POLITECNICO DI TORINO

DIPARTIMENTO DI INGEGNERIA MECCANICA E AEROSPAZIALE

Corso di Laurea Magistrale in Ingegneria Biomedica

Sleep apnea events recognition based on polysomnographic signals recordings: a machine learning approach

Supervisor:
Prof. Molinari Filippo
External supervisor (SUPSI):
Prof. Puiatti Alessandro

Candidate:
La Porta Nicolò

ACADEMIC YEAR 2021/2022

Contents

List of Figures

List of Tables

Acronyms

ΔI Delta Index.

AASM American Academy of Sleep Medicine.

ACC diagnostic accurracy.

AHI Apnea-Hypopnea Index.

ANN Artificial Neural Network.

APAP Auto-titrating Positive Airway Pressure.

ApEn approximate entropy.

ASV Adaptive Servo Ventilation.

AUC Area under the ROC Curve.

BPAP Bi-level Positive Airway Pressure.

CPAP Continuous Positive Airway Pressure.

CSA Central Sleep Apnea.

DWT Discrete Wavelet transform.

ECG electrocardiogram.

EEG electroencephalogram.

EMG electromyogram.

EOG electrooculogram.

EPAP expiratory positive airway pressure.

ESS Epworth sleepiness scale.

ICSD International classification of sleep disorders.

IPAP inhalation positive airway pressure.

LM Leg Movement.

LMA Leg Movement Arousal.

MAAs Mandibular Advancement Appliances.

MANOVA Multivariate ANalysis Of VAriance.

MRA Multiresolution Analysis.

MSA Mixed Sleep Apnea.

NEPAP Nasal Expiratory Positive Airway Pressure.

NPV negative predictive value.

NREM Non Rapid Eye Movement.

NSRR National Sleep Research Resource.

OAs Oral Appliances.

ODs Intraoral Devices.

OSA Obstructive Sleep Apnea.

PAP Positive Airway Pressure.

PCA Principal Component Analysis.

PPV positive predictive value.

PSD Power Spectral Density.

PSG Polysomnography.

PT Positional Therapy.

REM Rapid Eye Movement.

RIP respiratory inductance plethysmography.

RLS Restless Leg Syndrome.

ROC Receiver Operating Characteristic.

SAHS Sleep Apnea-Hypopnea Syndrome.

SENS sensitivity.

SPEC specificity.

SPLs Soft Palate Lifters.

STFT Short Time Fourier Transform.

TFR Time-Frequency Representation.

TMJ Temporo-Mandibular Joint.

TRDs Tongue Retaining Devices.

TST total sleep time.

WSC Wisconsin Sleep Cohort.

WSS Wide Sense Stationary.

WT Wavelet Transform.

Abstract

Sleep diseases are one of the major causes of physical and psychological problems for workers, resulting in high financial losses in terms of direct, indirect, related and intangible costs. This thesis will focus on a particular type of sleep-related disorder, the Sleep Apnea-Hypopnea Syndrome (SAHS) which causes numerous involuntary respiratory pauses during the night ("apneic events")leading to a drop of blood oxygen saturation with consequent subject awakening and reduction of sleep quality. There are mainly three forms of sleep apnea: 1)The Obstructive Sleep Apnea (OSA), which is characterized by an upper airway airflow reduction caused by the collapse of the soft tissues in the back of the throat and the tongue; 2) The Central Sleep Apnea (CSA), which is characterized by the absence of respiratory effort and, thus, the absence of airflow; 3) The Mixed Sleep Apnea (MSA), which is a combination of the previous two. Respectively, they represent the 84%, the 0.4% and the 15% of the total cases in U.S. and Europe. The gold standard diagnosis system for sleep apnea is Polysomnography (PSG) that is conducted in special sleep units during an entire night. Numerous physiological signals are recorded during the PSG, and then analyzed by experts that divide the entire signal into epochs and assign to each epoch a sleep stage and an apneic score to specify the subject's condition during sleep. Manual revision of the PSG recordings requires a considerable amount of time and effort, it is not error free, and together with the sleep units saturation elevates the economic burden of SAHS.

In this perspective, a system based on artificial intelligence could offer support to the experts. The main purpose of the present work is to develop an automatic sleep apnea detection algorithm based on only a subset of the signals recorded during a PSG . In particular, it includes only signals from non-invasive sensors that are: EEGs (C3-M2 and O1-M2), EOGs, ECG, SpO2, thoracic volume, audio recording and body position.

The dataset used for the present work is the Wisconsin Sleep Cohort DB. It includes more than 2'500 subjects' PSG (with an average length of 6.13h $\pm$ 0.97h) and 3 textual files with useful information. The dataset was cleaned removing potential outliers and a set of 130 features was computed for each subject using the Matlab® working environment.

Then, a MANOVA was performed in order to understand how many clusters could be distinguished among the 4 considered (normal, CSA, OSA, MSA), based on the data variability. Only 3 classes were distinguishable, since the MSA is very less represented and with features similar to those of OSA and CSA, therefore it was excluded from the dataset.

Finally, different machine learning models were trained, such as Decision Trees, Naïve Bayes, Discriminant Analysis, SVM and K-NN, and underwent hyperparameters optimization according to built-in MatLab routines.

In conclusion, among the models trained, the most efficient classifier manages to distinguish between normal sleep epochs and apnea epochs, while the classification performances deteriorated when hypopnea was considered since this condition is in the middle of the previous two. Moreover, a cascading classifier was tested to distinguish between different apnea forms after separating normal epochs from apnea epochs.

Chapter 1

Introduction

1.1 Physiological and clinical background

Sleep is a very important physiological condition for the whole body as it relaxes the muscles, allows cell turnover and tissue regeneration and especially strengthens the CNS since during sleep the maximum of neuronal plasticity is reached.
Sleep is fundamental for everybody, especially for those who carry out purely intellectual activities, allowing the formation of memories.

Subjects who have sleep disorders go to places called *sleep clinics* where they undergo analysis of the gradual transition from waking to resting condition, the maintenance of the condition of rest, and the return from the condition of rest to the condition of wakefulness.
In such sleep clinics, patients are sensorized by expert technicians in order to record the activity of most of the physiological systems in the form of biomedical signals successively used to analyze manifestations of sleep disorders.

As reported in Roebuck's review [Roebuck et al., 2013], sleep disorders were divided into eight categories by the International classification of sleep disorders (ICSD):

1. Insomnias: difficulty falling asleep, difficulty staying asleep, early awakening, or poor sleep quality;

2. Sleep-related breathing disorders;

3. Hypersomnias of central origin not due to a circadian rhythm sleep disorder, sleep-related breathing disorder, or other cause of disturbed nocturnal sleep;

4. Circadian rhythm sleep disorders;

5. Parasomnias: disorders that intrude into the sleep process and are manifestations of central nervous system activation;

6. Sleep-related movement disorders;

7. Isolated symptoms, apparent normal variants, and unresolved issues;

8. Other sleep disorders.

This thesis will discuss a particular type of sleep-related breathing disorder, the Sleep Apnea-Hypopnea Syndrome (SAHS), whose effects influence physical and psychological aspects of subjects' life and represent an important financial impact on society.

1.2 Costs of sleep disorders

In their work, Hossain and Shapiro [Hossain and Shapiro, 2002] identified four categories of the costs due to sleep disorders, which were reported in Roebuck's work [Roebuck et al., 2013] as:

- Direct costs: such as visits to health care professionals, diagnostic tests, treatments, and hospital services;

- Indirect costs: such as ambulatory care, work absenteeism, disability, reduction or loss of productivity, industrial and motor vehicle accidents, hospitalization, increase in medical costs comorbid conditions, and alcohol consumption;

- Related costs: such as accident-related property damages, travel costs to health care providers, and cost to family of additional care;

- Intangible costs: such as decreased quality of life, impaired schooling, and loss of activities of daily living.

The main aspects causing causes an elevation in the economic costs associated with the diagnosis of SAHS are:

- Manual revision of the PSG recording which requires a considerable amount of clinicians' time and effort;

- Waste of the personnel resources;

- Saturation of sleep units.

According to Huyett and Bhattacharyya, 2021, sleep disorders afflict 13.6 ± 0.6 million adults in the US. These subjects were found to have increased utilization of office visits (16.3 ± 0.8 vs 8.7 ± 0.3, P < .001), emergency room visits (0.52 ± 0.03 vs 0.37 ± 0.02, P < .001), and prescriptions (39.7 ± 1.2 vs 21.9 ± 0.4, P < .001) with respect to those without sleep disorders.
The additional incremental health care expenses for those with sleep disorders were increased in all examined measures: total health care expense ($6,975 ± $800, P < .001), total office-based expenditures ($1,694 ± $277, P < .001), total prescription expenditures ($2,574 ± $364, P < .001), and total self-expenditures for prescriptions ($195 ± $32, P < .001).
It is clear that sleep disorders are associated with significantly higher rates of health care utilization and expenditures.
In 2021, the overall incremental health care costs of sleep disorders in the US represents approximately $94.9 billions.
According to Streatfeild et al., 2021, the estimated overall cost of sleep disorders in Australia in 2019–2020 (population: 25.5 million) was $35.4 billions divided in $25.4 billions of nonfinancial costs and $10.0 billions of financial costs which comprise health system costs for $0.7 billions, productivity losses for $7.7 billions, informal care for $0.2 billions, other mainly non-medical accident costs for $0.4 billions and deadweight losses for $1.0 billions. For moderate to severe OSA syndrome, insomnia unrelated to other conditions and Restless Leg Syndrome (RLS), financial costs represented $16,717, $21,982, and $16,624 per adult with the condition for the year, respectively.

1.3 Sleep Apnea-Hypopnea Syndrome (SAHS)

Patients affected by SAHS experience numerous involuntary respiratory pauses during the night. The duration of these nocturnal respiratory events, called *apneic events*, differs from patient to patient and in order to be considered clinically significant must last from 10 seconds to 2 minutes. The common duration of the apneic event, though, is usually about 20 to 40 seconds.
Such apneic events lead to a drop in the oxygen saturation levels in arterial blood proportional to the causing airflow reduction which triggers an autonomic response that increases the subject's alertness level often causing neurophysiological awakening which breaks up the normal sleep structure interrupting a refreshing rest [Shokoueinejad et al., 2017].
Common symptoms during sleep are:

- Snoring (usually loud).
- Gasping or choking sounds.

- Breathing pauses observed by someone watching you sleep.
- Sudden or jerky body movements.
- Restless tossing and turning.
- Frequent awakenings from sleep.

Common symptoms while awake are:

- Daytime sleepiness.
- Morning headache.
- Dry or sore throat in the morning.
- Fatigue or tiredness throughout the day.
- Personality changes, such as mood swings and difficulty getting along with others.
- Problems with poor memory or inability to concentrate.

1.3.1 Forms of sleep apnea

In particular, sleep apnea is categorized into three forms: *Central Sleep Apnea (CSA)*, *Obstructive Sleep Apnea (OSA)*, and *Mixed Sleep Apnea (MSA)* (the combination of OSA and CSA), constituting 0.4%, 84%, 15% of cases, respectively, in United States and Europe [Morgenthaler et al., 2006].

CSA Characterized by the absence of respiratory effort and, thus, the absence of airflow during sleep. It may occur secondary to lesions that affect the sensory component, the integrative and executive neuronal systems, or the motor component (i.e., the lower motor neurons, nerves, and muscles) of the neural systems involved in respiratory control.

OSA Characterized by times during sleep in which air cannot flow normally into the lungs due to an obstruction usually caused by the collapse of the soft tissues in the back of the throat (upper airway) and tongue during sleep. OSA is more common in men, women after menopause, people over the age of 65 and can also occur in children.

Polysomnography (PSG)

Nowadays, the gold standard diagnosis system for SAHS recognition is Polysomnography (PSG). This test usually involves the following steps:

- The subject has to go to the sleep unit of medical center where he will spend the night in order to record several physiological signals;
- The resulting signals, which will be referred to as "polysomnographic recording" or "PSG", are then visually analyzed offline by the medical specialists.

The aim of this analysis is to identify sleep stages, score breathing-related events which occur during the night, and extract characteristic indices which describe the subject's sleep condition during that night (e.g. Apnea-Hypopnea Index, the number of events per hour of sleep).
The Standards of Practice Committee of the AASM develops and reviews indications for PSG in the diagnosis of commonly encountered sleep disorders, including sleep apnea.

One of the main problems of this syndrome is that patients are usually unaware of their own symptoms. Moerover, it is known that subject who suffers SAHS have an higher probability of having cardiac and cerebral infarcts or high arterial blood pressure, as well as arrhythmias and other dysfunctions of the cardiorespiratory system. [Alvarez-Estevez and Moret-Bonillo, 2015].

According to the US Institute of Medicine Committee on Sleep Medicine, about 50–70 million US adults suffer of sleep or wakefulness disorders and American Academy of Sleep Medicine estimates that more than 29 million US adults suffer from moderate to severe OSA, with an estimated 80% living unaware and undiagnosed. It is obvious to understand why SAHS is a public health and economic challenge. During the last 30 years, lots of researches has been conducted to clarify the incidence and the risk factors, such as the obesity epidemic and global prevalence of OSA. In fact, being OSA the most probable from of apnea, have been encouraged OSA studies [Shokoueinejad et al., 2017].

Although various devices have been used to measure physiological signals, detect apneic events and help treat sleep apnea, significant opportunities remain to improve the quality, efficiency, and affordability of sleep apnea care.

1.4 System requirements to diagnose and detect apneic events

American Academy of Sleep Medicine (AASM) digital task force identifies five basic tasks a system used to diagnose and detect breathing-related events must embrace [Penzel and Conradt, 2000]:

1. The system must allow to acquire and record data;
2. The system must allow to visualize the aforementioned data;
3. The system must allow data manipulation so that clinicians can visually assign a score to events;
4. The system must allow for data reduction. In particular, the final goal is to obtain useful diagnostic summary statistics for reporting starting from epochs;
5. The system must allow storage of relevant data and results.

It is important to say that does not exist a uniform standard for the upper listed processes.
Stages 3 and 4 are the most interesting ones in the light of this thesis.

AASM also indicates four types of sleep diagnostic devices which are better described in Table 1.1.

	TYPE 1	**TYPE 2**	**TYPE 3**	**TYPE 4**
Parameters	At least seven: • EEG (C4-A1 or C3-A2); • EOG; • chin EMG; • ECG; • airflow; • respiratory effort; • oxygen saturation	At least seven: • EEG (C4-A1 or C3-A2); • EOG; • chin EMG; • ECG or heart rate; • airflow; • respiratory effort; • oxygen saturation	At least four: • ventilation (at least two channels of respiratory movement, or respiratory movement and airflow); • ECG or heart rate; • oxygen saturation	At least one
Body position	Documented or measured	Not mandatory	Not mandatory	Not measured
Leg movement	EMG or motion sensor	EMG or motion sensor	Not mandatory	Not mandatory
Personnel	Always present	Not present	Not present	Not present
Possible Intervention	Yes	No	No	No

Table 1.1: Sleep diagnostic devices types according to American Academy of Sleep Medicine (AASM) [Ferber et al., 1994].
Type 1: Standard PSG system.
Type 2: Comprehensive portable PSG.
Type 3: Modified portable sleep apnea testing.
Type 4: Continuous single- or dual-bioparameter recording.

1.5 Current Treatment Modalities

Sleep apnea is a complex condition for which it has not yet been identified an efficient and comfortable treatment.
In this section will be discussed the treatment modalities which are currently used.

1.5.1 Assistive devices

The assistive devices are by far identified with the Positive Airway Pressure (PAP) devices whose aim is to apply external pressure to the patient's upper airway.
They usually consist of three major parts:

1. A positive pressure generator, such as a fan or turbine;
2. A nasal or oral interface, such as a mask;
3. A tube connecting the two parts.

PAP devices are usually divided into four categories:

1. Continuous Positive Airway Pressure (CPAP) devices: the basic type of PAP machines. They deliver constant pressure during the whole night and require manual laboratory titration prior to use. According to Becker et al., 2003, CPAP devices are effective in the treatment of apneas and hypopneas but the major problems are that over 40% of the patients with OSA are noncompliant [Weaver and Grunstein, 2008] and it may not suppress CSA, especially in patients who experienced heart failure [Arzt et al., 2007];
2. Bi-level Positive Airway Pressure (BPAP) devices: a pressure-controlled ventilation which delivers higher inhalation positive airway pressure (IPAP) and a lower expiratory positive airway pressure (EPAP) decreasing the subject's breathing effort since the timing to change from EPAP to IPAP is carefully clocked by a flow sensor. According to Berry et al., 2012, typical IPAP and EPAP ranges go from 4 to 30 cmH2O;
3. Auto-titrating Positive Airway Pressure (APAP) devices: such devices provide adjustable pressure to maintain airway patency and provide an adequate response to respiratory events. The data in the literature mostly shows that CPAP and APAP affect similarly sleep quality and that they can adjust and compensate for the pressure when leaks occur [Shokoueinejad et al., 2017]. However, Berry and Sriram, 2014, observed that some patients may be sensitive to the pressure changes and consequently feel less comfortable with APAP;

4. Adaptive Servo Ventilation (ASV) devices: these are devices that provide patients steady, minute ventilation based on the measurement of patient breaths.

Another type of PAP device is the Nasal Expiratory Positive Airway Pressure (NEPAP) device which consists of two nasal valves placed on the nostrils that provide variable resistance during breathing: low when the subject inhales and high when the subject exhales. Providing high resistance translates in applying a positive pressure through the airway that open the soft tissues up contrasting their collapse[Berry et al., 2011]. However, NEPAP devices are not side effects free, for instance subjects have reported dry mouth, headache, and others.

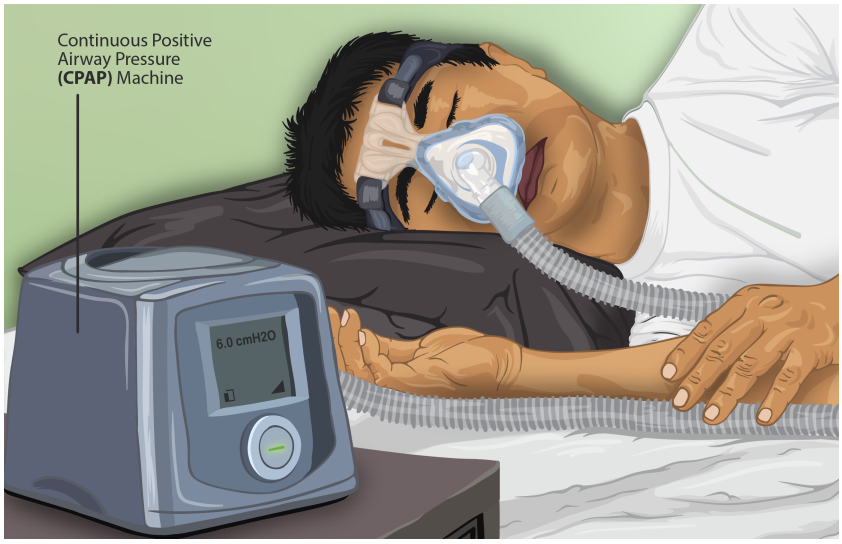


Figure 1.1: Depiction of a Sleep Apnea patient using a CPAP machine. Licensed under Creative Commons Attribution-Share Alike 4.0 International.

As reported in Shokoueinejad's review, "current continuous technological advances aim to improve the patient's comfort, adherence, and clinical benefits which could be categorized as respiration phase detection, ventilation estimation, CSA distinguishing, humidifiers, expiratory pressure relief (EPR), ramp, and automatic start and finish" [Shokoueinejad et al., 2017].

1.5.2 Therapeutical oral devices

Oral Appliances (OAs)

Oral Appliances (OAs), or Intraoral Devices (ODs), physically interact with mandible, tongue, or soft palate during sleep in order to prevent the collapse of upper airway muscles causing OSA. In their work, Shokoueinejad et al. reported that "ODs

are found to have higher adherence rates but still tend to be a secondary form of treatment for OSA to CPAP, which results more efficient". Moreover, ODs often resulted helpful for patients who become non-compliant to CPAP despite they have been shown to not strictly alter sleep apnea in patients [Shokoueinejad et al., 2017]. In the same review is indicated that ODs can be divided in:

- Soft Palate Lifters (SPLs): These devices need to be in contact with the most posterior region of the mouth and to be secured by a mouth guard or retainer, therefore are the most uncomfortable kind of OAs. This kind of OAs are less effective and less compliant than CPAP according to Barthlen et al., 2000, and the most common side effects are gagging, soft tissue irritation and choking.

- Tongue Retaining Devices (TRDs): These are the second most utilized form of OAs. They utilize negative pressure to secure the tongue but have lower efficacy and compliance rates than MAAs. These devices can be applied either in combination with MAAs or stand-alone if MAAs cause dental issues. These devices have less effectiveness but similar compliance compared to CPAP according to Deane et al., 2009, and the most common side effects are excess salivation, dryness of mouth, and soft tissue irritation.
  A sketch of such devices can be observed in figure 1.2.

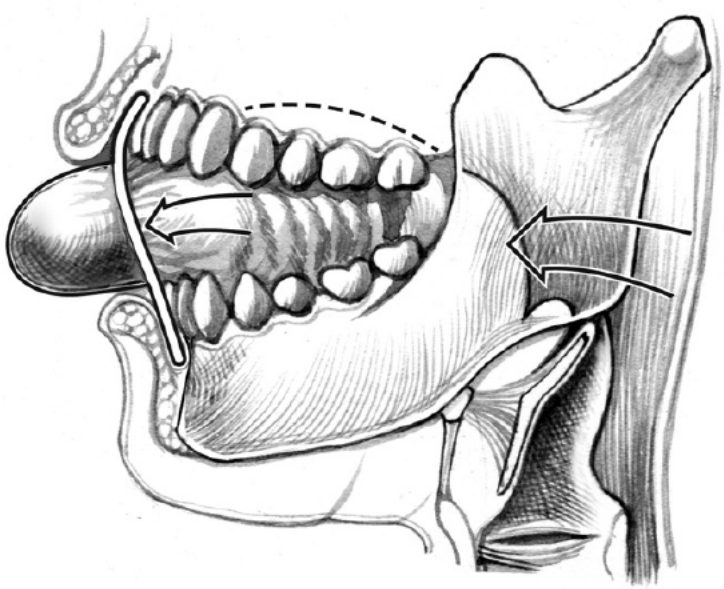
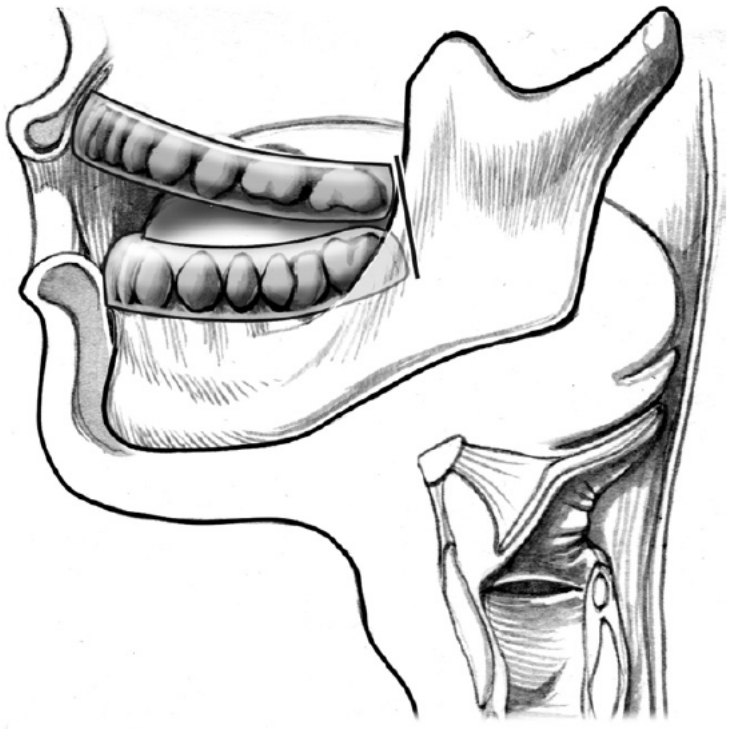
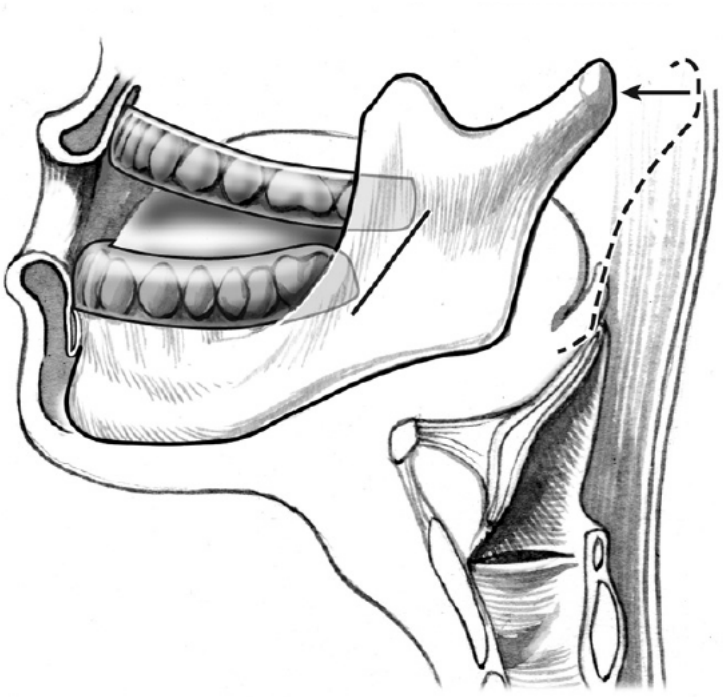


Figure 1.2: Sketch of tongue retaining device. [Bender, 2012]

- Mandibular Advancement Appliances (MAAs): These devices are by far the most used ODs to combat OSA. They are molded using dental impressions of the patient's mouth and fixed by attaching them to one or two dental arches. The working principle of all MAAs is to rotate the mandible downwards and protrude the mandible into an anterior position to force an increase in the upper airway as stated by Schmidt-Nowara et al., 1995.
  The major problem is determining how far protruded the mandible needs to be in order to reduce the AHI and to prevent Temporo-Mandibular Joint (TMJ) pain. There are several protocols that can be followed with the newer ones being automatized while the older ones need manual titration.
  A sketch of MAAs working principle can be observed in figure 1.3. A comparison between MAAs and CPAP showed that MAAs are less effective but more compliant than CPAP and the most common side effects are jaw soreness, as said before, and mouth dryness [Hoffstein, 2007].



(a)



(b)

Figure 1.3: a) Sketch of a mandibular advancement device in neutral position. b) Sketch of a mandibular advancement device with lower jaw protruded. [Bender, 2012]

1.5.3 Oral Pressure Therapy (OPT)

This is a more recent approach that consists in pulling the soft tissues forward preventing their collapse by applying a vacuum in the mouth. This method still requires an external device but tends to be less noisy than CPAP and does not need a full mask but only a small mouth guard.

1.5.4 Positional Therapy

Positional OSA is a particular subset of OSA characterized by a double increase in AHI while the patient is in a supine position compared to a non-supine position. It is stated that about half of all OSA are positional OSA. The aim of PT is to force the patient to sleep in a non-supine position; examples of PT are alarm systems, pillows with straps, and vibrating devices. This approach has big initial compliance which decreases over time to sub-optimal rates.

1.5.5 Other therapies

The approaches discussed until now are the most used. As reported in Shokoueinejad's review, there are also other therapies used as alternatives to CPAP, custom-made OA, and upper airway surgery such as the electrical stimulation of the hypoglossal and phrenic nerves and the use of inspired CO_2.

The hypoglossal nerve innervates the genioglossus muscle whose correct stimulation leads to increased inspiratory airflow (since the activation of this muscle is correlated with increased upper airway patency) without waking the patient. Schwartz et al., 2001, observed that these conditions can let patients go into deeper sleep stages. The stimulation is granted by an electrode chirurgically placed on the hypoglossal nerve, which stimulates tongue protrusion. The main problem is that surgery is required in order to implant the electrical stimulator.
Another electrical stimulation method has been investigated as a treatment for SAHS: the stimulation of the phrenic nerve. The key concept is to stimulate the diaphragm in order to restore a normal, physiological breathing pattern during sleep. Such stimulus is provided when the subject is recognized as being in normal sleeping and posture conditions trying to synchronize it with the inspiratory phase of respiration. Both phrenic nerve and hypoglossal nerve stimulation systems consist of a pulse generator, a stimulation lead, and a sensing lead.

The inspired CO_2 method is used method to control abnormal breathing. In particular, breathing in low CO_2 concentrations can help prevent periodic reductions in partial pressure of CO_2 below the apnea threshold which is usually 2 to 6 mmHg below the eucapnic sleeping partial pressure level of CO_2. This could reduce the number of apnea events.
However, the system required for exogenous CO_2 is impractical to keep at home. Another method of increasing CO_2 is the addition of about 500ml of dead space which does not contribute to gas exchange [A. Xie et al., 2001]. The main problem is that this practice could lead to cardiovascular problems.

1.6 Thesis Goal

The purpose of the present work is to extract a particular set of parameters, called features, from the polysomnographic record and uses these features in order to train some machine learning model to recognize apneic epochs. The concept of "sleep epoch" will be deepened in the next chapter.

As can be observed in table 1.1, some of the signals acquired by standard PSG systems and comprehensive portable PSGs (respectively type I and type II devices)

can be considered intrusive to a subject's sleep and can make sleep uncomfortable. Examples of such intrusive signals can be airflow signals and nasal pressure signal. The firsts are usually taken by airflow thermistors inserted inside the subject's nostrils (nasal airflow) or mouth (oral airflow) and whose wires need to be taped to the outer sides of the nose in order to not come out during rest. The latter is usually acquired by a nasal pressure transducer through a nasal cannula placed over the top of the airflow sensor transducers and whose wires are taped together with the wires of the airflow sensor. Moreover, these signals were excluded because also CPAP users were considered since the recording equipment results even more uncomfortable for them.

The thesis goal will be pursued trying to use low intrusive signals. In order to do so, it has been decided to use an alternative set of signals that includes some of the signals required by type II devices and some of the signals required by type III devices.In particular, the chosen signals are:

- *EEG and EOG signals*;
- *ECG or RR signal*;
- *SpO_2 signal*;
- *Thoracic volume signal*;
- *Audio recording*;
- *Body position signal*.

Such signals have been selected with the intention of having a simple-to-put-on and comfortable recording system for the patient that could be embedded in some wearable system and that could represent an alternative to portable PSG systems. The exception are EEG and EOG signals whose recording requires the placement of some electrodes; they were not excluded because considered very informative about the sleep condition of the patient being some of the pivotal signals for the assignment of sleep stages.
As reported in the manual of operations of WSC [Wisconsin Sleep Cohort, 2009], the data channels critical to the sleep analysis process are EOG and EEG channels, nasal pressure, respitrace, and oximetry.

Chapter 2

Database and signals description

Thanks to Professor Alessandro Puiatti (SUPSI), with whom this thesis project was carried out, who asked and obtained permission for the Wisconsin Sleep Cohort (WSC) database.

2.1 Wisconsin Sleep Cohort (WSC)

The Wisconsin Sleep Cohort (WSC) is an ongoing longitudinal study of the causes, consequences, and natural history of sleep disorders, particularly sleep apnea [G.-Q. Zhang et al., 2018; Young et al., 2009].

As reported in National Sleep Research Resource (NSRR) website: "the WSC collected overnight in-laboratory sleep studies (in-patient studies at the University of Wisconsin - Madison ICTR's CTRC). The provided data represent a limited portion of all potentially available Wisconsin Sleep Cohort data. Some Wisconsin Sleep Cohort data are presently withheld from NSRR due to one or more of the following":

- Constraints on sharing due to some participants refusing data sharing options at the point of obtaining informed consent;
- Data are being used to address currently-funded projects;
- Data have not yet been prepared for sharing;
- Data are not available in an electronically-shareable format (e.g., "old-school" paper-based polysomnographic recordings);

2.2 Database Details

The database was provided together with a dataset containing useful information about the subjects who underwent PSGs such as anthropometric parameters (e.g. body mass index (BMI), height, neck circumference, etc.), demographics parameters, clinical data (e.g. blood pressure, total cholesterol, etc.), sleep questionnaires scorings from different scales (e.g. Epworth sleepiness scale (ESS), modified Horne-Ostberg Morningness-Eveningness Questionnaire, etc.) or such as PSG derived indexes (e.g. Apnea-Hypopnea Index (AHI), total sleep time (TST) both REM and NREM, minimum blood saturation value, etc.).
Some important data are reportend in table 2.1.

	Male (n=848)	**Female (n=739)**
Age [years]	58 ± 8	56 ± 8
BMI [kg/m^2]	31 ± 6	33 ± 8
Neck circumference [cm]	41 ± 3	36 ± 4
TST [h]	5.9 ± 0.9	6.4 ± 0.9
NREM TST [h]	5 ± 0.8	5.3 ± 0.8
REM TST [h]	1 ± 0.4	1.1 ± 0.5
AHI* [events/h]	1.9 ± 1.4	1.5 ± 1.4
NREM AHI* [events/h]	1.7 ± 1.5	1.5 ± 1.2
REM AHI* [events/h]	2.4 ± 1.2	2.3 ± 1.3
ESS score	9 ± 4	9 ± 4
CPAP (n users)	51	33

Table 2.1: Comparison of demographic, clinical ans polysomnographic Characteristics of WSC database subjects. Data are presented as mean ± SD.
Abbreviations: Obstructive Sleep Apnea (OSA); body mass index (BMI); Epworth sleepiness scale (ESS); total sleep time (TST).
* Notice that since AHI, NREM AHI and REM AHI have similar to lognormal distributions, for these parameters mean and standard deviation are indicated in log scale.

For each subject in the database are provided 16 signals recorded during the night and three textual files. In particular, the available signals are:

- *2 EEG signals*: C3-M2 and O1-M2;
- *2 EOG signals*: E1 and E2;
- *2 EMG signals*: chin (mastoid) and linked legs;
- *1 ECG signal*;
- *1 microphone registration* (called "Snore");
- *2 flow signals*: oral and nasal flows;
- *1 nasal pressure*;
- *3 volume signals*: thoracic volume, abdominal volume and the sum of the previous;
- *1 body position signal*;
- *1 blood saturation signal*.

The textual files provided are:

- A *log file* in which are reported some useful information about the measurements of the signals during the PSG;
- A *staging file* with the sleep staging epoch per epoch made by an expert technician according to the R&K rules;
- A *scoring file* which contains information about the events which occurs during the sleep, such as OSA and CSA.

The log files contain information about the calibration process of the measurement system used to record all the signals. The purpose of doing both machines and biophysical calibrations is to validate the integrity of the input data. The first step is to check if there is data coming in on all channels and artifact-free being particularly careful with EOG, EEG, respitrace, and oximeter channels which are critical to the analysis process. In other words, a loss of data in these channels may cause sleep epochs to be unscorable.
Then the recording on the Grass system starts and the bio-calibration process begins with the isomaneuver: the patient has to twice take in a deep breath and exhale and after the second exhalation, have the patient hold the breath in order to apply nose clips and with mouth tightly closed make paradoxical movements of

the chest and abdomen, simulating an apneic event. This process is required in order to adjust the gains of the channels.
Once the gain has been set, the remaining bio-calibrations are performed:

- Eyes closed for 15 seconds;
- Eyes open for 15 seconds;
- Move eyes up and back to center;
- Move eyes down and back to center;
- Move eyes left and back to center;
- Move eyes right and back to center;
- Blink several times;
- Clench teeth and relax;
- Move tongue back and forth in the mouth;
- Flex left big toe;
- Flex right big toe;
- Count from 1 to 10;
- Simulate snore sounds;
- Put on nose clip and breathe through mouth;
- Remove nose clips, close mouth, and breathe through the nose;
- Take a deep breath, exhale, and hold breath as long as possible;
- Record peak SAO_2 and lowest value as a user-defined annotation;
- Close eyes and start steady state for 2 minutes;
- Open eyes and start steady state for 2 minutes;
- End steady state.

The "log" files contain other important information such as possible changes in the subject's position during sleep, possible gain changes of some signal channels, or if the subject spent time awake during the night because of different reasons (e.g. Needed to go to the bathroom, or needed to take some medicine...) indicating the epochs when the wake condition starts and ends. The epochs in which the lights are switched off and switched back on are also written in these files. The latter are epochs that will be useful to give the definition of "useful signals" and "sleep epochs" that will be given later in this chapter.

It is important to notice that each signal is divided into 30 seconds segments called "epochs" and at each epoch is assigned a numeric value indicating the sleep stage of the patient according to the rules that will be discussed in the following paragraph. The aforementioned 30 seconds epochs' length is a legacy of paper PSGs since 30 seconds of recording occupied a page when recorded at a speed of 10 mm/s (ideal for viewing alpha spindles).

2.2.1 Sleep Staging

"Sleep has been traditionally divided into non-REM (NREM) sleep and REM sleep. The sleep staging criteria were standardized in 1968 by Rechtschaffen and Kales [Rechtschaffen, 1968] (or R&K rules), based on EEG changes, dividing NREM sleep into four stages: stage I, stage II, stage III, stage IV. In 2007 was published a revised version of the R&K scoring system in which arousals, as well as respiratory, cardiac, and movement events, were also added to the scoring. This revised system is called the AASM Manual for the Scoring of Sleep and Associated Events and the most significant change was merging stages III and IV into stage N3" [Roebuck et al., 2013].

The sleep staging guidelines followed by WSC clinicians and the numerical encoding for sleep stages are resumed in table 2.2.

For all the epochs in which the patient was moving have been used the numeric code 6 and for all the remaining epochs which can not be categorized have been used the numeric code 7.
The sleep stage encoding will be essential for the definition of "useful signal" (see section 2.3).

STAGE	CODE	EEG	EOG	EMG
Wake	0	Alpha activity and/or low voltage mixed frequencies.	REMs and eye blinks.	High tonic EMG.
1	1	Relatively low voltage, mixed frequency EEG predominantly in the 2-7 Hz range, at about 50-75 μV. Vertex sharp waves up to 200 μV.	Slow eye movements of several secs duration. No REMs.	Tonic EMG levels below those of Stage W.
2	2	Sleep spindles and/or K-complexes (0.5 sec duration), and absence of sufficient high amplitude to define presence of Stage 3 or 4.		
3	3	20% but not more that 50% of epoch consists of 2 Hz or slower waves with amplitude greater than 75 μV peak to peak. Sleep spindles may or may not be present.		
4	4	More than 50% of epoch consists of 2 Hz waves or slower with amplitudes greater than 75 μV peak to peak. Sleep spindles may or may not be present.		
REM	5	Relatively low voltage, mixed frequency EEG. Presence of saw tooth waves in vertex and frontal areas. Alpha activity that is 1-2 Hz slower than wakefulness may be prominent. Absence of K-complexes and sleep spindles.	Tonic mental-submental EMG at lower level than preceding sleep stage.	REMs.

Table 2.2: Sleep staging: criteria and encoding in the Grass Heritage System.

2.2.2 Scoring of breathing events and epochs encoding

In the scoring files are reported tables containing useful information about the events which happen during the PSG such as the epoch in which the event starts, the length of the event in samples, the exact time of the event, the minimum SpO_2 value reached during the event, the recording channel from which the presence of a given event is recognized, a marker code and a marker text indicating what type of event it is (e.g. "OSA-100", "CSA-102, "Hypopnea-202", "LMA-405", "LM-412",...). On the base of the scoring files, was implemented a function that extracted, for each subject, a table with the events' information and that assigned returned a numeric encoding at each event as reported in Table 2.3.

EVENT	CODE
Normal sleep	1
Hypopnea	2
Obstructive Sleep Apnea (OSA)	3
Central Sleep Apnea (CSA)	4
Mixed Sleep Apnea (MSA)	5

Table 2.3: True Class, numerical encoding

In particular, in the manual of operation provided with the database are reported the definitions and the scoring procedures of apneas, hypopneas, and leg movements. The ones of interest for the goal of this thesis are the first two.
More accurate definitions of breathing events and scoring procedures for each event can be found in Appendix A.

2.2.3 Database dimension

Among the 2559 subjects provided with the WSC database, only 1827 subjects have been considered due to the lack of scoring and staging information needed in order to proceed with supervised learning algorithms training and due to the absence of occipital EEG channel.

2.3 Definition of useful signal

In order to normalize the pre-processing routines of all the signals used it has been decided to consider only the epochs in which the subject is effectively sleeping because the apnea events begin and develop during sleep. To do so a function that separates the sleep epochs from the non-sleep epochs was implemented. This function calculated the number of epochs in the signals as the length of a signal (all the signals for a specific subject must have all the same length) divided by the length of an epoch (30 seconds) and extracted from the staging files the relative code for each epoch, as stated in table 2.2, and make a double check extracting the lights-out and lights-on epochs from the log file and forcing the epochs before lights-out and after lights-on to be considered as non-sleep epoch.
In this way, it was possible to remove all the epochs in which the subject was not sleeping such as wake epochs, calibration epochs, or the epochs in which the subject was not in the bed. The latter epochs are usually characterized by an initial detachment of the sensors which results in abrupt changes in the signals until the subject return to bed with all the sensors correctly attached as shown in the figure 2.1.

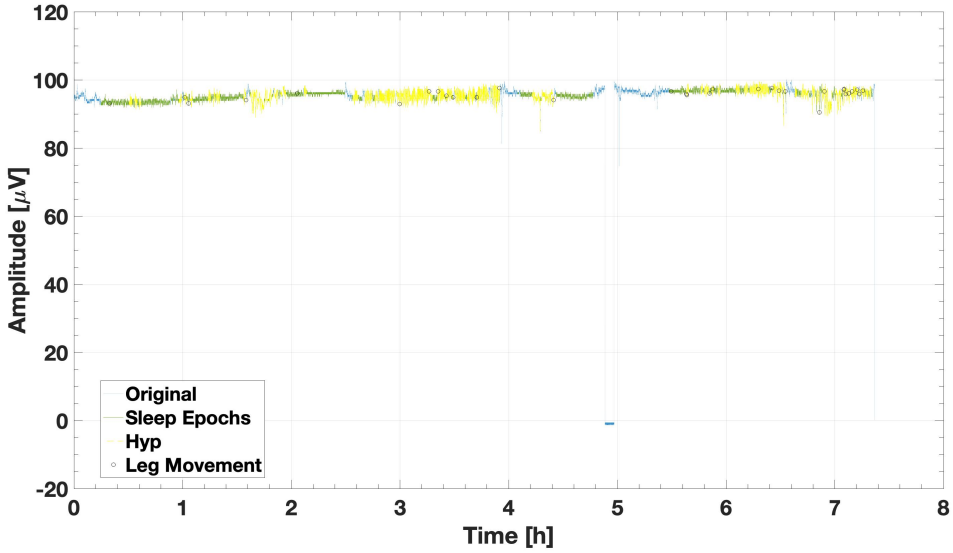


Figure 2.1: Example of SpO_2 signal considering only useful epochs.

Notice that major movement epochs which cause subjects to wake up were accounted for in the initialization of the sleep epochs. In conclusion, all the subjects without indications in the log file about the lights-out and lights-on epochs were discarded leaving 1587 subjects out of 2559 ready for features extraction.

2.4 Signals details

Before proceeding with the description of the processing routines for each signal, it is important to say that all the signals in the database have been recorded using the Grass Heritage system.
According to this system, all the signals have been sampled with a sampling frequency (fs) equal to 100Hz and have been pre-filtered by the hardware. In particular, the upper cut-off frequency for all signals is 30Hz while the lower cut-off frequency varies by signals: the EOG signals, the EEG signals, the ECG signal, the EMG signals, the audio record, the nasal and oral flow signals, and the nasal pressure signal have a lower cut-off frequency of 0.1Hz while the SpO_2 signal, the position signal, the thorax volume, abdomen volume and sum of volumes signals have a lower cut-off frequency of 0.01Hz.

In their work Penzel and Conradt indicated that "currently 100 Hz is regarded as the minimum acceptable sampling rate for EEG, EOG, and EMG in sleep recordings when using 35 Hz high filters while international recommendations for ECG require a sampling rate of at least 250 Hz for long-term ECG recordings and that at least two channels of ECG are recorded but if the ECG is recorded in a sleep laboratory and is only used to estimate the heart rate and not to investigate heart rate variability, then a sampling rate of 100 Hz may be sufficient. This results in an accuracy of 10msec when calculating RR intervals which is sufficient for clinical interpretation." [Penzel and Conradt, 2000].

The signals units of measure and pre-filter specifications are resumed in table 2.4.

In conclusion, the goal of signal processing for features extraction is to extract a table of features for each subject such that each row of this table represents a sleep epoch for that subject and each column represents a specific variable.
Finally, all the tables will be merged along columns into a unique table and two more columns will be added: the first identifies the subject while the second contains the true class to whom a specific epoch belongs (according to the numerical encoding reported in table 2.3).

SIGNAL	**PRE-FILTER**	**UNITS**
E1	BP: 0.1Hz - 30Hz	μV
E2	BP: 0.1Hz - 30Hz	μV
C3-M2	BP: 0.1Hz - 30Hz	μV
O1-M2	BP: 0.1Hz - 30Hz	μV
chin EMG	BP: 0.1Hz - 30Hz	μV
Linked Legs EMG	BP: 0.1Hz - 30Hz	μV
Snore	BP: 0.1Hz - 30Hz	μV
ECG	BP: 0.1Hz - 30Hz	μV
Nasal flow	BP: 0.1Hz - 30Hz	μV
Oral flow	BP: 0.1Hz - 30Hz	μV
Nasal Pressure	BP: 0.1Hz - 30Hz	μV
Thorax volume	BP: 0.01Hz - 30Hz	V
Abdomen volume	BP: 0.01Hz - 30Hz	V
Sum of volumes	BP: 0.01Hz - 30Hz	V
Position	BP: 0.01Hz - 30Hz	V
SpO_2	BP: 0.01Hz - 30Hz	Percentage

Table 2.4: Signals available in the WSC database with pre-filter settings and units of measure according to Grass Heritage system.
BP = band-pass.

Chapter 3

Signal preprocessing and features extraction

Theoretically, at this point in the workflow there should be the pre-processing of the signals in terms of filtering and artifact removal but, since the signals were supplied after a step of analog pre-filtering and artifacts rejection done by the recording system, will be discussed other pre-processing and features extraction algorithms implemented for the considered signals.
For each signal, was implemented a function that initializes, subject per subject, an mxn matrix such that m is the number of epochs for the considered subject and n is the number of features extracted from the considered signal. The initialization assigns NaN values along the rows which represent the no-sleep epochs and zeros along the rows which represent sleep epochs.

3.1 General Statistics features

A set of general statistics features was extracted from all the epochs of a great part of the signals. These are time-based statistics describing the signals' time series data distribution. They comprehend measures of central tendency, such as the mean and the median of the epoch's samples, and measures of dispersion, such as the range, the minimum (which is not a measure of dispersion itself but will be useful to give a statistical meaning as will be described later in this subsection), the standard deviation and a percentile range.
It was specified *percentile range* because for the blood saturation signal was decided to calculate the *5 to 95 percentile range* while for the rest of the signals have been calculated the *interquartile range*.

In particular, it was decided to calculate the *mean* because it uses all the data values and is, in a statistical sense, efficient but it is vulnerable to outliers while the *median* is not affected by outliers but is not statistically efficient, as it does not make use of all the individual data values.
Dispersion describes the spread or variability of the data, it has been decided to calculate the *range* because it is the simplest measure of variability; since it is physically described by a single number, it is not a statistical representation of the range (which is given by two numbers: the highest and the lowest of the series). In order to fix this, the *minimum* value has been added as a feature while the maximum value can be derived as range plus minimum.

3.2 BLOOD SATURATION signal

With the term "SpO_2" we refer to the percentage of oxygenated hemoglobin in the blood. Oxygen concentration is a great interest parameter to monitor during sleep studies because it may reduce dramatically during an apneic event causing low organ perfusion and, eventually, possible organ failure.
The device used for measuring SpO_2 is called "pulse oximeter" and it monitors non-invasively the peripheral oxygen saturation 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. Figure 3.1 shows a sketch of a pulse oximeter functioning principle.

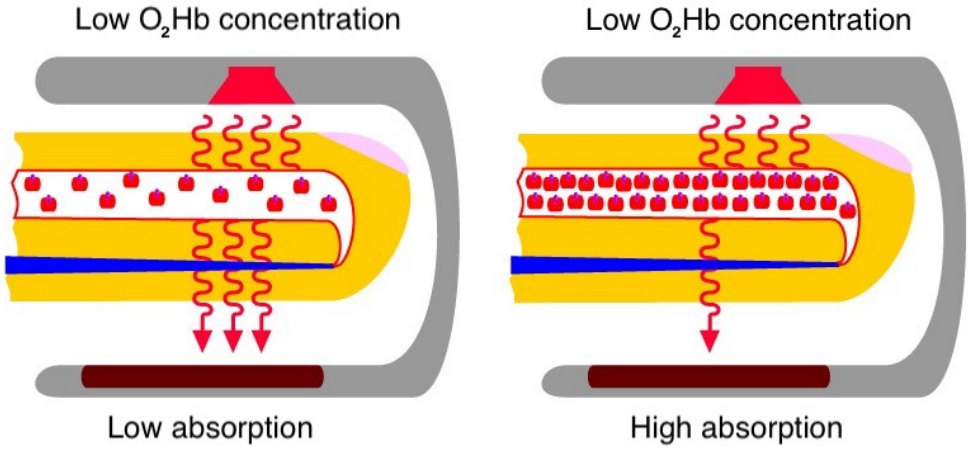


Figure 3.1: Sketch of a pulse oximeter functioning principle. Licensed under Creative Commons Attribution-Share Alike 4.0 International.

Regarding the SpO_2 signals present in the WSC database, they were recorded using an Ohmeda 3900 oximeter with a 3 seconds averaging rate.

In their review, Roebuck et al. indicated that normally SpO_2 ranges fall between 85–95% with a certain inter-subject variability [Roebuck et al., 2013] with the biggest limitation of pulse oximetry monitors being the high rate of false alarms caused by motion artifacts and poor sensor contact, as noticed by Chambrin, 2001. As described in section 2.3, major movements were already considered as part of non-useful epochs since they also greatly influence the other signals.
In particular, it was observed that movement artifacts generate false alarm rates between 70% and 80% [Petterson et al., 2007].
Since major movements cause the awake of the patient with the exit from a sleep stage, these events were already accounted for in the initialization of the sleep epochs on which will be calculated all the features as explained earlier in section 2.3.

Before proceeding with features extraction it was decided to downsample SpO_2 signals with a decimation factor of 20 (from 100Hz to 5Hz) in order to reduce the computational burden of the calculation of the features since they are slow time-variating signals.

3.2.1 Features extraction

For blood saturation signal the calculated features (or biomarkers) as indicated by Levy et al., 2021, can be divided in:

- *General statistic features*: time-based statistics describing the oxygen saturation time series data distribution;
- *Complexity features*: quantifies the presence of long-range correlations in non-stationary time series;
- *Hypoxic burden features*: time-based measures quantifying the overall degree of hypoxemia imposed on the heart and other organs during the recording period.

General statistic features

In particular, with respect to the general statistic features discussed in subsection 3.1, the following features were added:

Mx The percentage of the signal at least x% below median oxygen saturation with x=2% as suggested by Deviaene et al., 2018. In other words, it represents the percentage of SpO_2 signal at least 2% below median oxygen saturation.

5 to 95 percentile range Used instead of the interquartile range, in order to consider a bigger portion of the signal since desaturations caused by apnea events can abruptly lower the SpO_2 levels that want to be considered.

ZCx The number of zero-crossing points at the x% SpO_2 within the considered epoch. It is used to characterize the signal oscillations around a certain blood oxygen saturation level, x. In this case, as the baseline level was chosen the mean, as indicated by B. Xie and Minn, 2012. In particular, when two consecutive samples of the blood saturation signal are, respectively, lower and higher (or vice versa) than the baseline, a crossing point is detected.

As reported in Levy's work, "one should expect that a SpO_2 time series from an apneic patient as compared to that of a non-apneic patient will oscillate more around the baseline because of the presence of desaturations and then reach a higher value" [Levy et al., 2021]. ZCav is defined as:

$$ZC_{av} = \sum_{i=10}^{N_{SpO_2}-1} ZC_i(av),$$

$$ZC_i(av) = \begin{cases} 1 & \text{if } (SpO_{2_i} - av) * (SpO_{2_{i+1}} - av) < 0 \\ 0 & \text{else} \end{cases}$$

where N_{SpO_2} is the number of samples of the SpO_2 time series.

Delta Index (ΔI) It was defined by Pépin et al., 1991 as "the sum of the absolute variations between two successive points divided by the number of intervals. The original intuition was that SpO_2 oscillations, induced by repeated apnea resumption of ventilation sequence, will lead to a high ΔI, while prolonged desaturations or nearly constant SpO_2 values would lead to a low ΔI".
The ΔI index is defined as:

$$\Delta I = \frac{1}{N_{window}} \sum_{i=1}^{N_{window}} |SpO_{2window_{i+1}} - SpO_{2window_i}|,$$

where $SpO_{2_{window_i}}$ is the average of the level of oxygen saturation for the window i of length x s, and N_{window} is the number of windows. In this case it has been chosen x = 30 seconds.
Since for each current epoch was needed the previous and the following epochs for

calculating the Delta Index, it is obvious that no ΔI values have been calculated for the first and the last useful epochs since the previous and the following epochs, respectively, are initialized as NaN values as explained in section 2.3.

Complexity features

The only feature that have been considered is the *approximate entropy (ApEn)*, whose properties make it suitable for biomedical time series analysis.
This feature aims to quantify irregularity, as stated by Pincus, 2001. In particular, "ApEn evaluates both dominant and subordinant patterns in the data, and discriminates series for which clear feature recognition is difficult being almost unaffected by low level noise, robust to outliers, scale invariant and model independent" [Álvarez et al., 2007].

One limitation is that it is applicable to time series with at least 50 data points. This represent a lower limit to downsampling of the signal, in fact SpO_2 signal has been downsampled to 5Hz in order to have 150 data points for each epoch despite even if, in order to calculate the other features, it could have been downsampled more drastically (e.g. to 1Hz, which would have meant to have 30 data points for each epoch).
Pincus states also that "ApEn assigns a non-negative number to a sequence or time series, with larger values corresponding to greater apparent process randomness or serial irregularity, and smaller values corresponding to more instances of recognizable features or patterns in the data". Moreover, "two input parameters, a run length m and a tolerance window r, must be specified to compute ApEn. In other words, ApEn measures the logarithmic likelihood that runs of patterns that are close (within r) for m contiguous observations remain close (within the same tolerance width r) on subsequent incremental comparisons" [Pincus, 2001].
ApEn(m,r,N) can be defined as:

$$ApEn = \varphi^m(r) - \varphi^{m+1}(r),$$

$$\varphi^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \ln\left(\frac{N^m(i)}{N-m+1}\right),$$

where $N^m(i)$ is the number of windows of length m for which the distance from the window beginning at the index i is lower than or equal to r.

The criticality of ApEn is the choice of the values for m and r.
Actually there are no straight rules but approximately m should be linked to the length of signal fluctuations that one can expect. Since speaking about fluctuations is similar to speaking about spectral content of the signal, an empirical rule is to

choose m near to the mid-band frequency of the considered signal.
For r, instead, it is usually chosen a value between 0.1 and 0.25 times the standard deviation of the signal, as indicated by Pincus et al.

In this case, in order to have a more complete depiction of the differences among different couples of values of m and r it was chosen to calculate *three ApEn parameters*:

1. *ApEn010*: with m=1 and r=0.10·std(x),
2. *ApEn015*: with m=1 and r=0.15·std(x),
3. *ApEn025*: with m=1 and r=0.25·std(x),

where x is the considered epoch.

Hypoxic burden features

It was decided to calculate the *cumulative time*, that is the percentage of time spent below a certain oxygen saturation level. In particular, five cumulative times have been calculated:

1. *TSA70*: Percentage of the time spent below the 70% oxygen saturation level;
2. *TSA80*: Percentage of the time spent below the 80% oxygen saturation level;
3. *TSA85*: Percentage of the time spent below the 85% oxygen saturation level;
4. *TSA90*: Percentage of the time spent below the 90% oxygen saturation level;
5. *TSA95*: Percentage of the time spent below the 95% oxygen saturation level;

In conclusion, the subset of features calculated from the SpO_2 signal amount to 18 features: 10 general statistic features, 3 complexity features and 5 hypoxic burden features.

3.3 VOLUME signal

Within the WSC database there were three signals related to respiratory volumes: thoracic volume signal, abdominal volume signal and the sum of the previous two. All of these signals have been acquired by respiratory inductance plethysmography (RIP) which is a widely studied technique highly regarded by the AASM due to its several advantages such as great accuracy, sensitivity, and high patient safety [Z. Zhang et al., 2012].

Shokoueinejad et al., 2017 described the functioning principle of RIP: "it indirectly measures ventilation by recording changes in thoracic and abdomen cross-section". In other words, "it measures abdomen chest movements through coiled wires wrapped around a patient's chest that carry a low amplitude sine wave. Changes in chest and abdomen circumference alter the self-inductance of the wires, and therefore the frequency of the sine wave, which can be demodulated and processed to track changes in chest and abdomen size [Watson et al., 1988]. The wires were initially integrated into two elastic bands, one around the abdomen and one around the chest, but RIP has also been used with wires sewn into shirts, comfortably worn by patients. Other RIP devices are belts that utilize embedded piezocrystals that emit changes in voltage corresponding to movement of the thorax and abdomen during breathing". [Levin and Chauvel, 2019] It is important to remember that the posture of the patient greatly affects the measurement of tidal volume: if the RIP is calibrated with the patient in the upright position, sleeping and breathing in the supine position can lead to error.

With proper calibration, RIP can achieve a tidal volume measurement accuracy of 96% when compared with pneumotachography or spirometry [Gonzalez et al., 1984].

RIP is noted for its added benefit of helping to distinguish between OSA and CSA, as illustrated in figure 3.2. While during OSA, cessation of breathing occurs despite an ongoing effort to breathe, while CSA occurs when the brain does not properly send signals to the muscles controlling respiration. During CSA, the lack of effort by the muscles in the abdomen and chest can be noted with the use of RIP, aiding distinguishing CSA from OSA.

A disadvantage of RIP technique is that it has poor accuracy and precision in obese patients which can limit the accuracy of detecting hypopneas. However, RIP scoring has been shown to have increased sensitivity and specificity in overweight or obese patients when compared to the recommended and acceptable criteria for sleep scoring by the AASM [Kogan et al., 2016]. Although RIP accuracy and precision is significantly decreased during sleep in obese patients, RIP is still clinically useful [Cantineau et al., 1992].

In particular, the *Pro-Tech 2-Rip inductance plethysmography summation system* has been used to record the signals.

It was decided to use only the thoracic RIP signal because it has showed the greatest changes during most of the sleep apnea events compared with the others and in order to decrease the computational burden.

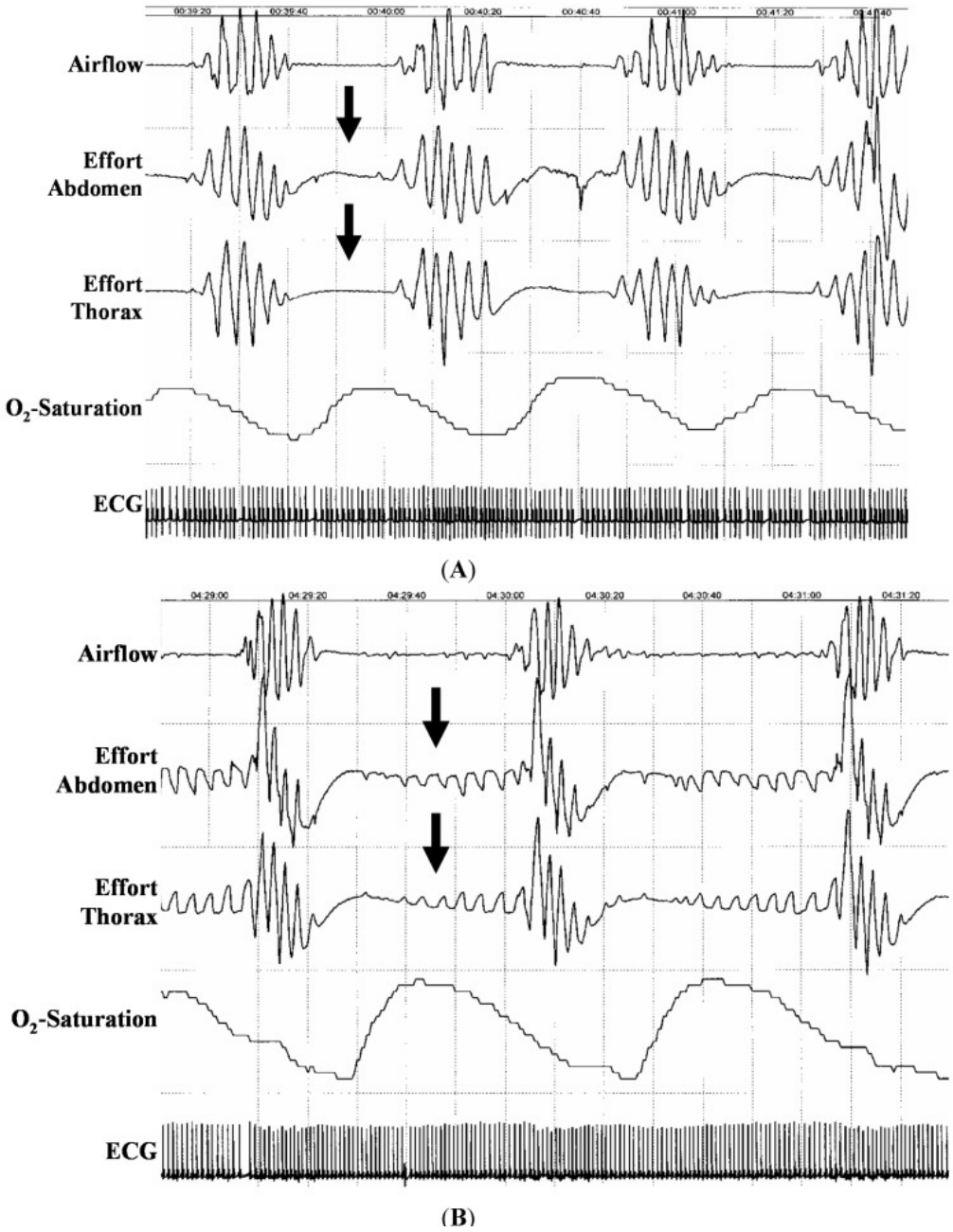


Figure 3.2: Polysomnography (PSG) (3 min) with central sleep apnea (A) and obstructive sleep apnea (B) Note the absence of chest efforts and abdominal movements in the absence of oronasal airflow in central sleep apnea but not in obstructive sleep apnea (arrows). Also, note the pronounced decrease in O2-saturation following each apnea episode [Grimm and Koehler, 2014]. Licensed under Creative Commons Attribution-Share Alike 4.0 International.

Due to the fact that signals contain noise generated from different sources such as subject movements, electrical inference and other disturbances, in his work Van Steenkiste identifies an upper cutoff frequency of about 0.7Hz under which all relevant respiration information in the frequency domain can be extracted [Van Steenkiste et al., 2018].
After segmenting the signal in 30 seconds epochs, each epoch was decomposed by performing a six levels wavelet decomposition using Daubechies 4 (*db4*) as mother wavelet. In particular, only the approximation coefficients were considered. This resulted in a reduction of the band of the reconstructed signal up to 0.78 Hz.
For a more detailed explanation of the Wavelet Transform see appendix B.
Then each epoch was rescaled to get the minimum section value to 0 and the maximum section value to 1 to ensures all characteristics in the respiration signal are equally visible regardless of the signal strength.

3.3.1 Features extraction

For thoracic volume signal the calculated feature can be divided in:

- *General statistic features*;
- *Frequency related features*;
- *Peaks related features.*

General statistic features

For this signal were calculated some of the features indicated in subsection 3.1 such as the *mean*, the *standard deviation*, the *median*, the *interquartile range* and were added the *Area under the absolute value of the signal* which is straightly related to the amount of air occupying the airways.

Frequency related features

Since respiration is a quasi periodic behaviour some frequency related feature were extracted too after computing the Power Spectral Density (PSD) of each section such as:

- *Mean frequency*;
- *Median frequency*;
- *Peak frequency*;
- *Total band-power*;

- *Energy in the filtered band.*

These features were chosen in order to account for the variation that an apneic event could lead the power spectrum of the signal to.
Finally, it has been added the *Renyi entropy* as a *complexity measure* with the intuition that a variation of the morphology of the signal can lead to a variation of entropy: for example, a lack of breathing effort (typical of CSA epochs) could lead to a reduction in the complexity of the signal indicating that the physiological system producing such a signal has reduced activity.

Peaks related features

This is a subset of time-domain features calculated after performing a peak detection to locate respiratory peaks. In particular, it was implemented a function which searches for peaks with a *minimum prominence* of 5% the range of the epoch amplitude (which is exactly 0.05 after the rescaling).
The prominence of a peak measures how much the peak stands out due to its intrinsic height and its location relative to other peaks. A low isolated peak can be more prominent than one that is higher but is an otherwise unremarkable member of a tall range.
To measure the prominence of a peak the following guidelines from MATLAB user's guide [The MathWorks Inc., 2021] were followed:

1. Place a marker on the peak;
2. Draw an horizontal line from the peak to the left and right until the line does one of the following:
   - Crosses the signal because there is a higher peak;
   - Reaches the left or right end of the signal.
3. Find the minimum of the signal in each of the two intervals defined in Step 2. This point is either a valley or one of the signal endpoints;
4. The higher of the two interval minima specifies the reference level. The height of the peak above this level is its prominence.

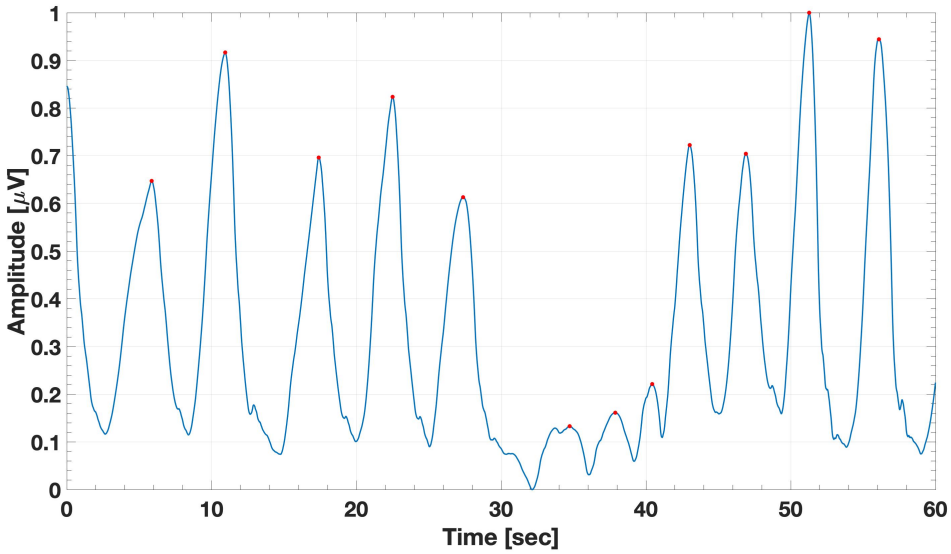
Then a double check over detected maxima was performed considering a 2 seconds window centered on each detected peak and verifying that the peak is a local maximum in that window in order to exclude that the peak in question derives from an artifact or that it is generated by noise.

Two examples are shown in figure 3.3: it is possible to appreciate the differences between an OSA event, in which respiratory effort is present but hindered by the collapse of soft tissues (in fact respiratory peaks results still detectable during the event), and a CSA event, in which, instead, there is no breathing effort (in fact it is not possible to detect respiratory peaks during the event).

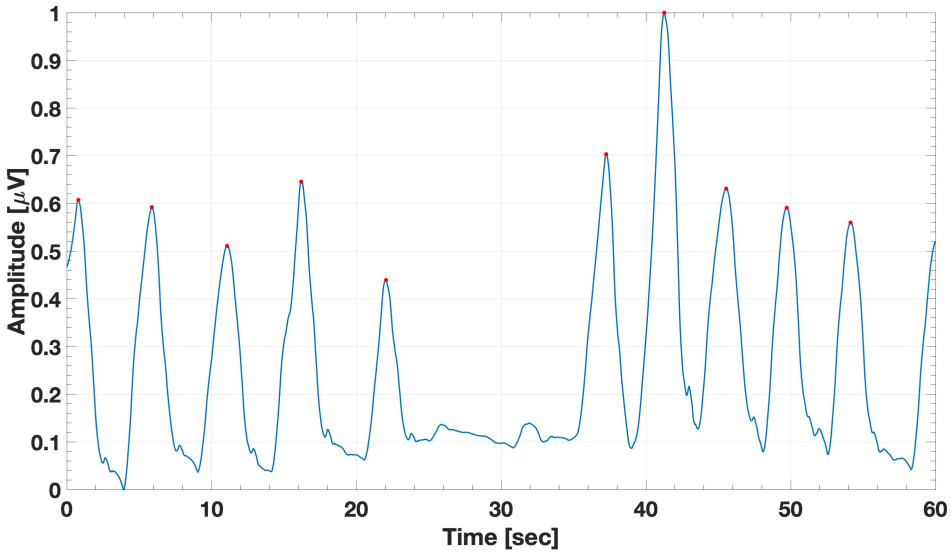
I was decided to calculate the following peaks related features which were indicated in Van Steenkiste's work [Van Steenkiste et al., 2018]:

- *Mean peak height*;
- *Standard deviation peak height*;
- *Skewness of peak height*;
- *Sum of the peak heights.*
- *Number of peaks*;
- *Mean inter-peak distance*;
- *Standard deviation of the inter-peak distance*
- *Skewness of the inter-peak distance*;

The first four characterize the amplitude of the breathing pattern, while the last four are related to the firsts in terms of conservation of inspired air volume. The intuition is that whenever an apneic event occurs, there will be a change in the amplitude of the signal (reduction in OSA, cessation in CSA) but, since the perfusion must be preserved, one can expect that at the lowering of the amplitude there will be an increase in number of breaths in a fixed time.



(a)



(b)

Figure 3.3: (a) Two consecutive epochs of thoracic volume signal containing a OSA event. (b) Two consecutive epochs of thoracic volume signal containing a CSA event. The respiratory peaks have been detected and marked in both cases.

In conclusion, the subset of features calculated from the thoracic volume signal amount to 19 features: 5 general statistics features, 5 frequency-related features, 1 complexity measure for each signal and 8 peaks related features

3.4 BODY POSITION signal

The body position signal is used to characterize whenever there is a change in position of the subject during sleep. Recording body position is important in patients with OSA since the apneic events occurring in the supine position are more severe than those occurring while sleeping in the lateral position. Thus, it is not only the number of apneic events that worsening the supine sleep position but, probably no less important, the nature of the apneic events themselves. [Oksenberg et al., 2000].

These signals in the WSC database are taken by a Natus Neurology DC body position sensor which is essentialy an accelerometer fasten to the chest of the subject by straps or by an adhesive patch. Since the signal is a slow time-variating signal, it was downsampled with a decimation factor of 20 from 100Hz to 5Hz in order to reduce computational burden of the features calculation. Then, a moving mean filter with a 15 samples window (which is $\frac{1}{10}$ of the length of the resampled signal) was applied with the intention of smoothing the signal's oscillation caused by noise or artifacts. In figure 3.4 is shown an epoch before and after the aforementioned preprocessing.

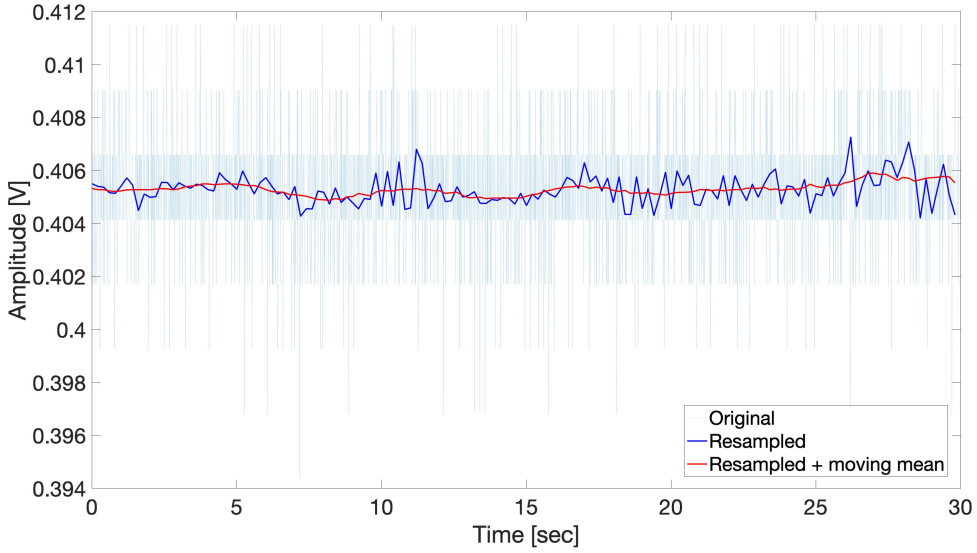


Figure 3.4: Body position signal epoch before and after preprocessing

Before talking about the feature extracted from the body position signal it is important to say that since major body movement affects negatively all the signals it was decided to implement a function that detects this major movements during sleep and discard the epochs where the events were found. To do this, keeping in mind that the signal was resampled and filtered in order to result almost flat, considering that all the major movements causes big oscillations with respect to the mean value of the signal, all epochs with peaks of prominence greater than 5 mV were excluded.

3.4.1 Features extraction

Because a sensorized subject cannot sleep in a prone position due to all the equipment he is connected to, the possible sleeping position are only three: supine, on the left side and on the right side. To identify the position in which a subject was sleeping during a certain epoch the numerical encoding in table 3.1 was used.

POSITION	**CODE**
Supine	1
Right side	2
Left side	3
Unidentified	4

Table 3.1: Numerical coding of possible sleep positions

An additional class has been added in order to account for all that epochs in which the subject position is not assignable to other classes such as the epochs in which the subject change position multiple times or whenever there is a movement artifact due to the sensor slip.

With the intuition that there could be epochs in which the subject changes position due to the possible occurrence of an apneic event, a *dispersion measure* has been added. In particular, it has been decided to calculate the *standard deviation* of the signal during the considered epoch.

3.5 EEG and EOG signals

As we can see from table 2.2, EEG and EOG signals (in combination with the EMG signal), are the are the key signals for assigning of sleep stages since they carry information about the cerebral activity of the subject during sleep, when there is the maximum of neuronal plasticity.

In particular, EOG signals are used to evaluate if the subject is in REM or NREM sleep and to remove eventual eyes motion artifacts from EEG signals. Since the sleep staging was already provided with the data, no eye motion artifacts rejection based on EOG signals was performed on EEG signals with the idea of characterizing the known REM epochs. Moreover, the acquiring system is already calibrated so that there are no 60 Hz artifacts or electrode popping artifacts neither on the EEG and on the EOG signals.

The positioning of the electrodes on the scalp of the subject has been carried out according to the international 10-20 system of electrode placement whose representation in shown in figure 3.5.

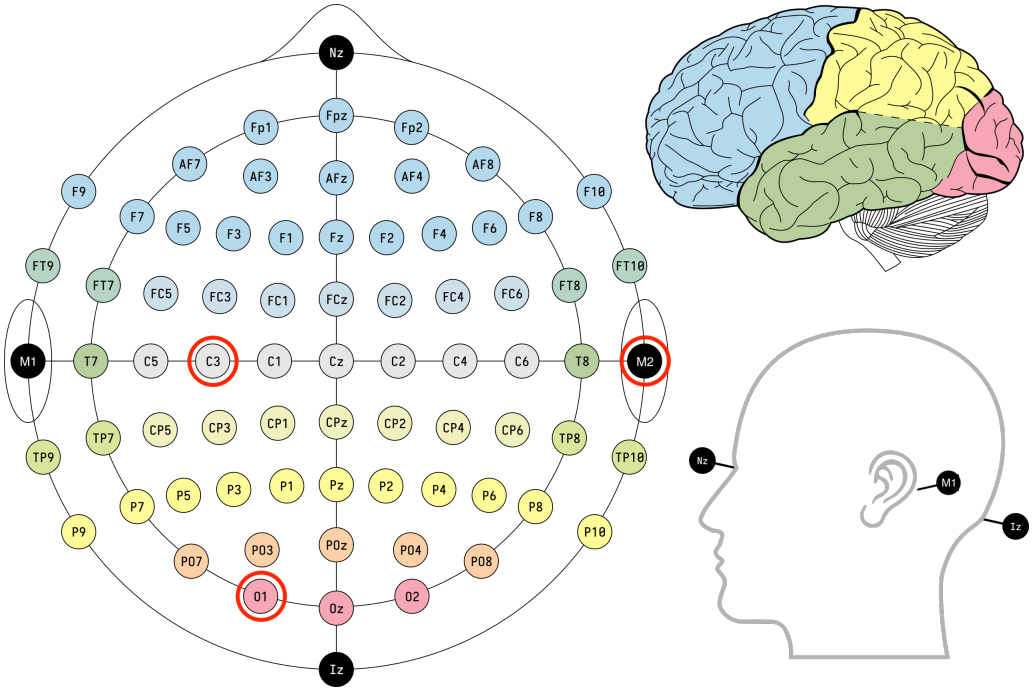


Figure 3.5: International 10-20 system of electrode placement. The available EEG channels electrodes are circled in red.

For each subject of the database are available 2 EEG signals and 2 EOG signals:

- *E1 (left EOG)*
- *E2 (right EOG)*
- *C3-M2 (left central EEG)*
- *O1-M2 (left occipital EEG)*

Where E1 is placed 2cm to the left and 2cm down from the outer canthus of the left eye, E2 is placed 2cm to the right and 2cm up from the outer canthus of the left eye. The M2 electrode, as shown in figure 3.5, is the *reference electrode* applied over the skin at the level of the right mastoid bone (behind the ear) which is a cephalic standard position. Its potential is subtracted from the potential of the sampling electrodes since the EEG signals are single-differentials. The EEG channels that have been acquired are positioned over the motility area (C3) and over the area of the sight (O1) with the idea of detect eventual EEG variation due to cerebral activity in these two fields.

For non sleep-related application, the EEG signals are usually acquired using soft helmets with electrode housings but, since during sleep this can be uncomfortable for the subject, in PSGs the sampling electrodes for the EEG signal are placed manually by the operator and held in place with an adhesive patch. This can add noise because of variability both inter- and intra-operator.

3.5.1 Features extraction

Since both EEG and EOG signals are stochastic processes (so much so that often they are referred as "interference signals"), they carry the great part of their information in the Fourier domain (or frequency domain), so it has been decided to calculate mostly frequency-related features.
To do so the PSD of the signals was estimated using a traditional method for the periodogram: the Welch method.
According to this method, in order to estimate the PSD of the signal the following steps were followed:

1. *Divide the considered portion of the signal into smaller overlapping segments.* In particular, in this case, the portion of the signal is one 30 seconds epoch that has been divided into 1 second overlapping segments of 50% of their length (this amount of overlap is the most used for biomedical signals), in order to reduce the number of point on which the PSD is estimated increasing the consistency of the estimator;

2. *Apply a smoothing window to each segment in order to reduce the polarization error of the estimator* (in other words reducing the *power leakage*. In this case it has been used an Hamming window, which is the most used window for biomedical signals being characterized by small side lobes (<-40dB);

3. *Estimate the periodogram for each segment*;

4. *Calculate the mean of all the periodograms obtained.*

A better explanation of Welch's use of Fast Fourier Transform (FFT) for the estimation of PSD can be found in Welch, 1967.

The EEG power spectrum can be divided into 5 bands (or rhythms) with different characteristics, as shown in table 3.2

RHYTHM	**FREQUENCY**	**AMPLITUDE**	**CONDITIONS**
Delta (δ)	0.5Hz - 3Hz	$20\mu V$ - $200\mu V$	Pathological or very deep sleep
Theta (θ)	3Hz - 7Hz	$5\mu V$ - $100\mu V$	Deep sleep
Alpha (α)	8Hz - 13Hz	$100\mu V$ - $200\mu V$	Mental relaxation
Beta (β)	14Hz - 30Hz	$1\mu V$ - $20\mu V$	Concentration, activated cortex
Gamma (γ)	$>$ 30Hz	$1\mu V$ - $20\mu V$	Concentration, activated cortex

Table 3.2: EEG rhythms with relative frequency and amplitude ranges and associated conditions [Molinari, 2021].

Since the signals have been sampled at 100Hz and each epoch has 3000 samples (30 seconds), considering that the windowing step reduce the number of samples in each section to 300 (1 second), the theoretical frequency resolution is about 333 mHz which is enough to distinguish the EEG rhythms and eventual small variations in the power spectrum.

In the figure 3.6 are reported comparisons between PSDs of C3-M2 and O1-M2 signals: the greatest part of signal power is brought by the delta and theta rhythms indicating that the subject is in a deep sleep condition. Moreover, it is observable that passing from a normal sleep condition to an apnea condition (either OSA or CSA) causes power to shift in the theta, low alpha rhythms indicating an increasing activity of the brain. In particular the greatest increase is observable during the CSA event while a milder increase is observable during the OSA event.

This is surely linked to the intrinsic difference between the two types of apnea: CSA characterized by absence of breathing effort and OSA characterised by presence of breathing effort but obstructed upper airways, as stated in subsection 1.3.1

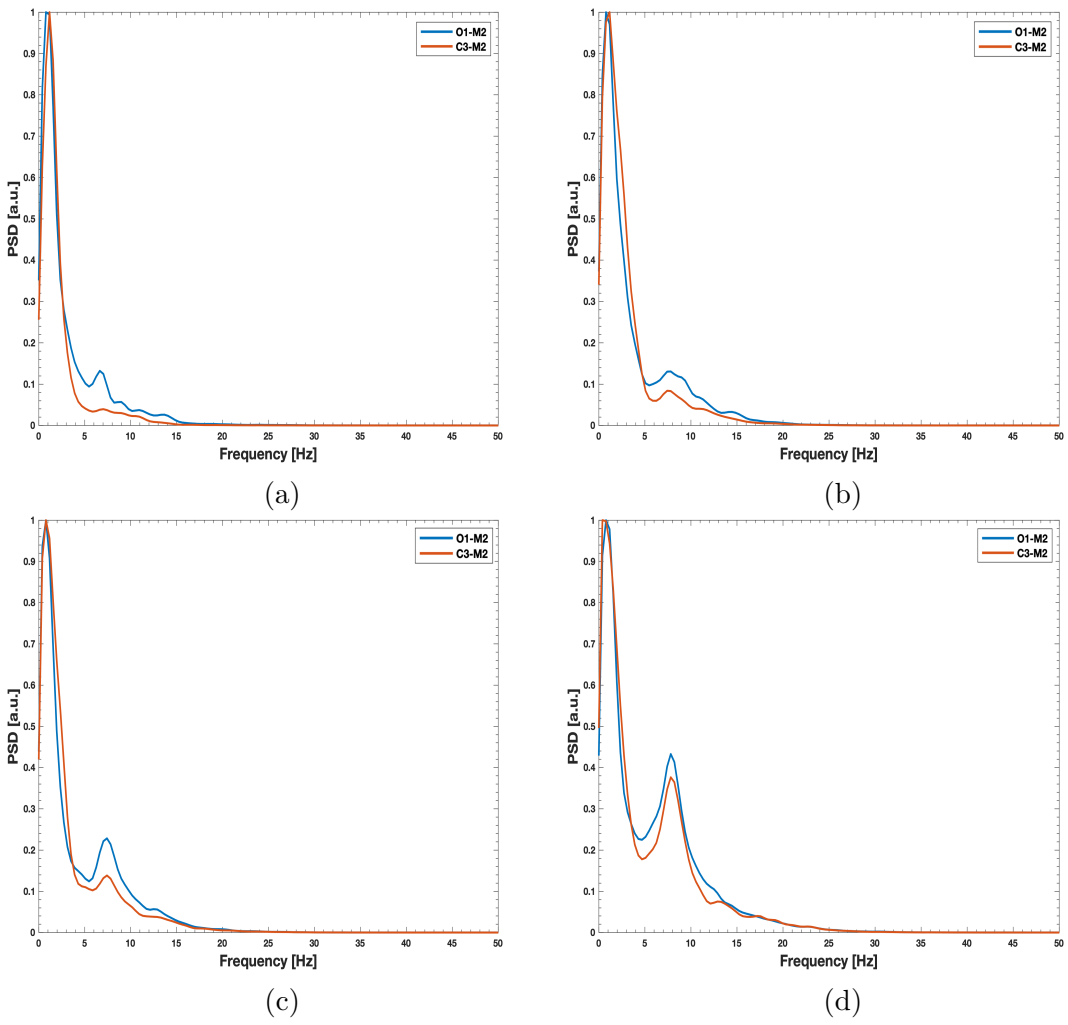


Figure 3.6: PSD comparison between C3-M2 and O1-M2 EEG signals.
(a) Normal sleep. (b) Hypopnea event. (c) OSA event. (d) CSA event.
The PSD are normalized over the respective maximum value in order to appreciate the relative power change in bands theta and alpha.

The following spectral features were derived from EEG and EOG signals [Alvarez et al., 2009]:

- *The relative power in the EEG classical frequency bands* reported in 3.2;
- *The total EEG power*;
- *The median frequency*;
- *3 ratio of different power values*:
  - *The ratio of power in the Delta band and the combined power in Theta and Alpha bands*;
  - *The ratio of power in the Theta band and the combined power in Delta and Alpha bands*;
  - *The ratio of power in the Alpha band and the combined power in Delta and Theta bands.*

While in their work Alvarez et al., 2009 indicate the spectral entropy as a plausible complexity measure for EEG signal, in this case it has been decided to calculate the *Renyi's entropy* in order to be consistent with other complexity measures calculated in other signals.
The intuition is that when the subject is in a deep stage of sleep the signal power will mainly be carried into the delta or theta rhythms, characterized by low frequency ranges, indicating that nervous central system is less active compared to when a subject is in a higher sleep stage which causes signal power to be mainly carried into the alpha rhythm, characterized by an higher frequency range.

Although Renyi's Entropy is a PSD-based complexity measure susceptible to all the PSD estimation problems, it was decided to calculate this measure of complexity rather than other non PSD-based complexity measures such as approximate entropy or sample entropy (which are more often used for EEG signal characterization) because of the high computational time required for the latter.

In conclusion, since for each signal have been extracted 10 features, the subset of features calculated from the 2 EEG and 2 EOG signals amount to 40 features: 9 frequency-related features and 1 complexity measure for each signal.

3.6 ECG signal

The electrocardiogram gives information about the electrical activity of the heart. The ECG signal is deterministic and is characterized by a distinctive waveform shown in figure 3.7. It normally consists of the P wave, QRS complex and the T wave. Among the several methods proposed for identification of sleep apnea based on a single signal, ECG is the most extensively studied signal and, in terms of feature extraction, ECG-based features are considered as ones of the most efficient features to detect sleep disorders and have been extracted using different methods.

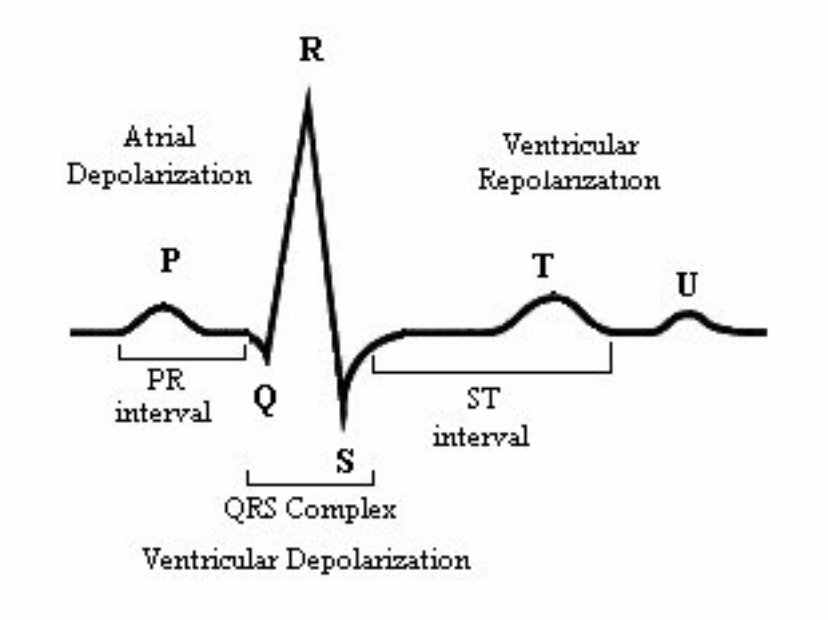


Figure 3.7: Points and elements of ECG signal [Saad et al., 2006].

3.6.1 Features extraction

General statistic features

In this case it has been decided to not extract the general statistic features described in subsection 3.1 because, as observed by Yılmaz et al., 2010, for ECG signals in particular, features like mean, standard deviation and range are particularly affected by outliers and thus classification performance deteriorates when these features are included in the analysis.
For instance, the ECG signal is greatly affected by the sleep position of the subject as can be observed in 3.8.

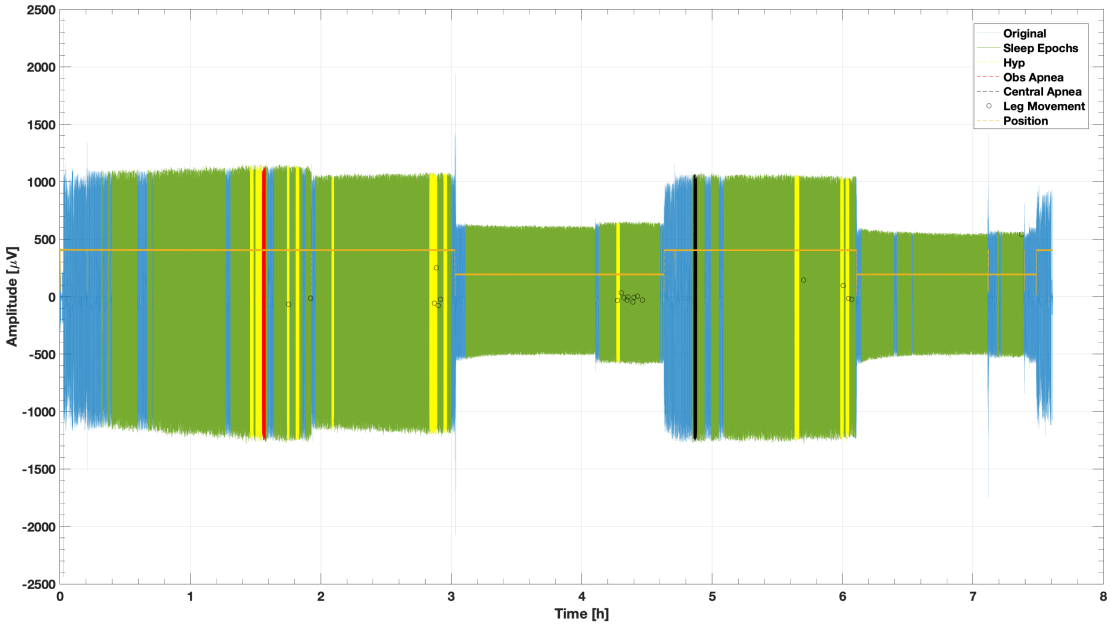


Figure 3.8: ECG signal with [Saad et al., 2006].

For ECG signals the calculated feature can be divided in:

- *RR interval related features*;
- *Frequency-related features.*

RR intervals related features

RR intervals are identified as the time intervals between two consecutive R peaks of ECG waveform which represent cyclic variations in the duration of a heartbeat. They have been associated with sleep apnea episodes in terms of bradycardia during apnea followed by tachycardia upon its cessation [Almazaydeh et al., 2012].

In order to distinguish the R waves from the other waves of the ECG signal, as proposed by Almazaydeh et al., 2012, the following two conditions were considered:

1. It has to be a local maximum, which is detected by a local max function within a window of 150ms;
2. The local max peaks must be at least 2 standard deviation above the mean.

In this case, with the intention of discarding eventual outliers generated by motion artifacts, it was decided to consider all the local max peaks at least 2 and a half standard deviation above the mean.

Once the R-peak was determined, RR intervals were computed as the peak to peak time period from two continuous peak signals:

$$RR(i) = R(i+1) - R(i), \text{ where } i = 1, 2, ..., n - 1$$

Where n - 1 and n is the number of peaks contained in the considered epoch and *R(i)* and *R(i+1)* are the time instants of two consecutive peaks.

Following Isa and Fanany's guidelines [Isa et al., 2010], the below listed ECG features which are most effective for apnea detection are calculated:

- *Mean epoch RR-interval*;
- *Standard deviation of the epoch RR intervals*;
- *The NN50 measure (variant 1)*, defined as the number of pairs of adjacent RR- intervals where the first RR interval exceeds the second RR- interval by more than 50 ms;
- *The NN50 measure (variant 2)*, defined as the number of pairs of adjacent RR-intervals where the second RR-interval exceeds the first RR interval by more than 50 ms;
- *Two pNN50 measures*, defined as each NN50 measure divided by the total number of RR-intervals;
- *The SDSD measures*, defined as the standard deviation of the differences between adjacent RR intervals;
- *The RMSSD measures*, defined as the square root of the mean of the sum of the squares of differences between adjacent RR- intervals;
- *Median of RR-intervals*;
- *Inter-quartile range*, defined as difference between 75th and 25th percentiles of the RR-interval value distribution;
- *Mean absolute deviation values*, defined as mean of absolute values obtained by the subtraction of the mean RR-interval values from all the RR-interval values in an epoch.

The first seven features are proposed by de Chazal et al., 2004, while the three latter features are proposed by Yılmaz et al., 2010, who claimed that RR interval mean, standard deviation, and range are sensitive to outliers, and thus classification performance deteriorates when only these features are included.

Frequency related features

In their work Rachim et al., 2014 proposed a five levels wavelet decomposition to decompose ECG into five detail coefficients and one approximation coefficient. In particular, since the relative power of an ECG wave is between two and 30 Hz, decomposing ECGs into five levels of signal is enough and, in order to do so, Debauches 4 (db4) was chosen as the mother wavelet for wavelet decomposition. A more extensive explanation of wavelet decomposition and Multiresolution Analysis (MRA) can be found in appendix B.

Following Rachim's guidelines, the below listed features were calculated 5 times, one for each detail coefficient. In this perspective, each of the following variables represents five features in the final dataset:

- *Interquartile range*;
- *Variance*;
- *Standard deviation*;
- *Mean Absolute Deviation (MAD)*, defined as:

$$MAD = mean(abs(detail_{coeff} - mean_{detail_{coeff}}))$$

In conclusion, the subset of features calculated from ECG signal amount to 33 features: 10 RR intervals related features, 22 frequency-related features and 1 complexity measure.

3.7 AUDIO signal

As mentioned in section 1.3, snoring and choking sounds are common apnea symptoms which can manifest during sleep. It is believed that they could be linked to NREM and REM sleep as well as upper airway collapse in patients with OSA [Akhter et al., 2018]. In this perspective it was decided to proceed with sleep audio recording analysis.

Audio signal is recorded using a microphone which is usually taped, at a point below the microphone itself, to the volunteer's forehead or near the trachea. The audio signals within the WSC database were recorded with a Pro-Tech Snore Sensor like the one shown in figure 3.9.

Figure 3.9: Pro-Tech Snore Sensor.

3.7.1 Features extraction

Before proceeding with the description of the calculated features it is important to remember that usually a audio signals are sampled with high sampling frequency, such as 44'100 Hz or 48'000 Hz. Nevertheless, the available signals are sampled at 100Hz which is three orders of magnitude lower than usual. Because of this it resulted impossible to precisely calculate some of the key features used to characterize audio recordings such as the pitch.

For audio signal the selected feature can be divided in:

- *General statistic features*;
- *Frequency related features*;
- *Complexity features*;

General statistic features

For this signal were calculated all the features indicated in subsection 3.1. While the central tendency measures give information about the trend the signal follows, the dispersion features can be indicative of the presence of some event such as snores or breathing-related sounds that generate some oscillations.

Frequency related features

Because of the limitation induced by the low sampling frequency and because potential snore episodes are known to have nonstationary and complex behaviors [Wang et al., 2017], it was decided to decompose the signal with a five level wavelet decomposition with a Debauches 4 (db4) type mother wavelet and to consider the energy of the six levels (five details levels and one approximation level) of decomposition. Moreover it was calculated the *median frequency* with the intuition that the power brought by higher frequencies increases when the subject is snoring or chocking.

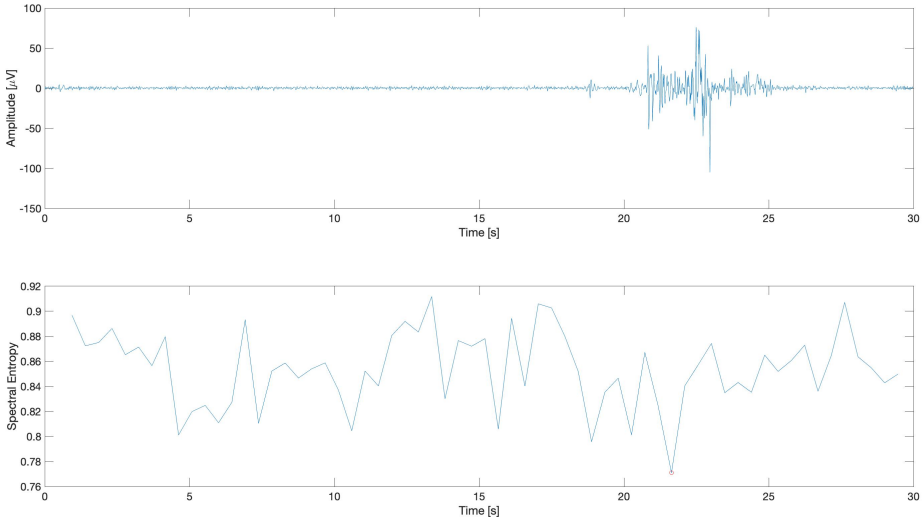
Complexity features

In their work, Wang et al. also indicate that *spectral entropy* is used to measure the flatness of PSD since spectral entropy is a complexity feature.
In this context it was decided to implement a function in order to get 4 spectral entropy related features. In particular, this function calculates the *instantaneous spectral entropy of the signal*(se), and search for spectral entropy peaks outside the range $mean(se) \pm 2 \cdot std(se)$. Then a control on nearby peaks is done in order to consider two peaks as part of the same respiratory event if there is no portion of the signal between them whose instantaneous spectral entropy is within the range $mean(se) \pm 2 \cdot std(se)$. In this way it is possible to detect events which increase signal complexity over a certain threshold.

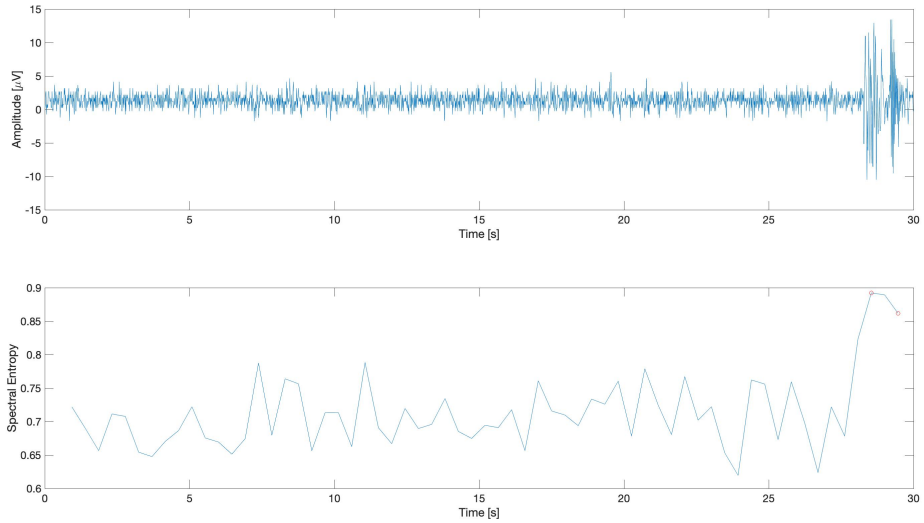
The features computed from this analysis are:

- *The number of SE peaks*, with the intention of identifying different events that occurs in the same epoch;
- *The SE mean value*, to have an overall vision of complexity;
- *The SE range*, to understand biggest event's severity in the epoch;
- *The time spent out of the SE threshold*, with the intuition that long events increase signal complexity for a longer time than short events helping discriminating from environmental noises and gasping sounds or snoring;

Examples of the upper algorithm are shown in figure 3.10.



(a)



(b)

Figure 3.10: (a) OSA epoch (upper image) and relative instantaneous spectral entropy (lower image) with red circled peak detected. (b) CSA epoch (upper image) and relative instantaneous spectral entropy (lower image) with red circled peak detected.

Finally, it was also calculated the *Renyi entropy* since this was the main complexity measure extracted from the previous signals.

In conclusion, the subset of features calculated from audio recording amount to 18 features: 6 general statistic features, 7 frequency-related features and 5 complexity features.

Chapter 4

Machine Learning algorithms preparation

Once features were extracted from processed signals, a dataset (matrix) of 1'514'301 epochs (along the rows) and 130 variables (along the columns) was formed.
In machine learning terms the epochs and the variables in the dataset are called, respectively, "observations" and "features" (or "predictors").
Since the present work aims to distinguish the sleep apnea condition, it is important to keep in mind that the data will be used in order to train supervised learning algorithms which are basically models that make predictions identifying patterns in data once trained upon available observations. In other words such models "learn" from observations whose class is already known.

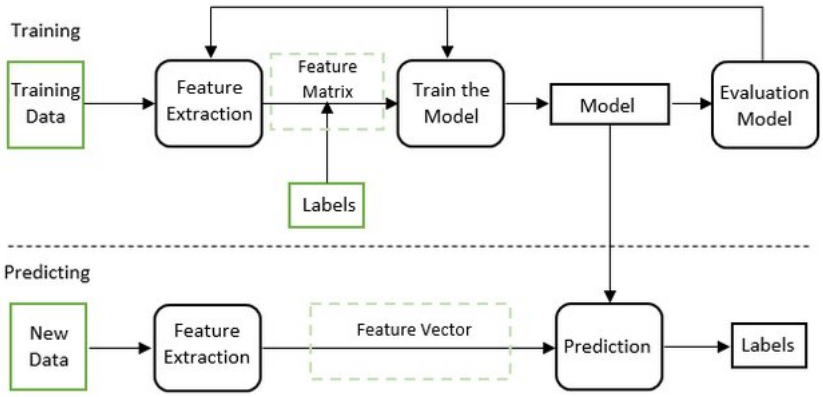


Figure 4.1: Supervised learning workflow [Ghareeb et al., 2022].

4.1 Data Cleaning

The first step for supervised learning is data preparation.
The dataset underwent a data cleaning process with the intention of eliminating outliers and artifacts that would have deteriorated the classification performances. Visual observation of different variables distributions, as expected, identified various behaviours. Apnea conditions (OSA, MSA and CSA) distributions showed with different trends: sometime such distribution were overlapped on each other and on normal distribution with proportionally the same central tendency and dispersion measures, while other times they were overlapped but with different central tendency and dispersion measures. Moreover, not all distributions were normal (e.g. some distribution showed a lognormal trend).

Because of the large number of features, it was decided to start cleaning up the dataset starting from general statistics features using previous knowledge about the physiological systems that generated the signals.

For instance, talking about the ECG signal, it is well known that during sleep the heart rate decreases and considering that patients which experience apneic events can have an increase of the heart rate caused by the hypercapnic condition generated by the reduction of blood saturation which lead to sudden arousals, an upper threshold of 150bpm (mean RR interval of 0.4 seconds) was imposed on the *mean RR interval length* feature. In the same perspective, a lower threshold was applied too, in particular all the epochs with a heart rate less than 30 bpm (mean RR interval of 2 seconds) were excluded. A double check was done with the same thresholds on the median value of RR intervals. Moreover, observing the standard deviation of the RR intervals it was decided to eliminate all the values above the 99.5 percentile of the OSA class, since this was the class with the higher values of standard deviation.

Considering, now, the blood saturation signal, different arguments have been made. First of all, as mentioned in section 2.2, together with the signal, was provided a dataset of clinicians verified parameters which describe the general condition of the subject and give useful information about the PSG itself such as the minimum blood saturation value that occurred during sleep. This value was used as a threshold to remove any epoch whom minimum value of SpO_2 was lower than what the sleep analyst has declared.
Moreover, knowing the SpO_2 range and the minimum values for each epoch it was possible to get the maximum value summing the previous two and since physiological blood saturation can not exceed 100% all the epochs with a maximum value above this threshold were excluded.

Another check was performed on the SpO_2 ranges. Considering one subject at a time, the intuition was that a spread distribution of SpO_2 ranges calculated over his epochs was indicative either of low quality signal recording or of multiple severe apneic events. This way it was possible to remove subject like the one shown in figure 4.2 which has a clearly poor quality blood saturation signal.

Finally, focusing on thoracic volume signal, since a normal subject has a respiration rate of 12-20 breath per minute, considering that when someone falls asleep the respiration rate slightly increase and considering that are also present CPAP users [Sleep Foundation, 2022], it was decided to eliminate all the epochs with more than 50 breaths (being an epoch 30 seconds long, it would have meant a respiratory rate of 100 breaths per minute).

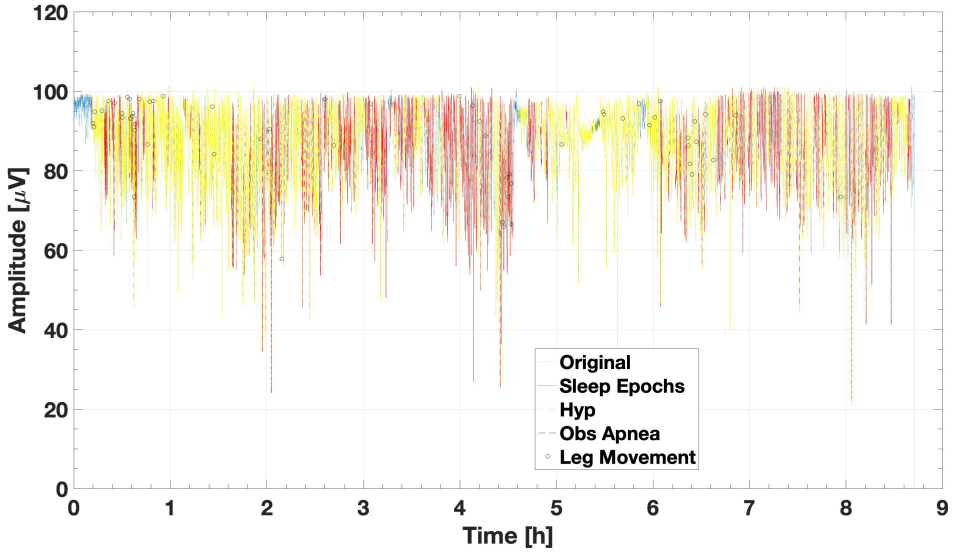


Figure 4.2: Example of poor quality SpO_2 signal.

After the data cleaning process, the useful observations left were only 973'571. It is significant to say that the discarded 540'730 epochs represent about the 35.71% of the total epochs which translates in more that a third of the overall size of the database (about 100GB out of more than 310GB) of useless information. This gives an impression of how big the data the processing system has to deal with.

4.2 Algorithms choice and fitting

The second step for supervised learning is choosing an algorithm and fitting a model to available data.

4.2.1 Algorithms choice

The algorithm choice is not univocal and requires searching for good tradeoffs between speed of training, memory usage and interpretability. Table 4.1 shows typical characteristics of the the aforementioned supervised learning algorithms.

Classifier	**Multiclass Support**	**Prediction Speed**	**Memory Usage**	**Interpretability**
Decision Trees	Yes	Fast	Small	Easy
Discriminant Analysis	Yes	Fast	Small (linear), large (quadratic)	Easy
Naïve Bayes	Yes	Medium (simple distributions), Slow (kernel distributions or high-dimensional data)	Small (simple distributions), Medium (kernel distributions or high-dimensional data)	Easy
SVM	No	Medium (linear), Slow (others)	Medium (linear), Large (binary)	Easy (linear), Hard (other kernels).
kNN	Yes	Slow (cubic), Medium (others)	Medium	Hard

Table 4.1: Typical characteristics of principal supervised learning algorithms [The MathWorks Inc., 2022].

In the present work it was decided to compare the previous models in order to identify the one with the best performances. Multiple dichotomous classifications were performed considering different couple of classes as will be described in the next chapter.

4.2.2 Training Set and Test Set creation

The aforementioned models were trained using only a portion of the initial data and were tested using another portion with different data. These subsets are called *Training Set* and *Test Set*. Training Set contained a balanced number epochs picked from the 70% of the database subjects randomly selected. Test Set contained a balanced number epochs picked from the remaining 30% of the database subjects. Because it is advisable to work with balanced sets, both training set and test set are constructed starting from the less numerous class in order to have a final set with a number of observations equal to the double of the latter. Attention was payed to keep the same classes proportions of the original dataset within the clusters considered for each analysis. Since the subjects of training set are randomly selected and do not have the same number of useful epochs, two different training sets or test sets will not have the same number of observation. In order to compare different models it is important to train them using the same training set and testing them with the same test set.

4.2.3 Models fitting

All the models were trained with training sets created as explained in previous subsection and it was specified to tune hyperparameters according to the built-in routines present in MATLAB© working environment limiting the tuning steps to 30 steps to try keeping training time below 6 hours since different trials needed to be done. Moreover, to obtain a better estimate of the predictive accuracy, a 5-KFold cross validation was done. It basically splits the training data into 5 parts at random (maintaining proportions between classes) and trains 5 models, each one on 4 parts holding the last part to examines the predictive accuracy.

4.3 Choice of the validation method

Then next step is the choice of the validation method to evaluate the accuracy of the fitted models. The three main methods are:

1. Examine the resubstitution error;
2. Examine the cross-validation error;
3. Examine the out-of-bag error (for bagged decision trees).

It was chosen to examine the resubstitution error (the error made by the model when it classifies the same data used for its training) having the lightest computational workload. In general, it is preferable to have low resubstitution error even though it does not guarantee good predictions for new data.

4.4 Prediction metrics

The last step consist in using the fitted model for making predictions on new data. Hereinafter are listed some important metrics used to compare trained models performances where TP, TN, FP and FN indicates, respectively, "True positives", "True negatives", "False positives" and "False negatives". Notice that "positive" and "negative" indicate belonging to different cluster which are not always normal and apneic condition therefore, for each of the following classification problems, it will be specified.

Sensitivity (SENS) It represents the ability to correctly identify positive epochs. It is defined as:

$$SENS = \frac{TP}{TP + FN}$$

Specificity (SPEC) It represents the ability to detect negative epochs correctly. It is defined as:

$$SPEC = \frac{TN}{TN + FP}$$

Positive Predictive Value (PPV) It represents a measure of the correct number of positive epochs with respect to the total number of positives epochs. It is defined as:

$$PPV = \frac{TP}{TP + FP}$$

Negative Predictive Value (NPV) It represents a measure of the correct number of negative epochs with respect to the total number of negative epochs. It is defined as:

$$NPV = \frac{TN}{TN + FN}$$

Accuracy (ACC) It represents the number of correct classified epochs, both positive and negative. It is defined as:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

For the same accuracy, it is preferable to have more FP than FN because an FP can undergo further investigations that can allow recognizing his negativity.

Chapter 5

Models performances

Below are discussed the classification performances obtained by the models on balanced Test Sets created according to subsection 4.2.2 guidelines for different classification problems.

In particular, for each classification problem, models performances are compared in terms of:

- Their resubstitution errors in percentage;
- Their prediction metrics (presented in section 4.4) supported by the relative confusion matrices and Receiver Operating Characteristic (ROC) curves.

The latter report classification performances at different thresholds and can be plotted because the considered classification problems are binary. On the vertical axis there is the *true-positive rate* (*sensitivity*) and on the horizontal axis there is the *false positive rate* (*1 - specificity*) and each point of the curve represent a different classifier. The current classifier is indicated in each ROC curve with a filled blue marker.
Another important metric is the *Area under the ROC Curve (AUC)* which is a value between 0 and 1 directly proportional to classification performances.

5.1 Normal & HYP vs Apnea - 1st Analysis

Before proceeding with the classification itself, to better understand how the variance of the available dataset can help discriminating through the different classes a Multivariate ANalysis Of VAriance (MANOVA) was performed.
From this analysis resulted that the variance of dataset was sufficient to distinguish four classes out of the five that were scored by clinicians.

In figure 5.1 it is possible to observe a dendrogram plot of the group means obtained from the MANOVA which clearly shows 2 main clusters: one containing normal and hypopnea epochs and one containing apnea epochs.

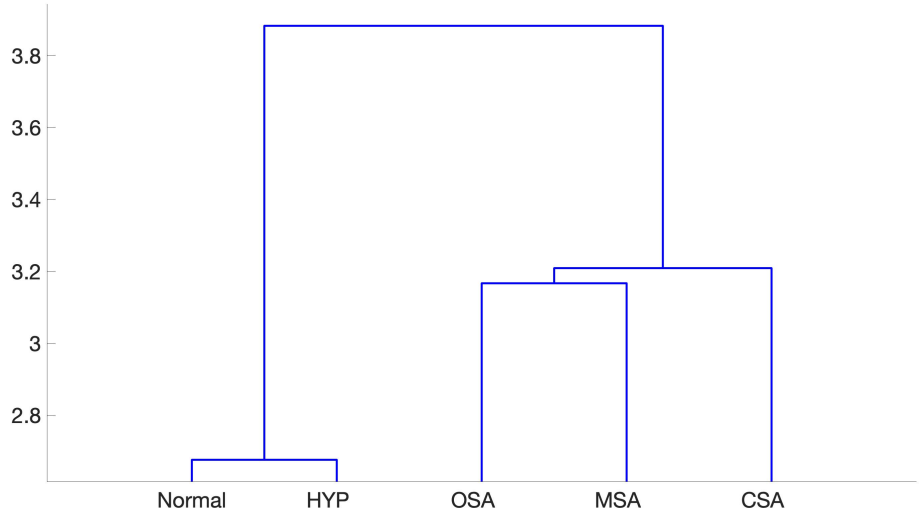


Figure 5.1: MANOVA dendrogram with 5 classes.

In order to numerically support this view, the distances between each pair of group means were reported in table 5.1.
From this table it is clear how near normal and hypopnea clusters are with respect to the apnea ones and how the apnea clusters are similar to each other. Because of this result it was firstly decided to consider the problem as a dichotomous classification problem between normal/hypopnea cluster and apnea cluster (OSA/MSA/CSA).

Normal	HYP	OSA	CSA	MSA
0	2.68	10.34	8.30	15.31
	0	3.88	3.98	8.28
		0	3.21	3.17
	symmetric		0	3.70
				0

Table 5.1: Distance between each pair of group means.

5.1.1 Dimensionality reduction

Since the observations are described by many features it was thought to apply some dimensionality reduction technique after data normalization.
In particular, all the features were normalized according to the z-score normalization. In this way, each features vector got null mean and standard deviation equal to 1.

Firstly, it was thought to apply a Principal Component Analysis (PCA), which is a feature transformation technique that generates principal components (PCs) orthogonal to each other (to avoid redundant information) making linear combinations of the original variables forming a new (orthogonal) basis for the space of the data. In particular, it was decided to consider a number of principal components that explained the 70% of total variance.
The basic idea is to check if great part of the variance can be explained by a limited number of PCs and use the latter to proceed with the classification.

The first 10 PCs are shown in the Pareto chart reported in figure 5.2 where the blue line indicates the cumulative sum of the variances explained by the principal components.
Looking at this chart it was obvious to conclude that there is a lot of redundant information since the first 10 PCs explained about the 60% of total variance and 17 PCs were needed to reach the aforementioned 70% of variance.

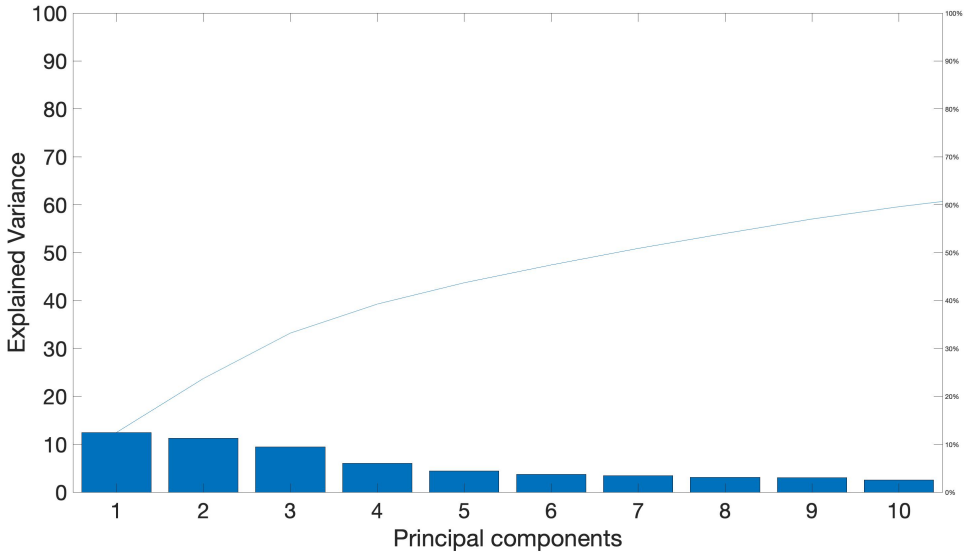


Figure 5.2: Pareto chart

To visualize the discriminating power of the features over the considered clusters a parallel coordinates chart was plotted.
It is reported in figure 5.3 and shows how the features median values (continuous lines) and interquartile ranges (dotted lines) vary between the two classes.
For visual purposes the plot is divided in 4 separate plots.
The idea was removing features whose median values and interquartile ranges are too overlapped on each other.

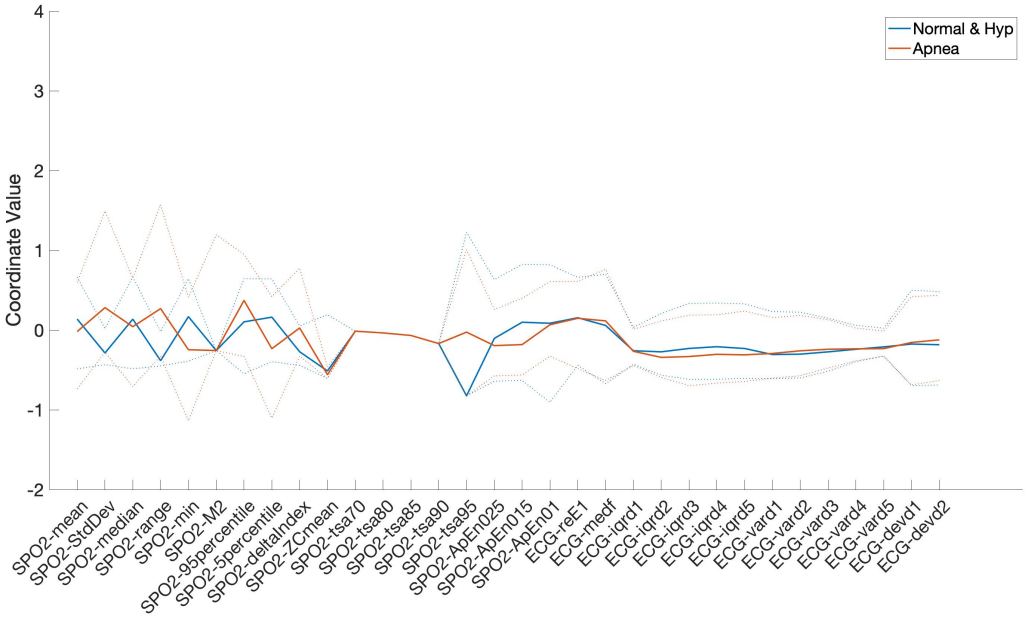
It was decided to remove from the dataset the following features:

- 5 blood saturation features: *TSA70, TSA80, TSA85 and TSA90* (4 hypoxic burden features) and *ApEn01* (1 complexity feature);
- 24 ECG features: from *median frequency* to *mad5* (21 frequency-related features) and *NN501, NN502* and *SDSD* (3 RR interval related features);
- 11 among EEG and EOG features: *EEG-O1M2medf, EEG-O1M2re, EEG-O1M2Ptot* and *EEG-O1M2PdPaPt* (4 from O1-M2 signal), *E1-Palpha, E1-Ptot* and *E1-PdPaPt* (3 from left EOG), *E2-Palpha, E2-Ptot, E2-PdPaPt* and *E2-PtPaPd* (4 from right EOG);
- 1 position feature: *standard deviation*;

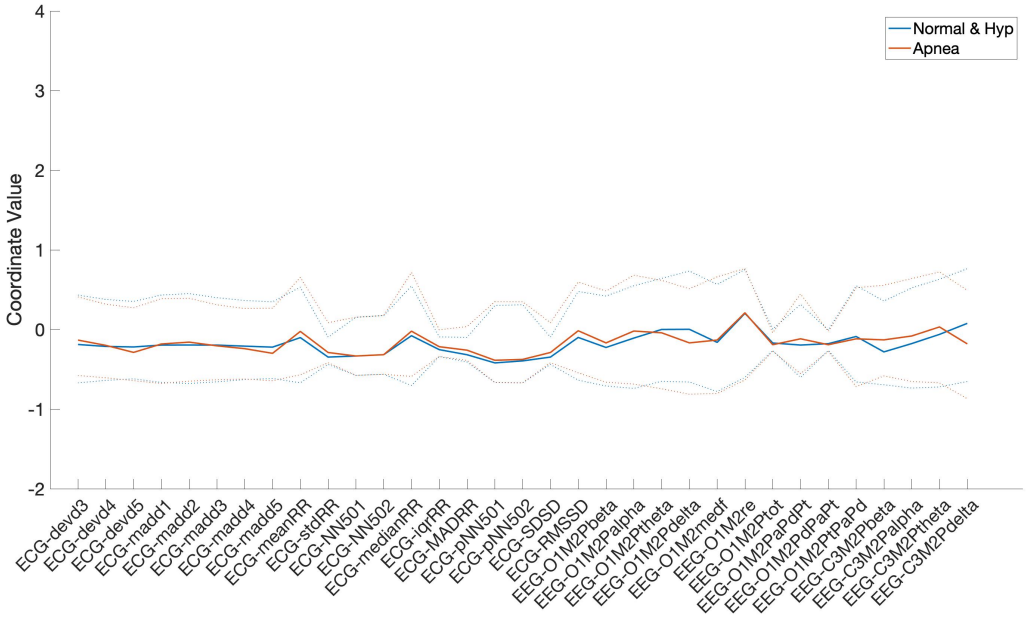
- 9 snore features: *energD3, energD2* and *median frequency* (3 frequency related features), *standard deviation, minimum value* and *interquartile range* (3 general statistics features), *re, SEpeaks* and *SEmean* (3 complexity measures);
- 2 thoracic volume features: *maxPeak* (1 peak related feature) and *energA* (1 frequency related feature).

A total of 52 features which represent the 40% of total were removed and then both MANOVA and PCA were performed again to understand how this removal influenced the variance of the data.

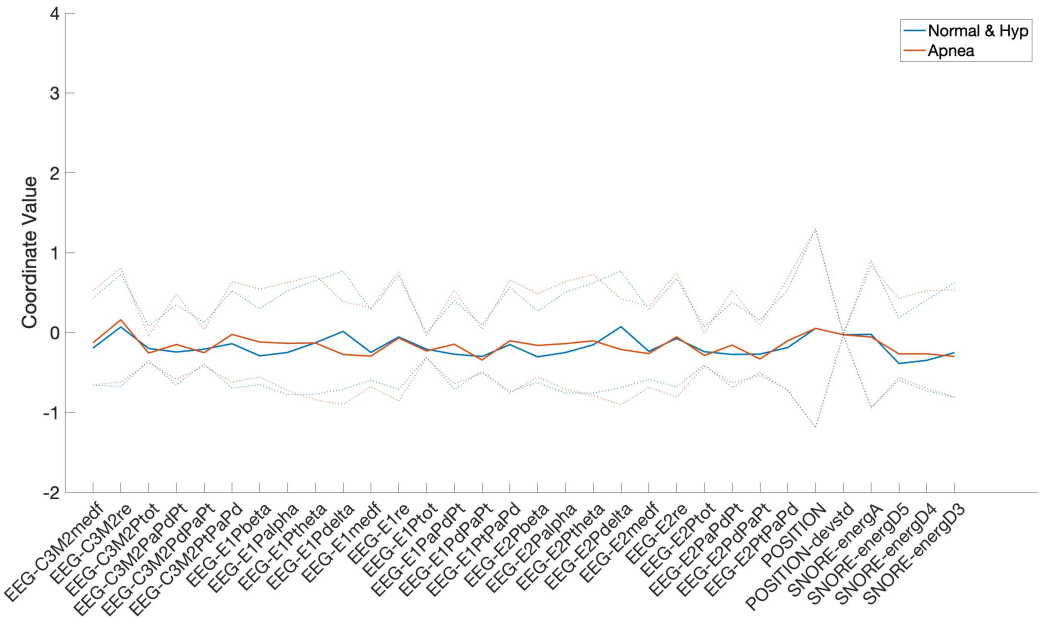
Qualitatively, one can expect that reducing the number of low discriminant features should increase the distance between clusters but in this case, since the normal cluster and the hypopnea cluster were merged as well as the three apnea forms clusters, it was expected a reduction of the distance between the two group means because from table 5.1 it is obvious that the hypopnea cluster has similar distances from normal, OSA and CSA clusters.



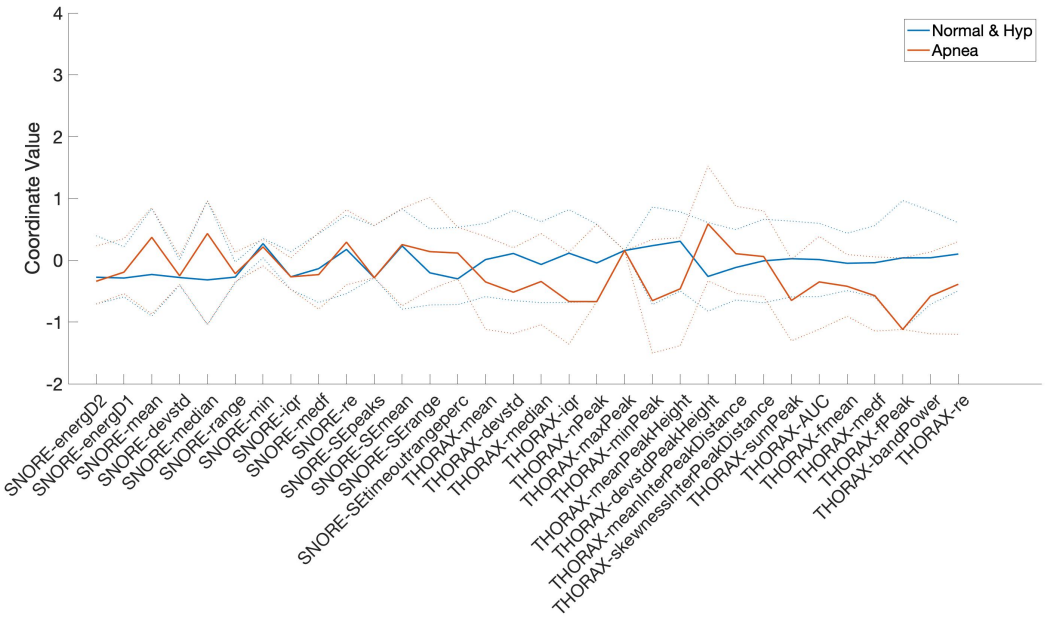
(a)


Normal & Hyp
Apnea
Coordinate Value
4
3
2
1
0
-1
-2
ECG-devd3
ECG-devd4
ECG-devd5
ECG-madd1
ECG-madd2
ECG-madd3
ECG-madd4
ECG-madd5
ECG-meanRR
ECG-stdRR
ECG-NN501
ECG-NN502
ECG-medianRR
ECG-iqrRR
ECG-MADRR
ECG-pNN501
ECG-pNN502
ECG-SDSD
ECG-RMSSD
EEG-O1M2Pbeta
EEG-O1M2Palpha
EEG-O1M2Ptheta
EEG-O1M2Pdelta
EEG-O1M2medf
EEG-O1M2re
EEG-O1M2Ptot
EEG-O1M2PaPdPt
EEG-O1M2PdPaPt
EEG-O1M2PtPaPd
EEG-C3M2Pbeta
EEG-C3M2Palpha
EEG-C3M2Ptheta
EEG-C3M2Pdelta

(b)


Normal & Hyp
Apnea
Coordinate Value
4
3
2
1
0
-1
-2
EEG-C3M2medf
EEG-C3M2re
EEG-C3M2Ptot
EEG-C3M2PaPdPt
EEG-C3M2PdPaPt
EEG-C3M2PtPaPd
EEG-E1Pbeta
EEG-E1Palpha
EEG-E1Ptheta
EEG-E1Pdelta
EEG-E1medf
EEG-E1re
EEG-E1Ptot
EEG-E1PaPdPt
EEG-E1PdPaPt
EEG-E1PtPaPd
EEG-E2Pbeta
EEG-E2Palpha
EEG-E2Ptheta
EEG-E2Pdelta
EEG-E2medf
EEG-E2re
EEG-E2Ptot
EEG-E2PaPdPt
EEG-E2PdPaPt
EEG-E2PtPaPd
POSITION
POSITION-devstd
SNORE-energA
SNORE-energD5
SNORE-energD4
SNORE-energD3

(c)



(d)

Figure 5.3: Parallel coordinates plots.

MANOVA results confirmed decreased distances between group means.

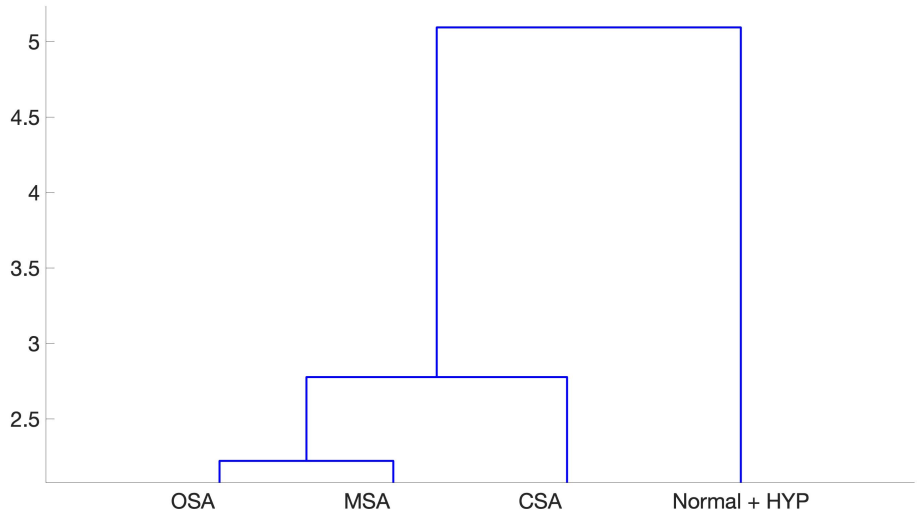


Figure 5.4: MANOVA dendrogram after features removal.

This is numerically observable in table 5.2.

Normal+HYP	**OSA**	**CSA**	**MSA**
0	5.90	5.09	9.33
	0	2.78	2.22
symmetric		0	2.79
			0

Table 5.2: Distance between each pair of group means.

From PCA no massive improvement was observed since the first 10 PCs explained about 65% of total variance and 11 PCs were needed to explain 70% of variance meaning that the removed features carried mainly redundant information.

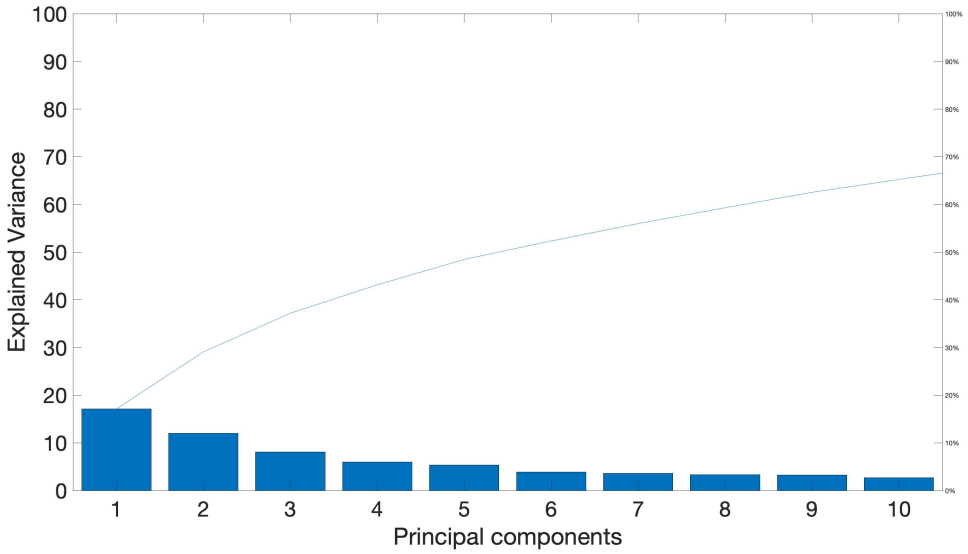


Figure 5.5: Pareto chart after features removal.

Before continuing with features removal, it was thought to proceed with a first classification test in order to get an idea of models performances on these data. Training set and test set were built according to the guidelines in subsection 4.2.2: since apnea cluster had less observations than normal/hypopnea cluster, it was created a training and a test set which contained all the possible apnea events of respective subjects and an equal number of randomly chosen epochs from the normal/hypopnea cluster paying attention to maintain the proportion between

normal and hypopnea clusters the same as in the original dataset.
In terms of number of observations (epochs) the training set and the test set contained, respectively, about 47'000 epochs and about 22'000 epochs.

5.1.2 Results

In the results reported below the negative class is indicated with number 1 and is referred to the normal/hypopnea cluster while the positive class is indicated with number 2 and is referred to the apnea cluster (OSA/MSA/CSA).
The resubstitution errors of the trained models are reported in table 5.3.

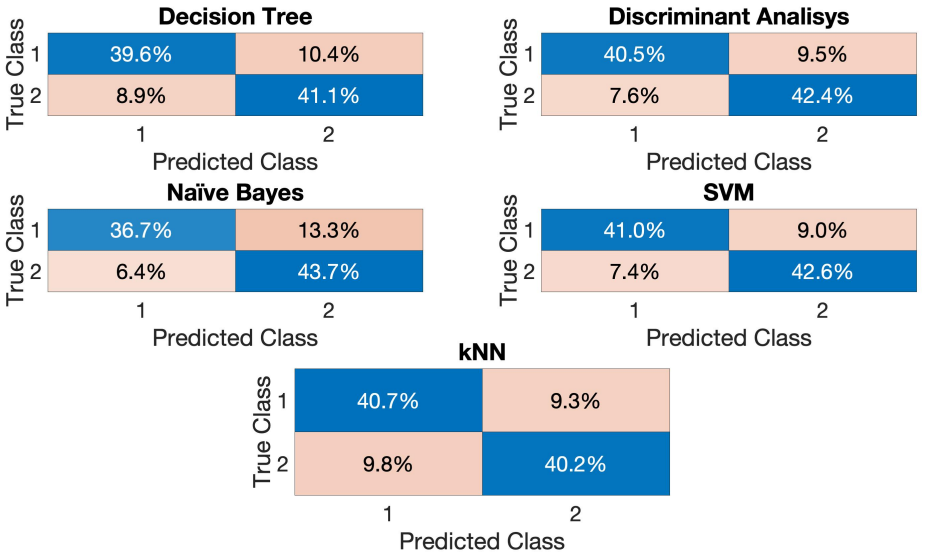
Model	Resubstitution Error
Decision Tree	15.53%
Discriminant Analysis	20.07%
Naïve Bayes	21.36%
SVM	19.51%
kNN	**10.86%**

Table 5.3: Resubstitution errors of the trained models for normal/hypoponea vs apnea classification considering 78 features.

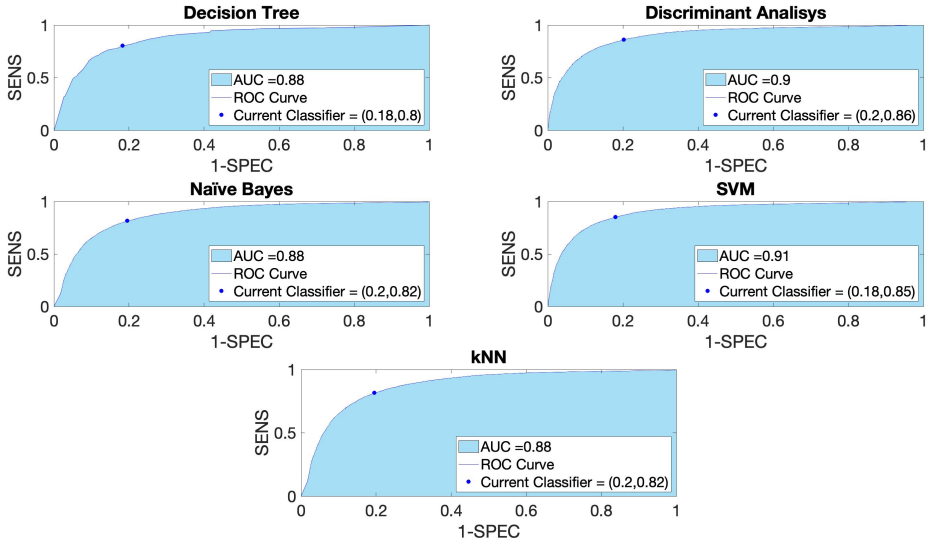
In table 5.4 are reported the prediction metrics and in figures 5.6a and 5.6b the relative confusion matrices and ROC curves.

Model	SENS	SPEC	PPV	NPV	ACC
Decision Tree	82.1%	79.3%	79.8%	81.6%	80.7%
Discriminant Analysis	84.8%	80.9%	81.6%	84.2%	82.9%
Naïve Bayes	**87.3%**	73.5%	76.7%	**85.3%**	80.4%
SVM	85.2%	**81.9%**	**82.5%**	84.7%	**83.6%**
kNN	80.4%	81.4%	81.2%	80.6%	80.9%

Table 5.4: Prediction metrics of the trained models for normal/hypoponea vs apnea classification considering 78 features.



(a)



(b)

Figure 5.6: Normal/hypopnea vs apnea results comparison.
a) Confusion matrices comparison (normalized over the total number of observations, in order to see the percentages of FP and FN).
b) ROC curves comparison.

From the previous analysis resulted that the best models is SVM since it had almost all prediction metrics above 82%, despite the model with the best resubstitution error was the kNN (10.86%).
In particular, Naive bayes had the highest SENS and NPV but lower ACC, SPEC and PPV than SVM. Moreover, the AUCs of all classifiers were similar with SVM's one being the best (0.91) and SVM's current classifier was slightly nearer to the top left corner. Furthermore, as observable in figure 5.6a, SVM model, in addition to having best diagnostic accuracy, returned a greater number of false positives rather than false negatives.

5.2 Normal vs Apnea - 1st Analysis

Results from the previous classification was considered not optimal so it was thought to analyze how merging normal and hypopnea clusters affected the classification, since hypopnea is basically an intermediate condition between normal and apneic ones that was thought to be prone to missclassification (see Appendix B for detailed definition and scoring procedures).
Therefore it was considered a scenario where the aim was to distinguish normal epochs from apnea epochs using the same features of previous analysis without the normal/hypopnea merging previously considered.

5.2.1 Dimensionality reduction

In this perspective both MANOVA and PCA were performed and, as expected, results from MANOVA revealed that the distance between group means increased considering the same features, as shown in table 5.5 and in figures 5.7.

Normal	**OSA**	**CSA**	**MSA**
0	17.25	11.62	22.54
	0	4.56	2.92
		0	4.53
symmetric			0

Table 5.5: Distance between each pair of group means.

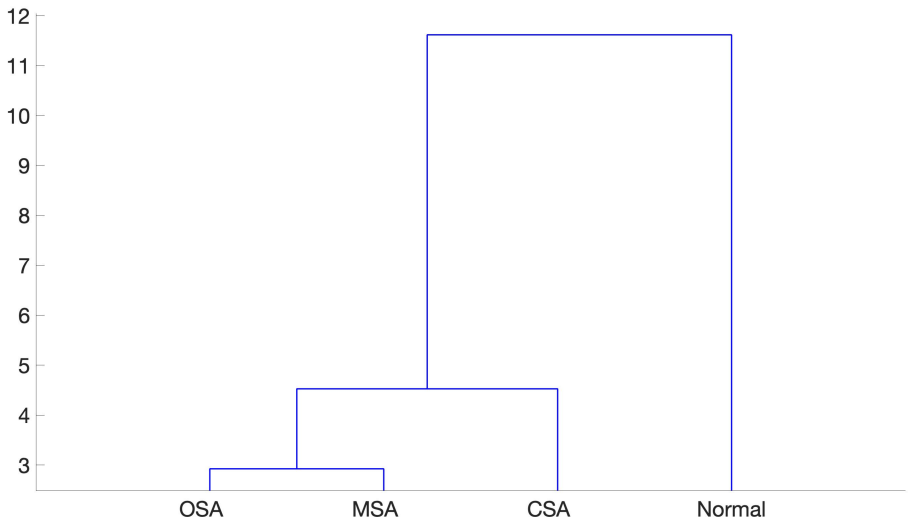


Figure 5.7: MANOVA dendrogram after features removal.

From PCA no significant improvement was observed since, similarly to the first analysis, 11 PCs were needed to explain 70% of variance.

5.2.2 Results

In the results reported below the negative class is indicated with number 1 and is referred to the normal cluster while the positive class is indicated with number 2 and is referred to the apnea cluster (OSA/MSA/CSA).
The classification errors of the trained models are reported in table 5.6.

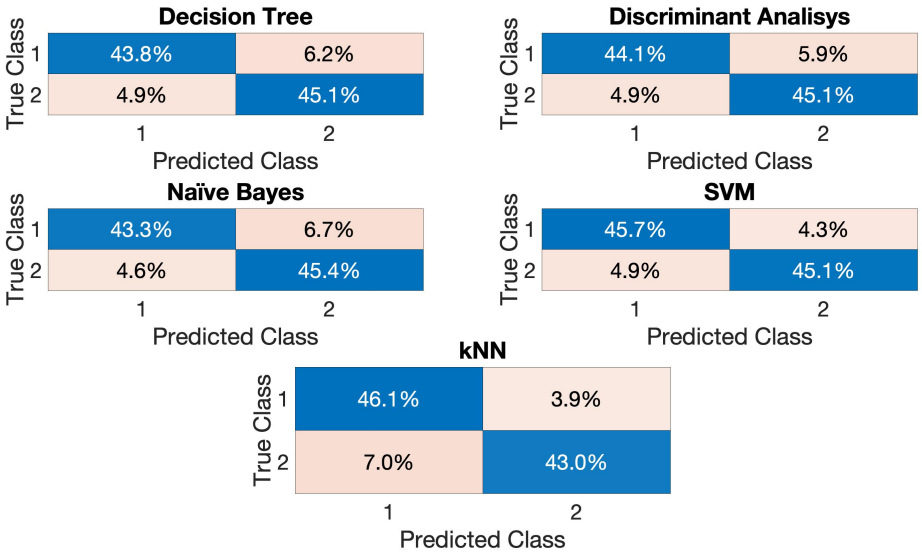
Model	Resubstitution Error
Decision Tree	8.45%
Discriminant Analysis	14.11%
Naïve Bayes	14.89%
SVM	12.20%
kNN	**5.95%**

Table 5.6: Resubstitution errors of the trained models for normal vs apnea classification considering 78 features.

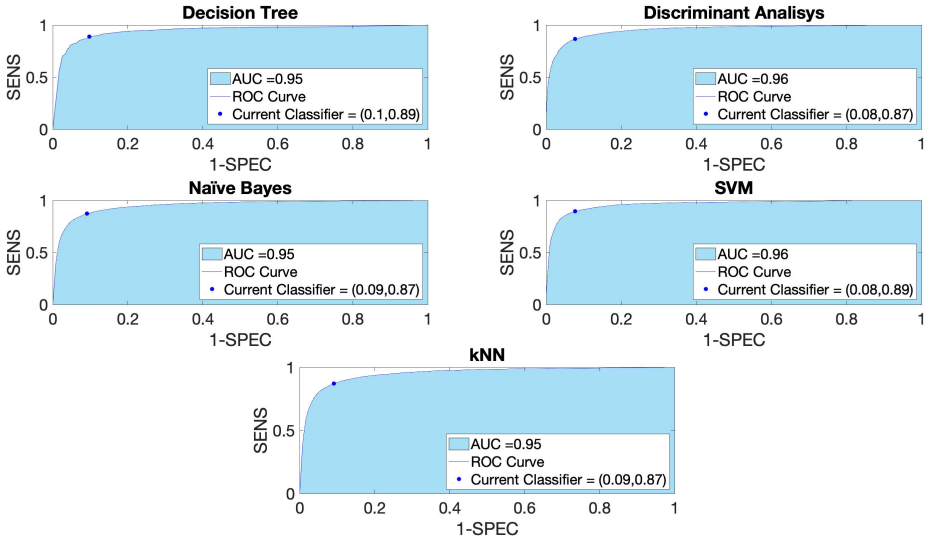
In table 5.7 are reported the prediction metrics and in figures 5.8a and 5.8b the relative confusion matrices and ROC curves.

Model	SENS	SPEC	PPV	NPV	ACC
Decision Tree	90.3%	87.7%	88.0%	90.0%	88.9%
Discriminant Analysis	90.1%	88.2%	88.4%	89.9%	89.2%
Naïve Bayes	**90.7%**	86.6%	87.1%	**90.3%**	88.7%
SVM	90.2%	91.4%	91.3%	**90.3%**	**90.8%**
kNN	85.9%	**92.2%**	**91.7%**	86.7%	89.1%

Table 5.7: Prediction metrics of the trained models for normal vs apnea classification considering 78 features.



(a)



(b)

Figure 5.8: Normal vs apnea results comparison.
a) Confusion matrices comparison (normalized over the total number of observations, in order to see the percentages of FP and FN).
b) ROC curves comparison.

From the previous analysis resulted that the best tradeoff was obtained by the SVM model since it had almost all prediction metrics above 90%, despite the model with the best resubstitution error was the kNN (5.95%). In particular, naïve bayes had the highest SENS and the same NPV but lower ACC, SPEC and PPV than SVM while the kNN had higher SPEC and PPV but lower SENS, NPV and ACC than SVM.
Moreover, the AUCs of all classifiers were similar with SVM's and discriminant analysis's ones being the highest (0.96) but with SVM's current classifier slightly nearer to the top left corner.
Unlike analysis of section 5.1, as observable in figure 5.8a, SVM classifier returned a greater number of false negatives rather than false positives.

The previous two analyses showed that SVM had the best performance tradeoff. In particular, after the hyperparameters optimization processes, in both cases, a linear kernel with a different scale (1.3299 for the first analysis and 0.1627 for the second analysis) was chosen, meaning that data are linearly separable using an hyperplane. Notice that training an SVM with a linear kernel is faster than with any other Kernel since less parameters need to be optimized.

5.3 Normal & HYP vs Apnea - 2nd Analysis

Looking at the results obtained from the analysis of section 5.1 it was thought to proceed with a more radical feature selection with the intention of keeping only the features that discriminates more according to the parallel coordinate chart of figure 5.3.

5.3.1 Dimensionality reduction

It was decided to remove all the features whose median values were almost superimposed and whose interquartile were too similar.
After this procedure remained the following 32 features:

- 9 blood saturation features: *standard deviation, range, min, 95th percentile, 5th percentile, delta index* (6 general statistics features), *TSA95* (1 hypoxic burden feature), *ApEn025 and ApEn015* (2 complexity features);
- 3 EOG features: *E1-Pbeta, E1-Pdelta* (from left EOG) and *E2-Pdelta* (from right EOG);
- 4 snore features: *mean and median* (2 general statistic features) and *SErange and SEtimeoutrangeperc* (2 complexity features);
- 16 thoracic volume features: *mean, standard deviation, median, interquartile range* (4 general statistic features), *nPeak, minPeak, meanPeakHeight, devstdPeakHeight, meanInterPeakDistance, sumPeak, AUC* (7 peak related features), *mean and median frequencies*, *peak frequency* and *band power* (4 frequency related features) and *Renyi's entropy* (1 complexity feature).

A total of 98 features which represent about the 75% of total were removed.
It is important to notice that all the EEG, ECG and position features were not considered.

Subsequently, both MANOVA and PCA were performed again to understand how this removal influenced the variance of the data.
From MANOVA resulted that distance between clusters' centroids slightly decreased with respect to the same distance in the case with 78 features (from 2.61 to 2.31). This reduction suggested that models' performances may have worsened.

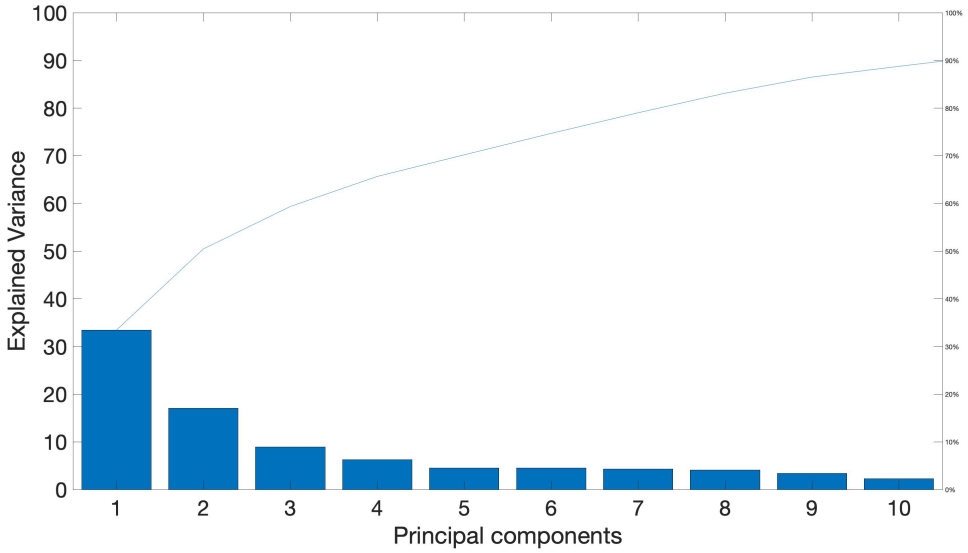


Figure 5.9: Pareto chart after features removal for normal/hypopnea vs apnea classification.

Obviously, from PCA an improvement was noted mainly due to the smaller number of features considered with the first 10 PCs explaining about the 90% of the total variance. In this case 5 PCs were needed to explain 70% of variance meaning that there still was redundant information within the data.

To clarify where redundant information was located, the absolute value of correlation coefficients between each pair of features was calculated and is shown in figure 5.10. Moreover, in order to highlight the highest correlations all the correlation coefficients less than 0.5 were capped to 0.
From this figure it is clear how features tend to be highly correlated to other features derived from the same signal as can be seen observing the cluster of blood saturation features and the cluster of thoracic volume features, which are the most numerous.

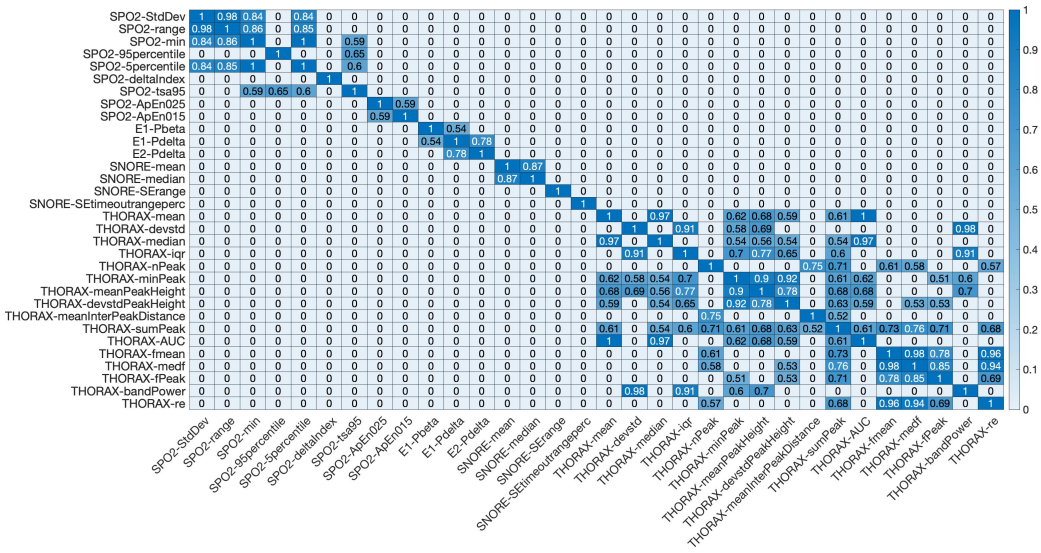
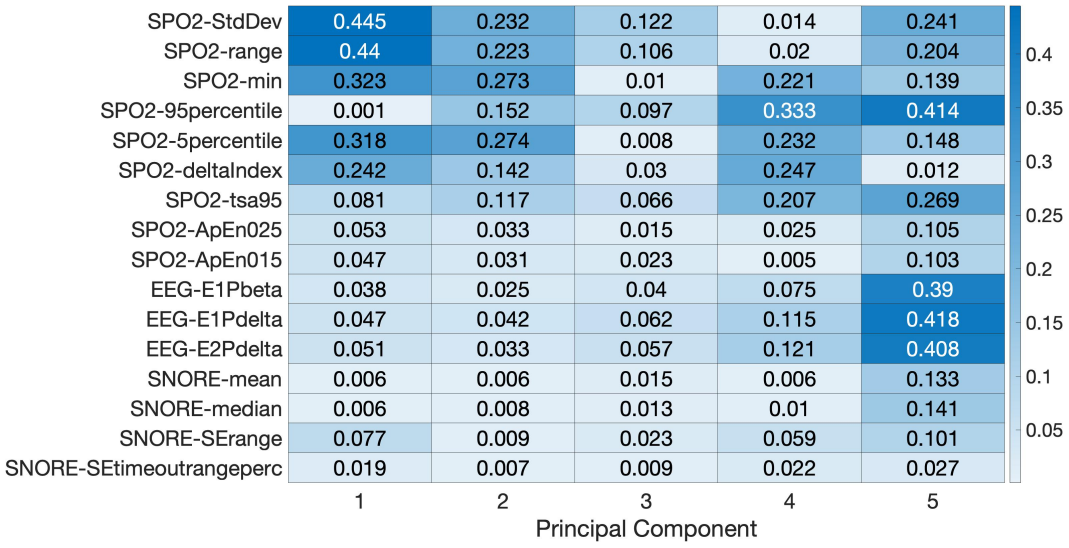
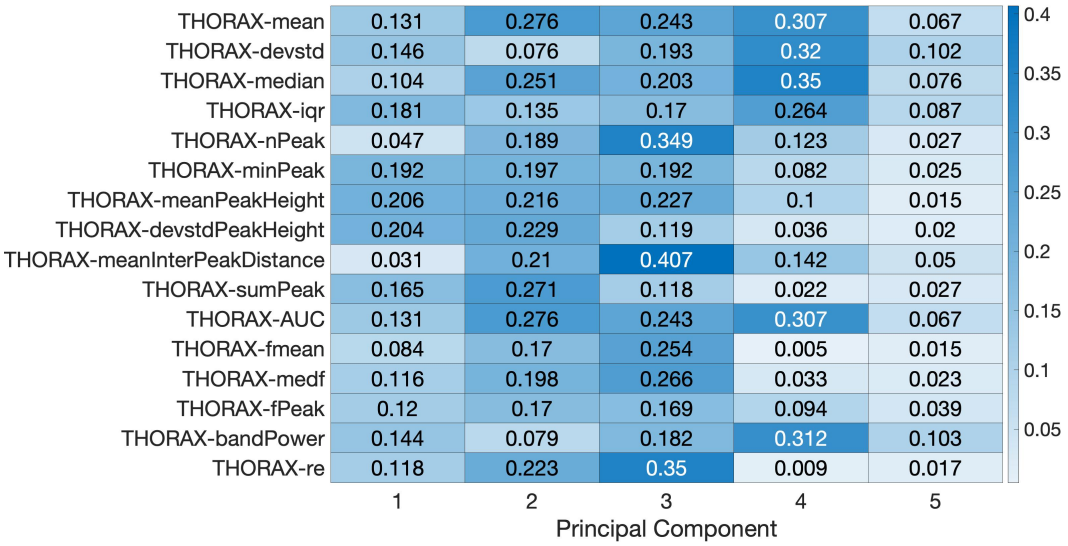


Figure 5.10: Correlation matrix in absolute values for normal/hypopnea vs apnea classification considering 32 features.

Moreover, it was decided to plot an heatmap, shown in figures 5.11 of the first 5 PCs in order to understand how much these features weight in terms of explained variance. For visual purposes the heatmap is splitted in half.



(a)



(b)

Figure 5.11: Heatmap of the first 5 principal components for normal/hypopnea vs apnea classification considering 32 features.

From figure 5.9 it is noticeable that the first 2 PCs explain about the 50% of variance and from the previous heatmap it is clear that these PCs are mainly influenced by SpO_2 general statistic features and thoracic volume features (more the second than the first) and less influenced by the other features.

To visually notice what the heatmap is trying to express numerically, it is helpful to look at the biplot of figure 5.12 where each point was labeled in order to recognize the relative feature.
To better visualize features spread it was decided to divide the biplot in four smaller biplots with the lower two containing the thoracic volumes features and the upper two containing the remaining features.
The points' projections on the axes represent the weight that that specific feature has within the principal component which lies on that axis.
Attention has to be payed to axis scale.

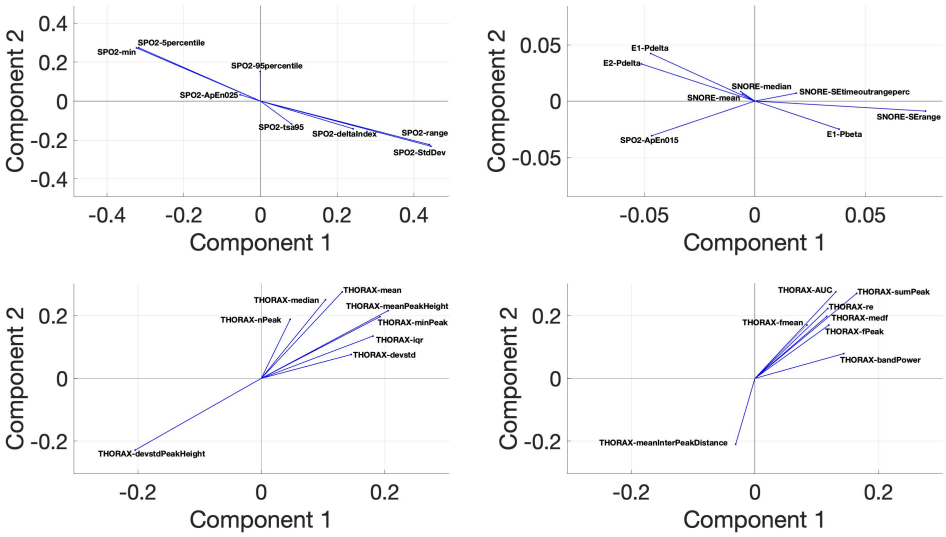


Figure 5.12: Biplot of the first 2 PCs for normal/hypopnea vs apnea classification considering 32 features.

In this condition it was decided to proceed with another classification test in order to examine model performances after this more substantial features removal.

5.3.2 Results

In this case, negative and positive class are the same of subsection 5.1.2.
The resubstitution errors of the trained models are reported in table 5.8.

Model	Resubstitution Error
Decision Tree	20.81%
Discriminant Analysis	22.29%
Naïve Bayes	22.92%
SVM	21.77%
kNN	**19.98%**

Table 5.8: Resubstitution errors of the trained models for normal/hypopnea vs apnea classification considering 32 features.

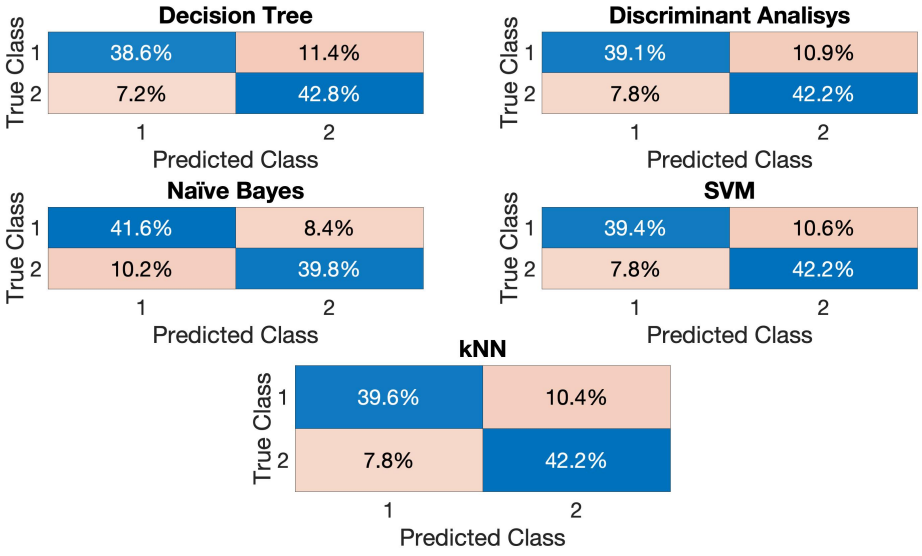
The resubstitution errors have worsened compared to the ones of table 5.3 with kNN being the most worsened but still the best.

In table 5.9 are reported the prediction metrics and in figures 5.13a and 5.13b the relative confusion matrices and ROC curves.

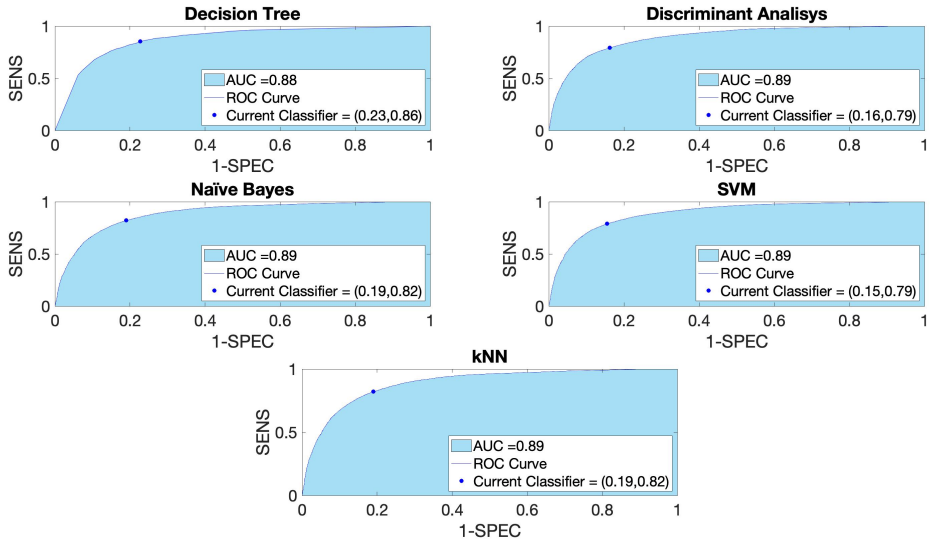
Model	**SENS**	**SPEC**	**PPV**	**NPV**	**ACC**
Decision Tree	**85.5%**	77.3%	79.0%	**84.2%**	81.4%
Discriminant Analysis	84.4%	78.2%	79.5%	83.4%	81.3%
Naïve Bayes	79.6%	**83.2%**	**82.6%**	80.3%	81.4%
SVM	84.3%	78.8%	79.9%	83.4%	81.6%
kNN	84.4%	79.2%	80.2%	83.5%	**81.8%**

Table 5.9: Prediction metrics of the trained models for normal/hypopnea vs apnea classification considering 32 features.

From the previous tables it is clear that models' performances were deteriorated with respect to models' performances of the first analysis. In this case the best tradeoff between prediction metrics was the kNN. It resulted the most accurate model and shown with seconds highest values for all the remaining metrics.
Comparing this model to the best of the first analysis (SVM) it is clear that all the prediction metrics decreased: SENS (-0.8%), SPEC (-2.7%), PPV (-2.3%), NPV (-1.2%) and ACC (-1.8%).
The AUCs of all models were similar with decision tree's one being slightly lower. Furthermore, as observable in figure 5.13a, kNN classifier, in addition to having best accuracy, returns a greater number of false positives (FP) rather than false negatives.



(a)



(b)

Figure 5.13: Normal/hypopnea vs apnea results comparison considering 32 features.
a) Confusion matrices comparison (normalized over the total number of observations, in order to see the percentages of FP and FN).
b) ROC curves comparison.

5.4 Normal vs Apnea - 2nd Analysis

This second analysis was carried out, like the one in section 5.2, to examine models' performances with the new 32-features dataset obtained in subsection 5.3.1 with the intention of distinguishing between normal and apnea clusters.

5.4.1 Dimensionality reduction

After removing the features identified in subsection 5.3.1 both MANOVA and PCA were performed to understand how this removal influenced the variance of the data. From MANOVA resulted that the distance between clusters' centroids was almost the double (4.23) of the same distance when normal and hypopnea clusters are merged (2.31).

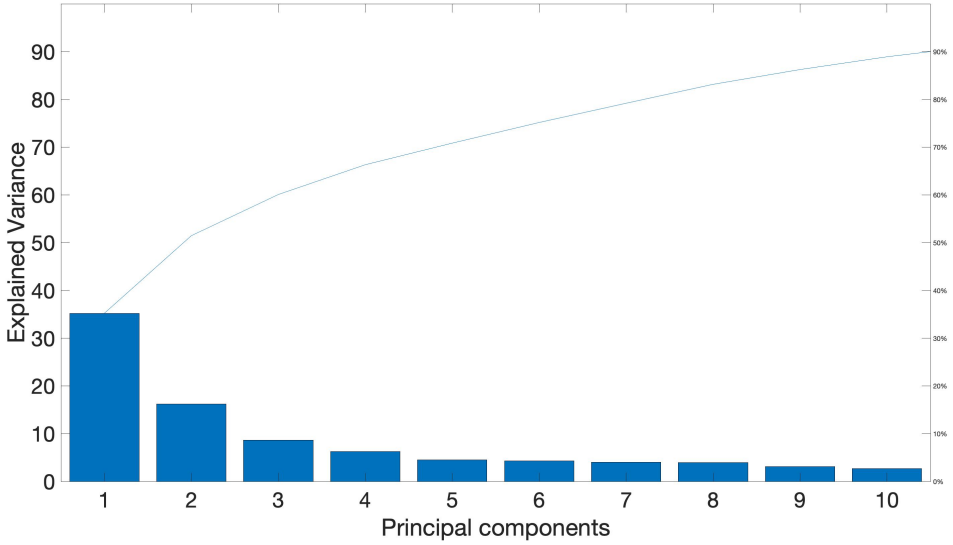


Figure 5.14: Pareto chart after features removal.

Similarly to what seen in subsection 5.3.1, from PCA little improvement was observed mainly due to the smaller number of features considered: the first 10 PCs explained almost the 90% of total variance and at least 5 PCs continued to be needed to explain 70% of variance meaning that there still was some redundant information left from the previous features removal.

5.4.2 Results

In this case, negative and positive class are the same of subsection 5.2.2.
The classification errors of the trained models are reported in table 5.10.

Model	Resubstitution Error
Decision Tree	**8.70%**
Discriminant Analysis	15.22%
Naïve Bayes	14.50%
SVM	13.26%
kNN	8.97%

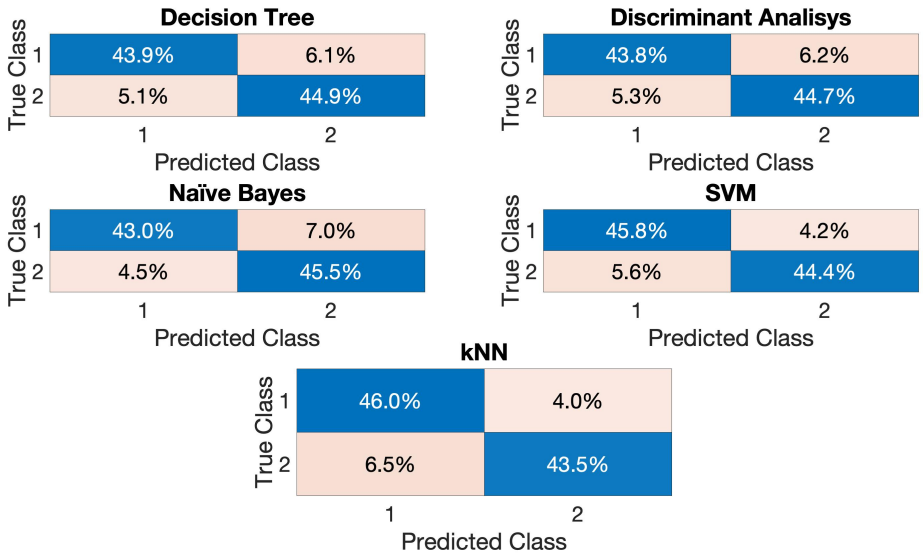
Table 5.10: Resubstitution errors of the trained models for normal vs apnea classification considering 32 features.

The resubstitution errors have worsened compared to the ones of table 5.6. kNN resulted the most worsened (from 5.95% to 8.97%) and decision tree resulted the best with an error of 8.70%. However this two models showed very similar and much lower resubstitution errors than other ones.

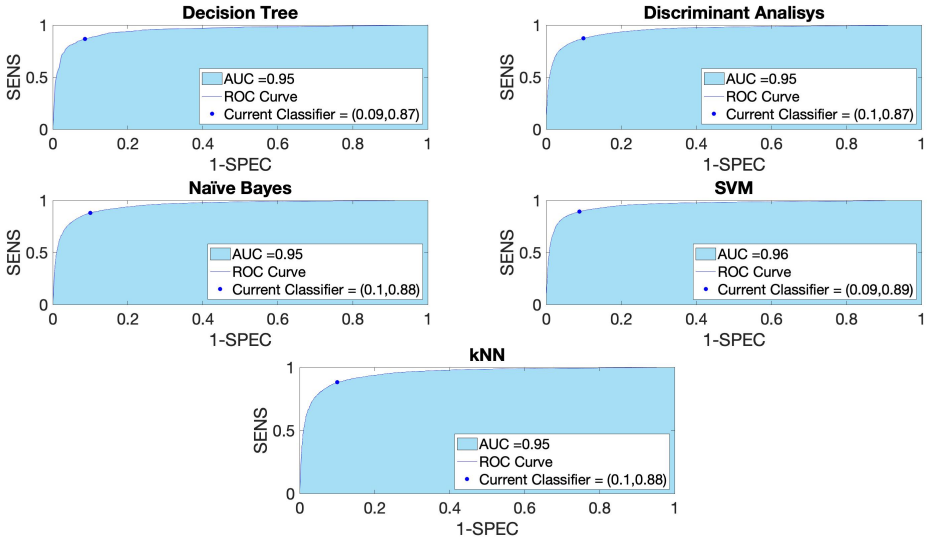
In table 5.11 are reported the models' prediction metrics and in figure 5.15 the relative confusion matrices and ROC curves.

Model	SENS	SPEC	PPV	NPV	ACC
Decision Tree	89.8%	87.7%	88.0%	89.6%	88.8%
Discriminant Analysis	89.3%	87.6%	87.8%	89.1%	88.5%
Naïve Bayes	**90.9%**	86.0%	86.6%	**90.5%**	88.5%
SVM	88.9%	91.5%	91.3%	89.2%	**90.2%**
kNN	86.9%	**92.0%**	**91.6%**	87.6%	89.5%

Table 5.11: Normal vs apnea prediction metrics considering 32 features.



(a)



(b)

Figure 5.15: Normal vs apnea results comparison considering 32 features.
a) Confusion matrices comparison (normalized over the total number of observations, in order to see the percentages of FP and FN).
b) ROC curves comparison.

From this analysis resulted that the model with the highest diagnostic accuracy is SVM (90.2%), despite not having any other prediction metric that outperform the other models and not having one of the best resubstitution errors (13.26%). In particular, naïve bayes had the highest SENS (90.9%) and NPV (90.5%) while kNN had the highest SPEC (92.0%) and PPV (91.6%).
Moreover, the AUCs of all classifiers were similar with SVM's one being the best (0.96) having current classifier was slightly nearer to the top left corner.
Like analysis of section 5.2, as observable in figure 5.15a, SVM classifier returned a greater number of false negatives rather than false positives.

The best performance tradeoff from the third and fourth analyses were obtained with two different models: kNN for the first one and SVM for the second one. Nevertheless, SVM's performances of the third analysis was not so worse than kNN's ones and from SVM's hyperparameters optimization process resulted a linear kernel. The same thing appened for the fourth analysis where the SVM model was the best. This led to think that features removal kept data linearly separable with an hyperplane, similarly to the first two analysis.

5.5 OSA & MSA vs CSA

So far there has been talk of classification tests whose aim is to distinguish between general apneic condition from another condition (normal or normal/hypopnea clusters).
In this section it was thought to examine classification performances of models specifically trained to discriminate apnea forms. Looking at figure 5.1 it was clear that OSA and MSA clusters were nearer to each other with respect to CSA, so it was decided to merge this two clusters and to consider the problem as a dichotomous classification problem between OSA/MSA cluster and CSA cluster.
It was decided to analyze this case because it was thought that, rather than training a single model to try distinguish between all the considerable classes, it could be more efficient to build a cascading structure of dichotomous classifiers where the first level is trained to recognize the presence or not of apnea and the second level is trained to distinguish between different apnea forms.

5.5.1 Dimensionality Reduction

Following the same reasoning of the previous cases, before proceeding with the classification test MANOVA and PCA were performed.
MANOVA results confirmed that the variance of dataset was enough among 2 classes out of three apnea classes.

The Pareto chart of figure 5.16 clarified that there was a lot of redundant information since the first 10 PCs explained about the 60% of total variance and 17 PCs were needed to reach the usual 70% of variance.

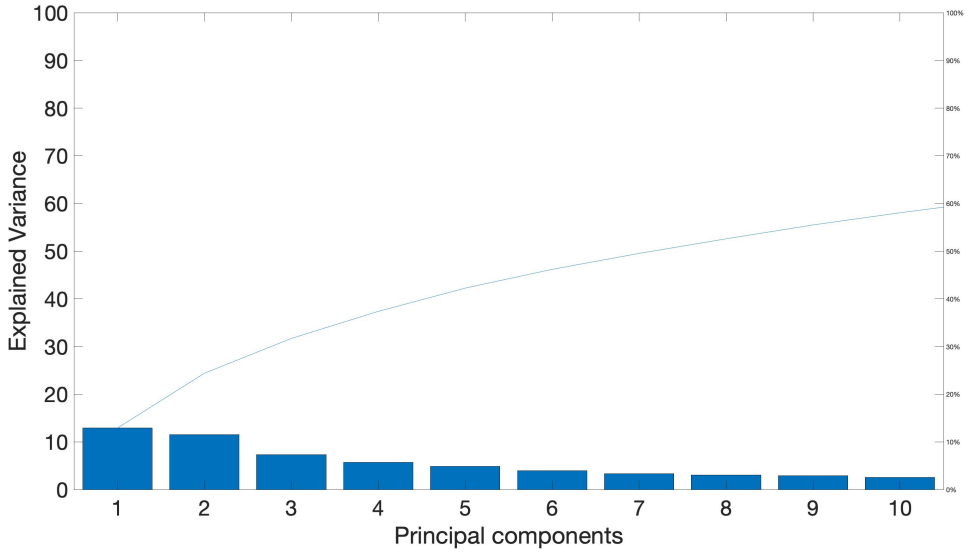


Figure 5.16: Pareto chart

As done before, in order to reduce the number of features a a parallel coordinates chart was plotted and as many low discriminating features (whose median values and interquartile ranges are too overlapped on each other) as possible were removed.

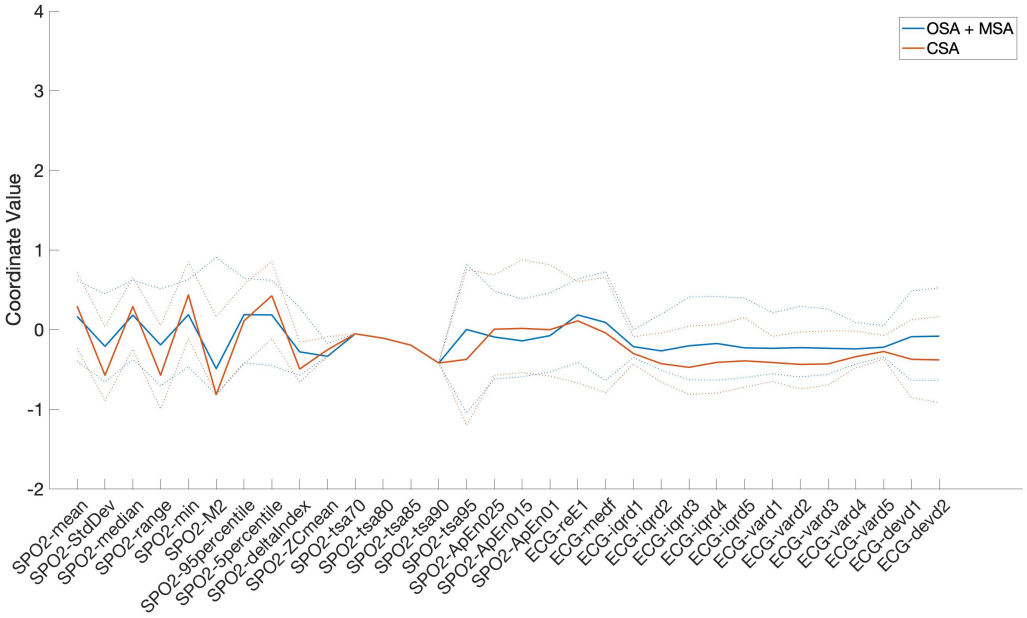
The following 44 features passed the removal process:

- 9 blood saturation features: *standard deviation, range, minimum value, M2, 5th percentile and delta index* (6 general statistics features), *TSA95* (1 hypoxic burden feature), *ApEn025 and ApEn015* (2 complexity features);
- 18 ECG features: *interquartile ranges of 3rd, 4th and 5th detail levels of wavelet decomposition, variance of 1st, 2nd and 3d detail levels of wavelet decomposition, standard deviation and MAD of all detail levels of wavelet decomposition* (16 frequency related features) and *mean and median of RR intervals* (2 RR-intervals related features;
- 1 position feature: *position numerical encoding*;
- 4 snore features: *energy in the 3rd and 4th detail levels of wavelet decomposition* (2 frequency related features) and *instantaneous spectral entropy range*

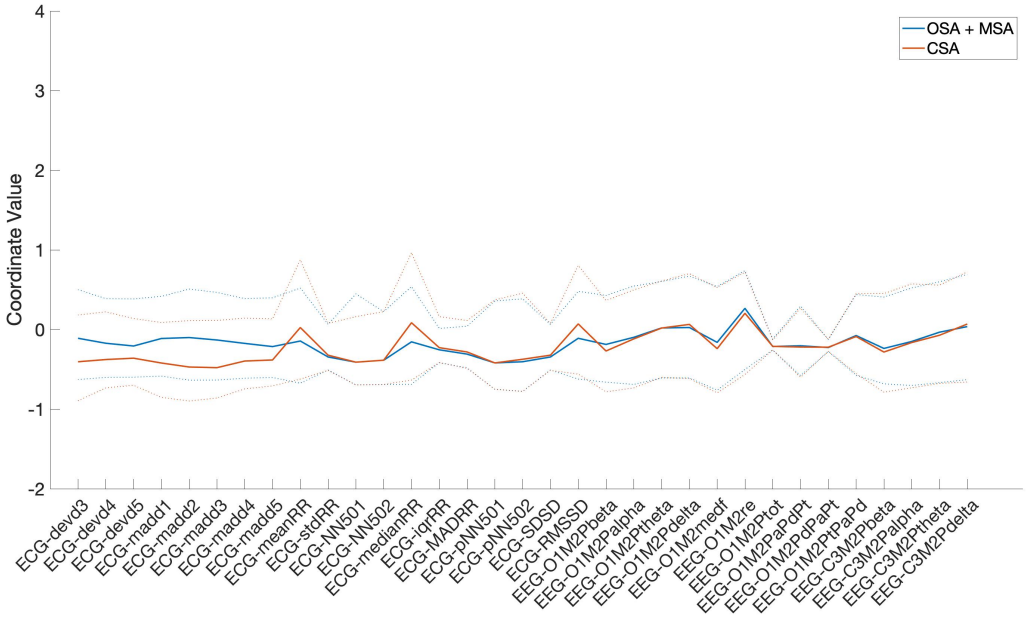
and Renyi's entropy (2 complexity features);

- 12 thoracic volume features: *standard deviation and interquartile range* (2 general statistic features), *number of peaks, minimum peak value, mean peak height, standard deviation of peaks height, mean inter-peak distance, skewness of inter-peak distance, sum of the peak values and AUC* (8 peak related features), *mean frequencies and band power* (2 frequency related features).

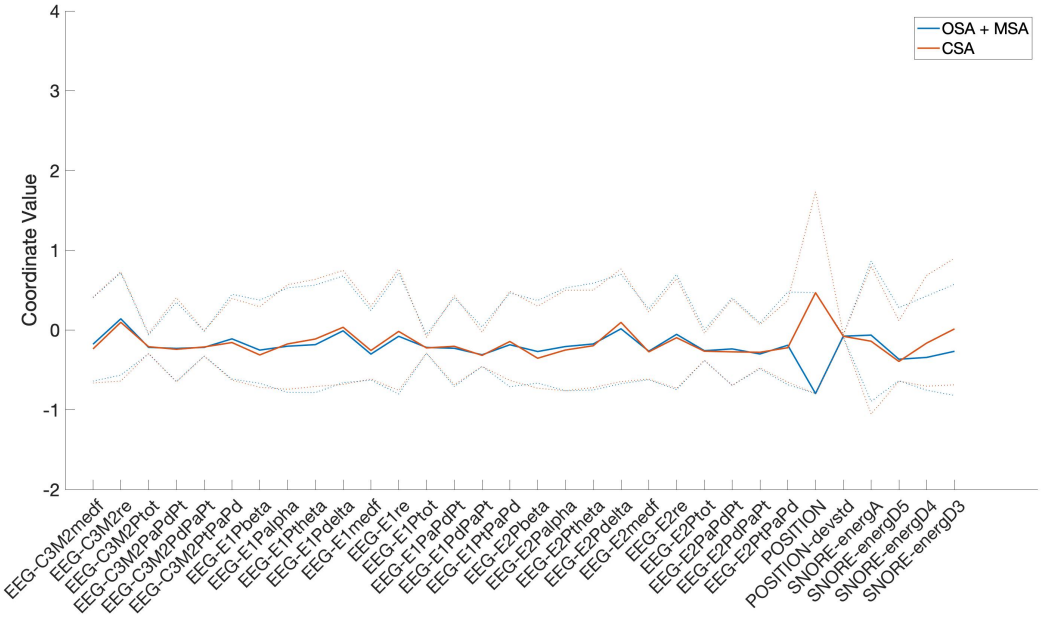
A total of 86 features which represent about the 66% of total were removed.
It is important to note that all the EEG and EOG features were excluded and, differently from all the previous cases, this time the ECG signal features resulted having an higher discriminating power, in fact almost all ECG features were considered.



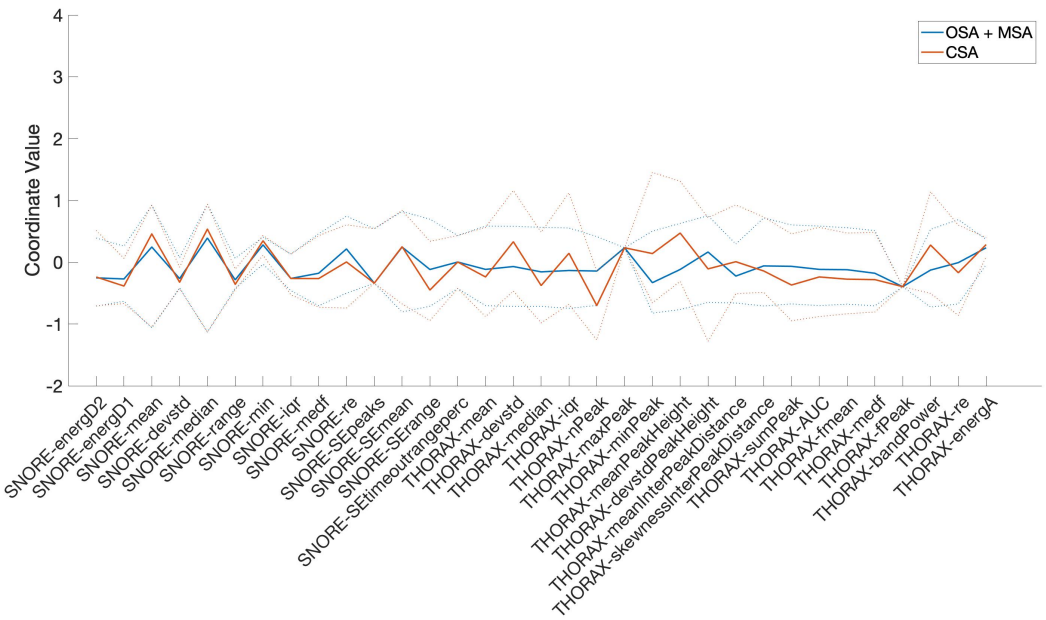
(a)


OSA + MSA
CSA
Coordinate Value
4
3
2
1
0
-1
-2
ECG-devd3
ECG-devd4
ECG-devd5
ECG-madd1
ECG-madd2
ECG-madd3
ECG-madd4
ECG-madd5
ECG-meanRR
ECG-stdRR
ECG-NN501
ECG-NN502
ECG-medianRR
ECG-iqrRR
ECG-MADRR
ECG-pNN501
ECG-pNN502
ECG-SDSD
ECG-RMSSD
EEG-O1M2Pbeta
EEG-O1M2Palpha
EEG-O1M2Ptheta
EEG-O1M2Pdelta
EEG-O1M2medf
EEG-O1M2re
EEG-O1M2Ptot
EEG-O1M2PaPdPt
EEG-O1M2PdPaPt
EEG-O1M2PtPaPd
EEG-C3M2Pbeta
EEG-C3M2Palpha
EEG-C3M2Ptheta
EEG-C3M2Pdelta

(b)


OSA + MSA
CSA
Coordinate Value
4
3
2
1
0
-1
-2
EEG-C3M2medf
EEG-C3M2re
EEG-C3M2Ptot
EEG-C3M2PaPdPt
EEG-C3M2PdPaPt
EEG-C3M2PtPaPd
EEG-E1Pbeta
EEG-E1Palpha
EEG-E1Ptheta
EEG-E1Pdelta
EEG-E1medf
EEG-E1re
EEG-E1Ptot
EEG-E1PaPdPt
EEG-E1PdPaPt
EEG-E1PtPaPd
EEG-E2Pbeta
EEG-E2Palpha
EEG-E2Ptheta
EEG-E2Pdelta
EEG-E2medf
EEG-E2re
EEG-E2Ptot
EEG-E2PaPdPt
EEG-E2PdPaPt
EEG-E2PtPaPd
POSITION
POSITION-devstd
SNORE-energA
SNORE-energD5
SNORE-energD4
SNORE-energD3

(c)



(d)

Figure 5.17: Parallel coordinates plots.

Subsequently, both MANOVA and PCA were performed again in order to understand how this removal influenced the variance of the data.
As shown in figure 5.18, from PCA resulted a little improvement since the first 10 PCs explained slightly more than 80% of total variance and 7 PCs were needed to reach the usual 70% of variance meaning that there still was some redundant information within the data.
Nevertheless, from MANOVA resulted that distance between clusters' centroids slightly decreased with respect to the same distance before features removal passing from 1.68 to 1.27.

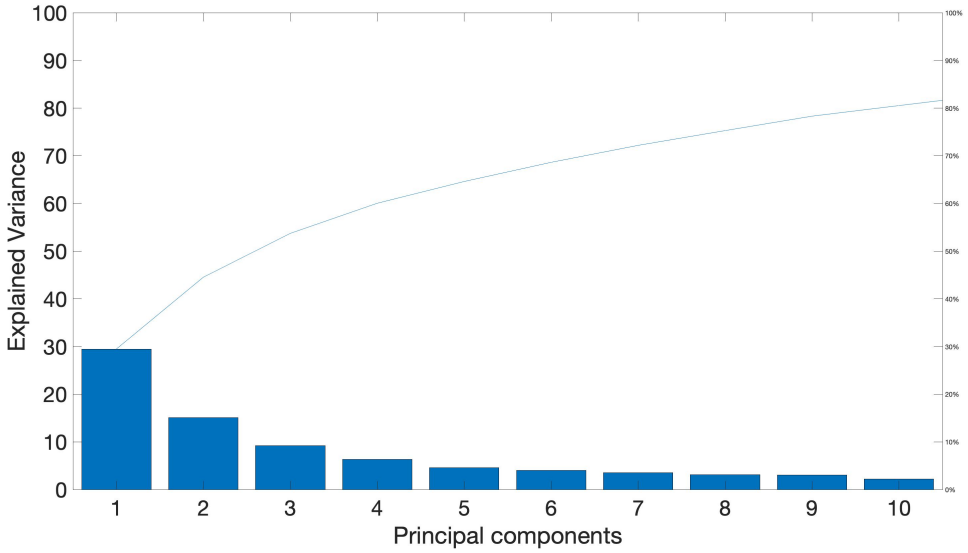


Figure 5.18: Pareto chart after feature removal

To clarify where redundant information was located, a correlation matrix as the one of subection 5.3.1 was calculated. It is shown in figure 5.19

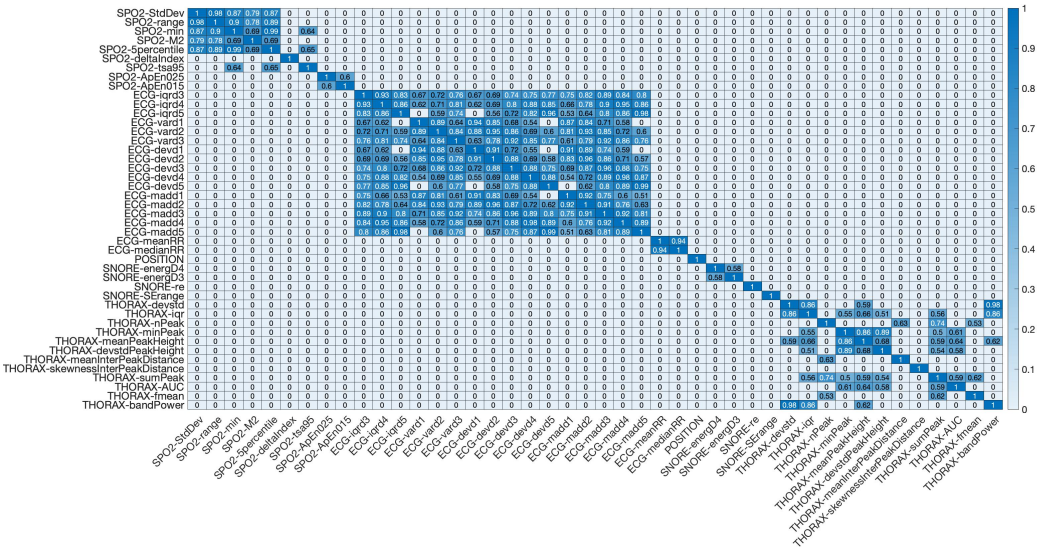
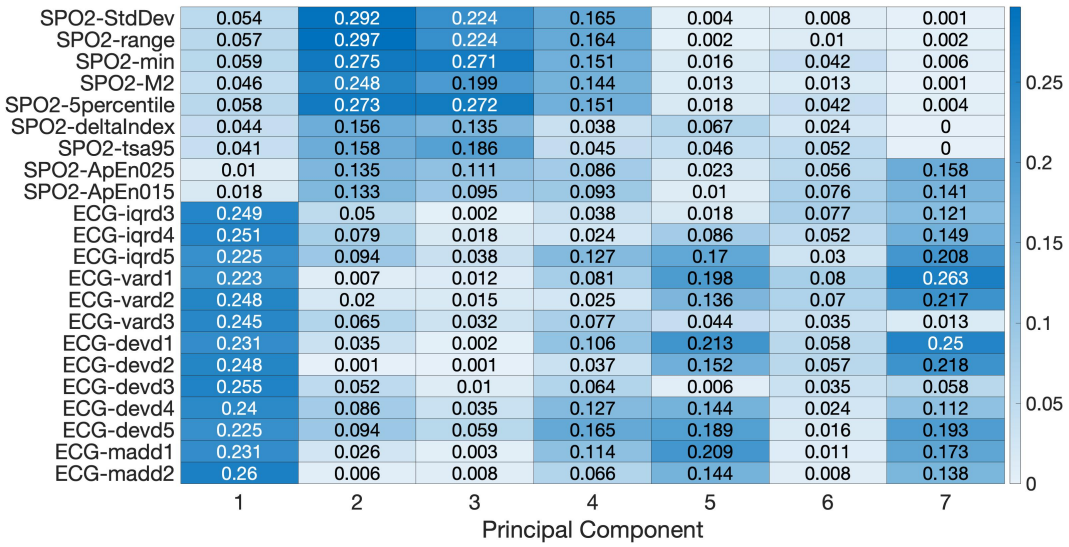


Figure 5.19: Correlation matrix in absolute values for OSA/MSA vs CSA classification considering 44 features.

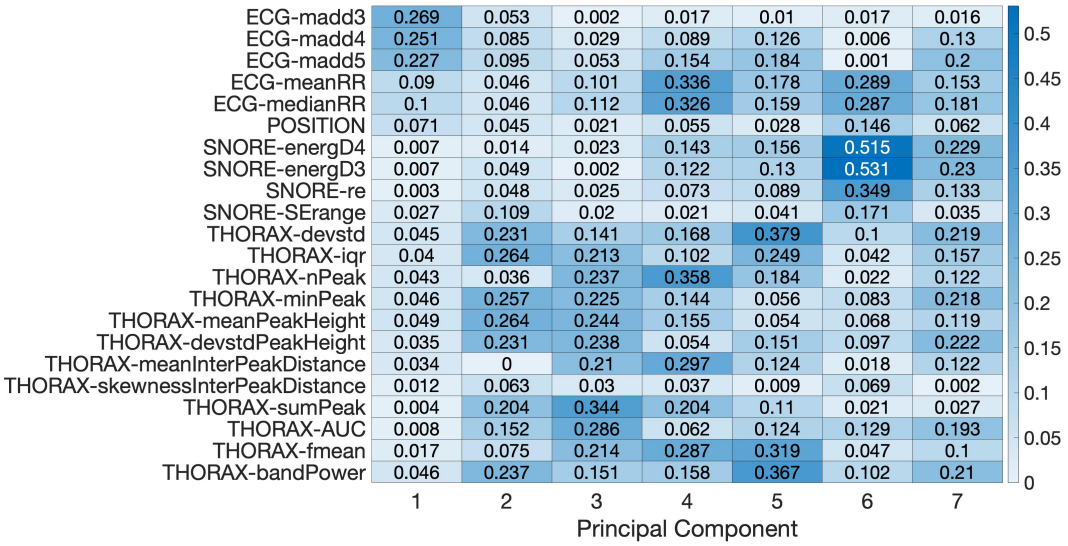
From this figure it is clear how features tend to be highly correlated to other features derived from the same signal. In particular, with respect to the considered classes, blood saturation and ECG features seems to be more correlated than thoracic volume features.

Moreover, it was decided to plot an heatmap, shown in figures 5.20 of the first 7 PCs in order to understand how much the selected features weighted in terms of explained variance. For visual purposes the heatmap is splitted in half and the values were rounded to the third decimal digit.

From figure 5.18 it is noticeable that the first 2 PCs explain about the 45% of total variance and from the heatmap it is clear that the first principal component (which explains about the 30% of total variance) is mainly influenced by ECG featuers while the second principal component (which explains about the 15% of total variance) is mainly influenced by blood saturation and thoracic volume features.



(a)



(b)

Figure 5.20: Heatmap of the first 2 principal components.

To visually notice what the heatmap is trying to express numerically, it is helpful to look at the biplot of figure 5.21.

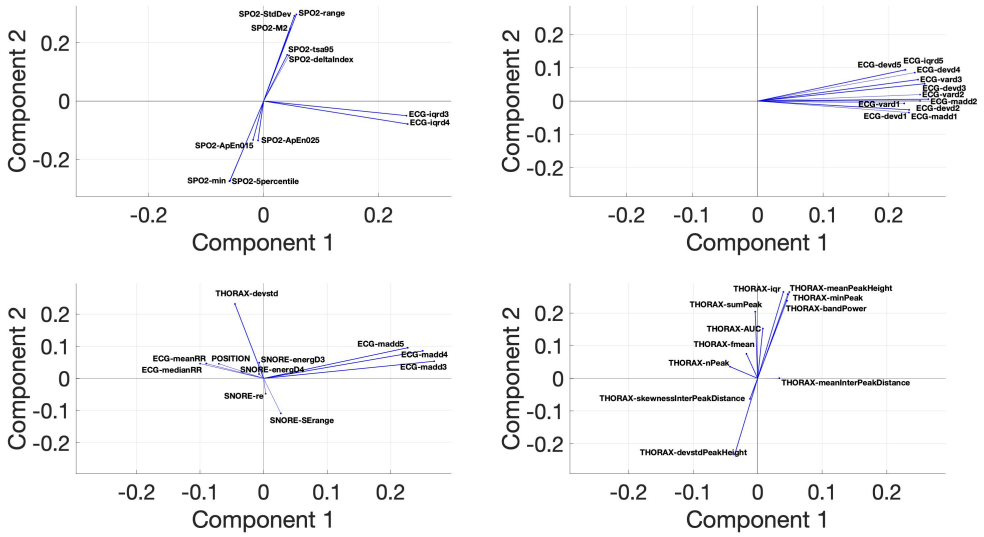


Figure 5.21: Biplot of the first 2 PCs for OSA/MSA vs CSA classification considering 44 features.

In this perspective it was decided to proceed with models training in order to get an idea of models' performances for this classification problem.

5.5.2 Results

In the results reported below the negative class is indicated with number 1 and is referred to the OSA/MSA cluster while the positive class is indicated with number 2 and is referred to the CSA cluster.
The classification errors of the trained models are reported in table 5.12.

Model	Resubstitution Error
Decision Tree	26.26%
Discriminant Analysis	28.00%
Naïve Bayes	34.03%
SVM	37.45%
kNN	**11.10%**

Table 5.12: Resubstitution errors of the trained models for OSA/MSA vs CSA classification considering 44 features.

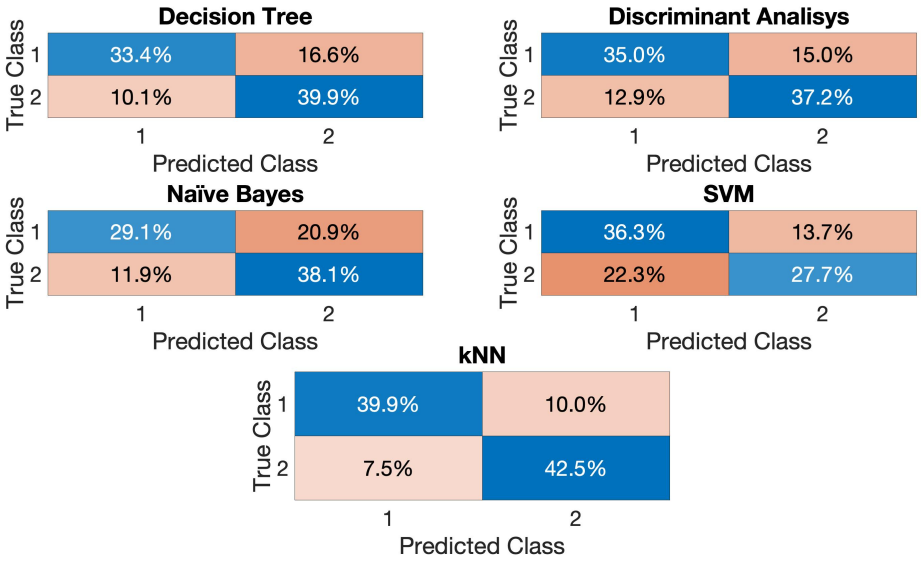
In table 5.13 are reported the prediction metrics and in figures 5.22a and 5.22b the relative confusion matrices and ROC curves.

Model	SENS	SPEC	PPV	NPV	ACC
Decision Tree	79.8%	66.9%	70.7%	76.8%	73.3%
Discriminant Analysis	74.3%	69.9%	71.2%	73.1%	72.2%
Naïve Bayes	76.2%	58.3%	64.6%	71.0%	67.2%
SVM	55.4%	72.6%	66.9%	62.0%	64.0%
kNN	**84.9%**	**79.9%**	**80.9%**	**84.1%**	**82.4%**

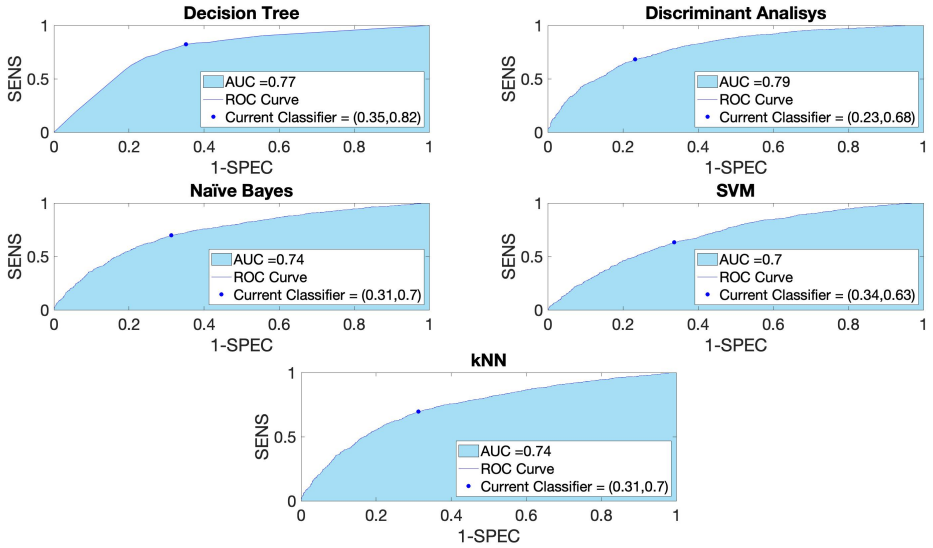
Table 5.13: OSA & MSA vs CSA prediction metrics considering 44 features.

From tables 5.12 and 5.13 it is clear that kNN model outperformed all the other models both in terms of resubstitution error and of prediction metrics with all values except specificity over 80%.

As observable from figure 5.22, despite having the best prediction metrics, kNN did not had the highest AUC with Discriminant analysis model having it.
Furthermore, as observable in figure 5.22a, kNN classifier, in addition to having best accuracy, returns a greater number of false positives (FP) rather than false negatives.



(a)



(b)

Figure 5.22: OSA/MSA vs CSA results comparison considering 44 features.
a) Confusion matrices comparison (normalized over the total number of observations, in order to see the percentages of FP and FN).
b) ROC curves comparison.

From these results it is clear that further investigation has to be done in order to reach acceptable prediction metrics values for this classification problem.
Nevertheless, it is important to keep in mind that all the features present in the initial dataset were calculated with the intention of discriminating an apneic condition from a non-apneic condition and not to discriminate between different forms of apnea.
An important aspect emerged from this analysis is that ECG features showed a strong discriminant power in apnea distinction but not in apnea detection.
Moreover, the fact kNN outperformed all the other models led to think that the data which represent the classes do not need to be mapped in another space being already informative in the current space. Another noticeable aspect is that the hyperparameters optimization process resulted in a small number of nearest neighbors (3) exhaustively searched (to find the nearest neighbors the distance values from all points were calculated) with a standardized euclidean distance metric (which is basically a classic euclidean distance scaled with standard deviation).

5.6 Comparison with related work

The first four analyses contained in the present work can be basically considered as screening approaches to recognize subjects' apneic condition. This can be useful to reduce sleep units saturation and to give an approximate idea of subject's apnea severity.
In order to get a fair comparison will be considered only the studies whose reference test is the same-night PSG, since for the present work was used the same reference.

For comparison reference, it was decided to look at Alvarez's review [Alvarez-Estevez and Moret-Bonillo, 2015], which summarized in table 4 different apneic event detection approaches.

It is also important to keep in mind that all the studies contained in table 4 of Alvarez's review were conducted on a much smaller number of subjects (18 out of 24 studies considered less than 50 subjects and the most extensive work comprised less than 250 subjects and no values of sensitivity and specificity are reported for the latter study) than the one considered by the present work (more than 1'550). Moreover, it has to be considered that:

- the present work considered also CPAP users;
- the present work was based on a larger subset of polysomnographic signals;
- The events definitions are not unique for all the studies.
- The detection unit (time considered for detection) was not unique.

Considering the studies conducted on the same detection unit (30 seconds), it is clear that the normal/hypopnea vs apnea analyses of the present work does not outperform the table's studies in terms of prediction metrics values while normal vs apnea analyses of the present work (which are more similar to the table's screening approaches since hypopnea epochs are not merged with normal epochs) have more balanced prediction metrics and show better prediction metrics values than table's works, in particular, they resulted on average or even better than studies conducted on different detection units.
Particular attention has to be paid in comparison to some multichannel studies with 30 seconds detection unit present in table 4.
The first is van Houdt's work, which is based on nasal airflow and thoracic and abdominal breathing volumes. The only prediction metrics reported are sensitivity and positive predictive value. In particular, sensitivity resulted better than normal/hypopnea vs apnea analyses but worse than normal vs apnea analyses and PPV resulted worse than any PPV obtained in the present work.

The second is Waxman's work, which is based on EEG, ECG, nasal pressure, oronasal temperature, right EOG, and EMG signals. The only prediction metrics reported are sensitivity and specificity which result slightly better than the present work normal/hypopnea vs apnea analyses but worse than the present work normal vs apnea analyses.
The third is Alvarez-Estevez's work, which is based on airflow, blood saturation, thoracic and abdominal breathing signals. The only prediction metrics reported are sensitivity and specificity which result slightly better than the present work normal/hypopnea vs apnea analyses but worse than the present work normal vs apnea analyses.
Another important comparison can be done with Taha's work, which was carried out on the WSC database using blood saturation and RIP signals. In particular, this work considered only 10 male subjects and obtained a sensitivity of 93.1% and a positive predictive value of 97% which are slightly higher than the ones of normal vs apnea analyses carried out in the present work.

Instead, for what concerns the apnea distinction analysis, it can be helpful to look at table 6 of Alvarez's review where validations of respiratory event classification approaches are reported. Since the present work's OSA/MSA vs CSA classification problem is not a full distinction will follow a qualitative argumentation. With particular attention to Taha's work (the same as the previous comparison), which was carried out on blood saturation and RIP signals of 10 male subjects picked from the WSC database, it is clear that the diagnostic accuracy is less than the diagnostic accuracy obtained in the present work and sensitivity and specificity are unbalanced for all the classes unlike what seen in the present work.
Moreover, all the studies considered few subjects (10 out of 12 on less than 30 subjects, and the most extensive embraced 66 subjects).

These comparisons help to gain an idea of how wide this problem is and how much work still needs to be done to better characterize such a problem.

Chapter 6

Conclusions

The present work has to be seen as a starting point for future projects since it evidenced that the features extracted from the chosen subset of polysomnographic signals are capable of characterizing subjects' condition during sleep with improvable prediction metrics' values.

There are some important conclusions:

- The selected features showed to carry redundant information and the simple feature selection performed did not manage to reduce enough this redundancy;
- Taking into account hypopnea causes performances to worsen because it is basically a transition condition between normal and apneic ones. Nevertheless, a condition of hypopnea not necessarily evolves into one of apnea: for instance there are some subjects within the WSC database that went through different hypopnea events during the night but did not experience a single apneic event. This is important to remember since some indexes like Apnea-Hypopnea Index (AHI) (number of events per hour of sleep) consider every breathing-related event (either hypopnea or apnea), so subjects with the same AHI do not necessarily face the same type and severity of events during the night;
- The more radical feature selection has slightly worsened the performance. This leads to thinking that a more precise feature selection should maintain better performances and may eliminate the need to pick up certain signals. For instance, the 32-features classification cases were characterized by the absence of EEG and ECG-derived features. This is significant mostly because recording EEG signals can be uncomfortable for the subject (above all for CPAP users) and this can simplify an eventual home recording system based on this subset of signals.

- The latest classification case (OSA/MSA vs CSA) brought to light that, in order to distinguish between apnea clusters, ECG-derived features have a more discriminative power than the other features. This is significant since brings to the conclusion that monitoring cardiac activity can help to discriminate between apnea forms but is not strictly necessary to understand if a subject is affected by SAHS since other signals' features resulted more informative.

6.1 Limits of present work

The non-optimal classification performances may be traced back to the big overlap of different clusters' feature distributions. There are multiple possible explanations for this phenomenon, such as:

- The chosen features resulted not enough representative about the sleep condition of the subjects since there was a lot of redundant information also after feature selection;
- The heterogeneity of subjects in the database. This represented a biasing factor for trained classifiers because subjects with different conditions (e.g. apnea severity, BMI...) as well as CPAP users were considered resulting in extensive statistical distributions of the features due to the manual events titration made by clinicians who consider different normal and apneic ranges from subject to subject;
- The provided database presented the following main limitations:
  - The sampling frequency used to record the signals resulted excessive for some slow time-varying signals such as blood saturation and body position and insufficient for others, for example, it did not allow to correctly analyze the snore signal which is an important non-invasive to record signal that is usually sampled at much higher frequencies (see subsection 3.7.1) making possible to extract more interesting features;
  - The hardware pre-filtering eliminated frequencies above 30 Hz cutting off a great part of spectral information from different signals such as ECG (whose band goes up to 125 Hz);
  - It was observed that some subjects have changes in the gain of some recording channels during the registration that makes it difficult to extract useful features since these changes could make features appear as outliers within the dataset;
  - Despite the scoring procedures for breathing-related events being standardized, it cannot be excluded residual inter- and intra-operator variability which affects the data.

Moreover, there are two other big limitations that affect this kind of study. The first one consists in considering 30 seconds epochs and calculating parameters that try to explain the subject's condition during the duration of the epoch. This represents a problem since an event may not start when an epoch starts but that epoch, according to the standardized scoring procedure will be considered pathological even if the even occupies only a small portion. It could interesting and maybe more efficient to investigate real-time sleep analysis or sleep analysis based on smaller detection units rather than analysis based on detection units. Moreover, another big limitation is represented by computational workload since the step of signal processing features extraction and training and evaluation of models' performances required a very long time. Just think that the elapsed time for one cycle of features extraction for all the signals was above 200 hours (being the EEG processing the longest requiring about 4 days to complete) and the elapsed time for one run of models training and testing was above 100 hours.
These numbers give an idea of how fundamental is to have high-performance computers to work on.

6.2 Future developments

The present work lays the foundations for future studies on sleep conditions classification and SAHS screening approaches.

Further studies could follow the same path and refine the feature selection and hyperparameter optimization processes in order to investigate more deeply the potentialities of this dataset.
Furthermore, other studies could focus on extracting more informative features for specific classification problems keeping only the most discriminant ones identified in the present work before proceeding with model training.
Because of the large number of observations in the dataset, it is plausible to think about using some kind of Artificial Neural Network (ANN) in future works.

In order to increase diagnostic accuracy, it could be useful to train different models using a group of signals recorded from subjects with similar apnea severity, and similar central tendency measures of the main general statistic features rather than considering all the subjects together. This way it should be possible to narrow statistical distributions making more accurate predictions since normal and apneic characteristic features ranges should be better distinguishable.
For instance, classifiers with different blood saturation baselines could be trained and subjects could be asked for measuring blood saturation recording for several

nights before the polysomnography (or before undergo home-PSG) in order to get information about the general condition during sleep and assign them to the most appropriate classifier (using a fuzzy inference system for example).

Chapter 7

Acknowledgements

This Wisconsin Sleep Cohort Study was supported by the U.S. National Institutes of Health, National Heart, Lung, and Blood Institute (R01HL62252), National Institute on Aging (R01AG036838, R01AG058680), and the National Center for Research Resources (1UL1RR025011). The National Sleep Research Resource was supported by the U.S. National Institutes of Health, National Heart Lung and Blood Institute (R24 HL114473, 75N92019R002).

It is my pleasure to write the next thanks in my native language.

Ringrazio il Professore Filippo Molinari, perché è grazie al suo corso di Elaborazione dei Segnali Biomedici che ho acquisito interesse, poi evoluto in passione, verso gli argomenti trattati in questo progetto di tesi.

Ringrazio il Professore Alessandro Puiatti, relatore esterno, per avermi offerto la possibilità di svolgere questo progetto in collaborazione con lui e per tutte le ore trascorse insieme durante le quali mi ha insegnato tanto.

Ringrazio mia madre e mio padre: una per avermi insegnato ad apprezzare i cambiamenti e ad impegnarmi sempre in ogni cosa, l'altro per avermi indirizzato verso una forma mentis che mi ha permesso di approcciarmi correttamente ai problemi che finora ho affrontato, entrambi per non avermi mai fatto mancare nulla e per avermi permesso di studiare e di esprimermi attraverso il mio lavoro.

Ringrazio Federica per avermi sempre supportato in ogni decisione, per avermi insegnato ad apprezzare i momenti fugaci e per stare al mio fianco da anni. Ovunque tu sia, qualunque cosa tu scelga.

Ringrazio i miei parenti più stretti che hanno dimostrato il loro affetto non dimenticandosi di me nonostante la distanza.

Ringrazio gli amici di una vita: Biagio, Dario, Peppe, Salvo e Sergio, per tutte le avventure insieme e per avermi fatto trascorrere momenti che non dimenticherò mai. Siete i fratelli che non ho mai avuto.

Ringrazio gli amici che ho più recentemente scoperto di avere per il continuo confronto, supporto ed incoraggiamento. Siete stati preziosi in questa parte finale della mia avventura.

Ringrazio chi non c'è più per i ricordi che mi avete lasciato e perché ho fatto tesoro dei vostri insegnamenti. Mi mancate, non sarete mai dimenticati.

Infine ringrazio me stesso per non essermi mai arreso davanti alle difficoltà, per essere uscito più volte dalla mia comfort zone e per essermi sempre messo in gioco. Nella consapevolezza di perseguire un continuo miglioramento riporto il motto che mi ispira ogni giorno:

"Non so se vinceremo, ma so che non ci daremo mai per vinti."

Appendix A

Definitions and scoring of breathing events

In this appendix will be deepened the definitions and the scoring precedures of apneas and hypopnea reported in the WSC manual of operations [Wisconsin Sleep Cohort, 2009] followed by clinicians to score the breathing events using the database signals.

A.1 Apneas

Definitions

Apneas are characterized by no indication of airflow in nasal pressure, no detectable breathing pattern in the thermistor and a clear amplitude reduction in effort, followed by an associated desaturation.
The different types of apnea have been distinguished observing the thermocouple and the respitrace signals:

OSA *No indication of airflow by thermocouple and an indication of effort in respitrace channels.*

CSA *No indication of airflow by thermocouple and no indication of effort in respitrace channels.*

MSA *No indication of airflow by thermocouple and areas of no effort followed by effort in respitrace channels.*

Scoring Procedure

1. Determine if there is flow or no flow. Criteria for NO flow:
   - Does not follow previous pattern of flow **and/or**
   - is <20% of amplitude of the largest previous breath (determined by μV/mm of unclipped air flow sensitivity, if necessary) **and**
   - has an interruption of airflow that is > 10 seconds in duration.
2. Determine if there is effort or no effort. Criteria for No effort (from Respitrace):
   - Does not follow previous pattern of breathing **and**
   - has no discernable amplitude of the signal in the respitrace.
3. Measure duration of event:
   - Measure from the beginning of the last expiration on the air flow channels to the beginning of the next inspiration on the air flow channels to determine the 10 second criterion;
   - If 10 seconds, measure the duration of the event from the beginning of the last expiration to the beginning of the next inspiration on the SUM channel (sum of volumes) of the respitrace that best corresponds to the points of measurement of duration in the airflow channels;
   - If not 10 seconds, determine if the event meets the criterion for a hypopnea : 4% desaturation. If it does not, then the event is ignored and not scored;
   - NOTE: When the event is obviously an apnea and is between 9.5 and 10 seconds, round the duration up to 10 seconds and score.

A.2 Hypopnea

Definition

A discernable decrease in flow in nasal pressure channel and/or thermistor with an associated oxygen desaturation of 4% or greater indicated in the SpO_2 channel beginning in sleep.

Scoring Procedure

1. Use a display view of at least a 120 second window;
2. Determine a discernable decrease in the SUM channel (sum of volumes) defined as a >50% decrease in the mean amplitude of the three largest breaths preceding the onset of the event, or a clear reduction in amplitude that is <50% with an associated oxygen desaturation of > 4%;
3. Measure the duration of the event:
   - Measure from the beginning of the last expiration on the SUM to the beginning of the next inspiration on the SUM to determine the 10-second criterion (from the beginning to the end of the event);
   - If not 10 seconds, delete the event mark;
   - Mark the desaturation event on the SpO_2 channel following the respiratory event, beginning within 30 seconds of the end of the respiratory event;
   - Delete the desaturation event for a hypopnea if the desaturation is <4%;
   - Determine that the desaturation occurs in sleep. *Events that begin in sleep and end in wake are always scored. Events that begin and end in wake are never scored*;
   - Mark the beginning and end of the event in the SpO_2 channel corresponding to the desaturation. Duration of desaturation events should not be greater than 120 sec.
4. Mark the corresponding event in the SUM channel as Hypopnea if associated with a desaturation of > 4%;
5. Without the presence of an adequate signal in the nasal pressure channel, use the nasal/oral thermistor channel for determination of flow.

Appendix B

Wavelet Transform (WT)

The Wavelet Transform is a linear Time-Frequency Representation (TFR): *time-frequency representation* because it aims to describe how the spectral content of a signal is changing in time (non Wide Sense Stationary (WSS) signals) and *"linear"* because it satisfies the superposition principle: *"if x(t) is a linear combination of some signal components, the its TFR is the linear combination of the TFRs of all components"*.[Mesin, 2017]

The wavelet analysis is used because it overpass the limitation of another linear TFR, the Short Time Fourier Transform (STFT) which was introduced in order to overcome the difficulties that Fourier analysis finds in representing non-stationary signals with the basic idea of analyzing successive portions of signal, approximating them to stationary [Masotti, 2001].
While for the latter time resolution and frequency resolution are constants and depend on the length of the analysis window, instead for the WT frequency resolution is allowed to vary in the time-frequency plane. In general, there is a linear relation between the length of the analysis window and the frequency resolution as there is a linear relation between time resolution and the central frequency of the considered portion of the spectrum. Therefore, considering the windowing function as a band-pass filter, by scaling it with frequency, band-pass filters with constant relative bandwidth are used.

B.1 Multiresolution analysis (MRA)

Another approach for describing wavelet analysis is based on the choice of a prototype function, called *wavelet*, which can represent the signal in a compact way if it is chosen in order to approximate well the different components of the signal.

The independent variables of the WT (time and scale) can be either continuous or discrete [Mesin, 2017].

In other words, Fourier Transform decomposes signals into a sum of sine and cosine (with infinite support) while Wavelet Transform decomposes signals considering two functions with compact supports, together with their scaled versions, called *Mother Wavelet* and *Father Wavelet*, which represent different frequency components of the original signal and can be scaled. *Father wavelet will* be representative of the smooth and LOW-frequency part of the signal giving an *approximate* representation of the signal, while *Mother wavelet* will be representative of the detail and HIGH-frequency part of the signal.

Therefore, WT can be used to decompose signals into low-frequency components and high-frequency components using low-pass filter (LPF) and high-pass filter (HPF), respectively: the output of the HPF are the detail coefficients while the output of the LPF are approximation coefficients as shown in figure B.1.

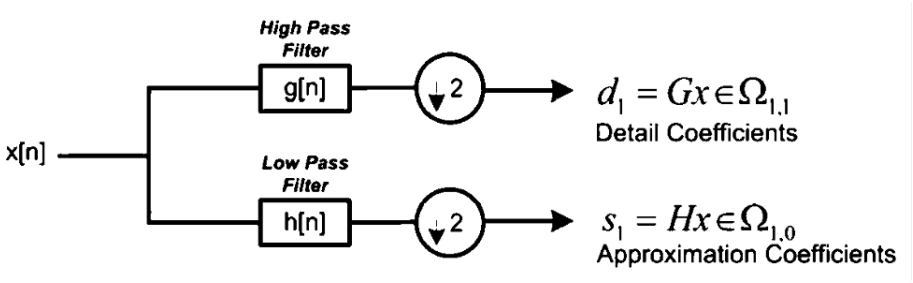


Figure B.1: First step of wavelet decomposition.

This kind of decomposition can be perpetuated in order to obtain more families from different Father wavelet and Mother wavelet as con be observed in figure B.2.

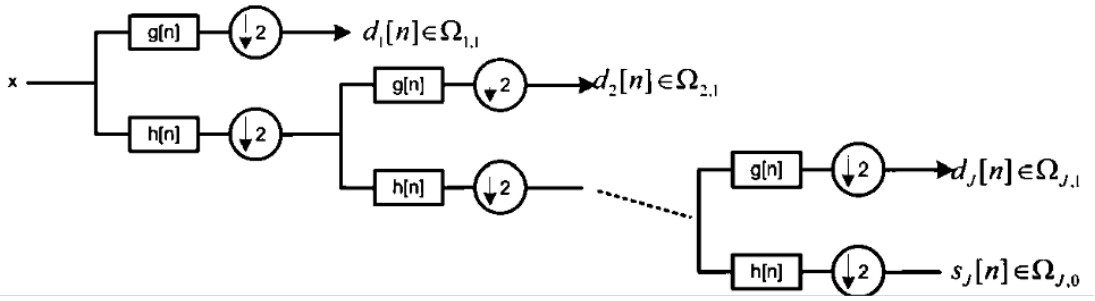


Figure B.2: Cascade of filter banks.

WT divides a signal in two parts and then, recursively, divides the low-pass part into another two parts. In other words, it keeps decomposing only the left side: if the band of the smallest low-pass part is B, the band of the father of this part will be 2B and the band of the father's father will be 4B, as shown in figure B.3.
Each time the signal is divided there is a shrinkage in time: the LP takes about a half of the samples of the original signal while the HP takes the other half. In this way, using the WT you keep decomposing in frequency and at the same time you are shrinking the size in the time domain as well.

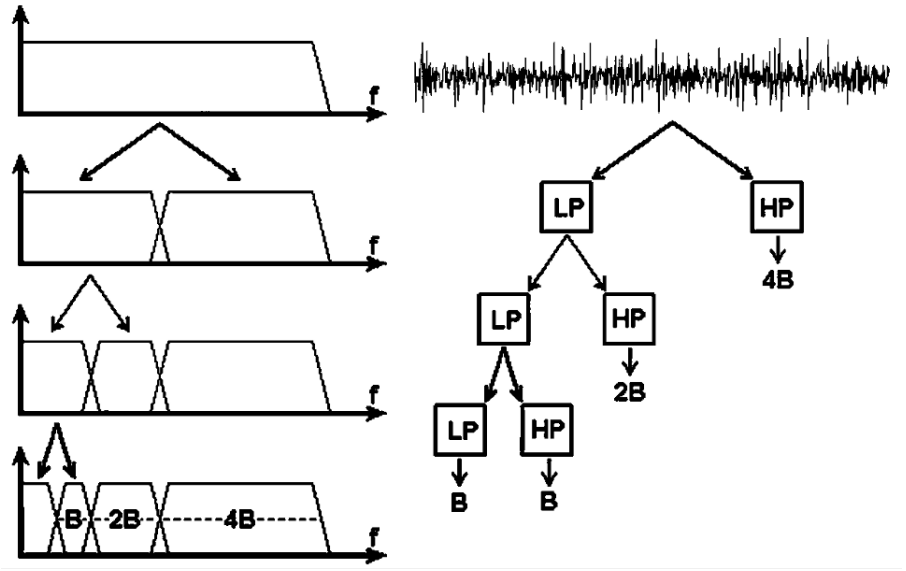


Figure B.3: The left side of the power spectrum continues to be divided in half as the number of decomposition levels increase.

The previous treatment described the concept of Multiresolution Analysis (MRA) based on DWT, which is by far the most used since in the real life problems the signals to be analyzed are finite.

The main problem of the MRA is that the choice of the wavelet type is arbitrary. There are several types of wavelets families (we talk about families because we refer to all the mother wavelets and their scaled and shifted versions) as shown in figure B.4.

The empiric rule that guides the choice is that the wavelet must approximate well the different components of the signal. For instance, if the signal to be analyzed is a random process it is reasonable to choose a Daubechies (Db) or a Symlet (variation of the Daubechies) because the dilated and translated version of this wavelets

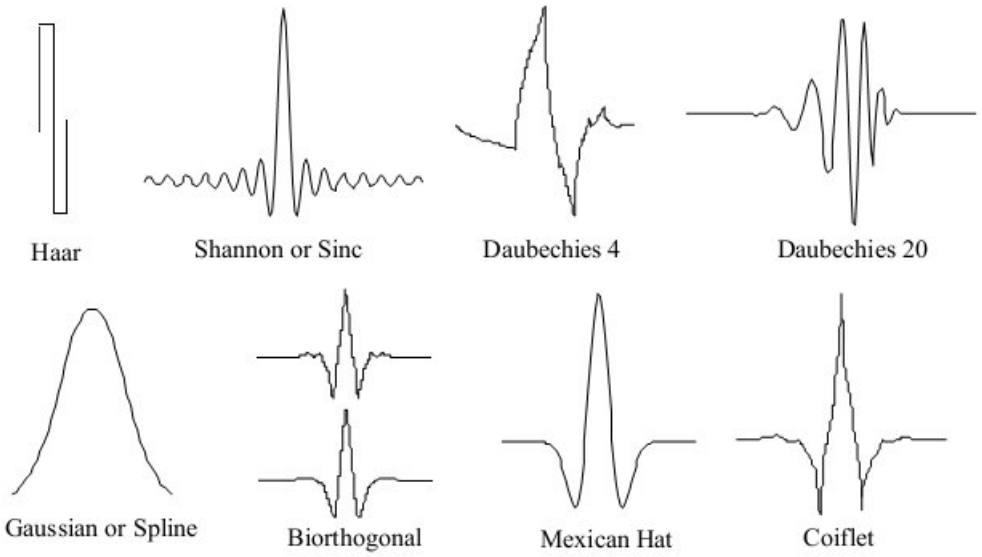


Figure B.4: Examples of types of wavelets [Al-Geelani et al., 2016].

can "match" the signal to the hidden event and thus discover its frequency and location in time. Another example could be the use of the Haar wavelet to match an abrupt discontinuity or the use of the Db20 to match a chirp signal [Al-Geelani et al., 2016].

Moreover, the MRA represent an high computational burden above all if there are many signals to analyze and if these signals are long.

Bibliography

Akhter, S., Abeyratne, U. R., Swarnkar, V., & Hukins, C. (2018). Snore sound analysis can detect the presence of obstructive sleep apnea specific to nrem or rem sleep. *Journal of Clinical Sleep Medicine*, *14*(6), 991–1003.

Al-Geelani, N. A., Piah, M. A. M., Saeh, I., Othman, N. A., Muhamedin, F. L., & Kasri, N. (2016). Identification of acoustic signals of internal electric discharges on glass insulator under variable applied voltage. *International Journal of Electrical and Computer Engineering*, *6*(2), 827.

Almazaydeh, L., Elleithy, K., & Faezipour, M. (2012). Detection of obstructive sleep apnea through ecg signal features. *2012 IEEE International Conference on Electro/Information Technology*, 1–6.

Alvarez, D., Hornero, R., Marcos, J. V., Del Campo, F., & Lopez, M. (2009). Spectral analysis of electroencephalogram and oximetric signals in obstructive sleep apnea diagnosis. *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 400–403.

Álvarez, D., Hornero, R., Garcıa, M., del Campo, F., & Zamarrón, C. (2007). Improving diagnostic ability of blood oxygen saturation from overnight pulse oximetry in obstructive sleep apnea detection by means of central tendency measure. *Artificial intelligence in medicine*, *41*(1), 13–24.

Alvarez-Estevez, D., & Moret-Bonillo, V. (2015). Computer-assisted diagnosis of the sleep apnea-hypopnea syndrome: A review. *Sleep disorders*, *2015*.

Arzt, M., Floras, J. S., Logan, A. G., Kimoff, R. J., Series, F., Morrison, D., Ferguson, K., Belenkie, I., Pfeifer, M., Fleetham, J., et al. (2007). 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*, *115*(25), 3173–3180.

Barthlen, G. M., Brown, L. K., Wiland, M. R., Sadeh, J. S., Patwari, J., & Zimmerman, M. (2000). Comparison of three oral appliances for treatment of severe obstructive sleep apnea syndrome. *Sleep medicine*, *1*(4), 299–305.

Becker, H. F., Jerrentrup, A., Ploch, T., Grote, L., Penzel, T., Sullivan, C. E., & Peter, J. H. (2003). Effect of nasal continuous positive airway pressure treatment on blood pressure in patients with obstructive sleep apnea. *circulation*, *107*(1), 68–73.

Bender, S. D. (2012). Oral appliance therapy for sleep-related breathing disorders. *Operative Techniques in Otolaryngology-Head and Neck Surgery*, *23*(1), 72–78.

Berry, R. B., Budhiraja, R., Gottlieb, D. J., Gozal, D., Iber, C., Kapur, V. K., Marcus, C. L., Mehra, R., Parthasarathy, S., Quan, S. F., et al. (2012). 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. *Journal of clinical sleep medicine*, *8*(5), 597–619.

Berry, R. B., Kryger, M. H., & Massie, C. A. (2011). A novel nasal expiratory positive airway pressure (epap) device for the treatment of obstructive sleep apnea: A randomized controlled trial. *Sleep*, *34*(4), 479–485.

Berry, R. B., & Sriram, P. (2014). Auto-adjusting positive airway pressure treatment for sleep apnea diagnosed by home sleep testing. *Journal of Clinical Sleep Medicine*, *10*(12), 1269–1275.

Cantineau, J. P., Escourrou, P., Sartene, R., Gaultier, C., & Goldman, M. (1992). Accuracy of respiratory inductive plethysmography during wakefulness and sleep in patients with obstructive sleep apnea. *Chest*, *102*(4), 1145–1151.

Chambrin, M.-C. (2001). Alarms in the intensive care unit: How can the number of false alarms be reduced? *Critical Care*, *5*(4), 1–5.

Deane, S. A., Cistulli, P. A., Ng, A. T., Zeng, B., Petocz, P., & Darendeliler, M. A. (2009). Comparison of mandibular advancement splint and tongue stabilizing device in obstructive sleep apnea: A randomized controlled trial. *Sleep*, *32*(5), 648–653.

de Chazal, P., Penzel, T., & Heneghan, C. (2004). Automated detection of obstructive sleep apnoea at different time scales using the electrocardiogram. *Physiological measurement*, *25*(4), 967.

Deviaene, M., Testelmans, D., Buyse, B., Borzée, P., Van Huffel, S., & Varon, C. (2018). Automatic screening of sleep apnea patients based on the spo 2 signal. *IEEE journal of biomedical and health informatics*, *23*(2), 607–617.

Ferber, R., Millman, R., Coppola, M., Fleetham, J., Friederich Murray, C., Iber, C., McCall, W. V., Nino-Murcia, G., Pressman, M., Sanders, M., et al. (1994). Portable recording in the assessment of obstructive sleep apnea. *Sleep*.

Ghareeb, S., Hussain, A. J., Al-Jumeily, D., Khan, W., Al-Jumeily, R., Baker, T., Al Shammaa, A., & Khalaf, M. (2022). Evaluating student levelling based on machine learning model's performance. *Discover Internet of Things*, *2*(1), 1–25.

Gonzalez, H., Haller, B., Watson, H. L., & Sackner, M. A. (1984). Accuracy of respiratory inductive plethysmograph over wide range of rib cage and abdominal compartmental contributions to tidal volume in normal subjects and in patients with chronic obstructive pulmonary disease. *American Review of Respiratory Disease*, *130*(2), 171–174.

Grimm, W., & Koehler, U. (2014). Cardiac arrhythmias and sleep-disordered breathing in patients with heart failure. *International journal of molecular sciences*, *15*(10), 18693–18705.

Hoffstein, V. (2007). Review of oral appliances for treatment of sleep-disordered breathing. *Sleep and Breathing*, *11*(1), 1–22.

Hossain, J. L., & Shapiro, C. M. (2002). The prevalence, cost implications, and management of sleep disorders: An overview. *Sleep and Breathing*, *6*(02), 085–102.

Huyett, P., & Bhattacharyya, N. (2021). Incremental health care utilization and expenditures for sleep disorders in the united states. *Journal of Clinical Sleep Medicine*, *17*(10), 1981–1986.

Isa, S., Fanany, M., Jatmiko, W., & Murini, A. (2010). Feature and model selection on automatic sleep apnea detection using ecg. *International Conference on ComputerScience and Information Systems*, 357–362.

Kogan, D., Jain, A., Kimbro, S., Gutierrez, G., & Jain, V. (2016). Respiratory inductance plethysmography improved diagnostic sensitivity and specificity of obstructive sleep apnea. *Respiratory Care*, *61*(8), 1033–1037.

Levin, K. H., & Chauvel, P. (2019). *Clinical neurophysiology: Basis and technical aspects: Handbook of clinical neurology series*. Elsevier.

Levy, J., Álvarez, D., Rosenberg, A. A., Alexandrovich, A., del Campo, F., & Behar, J. A. (2021). Digital oximetry biomarkers for assessing respiratory

function: Standards of measurement, physiological interpretation, and clinical use. *npj Digital Medicine*, *4*. https://doi.org/10.1038/s41746-020-00373-5

Masotti, M. (2001). *Analisi wavelet multirisoluzione: Dalla teoria matematica ad una possibile applicazione nell'ambito dell'esperimento alice* (Doctoral dissertation).

Mesin, L. (2017). *Introduction to biomedical signal processing*.

Molinari, F. (2021). *Corso di elaborazione dei segnali biomedici*. Politecnico di Torino.

Morgenthaler, T. I., Kagramanov, V., Hanak, V., & Decker, P. A. (2006). Complex sleep apnea syndrome: Is it a unique clinical syndrome? *Sleep*, *29*(9), 1203–1209.

Oksenberg, A., Khamaysi, I., Silverberg, D. S., & Tarasiuk, A. (2000). Association of body position with severity of apneic events in patients with severe nonpositional obstructive sleep apnea. *Chest*, *118*(4), 1018–1024.

Penzel, T., & Conradt, R. (2000). Computer based sleep recording and analysis. *Sleep medicine reviews*, *4*(2), 131–148.

Pépin, J. L., Lévy, P., Lepaulle, B., Brambilla, C., & Guilleminault, C. (1991). Does oximetry contribute to the detection of apneic events?: Mathematical processing of the sao2 signal. *Chest*, *99*(5), 1151–1157.

Petterson, M. T., Begnoche, V. L., & Graybeal, J. M. (2007). The effect of motion on pulse oximetry and its clinical significance. *Anesthesia & Analgesia*, *105*(6), S78–S84.

Pincus, S. M. (2001). Assessing serial irregularity and its implications for health. *Annals of the New York Academy of Sciences*, *954*(1), 245–267.

Rachim, V. P., Li, G., & Chung, W.-Y. (2014). Sleep apnea classification using ecg-signal wavelet-pca features. *Bio-medical materials and engineering*, *24*(6), 2875–2882.

Rechtschaffen, A. (1968). A manual for standardized terminology, techniques and scoring system for sleep stages in human subjects. *Brain information service*.

Roebuck, A., Monasterio, V., Gederi, E., Osipov, M., Behar, J., Malhotra, A., Penzel, T., & Clifford, G. (2013). A review of signals used in sleep analysis. *Physiological measurement*, *35*(1), R1.

Saad, N. M., Abdullah, A. R., & Low, Y. F. (2006). Detection of heart blocks in ecg signals by spectrum and time-frequency analysis. *2006 4th Student Conference on Research and Development*, 61–65.

Schmidt-Nowara, W., Lowe, A., Wiegand, L., Cartwright, R., Perez-Guerra, F., & Menn, S. (1995). Oral appliances for the treatment of snoring and obstructive sleep apnea: A review. *Sleep*, *18*(6), 501–510.

Schwartz, A. R., Bennett, M. L., Smith, P. L., De Backer, W., Hedner, J., Boudewyns, A., Van de Heyning, P., Ejnell, H., Hochban, W., Knaack, L., et al. (2001). Therapeutic electrical stimulation of the hypoglossal nerve in obstructive sleep apnea. *Archives of Otolaryngology–Head & Neck Surgery*, *127*(10), 1216–1223.

Shokoueinejad, M., Fernandez, C., Carroll, E., Wang, F., Levin, J., Rusk, S., Glattard, N., Mulchrone, A., Zhang, X., Xie, A., Teodorescu, M., Dempsey, J., & Webster, J. (2017). Sleep apnea: A review of diagnostic sensors, algorithms, and therapies. *Physiological Measurement*, *38*, R204–R252. https://doi.org/10.1088/1361-6579/aa6ec6

Sleep Foundation. (2022). *Sleep respiratory rate*. https://www.sleepfoundation.org/sleep-apnea/sleep-respiratory-rate (Retrieved June 27 2022)

Streatfeild, J., Smith, J., Mansfield, D., Pezzullo, L., & Hillman, D. (2021). The social and economic cost of sleep disorders. *Sleep*, *44*. https://doi.org/10.1093/sleep/zsab132

The MathWorks Inc. (2021). *User's guide for matlab r2021b*. https://it.mathworks.com/help/signal/ref/findpeaks.html (Retrieved November 17 2021)

The MathWorks Inc. (2022). *User's guide for matlab r2021b*. https://it.mathworks.com/help/stats/supervised-learning-machine-learning-workflow-and-algorithms.html#bswluhd (Retrieved June 27 2022)

Van Steenkiste, T., Groenendaal, W., Ruyssinck, J., Dreesen, P., Klerkx, S., Smeets, C., de Francisco, R., Deschrijver, D., & Dhaene, T. (2018). Systematic comparison of respiratory signals for the automated detection of sleep apnea. *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 449–452.

Wang, C., Peng, J., Song, L., & Zhang, X. (2017). Automatic snoring sounds detection from sleep sounds via multi-features analysis. *Australasian physical & engineering sciences in medicine*, *40*(1), 127–135.

Watson, H. L., Poole, D. A., & Sackner, M. A. (1988). Accuracy of respiratory inductive plethysmographic cross-sectional areas. *Journal of Applied Physiology*, *65*(1), 306–308.

Weaver, T. E., & Grunstein, R. R. (2008). Adherence to continuous positive airway pressure therapy: The challenge to effective treatment. *Proceedings of the American Thoracic Society*, *5*(2), 173–178.

Welch, P. (1967). The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics*, *15*(2), 70–73.

Wisconsin Sleep Cohort. (2009). *WSC Manual of Operations*.

Xie, A., Skatrud, J. B., & Dempsey, J. A. (2001). Effect of hypoxia on the hypopnoeic and apnoeic threshold for co2 in sleeping humans. *The Journal of Physiology*, *535*(1), 269–278.

Xie, B., & Minn, H. (2012). Real-time sleep apnea detection by classifier combination. *IEEE Transactions on information technology in biomedicine*, *16*(3), 469–477.

Yılmaz, B., Asyalı, M. H., Arıkan, E., Yetkin, S., & Özgen, F. (2010). Sleep stage and obstructive apneaic epoch classification using single-lead ecg. *Biomedical engineering online*, *9*(1), 1–14.

Young, T., Palta, M., Dempsey, J., Peppard, P. E., Nieto, F. J., & Hla, K. M. (2009). Burden of sleep apnea: Rationale, design, and major findings of the wisconsin sleep cohort study. *WMJ: official publication of the State Medical Society of Wisconsin*, *108*(5), 246.

Zhang, G.-Q., Cui, L., Mueller, R., Tao, S., Kim, M., Rueschman, M., Mariani, S., Mobley, D., & Redline, S. (2018). The national sleep research resource: Towards a sleep data commons. *Journal of the American Medical Informatics Association*, *25*(10), 1351–1358.

Zhang, Z., Zheng, J., Wu, H., Wang, W., Wang, B., & Liu, H. (2012). Development of a respiratory inductive plethysmography module supporting multiple sensors for wearable systems. *Sensors*, *12*(10), 13167–13184.