



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

DESIGN AND IMPLEMENTATION OF AN ALGORITHM FOR THE
SCREENING OF OBSTRUCTIVE SLEEP APNEA IN CHILDREN
UNDER 15 YEARS OLD

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERA CIVIL
INDUSTRIAL

LILÍ ERAZO GAJARDO

PROFESOR GUÍA:
SEBASTIÁN A. RÍOS PÉREZ

MIEMBROS DE LA COMISIÓN:
CLAUDIO I. ESTÉVEZ MONTERO
IGNACIO CALISTO LEIVA

SANTIAGO DE CHILE
MARZO 2014

A mi incondicional hermana

Agradecimientos

Quisiera empezar agradeciendo a mi familia, a mis hermanos por escuchar mis reclamos y celebrar mis victorias, y a mis padres por siempre bendecirme con sus consejos llenos de amor. Y por supuesto a mi familia extendida, a Verónica, Joaquín y José.

A los amigos que tengo la suerte de tener: a Maca, quién siempre tuvo tiempo para mí y para quién siempre tendré tiempo; y a Letsan, por acompañarme desde el principio.

Estos, fueron sin duda seis años muy importantes en mi vida y agradezco haber conocido aquí a quienes serán los amigos que me acompañarán en la vida. A riesgo de dejar a alguien en el tintero: Elisa, agradezco que hayas sido la primera persona que conocí en la universidad y atesoro cada almuerzo contigo, cada tarde de estudio, cada conversación. Topo, Cami, Gustavo, Juan, Pablo, Jacky, Fano, Dani, MI, Vale, Coni y a todos los que olvido mencionar, muchas gracias por aparecer y permanecer en mi vida.

Por último, agradezco a mi profesor guía Sebastián por su ayuda y por su preocupación constante.

Resumen

Los Trastornos Respiratorios del Sueño (TRS) son un grupo de enfermedades que afectan la función respiratoria durante la noche, desde el ronquido primario a la apnea del sueño –la más grave de ellas–.

Los niños que padecen apnea del sueño pueden desarrollar desde trastornos de crecimiento hasta deficiencias cognitivas de largo plazo. Sin embargo, una vez que han sido diagnosticados, el tratamiento es efectivo en la mayoría de los casos, mejorando su calidad de vida y evitando consecuencias en su desarrollo cognitivo.

El *gold-standard* (mejor prueba clínica) para el diagnóstico de TRS es la polisomnografía, que consiste en la recolección de señales biomédicas durante el sueño. El problema es que este examen es invasivo, costoso y difícil de practicar en niños pequeños. Por lo tanto existe la necesidad de una forma de prediagnosticar TRS, específicamente apnea para aumentar el número de niños diagnosticados, y, como resultado, el número de niños tratados.

Para esto, el objetivo general planteado es prediagnosticar apnea obstructiva del sueño a partir de señales recogidas por polisomnografía en niños menores de 15 años usando técnicas de minería de datos.

El resultado esperado de este trabajo es un algoritmo capaz de clasificar infantes, usando menos información que la polisomnografía, en dos grupos: población en riesgo de padecer apnea y población fuera de riesgo (o muy bajo riesgo).

El trabajo está dividido en dos etapas, la primera de ellas son modelos basados en una señal, la segunda consta de modelos basados en más de una señal. Los resultados de la primera etapa muestran modelos de buena calidad aunque sólo están basados en una señal; algunos con sensibilidad y especificidad por sobre el 85% y podrían sentar las bases de un modelo válido de prediagnóstico. En la segunda etapa se identificaron, mediante técnicas de reducción de información, las señales que tienen mayor poder predictivo para realizar el prediagnóstico, los modelos basados en estas señales alcanzaron hasta el 100% de precisión.

Summary

Sleep Disordered Breathing (SDB) are a group of diseases that affect the normal respiratory function during sleep, from primary snoring to obstructive sleep apnea (OSA) –the most severe of the diseases–.

Children affected by OSA may develop growing disorders and even long-term cognitive disadvantages. However, once they have been diagnosed, treatment is effective in most of the cases improving their quality of life and avoiding consequences in their cognitive development.

The gold-standard (best available clinical test) for the diagnosis of SDB is the polysomnography, which consists in collecting biomedical signals during sleep. The problem is that this test is invasive, expensive and hard to perform in young children. Therefore it is necessary to design a way to screen SDB, specifically apnea, in order to increase the number of diagnosed children, and, as a result the number of treated children.

To accomplish this purpose, the proposed general objective is to screen obstructive sleep apnea from signals collected by polysomnography in children under 15 years old using data mining techniques.

The expected result of this thesis is an algorithm able to classify children, using less information than a polysomnography, into two groups: the group in risk of having a SDB, and the group with no risk (or very little risk) of having a SDB.

This work is divided in two phases, first of them are models based on one signal, and second of them are models based on more than one signal. Results of the first phase show a model of high levels of quality based on just one signal; some of them show sensitivity or specificity over 85% and may set the base to a valid screening method. In the second phase, using information reduction techniques, signals with best predictive power for screening were identified, models based on this signals reached 100% of precision.

Contents

Agradecimientos	ii
Resumen	iii
Summary	iv
1 Introduction	1
1.1 Problem Understanding	1
1.1.1 Terminology	2
1.2 Rationale	2
1.3 Work Description	4
1.3.1 General Objectives	4
1.3.2 Specific Objectives	4
1.3.3 Expected Results	5
1.4 Methodology	5
1.5 Thesis Structure	9
2 Gold-Standard in OSA Detection	10
2.1 Preprocessing and Signal Classification	10
2.2 Data Understanding	12
3 Classic Models Construction	14
3.1 Data Preprocessing	14
3.2 Classification Models	16
3.2.1 Support Vector Machines	16
3.2.2 Neural Networks	17
3.2.3 Logit Model	17
3.3 Preliminary Results	17

4	Metamodel Construction	19
4.1	Optimization Problem	20
4.2	Implementation	21
5	Minimal Signal Selection	23
5.1	Data Preprocessing	23
5.2	Feature Extraction	24
5.2.1	ECG Feature Extraction	24
5.2.2	EEG Feature Extraction	25
5.2.3	Abdominal and Thoracic Effort Feature Extraction	28
5.2.4	Air Flow Feature Extraction	30
5.2.5	Leg Movement Feature Extraction	30
5.2.6	Body Position Feature Extraction	32
5.2.7	Electromyogram Feature Extraction	33
5.2.8	Electro-oculogram Feature Extraction	33
5.2.9	Pulse Feature Extraction	34
5.2.10	Snore Feature Extraction	35
5.2.11	Oxygen Saturation Feature Extraction	36
5.3	Feature Selection	37
5.4	Classification Models	39
5.5	Preliminary Results	40
6	Results and Discussion	42
6.1	Classic Models Stage Results	42
6.1.1	Support Vector Machine	42
6.1.2	Neural Network	43
6.1.3	Logit	45
6.2	Metamodel Stage Results	46
6.3	Discussion	46
6.3.1	Recommendations and Further Work	48
	References	49

Appendix	53
A Classic Models Stage: Results for Support Vector Machine applied to Air Flow signal	53
B Classic Models Stage: Results for Neural Network applied to Air Flow signal	54
C Classic Models Stage: Results for Logit Model applied to Air Flow signal . .	55
D Classic Models Stage: Results for Support Vector Machine applied to Electrocardiography signal	56
E Classic Models Stage: Results for Support Vector Machine applied to Electrocardiography signal	57
F Classic Models Stage: Results for Neural Network applied to Electrocardiography signal	58
G Classic Models Stage: Results for Neural Network applied to Electrocardiography signal	59
H Classic Models Stage: Results for Support Vector Machine applied to Oximetry signal	60
I Classic Models Stage: Results for Neural Network applied to Oximetry signal	61
J Classic Models Stage: Results for Logit Model applied to Oximetry signal . .	62

List of Tables

1	Previous Intents of Automated OSA Screening Methods in Adult Patients	11
2	Classification Performance of Support Vector Machine, Neural Network and Logit Models for Different Wavelet Decomposition Functions and for ECG, Airflow and Oxygen Saturation	17
3	Models Selected for Metamodel Construction for Each signal	19
4	Summary of features resulting from Feature Extraction Process divided by signal for a complete PSG	38
5	Summary of features resulting from Feature Selection Process divided by signal	40
6	Classification Performance of Support Vector Machine using a feature subset from PSG as input	40
7	Classification Performance of Neural Network using a feature subset from PSG as input	41
8	Classification Performance of Logit Regression using a feature subset from PSG as input	41
9	Summary of performance indicators for Support Vector Machine applied to Air Flow signal	42
10	Summary of performance indicators for Support Vector Machine applied to ECG signal	43
11	Summary of performance indicators for Support Vector Machine applied to Oximetry signal	43
12	Summary of performance indicators for Neural Network applied to Air Flow signal	43
13	Summary of performance indicators for Neural Network applied to ECG signal	44
14	Summary of performance indicators for Neural Network applied to Oximetry signal	44

15	Summary of performance indicators for Logit Model applied to Air Flow signal	45
16	Summary of performance indicators for Logit Model applied to ECG signal .	45
17	Summary of performance indicators for Logit Model applied to Oximetry signal	46
18	Classification Performance of Metamodel	46

List of Figures

1	Cycle of phases for Cross Industry Standard Process for Data Mining	6
2	Cross Industry Standard Process Phases	7
3	Project Plan phases	7
4	Classic Models Stage Design	14
5	Minimal Signal Selection Stage Design	23
6	ECG Sample	24
7	ECG Wave Characteristics	24
8	Electrodes position for an EEG	26
9	EEG Sample corresponding to patient A0001945	26
10	Spectrogram Sample corresponding to patient A0001945	27
11	Processed EEG signal corresponding to patient A0001945	28
12	Abdominal Thoracic Signal (blue) and Abdominal Effort Signal (red) corresponding to patient A0000678	29
13	Different Wavelet Decompositions (blue) for the Thoracic Effort Signal (red) corresponding to patient A0000678	29
14	Air Flow Signal corresponding to patient A0000739	30
15	(a)Leg1 Signal corresponding to patient A0001484. (b)Leg2 Signal corresponding to patient A0001484. (c)Legs Signal corresponding to patient A0011947.	31
16	Prone and Supine positions	32
17	(a)Body Position Signal corresponding to patient A0001460. (b)Auxiliary signal derived from Body Position Signal for patient A0001460	32
18	EMG Chin Signal corresponding to patient A0002197	33
19	EOG Sinals of Left (blue) and Right (Red) ayes corresponding to patient A0006225	34

20	(a)Pulse signal corresponding to patient A0000294. (b)Filtered pulse signal. (c)Absolute Pulse Rate Increase (APRI) derived from pulse signal. (d)Absolute Pulse Rate Decrease (APRD) derived from pulse signal.	35
21	Snore Signal corresponding to patient A0000884	36
22	Oxygen Saturation Signal Signal corresponding to patient A0001183	37

Chapter 1

Introduction

1.1 Problem Understanding

Sleep Disordered Breathing (SDB) are a group of diseases that affect the normal respiration process during the night. These can affect people at any age because of different causes: in newborns and young children it is related to congenital defects or premature birth; in older children and adults it may be related to obesity, morphological causes, and hypertension [14, 15, 19, 23] .

The diseases considered as SDB are: primary snoring, upper airway resistance, and obstructive sleep apnea (OSA) -from less to most severe-. These diseases have degrees of severity. For example, a newborn with OSA can experience episodes of apnea (this is the absence of airflow for ten seconds or more) during night and have a relatively normal life, while a severe OSA may lead to sudden death [13].

Symptoms of SDB are abnormal day sleepiness, sudden naps during day (for example, in a red light while driving), general tiredness, fatigue, trouble sleeping, and some other related diseases [9].

The diagnosis of SDB is accomplished through a clinical study called polysomnography, which collects around twenty different biomedical signals during sleep, including: electrocardiography, electroencephalography, electromiography, plethysmography, oronasal airflow, chest movement, abdominal movement, legs movement, among others.

This work is based on data collected by Pablo E. Brockmann M.D. and his research team. This dataset consists in 78 whole night polysomnographies from patients under 15 years old. This is the largest dataset of these characteristics used for OSA screening.

1.1.1 Terminology

- **Apnea:** 10 or more seconds of air flow complete absence.
- **Hypopnea:** Decrease of air flow of 50% or more, but not the absence of it.
- **OSA:** Initials of Obstructive Sleep Apnea, a respiratory disease that affect normal respiratory function during sleep.
- **SDB:**Initials of Sleep Disordered Breathing, a group of diseases associated to abnormal respiration during sleep.
- **PSG:** (Polysomnography) Clinical test used to detect and diagnose SDB's. Consists in over 20 biomedical signals collected during sleep.
- **ECG:** (Electrocardiogram) Clinical test used to register heart activity based on electrical impulse registration.
- **EEG:** (Electroencephalogram) Clinical test used to register brain activity based on electrical impulse registration.
- **QRS complex:** ECG signals can be characterized by 5 points: P, Q, R, S and T. each of them refer a particular point of the ECG wave. The QRS complex is the area formed by points Q, R and S.
- **HRV:** Initials of Heart Rate Variability, it is a signal derived from ECG, constructed from R-R intervals.
- **EDR:** Initials of ECG Derived Respiratory Signal, it is a signal derived from ECG, constructed from QRS amplitud, this is the vertical distance between Q and R.

1.2 Rationale

SDB are suffered by 15% of the pediatric population in the world, this means children under 15 years old. This percentage may be even bigger in Latin American children, mostly because of morphological reasons.

In Chile the situation is not any better. OSA affects 4% of adult population and the prevalence increases with age. It is most common in males (2% to 4% in males, and 1% to 2% in females), and the risk population reaches 7% in adult males and 4% in adult females [17].

A study carried in Argentina showed prevalence in children is estimated in 2% to 3%, but it may increase to 13% to 66% in obese children and teenagers [5].

These are the only two studies conducted in Latin America mainly because this clinical study requires the patient to spend the night in the hospital, connected to a variety of sensors, in order to register the sleeping patterns that will lead to the diagnosis. After the study is completed, the specialist must review records of the whole night and manually tag episodes of abnormal activity during sleep. This process may take two hours for average doctors, and at least forty minutes for experienced doctors, so conducting an study is hard and often less consuming methods are preferred over polysomnography based studies [2, 3].

This is why polysomnography is not always available, it is resource consuming, it is expensive, and it is hard to perform in little children. Also, this may not reflect a normal night of sleep -because it is performed under a controlled environment- and in some cases a follow up is performed [13–15].

Studies in Latin America in general use other methods, like questionnaires to predict who has or doesn't have the disease. The standard process is as follows: when a pediatric patient arrives at a clinical center, parents must answer a questionnaire about the child sleeping habits, also medical indicators like body weight, age, and height are considered. Based on the answers, the patient is scored and only if the score exceeds a predefined threshold, the child is put through polysomnography test to arrive to the final diagnosis, this means the only screening method used so far is questionnaires, no medical test is used for this matter before polysomnography [4].

Specifically in children, studies show that prevalence of SDB are associated with prevalence of other diseases and cognitive disorders. In Germany there are some attempts to use the pulse signal as a screening method, but on a very low scale [21].

This is why a method to assist the diagnosis of SDB, and reduce the amount of in-

formation needed to perform it, may contribute to the widespread of diagnostics and follow up of sleep quality in children and teenagers.

There is a need of a way to pre-diagnose SDB without depending on a polysomnography, lesser time consuming, lesser medical resources needed, and lesser expensive to the patients. This work intends to design and implement an algorithm, using less information than a full polysomnography and able to classify children's signals into two groups: OSA and no-OSA.

1.3 Work Description

1.3.1 General Objectives

The main objective of this thesis is to design and develop a set of algorithms able to screen OSA from signals collected by the polysomnography in children under 15 years old using data mining techniques for signal classification.

1.3.2 Specific Objectives

- Research the state-of-the-art algorithms for sleep apnea screening, arousals detection and detection of snoring episodes in pediatric population.
- Build a high quality repository of pediatric polysomnographic data needed to apply the developed methods and the process needed to collect new data.
- Research the minimum signals needed to perform a screening in children that suffer of OSA.
- Develop a benchmark library of existing methods for adult signal classification and improve them to screen children polysomnographic data.
- Design and develop a set of algorithms able to classify signals in order to perform apnea screening and evaluate its performance.

1.3.3 Expected Results

- A high quality repository of polysomnographic data tagged by qualified clinicians.
- Description of the process needed to collect new data in order to apply the developed screening methods.
- A benchmark library of existing methods for OSA screening in adults.
- A set of algorithms able to perform OSA screening.
- Performance evaluation of methods developed to perform OSA screening.

1.4 Methodology

The methodology selected for this investigation is CRoss Industry Standa Process for Data Mining (CRISP-DM). CRISP-DM provides a framework for data mining projects, aims to make large data mining projects less costly, more reliable, more repeatable, more manageable, and faster [25].

CRISP-DM is organized into a four-level breakdown: Phases, Generic Tasks, Specialized Tasks, and Process Instances. Phases are understood as a virtuous cycle for data mining projects: the work is not finished once the results are deployed, some lessons learned during the process can trigger new and deeper business questions [25].

The initial phase correspond to the **Business Understanding**, it focuses on understand the business perspective of the problem in order to design (to convert it into) a Data Mining problem. In the **Data Understanding** phase is expected to collect the initial data and get familiar with it, identify potential problems, and perhaps form preliminary hypothesis. These two initial phases are related because Data Understanding is needed to form initial hypothesis and have a complete Business Understanding of the Data Mining problem. As output of these two, is expected a preliminary project plan. Next phase is **Data Preparation**, it corresponds to any activity needed to complete the final repository from the initial raw data.

In the **Modeling** phase various modeling techniques are applied and calibrated. As

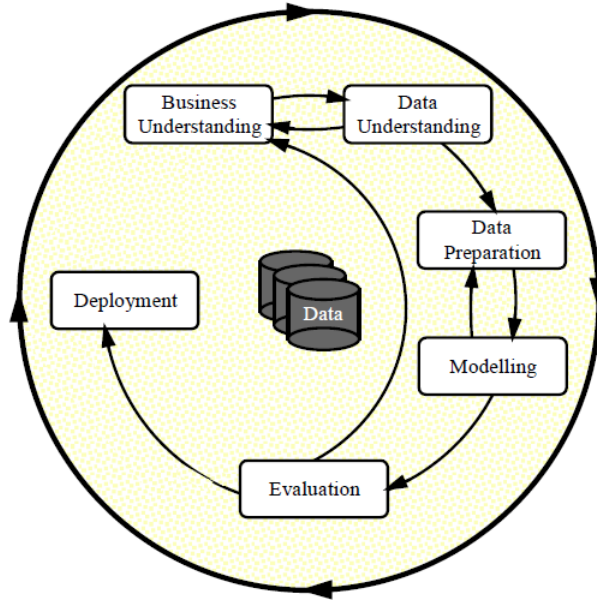


Figure 1: Cycle of phases for Cross Industry Standard Process for Data Mining

many models can solve the same problem in Data Mining, and some of them required specifically data formats, this phase is closely related to the Data Preparation phase.

After models have been constructed, in the **Evaluation** phase every objective is reviewed in order to detect problems in the process, or objectives that the model haven't achieve satisfactorily.

Finally, in the **Deployment** phase, the model is presented in a way that can solve the initial business problem, this can be as easy as generate a report, or more complex, like a repeatable data mining process. This phase considers that the user, not the analyst will be in charge of using the model, so the deployment has to be independent of the analyst [25].

In the following figure, each phase is shown with its corresponding tasks and expected outputs:

This generic phases may be translate into a custom project plan for every data mining project.

In the first stage, Business Understanding, a Project Plan is expected as an output.

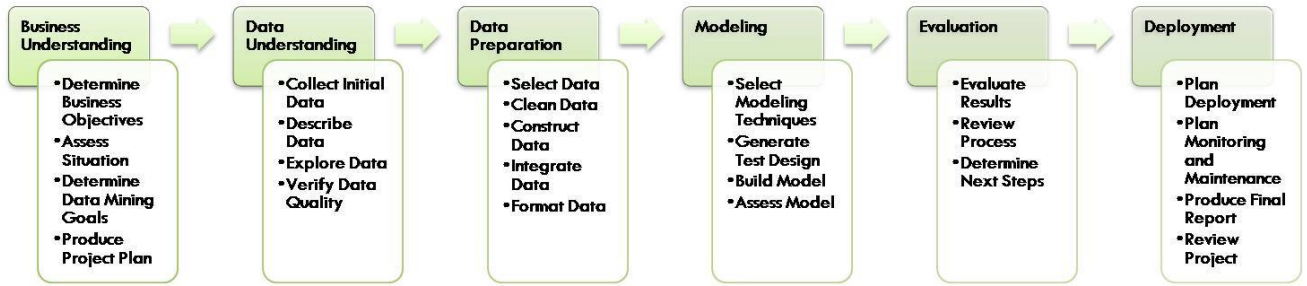


Figure 2: Cross Industry Standard Process Phases

Figure 3 is a diagram of the custom methodology design for this project and the following paragraphs is a detailed explanation of every stage.

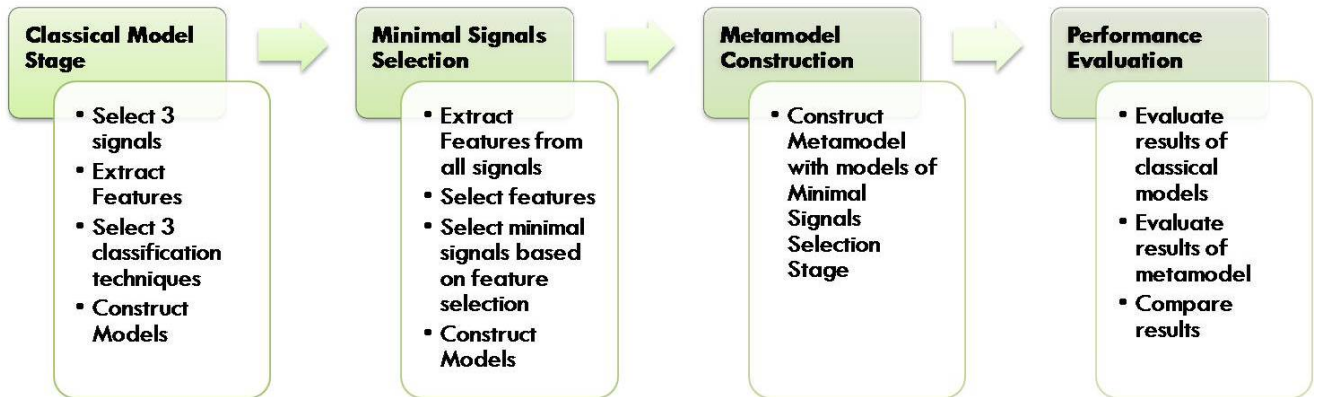


Figure 3: Project Plan phases

Classical Models Stage aims to construct basic models based on classification techniques for just one signal described in bibliography. Most of this models have only been tested in adults polysomnographic data. This is why this stage will set the minimum levels of sensitivity, specificity and accuracy expected for more complex models.

Signals and models selected for this stage are based on what has already been done, and showed satisfactory results. Signals selected are: ECG, Oximetry and Air Flow. Models selected are: Support Vector Machine (SVM), Neural Network (NN) and Logit.

Minimal Signals Selection Stage aims to find signals with best predictive power in order to construct a high performance model for screening.

In order to complete this task, every signal will be preprocessed alone to extract its

features. Then a filter-based approach will be used to select relevant features and used as input for the three models mentioned before. As a result of this stage, three multi-signal models will be constructed.

Metamodel Construction Stage will use as input the three models from previous stage, and will assign a score to each model credibility in order to reach the best performance possible with the feature set selected.

Last stage, Performance Evaluation, aims to evaluate results from stages one and three, and describe its strengths and weaknesses.

Indicators selected for this tasks are Sensitivity, Specificity and Accuracy (commonly used to assess clinical tests performance); and Precision, Recall and F-score (to assess the classification performance of algorithms). They will be calculated as follows:

$$\begin{aligned} \text{Recall} = \text{Sensitivity} &= \frac{TP}{TP + FN} \\ \text{Specificity} &= \frac{TN}{TN + FP} \\ \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN} \\ \text{Precision} &= \frac{TP}{TP + FP} \\ \text{F-score} &= \frac{2 * TP}{2 * TP + FN + FP} \end{aligned}$$

Where TP corresponds to the number of true positive results, this is if the algorithm classified a patient as sick (a positive case) and the patient was actually sick; on the other hand, FP corresponds to the number of false positive results (patients classified as sick who where actually healthy). TN correspond to the number of true negative results, this is patients classified as healthy (this is a negative case) by the algorithm who where actually healthy; and last, FN corresponds to the number of patients mistakenly classified as healthy.

Interpretation of this indicators is based on the ability to detect sick people, this is because it is more important to correctly classify sick children as sick than healthy children as healthy.

Sensitivity is the ability to detect sick people, Specificity is the ability to detect healthy people, Accuracy is the ability to correctly classify, that is, how many children were

correctly classified.

On the other hand, Precision is the proportion of the sick people group that are correctly classified. Finally, F is a measure of how accurate is the model, in a range of 0 to 1, where 1 is the best model possible.

First two measures are the standard evaluators for clinical tests results, and last three measures correspond to the standard evaluators for classification algorithms in general. In order to assess correctly the overall performance of the algorithm, all five measures were included in the Performance Evaluation stage.

1.5 Thesis Structure

This document present results and discussion of the work described above. It is structured as follows:

First, in Chapter 2 a bibliographic review of OSA screening methods is presented, also a sistematic review of preprocessing and signal classification approaches for this matter in included.

This chapter also includes a detailed description of the dataset.

Chapter 3 describes a more detailed step-by-step methodology for Classical Models Stage, some preliminary results are included.

Chapter 4 includes a detailed methodology for Minimal Signals Selection and Meta-model Construction Stages. This chapter ends with some preliminary results derived from each model, and the final metamodel.

In Chapter 5 all results are reviewed in detail. It also presents the conclusion of this work, its main results and further work derived from it.

At the end of this document, References and Appendix can be found.

Chapter 2

Gold-Standard in OSA Detection

2.1 Preprocessing and Signal Classification

Based on previous studies, It is known that some signals collected by PSG have enough predictive power to perform an screening [1, 10, 11].

One signal screening methods have been succesfully tested in adults and children [22,24], but those methods still require attending personnel and a whole night dedicated systems. Most of them, also required a medical evaluation afterwards.

Automated classification methods aim to avoid unnecessary resource consuming screening methods. To this subject, most important contribution was made by Computers in Cardiology Challenge of 2000, where the task was to automatically tag minute-by-minute a single ECG signal as OSA or no-OSA and get to a final diagnosis: OSA or no-OSA for every record [8, 16, 18, 20] .

Some of this methods reached a 100% of precision in the binary diagnosis and over 85% in minute-by-minute tagging.

Table 1 summarizes the results and different approaches of the investigations mentioned before. In the comments column can be seen that none of these studies have been tested in children, and most of them were trained and tested on databases specially designed for this task. Specially models tested on the Computers in Cardiology 2000 Database can not be compared, because the were developed using extremely clean data, which is not realistic.

Table 1: Previous Intents of Automated OSA Screening Methods in Adult Patients

Author, Year	Brief Description	Results	Comments
Álvarez-Estévez et al., 2009	Detection of apneic events using fuzzy reasoning and based on Air flow signal and Oxygen Saturation signal	Sens=87%, Spec=89%	Tested on a 12 patients database (Sleep Heart Health Study), it does not diagnose OSA, it only detects apneic events. All registers corresponding to adults.
De Chazal et al., 2003	Apnea diagnosis based on ECG signal. Features extracted using black-box methods. Classification methods: linear discriminant, quadratic discriminant.	Acc=100%	Tested on Computers in Cardiology 2000 database. 75 records tagged minute-by-minute. All registers corresponding to adults.
Driver et al., 2011	Validation of MediByte (portable monitor) as screening method.	Sens=80%, Spec=87%	This monitor requires a clinician to perform the diagnosis.
Jarvis and Mitra, 2000	Apnea diagnosis based on ECG signal. Features used derived from spectral analysis. Classification based on threshold criteria.	Acc=100%	Tested on Computers in Cardiology 2000 database. 75 records tagged minute-by-minute. All registers corresponding to adults.
Khandoker et al., 2009	Apnea binary diagnosis (OSA+, OSA-) based on ECG signal. Features extracted: Wavelet Transform coefficients. Classification method: Support Vector Machine.	Acc=100%	Tested on three combined databases: Sleep Research Unit Database, Physionet Apnea ECG Database and Saint Vincent's University Hospital/University College Dublin Sleep Apnea Database. All 125 registers corresponding to adults.
Mendez et al., 2007	Apnea binary diagnosis based on ECG signal. Features extracted: derived signals (EDR, HRV) coefficients. Classification method: K-Nearest Neighbor.	Acc=85,5%, Sens=83,9%, Spec=88,5%	Tested on Physionet Apnea ECG Database (50 registers) tagged as OSA+ or OSA-.

2.2 Data Understanding

Data used in this study was collected in Sleep Study Center of Clinical Hospital of Catholic University of Chile by Dr. Pablo E. Brockmann and his fellows.

All patients included in this study were referred to the Sleep Study Center mentioned in one of its three locations: *San Jorge*, *San Carlos* or *Centro del Sueño (Marcoleta)*.

All children included in this study came to the Sleep Center because they parents or doctors had concerns about the quality of their breathing during sleep. Patients parents has to answer an standard questionnaire in order to evaluate the severity of the potential disorders. Only children consider as risk population went under polysomnographic test.

Each patient spent a whole night in the hospital where 20 or 21 biomedical signals were taken (approximately 8 hours of records¹), including Electroencephalography (11 channels: EEG F3-A1, EEG C3-A1, EEG P3-A1, EEG O1-A1, EEG F4-A2, EEG C4-A2, EEG P4-A2, EEG O2-A2, EEG A1-A2, EEG Fp1-A1, EEG Fp2-A2), Electrocardiography (ECG or ECGI channel), Electroculography (EOG Right and EOG Left), Electromiography (EMG Chin), Air Flow (Patient Airflow channel) and Oxygen Saturation (SPO2 channel) among others².

After the test were taken each patient parents consent the use of the data collected during the polysomnographic test for later studies. Also, the Catholic Clinic Hospital Ethics Committee approve the use of this data.

PSG was collected in Sleep Study Center by Alice® 5 Diagnostic Sleep System from Respironics (Philips) and transformed using the corresponding software (Alice® Sleepware) to the generally accepted format for biomedical time series European Data Format (EDF).

Afterwards each test was manually scored by Dr. Pablo E. Brockmann or one of his fellows according to AASM (American Academy of Sleep Medicine) criteria [12]. Apnea, hypopnea, arousals, leg movement, and any other event of interest were detected.

All this steps lead to put together the following database: Data corresponds to 78

¹originally records were almost 9 hours long, but after data cleaning, removing first minutes when electrodes were placed and last minutes with no records, they were left in 8 hours approximately

²detailed information of each signal might be found in Chapter 4

complete night PSG from 78 different patients of ages from 2 to 16 years old (mean \pm SD: $9,7 \pm 3,6$ years). 64% of them are girls.

The commonly use criterion to diagnose OSA is based on the number of events register during a period of time, this index is known as Apnea Hypopnea Index. The mean index for this group was 2,6.

According to AASM criteria, children are diagnosed with OSA with an Apnea Hypopnea index equal or greater than 1 [12]. Using this specification, 43% of the PSG correspond to children with OSA.

No PSG was excluded because of other diseases presented in patients.

It is important to mention that most studies of OSA screening have small databases, no more than 20 registers. Even though some bigger databases are available, in children data, our database is the biggest documented (to the best of our knowledge).

Chapter 3

Classic Models Construction

This stage aims to construct a set of algorithms able to classify children data to perform OSA screening. However, these models are constructed from existing methods for adults. As said before, children and adults respiratory functions are not the same, so results from this stage should not be good enough to perform screening alone.

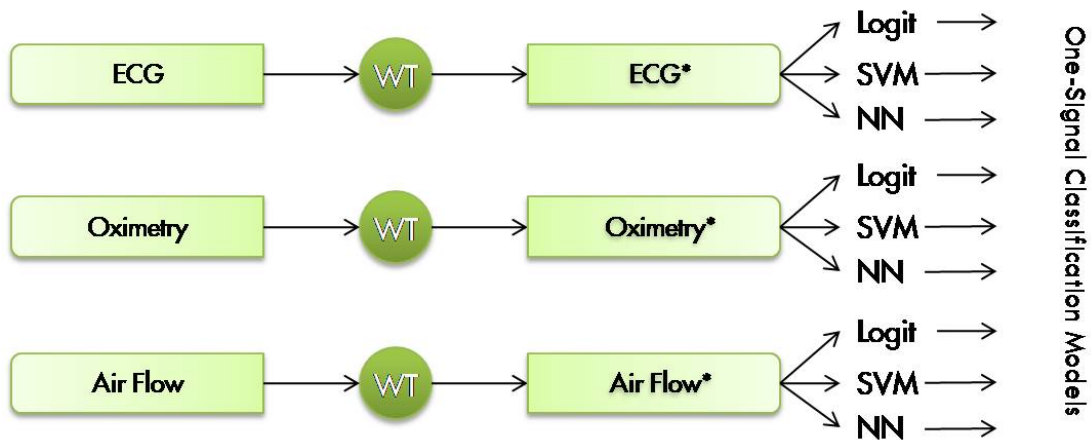


Figure 4: Classic Models Stage Design

3.1 Data Preprocessing

As expected, in different studies different features were used to characterize signals. As the objective of this study is not to replicate adult results in children, but to compare the classification performance, all algorithms had to be trained and tested with the same features.

To accomplish this, a blackbox feature extraction method was adopted. As algorithm

selection, the feature extraction method selection was based on most often used in references. According to these criteria, Wavelet Transform (WT) was adopted (in Figure 4 corresponds to the WT circles).

Wavelet Transform is used in 1D nonstationary signals to process time series as the ones collected by PSG. It allows to collect frequency information contained in the time series. Also, there is no requirement of the frequency stability, thus it is frequently used for ECG feature extraction.

Several wavelets were tested in order to extract features. Daubechies and Symlet of levels from 1 to 10 were tested, in addition to Haar function. This means that every signal was processed 21 times under each of these 21 functions of Wavelet Transform (Daubechies1, Daubechies2, Daubechies3, and so on).

To process the dataset, Wavelet Toolbox for Matlab was selected. With this objective, EDF's had to be transformed into a compatible Matlab format.

Each signal processed with Wavelet Transform was decomposed to its 14th detail level, obtaining 14 time series for each decomposition function, i.e. 14 time series for ECG processed with WT with Haar function, 14 time series for ECG processed with WT with Daubechies of level 1 function, and so on.

In order to reduce data and extract features, three features were calculated for each time serie, these were: mean, variance and energy; calculated as follows:

$$\begin{aligned} \text{Mean:} \quad \bar{x} &= \frac{1}{N} \sum_{i=1}^N x_i \\ \text{Variance:} \quad \sigma^2 &= \sum_{i=1}^N (\bar{x} - x_i)^2 \\ \text{Energy:} \quad E &= \sum_{i=1}^N x_i^2 \end{aligned}$$

Coefficients obtained of the preprocessing were named as $Db1_{mean1}$, $Db1_{var1}$, $Db1_{ene1}$ for the mean, variance and energy of first detail level for Daubechies1 WT function. The same process was applied for every WT function and to every signal. This way, 21 different databases were generated for every signal -one for each WT function-, each of them containing 42 features: $mean1$, $var1$, $ene1$, $mean2$, $var2$, $ene2$ up to $mean14$, $var14$, $ene14$.

3.2 Classification Models

In order to make comparable the three methods selected, all algorithms were trained and tested with the same feature datasets.

With all 78 records available, training datasets were constructed selecting 70% of records that were tagged as OSA (as described in the Data Understanding section, 43% of patients were tagged as OSA), and 70% of records tagged as no-OSA. the remaining 30% was used to test performance on every case. Although as said before, the database had enough OSA and no-OSA cases, this way ensures a balanced train and testset for the methods.

Doing this iteratively, thirty of this training and testing dataset were created and used as input for all classification methods. Metrics of performance for every method were calculated as the average performance of these 30 iterations.

All algorithms were tested in its classification performance and its predictive power using the same metrics: sensitivity, specificity, accuracy, precision and F score (An example of these metrics can be seen in Table 9) on page 42.

Classification methods used are briefly mentioned in the following subsections.

3.2.1 Support Vector Machines

In latest years Support Vector Machines has gain importance in classification problems. Specifically in biomedical classification signals, good quality performance models have been based in SVM.

In this particular case, the objective of the method is to find the optimal hyperplane to obtain the best separation between two groups: OSA and no-OSA. After testing the method with some kernel functions, Gaussian Radial Basis was selected based on its superior performance with these datasets.

As mentioned, all three measures were used to assess the performance of SVM.

3.2.2 Neural Networks

Neural networks are commonly used in problems like this because of its learning ability. Basically a neural networks learn the patterns that determine a later conduct, for this problem, the target was OSA or no-OSA, and the patterns were determined by the feature set.

A one hidden layer neural network was selected, and after testing its performance, 10 neurons were used in this layer.

Performance was assess by five metrics mentioned.

3.2.3 Logit Model

A Logit Model is a specific application of a linear regression where the value predicted corresponds to a probability, in this case it wouls correspond to the probability of having OSA, so the closer the value predicted is to 1, the more probable is the patient has the disease.

Logit Model inclusion was decided based on its simplicity because it allows to compare more complex machine learning methods -SVM, NN- with simpler classification techniques. The objective was to evaluate how much performance improved with machine learning.

3.3 Preliminary Results

Table 2: Classification Performance of Support Vector Machine, Neural Network and Logit Models for Different Wavelet Decomposition Functions and for ECG, Airflow and Oxygen Saturation

Method	WT Func.	ECG			Air Flow			SpO2		
		Sens	Spec	Acc	Sens	Spec	Acc	Sens	Spec	Acc
SVM	Db1	8,33%	82,05%	50,00%	75,64%	26,33%	54,20%	9,67%	82,56%	50,87%
	Db9	2,67%	88,21%	51,01%	75,38%	28,67%	55,07%	15,00%	79,74%	51,59%
	Haar	10,00%	79,74%	49,42%	70,77%	32,67%	54,20%	13,33%	77,18%	49,42%
	Sym8	5,33%	85,13%	50,43%	75,13%	27,00%	54,20%	12,33%	82,05%	51,74%
NN	Db3	34,33%	60,26%	48,99%	84,36%	19,00%	55,94%	30,00%	75,13%	55,51%
	Db4	36,33%	63,85%	51,88%	80,51%	21,67%	54,93%	19,67%	77,69%	52,46%
	Haar	25,67%	65,90%	48,41%	80,26%	26,33%	56,81%	38,67%	60,00%	50,72%
	Sym1	28,67%	67,69%	50,72%	80,26%	24,33%	55,94%	41,00%	63,33%	53,62%
Logit	Db3	44,56%	54,54%	50,17%	18,89%	74,62%	43,65%	44,52%	51,37%	48,41%
	Db4	41,92%	52,99%	48,21%	19,25%	78,89%	44,93%	55,11%	61,94%	59,03%
	Haar	43,58%	58,36%	51,81%	39,29%	58,13%	46,29%	47,00%	55,79%	52,05%
	Sym7	43,85%	50,99%	47,89%	19,10%	75,59%	43,38%	52,23%	57,03%	55,02%

Table 2 shows a sample of the results obtained from the implemented methods. As seen, the overall performance is not good enough for a screening method to be implemented but, some of the models may set the base for a good quality screening test.

As said before, the desirable model should have high sensitivity and high accuracy, this is, it has a good over all performance and its ability to detect sick people is high.

No model by itself has these characteristics, but models using Air Flow signal had an outstanding sensitivity, over 80% in the Neural Network models, and regular accuracy. ECG based models, on the other hand showed high Specificity, this is, they have the ability to detect healthy people. This suggests a combination of this signals may lead to a desirable model.

As seen on Table 1 many one-signal based models reached 100% of accuracy in adults, this is a clear indicator of the need of a method able to classify accurately children into OSA+ and OSA- groups.

The following Chapter describes a novel approach to this task, from a purely mathematical point of view, signal selection is performed in order to construct a minimal signal model able to screen satisfactorily pediatric population.

Chapter 4

Metamodel Construction

As seen in the preliminary results of Classic Models, none of them by itself can accomplish a satisfactory performance.

In an attempt to integrate advantages from all models a metamodel concept was design. The underlying idea is to assign credibility degrees to a set of models in order to capture the best of each one.

The first approach to this task was through three models. The selection of these models was based on the precision: for each signal, the model that reached the higher precision was selected (*see Appendix*). Table 3 shows models selected for the metamodel construction.

Table 3: Models Selected for Metamodel Construction for Each signal

Signal	Model
Air Flow	ANN Daubechies 2
ECG	SVM Daubechies 8
SpO2	Symlet 2

In order to accomplish the best performance possible the metamodel design was based on a convex combination of the three models described in the previous section. The metamodel is based on a K value used as strike to classify using as input the continuous output derived from the Air Flow, ECG and Oximetry (SpO2) models.

This is, for i patients and j models, the value $x_{i,j}$ represents the output of the model j for the register i , this means the convex combination of models is determined by

parameters α_j depending on the model as follows:

$$X_i = \sum_j \alpha_j x_{i,j} \quad (4.1)$$

The resulting value X_i is a continuous output to be assigned to the classes OSA (1) or no-OSA (0) using the value K . This means, if the output value X_i is higher than K , the register i is assign to class 1, otherwise, to class 0.

The values for α and K were selected based on an optimization model described as follows:

4.1 Optimization Problem

Data:

$x_{i,j}$ = output of the model j for patient i , $x_{i,j} \in [0, 1]$

r_i = real class observed for patient i , $r_i \in \{0, 1\}$

C_{FP} = cost of a False Positive diagnosis

C_{FN} = cost of a False Negative diagnosis

Variables:

α_j = credibility assigned to model j , $\alpha \in [0, 1]$

K = strike value to assign classes to patients based on X_i value, $K \in [0, 1]$

Y_i = class corresponding to patient i , $Y_i \in \{0, 1\}$

Target Function:

$$\text{Min Cost} = \sum_{i,j} r_i(r_i - Y_i)C_{FN} + (1 - r_i)(Y_i - r_i)C_{FP}$$

Restrictions:

1. Convex Combination : $\sum_j \alpha_j = 1$
2. Class assignment (1): $Y_i \geq X_i - K$

3. Class assignment (2): $Y_i \leq 1 + X_i - K$
4. X_i definition: $X_i = \sum_j x_{i,j}$

4.2 Implementation

In the previous section the three models used as an input for this metamodel were presented. Yet, these models can not be used directly as an input.

In the Classification Models section the methodology used to determine the performance of each model was described. In order to obtain robust results, every model was trained and tested 30 times **with 30 different training sets**, this means thirty different models were obtained, because every particular models calibrates its parameters based on examples extracted from the training set, so if the training set changes in every iteration, the model parameters are expected to change with every iteration.

To overcome this fact, in every iteration the accuracy of the model was evaluated as follows:

First, the first iteration model was saved, and an auxiliary variable saved the accuracy of that model as well

Second, for the second iteration the accuracy of the current model was compared with the saved accuracy (from the auxiliary variable), if the current model accuracy was higher, the current model was saved replacing the one saved before

Third, the second step was repeated for every other iteration

In every iteration a new model was trained and tested. Resulting from this process, 30 models were constructed. The model that showed best accuracy was chosen as representative for its class.

To train and test the metamodel, a cross-validation methodology was implemented. On every iteration a random balanced training set, corresponding to 70% of the register was selected, used to train the metamodel, and then the performance of the metamodel was assess with the remaining 30% of the registers. This process was repeated 30 times, and the resulting parameters and performance correspond to the average obtained.

Evidently, the accuracy of the selected models was higher than the accuracy obtained in the Preliminary Results section. But, unexpectedly, the best model obtained from the Oximetry (SpO2) signal had an accuracy of 100%, this fact affects directly the metamodel, the expected α_{SpO2} value is close to 1.

As expected, values of α 's reflected the performance obtained in the Preliminary Results section, giving full credibility to the Oximetry model, although Air Flow and ECG models has acceptable performances as well.

Also value of K is very close to 0, this might be explained by the fact that variable X_i always took a value of 0 or 1, so there was no need of an strike value to decide in the continuous output case, this, of course, is consequence of the values of the α 's.

Chapter 5

Minimal Signal Selection

This phase aims to construct a more complex model able to classify data. As a result, a multi-signal classification algorithm is expected.

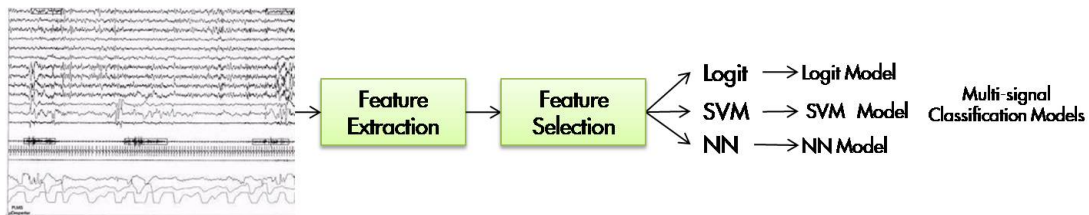


Figure 5: Minimal Signal Selection Stage Design

Figure 5 describes the process to construct first multi-signal classification models.

A detailed explanation of the Feature Extraction and Feature Selection process can be found in the next two sections.

5.1 Data Preprocessing

In order to extract features from each signal alone, polysomnographic data has to be separated into one-signal data files.

Records with more than one signal (like EEG for example) were separated as well. So every signal has its corresponding feature set.

Another possible approach would be to treat each record as one, and extract features from all signals in an EEG, following the example.

5.2 Feature Extraction

5.2.1 ECG Feature Extraction

In order to extract relevant features from an ECG signal, it is important to understand how it is analyzed by clinicians.

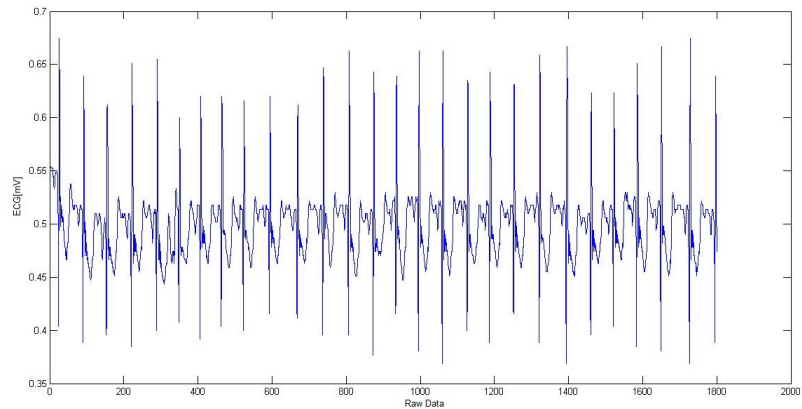


Figure 6: ECG Sample

Figure 6 shows a sample of an ECG signal corresponding to the A0011823 patient.

A regular ECG wave can be characterized according to some points. As seen on Figure 7 points P, Q, R, S and T describe a complete pulse of an ECG.



Figure 7: ECG Wave Characteristics

The relevant characteristics for respiratory functions are associated to the QRS complex. This is the name of the area formed by Q, R and S points. Commonly the features

are extracted from derived signals from the ECG recording. This signals are: Heart Rate Variability (HRV) and ECG Derived Respiratory Signal (EDR).

HRV corresponds to the time serie constructed from R-R intervals, this is the time between successive R points. EDR corresponds to the time serie constructed from QRS amplitude, this is the vertical distance between Q and R points.

Before extracting features, all signals had to be normalized to ignore the variance associated to variation among people. While a certain heart rate might be normal to a person, it could be accelerated to someone else, so normalization is the standard proceeding when working with biological signals.

Common features extracted from HRV and EDR signal are:

$$\begin{aligned} \text{Mean:} \quad \bar{x} &= \frac{1}{N} \sum_{i=1}^N x_i \\ \text{Variance:} \quad \sigma^2 &= \frac{1}{N} \sum_{i=1}^N (\bar{x} - x_i)^2 \end{aligned}$$

Also, some time domain features were extracted from HRV signal. In order to do this a new time serie had to be constructed, corresponding to the successive differences of the HRV signal. These features were:

$$\begin{aligned} \text{Root Mean Square of Successive Differences:} &= \sqrt{\sum_{i=1}^N (x_i - x_{i-1})^2} \\ \text{Standard Deviation:} &= \sqrt{\frac{1}{N} \sum_{i=1}^N (\bar{x} - x_i)^2} \end{aligned}$$

With all these the resulting feature set to characterize the ECG signal is: *meanHRV*, *varHRV*, *meanEDR*, *varEDR*, *RMSSD* and *SDSD*.

5.2.2 EEG Feature Extraction

The EEG performed during a PSG has, as output, eleven signals: *EEGFp1 – A1*, *EEGFp2 – A2*, *EEGA1 – A2*, *EEGF3 – A1*, *EEGF4 – A2*, *EEGC3 – A1*, *EEGC4 – A2*, *EEGP3 – A1*, *EEGP4 – A2*, *EEGO1 – A1* and *EEGO2 – A2*. Each of them characterized by the position of the electrode in the head, where *A1* corresponds to the left side, and *A2* to the right side, as seen on Figure 8.

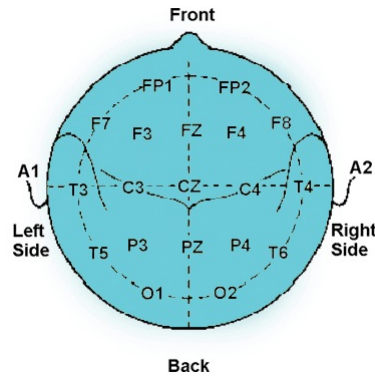


Figure 8: Electrodes position for an EEG

31 registers of the database did not contain signals $EEGA1 - A2$, $EEGFp1 - A1$, $EEGFp2 - A2$, $EEGO1 - A1$ y $EEGO2 - A2$. This is mainly because in small children (under 2 years old) not all the electrodes are placed in the head.

Although, some techniques propose to process all signals together (by adding the signals or applying Principal Components Analysis for example), the decision made was to process them separately. This was decided for two reasons:

First, clinicians analyze them separately and

Second, if there is a combination that allows to process them all together these should be reflected in the feature selection. This is, if the signals are giving redundant information, the feature selection will detect this effect and eliminate it.

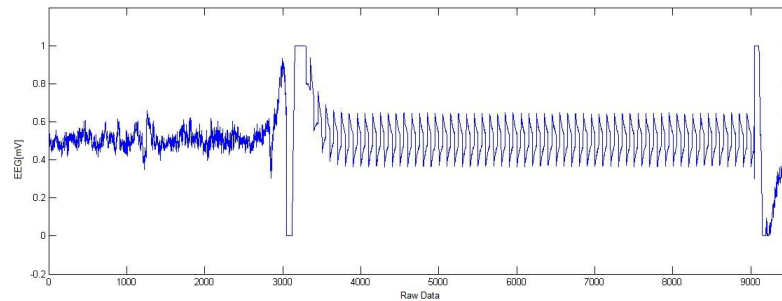


Figure 9: EEG Sample corresponding to patient A0001945

Figure 9 shows a sample of the $EEGF4 - A2$ signal corresponding to patient A0001945. When a PSG is processed, clinicians look into EEG signals to tag time windows according to the sleep stage the patient is in. For example in Figure 9 is easy to distinguish

two stages. So, the main task with these signals is to detect variations in the wave frequencies, in order to characterize those different stages. What matters is to determine how much a wave varies during night. To do this, a methodology was developed.

First, a spectrogram was constructed. Spectrograms are frequently used to analyze sound waves or images. They are a graphic representation of the variety of frequencies in a wave as time varies.

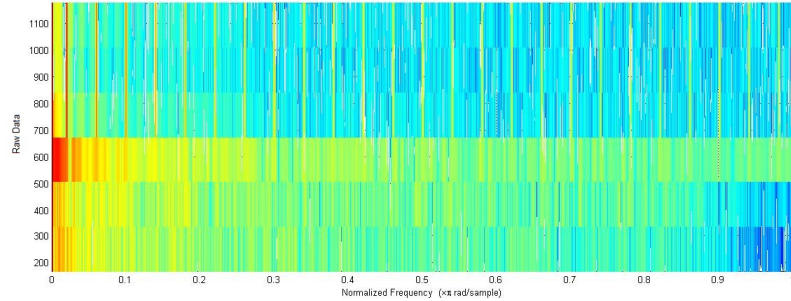


Figure 10: Spectrogram Sample corresponding to patient A0001945

Figure 10 shows an example of how spectrograms look like. On the X-axis a spectrum of frequencies is represented. On the Y-axis the time is divided in small windows for analysis, and the color scale shows how strong is the presence of a frequency in an specific time window.

The Spectrogram showed is the resulting image from processing the signal in Figure 9, even to the naked eye is easy to see two different stages.

If a column is analyzed separately, the variation of an specific frequency can be observed. Based on this, the features to be extracted characterize the number of times a *big* variation occurs, because, when this happens, a change in the sleep stage is suspected.

If smaller time windows are used, smaller changes can be detected, but this effect is not desired, because, changes in the sleep stage are characterized by big changes, so only eight time windows will be used for the analysis.

In order to do this, an auxiliary time series is defined as the difference of a frequency from one time window to the next. This is performed for every frequency considered in the spectrogram.

The main idea is to detect significant variations, so the relative differences were calcu-

lated as the difference divided by the initial value. With this, a new time series was constructed for each signal. Later, this signal was filtered to make easier to detect *big* variations.

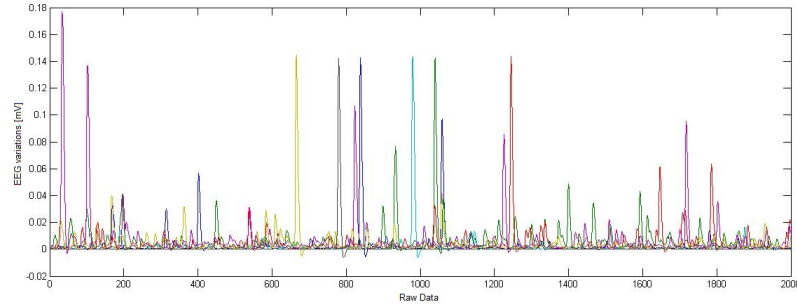


Figure 11: Processed EEG signal corresponding to patient A0001945

Figure 11 shows the processed signal corresponding to the sample in Figure 9, every frequency analyzed is showed in a different color, the variation in time is represented graphically. This means, peaks represent the big changes indicating a change in the sleep stage is suspected.

Finally features extracted from each EEG signal correspond to the number of local minima and the number of local maxima from every frequency filtered signal.

5.2.3 Abdominal and Thoracic Effort Feature Extraction

Both, Abdominal and Thoracic Effort Signals correspond to a band that goes around the abdomen and chest, respectively, that register the movement the patient does while sleeping.

The main idea was to detect peaks of movement, this is, when a patient has a regular respiratory function, no abnormality should appear in the signal. In Figure 12 for example, an abnormality appears at the end of both signals (time ~ 8500). In order to detect peaks wavelet decomposition was used for its ability to show this peaks in the signal. The methodology implemented is the same that in the Classic Models section, this is, features extracted from the signal correspond the mean, standard deviation and energy of the detail signals derived from the decomposition.

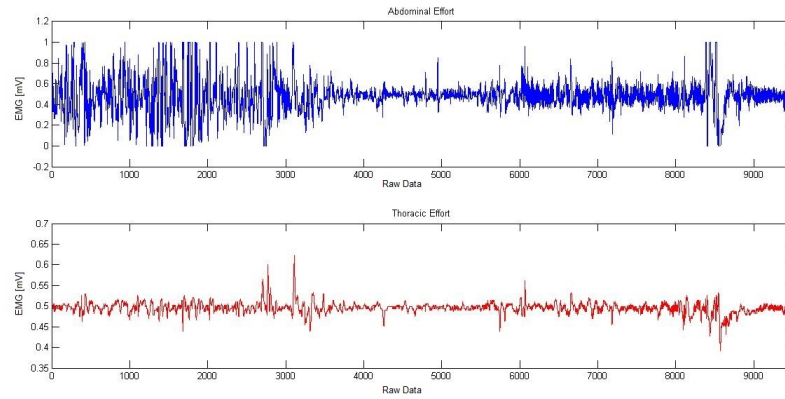


Figure 12: Abdominal Thoracic Signal (blue) and Abdominal Effort Signal (red) corresponding to patient A0000678

As said in the previous chapter, many families of functions can be used with this decomposition, to select the signal that best fits the objectives of this tasks, a visual inspection was done.

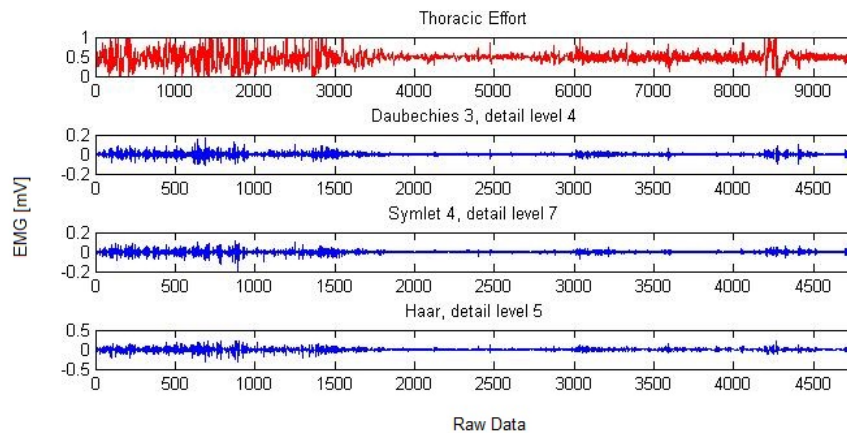


Figure 13: Different Wavelet Decompositions (blue) for the Thoracic Effort Signal (red) corresponding to patient A0000678

In Figure 13, three of the wavelet functions tested are shown. Although three functions seem to detect variations equally, Daubechies 3 had a better performance amplifying small variations, as the one seen at the end of the time window shown in the figure 13, this is why the function selected to process both signals was Daubechies 3 of level 4, so the features resulting from this preprocessing were: $Db3_{mean1}$, $Db3_{var1}$, $Db3_{ene1}$, $Db3_{mean2}$, $Db3_{var2}$, $Db3_{ene2}$, $Db3_{mean3}$, $Db3_{var3}$, $Db3_{ene3}$, $Db3_{mean4}$, $Db3_{var4}$ y $Db3_{ene4}$.

5.2.4 Air Flow Feature Extraction

The Air Flow Signal registers the *air movement* through the nose and the mouth. So, to detect the presence of OSA, one must detect when an abnormal flow occurs, this is, the is complete absence of flow or a *big decrease* in it.

Figure 14 shows a complete night record, as seen, the task is to detect the flat areas in the signal.

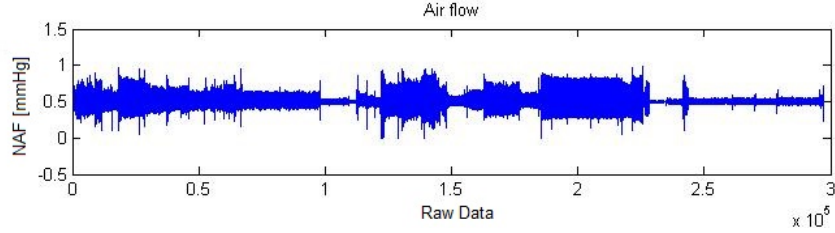


Figure 14: Air Flow Signal corresponding to patient A0000739

The same task was approached in the Classic Models Stage with Wavelet Transform, and the features extracted were used as input for one-signal models. This resulted in high performance models.

So, the decision made was to use the same features collected for that stage, and the function selected was Daubechies 3 of level 14, for its performance on the first stage.

This means, the resulting features were $Db3_{mean1}$, $Db3_{mean2}$, $Db3_{mean3}$, $Db3_{mean4}$, $Db3_{mean5}$, $Db3_{mean6}$, $Db3_{mean7}$, $Db3_{mean8}$, $Db3_{mean9}$, $Db3_{mean10}$, $Db3_{mean11}$, $Db3_{mean12}$, $Db3_{mean13}$, $Db3_{mean14}$, $Db3_{var1}$, $Db3_{var2}$, $Db3_{var3}$, $Db3_{var4}$, $Db3_{var5}$, $Db3_{var6}$, $Db3_{var7}$, $Db3_{var8}$, $Db3_{var9}$, $Db3_{var10}$, $Db3_{var11}$, $Db3_{var12}$, $Db3_{var13}$, $Db3_{var14}$, $Db3_{ene1}$, $Db3_{ene2}$, $Db3_{ene3}$, $Db3_{ene4}$, $Db3_{ene5}$, $Db3_{ene6}$, $Db3_{ene7}$, $Db3_{ene8}$, $Db3_{ene9}$, $Db3_{ene10}$, $Db3_{ene11}$, $Db3_{ene12}$, $Db3_{ene13}$ and $Db3_{ene14}$.

5.2.5 Leg Movement Feature Extraction

The leg signal registers the movement of one or both legs during the sleeping time.

From the 78 patients, 36 had only one signal that recorded their legs movement during sleep, 41 of them had one record for each leg, this is two signals recorded their legs movement, and one patient had no record.

As seen on Figure 15 both legs are highly correlated, so no loss of information is suspected in the case of only one record. Yet, both signals were included in the analysis, if the information results redundant, the Feature Selection Stage should detect it.

All three types of signals were processed using the same method.

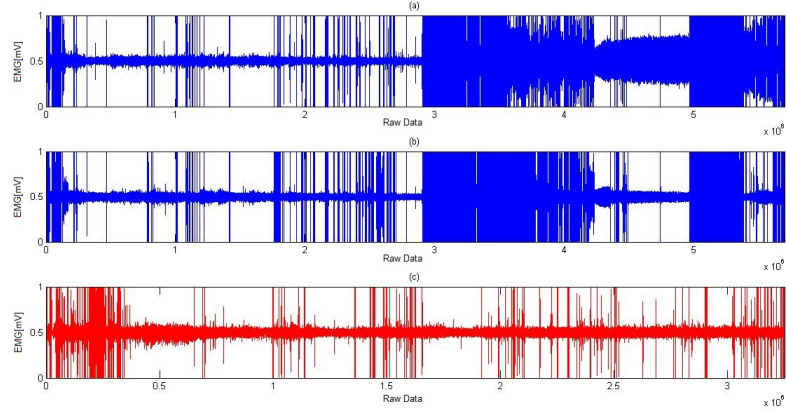


Figure 15: (a) Leg1 Signal corresponding to patient A0001484. (b) Leg2 Signal corresponding to patient A0001484. (c) Legs Signal corresponding to patient A0011947.

The legs signal is often used to diagnose a neurological disease called Restless Legs Syndrome (RLS), also known as Willis-Ekbom disease (WED). Although this is not an SDB, is diagnosed using polysomnography as well. So, even though clinicians use this signal in order to detect movement that may be interfering in other records, it is not relevant in the diagnosis of SDB, particularly not relevant in the detection of OSA.

The characterization of the signal had to be able to represent the variation of the legs movement, this is, detect abnormal movement or *big movement* episodes.

As used before, Wavelet Transform has the ability to extract these features from the signal.

After testing a variety of functions, using the same criterion described before (the ability of detect big variations) Daubechies 8, of level 6 was selected.

As previous uses of this signal, mean, variance and energy of every detail level were calculated, obtaining the following features: $Db8_{mean1}$, $Db8_{var1}$, $Db8_{ene1}$, $Db8_{mean2}$, $Db8_{var2}$, $Db8_{ene2}$, $Db8_{mean3}$, $Db8_{var3}$, $Db8_{ene3}$, $Db8_{mean4}$, $Db8_{var4}$, $Db8_{ene4}$, $Db8_{mean5}$, $Db8_{var5}$, $Db8_{ene5}$, $Db8_{mean6}$, $Db8_{var6}$ and $Db8_{ene6}$.

5.2.6 Body Position Feature Extraction

Body Position signal detects the changes of position of the body, from supine to prone. Prone is the position where the body is facing down completely, chest is the main support for body weight and the head is either resting on the left or right ear. Supine, on the other hand, is the position where the body is completely extended with the back facing down (see *Figure 16*).

During sleeping the body can take many other positions besides supine and prone, but in order to simplify the collection of data, only changes from these two positions are considered as a measure of the movement during the night.

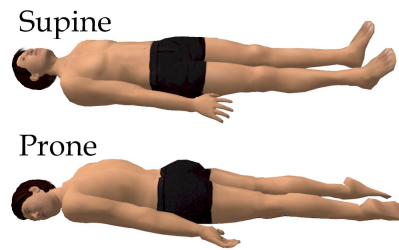


Figure 16: Prone and Supine positions

The sensor used for this assigns a number to each position. The number in this case, does not mean anything, the underlying idea is a nominal variable that detects changes in position.

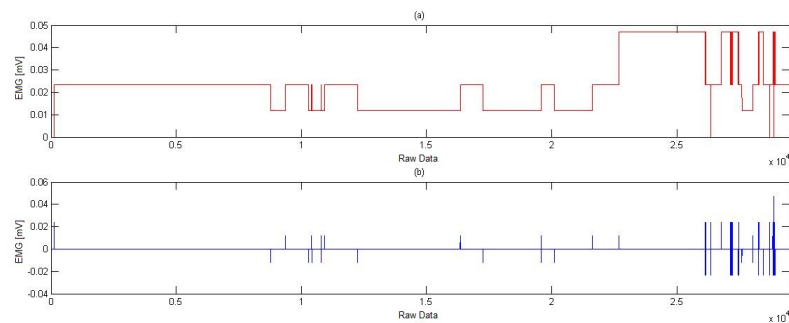


Figure 17: (a) Body Position Signal corresponding to patient A0001460. (b) Auxiliary signal derived from Body Position Signal for patient A0001460

This signal is used by clinicians to detect the number of times a patient moves during the night. In order to do this an auxiliary signal was constructed as the successive differences of the signal, so only the changes in position would appear a different from

zero as seen on Figure 17. With this, the feature extracted was an index of the number of changes divided by the time of sleep.

5.2.7 Electromyogram Feature Extraction

The Electromyogram signal (EMG) registers the electrical potential generated by muscle cells, in a PSG in particular is used to register movement of the chin during sleep. Chin movement is important in the diagnosis of SDB, and OSA, because respiration may occur through nose or mouth, thus chin movement is used as an indicator of mouth movement.

On the other hand, tension in the chin might indicate snoring or another abnormalities during sleep, like bruxism for example.

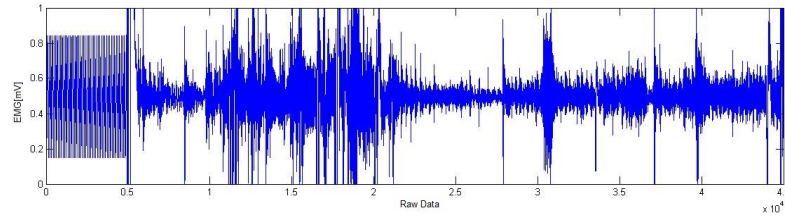


Figure 18: EMG Chin Signal corresponding to patient A0002197

As other signals that register movement, the process of the signal was accomplished through Wavelet Transform. The same selection process was performed, the chosen function was Daubechies 7 of level 8, and the resulting features were: $Db7_{mean1}$, $Db7_{var1}$, $Db7_{ene1}$, $Db7_{mean2}$, $Db7_{var2}$, $Db7_{ene2}$, $Db7_{mean3}$, $Db7_{var3}$, $Db7_{ene3}$, $Db7_{mean4}$, $Db7_{var4}$, $Db7_{ene4}$, $Db7_{mean5}$, $Db7_{var5}$, $Db7_{ene5}$, $Db7_{mean6}$, $Db7_{var6}$, $Db7_{ene6}$, $Db7_{mean7}$, $Db7_{var7}$, $Db7_{ene7}$, $Db7_{mean8}$, $Db7_{var8}$ and $Db7_{ene8}$.

5.2.8 Electro-oculogram Feature Extraction

The Electro-oculogram (EOG) registers the movement of the eyes by placing electrodes around them. The result of this exam is one signal for each eye. So, in a complete PSG, two signals correspond to EOG, and they are tagged as EOG Right, and EOG Left for the right eye, and the left eye respectively.

Even though these are separated signals, a high correlation is expected, thus eye movements are not independent from each other.

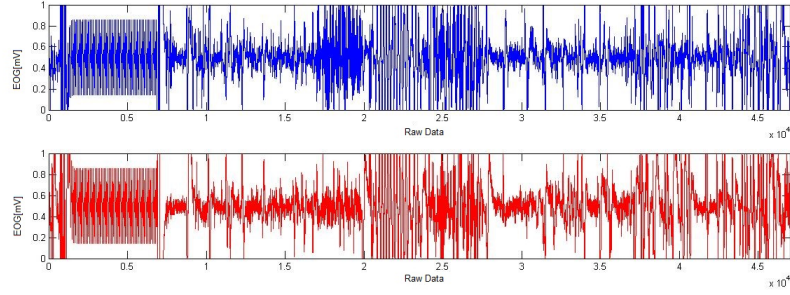


Figure 19: EOG Signals of Left (blue) and Right (Red) eyes corresponding to patient A0006225

In Figure 19 is easy to see the correlation between both left and right eye of patient A0006225. Same correlation is observed in every record from this signal. Yet, every signal was processed because feature extraction process should be able to detect this highly correlated features.

In order to be consistent in the treatment of signals, all movement signals were processed with Wavelet Transform. The function used for EOG feature extraction was Daubechies 7 of level 8. So the features extracted for each signal (right and left) were: $Db7_{mean1}$, $Db7_{var1}$, $Db7_{ene1}$, $Db7_{mean2}$, $Db7_{var2}$, $Db7_{ene2}$, $Db7_{mean3}$, $Db7_{var3}$, $Db7_{ene3}$, $Db7_{mean4}$, $Db7_{var4}$, $Db7_{ene4}$, $Db7_{mean5}$, $Db7_{var5}$, $Db7_{ene5}$, $Db7_{mean6}$, $Db7_{var6}$, $Db7_{ene6}$, $Db7_{mean7}$, $Db7_{var7}$, $Db7_{ene7}$, $Db7_{mean8}$, $Db7_{var8}$ and $Db7_{ene8}$.

5.2.9 Pulse Feature Extraction

Pulse signal as predictor for OSA prevalence has been used in other studies, because, as expected, an abnormal pulse rate might be a good indicator of an apnea episode.

In particular, Noehren et al. [21] uses four time series derived from the pulse signal:

- APRD: Absolute Pulse Rate Decrease
- APRI: Absolute Pulse Rate Increase
- RPRD: Relative Pulse Rate Decrease
- RPRI: Relative Pulse Rate Increase

For the purpose of this study, only the first two were used.

First the pulse signal was normalized to avoid variation among patients(see *Figure 20 (a)*). Then a butter filter was used to eliminate the noise from the signal and detect easily peaks of local minima and local maxima (see *Figure 20 (b)*).

The APRD length corresponded to the number of peaks found, and each value was calculated as the difference of a local minimum and the local maximum immediately before (see *Figure 20 (d)*). The same procedure was applied to construct the APRI (see *Figure 20 (c)*), calculating the difference between a local maximum and the local minimum immediately before.

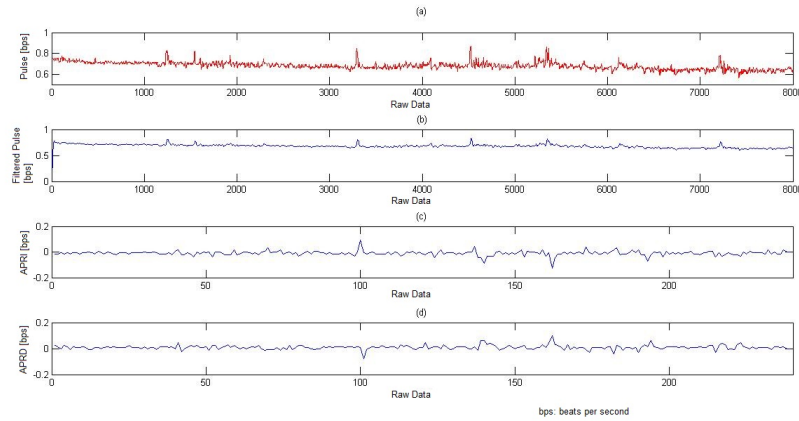


Figure 20: (a)Pulse signal corresponding to patient A0000294. (b)Filtered pulse signal. (c)Absolute Pulse Rate Increase (APRI) derived from pulse signal. (d)Absolute Pulse Rate Decrease (APRD) derived from pulse signal.

After this, the result are two time series, so the decision was to extract, from both of them: mean, standard deviation, median, maximum and minimum. So, the resulting feature set was: $APRD_{mean}$, $APRD_{std}$, $APRD_{median}$, $APRD_{max}$, $APRD_{min}$, $APRI_{mean}$, $APRI_{std}$, $APRI_{median}$, $APRI_{max}$ and $APRI_{min}$.

5.2.10 Snore Feature Extraction

The snore signal is used to diagnose not only OSA, but other SDB's, like primary snoring. So, although is very useful for diagnosis, the final classification into the groups OSA and no-OSA can not rely only on this signal.

Snoring is measure by a sensor located on the throat that detects vibration. As snoring is produced by an obstruction of the airway, vibration is produced, and it can be measure this way.

The resulting signal is normalized because the importance is in the variations among it, so as an output for the PSG, a signal from 0 to 1 results.

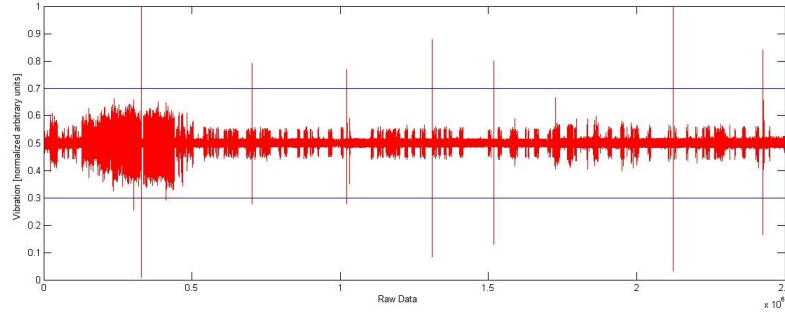


Figure 21: Snore Signal corresponding to patient A0000884

The task in the preprocessing of this signal was to determine when an abnormal *vibration episode* occurred. The signal provided by the PSG is already normalized, so an abnormal episode was considered when the signal cross over 0.7 or under 0.3 (blue lines on Figure 21), this selection was based on preliminary analysis of the signal behavior. The features extracted corresponded to the number of abnormal episodes *up* and abnormal episodes *down*. The resulting feature set was: $Snore_{up}$ and $Snore_{down}$.

5.2.11 Oxygen Saturation Feature Extraction

PSG also includes a sensor that measures during the whole night the oxygen saturation in blood. Regularly, saturation is measured periodically, so PSG is one of the few clinical tests that allows a Saturation Signal.

Evidently this is one of the most important signals to determine a diagnosis, this, because is directly related to apnea episodes. A desaturation episode is a clear symptom of an abnormal respiratory function, in particular, it might be the perfect indicator of an apnea episode. Figure 22, shows a time window of the Oxygen Saturation Signal collected by PSG from patient A0001183 (a healthy patient).

Many indicators are used for processing this signal. For this study, DEI and ODI were selected because of its previous use in continuous positive airway pressure devices (CPAP) tests and previous experiences in diagnosing apnea episodes [6, 7].

DEI are the initials for Desaturation Event Index and its defined as a saturation signal

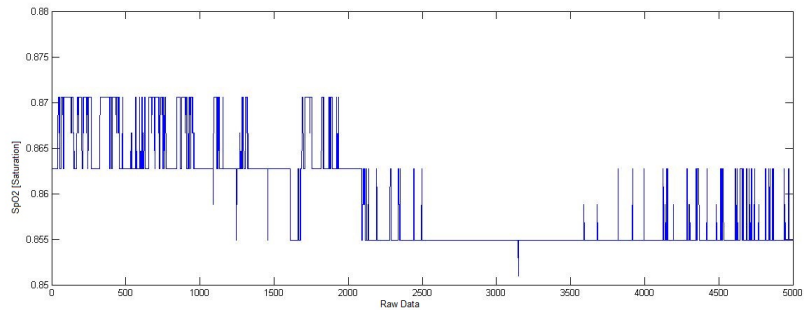


Figure 22: Oxygen Saturation Signal corresponding to patient A0001183

drop in four or more points (in scale of 1 to 100). ODI are the initials of Oxygen Desaturation Index, also known as CT90 (Cumulative Time 90), it corresponds to the amount of time the patient has a saturation under 90%.

So, as result, the feature set was: *DEI* and *ODI* for each signal.

5.3 Feature Selection

Resulting from the Feature Extraction process, 355 features are used to resume information collected in a PSG.

The hypothesis of this stage is not all that information is needed to perform a high quality screening. If this is true, a less resource consuming clinical test may be develop to prediagnose children with suspected OSA, which is the main expected result of this investigation.

The first step to accomplish this is to select features. this selection should imply a reduction in the signals needed to perform a screening. Table 4 summarizes the features resulting from the previous section. In the table, features are grouped by signal, this approach was used because if one feature corresponding to one signal was considered important for the resulting subset, all the other features corresponding to the same signal were included in this first step (because the signal is needed anyway).

A filter approach was selected to perform Feature Selection. This allows to use the same subset with all three models and allows comparison under the same conditions for the Preliminary Results Section.

Embedded methods were discarded because of its dependency on the model selected,

Table 4: Summary of features resulting from Feature Extraction Process divided by signal for a complete PSG

Signal	Features Extracted	Total
ECG	$mean_{HRV}, var_{HRV}, mean_{EDR}, var_{EDR}, RMSSD$ and $SDSD$.	6
EEG (11)	$max_1, max_2, max_3, max_4, max_5, max_6, max_7, min_1, min_2, min_3, min_4, min_5, min_6$ and min_7 .	154
Abdominal Effort	$Db3_{mean1}, Db3_{var1}, Db3_{ene1}, Db3_{mean2}, Db3_{var2}, Db3_{ene2}, Db3_{mean3}, Db3_{var3}, Db3_{ene3}, Db3_{mean4}, Db3_{var4}$ y $Db3_{ene4}$.	12
Thoracic Effort	$Db3_{mean1}, Db3_{var1}, Db3_{ene1}, Db3_{mean2}, Db3_{var2}, Db3_{ene2}, Db3_{mean3}, Db3_{var3}, Db3_{ene3}, Db3_{mean4}, Db3_{var4}$ y $Db3_{ene4}$.	12
Air Flow	$Db3_{mean1}, Db3_{var1}, Db3_{ene1}, Db3_{mean2}, Db3_{var2}, Db3_{ene2}, Db3_{mean3}, Db3_{var3}, Db3_{ene3}, Db3_{mean4}, Db3_{var4}, Db3_{ene4}, Db3_{mean5}, Db3_{var5}, Db3_{ene5}, Db3_{mean6}, Db3_{var6}, Db3_{ene6}, Db3_{mean7}, Db3_{var7}, Db3_{ene7}, Db3_{mean8}, Db3_{var8}, Db3_{ene8}, Db3_{mean9}, Db3_{var9}, Db3_{ene9}, Db3_{mean10}, Db3_{var10}, Db3_{ene10}, Db3_{mean11}, Db3_{var11}, Db3_{ene11}, Db3_{mean12}, Db3_{var12}, Db3_{ene12}, Db3_{mean13}, Db3_{var13}, Db3_{ene13}, Db3_{mean14}, Db3_{var14}$ and $Db3_{ene14}$.	48
Leg Movement (2)	$Db8_{mean1}, Db8_{var1}, Db8_{ene1}, Db8_{mean2}, Db8_{var2}, Db8_{ene2}, Db8_{mean3}, Db8_{var3}, Db8_{ene3}, Db8_{mean4}, Db8_{var4}, Db8_{ene4}, Db8_{mean5}, Db8_{var5}, Db8_{ene5}, Db8_{mean6}, Db8_{var6}$ and $Db8_{ene6}$.	36
Body Position	CI	1
EMG	$Db7_{mean1}, Db7_{var1}, Db7_{ene1}, Db7_{mean2}, Db7_{var2}, Db7_{ene2}, Db7_{mean3}, Db7_{var3}, Db7_{ene3}, Db7_{mean4}, Db7_{var4}, Db7_{ene4}, Db7_{mean5}, Db7_{var5}, Db7_{ene5}, Db7_{mean6}, Db7_{var6}, Db7_{ene6}, Db7_{mean7}, Db7_{var7}, Db7_{ene7}, Db7_{mean8}, Db7_{var8}$ and $Db7_{ene8}$.	24
EOG (2)	$Db7_{mean1}, Db7_{var1}, Db7_{ene1}, Db7_{mean2}, Db7_{var2}, Db7_{ene2}, Db7_{mean3}, Db7_{var3}, Db7_{ene3}, Db7_{mean4}, Db7_{var4}, Db7_{ene4}, Db7_{mean5}, Db7_{var5}, Db7_{ene5}, Db7_{mean6}, Db7_{var6}, Db7_{ene6}, Db7_{mean7}, Db7_{var7}, Db7_{ene7}, Db7_{mean8}, Db7_{var8}$ and $Db7_{ene8}$.	48
Pulse	$APRD_{mean}, APRD_{std}, APRD_{median}, APRD_{max}, APRD_{min}, APRI_{mean}, APRI_{std}, APRI_{median}, APRI_{max}$ and $APRI_{min}$.	10
Snore	$Snore_{up}$ and $Snore_{down}$	2
Oxygen Saturation	DEI and ODI	2
TOTAL		355

and all three models would have ended with three different feature subsets. Wrapper methods were discarded also, because they did not allow to control the number of resulting features in the subset.

Before starting the selection, the quality of the data was checked. Features corresponding to pulse signal were eliminated from the starting Features Set because only 6 patients had complete pulse records derived from PSG.

This means, the starting Features Set excluded $APRD_{mean}, APRD_{std}, APRD_{median}, APRD_{max}, APRD_{min}, APRI_{mean}, APRI_{std}, APRI_{median}, APRI_{max}$ and $APRI_{min}$.

The selection procedure was based on correlation analysis and Principal Component

Analysis. The step by step methodology is described in the following:

First, Principal Component Analysis (PCA) was computed over all the Features included, using Varimax rotation criterion. The main idea was to extract variables able to explain enough of the variance to incorporate as much information as possible in the modelling section.

Second, first five components resulting from Factorial Analysis were analysed, as the explained 89,9% of the variance, and the sixth component added only 2% to the variance explanation.

Third, variables eigenvalues were extracted, and only variables with an eigenvalue of 0,05 (5%) or more for the first five principal components were kept.

Fourth, as said before, all features corresponding to selected signals were kept. Resulting from these steps, features corresponding to the following signals formed the preliminary subset: EMG Chin, EOG (Left and Right), Snore and Legs. Preliminary subset had 129 features.

Fifth, Correlation Matrix was computed over the preliminary subset.

Sixth, every pair of signals was analysed, if a correlation of over 0,6 was found, only one of the pair of signals was kept.

Seventh, resulting from these steps, 51 features remain in the subset. This is the definitive Features Subset.

Table 5 shows the resulting subset of features from Feature Selection process.

5.4 Classification Models

Models used to classify patients into the two groups defined are the same used in the first stage. These are: Support Vector Machine, Neural Networks and Logit Regression. Also, the same validation methodology was implemented.

Thirty different training and testing balanced sets were created and used to train and test every model. Performance metrics were calculated each time.

Table 5: Summary of features resulting from Feature Selection Process divided by signal

Signal	Features Selected	Total
Leg 1	$Db8_{mean1}, Db8_{mean2}, Db8_{var2}, Db8_{mean3}, Db8_{mean4}, Db8_{mean5}, Db8_{var5}$ and $Db8_{mean6}$.	8
Leg 2	$Db8_{mean3}, Db8_{mean4}, Db8_{mean5}$ and $Db8_{var5}$.	4
Legs	$Db8_{mean1}, Db8_{var1}, Db8_{mean2}, Db8_{mean3}, Db8_{mean4}$ and $Db8_{mean5}$.	6
EOG Right	$Db7_{mean1}, Db7_{mean2}, Db7_{mean3}, Db7_{mean4}, Db7_{mean5}, Db7_{mean6}, Db7_{mean7}$ and $Db7_{var7}$.	8
EOG Left	$Db7_{mean1}, Db7_{var1}, Db7_{mean2}, Db7_{mean3}, Db7_{var3}, Db7_{mean4}, Db7_{mean5}, Db7_{mean6}, Db7_{mean7}, Db7_{var7}$ and $Db7_{mean8}$.	11
EMG Chin	$Db7_{mean1}, Db7_{var1}, Db7_{mean2}, Db7_{mean3}, Db7_{mean4}, Db7_{mean5}, Db7_{mean6}, Db7_{mean7}, Db7_{var7}, Db7_{mean8}$ and $Db7_{var8}$.	11
TOTAL		51

The resulting performance corresponds to the average of the metrics resulting from every iteration (this is, the result from one training set and one testing set).

Metrics selected to evaluate models performance are the same used in Classic Models Stage: Sensitivity, Specificity, Accuracy, Precision, Recall, and F-score.

5.5 Preliminary Results

The following tables describe the results obtained from every classification model separately.

Table 6: Classification Performance of Support Vector Machine using a feature subset from PSG as input

Model	Sensitivity	Specificity	Accuracy	Precision	Recall	F-Score
SVM	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Support Vector Machine showed an outstanding performance when the minimal signal selected are used as input. This result is specially unexpected because, as far as this investigation went, no other published investigation used EMG, EOG and leg movement as input signals. In fact, clinicians use these signals for auxiliary information, and never as the main source of information for diagnosis.

These two models (Artificial Neural Networks and Logit) showed poorer performance

Table 7: Classification Performance of Neural Network using a feature subset from PSG as input

Model	Sensitivity	Specificity	Accuracy	Precision	Recall	F-Score
NN	28,67%	100,00%	28,67%	100,00%	28,67%	39,91%

Table 8: Classification Performance of Logit Regression using a feature subset from PSG as input

Model	Sensitivity	Specificity	Accuracy	Precision	Recall	F-Score
LOGIT	47,30%	100,00%	47,30%	100,00%	47,30%	62,69%

compared to Support Vector Machine, but they also showed unexpected Precision results.

Chapter 6

Results and Discussion

6.1 Classic Models Stage Results

6.1.1 Support Vector Machine

SVM applied to the airflow signal had an average sensitivity of $71,14\% \pm 3,34\%$, specificity of $33,08\% \pm 4,71\%$ and accuracy of $54,59\% \pm 0,98\%$. As the method is being tested in order to implement an screening method, specificity is the most important metric, because it measures the ability of the algorithm to detect OSA patients, this is, it does not classify a patient as no-OSA when is OSA (*see Table 9*).

Table 9: Summary of performance indicators for Support Vector Machine applied to Air Flow signal

	Sensitivity	Specificity	Accuracy	Precision	Recall	F
mean	71,14%	33,08%	54,59%	58,12%	71,14%	62,98%
sd	3,34%	4,71%	0,98%	0,83%	3,34%	1,31%
max	76,41%	40,67%	56,52%	59,02%	76,41%	65,66%
min	66,15%	25,33%	52,46%	56,20%	66,15%	60,74%

The maximum performance in specificity for airflow, is 40,67% for Symlet3 decomposition function. Although, sensitivity has some good quality results, with a maximum of 76,41% for Symlet7 decomposition function, this is a poor performance for an screening method considering an average clinic test has a performance of 95% or higher.

On the ECG signal, SVM had an average performance of $8,79\% \pm 4,50\%$ for sensitivity, $83,49\% \pm 2,96\%$ for specificity and $51,01\% \pm 2,42\%$ for accuracy. most of the models

Table 10: Summary of performance indicators for Support Vector Machine applied to ECG signal

	Sensitivity	Specificity	Accuracy	Precision	Recall	F
media	8,79%	83,49%	51,01%	28,04%	8,79%	11,92%
sd	4,50%	2,96%	2,42%	12,32%	4,50%	6,13%
max	22,00%	88,21%	58,55%	60,40%	22,00%	30,41%
min	2,67%	78,21%	47,54%	13,57%	2,67%	3,90%

had good performance in specificity (over 80%) but bad results in sensitivity (under 10%) resulting in accuracy of no more than 60%.

Table 11: Summary of performance indicators for Support Vector Machine applied to Oximetry signal

	Sensitivity	Specificity	Accuracy	Precision	Recall	F
media	15,83%	79,74%	51,95%	36,67%	15,83%	19,98%
sd	6,51%	3,78%	1,87%	8,94%	6,51%	7,27%
max	29,33%	89,49%	55,07%	48,05%	29,33%	33,93%
min	4,00%	73,85%	48,12%	13,17%	4,00%	5,57%

SVM tested with the SpO2 dataset resulted in $15,83\% \pm 6,51\%$ of sensitivity, $79,74\% \pm 3,78\%$ of specificity and $51,95\% \pm 1,87\%$ of accuracy. Because of the nature of the signal and its relevant features, model did not perform good. Good sensitivity and accuracy close to 50% were not expected, this may be due to random effect, or just because as features are not relevant, the model classifies 60% as no-OSA, because of the dataset balance.

6.1.2 Neural Network

Table 12: Summary of performance indicators for Neural Network applied to Air Flow signal

	Sensitivity	Specificity	Accuracy	Precision	Recall	F
mean	76,36%	26,30%	54,60%	57,60%	76,36%	64,09%
sd	4,35%	3,78%	1,76%	1,43%	4,35%	2,67%
max	84,36%	34,67%	57,97%	60,68%	84,36%	67,77%
min	69,23%	19,00%	51,16%	55,30%	69,23%	58,31%

For the airflow signal, the average sensitivity was $76,36\% \pm 4,35\%$, the average specificity $26,30\% \pm 3,78\%$ and the average accuracy $54,60\% \pm 1,76\%$. This particular

method has a maximum sensitivity of 84,36% for Daubechies3 decomposition signal. But the Specificity for that same signal is below 20%. This too is a bad method by itself to be used as screening.

Table 13: Summary of performance indicators for Neural Network applied to ECG signal

	Sensitivity	Specificity	Accuracy	Precision	Recall	F
media	32,68%	64,32%	50,57%	41,30%	32,68%	34,34%
sd	4,92%	3,22%	1,91%	4,05%	4,92%	4,37%
max	43,00%	73,33%	55,51%	51,19%	43,00%	44,49%
min	24,00%	57,69%	48,41%	34,05%	24,00%	27,56%

On the ECG signal, poor results were achieved. Sensitivity had an average performance of $32,68\% \pm 4,92\%$, while specificity and accuracy were $64,32\% \pm 3,22\%$ and $50,57\% \pm 1,91\%$ respectively. Best model was based on Symlet9 decomposition signal, with sensitivity of 43,00%, specificity of 65,13% and accuracy of 55,51%. Although this is the best model, its accuracy shows that is not much better than flipping a coin.

Table 14: Summary of performance indicators for Neural Network applied to Oximetry signal

	Sensitivity	Specificity	Accuracy	Precision	Recall	F
media	24,35%	76,04%	53,57%	46,08%	24,35%	27,59%
sd	7,99%	7,83%	2,25%	7,40%	7,99%	7,00%
max	41,00%	87,18%	56,52%	57,47%	41,00%	42,08%
min	14,00%	58,72%	47,83%	29,87%	14,00%	18,36%

Neural Network applied to SpO2 signal resulted in good specificity and bad sensitivity models, the metrics obtained were: $24,35\% \pm 7,99\%$, $76,04\% \pm 7,83\%$ and $53,57\% \pm 2,25\%$ for sensitivity, specificity and accuracy.

Although general results were poor, Symlet2 decomposition signal model, resulted in sensitivity of 16,33%, specificity 87,18% and accuracy 56,38%.

6.1.3 Logit

Logit Model had a poorer performance compared to machine learning algorithms.

Table 15: Summary of performance indicators for Logit Model applied to Air Flow signal

	Sensitivity	Specificity	Accuracy	Precision	Recall	F
mean	28,66%	69,69%	46,14%	54,13%	28,66%	31,32%
sd	7,10%	4,84%	3,46%	3,76%	7,10%	7,28%
max	43,36%	78,89%	52,14%	60,05%	43,36%	47,24%
min	18,36%	58,13%	39,93%	47,89%	18,36%	20,88%

For airflow signal, the average performances were: 28,66% \pm 7,10% for sensitivity, 69,69% \pm 4,84% for specificity and 46,14% \pm 3,46% for accuracy.

Best performance was achieved with Daubechies4 decomposition function, with sensitivity of 19,25%, specificity of 78,89% and accuracy of 44,93%.

It is important to notice that even though its performance was poorer, the best specificity indexes were achieved using Logit, this could mean that Logit Model classify almost every record as OSA, but its accuracy is similar to the other methods.

Table 16: Summary of performance indicators for Logit Model applied to ECG signal

	Sensitivity	Specificity	Accuracy	Precision	Recall	F
media	42,37%	54,07%	49,05%	41,32%	42,37%	40,54%
sd	5,89%	3,58%	2,75%	4,41%	5,89%	5,15%
max	49,63%	61,44%	54,19%	49,00%	49,63%	48,11%
min	26,55%	47,37%	43,24%	29,68%	26,55%	26,08%

the average performance for ECG signal were 42,37% \pm 5,89%, 54,07% \pm 3,58% and 49,05% \pm 2,75% for sensitivity, specificity and accuracy respectively.

The best model was Daubechies10 with metrics of 38,11%, 60,97% and 50,81%. For this particular signal, Logit model showed no advantage over other more complex models.

Applied to SpO2 signal, results were the following: 47,88% \pm 5,33% of sensitivity,

Table 17: Summary of performance indicators for Logit Model applied to Oximetry signal

	Sensitivity	Specificity	Accuracy	Precision	Recall	F
media	47,88%	54,89%	51,86%	45,04%	47,88%	45,28%
sd	5,33%	4,03%	3,97%	4,49%	5,33%	5,00%
max	56,56%	61,94%	59,03%	52,53%	56,56%	53,13%
min	37,89%	46,20%	42,15%	35,32%	37,89%	35,56%

54, 89% \pm 4, 03% of specificity and 51, 86% \pm 3, 97% of accuracy.

6.2 Metamodel Stage Results

Although, the Metamodel implemented did not increase performance of the models, because one of the models showed 100% performance in every metric used for this investigation, the results of the Metamodel implemented are shown in Table 18.

Table 18: Classification Performance of Metamodel

Model	Sesitivity	Specificity	Accuracy	Precision	Recall	F-Score
METAMODEL	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

As expected, the Metamodel had the same performance as the model with an α (trust coefficient) of 1 (Results of every model separately are shown in Section 5.5).

models from this stage reached the performance reported in adults studies of 100% accuracy (see Table 1), although those models use only one signal, the performance of three-models signal largely overcomes the expected results in children (no documented experience was found in children).

6.3 Discussion

The first phase showed methods for screening of OSA based on a black-box approach to the feature selection, also, in order to make them comparable all feature extraction process was made through Wavelet Transform, that lead to good quality features - relevant for the diagnosis- in two of the three signals: ECG and airflow. But no relevant features were extracted in SpO2 signal.

Resulting from this phase several models with quality not good enough to be considered

as acceptable screening methods were obtained. The main conclusion of this first approach is that **no signal alone can be a good predictor for OSA in children**, at least not with this black-box feature extraction approach.

In a second attempt to use these commonly used signals, a metamodel was constructed, but as the selection criterion was only based on the accuracy of the models, no major conclusion can be drawn from its implementation. Although it sets a base to integrate a various models into a higher performance robust model.

A derived conclusion from this, is the suspicion that a more physiological approach to feature extraction may lead to good quality classification methods.

Evidence in adults and some studies conducted in children showed good-quality screening methods were mainly based on physiological feature extraction methods. Most of these studies also, were based on one-signal only. The few of them that were based on more than one signal did not show any criterion for the signal selection beyond clinicians criteria.

This is why, the approach presented in the second phase is a novel method for OSA screening. Signal selection was based only on mathematical criteria, but also, the secondary objective of a massifiable screening methods, was always present.

Resulting from this process, three signals were selected: EMG-Chin, EOG and Leg Movement. Surprisingly none of these signals have been used by itself, or in combination with other signals, for screening in any documented study reviewed in this investigation.

Also, all of this signals are collected with ordinary electrodes, widely available in the market. This fact makes possible to think in this algorithm as an actual screening method for future investigations or even to be tested in patients with suspected OSA.

Regarding the objectives of this investigation, they are consider accomplished. An algorithm able to screen OSA with 100% of precision was developed using signals collected by polysomnography and information needed was reduce using data mining techniques, so only three signals are needed to perform a high quality prediagnosis.

Also, resulting from this work, a high quality repository of polysomnographic children data tagged by qualified clinicians is available for further work.

6.3.1 Recommendations and Further Work

For signals like ECG or Air Flow, widely used in screening methods for adults, the results showed no relevance or correlation with the target variable in (feature selection process, see Section 5.3). Clinicians base their diagnosis in this signals (although they use all information available from a polysomnography) so the intuitive reasoning lead to think these signals should appear as relevant in the feature selection. This fact may mean two things: (1) signals are not relevant for diagnosis or (2) features extracted are not representative enough of the information contained in the signal. (1) is clearly not correct because, as said before, when the diagnosis process is performed by clinicians, these signals are the basis of the diagnosis.

This reasoning leads to perform feature extraction with more adequate methods for each signals. Some investigations devote most of the time to perform feature extraction correctly, so a comparison of different feature extraction methods is recommended in order to correctly represent signals for the feature selection process.

If this algorithm, or any other, is intended to be implemented as a valid screening method, is necessary it meets some requirements. First, and most evident of them, is that it has enough predictive power to be trusted with the prediagnosis. Also it has to be easier to perform than the gold-standard, less resource consuming and independent of clinician supervision. The proposed method in this investigation actually meets these requirements. But there is a third requirement, not less important: clinicians must trust in the screening test in order to actually prescribe it.

As mentioned before, at the best of our knowledge, no documented work uses any of the resulting signals (EMG Chin, EOG, Leg Movement) alone or combined with others for prediagnosis. This is mainly because these signals are not the main source of information in regular diagnosis process, so is pending to check clinicians opinion and disposition to actually prescribe this kind of clinical test.

References

- [1] Diego Álvarez-Estévez and Vicente Moret-Bonillo. Fuzzy reasoning used to detect apneic events in the sleep apnea-hypopnea syndrome. *Expert Systems with Applications*, 36(4):7778–7785, 2009.
- [2] Peter Anderer, A Moreau, M Woertz, M Ross, G Gruber, S Parapatics, E Loretz, E Heller, A Schmidt, M Boeck, et al. Computer-assisted sleep classification according to the standard of the american academy of sleep medicine: validation study of the aasm version of the somnolyzer 24× 7. *Neuropsychobiology*, 62(4):250–264, 2010.
- [3] Richard B Berry, Rohit Budhiraja, Daniel J Gottlieb, David Gozal, Conrad Iber, Vishesh K Kapur, Carole L Marcus, Reena Mehra, Sairam Parthasarathy, Stuart F Quan, et al. Rules for scoring respiratory events in sleep: Update of the 2007 aasm manual for the scoring of sleep and associated events: Deliberations of the sleep apnea definitions task force of the american academy of sleep medicine. *Journal of clinical sleep medicine: JCSM: official publication of the American Academy of Sleep Medicine*, 8(5):597, 2012.
- [4] Pablo E Brockmann, Pablo Bertrand, Trinidad Pardo, Jaime Cerda, Bernardita Reyes, and Nils L Holmgren. Prevalence of habitual snoring and associated neurocognitive consequences among chilean school aged children. *International Journal of Pediatric Otorhinolaryngology*, 2012.
- [5] Carolina Caminiti, Patricia Evangelista, Vivian Leske, Yanina Loto, and Carmen Mazza. Síndrome de apnea obstructiva del sueño en niños obesos sintomáticos: confirmación polisomnográfica y su asociación con trastornos del metabolismo hidrocarbonado. *Archivos argentinos de pediatría*, 108(3):226–233, 2010.

- [6] Li Chang, Jianxin Wu, and Ling Cao. Combination of symptoms and oxygen desaturation index in predicting childhood obstructive sleep apnea. *International journal of pediatric otorhinolaryngology*, 2012.
- [7] Frances Chung, Pu Liao, Hisham Elsaid, Sazzadul Islam, Colin M Shapiro, and Yuming Sun. Oxygen desaturation index from nocturnal oximetry: a sensitive and specific tool to detect sleep-disordered breathing in surgical patients. *Anesthesia & Analgesia*, 114(5):993–1000, 2012.
- [8] Philip De Chazal, Conor Heneghan, Elaine Sheridan, Richard Reilly, Philip Nolan, and Mark O’Malley. Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnoea. *Biomedical Engineering, IEEE Transactions on*, 50(6):686–696, 2003.
- [9] María de la Luz Alonso Álvarez, Joaquín Terán Santos, José Cordero Guevara, Mónica González Martínez, Luis Rodríguez Pascual, José Luis Viejo Bañuelos, and Ángel Marañón Cabello. Reliability of home respiratory polygraphy for the diagnosis of sleep apnea-hypopnea syndrome. analysis of costs. *Archivos de Bronconeumología ((English Edition))*, 44(1):22–28, 2008.
- [10] Helen S Driver, Effie J Pereira, Kathryn Bjerring, Fern Toop, Steven C Stewart, Peter W Munt, and Michael F Fitzpatrick. Validation of the medibyte® type 3 portable monitor compared with polysomnography for screening of obstructive sleep apnea. *Canadian respiratory journal: journal of the Canadian Thoracic Society*, 18(3):137, 2011.
- [11] W Ward Flemons, Michael R Littner, James A Rowley, Peter Gay, W McDowell Anderson, David W Hudgel, R Douglas McEvoy, and Daniel I Loubé. Home diagnosis of sleep apnea: A systematic review of the literaturean evidence review cosponsored by the american academy of sleep medicine, the american college of chest physicians, and the american thoracic society. *CHEST Journal*, 124(4):1543–1579, 2003.
- [12] AASM GENERAL. International classification of sleep disorders, 2nd edn.: Diagnostic and coding manual. westchester, illinois: American academy of sleep medicine, 2005. broughton r. pathological fragmentary myoclonus, intensified sleep

- starts and hypnagogic foot tremor: three unusual sleep related disorders. *Sleep*, 28:113–21, 2005.
- [13] Nira A Goldstein, Vasanthi Pugazhendhi, Sudha M Rao, Jeremy Weedon, Thomas F Campbell, Andrew C Goldman, J Christopher Post, and Madu Rao. Clinical assessment of pediatric obstructive sleep apnea. *Pediatrics*, 114(1):33–43, 2004.
- [14] James L Goodwin, Sardar I Babar, Kris L Kaemingk, Gerald M Rosen, Wayne J Morgan, Duane L Sherrill, and Stuart F Quan. Symptoms related to sleep-disordered breathing in white and hispanic children—the tucson children’s assessment of sleep apnea study. *CHEST Journal*, 124(1):196–203, 2003.
- [15] James L. Goodwin, Kris L. Kaemingk, Ralph F. Fregosi, Gerald M. Rosen, Wayne J. Morgan, Duane L. Sherrill, and Stuart F. Quan. Clinical outcomes associated with sleep-disordered breathing in caucasian and hispanic children—the tucson children’s assessment of sleep apnea study (tucasa). *Sleep Disordered Breathing*, 26(5):587–591, 2003.
- [16] MR Jarvis and PP Mitra. Apnea patients characterized by 0.02 hz peak in the multitaper spectrogram of electrocardiogram signals. In *Computers in Cardiology 2000*, pages 769–772. IEEE, 2000.
- [17] Jorge Jorquera. Síndrome de apnea obstructiva del sueño. *Boletín de Medicina UC, Pontificia Universidad Católica*, 32:83–88, 2007.
- [18] Ahsan H Khandoker, Marimuthu Palaniswami, and Chandan K Karmakar. Support vector machines for automated recognition of obstructive sleep apnea syndrome from ecg recordings. *Information Technology in Biomedicine, IEEE Transactions on*, 13(1):37–48, 2009.
- [19] P Levy, MR Bonsignore, and J Eckel. Sleep, sleep-disordered breathing and metabolic consequences. *European Respiratory Journal*, 34(1):243–260, 2009.
- [20] Martin O Mendez, Davide D Ruini, Omar Paolo Villantieri, Matteo Matteucci, Thomas Penzel, Sergio Cerutti, and Anna Maria Bianchi. Detection of sleep apnea from surface ecg based on features extracted by an autoregressive model. In

Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE, pages 6105–6108. IEEE, 2007.

- [21] Anke Noehren, Pablo E Brockmann, Michael S Urschitz, Christiane Sokollik, Martin Schlaud, and Christian F Poets. Detection of respiratory events using pulse rate in children with and without obstructive sleep apnea. *Pediatric pulmonology*, 45(5):459–468, 2010.
- [22] F Roche, V Pichot, E Sforza, D Duverney, F Costes, M Garet, JC Barthélémy, et al. Predicting sleep apnoea syndrome from heart period: a time-frequency wavelet analysis. *European Respiratory Journal*, 22(6):937–942, 2003.
- [23] Carol L Rosen, Tonya M Palermo, Emma K Larkin, and Susan Redline. Health-related quality of life and sleep-disordered breathing in children. *Sleep*, 25(6):657–666, 2002.
- [24] Chih-Min Tsai, Chia-Hao Kang, Mao-Chang Su, Hsin-Ching Lin, Eng-Yen Huang, Chih-Cheng Chen, Jui-Chieh Hung, Chen-Kuang Niu, Da-Ling Liao, and Hong-Ren Yu. Usefulness of desaturation index for the assessment of obstructive sleep apnea syndrome in children. *International journal of pediatric otorhinolaryngology*, 2013.
- [25] Rüdiger Wirth and Jochen Hipp. Crisp-dm: Towards a standard process model for data mining. In *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, pages 29–39. Citeseer, 2000.

Appendix

A Classic Models Stage: Results for Support Vector Machine applied to Air Flow signal

Func.	WT	DB	SENS	SPEC	ACC	PREC	RECALL	F
Daubechies 1	db1		75,64%	26,33%	54,20%	57,35%	75,64%	0,64
Daubechies 2	db 2		67,95%	38,33%	55,07%	58,50%	67,95%	0,62
Daubechies 3	db3		68,46%	38,00%	55,22%	58,67%	68,46%	0,62
Daubechies 4	db4		67,44%	37,00%	54,20%	58,64%	67,44%	0,62
Daubechies 5	db5		70,77%	28,67%	52,46%	56,20%	70,77%	0,62
Daubechies 6	db6		75,13%	28,33%	54,78%	58,11%	75,13%	0,65
Daubechies 7	db7		69,74%	36,33%	55,22%	58,98%	69,74%	0,63
Daubechies 8	db8		66,67%	39,33%	54,78%	59,01%	66,67%	0,61
Daubechies 9	db9		75,38%	28,67%	55,07%	57,85%	75,38%	0,65
Daubechies 10	db10		74,87%	25,33%	53,33%	56,46%	74,87%	0,64
Haar	haar		70,77%	32,67%	54,20%	58,31%	70,77%	0,62
Symlet 1	sym1		73,08%	31,33%	54,93%	58,30%	73,08%	0,64
Symlet 2	sym2		73,08%	30,00%	54,35%	57,99%	73,08%	0,64
Symlet 3	sym3		66,15%	40,67%	55,07%	58,79%	66,15%	0,61
Symlet 4	sym4		70,00%	38,67%	56,38%	59,02%	70,00%	0,63
Symlet 5	sym5		67,18%	34,33%	52,90%	57,42%	67,18%	0,61
Symlet 6	sym6		70,00%	32,67%	53,77%	57,25%	70,00%	0,62
Symlet 7	sym7		76,41%	30,67%	56,52%	58,96%	76,41%	0,66
Symlet 8	sym8		75,13%	27,00%	54,20%	57,42%	75,13%	0,64
Symlet 9	sym9		71,54%	33,67%	55,07%	58,77%	71,54%	0,64
Symlet 10	sym10		68,46%	36,67%	54,64%	58,47%	68,46%	0,62

B Classic Models Stage: Results for Neural Network applied to Air Flow signal

Func. WT	DB	SENS	SPEC	ACC	PREC	RECALL	F
Daubechies 1	db1	69,23%	34,67%	54,20%	58,27%	69,23%	0,61
Daubechies 2	db 2	79,74%	29,67%	57,97%	60,68%	79,74%	0,67
Daubechies 3	db3	84,36%	19,00%	55,94%	57,61%	84,36%	0,68
Daubechies 4	db4	80,51%	21,67%	54,93%	57,17%	80,51%	0,66
Daubechies 5	db5	78,72%	28,00%	56,67%	59,15%	78,72%	0,67
Daubechies 6	db6	79,49%	26,33%	56,38%	58,76%	79,49%	0,66
Daubechies 7	db7	74,10%	24,67%	52,61%	55,82%	74,10%	0,62
Daubechies 8	db8	79,49%	20,67%	53,91%	56,58%	79,49%	0,65
Daubechies 9	db9	78,72%	24,33%	55,07%	59,15%	78,72%	0,65
Daubechies 10	db10	79,23%	24,67%	55,51%	58,48%	79,23%	0,65
Haar	haar	80,26%	26,33%	56,81%	59,67%	80,26%	0,67
Symlet 1	sym1	80,26%	24,33%	55,94%	58,19%	80,26%	0,66
Symlet 2	sym2	77,44%	21,67%	53,19%	56,20%	77,44%	0,64
Symlet 3	sym3	73,59%	30,33%	54,78%	56,82%	73,59%	0,63
Symlet 4	sym4	74,10%	26,33%	53,33%	56,91%	74,10%	0,63
Symlet 5	sym5	69,49%	28,33%	51,59%	55,30%	69,49%	0,58
Symlet 6	sym6	69,74%	27,00%	51,16%	55,53%	69,74%	0,59
Symlet 7	sym7	76,92%	28,00%	55,65%	58,35%	76,92%	0,65
Symlet 8	sym8	71,54%	29,00%	53,04%	56,51%	71,54%	0,62
Symlet 9	sym9	75,90%	25,67%	54,06%	57,44%	75,90%	0,64
Symlet 10	sym10	70,77%	31,67%	53,77%	56,98%	70,77%	0,61

C Classic Models Stage: Results for Logit Model applied to Air Flow signal

Func. WT	DB	SENS	SPEC	ACC	PREC	RECALL	F
Daubechies 1	db1	43,36%	62,88%	52,10%	59,29%	43,36%	0,47
Daubechies 2	db 2	28,14%	65,53%	44,68%	49,59%	28,14%	0,31
Daubechies 3	db3	18,89%	74,62%	43,65%	47,89%	18,89%	0,22
Daubechies 4	db4	19,25%	78,89%	44,93%	52,91%	19,25%	0,22
Daubechies 5	db5	25,53%	71,93%	43,04%	52,94%	25,53%	0,27
Daubechies 6	db6	28,37%	72,11%	47,57%	54,95%	28,37%	0,31
Daubechies 7	db7	28,37%	71,56%	48,24%	54,78%	28,37%	0,31
Daubechies 8	db8	25,75%	66,04%	40,90%	52,37%	25,75%	0,29
Daubechies 9	db9	35,54%	68,29%	52,14%	60,05%	35,54%	0,36
Daubechies 10	db10	33,12%	66,28%	48,55%	55,08%	33,12%	0,36
Haar	haar	39,29%	58,13%	46,29%	53,24%	39,29%	0,42
Symlet 1	sym1	35,47%	67,79%	50,21%	57,08%	35,47%	0,39
Symlet 2	sym2	38,64%	66,20%	49,93%	59,96%	38,64%	0,42
Symlet 3	sym3	29,74%	65,57%	44,30%	53,50%	29,74%	0,32
Symlet 4	sym4	18,36%	73,33%	39,93%	49,26%	18,36%	0,21
Symlet 5	sym5	33,09%	72,85%	49,91%	59,76%	33,09%	0,36
Symlet 6	sym6	24,33%	69,80%	43,80%	49,23%	24,33%	0,26
Symlet 7	sym7	19,10%	75,59%	43,38%	50,44%	19,10%	0,22
Symlet 8	sym8	25,72%	74,79%	46,92%	55,07%	25,72%	0,26
Symlet 9	sym9	27,10%	71,74%	43,96%	57,41%	27,10%	0,31
Symlet 10	sym10	24,67%	69,56%	44,45%	51,95%	24,67%	0,28

D Classic Models Stage: Results for Support Vector Machine applied to Electrocardiography signal

Func. WT	DB	SENS	SPEC	ACC	PREC	RECALL	F
Daubechies 1	db1	8,33%	82,05%	50,00%	21,99%	8,33%	0,11
Daubechies 2	db 2	9,33%	78,21%	48,26%	21,24%	9,33%	0,12
Daubechies 3	db3	7,33%	78,46%	47,54%	25,56%	7,33%	0,10
Daubechies 4	db4	7,67%	79,74%	48,41%	18,86%	7,67%	0,10
Daubechies 5	db5	8,33%	86,41%	52,46%	41,45%	8,33%	0,13
Daubechies 6	db6	4,67%	84,62%	49,86%	15,19%	4,67%	0,07
Daubechies 7	db7	6,33%	84,10%	50,29%	16,92%	6,33%	0,08
Daubechies 8	db8	22,00%	86,67%	58,55%	60,40%	22,00%	0,30
Daubechies 9	db9	2,67%	88,21%	51,01%	13,57%	2,67%	0,04
Daubechies 10	db10	4,33%	85,13%	50,00%	22,79%	4,33%	0,06
Haar	haar	10,00%	79,74%	49,42%	32,21%	10,00%	0,14
Symlet 1	sym1	8,67%	84,62%	51,59%	32,35%	8,67%	0,12
Symlet 2	sym2	11,00%	84,87%	52,75%	39,22%	11,00%	0,15
Symlet 3	sym3	5,33%	82,82%	49,13%	17,76%	5,33%	0,07
Symlet 4	sym4	12,00%	83,85%	52,61%	39,19%	12,00%	0,17
Symlet 5	sym5	16,67%	80,26%	52,61%	41,57%	16,67%	0,22
Symlet 6	sym6	11,67%	86,41%	53,91%	41,59%	11,67%	0,17
Symlet 7	sym7	3,33%	87,18%	50,72%	16,10%	3,33%	0,04
Symlet 8	sym8	5,33%	85,13%	50,43%	17,70%	5,33%	0,08
Symlet 9	sym9	11,00%	84,10%	52,32%	33,71%	11,00%	0,14
Symlet 10	sym10	8,67%	80,77%	49,42%	19,53%	8,67%	0,11

E Classic Models Stage: Results for Support Vector Machine applied to Electrocardiography signal

Func. WT	DB	SENS	SPEC	ACC	PREC	RECALL	F
Daubechies 1	db1	8,33%	82,05%	50,00%	21,99%	8,33%	0,11
Daubechies 2	db 2	9,33%	78,21%	48,26%	21,24%	9,33%	0,12
Daubechies 3	db3	7,33%	78,46%	47,54%	25,56%	7,33%	0,10
Daubechies 4	db4	7,67%	79,74%	48,41%	18,86%	7,67%	0,10
Daubechies 5	db5	8,33%	86,41%	52,46%	41,45%	8,33%	0,13
Daubechies 6	db6	4,67%	84,62%	49,86%	15,19%	4,67%	0,07
Daubechies 7	db7	6,33%	84,10%	50,29%	16,92%	6,33%	0,08
Daubechies 8	db8	22,00%	86,67%	58,55%	60,40%	22,00%	0,30
Daubechies 9	db9	2,67%	88,21%	51,01%	13,57%	2,67%	0,04
Daubechies 10	db10	4,33%	85,13%	50,00%	22,79%	4,33%	0,06
Haar	haar	10,00%	79,74%	49,42%	32,21%	10,00%	0,14
Symlet 1	sym1	8,67%	84,62%	51,59%	32,35%	8,67%	0,12
Symlet 2	sym2	11,00%	84,87%	52,75%	39,22%	11,00%	0,15
Symlet 3	sym3	5,33%	82,82%	49,13%	17,76%	5,33%	0,07
Symlet 4	sym4	12,00%	83,85%	52,61%	39,19%	12,00%	0,17
Symlet 5	sym5	16,67%	80,26%	52,61%	41,57%	16,67%	0,22
Symlet 6	sym6	11,67%	86,41%	53,91%	41,59%	11,67%	0,17
Symlet 7	sym7	3,33%	87,18%	50,72%	16,10%	3,33%	0,04
Symlet 8	sym8	5,33%	85,13%	50,43%	17,70%	5,33%	0,08
Symlet 9	sym9	11,00%	84,10%	52,32%	33,71%	11,00%	0,14
Symlet 10	sym10	8,67%	80,77%	49,42%	19,53%	8,67%	0,11

F Classic Models Stage: Results for Neural Network applied to Electrocardiography signal

Func. WT	DB	SENS	SPEC	ACC	PREC	RECALL	F
Daubechies 1	db1	24,00%	73,33%	51,88%	44,10%	24,00%	0,29
Daubechies 2	db 2	35,33%	62,31%	50,58%	41,00%	35,33%	0,37
Daubechies 3	db3	34,33%	60,26%	48,99%	39,38%	34,33%	0,35
Daubechies 4	db4	36,33%	63,85%	51,88%	44,92%	36,33%	0,37
Daubechies 5	db5	28,33%	63,85%	48,41%	36,16%	28,33%	0,30
Daubechies 6	db6	38,00%	67,44%	54,64%	49,47%	38,00%	0,41
Daubechies 7	db7	28,33%	65,13%	49,13%	36,45%	28,33%	0,30
Daubechies 8	db8	31,67%	65,90%	51,01%	43,68%	31,67%	0,34
Daubechies 9	db9	34,33%	64,36%	51,30%	41,08%	34,33%	0,36
Daubechies 10	db10	32,00%	62,05%	48,99%	41,29%	32,00%	0,33
Haar	haar	25,67%	65,90%	48,41%	34,05%	25,67%	0,28
Symlet 1	sym1	28,67%	67,69%	50,72%	38,07%	28,67%	0,31
Symlet 2	sym2	38,67%	57,69%	49,42%	41,50%	38,67%	0,38
Symlet 3	sym3	28,33%	63,85%	48,41%	38,36%	28,33%	0,30
Symlet 4	sym4	34,67%	62,31%	50,29%	39,61%	34,67%	0,36
Symlet 5	sym5	31,67%	66,67%	51,45%	41,81%	31,67%	0,34
Symlet 6	sym6	38,33%	60,51%	50,87%	42,14%	38,33%	0,39
Symlet 7	sym7	33,33%	64,36%	50,87%	39,91%	33,33%	0,35
Symlet 8	sym8	35,33%	62,56%	50,72%	43,35%	35,33%	0,36
Symlet 9	sym9	43,00%	65,13%	55,51%	51,19%	43,00%	0,44
Symlet 10	sym10	26,00%	65,64%	48,41%	39,86%	26,00%	0,29

G Classic Models Stage: Results for Neural Network applied to Electrocardiography signal

Func. WT	DB	SENS	SPEC	ACC	PREC	RECALL	F
Daubechies 1	db1	36,48%	61,44%	49,99%	43,39%	36,48%	0,38
Daubechies 2	db 2	34,24%	52,53%	44,70%	34,48%	34,24%	0,33
Daubechies 3	db3	44,56%	54,54%	50,17%	43,42%	44,56%	0,43
Daubechies 4	db4	41,92%	52,99%	48,21%	40,47%	41,92%	0,41
Daubechies 5	db5	26,55%	54,76%	43,24%	29,68%	26,55%	0,26
Daubechies 6	db6	42,32%	52,44%	48,10%	39,07%	42,32%	0,40
Daubechies 7	db7	49,04%	58,23%	54,19%	49,00%	49,04%	0,48
Daubechies 8	db8	38,30%	55,71%	47,85%	39,89%	38,30%	0,37
Daubechies 9	db9	40,54%	49,95%	45,93%	37,62%	40,54%	0,38
Daubechies 10	db10	38,11%	60,97%	50,81%	43,80%	38,11%	0,40
Haar	haar	43,58%	58,36%	51,81%	45,09%	43,58%	0,42
Symlet 1	sym1	45,60%	47,37%	VM	39,41%	45,60%	0,42
Symlet 2	sym2	45,37%	50,35%	48,18%	41,61%	45,37%	0,43
Symlet 3	sym3	34,86%	54,93%	45,64%	36,35%	34,86%	0,34
Symlet 4	sym4	43,30%	55,66%	50,22%	43,76%	43,30%	0,42
Symlet 5	sym5	47,04%	52,46%	50,10%	43,36%	47,04%	0,43
Symlet 6	sym6	47,96%	50,70%	49,36%	41,20%	47,96%	0,44
Symlet 7	sym7	43,85%	50,99%	47,89%	39,75%	43,85%	0,41
Symlet 8	sym8	47,91%	53,35%	51,03%	45,42%	47,91%	0,45
Symlet 9	sym9	49,63%	55,60%	53,04%	47,15%	49,63%	0,48
Symlet 10	sym10	48,67%	52,05%	50,58%	43,76%	48,67%	0,45

H Classic Models Stage: Results for Support Vector Machine applied to Oximetry signal

Func. WT	DB	SENS	SPEC	ACC	PREC	RECALL	F
Daubechies 1	db1	9,67%	82,56%	50,87%	30,68%	9,67%	0,13
Daubechies 2	db 2	12,33%	78,46%	49,71%	27,08%	12,33%	0,16
Daubechies 3	db3	11,33%	79,23%	49,71%	30,77%	11,33%	0,15
Daubechies 4	db4	27,00%	76,67%	55,07%	45,33%	27,00%	0,32
Daubechies 5	db5	14,33%	77,95%	50,29%	38,30%	14,33%	0,18
Daubechies 6	db6	4,00%	87,18%	51,01%	13,17%	4,00%	0,06
Daubechies 7	db7	21,67%	76,15%	52,46%	44,45%	21,67%	0,27
Daubechies 8	db8	21,33%	78,21%	53,48%	43,81%	21,33%	0,27
Daubechies 9	db9	15,00%	79,74%	51,59%	37,33%	15,00%	0,19
Daubechies 10	db10	6,67%	89,49%	53,48%	31,47%	6,67%	0,10
Haar	haar	13,33%	77,18%	49,42%	31,80%	13,33%	0,17
Symlet 1	sym1	11,33%	76,41%	48,12%	24,25%	11,33%	0,14
Symlet 2	sym2	25,67%	77,18%	54,78%	47,81%	25,67%	0,31
Symlet 3	sym3	18,33%	77,95%	52,03%	43,44%	18,33%	0,23
Symlet 4	sym4	29,33%	73,85%	54,49%	48,05%	29,33%	0,34
Symlet 5	sym5	15,67%	78,46%	51,16%	33,78%	15,67%	0,20
Symlet 6	sym6	12,67%	83,08%	52,46%	35,72%	12,67%	0,18
Symlet 7	sym7	19,67%	78,97%	53,19%	46,72%	19,67%	0,24
Symlet 8	sym8	12,33%	82,05%	51,74%	45,05%	12,33%	0,16
Symlet 9	sym9	17,67%	80,00%	52,90%	37,22%	17,67%	0,22
Symlet 10	sym10	13,00%	83,85%	53,04%	33,91%	13,00%	0,17

I Classic Models Stage: Results for Neural Network applied to Oximetry signal

Func. WT	DB	SENS	SPEC	ACC	PREC	RECALL	F
Daubechies 1	db1	40,67%	58,72%	50,87%	42,84%	40,67%	0,40
Daubechies 2	db 2	18,67%	81,79%	54,35%	36,24%	18,67%	0,21
Daubechies 3	db3	30,00%	75,13%	55,51%	55,95%	30,00%	0,34
Daubechies 4	db4	19,67%	77,69%	52,46%	39,47%	19,67%	0,23
Daubechies 5	db5	32,67%	74,87%	56,52%	53,38%	32,67%	0,37
Daubechies 6	db6	24,00%	71,28%	50,72%	39,81%	24,00%	0,27
Daubechies 7	db7	21,33%	82,05%	55,65%	54,35%	21,33%	0,26
Daubechies 8	db8	21,33%	78,97%	53,91%	44,70%	21,33%	0,26
Daubechies 9	db9	20,33%	78,46%	53,19%	47,70%	20,33%	0,24
Daubechies 10	db10	22,33%	67,44%	47,83%	29,87%	22,33%	0,23
Haar	haar	38,67%	60,00%	50,72%	44,36%	38,67%	0,39
Symlet 1	sym1	41,00%	63,33%	53,62%	45,53%	41,00%	0,42
Symlet 2	sym2	16,33%	87,18%	56,38%	57,47%	16,33%	0,22
Symlet 3	sym3	24,67%	79,23%	55,51%	52,29%	24,67%	0,31
Symlet 4	sym4	18,33%	79,23%	52,75%	36,68%	18,33%	0,21
Symlet 5	sym5	19,67%	83,59%	55,80%	57,20%	19,67%	0,25
Symlet 6	sym6	27,67%	74,87%	54,35%	44,57%	27,67%	0,31
Symlet 7	sym7	15,00%	79,23%	51,30%	46,19%	15,00%	0,18
Symlet 8	sym8	22,33%	78,72%	54,20%	45,05%	22,33%	0,26
Symlet 9	sym9	14,00%	85,38%	54,35%	50,83%	14,00%	0,19
Symlet 10	sym10	22,67%	79,74%	54,93%	43,21%	22,67%	0,26

J Classic Models Stage: Results for Logit Model applied to Oximetry signal

Func. WT	DB	SENS	SPEC	ACC	PREC	RECALL	F
Daubechies 1	db1	48,15%	54,98%	52,23%	44,20%	48,15%	0,45
Daubechies 2	db 2	47,90%	50,16%	49,17%	42,64%	47,90%	0,45
Daubechies 3	db3	44,52%	51,37%	48,41%	41,06%	44,52%	0,42
Daubechies 4	db4	55,11%	61,94%	59,03%	52,53%	55,11%	0,53
Daubechies 5	db5	37,89%	50,96%	45,95%	37,73%	37,89%	0,36
Daubechies 6	db6	48,22%	57,07%	53,18%	45,96%	48,22%	0,46
Daubechies 7	db7	50,93%	53,48%	52,38%	46,37%	50,93%	0,48
Daubechies 8	db8	50,89%	57,52%	54,61%	48,58%	50,89%	0,49
Daubechies 9	db9	45,29%	54,57%	50,48%	42,99%	45,29%	0,42
Daubechies 10	db10	45,96%	51,25%	48,84%	41,34%	45,96%	0,43
Haar	haar	47,00%	55,79%	52,05%	44,53%	47,00%	0,44
Symlet 1	sym1	39,14%	57,18%	49,09%	41,55%	39,14%	0,38
Symlet 2	sym2	53,96%	59,48%	57,08%	52,35%	53,96%	0,52
Symlet 3	sym3	51,81%	50,01%	50,77%	44,62%	51,81%	0,47
Symlet 4	sym4	56,56%	54,99%	55,65%	49,47%	56,56%	0,52
Symlet 5	sym5	53,91%	54,51%	54,37%	47,47%	53,91%	0,49
Symlet 6	sym6	43,11%	53,55%	48,99%	41,52%	43,11%	0,41
Symlet 7	sym7	52,23%	57,03%	55,02%	48,83%	52,23%	0,50
Symlet 8	sym8	38,27%	46,20%	42,15%	35,32%	38,27%	0,36
Symlet 9	sym9	47,67%	61,81%	55,69%	49,71%	47,67%	0,48
Symlet 10	sym10	47,04%	58,95%	53,97%	47,02%	47,04%	0,46