University of Crete
Department of Medicine
MSc in Biomedical Engineering

Thesis

Automatic Obstructive Sleep Apnea Detection

Bantra Konstantina

Supervisor: Professor Sakkalis Vangelis

16 Ιανουαρίου 2023

Abstract

Over the years, ballistocardiography (BCG) has emerged as a unobstructive, non-invasive, and safe technique that can be used for calculating the heart rate of a patient. As the heart rate is an essential element not only for the differential diagnosis process but also for monitoring patients, various scholars have conducted research on possible applications and use for the BCG signal. Among these, sleep apnea has gained a lot of interest as it benefits greatly from the unobstructive nature of the BCG recording. Compared to Polysomnography (PSG), the golden standard used today, BCG bears the advantage of recording the heart rate and the respiration rate without physically touching the patient, thus removing the discomfort of the multiple attached wires and sensors which are needed during the PSG. This is crucial, as for an accurate diagnosis, multiple hours of recordings have to be acquired and most of the time the procedure has to be repeated for more than one night. By removing the discomfort for the patient we could obtain more accurate results and people would be less hesitant to undergo the diagnosis procedure. Furthermore, as the BCG sensor is a very small and portable device, at home monitoring could also be a possibility. With the present thesis we aim to assess the possibility of monitoring the heart rate and respiration rate of a patent using solely a BCG recording. Additionally, the possibility of developing a system which can utilize the calculated physiological signals and automatically detect Obstructive Sleep Apnea Events during patient's sleep, using machine learning and Artificial Intelligence techniques is also explored hoping that the system will, in the long run, help in reducing diagnostic ambiguity and also in creating more concise and realizable results.

Contents

List of Figures

Chapter 1

Introduction

1.1 Motivation and Objectives

Over the past few decades research has recognized Obstructive Sleep Apnea (OSA) as a major risk factor for numerous health problems, including hypertension, stroke, cardiomyopathy, heart failure, diabetes, and heart attacks [5]. Unfortunately, to this day, OSA and sleep disorders are generally discarded and not given proper attention. In a survey conducted in the USA in 2002, it was estimated that even though OSA affects approximately 20% of the population, 90% of them are undiagnosed [6]. However, as in the case of any disease, early diagnosis and treatment can play a critical role in the success of a treatment, faster recovery, and offer more chances of longer survival. This highlights the need for simpler, specialized diagnostic techniques that can aid the doctors during a diagnostic procedure by providing more data and thus more insight into a patient's health condition.
One of the most promising techniques for diagnosing OSA is ballistocardiography. The technique is aptly named; it is comprised of the word βαλλισμός (ballism) which is used to describe the spasmodic movement, καρδιά (heart), and γράφος (graph). It is a technique for producing a graphical representation of repetitive motions of the human body arising from the sudden ejection of blood into the great vessels with each heartbeat. Some of the main advantages of this method are that the recording device is unobstructive, simple to use, and of low cost. Most importantly though, it has the potential to offer highly

accurate results for OSA detection as the signal holds information for both the respiration, as well as the heat rate of the patient.
Heretofore, the golden standard for diagnosing sleep apnea is PSG, which is predominantly based on the visual observation of the recorded signals by well-trained doctors. This means that the patient will have to stay overnight at the hospital with a couple of sensors and electrodes attached as well as a nose tube. It is often that the discomfort of the process deters patients from seeking diagnosis and of course, lurks the possibility of interfering with the results. To overcome this problem, research has focused on alternative recording techniques which can produce reliable results. Unfortunately, to this day there is no established standardization in the recording device nor the measured parameters in the BCG signal or other signals, and for that reason, it is a very active research field. The rapid advancement of technology in recent years has offered more interesting signal processing techniques and maybe in the future home diagnosis will be possible.
The topic of this thesis is the digital processing of the BCG signal to accurately compute the heart rate and the respiration rate of the patient. Additionally, the possibility of identifying sleep apnea events using machine learning and AI techniques is explored to assess BCG as a potential solution for monitoring a patient's sleep that in the future, could be an alternative to PSG.

1.2 The Golden Standard

Polysomnography, also known as a sleep study, is a diagnostic procedure used for diagnosing and monitoring sleeping disorders. During the procedure, the brainwaves, the oxygen levels in a patient's blood, heart rate, breathing as well as eye and leg movements are recorded by multiple sensors attached to the patient. PSG is conducted in hospitals or specialized sleep clinics. During PSG a specialized person monitors:

- Brain waves (EEG)
- Eye movements (EOG)
- Chin muscle activity (EMG)

- Heart rate (ECG)
- Breathing pattern
- Blood oxygen level
- Body position
- Chest and abdominal movement
- Limb movement (EMG)
- Snoring and other noise that a patient may make during sleep

To record this data, the technician places a series of electrodes on the patient's scalp, temples, chest, and legs. Additionally, an elastic belt is placed around the patient's chest and stomach area, and a small oximeter is clipped on the patient's finger. Finally, the recordings of the multiple physiological signals are evaluated by a doctor or a sleep technician who annotates them as different sleep stages and sleep disorder events. To identify sleep apnea, the sleep specialist will check the frequency of the apnea episodes, meaning the frequency of episodes where the patient's breath stopped for more than 10 seconds, or the frequency of hypopnea episodes, meaning the frequency of episodes where the patient's breathing was partially blocked. The frequency of these events is then used to calculate the apnea-hypopnea index (AHI) which is used to classify the severity of the sleeping disorder. Particularly,

- AHI<5, normal breathing
- 5<AHI<15, mild sleep apnea
- 15<AHI<30, moderate sleep apnea
- 30<AHI, severe sleep apnea

Some of the advantages that PSG has are the following,

- A technician is always present and can adjust sensors for optimal recording
- The whole sleep is recorded
- Other conditions may be observed

However, the PSG has also many disadvantages. Some of them are:

- The recordings require multiple wires,belts, electrodes and sensors attached to the patient's body which may cause discomfort
- The recording is performed in an unfamiliar environment
- It is an expensive procedure
- The waiting list for an appointment can be very long, especially during pandemic periods

It is because of these disadvantages that people have looked into alternative diagnostic procedures for OSA. A huge effort has focused on identifying the smallest subgroup of recorded signals than can be used to accurately identify sleep apnea. Limited Channel Monitoring is one of the suggested approaches and focuses only on breathing and blood oxygen levels. This test also has the advantage that it can be performed at home, but it has a greater failure rate.
The BCG was first suggested by Isaac Starr [7] in 1939. When it was first proposed, it attracted a lot of interest from the scientific society, however, it was soon abundant due to a lack of physiological interpretation of the signal and the development of the ECG, a more reliable and easier to interpret diagnostic device.
Recently, the BCG has resurfaced and has once again gained a lot of interest due to recent technological developments in the biomedical field. New types of sensors combined with more powerful signal processing techniques show promising results and they open up the road to new applications.

1.2.1 BCG Signal

The BCG signal is defined as a measure of the ballistic forces generated by the blood being pumped by the heart. A typical BCG signal is characterized by 3 distinct peaks, I, J, and K which help characterize the functional waves of the heart. These three describe the systolic wave and they can also be preceded by a point G which marks the beginning of the pre-systolic wave and they can be followed by two more points, L and M which show the diastolic wave. The three points

that mark the systolic wave are the ones that we are most interested in and can be correlated to the QRS peaks of an ECG signal.
The characteristic morphology of a BCG signal is presented in the following schematic alongside the corresponding ECG signal.

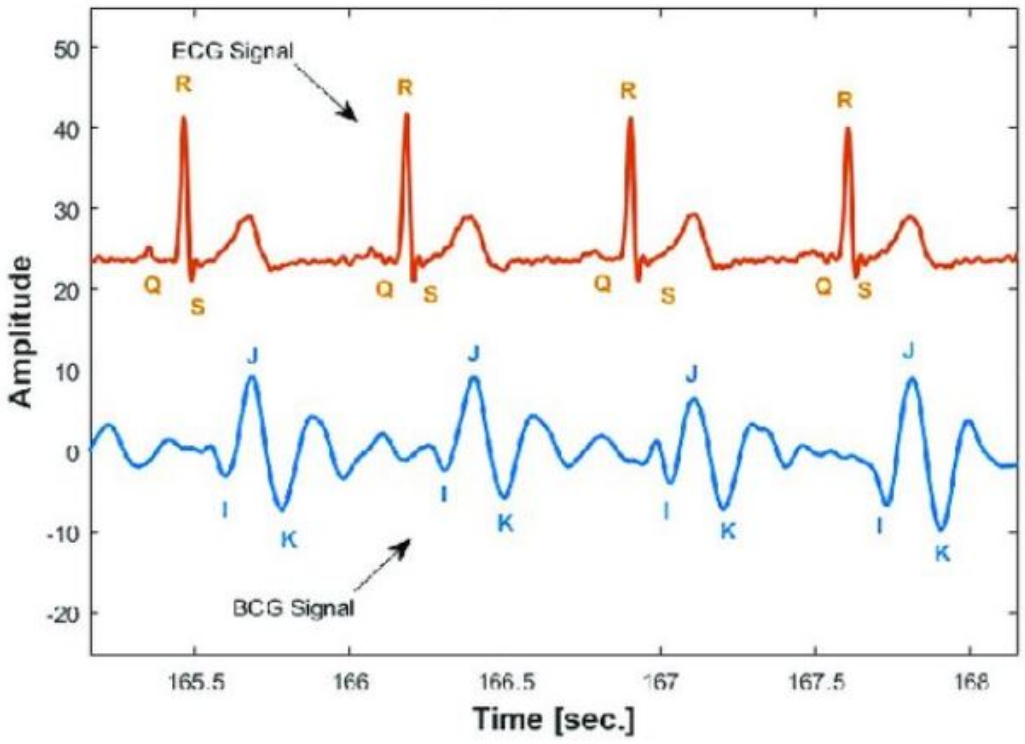


Figure 1.1: Electrocardiogram (ECG) and Ballistocardiogram (BCG) signals. BCG lags ECG because the electrical activity causes mechanical action. Image acquired by [1]

1.2.2 BCG Sensors

A wide variety of sensors can be used to record the BCG signal. The most commonly used ones are Piezoelectric Polyvinylidene Fluoride-base sensors, Electromechanical film-based sensors, Pneumatic-based sensors, Strain-gauge based sensors, Hydraulic-based sensors, and Fiber Optic-based sensors [8]. The sensors are usually integrated into various everyday used items like beds, chairs, and weighing scales so that their use is easier.

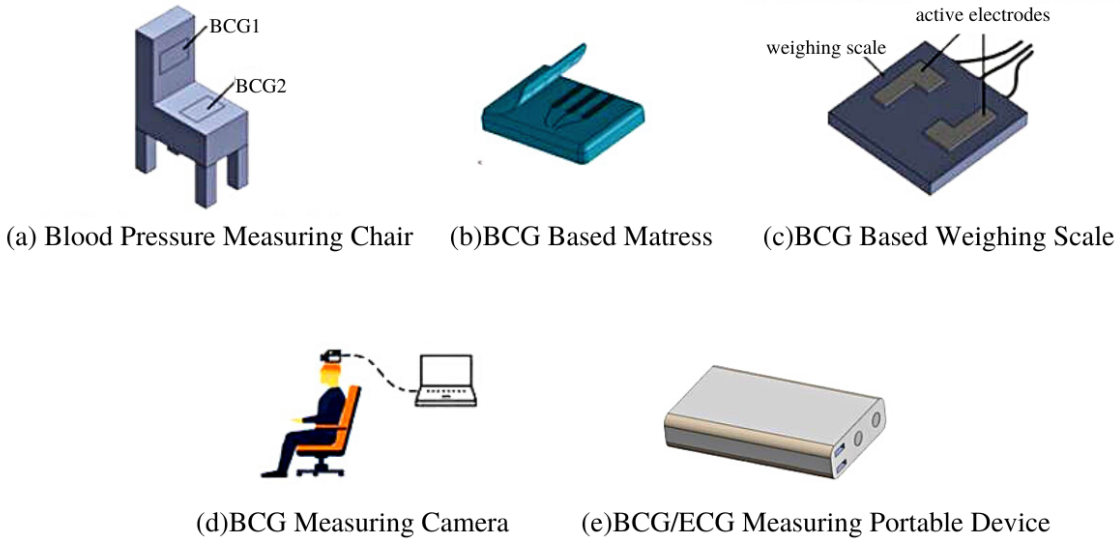


Figure 1.2: Devices that can be used for the recording BCG. Image acquired by [2]

Each type of sensor presents different advantages and disadvantages and of course, requires different preprocessing steps to improve the quality of the signal.
It is important to note that though there is a very wide variety of sensors, only a few of them have FDA approval or CE marking, which is very important when developing biomedical sensors.
MuRata company is one of the few companies that offers FDA-approved sensors for recording the BCG signal. The company first developed the 3-axis inclinometer BCGMCU. Recently, they launched the second generation of the sensor which is now offered as part of the preprogrammed microcontroller BMGMCU-D01, which can extract heart rate and other vital signs from the signal.
Casana is another start-up company that is currently undergoing the FDA clearance process for the heart seat device that they have developed. The company has integrated into a toilet seat an ECG, a BCG, and a PPG sensor that can monitor a patient's health without interrupting it. The sensors are powered by a battery that can last for

several years without recharging and they upload the recorded signals to the cloud using a WiFi connection.
Drowzle company has developed a very innovative mobile app that can record and analyze sleep breathing patterns. Though it does not offer the possibility to diagnose sleep apnea, it can help detect cases when a consultation with a doctor is needed. The app has a monthly subscription fee of 400$ and can monitor up to 5 patients with that subscription.
WatchPAT company is the only company at the moment which offers an at-home solution for diagnosing sleep apnea. To do so, the company supplies the patient with a device that is comprised of three parts, one sensor placed on the patient's chest, one placed on their finger, and the main component of the device which is placed on the patient's wrist. The company promises an accuracy of 89% and the device has a cost of approximately 400$.

1.3 Thesis Innovation

As the medical field embraces new applications of computer science, digital signal processing can produce innovative analysis methods which can have a huge impact on the medical field. The present thesis aims to contribute to the further development of the field by exploring the potential use of BCG signals obtained during patient's sleep to monitor physiological signals and identify apneic events using machine learning and artificial intelligence techniques.

1.4 Outline Of Thesis

The present thesis is divided into 4 chapters. Chapter 2 presents the literature review on current research conducted on detecting sleep apnea. In Chapter 3 we present the algorithmic pipeline for the evaluation of the heart rate and the respiration rate and finally, in Chapter 4 we present the acquired results for the proposed methodology and the machine learning models.

Chapter 2

State Of The Art

In this chapter I introduce the bibliographic research conducted on the processing and analysis of the BCG signal. We also include a summary of the theoretical background for each studied method, as well as their main advantages and disadvantages.
Our survey started more generally by looking into research papers focusing on detecting OSA events using PSG signals. Then we narrowed the research to methodologies developed for detecting OSA using BCG. This included more traditional signal processing approaches, as well as more modern ones that employ machine learning and artificial intelligence techniques.

2.1 Proposed Methodologies Using PSG Signals

2.1.1 PSG Signals

The work of Koley et al. [3], was very useful for this part of the bibliographical research. The authors of the paper proposed a methodology based on the oronasal airflow signal of the patient. The researchers segmented the signal using 8s windows with a 2s step, creating smaller, overlapping segments of the signal. For each segment, 24 features were calculated, both in the time and frequency domain.

LIST OF THE EXTRACTED FEATURES

Label	Description	Type
F1, F2, F3 & F4	Minimum, maximum, mean and variance of IRA	Time-domain
F5, F6, F7 & F8	Minimum, maximum, mean and variance of IRI	
F9 & F10	Area and length of respiration signal	
F11	90^{th} Percentile	
F12 & F13	Logarithm of total power and power in HF band	Frequency domain
F14 & F15	Relative powers in LF and HF band	
F16	Respiratory frequency	
F17	Logarithm of power in respiratory frequency	
F18 & F19	Mean and variance of PSD in 0.05-0.5 Hz	
F20	Approximate Entropy	Non-linear
F21	Lempel-Ziv Complexity	
F22	Largest Lyapunov exponent	
F23	Higuchi Fractal Dimension	
F24	Correlation Dimension	

Figure 2.1: List of the extracted features Koley et al. Table acquired by [3]

Then, using an SVM classifier, they classified the segments into normal breathing ones and segments where an abnormal breathing event takes place. Afterwards, using a second SVM classifier, they distinguished the segments between apnea and hypopnea events. Their algorithm was assessed using a dataset which comprised of 28 air recordings, out of which, 18 were used for training the algorithm and 10 were used for external validation. Additionally, a separate group of 8 patients was used for online testing. The results were very impressive as they managed to achieve an accuracy of 94.9% and 91.8% accuracy for the online classification. Furthermore, their approach allows real-time monitoring and it can adapt to different users automatically.

Another promising approach is that of Lazazzera et al. [9]. The authors selected the oxygen saturation and the photoplethysmography (PPG) signals to develop their methodology. Both signals are acquired using an oxygen-desaturation and a DAP detector, respectively, placed

on the index finger of the patients. Initially, both signals are preprocessed by downsampling and detrending. The PPG is then further processed and the envelope of the signal is calculated, based on which the researchers can identify sudden drops in the signal based on an adaptive threshold. This way they are able to identify DAP events (reductions in amplitude fluctuation of PPG) and also they can calculate the pulse rate variability. As for the SpO2 signal, initially, outlier values are removed. Afterwards, in an extended window around the DAP event, a Sleep Disruptive Breathing Event (SDBE) is detected if there is a difference of 1% - 3% between the maximum and the minimum value of the SpO2 signal in that window. The two signals are then segmented into 1-minute long segments and 37 features are calculated which include PPG features, SpO2 features, pulse rate time domain features and pulse rate frequency domain features (computed by using the smooth pseudo Wigner-Ville distribution (SPWV) and the Lomb periodogram). The researchers compared the performance of 25 machine learning algorithms on a dataset of 96 overnight recordings. Fine Gaussian SVM performed best, with an accuracy of 92.6% for separating central from obstructive apnea, 83.7% for central apnea and central hypopnea, and 82.7% for obstructive apnea and obstructive hypopnea. A major advantage of this method is that it can classify 4 possible breathing abnormalities using very low-cost equipment.
In 2020, Yonn et al. [10] proposed a methodology based on the oxygen saturation level drops of a patient. In particular, the researchers placed a MARS, type 2001 pulse oximeter on the index finger of the non-dominant hand of the patients to measure the SpO2 levels. The digitalized SpO2 values were provided every 1 sec. with 1% resolution and is processed using the Motion Artifact Rejection System (MARS) algorithm. Afterwards, the researchers defined 3 distinct points on the acquired signal. The first one, point A, refers to a signal drop rate of 1-3% per second. The second one, point B, refers to a signal drop rate of at least 3% compared to point A. The third one, point C, refers to a drop rate by 1% smaller than that of point A or 3% higher than point B. The time period between point A and point I should be between 10 and 90 sec. Finally, the signal is segmented into 1 min parts, and using a regression model which employed the Hill function, the researchers

calculated the AHI in the total recording. To test their proposed algorithm, the researchers used a dataset of 15 overnight recordings and they calculated the accuracy of the approach at 87.5%. A major benefit of their proposal is that they can have a near real-time detection of hypopnea events.

Another interesting approach to detecting OSA events is that of Tom Van Steenkiste et al. [11], published in 2018. The methodology was very innovative as they proposed the use of a Long Short Term Memory Neural Network, which they trained using as few PSG signals as possible. The selected signals were the ECG recording of the patient and the respiration signal which was recorded by placing an elastic belt on the chest of the patient that records its movement. The signals were pre-processed using a low pass filter to remove the noise, as well as a moving average filter to remove outlier values caused by limb movement. Then the signals are segmented into 30 sec long signals using a step of 1 sec to create overlapping segments. A sleep technician was responsible for labeling the smaller segments. The problem with this approach was that they created a very unbalanced dataset as the duration of apneic events is significantly smaller than normal breathing. To tackle this problem, the researchers performed balanced bootstrapping, which uses all of the segments with a sleep apnea label and an equal number of segments of normal breathing, selected randomly. The researchers used a dataset of 1008 female and 1092 male subjects with 6 h or more of useful recording. Some of the advantages that the proposed method has have been that the LSTM model is very robust to noise and the granularity of the detection can be increased to a per-second basis. The final accuracy result they were able to achieve was 77.2%, which outperformed the ANN, LR and RF, and it was achieved when the respiratory belt was placed below the lower edge of the left rib cage. A disadvantage of this method is that because the decision is made on a per sec basis, shorter respiratory events may be identified as apneas. Additionally, as the annotations are performed by a sleep technician, the positions may not be labeled with a granularity precision to seconds.This can cause the model to

detect apnea events with a time delay.

A recent and promising paper that came up during the bibliographic research was by Moridani et al. [4] published in 2019, who also used neural networks to detect apnea events. The researchers used the EEG, ECG, and EMG signals recorded during a PSG and performed wavelet analysis for 3 levels on all 3 signals. For each signal, the following 8 features were calculated and normalized.

No	Feature Type	Formula
1	Mean	$M = \frac{\sum_{n=1}^{N} x(n)}{N}$
2	Standard Deviation	$\text{STD}= \sqrt{\frac{\sum_{n=1}^{N}(x(n)-M)^2}{N-1}}$
3	Kurtosis	$\text{KURT}= \sqrt{\frac{\sum_{n=1}^{N}(x(n)-M)^2}{(N-1)STD^4}}$
4	Root Mean Square	$RMS = \sqrt{\frac{\sum_{n=1}^{N} x(n)^2}{N}}$
5	Energy	$En = \sum_{n=1}^{N} x(n)^2$
6	Harmonic Mean	$Harmo = \frac{N}{\sum_{n=1}^{N} \frac{1}{x(n)}}$
7	Skewness	$SKEW = \frac{\sum_{n=1}^{N}(x(n)-M)^3}{STD^3}$
8	Correlation Coefficient	$CC_{x,y} = \frac{cov(x,y)}{\sigma_x \sigma_y}$

Figure 2.2: List of the extracted features Moridani et al. Table acquired by [4]

To reduce the feature dimensions, the researchers used PCA and the selected features were entered into an MLP neural network. The proposed algorithm was tested on a dataset of 14 overnight recordings and it performed with a 96.87% accuracy. A major advantage of this algorithm is that it can achieve a near real-time detection of apnea events.

To make it easier to summarize and compare the aforementioned studies, we created the following table.

Author	Year	Signals Used	Objective	Accuracy
Koley et al.	2013	Oronasal Airflow Signal	OSA Classification	94.9%
Lazazzera et al.	2021	SpO2 and PPG	OSA Classification	92.%
Yonn et al.	2020	SpO2	OSA Classification	87.5%
Tom Van Steenkiste et al	2018	PSG Signals	OSA Classification	77.2%
Moridani et al	2019	EEG, ECG, and EMG	OSA Classification	96.87%

Though all the aforementioned proposals present very promising results, they are based on signals recorded during a PSG examination. This means that these signals are measured intrusively, causing discomfort to the patient and interfering with the examination results. This is probably why the datasets in which the algorithms were tested were relatively small. Additionally, they require very expensive equipment for the recording, as well as a trained person who monitors them throughout the duration of the examination. A promising solution is the use of the BCG signal to identify apnea events.

2.2 Proposed Methodologies Using The BCG Signal

In this section of the report, we will present proposed methodologies in the literature which used the BCG signal as the primary source of information. The vast majority of the methodologies have focused on accurately detecting the heart rate of a patient from the BCG, as to this day this is still an open issue. Other studies have focused on

calculating the respiration rate based on the BCG signal, while very few have moved on to try to detect OSA events.

2.2.1 Heart Rate Detection

Cardiovascular changes accompany every apnea event, which highlights the need for calculating the heart rate using a BCG signal. In particular, studies have shown that during apnea, relative bradycardia can be observed which is immediately followed by tachycardia as the respiration of the patient reverts.

Kortelainen et al. in their work [12] used a method commonly found in voice analysis applications to detect the heartbeat of a patient from a BCG signal. In particular, their proposal is based on the calculation of the Cepsrum, which is the inverse Fourier transform of the logarithm of the signal's spectrum. This way they were able to calculate the periodicity of the signal and extract the heart rate. The algorithm was tested using 15-night recordings from 6 male subjects and achieved a relative error of 0.35%. A major drawback of the method is that due to the patient's movement during sleeping, a lot of data is discarded and not handled in a different way.

An interesting approach was suggested by He et al. [13] in 2019 to determine if the time delay between the R peak of the ECG signal and the J peak of the BCG signal can be used to as an alternative to the pulse transit time (PTT). This way, a cuffless blood pressure estimation could be possible. The researchers used the ECG (dry electrodes was mounted on both the left and right index fingers), BCG (capacitive wristband placed on the left wrist), PPG sensors (optical sensor placed on the left patient's wrist) and the reference continuous BP cuffs to record the respective signals for 10 healthy individuals.
Initially all of the signals are filtered using a Type I band-pass filter with different band ranges for each signal in order to remove out-of-band components. Additionally, a 60Hz notch filter was used to remove the powerline noise.
The researchers used a peak detection algorithm based on local max-

ima to identify peaks on the ECG and PPG signals. Finally, to determine the peaks in the BCG signal, they applied a J peak detection algorithm which uses the J peaks detected in the ECG signal as reference points. In particular, the algorithm works in the following way:

- Peak detection is performed based on local maxima
- Based on estimated pulse duration, the BCG is segmented and the local maxima of each segment is calculated.
- The ECG R peak is used as an indicator to detect the J peak of the BCG, as it usually follows the QRS complex with a time delay of 0.22 to 0.26 sec.
- RJ intervals, meaning the time delay between an R peak of the ECG and J peak of the BCG, and PTT, meaning the time delay between the R peak of ECG and P peak of PPG are calculated.

Based on these calculations, the researchers calculated the SBP of each subject which can provide information on the BP variability. The method was tested on a dataset of 15-minute recordings of 10 healthy subjects and showed very promising results. The researchers used the correlation coefficient, mean absolute difference (MAD) as well as root mean standard deviation (RMSD) to assess the performance of the suggested methodology and the mean results they acquired were 0.626, 2,819 and 3.465 respectively. However, as the calculations were optimized for each subject in the dataset, the researchers would like to use machine learning models in the future to have a better estimation of the accuracy of the algorithm.

A very recent proposal is that of Zaid et Al. [14] in 2021. The authors, after pre-processing the BCG signal with a low-pass filter, used a 2^{nd} degree polynomial to smooth the BCG signal. Then, they identified the R peaks on the ECG recording using the Pan-Tompkins algorithm and for each time interval identified by two consecutive peaks, a maximum peak was identified in the BCG signal. This way they were able to calculate the time interval between R peaks in ECG and J peaks in BCG, commonly known as TEB, which is an important marker for

ventricular contractility. The researchers tested their algorithm on a dataset of 3 10-minute recordings from 8 subjects out of which 6 were patients of SICU and 2 were healthy individuals. The method proved to be robust to the position of the patient on the bed and is very promising for monitoring patients in ICUs. However, the only assessing method conducted by the researchers was the use of boxplots to determine the outlier values of the distance of the highest BCG value to the first R peak of the ECG signal (TEB).

The previously mentioned works present very interesting results using the BCG signal. However, the analysis of the BCG signal heavily relies on the corresponding ECG recording or it is not used to detect possible OSA events. In the present thesis, the ECG recording is only used as ground truth to assess the produced results. Once the methodology is established and properly assessed, it can run completely independent of the ECG and the rest of the recorded PSG signals offering truly an invasive way of monitoring sleep apneas.

2.2.2 Respiration Detection

Tavakolian et al. [15] focused on the respiration information that the BCG signal holds. The objective of their study was to produce a methodology for averaging the recorded BCG signal, in order to provide a better signal that could be used as a device output. The researchers based their methodology on the averaging of the BCG signal. In particular, the inspiration and expiration points were first identified as well as the R peaks on the recorded ECG. R-R intervals which were shorter by 75% of the average R-R were removed. Then, a linear regression line was fitted to each detected respiration cycle. Finally, the inspiration and the expiration points in the cycle were averaged, thus creating an inspiration and expiration BCG template. The proposed algorithm was tested on a dataset containing 45 recordings of subjects without any heart-related pathologies before and after running on a treadmill. To evaluate the algorithm, a measure of similarity was established by cross-correlating the inspiration and the expiration cycles of all subjects. After performing statistical analysis, a p-value smaller than 0.01 was calculated between the respiration and the expiration

averages.

Another interesting work is that of Liu et Al. [16] presented in 2016. The researchers used the BCG signal to identify respiratory patterns which then were used to identify OSA events. After preprocessing the BCG signal, they used the 7th wavelet approximation to detect signal arousals. Based on them, they identified potential events and segmented the BCG signal accordingly. For each segment, a feature extraction step was performed and out of the calculated features, 5 fine-grained were selected to classify the segments. In particular, the features used were: Sample Entropy (SampEn), Zero Crossing Rate (ZCR), Mean Number of Extreme Points (MNEP), Average Signal Turns Count (ASTC), and Average Cumulative Amplitude Change (ACAC). Finally, the classification was performed using a back propagation neural network. The researchers tested their proposal on a dataset that was comprised of one night's sleep recordings from 38 subjects and were able to achieve a 94.6% accuracy when they validated their model using 10-fold classification.

2.2.3 Heart Rate and Respiration Rate

Wang et al. [17] was one of the first published papers that focused on the calculation of a patient's heart rate from a BCG signal. The researchers developed a new PVDF sensor and acquired 1-hour recording of the BCG signal. To process the signal, the researchers used Wavelet decomposition of the signal to separate the heartbeat and respiration components. In particular, the researchers used the 6th approximation is used to detect respiration peaks using a time-variant adaptive threshold. As for the detection of the heartbeat, the researchers calculated the first level approximation of the signal. After squaring the signal, the envelope of the signal is calculated using a moving average smoothing algorithm, and using an adaptive threshold, the peaks of the signal are identified. To improve the peak detection, a judgment method is introduced based on which if two consecutive peaks are identified with a time delay smaller than 0.25 sec, then the one with the highest value is considered the true positive and the second is discarded. An error rate of 1.25% is calculated, however,

the dataset is very small and further testing should be performed to accurately access the performance of the algorithm.

A more recent work that was published in 2021, is that of Nasr et al. [18]. The researchers used machine learning techniques to access a patient's sleep. In particular, based on their previously published work, after calculating the heart rate and the respiratory rate of the patient, they calculated two additional features: the spectral flatness measure and the spectral centroid. To achieve this, the BCG signal was segmented using equal length segmentation and then the Gaussian Mixture Model to classify the data into 7 classes: normal activity in a still position, coughing, post-coughing, holding breath, expiration, movement, and others. A binary classification method that used KNN was then used to determine a relation between the multiple classes and the functions of the human body. This way two binary flags were created, a CAD/NoCAD, to describe the cardiac activity, and RAD/NoRAD, to describe the respiratory activity. The proposed algorithm was tested on 5 min recordings of 3 male and 3 female subjects. The algorithm was evaluated and a 98% accuracy was calculated for classifying CAD events, however, a 33% misclassification for NoCAD events, a 95% for classifying RAD events, and a misclassification of 28% for the case of NoRAD events.

To make it easier to summarize and compare the aforementioned studies, we created the following table, which contains the basic characteristics of each study. It is also worth noting that most of them were tested in very small populations using a small dataset.

Author	Year	Signals	Objective	Accuracy
Kortelainen et al.	2007	BCG	HR Calculation	94.9%
He et al.	2019	ECG & BCG	Blood Pressure estimation	r:0.626 MAD:2,819 , RMSD:3.465
Zaid et Al.	2021	BCG	TEB Consistency	not calculated
Tavakolian et al.	2008	BCG & ECG	BCG Recording improvement	p-value:0.01
Liu et al.	2016	BCG	OSA Classification	94.6%
Wang et al.	2017	BCG	HR Calculation	98.75%
Nasr et al.	2021	BCG	OSA Classification	98%

Chapter 3

Proposed Methodology for Evaluation of BCG Signal

The methodology proposed for this thesis addresses two separate points. The first one is to assess the accuracy of the calculated physiological features, the heart rate, and the respiration rate from the BCG signal. As the BCG signal record is affected by every single motion of the patient during their sleep, such as limp motion, respiration, heart beating, etc, it is important to first examine the precision of the extracted physiological signals. The reason for this is that currently, there is not a determined protocol on how to record and process the BCG signal in order to have physiological values extracted from it. Additionally, during apneic events, the patient presents bradycardia followed by tachycardia as the respiration returns to normal patterns [19], a pattern which would be very useful to utilize when performing the OSA event detection. For that reason, it is only after establishing the accurate calculation of the physiological features (HR and RR) that the second part uses these as well as a number of other selected features to perform classification of the signal into normal and apneic events.

3.1 The Datasets

For the development and testing of the proposed methodology, multiple datasets were used.
The first dataset that we used is provided by the Scientific Challenge

Competition of the Health Informatics Working Group of the IFMBE (https://www.icbhi2022.com/call-for-sc/) and will be referred to as dataset 1. It contains 4-5 min. ECG and BCG recordings for 47 healthy volunteers and patients suffering from Atrial fibrillation. For the recording of the signals, the subjects were sitting in an EMFi-fitted chair for measuring BCG and wearing 3 ECG electrodes for a 1-derivation ECG record. Both signals were acquired simultaneously. In the dataset, both the ECG and the BCG signals are provided and additionally, the locations of the ECG R peaks are annotated.
The second dataset used is an open access dataset published on the IEEE Data Port [20], created under Kansas State University IRB protocol 9386 and will be referred to as dataset 2. The dataset was originally used by the researchers in order to study the possibility of monitoring changes in systolic blood pressure and stroke volume using the BCG signal. The BCG signals were collected from a custom, bed-based ballistocardiographic system comprised of four electromechanical films and four load cells. Affiliated cardiopulmonary signals were acquired using a GE Datex CardioCap 5 patient monitor (which collected ECG and PPG data) and a Finapres Medical Systems Finometer PRO. To accurately measure these changes the researchers used the ECG's J peaks positions and identified the most prominent peak in the BCG signal in a window of 100 and 400 ms after the J peak, making their developed methodology highly dependent on the ECG signal.
The first and second datasets were solely used for assessing the calculation of the HR using the BCG signal compared to the calculated HR of the ECG signal.
Dataset 3 [21] was recorded for the work of Wang Z et al. [17]. The public dataset contains 136 overnight BCG sleeping recordings of 41 healthy individuals, 23 mild sleep apnea patients, 34 moderate sleep apnea patients, and 38 severe sleep apnea patients. The recordings have a minimum duration of 6h and a maximum of 9h. For the recording, a micro-movement sensitive mattress was used (MSM), an AD converter, and a terminal PC, while the sampling frequency is set to 100Hz. For each recording, information on the duration and the position of the apnea event is also available. The additional information that it brings is the labeling of apnea events on the BCG signal, which

was vital for classifying the BCG segments into apneic and normal. Finally, the methodology was also tested on data collected from the FORTH institute which will be referred to as dataset 4. The dataset is comprised of 16 PSG examinations of adult patients. The recordings were performed in the sleep lab of the University General Hospital of Heraklion "PAGNI" and were approved by the hospital's Ethical Committee, while all of the patients gave signed consent for the use of their data. Alongside the PSG, a BCG recording is also available for each patient. The dataset, though small, is the most complete for the proposed work, as it offers both the PSG signals used for ground truth as well as apnea events labeling.

3.2 HR & RR Calculation

3.2.1 Signal Preprocessing

Due to the different acquisition techniques used for the recording of the BCG signal, it was necessary to preprocess each dataset in a different way in order to achieve the best possible results. For the calculation of the heart rate, the preprocessing methodology described by flowchart A was used for datasets 1-3, while for dataset B, the preprocessing steps are described by flowchart B in the figure below.

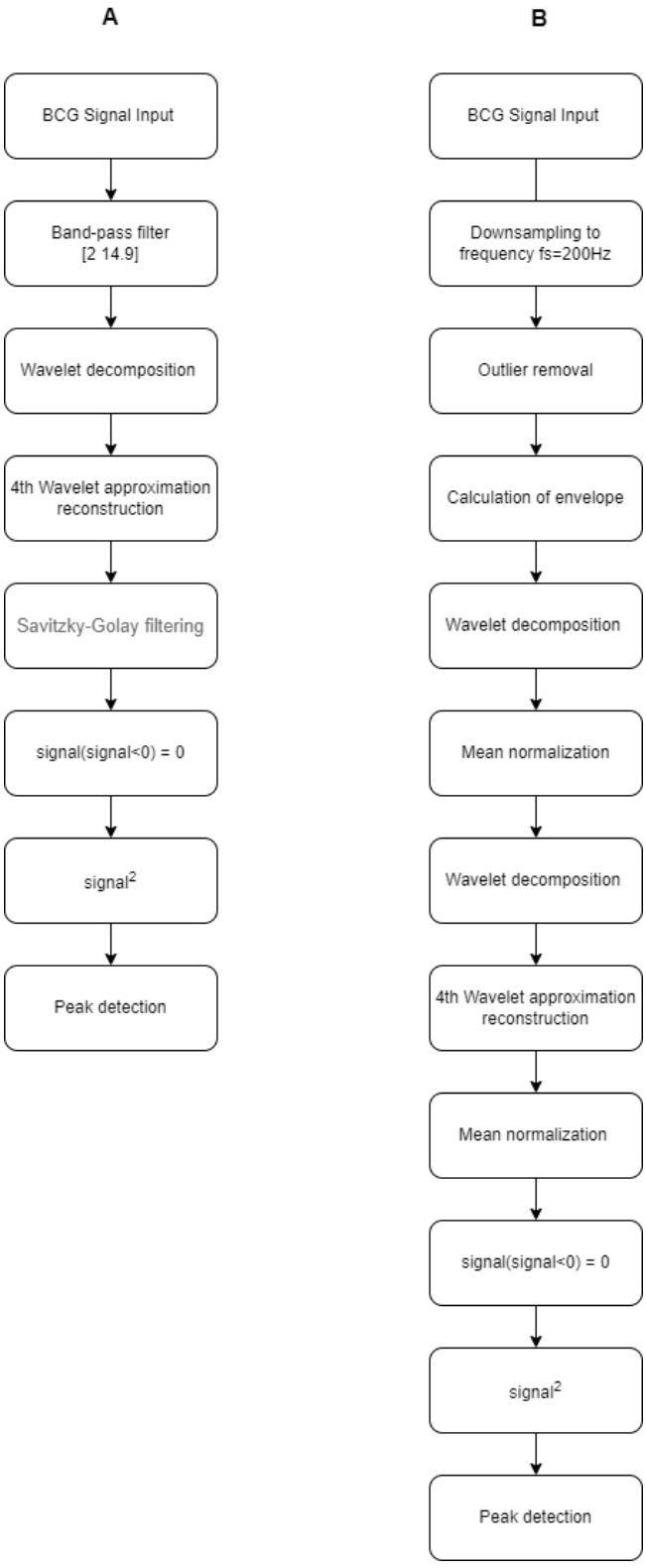


Figure 3.1: A:preprocessing flowchart of datasets 1-3 B:preprocessing flowchart of dataset 4

The datasets contained BCG recordings of the following form:

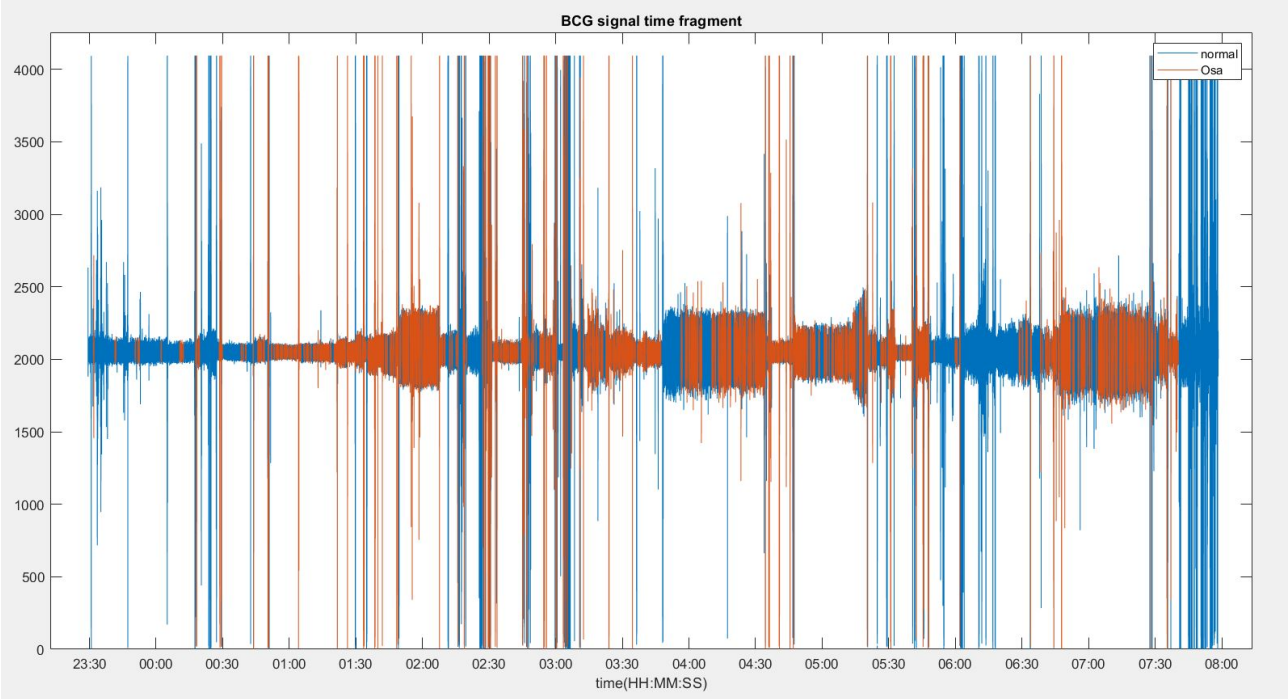


Figure 3.2: Original BCG signal with OSA annotations

Outlier Removal

A very common problem with the BCG signal is the existence of outlier values which are caused by limb movement. For the given dataset these values represent approximately 2.5% of the signal. Given that the percentage of these values is relatively low, a moving median filter is used to remove them. The window length of the filter is set to 15 minutes, while an outlier value is defined as any value greater than 3 stds. After applying the filter the outlier percentage is calculated to be lower than 0.8%.

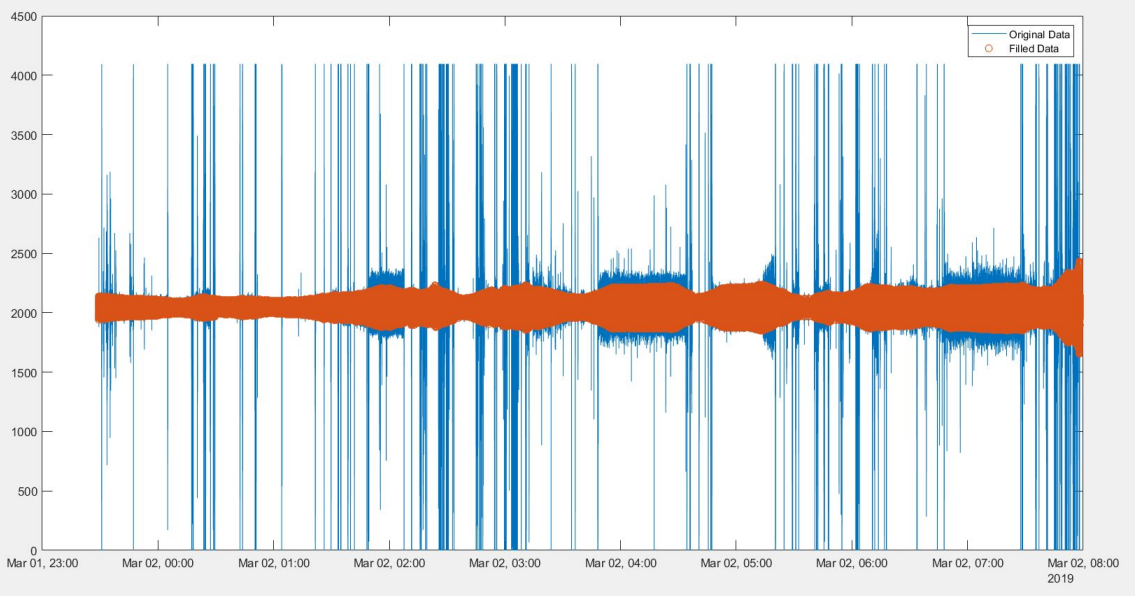


Figure 3.3: Outlier removal results Blue: Original signal, Orange: Signal after filtering

Butterworth Filtering

For datasets 1-3, the BCG signal is firstly filtered using a Butterworth pass band filter with $f_{stop}^{lower} = 2Hz$ and $f_{stop}^{higher} = 15Hz$. After these preprocessing steps, the BCG signal has the following form

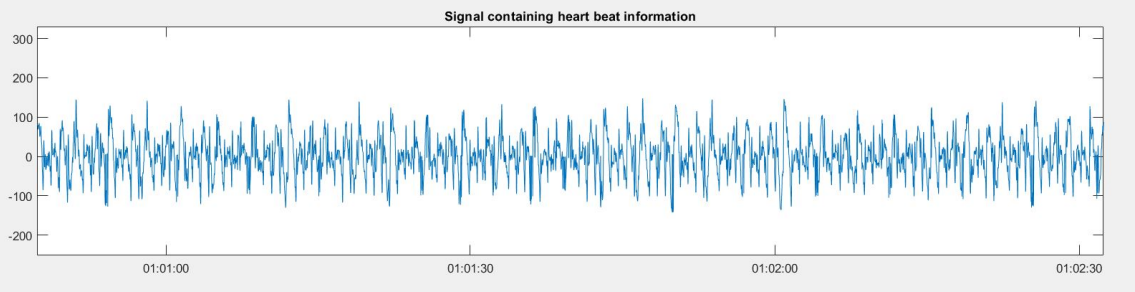


Figure 3.4: Small time frame of preprocessed signal and signal after band-pass filtering

3.2.2 Heart Rate Calculation

It is quite obvious that the current form is not ideal for performing peak detection especially for calculating the heartbeat of a patient. Based on the fact that the IJK complex of the BCG signal bears high

resemblance to the QRS complex of the ECG recording which in its turn resembles the 'sym4' wavelet, we used the maximal overlap discrete wavelet transform (MODWT) to enhance the J peaks in the BCG waveform in order to improve the peak detection. Figure 3.5 shows the scales 1-6 of the BCG signal. The level 4 approximation has the closest morphology to an ECG signal and that is why it is used for the reconstruction of the BCG signal.

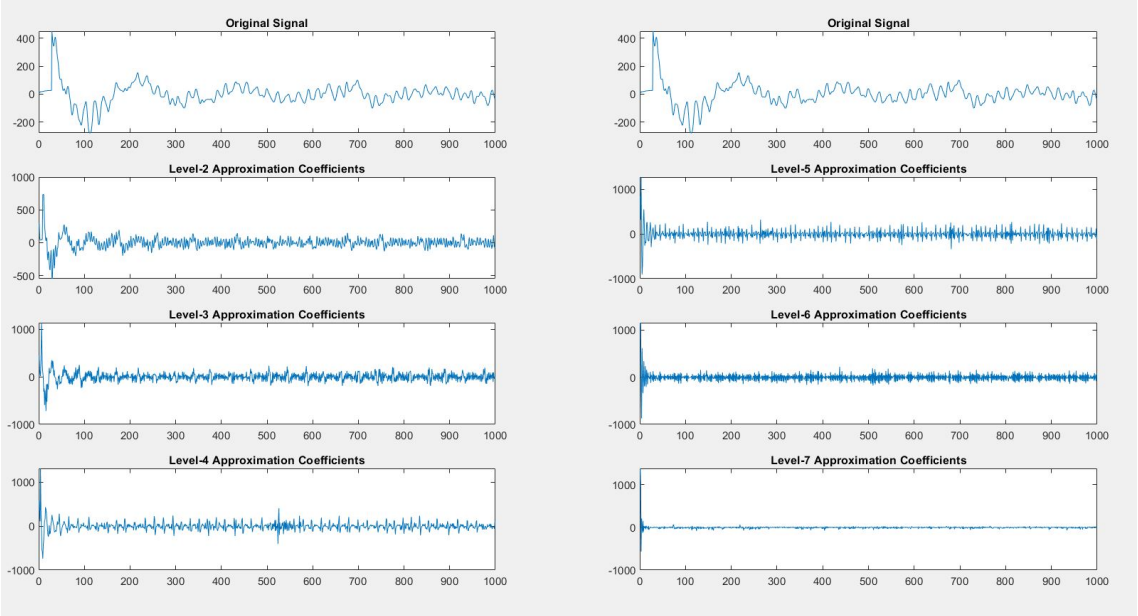


Figure 3.5: Wavelet analysis of signal

Finally, the negative values of the signal are set to zero and the signal is then squared in order to use a peak detection algorithm.

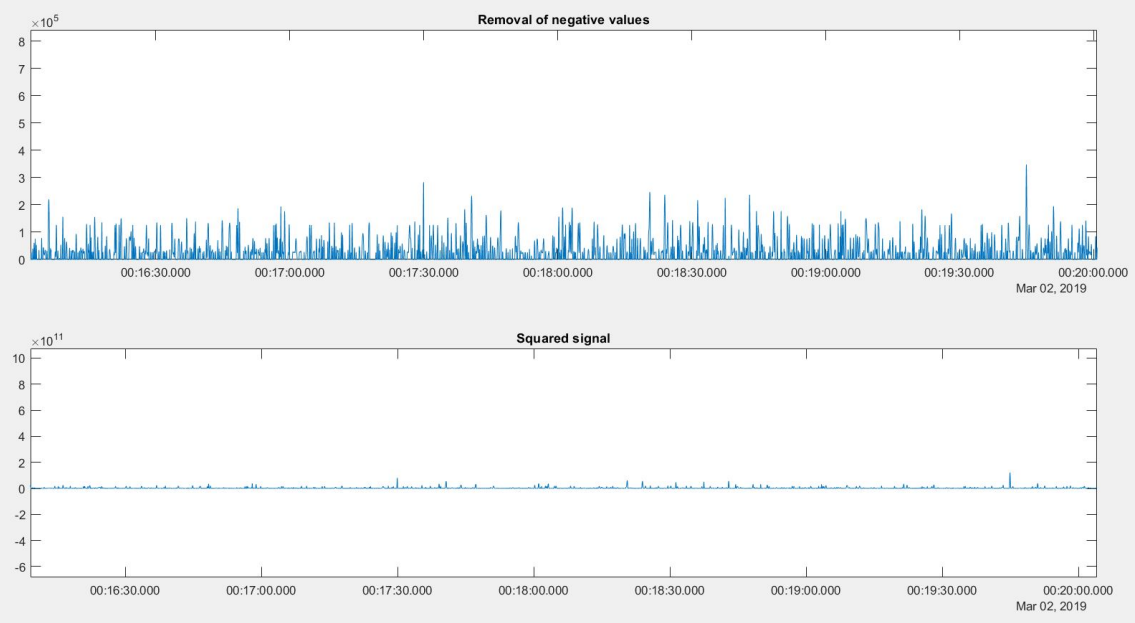


Figure 3.6: BCG signal with removed negative values and squared

For the peak detection, the minimum value is set equal to the mean value of the signal and the minimum distance between two consecutive peaks is set equal to half of the sampling frequency.

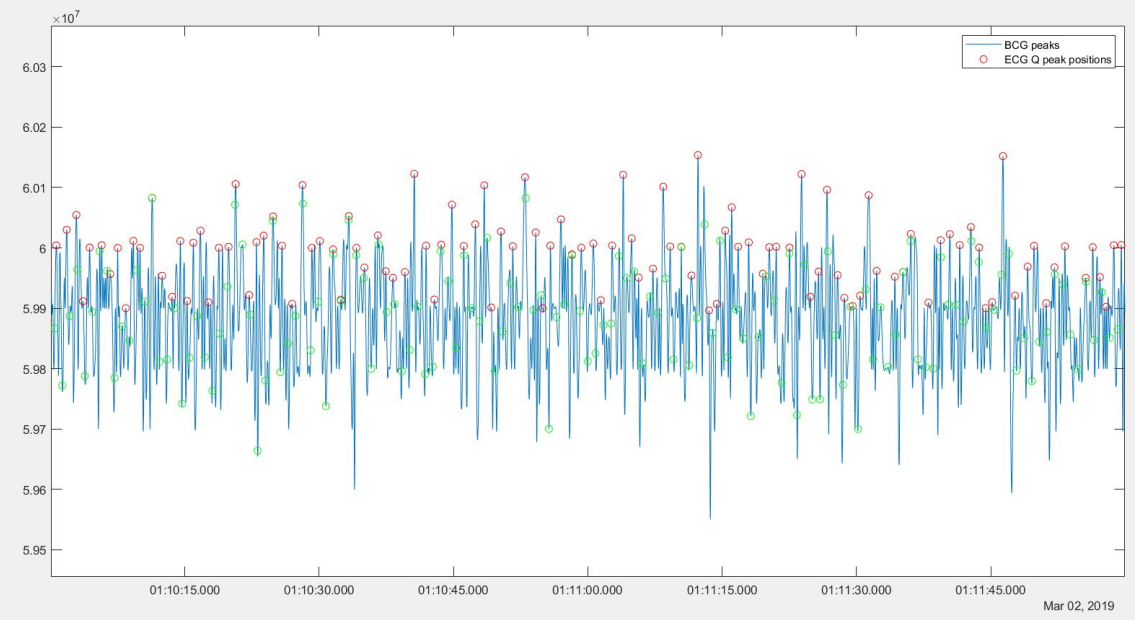


Figure 3.7: BCG detected peaks (red) vs Q-peak positions of ECG signal (green)

As for dataset 4, because the recordings were noisier, a different approach was used in order to identify the heart rate. Firstly, the BCG

recordings are downsampled from the original frequency of 1 kHz to 200Hz, as the respective ECG also has a sampling frequency of 200Hz. Then, outlier values, meaning points with a value more than three scaled median absolute deviations (MAD) away from the median of the signal, are removed using a moving median filter with a window length of 15 minutes. The upper envelope of the signal is afterward calculated, using spline interpolation over local maxima separated by at least 15 samples. Finally, the signal is reconstructed using the 4th Wavelet approximation and is normalized so that the mean value is equal to zero before applying a peak detection algorithm. For peak detection, the parameters are the same as for datasets 1-3.
After detecting the peaks, the time intervals between the peaks were calculated in order to calculate the detected heart rate. In particular, given that a normal heart rate is between 60 and 100 bpm, then if a time interval between two convective peaks is smaller than 1 bps then the segment between the peaks is labeled as tachycardia, if it is between 1 and 1.67 bps it is labeled as normal and finally if it is greater than 1.67 then it is labeled as bradycardia.
An extra step is then taken in order to eliminate the detection of very small segments of abnormal heart rate (about 2 to 3 consecutive peaks). In particular, we detect a pattern of normal heart rate, abnormal heart rate, and then normal heart rate again. If the abnormal heart rate is smaller than 25% of the total duration of that pattern, then it is discarded and considered as peak misidentification.
The signal with the annotated heart functionality eventually takes the following form.

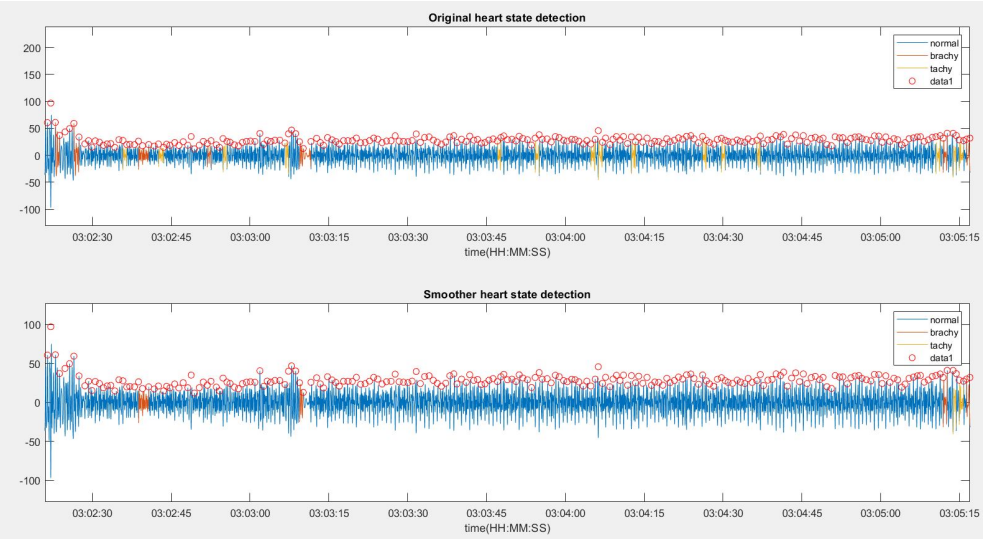


Figure 3.8: BCG signal with annotated heart functionality and detected peaks

Finally, the peak detection was assessed for datasets 1,2, and 4 with the following formula as described by the Scientific Challenge Competition of the Health Informatics Working Group of the IFMBE (www.icbhi2022.com). The locations of the R peaks are provided by datasets 1 and 2 and the ΔR denotes the distance between two consecutive peaks. The J peaks are the ones detected on the BCG signal based on the methodology described and ΔJ is the distance of two consecutive J peaks.

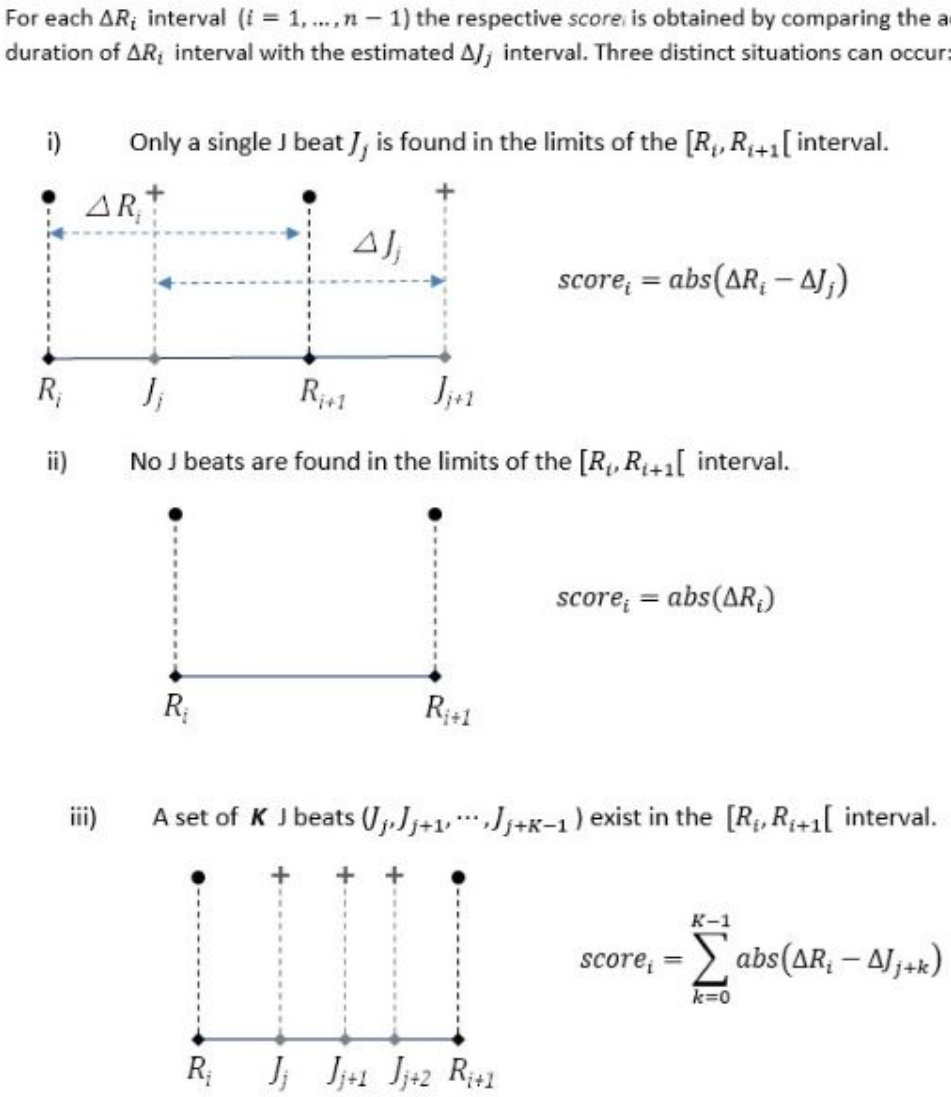


Figure 3.9: Assessment of peak detection

Basically, to assess the peak detection of the BCG signal we compare the durations of the ΔR to the estimated ΔJ, taking into consideration the three scenarios as depicted in the schematic above. The goal is to minimize the calculated score which can be achieved by detecting the same number of J peaks as R peaks with the same distance in between the peaks. The reason for using this way of assessment is the lag that the BCG signal presents compared to the ECG signal which makes it impossible to detect peaks at the same position. The global score is calculated using the formula

$$score = \frac{1}{n-1} \sum_{i=1}^{n-1} score_i.$$

The number of the detected peaks in each 30 sec segments multiplied by 2 represents the heart rate of the patient and can be used as a feature in order to identify sleep apnea events.

3.2.3 Respiration Rate Calculation

During the polysomnography procedure, one of the recorded signals is the oronasal airflow measurement using a thermistor. Alongside the use of respiratory inductive plethysmography (RIP) belts, it is currently the golden standard for monitoring the respiration of patients during their sleep. A recent study [22] showed promising results for calculating the respiration rate (RR) of the patient using only the flow recordings. In particular, the proposed method filters the flow signal using a bandpass filter and segments it into 30-sec intervals, and calculates the most predominant frequency in its frequency spectrum.
To estimate the accuracy of the calculation of the RR from solely the BCG signal, we will compare the results of our proposed method against the results of using the flow signal.
The lower frequency components of a BCG signal have been associated with the respiration of a patient [23], as the normal RR for adult people is between 12 to 20 breaths per minute. Those rates correspond to a frequency bandwidth of 0.13–0.5 Hz, however, in our proposed methodology the BCG signal is filtered with a band-pass filter of 0.1 to 2 Hz. Afterward, the BCG signal is downsampled to a new frequency of 200Hz, a moving average is used to smooth the signal, and then the signal is segmented into 30-sec segments. Subsequently, the signal is centered around zero, its negative values are set to zero and the rest are squared. Finally, a peak detection algorithm is used and the RR is calculated as the number of peaks detected in the segment multiplied by 2 as usually the RR is measured in breaths per minute.

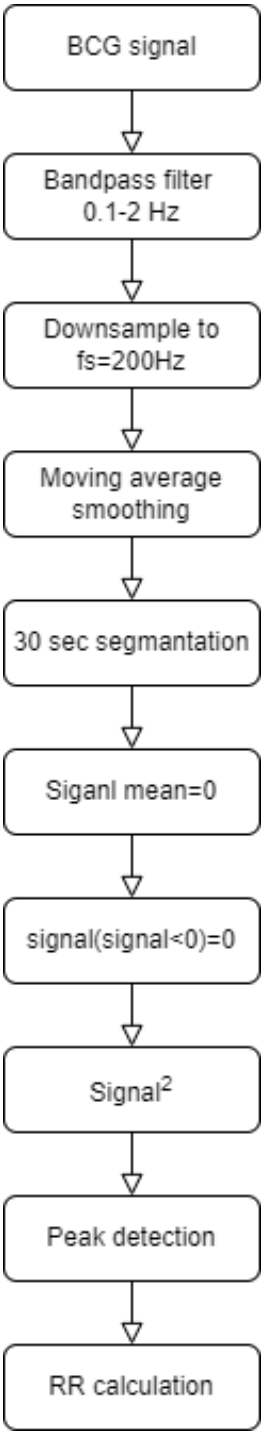


Figure 3.10: Respiration Rate calculation flowchart

The figure below shows the changes in the signal before performing the peak detection to calculate the respiration rate.

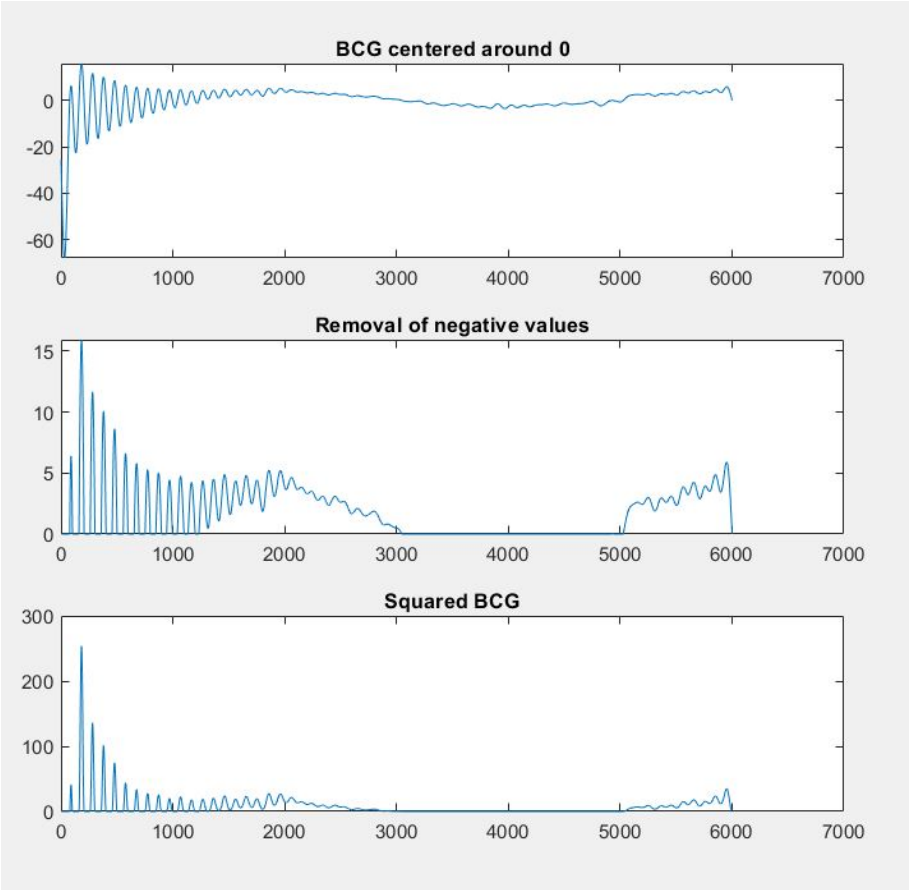


Figure 3.11: Respiration Rate Calculation steps

Finally, below follows an image depicting the original BCG signal with the OSAs annotated and the filtered BCG signal with the annotations of segments with slow (<12 breaths per min.), normal (12-20 breaths per min.), and fast breathing segments (>20 breaths per min.).

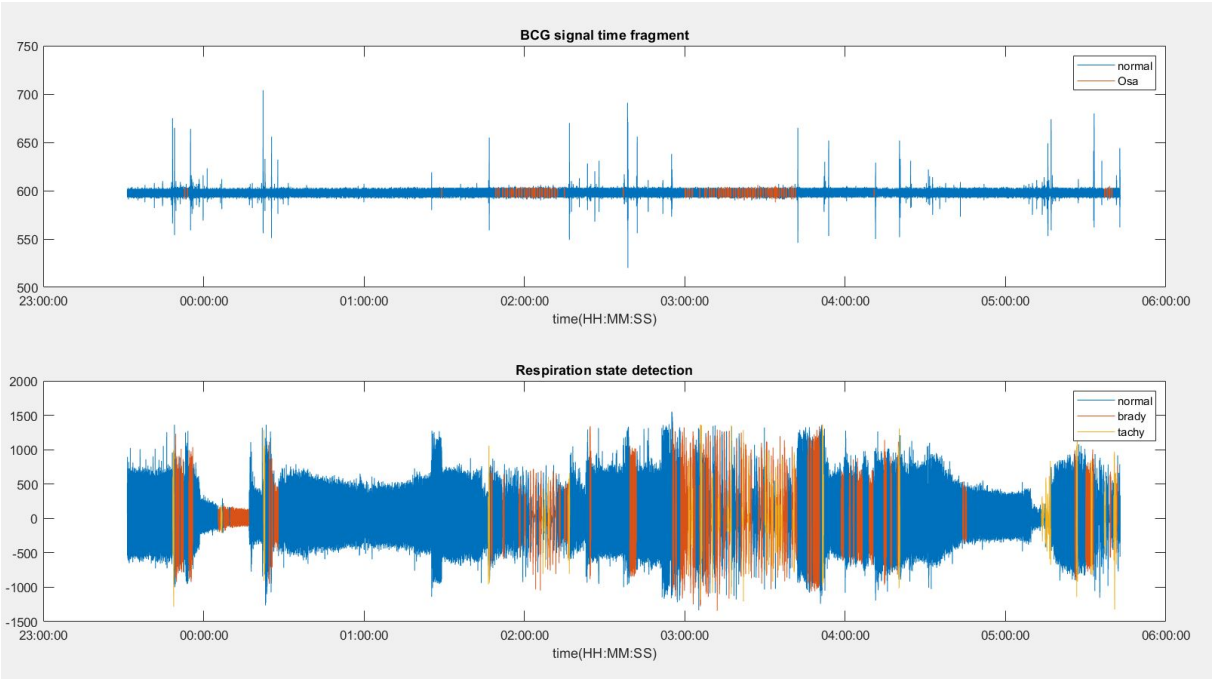


Figure 3.12: Calculated Respiration Rate using BCG signal

To assess the respiration rate the following formula was used:

$$score = \frac{1}{N} \sum_{i=0}^{N-1} |RR_{flow} - RR_{bcg}|$$

where N is the number of segments.

3.3 OSA Classification

3.3.1 Signal Segmentation

In dataset 3, the duration of the labeled OSA events has the following distribution 3.13, with a mean duration of 28 sec. For the calculation of the selected features, each recorded BCG signal is divided into 30sec non-overlapping segments. Initially, if more than 75% of the segment is in an annotated OSA event, the whole segment was labeled as 1, otherwise, it is labeled as 0. Afterward, we tested the performance of the models by changing the percentage of the OSA duration laying in each segment and the results are presented in table 4.2.1.

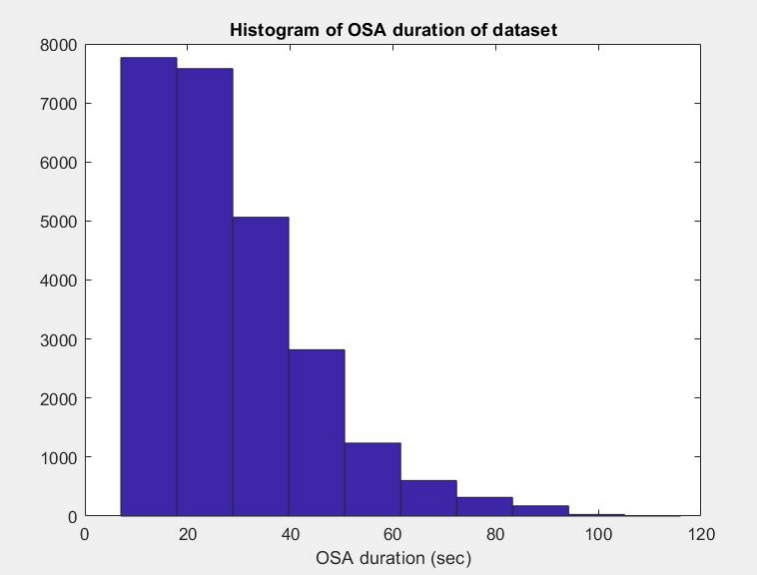


Figure 3.13: Length of OSA events

3.3.2 Feature Extraction

Unfortunately, after a statistical analysis of the OSA events and non-OSA events, it became evident that statistical features of the signal could not be used to identify OSA events because they are very similar to the non-OSA, which would lead to an inaccurate classification of the events. For that reason, the use of frequency features was explored as well as non-linear features such as entropy.

3.3.3 PSD Of OSA & NON-OSA

A very promising feature of non-stationary biological signals is the Power Spectrum Density. As it is evident from the following figure, the PSD of OSA events tends to have a lower amplitude compared to the normal segments. For that reason, there are a few PSD-based features that could be used for the successful classification of the events.

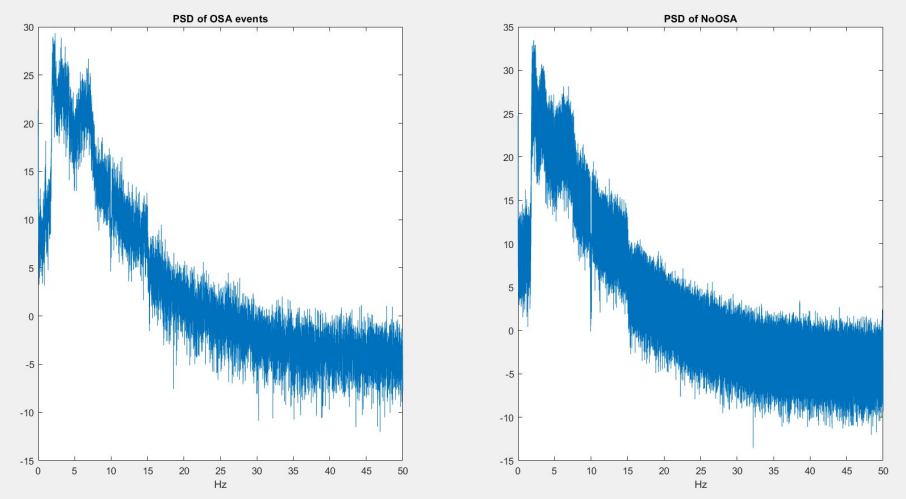


Figure 3.14: PSD of OSA and NoOSA events

A total of 12 features are calculated for each segment. In particular, based on the heart component of the BCG signal, the following were calculated

- Signal mean and std
- Skewness
- Kurtosis
- Entropy
- Approximate Entropy
- Sample Entropy
- Hurst Exponent

Then, some additional features based on the PSD of that signal were also calculated:

- $PSD_{low/total}$ ratio.
- $PSD_{high/total}$ ratio.

where,

- PSD_{low} for frequency range 0-1 Hz.

- PSD_{high} for frequency range 10-50 Hz.
- PSD_{total} for frequency range 0-50 Hz.

The final 2 features that were calculated were the heart rate variability and the respiration rate detected in the segment. Heart Rate Variability was calculated as the standard deviation of the distances between consecutive peaks in the segment while respiration rate is the number of respirations detected in the segment multiplied by 2 in order to have the breaths per minute.

3.3.4 Dataset Balancing

Since the OSA events have a significantly smaller duration compared to the healthy segments in a BCG recording, it was necessary to balance the dataset before training the models to avoid overfitting. To achieve that, we randomly select a number of healthy segments equal to the number of the OSA labeled events to create the total observations used for the classification.

3.3.5 Machine Learning

Machine Learning Models

The selected features were used to train 23 machine learning models which include:

- Fine Tree
- Medium Tree
- Coarse Tree
- Linear Discriminant
- Quadratic Discriminant
- Logistic Regression
- Linear SVM
- Quadratic SVM
- Cubic SVM

- Fine Gaussian SVM
- Medium Gaussian SVM
- Coarse Gaussian SVM
- Fine KNN
- Medium KNN
- Coarse KNN
- Cosine KNN
- Cubic KNN
- Weighted KNN
- Boosted Trees
- Bagged Trees
- Subspace Discriminant
- Subspace KNN
- RUSBoosted Trees

Machine Learning Validation

K fold cross-validation is a statistical method commonly used to assess the performance of the machine learning models. The parameter k determines the number of groups that the dataset is split into. Then, you run k learning experiments where each time a different group is used as the testing set and the rest k-1 groups are used for the training of the model. Finally, the acquired results are averaged to obtain the accuracy of the model.
In the present thesis, 5-fold cross-validation was used to validate the performance of the models. The reason for using cross validation is that the datasets are relatively small and also we wanted accomplish a less biased model as every observation has the chance of appearing in both train and test sets. This is very important as our dataset is comprised of data coming from different patients.
It is important to note that for a segment to be labeled as OSA event, a

specific percentage of it had to overlap with an annotated OSA event. Various thresholds were tested for that percentage and the accuracy of the models was calculated for each one of them. This resulted in also having a different size of a dataset as out of the total observations used, half of them correspond to normal recordings, and the other half to OSA events. The bigger the percentage necessary to label a segment as OSA event, the smaller the acquired dataset.

Chapter 4

Discussion

In the present chapter, we present the results of the methodology per dataset as well as the main conclusions and future work suggestions.

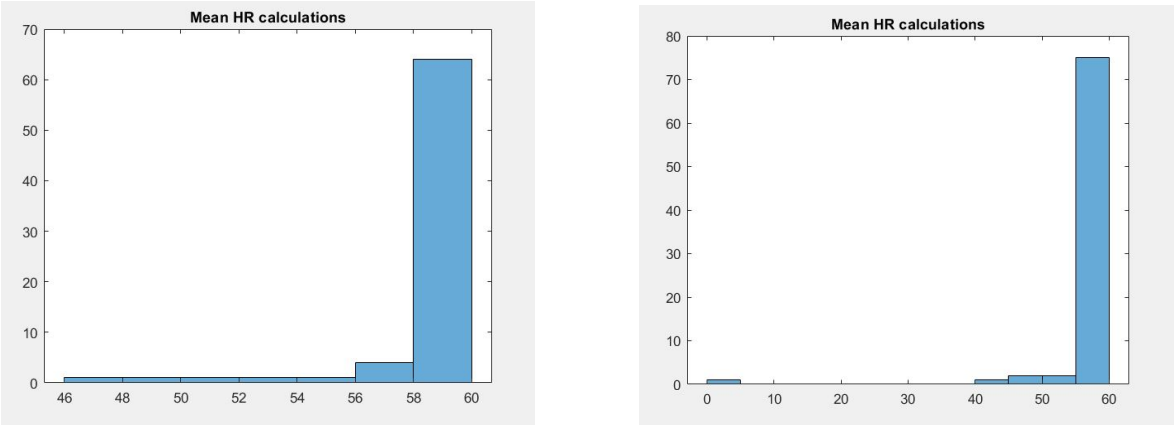
4.1 Dataset 1 and 2

4.1.1 HR Calculation

The first and second datasets were only used for the assessment of the HR calculation. The following results were obtained when using the scoring algorithm described in figure 3.9.

Dataset	Mean Score	Score std
1	43.88	31.58
2	45.26	65.96

The mean score of the first dataset was calculated at 43.88. When divided by the sampling frequency of the dataset $fs = 100Hz$, it gives a mean difference of 0.44 sec in the duration of the detected ΔJ segments compared to the ΔR segments.
The histograms of the mean HR calculated for each signal in the dataset are depicted below in figure 4.1.



(a) Dataset 1 (b) Dataset 2

Figure 4.1: Histogram of Mean HR Calculated

As we can see from the histograms, the results are very satisfactory. The mean HR calculation varies within normal ranges, especially for dataset 1. For dataset 2, the results are not as accurate as in the case of dataset 1. This is because the algorithm can detect peaks that are very close to each other, thus giving a very small heart rate value, increasing the variation of the results. This could be improved by adding further preprocessing steps to the BCG signal, however, as we wanted to keep the process as similar as possible, no further investigation to improve the results was made. Finally, we have to keep in mind that for the collection of the BCG signals, different recording techniques were used for the two datasets, thus we expect a difference in the results.

4.2 Dataset 3

4.2.1 OSA Classification Results

Due to the fact that the third dataset didn't have the corresponding ECG and flow recordings, we weren't able to assess the HR and the RR calculations.
However, compared to Datasets 1 and 2, Dataset 3 provided labelings for the OSA and NON-OSA events on the BCG signal, thus it was used for the binary classification of the signal segments.

One of the biggest challenges when developing the proposed algorithm was determining the percentage overlap of a labeled OSA event with the fixed-length segments under investigation in order to classify it as an OSA event. For that reason, we decided to test different overlapping percentages and acquire accuracy results. The biggest disadvantage that is presented as the overlapping percentage increases, the size of the dataset decreases. The reason for that is that for a larger overlapping percentage, less fixed-length segments are labeled as OSA events, and respectively, less healthy fixed-length healthy segments are selected to create the training and testing dataset.
Below follow the accuracy results calculated for the 23 machine learning models.

	Accuracy			
Model	label 20%	label 30%	label 40%	label 50%
Fine Tree	65.2%	66.3%	63.8%	63.2%
Medium Tree	64.3%	65.6%	63.5%	63.4%
Coarse Tree	62.8%	64.3%	62.0%	60.9%
Linear Discriminant	59.8%	61.8%	59.3%	59.6%
Quadratic Discriminant	51.3%	51.9%	51.4%	51.2%
Logistic Regression	58.8%	63.5%	59.2%	60.3%
Linear SVM	61.4%	64.1%	61.3%	60.8%
Quadratic SVM	62.6%	48.7%	64.3%	63.5%
Cubic SVM	49.5%	49.9%	49.7%	49.8%
Fine Gaussian SVM	66.6%	67.3%	65.7%	65.4%
Medium Gaussian SVM	65.3%	66.4%	64.6%	64.0%
Coarse Gaussian SVM	63.5%	65.5%	63.0%	62.6%
Fine KNN	59.7%	59.5%	59.7%	58.7%
Medium KNN	63.9%	64.0%	63.6%	62.7%
Coarse KNN	65.2%	66.0%	64.6%	64.1%
Cosine KNN	64.2%	63.9%	63.3%	62.5%
Cubic KNN	63.2%	63.8%	63.4%	62.5%
Weighted KNN	63.4%	63.8%	65.1%	62.6%
Boosted Trees	65.3%	66.5%	65.8%	64.6%
Bagged Trees	65.6%	66.8%	59.4%	64.8%
Subspace Discriminant	59.5%	61.8%	62.8%	59.5%
Subspace KNN	63.7%	64.2%	62.8%	62.5%
RUSBoosted Trees	64.3%	65.6%	63.5%	63.4%

	Accuracy			
Model	label 60%	label 70%	label 75%	label 80%
Fine Tree	61.1%	59.7%	63.8%	58.6%
Medium Tree	60.7%	59.2%	63.5%	58.7%
Coarse Tree	59.7%	57.5%	62.0%	58.2%
Linear Discriminant	60.0%	58.4%	59.3%	56.7%
Quadratic Discriminant	51.3%	51.7%	51.4%	51.2%
Logistic Regression	60.3%	59.4%	59.2%	57.1%
Linear SVM	60.8%	59.4%	61.3%	56.5%
Quadratic SVM	62.6%	61.6%	64.3%	59.5%
Cubic SVM	50.0%	49.6%	49.7%	49.3%
Fine Gaussian SVM	62.6%	62.5%	65.7%	60.6%
Medium Gaussian SVM	63.1%	62.4%	64.6%	60.5%
Coarse Gaussian SVM	62.0%	59.6%	63.0%	57.5%
Fine KNN	57.7%	57.0%	59.7%	56.5%
Medium KNN	61.6%	61.0%	63.6%	58.8%
Coarse KNN	62.6%	62.4%	64.6%	59.6%
Cosine KNN	61.6%	60.5%	63.3%	59.4%
Cubic KNN	61.3%	61.1%	63.4%	58.7%
Weighted KNN	61.2%	61.0%	65.1%	59.1%
Boosted Trees	63.5%	61.3%	65.8%	60.7%
Bagged Trees	62.8%	62.5%	59.4%	61.3%
Subspace Discriminant	59.7%	58.0%	62.8%	56.6%
Subspace KNN	61.5%	60.3%	62.8%	58.5%
RUSBoosted Trees	60.7%	59.1%	63.5%	58.9%

Fine Gaussian SVM outperforms the other models in most cases and it is the one that also performs best for the segmentation that labels the segment as an OSA event when 30% of it is labeled as OSA. It is important to note that as the normal segments are randomly selected, the accuracy results might slightly vary on consecutive runs.

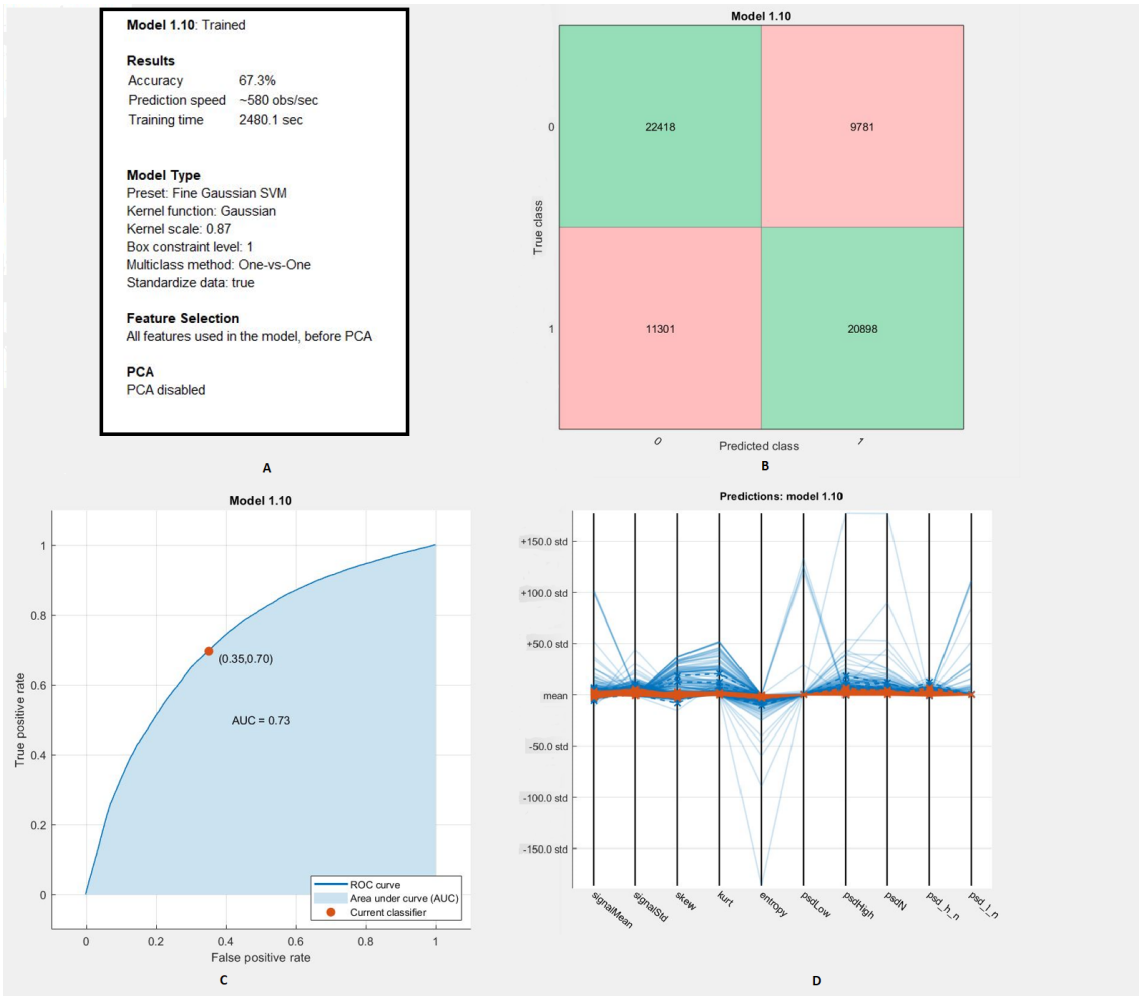


Figure 4.2: SVM model results A:Model Parameters B: Confusion Matrix Plot C:ROC Curve D: Parallel Coordinates

The parallel coordinates plot helps us compare the significance of the features in the classification process, as the data is normalized. With orange are depicted OSA labeled segments and with blue are labeled healthy segments. As it is evident, the healthy segments tend to have bigger values for almost all of the selected features, with the exception of the PSD total and the ratio of the PSD high to the total. This gives a good indication that the selected features are appropriate for the final classification. However, we could consider reducing the feature space by two.

One of the possible reasons for the SVM model to outperform the other ones is that generally, the SVM models perform well in binary classification problems, such as the present one. Additionally, all of the fea-

tures are real calculated values, meaning that no categorical features were used for the classification. One of the most important advantages that the SVM models have is that because their decision boundaries are simple, there is no problem with overfitting the model. This is vital as this is a medical application problem and we would like to perform the same regardless of the patient that the recordings come from.
What we can see from the ROC curve, the AUC is calculated at 0.73 which means that it performs better than a random classifier which would have an AUC of 0.5. As this classification refers to a medical classification problem, we also have to consider this result and not focus only on the accuracy obtained. The reason for that is that the ROC curve depicts the True Positive Rate (TPR) to the False Positive Rate (FPR), two very important parameters when assessing a medical application. The TPR, also referred to as sensitivity, is given by TP/TP+FN and expresses the probability that an actual positive will test positive. Respectively, the FPR, is given by FP/FP+TN and expresses the probability that an actual negative will test positive. We could fine-tune the model in a way that minimizes as much as possible the false negative results. This would mean that we aim for a higher true positive rate (TPR), which would have as a trade-off a higher False Positive Rate (FPR).
Below, we present the results of the SVM model more analytically. Additionally, we present the calculated accuracy for each different percentage for segmenting the label and the number of observations detected. Out of the total observations, half of them correspond to normal recordings, and the other half to OSA events. As we can see in the figure below, the greatest accuracy results are calculated for the largest number of observations. Additionally, we can see that the accuracy results present an almost linear relationship to the size of the dataset.

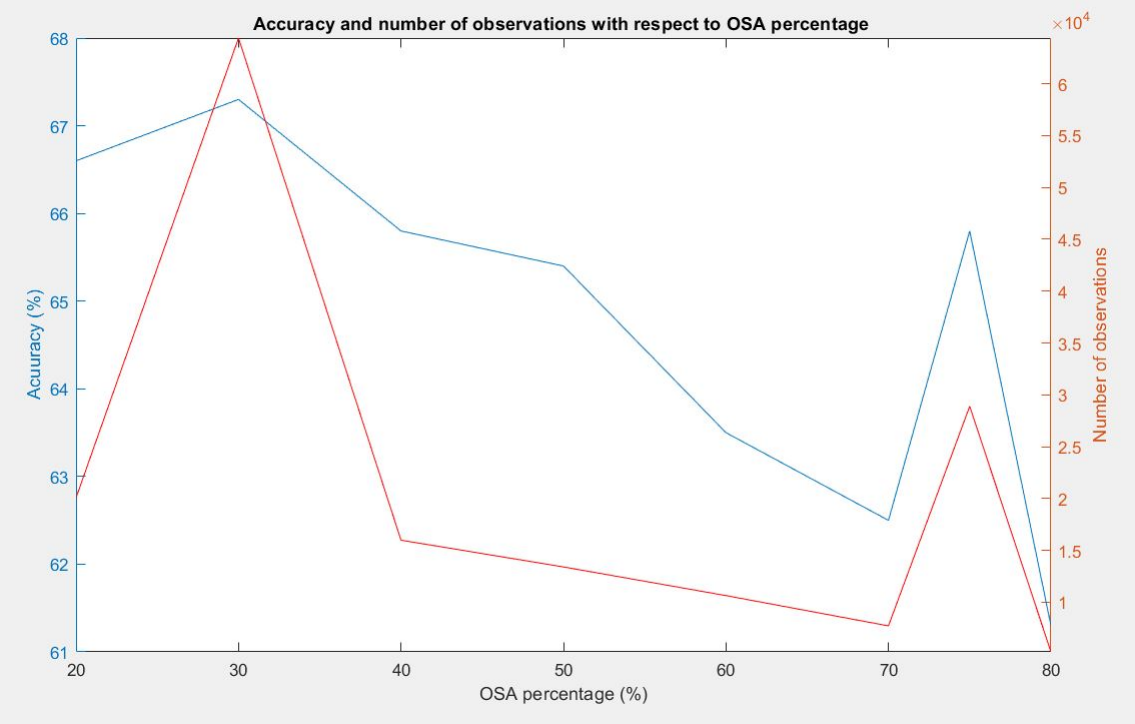


Figure 4.3: Accuracy and number of observations with respect to OSA percentage

4.3 Dataset 4

The fourth dataset was used for assessing the HR calculation as well as the RR calculation, as it was the only one out of the 4 datasets which had available data on the ECG, the airflow of the patient, the OSA labelings and the apnea events classification. The only disadvantage of this dataset is that it is significantly smaller compared to the other 3 ones.

4.3.1 HR calculation

For the HR calculation (based on the process described in section 3.2.2), the results were a mean difference of 0.43 sec in the duration of the detected ΔJ segments compared to the ΔR segments.
Below follows the histogram of the mean HR calculated for each signal in the dataset.

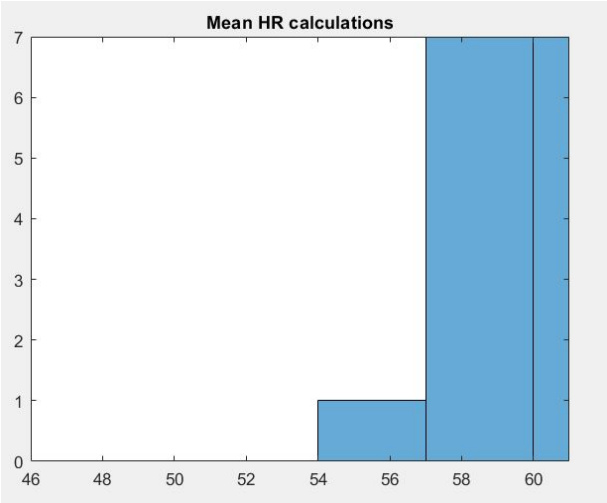


Figure 4.4: Dataset 3 Mean HR calculations

4.3.2 RR calculation

For the RR calculation, when assessed according to 3.11, the final score had a mean difference of an additional 0.06 detected respirations in the BCG signal compared to the airflow signal of the patient. This would correspond to a maximum of one extra breath detected to a segment, which is an adequate performance.
The calculated respiration rate of the signal can also be observed on the following histogram. Given that for a healthy adult the normal respiration rate is 12-20 bpm, the results are satisfactory.

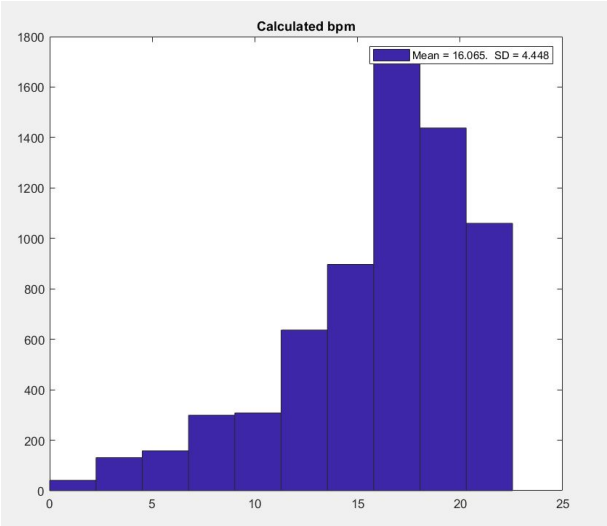


Figure 4.5: Histogram of Respiration Rate

4.3.3 OSA Classification Results

The calculated accuracy results of the 23 machine learning models are presented in the following table. Since we had already established that the label that leads to the highest accuracy results is the one that contains at least 30% OSA event, we only tested the algorithm for that percentage. Using this percentage, a total of 1996 samples was used to train the models, out of which 998 were normal segments and 998 were OSA-labeled segments.

Model	label 30%
Fine Tree	75.6%
Medium Tree	77.4%
Coarse Tree	79.6%
Linear Discriminant	74.8%
Quadratic Discriminant	65.9%
Logistic Regression	74.8%
Linear SVM	76.9
Quadratic SVM	79.3
Cubic SVM	78.7
Fine Gaussian SVM	75.4%
Medium Gaussian SVM	80.5%
Coarse Gaussian SVM	78.7
Fine KNN	73.4
Medium KNN	78.7
Coarse KNN	79.1
Cosine KNN	77.7
Cubic KNN	78.3
Weighted KNN	79.4%
Boosted Trees	74.0%
Bagged Trees	79.6%
Subspace Discriminant	74.0%
Subspace KNN	62.5%
RUSBoosted Trees	77.8%

It is clear that Medium Gaussian SVM performed best compared to the

rest of the models, with a training time of 12.78 sec and accuracy that reached 80.5%.
The Fine Gaussian SVM and the Medium Gaussian SVM are both kernel-based Gaussian SVM Models. Their difference is the size of the kernel. In the case of Fine Gaussian SVM, the kernel size is set at $\sqrt{P}/4$, while for the Medium Gaussian, the kernel size is set at $\sqrt{P}$, where P is the number of predictors. Both of these models have the disadvantage of having very hard interpretability, however, they are both much more flexible compared to the Linear SVM model.

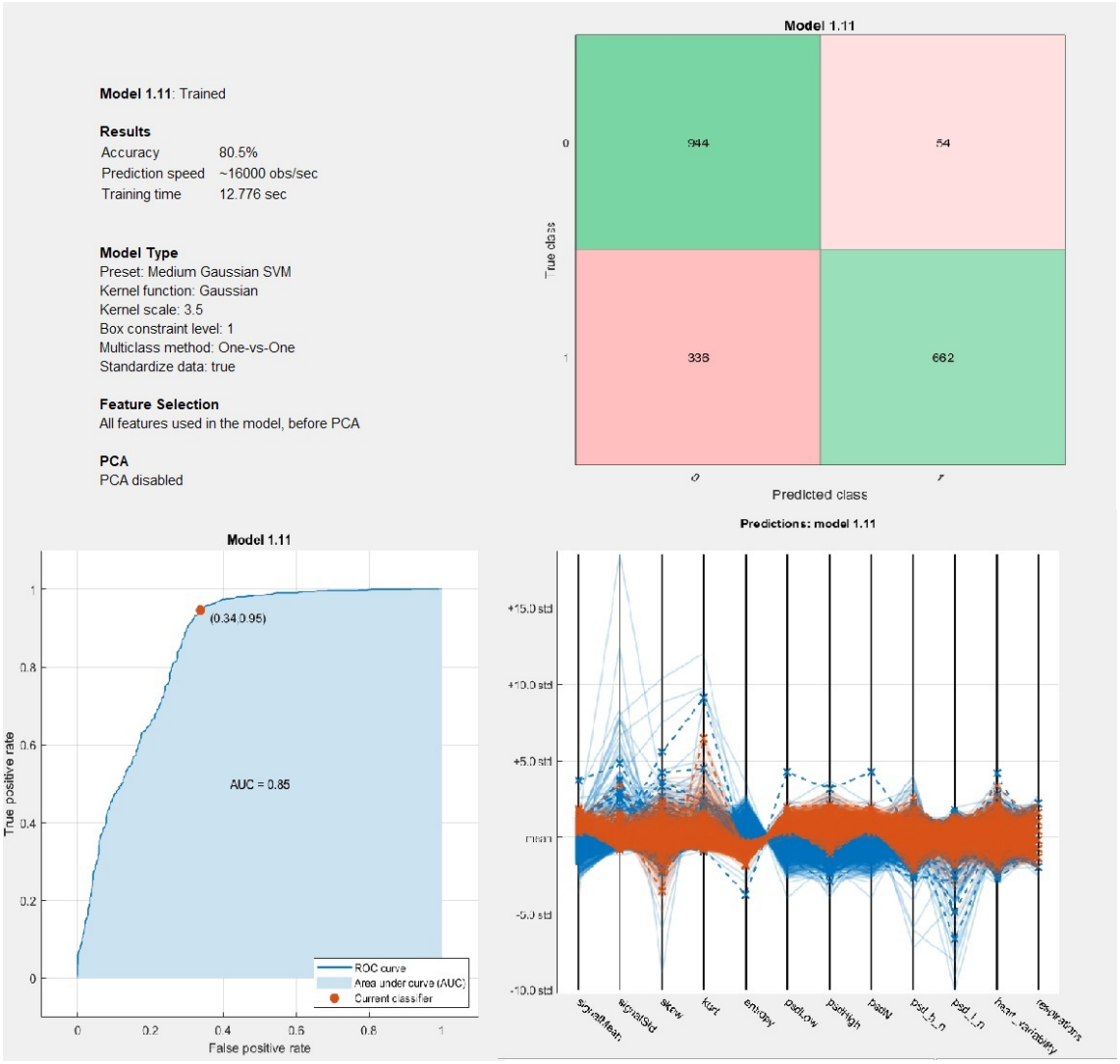


Figure 4.6: Medium Gaussian SVM model results A:Model Parameters B: Confusion Matrix Plot C:ROC Curve D: Parallel Coordinates

In this case, we can see that the AUC has a value of 0.85 which is better than the 0.73 that we got after training the models with dataset 3. We could again choose to fine-tune the model in a way that minimizes as much as possible the false negative results because it refers to a medical application. We can also observe that in this case, the curve is steeper which is generally preferred as it is better to have very low FPR for a reasonably high TPR.

4.4 Conclusions

It is evident that the results for the FORTH dataset far surpass the results of dataset 3. This was expected as the FORTH dataset also comes with information from the PSG examination of the patient and we were able to access the accuracy of the calculated features, while for dataset 3 we could only qualitatively assess them using well-established medical standards. Unfortunately, the FORTH dataset is a very small dataset and ideally, the methodology should be further tested using more subjects.
Generally, the acquired results for the 4th dataset are satisfying but could be further improved. However, the options for improving the results with such a small dataset are very limited. Additionally, one of the major challenges is the fact that there are many types of apnea that can be manifested in a couple of different ways and can interfere with the BCG recording. Apnea, which is a period of at least 10 seconds during which there is a complete or near complete pause in breathing. Hypopnea, which is a decrease in airflow lasting at least 10 seconds as well and respiratory effort related arousal (RERA), which is a limitation in breathing that results in increased respiratory effort and culminates in an arousal [24]. Additionally, a patient can be suffering from Mixed Sleep Apnea, then showcase both OSA symptoms as well as Central Sleep Apnea Symptoms. The main difference of the second one is that is caused by a central nervous system disorder where the brain fails to trigger the signal to inhale and exhale [25]. Both cases can influence the results and in future work, these cases could be further explored and for example, instead of performing a binary classification, we could classify apneas in more detail.

4.5 Proposed Future work

As in every case of machine learning applications, one of the most important aspects is the size of the dataset. Further testing of the methodology would be of great importance, ideally on a set with a structure similar to the FORTH dataset, meaning having both the BCG as well as PSG recordings. Furthermore, as there are many different types of apneas, the dataset could provide further information on the type of apnea and the classification could extend to multiple classes instead of the binary case that this thesis studies. Additionally, by having a larger dataset, the use of neural networks could be explored in identifying apneic episodes. Moreover, body measurements of the patients could also be very useful and could be recorded, as the BMI of a person has been strongly linked to them suffering from obstructive apneas [26].

For the current state of the work, as this is intended to be a medical purpose application, it is important to further reduce the false negatives. By observing the confusion matrix of the trained models 4.3 and 4.6, we can see that the false negatives are quite high, compared to the other classes. Data augmentation and the use of probabilistic classifications in combination with multiple thresholds for multiple different actions may produce a better outcome.

Another point that could be improved is the segmentation of the signal. As the apnea events have a very short duration compared to an overnight recording, the produced segments rarely are containing solely apnea or normal recording. By using a more targeted segmentation method, for example by first identifying possible points of an OSA event and then segmenting the signal, a better result could be achieved.

Finally, one of the major problems of BCG applications is the lack of standardization in the recording settings, which is very unfortunate considering that the BCG recording is directly affected by the placement of the sensor. Further studying should be conducted on best practices when acquiring a BCG signal. This would also minimize the need for different preprocessing methods, as the registered data would be more uniform.

Abbreviations and acronyms

The next list describes several abbreviations and acronyms that were used within the body of the document

AI Artificial Intelligence

BCG Ballistocardiography

HR Heart Rate

HRV Heart Rate Variability

OSA Obstructive Sleep Apnea

PSG Polysomnography

RR Respiration Rate

Bibliography

[1] E. Pino, J. Chávez, and P. Aqueveque, "Bcg algorithm for unobtrusive heart rate monitoring," 11 2017, pp. 180–183.

[2] X. Han, X. Wu, J. Wang, H. Li, K. Cao, H. Cao, K. Zhong, and X. Yang, "The latest progress and development trend in the research of ballistocardiography (bcg) and seismocardiogram (scg) in the field of health care," *Applied Sciences*, vol. 11, no. 19, 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/19/8896

[3] B. L. Koley and D. Dey, "Real-time adaptive apnea and hypopnea event detection methodology for portable sleep apnea monitoring devices," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 12, pp. 3354–3363, 2013.

[4] M. K. Moridani, M. Heydar, and S. S. Jabbari Behnam, "A reliable algorithm based on combination of emg, ecg and eeg signals for sleep apnea detection : (a reliable algorithm for sleep apnea detection)," *2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI)*, pp. 256–262, 2019.

[5] Tietjens JR, Claman D, Kezirian EJ, De Marco T, Mirzayan A, Sadroonri B, Goldberg AN, Long C, Gerstenfeld EP, Yeghiazarians Y., "Obstructive sleep apnea in cardiovascular disease: A review of the literature and proposed multidisciplinary clinical management strategy." *J Am Heart Assoc*, 1 2019.

[6] Young T, Peppard PE, Gottlieb DJ., "Epidemiology of obstructive sleep apnea: a population health perspective." *Am J Respir Crit Care Med.*, 5 2002.

[7] I. Starr and F. ClarkWood. (1939, 7) Studies on the estimation of cardiac output in man, and of abnormalities in cardiac function, from the heart's recoil and the blood's impacts; the ballistocardiogram.

[8] I. Sadek, J. Biswas, and B. Abdulrazak, "Ballistocardiogram signal processing: a review," *Health Information Science and Systems*, vol. 7, 05 2019.

[9] R. Lazazzera, M. Deviaene, C. Varon, B. Buyse, D. Testelmans, P. Laguna, E. Gil, and G. Carrault, "Detection and classification of sleep apnea and hypopnea using ppg and spo<inline-formula><tex-math notation="latex">$_2$</tex-math></inline-formula> signals," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 5, pp. 1496–1506, 2021.

[10] H. Yoon, J. H. Choi, and H. Jae Baek, "Apneic event estimation only using spo2 dynamics in sleep apnea patients," pp. 5335–5338, 2020.

[11] T. Van Steenkiste, W. Groenendaal, D. Deschrijver, and T. Dhaene, "Automated sleep apnea detection in raw respiratory signals using long short-term memory neural networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 6, pp. 2354–2364, 2019.

[12] J. M. Kortelainen and J. Virkkala, "Fft averaging of multichannel bcg signals from bed mattress sensor to improve estimation of heart beat interval," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2007, pp. 6685–6688.

[13] S. He, H. R. Dajani, R. D. Meade, G. P. Kenny, and M. Bolic, "Continuous tracking of changes in systolic blood pressure using bcg and ecg," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 6826–6829.

[14] M. Zaid, S. Ahmad, A. Suliman, M. Camazine, I. Weber, J. Sheppard, M. Popescu, J. Keller, L. Despins, M. Skubic, and

G. Guidoboni, "Noninvasive cardiovascular monitoring based on electrocardiography and ballistocardiography: a feasibility study on patients in the surgical intensive care unit," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, 2021, pp. 951–954.

[15] K. Tavakolian, B. Kaminska, A. Vaseghi, and H. Kennedy-Symonds, "Respiration analysis of the sternal ballistocardiograph signal," in *2008 Computers in Cardiology*, 2008, pp. 401–404.

[16] F. Liu, X. Zhou, Z. Wang, T. Wang, H. Ni, and J. Yang, "Identifying obstructive sleep apnea by exploiting fine-grained bcg features based on event phase segmentation," in *2016 IEEE 16th International Conference on Bioinformatics and Bioengineering (BIBE)*, 2016, pp. 293–300.

[17] Z. Wang, X. Zhou, W. Zhao, F. Liu, H. Ni, and Z. Yu, "Assessing the severity of sleep apnea syndrome based on ballistocardiogram," *PLOS ONE*, vol. 12, no. 4, pp. 1–24, 04 2017. [Online]. Available: https://doi.org/10.1371/journal.pone.0175351

[18] M. C. Ben Nasr, S. Ben Jebara, S. Otis, B. Abdulrazak, and N. Mezghani, "A spectral-based approach for BCG signal content classification," *Sensors (Basel)*, vol. 21, no. 3, p. 1020, Feb. 2021.

[19] C. Zwillich, T. Devlin, D. White, N. Douglas, J. Weil, and R. Martin, "Bradycardia during sleep apnea. characteristics and mechanism," *J. Clin. Invest.*, vol. 69, no. 6, pp. 1286–1292, Jun. 1982.

[20] C. Carlson, V.-R. Turpin, A. Suliman, C. Ade, S. Warren, and D. E. Thompson, "Bed-based ballistocardiography dataset," 2020. [Online]. Available: https://dx.doi.org/10.21227/77hc-py84

[21] Z. Wang, "Assessing the severity of sleep apnea syndrome based on ballistocardiogram (bcg)," Apr 2017. [Online]. Available: https://figshare.com/articles/dataset/Assessing_the_Severity_of_Sleep_Apnea_Syndrome_Based_on_Ballistocardiogram_BCG_/4868102/3

[22] M. Pediaditis, C. Farmaki, S. Schiza, N. Tzanakis, E. Galanakis, and V. Sakkalis, "Contactless respiratory rate estimation from video in a real-life clinical environment using eulerian magnification and 3d cnns," in *2022 IEEE International Conference on Imaging Systems and Techniques (IST)*, 2022, pp. 1–6.

[23] M.-C. Huang, W. Xu, J. Liu, L. Samy, A. Vajid, N. Alshurafa, and M. Sarrafzadeh, "Inconspicuous on-bed respiratory rate monitoring," 05 2013.

[24] V. Tsara, A. Amfilochiou, M. J. Papagrigorakis, D. Georgopoulos, and E. Liolios, "Guidelines for diagnosis and treatment of sleep-related breathing disorders in adults and children. definition and classification of sleep related breathing disorders in adults: different types and indications for sleep studies (part 1)," *Hippokratia*, vol. 13, no. 3, pp. 187–191, Jul. 2009.

[25] D. J. Eckert, A. S. Jordan, P. Merchia, and A. Malhotra, "Central sleep apnea: Pathophysiology and treatment," *Chest*, vol. 131, no. 2, pp. 595–607, Feb. 2007.

[26] S. Jehan, F. Zizi, S. R. Pandi-Perumal, S. Wall, E. Auguste, A. K. Myers, G. Jean-Louis, and S. I. McFarlane, "Obstructive sleep apnea and obesity: Implications for public health," *Sleep Med. Disord.*, vol. 1, no. 4, Dec. 2017.