Measuring short-term changes in body weight can be the most reliable and straightforward indication of short-term changes in fluid status [213]. In long-term care facilities, PSM could monitor individuals' weight on a nightly basis in order to alert caretakers/doctors if sudden weight changes occur.

Moreover, a longitudinal study in [214] has shown that there is a relationship between weight gain and increased sleep-disordered breathing (SDB) severity. Not only weight gain predicted the development of moderate-to-severe SDB in those individuals who initially had mild or no SDB, but also weight loss was associated with a reduced SDB severity and the possibility of

developing SDB. These findings emphasize the importance of preventing weight gain in normal-weight persons to avoid the development or progression of SDB. For this purpose, simultaneously with the SDB treatment monitoring, the individuals' weight can be monitored.

8.2.3 Sleep Stage Determination

During sleep, people continuously switch between two states, known as rapid eye movement (REM) and non-rapid eye movement (NREM). These states alternate throughout the sleep bout, producing cycles of 90–110 minutes in duration. The study of the alternations among wake-NREM-REM stages provides useful information related to sleep quality since sleep disorders profoundly affect this alternating pattern [215]. Since each sleep cycle is also associated with other physiological oscillations (e.g. respiratory and cardiovascular physiology, body movements, etc.), tracking these factors as indirect markers through sleep, the timing and duration of sleep bouts can be obtained with reasonable accuracy [216].

In [241], a PSM-based rollover detection method has been used. Rollovers do not occur during REM sleep and may be useful for sleep staging. As determined by the algorithm of wrist-worn actigraphy, rollover frequency was used in [167] to infer deep vs. light sleep. Many commercially available actigraphs (i.e. fitness trackers) yield standard sleep-related metrics such as total sleep time, a distinction between sleep and wake, an "arousal index", and sleep latency based on episodes of movement during presumed sleep time [137].

In addition to rollovers and movements, the respiration rate can be an indicator for estimating sleep phases [219]. NREM is the most comforting and most restful period of deep sleep. During NREM sleep, blood pressure, respiratory rate, and basal metabolic rate (BMR) decrease by 10–30% [190], [220]. Interestingly, CSA occurs most commonly during NREM sleep. This may be attributed to increased ventilatory motor output during REM sleep relative to NREM sleep.

Additionally, CSA may accompany the transition from wakefulness to NREM sleep. Sleep onset, determined as a transition from alpha waves (neural oscillations occurring during wakeful relaxation with closed eyes) to theta waves (neural oscillations occurring most often in sleep), is associated with an increase in breath duration which may appear as central apnea in some individuals [154]. Here, in comparison with actigraphy, incorporating detected movements and rollovers with respiration-related activity and/or the interval length between postural changes may significantly improve movement-based sleep staging classification, as suggested in [221]. All types of information are present in the PSM signal in a more unobtrusive manner.

8.2.4 Other Sleep-related Disorders

The PSM's ability to detect limb movement and motion is unequivocal. By extracting PSM signals for the specified region of interest, PSM could also be applied to detect and monitor other sleep disorders such as periodic limb movement disorder (PLMD) and restless leg syndrome. Currently, in order to diagnose their disorder, patients need to do an overnight sleep study, i.e., PSG.

8.2.5 Compliance with Treatment

Results obtained in this study shows that PSM could solely be used to detect CSA and monitor breathing rate in the sleep lab and individuals' homes. A real-time adaptive methodology of the proposed method could be extended to point-of-care applications. For instance, for those sleep apnea patients who are being treated, the CPAP device could be adjusted adaptively from the output of the classification model, using PSM data.

References

[1] "Clinical Guidelines for the Use of Unattended Portable Monitors in the Diagnosis of Obstructive Sleep Apnea in Adult Patients," *J. Clin. Sleep Med. JCSM Off. Publ. Am. Acad. Sleep Med.*, vol. 3, no. 7, pp. 737–747, Dec. 2007.

[2] C. A. Kushida *et al.*, "Practice parameters for the indications for polysomnography and related procedures: an update for 2005," *Sleep*, vol. 28, no. 4, pp. 499–521, Apr. 2005.

[3] O. Le Bon *et al.*, "The first-night effect may last more than one night," *J. Psychiatr. Res.*, vol. 35, no. 3, pp. 165–172, Jun. 2001.

[4] C. Brüser, C. H. Antink, T. Wartzek, M. Walter, and S. Leonhardt, "Ambient and Unobtrusive Cardiorespiratory Monitoring Techniques," *IEEE Rev. Biomed. Eng.*, vol. 8, pp. 30–43, 2015, doi: 10.1109/RBME.2015.2414661.

[5] H. Azimi, M. Bouchard, R. Goubran, and F. Knoefel, "Smart Bed Technology For Breathing Monitoring," presented at the Medical Devices Innovation Institute (MDII) and NSERC CREATE-BEST Poster Day, Ottawa, Sep. 2017.

[6] H. Azimi, S. S. Gilakjani, M. Bouchard, S. Bennett, R. A. Goubran, and F. Knoefel, "Breathing signal combining for respiration rate estimation in smart beds," in *2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, May 2017, pp. 303–307, doi: 10.1109/MeMeA.2017.7985893.

[7] S. S. Gilakjani, H. Azimi, M. Bouchard, R. A. Goubran, and F. Knoefel, "Improved sensor selection method during movement for breathing rate estimation with unobtrusive pressure sensor arrays," in *2018 IEEE Sensors Applications Symposium (SAS)*, Mar. 2018, pp. 1–6, doi: 10.1109/SAS.2018.8336769.

[8] H. Azimi, S. S. Gilakjani, M. Bouchard, R. A. Goubran, and F. Knoefel, "Automatic apnea-hypopnea events detection using an alternative sensor," in *2018 IEEE Sensors Applications Symposium (SAS)*, Mar. 2018, pp. 1–5, doi: 10.1109/SAS.2018.8336762.

[9] H. Azimi, M. Bouchard, R. A. Goubran, and F. Knoefel, "Apnea Event Detection Methodology using Pressure Sensors," presented at the 2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Jun. 2019.

[10] H. Azimi, M. Bouchard, R. Goubran, and F. Knoefel, "Smart Home And Sleep Apnea: Application To Eldercare," presented at the Engineering and Computer Science Graduate Poster Competition, Ottawa, Mar. 2019.

[11] H. Azimi, P. Xi, M. Bouchard, R. Goubran, and F. Knoefel, "Machine Learning-Based Automatic Detection of Central Sleep Apnea Events From a Pressure Sensitive Mat," *IEEE Access*, vol. 8, pp. 173428–173439, 2020, doi: 10.1109/ACCESS.2020.3025808.

[12] H. Azimi, M. Bouchard, R. Goubran, and F. Knoefel, "Unobtrusive Screening of Central Sleep Apnea From Pressure Sensors Measurements: A Patient-Specific Longitudinal Study," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 3282–3296, Jun. 2020.

[13] F. Knoefel, H. Azimi, M. Bouchard, and R. Goubran, "Detecting Apnea in Community Dwelling Older Adults," presented at the International Society of Gerontechnology (ISG), Trondheim, Norway, Oct. 2020.

[14] H. Azimi *et al.*, "Real Time Data Analytics using IBM Cloud for Sleep Apnea," presented at the AGE-WELL's 5th Annual Conference, Moncton, NB, Canada, Oct. 22, 2019.

[15] H. Azimi *et al.*, "Cloud Processing of Bed Pressure Sensor Data to Detect Sleep Apnea Events," in *2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, Bari, Italy, 2020, p. 5.

[16] S. S. Gilakjani, S. Bennett, R. A. Goubran, H. Azimi, M. Bouchard, and F. Knoefel, "Movement detection with adaptive window length for unobtrusive bed-based pressure-sensor array," in *2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, May 2017, pp. 355–360, doi: 10.1109/MeMeA.2017.7985902.

[17] R. B. Berry *et al.*, "Rules for scoring respiratory events in sleep: update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. Deliberations of the Sleep Apnea Definitions Task Force of the American Academy of Sleep Medicine," *J. Clin. Sleep Med. JCSM Off. Publ. Am. Acad. Sleep Med.*, vol. 8, no. 5, pp. 597–619, Oct. 2012, doi: 10.5664/jcsm.2172.

[18] Sleep Heart Health Study SRC, "Sleep Heart Health Study. Reading center manual of operations." Tech. Rep. VMLA-039- 02, Case Western Reserve University, 2002.

[19] W. A. Whitelaw and K. R. Burgess, "Diagnosis of sleep apnoea: some critical issues," *INDIAN J MED RES*, p. 13, 2010.

[20] B. A. Staats, H. W. Bonekat, C. D. Harris, and K. P. Offord, "Chest wall motion in sleep apnea," *Am. Rev. Respir. Dis.*, vol. 130, no. 1, pp. 59–63, Jul. 1984, doi: 10.1164/arrd.1984.130.1.59.

[21] J. Hammer and C. J. L. Newth, "Assessment of thoraco-abdominal asynchrony," *Paediatr. Respir. Rev.*, vol. 10, no. 2, pp. 75–80, Jun. 2009, doi: 10.1016/j.prrv.2009.02.004.

[22] D. Kogan, A. Jain, S. Kimbro, G. Gutierrez, and V. Jain, "Respiratory Inductance Plethysmography Improved Diagnostic Sensitivity and Specificity of Obstructive Sleep Apnea," *Respir. Care*, vol. 61, no. 8, pp. 1033–1037, Aug. 2016, doi: 10.4187/respcare.04436.

[23] N. Sezgin and M. Emin Tagluk, “Energy based feature extraction for classification of sleep apnea syndrome,” *Comput. Biol. Med.*, vol. 39, no. 11, pp. 1043–1050, Nov. 2009, doi: 10.1016/j.compbiomed.2009.08.005.

[24] “Clinical Guideline for the Evaluation, Management and Long-term Care of Obstructive Sleep Apnea in Adults,” *J. Clin. Sleep Med. JCSM Off. Publ. Am. Acad. Sleep Med.*, vol. 5, no. 3, pp. 263–276, Jun. 2009.

[25] A. S. Gami, S. M. Caples, and V. K. Somers, “Obesity and obstructive sleep apnea,” *Endocrinol. Metab. Clin. North Am.*, vol. 32, no. 4, pp. 869–894, Dec. 2003.

[26] A. R. Schwartz, S. P. Patil, A. M. Laffan, V. Polotsky, H. Schneider, and P. L. Smith, “Obesity and Obstructive Sleep Apnea,” *Proc. Am. Thorac. Soc.*, vol. 5, no. 2, pp. 185–192, Feb. 2008, doi: 10.1513/pats.200708-137MG.

[27] A. S. Gami *et al.*, “Obstructive sleep apnea, obesity, and the risk of incident atrial fibrillation,” *J. Am. Coll. Cardiol.*, vol. 49, no. 5, pp. 565–571, Feb. 2007, doi: 10.1016/j.jacc.2006.08.060.

[28] D. D. Sin, F. Fitzgerald, J. D. Parker, G. Newton, J. S. Floras, and T. D. Bradley, “Risk factors for central and obstructive sleep apnea in 450 men and women with congestive heart failure,” *Am. J. Respir. Crit. Care Med.*, vol. 160, no. 4, pp. 1101–1106, Oct. 1999, doi: 10.1164/ajrccm.160.4.9903020.

[29] R. J. Schwab *et al.*, “Identification of upper airway anatomic risk factors for obstructive sleep apnea with volumetric magnetic resonance imaging,” *Am. J. Respir. Crit. Care Med.*, vol. 168, no. 5, pp. 522–530, Sep. 2003, doi: 10.1164/rccm.200208-866OC.

[30] T. Shintani, K. Asakura, and A. Kataura, “Adenotonsillar hypertrophy and skeletal morphology of children with obstructive sleep apnea syndrome,” *Acta Oto-Laryngol. Suppl.*, vol. 523, pp. 222–224, 1996.

[31] E. P. Hamans, E. A. Van Marck, W. A. De Backer, W. Creten, and P. H. Van de Heyning, “Morphometric analysis of the uvula in patients with sleep-related breathing disorders,” *Eur. Arch. Oto-Rhino-Laryngol. Off. J. Eur. Fed. Oto-Rhino-Laryngol. Soc. EUFOS Affil. Ger. Soc. Oto-Rhino-Laryngol. - Head Neck Surg.*, vol. 257, no. 4, pp. 232–236, 2000.

[32] K. L. Do, H. Ferreyra, J. F. Healy, and T. M. Davidson, “Does tongue size differ between patients with and without sleep-disordered breathing?,” *The Laryngoscope*, vol. 110, no. 9, pp. 1552–1555, Sep. 2000, doi: 10.1097/00005537-200009000-00027.

[33] B. Cakirer, M. G. Hans, G. Graham, J. Aylor, P. V. Tishler, and S. Redline, “The relationship between craniofacial morphology and obstructive sleep apnea in whites and in African-Americans,” *Am. J. Respir. Crit. Care Med.*, vol. 163, no. 4, pp. 947–950, Mar. 2001, doi: 10.1164/ajrccm.163.4.2005136.

[34] E. Simou, J. Britton, and J. Leonardi-Bee, “Alcohol and the risk of sleep apnoea: a systematic review and meta-analysis,” *Sleep Med.*, vol. 42, pp. 38–46, Feb. 2018, doi: 10.1016/j.sleep.2017.12.005.

[35] D. W. Wetter, T. B. Young, T. R. Bidwell, M. S. Badr, and M. Palta, “Smoking as a risk factor for sleep-disordered breathing,” *Arch. Intern. Med.*, vol. 154, no. 19, pp. 2219–2224, Oct. 1994.

[36] N. M. Punjabi, “The Epidemiology of Adult Obstructive Sleep Apnea,” *Proc. Am. Thorac. Soc.*, vol. 5, no. 2, pp. 136–143, Feb. 2008, doi: 10.1513/pats.200709-155MG.

[37] C. Sauter *et al.*, “Excessive daytime sleepiness in patients suffering from different levels of obstructive sleep apnoea syndrome,” *J. Sleep Res.*, vol. 9, no. 3, pp. 293–301, Sep. 2000.

[38] G. Slater and J. Steier, “Excessive daytime sleepiness in sleep disorders,” *J. Thorac. Dis.*, vol. 4, no. 6, pp. 608–616, Dec. 2012, doi: 10.3978/j.issn.2072-1439.2012.10.07.

[39] P. Bernstein and J. H. Ebba, “Snoring versus obstructive sleep apnea: a case report,” *Perm. J.*, vol. 10, no. 1, pp. 21–23, 2006.

[40] C. Guilleminault, A. Tilkian, and W. C. Dement, “The sleep apnea syndromes,” *Annu. Rev. Med.*, vol. 27, pp. 465–484, 1976, doi: 10.1146/annurev.me.27.020176.002341.

[41] A. Kales *et al.*, “Severe obstructive sleep apnea--I: Onset, clinical course, and characteristics,” *J. Chronic Dis.*, vol. 38, no. 5, pp. 419–425, 1985.

[42] N. F. Watson, “Health Care Savings: The Economic Value of Diagnostic and Therapeutic Care for Obstructive Sleep Apnea,” *J. Clin. Sleep Med. JCSM Off. Publ. Am. Acad. Sleep Med.*, vol. 12, no. 8, pp. 1075–1077, Aug. 2016, doi: 10.5664/jcsm.6034.

[43] Public Health Agency of Canada, *What is the impact of sleep apnea on Canadians?: fast facts from the 2009 Canadian community health survey - sleep apnea rapid response.* Ottawa: Public Health Agency of Canada, 2010.

[44] W. R. Ruehland, P. D. Rochford, F. J. O’Donoghue, R. J. Pierce, P. Singh, and A. T. Thornton, “The New AASM Criteria for Scoring Hypopneas: Impact on the Apnea Hypopnea Index,” *Sleep*, vol. 32, no. 2, pp. 150–157, Feb. 2009.

[45] “Sleep-Related Breathing Disorders in Adults: Recommendations for Syndrome Definition and Measurement Techniques in Clinical Research | Sleep | Oxford Academic.” https://academic.oup.com/sleep/article/22/5/667/2726040 (accessed Jan. 23, 2019).

[46] M. Shokoueinejad *et al.*, “Sleep apnea: a review of diagnostic sensors, algorithms, and therapies,” *Physiol. Meas.*, vol. 38, no. 9, pp. R204–R252, Aug. 2017, doi: 10.1088/1361-6579/aa6ec6.

[47] N. A. Collop *et al.*, “Obstructive sleep apnea devices for out-of-center (OOC) testing: technology evaluation,” *J. Clin. Sleep Med. JCSM Off. Publ. Am. Acad. Sleep Med.*, vol. 7, no. 5, pp. 531–548, Oct. 2011, doi: 10.5664/JCSM.1328.

[48] D. Alvarez-Estevez and V. Moret-Bonillo, “Computer-Assisted Diagnosis of the Sleep Apnea-Hypopnea Syndrome: A Review,” *Sleep Disord.*, vol. 2015, p. 237878, 2015, doi: 10.1155/2015/237878.

[49] M. B. Uddin, C. M. Chow, and S. W. Su, “Classification methods to detect sleep apnea in adults based on respiratory and oximetry signals: a systematic review,” *Physiol. Meas.*, vol. 39, no. 3, p. 03TR01, Mar. 2018, doi: 10.1088/1361-6579/aaafb8.

[50] J. A. Waxman, D. Graupe, and D. W. Carley, “Automated prediction of apnea and hypopnea, using a LAMSTAR artificial neural network,” *Am. J. Respir. Crit. Care Med.*, vol. 181, no. 7, pp. 727–733, Apr. 2010, doi: 10.1164/rccm.200907-1146OC.

[51] D. Álvarez-Estévez and V. Moret-Bonillo, “Fuzzy reasoning used to detect apneic events in the sleep apnea-hypopnea syndrome,” *Expert Syst. Appl.*, vol. 36, no. 4, pp. 7778–7785, May 2009, doi: 10.1016/j.eswa.2008.11.043.

[52] Y. Maali and A. Al-Jumaily, “Genetic Fuzzy Approach based Sleep Apnea/Hypopnea Detection,” *Int. J. Mach. Learn. Comput.*, pp. 685–688, 2012, doi: 10.7763/IJMLC.2012.V2.215.

[53] A. Otero, P. Félix, and M. R. Álvarez, “Algorithms for the analysis of polysomnographic recordings with customizable criteria,” *Expert Syst. Appl.*, vol. 38, no. 8, pp. 10133–10146, Aug. 2011, doi: 10.1016/j.eswa.2011.02.081.

[54] H. M. Al-Angari and A. V. Sahakian, “Automated recognition of obstructive sleep apnea syndrome using support vector machine classifier,” *IEEE Trans. Inf. Technol. Biomed. Publ. IEEE Eng. Med. Biol. Soc.*, vol. 16, no. 3, pp. 463–468, May 2012, doi: 10.1109/TITB.2012.2185809.

[55] P. Várady, T. Micsik, S. Benedek, and Z. Benyó, “A novel method for the detection of apnea and hypopnea events in respiration signals,” *IEEE Trans. Biomed. Eng.*, vol. 49, no. 9, pp. 936–942, Sep. 2002, doi: 10.1109/TBME.2002.802009.

[56] C. Avcı and A. Akbaş, “Sleep apnea classification based on respiration signals by using ensemble methods,” *Biomed. Mater. Eng.*, vol. 26 Suppl 1, pp. S1703-1710, 2015, doi: 10.3233/BME-151470.

[57] Y. Maali and A. Al-Jumaily, “Multi Neural Networks Investigation based Sleep Apnea Prediction,” *Procedia Comput. Sci.*, vol. 24, pp. 97–102, Jan. 2013, doi: 10.1016/j.procs.2013.10.031.

[58] J. H. Choi, B. Lee, and S.-H. Hwang, “Association of Respiratory Mechanic Instability and Respiratory Parameters among Adults with Obstructive Sleep Apnea,” *Otolaryngol.--Head Neck Surg. Off. J. Am. Acad. Otolaryngol.-Head Neck Surg.*, p. 194599818825468, Jan. 2019, doi: 10.1177/0194599818825468.

[59] B. H. Taha *et al.*, “Automated detection and classification of sleep-disordered breathing from conventional polysomnography data,” *Sleep*, vol. 20, no. 11, pp. 991–1001, Nov. 1997.

[60] J. Y. Tian and J. Q. Liu, "Apnea Detection Based on Time Delay Neural Network," in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, Jan. 2005, pp. 2571–2574, doi: 10.1109/IEMBS.2005.1616994.

[61] D. Alvarez, G. C. Gutierrez, J. V. Marcos, F. Del Campo, and R. Hornero, "Spectral analysis of single-channel airflow and oxygen saturation recordings in obstructive sleep apnea detection," *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2010, pp. 847–850, 2010, doi: 10.1109/IEMBS.2010.5626861.

[62] W. Huang, B. Guo, Y. Shen, and X. Tang, "A novel method to precisely detect apnea and hypopnea events by airflow and oximetry signals," *Comput. Biol. Med.*, vol. 88, pp. 32–40, 01 2017, doi: 10.1016/j.compbiomed.2017.06.015.

[63] A. Otero, P. Félix, S. Barro, and C. Zamarrón, "A structural knowledge-based proposal for the identification and characterization of apnoea episodes," *Appl. Soft Comput.*, vol. 12, no. 1, pp. 516–526, Jan. 2012, doi: 10.1016/j.asoc.2011.08.009.

[64] B. V. Vaughn and P. Giallanza, "Technical review of polysomnography," *Chest*, vol. 134, no. 6, pp. 1310–1319, Dec. 2008, doi: 10.1378/chest.08-0812.

[65] B. L. Koley and D. Dey, "Real-Time Adaptive Apnea and Hypopnea Event Detection Methodology for Portable Sleep Apnea Monitoring Devices," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 12, pp. 3354–3363, Dec. 2013, doi: 10.1109/TBME.2013.2282337.

[66] H. Nazeran, A. Almas, K. Behbehani, J. Burk, and E. Lucas, "A fuzzy inference system for detection of obstructive sleep apnea," in *2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Oct. 2001, vol. 2, pp. 1645–1648 vol.2, doi: 10.1109/IEMBS.2001.1020530.

[67] A. A. Morsy and K. M. Al-Ashmouny, “Sleep Apnea Detection Using an Adaptive Fuzzy Logic Based Screening System,” in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, Jan. 2005, pp. 6124–6127, doi: 10.1109/IEMBS.2005.1615891.

[68] C. Massaroni, A. Nicolò, D. Lo Presti, M. Sacchetti, S. Silvestri, and E. Schena, “Contact-Based Methods for Measuring Respiratory Rate,” *Sensors*, vol. 19, no. 4, p. 908, Jan. 2019, doi: 10.3390/s19040908.

[69] Z. Zhang, J. Zheng, H. Wu, W. Wang, B. Wang, and H. Liu, “Development of a Respiratory Inductive Plethysmography Module Supporting Multiple Sensors for Wearable Systems,” *Sensors*, vol. 12, no. 10, pp. 13167–13184, Sep. 2012, doi: 10.3390/s121013167.

[70] J. P. Cantineau, P. Escourrou, R. Sartene, C. Gaultier, and M. Goldman, “Accuracy of respiratory inductive plethysmography during wakefulness and sleep in patients with obstructive sleep apnea,” *Chest*, vol. 102, no. 4, pp. 1145–1151, Oct. 1992.

[71] A. Thommandram, J. M. Eklund, and C. McGregor, “Detection of apnoea from respiratory time series data using clinically recognizable features and kNN classification,” *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2013, pp. 5013–5016, 2013, doi: 10.1109/EMBC.2013.6610674.

[72] M. E. Tagluk and N. Sezgin, “Classification of sleep apnea through sub-band energy of abdominal effort signal using Wavelets + Neural Networks,” *J. Med. Syst.*, vol. 34, no. 6, pp. 1111–1119, Dec. 2010, doi: 10.1007/s10916-009-9330-5.

[73] M. Emin Tagluk, M. Akin, and N. Sezgin, “Classıfication of sleep apnea by using wavelet transform and artificial neural networks,” *Expert Syst. Appl.*, vol. 37, no. 2, pp. 1600–1607, Mar. 2010, doi: 10.1016/j.eswa.2009.06.049.

[74] P. Varady and S. Bongar, “Detection of airway obstruction and sleep apnea by analyzing the phase relation of respiration movement signals,” in *IMTC 2001. Proceedings of the 18th IEEE Instrumentation and Measurement Technology Conference. Rediscovering Measurement in the Age of Informatics (Cat. No.01CH 37188)*, May 2001, vol. 1, pp. 185–190 vol.1, doi: 10.1109/IMTC.2001.928810.

[75] M. T. Bianchi, T. Lipoma, C. Darling, Y. Alameddine, and M. B. Westover, “Automated sleep apnea quantification based on respiratory movement,” *Int. J. Med. Sci.*, vol. 11, no. 8, pp. 796–802, 2014, doi: 10.7150/ijms.9303.

[76] K. D. McClatchey, *Clinical Laboratory Medicine*. Lippincott Williams & Wilkins, 2002.

[77] “Medical Instrumentation: Application and Design, 4th Edition,” *Wiley.com*. .

[78] N. Netzer, A. H. Eliasson, C. Netzer, and D. A. Kristo, “Overnight pulse oximetry for sleep-disordered breathing in adults: a review,” *Chest*, vol. 120, no. 2, pp. 625–633, Aug. 2001.

[79] K. P. Strohl and M. D. Altose, “Oxygen saturation during breath-holding and during apneas in sleep,” *Chest*, vol. 85, no. 2, pp. 181–186, Feb. 1984.

[80] C. Zamarrón, P. V. Romero, F. Gude, A. Amaro, and J. R. Rodriguez, “Screening of obstructive sleep apnoea: heart rate spectral analysis of nocturnal pulse oximetric recording,” *Respir. Med.*, vol. 95, no. 9, pp. 759–765, Sep. 2001, doi: 10.1053/rmed.2001.1128.

[81] J. V. Marcos, R. Hornero, I. T. Nabney, D. Álvarez, and F. D. Campo, “Analysis of nocturnal oxygen saturation recordings using kernel entropy to assist in sleep apnea-hypopnea diagnosis,” in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2011, pp. 1745–1748, doi: 10.1109/IEMBS.2011.6090499.

[82] D. Alvarez, R. Hornero, J. V. Marcos, F. del Campo, and M. López, "Obstructive sleep apnea detection using clustering classification of nonlinear features from nocturnal oximetry," *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2007, pp. 1937–1940, 2007, doi: 10.1109/IEMBS.2007.4352696.

[83] J. V. Marcos, R. Hornero, D. Alvarez, F. Del Campo, and M. Aboy, "Automated detection of obstructive sleep apnoea syndrome from oxygen saturation recordings using linear discriminant analysis," *Med. Biol. Eng. Comput.*, vol. 48, no. 9, pp. 895–902, Sep. 2010, doi: 10.1007/s11517-010-0646-6.

[84] J. V. Marcos, R. Hornero, D. Alvarez, F. Del Campo, C. Zamarrón, and M. López, "Utility of multilayer perceptron neural network classifiers in the diagnosis of the obstructive sleep apnoea syndrome from nocturnal oximetry," *Comput. Methods Programs Biomed.*, vol. 92, no. 1, pp. 79–89, Oct. 2008, doi: 10.1016/j.cmpb.2008.05.006.

[85] J. Víctor Marcos, R. Hornero, D. Alvarez, F. Del Campo, C. Zamarrón, and M. López, "Single layer network classifiers to assist in the detection of obstructive sleep apnea syndrome from oximetry data," *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2008, pp. 1651–1654, 2008, doi: 10.1109/IEMBS.2008.4649491.

[86] H. Rauscher, W. Popp, and H. Zwick, "Computerized detection of respiratory events during sleep from rapid increases in oxyhemoglobin saturation," *Lung*, vol. 169, no. 6, pp. 335–342, 1991.

[87] F. Portier *et al.*, "Evaluation of Home versus Laboratory Polysomnography in the Diagnosis of Sleep Apnea Syndrome," *Am. J. Respir. Crit. Care Med.*, vol. 162, no. 3, pp. 814–818, Sep. 2000, doi: 10.1164/ajrccm.162.3.9908002.

[88] S. Kristiansen, M. S. Hugaas, V. Goebel, T. Plagemann, K. Nikolaidis, and K. Liestøl, "Data Mining for Patient Friendly Apnea Detection," *IEEE Access*, vol. 6, pp. 74598–74615, 2018, doi: 10.1109/ACCESS.2018.2882270.

[89] K. S. Park and S. H. Choi, "Smart technologies toward sleep monitoring at home," *Biomed. Eng. Lett.*, Jan. 2019, doi: 10.1007/s13534-018-0091-2.

[90] A. Zaffaroni, P. de Chazal, C. Heneghan, P. Boyle, P. R. Mppm, and W. T. McNicholas, "SleepMinder: an innovative contact-free device for the estimation of the apnoea-hypopnoea index," *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2009, pp. 7091–7094, 2009, doi: 10.1109/IEMBS.2009.5332909.

[91] P. de Chazal, E. O'Hare, N. Fox, and C. Heneghan, "Assessment of sleep/wake patterns using a non-contact biomotion sensor," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2008, pp. 514–517, doi: 10.1109/IEMBS.2008.4649203.

[92] M. B. Norman, S. Middleton, O. Erskine, P. G. Middleton, J. R. Wheatley, and C. E. Sullivan, "Validation of the Sonomat: a contactless monitoring system used for the diagnosis of sleep disordered breathing," *Sleep*, vol. 37, no. 9, pp. 1477–1487, Sep. 2014, doi: 10.5665/sleep.3996.

[93] D. C. Mack, M. Alwan, B. Turner, P. Suratt, and R. A. Felder, "A Passive and Portable System for Monitoring Heart Rate and Detecting Sleep Apnea and Arousals: Preliminary Validation," in *1st Transdisciplinary Conference on Distributed Diagnosis and Home Healthcare, 2006. D2H2.*, Apr. 2006, pp. 51–54, doi: 10.1109/DDHH.2006.1624795.

[94] R. Crupi, T. Faetti, and R. Paradiso, “Preliminary evaluation of Wearable Wellness System for Obstructive Sleep Apnea detection,” in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug. 2015, pp. 4142–4145, doi: 10.1109/EMBC.2015.7319306.

[95] J. H. Shin, Y. J. Chee, D. Jeong, and K. S. Park, “Nonconstrained Sleep Monitoring System and Algorithms Using Air-Mattress With Balancing Tube Method,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 1, pp. 147–156, Jan. 2010, doi: 10.1109/TITB.2009.2034011.

[96] C. Yang, G. Cheung, V. Stankovic, K. Chan, and N. Ono, “Sleep Apnea Detection via Depth Video and Audio Feature Learning,” *IEEE Trans. Multimed.*, vol. 19, no. 4, pp. 822–835, Apr. 2017, doi: 10.1109/TMM.2016.2626969.

[97] J. Abad *et al.*, “Automatic Video Analysis for Obstructive Sleep Apnea Diagnosis,” *Sleep*, vol. 39, no. 8, pp. 1507–1515, 01 2016, doi: 10.5665/sleep.6008.

[98] O. Faust, Y. Hagiwara, T. J. Hong, O. S. Lih, and U. R. Acharya, “Deep learning for healthcare applications based on physiological signals: A review,” *Comput. Methods Programs Biomed.*, vol. 161, pp. 1–13, Jul. 2018, doi: 10.1016/j.cmpb.2018.04.005.

[99] N. Ganapathy, R. Swaminathan, and T. M. Deserno, “Deep Learning on 1-D Biosignals: a Taxonomy-based Survey,” *Yearb. Med. Inform.*, vol. 27, no. 1, pp. 98–109, Aug. 2018, doi: 10.1055/s-0038-1667083.

[100] C. Xiao, E. Choi, and J. Sun, “Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review,” *J. Am. Med. Inform. Assoc. JAMIA*, vol. 25, no. 10, pp. 1419–1428, Oct. 2018, doi: 10.1093/jamia/ocy068.

[101] R. Zemouri, N. Zerhouni, and D. Racoceanu, “Deep Learning in the Biomedical Applications: Recent and Future Status,” *Appl. Sci.*, vol. 9, no. 8, Art. no. 8, Jan. 2019, doi: 10.3390/app9081526.

[102] U. Erdenebayar, Y. J. Kim, J.-U. Park, E. Y. Joo, and K.-J. Lee, “Deep learning approaches for automatic detection of sleep apnea events from an electrocardiogram,” *Comput. Methods Programs Biomed.*, vol. 180, p. 105001, Oct. 2019, doi: 10.1016/j.cmpb.2019.105001.

[103] S. J. Pan and Q. Yang, “A Survey on Transfer Learning,” *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010, doi: 10.1109/TKDE.2009.191.

[104] A. A. Gharbali, S. Najdi, and J. M. Fonseca, “Transfer Learning of Spectrogram Image for Automatic Sleep Stage Classification,” in *Image Analysis and Recognition*, 2018, pp. 522–528.

[105] L. Bote-Curiel, S. Muñoz-Romero, A. Gerrero-Curieses, and J. L. Rojo-Álvarez, “Deep Learning and Big Data in Healthcare: A Double Review for Critical Beginners,” *Appl. Sci.*, vol. 9, no. 11, p. 2331, Jan. 2019, doi: 10.3390/app9112331.

[106] A. Vilamala, K. H. Madsen, and L. K. Hansen, “Deep Convolutional Neural Networks for Interpretable Analysis of EEG Sleep Stage Scoring,” *ArXiv171000633 Cs Stat*, Oct. 2017, Accessed: Sep. 09, 2019. [Online]. Available: http://arxiv.org/abs/1710.00633.

[107] S. Arslan Tuncer, B. Akılotu, and S. Toraman, “A deep learning-based decision support system for diagnosis of OSAS using PTT signals,” *Med. Hypotheses*, vol. 127, pp. 15–22, Jun. 2019, doi: 10.1016/j.mehy.2019.03.026.

[108] D. Verstraete, A. Ferrada, E. L. Droguett, V. Meruane, and M. Modarres, “Deep Learning Enabled Fault Diagnosis Using Time-Frequency Image Analysis of Rolling Element

Bearings," *Shock and Vibration*, Oct. 09, 2017. https://www.hindawi.com/journals/sv/2017/5067651/ (accessed Nov. 16, 2020).

[109] C.-H. Kang, U. Erdenebayar, J.-U. Park, and K.-J. Lee, "Multi-Class Classification of Sleep Apnea/Hypopnea Events Based on Long Short-Term Memory Using a Photoplethysmography Signal," *J. Med. Syst.*, vol. 44, no. 1, p. 14, Dec. 2019, doi: 10.1007/s10916-019-1485-0.

[110] H. T. Chaw, S. Kamolphiwong, and K. Wongsritrang, "Sleep apnea detection using deep learning," *Teh. Glas.*, vol. 13, no. 4, pp. 261–266, Dec. 2019, doi: 10.31803/tg-20191104191722.

[111] C. E. Sullivan, F. G. Issa, M. Berthon-Jones, and L. Eves, "Reversal of obstructive sleep apnoea by continuous positive airway pressure applied through the nares," *Lancet Lond. Engl.*, vol. 1, no. 8225, pp. 862–865, Apr. 1981.

[112] M. Arzt *et al.*, "Suppression of central sleep apnea by continuous positive airway pressure and transplant-free survival in heart failure: a post hoc analysis of the Canadian Continuous Positive Airway Pressure for Patients with Central Sleep Apnea and Heart Failure Trial (CANPAP)," *Circulation*, vol. 115, no. 25, pp. 3173–3180, Jun. 2007, doi: 10.1161/CIRCULATIONAHA.106.683482.

[113] V. Hoffstein, "Review of oral appliances for treatment of sleep-disordered breathing," *Sleep Breath. Schlaf Atm.*, vol. 11, no. 1, pp. 1–22, Mar. 2007, doi: 10.1007/s11325-006-0084-8.

[114] C. A. Kushida *et al.*, "Practice parameters for the treatment of snoring and Obstructive Sleep Apnea with oral appliances: an update for 2005," *Sleep*, vol. 29, no. 2, pp. 240–243, Feb. 2006.

[115] P. R. Eastwood *et al.*, “Treating obstructive sleep apnea with hypoglossal nerve stimulation,” *Sleep*, vol. 34, no. 11, pp. 1479–1486, Nov. 2011, doi: 10.5665/sleep.1380.

[116] P. J. Strollo *et al.*, “Upper-Airway Stimulation for Obstructive Sleep Apnea,” *N. Engl. J. Med.*, vol. 370, no. 2, pp. 139–149, Jan. 2014, doi: 10.1056/NEJMoa1308659.

[117] W. T. Abraham *et al.*, “Phrenic nerve stimulation for the treatment of central sleep apnea,” *JACC Heart Fail.*, vol. 3, no. 5, pp. 360–369, May 2015, doi: 10.1016/j.jchf.2014.12.013.

[118] S. Joseph and M. R. Costanzo, “A novel therapeutic approach for central sleep apnea: Phrenic nerve stimulation by the remedē® System,” *Int. J. Cardiol.*, vol. 206 Suppl, pp. S28-34, Mar. 2016, doi: 10.1016/j.ijcard.2016.02.121.

[119] R. B. Berry, M. H. Kryger, and C. A. Massie, “A novel nasal expiratory positive airway pressure (EPAP) device for the treatment of obstructive sleep apnea: a randomized controlled trial,” *Sleep*, vol. 34, no. 4, pp. 479–485, Apr. 2011.

[120] A. Mulchrone, M. Shokoueinejad, and J. Webster, “A review of preventing central sleep apnea by inspired CO2,” *Physiol. Meas.*, vol. 37, no. 5, pp. R36-45, 2016, doi: 10.1088/0967-3334/37/5/R36.

[121] M. Bruyneel and V. Ninane, “Unattended home-based polysomnography for sleep disordered breathing: current concepts and perspectives,” *Sleep Med. Rev.*, vol. 18, no. 4, pp. 341–347, Aug. 2014, doi: 10.1016/j.smrv.2013.12.002.

[122] N. J. Douglas, “Home diagnosis of the obstructive sleep apnoea/hypopnoea syndrome,” *Sleep Med. Rev.*, vol. 7, no. 1, pp. 53–59, Feb. 2003, doi: 10.1053/smrv.2001.0205.

[123] J. M. Kelly, R. E. Strecker, and M. T. Bianchi, “Recent Developments in Home Sleep-Monitoring Devices,” *International Scholarly Research Notices*, 2012. https://www.hindawi.com/journals/isrn/2012/768794/ (accessed Mar. 22, 2019).

[124] A. Y. Ng and M. I. Jordan, “On Discriminative vs. Generative Classifiers: A comparison of logistic regression and naive Bayes,” in *Advances in Neural Information Processing Systems 14*, T. G. Dietterich, S. Becker, and Z. Ghahramani, Eds. MIT Press, 2002, pp. 841–848.

[125] J. H. Porter, “Differential fiber optic proximity sensor,” US4358960A, Nov. 16, 1982.

[126] J.-H. Kim, R. Roberge, J. B. Powell, A. B. Shafer, and W. J. Williams, “Measurement Accuracy of Heart Rate and Respiratory Rate during Graded Exercise and Sustained Exercise in the Heat Using the Zephyr BioHarness™,” *Int. J. Sports Med.*, vol. 34, no. 6, pp. 497–501, Jun. 2013, doi: 10.1055/s-0032-1327661.

[127] R. Mueller, “Sleep Data - National Sleep Research Resource - NSRR.” https://sleepdata.org/ (accessed Feb. 15, 2019).

[128] G. Parati, G. Bilo, C. Lombardi, and G. Mancia, “Chapter 11 - Secondary Hypertension: Sleep Apnea,” in *Hypertension*, H. R. Black and W. J. Elliott, Eds. Philadelphia: W.B. Saunders, 2007, pp. 134–143.

[129] Z. Ren, “Patient Mobility Monitoring Through the Analysis of Bed Exit Centre of Pressure,” Text, Carleton University, 2015.

[130] Goldberger Ary L. *et al.*, “PhysioBank, PhysioToolkit, and PhysioNet,” *Circulation*, vol. 101, no. 23, pp. e215–e220, Jun. 2000, doi: 10.1161/01.CIR.101.23.e215.

[131] A. Kales, A. Rechtschaffen, University of California, Los Angeles, and NINDB Neurological Information Network (U.S.), Eds., *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects: Allan Rechtschaffen and Anthony Kales, editors*. Bethesda, Md: U. S. National Institute of Neurological Diseases and Blindness, Neurological Information Network, 1968.

[132] F. Q. Al-Khalidi, R. Saatchi, D. Burke, H. Elphick, and S. Tan, “Respiration rate monitoring methods: a review,” *Pediatr. Pulmonol.*, vol. 46, no. 6, pp. 523–529, Jun. 2011, doi: 10.1002/ppul.21416.

[133] A. Arcelus, M. H. Jones, R. Goubran, and F. Knoefel, “Integration of Smart Home Technologies in a Health Monitoring System for the Elderly,” in *21st International Conference on Advanced Information Networking and Applications Workshops (AINAW’07)*, May 2007, vol. 2, pp. 820–825, doi: 10.1109/AINAW.2007.209.

[134] D. I. Townsend, M. Holtzman, R. Goubran, M. Frize, and F. Knoefel, “Measurement of Torso Movement With Delay Mapping Using an Unobtrusive Pressure-Sensor Array,” *IEEE Trans. Instrum. Meas.*, vol. 60, no. 5, pp. 1751–1760, May 2011, doi: 10.1109/TIM.2010.2092872.

[135] M. Holtzman, D. Townsend, R. Goubran, and F. Knoefel, “Breathing sensor selection during movement,” *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2011, pp. 381–384, 2011, doi: 10.1109/IEMBS.2011.6090123.

[136] M. Holtzman, D. Townsend, R. Goubran, and F. Knoefel, “Validation of pressure sensors for physiological monitoring in home environments,” in *2010 IEEE International Workshop on Medical Measurements and Applications*, Apr. 2010, pp. 38–42, doi: 10.1109/MEMEA.2010.5480218.

[137] D. Townsend, M. Holtzman, R. Goubran, M. Frize, and F. Knoefel, “Relative Thresholding With Under-Mattress Pressure Sensors to Detect Central Apnea,” *IEEE Trans. Instrum. Meas.*, vol. 60, no. 10, pp. 3281–3289, Oct. 2011, doi: 10.1109/TIM.2011.2123250.

[138] M. Holtzman, R. Goubran, and F. Knoefel, “Maximal ratio combining for respiratory effort extraction from pressure sensor arrays,” in *2011 IEEE International Symposium on Medical*

Measurements and Applications, Bari, Italy, May 2011, pp. 88–92, doi: 10.1109/MeMeA.2011.5966674.

[139] L. Litwin, "FIR and IIR digital filters," *IEEE Potentials*, vol. 19, no. 4, pp. 28–31, Oct. 2000, doi: 10.1109/45.877863.

[140] H. Schulz, "Rethinking Sleep Analysis," *J. Clin. Sleep Med. JCSM Off. Publ. Am. Acad. Sleep Med.*, vol. 4, no. 2, pp. 99–103, Apr. 2008.

[141] K. Niizeki, I. Nishidate, K. Uchida, and M. Kuwahara, "Unconstrained cardiorespiratory and body movement monitoring system for home care," *Med. Biol. Eng. Comput.*, vol. 43, no. 6, pp. 716–724, Nov. 2005.

[142] H. F. M. Van der Loos, N. Ullrich, and H. Kobayashi, "Development of Sensate and Robotic Bed Technologies for Vital Signs Monitoring and Sleep Quality Improvement," *Auton. Robots*, vol. 15, no. 1, pp. 67–79, Jul. 2003, doi: 10.1023/A:1024444917917.

[143] J. L. Rodgers and W. A. Nicewander, "Thirteen Ways to Look at the Correlation Coefficient," *Am. Stat.*, vol. 42, no. 1, pp. 59–66, Feb. 1988, doi: 10.1080/00031305.1988.10475524.

[144] D. G. Brennan, "Linear Diversity Combining Techniques," *Proc. IRE*, vol. 47, no. 6, pp. 1075–1102, Jun. 1959, doi: 10.1109/JRPROC.1959.287136.

[145] B. D. Van Veen and K. M. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP Mag.*, vol. 5, no. 2, pp. 4–24, Apr. 1988, doi: 10.1109/53.665.

[146] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*. Springer Science & Business Media, 2008.

[147] D. Waltisberg, O. Amft, and G. Tröster, "Accuracy-coverage Tradeoff of Nocturnal Vital Sign Estimation in Smart Beds," in *Proceedings of the 2014 ACM International Joint*

Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, New York, NY, USA, 2014, pp. 1249–1256, doi: 10.1145/2638728.2638810.

[148] M. H. Jones, R. Goubran, and F. Knoefel, "Reliable Respiratory Rate Estimation from a Bed Pressure Array," in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2006, pp. 6410–6413, doi: 10.1109/IEMBS.2006.260164.

[149] "St. Vincent's University Hospital / University College Dublin Sleep Apnea Database." https://physionet.org/pn3/ucddb/ (accessed Dec. 15, 2018).

[150] A. Mesaros, T. Heittola, and T. Virtanen, "Metrics for Polyphonic Sound Event Detection," *Appl. Sci.*, vol. 6, no. 6, p. 162, Jun. 2016, doi: 10.3390/app6060162.

[151] S. Devuyst, T. Dutoit, P. Stenuit, and M. Kerkhofs, "Automatic sleep spindles detection--overview and development of a standard proposal assessment method," *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2011, pp. 1713–1716, 2011, doi: 10.1109/IEMBS.2011.6090491.

[152] B. Xie and H. Minn, "Real-Time Sleep Apnea Detection by Classifier Combination," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 3, pp. 469–477, May 2012, doi: 10.1109/TITB.2012.2188299.

[153] C. R. Dell'Aquila, L. S. Correa, R. Correa, G. E. Cañadas, and E. Laciar, "New insights into the scoring of respiratory events based on alternative sensors: A comparative effectiveness study," in *VII Latin American Congress on Biomedical Engineering CLAIB 2016, Bucaramanga, Santander, Colombia, October 26th -28th, 2016*, 2017, pp. 605–608.

[154] M. S. Badr and S. Javaheri, "Central Sleep Apnea: a Brief Review," *Curr. Pulmonol. Rep.*, Mar. 2019, doi: 10.1007/s13665-019-0221-z.

[155] S. Bai, J. Z. Kolter, and V. Koltun, “An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling,” *ArXiv180301271 Cs*, Apr. 2018, Accessed: Jan. 09, 2020. [Online]. Available: http://arxiv.org/abs/1803.01271.

[156] D. Townsend, R. Goubran, F. Knoefel, and J. Leech, “Validation of Unobtrusive Pressure Sensor Array for Central Sleep Apnea Screening,” *IEEE Trans. Instrum. Meas.*, vol. 61, no. 7, pp. 1857–1865, Jul. 2012, doi: 10.1109/TIM.2012.2192342.

[157] N. A. Collop, “Scoring variability between polysomnography technologists in different sleep laboratories,” *Sleep Med.*, vol. 3, no. 1, pp. 43–47, Jan. 2002, doi: 10.1016/s1389-9457(01)00115-0.

[158] S. Lord *et al.*, “Interrater Reliability of Computer-Assisted Scoring of Breathing during Sleep,” *Sleep*, vol. 12, no. 6, pp. 550–558, Nov. 1989, doi: 10.1093/sleep/12.6.550.

[159] S. Soleimani, “Long-Term Sleep Assessment by Unobtrusive Pressure Sensor Arrays,” Thesis, Université d’Ottawa / University of Ottawa, School of Electrical Engineering and Computer Science, 2018.

[160] M. Pouliot, V. Joshi, R. Goubran, and F. Knoefel, “Bed occupancy monitoring: Data processing and clinician user interface design,” in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Diego, CA, Aug. 2012, pp. 5810–5814, doi: 10.1109/EMBC.2012.6347315.

[161] D. Dorran, R. Lawlor, and E. Coyle, “A comparison of time-domain time-scale modification algorithms,” 2006.

[162] C. A. Robles-Rubio, G. Bertolizio, K. A. Brown, and R. E. Kearney, “Scoring Tools for the Analysis of Infant Respiratory Inductive Plethysmography Signals,” *PloS One*, vol. 10, no. 7, p. e0134182, 2015, doi: 10.1371/journal.pone.0134182.

[163] H. Byun and S.-W. Lee, “A survey on pattern recognition applications of support vector machines,” *Int. J. Pattern Recognit. Artif. Intell.*, vol. 17, no. 03, pp. 459–486, May 2003, doi: 10.1142/S0218001403002460.

[164] F. Melgani and L. Bruzzone, “Classification of hyperspectral remote sensing images with support vector machines,” *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004, doi: 10.1109/TGRS.2004.831865.

[165] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, “Gene Selection for Cancer Classification using Support Vector Machines,” *Mach. Learn.*, vol. 46, no. 1, pp. 389–422, Jan. 2002, doi: 10.1023/A:1012487302797.

[166] A. H. Khandoker, M. Palaniswami, and C. K. Karmakar, “Support Vector Machines for Automated Recognition of Obstructive Sleep Apnea Syndrome From ECG Recordings,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 1, pp. 37–48, Jan. 2009, doi: 10.1109/TITB.2008.2004495.

[167] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, Eds., *Feature Extraction: Foundations and Applications*. Berlin Heidelberg: Springer-Verlag, 2006.

[168] A. Y. Abdelaziz, S. F. Mekhamer, A. M. Ibrahim, and M. Ramadan, “Transmission Line Shunt Faults Detection and Classification using Support Vector Machine,” p. 7.

[169] B. Hjorth, “EEG analysis based on time domain properties,” *Electroencephalogr. Clin. Neurophysiol.*, vol. 29, no. 3, pp. 306–310, Sep. 1970, doi: 10.1016/0013-4694(70)90143-4.

[170] S. J. Redmond and C. Heneghan, “Cardiorespiratory-based sleep staging in subjects with obstructive sleep apnea,” *IEEE Trans. Biomed. Eng.*, vol. 53, no. 3, pp. 485–496, Mar. 2006, doi: 10.1109/TBME.2005.869773.

[171] S. Charbonnier, L. Zoubek, S. Lesecq, and F. Chapotot, “Self-evaluated automatic classifier as a decision-support tool for sleep/wake staging,” *Comput. Biol. Med.*, vol. 41, no. 6, pp. 380–389, Jun. 2011, doi: 10.1016/j.compbiomed.2011.04.001.

[172] J. Ben Ali, L. Saidi, S. Harrath, E. Bechhoefer, and M. Benbouzid, “Online automatic diagnosis of wind turbine bearings progressive degradations under real experimental conditions based on unsupervised machine learning,” *Appl. Acoust.*, vol. 132, pp. 167–181, Mar. 2018, doi: 10.1016/j.apacoust.2017.11.021.

[173] K. Susmáková and A. Krakovská, “Discrimination ability of individual measures used in sleep stages classification,” *Artif. Intell. Med.*, vol. 44, no. 3, pp. 261–277, Nov. 2008, doi: 10.1016/j.artmed.2008.07.005.

[174] M. Bosi and R. E. Goldberg, *Introduction to Digital Audio Coding and Standards*. Springer US, 2003.

[175] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.

[176] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[177] Y. Zhang, Q. Liu, and L. Song, “Sentence-State LSTM for Text Representation,” *ArXiv180502474 Cs Stat*, May 2018, Accessed: Jan. 31, 2020. [Online]. Available: http://arxiv.org/abs/1805.02474.

[178] P. Badjatiya, L. J. Kurisinkel, M. Gupta, and V. Varma, “Attention-Based Neural Text Segmentation,” in *Advances in Information Retrieval*, Cham, 2018, pp. 180–193, doi: 10.1007/978-3-319-76941-7_14.

[179] H. He and E. A. Garcia, “Learning from Imbalanced Data,” *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009, doi: 10.1109/TKDE.2008.239.

[180] P. Lingras and C. J. Butz, “Precision and Recall in Rough Support Vector Machines,” in *2007 IEEE International Conference on Granular Computing (GRC 2007)*, Nov. 2007, pp. 654–654, doi: 10.1109/GrC.2007.77.

[181] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *ArXiv14126980 Cs*, Jan. 2017, Accessed: Jan. 08, 2020. [Online]. Available: http://arxiv.org/abs/1412.6980.

[182] F. Ahamed, S. Shahrestani, and H. Cheung, “Intelligent Fall Detection with Wearable IoT,” in *Complex, Intelligent, and Software Intensive Systems*, Cham, 2020, pp. 391–401, doi: 10.1007/978-3-030-22354-0_35.

[183] S. Saha *et al.*, “Portable diagnosis of sleep apnea with the validation of individual event detection,” *Sleep Med.*, Jan. 2020, doi: 10.1016/j.sleep.2019.12.021.

[184] M. Bsoul, H. Minn, and L. Tamil, “Apnea MedAssist: Real-time Sleep Apnea Monitor Using Single-Lead ECG,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 3, pp. 416–427, May 2011, doi: 10.1109/TITB.2010.2087386.

[185] A. Burgos, A. Goñi, A. Illarramendi, and J. Bermúdez, “Real-time detection of apneas on a PDA,” *IEEE Trans. Inf. Technol. Biomed. Publ. IEEE Eng. Med. Biol. Soc.*, vol. 14, no. 4, pp. 995–1002, Jul. 2010, doi: 10.1109/TITB.2009.2034975.

[186] M. Längkvist, L. Karlsson, and A. Loutfi, “A review of unsupervised feature learning and deep learning for time-series modeling,” *Pattern Recognit. Lett.*, vol. 42, pp. 11–24, Jun. 2014, doi: 10.1016/j.patrec.2014.01.008.

[187] P. M. Kulkarni *et al.*, “A deep learning approach for real-time detection of sleep spindles,” *J. Neural Eng.*, vol. 16, no. 3, p. 036004, Jun. 2019, doi: 10.1088/1741-2552/ab0933.

[188] B. Rim, N.-J. Sung, S. Min, and M. Hong, “Deep Learning in Physiological Signal Data: A Survey,” *Sensors*, vol. 20, no. 4, p. 969, Jan. 2020, doi: 10.3390/s20040969.

[189] W. W. Flemons *et al.*, “Home diagnosis of sleep apnea: a systematic review of the literature. An evidence review cosponsored by the American Academy of Sleep Medicine, the American College of Chest Physicians, and the American Thoracic Society,” *Chest*, vol. 124, no. 4, pp. 1543–1579, Oct. 2003.

[190] M. Kayabekir, “Sleep Physiology and Polysomnogram, Physiopathology and Symptomatology in Sleep Medicine,” *Sleep Med. Clin. Neurol.*, Feb. 2019, doi: 10.5772/intechopen.82754.

[191] R. D. Chervin, “The multiple sleep latency test and Epworth sleepiness scale in the assessment of daytime sleepiness,” *J. Sleep Res.*, vol. 9, no. 4, pp. 399–401, Dec. 2000.

[192] N. L. Bragazzi, O. Guglielmi, and S. Garbarino, “SleepOMICS: How Big Data Can Revolutionize Sleep Science,” *Int. J. Environ. Res. Public. Health*, vol. 16, no. 2, Jan. 2019, doi: 10.3390/ijerph16020291.

[193] “What Are Longitudinal Data? | National Longitudinal Surveys.” https://www.nlsinfo.org/content/getting-started/what-are-longitudinal-data (accessed Nov. 16, 2020).

[194] N. K. Y. Tang and A. G. Harvey, “Altering misperception of sleep in insomnia: behavioral experiment versus verbal feedback,” *J. Consult. Clin. Psychol.*, vol. 74, no. 4, pp. 767–776, Aug. 2006, doi: 10.1037/0022-006X.74.4.767.

[195] A. Sadeh and C. Acebo, “The role of actigraphy in sleep medicine,” *Sleep Med. Rev.*, vol. 6, no. 2, pp. 113–124, Apr. 2002.

[196] O. Le Bon *et al.*, "Mild to moderate sleep respiratory events: one negative night may not be enough," *Chest*, vol. 118, no. 2, pp. 353–359, Aug. 2000.

[197] J. Newell, O. Mairesse, P. Verbanck, and D. Neu, "Is a one-night stay in the lab really enough to conclude? First-night effect and night-to-night variability in polysomnographic recordings among different clinical population samples," *Psychiatry Res.*, vol. 200, no. 2–3, pp. 795–801, Dec. 2012, doi: 10.1016/j.psychres.2012.07.045.

[198] D. J. Levendowski *et al.*, "Assessment of the test-retest reliability of laboratory polysomnography," *Sleep Breath. Schlaf Atm.*, vol. 13, no. 2, pp. 163–167, May 2009, doi: 10.1007/s11325-008-0214-6.

[199] P. Cao, D. Zhao, and O. Zaiane, "An Optimized Cost-Sensitive SVM for Imbalanced Data Learning," in *Advances in Knowledge Discovery and Data Mining*, 2013, pp. 280–292.

[200] H. Byun and S.-W. Lee, "Applications of Support Vector Machines for Pattern Recognition: A Survey," in *Pattern Recognition with Support Vector Machines*, 2002, pp. 213–236.

[201] J. Zhang and I. Mani, "KNN Approach to Unbalanced Data Distributions: A Case Study Involving Information Extraction | BibSonomy." .

[202] A. J. K. Phillips *et al.*, "Irregular sleep/wake patterns are associated with poorer academic performance and delayed circadian and sleep/wake timing," *Sci. Rep.*, vol. 7, Jun. 2017, doi: 10.1038/s41598-017-03171-4.

[203] J. R. Lunsford-Avery, M. M. Engelhard, A. M. Navar, and S. H. Kollins, "Validation of the Sleep Regularity Index in Older Adults and Associations with Cardiometabolic Risk," *Sci. Rep.*, vol. 8, no. 1, p. 14158, Sep. 2018, doi: 10.1038/s41598-018-32402-5.

[204] Institute of Medicine (US) Committee on Sleep Medicine and Research, *Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem*. Washington (DC): National Academies Press (US), 2006.

[205] T. Young, M. Palta, J. Dempsey, J. Skatrud, S. Weber, and S. Badr, "The Occurrence of Sleep-Disordered Breathing among Middle-Aged Adults," *N. Engl. J. Med.*, vol. 328, no. 17, pp. 1230–1235, Apr. 1993, doi: 10.1056/NEJM199304293281704.

[206] Redline Susan *et al.*, "The Scoring of Respiratory Events in Sleep: Reliability and Validity," *J. Clin. Sleep Med.*, vol. 03, no. 02, pp. 169–200, Mar. 2007, doi: 10.5664/jcsm.26818.

[207] W. W. Flemons, N. J. Douglas, S. T. Kuna, D. O. Rodenstein, and J. Wheatley, "Access to Diagnosis and Treatment of Patients with Suspected Sleep Apnea," *Am. J. Respir. Crit. Care Med.*, vol. 169, no. 6, pp. 668–672, Mar. 2004, doi: 10.1164/rccm.200308-1124PP.

[208] M. Holtzman, A. Arcelus, R. Goubran, and F. Knoefel, "Breathing Signal Fusion in Pressure Sensor Arrays," in *2008 IEEE International Workshop on Medical Measurements and Applications*, Ottawa, ON, Canada, May 2008, pp. 71–76, doi: 10.1109/MEMEA.2008.4543001.

[209] M. H. Jones, R. Goubran, and F. Knoefel, "Identifying Movement Onset Times for a Bed-Based Pressure Sensor Array," in *IEEE International Workshop on Medical Measurement and Applications, 2006. MeMea 2006.*, Apr. 2006, pp. 111–114, doi: 10.1109/MEMEA.2006.1644474.

[210] R. D. Cartwright, "Effect of sleep position on sleep apnea severity," *Sleep*, vol. 7, no. 2, pp. 110–114, 1984.

[211] B. A. Phillips, J. Okeson, D. Paesani, and R. Gilmore, "Effect of sleep position on sleep apnea and parafunctional activity," *Chest*, vol. 90, no. 3, pp. 424–429, Sep. 1986.

[212] C. F. George, T. W. Millar, and M. H. Kryger, “Sleep apnea and body position during sleep,” *Sleep*, vol. 11, no. 1, pp. 90–99, Feb. 1988.

[213] “Consensus recommendations for the management of chronic heart failure. On behalf of the membership of the advisory council to improve outcomes nationwide in heart failure,” *Am. J. Cardiol.*, vol. 83, no. 2A, pp. 1A-38A, Jan. 1999.

[214] P. E. Peppard, T. Young, M. Palta, J. Dempsey, and J. Skatrud, “Longitudinal study of moderate weight change and sleep-disordered breathing,” *JAMA*, vol. 284, no. 23, pp. 3015–3021, Dec. 2000.

[215] M. Migliorini *et al.*, “Automatic sleep staging based on ballistocardiographic signals recorded through bed sensors,” *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2010, pp. 3273–3276, 2010, doi: 10.1109/IEMBS.2010.5627217.

[216] E. C. Winnebeck, D. Fischer, T. Leise, and T. Roenneberg, “Dynamics and Ultradian Structure of Human Sleep in Real Life,” *Curr. Biol. CB*, vol. 28, no. 1, pp. 49-59.e5, 08 2018, doi: 10.1016/j.cub.2017.11.063.

[217] D. I. Townsend, R. Goubran, M. Frize, and F. Knoefel, “Preliminary results on the effect of sensor position on unobtrusive rollover detection for sleep monitoring in smart homes,” in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Sep. 2009, pp. 6135–6138, doi: 10.1109/IEMBS.2009.5334690.

[218] H. Miwa, S. Sasahara, and T. Matsui, “Roll-over detection and sleep quality measurement using a wearable sensor,” *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2007, pp. 1507–1510, 2007, doi: 10.1109/IEMBS.2007.4352587.

[219] G. S. Chung, B. H. Choi, J.-S. Lee, J. S. Lee, D.-U. Jeong, and K. S. Park, “REM sleep estimation only using respiratory dynamics,” *Physiol. Meas.*, vol. 30, no. 12, pp. 1327–1340, Dec. 2009, doi: 10.1088/0967-3334/30/12/003.

[220] J. Paalasmaa, L. Leppäkorpi, and M. Partinen, “Quantifying respiratory variation with force sensor measurements,” *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2011, pp. 3812–3815, 2011, doi: 10.1109/IEMBS.2011.6090773.

[221] B. H. Jansen and K. Shankar, “Sleep staging with movement-related signals,” *Int. J. Biomed. Comput.*, vol. 32, no. 3, pp. 289–297, May 1993, doi: 10.1016/0020-7101(93)90021-W.

Appendix A

The CAI per night and duration of detected events for the other eight patients are presented in this section.

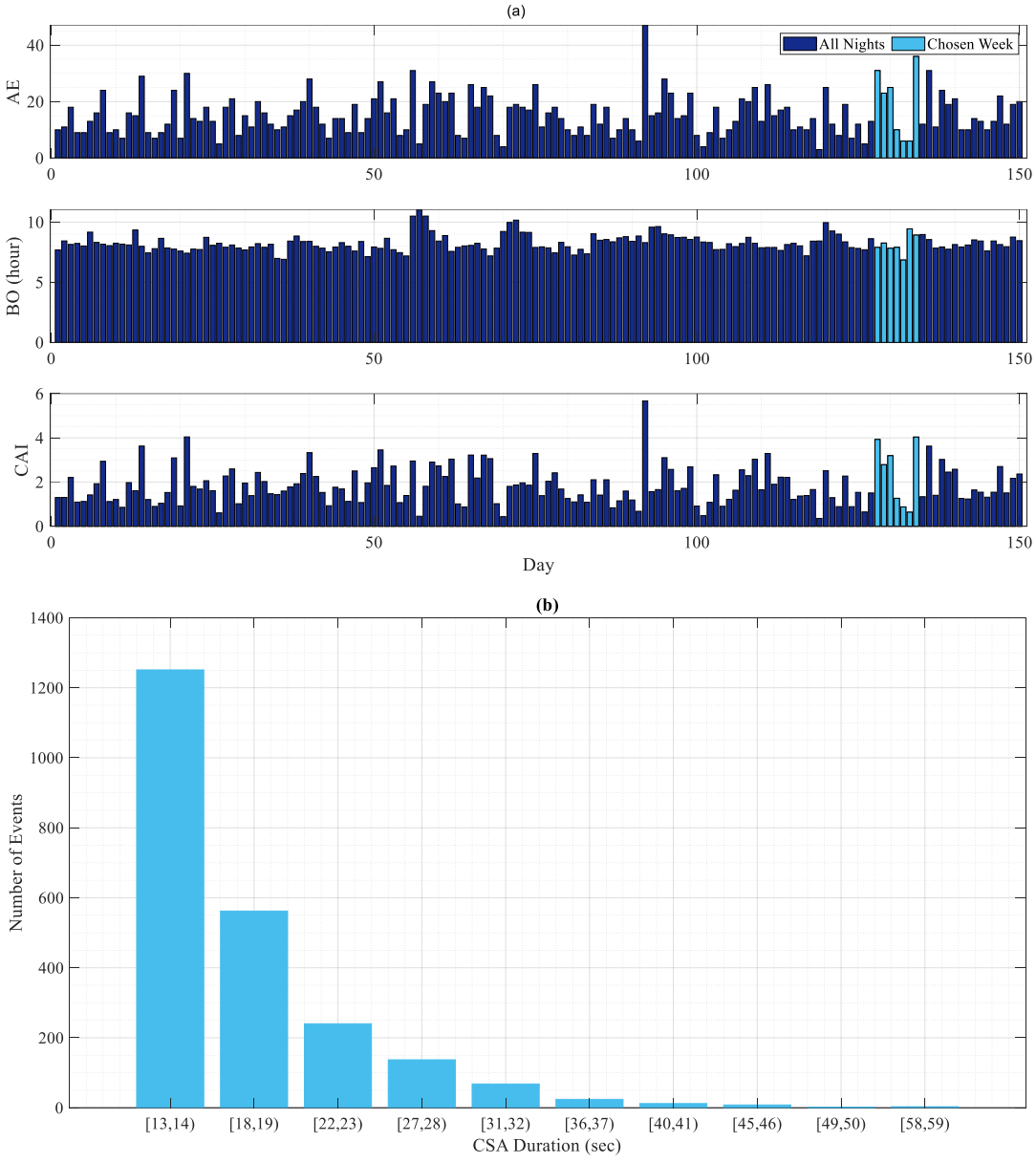


Figure A. 1 Subject 116, a) CAI per night and b) duration of detected events.

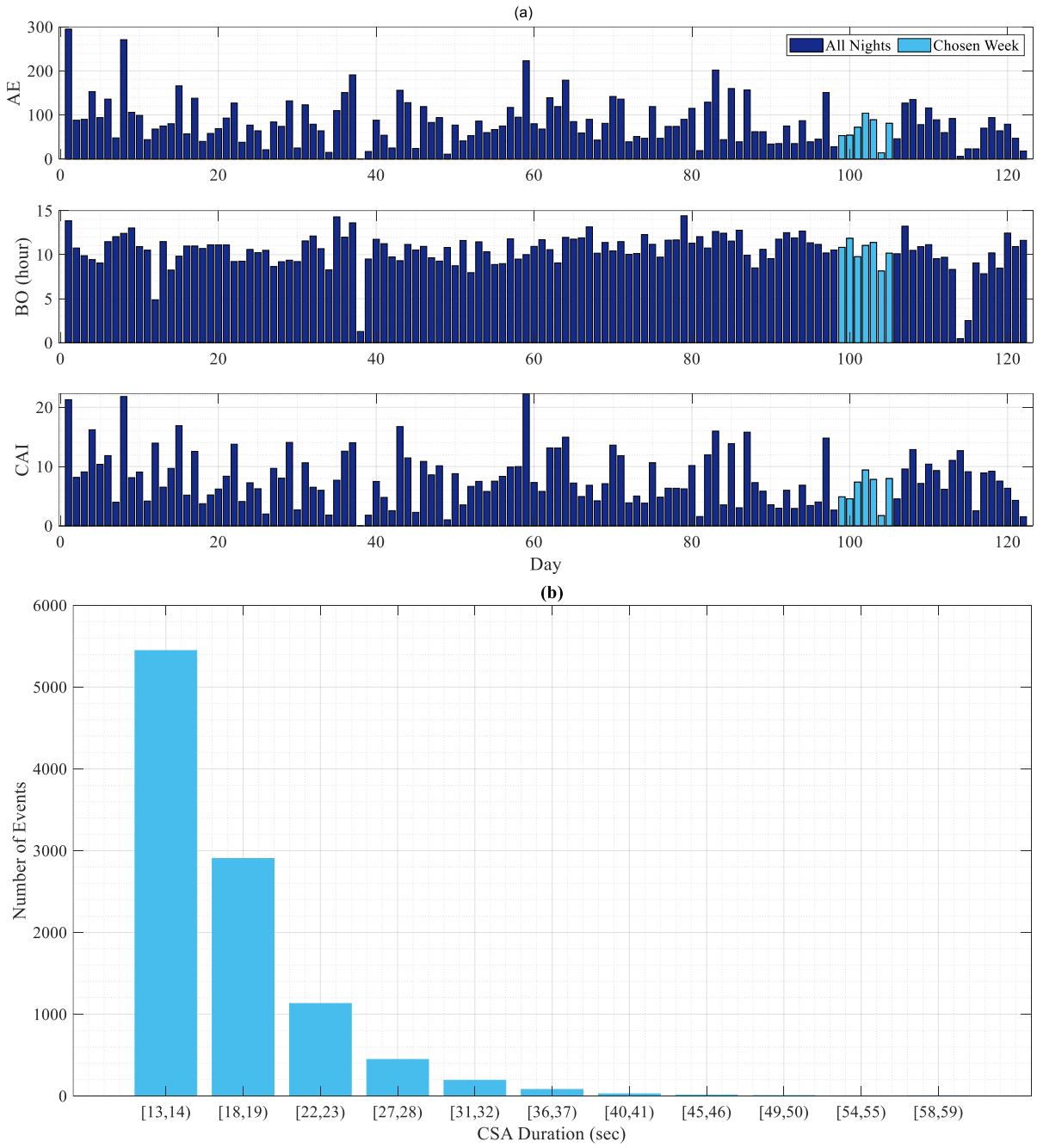


Figure A. 2 Subject 139, a) CAI per night and b) duration of detected events.

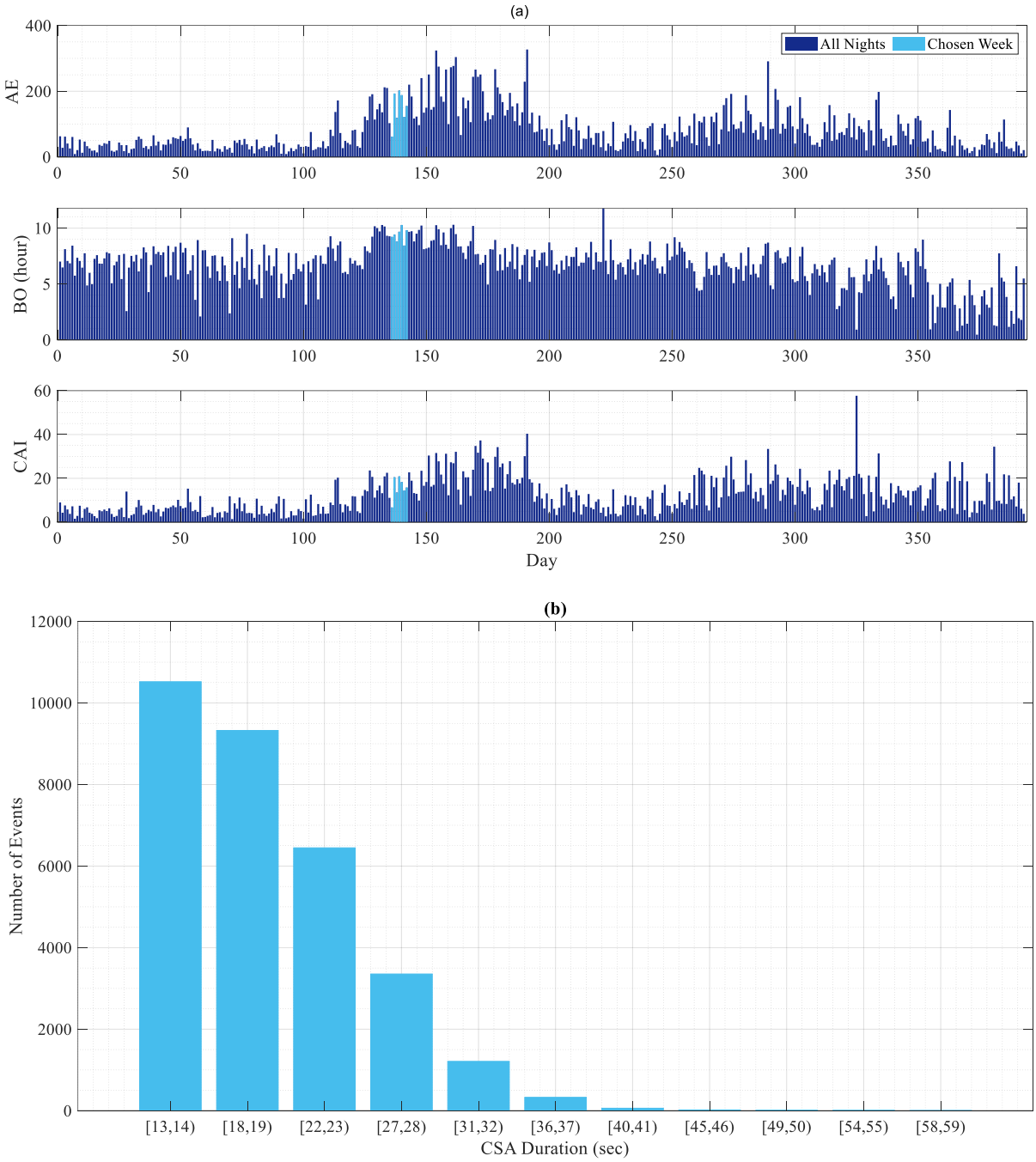


Figure A. 3 Subject 206, a) CAI per night and b) duration of detected events.

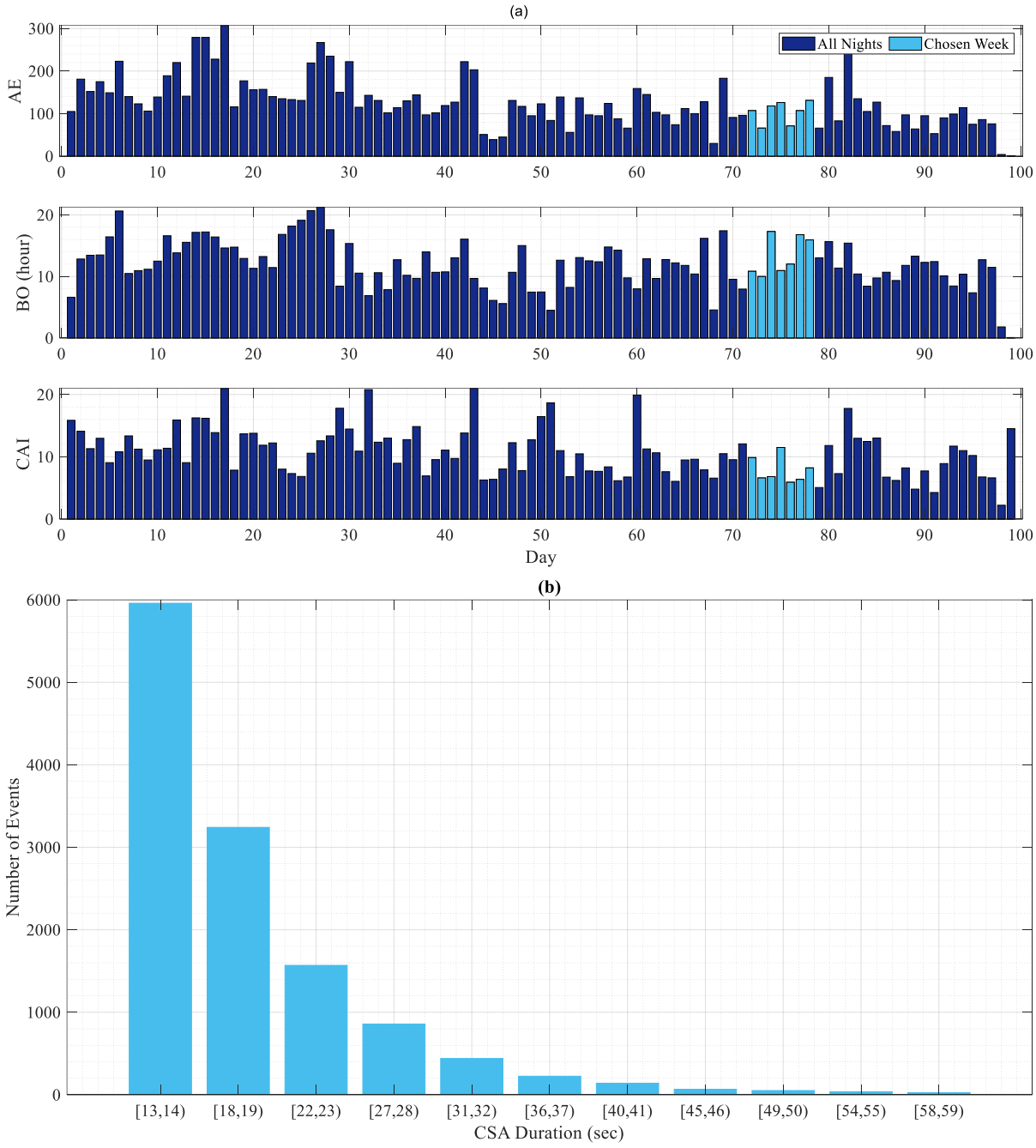


Figure A. 4 Subject 226, a) CAI per night and b) duration of detected events.

Figure A. 5 Subject 231, a) CAI per night and b) duration of detected events.

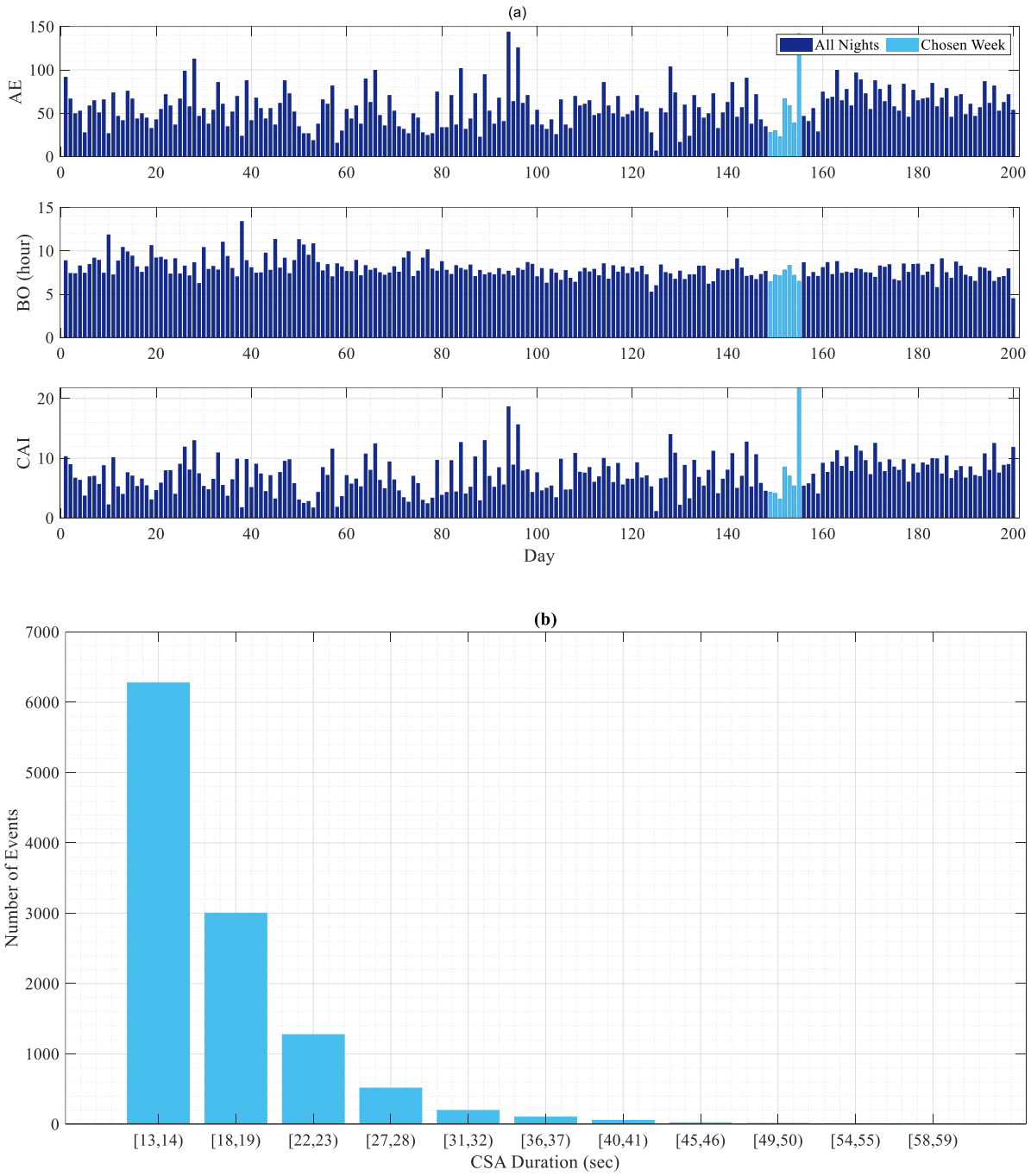


Figure A. 6 Subject 259, a) CAI per night and b) duration of detected events.

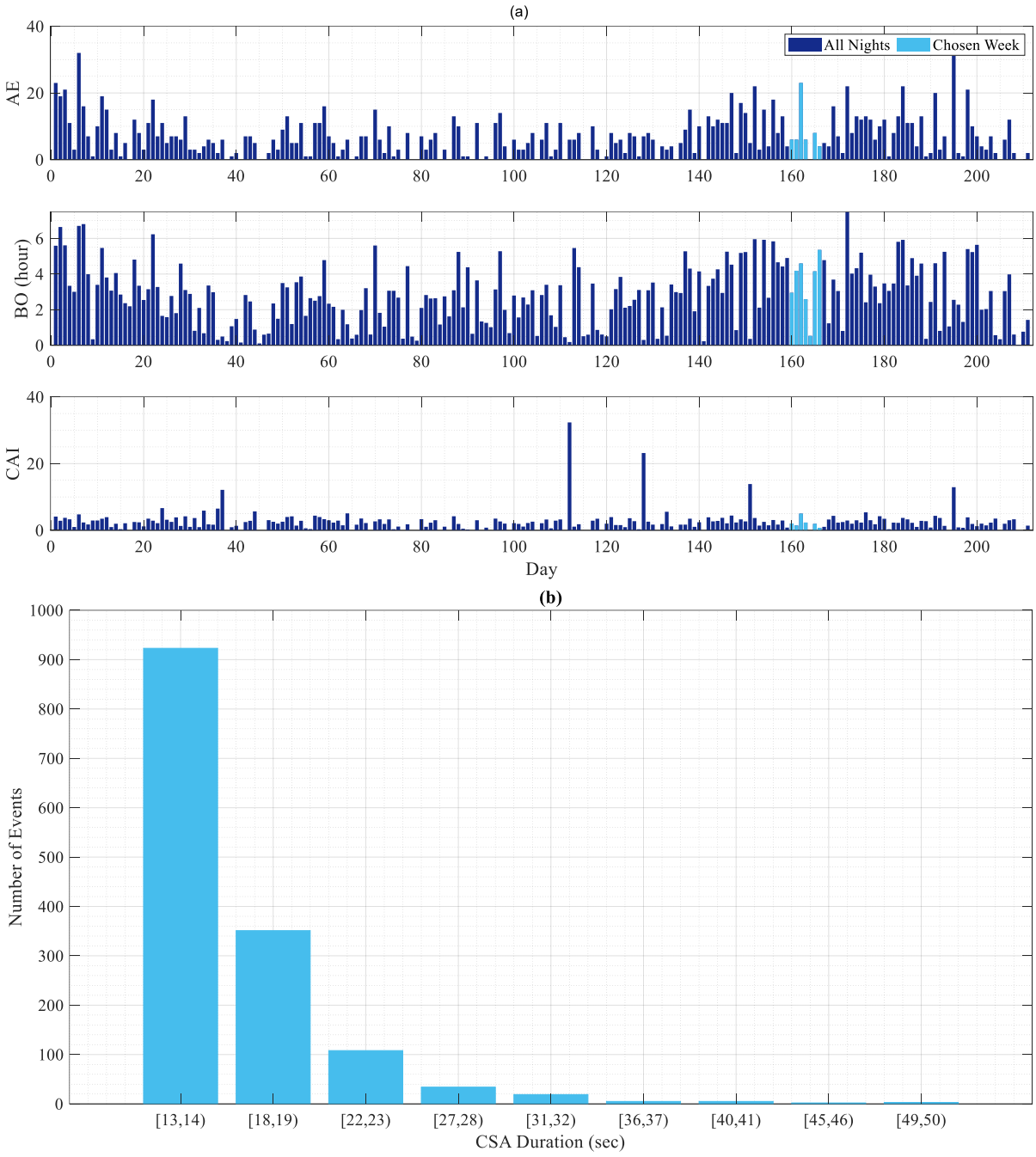


Figure A. 7 Subject 442, a) CAI per night and b) duration of detected events.

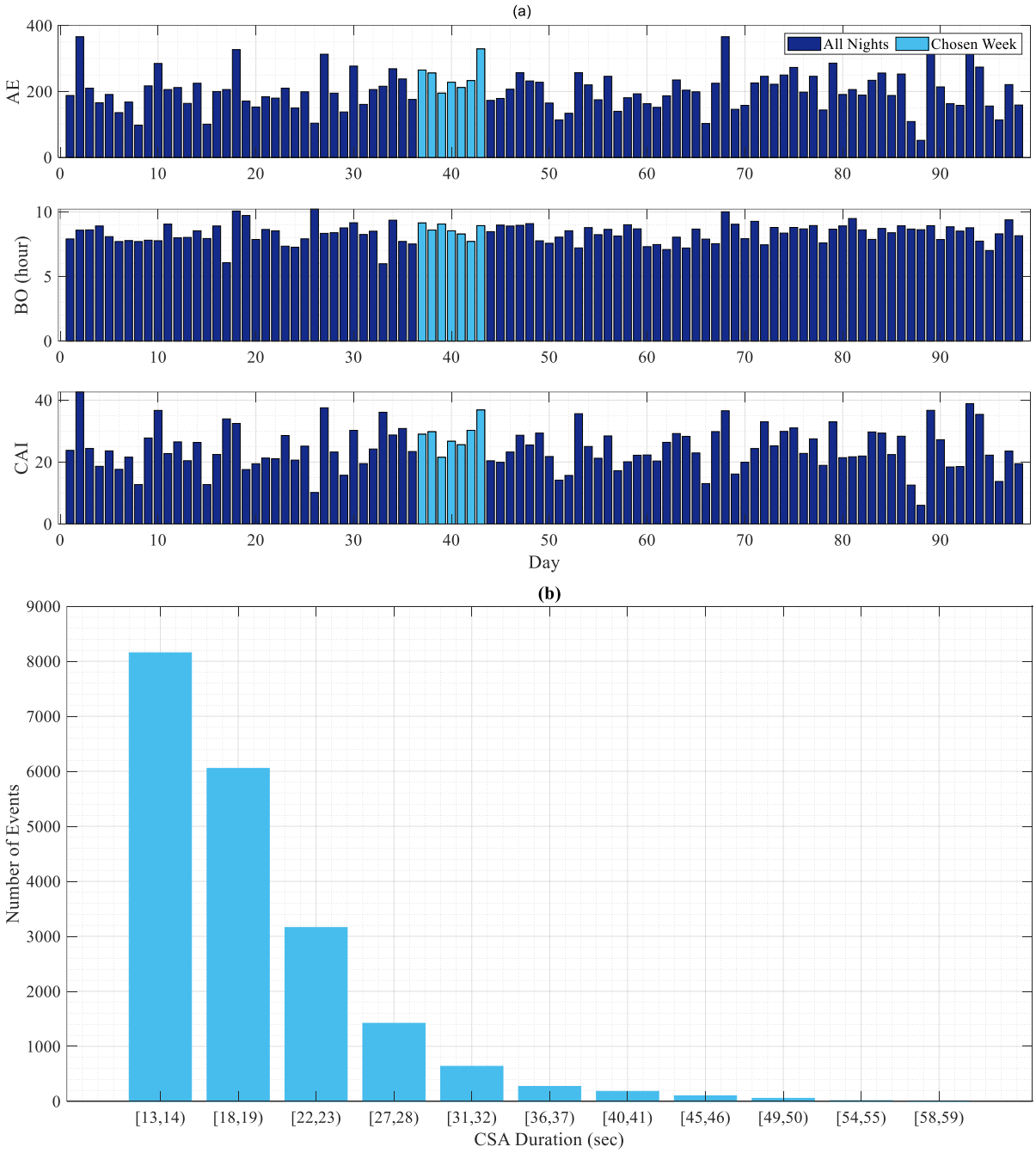


Figure A. 8 Subject 202, a) CAI per night and b) duration of detected events.

Appendix B

Cloud Processing of Bed Pressure Sensor Data to Detect Sleep Apnea Events

Hilda Azimi,
School of Electrical Engineering and Computer Science,
University of Ottawa
Ottawa, ON, Canada

Haoyang Liu
Systems and Computer Engineering; Carleton University
SAM3 Innovation Hub; Ottawa, Canada

Martin Bilodeau
School of Rehabilitation Sciences
University of Ottawa
Bruyere Research Institute
Ottawa, Canada

Bruce Wallace
Systems and Computer Engineering, Carleton University
Bruyere Research Institute
SAM[3] Innovation Hub; Ottawa, Canada

Martin Bouchard
School of Electrical Engineering and Computer Science,
University of Ottawa
Ottawa, ON, Canada

Rafik Goubran
Systems and Computer Engineering, Carleton University
Bruyere Research Institute
Ottawa, Canada

Frank Knoefel
Systems and Computer Engineering, Carleton University
Bruyere Research Institute
Faculty of Medicine, University of Ottawa
Ottawa, Canada

***Abstract*—Home-based monitoring of well-being through sensors that are either ambient or worn by the resident is a source of significant data about the resident. However, wide-scale use of such data creates a significant challenge in that the data has to be processed in a scalable way across many residents and residences. Cloud-based classification and machine learning tools, such as the IBM Watson cloud services, are now available as a potential deployment model. The assessment of sleep apnea is one example application that could benefit from residential assessment as a sleep lab assessment is costly and inconvenient. Home-based clinical information could allow physicians to triage patients for the costlier hospital-based testing. In this work, the performance of a sleep apnea algorithm is compared between the Matlab classification environment that was used for its development and IBM Watson cloud classification services. The work shows that similar but slightly different performance was achieved between the two systems with the IBM Watson tools having lower accuracy (mean 92.7% compared to 92.9%), higher precision (92.7% compared to 92.5%), lower recall (92.7% compared to 93.4%), and lower weighted F_1 measure (92.7% compared to 93.0%). This slight difference demonstrates the portability of Matlab lab results into a cloud solution, thus facilitating scalability.**

This work was supported by the Natural Sciences and Engineering Research Council (NSERC) Discovery Grant and ENGAGE Grant programs. IBM supported the work by providing access to their software applications and cloud services in addition to technical support and expertise. This work was supported by AGE-WELL NCE Inc., a member of the Networks of Centre of Excellence program, a Government of Canada program that funds partnerships between universities, industry, government and not for-profit organizations.

Keywords— central apnea detection; unobtrusive sensors; apnea classification; pressure sensor array; cloud processing.

I. Introduction

Continuing to live independently while maintaining an active lifestyle provides the best quality of life for aging adults and by extending independence, the demand for health care system resources can be delayed and ideally reduced. One mechanism that has been proposed to support independence is ongoing wellness monitoring through ambient and/or sensors worn by the aging adult [1]–[3].

Ambient sensing through the placement of sensors within the bed has been shown to be a significant potential application as the daily use of the bed for extended periods of time allows for ongoing longitudinal assessment. The deployment of a pressure-sensitive mat (PSM) under the mattress has been shown to enable the assessment of movements [4], [5] and specifically the transition from sitting to standing associated with a bed exit [6]. Different technologies are available for measuring and monitoring the respiration rate as a vital sign. Such technologies may be classed in contact-based and contactless [7]. The use of PSM as a contactless sensor has been extended to measure the respiration rate [8] since the pressure sensor is able to detect the small pressure changes associated with the breathing motion.

Sleep apnea is a breathing-related disorder experienced during sleep where the breathing rate is reduced or has stopped completely for intervals leading to disruptions of sleep [9]. This

is diagnosed typically in a sleep laboratory through the use of polysomnography (PSG) as observed by a technician. The PSG is then analyzed by a medical specialist to diagnose sleep apnea. Despite its high accuracy, PSG is costly and time-consuming. Moreover, the artificial and supervised environment of the sleep laboratory might change the nocturnal characteristics of sleep compared with the comfortable home environment [10]. One potential application for a PSM is apnea assessment within the home. Information about the frequency and duration of apnea episodes could assist physicians in triaging which patients require further sleep lab assessment. In recent years, in order to achieve this end and enhance sleep apnea diagnosis, besides PSM, other environmental sensors and ambient sensing have been taken into consideration. Non-contact radio-frequency sensors for measuring the bio-motion due to breathing and body movement [11], [12], Microsoft Kinect camera [13] and Digital video cameras for measuring the volume of air circulating into the lungs [14], are examples of these sensors. However, these devices can be only sensitive to certain positions of the body with a strong influence of body movements [15] or they may violate patients' privacy through the use of video.

The automation of the detection of central sleep apnea has recently been made possible through signal processing and classification of the respiration signal, derived from PSMs, showing the potential for this new sensing and assessment method [16]–[18]. For optimal assessment, a method is required to deploy PSM sensors and associated processing into the home. One architecture for this deployment is the placement of only the sensors in the home and using cloud processing services for analysis. This has the benefit of reducing the complexity and quantity of equipment required in each residence. The previous proof of concept work has shown the potential for this architecture to be implemented [19] and the ability for low cost embedded computing platforms to provide the in-residence processing [20].

To achieve cloud-based processing for classification of the sensor signals to detect apnea events, the algorithms previously created and validated within Matlab [16], [17] have to be ported to a cloud-based classification system such as the IBM Watson [19] platform. This represents a complete change in the machine learning methodology including the training of the classifier and subsequent use of the classifier. This paper specifically focuses on the comparison of performance between the Matlab and IBM Watson implementations for sleep apnea classification using identical datasets for each of the two systems.

II. Methods

A. Data Collection

Fig. 1 and 2 show two alternative models for PSM data collection. In Fig. 1, the data is collected locally in a device near the bed and recorded onto media that can then be periodically collected. Fig. 2 instead shows an alternative model that allows for the data to be streamed in real-time and stored within the cloud [20]. This is the target architecture for home monitoring. In this work, the chosen datasets were captured a priori, allowing them to be used in two classification methods in parallel. They were collected using the model shown in Fig. 1, but the methodology is intended to allow the Fig. 2 model when real-time capture is used [20]. Each set of sensor arrays has 72 fiber optic pressure sensors, in 8x9 evenly spaced array 10 cm apart. PSM is placed under the bed mattress to sense the area from about the patient's head to their hips.

The project had ethics approval from the Carleton University and Bruyère Continuing Care Research Ethics boards. Nine volunteers from different communities participated in the study. They commonly were:

- Sixty-five years of age or more;
- Community-dwelling older adults living in subsidized seniors housing;
- Discharged from the Geriatric Rehabilitation Unit at Élisabeth Bruyère Hospital.

The monitoring time was 8-12 months. Except for one visit every month, the rest of the data collection was not supervised. One week of data collected from each subject (in total 63 days) was randomly selected and employed in this study.

Start points and endpoints of events suspected to be apneic were marked manually by a technician, being aware of the American Academy of Sleep Medicine (AASM) rules [21] and trained to detect CSA events from PSM data. The definition used was a complete cessation of respiratory movements captured by PSM for a duration of at least 10 s.

B. Pre-processing and feature extraction

The resulting processing models are shown in Fig. 3 where Fig. 3A shows the model for processing using the offline Matlab processing method [16], while Fig. 3B shows the processing model of the IBM cloud tools [19]. The Matlab based model is limited to the post-processing of data that has been captured previously. The IBM cloud tools model enables future real-time processing through scalable cloud infrastructure [19], [20] where the processing and feature extraction is performed with Python-based software. In this work, to allow direct comparison of the classifier performance and function, both classifiers were provided the same datasets of labeled features generated with Matlab code.

The processing of the sleep data within the Matlab environment followed the flow process shown in Fig. 3A where the previously collected sleep data (CSV format) was loaded into a Matlab Data structure. The data were concatenated on a day-by-day basis for 24 hours, from noon to noon for sleep assessment and subsequently processed for various filtering, noise removal and ultimately feature extraction previously reported in [16] as follows:

Occupancy Extraction to detect and remove time periods that the bed was not occupied based on the algorithm in [22] and therefore discard irrelevant data.

Bandpass filtering with a passband range of (0.07-0.8) Hz corresponding to extreme breathing conditions of (4-48) breath per minute (bpm) to filter each of the 72 sensors' signals.

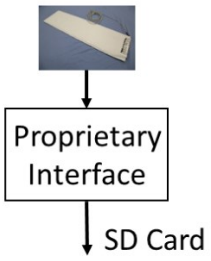


Fig. 1. Matlab Data Capture and recording flow diagram using capture to SD card and subsequent manual collection.

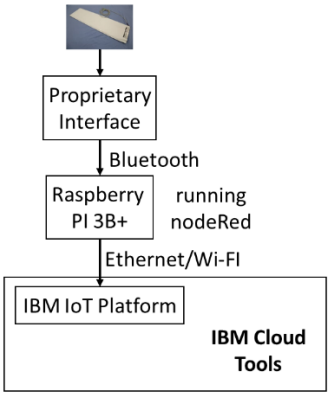


Fig. 2. IBM Cloud Tools Data Capture and recording flow diagram using the ability to flow data into the cloud in real-time [20].

It should be noted that the typical respiratory rate for a healthy adult at rest is 12 to 20 bpm [7]. However in this work in order to prevent possible loss of information, the complete range of breathing rates (i.e., 4-48 bpm) was carefully considered for signal filtering [23].

Combining and concatenating signals to weight all 72 signals from a PSM based on the quality of their information and then combine the weighted signals to generate a single output signal *y* with better signal quality. A comprehensive explanation of the applied combining method can be found in [8].

Signal normalization to provide a more accurate estimate of respiratory depth and volume. Body movements and different body postures might affect PSM signal amplitude [24]. Body movements cause large fluctuations in *y*. These fluctuations were detected as a value that is more than three scaled median absolute deviations (MAD) away from the median of *y* [16]. After detecting movements, each part of the signal between any two detected movements, *z,* was normalized in two consecutive steps. First, the average of *z* was removed. Next, *z* was divided by the maximum absolute value of *z*, excluding the highest 5% and lowest 5% of the values.

The output of the pre-processing pipeline, *y*, is a single normalized signal per day. *y* was segmented into smaller 10-s sequences with a 50% overlap classification of classes, i.e., not-

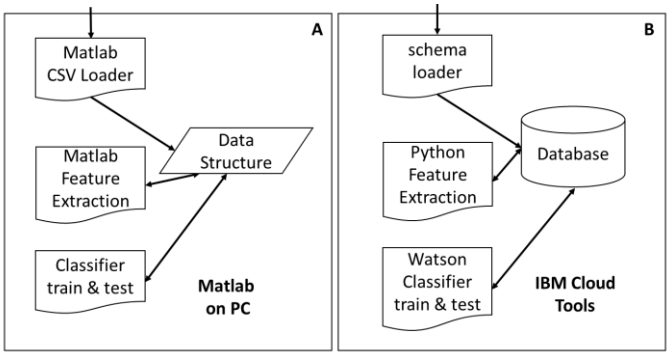


Fig. 3. Comparison of a processing model for Matlab and IBM. A: Matlab: data stored within a data structure while Matlab code performs feature extraction and classification training/test. B: IBM cloud tools: data stored within a database and this is used by Python code to perform feature extraction and Watson used for classification training/test.

apneic "N" or apneic "A". "A" labeled segments are those that have at least 50% of their length falling within a start-to-end time of apnea events. Unlike [16], those segments that had overlap with detected movements from the preprocessing step were classified as outliers and they were completely removed from the whole data-set to simplify the procedure and reduce the initial complexity of the models being tested as the goal of this work was the comparison of the two methods.

Thirty-four (34) features from the time and frequency domain were extracted from the remaining segments to be fed to Linear SVM [16]. The resulting features were then added to the data structure. Matlab classification tools were then used to train and test the classifier to provide a direct comparison of the performance with the IBM Watson environment.

The processing within the IBM Watson cloud tools environment requires fundamental changes to the processing steps as these tools do not support simple data structures similar to Matlab and instead make use of a database for storage. In our work, a DB2 relational database was chosen as it is the one that has the appropriate features for the work including enabling real-time processing in the cloud in the future. This resulted in the need to create a processing script to load the collected sleep data into the database according to the chosen schema design.

Normally in the IBM Watson cloud tool model, the processing of the mat sleep data for filtering, noise removal, and feature extraction would have to be ported to Python but otherwise performing the same mathematical processing as this is the language supported in the cloud tools along with modifications to reflect the need to interface to the database. In this work, as we wanted to focus on the differences, if any, between the classifier training and performance results, the Matlab features and labeled datasets were exported and then loaded into the DB2 database with an appropriate schema. This allowed the direct use of the IBM Watson machine learning system for the training and testing of the resulting classifiers.

C. Classification

The number of observations in two classes of "N" and "A" is imbalanced. In comparison to class "A", class "N" is significantly over-represented in the dataset. A simple way to balance the

datasets and consequently prevent "accuracy paradox", where the accuracy is only reflecting the underlying class distribution and not the actual performance is resampling of the classes; i.e. oversampling the minority class or undersampling the majority class [25]. Here, we randomly undersampled class "N" instances to the number of instances in class "A". To examine whether the models in each tool would be generalized and not data-dependent, the undersampling of class "N" instances was repeated 10 times, resulting in 10 different datasets. Each of these datasets containing 30092 observations (15046 instances from each class) was loaded into the respective classification tools leading to a labeled data structure of features within the Matlab environment and a DB2 database containing the same labeled feature set. The results in [16] had shown that Linear SVM provided the best performance for the Matlab case and given that Linear SVM was available in both environments, it was chosen as the comparative classification method.

The training method for both of the classifiers was chosen to match as closely as possible and specifically for both systems. For both environments, the Linear SVM kernel was chosen and the cost for classification errors was equal for false negative and false positive errors. Both used a training and testing model where 80% of the data was used to train, while 20% was used to test the resulting classifiers. Within the Matlab training model, this is explicitly described as 5-fold validation where the dataset is randomly split into 5 groups (20% each) and the training/test process is performed using each of the groups as the test set and reporting the average results for these 5 folds of training. The IBM Watson toolset only describes and performs the test/train step once. One subtle difference in the two training and test methods is that they each control the random split of the data into training and test groups causing a difference between the platforms in the data used for training and subsequent test. Although this reporting of performance during classifier design is different, should a chosen model be exported for subsequent use, it is normally based on re-training using 100% of the dataset leading to both exported models having the same training set.

III. Results

The comparative test results for the Matlab and IBM cloud tools are shown in Table I and II respectively in terms of accuracy (1), recall (2), precision (3) and F_1 measure (4),

$$\text{Accuracy}=\frac{\text{TP}+\text{TN}}{(\text{TP}+\text{TN}+\text{FP}+\text{FN})}\times 100\% \quad (1)$$

$$\text{Recall}=\frac{\text{TP}}{(\text{TP}+\text{FN})}\times 100\% \quad (2)$$

$$\text{Precision}=\frac{\text{TP}}{(\text{TP}+\text{FP})}\times 100\% \quad (3)$$

$$F_1=\frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \quad (4)$$

where TP, TN, FP, and FN are abbreviations for true positive, true negative, false positive and false negative respectively.

The results presented in both tables are consistent across the 10 datasets for the two methods as there is limited variance in each of the performance measures within the tables.

Both tables illustrate the results for a Linear SVM classifier within each of the respective tools and although the same classifier was chosen, they are based on the two completely different implementations of the training and evaluation algorithms including the randomized segmentation of the dataset. The result is that the IBM tool chain reports a slightly lower accuracy across the datasets with the mean dropping from 92.9% to 92.7 while the weighted precision has corresponding increased to 92.7% from 92.5%. The weighted recall performance for the IBM tools is slightly lower at 92.7% from 94.3% for Matlab with the combined F_1 measure for IBM being slightly lower at 92.7% compared to 93.0%. Overall, the results for the two implementations are remarkably consistent.

IV. Discussion

This work has shown the potential for sleep apnea detection algorithms to be implemented within a cloud-capable data capture, noise and other filtering processing, feature extraction, and subsequent machine learning training and classification. The results for Matlab based models have been shown to be replicable within the IBM cloud tool environment. This provides confidence that algorithms developed with a Matlab based research environment where these tools are used for their high flexibility to evaluate potential algorithms and methods can be ported into a cloud-based framework with limited change in performance.

One challenge noted in the work were differences between the classifiers supported within the Matlab toolset and those supported within IBM Watson. Specifically, we noted that the IBM and Matlab environments had differing names for some classifiers although all methods were supported by both. Within the methods there were differences in the configuration

TABLE I. MATLAB PERFORMANCE RESULTS FOR EACH OF THE 10 DATASETS SHOWING THE PERFORMANCE ACHIEVED FOR APNEA DETECTION USING THE IBM WATSON CLASSIFICATION TRAINING AND TEST PROCESS.

DATASET#	ACCURACY	WEIGHTED PRECISION	WEIGHTED RECALL	WEIGHTED F_1 MEASURE
1	93.1%	92.6%	93.6%	93.1%
2	92.9%	92.3%	93.6%	92.9%
3	93.0%	92.6%	93.6%	93.1%
4	93.0%	92.7%	93.4%	93.0%
5	92.9%	92.4%	93.5%	92.9%
6	92.9%	92.5%	93.4%	93.0%
7	92.9%	92.4%	93.4%	92.9%
8	92.7%	92.3%	93.2%	92.8%
9	93.0%	92.7%	93.4%	93.0%
10	92.9%	92.5%	93.4%	93.0%
MEAN	92.9%	92.5%	93.4%	93.0%

TABLE II. IBM CLOUD TOOLS PERFORMANCE RESULTS FOR EACH OF THE 10 DATASETS SHOWING THE PERFORMANCE ACHIEVED FOR APNEA DETECTION USING THE IBM WATSON CLASSIFICATION TRAINING AND TEST PROCESS.

DATASET#	ACCURACY	WEIGHTED PRECISION	WEIGHTED RECALL	WEIGHTED F_1 MEASURE
1	92.7%	92.7%	92.7%	92.7%
2	92.8%	92.8%	92.8%	92.8%
3	92.9%	92.9%	92.9%	92.9%
4	92.8%	92.8%	92.8%	92.8%
5	92.7%	92.7%	92.7%	92.7%
6	92.7%	92.7%	92.7%	92.7%
7	92.5%	92.5%	92.5%	92.5%
8	92.8%	92.8%	92.8%	92.8%
9	92.6%	92.6%	92.6%	92.6%
10	92.9%	92.9%	92.9%	92.9%
MEAN	92.7%	92.7%	92.7%	92.7%

parameters provided and the default values for these, which could introduce differences during migration to the cloud using the IBM Watson infrastructure. For this work, Linear SVM was chosen as it was supported in both and was known to provide good performance. We ensured the training and test method was as similar as possible and the one difference has been noted in the paper.

V. Conclusion & Future Work

The cloud-based framework, such as the IBM Watson environment reported in this work, has been shown to provide an implementation model for well-being assessment algorithms such as sleep apnea. The portability of lab bench Matlab results to cloud based implementation models such as IBM Watson provides scalable performance models that can be scaled to support many users, each living independently within their own residence or even within care settings. It can also now allow researchers with the ability to explore larger data sets that are beyond desktop computation capability.

Future explorations will include the cloud-based model enabling many new exploration options for well-being monitoring and assessment through both the implementation of real-time processing of complex sensor data from many residences and the processing of historically captured datasets of well-being data through new tools that allow for much more extensive processing than could have been accomplished without the scale of cloud processing.

References

[1] P. Chaurasia *et al.*, "Modelling assistive technology adoption for people with dementia," *Journal of Biomedical Informatics*, vol. 63, pp. 235–248, Oct. 2016.

[2] H. Megges, S. D. Freiesleben, N. Jankowski, B. Haas, and O. Peters, "Technology for home dementia care: A prototype locating system put to the test," *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, vol. 3, no. 3, pp. 332–338, Sep. 2017.

[3] G. Gibson, C. Dickinson, K. Brittain, and L. Robinson, "The everyday use of assistive technology by people with dementia and their family carers: a qualitative study," *BMC Geriatrics*, vol. 15, no. 1, p. 89, Jul. 2015.

[4] D. I. Townsend, M. Holtzman, R. Goubran, M. Frize, and F. Knoefel, "Measurement of Torso Movement With Delay Mapping Using an Unobtrusive Pressure-Sensor Array," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 5, pp. 1751–1760, May 2011.

[5] M. H. Jones, R. Goubran, and F. Knoefel, "Identifying Movement Onset Times for a Bed-Based Pressure Sensor Array," in *IEEE International Workshop on Medical Measurement and Applications, 2006. MeMea 2006.*, Benevento, Italy, 2006, pp. 111–114.

[6] Zhaofen Ren *et al.*, "Analyzing center of pressure progression during bed exits," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Chicago, IL, 2014, pp. 1786–1789.

[7] C. Massaroni, A. Nicolò, D. Lo Presti, M. Sacchetti, S. Silvestri, and E. Schena, "Contact-Based Methods for Measuring Respiratory Rate," *Sensors*, vol. 19, no. 4, p. 908, Feb. 2019, doi: 10.3390/s19040908.

[8] H. Azimi, S. Soleimani Gilakjani, M. Bouchard, S. Bennett, R. A. Goubran, and F. Knoefel, "Breathing signal combining for respiration rate estimation in smart beds," in *2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, Rochester, MN, USA, 2017, pp. 303–307.

[9] W. W. Flemons *et al.*, "Home Diagnosis of Sleep Apnea: A Systematic Review of the Literature: An Evidence Review Cosponsored by the American Academy of Sleep Medicine, the American College of Chest Physicians, and the American Thoracic Society," *CHEST*, vol. 124, no. 4, pp. 1543–1579, Oct. 2003.

[10] S. Doering, J. A. Boeckmann, S. Hugger, and P. Young, "Ambulatory polysomnography for the assessment of sleep bruxism," *Journal of Oral Rehabilitation*, vol. 35, no. 8, pp. 572–576, 2008.

[11] A. Zaffaroni, P. de Chazal, C. Heneghan, P. Boyle, P. R. Mppm, and W. T. McNicholas, "SleepMinder: An innovative contact-free device for the estimation of the apnoea-hypopnoea index," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2009, pp. 7091–9094.

[12] R. Nandakumar, S. Gollakota, and N. Watson, "Contactless Sleep Apnea Detection on Smartphones," in *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services - MobiSys '15*, Florence, Italy, 2015, pp. 45–57.

[13] C. Yang, G. Cheung, V. Stankovic, K. Chan, and N. Ono, "Sleep Apnea Detection via Depth Video and Audio Feature Learning," *IEEE Trans. Multimedia*, vol. 19, no. 4, pp. 822–835, Apr. 2017.

[14] J. Abad *et al.*, "Automatic Video Analysis for Obstructive Sleep Apnea Diagnosis," *Sleep*, vol. 39, no. 8, pp. 1507–1515, Aug. 2016.

[15] "Recent development of respiratory rate measurement technologies - IOPscience." [Online]. Available: https://iopscience.iop.org/article/10.1088/1361-6579/ab299e. [Accessed: 16-Mar-2020].

[16] H. Azimi, M. Bouchard, R. A. Goubran, and F. Knoefel, "Apnea Event Detection Methodology using Pressure Sensors," in *2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, Istanbul, Turkey, 2019, pp. 1–6.

[17] H. Azimi and F. Knoefel, "Unobtrusive Screening of Central Sleep Apnea from Pressure Sensors Measurements: A Patient-Specific Longitudinal Study," *in review IEEE Transactions on Instrumentation and Measurement (TIM)*, p. 15, 2020.

[18] H. Azimi, S. S. Gilakjani, M. Bouchard, R. A. Goubran, and F. Knoefel, "Automatic apnea-hypopnea events detection using an alternative sensor," in *2018 IEEE Sensors Applications Symposium (SAS)*, Seoul, 2018, pp. 1–5.

[19] R. B. Wallace, H. Liu, R. Goubran, M. Bilodeau, and F. Knoefel, "Cloud based Artificial Intelligence Processing of Ambient Home Sensors," in *International Society for Professional Innovation Management (ISPIM)*, 2019, pp. 1–12.

[20] R. B. Wallace, H. Liu, R. Goubran, M. Bilodeau, and F. Knoefel, "Streaming Pressure Sensor Data for Cloud Processing," in *2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, Istanbul, Turkey, 2019, pp. 1–5.

[21] R. B. Berry *et al.*, "Rules for scoring respiratory events in sleep: update of the 2007 AASM Manual for the Scoring of Sleep and Associated

Events. Deliberations of the Sleep Apnea Definitions Task Force of the American Academy of Sleep Medicine," *J Clin Sleep Med*, vol. 8, no. 5, pp. 597–619, Oct. 2012, doi: 10.5664/jcsm.2172.

[22] S. Soleimani, "Long-Term Sleep Assessment by Unobtrusive Pressure Sensor Arrays," Thesis, Université d'Ottawa / University of Ottawa, 2018.

[23] M. Holtzman, R. Goubran, and F. Knoefel, "Maximal ratio combining for respiratory effort extraction from pressure sensor arrays," in *2011 IEEE International Symposium on Medical Measurements and Applications*, Bari, Italy, 2011, pp. 88–92, doi: 10.1109/MeMeA.2011.5966674.

[24] D. Townsend, M. Holtzman, R. Goubran, M. Frize, and F. Knoefel, "Relative Thresholding With Under-Mattress Pressure Sensors to Detect Central Apnea," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 10, pp. 3281–3289, Oct. 2011.

[25] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *1*, vol. 16, pp. 321–357, Jun. 2002.