UNIVERSIDADE DE LISBOA

INSTITUTO SUPERIOR TÉCNICO

Signal Processing Approaches for Sleep Quality Analysis in Suspected Sleep Disorder Patients

Fábio Rúben Silva Mendonça

Supervisors: Doctor Fernando Manuel Rosmaninho Morgado Ferrão Dias
Doctor João Paulo Salgado Arriscado Costeira

Co-supervisor: Doctor Antonio Gabriel Ravelo García

Thesis approved in public session to obtain the PhD Degree in

Electrical and Computer Engineering

Jury Final Classification: **Pass with Distinction and Honour**

2021

UNIVERSIDADE DE LISBOA

INSTITUTO SUPERIOR TÉCNICO

Signal Processing Approaches for Sleep Quality Analysis in Suspected Sleep Disorder Patients

Fábio Rúben Silva Mendonça

Supervisors: Doctor Fernando Manuel Rosmaninho Morgado Ferrão Dias
Doctor João Paulo Salgado Arriscado Costeira
Co-supervisor: Doctor Antonio Gabriel Ravelo García

Thesis approved in public session to obtain the PhD Degree in

Electrical and Computer Engineering

Jury Final Classification: **Pass with Distinction and Honour**

Jury

Chairperson: Doctor Isabel Maria Martins Trancoso, Instituto Superior Técnico, Universidade de Lisboa.

Members of the Committee:
Doctor Thomas Walter Friedel Penzel, Interdisciplinary Center of Sleep Medicine, Charité-Universitätsmedizin Berlin, Alemanha.
Doctor Mário Alexandre Teles de Figueiredo, Instituto Superior Técnico, Universidade de Lisboa.
Doctor Jorge Manuel Oliveira Henriques, Faculdade de Ciências e Tecnologia, Universidade de Coimbra.
Doctor Fernando Manuel Rosmaninho Morgado Ferrão Dias, Faculdade de Ciências Exatas e da Engenharia, Universidade da Madeira.
Doctor Ana Luísa Nobre Fred, Instituto Superior Técnico, Universidade de Lisboa.
Doctor Amâncio Lucas de Sousa Pereira, Instituto Superior Técnico, Universidade de Lisboa.

Funding Institutions

2021

I declare that this document is composed of original work with my own authorship, fulfilling the requirements of the Code of Conduct and Good Practices from Universidade de Lisboa.

Abstract

Sleep is part of the circadian rhythm and is characterized by sequences of stages with autonomous nervous system functions that are related to it. It is a complex physiological process inherent to each individual and commonly covers nearly one-third of the lifespan. Sleep quality is one of the most relevant factors that affects physical and mental health and sleep-related complaints are the second most common cause to seek medical attention, superseded only by the feel of pain. Non-restorative sleep is commonly associated with the presence of a sleep-related disorder such as sleep apnea. Full night polysomnography is the standard for sleep studies, requiring monitoring multiple physiological signals that are commonly analyzed by visual examination to score the results. However, this process is slow and expensive, prone to errors due to fatigue of the scorer, and unavailable to a large portion of the world population. Another difficulty in sleep analysis is the lack of a definitional consensus about what is sleep quality while the term is widely used by clinicians, researchers, and the public. It is commonly based on self-rating indexes, the duration of sleep, environmental factors, physiologically derived indices, pharmacologic interventions, polysomnographic parameters, and occurrence of sleep disorders. Nevertheless, correspondence between objective sleep measurements and the person's subjective assessment of the sleep quality is considerably low. To address these constraints, this research focuses on developing automatic algorithms to estimate sleep quality and implement the most suited approaches in cost-effective devices. Therefore, literature reviews were conducted and three approaches were identified as suitable paths to implement the sleep quality examination. The first considered the examination of sleep metrics and the electroencephalogram cyclic alternating pattern related metrics were selected as the most suitable. The second approach was to assess the presence of sleep related disorders, theorizing that the occurrence of such disorders can be a key contributor to poor sleep quality. Therefore, sleep apnea was considered in this work as the most relevant disorder to be addressed since it is one of the most prevalent disorders that is majorly undiagnosed. The last approach was hypothesized in this work and considered that the combination of sleep quality metrics and the detection of sleep related disorders could provide a better estimation of global sleep quality. Several algorithms, based on machine learning for pattern recognition, were developed to estimate the studied metrics, testing multiple sensors to measure the physiological signals. Specifically, models were developed for the examination of the electroencephalogram and electrocardiogram signals, proposing new approaches for sleep quality analysis. The best models were then implemented in the developed home monitoring devices. It was verified that the algorithms' performance is in the same range as the average specialist agreement, indicating that the developed algorithms could be useful for medical diagnosis.

Keywords

Sleep quality; cyclic alternating pattern; sleep apnea; machine learning; home monitoring devices.

Resumo

O sono faz parte do ritmo circadiano e é caracterizado por sequências de estados associados a funções do sistema nervoso autónomo. É um processo fisiológico complexo inerente a cada indivíduo e usualmente cobre aproximadamente um terço da vida. A qualidade do sono é um dos fatores mais relevantes que afetam a saúde física e mental e as queixas relacionadas ao sono são das causas mais comuns que levam à procura de atendimento médico, sendo apenas superadas pela sensação de dor. O sono não restaurador é tipicamente associado à presença de um distúrbio relacionado ao sono, como a apneia do sono. A polissonografia é o exame padrão para estudos do sono, sendo necessária a monitorização de vários sinais fisiológicos que são tipicamente visualmente analisados para avaliar os resultados. No entanto, este processo é lento e dispendioso, propenso a erros, devido à fadiga, e indisponível a uma grande parte da população mundial. Outra dificuldade na análise do sono é a falta de consenso relativo à definição da qualidade do sono, enquanto o termo é amplamente utilizado por médicos, pesquisadores e pelo público. Esta é tipicamente baseada em questionários, quantidade de sono, fatores ambientais, índices derivados de fatores fisiológicos, intervenções farmacológicas, parâmetros polissonográficos e presença de distúrbios do sono. No entanto, a correspondência entre medidas objetivas do sono e a avaliação subjetiva da qualidade do sono é consideravelmente baixa. De forma a obter alternativas para lidar com essas restrições, esta pesquisa concentra-se no desenvolvimento de algoritmos de classificação automática para estimar a qualidade do sono e na implementação das abordagens mais adequadas, em dispositivos económicos. Para tal, revisões de literatura foram realizadas e três abordagens foram identificadas como caminhos adequados para a examinação da qualidade do sono. O primeiro considerou o exame de métricas do sono e as métricas relacionadas com o padrão alternante cíclico de eletroencefalograma foram selecionadas como as melhores para estimar a qualidade do sono. A segunda abordagem considerou a avaliação da presença de distúrbios relacionados com o sono, teorizando que a ocorrência de tais distúrbios pode ser um dos principais contribuintes para a má qualidade do sono. Desta forma, a apneia do sono foi considerada neste trabalho como o distúrbio mais relevante a ser abordado, por ser um dos distúrbios mais prevalentes e que na maior parte dos casos não é diagnosticado. A última abordagem foi proposta neste trabalho e considerada que a combinação de métricas de qualidade do sono e a deteção de distúrbios relacionados ao sono podem fornecer uma melhor estimativa da qualidade global do sono. Vários algoritmos, baseados em aprendizagem de máquina para reconhecimento de padrões, foram desenvolvidos para estimar as métricas estudadas, testando múltiplos sensores para medir os sinais fisiológicos. Especificamente, foram desenvolvidos modelos para o exame dos sinais do eletroencefalograma e eletrocardiograma, propondo novas abordagens para a análise da qualidade do sono. Os melhores modelos foram posteriormente implementados nos dispositivos de monitorização desenvolvidos. Verificou-se que o desempenho dos algoritmos está em linha com o valor médio do acordo entre especialistas, analisando os mesmos dados, indicando que os algoritmos desenvolvidos podem ser úteis para o diagnóstico médico.

Palavras-chave

Qualidade do sono; padrão alternante cíclico; apneia do sono; aprendizagem de máquina; dispositivos de monitoramento doméstico.

To

Ana and my parents

Acknowledgments

Acknowledgment to ARDITI – Agência Regional para o Desenvolvimento da Investigação, Tecnologia e Inovação under the scope of the Project M1420-09-5369-FSE-000001 - PhD Studentship.

Acknowledgment to the Portuguese Foundation for Science and Technology for their support through Projeto Estratégico LA 9 - UID/EEA/50009/2019.

Acknowledgement to the Project MITIExcell co-financed by Regional Development European Funds, for the Operational Programme "Madeira 14-20" – EIXO PRIORITÁRIO 1, of Região Autónoma da Madeira, with no. M1420-01-0145-FEDER-000002.

Contents

List of figures

List of tables

List of relevant acronyms

1D-CNN	One Dimensional Convolutional Neural Network
AASM	American Academy of Sleep Medicine
Acc	Accuracy
AdaBoost	Adaptive Boosting
ADC	Analog to Digital Conversion
AHI	Apnea-Hypopnea Index
AI	Arousal Index
AIC	Akaike Information Criterion
ARM	Autoregressive Model
AUC	Area Under the receiver operating characteristic Curve
Aut	Automatisation
AveP	Average Pooling
AWGN	Additive White Gaussian Noise
BCG	Ballistocardiogram
BIC	Bayesian Information Criterion
BLSTM	Bidirectional Long Short-Term Memory
BS	Bootstrap
CAP	Cyclic Alternating Pattern
CAPSD	Cyclic Alternating Pattern Sleep Database
CART	Classification And Regression Trees
CFNN	Cascade-Forward Neural Network
CNN	Convolutional Neural Network
CO	Combined Objective
CPC	Cardiopulmonary Coupling
CV	Cross-Validation
DC	Direct Current
DET	Percentage of determinism
Dmax	Maximal line length in the diagonal direction
DrNUH	Dr. Negrín University Hospital
DSAE	Deep Stacked Autoencoder
EB	Epoch Based
ECG	Electrocardiogram
EDR	Electrocardiogram Derived Respiratory
EEG	Electroencephalogram
EL	Evidence Level
EM	Expectation-Maximization
EMG	Electromyogram
ENT	Shannon entropy of the frequency distribution of the diagonal line lengths
EOG	Electrooculogram
ET	Ensemble of decision Trees
FFNN	Feed-Forward Neural Network
FN	False Negatives
FP	False Positives
FPGA	Field-Programmable Gate Array
FSM	Finite State Machine
GL	Group of Layers
GELM	Graph regularized Extreme Learning Machine

GMM	Gaussian Mixture Model
GRU	Gated Recurrent Unit
GUI	Graphical User Interface
HF	High Frequency
HL	Hidden Layers
HMD	Home Monitoring Device
HMM	Hidden Markov Model
HOSA	Heuristic Oriented Search Algorithm
HQC	Hannan–Quinn information Criterion
HRV	Heart Rate Variability
kMC	K-Means Clustering
kNN	K-Nearest Neighbor
LDA	Linear Discriminant Analysis
LF	Low Frequency
LOOCV	Leave One Out Cross-Validation
LR	Logistic Regression
LSTM	Long Short-Term Memory
MaxP	Maximum Pooling
MoL	Matrix of Lags
mRMR	Minimal-Redundancy-Maximal-Relevance
MrOSSS	MrOS Sleep Study
MWI	Moving-Window Integration
N-N series	Normal-to-normal sinus intervals
NREM	Non-Rapid-Eye Movement
OSA	Obstructive Sleep Apnea
PCA	Principal Component Analysis
PDF	Probability Density Function
PPG	Photoplethysmography
PSD	Power Spectral Density
PSG	Polysomnography
PSM	Probabilistic Sleep Model
PSQI	Pittsburgh Sleep Quality Index
QDA	Quadratic Discriminant Analysis
QR	Quality Rating
RANSAC	Random Sample Consensus
rec	Recordings
REM	Rapid-Eye Movement
ReLU	Rectified Linear Unit
RF	Random Forest
RNN	Recurrent Neural Networks
RQA	Recurrence Quantification Analysis
RR	Recurrence Rate
SB	Subject Based
SBD	Sleep-related Breathing Disorder
SBS	Sequential Backward Selection
SDPR	Second Degree Polynomial Regression
SE	Sleep Efficiency
SELU	Scaled Exponential Linear Unit

Sen	Sensitivity
SEnt	Spectral Entropy
SFS	Sequential Forward Selection
SFSe	Sequential Feature Selection
SI	Spectrographic Image
SNR	Signal-to-Noise Ratio
SOL	Sleep Onset Latency
SOM	Self-Organizing Map
Spe	Specificity
SpO2	Blood oxygen saturation
SSA	Self-rating Sleep and Awakening quality scale
SVM	Support Vector Machine
sub	Subjects
TDPR	Third Degree Polynomial Regression
TEO	Teager Energy Operator
TFCV	Two Fold Cross-Validation
TN	True Negatives
TND	Trend
TP	True Positives
TST	Total Sleep Time
TW	Time Window
UART	Universal Asynchronous Receiver/Transmitter
UCDSAD	University College Dublin Sleep Apnea Database
VHDL	Very high speed integrated circuits Hardware Description Language
VLF	Very Low Frequency
WASO	Wake After Sleep Onset

1. Introduction

1.1. Motivation

1.2. Objectives

1.3. Research questions

1.4. Main contributions

1.5. Collaborations

1.6. Outline

Sleep is a complex physiological process that is essential for each individual, covering approximately one third of the lifespan. The daily wear of the body systems, such as the respiratory, the circulatory, the central nervous system, and the musculoskeletal is repaired during sleep [1]. Thus, a continuous deficiency of this process increases the risk of cardiovascular pathologies, obesity, hypertension, diabetes, and metabolic deregulation, which can lead to a decline in the immune system efficiency [2]. Sleep processes are also involved in the physical development, consolidation of memories, learning, and emotion regulation [3]. It is also directly related to the quality of life concept since physical health was considered, by the World Health Organization, as one of the four domains that define this concept and sleep-related complaints are the second most frequent reason for pursuing medical attention, only superseded by the feel of pain [4].

It was estimated that a reduction in the workplace productivity, associated with poor sleep, will lead to annual economic losses, by the year 2020, in the range of billions of dollars in the United States of America ($299 to $433 billion), Japan ($94 to $146 billion), United Kingdom ($40 to $54 billion), Germany ($41 to $62 billion), and Canada ($14 to $22 billion), with an increase in the economic losses in the following years [5]. Therefore, improving the quality of sleep is not only relevant for the person's quality of life but also for society. An increment in the prevalence of sleep disturbances is also expected with the growth of the elderly population. Therefore, it is predictable that sleep quality will become a major medical diagnosis element since nearly half of the older adults report poor quality of sleep with a lower prevalence in healthy adults [6]. Also, women are more likely to report poor sleep quality as the age progresses [7].

However, there is a lack of definitional consensus regarding what is sleep quality while it is widely used by researchers, clinicians, and the public. This is most likely due to the uncertainty associated with the definition of what quality is [8]. Two approaches have been proposed to address this issue. The first is the most used and employs self-rating indexes, which reflect the individual satisfaction with sleep, producing a subjective measurement. It is considered the most economical and less intrusive approach. However, during sleep the subject is in a state of loss of consciousness, making the individual a poor self-observer of the behaviors; thus, the accuracy of this approach is subject to the individual's recall. Polysomnography (PSG) based metrics are used by the second approach, defining sleep quality with objective measures that are not biased by the personal sleep experience [9] and can be grouped into four domains [10]: duration; continuity; intensity; stability. It was verified that the most commonly employed metrics to classify the sleep quality (based on continuity, duration or intensity of sleep) have a low correlation with the subjective appraisal of sleep [11]. Thus, stability metrics could be the most relevant for medical diagnosis [12] and the electroencephalogram (EEG) Cyclic Alternating Pattern (CAP) metrics are particularly relevant as some are characterized by a low night-to-night intra-individual variability with defined age-related percentages, and has characteristic behaviors associated with specific sleep disorders [10].

The quality of sleep can also be significantly affected by the presence of sleep-related disorders. More than sixty have been acknowledged by the International Classification of Sleep Disorders, divided into seven categories [13]: insomnia; Sleep-related Breathing Disorders (SBD); central disorders of hypersomnolence; circadian rhythm sleep-wake

disorders; parasomnias; sleep-related movement disorders; other sleep disorders. The second category groups the most prevalent disorders that are usually undiagnosed and considers: Obstructive Sleep Apnea (OSA); central sleep apnea; sleep-related hypoxemia disorder; sleep-related hypoventilation disorders. OSA is the most common in this group, and it was estimated, in population-based studies, to have a prevalence of 10% and 3% for 30-49 year-olds males and females, respectively, increasing to 17% and 9% for 50-70 year-olds, respectively [14]. However, most subjects are typically undiagnosed due to the person's neglect and the lack of availability to perform a PSG, the standard exam for sleep analysis [15].

PSG measures multiple sensors to record, at least, the respiratory and body movements, the breath airflow, blood oxygen saturation (SpO2), EEG, electromyogram (EMG), electrooculogram (EOG) and electrocardiogram (ECG). Thus, a detailed diagnostic can be achieved by analyzing these signals [16]. Nonetheless, the process is expensive and slow since it frequently involves the supervision of a specialized technician to monitor the patient attendance at a sleep laboratory and the manual scoring of the recorded signals to generate the clinical reports. This process is unavailable to a large part of the world population, and the employment of all the necessary sensors results in an uncomfortable experience that can significantly affect the results, an issue known as the first night effect [17].

Home Monitoring Devices (HMD) are increasing their relevance in the new health care perspective, which is changing the focus from primary and specialty care to wellness and prevention. They can assist in the monitoring and detection of pathologies with significantly lower cost, providing personalized health data with less disturbance, and increasing the effectiveness of behavior change interventions [18]. Therefore, the HMD can be seen as an alternative to PSG to determine the presence of sleep disorders.

1.1. Motivation

Several HMDs have been developed and can be found in the state of the art, to analyze sleep disorders, using fewer sensors than a PSG and usually employing automatic algorithms to evaluate the signals. However, the achieved performance is commonly not enough to produce a clinical diagnosis and a compromise between the complexity of the device (with a certain number of considered sensors) and the performance detecting sleep breathing events as been established. Taking into consideration that in most cases, poor quality of sleep is directly associated with the occurrence of a sleep related disorder [13], a possible way to improve the accuracy of the disorder diagnosis is to introduce a sleep quality appraisal into the analysis.

The CAP rate is a prime candidate for the sleep quality metric since it takes into consideration the age [11] and presents characteristic behaviors related to the presence of sleep disorders. This proposal takes into consideration that sleep is a multifaceted construct that is difficult to characterize by considering a single measure [12]. However, there is a significant gap in the literature regarding the implementation of objective sleep quality metrics in HMDs. Moreover, it is not possible to address all sleep-related disorders that have been identified without using a significant number of sensors, jeopardizing the

utility of HMD in relation to PSG. Therefore, a compromise between the desired capacity of the HMD and the number of sensors should be taken into consideration.

1.2. Objectives

The development of methods for sleep quality estimation is the main objective of this thesis. OSA was selected as the sleep disorder to be investigated since it is considered to be the most prevalent and undiagnosed disorder that is strongly related to CAP. All the algorithms were tested on the same database (the CAP sleep Database from PhysioNet [19] for CAP analysis or a database from Hospital Universitario de Gran Canaria Dr. Negrín for OSA examination) to allow a fair comparison between the achieved results. In the end, several HMDs were implemented to address each of the three approaches identified in the state of the art as suitable to perform the sleep quality examination (the first considered the examination of sleep metrics, the second assess the presence of a sleep related disorder, and the considered a combination of sleep quality metrics and the detection of sleep related disorders). The main objectives can be summarized as:

- Review the state of the art for both methods and devices that can assess the sleep quality and the presence of OSA.
- Develop methods to examine the CAP from one EEG monopolar derivation signal.
- Produce a new sleep model for the sleep microstructure analysis.
- Develop methods to indirectly estimate CAP using the signal from a single-lead ECG.
- Develop methods for OSA estimation, one based on the SpO2 signal and the other based on the ECG signal.
- Propose a sleep quality estimation model based on the ECG signal.
- Create multiple prototypes of HMD that can implement the developed algorithm.

1.3. Research questions

A group of research questions were developed to address the identified constraints:

- Are the self-rating indexes of sleep quality the best way to estimate the quality of sleep or should objective measures, such as CAP rate, be used?
- Can CAP be reliably assessed by analyzing the signal from one EEG monopolar derivation?
- Can the sleep quality be assessed by considering an indirect estimation of the CAP rate?
- Can a cost-effective home monitoring device be developed to perform both the estimation of the sleep quality and the presence of a sleep disorder?

The chosen path to provide the answer to the questions is divided into six stages and an overview of the followed path with key findings is presented in Figure 1.1.

- First stage:
    - Reviewed the sleep quality metrics presented in the state of the art.

    - Reviewed the developed methods and devices to determine sleep quality metrics and OSA presented in the state of the art.
- Second stage:
    - Reviewed the features and classifiers of the state of the art methods to perform the detection of the CAP cycles and activation phases.
    - Developed models (based on handcrafted features and methods without an explicit feature extraction process) for CAP estimation, from the signal of one EEG monopolar derivation, and sleep quality assessment.
    - Developed a model for sleep analysis, based on the sleep microstructure, using the signal of one EEG monopolar derivation.
    - Developed models (based on handcrafted features and methods without an explicit feature extraction process) for the A phase subtype estimation from the signal of one EEG monopolar derivation, and performed a characterization analysis for the A phase subtypes.
- Third stage:
    - Developed algorithms to indirectly determine the CAP rate and the quality of sleep using the signal from a single-lead ECG.
    - Developed a tool for time series analysis, suitable for the indirect estimation of the CAP using the signal from a single-lead ECG.
- Fourth stage:
    - Developed algorithms to estimate the OSA events and AHI, either from the SpO2 signal or from the ECG signal.
- Fifth stage:
    - Proposed a sleep quality metric based on the examination of the single-lead ECG signal.
    - Developed a sleep quality model that analyzes the single-lead ECG signal.
- Sixth stage:
    - Created two HMD to perform the OSA assessment, one intended for personal use and one intended for clinical use.
    - Created two HMD that can perform the estimation of the quality of sleep, the first examined the EEG signal while the second evaluated the ECG signal.
    - Create one HMD that can estimate both the quality of sleep and the AHI.

1.4. Main contributions

The developed work was published in the following journals and international conferences:

- **Journals:**
    - Mendonça, F., Mostafa, S., Ravelo-García, A., Morgado-Dias, F., and Penzel, T. (2018). Devices for Home Detection of Obstructive Sleep Apnea: A Review. Sleep Medicine Reviews 41: 149-160. (IF=10.517, Q1) https://doi.org/10.1016/j.smrv.2018.02.004.
    - Mendonça, F., Fred, A., Mostafa, S., Morgado-Dias, F., and Ravelo-García, A. (2018). Automatic detection of cyclic alternating pattern.

Neural Computing and Applications. (IF=4.664, Q1) https://doi.org/10.1007/s00521-018-3474-5.

- Mendonça, F., Mostafa, S., Morgado-Dias, F., Navarro-Mesa, J., Julia-Serda, G., and Ravelo-Garcia, A. (2018). A portable wireless device based on oximetry for sleep apnea detection. Computing 100(11): 1203-1219. (IF=2.063, Q2) https://doi.org/10.1007/s00607-018-0624-7.
- Mendonça, F., Mostafa, S., Morgado-Dias, F., and Ravelo-Garcia, A. (2018). Sleep quality estimation by cardiopulmonary coupling analysis. IEEE Transactions on Neural Systems and Rehabilitation Engineering 26(12): 2233-2239. (IF=3.478, Q1) https://doi.org/10.1109/TNSRE.2018.2881361.
- Mendonça, F., Mostafa, S., Ravelo-García, A., Morgado-Dias, F., and Penzel, T. (2019). A Review of Obstructive Sleep Apnea Detection Approaches. IEEE Journal of Biomedical and Health Informatics 23(2): 825-837. (IF=4.217, Q1) https://doi.org/ 10.1109/JBHI.2018.2823265;
- Mendonça, F., Mostafa, S., Ravelo-García, A., Morgado-Dias, F., and Penzel, T. (2019). A Review of Approaches for Sleep Quality Analysis. IEEE Access 7: 24527-24546. (IF=4.098, Q1) https://doi.org/10.1109/ACCESS.2019.2900345.
- Mendonça, F., Mostafa, S., Morgado-Dias, F., Ravelo-García, A., and Penzel, T. (2019). Sleep quality of subjects with and without sleep-disordered breathing based on the cyclic alternating pattern rate estimation from single-lead ECG. Physiological Measurement 40(10): 1-14. (IF=2.246, Q3) https://doi.org/10.1088/1361-6579/ab4f08.
- Mendonça, F., Mostafa, S., Morgado-Dias, F., and Ravelo-García, A. (2019). A Portable Wireless Device for Cyclic Alternating Pattern Estimation from an EEG Monopolar Derivation. Entropy 21(12). (IF=2.419, Q2) https://doi.org/10.3390/e21121203.
- Mendonça, F., Mostafa, S., Morgado-Dias, F., and Ravelo-García, A. (2020). Matrix of Lags: a tool for Analysis of Multiple Dependent Time Series applied for CAP scoring. Computer Methods and Programs in Biomedicine 189(1). (IF=3.424, Q1) https://doi.org/10.1016/j.cmpb.2020.105314.
- Mendonça, F., Mostafa, S., Morgado-Dias, F., and Ravelo-García, A. (2020). An Oximetry Based Wireless Device for Sleep Apnea Detection. Sensors 20(3). (IF=3.031, Q1) https://doi.org/10.3390/s20030888.
- Mendonça, F., Mostafa, S., Morgado-Dias, F., Juliá-Serdá, G., and Ravelo-García, A. (2020). A Method for Sleep Quality Analysis Based on CNN Ensemble With Implementation in a Portable Wireless Device. IEEE Access 8(1): 158523-158537. (IF=4.098, Q1) https://doi.org/10.1109/ACCESS.2020.3019734.
- Mendonça, F., Mostafa, S., Morgado-Dias, F., and Ravelo-García, A. (2020). Cyclic alternating pattern estimation based on a probabilistic model over an EEG signal. Biomedical Signal Processing and Control 62(1). (IF=3.137, Q2) https://doi.org/10.1016/j.bspc.2020.102063.
- Mendonça, F., Mostafa, S., Morgado-Dias, F., and Ravelo-García, A. (2020). On the use of patterns obtained from LSTM and feature-based

methods for time series analysis: application in automatic classification of the CAP A phase subtypes. Journal of Neural Engineering. (IF=4.141, Q1) https://doi.org/10.1088/1741-2552/abd047.
    - Mendonça, F., Mostafa, S., Morgado-Dias, F., and Ravelo-García, A. (2021). A Method based on Cardiopulmonary Coupling Analysis for Sleep Quality Assessment with FPGA Implementation. Artificial Intelligence in Medicine. (IF=4.383, Q1) https://doi.org/10.1016/j.artmed.2021.102019.
- **Conferences:**
    - Mendonça, F., Mostafa, S., Morgado-Dias, F., Navarro-Mesa, J., Julia-Serda, G., and Ravelo-Garcia, A. (2017). A minimally invasive portable system for sleep apnea detection. Bioinspired Intelligence (IWOBI), 2017 International Conference and Workshop on, IEEE. https://doi.org/10.1109/IWOBI.2017.7985540.
    - Mendonça, F., Fred, A., Mostafa, S., Morgado-Dias, F., and Ravelo-García, A. (2018). Automatic Detection of a Phases for CAP Classification. Pattern Recognition Applications and Methods (ICPRAM), 7th International Conference on, SCITEPRESS. https://doi.org/10.5220/0006595103940400.
    - Mendonça, F., Mostafa, S., Morgado-Dias, F., and Ravelo-Garcia, A. (2018). Sleep Quality Analysis with Cardiopulmonary Coupling. Biomedical Engineering and Applications (ICBEA), 2018 International Conference on, IEEE. https://doi.org/10.1109/ICBEA.2018.8471727.
    - Mendonça, F., Mostafa, S., Morgado-Dias, F., and Ravelo-Garcia, A. (2019). Cyclic Alternating Pattern Estimation from One EEG Monopolar Derivation Using a Long Short-Term Memory. Engineering Applications (ICEA), 2019 International Conference on, IEEE. https://doi.org/10.1109/CEAP.2019.8883470.

The following awards were received:

- Best student paper award in the 2017 International Conference and Workshop on Bioinspired Intelligence (IWOBI).
- 2017 Cátedra Telefónica da Universidad de Las Palmas de Gran Canária.

1.5. Collaborations

This work involved the collaboration of one hospital, Hospital Universitario de Gran Canaria Dr. Negrín, and four universities: Charité Universitatsmedizin; Instituto Superior Técnico – Universidade de Lisboa; Universidad de Las Palmas de Gran Canaria; Universidade da Madeira.

1.6. Outline

The outline of this thesis is the following:

- **Chapter 2** presents the physiological contextualization of the sleep process.
- **Chapter 3** reviews the state of the art regarding the developed methods and HMD for sleep quality assessment and OSA estimation.

- **Chapter 4** presents the materials employed in this research.
- **Chapter 5** examines the proposed algorithms for CAP estimation using the signal from one EEG monopolar derivation.
- **Chapter 6** analyzes the developed methods for indirect CAP rate detection from a single-lead ECG signal.
- **Chapter 7** studies the proposed methods for OSA and AHI estimation.
- **Chapter 8** presents the new sleep quality metric and model.
- **Chapter 9** presents the implemented HMDs.
- **Chapter 10** concludes this report, providing an overview of the developed work and targets for future work.

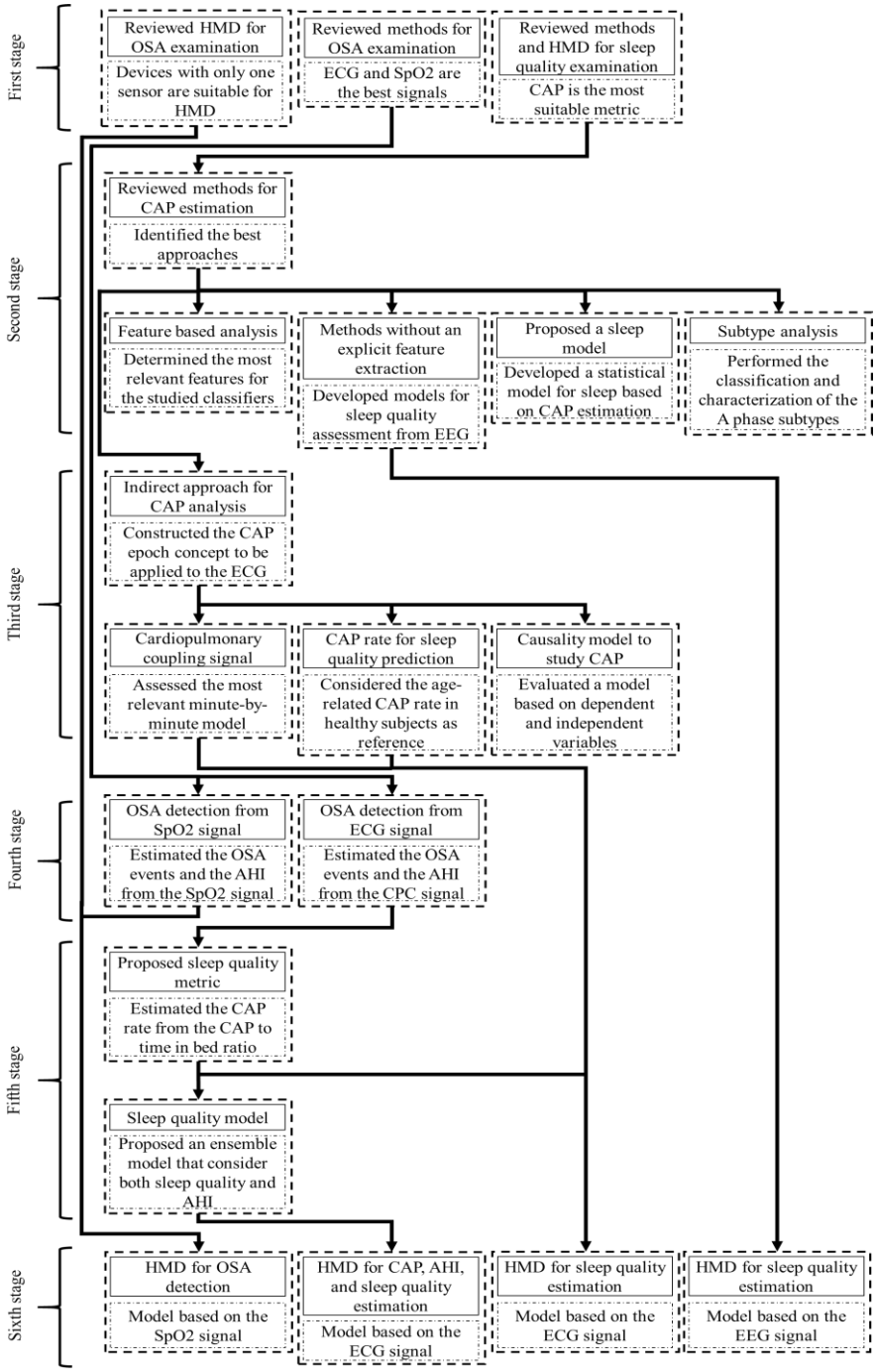


Figure 1.1. Overview of the followed path.

2. Physiological contextualization

2.1. Sleep quality contextualization

2.1.1. Sleep analysis

2.1.2. CAP definition

2.2. Contextualization for the indirect CAP assessment

2.3. OSA contextualization

2.4. Key remarks

A physiological contextualization was carried out to describe the sleep processing associated with CAP and the OSA from the physiological/clinical perspective, with the goal of further justifying the followed approaches in this work.

2.1. Sleep quality contextualization

2.1.1. Sleep analysis

Multiple imaging techniques have been developed to analyze the human body, working as auxiliary diagnosis elements. The electroencephalography belongs to the electrobiological measurements group and is based on the measure of the electrical activity produced by the brain. EEG is one of the most employed techniques in this field, recording the alternating electrical activity at the scalp surface using electrodes and conductive media [20]. The electrodes distribution frequently follows the 10-20 electrode placement standardization presented in Figure 2.1. The electrodes cover the frontal (F), temporal (T), parietal (P), and occipital (O) regions, the central (C), and ears (A) lactations are also marked. The electrodes positions are referenced by odd numbers on the left side, even numbers on the right side, and a Z in the middle line [21].

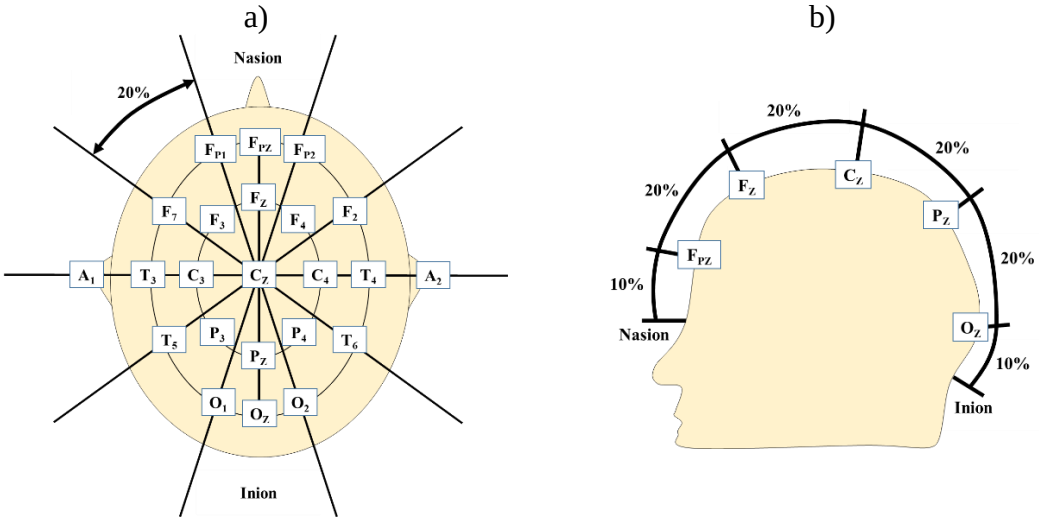


Figure 2.1. View of the 10-20 electrode placement standardization form a) the top and b) from the side [22].

The American Academy of Sleep Medicine (AASM) provided the guidelines for sleep analysis [21]. The considered architecture is composed of a macrostructure and a microstructure defined by the EEG signals that are commonly segmented into four frequency bands [23]: delta (0.5–4 Hz); theta (4–8 Hz); alpha (8–13 Hz); beta (13–30 Hz). The macrostructure is composed of alternating cycles of Rapid-Eye Movement (REM) and Non-REM (NREM) sleep (each sleep cycle usually lasts approximately 90 minutes). The NREM is composed of three stages (N1, N2, and N3). Therefore, the sleep period was divided into five phases (Wake, N1, N2, N3, and REM), considering 30 s as the standard epoch duration for classification [21]. Each stage is characterized by a predominance of energy in specific frequency bands and characteristic shapes of the EEG waves such as [23]:

- Wake: alpha and beta bands.

- N1: theta band.
- N2: theta band with the presence of k-complexes and sleep spindles.
- N3: delta band.
- REM: theta, alpha, and beta bands with the presence of sawtooth waves.

Typically, the sleep stages are scored by visual examination, requiring the analysis of an expert for multiple hours. It is a slow and expensive process that is prone to errors, due to fatigue of the scoring expert, with an average agreement among experts scoring the same signals for the total duration of the sleep macrostructure analysis, lower than 90%, [24]. Multiple approaches were proposed in the state of the art to address these issues by performing an automatic sleep staging [25]. Another relevant aspect is that the quality of sleep can be assessed by evaluating the macrostructure's based metrics recommended by the AASM manual for scoring sleep and associated events [21]. Conversely, some subjects have sleep related complaints while having comparable values for the metrics as those seen in non-complaining individuals [12]. Therefore, the basis for such sleep complaints may not be related to the architecture, timing, or amount of sleep but rather to differences in the sleep experience itself. As a result, it was conceptualized that stability analysis could conceivably offer a more comprehensive and sensitive measure for the quality of sleep. It is also relevant to notice that arousals are also associated with the restorative properties of sleep [10].

2.1.2. CAP definition

The regular organization of arousals is known as CAP and describes the microstructure of sleep while measuring the amount of unstable sleep. The microstructure is composed of transient and phasic events that have a shorter duration than the conventional scoring epoch. These events appear as abrupt frequency shifts and/or amplitude changes that can clearly be distinguishable from the background activity [19].

The periodic activities can be grouped by considering three parameters: a repetitive activity that is distinguishable from the background and constitutes an activation phase (A phase); return to background activity (B phases); recurrence rate (total duration of the A phase and the subsequent B phase) [19]. In the NREM sleep, this information can be combined to define the CAP, a cyclic activity in which both phases (A and B) range between 2 and 60 s, and each cycle is composed of the A phase and the subsequent B phase. Figure 2.2 presents an example of a CAP sequence that is created when more than two consecutive cycles occurred. A period is considered to be non-CAP if there is a lack of a CAP cycle for more than 60 s. The phasic activities which initiate an A phase must have an amplitude of at least one third higher than the background voltage. The A phase waveforms includes [19]:

- Vertex sharp transients: EEG potentials which last between 50 and 200 ms and present variable amplitude, with a maximum of 250 μV.
- K-complex sequences with or without spindles: a sequence of at least two consecutive k-complexes, ranging from 0.5 to 2 s, which can be mixed with or followed by a sleep spindle, and each complex presents a bi/tri-phasic pattern comprising an initial rapid negative component which is followed by a slower positive wave.

- K-alpha: K-complex followed directly by an alpha burst.
- Delta bursts: a sequence of at least two waves, in the delta band, whose amplitude is at least one third higher than the background activity.
- Intermittent alpha: a sequence of activity in the alpha band.
- Polyphasic bursts: clusters of high-voltage delta waves which are mixed with theta, alpha or beta rhythms.
- EEG Arousals: abrupt frequency shifts toward faster rhythms (but not spindles), interrupting sleep continuity for at least 3 s.

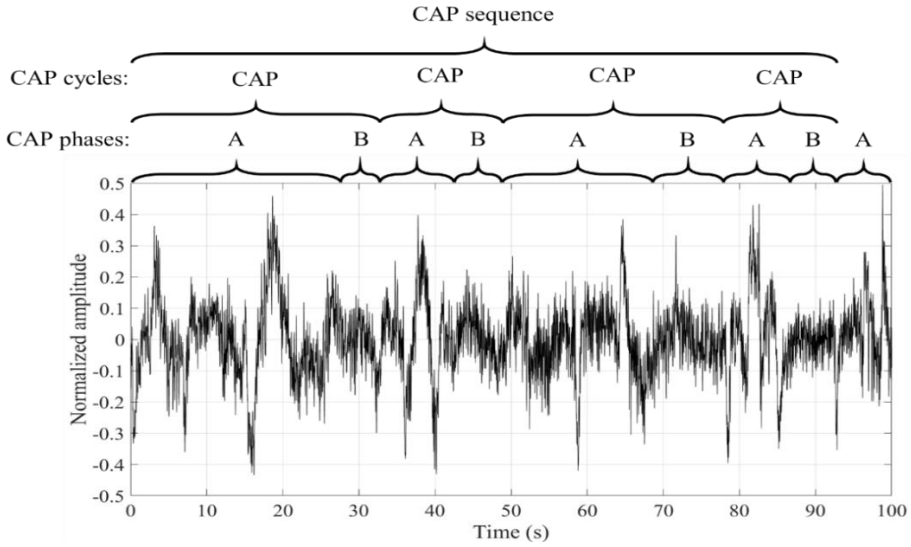


Figure 2.2. Example of a CAP sequence, composed of multiple CAP cycles, which are composed of CAP phases [26].

The CAP can be linked to sleep instability and/or sleep disturbance. This pattern can occur spontaneously in NREM sleep or in association with an identifiable sleep pathophysiology such as sleep-disordered breathing. Therefore, the CAP concept comprises both the process of sleep fragmentation and maintenance [19]. An alternative view to the arousal process was provided by the three A phase subtypes (A1, A2, and A3), characterized by specific spectral and amplitude contents of the EEG signal [27] [28]:

- A1: Associated with minor or mild polygraphic variations and characterized by high-voltage slow waves (known as synchronized EEG patterns), such as intermittent alpha rhythm in the N1 sleep stage, and sequences of K-complexes or delta bursts in the N2 and N3 sleep stages. Low-amplitude fast rhythms (known as desynchronized EEG pattern) can occur in this subtype but must account for less than 20% of the whole activation time.
- A2: Related to an increase of muscle tone and/or increase of cardiorespiratory rate phases, with desynchronized EEG patterns which are either mixed with or are preceded by slow high-voltage waves (such as K-complexes with alpha and beta activities, and arousals with slow wave synchronization). 20% to 50% of the total activation time must have rapid activities.
- A3: Linked with the existence of desynchronized EEG patterns whose duration surpasses two thirds of the total activation duration. These patterns are coupled

with a significant increase in muscle tone and/or increase in cardiorespiratory rate phases. As a result, more than the 50% of the total activation period has rapid activities (especially in the beta band).

High-amplitude EEG bursts (encompassed by subtypes A1) are associated with sleep instability, marking the brain's attempt to maintain sleep. Conversely, if the preservation attempt fails or if sleep becomes too unstable then, an EEG arousal will either replace or accompany the high-amplitude slow activity. Therefore, subtypes A2 and A3 comprise central nervous system arousal. It was theorized that EEG synchrony is build-up and preserved by a fluctuating process of slow activities associated with the A1 subtype, which are combined with a powerful inhibition of rapid EEG shifts, linked to the A2 and A3 subtypes. It was also observed that the A1 subtype is the most prevalent through the sleep cycles in the descending branch (92% of the A phases are related to this subtype) while the A2 and A3 subtypes are more frequent in the ascending branch (45% and 19% of the A phases are related to the A2 and A3 subtypes, respectively) [29]. It is also relevant to notice that different A phase subtypes can occur within the same CAP sequence, and a CAP cycle can spawn between two adjacent sleep stages [19]. An example of the CAP phases is presented in Figure 2.3.

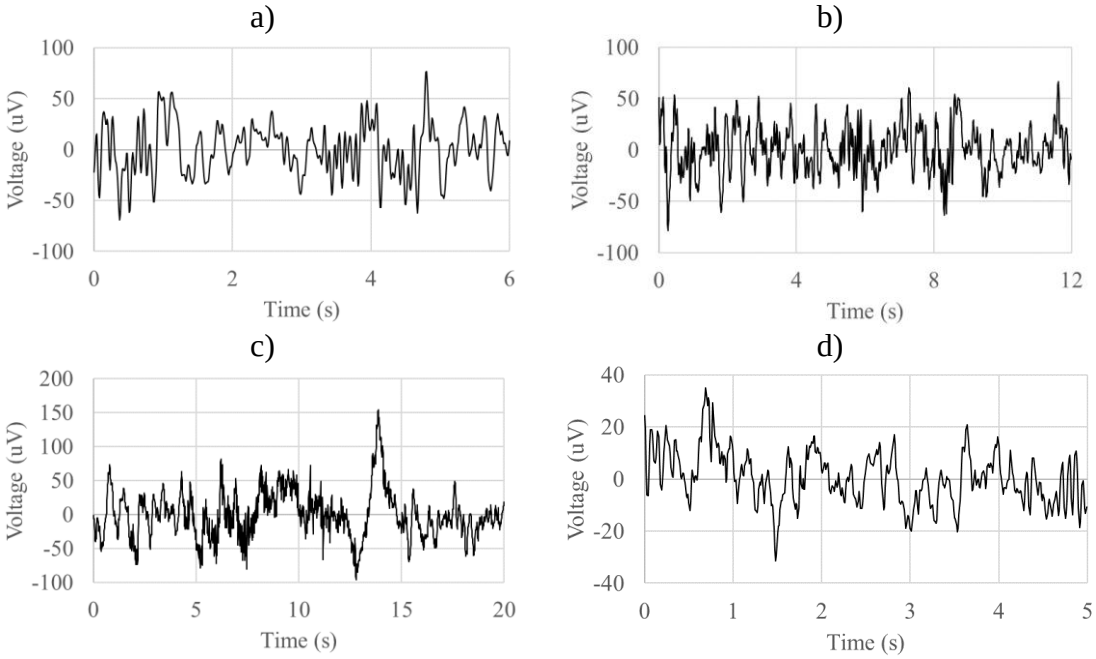


Figure 2.3. Example of CAP a) A phase subtype A1, b) A phase subtype A2, c) A phase subtype A3 and d) B phase [30].

It was conceptualized that the regulatory mechanisms underlying the CAP recurrent fluctuations can possibly turn into the pathophysiological source for disordered sleep [31]. Hence, changes in the A phase subtypes can be an indicator of the manifestation of sleep related disorders such as: periodic limb movements [32]; sleep apnea [33]; narcolepsy [34]; insomnia [35]; nocturnal frontal lobe epilepsy [36].

CAP was found to be a prime indicator of sleep's stability, and was considered to be a sleep quality marker [37]. In healthy adults, a CAP sequence is expected to last, on average, 2 min and 33 s, comprising 5.6 CAP cycles. The mean duration of a CAP cycle is 26.9 ± 4.1 s [38]. A disturbance in sleep produces alterations in the CAP rate, defined

as the ratio of the total CAP time to the total NREM time. The CAP rate was found to be the most exploited microstructural parameter for clinical purposes, and it is possibly the most relevant metric to define sleep quality based on the sleep stability [37]. It is characterized by a low night-to-night intra-individual variability (in normal subjects) with defined age-related percentages [10], thus, it can take into consideration the person's age to define what can be considered as the normal value for this matric [11]. The CAP rate increases when the sleep process is disturbed by external or internal factors, and its variations reflect the perception of sleep quality. A poorer sleep quality is related to higher values of this ratio and, when considering the normal age-related values for the CAP rate [10] [37], it is possible to assess the quality of sleep.

It was identified that the A phase subtypes prevalence is correlated with age [27]. For healthy subjects, it was observed that the A1 subtype is the most prevalent, while A3 is the less predominant. Nevertheless, the incidence of A1 decreases as a person progresses to an older age, while A3 increases. The normal age-related values for the A1, A2, and A3 subtypes are [27]: 71%, 20%, and 9% for adolescence, respectively; 61%, 28%, and 11% for young adults, respectively; 62%, 27%, and 11% for mature adults, respectively; 47%, 35%, and 18% for senescence, respectively. As a result, through the person's development, the CAP rate undergoes through complex variations and the normal age-related values are [10]: 12.9% for infants; 25.9% for preschool-aged children; 33.4% for school-aged children; 62.1% for peripubertal children; 43.4% for teenagers; 31.9% for young adults; 37.5% for middle aged subjects; 55.3% for elderly persons. By considering the CAP rate characteristics and the regression analysis presented by Parrino et al. [39] (fitted a curve with the average value of CAP rate for the person's age), it was conceptualized that the quality of sleep can be assessed by comparing the estimated CAP rate with the CAP rate in normal sleep (according to the value for the subjects age). Therefore, if the predicted CAP rate was higher than the expected value then the quality of sleep was considered poor, otherwise, it was considered good.

2.2. Contextualization for the indirect CAP assessment

It was verified that sleep instability has manifestations in multiple physiological signals, with repercussions for vegetative and motor activities, which fluctuate during CAP while remaining quiescent during the absence of CAP. The heart rate is controlled by efferent sympathetic and vagal activities which are directed to the sinus node, and these are modulated by central brainstem and peripheral oscillators [40]. Therefore, it was observed that sleep instabilities have an impact on cardiac autonomic regulation, which can be evaluated by spectral analysis of Heart Rate Variability (HRV). In this spectral evaluation, vagal activity was identified as the major contributor to the high-frequency (0.15–0.4 Hz) component, while sympathetic influences (possibly with vagal influences) where associated with the low-frequency (0.04–0.15 Hz) component [40]. These observations lead to the proposal of an extended concept for CAP where it can be viewed in a broader sense, designated as the stability of sleep [41], and the Cardiopulmonary Coupling (CPC) analysis was proposed.

CPC measures the degree of coherent coupling between variations of the R-wave amplitude, produced by modulations of the respiratory tidal volume, which is known as ECG Derived Respiratory (EDR), and corrected HRV, denoted normal-to-normal sinus

intervals (N-N series) [40]. Two key properties are examined by CPC, specifically, if the evaluated signals (EDR and N-N series) are coupled and synchronized (property assessed by the coherence, requesting that the phases of both signals are aligned with the aim of keeping the phase relationship constant), and if both signals oscillate at a similar frequency (property evaluated by cross-spectral analysis) [41]. It was observed that High Frequency (HF) coupling (0.1–0.4 Hz) is linked to periods of stable sleep (absence of CAP) as the vagal modulations induced by respiration, denoted as respiratory sinus arrhythmia, prevail in the cardiac autonomic regulation. On the other hand, Low Frequency (LF) coupling (0.01–0.1 Hz) is related to periods of unstable sleep (presence of CAP). Very Low Frequency (VLF) coupling (0–0.01 Hz) was found to be associated with wake or REM sleep periods. As a result, it was possible to view sleep as an oscillation between stable, unstable, and wake or REM periods [40] [41]. An example of CPC analysis is presented in Figure 2.4.

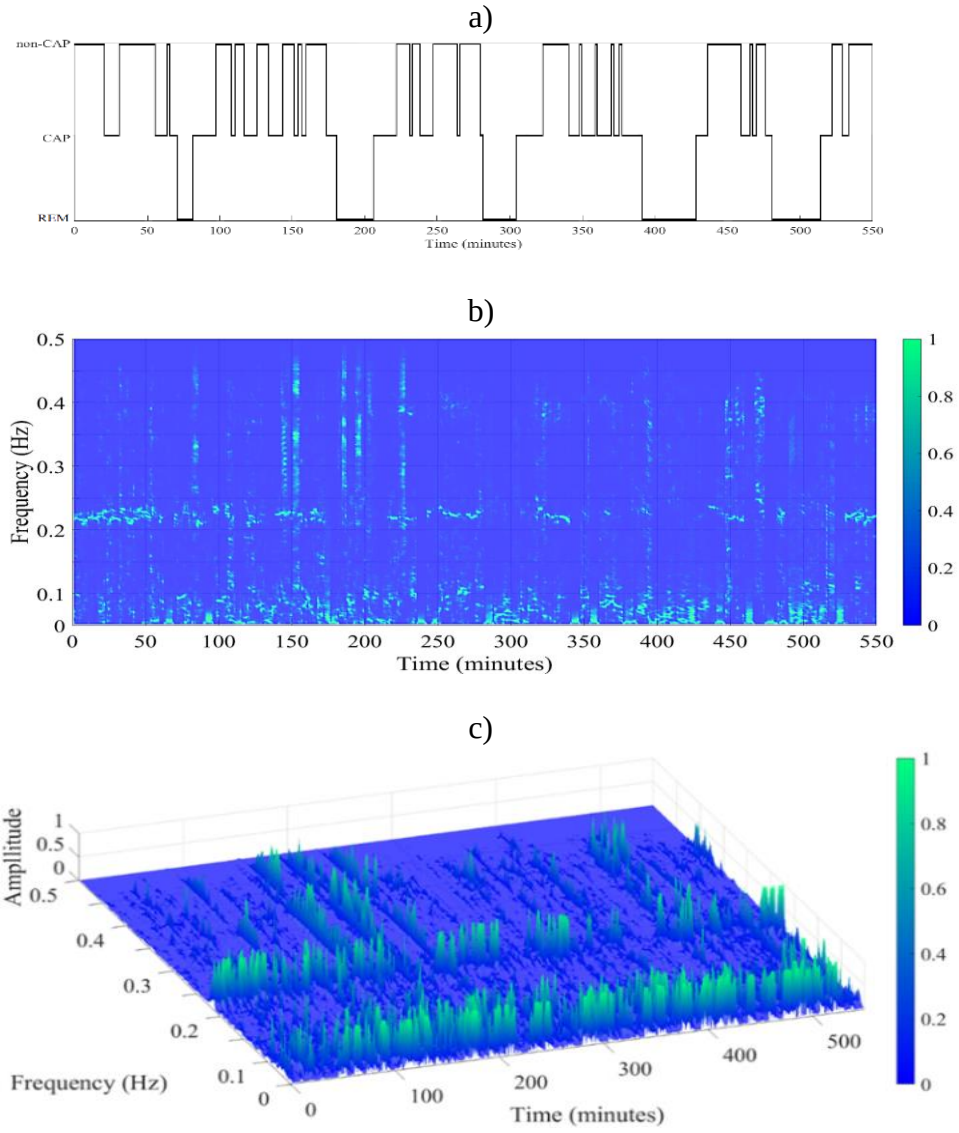


Figure 2.4. Example of CPC analysis during periods of wake or REM, instable (occurrence of CAP), and stable (absence of CAP) sleep, presenting the a) sleep information, b) top view from CPC spectrographic information, and c) side view from CPC spectrographic information [42].

A simplification of this concept was proposed in this work, to view the sleep process as oscillations between unstable or stable sleep according to the occurrence or absence of the CAP. This proposed model can be evaluated for both conventional and extended concepts for CAP. Specifically, for the conventional denotation of CAP, a new sleep model can be proposed to describe sleep with an approximation to continuous traces by evaluating epochs with the standard scoring duration for the microstructure (one second). This model can be interpreted as a new view of the sleep process, which oscillates between stable and unstable periods instead of the conventional sleep macrostructure stages. A similar concept can be proposed for the extended concept for CAP, where the oscillation between the stable and unstable is assessed by evaluating the indirect effect of CAP in other physiological signals.

It was indicated that the standard epoch duration for CAP scoring is 60 seconds [41]. However, a difficulty remains for the minute-by-minute analysis, which is how to decide what should be the minimum CAP time (duration of a CAP event) to designate a minute epoch as CAP. This issue is related to the fact that the sleep microstructure is scored considering epochs which last one second. Thus, a CAP cycle can be divided into two minute based epochs. An approach to address this issue was proposed in this work by considering the CAP epoch concept for the indirect CAP analysis, formally introduced as a period where more than a defined percentage of the 60 s of data was scored as a CAP cycle. Therefore, a threshold based methodology defined the one minute epoch as either unstable or stable sleep. However, this threshold should be properly tuned to remove the very short CAP related events which are too small to manifest in this indirect CAP analysis properly.

It was conceptualized that an alternative analysis to CPC can be performed by evaluating the causality between the EDR and the N-N series as in phenomena such as respiratory sinus arrhythmia, the heart rate is modulated by the breathing pattern. This approach could lead to the development of alternative ways of exploring the indirect CAP assessment for sleep quality estimation. It is also likely for this causality analysis to be affected by the occurrence of OSA as this disorder is characterized by repetitive breathing pauses, caused by the upper airway collapse during sleep, which impairs the normal ventilation of the lungs [43]. However, the methods proposed in the state of the art for OSA examination, based on the ECG signal, usually only consider isolated features from the HRV or from the EDR, instead of evaluating the information that the causality between the two signals can provide [44]. As a result, this causality model may be relevant for sleep quality analysis in subjects suffering from OSA.

2.3. OSA contextualization

Sleep apnea can be obstructive, central, or complex [45]. The first, denoted as OSA, is characterized by recurrent collapse and obstruction of the upper airway, impairing the normal ventilation of the lungs during sleep. The second is characterized by a lack of drive to breathe during sleep, which leads to repetitive periods of insufficient ventilation. However, during the cessations of airflow, in OSA, it is observed an ongoing respiratory effort while in central apnea, there is a lack of respiratory effort [46]. Complex apnea exhibit breathing patterns similar to central apnea but has clinical features that resemble OSA [47]. OSA was found to be the most prevalent of the three types of apnea, and it is

also related to the occurrence of CAP in subjects suffering from SBD. Thus, it was the sleep related disorder evaluated in this work. The severity of this disorder is assessed by the Apnea-Hypopnea Index (AHI), that calculates the average number of apneas and hypopneas per hour of sleep, and four thresholds are considered [45]: mild if $5 \leq AHI < 15$; moderate if $15 \leq AHI < 30$; severe if $AHI \geq 30$.

Hypoventilation can lead to a decrease in the oximetry levels, depending on the oxyhemoglobin desaturation dynamic, which can be assessed by the SpO2 sensor. As a result, oximetry is considered to be an indirect surrogate method for OSA screening [48]. Although the repetitive oxygen desaturation is usually associated with apnea, it is common for short respiratory pauses to not display a distinctive pattern in the oximetry signal. Such occurrence can possibly be related to the oxygen-hemoglobin dissociation curve, indicating that a distinct decrease in the partial oxygen pressure did not happen [49]. It is also relevant to notice that some desaturations may not be related to breathing pauses but may be produced by alveolar hypoventilation or chronic obstructive pulmonary disease.

It was observed that OSA events presented progressive bradycardia which is followed by abrupt tachycardia during the recommencement of breathing [50]. These events can be observed in the example presented in Figure 2.5. The N-N series shows the occurrence of bradycardia before the OSA. Simultaneously, the amplitude variation of EDR decreases, indicating a breathing reduction which is manifested in the SpO2 signal's desaturation. Multiple periods of intermittent breathing occur during the OSA event, leading abrupt tachycardia, visible in the N-N series, during the recommencement of breathing, depicted by the increase in EDR's amplitude and in the SpO2 saturation. As a result, it was concluded that changes in the ECG observed during apnea are mediated by the autonomic nervous system [50], leading to the conclusion that the HRV is a suitable signal for OSA evaluation.

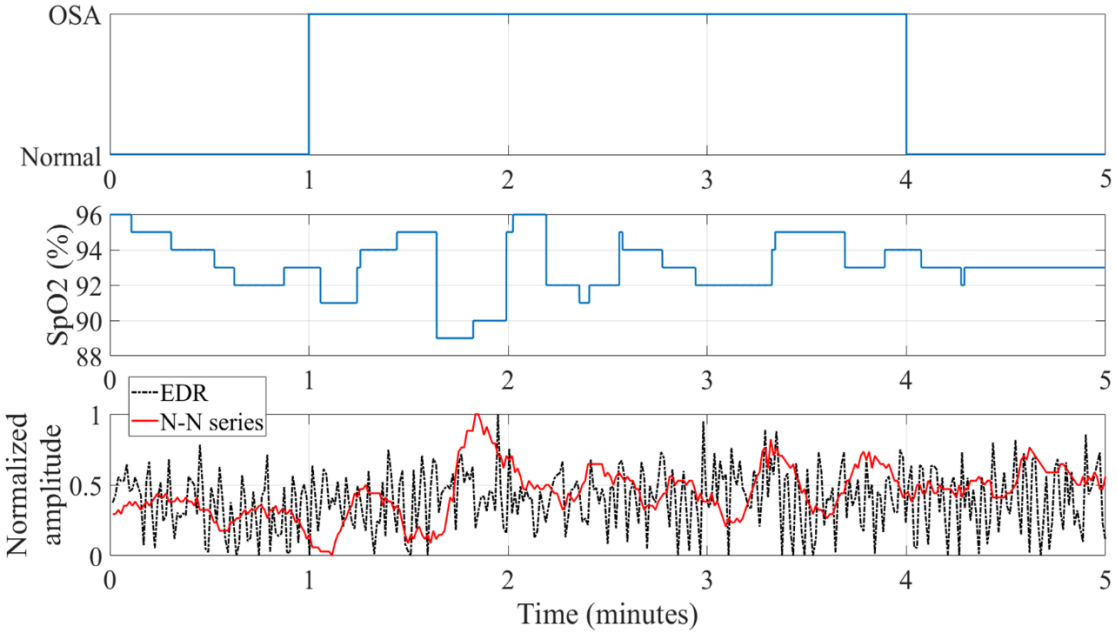


Figure 2.5. Example presenting the variation in the oximetry, EDR and N-N series signals during an OSA event.

The CPC analysis presented a special characteristic in the LF coupling, denoted elevated LF coupling, which was found to be related to apnea and hypopnea periods [51]. A particularly strong correlation was found between the presence of CAP, and the occurrence of OSA as increased amounts of arousals are usually found in patients suffering from this disorder [10] [52]. It was also observed that sleep instability, related to the presence of CAP, showed a progressive augmentation from normal subjects to mild and moderate-severe OSA patients [10]. Therefore, OSA was found to be the most suitable disorder to be examined in this work (for the second and third evaluated approaches for sleep quality examination) since it is one of the most prevalent sleep related disorders which is mostly undiagnosed, it is well correlated with CAP, and can be estimate from the same signal as the indirect CAP assessment (the ECG sensor).

2.4. Key remarks

A contextualization analysis was performed in this chapter, explaining the physiological/clinical perspective of the work. It was noted that the sleep process has manifestations in multiple physiological signals, and that sleep quality is a problematic concept to be properly defined. Nonetheless, the stability analysis of sleep, based on CAP evaluation, is likely to be the most relevant to appraise the sleep quality. It was conceptualized that the CAP concept can be interpreted in a broader sense as an indicator of sleep instability periods, leading to the development of new ways of interpreting the sleep process.

The manual CAP scoring is a complicated task for physicians to perform since it is a tedious and labor intensive process, making the classification prone to error. Particularly, the specialist agreement for CAP analysis, examining the same EEG results, ranges from 69% to 78% [53] (getting closer to the lower bound as the number of physicians performing the scoring increases [54]). Taking into account the available dataset for CAP analysis, it is possible to develop models for automatic CAP examination by employing machine learning models. However, there is a need to test multiple combinations of features and types of classifiers to assess which model is more suitable. Deep leaning approaches can also be tested to proposed models which learn the relevant patterns directly from the data.

3. State of the art

3.1. Approaches for sleep quality estimation

3.2. Overview of methods for OSA detection

3.3. Review of developed HMD for OSA detection

3.4. Key remarks

In this chapter, several reviews of the state of the art were performed to assess the proposed methodologies and devices for sleep quality and OSA examination. This research was performed to contextualize the developed work and orientate the research towards the identified gaps in the literature.

Several sensors and devices were proposed in the literature to examine the quality of sleep metrics and/or OSA assessment. Therefore, three reviews were conducted, the first aims to determine the most suitable approach for sleep quality examination [55] while the second examined the best methods and sensors for OSA detection [56]. The last review evaluated the developed devices that allow the OSA detection at the patient's home [57].

3.1. Approaches for sleep quality estimation

Multiple subjective measurements were proposed through the years to define the sleep quality, and are either based on paper and pencil sleep diaries or on sleep questionnaires. The sleep diaries, such as the Karolinska sleep diary [58], are commonly filled in the subsequent morning and present sleep appraisal questions to the subject or ask the subject to describe the sleep experience. However, a questionnaire's employment is a faster and easier approach that commonly elaborates questions whose answers are translated to a self-rating index [59].

The Pittsburgh Sleep Quality Index (PSQI) is a commonly used questionnaire to evaluate sleep quality with questions that can be categorized into seven domains (use of sleep medication, sleep efficiency, sleep duration, sleep latency, sleep disturbance, daytime dysfunction, and subjective sleep quality) [60]. A similar approach is used by the Self-rating Sleep and Awakening (SSA) quality scale, which produces three sub-scores: subjective sleep quality; somatic complaints; subjective awakening quality. The information is then used to compute a total score that reflects the sleep experience [61]. A new single-item sleep quality scale [62] was recently proposed, asking the subject to rate (from terrible to excellent) the quality of sleep over a 7-day recall period.

However, it was verified that sleep complaints are usually more related to general dissatisfaction than to a specific disorder [63]. Furthermore, the correspondence between objective measurements and the person's subjective assessment of the overall quality of sleep quality is significantly low, with a maximum correlation of 35% when considering the SSA score and standard sleep metrics [63]. Therefore, a systematic review was conducted with the goal of analyzing the methods and devices presented in the literature, which employ objective measures to define the quality of sleep. The research period ranges from 2000 to 2018, with the distribution of reviewed articles by year of publication presented in Figure 3.1. Two research questions were formulated [55]: What methods for sleep quality assessment were developed?; What kind of measures are employed by the devices that were developed to estimate sleep quality?. The review was conducted in the ScienceDirect, Web of Science, IEEE explorer, Google Scholar, and PubMed databases considering the search keywords "sleep quality AND device" and "sleep quality AND method". In total, 10594 articles were found. Inclusion and exclusion criteria were employed to assess the relevance of the studies.

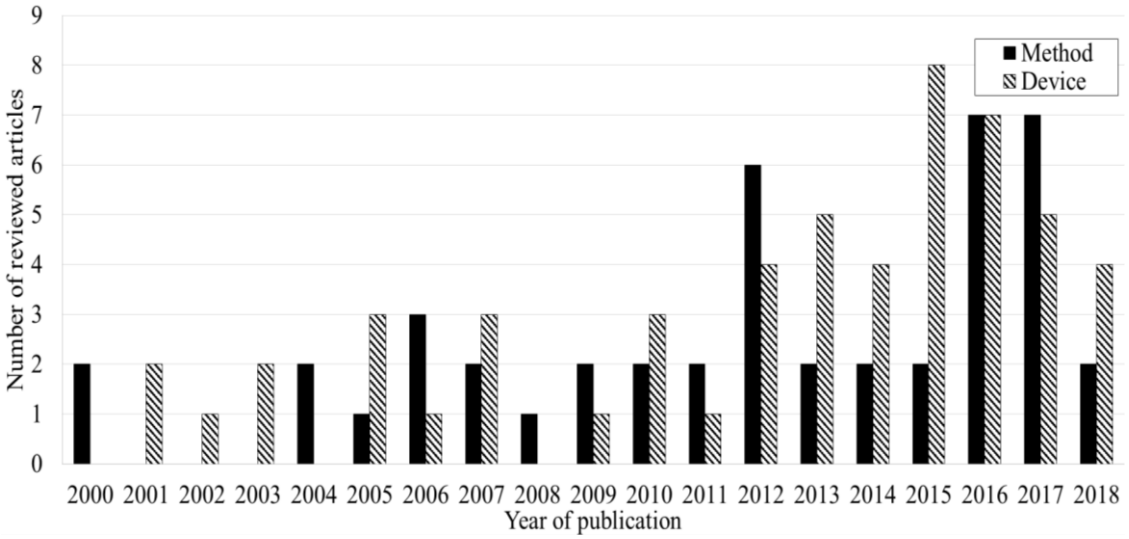


Figure 3.1. Distribution of the reviewed articles, in the sleep quality review, by year of publication [55].

Specifically, the inclusion criteria for the articles which presented a method were: presentation of a method to measure sleep quality; specifically mention the usability of the method; a study published in a scientific conference or scientific journal. For the devices, the inclusion criteria were: presentation of a device capable of assessing the sleep quality; validation of the research project or commercial device analyzing sleep quality metrics; specifically mention the usability of the device for sleep quality analysis. The exclusion criteria for both methods and devices were: article not written in English; lack of description of the measurement method or sleep quality metric; the developed device is not suitable for home detection; presentation of an application that only uses the smartphone sensors. The last exclusion criteria was used since the smartphone applications have already been examined in a review [64].

Ninety articles were selected for the review after removing the duplicated articles and applying the selection criteria. A compilation of the assessed PSG based sleep quality metrics is presented in Table 3.1, and the metrics analysis is presented in Table 3.2.

Table 3.1: PSG based sleep quality metrics [55].

Group	*Description*	*Measure*	*Simplified formula*
Duration	Metric based on time duration	-Lights out to N1 (LN1)	- LN1 = Σ(minutes from lights out to first N1 sleep stage)
		-Lights out to N2 (LN2)	- LN2 = Σ(minutes from lights out to first N2 sleep stage)
		-Lights out to SWS (LSWS)	- LNSWS = Σ(minutes from lights out to first N3 sleep stage)
		-Lights out to REM (LR)*	- LR = Σ(minutes from lights out to first REM sleep)
		-Maximal sustained N1 (MN1)	- MN1 = max(sustained N1 period)
		-Maximal sustained N2 (MN2)	- MN2 = max(sustained N2 period)
		-Maximal sustained N3 (MN3)	- MN3 = max(sustained N3 period)
		-Maximal sustained REM (MR)	- MR = max(sustained REM period)
		-Maximal sustained Wake (MW)	- MW = max(sustained wake period)
		-Sleep Onset Latency (SOL)*	- SOL = Σ(recording minutes)-TST

		-REM latency (REML)	- REML = TST-Σ(minutes after start first REM until final awake)
		-Time Attempting to Sleep After the Final Awakening (TASAFA) episode	- TASAFA = Σ(minutes after the final awakening)
		-Time in Bed (TIB)	- TIB = Σ(minutes from lights out to the end of recording)
		-Total Sleep Time (TST)*	- TST = Σ(sleep minutes)
		-Total Wake Time (TWT)	- TWT = Σ(wake minutes after start the sleep until final awakening)
		-Total N1 sleep (TN1)*	- TN1 = Σ(N1 minutes)
		-Total N2 sleep (TN2)*	- TN2 = Σ(N2 minutes)
		-Total N3 sleep (TN3)*	- TN3 = Σ(N3 minutes)
		-Total NREM minutes (TNR)	- TNR = Σ(NREM minutes)
		-Total REM minutes (TR)*	- TR = Σ(REM minutes)
		-Total Wake minutes (TW)	- TW = Σ(wake minutes)
		-Wake After Final Awakening (WAFA)	- WAFA = Σ(minutes from final awakening to the end of recording)
Intensity	Percentage of time spent on a specific sleep stage	-N1 percentage (N1%)*	- N1% = Σ(N1 minutes)/TST
		-N2 percentage (N2%)*	- N2% = Σ(N2 minutes)/TST
		-REM percentage (REM%)*	- REM% = Σ(REM minutes)/TST
		-REM to NREM ratio (RNR%)	- RNR% = Σ(REM minutes)/ Σ(NREM minutes)
		-Stage Shift Index (SSI)	- SSI = Σ(sleep stage shifts)/TST
		-SWS percentage (SWS%)*	- SWS% = Σ(N3 minutes)/TST
		-SWS to Light sleep ratio (SL%)	- SL%= Σ(N3 minutes)/[Σ(N1 minutes)+ Σ(N2 minutes)]
		-SWS to NREM ratio (SNR%)	- SNR%= Σ(N3 minutes)/Σ(NREM minutes)
		-SWS to REM ratio (SR%)	- SR%= Σ(N3 minutes)/Σ(REM minutes)
Continuity	Degree of sleep fragmentation	-Arousal Index (AI)*	AI = Σ(number of arousals)/hour of sleep
		-Awakening (A)	- A = Σ(number of awakenings lasting more than a defined period)
		-Awakening Index (AwI)	AwI = Σ(number of awakenings)/hour of sleep
		-Deep sleep Efficiency (DEI)	- DEI = Σ(N3 minutes)/Σ(minutes in bed)
		-Frequency of Stage Shifts (FSS)	- FSS = Σ(number of sleep stage shifts)
		-Frequency of Sights from SWS to N1 or N2 (FSN12)	- FSN12 = Σ(number of sleep stage shifts from SWS to N1 or N2)
		-Number of Stage Shifts (NSS)	- NA = Σ(number of arousals)
		-Number of Arousals (NA)*	- NSS = Σ(number of sleep stage shifts)
		-Sleep Efficiency (SE)*	- SE = TST/Σ(minutes in bed)
		-Wake After Sleep Onset (WASO)*	- WASO = Σ(wake minutes after start sleep)/TST
Stability	Sleep disruptive events	-A1 phase percentage (A1%)	- A1% = Σ(number of A1 phases)/ Σ(number of A phases)
		-A2 phase percentage (A2%)	- A2% = Σ(number of A2 phases)/ Σ(number of A phases)
		-A3 phase percentage (A3%)	- A3% = Σ(number of A3 phases)/ Σ(number of A phases)
		-A phase Index (ApI)	- ApI = Σ(number of A phases)/hour of NREM sleep
		-A1 phase Index (A1pI)	- A1pI = Σ(number of A1 phases)/hour of NREM sleep
		-A2 phase Index (A2pI)	- A2pI = Σ(number of A2 phases)/hour of NREM sleep
		-A3 phase Index (A3pI)	- A3pI = Σ(number of A3 phases)/hour of NREM sleep

		-Arousals percentage (A%)	- A% = Σ(arousal events)/TST
		-B phase Index (BpI)	- BpI = Σ(number of B phases)/hour of NREM sleep
		-CAP cycle Index (CAPI)	- CAPI = Σ(number of CAP cycles)/hour of sleep
		-CAP Rate (CAPR)	- CAPR = Σ(CAP minutes)/Σ(NREM minutes)
		-EEG spectral indices (e.g. energy, in NREM and REM, of the frequency bands: alpha, beta, delta, high-frequency sigma, low frequency-sigma, sigma, theta)	
		-Mean A phase Duration (MAD)	- MAD = Σ(A phases minutes)/ Σ(number of A phases)
		-Mean A1 phase Duration (MA1D)	- MA1D = Σ(A1 phases minutes)/ Σ(number of A phases)
		-Mean A2 phase Duration (MA2D)	- MA2D = Σ(A2 phases minutes)/ Σ(number of A phases)
		-Mean A3 phase Duration (MA3D)	- MA3D = Σ(A3 phases minutes)/ Σ(number of A phases)
		-Mean B phase Duration (MBD)	- MBD = Σ(B phases minutes)/ Σ(number of B phases)
		-Number of CAP Cycles (CAPC)	- CAPC = Σ(number of CAP cycles)
		-Number of Periodic Limb Movements (NPLM)*	- NPLM = Σ(number of periodic limb movements)
		-Periodic Limb Movements Index (PLMI)*	- PLMI = Σ(NPLM*60)/TST
		-Quantity or presence of pathological events	
		-Total CAP Time (CAPT)	- CAPT = Σ(CAP minutes)
		-Total number of A phases (TA)	- TA = Σ(number of A phases)
		-Total number of A1 phases (TA1)	- TA1 = Σ(number of A1 phases)
		-Total number of A2 phases (TA2)	- TA2 = Σ(number of A2 phases)
		-Total number of A3 phases (TA3)	- TA3 = Σ(number of A3 phases)
		-Total number of B phases (TB)	- TB = Σ(number of B phases)
Frequency	Number of occurrences of a sleep stage	-Number of times in the N1 stage (NN1)	- NN1 = Σ(number of times in the N1 stage)
		-Number of times in the N2 stage (NN2)	- NN2 = Σ(number of times in the N2 stage)
		-Number of REM cycles (NR)	- NR = Σ(REM cycles)
		-Number of times in the N3 stage (NS)	- NS = Σ(number of times in the SWS stage)
Sleep episodes	Description of the nap periods	-Number of Naps (NN)	- NN = Σ(number of naps in 24 h)
		-Nap Duration (ND)	- ND = Σ(nap minutes)/NN
		-Nap Frequency (NF)	- NF = Σ(number of days in the past week that a nap occur)

* Recommended by the AASM manual for the scoring of sleep and associated events to be a reported parameter [21].

Table 3.2: Analysis of the PSG based sleep quality metrics [55].

Sleep structure	*A higher value indicates better sleep quality*	*A higher value indicates low sleep quality*	*Specified optimal value for a healthy subject*	*Unspecified optimal value for a healthy subject*
Macrostructure	DEI	LN1	NN1	FSS
	MR	LN2	NN2	FSN12
	SE*	LR*	ND	SR%
	REM%*	LSWS	NR	MN1
	RNR%	MW	NS	MN2
	SL%	N1%*	NSS	MN3
	SNR%	N2%*	TIB	
	SWS%*	NF	TN1*	
	TR*	NN	TN2*	
		REML	TN3*	
		SOL*	TNR	

	SSI	TST*	
	TASAFA		
	WASO*		
	TW		
	TWT		
	WAFA		
Microstructure	A	CAPC	A1%
	A%	MAD	A2%
	AI*		A3%
	AwI		ApI
	CAPI		A1pI
	CAPR		A2pI
	CAPT		A3pI
	NA*		BpI
	TA		MA1D
	TA1		MA2D
	TA2		MA3D
	TA3		MBD
	TB		

* Recommended by the AASM manual for the scoring of sleep and associated events to be a reported parameter [21].

The proposal of new methods was analyzed in the first part of the review and the summary of the reviewed articles is presented in Table 3.3. The goal for this table's analysis is to assess the developments in the field of sleep quality examination, indicating the key elements of the proposals and a description of the methods. On the other hand, Table 3.4 presents the analysis of the developed devices (reviewed in the second part). The objective for the development of this table is to assess what are the most frequently used sensors for sleep quality examination, and what type of metrics these devices employ.

Table 3.3: Review of proposed methods for sleep quality analysis [55].

Approach	*Article*	*Key element*	*Method description*
Based on sleep macrostructure parameters	[65]	EEG Slow-wave (Delta band) microcontinuity	Quantify the sleep depth by determining the fraction of the present delta band waves that continues to the near-future of the signal.
	[66]	Sleep restorative ability	Metric evaluated every 30 s epoch that is incremented if the sleep stage is N3, remain the same in REM and reduced in the other stages.
	[67]	Chaos analysis of HRV	Evaluated a two-dimensional map that was composed of the largest Lyapunov exponent and a correlation dimension to estimate the number and periodicity of the sleep cycles.
	[68]	Sleep or wake detection	Repeating breathing patterns were determined by analyzing the period, intensity and consistency of each 12 s interval of the respiratory signal. Snore properties were assessed by estimating the maximum snore likelihood scores in a 30 s window, and the number of snores per hour (snore index). Sleep or wake detection was performed using an adaptive boosting classifier that was fed with features from the current and the two previous windows.
	[69]	Sleep or wake detection	Analyzed the repeating breathing patterns and snore properties to feed a two-state Hidden Markov Model (HMM). A probability density function was determined, by integrating the transition probability, to estimate a sleep quality score every 5 s.
	[70]	Sound events	Sounds related to sleep events were clustered using Kullback–Leibler kernel self-organizing map and afterwards categorized by hierarchical clustering. The output was classified as either good or bad sleep using a multinomial HMM.

	[71]	Sleep stage detection	Time and frequency based features were extracted from HRV and surrogate ECG Derived Respiratory (EDR). The features were then fed to a multi-stage Support Vector Machine (SVM) with a Gaussian kernel.
	[72]	Normal to normal heart rate	Analysis of the normal to normal heart rate standard deviation average over a 5 minutes window to estimate the sleep stage.
	[73]	HRV from photoplethysmography (PPG)	Estimation of sleep or wake periods by analyzing the power in the low-frequency band of the HRV estimated from the PPG signal.
	[74]	Body movements and HRV	Body movements estimated from actigraphy and PPG-derived HRV were fed to a linear discriminant analysis for classification of sleep or wake periods.
	[75]	Sleep or wake periods	Actigraphy signals were employed to estimate the vigor of motion, movement frequency and time spent in motion to estimate the sleep or wake periods in one minute windows.
	[76]	Sleep or wake periods	Determine the sleep time by analyzing actigraphy signals that were fed to a convolutional neural network.
	[77]	Arterial baroreflex	Ballistocardiogram (BCG) and ECG and signals were used to estimate the R–J interval (the period between the R peak from ECG and the J peak from BCG). The presence of a negative correlation between the HRV and the detrended R–J interval fluctuations, for more than 120 s, indicated the beginning of sleep since the arterial baroreflex produces a characteristic influence on the HRV control at the beginning of sleep.
	[78]	BCG peak detection	Nocturnal awakening periods were detected by performing BCG peak detection.
Using parameters from sleep microstructure	[30]	CAP estimation	Features produced from EEG monopolar derivations, specifically, the Power Spectral Density (PSD) in the beta and theta bands, Teager energy operator, autocovariance and Shannon entropy, were fed to a Feed-Forward Neural Network (FFNN) to estimate the presence of A phases and a finite state machine was employed to determine the CAP cycles.
	[79]	A phase detection	Employed band descriptors, Hjorth activity and differential variance on EEG monopolar derivations to produce features that were feed to a FFNN to determine the presence of the A phases.
	[41]	CAP estimation	Analyzed the Cardiopulmonary Coupling (CPC) of normal-to-normal sinus intervals (N-N series) and EDR. A quantitative index was produced by multiplying the crossspectral power by the coherence and a predominance of low-frequency coupling (0.01 to 0.1 Hz) was associated with the presence of CAP.
	[80]	CAP rate estimation from a single-lead ECG	A spectrographic representation of CPC, from EDR and N-N series, was fed to a Deep Stacked Autoencoders (DSAE) to detect the NREM and CAP periods.
	[81]	CAP rate estimation from a single-lead ECG	The spectrographic representation of CPC was fed to the DSAE and tuned thresholds were employed in the CAP classification.
Analyzing a combination of parameters from the sleep structure	[82]	PSD features extracted from the EEG	62-dimension PSD features were produced from the EEG bands and minimal-redundancy-maximal-relevance was employed for feature selection. The selected features and the energy in the brain topographic map were fed to a discriminative graph regularized extreme learning machine.
	[12]	Combination of multiple measures	A combination of NREM spectral EEG indices, CAP rate and traditional PSG indices was used to produce groups of features and thresholds were employed to define the relevance of each group. A final threshold was then employed to perform the classification of sleep quality.
	[83]	Crosscorrelation functions	A crosscorrelation matrix was produced by analyzing the normalized crosscorrelation coefficients between two periods and employed to calculate an error function. The local minima of the function represents the transition in sleep.
Using models based on sleep microstates	[84]	Probabilistic Sleep Model (PSM)	Defined sleep through posterior probabilities of twenty sleep microstates, forming the PSM, and identified the ones that are related to the presence of poor sleep quality periods.
	[63]	PSM	Describe sleep using the PSM and identified the ones that are related to standard sleep variables from PSG.
	[85]	PSM	Employed the PSM to estimate the relative time spent in a microstate and the number of transitions between microstates.
	[86]	Model based on commonly not directly observed latent variables	Grouped a wide set of metrics into a smaller parsimonious set of generally not directly observed latent variables to produce a sleep quality index.
	[87]	HMM	Employ HMM to analyze the data from a single EEG signal with one second epochs.
	[88]	HMM	Analyzed the EEG signal with a HMM and defined metrics based on probabilistic traces to define the quality of sleep.

Non sleep structure features	[89]	Rules based on attributes	Defined sets of rules that combine multiple attributes and selected the most significant set by examining the attributes that are most commonly related to poor sleep quality.
	[90], [91], [92]	Multimodality sensor	Features from multiple sensors were fed to classifiers (one SVM for each sensor). The outputs of each classifier were fused to deduce the sleep quality.
	[93]	Motion artifacts	Employed electromechanical film sensors to detect periodic movements during sleep-related to period of poor sleep quality.
Proposed new metrics	[94]		Proposed two sleep quality ratios.
	[95]		Developed a sleep fragmentation index.
	[96]		Suggested a weighted-transition sleep fragmentation index, where the weight was related to the possible sleep stage transitions.
	[97]		Proposed a sleep fragmentation index.
	[98]		Defined a mathematical diagnosis for the sleep fragmentation.
	[99]		Presented two entropy-based metrics.
	[100], [101]		Developed a sleep diversity index based on Shannon entropy.
	[102]		Combined the NREM sleep duration, sleeping position and presence of sleeping disorders into a sleep quality index.
	[103]		Apply a threshold to classify a metric based on the frequency of roll-overs.
	[104]		Developed a metric based on the acceleration of the movements and the sleeping position.
	[105]		Developed a metric to quantify the bed occupancy by estimating the number of times the subject left the bed.
	[106]		Determined the bed occupancy by analyzing the radiated temperature.
	[107]		Developed three metrics of sleep continuity.

Table 3.4: Review of the developed devices for sleep quality analysis [55].

Type of device	*Name of the device*	*Article*	*Sensors*	*Sleep quality measurements used in the article*
Proposed by researchers		[90]	Camera; passive infrared sensor; heart-rate sensor	SOL; TST; SE
		[91]	Passive infrared sensor; heart-rate sensor; Microphone	SOL; TST; SE
		[108]	EEG	Developed metric
		[109]	EEG	TN1; TN2; TN3; SE
		[72]	PPG	SWS%; developed metric
	Sleep MedAssist	[71]	ECG	Developed metric
		[110]	ECG; actigraph; oximeter; microphone	SE; SOL; A; WASO; PLMI; N1%; N2%; SWS%; REM%; REML; developed metric
		[111]	Actigraph	TST; SE; WASO; SOL
	RAHAR	[76]	Actigraph	SE
		[102]	Actigraph; pressure sensor	Developed metric
		[78]	Pressure sensor	SE
		[112]	Pressure sensor	Developed metric
		[113]	Pressure sensor	Developed metric
		[114]	Pressure sensor	TIB; developed metric
		[115]	Pressure sensor	Developed metric
		[77]	Pressure sensor	SOL
		[116]	Pressure sensor	TST; TN3; SE; developed metric
	DoppleSleep	[117]	Doppler radar	SOL; A; TST; SE
	Wi-Sleep	[118]	Transmitter and receiver antennas	Developed metric
		[119]	Magnetometer	Developed metric
		[120]	Thermometer; humidity sensor; microphone; luminosity sensor; micro-vibration sensor	Developed metric

Commercial	Actical	[121]	Actigraph	TST; SE
	Actillume	[122]	Actigraph	TST; developed metric
	Actigraph	[123]	Actigraph	Developed metric
	Actigraph GT9X	[124]	Actigraph	TST; SE; TIB; WASO
	Actiwatch	[121]	Actigraph	TST; SE
	Actiwatch 2	[125]	Actigraph	TST; SE; WASO; SOL
	Actiwatch-16	[126]	Actigraph	TST; SE; TIB; SOL; TWT
	Actiwatch 64	[127]	Actigraph	TST; SOL; WASO; SE
	Actiwatch-L	[128]	Actigraph	TST; SE; SOL; TWT
	Actiwatch Spectrum	[129]	Actigraph	TST; SE; developed metric
	AW4	[130]	Actigraph	TST; SE; A
	Basis Health Tracker	[129]	Actigraph	TST; SE; developed metric
	Basis Peak	[124]	Actigraph	TST; SE; TIB; WASO
	Fitbit	[131]	Actigraph	TST; SE
	Fitbit Charge 2	[132]	Actigraph	TST; TIB; SE
	Fitbit Charge HR	[133]	Actigraph	TST; TIB; developed metric
	Fitbit Flex	[129]	Actigraph	TST; SE; developed metric
	Fitbit One	[134]	Actigraph	TST; SE; SOL; developed metric
	Fitbit Ultra	[135]	Actigraph	TST; SE; WASO
	GT3X+	[127]	Actigraph	TST; SOL; WASO; SE
	IST Vivago	[136]	Actigraph	TIB; developed metric
	Jawbone UP	[137]	Actigraph	TST; SE; TIB; TW; SOL; WASO
	Jawbone UP3	[124]	Actigraph	TST; SE; TIB; WASO
	MicroMini-Motionlogger	[138]	Actigraph	TST, SE, SOL; A
	Mini Motionlogger	[139]	Actigraph	TST; SE; SOL; A
	Misfit Shine	[129]	Actigraph	TST; SE; developed metric
	SenseWear Armband	[140]	Actigraph	TST; TIB; SE
	SenseWear Armband Mini	[124]	Actigraph	TST; SE; TIB; WASO
	SenseWear Pro2 Armband	[103]	Actigraph	Developed metric
	SenseWear Pro3 Armband	[125]	Actigraph	TST; SE; WASO; SOL
	Sleepwatch	[121]	Actigraph	TST; SE
	Sleepwatch-O	[75]	Actigraph	TST; SE; SOL; WASO
	Vivago WristCare	[141]	Actigraph	TST
	Vivosmart	[124]	Actigraph	TST; SE; TIB; WASO
	Withings Pulse O2	[129]	Actigraph	TST; SE; developed metric
	Z80-32K V1	[142]	Actigraph	TST; SE; SOL; TR
	Ōura ring	[143]	Accelerometer; infrared sensor	TST; SOL; WASO; TN1; TN2; TN3; TR
	Zeo Sleep Manager Pro	[144]	EEG	TST; SE; SOL; WASO; A; TNR; TR; TW; REML; developed metric
	NeuroOn Open	[145]	EEG; EOG; oximeter; thermometer; accelerometer	TST; SE; A; WASO; SOL; TN1; TN2; TN3; TR
	SleepImage	[146]	ECG	Developed metric
	M1	[147]	ECG; accelerometer; vibration sensor	TST; WASO; developed metric
	Beddit Pro	[134]	Thermometer; microphone; luminosity sensor; force sensor	TST; SE; A
	EarlySense	[148]	Pressure sensor	TST

It was verified that a comparison between the analyzed articles was not feasible due to the employment of different metrics, databases, and application conditions in the experiments. Thus, the performed analysis was focused on studying what type of metrics were used by the developed methods and devices. The summary of this study is presented in Figure 3.2.

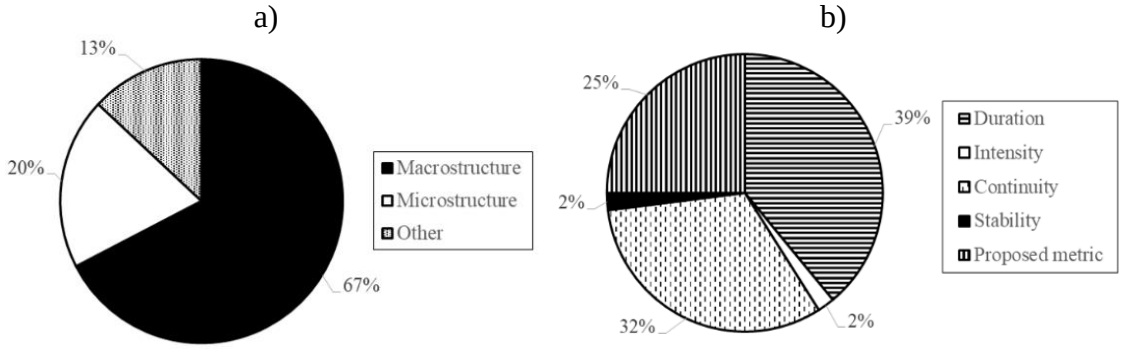


Figure 3.2. Resume of the reviewed articles regarding the a) proposed metrics, and b) the developed devices [55].

It was verified that primarily the proposed metrics are based on characteristics of the sleep macrostructure while microstructure was only employed by 20% of the methods. Thus, it is plausible to institute a tendency in the research for metrics that are based or related to the sleep macrostructure. This is likely due to the fact that sleep macrostructure has well-established definitions that help to propose solid metrics, while the microstructure is usually employed to assess the presence of sleep disorders. However, as the person progresses into an older age, the sleep structure undergoes through significant changes, increasing the subjective complaints that are associated with poor sleep quality [7].

These changes produce characteristic variations in the microstructure. Hence, the driven metrics from this structure can provide strong indications of sleep instability, and provide a sleep quality estimation which has a higher correlation with the self-ratings [37]. Taking into consideration the inherent difficulties related to the second by second examination required for the microstructure analysis, it was conceptualized that the proposal of a new sleep model, based on the microstructure, could be relevant for the future research in this field. Such model can be based on the instability analysis, which works as a foundation for CAP examination, and might provide a framework that can be significant for future clinical analysis.

Most of the devices employed metrics based on duration or continuity of sleep. New metrics were proposed by 26% of the studied cases, while 4% used either stability or intensity metrics. These results are due to the fact that actigraphs dominate the home health care market, since they are widely available, simple to be used, and non-invasive (commonly they are mounted on a bracelet that is worn on the wrist). The taxonomy of sleep quality analysis that was attained from this review is presented in Figure 3.3.

However, it was verified that duration and continuity metrics provide a small contribution to subjective ratings regarding prior-night sleep quality [11]. It was also assessed, by the AASM, that sleep logs and actigraphy provide significantly different results in the estimation of Sleep Efficiency (SE), Sleep Onset Latency (SOL) and Total

Sleep Time (TST) with a comparable estimation of Wake After Sleep Onset (WASO). Actigraphy and PSG measures have a poor agreement in the SE and WASO estimation with a reasonable agreement for the SOL and TST [149]. However, as it was previously observed, metrics based on the microstructure provide a sleep quality estimation which was found to be more correlated with self-ratings [37]. Particularly, it was conceptualized (in Chapter 2) that stability analysis (based on the CAP assessment) could conceivably offer a more comprehensive and sensitive measure for the quality of sleep. Taking into consideration that only 2% of the devices used stability metrics, it was theorized that there is a need for the development of new prototypes capable of performing CAP assessment at the patients' home. Such devices could improve the availability of sleep quality analysis for the general population, having a significantly lower cost than a full night PSG, hence, becoming a valuable tool for the future of healthcare.

From the overview of the analysis, it is possible to infer that the development of devices which employ metrics capable of measuring the stability of sleep, such as CAP rate, can become significant tools for medical examinations that require sleep quality analysis.

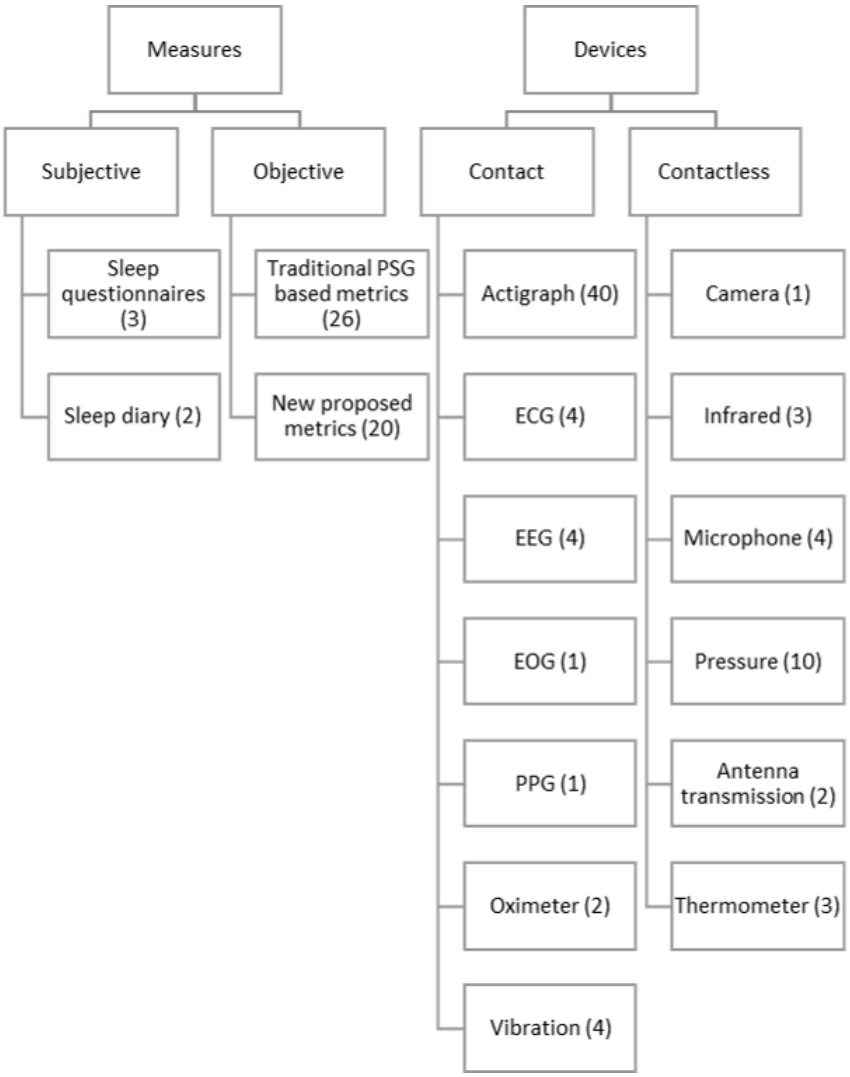


Figure 3.3. Taxonomy of sleep quality analysis (number of articles for each category is indicated between brackets) [55].

3.2. Overview of methods for OSA detection

A literature review was conducted to determine the most suitable sensors, according to the performance of the metric that analyzed the signals from the sensor, for OSA detection [56]. Specifically, the purpose of this review was to assess what sensors are likely to be the most relevant for OSA examination, according to the reported performance. The methods based on the combination of two or more sensors were also examined to verify if such combination can lead to a significant improvement of the performance when comparing with the methods based on a single source sensor. The search covered the period from 2003 to 2017 and was conducted on IEEE explorer, Web of Science, PubMed, various journals, and cited literature in the included articles. The search keywords were: “algorithm AND sleep apnea”; “apnea AND deep”; “ECG AND apnea”; “oximetry AND apnea”; “Respiration analysis AND apnea”; “sound AND apnea”; “snoring AND apnea”.

The inclusion criterion was the presentation of a method, validated using data collected on a hospital or available in a database, which had not been implemented on hardware. The exclusion criterion was the lack of diagnostic elements to evaluate the capability of the method. A group of 84 articles that present methods with the potential to be promising diagnostic tools were selected and divided into five categories according to the employed source sensor: oximetry; ECG; respiration; sound; the combination of two or more source sensors (oximetry, ECG, respiration, and sound).

The distribution of the reviewed articles according to the year of publication is presented in Figure 3.4. The summary of the analyzed articles is presented in Table 3.5, indicating for each method the population, either subjects (sub) if the data was collected in a hospital or recordings (rec) if a database was used, employed to test the method, how the data was acquired, the Time Window (TW) of the algorithm, and the performance metrics that were considered, specifically: Epoch Based (EB) accuracy (Acc), Area Under the receiver operating characteristic Curve (AUC), sensitivity (Sen), and specificity (Spe); Acc of the global classification of the disorder; Subject Based (SB) AUC, Sen, and Spe. The SB approach measures globally every subject while the EB classifies individually every epoch for all subjects.

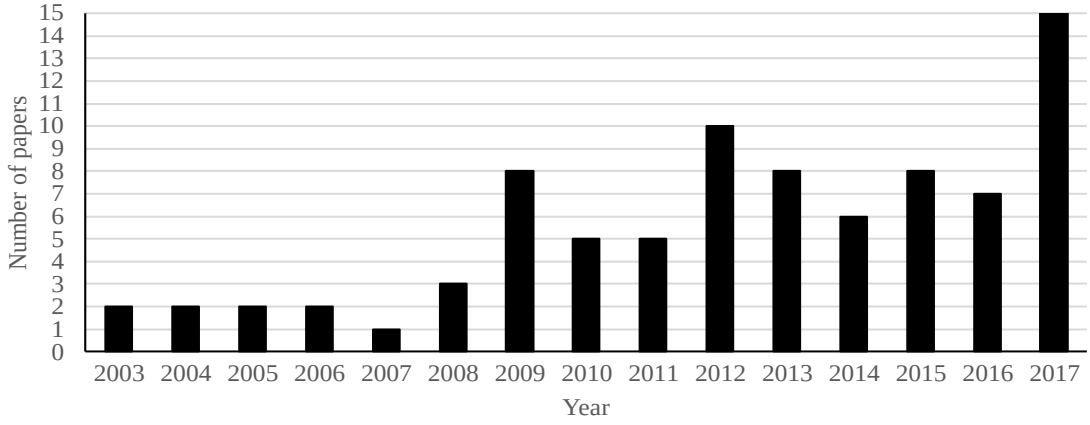


Figure 3.4. Distribution of the reviewed articles for OSA detection by year of publication [56].

Table 3.5: Review of the methods for OSA detection [56].

Source sensor	*Article*	*Population*	*Data acquisition*	*EB Acc (%)*	*EB AUC (%)*	*EB Sen (%)*	*EB Spe (%)*	*Global Acc (%)*	*SB AUC (%)*	*SB Sen (%)*	*SB Spe (%)*	*TW (s)*
Oximetry	[150]	187 sub	Hospital	-	-	-	-	87	92	90	83	120
	[151]	83 sub	Hospital	-	-	-	-	86	91	91	79	-
	[152]	113 sub	Hospital	-	-	-	-	88	93	91	83	-
	[153]	129 sub	Hospital	-	-	-	-	93	95	97	79	120
	[154]	148 sub	Hospital	-	-	-	-	90	97	92	85	30
	[155]	144 sub	Hospital	-	-	-	-	87	-	92	77	-
	[156]	8 rec	Database*	93	-	88	100	-	-	-	-	-
	[157]	21 sub	Hospital	70	78	82	69	87	-	100	71	40
	[158]	115 sub	Hospital	-	-	-	-	94	96	92	96	60
	[159]	36 sub	Hospital	-	-	-	-	85	88	88	84	120
	[160]	127 sub	Hospital	-	-	-	-	90	-	94	70	-
	[161]	25 rec	Database+	85	-	60	92	-	-	-	-	60
	[162]	79 sub	Hospital	-	-	-	-	94	-	97	79	-
	[163]	8 rec	Database*	96	98	-	-	-	-	-	-	60
	[164]	92 sub	Hospital	91	-	83	89	97	99	98	95	60
	[165]	8 rec	Database*	98	-	97	99	-	-	-	-	60
ECG	[166]	147 sub	-	-	-	-	-	91	-	92	90	-
	[167]	35 rec	Database*	-	-	92	95	-	-	-	-	60
	[168]	35 rec	Database*	90	-	89	91	89	-	-	-	60
	[169]	35 rec	Database*	84	-	79	87	-	-	-	-	60
	[170]	5 rec	Database#	-	-	-	-	-	-	70	44	30
	[171]	25 rec	Database*	86	-	84	89	-	-	-	-	420
	[172]	25 rec	-	-	-	-	-	88	-	89	86	60
	[173]	35 rec	Database*	93	-	-	-	-	-	-	-	180
	[174]	16 sub	Hospital	-	-	-	-	95	-	-	-	60
	[175]	42 sub	-	-	-	-	-	93	-	-	-	-
	[176]	30 rec	Database*	93	-	90	100	100	-	-	-	60
	[177]	17 sub	Hospital	-	-	-	-	87	-	-	-	30
	[178]	35 rec	Database*	76	-	-	-	-	-	-	-	60
	[179]	60 rec	Database*	-	-	-	-	100	-	100	100	60
	[180]	35 rec	Database*	81	-	-	-	-	-	100	83	-
	[16]	32 rec	Database*	-	-	-	-	97	-	93	100	15
	[181]	35 rec	Database*	-	89	74	86	93	-	-	-	60
	[182]	35 rec	Database*	99	-	-	-	-	-	-	-	-
	[183]	97 sub	Hospital	-	-	-	-	-	94	89	83	30
	[184]	69 rec	Database*	-	-	-	-	-	93	87	88	60
	[185]	35 rec	Database*	85	-	86	83	-	-	-	-	60
	[186]	35 rec	Database*	94	-	95	93	94	-	-	-	60
	[187]	188 sub	Hospital	-	-	-	-	72	89	80	59	-
	[188]	35 rec	Database*	84	-	-	-	-	-	-	-	60
	[189]	35 rec	Database*	85	92	75	91	-	-	-	-	60
	[190]	70 rec	Database*	-	-	-	-	93	-	97	99	60
	[191]	35 rec	Database*	85	91	83	82	-	-	-	-	60
	[192]	35 rec	Database*	87	-	82	91	-	-	-	-	60
	[193]	35 rec	Database*	86	94	83	88	97	100	96	100	60
	[194]	35 rec	Database*	89	-	88	91	-	-	-	-	60
	[195]	9 rec	Database+	-	-	-	-	95	-	100	80	-
	[196]	35 rec	Database*	85	92	82	87	97	-	-	-	60
	[197]	69 rec	Database*	-	-	-	-	98	-	98	100	-
	[198]	10 rec	Database*	-	-	-	-	98	-	-	-	-
	[199]	17 rec	Database*	100	-	-	-	-	-	-	-	60
Respiration	[200]	41 sub	Hospital	-	-	-	-	-	88	81	95	300
	[201]	12 sub	-	89	-	87	90	-	-	-	-	30
	[202]	148 sub	Hospital	-	-	-	-	82	90	88	71	-
	[203]	14 rec	Database#	-	-	-	-	93	-	-	-	60
	[204]	100 rec	Database^	-	-	-	-	-	-	84	-	60
	[205]	70 rec	Database*	91	96	88	96	-	-	-	-	60
	[206]	4 sub	Hospital	82	-	86	81	96	-	-	-	-
	[207]	8 rec	Database*	99	-	-	-	-	-	-	-	60
	[208]	6 sub	-	88	-	91	77	-	-	-	-	40
	[209]	8 rec	Database*	99	-	-	-	-	-	-	-	-

	[210]	100 rec	Database~	75	-	-	-	-	-	-	-	30
Sound	[211]	40 sub	Hospital	-	-	88	82	-	-	-	-	-
	[212]	80 sub	Hospital	-	-	-	-	81	-	78	85	-
	[213]	87 sub	-	-	-	-	-	-	-	81	83	60
	[214]	41 sub	Hospital	-	-	-	-	90	97	89	92	-
	[215]	50 sub	-	97	-	-	-	-	-	-	-	-
	[216]	40 sub	Hospital	-	-	-	-	-	-	85	75	-
	[217]	186 sub	Hospital	-	-	-	-	86	-	-	-	-
	[218]	33 sub	Hospital	-	-	-	-	76	-	-	-	-
	[219]	10 sub	-	-	-	-	-	-	-	93	100	-
Combined approaches	[220]	120 sub	Hospital	-	-	-	-	89	-	94	82	-
	[221]	15 sub	-	-	-	91	86	-	-	-	-	-
	[222]	83 sub	-	-	-	-	-	95	97	92	97	-
	[223]	148 sub	Hospital	-	-	-	-	89	-	91	83	-
	[224]	66 sub	Hospital	-	-	-	-	-	95	83	91	-
	[225]	106 sub	Hospital	-	-	-	-	-	-	81	98	-
	[226]	66 sub	Hospital	-	-	-	-	-	96	90	86	-
	[227]	25 rec	Database+	82	-	84	81	-	-	-	-	60
	[228]	100 sub	Database^	82	-	70	91	95	-	92	98	60
	[229]	285 sub	Hospital	-	-	-	-	72	73	73	65	-
	[49]	70 sub	Hospital	87	92	73	92	100	-	-	-	300
	[230]	8 rec	Database*	-	-	-	-	-	-	97	-	15
	[231]	35 rec	Database*	-	-	-	-	97	-	-	-	-

* PhysioNet apnea-ECG database [43]

MIT-BIH polysomnography database [232]

+ University college of Dublin sleep apnea database [233]

^ Sleep heart health study database [234]

~ Scaling up scientific discovery in sleep database [235]

By analyzing the table, it is possible to assess that the highest EB accuracy (100%) was attained by evaluating the ECG signal [199]. A similar conclusion was reached for the global classification (100%), although one of the three works which reported the highest performance also used oximetry analysis [49] [176] [179]. The highest EB sensitivity (97%) was reported by a method using oximetry analysis [165], while the highest SB sensitivity (100%) was attained by either oximetry [157] or ECG [179] [180] [195]. The maximum EB specificity (100%) was also attained using either oximetry [156] or ECG [176]. Regarding the best SB specificity (100%), the best results were achieved using either ECG [179] [193] [197] or sound [219]. As a result of the table's analysis it was observed that ECG achieved, on average, the best performance (for most of the metrics) while the methods based on oximetry, respiration and sound, attained the second, third, and fourth best results, respectively.

However, it is relevant to notice that most ECG methods were tested in public databases, which are likely to have cleaner signals that might contribute to increase the algorithm's diagnostic capability [56]. It was observed that higher noise is usually present in the respiration signals. Hence, this can possibly be the reason why the performance for this sensor was lower, although the measured signal is in the foundation of the OSA scoring protocol. It was also verified that a combination of source sensors did not contribute to a significant increment of the classification capability, possibly indicating that one of the sensors is dominating the analysis.

The majority of the methods detect OSA by employing supervised learning algorithms such as k-Nearest Neighbor (kNN), Support Vector Machine (SVM), and Feed-Forward Neural Network (FFNN). However, the methods should have a good performance to complexity ratio to allow an efficient hardware implementation.

3.3. Review of developed HMD for OSA detection

A systematic review, covering the period between 2002 to 2017, was performed to assess the performance of the developed commercial and research based HMD for OSA detection, with the goal of identifying the trends in the field regarding the composition of the devices [57]. The search was conducted in IEEE explorer, PubMed Web of Science, multiple journals, and cited literature in the selected articles.

The considered keywords were: “automated AND sleep apnea”; “home AND monitoring AND apnea”; “validation AND device AND apnea”; “portable AND apnea”. The inclusion criterion was the presentation of a research project or the analysis of commercial devices that have been validated by independent researchers, while the exclusion criterion was the absence of diagnostic elements that allow to evaluate the capability of the device for OSA detection. A total of 117 articles were selected, comprising 25 research projects and 50 commercial devices. The distribution of the review articles by year of publication is presented in Figure 3.5.

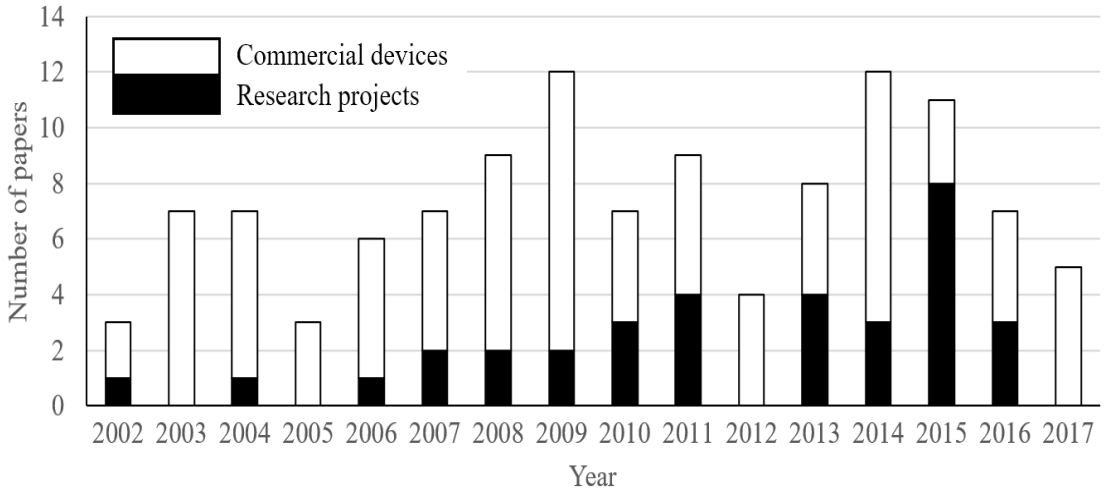


Figure 3.5. Distribution of reviewed articles, in the review of HMD for OSA detection, by year of publication [57].

The articles were divided into the same five categories according to the source sensor, employed in the overview of methods for OSA detection, and the same performance metrics were considered. Evaluation of the articles was performed by considering ten indicators [236]:

- Blind comparison: reference analysis, PSG, and HMD were scored separately and without awareness of the outcomes of each other.
- Consecutive patients: investigators did not participate in the subject’s selection.
- Reference standard performed on all subjects: all subjects were tested by both HMD and the reference test.
- Prospective recruitment of subjects: data that was analyzed by the HMD and reference tests were obtained directly from the subjects.
- Random order of testing: subjects were randomly assigned to perform the tests.
- Small loss of data: less than 10% of data was removed.

- High percentage completed: more than 90% of the subjects concluded the study's protocol.
- Methodology/definitions fully described: the used equipment must be characterized, and the methodology employed to classify the events must be indicated.
- HMD methodology/definitions fully described: the used equipment must be characterized, and the methodology employed to classify the events must be indicated.
- HMD scoring fully described: indication if the scoring was automated or manual.

The first three indicators were used do define the Evidence Level (EL), qualified from I to IV according to the number of criteria that were fulfilled: I- meets three; II- meets the first and third; III- meets the first and second; IV- PSG was not applied independently or blindly. The remaining seven indicators specify the Quality Rating (QR), defined as either a (the study meets all or misses only one of the seven criteria), b (the study does not meet two of the seven criteria), c (the study does not meet three of the seven criteria) or d (the study does not meet four or more of the seven criteria).

Due to the large number of sensors combinations that a device can employ, a descriptive characterization based on six categories, named SCOPER, was employed [237]:

- 1-Sleep: (S_1) EOG, three EEG channels and EMG; (S_2) less than three EEG channels, including or not EOG or EMG; (S_3) sleep surrogate, typically using actigraphy; (S_4) other sleep measures.
- 2-Cardiovascular: (C_1) more than one ECG lead; (C_2) peripheral arterial tonometry; (C_3) one ECG lead; (C_4) derived pulse, usually from oximetry; (C_5) other cardiac measures.
- 3-Oximetry: (O_1) oximetry, either on finger or ear, with sampling indication; (O_{1x}) oximetry, either on finger or ear, without sampling indication; (O_2) oximetry on alternative site; (O_3) other oximetry.
- 4-Position: (P_1) video or visual position measurement; (P_2) nonvisual position measurements.
- 5-Effort: (E_1) two respiratory inductance plethysmography belts; (E_2) one respiratory inductance plethysmography belt; (E_3) derived effort; (E_4) other effort measurements.
- 6-Respiratory: (R_1) nasal pressure and thermal device; (R_2) nasal pressure; (R_3) thermal device; (R_4) end-tidal CO2; (R_5) other respiratory measurements.

A new category, 7, was introduced in the review to cover sound recording devices [57].

- 7-Audio: (A_1) recorded with a microphone; (A_2) other sound recording device.

When the article does not specify the required information for the full categorization, an 'x' replaces the category number. The device automatization (Aut) was also introduced as a categorization metric, defining as [57]: Full-Automated diagnosis (FA); automated diagnosis with some level of manual intervention (SA); Manual Diagnosis (MA). The

summary of the analyzed articles is presented in Tables 3.6 (for research projects) and 3.7 (for commercial HMD).

Table 3.6: Review of the HMD research projects for OSA detection [57].

Article	*EL*	*QR*	*Cat*	*Pop*	*AHI*	*EB Acc (%)*	*EB AUC (%)*	*EB Sen (%)*	*EB Spe (%)*	*Global class (%)*	*SB AUC (%)*	*SB Sen (%)*	*SB Spe (%)*	*TW (s)*	*Aut*
[238]	IV	c	A_1	30	-	-	-	-	-	95	-	-	-	-	FA
[239]	IV	b	A_1	383	5	-	-	-	-	-	94	93	67	-	FA
					15	-	-	-	-	-	97	79	95		
[240]	II	b	E_4	13	-	-	-	89	95	-	-	-	-	30	SA
[241]					5					-	-	94	100		
	II	a	C_3O_1	59	10	85	-	51	87	-	-	82	91	30	FA
					15					-	-	74	96		
[242]	IV	c	C_3	60*	-	-	-	-	-	95	-	-	-	60	FA
[104]					5	-	-	-	-	-	86	89	92		
	I	a	O_1A_1	40	10	-	-	-	-	-	88	82	91	-	FA
					20	-	-	-	-	-	96	100	97		
[243]	IV	c	O_1	3*	-	93	99	92	94	-	-	-	-	60	FA
[244]	IV	b	E_4	13	-	-	-	88	93	-	-	-	-	30	FA
[245]	IV	d	C_3	-	-	78	-	-	-	-	-	-	-	100	FA
[246]	IV	c	C_3	35*	-	89	-	96	85	-	-	-	-	60	FA
[247]	II	b	A_1	42	5	-	-	-	-	-	-	90	92	-	FA
[248]	IV	d	$C_3O_{1x}R_3$	3	-	-	-	81	81	-	-	-	-	-	FA
[249]	II	b	O_1	40	-	-	-	96	90	90	-	-	-	150	FA
[250]	II	a	O_1	68	5	-	-	-	-	88	-	80	95	90	FA
[251]	IV	b	$S_3O_{1x}A_1$	15	-	-	-	-	-	-	-	100	86	60	SA
[252]					5					96	98	100	75		
	II	b	E_4	26	15	86	-	73	91	96	99	100	92	60	FA
					20					92	98	92	92		
[253]	II	c	R_5	30	-	-	-	89	100	-	-	-	-	10	FA
[254]	IV	c	C_1	70*	-	-	91	-	-	-	-	-	-	60	FA
[255]	II	a	O_{1x}	160	-	79	87	79	78	-	-	-	-	60	FA
[256]	II	b	R_2	5	2	-	-	-	-	-	-	100	86	100	FA
[257]	I	d	E_4	-	-	-	-	98	84	-	-	-	-	30	SA
[258]	II	b	A_1	37	-	-	-	-	-	86	-	-	-	30	FA
[259]	IV	c	$S_3C_2A_1$	121	-	98	94	81	100	88	-	-	-	120	FA
[102]	IV	d	E_4	10	-	-	-	-	-	98	-	-	-	60	FA
[260]	IV	c	C_3	70*	-	83	-	-	-	-	-	-	-	300	FA

* PhysioNet apnea-ECG database [43]

Table 3.7: Review of the commercial HMD for OSA detection [57].

Device	*Paper*	*EL*	*QR*	*Categorization*	*Pop*	*AHI*	*Global Acc (%)*	*SB AUC (%)*	*SB Sen (%)*	*SB Spe (%)*	*Aut*
Apnoescreen-I	[261]	II	a	$S_3C_4O_{1x}P_2R_3$	44	10	75	86	-	-	FA
SleepStrip	[262]	I	a	R_3	288	10	-	81	86	57	FA
						20	-	84	80	70	
						40	-	92	80	86	
NovaSom QSG	[263]	II	a	$C_4O_{1x}E_xR_5A_1$	44	15	-	-	91	83	FA
BREAS SC-20	[264]	II	b	$S_3C_4O_{1x}P_xE_4R_5A_x$	60	5	-	89	98	70	SA
						15	-	98	97	92	
						30	-	95	79	100	
Embletta	[265]	I	a	$O_{1x}P_2E_4R_2$	39	15	-	-	60	-	SA
Apnoescreen-II	[266]	I	b	$S_3O_1P_2E_4R_3$	68	5	79	90	83	89	MA
						10	84	91	79	98	
						15	80	86	68	95	
SNAP	[267]	I	a	$C_4O_{1x}E_xR_5A_1$	60	5	88	95	98	40	SA
						10	83	90	88	74	
						15	80	87	84	76	
NGP 140	[268]	II	a	$S_3C_4O_{1x}P_2E_4R_3A_2$	92	10	93	97	97	85	SA
						15	94	99	92	97	
						30	95	99	98	93	
CPS Nx-301	[269]	II	b	$C_1O_{1x}R_5$	49	5	87	-	100	71	SA
						10	87	-	87	87	
						15	83	-	64	95	
Watch PAT100	[270]	I	a	$S_3C_2O_{1x}$	32	5	-	-	94	80	SA
						15	-	-	96	79	
						35	-	-	83	72	
Remmers Sleep Recorder	[271]	I	a	$O_{1x}P_2R_5A_x$	94	5	77	85	75	81	FA
						10	77	90	68	87	

						15	81	91	63	96		
Edentrace II	[272]	I	a	$C_4O_{1x}P_2E_4R_3A_1$	45	10	-	88	71	92	SA
RUSleeping	[273]	II	b	R_2	25	5	-	94	-	-	FA
						15	-	88	-	-	
						30	-	91	-	-	
ARES Unicorder	[274]	II	a	$S_3C_4O_2P_2E_3R_2A_1$	92	5	-	-	98	84	SA
						10	-	-	97	85	
						15	-	-	92	95	
Flow Wizard	[275]	II	a	R_2	31	8	-	-	100	43	SA
						12	-	-	96	71	
						18	-	-	92	86	
SpiderView	[276]	II	b	C_3R_3	19	20	-	85	71	100	SA
Apnomonitor 5	[277]	II	a	$O_{1x}P_2E_xR_3A_1$	22	15	-	-	95	-	MA
ApneaLink	[278]	II	a	$C_4O_{1x}R_2A_x$	50	5	-	100	100	100	SA
						10	-	100	98	100	
						20	-	100	97	100	
Lifeshirt	[279]	II	a	$S_1C_4O_{1x}R_5$	48	5	-	76	85	67	SA
						10	-	90	92	88	
						20	-	90	85	94	
Somnocheck	[280]	II	b	$C_4O_{1x}P_2R_2A_x$	121	5	-	96	96	65	SA
						10	-	92	91	83	
						15	-	91	81	83	
WristOx 3100	[281]	II	a	O_1	154	5	-	-	89	94	FA
						10	-	-	88	94	
						15	-	-	88	90	
Stardust II	[282]	II	a	$C_4O_{1x}P_2E_4R_2$	80	5	87	95	95	62	SA
						15		95	86	78	
						30		96	74	96	
SleepMinder	[283]	II	b	A_2	157	5	91	86	86	46	FA
						10		94	84	84	
						15		97	89	92	
Morpheus	[284]	II	b	$C_4O_{1x}E_xR_3$	83	20	-	87	86	81	SA
Embletta PDS	[277]	II	b	$C_4O_{1x}P_2E_xR_2A_x$	47	5	-	88	91	60	SA
						10	-	-	75	87	
						15	-	-	63	93	
MediByte	[285]	II	a	$C_4O_{1x}P_2E_4R_2$	73	5	-	94	97	67	MA
						10	-	94	84	91	
						20	-	89	80	95	
ApneaLink Plus	[286]	II	b	$C_3O_{1x}E_xR_2$	25	5	-	94	86	83	SA
						10	-	100	100	90	
ResCare AutoSet	[287]	II	b	$O_{1x}R_2$	452	15	-	89	96	73	FA
SD-101	[288]	II	b	P_2E_4	53	5	-	-	95	60	FA
						15	-	96	88	86	
Watch PAT200	[289]	II	b	$S_3C_2O_{1x}$	75	5	-	91	96	43	SA
						10	-	95	90	69	
						15	-	92	92	77	
Sleep&Go	[290]	II	a	$O_{1x}P_2E_4R_3$	55	5	-	96	92	67	SA
						15	-	85	95	56	
NOX T3	[291]	II	b	$S_3O_{1x}P_2E_1R_2A_x$	32	5	-	-	100	70	FA
						15	-	-	92	85	
Sonomat	[292]	II	a	E_4A_2	60	5	90	94	94	77	SA
						15	90	97	88	91	
						30	97	100	100	96	
SleepView	[293]	I	a	$C_4O_{1x}R_2A_x$	93	5	-	92	80	95	SA
						15	-	92	87	85	
						30	-	98	95	93	
Micromovement Sensitive Mattress	[294]	II	a	$O_{1x}E_4$	131	5	-	98	95	100	SA
						15	-	98	90	97	
						30	-	98	90	95	
Alice PDx	[295]	II	a	$S_3O_{1x}E_4R_1$	71	15	-	80	69	87	MA
						30	-	82	87	66	
Sleep Design APNiA	[296]	II	a	A_1	30	26	91	-	71	93	FA
	[297]	II	b	$S_3C_4O_1P_2E_4R_2A_2$	28	5	82	96	88	73	FA
						15	86	97	70	94	
						30	93	100	100	93	

By analyzing the tables, it was possible to conclude that commercial devices use on average, three or more sensors, while the research HMD typically employs only one sensor. These observations point to a new trend in the research, developing devices that are simple to self-assembly, less invasive, cheaper to build (due to the lower number of sensors), and thus achieve a higher performance-complexity ratio. It was also verified that the combination of oximetry and sound provided the best results in the research based

devices, followed by the respiration examination that was employed, either in combination or separately, by all the commercial devices that have reported the highest diagnostic performance.

By the general analysis of this review, it was possible to infer that the performance of portable monitors (denoting HMD) for OSA classification is consistent and high enough so they can possibly be used as an initial tool for OSA diagnostic, which can substantially improve the accessibility of the general population to OSA examination. This observation is particularly relevant as OSA is one of the most prevalent sleep related disorders, which is commonly undiagnosed.

However, the validation studies for the research based devices presented multiple methodological limitations. Figure 3.6 presents the percentage of articles that report the same EL and QR from both commercial and research based HMD. The articles that analyzed commercial devices mainly have an EL of II, while the studies that have proposed research projects vary between II and IV. Regarding the QR, the works that studied commercial devices are mostly in the a and b categories, while the research based HMD are distributed across the four categories. Consequently, the results of the publications that analyzed commercial HMD devices are more reliable, pointing out the need for the research projects to improve the validation procedure to improve the EL and QR [57].

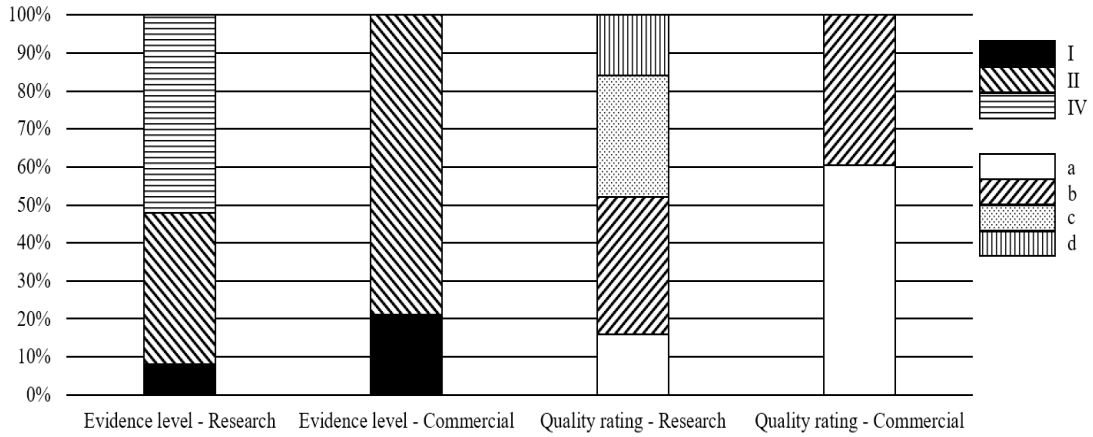


Figure 3.6. Analysis of the EL and QR of the articles [57].

3.4. Key remarks

Taking into consideration the lack of consensus regarding the sleep quality definition, three main approaches were identified as suitable paths to perform the sleep quality examination. The first is the most conventional and considers the examination of sleep quality metrics. From the analysis performed in this chapter it was decided to evaluate the CAP metrics for sleep quality assessment since it is a sleep stability metric, highly correlated to the occurrence of OSA in SBD patients. Therefore, approaches based on the EEG sensor must be evaluated since CAP is a characteristic pattern from this sensor. It was also observed a gap in the literature regarding the implementation of CAP based analysis in HMD. It was conceptualized that this gap can be addressed by developing a HMD for CAP assessment based on the EEG monopolar derivation signal analysis. Using

only one signal reduces the number of sensors that need to be assemble in the patient's body, providing a more conformable experience during the test and also possibly allowing the subjects to easily self-assemble the HMD.

The second approach was to evaluate the presence of sleep related disorders, theorizing that the presence of such disorders may be the main contributor for poor sleep quality. OSA is a prime candidate to be examined in this study as it was identified as one of the most prevalent and frequently undiagnosed sleep related disorders which can significantly affect the quality of sleep. From the reviewed literature it was possible to conclude that research based devices are focusing the analysis in the oximetry and sound signals while most of the commercial HMD use respiration signals. However, it was verified, on the methods review, that ECG and oximetry based algorithms achieved the best performance. Therefore, a compromise in the research is the proposal of methods to examine the ECG or oximetry signals, having the developed algorithms implemented in HMD whose validation process agrees with all the ten indicators that define the EL and QR.

The last approach is to combine the estimation of sleep quality metrics and the detection of sleep related disorders to provide a better view of global sleep quality. It was theorized that a combination of both CAP and OSA assessment can possibly be attained using only one sensor that has relevant information for both calculations. Such analysis can possibly be performed by evaluating the ECG signal, considering CAP in a broader context (designating instability of sleep), through the proposal of a new sleep quality model. The development of an algorithm based on this new concept can be implemented in a HMD, providing a simple tool which can be significant for the future of healthcare since several sleep quality metrics and sleep related disorders can be combined and evaluated in a single examination.

The main limitation of these reviews was the impossibility to examine all published literature, such as the articles which were not published in English or the articles outside the search period. The comparison between the results reported by the different articles is challenging to perform since these works have examined diverse populations and reported different performance metrics.

4. Materials for research

A summary of the employed materials is indicated in this chapter. Specifically, this chapter presents the evaluated databases, the classifiers which were used in the developed methods, the employed feature selection methods, and the methodologies and metrics used for the performance assessment.

4.1. Evaluated databases

A total of four databases were examined in this work. The public database for CAP assessment (and also used by the reviewed articles) was the CAP Sleep Database (CAPSD) from PhysioNet [19]. This database has annotations regarding the macro and microstructure of sleep (annotation indicating the occurrence of an A phase, the corresponding subtype, and the duration of the event) made by a team of neurologists from the Ospedale Maggiore of Parma, Italy. EEG was recorded using the 10-20 international system and one of the monopolar derivations signals (C4-A1 or C3-A2) was used by all developed methods for CAP estimation from the EEG signal. The single-lead ECG signal is also available for several subjects and was employed by all developed methods for indirect CAP estimation from the ECG signal.

Three other databases were evaluated by the method proposed in section 6.2 to allow a comparison between standard sleep quality metrics and the indirect CAP estimation. These databases are: St. Vincents university hospital / University College Dublin Sleep Apnea Database (UCDSAD) from physionet [19] [233]; MrOS Sleep Study (MrOSSS) from National Sleep Research Resource (the database has only recordings from males with 65 years or older thus, 50 recordings were randomly chosen from the 2911 available) [235] [298]; database of collected PSG signals, recorded at Dr. Negrín University Hospital (DrNUH). The last database (DrNUH) was also used for the development of the methods for OSA detection, presented in chapter 7.

General characteristics of the examined databases' subjects are presented in Table 4.1. Further details for each subject of the CAPSD are presented in Table 4.2 as this database is the most relevant for the developed methods. In this table, subjects 1 to 16 are indicated as normal (do not suffer from sleep related disorders) while subjects 17 to 20 suffer from sleep-disordered breathing. The CAP rate estimated from the provided CAP A phase labels was compared with the age-related CAP rate percentages in healthy subjects to assess the sleep quality.

Table 4.1: General characteristics of the evaluated databases' subjects (metrics based on the average of all subjects).

Database	*Measure*	*Mean ± standard deviation*	*Range (minimum – maximum)*
CAPSD (9 females and 11 males)	Age (years)	40.00 ± 16.59	23 – 78
	NREM sleep (min)	338.05 ± 52.96	219 – 444
	REM sleep (min)	96.28 ± 42.56	8 – 190
UCDSAD (4 females and 21 males)	Age (years)	71.30 ± 6.30	65 – 78
	NREM sleep (min)	261.20 ± 39.80	196 – 314
	REM sleep (min)	60.32 ± 31.06	10 – 97
MrOSSS (50 males)	Age (years)	75.30 ± 5.47	68 – 90
	NREM sleep (min)	278.58 ± 68.75	124 – 531
	REM sleep (min)	69.04 ± 31.77	0 – 143
DrNUH (16 females and 54 males)	Age (years)	52.29 ± 12.31	18 – 78
	NREM sleep (min)	247.54 ± 52.20	146 – 398
	REM sleep (min)	53.17 ± 30.23	0 – 114

The A phase annotations were used to perform the A phase binary classification, considering the events as either A (when the examined epoch corresponds to an A phase annotated in the database) or not-A (when the evaluated epoch corresponds to periods without any A phase annotation in the database). By examining Table 4.2 it is notorious that the number of seconds scored as an activation event is significantly lower than the TST (given by the total NREM and REM sleep). Such occurrence leads to an unbalanced dataset with substantially more not-A than A epochs. This unbalance is characteristic in CAP analysis, although the subjects suffering from SBD usually have a longer A phase duration than the subjects without a sleep related disorder.

Table 4.2: Characteristics of the evaluated subjects from CAPSD. The average value for each metric is presented.

Subject	*Age (years)*	*Gender (male, M, or female, F)*	*NREM sleep (s)*	*REM sleep (s)*	*Number of A1 subtypes*	*A1 subtypes duration (s)*	*Number of A2 subtypes*	*A2 subtypes duration (s)*	*Number of A3 subtypes*	*A3 subtypes duration (s)*	*Number of A phases*	*A phase duration (s)*	*CAP cycle duration (s)*	*CAP rate*	*Sleep quality (good, G, or poor, P)*
1	37	F	26040	7170	363	2217	94	747	80	1135	537	4099	12139	0.47	P
2	34	M	21180	4530	186	1188	72	688	94	1239	352	3115	7642	0.36	G
3	35	F	20310	5640	141	656	106	631	108	1043	355	2330	7167	0.35	G
4	25	F	17370	6270	192	986	43	356	52	893	287	2235	6165	0.35	G
5	35	F	22980	6960	462	2863	24	328	60	784	546	3975	11601	0.50	P
6	31	M	21510	7950	303	1871	115	976	94	1414	512	4261	12196	0.57	P
7	31	M	20160	7440	307	1616	99	565	42	480	448	2661	9043	0.45	P
8	42	F	20310	5940	205	949	79	465	160	1868	444	3282	9310	0.46	P
9	31	M	20520	6780	212	1036	56	377	48	678	316	2091	6226	0.30	G
10	23	M	17280	6540	164	1489	34	336	61	922	259	2747	5000	0.29	G
11	28	F	18480	11430	235	1724	80	583	50	796	365	3103	8554	0.46	P
12	29	M	19830	8940	187	1064	28	153	46	573	261	1790	4453	0.22	G
13	24	F	17790	5490	293	1628	118	1040	76	1041	487	3709	9618	0.54	P
14	35	F	19380	4950	195	1035	128	1234	74	1209	397	3478	7615	0.39	P
15	34	M	22200	5940	249	1449	128	1046	95	1244	472	3739	10059	0.45	P
16	41	F	22080	7440	322	2252	137	1138	71	891	530	4281	12765	0.57	P
17	65	M	13260	3270	208	1059	10	73	57	779	275	1911	5744	0.43	G
18	77	M	24630	1140	102	684	81	846	352	6472	535	8002	18799	0.76	P
19	78	M	16950	480	79	557	90	1297	222	5698	391	7552	13206	0.78	P
20	65	M	27180	2040	316	1858	90	781	438	7915	844	10554	23306	0.86	P
Mean	40.00	-	20472.00	5817.00	236.05	1409.05	80.60	683.00	114.00	1853.70	430.65	3945.75	10030.40	0.48	-
Standard deviation	16.59	-	3192.77	2545.16	89.89	586.05	36.62	349.80	103.37	2086.68	134.79	2195.16	4528.12	0.16	-
Minimum	23.00	-	13260.00	480.00	79.00	557.00	10.00	73.00	42.00	480.00	259.00	1790.00	4453.00	0.22	-
Maximum	78.00	-	27180.00	11430.00	462.00	2863.00	137.00	1297.00	438.00	7915.00	844.00	10554.00	23306.00	0.86	-
Total	-	-	20472.00	5817.00	4721	28181	1612	13660	2280	37074	8613	78915	200608	-	-

The rules defined Terzano's reference atlas [19] were employed to produce the CAP cycle annotations, from the A phase annotations provided by the database, using a Finite State Machine (FSM) developed in this work. However, for the indirect CAP assessment a minute-by-minute classification was performed. Thus, a threshold must be used to define the minimum CAP time (duration of a CAP event) to designate a minute epoch as CAP. The data flow diagram of the employed algorithm for label creation is presented in Figure 4.1. *Database_Label* denotes the labels provided by the database, *CAP_cycle* are the one second epochs scored as a CAP cycle (used for CAP analysis based on the examination of EEG signals), *Threshold* is the minimum CAP time duration, and *CAPm* has the labels for the minute-by-minute classification.

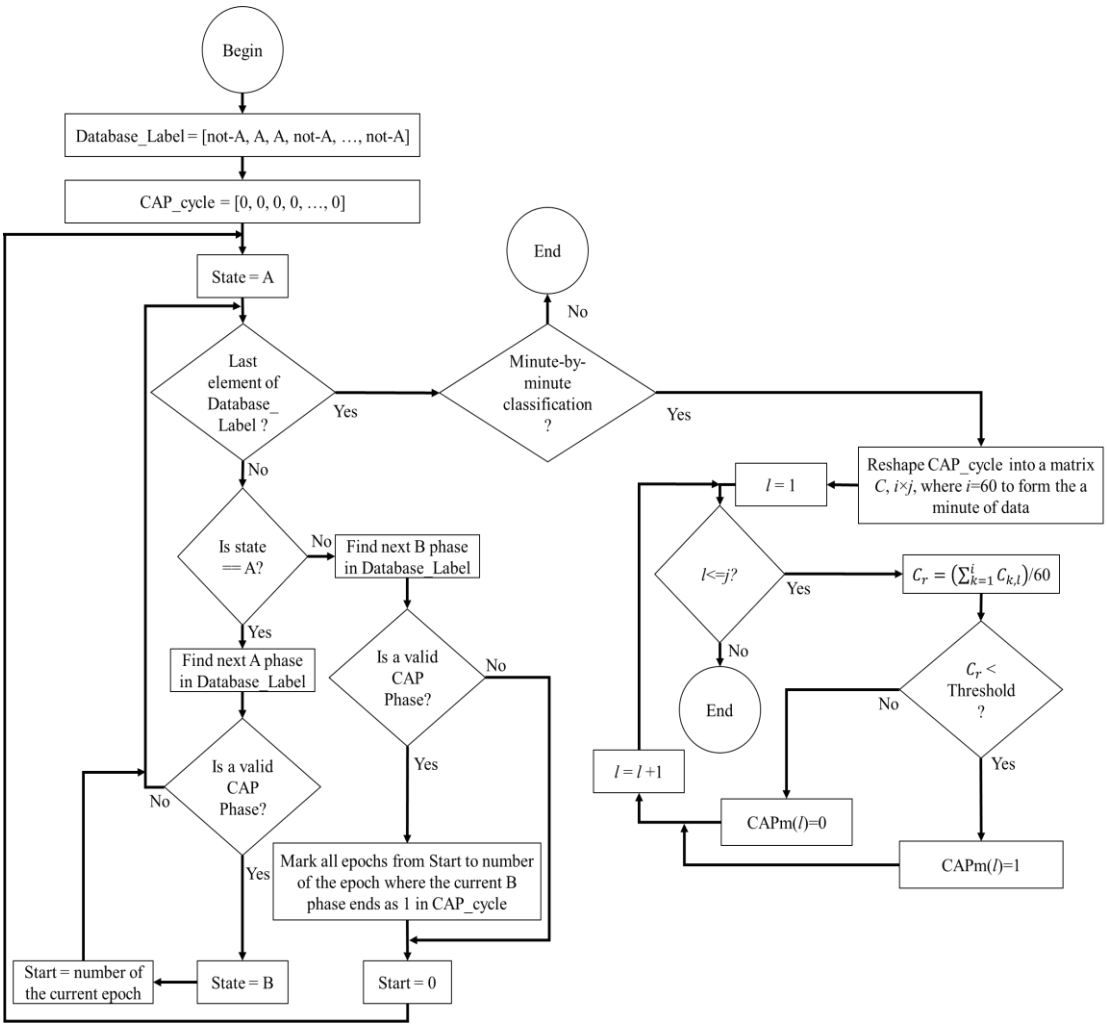


Figure 4.1. Flow diagram of the algorithm developed to create the global sleep quality labels. Adapted from Mendonça et al. [299].

The flow diagram of the algorithm developed to create the global sleep quality labels is presented in Figure 4.2. These labels were produced by comparing the CAP rate of the subject with the age-related CAP rate percentages in healthy subjects. If the subject's CAP rate was higher, then the sleep quality was considered as poor. Otherwise, it was considered as good. This analysis was validated by the fact that the CAP rate is characterized by a low night-to-night intra-individual variability, despite the complex changes this metric undertakes through the life period of a person [10].

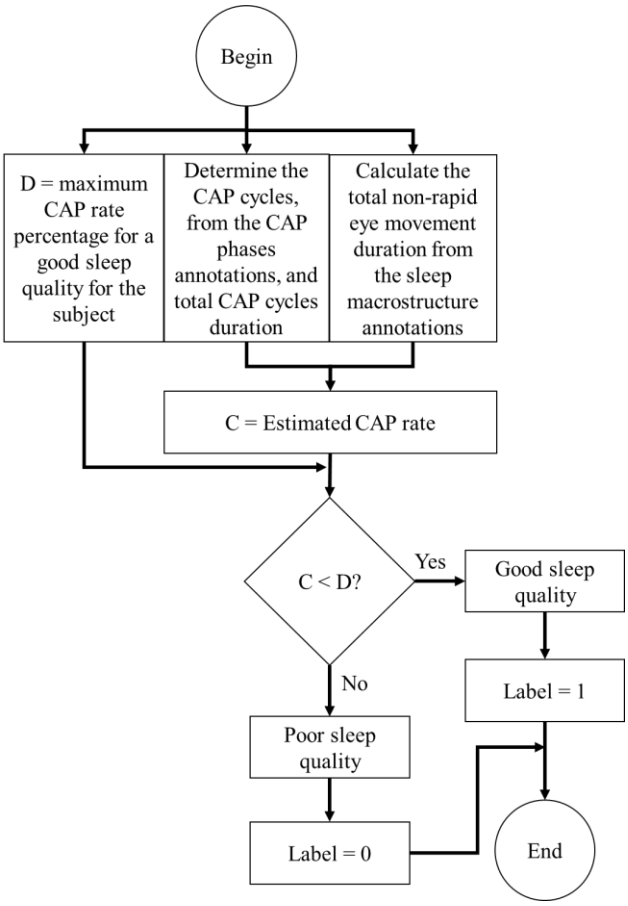


Figure 4.2. Flow diagram of the algorithm developed to create the global sleep quality labels [299].

A total of five versions of the CAPSD were evaluated by different developed methods:

- The first database was composed of EEG signals from nine normal subjects (subjects 1, 2, 3, 5, 6, 7, 9, 10, 11), four subjects with sleep-disordered breathing (subjects 17 to 20), and from one extra subject with bruxism. The concept for this database was to have a more balanced representation between subjects with and without a sleep related disorder. It was used by the methods presented in sections 5.2 and 5.3.
- The second database was composed of EEG signals from 15 normal subjects (subject 12 was not used because the synchronization between the labels and the events in the EEG signal had errors). This database was used to assess the viability of the first deep learning model proposed in section 5.3, the performance of the sleep model developed in section 5.4 (subjects suffering from sleep-disordered breathing were not considered because their signals have a longer CAP cycle duration, leading to significant overestimation of the CAP cycles in the subjects without sleep related disorders), and the viability of the models proposed in sections 5.5 and 5.6.
- The third database was composed of EEG signals from 15 normal subjects (excluding subject 12) and from the four subjects with sleep-disordered breathing (subjects 17 to 20). This database was employed to assess the performance of the methods without an explicit feature creation procedure proposed in section 5.3

(with exception of the first presented model), and the performance of the models proposed in sections 5.5 and 5.6.

- The fourth database was composed of ECG signals from the same subjects of the first database (following the same concept as the first database regarding the most balanced representation) and was employed to assess the performance of the methods proposed in chapter 6.
- The fifth database was composed of ECG signals from 15 normal subjects (subject 16 does not have the ECG signal available) and from the four subjects with sleep-disordered breathing (subjects 17 to 20). This database was employed to assess the performance of the sleep quality model proposed in chapter 8.

The sampling frequency of the examined records varied between 100 Hz and 512 Hz. Hence, for the methods without an explicit feature creation procedure, employed for the EEG signal examination, the records were resampled (in the preprocessing procedure) at the resolution of the subject with the lowest sampling frequency (present in the examined database) to allow the development of a device-independent estimation of the EEG signal [300]. The resampling was performed by decimation to avoid aliasing [301], using a constant reduction factor for the sampling rate, *k*, and a standard low-pass filter, implemented by an order 8 Chebyshev type one filter, with passband ripple of 0.05 dB and a normalized cutoff frequency of 0.8/*k*.

Afterwards, the resampling procedure selected each *kth* point from the filtered signal to produce the resampled signal. This way, a uniform input was attained. All the single-lead ECG signals examined in this work were resampled at 200 Hz since this frequency was found to be suitable for the QRS examination using the algorithm proposed by Pan and Tompkins [302]. The resampling was performed by decimation or interpolation.

4.2. Examined classifiers

Several classifiers were evaluated with the proposed methods, comprising a mixture between the classifiers suggested in the state of the art (identified as suitable for the intended classification) and classifiers which were not previously examined by any work in the state of the art (regarding CAP analysis) but were tested in this work as they have attained good performance in other fields. The studied classifier were:

- Linear Discriminant Analysis (LDA) in sections 5.2 and 6.1.
- Quadratic Discriminant Analysis (QDA) in section 6.1.
- Logistic Regression (LR) in sections 5.2, 6.1, and 7.1.
- Ensemble of decision Trees (ET) in sections 5.2 and 6.1.
- SVM in sections 5.2, 6.1, and 6.4.
- FFNN in sections 5.2, 5.3, and 6.1.
- Cascade-Forward Neural Network (CFNN) in section 5.2.
- kNN in sections 5.2 and 6.1.
- k-Means Clustering (kMC) in sections 5.2 and 6.1.
- Self-Organizing Map (SOM) in sections 5.2 and 5.4.
- Deep Stacked Autoencoder (DSAE) in sections 5.3 and 6.1.
- Long Short-Term Memory (LSTM) in section 5.3, 5.5, and 5.6.

- Gated Recurrent Unit (GRU) in section 5.3.
- Convolutional Neural Network (CNN) with one dimensional input (1D-CNN) in sections 5.3, 5.6, 6.2, 7.2, and 8.1.

4.2.1. Discriminant Analysis

LDA, a supervised learning classifier, was used for the A phase classification by evaluating features crafted from the EEG signal. This classifier first computes the average of each class and then determines the covariance, Σ. Thus, each class has a different mean but the same covariance matrix [303]. For the binary case, the decision boundary between the two classes is given by [303]

$$P(y = 1|\boldsymbol{X}, \theta) = sigm\left(\boldsymbol{W}^T(\boldsymbol{X} - x_0)\right) \quad 4.1$$

where *sigm* is the sigmoid function, defined as [303]

$$\sigma(\beta) = \frac{1}{1 + e^{-\beta}} \quad 4.2$$

y is the output class, $\boldsymbol{X}$ are the inputs, $\boldsymbol{\theta}$ are the model parameters, $\boldsymbol{W}$ is the projection vector, defined as [303]

$$\boldsymbol{W} = \boldsymbol{\Sigma^{-1}}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) \quad 4.3$$

and x_0 is the amount of displacement for the classification [303]

$$x_0 = \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_0) - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)\frac{\log\left(\frac{P(y=1)}{P(y=0)}\right)}{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T\boldsymbol{\Sigma^{-1}}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)} \quad 4.4$$

Therefore, the decision rule is to shift $\boldsymbol{X}$ by x_0, project onto the line defined by $\boldsymbol{W}$, and verify if the result is either positive or negative. QDA is similar to LDA but a covariance matrix is separately estimated for each class [303].

4.2.2. Based on the logistic function

LR considers that the posterior probability of a class, C, is defined as the logistic sigmoid function, σ, action on a linear function of the feature vector $\boldsymbol{\varphi}$ such that [304]

$$P(C = 1|\boldsymbol{\varphi}) = \sigma(\boldsymbol{W}^T\boldsymbol{\varphi}) \quad 4.5$$

where $\boldsymbol{W}$ is the model's weight vector.

4.2.3. Examination of trees

The ET, also named Classification and Regression Trees (CART), although only the Classification Trees (CT) were used, are characterized by a recursive partition of the input space, producing a local model in each resulting region of the input space. Therefore, a CART can be seen as a tree, with one leaf per region, forming an adaptive basis-function model specified as [303]

$$f(x) = \sum_{m=1}^{M} W_m \phi(x, v_m) \quad 4.6$$

where the basis functions, ϕ, describe the region m, and the weights, W_m, identify the response value of the region with the parameters v. Each CART can be interpreted as a weak learner and grouped to form an ensemble of decision trees using an ensemble algorithm. Three ensemble methods were tested, specifically:

- Bootstrap aggregation (bagging), where every tree was grown on an independently drawn bootstrap replica of the input, with the random forest algorithm (searches for the best features, amongst a random subset of features, for splitting a node).
- Adaptive Boosting (AdaBoost), trains the weak learners sequentially for all observations, N, and the weighted classification error is determined for each learner, e, by [303]

$$W_e = \sum_{n=1}^{N} w_n{}^{(i)} R(y_n \neq c_e) \quad 4.7$$

considering $w_n{}^{(i)}$ the weight of observation n at the iterative step i, and R is the binary indicator function that is 1 when the true class of the observation, y_n, differs from the predicted class of the learner c_e, and 0 otherwise. The algorithm increases the weights of the misclassified observations, by the learner e, and reduces for correctly classified. The next learner is then trained with the updated weights. The algorithm estimates the classes for new data using [303]

$$f(x) = \frac{1}{2} \sum_{e=1}^{E} \log\left(\frac{1 - W_e}{W_e}\right) c_e \quad 4.8$$

Linear programming boost iteratively maximizes the minimal margin (classification margin defined by the difference between the predicted the true class) through a sequence of linear programming problems [305].

- TotalBoost iteratively maximizes the minimal margin in the training set using quadratic programming [306].

4.2.4. Support vectors

SVM represents the input in a multi-dimensional space to be classified by the discriminant hyperplane [303]

$$f(\boldsymbol{X}) = \boldsymbol{W}^T \boldsymbol{X} + b \quad 4.9$$

where b is the bias. A soft margin approach was used since the data does not allow to define a hyperplane that fully separates the two classes. Hence, slack variables, $\boldsymbol{\zeta}$, and a regularization constant, C, were used in the optimization objective, relaxing the hard-margin constraints, and thus, allowing for some classification errors. The analyzed problem can be defined as [303]

$$\min_{w,b,\zeta}\left(\frac{1}{2}\boldsymbol{W}^T\boldsymbol{W}+C\sum_{i=1}^{I}\boldsymbol{\zeta}_i\right) \quad 4.10$$

and can be simplified using Lagrange multipliers, α. Finding a stationary point, and setting the gradient to zero, produced the dual form [307]

$$\max_{\alpha}\left[\sum_{i=1}^{I}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{I}y_iy_j\,\alpha_i\alpha_jK\left({x_{i\{1\dots L\}}}^T,x_{j\{1\dots L\}}\right)\right] \quad 4.11$$

under the constraints [307]

$$\begin{cases}\sum_{i=1}^{I}\alpha_iy_i=0\\ 0\le\alpha_i\le C,i=1,\dots,I\end{cases} \quad 4.12$$

where K is the kernel that maps the data into other space. Two kernels were tested, the Gaussian radial basis function, [307]

$$K=e^{\left(\frac{-\left\|x_{i\{1\dots L\}}-x_{j\{1\dots L\}}\right\|^2}{c}\right)} \quad 4.13$$

with c specifying a constant, and the linear [307]

$$K={x_{i\{1\dots L\}}}^Tx_{j\{1\dots L\}} \quad 4.14$$

4.2.5. Nearest neighbors

kNN is a non-parametric classifier that determines the K points in the training set that are closer to the test input x to classify the data into the class c with a probability given by [303]

$$P(y=c|x,D,K)=\frac{1}{K}\sum_{i\in N_K(x,D)}R(y_i=c) \quad 4.15$$

defining N_K as the indices of the K nearest points to x in the training dataset D and R is a binary indicator variable that is 1 if y_i belongs to the class c and 0 otherwise.

4.2.6. Cluster analysis

kMC identifies groups (named clusters) of data points $\boldsymbol{X}$ in a multidimensional space by analyzing an objective function, given by [304]

$$J=\sum_{n=1}^{N}\sum_{k=1}^{K}r_{nk}\|x_n-\mu_k\|^2 \quad 4.16$$

where K is the number of clusters, N is the number of observations, μ represents the cluster center, and r is a set of binary indicator variables such that r_{nk} is 1 if x_n is assigned to the

cluster *k*. Otherwise, r_{nk} is 0. *J* represents the sum of the squares of the distances of each x_n to the assigned cluster. Therefore, the objective is to minimize *J* through an iterative procedure by finding the optimal values for r_{nk} and μ_k.

4.2.7. Neural networks with supervised learning

FFNN is composed of an input layer, where inputs are fed into the model, an output layer, which provides the final manipulation of the data, and Hidden Layers (HL) between the input and output layers. Multiple HL can be sequentially stacked, and each layer receives the data from the previous layer, manipulates the data, and fed the output to the next layer. Each neuron applies a transfer function, φ, that takes into consideration the number of inputs, *n*, that are connected to the neuron, the weight of each connection, *w*, and the bias. Therefore, the neuron output is given by [308]

$$y = \varphi\left(\sum_{j=1}^{n} x_j \times w_j\right) + b \tag{4.17}$$

Both the hyperbolic tangent transfer function, defined as [308]

$$\varphi(\beta) = \frac{2}{(1 + e^{-2\beta})} - 1 \tag{4.18}$$

and the sigmoid transfer function were tested. Two training algorithms were studied, specifically, the Levenberg-Marquardt algorithm [304] and the scaled conjugate gradient [309]. CFNN is similar to the FFNN, but employs a direct connection from the input layer to the output layer [310].

The LSTM and GRU are variations of the Recurrent Neural Networks (RNN). Each memory cell of the LSTM model, at time step *t*, is controlled by three gates specifically, the forget gate (*F*), the input gate, and the output gate (*O*). Each input signal, x_t, was placed in the cell candidate, *P*, so that it could update the cell state, c_t, through [311]

$$c_t = F_t c_{t-1} + I_t P_t \tag{4.19}$$

and the hidden state of the model is defined by

$$h_t = O_t tanh(c_t) \tag{4.20}$$

where *tanh* is the hyperbolic tangent function [304]. Each gate and cell candidate has a specific bias (*B*) and weights for the input, *W*, and for the recurrence, *R*.

The input and output gates, respectively, perform the operations [311]

$$I_t = \sigma(W_I x_t + R_I h_{t-1} + B_I) \tag{4.21}$$

$$O_t = \sigma(W_O x_t + R_O h_{t-1} + B_O) \tag{4.22}$$

where σ is the sigmoid function [304], to control the flow of activations into the cell, and from the cell to the rest of the network. The forget gate adaptively resets the memory and perform a scaling of the internal state of the cell through

$$F_t = \sigma(W_F x_t + R_F h_{t-1} + B_F) \tag{4.23}$$

The cell candidate, used as input of the cell state update, also performs a scaling of the internal state by

$$P_t = \sigma(W_P x_t + R_P h_{t-1} + B_P) \quad 4.24$$

The GRU is composed of reset and update gates. The first gate decides if the previous hidden state should either be used or ignored (dropping the information to provide a more compact representation) while the second gate regulates the quantity of information that will be considered in the current hidden state (from the previous hidden state). Both gates perform a scaling operation and the activation vectors are respectively defined by [312]

$$S_t = \sigma(W_S x_t + R_S h_{t-1} + B_S) \quad 4.25$$

$$U_t = \sigma(W_U x_t + R_U h_{t-1} + B_U) \quad 4.26$$

A candidate activation, *M*, is composed of network's inputs and the learned information through [312]

$$M_t = tanh(W_M x_t + R_M S_t h_{t-1} + B_M) \quad 4.27$$

where [312]

$$h_t = (1 - U_t) h_{t-1} + U_t M_t \quad 4.28$$

The group of Group of Layers (GL) concept was proposed in this work as a way of reducing the amount of simulation required to optimize the CNN architecture, where each GL was either composed of a convolution, followed by a normalization, and a pooling operation or a convolution, followed by a subsampling (pooling) operation. This way, the developed searching algorithms for CNN optimizations evaluated the introduction of GL instead of introducing individual layers.

The 1D-CNN convolution operations, executed in the convolution layers, can be represented by [308]

$$c_d = \varphi(K_d \circledast \mathrm{X} + B_d) \quad 4.29$$

were *d* is the number of filters (*k*) employed in the layer, *X* are the inputs, and φ is the activation function that was either the Scaled Exponential Linear Unit (SELU) [313] or the Rectified Linear Unit (ReLU) [304]. For some of the developed classifier, the output of the previous layer was normalized at the batch normalization layer to continue a mean activation close to zero with a nearly unitary standard deviation [308].

A pooling operation was executed by the last layer of the GL to decrease the dimensionality of the data. Both the maximum (MaxP) and the average (AveP) operations were tested [308]. The final layer of the network was fully connected (dense) and performed the classification by [308]

$$Y = \varphi(W * \mathrm{X} + B) \quad 4.30$$

where the soft-max function, defined as [303]

$$\sigma(y_i) = \frac{e^{y_i}}{\sum_{m=1}^{R} e^{y_m}} \quad 4.31$$

which considers the probability distribution *y* for the input *i* over the *R* possible results, was used for classification. Some of the examined networks employed another dense layer before the output layer to increase the learning ability of the nonlinear parameters.

4.2.8. Neural networks with unsupervised learning

SOM is a type of neural network that evaluates topographic relationships of the input data, fed to an input node, which is mapped into the output units (or output node) that,

usually, have a pre-defined topology. Each input is fed to each output unit with an associated weight, *W*, and the squared Euclidean distance, between the nodes (n_1, n_2) and a specific point, *i*, from the input vector, *j*, was considered, given by [314]

$$E_d(n_1, n_2, i) = \sum_i \left(w_i(n_1, n_2) - x(j, i)\right)^2 \quad 4.32$$

Each input point is assigned to the node with the minimum E_d, known as the winning node. The weights that are associated with this node, and the weights that are related to the neighborhood nodes are updated according to the Kohonen rule [314]

$$w_i(n_1, n_2) = w_i + l_r F(n_1 - o_1, n_2 - o_1)(x(j, i) - w_i) \quad 4.33$$

where *o* are the data point indices, l_r is the learning rate, and *F* is the neighborhood function that is 1 at the wining node and gradually reduces proportionally to the distance that the neighborhood node is to the wining node until the maximum neighborhood distance.

In this work, the DSAE was composed of stacked autoencoders and an output layer to perform the classification. The model input has *D* samples from $\boldsymbol{X}$, and each autoencoder is composed of an encoder and a decoder. The decoder maps the encoded representation, *Z*, into an estimate of the encoder input vector, *x*, by [308]

$$y = \varphi\left(\sum_{i=1}^{D} z_i \times w_i + b\right) \quad 4.34$$

where φ is transfer function of the decoder, $\boldsymbol{W}$ is the weight matrix of the decoder, and *b* is the bias vector of the decoder. The same equation can be used to produce the encoded representation, replacing *z* by *x*, and considering the transfer function, weighs, and bias of the encoder. Sparsity can be introduced in the model by adding a regularizer to the cost function [308]

$$C = \frac{1}{D}\sum_{i=1}^{D}\sum_{j=1}^{J}\left(x_{ji} - \hat{x}_{ji}\right)^2 + \Gamma \times \Omega_w + \gamma \times \Omega_s \quad 4.35$$

considering each training example, *j*, the L_2 and sparsity regularization terms, respectively Ω_w and Ω_s, and the coefficient for each regularization term, respectively Γ and γ.

The sparsity regularization, defined by the Kullback-Leibler divergence (measures how different two distributions are, having a null value when they are equal and increases its value when they diverge from each other) [315]

$$\Omega_s = \sum_{i=1}^{D}\left[\beta \log\left(\frac{\beta}{\hat{\beta}_i}\right) + (1 - \beta)\log\left(\frac{1 - \beta}{1 - \hat{\beta}_i}\right)\right] \quad 4.36$$

that considers the average activation value of a neuron, given by [315]

$$\hat{\beta}_i = \frac{1}{J}\sum_{j=1}^{J}\varphi\left({w_i}^T \times x_j + b_j\right) \quad 4.37$$

and the desired activation value β, can make each neuron specialized in recognizing specific patterns by responding to some characteristics that only occur in a small subset of the training data. However, a decrease of z associated with an increase of w can lead to a small Ω_s that reduces the effectiveness of the sparsity regularization. Thus, the L_2 regularization, defined as [315]

$$\Omega_w = \frac{1}{2}\sum_{l=1}^{L}\sum_{j=1}^{J}\sum_{i=1}^{D}\left(w_{ij}{}^{(l)}\right)^2 \qquad 4.38$$

was used in $\boldsymbol{W}$ to address this issue. The output layer employed the soft-max function for classification.

4.3. Feature selection

Two feature selection methods were considered in this work for the proposed feature based methods. The first was Sequential Feature Selection (SFSe) which is a classifier dependent method, and two variations of this algorithm were examined. The first was Sequential Forward Selection (SFS) algorithm that initiates with two sets of variables, one empty (the first set) and one with all the features in a random order (the second set). The goal is to order, in the first set, the features by their relevance to the obtain maximization of the considered performance metric.

During the first iteration, the feature that achieved the highest value for the considered performance metric was moved for the first vector and placed in the first position. In the second iteration, the algorithm determined the second most relevant feature as the one that has the best compatibility with the feature on the first set, achieving the highest value for the considered performance metric, and moves the feature to the second position of the first set (placed after the first feature). This process was iteratively repeated until the second set is empty and the first set is full, producing a feature set (first set) ordered according to the feature relevance. The second variation was Sequential Backward Selection (SBS) which starts with a set of all features, selecting the least relevant feature at the end of each iteration (feature that produced the lowest average for the considered performance metric) which was moved from the feature set to the organized-features set that is organized from the least to the most relevant feature. For both variations, the optimal number of features to use was determined by analyzing in which iteration of the algorithm the average for the considered performance metric was the highest.

The second feature selection method was the minimal-Redundancy-Maximal-Relevance (mRMR) algorithm, which is a classifier independent method that evaluates the maximal statistical dependency criterion, based on mutual information M [316]. It starts with a feature set, F, with all features $\{x_i, i=1,\ldots,m\}$ and it is intended to find the highest dependency, among the features, in the target class T. The maximal relevance was attained by ordering the features by the condition $max\ R_l(F,T)$, where [316]

$$R_l = \frac{1}{|F|}\sum_{x_i \in F} M(x_i; T) \qquad 4.39$$

Since the chosen features are likely to have a large dependency among them, then the minimal redundancy condition, $min\ R_d(F)$, was also applied where [316]

$$R_d = \frac{1}{|F|^2} \sum_{x_i, x_j \in F} M(x_i, x_j) \quad 4.40$$

Finally, the mRMR criterion combines the two constraints by maximizing the operator R that is given by [316]

$$R = R_l - R_d \quad 4.41$$

Therefore, the algorithm simultaneously maximizes the relevance and minimizes the redundancy.

These two feature selection methods were chosen to cover the two most frequently evaluated approaches in the state of the art (classifier dependent and classifier independent methods). Although it is expected for the SFSe to attain a better performance (since the selected features are the ones which maximized the evaluated performance metric for the examined classifier), the required time to identify the relevant features, from the total number of features F_n, is expected to be approximately (F_n+1)/2 higher than the mRMR [317].

As a result, the SFSe can be significantly slower than the mRMR for models with a higher number of features or that employ classifiers whose training is computationally demanding. However, it was suggested that the performance difference between the features selected by the two methods might not be relevant [317]. Therefore, SFSe was used for the methods presented in sections 5.2 and 6.3, while mRMR was employed for the algorithms presented in sections 5.5 and 6.1.

4.4. Performance assessment

Four methods for performance assessment were considered in this work. The first was Cross-Validation (CV) with Two Folds (TFCV) where, in each iteration, half of the subjects were used to create the training dataset and the other half used as testing dataset. The subjects which composed the datasets were randomly selected at each iteration. Subject independent results were ensured by only using the data from a subject either in the training or in the testing dataset in each iteration. This methodology was used for the performance estimation of the methods developed in sections 5.2, 6.1, and 6.3.

A CV model with five folds was the second performance assessment methodology, used in section 5.3 since it was a recommended procedure (in the literature) for deep learning at the time when the method was proposed. However, it was subsequently observed that Leave One Out CV (LOOCV) could provide less biased results for classifiers with few samples [318]. Such is particularly relevant for the deep learning models. As a result, the third performance assessment methodology was LOOCV. At each iteration, the signals from all subjects excluding one (left out) were used to create the training dataset, and the left out subject was employed to produce the testing dataset. The data from each subject were either used in the testing dataset or the training dataset (ensuring subject independence) and each subject was only once selected to create the testing dataset. This method is also suitable for global assessment since only one subject at a time compose the testing dataset. The methods developed in sections 5.4, 6.1, and 6.2 employed this approach.

A combination of TFCV and LOOCV was employed for the development of the algorithms proposed in sections 5.3, 5.5, 5.6, and chapter 8, using TFCV as a faster method for the hyperparameters' tuning and LOOCV as a more reliable method for the performance assessment of the model.

The bootstrap (BS) method (using the .632 rule) was the last performance assessment methodology employed by the methods proposed in section 6.1 as BS was found to outperform LOOCV is small datasets [319]. This method produces a dataset with size *m* by creating bootstrap samples through sampling *m* instances uniformly from the data with replacement.

Cost-sensitive learning (apply a lower cost to the misclassification of a majority class element when compared to a minority class element) was employed to minimize the effect of the data unbalance in all supervised learning classifiers (also used during the fine tuning of the weights, performed using supervised learning, at the end of the training of the DSAE). For the models based on unsupervised learning developed in sections 5.2 and 6.1, the training dataset was balanced by oversampling the minority class to have an equal number of examples from each class. Undersampling was employed by one algorythm presented in section 5.3. No balancing operation was used for the development of the proposed sleep model (section 5.4) in order to maintain the original data distribution.

Each iteration of any of the performance assessment methodologies was repeated 50 times to achieve statistically significant results. The evaluated performance metrics consider the True Negatives (TN), True Positives (TP), False Negatives (FN), and False Positives (FP) to estimate the Acc, Sen, and Spe. These metrics are defined by [320]

$$\text{Acc} = \frac{TN + TP}{TN + TP + FN + FP} \tag{4.39}$$

$$\text{Sen} = \frac{TP}{FN + TP} \tag{4.40}$$

$$\text{Spe} = \frac{TN}{FP + TN} \tag{4.41}$$

The diagnostic capacity of the classifiers was assessed by the AUC as it designates how likely the classifier is to rank an arbitrarily chosen positive example higher than an arbitrarily chosen negative example [321].

An early stopping procedure was employed to lessen the simulation time and avoid overfitting the classifier. For all classifications, the training procedure was stopped (before the end of the maximum number of training cycles, defined as 50) if no improvement in the reference performance metric of the validation set was reached within 5 consecutive epochs. Three reference performance metrics were examined. The Combined Objective (CO), defined as (Acc+Sen+Spe)/3, was the first to be evaluated (employed by the methods presented in sections 5.2 and 5.3) with the goal of attaining balanced results by simultaneously maximizing the Acc, Sen, and Spe.

The AUC (second examined reference performance metric) was found to be a better solution as it produced more balanced results (similar Sen and Spe) than the CO in strongly unbalanced dataset (because for CAP analysis an increase in the Spe leads to larger increase in the Acc, when compared with the effect that an increase in the Sen has in the Acc, leading the model to favor a higher Spe).

As a result, AUC was used for classifiers of the remaining chapters performing the epoch-based assessment. The third examined reference performance metric was the Acc and was used for the global assessments performed in sections 6.2 and 8.1. This metric was employed with the goal of maximizing the accuracy of the global sleep quality classification.

4.5. Key remarks

The use of the CAPSD for the development of all methods based on CAP analysis allowed to perform a comparison between the attained results from the different methods. The same observation is valid for the methods for OSA detection as they were developed using the DrNUH. By examining multiple machine learning classifiers, it was possible to cover several potential solutions to identify the optimal methodology. It is relevant to notice that both supervised and unsupervised learning algorithms were tested. Also, conventional machine learning methods and deep learning were examined.

The performance assessment was based on state of the art algorithms and the 50 times repetition of each iteration allowed to achieve statistically significant results. The reported performance metrics are also frequently used in the state of the art, allowing to properly compare the attained results with the results of other works in the field.

Three limitations of the developed work were identified. The first was the multiple variants of the CAPSD, and the different performance assessment methods which were used by the developed works. These variations occurred as the works were published in different journals and conferences to comply with the requests by the reviewers. It was also due to the availability of better graphics processing units, which allowed the deep learning models to be developed in a reasonable time and using more data.

The second limitation was related to the CAPSD where there is a low number of subjects suffering from sleep-disordered breathing in comparison to the number of subjects without a sleep related disorder. This limitation was mitigated by the fact that the recordings were performed during a full night of sleep, hence, the total number of A phase examples from the subjects suffering from sleep-disordered breathing was 2045 with a total duration of 28019 s (each second has at least 100 samples).

The last limitation is also related to the database as the recording was performed with different sampling frequencies, ranging from 100 to 512 Hz. Therefore, a resampling process had to be performed for the models without an explicit feature creation procedure, which can influence the results.

5. Estimation of CAP from EEG signals

The goal for this chapter is to present algorithms for CAP assessment based on the examination of the EEG signal (as CAP is composed of EEG patterns), to assesses sleep quality metrics (based on CAP characteristics) which allow the sleep quality estimation according to the first approach for sleep quality examination, theorized in this work from the state of the art analysis. Two main approaches for CAP assessment are presented in this chapter. The first one is based on features (which were either identified by the performed survey of the methods for CAP analysis or were features proposed in this work), while the one second is based on methods without an explicit feature extraction process.

It was verified that although there is a strong correlation between the macrostructure and the microstructure, the standard time resolution for the macrostructure does not allow to perform a direct evaluation of the microstructure by studying the sleep stages. Therefore, a new sleep model was proposed as a framework for sleep microstructure analysis based on the CAP assessment. The subsequent examination was the A phase subtypes characterization and classification, as these have been considered to be highly relevant for sleep disorder diagnosis. The last examination was the proposal of methods for sleep quality examination based on the CAP assessment. All works evaluated the signal from one EEG monopolar derivation (C3–A2 or C4–A1).

5.1. Methods for CAP assessment

A systematic review was conducted to assess the developed methods to classify the CAP phases or the CAP cycles. The search was performed on IEEE explorer, PubMed databases, Web of Science, various journals, and cited literature in the included articles, using the search keywords: “classification AND cyclic alternating pattern”; “automatic AND CAP AND detection”; “detection AND A-phase AND cyclic alternating pattern”. All published articles to the time when the review was performed were considered. Therefore, the covered period was between 1998 and 2017.

The inclusion criteria were the presentation of an algorithm capable of detecting and classifying the CAP cycles or the A-phase of the CAP, while the exclusion criterion was the absence of diagnostic elements that allow to evaluate the performance. A total of 17 original research articles were selected. A summary of the analysis is presented in Table 5.1 with a description of the employed method, Number of Features (NF), and what is the classification (CL), either the A phase, the subtypes of the A phase or the CAP cycles.

Three main fields were identified for CAP analysis, specifically, the A phase subtype detection (using multiclass classification, either considering only the subtypes A1, A2, and A3, or the subtypes and the absence of activation, not-A), the A phase detection (considering each epoch as either A or not-A), and the CAP cycle estimation (seeing each epoch as either CAP or non-CAP). Two main approaches were identified from the review for CAP classification. The first is based on the detection of the A phases and, afterwards, employ a FSM to estimate the CAP cycles. The second directly classifies the CAP cycles, either by considering created EEG features or directly feeding the EEG signal to a classifier. The first method was found to be more suitable for CAP analysis, in the clinical perspective, as it can provide information about both CAP phases and CAP cycles. The second was found to be significant for the development of the proposed sleep model as it

directly predicts the cycles of instability in sleep. All works which have reported the evaluated EEG channel used a monopolar derivation, and the average studied population was 10 subjects. The reported TW ranged from 0.5 s to 70 s, although 1 s was the most frequently used.

Table 5.1: Review of the methods for CAP estimation from EEG signals [30].

Article	*Population*	*EEG channel*	*CL*	*NF*	*EB Acc (%)*	*EB Sen (%)*	*EB Spe (%)*	*TW (s)*	*Method description*
[322]	10	C3-A2 or C4-A1	A1; A2; A3	4	95; 85; 60	-	-	3 or 7	Manual intervention to define a point inside each A-phase, and inferential statistics to define the phase duration.
[323]	10	F4-C4	A	5	77	84	80	0.5	EEG signal was divided into 5 bands and a descriptor, given by the signal amplitude ratio (2 s – 64 s) / 64 s, was computed. Recognition of the phases used both amplitude and length thresholds.
[324]	10	F4-C4	A	5	84	-	-	0.5	Employed the same algorithm described by Navona et al. [323].
[28]	8	C3-A2 or C4-A1	A	6	69	59	71	1	Produced features using the same algorithm described by Navona et al. [323] and employed tuned thresholds for classification.
				2	72	70	72	3	Created features based on Hjorth parameters applied in the delta band, and employed tuned thresholds for classification.
				1	72	52	76	1	Computed the differential variance (the difference between the current one second window and the previous one), and employed tuned thresholds for classification.
[325]	30*	F3, C3, and A1	A1; A2; A3	6	-	80; 77; 68	83; 73, 74	1	Teager Energy Operator (TEO) was applied to the five EEG bands, and a tune threshold was used for each band.
[326]	6*	C3-A2 or C4-A1	A	-	81	76	81	-	A growing window was applied on the delta band signal (changing from 2 to 60 s), and the similarity of the windowed signal and reference windows were measured using statistical behavior of local extrema features.
[327]	5	C3-A2 or C4-A1	A1; A2; A3	5	90; 43; 80	-	-	2	Temporal, energy, and complexity measures fed a kNN classifier where the number of neighbors and features were chosen by employing a forward feature selection procedure with leave-one-out cross validation.
[328]	4	F4-C4 and, C4-A1	A	7	84	74	86	1	A combination of all the features tested by Mariani et al. [28] were fed to a soft-margin SVM with a Gaussian kernel.
[79]	4	C3-A2 or C4-A1	A	7	82	76	83	1	A combination of all the features tested by Mariani et al. [28] were fed to a FFNN.
[329]	16*	C3-A2 or C4-A1	A	7	86	67	90	-	A combination of all the features tested by Mariani et al. [28], but with a variable window, was fed to three LDA (discriminated three classes: background; A1; A2 and A3), and classification vectors were combined to classify the data.
[330]	8	C3-A2 or C4-A1	A	7	85; 82; 79; 82	73; 70; 69; 73	87; 84; 79; 82	1	A combination of all the features tested by Mariani et al. [28] were fed to: LDA; SVM; AdaBoost; FFNN.
[331]	30*	C4-A1	A	30	68; 70; 71	-	-	1	Developed features from the macro-micro structure descriptor, TEO, zero-crossing, Lempel-Ziv complexity, empirical mode decomposition, variance, and Shannon entropy to feed the classifiers: LDA; kNN; SVM.
[332]	4^	-	CAP	-	88	-	-	-	Employed thresholds to the EEG signal for the A phase detection, and a FSM to estimate the CAP cycles.

[333]	6[-]	-	CAP	7	67	-	-	-	The output of the fast discrete wavelet transform was employed as input to two moving averages that were compared to thresholds, tuned by a genetic algorithm, for classification.
[334]	8^	-	CAP	4	81	85	78	70	Employed the same method as Largo et al. [333].
[335]	4^	C4-A1	CAP	4	90	90	95	1	Employed the same method as Lima and Rosa [332].
[336]	8*[+]	C4-A1	CAP	3	79; 79; 77	-	-	1	Estimated the Kolmogorov entropy, sample entropy, and Shannon entropy to feed the classifiers: LDA; SVM; kNN.

* The CAP Sleep Database – PhysioNet [19]

[+] Average of the reported results

^ The indicated performance is based on a proposed metric which is related to the metrics examined in this work

[-] Acc excluding the B phase events

It was verified that the most relevant features for A phase detection were: Hjorth activity; differential variance; Lempel-Ziv complexity; Teager Energy Operator (TEO); five frequency band descriptors; Shannon entropy; empirical mode decomposition; EEG energy; power in the beta band. Regarding the classifiers, LDA, FFNN, SVM, and kNN were the most relevant. It is also important to notice that no deep learning model was proposed at the time the review was conducted. The majority of the analyzed methods manually removed the REM periods from the EEG signal, leading to an increase of the classification performance. However, this methodology was not used in this work to provide algorithms capable of performing fully automatic analysis with implementation in a HMD.

5.2. Feature based methods

5.2.1. Evaluation of features for CAP analysis

In the first approach, the LDA classifier (implemented in MATLAB) was employed to detect the A and not-A phases (using TFCV with cost-sensitive learning for performance assessment, and performing 50 repetitions of each iteration to achieve statistically significant results), and afterwards a FSM determined the CAP cycles. The first CAPSD was used in this examination (nine normal subjects and four subjects with sleep-disordered breathing). The features that were assessed by the review as the most relevant for A phase detection were tested using a two second time window (selected since it is the minimum A phase duration) of the EEG signal.

Two of these features are Shannon entropy and TEO that are, respectively, defined by [337]

$$\mathbb{H}_{\boldsymbol{S}}(x) = -\sum_{i=0}^{N-1} \big(P_i(x)\big)^2 \big(\log_2\big(P_i(x)\big)\big)^2 \qquad 5.1$$

where x is the random variable, N is the number of points in the variable, and [325]

$$\Phi\big(x(t)\big) = \dot{x}(t)^2 - x(t)\ddot{x}(t) \qquad 5.2$$

Both features presented good discriminatory capabilities. Entropy is expected to be significant for CAP analysis as it is a measure of uncertainty which quantifies the degree of complexity in a signal. Phasic and transient events occur during the A phase, leading to significant variations in the EEG signal, when compared to the background activity,

which can be detected by entropy analysis. Taking into consideration that EEG is a complex, non-stationary, and non-linear signal, it is also expected for TEO to be relevant for this analysis as it is a non-linear energy-tracking operator.

A finite impulse response filter of order 31, with a Kaiser window, was employed in the decomposition the EEG signal [329] in the frequency bands defined by Mariani et al. [328] (specifically: delta from 0.5 to 4 Hz; theta from 4 to 8 Hz; alpha from 8 to 12 Hz; sigma from 12 to 15 Hz; beta from 15 to 30 Hz). The relevance of the frequency based metrics for CAP analysis is associated with the fact that characteristic patterns of the sleep microstructure occur at specific frequencies. However, by testing both the five band descriptors and the Power Spectral Density (PSD) of the five frequency bands, it was observed that PSD attained a higher accuracy. The PSD features were proposed in this work to provide a similar information as produced by the band descriptors without requiring estimating the mean power in a window of 64 seconds (used by the band descriptors). Therefore, the PSD features were chosen and estimated by the standard Welch's method where the PSD estimation at a given frequency, *f*, is given by [338] [339]

$$\Gamma(f) = \frac{1}{K}\sum_{i=1}^{K} S_i(f) \tag{5.3}$$

where *K* is the number of segments, specified by [339]

$$K = floor\left[\frac{L - \sigma M}{(1-\sigma)M}\right] \tag{5.4}$$

where *floor* is the floor function, *L* is the total number of points in the examined segment, σ it the amount of overlapping (50% overlap was employed), and *M* is the examined segment's length. The Hanning window, *W*, was applied to the segments. *S* is the windowed periodogram calculation, applied for all *i* (1, 2, …, *K*) and computed as [338] [339]

$$S_i(f) = \frac{1}{QM}\left|\sum_{n=0}^{M-1} x_i(n) W_M(n) e^{-j2\pi f n}\right|^2 \tag{5.5}$$

where *Q* is the normalization factor given by [338] [339]

$$Q = \frac{1}{M}\sum_{n=0}^{M-1} [W_M(n)]^2 \tag{5.6}$$

A similar conclusion was attained when comparing the differential variance (feature proposed in the state of the art) with the autocovariance (feature proposed in this work, computing the covariance of a process with itself, at pairs of time points), given by the second moment product [340]

$$\gamma_x(s,t) = E[(x_s - \mu_s)(x_t - \mu_t)] \tag{5.7}$$

for all time points *s* and *t*, and respective average, μ. The autocovariance was examined to help identify the regularity of the current epoch, considering the hypothesis that the transient and phasic events present in the A phases, will lead to lower autocovariance values than the background activity. Time series analysis based on the standard deviation and the average power were found to have a good correlation with the occurrence of the

A phases. The hypothesis is that the transient and phasic events will lead to significant amplitude variations in the EEG signal when compared to the background activity.

The log-energy entropy, given by [337]

$$\mathbb{H}_{LE}(x) = -\sum_{i=0}^{N-1}\left(\log_2\left(P_i(x)\right)\right)^2 \qquad 5.8$$

was also employed since it was considered to be a relevant descriptor for EEG analysis in other research fields [337]. It was also observed that this feature achieved a better performance in the A phase detection than Hjorth activity, empirical mode decomposition, and Lempel-Ziv complexity. The selected features, with a respective identification number, are indicated in Table 5.2.

Table 5.2: Chosen features and the respective identification number [30].

Feature	*Identification number*
Average power	1
Standard deviation	2
Shannon entropy	3
Log-energy entropy	4
Autocovariance	5
TEO	6
PSD in the delta band	7
PSD in the beta band	8
PSD in the alpha band	9
PSD in the sigma band	10
PSD in the theta band	11

A post-processing algorithm was implemented to reduce the classification outliers by considering an A or B phase, lasting only two seconds, which occur between two opposite phases as misclassified data. Thus, it was converted into the opposite phase.

The relevance of the features was assessed using SFS, ordering the features by their relevance to the maximization of the CO. Table 5.3 presents the features ordered by SFS for the A phase detection with LDA, and Figure 5.1 displays the performance metrics according to the features selected, where the value of the CO is indicated by Total.

Table 5.3: Features ordered by SFS for the LDA [22].

Feature	*Selected order*	*Identification number*
PSD in the beta band	1	8
Average power	2	1
PSD in the theta band	3	11
TEO	4	6
Standard deviation	5	2
PSD in the alpha band	6	9
PSD in the sigma band	7	10
Shannon entropy	8	3
Log-energy entropy	9	4
Autocovariance	10	5
PSD in the delta band	11	7

SFS is a classifier dependent method. Consequently, both the order and the selected features for one classifier are likely not to be the best for other classifiers. Therefore, a classifier independent method for dimensionality reduction, Principal Component Analysis (PCA), was also tested. The flowchart of the tested models is presented in Figure

5.2. The features employed by Mariani et al. [330] were tested since the same classifier was used in this work. The best results were attained by using the six more relevant features selected by SFS and the first three components (variance of 78%) of PCA. The achieved results (average value ± standard deviation) of the implemented classifier with the different features are presented in Table 5.4.

Table 5.4: Achieved results of the LDA with the different features [22].

Employed features	*Acc (%)*	*Sen (%)*	*Spe (%)*	*AUC*	*CAP Acc (%)*
Selected by SFS	75 ± 5	78 ± 2	74 ± 7	0.76 ± 0.02	75 ± 7
Produced by PCA	74 ± 6	71 ± 5	75 ± 8	0.73 ± 0.02	76 ± 6
Proposed by Mariani et al. [330]	67 ± 3	79 ± 15	64 ± 4	0.71 ± 0.07	68 ± 6

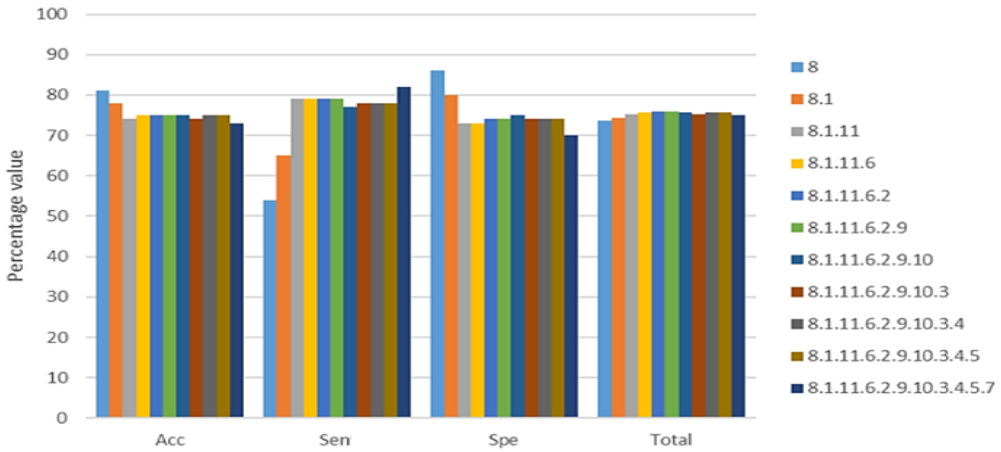


Figure 5.1. Performance of the A phase classification with the features ordered by SFS for the LDA [22].

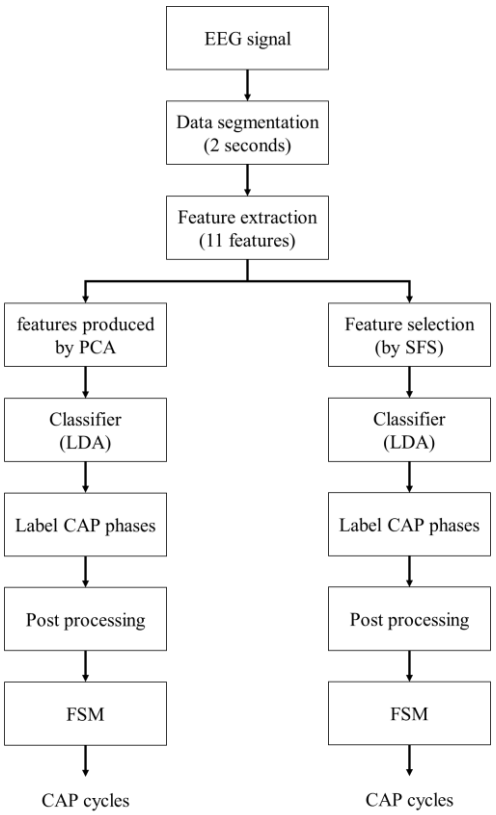


Figure 5.2. Flux diagram of the tested models considering LDA as the classifier.

By analyzing the table, it is possible to verify that the highest Acc and AUC were achieved using SFS while the best Spe and CAP Acc was provided by PCA. Due to the unbalance of the data (more not-A than A epochs) a better Spe leads to a higher CAP Acc. The features indicated by Mariani et al. [330] provided the greatest Sen with a large standard deviation. This variation in Sen is in line with the deviation presented by Mariani et al. [330]. Therefore, the model based on features selected by SFS achieved the most balanced results.

5.2.2. Further examination of methods for feature based CAP analysis

Nine more classifiers (LR, CT, ET, SVM, FFNN, CFNN, kNN, kMC, and SOM, implemented in MATLAB) were examined to assess if another combination of features and classification procedure could yield a better performance for CAP analysis. The evaluation was performed using the first CAPSD and TFCV for hyperparameter optimization and performance assessment (performing 50 repetitions of each iteration to achieve statistically significant results). The classifier selection concept was to cover multiple solutions, from supervised to unsupervised learning, including recommendations from the state of the art, and classifiers proposed in this work considering that they have attained a good performance in other fields.

Cost-sensitive learning was employed to minimize the effect of the data unbalance in all supervised learning classifiers while oversampling (of the minority class) was applied to the training dataset for the unsupervised learning methods. The performance of the classifiers was analyzed in two tests, one using the features selected by SFS and another with features produced by PCA. Table 5.5 presents the number of features and order (taking into account the identification number of the feature, presented in Table 5.2) chosen by SFS for each classifier. The distribution of the selected features by the analyzed classifiers is presented in Figure 5.3.

Table 5.5. Identification of the features selected by SFS for each classifier [30].

Classifier	*SFS order*
LR	8, 1, 9, 6
CT	7, 8, 11, 10, 9, 5, 4, 3
ET	9, 7, 11, 10, 4, 8, 6, 5, 3
SVM	1, 8, 11, 4, 6, 2, 10, 5
FFNN	8, 3, 6, 5, 11
CFNN	8, 3, 6, 9, 11, 5, 10
kNN	11, 8, 7, 9, 5, 3, 1, 4, 2
kMC	10, 5, 6, 3, 9, 8, 2
SOM	10, 1, 8, 6, 11, 5, 2

It was possible to assess that PSD in the beta band was the feature most frequently chosen as the most relevant while the opposite occurred for the PSD in the delta band. It was also verified that the first component of PCA is strongly correlated with the PSD in the beta band, as it can be assessed by analyzing the example in Figure 5.4.

Several observations were attained for each classifier:

- The effect of regularization in the LR was tested and it was verified that it did not improve the performance.

- The number of trees which composed the ET was varied between 3 and 30. It was verified that 10 trees achieved the best results using TotalBoost.
- A linear kernel, with a scale of 2, and 10% outlier fraction was assessed to be the best configuration for the SVM.
- Hyperbolic tangent was verified to be the most suitable activation function for both FFNN and CFNN, using the Levenberg-Marquardt algorithm for training. A single HL was employed for both networks, and the number of neurons was incremented, from 20 to 400, in steps of 10. The best results were achieved using 280 neurons for the FFNN and 270 for the CFNN.
- The number of nearest neighbors for the kNN was continuously incremented from 1 to 10, and 4 produced the best results considering Euclidean distance as the metric to determine the distance.
- A linear topology was tested for the SOM, changing the dimension size from 2 to 10. The best result was produced by employing a dimension size of 6.

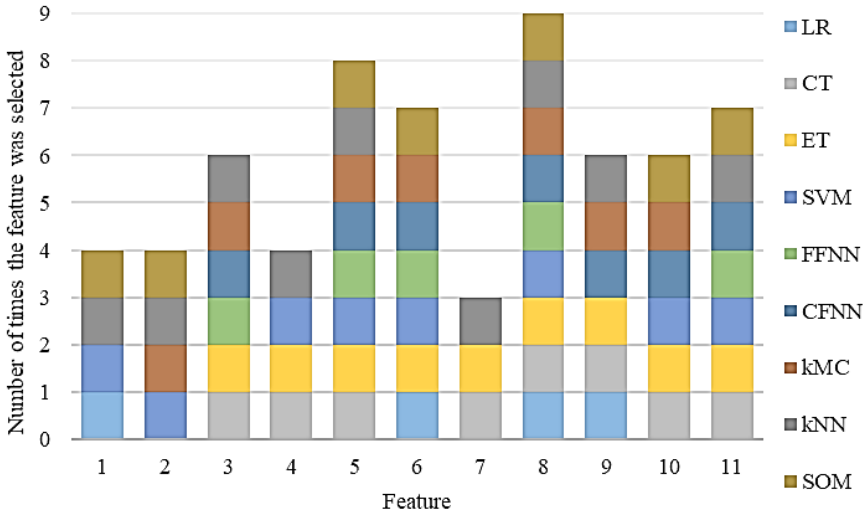


Figure 5.3. Number of times each feature was selected by the classifiers [30].

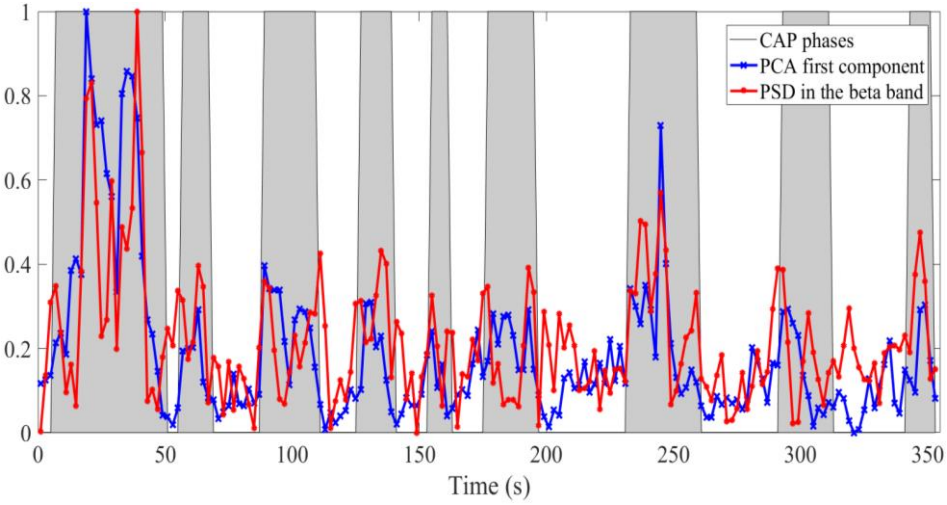


Figure 5.4. Example to show that the first component of PCA and the PSD in the beta band are strongly correlated with each other as the CAP phases (A phase is marked in grey) change [30].

Table 5.6 presents the results of these analyses. By analyzing the table, it is possible to confirm that FFNN, with features selected by SFS, achieved the highest average Acc and AUC. The highest Sen was attained by SOM with features produced by PCA. However, it has the lowest Spe which was the best metric achieved by the CT, with the features produced by PCA. Taking into consideration that FFNN, with features selected by SFS, also achieved the highest average CO (78.3%), it was therefore considered to be the best method.

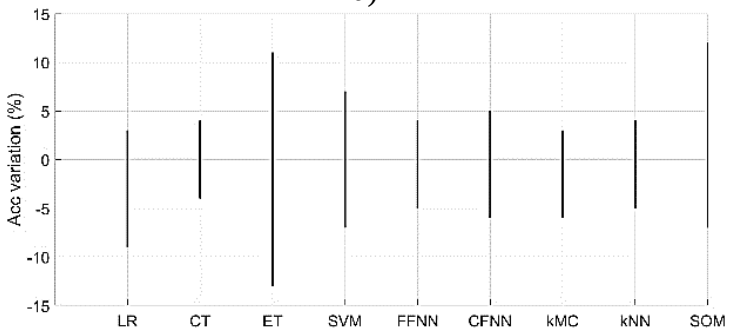
Table 5.6. Average results of the analyzed classifiers using the features selected by SFS and the features produced by PCA [30].

Features	*Classifier*	*Acc (%)*	*Sen (%)*	*Spe (%)*	*AUC*	*CAP Acc (%)*
Selected by SFS	LR	76	80	75	0.77	78
	CT	70	58	73	0.66	64
	ET	70	64	71	0.67	70
	SVM	72	80	70	0.76	75
	FFNN	79	76	80	0.78	79
	CFNN	76	77	76	0.76	77
	kNN	72	70	72	0.71	78
	kMC	78	67	81	0.74	70
	SOM	67	79	66	0.73	68
Produced by PCA	LR	67	78	65	0.71	69
	CT	74	51	82	0.62	68
	ET	74	63	77	0.70	76
	SVM	68	84	66	0.74	71
	FFNN	75	76	75	0.75	76
	CFNN	74	76	74	0.75	76
	kNN	69	65	70	0.67	61
	kMC	61	62	61	0.61	66
	SOM	22	90	08	0.49	60

Regarding the CAP Acc, the best results were attained using the FFNN for the classification of the CAP phases. Therefore, it was verified that FFNN, with features selected by SFS, was also the best classifier, and PCA generated the more unbalanced results. The variation of the performance metrics, considering the models with features selected by SFS, for the detection of the CAP phases and for the CAP cycle estimation is presented, respectively, in Figures 5.5 and 5.6. It was verified that the CFNN has the overall lowest variation, while ET has the highest.

Additive White Gaussian Noise (AWGN) was introduced in the EEG signal to estimate the effect that the noise has in the model that achieved the best results (FFNN with features selected by SFS). The Signal-to-Noise Ratio (SNR) was varied and the CO was used as the reference measurement. Figure 5.7 presents the results and it was assessed that a SNR lower than 20 dB starts to deteriorate the performance of the method. The lowest acceptable SNR was 0.5 dB (CO of 70%).

a)



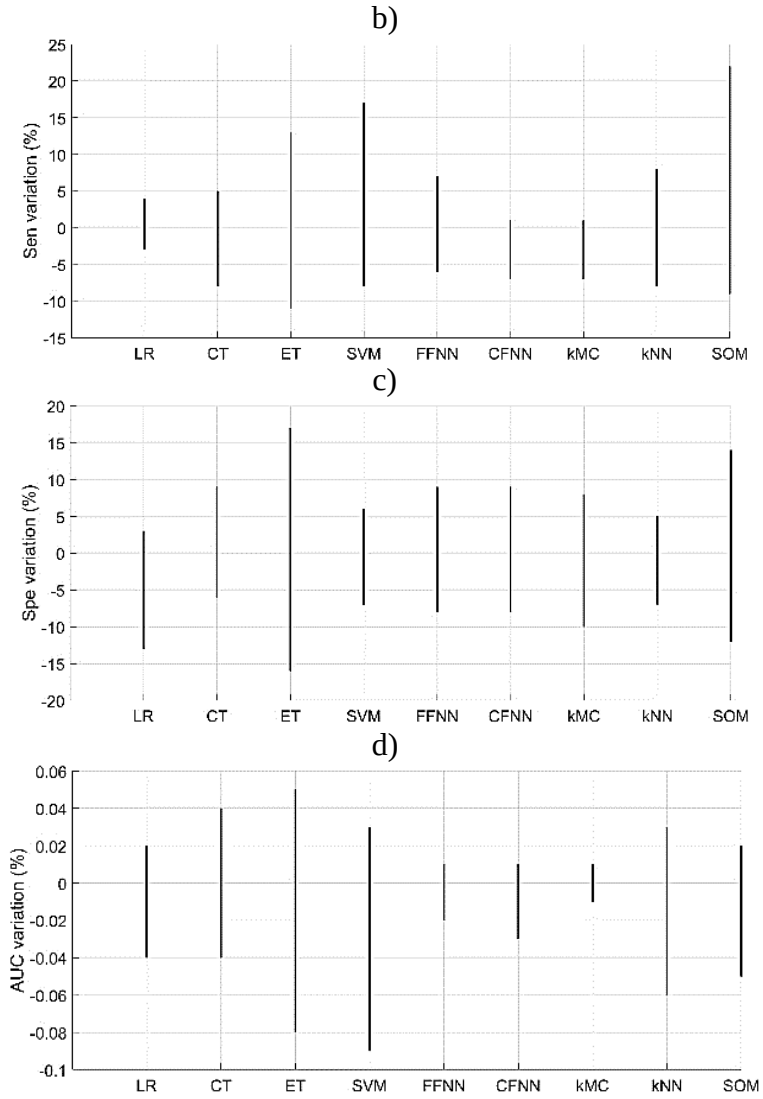


Figure 5.5. Variation of the a) Acc, b) Sen, c) Spe and d) AUC of each classifier using the features chosen by SFS [30].

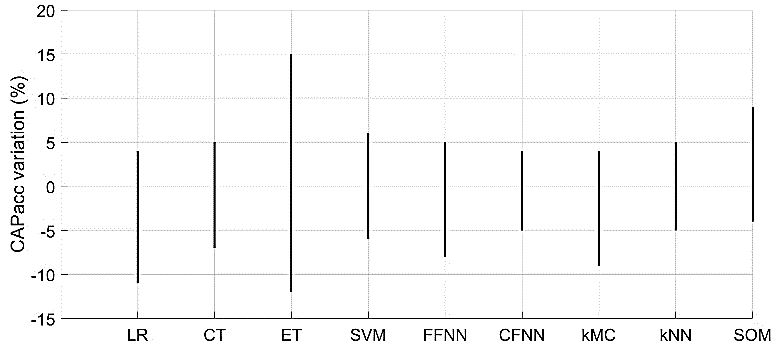


Figure 5.6. Variation of the CAPacc using the classifiers output to feed the FSM [30].

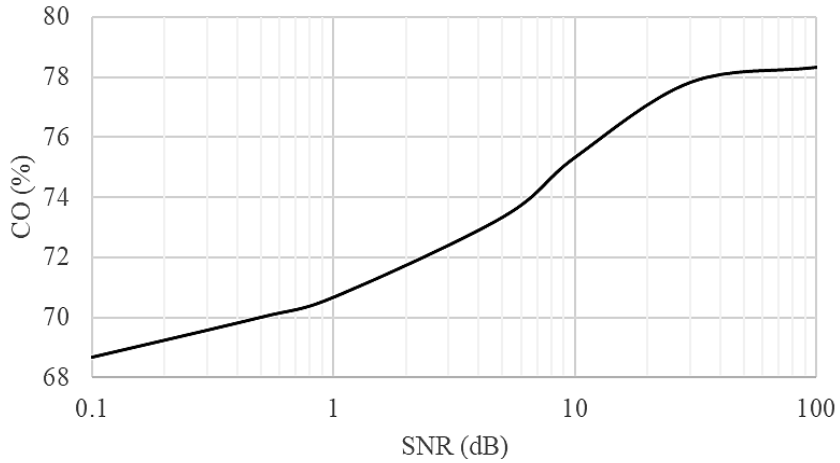


Figure 5.7. Effect of introducing AWGN in the performance of the FFNN with features selected by SFS [30].

5.3. Methods without an explicit feature extraction process

5.3.1. Deep and shallow networks combination for CAP estimation

The development of methods based on crafted features has the difficulty of finding the best set of features that are the most relevant for the description of the event. This issue can be addressed by employing a method without an explicit feature extraction process that automatically learns the most relevant characteristics of the signal.

The first model developed to examine this methodology (implemented in MATLAB) employed a DSAE fed with the EEG signal to automatically classify the A phases (A phase or not-A phase, named nAphase) considering epochs with two seconds. The output of the network was stored in a memory buffer, and its output was used as input of a FFNN (shallow neural network), a Multilayer Perceptron (MLP) to classify the CAP cycles, replacing the FSM. The implemented model is presented in Figure 5.8. A memory buffer was employed since the CAP classification required a longer data duration than the A phase (A phase duration ranges from 2 to 60 seconds while the CAP cycle ranges from 4 to 120 seconds).

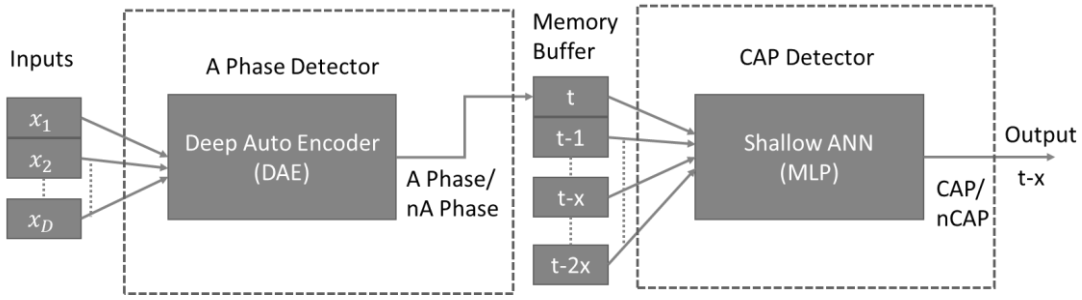


Figure 5.8. Model for A phase and CAP cycle classification methods without an explicit feature extraction process [341].

The DSAE was composed of two stacked autoencoders followed by an output layer to perform the classification. Each autoencoder was individually trained using an unsupervised method that encodes the input, using the sigmoid function, in a new

representation, and then attempts to attain the original input by decoding the information, minimizing the mean square error [308]. Afterwards, a deep network was formed by linking the input signal to the encoder of the first autoencoder (layer 1), and then connecting the encoder's output to the encoder of the second autoencoder (layer 2) which, in turn, the output (of the encoder) was linked to the output layer. A representation of the CAP phase classifier is presented in Figure 5.9. The DSAE outputs were stored in a memory buffer to feed a FFNN, which performed the CAP cycle estimation, and was fine-tuned using the Levenberg-Marquardt algorithm [304].

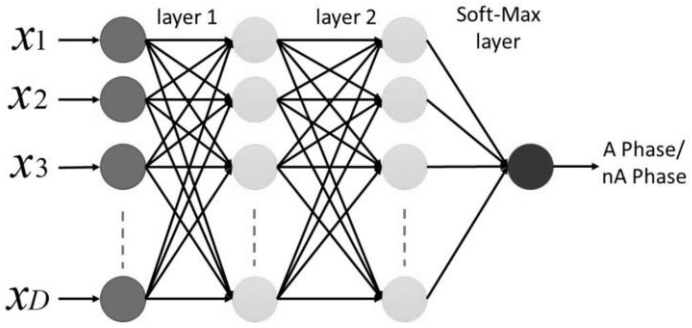


Figure 5.9. CAP phase classifier based on a DSAE [341].

The performance of the developed method was assessed using the first CAPSD and a fivefold CV scheme (repeated 50 times to achieve statistical significance), ensuring subject independence by using the data from each subject only in one of the folds. This procedure was used as it was recommended in the literature for deep learning at the time when the method was proposed. Underfitting was used in this model to have an equal number of A and not-A events in each fold used for training, without changing the fold used for testing. The balancing operation allowed the model to improve the performance by increasing the CO. A grid search strategy was employed to find the best number of neurons for both layers. It was verified, by performing multiple simulations, that good results were attained when layer 2 has half of the number of neurons from layer 1, which were varied from 64 to 768 in steps of 64. The variation of the performance metrics and the CO for the best approach is presented in Figure 5.10.

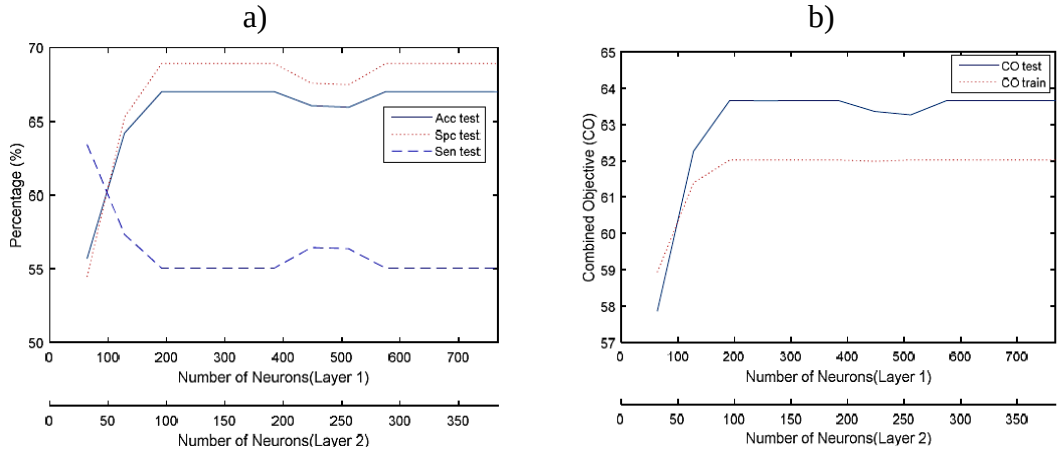


Figure 5.10. Variation of a) the performance metrics, and b) the CO for the best models of the DSAE [341].

It was verified that the best performance for the A phase detection was achieved using 192 and 96 neurons, in layers 1 and 2 of the DSAE, respectively. The average Acc, Sen, and Spe was 67.02 %, 55.03%, and 68.93%, respectively. The output of the DSAE was stored in a memory buffer whose size was the same as the input layer of the FFNN that performs the CAP cycle classification. The FFNN (performing the CAP cycle detection) had a single HL, and the optimal number of neurons of both input and HLs was found using a grid search strategy.

The best results were achieved by using 350 neurons in the HL. The buffer size was varied between 2 and 180. It was verified that the model attained the best results using 180 cells in the buffer (the model evaluates the 179 previous classifications from the A phase detector and the current A phase classification to perform the current CAP classification). The buffer size was not further incremented since the performance peak was reached. The average Acc, Sen, and Spe for the CAP cycle detection was, respectively, 61.5%, 66.64% and 58.72%.

5.3.2. Examination of methods without an explicit feature extraction process

The results attained by the first method without an explicit feature extraction process were lower than the results produced by the feature based approaches. Nevertheless, this approach has the main advantages of not requiring domain-specific knowledge to produce the features, and a feature selection method is not needed to improve the results. Therefore, two new classifier types were considered to verify if the performance can be improved. The first type is based on RNN, specifically the LSTM and GRU, and was selected since these networks allow the recognition of both long and short-term correlations in time-series that commonly occur in physiological signals [311]. The second approach was based on the employment of a CNN with one-dimensional input (1D-CNN), chosen because it is one of the best models for automatic feature extraction from complex signals [342].

The LSTM was the first model to be tested, and the block diagram of the developed algorithm is presented in Figure 5.11. Recordings from the second CAPSD (15 control subjects) were analyzed in this test. Performance assessment was performed by LOOCV (performing 50 repetitions of each iteration to achieve statistically significant results), and using cost-sensitive learning, while the best hyperparameters were identified by TFCV. The pre-processed EEG signal (EEG signal resampled at 100 Hz, to uniform the database, subsequently standardized, by subtracting the mean and dividing by the standard deviation, and segmented into one second epochs, without overlapping) was fed to the LSTM to categorize each epoch, as either an A or a not-A. This information was then employed by the previously developed FSM to determine the CAP cycles.

On the first test, it was intended to verify if the LSTM was a viable solution. Thus, the network architecture was fixed (with one HL) and the number of hidden units was chosen by a grid search algorithm (checking from 100 to 500, in steps of 100), performing multiple runs with cross-validation, considering the AUC as the reference metric, using the ADAM algorithm [343] for optimization (implementing the classifier in Python 3

using the Keras library). The performance of the developed algorithm is presented in Table 5.7.

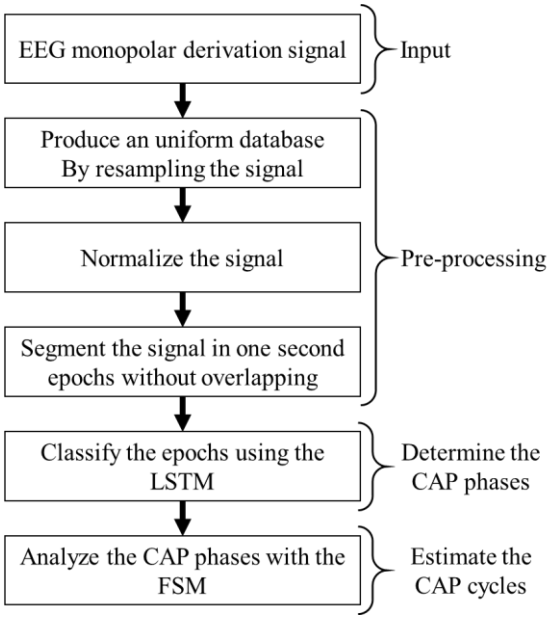


Figure 5.11. Block diagram of the first model based in the LSTM [344].

By examining Table 5.7 results', it is possible to verify that the best performance was attained using 400 hidden units and that the sensitivity has the highest standard deviation. However, by comparing with the model based on the DSAE it is possible to verify that both Acc and Spe were higher when using the LSTM. It was observed that the majority of the correctly classified CAP phases occur when the phases have a duration greater than 21 seconds. Conceivably, the longer phases have a characteristic pattern that was identified by the LSTM, thus justifying the need to further analyze the network architecture in the second test.

Table 5.7: Performance of the developed algorithm (mean ± standard deviation) based on an LSTM with a fixed architecture when the number of hidden units was changed [344].

Classification	*Number of hidden units*	*Acc (%)*	*Sen (%)*	*Spe (%)*	*AUC*
CAP phases (LSTM)	500	68.9±5.9	50.9±12.9	81.3±4.3	0.660±0.068
	400	69.7±5.9	51.2±12.8	81.1±4.3	0.663±0.068
	300	68.3±5.7	48.7±12.3	83.6±3.2	0.656±0.065
	200	66.5±5.8	49.9±10.2	82.7±3.2	0.656±0.062
	100	65.8±6.7	47.9±11.9	83.2±3.0	0.655±0.068
CAP cycles (LSTM and FSM)	500	68.1±5.0	48.9±8.9	89.9±3.9	0.696±0.048
	400	67.9±5.1	50.1±9.5	89.5±3.7	0.703±0.049
	300	67.8±5.3	48.6±9.1	90.9±2.7	0.694±0.051
	200	67.3±5.3	48.9±8.9	90.3±3.0	0.693±0.050
	100	67.2±5.0	47.4±8.5	90.7±2.7	0.690±0.048

It was also verified that most of the misclassified epochs occur either in the beginning or in the end of the A phases, as can be verified in the exampled presented in Figure 5.12, advocating the need to properly classify the phase's boundaries. This observation is in agreement with the results reported by Largo et al. [54].

The block diagram of the model that was employed in the second test is presented in Figure 5.13. The process is similar to the introduction of the post-processing procedure, which was previously presented in section 5.2, to reduce the misclassifications.

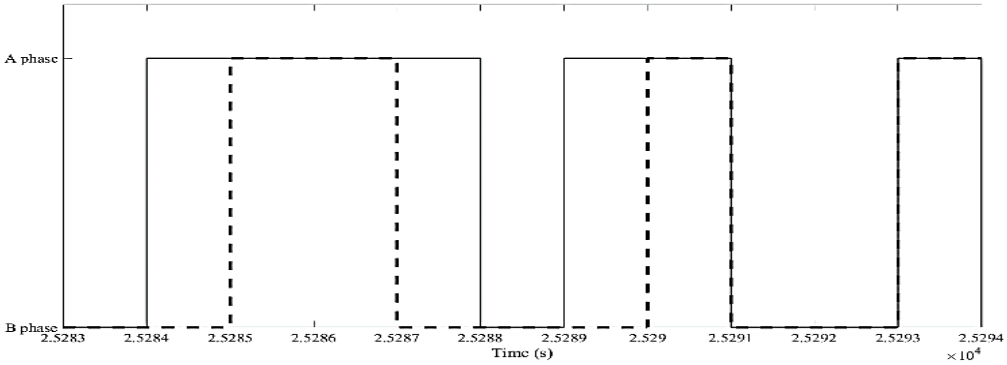


Figure 5.12. Example of the CAP phase detection where the dashed line is the CAP phase provided by the database label and the full line is the model prediction [344].

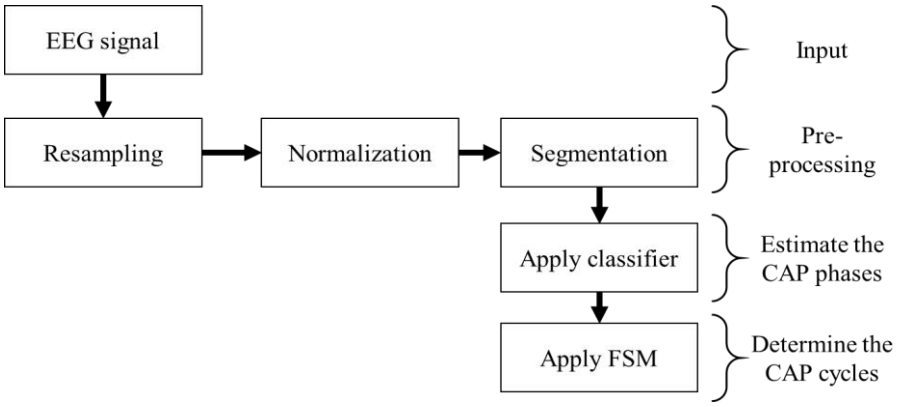


Figure 5.13. Block diagram of the second model that was used on the second test using deep neural networks [26].

Three machine learning based classifiers were analyzed in the second test, specifically, the LSTM, the GRU which is a simplification of the LSTM that also allows the detection of temporal correlation components [345], and the 1D-CNN that applies convolution kernels to implement a transformation of the inputs that augments the relevant patterns of the physiological signals [346].

Recordings from the third CAPSD (15 control subjects and 4 subjects with sleep-disordered breathing) were analyzed in this test, using LOOCV for performance assessment (performing 50 repetitions of each iteration to achieve statistically significant results). The hyperparameters of the three classifiers were chosen by running multiple simulations of different network configurations with TFCV. The effect of data unbalance was minimized by using cost-sensitive learning.

The developed Heuristic Oriented Search Algorithm (HOSA) determines the best architecture for the specified classifier, C (0 for LSTM, 1 for GRU, and 2 for 1D-CNN), by performing a simulation for each combination of the hyperparameters (each simulation was repeated 10 times). The number of neurons, N_I, for the LST or GRU layers was varied from 50 (N_{Istart}) to N_{Imax} (chosen to be 500), in steps, N_{Istep}, of 50. The evaluated subsequent layers were other recurrent layers with the same N_I as the first layer. The last recurrent layer could be followed by a fully connected layer (F_C) whose number of neurons, N_S, was varied from 10 (N_{Sstart}) to N_{Smax} (chosen to be 100), in steps, N_{Sstep}, of 10.

The number of kernels, K, used by the convolution layers was selected to be a power of two, ranging from 8 to 256 (step of 2^N where $3 \leq N \leq K_{max}$, and K_{max} was 8), and the filter length, F, was varied from 1 to F_{max}, chosen to be 10, in steps (F_{step}) of 1, with a stride of 1. The output of all networks was a fully connected layer that used the soft-max function for classification. The GL concept was used for the 1D-CNN architecture optimization, using groups composed of one convolution layer, followed by batch normalization, and a pooling layer. The evaluated pooling functions were the maximum (*MaxP*) or the average (*AveP*). However, the last pooling layer was chosen to be global. Hence, this GL was named GL2. The examined activation functions were *ReLU* and *SELU*. The pooling size was the same as F.

The number of layers (or GL for the 1D-CNN), G, for the algorithm to examine was iteratively incremented until a maximum value, G_{max}, selected to be three or until no relevant improvement was attained for the AUC (considering a threshold, t_h, of 1% to define the minimum relevant improvement) when introducing the subsequent layer. The HOSA pseudo code is:

HOSA (*Data*, G_{max}, K_{max},, F_{step}, F_{max}, N_{Istart}, N_{Istep}, N_{Imax}, N_{Sstart}, N_{Sstep}, N_{Smax}, t_h, C)
G = [1, 2, …, G_{max}]
if C == 0 or C == 1
  N_I = [N_{Istart}, N_{Istart} + N_{Istep}, …, N_{Imax}]
  N_S = [N_{Sstart}, N_{Sstart} + N_{Sstep}, …, N_{Smax}]
  if C == 0
    RecLayer = *LSTM*
  else
    RecLayer = *GRU*
  for g = 1 to length (G)
    for n = 1 to length (N_I)
      $Net_{0,n,0,0} \leftarrow I_{pt}$ (*Data*)
      if g == 1
        $Net_{g,n,0,0} \leftarrow Net_{0,n,0,0}$ + *RecLayer* (N_I (n))
        for m = 1 to length (N_S)
          $Net_{g,n,m,0} \leftarrow Net_{g,n,0,0} + F_C$ (N_S (m)) + F_C (2)
          $AUC_{g,n,m,0} \leftarrow$ test (train ($Net_{g,n,m,0}$))
        $Net_{g,n,0,1} \leftarrow Net_{g,n,0,0} + F_C$ (2)
        $AUC_{g,n,0,1} \leftarrow$ test (train ($Net_{g,n,0,1}$))
      else
        for z = 1 to g
          $Net_{z,n,0,0} \leftarrow Net_{z-1,n,0,0}$ + *RecLayer* (N_I (n))
        for m = 1 to length (N_S)
          $Net_{g,n,m,0} \leftarrow Net_{z,n,0,0} + F_C$ (N_S (m)) + F_C (2)
          $AUC_{g,n,m,0} \leftarrow$ test (train ($Net_{g,n,m,0}$))
        $Net_{g,n,0,1} \leftarrow Net_{z,n,0,0} + F_C$ (2)
        $AUC_{g,n,0,1} \leftarrow$ test (train ($Net_{g,n,0,1}$))
    $AUC_{g,n,0,1,max} = max(AUC_{g,n,0,1})|_{for\ all\ n}$
    $AUC_{g,n,0,0,max} = max(Net_{Dg,n,m,0})|_{for\ all\ n,m}$
    if g > 1
      if $AUC_{g,n,0,0,max} > AUC_{g,n,0,1,max}$
        if $AUC_{g,n,0,0,max} - AUC_{g-1,n,0,1,max} \leq t_h$
          if $AUC_{g,n,0,0,max} - AUC_{g-1,n,0,0,max} \leq t_h$

```
                if AUC_{g-1,n,0,1,max} > AUC_{g-1,n,0,0,max}
                    BestNet = Net_{g-1,n,0,1}|AUC_{g-1,n,0,1,max}
                else
                    BestNet = Net_{g-1,n,0,0}|AUC_{g-1,n,0,0,max}
                break
            else
                BestNet = Net_{g,n,0,0}|AUC_{g,n,0,0,max}
        else
            BestNet = Net_{g,n,0,0}|AUC_{g,n,0,0,max}
    else
        if AUC_{g,n,0,1,max} − AUC_{g-1,n,0,1,max} ≤ t_h
            if AUC_{g,n,0,1,max} − AUC_{g-1,n,0,0,max} ≤ t_h
                if AUC_{g-1,n,0,1,max} > AUC_{g-1,n,0,0,max}
                    BestNet = Net_{g-1,n,0,1}|AUC_{g-1,n,0,1,max}
                else
                    BestNet = Net_{g-1,n,0,0}|AUC_{g-1,n,0,0,max}
                break
            else
                BestNet = Net_{g,n,0,1}|AUC_{g,n,0,1,max}
        else
            BestNet = Net_{g,n,0,1}|AUC_{g,n,0,1,max}
else
    K = 2^N where 3 ≤ N ≤ K_max
    F = [1, 1 + F_step, …, F_max]
    P = [MaxP, AveP]
    A = [ReLU, SELU]
    for g = 1 to length (G)
        for k = 1 to length (K)
            for f = 1 to length (F)
                for p = 1 to length (P)
                    for a = 1 to length (A)
                        Net_0 ← I_pt (Data)
                        if g > 1
                            for z = 1 to g – 1
                                Net_z ← Net_{z-1} + GL (K (k), F (f), P (p), A (a))
                            Net_{g,k,f,p,a} ← Net_z + GL2 (K (k), F (f), P (p), A (a)) + F_C (2)
                        else
                            Net_{g,k,f,p,a} ← Net_0 + GL2 (K (k), F (f), P (p), A (a)) + F_C (2)
                        AUC_{g,k,f,p,a} ← test (train (Net_{g,k,f,p,a}))
        AUC_{g,k,f,p,a,max} = max(AUC_{g,k,f,p,a})|_{for all k,f,p,a}
        if g > 1
            if AUC_{g,k,f,p,a,max} − AUC_{g-1,k,f,p,a,max} ≤ t_h
                if AUC_{g,k,f,p,a,max} > AUC_{g-1,k,f,p,a,max}
                    BestNet = Net_{g,k,f,p,a}|AUC_{g,k,f,p,a,max}
                else
                    BestNet = Net_{g-1,k,f,p,a}|AUC_{g-1,k,f,p,a,max}
                break
            else
                BestNet = Net_{g,k,f,p,a}|AUC_{g,k,f,p,a,max}
return BestNet
```

Five optimization algorithms, commonly employed for deep learning, were tested specifically: RMSprop [347]; ADADELTA [347]; stochastic gradient descent with and without momentum [308]; ADAM [343]; adaptive gradient [348]. In the performed test, it was verified that ADAM outperformed the other algorithms, although RMSprop attained a similar performance in the recurrent networks. All the models were developed in Python 3 using the Keras library.

The optimal architecture of the analyzed networks is presented in Table 5.8. It was concluded that both LST and GRU attained the best AUC using 3 layers while 8 layers were used for the 1D-CNN. The total number of created networks for the LSTM, GRU, and 1D-CNN was 400, 400, and 720, respectively. Each network was simulated 10 times, hence, the total number of examined classifiers was 15200.

Table 5.8: Architecture of the classifiers that provided the highest AUC [26].

Type of network	*Architecture*
LSTM	-Input (100 neurons) -LSTM (400 cells) -fully connected (30 neurons) SELU -Output Soft-max
GRU	-Input (100 neurons) -GRU (200 cells) -fully connected (20 neurons) SELU -Output Soft-max
1D-CNN	-Input (100) -Convolution (64 filters with length 3) -Batch normalization (64 channels) ReLU -MaxP (pool size = 3) -Convolution (128 filters with length 3) -Batch normalization (128 channels) ReLU -Global AveP -Output Soft-max

The results for the CAP phases and cycles detection are presented in Table 5.9. By analyzing the table, it is possible to determine that the LSTM achieved the highest Acc and AUC (indicating balanced results) thus, it was considered to be the best model. These results advocate that both the CAP cycles and phases have a temporal dependency, agreeing with the observation of Terzano et al. [19].

Table 5.9: Performance of the developed algorithms (mean ± standard deviation) for CAP detection based on the EEG signal analysis [26].

Classification	*Classifier*	*Acc (%)*	*Sen (%)*	*Spe (%)*	*AUC*
CAP phases (classifier)	LSTM	76.0±6.1	74.6±10.8	76.6±6.1	0.752±0.012
	GRU	80.7±8.1	65.5±12.2	83.0±15.7	0.742±0.100
	1D-CNN	74.4±3.7	71.2±4.0	74.8±4.8	0.729±0.004
CAP cycles (classifier and FSM)	LSTM	76.3±7.6	71.4±11.3	84.2±10.2	0.778±0.064
	GRU	71.9±9.2	59.9±14.1	87.8±13.5	0.738±0.070
	1D-CNN	72.5±4.3	65.8±3.9	83.3±4.6	0.745±0.013

The 1D-CNN presented the lowest performance. Nevertheless, the results are still significant and moreover, it has the lowest standard deviation, suggesting that significant

patterns are present in the EEG signal as proposed by Mariani et al. [329]. The highest standard deviation and difference between Sen and Spe were presented by GRU, suggesting unbalanced results that are not suitable for clinical diagnosis.

5.4. Proposal of a new sleep model

5.4.1. Specification of the model

The AASM manual [21] defines the standards for sleep scoring and assigns a sleep stage or awake period to each 30 s epoch. The scoring rules emphasis visually identifiable features that allow the physicians to produce a similar assignment when performing the visual scoring of the PSG. However, significant variations occur between different experts with an estimated agreement of 83% when scoring the sleep stages [300]. This variation could possibly be due to the difficulty in identifying the necessary features since the scoring rules are unclearly defined [85]. It is also relevant the fact that the segmentation of sleep, that is a continuous process, into a discreet number of stages was founded on the understanding about the sleep processes when the rules were created. Therefore, the new developments in the understanding of sleep are left aside [349], such as the verification that the 30 s epochs have significant limitations to find a practical relationship to the physiological reality of sleep, and are currently unnecessary since the signals are digitally available and can be analyzed by computer algorithms [350].

Following this observations, one-second epochs were considered to allow a significant increase in the time resolution, applying clustering analysis on the EEG monopolar derivation signal to produce clusters that can be interpreted as new sleep stages. However, these new sleep stages do not have a correlation with the macrostructure but have a time resolution which allows the phasic and transient events detection. These events characterize the microstructure and, in turn, the CAP.

Therefore, the goal is to test the hypothesis that CAP can be assessed with a stochastic model, in this case an Hidden Markov Model (HMM), by using symbols created from the EEG clusters (which can be viewed as new sleep stages) as a sequence of observations that are associated with two hidden states (non-CAP or CAP). Consequently, this hypothesis explores the previous conclusion that the CAP cycles have a temporal dependency which could possibly be identified by the HMM, allowing to describe sleep with an approximation to continuous traces.

This model takes into consideration that CAP is a marker of sleep instability hence, the developed model provides a new view for the sleep process, where it oscillates between stable (non-CAP) and unstable (CAP) periods according to the transitions between the new sleep stages (lasting one second and without any particular relation with the conventional sleep stages scored in the sleep macrostructure). Hence, the proposed sleep model is composed of new sleep stages created automatically by the clustering analysis without imposing any defined set of rules (approach employed by the conventional sleep scoring methods defined by the AASM manual [21]), and the transitions between these stages reflect the stability of sleep. The conventional sleep model, defined by the AASM manual [21], considers a limited number of sleep stages, scored every 30 s by a defined set of rules, and each stage reflect the sleep state as either

light, deep, or REM. Therefore, the proposed sleep model envisions the sleep process as oscillations between stable or unstable sleep, while the conventional sleep model denotes the sleep process as cycles of transitions between sleep stages.

However, taking into consideration that the proposed model is based on CAP analysis and that CAP is only defined for the NREM sleep, it is likely that misclassification will occur during the periods where the conventional sleep model defines as either REM or wake after sleep onset. Nevertheless, the proposed method was examined as a proof of concept to assess if the model can properly identify the occurrence of sleep instability periods (related to CAP) by evaluating the transitions between the new sleep stages (assessing if the automatic clustering approach is suitable to create these new sleep stages for the proposed sleep model).

The proposed model can be used to estimate the quality of sleep as poor sleep quality, estimated by the CAP rate, is associated with higher amounts of total sleep under instability (CAP) periods. Thus, the ratio of total duration of CAP to the total duration of non-CAP (estimated by the proposed model) can possibly reflect the quality of sleep, similarly as it is assessed by the CAP rate.

The block diagram of the implemented model is presented in Figure 5.14, and can be divided into four phases. The first is the preprocessing of the EEG signal, and applies the same procedure as previously described for the methods without an explicit feature extraction process. The creation of the clusters was performed in the second phase where two clustering algorithms were tested, specifically, the SOM and the Gaussian Mixture Model (GMM).

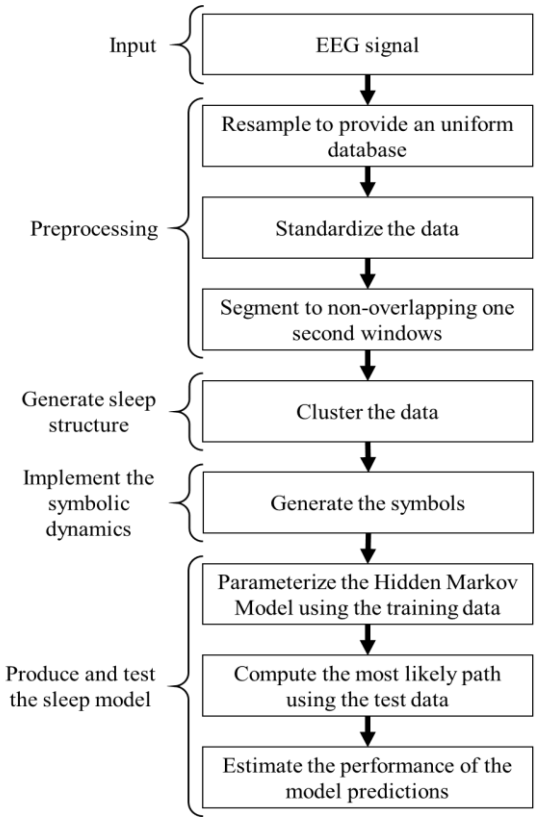


Figure 5.14. Block diagram of the proposed sleep model [351].

The SOM is essentially a neural network of spatially related nodes that have been trained using competitive learning to cluster the input signal, x, into a discrete map which, in this case, had either one or two dimensions. The GMM is composed of multivariate Gaussian distributions, and each can be defined by the mean, μ, and covariance, R [304]. The mixing proportions of all the distributions were further defined by the vector π, and the GMM parameters initialization was performed using the k-means++ algorithm [352], where k is the number of clusters. Initially, π was considered to be the same for all clusters with each Gaussian having identical and diagonal covariance matrices. The Gaussians center, $\boldsymbol{\mu}$, was uniformly distributed along $\boldsymbol{x}$. Subsequently, an iterative process determined each μ_k in two steps that are repeated until the selected μ no longer changes: calculate the distances (the Mahalanobis distances [304] was used) from each point in $\boldsymbol{x}$ to $\boldsymbol{\mu}$, and assign the closest μ_k to each point; choose the new μ_k with a probability that is proportional to the distance, D, from itself to the closest μ_k that was already chosen through [352]

$$P_d\big(x(j,i)\big) = \frac{D\big(x(j,i)\big)^2}{\sum_i D\big(x(j,i)\big)^2} \tag{5.9}$$

The Expectation-Maximization (EM) algorithm was then used to fit the GMM to $\boldsymbol{x}$ where the Probability Density Function (PDF) of each observation was defined as [353]

$$P(\boldsymbol{x};\Theta) = \sum_k \pi_k f_k\,(\boldsymbol{x};\theta_k) \tag{5.10}$$

where $\Theta = (\pi_1, \ldots, \pi_k, \theta_1, \ldots, \theta_k)$, and f_k is the PDF, of a multivariate Gaussian distribution, parameterized by θ. The complete log-likelihood function is given by [353]

$$\begin{aligned}\log P(\Theta) = \sum_i \sum_k a_k^i \Big[&\ln(\pi_k) - 0.5\ln(|R_k|) \\ &- 0.5tr\big[{R_k}^{-1}(x(j,i)-\mu_k)(x(j,i)-\mu_k)^T\big]\Big] + C\end{aligned} \tag{5.11}$$

where a_i^k is the indicator function (1 if i was created by the Gaussian k and 0 otherwise), C is a constant, and tr is the trace operator.

The EM iterates over two steps (iteratively repeated until convergence is achieved) specifically, the expectation (first step) that evaluates the parameters by computing the conditional probabilities [353]

$$\langle a_k^i\rangle \coloneqq P\big(a_k^i = 1 | x(j,i);\Theta^{old}\big) = \frac{\pi_k f_k(x(j,i);\theta_k)}{\sum_k \pi_k f_k\,(x(j,i);\theta_k)} \tag{5.12}$$

and the maximization (second step) that re-estimate the parameters of Θ by [353]

$$\begin{cases} \widehat{\pi_k} = \dfrac{1}{I}\displaystyle\sum_i \langle a_k^i\rangle \\ \widehat{\mu_k} = \left(\displaystyle\sum_i \langle a_k^i\rangle\right)^{-1} \displaystyle\sum_i \langle a_k^i\rangle x(j,i) \\ \widehat{R_k} = \left(\displaystyle\sum_i \langle a_k^i\rangle\right)^{-1} \displaystyle\sum_i \langle a_k^i\rangle (x(j,i)-\widehat{\mu_k})(x(j,i)-\widehat{\mu_k})^T \end{cases} \tag{5.13}$$

The number of the cluster that will further be employed as output for each epoch was defined by the index of the output unit that was more frequently chosen, for the SOM, and by the number of the Gaussian that achieved the highest π, for the GMM. This information was further applied in the third phase where the symbols were created by applying symbolic dynamics which creates a transformation of the sequence of cluster numbers into a sequence of symbols of the alphabet that was created so that the number of the cluster matched the symbol.

The next step was the performance assessment of the proposed model for the CAP assessment, bearing in mind that a good performance denotes that the HMM was capable of assessing the periods of stable and unstable sleep by examining the proposed sleep stages' transitions.

Three SOM topologies were tested, specifically, one with a one dimensional map (line), and two with the two dimensional map (square and hexagon). The number of elements in the map varied from 1×1 to 4×4. Therefore, the SOM alphabet employed the symbols {1, 2, …, 36}. For the GMM the number of Gaussians was varied from 1 to 4 thus, the symbols {1, 2, 3, 4} were used for the GMM alphabet. The words of the SOM model were composed of only one symbol since the possible number of combinations that could form a word is too large if two or more symbols were considered, and the used dataset does not allow to suitably determine the probabilities for all the transitions. As an example, for a word composed of two symbols, the model with 25 output units has 625 possible words. For the GMM a total of six possible word structures were considered: formed by 1 symbol; formed by 2 symbols overlapping with the previous or the next symbol; formed by 3 symbols; formed by 4 symbols overlapping with the previous 2 and next symbols or with the previous and next 2 symbols. Figure 5.15 presents an example of how the words were created for the model based on the GMM with 3 Gaussians, and a different number of symbols to compose the word.

The word sequence was later encoded by associating a unique number to each word, and the result was employed as the sequence of observations of the first order HMM in the final phase. During the training of the model, the maximum likelihood was used to estimate the transition, *a*, and emission, *b*, probabilities of the training dataset by considering the words generated with the symbolic dynamics as the sequence of observations and the CAP cycles (related to the training dataset) as the hidden states. The maximum likelihood estimators for the emission and transition matrix were assessed by counting the number of times an emission or transition occurred since both the states and observations are known in the training dataset [354]. The model was tested by calculating the most likely state path of the sequence of observations, **y**, using the Viterbi algorithm [303], which is similar to the calculation of the shortest path through a trellis diagram where the nodes are the possible states, for each input point, and the nodes and edges weights are log-probabilities. The weight of a path can be defined as [303]

$$\sigma_t(r) = b_r(y_t)\max_s \sigma_{t-1}(s)a_{sr} \quad 5.14$$

where *r* is the state at the current observation, *t*, that involves the most likely path to the state *s* of the previous observation, *t*−1. During the initialization, the initial distribution ***d*** was set to [303]

$$\sigma_1(s) = d_s b_s(y_1) \quad 5.15$$

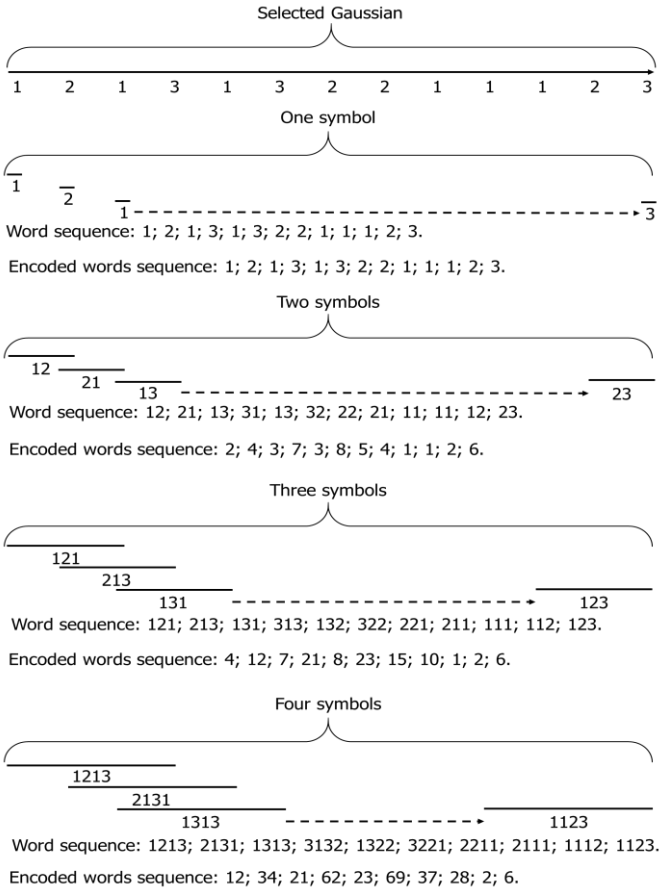


Figure 5.15. Explanatory example indicating how the words were produced from the number of the selected Gaussian, and how they were encoded [351].

A representation of the used methodology to train and test the HMM is presented in Figure 5.16. An example of how the sequence of observations was produced is presented in Figure 5.17. In the first figure, 10 epochs were randomly selected, and further reshaped into a matrix in the second figure (each cell of the matrix indicates the EEG signal's amplitude). The next figure presents the output of applying the GMM with 3 Gaussians to this matrix, that first determined the probability of each cell to belong to each Gaussian, and then selects the Gaussian with the highest probability to define the symbol of the epoch. Afterwards, words with three symbols were created and encoded to produce the sequence of observations for the HMM.

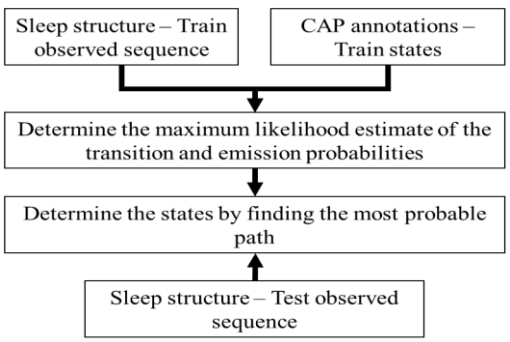


Figure 5.16. Block diagram of used methodology for training and testing the HMM [351].

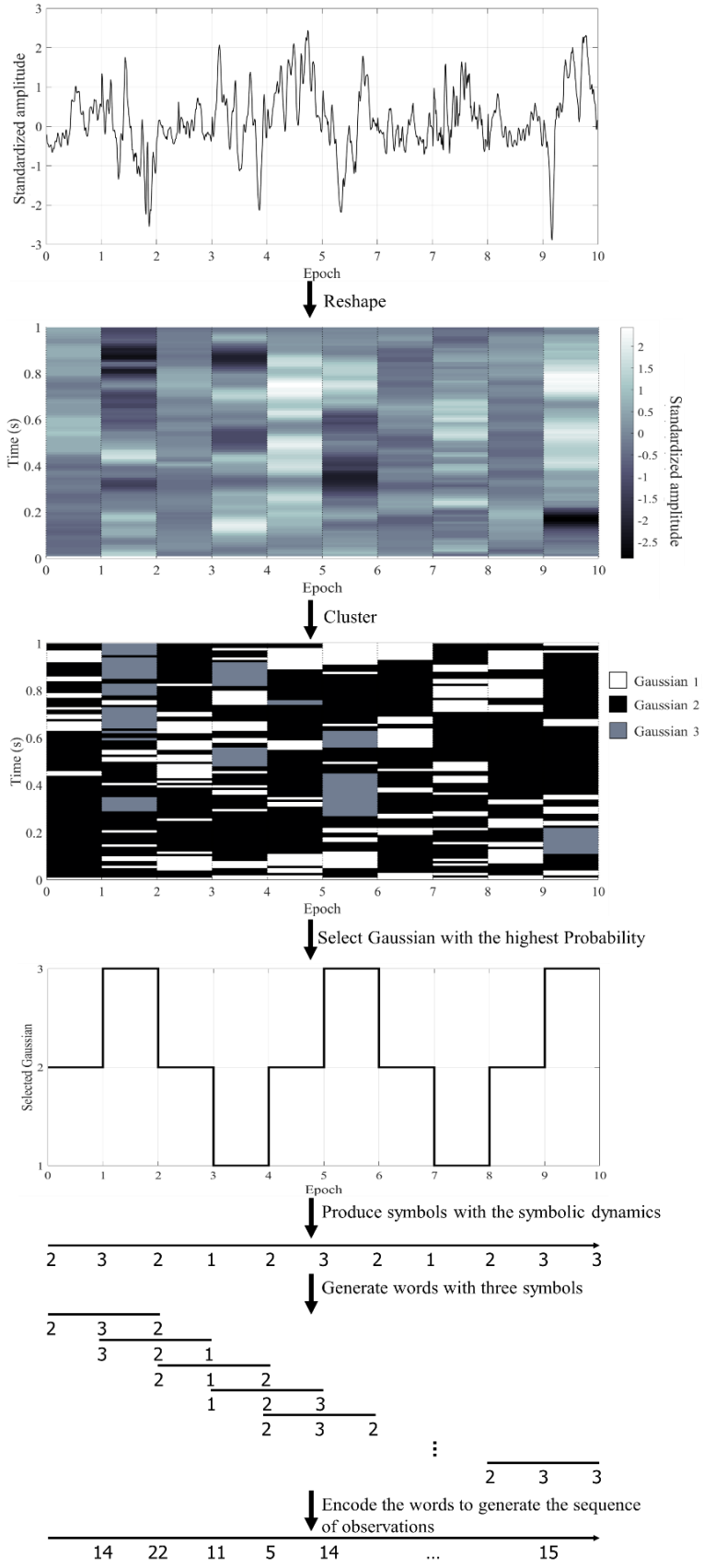


Figure 5.17. Example of how the sequence of observations that were fed to the HMM were produced [351].

5.4.2. Performance assessment

The performance of the developed method, implemented in MATLAB, was assessed by LOOCV (performing 50 repetitions of each iteration to achieve statistically significant results) and using the second CAPSD. The results for the model based on the SOM and on the GMM are presented, respectively, in Tables 5.10 and 5.11. By analyzing Table 5.10 it is possible to verify that the best performance was attained by the SOM with the hexagon topology and 25 units (5×5), while the lowest results were produced by the hexagon topology with 9 units (3×3).

Table 5.10: Performance of the developed algorithms (mean and standard deviation) for CAP detection based on the SOM [351].

Topology	*Number of output units*	*Acc (%)*	*Sen (%)*	*Spe (%)*	*AUC*
Line	9	70.97±3.53	48.24±13.55	84.76±5.13	0.665±0.056
	16	70.55±4.35	52.22±13.09	81.79±6.51	**0.67±0.057**
	25	71.99±4.64	45.5±13.86	88.29±4.35	0.669±0.066
	36	70.95±4.47	40.33±12.72	89.89±3.90	0.651±0.063
Square	9	61.69±2.91	19.96±4.83	87.75±1.97	0.539±0.02
	16	70.42±4.26	47.28±12.34	84.55±4.19	0.659±0.059
	25	74.24±3.89	41.43±12.11	94.57±3.16	**0.68±0.057**
	36	71.27±4.27	42.23±11.25	89.33±3.93	0.658±0.057
Hexagon	9	61.53±2.89	19.87±3.53	87.64±1.89	0.538±0.017
	16	69.28±3.11	42.91±13.99	85.65±9.09	0.643±0.042
	25	73.71±4.56	55.95±12.09	84.77±4.29	**0.701±0.062**
	36	72.97±4.87	49.58±14.35	87.19±4.47	0.684±0.068

Table 5.11: Performance of the developed algorithms (mean and standard deviation) for CAP detection based on the GMM [351].

Number of Gaussians	*Symbols in the word*	*Acc (%)*	*Sen (%)*	*Spe (%)*	*AUC*
2	1	71.01±8.73	59.72±22.24	77.12±8.12	0.684±0.111
	2*	72.47±6.98	63.23±20.16	77.29±8.53	**0.703±0.091**
	2+	72.16±7.12	62.95±20.36	76.96±8.39	0.700±0.093
	3	71.88±7.33	64.18±19.55	75.74±9.59	0.700±0.09
	4^	71.79±7.49	64.18±19.77	75.59±9.40	0.699±0.091
	4#	71.66±7.53	64.04±19.86	75.47±9.38	0.698±0.092
3	1	70.77±7.64	66.30±20.37	72.31±14.73	0.693±0.084
	2*	71.63±7.21	66.92±20.02	73.40±13.71	0.702±0.082
	2+	70.81±7.33	65.74±20.39	72.82±13.00	0.693±0.086
	3	72.23±6.24	65.74±20.33	75.23±10.47	**0.710±0.082**
	4^	71.27±6.71	65.84±20.30	73.63±9.26	0.697±0.089
	4#	71.27±6.69	65.89±20.28	73.62±9.28	0.698±0.087
4	1	68.95±8.72	65.60±21.58	69.79±14.28	0.677±0.010
	2*	70.55±7.52	67.02±20.05	71.55±13.72	0.693±0.085
	2+	69.92±7.71	65.96±20.68	71.24±12.86	0.686±0.090
	3	71.06±6.93	66.20±19.88	72.97±12.28	**0.696±0.082**
	4^	70.98±7.04	65.61±21.01	73.20±10.99	0.694±0.087
	4#	70.76±7.16	66.04±21.09	72.61±10.20	0.693±0.090

* Overlapping with the previous symbol
\+ Overlapping with the next symbol
^ Overlapping with the previous two symbols and the next symbol
\# Overlapping with the previous symbol and the next two symbols

Therefore, an increase in the number of units is beneficial for this topology, conceivably due to the possibility of better exploring the connections between the units. A similar result was produced by the square topology while the linear topology achieved a similar AUC for all structures, and was the best model when less than 25 units were used. The results of the three studied GMM, presented in Table 5.11, suggest that an improvement in the performance is achieved by using words composed of more than one symbol.

The best performance was attained by using 3 Gaussian with words composed of 3 symbols, while the model with 4 Gaussians and words with 1 symbol presented the lowest performance. No significant difference was identified by analyzing where the overlapping occurred (the previous or the next symbol). By comparing the models which attained the best results (based on SOM and GMM) it is possible to verify that a comparable accuracy and AUC was achieved. However, the GMM results are the most balanced (similar Sen and Spe); thus, this model is more suitable for medical analysis.

The average probability distribution regarding the presence or absence of CAP over the SOM with 25 hexagon units (the model that achieved the highest AUC) is presented in Figure 5.18. By analyzing the figure, it is possible to conclude that the presence of CAP generates a spread pattern where the highest probabilities occur in the extremities of the topology, while in the absence of CAP the highest probabilities are found in the center of the topology. A similar analysis was performed for the GMM with 3 Gaussians and words composed of 3 symbols (the model that achieved the highest AUC) and the obtained results are presented in Figure 5.19. By examining the figure, it is possible to verify that the Gaussian number 1 is the most frequently chosen in the presence of CAP while Gaussian number 2 is the less frequently selected. Therefore, these results further advocate the validity of the developed method since the presence of CAP is related to characteristic patterns in both SOM and GMM.

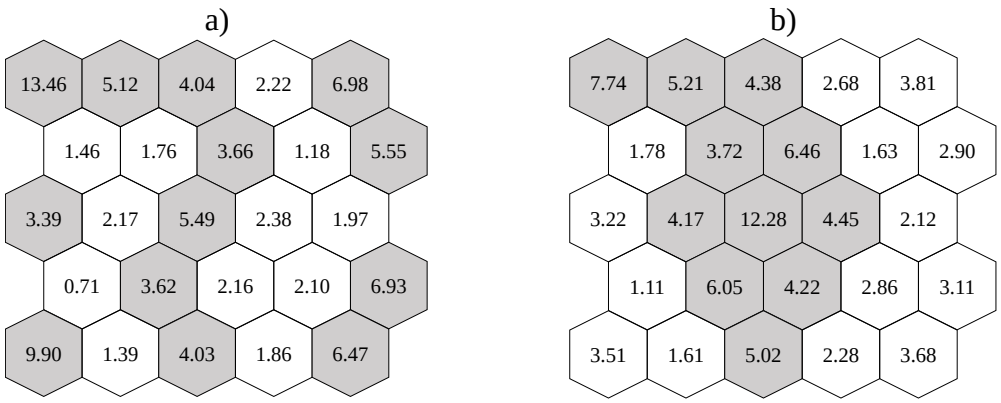


Figure 5.18. Probability distribution, over the hexagon shape topology of the SOM with 25 units, in the a) CAP and b) non-CAP periods.

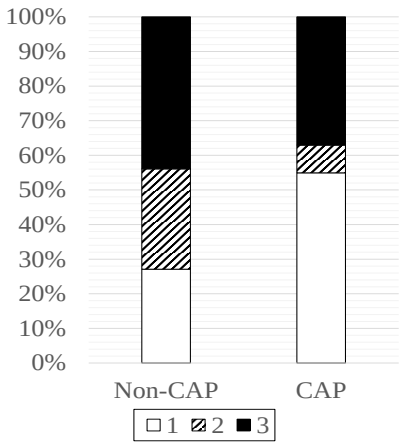


Figure 5.19. Probability distribution of the CAP cycles over the GMM with 3 Gaussians [351].

It was verified that the model has the tendency to overestimate the CAP cycle duration. Consequently, most misclassifications occurred in the CAP cycles boundaries, leading to a lower sensitivity. Figure 5.20 presents an example of such an event. Therefore, an improvement of the boundary recognition can significantly improve the performance of the developed model. These difficulties regarding the CAP boundary assessment were also reported by Largo et al. [54], where it was concluded that this is the main struggle for the clinical application of CAP analysis.

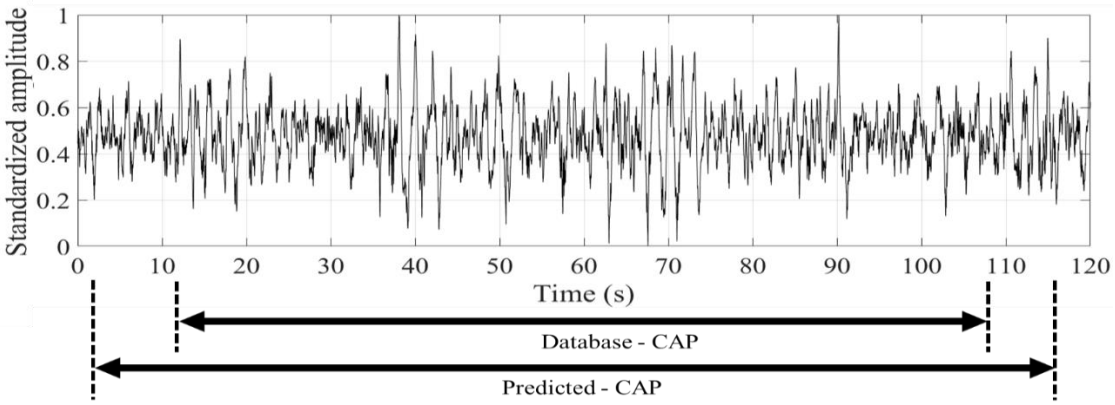


Figure 5.20. Example regarding the overestimation of the CAP cycles boundaries by the GMM with 3 Gaussians [351].

The HMM of the method which attained the highest AUC was further analyzed. The probability of transition from CAP to CAP was 97.42% while the transition probability from non-CAP to non-CAP was 88.31%. Figure 5.21 presents the emission probability of the observations while the normalized emission probability is presented in Figure 5.22. The model has 27 clusters that were created by encoding the GMM with 3 Gaussians. Dendrogram plots, presented in Figure 5.23, were produced to assess the relevance of the observations by creating an agglomerative hierarchical cluster over the utmost distance between the emission probabilities.

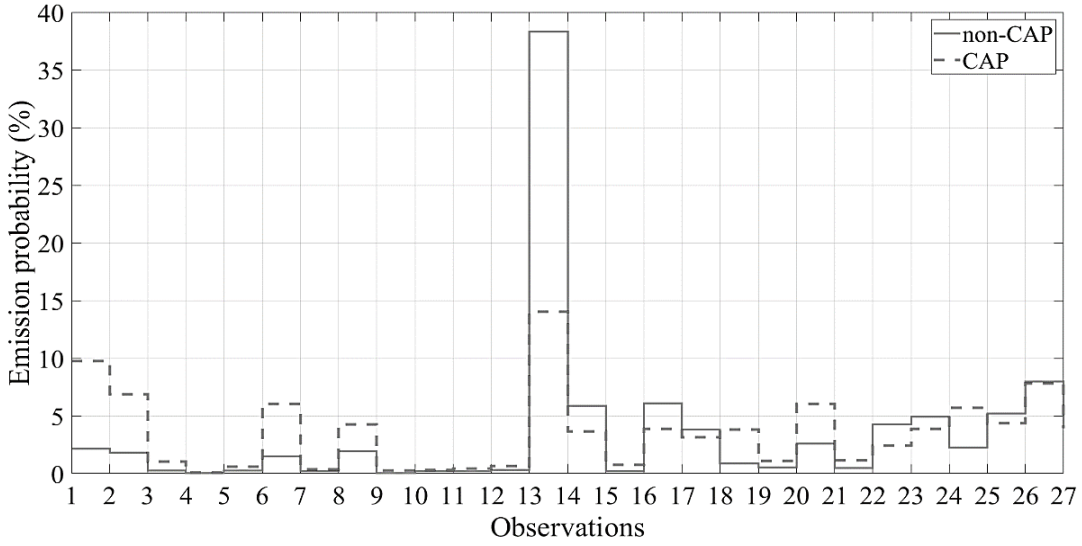


Figure 5.21. Emission probability of the observations of the model that attained the highest AUC, during CAP, and non-CAP cycles [351].

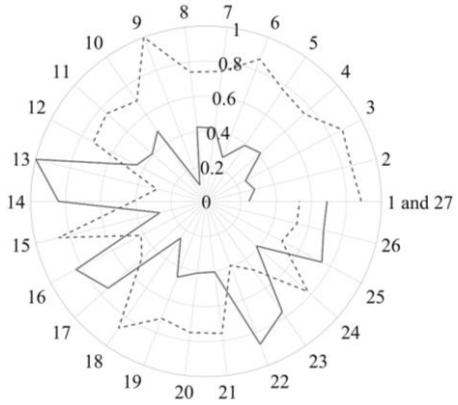
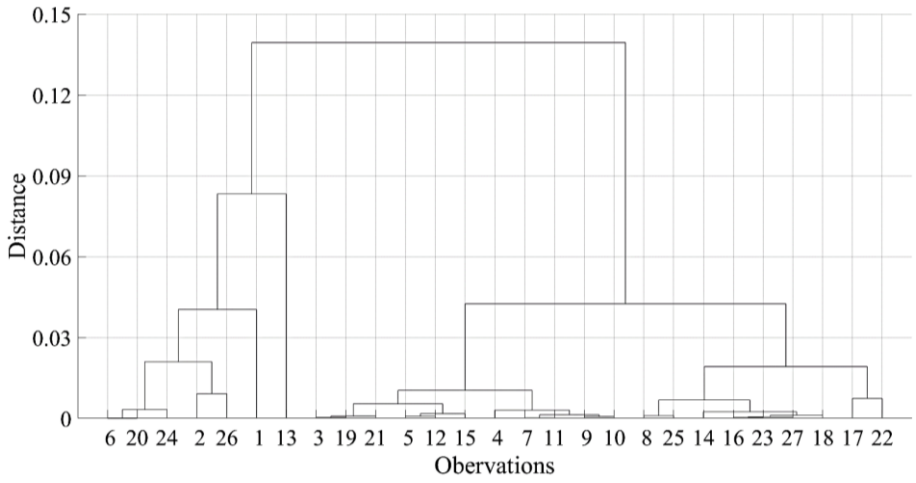


Figure 5.22. Normalized emission probability of the observations of the model that attained the highest AUC, during CAP (dashed line), and non-CAP (solid line) cycles [351].

a)



b)

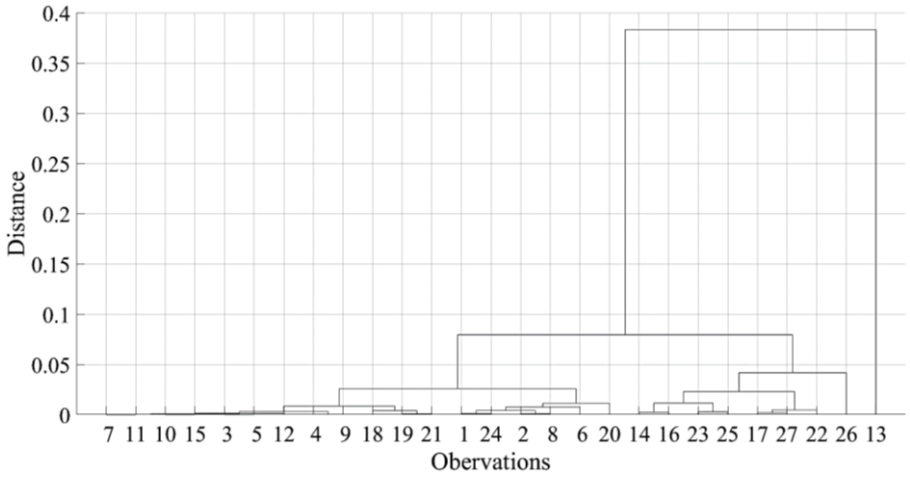


Figure 5.23. Dendrogram plot of the model that attained the highest AUC, during a) CAP, and b) non-CAP cycles [351].

By combining the information from Figures 5.22 (an observation was considered relevant for non-CAP if the normalized probability was higher for non-CAP when compared to CAP and vice-versa) and 5.23 (to order the features by their significance) it was possible to determine the observations relevance. For the CAP cycles, the relevant

features (ordered from most to less significant) were: 1; 2; 24; 20; 6; 3; 19; 21; 5; 12; 15; 4; 7; 11; 9; 10; 8; 18. For the non-CAP cycles the relevant features were: 13; 26; 22; 27; 17; 25; 23; 16; 14.

Therefore, it is likely to be possible to assess the quality of sleep by observing not only the oscillations between stable (indicated by non-CAP) and unstable (indicated by CAP) sleep, estimated by the HMM, but also the observations sequence that fed the HMM. Specifically, if the observations are the most common and relevant during the non-CAP epochs, it is conceivable to infer that good sleep is occurring. The opposite happens when the observations are the most common and relevant during the CAP epochs. These observations can possibly lead to a further characterization of CAP.

Although the models previously developed achieved better results, the proposed method allowed the creation of a statistical model for CAP, and can possibly provide a framework for sleep microstructure analysis. It can also be a relevant step towards the development of a continuous model that is closer to the physiological process of sleep. This assessment can conceivably allow further research in the physiological process that are associated with the production of CAP. Such analysis can be considered problematic to perform in more abstract models such as a deep learning classifier, where the features are usually automatically created and, consequently, difficult to interpret. These assertions are further corroborated by the results reported by Mendez et al. [327] where it was projected that exist 25% of subjectivity and ambiguity in the manual classification of the CAP phases, stressing the difficulties inherent to CAP analysis and suggesting the need for a mathematical model.

It is also important to bear in mind that the A phases which compose the CAP are a simplified approach for sleep microstructure analysis. As a result, the proposed model can be seen a further simplification of the CAP concept, where it is viewed as the sleep instability marker which denotes the oscillations between stable and unstable sleep. Such simplification can possibly be relevant for future sleep analysis by allowing the development of a sleep model which is focused on depicting the quality of sleep instead of providing a rigid representation of sleep. The developed model worked as a proof of concept for the proposed sleep model. Nevertheless, it is relevant to study in a future work if a third hidden state (for the HMM) related to the patterns where the conventional sleep model scores as either REM or wake after sleep onset periods can lead to a better performance for the CAP estimation, since CAP is not defined in the REM sleep or in the wake periods.

5.5. Subtype assessment and characterization analysis

The A phase subtypes were found to be strongly related with the dynamic organization of sleep [28] [19]. Hence, alterations of these subtypes can be an indicator of the existence of sleep disorders [37]. Therefore, the subtype's assessment allows the sleep quality estimation with two different approaches as theorized in this work from the state of the art. The first is by examining conventional sleep quality metrics (in this case, related to CAP) while the second is by identifying the presence of sleep related disorders.

Taking into consideration that the subtypes compose the A phase which, in turn, composes the CAP, hence, a CAP characterization analysis can be performed by

evaluating the subtypes, and the B phases. Recurrence Quantification Analysis (RQA) was used to perform this evaluation since the EEG signal has nonlinear temporal properties. The RQA metrics are [355]: Recurrence Rate (RR); percentage of determinism (DET); maximal line length in the diagonal direction (Dmax); Shannon entropy of the frequency distribution of the diagonal line lengths (ENT); trend (TND).

The objective of this investigation is to perform an event based analysis by examining changes of the dynamics within the EEG signal. Although multiple works have proposed suitable values for the RQA parameters [356], these mostly apply to windows with a fixed length. Taking into consideration that a CAP phase can last between two and 60 s, this methodology is not suitable. Hence, a new methodology was used in this work where the length of the evaluated windows changes according to the evaluated CAP phase length. As a consequence, the RQA analysis parameters are required to be appropriately selected for each examined window. The estimation of the recurrence analysis' delay parameter was performed by examining the mutual average information function [357], selecting the first local minima from multiple possible solutions [358]. The suitable embedding dimension was found by finding the first local minimum of the false-nearest neighbor algorithm, proposed by Kennel et al. [359], evaluating the number of false neighbors (present in the phase-space) as a function of the number of embedding dimensions.

The recurrence threshold is considered to be the most challenging parameter to tune. The employed value should not have either too few or too many recurrences, as it will lead the recurrences to mostly occur due to oscillations caused by noise or to include too many points of the neighborhood which are just consecutive points on the trajectory (hiding the recurrence structure), respectively [360]. For physiological time-series, such as the EEG signal, where noise can be a significant factor, it was recommended to use five times the signal's standard deviation [361] [362]. Hence, this methodology was followed in this work.

The significance of the results produced by the RQA analysis was examined by surrogate testing using the iterative amplitude adjusted Fourier transform surrogates method [363]. The surrogate data created by this algorithm conserves both amplitude distribution, and power spectrum of the original time series. The null hypothesis was that there is no significant difference between the average of the considered RQA metric estimated from the surrogate or from the original EEG signal. The alternative hypothesis was selected when the p-value was less than 0.05, and considered that the average of the considered RQA metric was higher than the metric estimate using the surrogate data (as indicated by Stam et al. [364], it is not expected that the original EEG signal will be less predictable than the surrogate data).

The Spectral Entropy (SEnt) was also computed to estimate the spectral power distribution of the CAP phases. This metric measures how regular the power spectrum is during a particular period of time, where high entropy is probably found in a broader spectrum, with several relevant frequencies, while a low entropy specifies the occurrence of few spikes where the energy is concentrated [365]. The average (normalized by the maximum value) of the RQA metrics and SEnt for each subtype and B phase are presented in Table 5.12.

Table 5.12: Average (normalized by the maximum value) of the RQA metrics and SEnt for each subtype and B phase. For the RQA metrics it is presented, between brackets, the number of subjects which support the alternative hypothesis, a dash, and the average p-value [366].

CAP phase	*RR*	*DET*	*Dmax*	*ENT*	*TND*	*SEnt*
A1	0.51 (13 – 0.10)	1.00 (19 – 0.01)	0.56 (17 – 0.02)	1.00 (18 – 0.03)	0.90 (12 – 0.09)	0.75
A2	0.57 (8 – 0.12)	0.80 (17 – 0.03)	0.65 (16 – 0.03)	0.86 (16 – 0.03)	0.84 (13 – 0.07)	0.86
A3	1.00 (2 – 0.26)	0.62 (19 – 0.00)	1.00 (18 – 0.01)	0.89 (18 – 0.01)	1.00 (18 – 0.02)	1.00
B	0.62 (10 – 0.06)	0.88 (19 – 0.00)	0.65 (19 – 0.00)	0.94 (19 – 0.00)	0.90 (19 – 0.01)	0.83

From Table 5.12 results, it is possible to perform a characterization analysis of the CAP phases. It was observed that the *RR* increased from the A1 to the A3 subtype, suggesting that the probability of a specific state to occur increased from A1 to the A3 subtype. Nonetheless, this metric attained the lowest support for the alternative hypothesis with the highest average p-value. Hence, these results may not be conclusive. On the other hand, the *DET* presented the highest support for the alternative hypothesis with the highest average p-value, denoting that the A1 subtype is more likely to be associated with periodic behaviors (attained the highest average value) while the A3 subtype is more prone to chaotic processes (presented the lowest average value).

Taking into consideration that *Dmax* is associated with the divergence of the trajectory segments, and that the A1 subtype presented the lowest average value for this metric, it may suggest that this subtype depicted the most divergent trajectories while the opposite occurred for the A3 subtype. However, it is important to notice that the used methodology for the RQA metrics calculation evaluated the full duration of each individual activation, to allow a comparative analysis between the A phase subtypes. As the average duration of each A phase subtype is considerably different, therefore, the assessed metrics will be affected by this factor, which is particularly relevant for *Dmax* as it designates how long was the longest diagonal line segment.

A higher *ENT* suggests a higher complexity, with a wider distribution of diagonal line lengths. Therefore, the A1 subtype is likely to have the highest complexity of all subtypes. It was also observed that the A3 subtype presented the highest *TND*, signifying greater nonstationary dynamics. The support for the alternative hypothesis was excellent for all subtypes regarding the *DET*, *Dmax*, *ENT*, and *TND*, and it was fair for the *RR*. However, A1 was the subtype which attained the highest support for the alternative hypothesis while A2 presented the lowest. The results attained for the RQA metrics are possibly related to the A phase subtypes definition, where the A1 subtypes are characterized by synchronized EEG patterns, composed of mild or minor polygraphic variations with high-voltage slow waves, while the A3 subtypes are characterized by desynchronized EEG patterns, with a predominance of low-amplitude fast rhythms. The A2 subtype characteristics are in between the other two subtypes.

By examining the *SEnt* results, it was observed that A1 subtypes have the lowest average value while the opposite is true for the A3 subtypes, signifying the occurrence of a broader spectrum. Once again, the A2 subtypes presented a behavior which was between the other two subtypes, having a lower value than the A3 subtypes but higher than the A1 subtypes. These results are possibly explained by the A phase subtypes definition, which goes from synchronized to desynchronized EEG patterns when going from the A1 to the A3 subtypes.

Regarding the B phase examination, it was verified that both RQA metrics and *SEnt* are typically between the A1 and A2 subtypes, possibly suggesting that the B phase events have characteristics more similar to the A2 and A3 subtypes than the A1 subtypes. These results can also suggest that A1 subtype assessment will be easier to perform than the A2 and A3 subtypes classification.

It was observed that all works presented in the state of the art for the automatic A phase subtypes classification either employ tuned thresholds or a machine learning classifier using multiclass methods. Nonetheless, the threshold-based methodologies are likely to be problematic when trying to generalize for a broader population, and the multiclass machine learning methods, presented in the state of the art, have all reported a poor sensitivity for at least one class.

Since each subtype has relevant information for either the estimation of a sleep quality metric or the assessment of a sleep related disorder, the multiclass classification output is usually transformed to three binary time series containing the information of a subtype in the one versus all representation. As a result, a new approach was employed in this work, using three individual binary classifications (one for each subtype) to examine if by allowing the model to only focus on optimizing the recognition of one class can lead to better performance, while still providing the same information as a multiclass output to be examined by the physicians.

By examining the state of the art it was observed that most works employ feature based analysis using features created by the researchers. Conversely, classifiers such as LSTM are capable of automatically finding relevant patterns in the input signal, which were are likely to have not been emphasized by any feature indicated in the state of the art. Hence, two methodologies were evaluated in this work. The first employed the concept of the methods without an explicit feature creation procedure, feeding the pre-processed signal to the classifier. The second follows the feature based methods concept where features selected by a ranking procedure fed the classifier. The followed experimental procedure for the A phase subtypes assessment (performed by three binary classifiers executing a one versus all analysis) is presented in Figure 5.24.

The first examination was a viability assessment using the subjects from the second CAPSD and TFCV for performance assessment. The algorithms were implemented in Python 3. The network's error optimization was performed by the Adam [343] algorithm and cost-sensitive learning was employed to lessen the influence of class unbalance. The post-processing procedure, previously presented in section 5.2, was applied after the classification to decrease the number of misclassifications.

Three categories of features were considered for the feature based methods. The first was based on the evaluation of the amplitude variation. Specifically, symbolic dynamics analysis was performed through a segmentation method proposed in this work, which implemented a transformation of the input signal into a sequence of symbols by evaluating multiple thresholds for the signal's amplitude (each threshold is a multiple of the standard deviation, σ, of the input signal). An example of the segmentation procedure is presented in Figure 5.25.

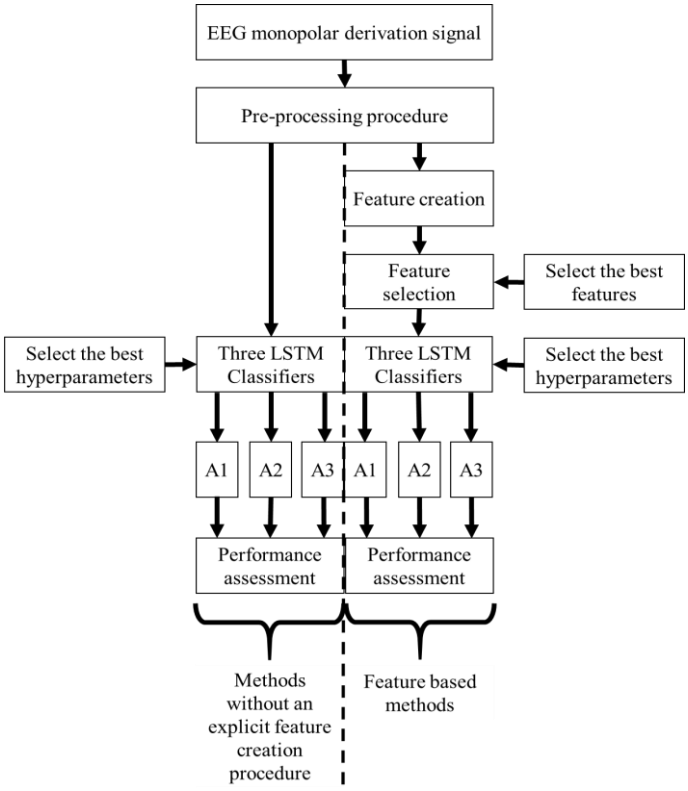


Figure 5.24. Experimental procedure for the A phase subtypes assessment [366].

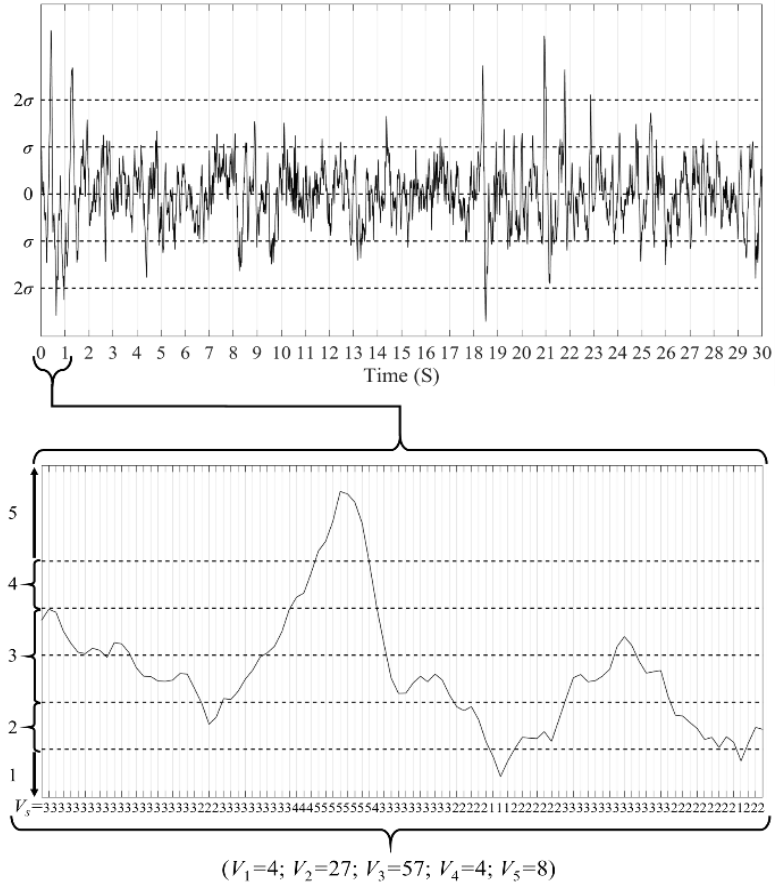


Figure 5.25. Example of the symbolic dynamics procedure, implemented by five thresholds (M=5), to create the vector which contained the sequence of symbols [366].

The most relevant parameter of the employed symbolic dynamics analysis is the multiplier, *M*, which will define the number of thresholds. The output's vector contains the sequence of symbols, V_s, from which it is possible to evaluate multiple statistical features [326] [367]. Therefore, the analysis was performed by counting the number of incidences of each symbol since this information can be used by the LSTM to assess the time based variations over multiple time steps. The last feature evaluated in this category was the amplitude variation metric, A_v, a metric proposed in this work given by

$$A_V = max(E) - [max(E-1) - max(E-2)] \quad 5.16$$

where *max* is the highest value of the epoch *E*. The features proposed in this category can detect the amplitude oscillations, which are characteristic of the A phase, and work as markers for the A phase onset transitions since the protocol for CAP analysis dictates that the phasic activity, which initiates an A phase, must have an amplitude at least 33.3% higher than the background activity of the two previous one second epochs [10].

The second category of features examined the power in frequency bands (using the PSD calculated by the Welch's method) following the procedure employed for the feature based methods presented in section 5.2. The third category of evaluated features combined concepts of the two previous categories by evaluating the ratio of the highest value of the epoch to the measured PSD for all characteristic EEG frequency bands. These features were also proposed in this work and follow a similar concept employed for the proposal of the macro-micro structure descriptors [368]. However, the proposed metrics combined information of both time (by evaluating the maximum value of the epoch) and frequency (considering the PSD of the frequency band) instead of appraising the mean power of the frequency bands as the macro–micro structure descriptors do. As a result, the proposed metric can differentiate the A phase subtypes since the A1 subtype is mostly composed of high-amplitude fast rhythms while the A3 subtype is mostly composed of low-amplitude fast rhythms (the A2 subtypes is in between the A1 and A3 subtypes). The relevance of the features for each subtype classification was assessed by the mRMR algorithm, since the computational time to select the features using SFSe is not reasonable when using the LSTM for classification.

Since the classifiers are implementing a one-versus-all evaluation, an error matrix was estimated to evaluate if the misclassifications were occurring because the classifier was confusing an activation phase with the background activity or because a manifestation of another subtype is leading the classifier to misclassify the background activity as an activation phase. Five counting variables were created for each subtype classification, considering one for the NREM periods, one for REM periods, and one for wake periods. When a false positive was identified, the algorithm checked if an activation phase, which was related to another subtype, happened. If so, then the counting variable associated with that subtype was incremented. If not, then the classifier confused the background activity with an activation phase hence, the counting variable, associated with the macrostructure label of the examined epoch, was increased. A similar procedure was employed when a false negative was identified. Afterwards, the value of the counting variables was divided by the total number of misclassifications.

A variation of the HOSA, named HOSA-L, was developed to select the LSTM structure by examining multiple possible architectures. After the input layer, I_{pt}, the following layer of the network was either a LSTM or a Bidirectional LSTM (BLSTM).

The subsequent layers were equal to the first recurrent layer. The last recurrent layer can be followed by a fully connected layer (F_C). The number of layers, G, was increased until the chosen maximum, G_{max}, number of five or stopped earlier if there was no significant improvement in the AUC (considering a threshold, t_h, of 1% to define the minimum relevant improvement) from the previous iteration (with G – 1 layers). The number of time steps, T, used by the recurrent layers, started with five (T_{start}) and were increased in steps, T_{step}, of 10 until T_{max}, selected to be 35 (for this work, after 35 the increase in time steps was found to be counterproductive, decreasing the performance).

The number of hidden units, N, employed by the first recurrent layer started at 100 (N_{start}), and was increased in steps, N_{step}, of 100 until the maximum, N_{max}, of 400 (in this work, the increase in performance saturated at 400). The number of hidden units used by the F_C layer was chosen to be either half (using the *floor* function to round the result of the division), the same, or twice the number of hidden units used by the previous recurrent layer. For the models with a cascade of LSTM layers, the number of hidden units employed by the second LSTM layer was the same as the number of hidden units used by the first LSTM layer. The Output layer employed the soft-max function. The HOSA-L follows the subsequent pseudo code:

```
HOSA-L (Data, G_max, T_step, T_max,, N_start, N_step, N_max, t_h)
G = [1, 2, …, G_max]
T = [T_start, T_start + T_step, …, T_max]
N = [N_start, N_start + N_step, …, N_max]
L = [LSTM, BLSTM]
for t = 1 to length (T)
   for n = 1 to length (N)
      for g = 1 to length (G)
         for l = 1 to length (L)
            Layer = L (l)
            for m = 1 to 4
               Net_0,l,t,n,0,m ← I_pt (Data, T (t))
               for z = 1 to g
                  Net_z,l,t,n,0,m ← Net_z-1,l,t,n,0,m + Layer (N (n))
               if m == 1
                  N_prev = floor (N (n) / 2 + 1 / 2)
                  Net_g,l,t,n,1,m ← Net_g,l,t,n,0,m + F_C (N_prev) + F_C (2)
               else
                  if m == 2
                     N_prev = N (n)
                     Net_g,l,t,n,1,m ← Net_g,l,t,n,0,m + F_C (N_prev) + F_C (2)
                  else
                     if m == 3
                        N_prev = N (n) × 2
                        Net_g,l,t,n,1,m ← Net_g,l,t,n,0,m + F_C (N_prev) + F_C (2)
                     else
                        Net_g,l,t,n,1,m ← Net_g,l,t,n,0,m + F_C (2)
               AUC_g,l,t,n,m ← test (train (Net_g,l,t,n,1,m))
         AUC_g,l,t,n,m,max = max(AUC_g,l,t,n,m)|for all l,m
         if g > 1
            if AUC_g,l,t,n,m,max – AUC_g-1,l,t,n,m,max ≤ t_h
```

```
            if AUC_{g,l,t,n,m,max} > AUC_{g-1,l,t,n,m,max} ≤ t_h
                BestNet_{t,n} = Net_{g,l,t,n,1,m}|AUC_{g,l,t,n,m,max}
            else
                BestNet_{t,n} = Net_{g-1,l,t,n,1,m}|AUC_{g-1,l,t,n,m,max}
            break
        else
            BestNet_{t,n} = Net_{g,l,t,n,1,m}|AUC_{g,l,t,n,m,max}
            else
                BestNet_{t,n} = Net_{g,l,t,n,1,m}|AUC_{g,l,t,n,1,m,max}
return BestNet_{t=1:length(T),n=1:length(N)}
```

For the methods without an explicit feature creation procedure, it was verified that the architecture with the best performance to complexity (number of parameters) ratio was composed of one LSTM layer followed by one dense layer (using half of the LSTM's number of hidden units). A similar conclusion was reported by Yadav et al. [369] which have concluded that one LSTM layer outperformed models with multiple layers. The replacement of LSTM by the BLSTM had an increase of less than 1% in the AUC, and the cascade LSTM architecture decreased the average AUC. Therefore, a total of 256 networks were created for each subtype and each network was simulated 10 times. As a result, the total number of examined classifiers was 7680. The classifiers found to be the most suitable for each combination of T and N were further examined, performing 50 repetitions of each iteration to achieve statistically significant results, and the results (AUC and Acc) of these analyses are presented in Figure 5.26.

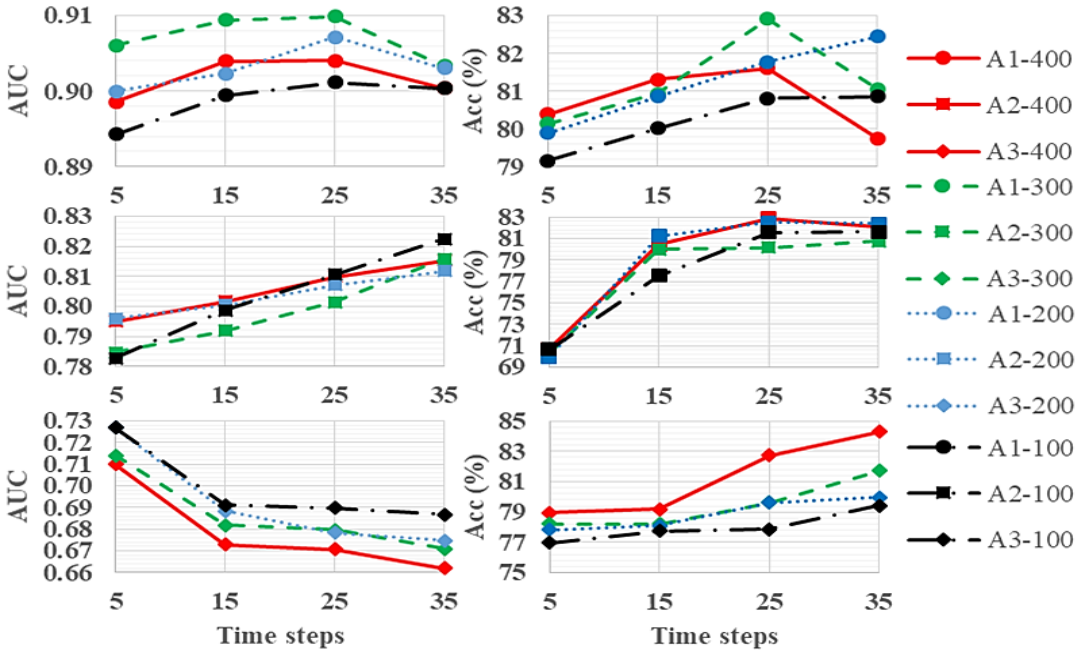


Figure 5.26. Variation of both AUC and Acc, of the network with the best architecture for the models without an explicit feature creation procedure for different time steps. The activation subtype is specified in the legend, followed by the employed number of hidden units of the LSTM [366].

It was observed that the A1 subtype classification reached the best AUC using 300 hidden units in the LSTM with 25 time steps (this subtype was the least sensitive to the number of time steps). The A2 subtype had significant variations by changing the number

of time steps, providing the best AUC when using 100 hidden units in the LSTM, and 35 time steps. The A3 subtype classification was the most challenging to be performed since the LSTM was not capable of effectively evaluate the information provided by the time steps, reaching the best AUC using 100 hidden units in the LST and five time steps.

The learning curves of the best classifiers (for each subtype) are presented in Figure 5.27. By evaluating the linear tendency line, it is possible to conclude that the inclusion of more data could lead to an improvement of the performance, although it might not be significant as the learning curves for the A2 and A3 subtypes already started to reach a saturation point.

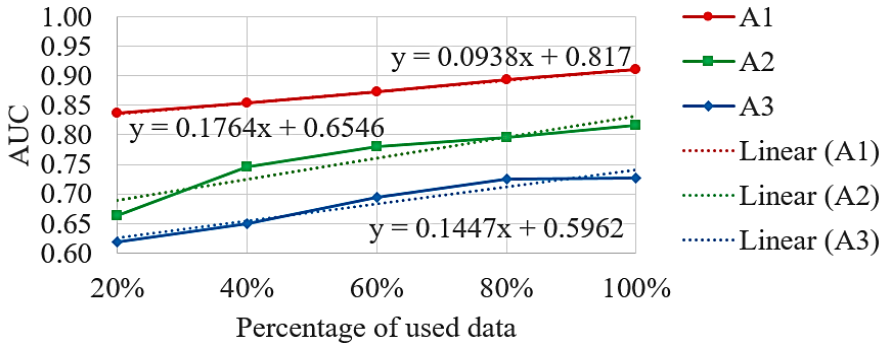


Figure 5.27. Learning curves for the best classifiers (for each subtype) with the linear tendency line and its respective equation [366].

Table 5.13 presents the performance of the best architecture for the models without an explicit feature creation procedure, where it was concluded that all classifiers attained a similar accuracy. Nevertheless, the A1 subtype reached a significantly better AUC, denoting a balanced performance, while the A3 subtype evaluation presented the lowest AUC, demonstrating unbalanced results. The error matrix of these results is presented in Table 5.14. It was observed that the A1 subtypes were mostly misclassified with the background activity (usually during NREM). Nearly all misclassification happened at the end of the activation phase, where the classifier predicted a longer duration of the activation (frequently by one or two epochs). Consequently, the main difficulty for the classification of these subtypes was the correct recognition of the activation phase offset (end of the activation boundary). This difficulty was present in all subtypes classifications.

Table 5.13: Performance (average ± standard deviation) of the viability assessment for the methods without an explicit feature creation procedure for each A phase subtype [366].

CAP phase	*Acc (%)*	*Sen (%)*	*Spe (%)*	*AUC*
A1	82.92±3.65	83.11±7.22	82.90±4.15	0.910±0.015
A2	81.66±5.90	63.97±11.23	82.03±6.19	0.822±0.029
A3	76.97±7.09	56.03±8.63	77.69±7.58	0.727±0.025

Table 5.14: Error matrix of viability assessment for the methods without an explicit feature creation procedure for each A phase subtype [366].

	Miss-classified during an activation phase (%)			*Miss-classified during background activity (%)*		
CAP phase	*A1*	*A2*	*A3*	*NREM*	*REM*	*Wake*
A1	-	2.11	1.91	88.85	1.69	5.44
A2	7.61	-	2.87	76.84	5.22	7.46
A3	3.84	2.30	-	49.01	14.72	30.13

It was verified that some of the A2 subtypes misclassifications happened during the occurrence of an activation phase related to another subtype, suggesting that the classifier was occasionally confusing the A2 subtype with other activation from other subtypes. This is possibly related to the A2 subtype definition, where it has characteristic patterns of both A1 and A3 subtypes. Another relevant factor was the increase of misclassifications associated with the wake (after the sleep onset) or REM periods. This tendency was further manifested in the A3 subtypes, denoting that it is likely for several of the patterns associated with both A2 and A3 subtypes to be related to the background activity during these periods. Therefore, the recognized difficulties for the A2 subtype classification are the accurate discrimination between the other subtypes, and the offset boundary detection.

The main difficulty for the A3 subtypes classification was the high number of misclassifications related to the background activity periods since more than a third of the REM periods and nearly all wake (after the sleep onset) periods were misclassified as an A3 subtype, supporting the need to use distinctive features to help discriminate the occurrence of the activations related to the A3 subtypes. It was also verified that some misclassifications happened when either the A1 or A2 subtypes presented desynchronized EEG patterns. As a result, the main difficulties for the A3 subtype classification were offset boundary detection, correctly differentiate the occurrence of an A3 activation and the presence of background activity, and properly differentiate the desynchronized EEG patterns associated with the A3 subtypes when comparing to the patterns of other subtypes.

The mRMR algorithm was employed for ranking the relevance of the features, whose specification and associated identification number are presented in Table 5.15. The ranking order of the features is presented in Figure 5.28. The optimal number of thresholds to use for the symbolic dynamics was assessed by increasing the number until these features (related to the higher thresholds) were ranked by the mRMR algorithm as the less relevant. It was observed that such occurred using nine thresholds (M=9) as the features with the identification number 1 and 9 (V_1 and V_9, respectively) are the less relevant for the A1 subtypes.

Table 5.15: Examined features for the A phase subtypes examination [366].

Feature identification number	*Description*	*Denomination*
1	Number of occurrences of symbol 1	V_1
2	Number of occurrences of symbol 2	V_2
3	Number of occurrences of symbol 3	V_3
4	Number of occurrences of symbol 4	V_4
5	Number of occurrences of symbol 5	V_5
6	Number of occurrences of symbol 6	V_6
7	Number of occurrences of symbol 7	V_7
8	Number of occurrences of symbol 8	V_8
9	Number of occurrences of symbol 9	V_9
10	Amplitude variation metric	A_v
11	PSD of the Delta band	PSD_D
12	PSD of the Theta band	PSD_T
13	PSD of the Alpha band	PSD_A
14	PSD of the Sigma band	PSD_S
15	PSD of the Beta band	PSD_B
16	Ratio of the maximum value to the PSD of the Delta band	R_D
17	Ratio of the maximum value to the PSD of the Theta band	R_T
18	Ratio of the maximum value to the PSD of the Alpha band	R_A
19	Ratio of the maximum value to the PSD of the Sigma band	R_S
20	Ratio of the maximum value to the PSD of the Beta band	R_B

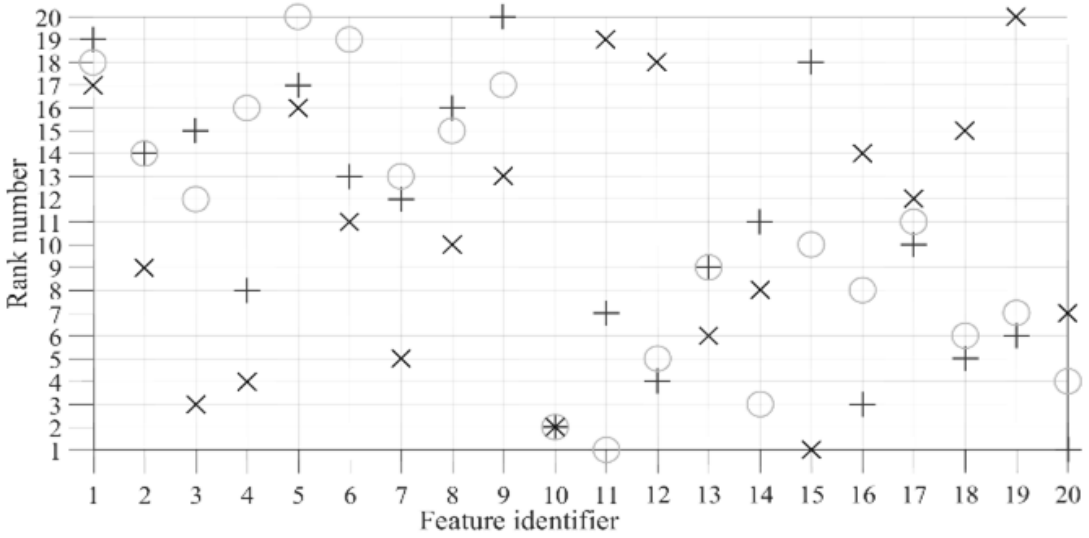


Figure 5.28. Features evaluated for each A phase subtype assessment ranked by the mRMR algorithm. A1 is designated by '+', A2 is designated by 'o', A3 is designated by 'x'. The feature identification number identifies the feature [366].

The optimal number of features for each A phase subtype classifier was found by evaluating 20 feature sets, where the first set was only composed of the most relevant feature (the lowest mRMR ranking), the second set was composed of the two most relevant features (the two features with the lowest mRMR ranking), and so one until the last set which was composed of all features ordered by the mRMR ranking.

Though the grid search procedure, employed for the methods without an explicit feature creation procedure, it was observed that a network composed of one LSTM layer (with 300 hidden units), followed by one dense layer (with half of the LSTM's hidden units), attained the best performance to complexity ratio for the A1 subtype classification. Taking into consideration that the performance for the A2 and A3 subtypes with 300 LSTM's hidden units in the LSTM layer was similar to the models with 100 LSTM's hidden units (best performance for the methods without an explicit feature creation procedure). Hence, all classifiers tested for the feature based methods employed the same architecture (one LSTM layer with 300 hidden units, followed by one dense layer with half of the LSTM's hidden units) in order to reduce the number of tested simulations.

The variation of both AUC and Acc, according to the number of features, for the feature based method is presented in Figure 5.29. By evaluating the results, it was concluded that the best AUC for all classifiers was attained using 25 time steps, suggesting that the classifiers were capable of recognizing the temporal information present in the time steps. The sequence of selected features for each subtype is presented in Table 5.16, while the performance of the best models is presented in Table 5.17.

Table 5.16: Sequence of selected features for the features based methods for the A phase subtype assessment.

CAP phase	*Feature sequence*
1	R_B; A_v; R_D; PSD_T; R_A; R_S; PSD_D; V_4; PSD_A; R_T; PSD_S; V_7; V_6; V_2; V_3; V_8; V_5; PSD_B
2	PSD_D; A_v; PSD_S; R_B; PSD_T; R_A; R_S; R_D; PSD_A; PSD_B; R_T; V_3; V_7
3	PSD_B; A_v; V_3; V_4

Table 5.17: Performance (average ± standard deviation) of the viability assessment for the feature based methods for each A phase subtype [366].

CAP phase	*Acc (%)*	*Sen (%)*	*Spe (%)*	*AUC*
A1	81.91±2.43	87.83±3.94	81.60±2.75	0.921±0.011
A2	79.65±5.69	81.07±7.26	79.59±5.96	0.884±0.025
A3	84.61±5.31	70.37±8.41	85.07±5.72	0.855±0.034

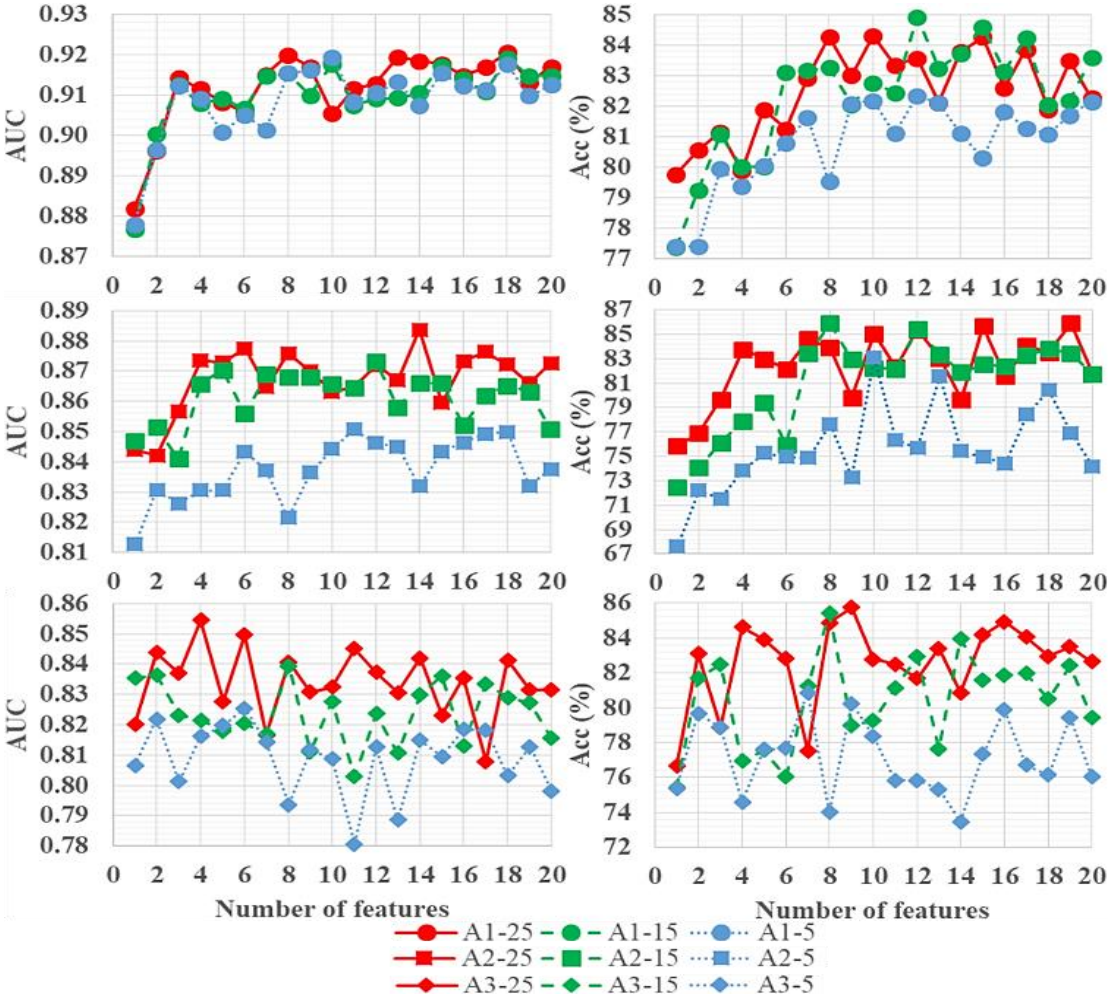


Figure 5.29. Variation of both AUC and Acc, of the network with the best architecture for the feature based models for different number of features selected. The activation subtype is specified in the legend, followed by the employed number of time steps of the LSTM [366].

By evaluating the features based methods results, it was concluded that the A1 subtype classification attained the best AUC, denoting balanced results, while the lowest AUC was reported by the A3 subtype classification, suggesting that the A3 assessment is once more the most challenging. When comparing with the results of the methods without an explicit feature creation procedure, presented in Table 5.13, it is possible to observe an increase of 18% for the AUC of the A3 subtype classification, caused by the improvement of about 25% and 10% for the Sen and Spe, respectively. These results support the viability of using feature based models for the A3 subtype classification, as the methods without an explicit feature creation procedure were not capable of extracting all relevant information when directly fed by the EEG signal. For the A2 subtype

classification it was observed an increase of 7% in the AUC by using the feature-based method. The A1 subtype classification attained a similar AUC for both methods. Another relevant indication for the feature-based models was the reduction of the standard deviation of the performance metrics. As a result, it was concluded that the features based methods are likely to be the most suitable for the A phase subtype assessment.

The error matrix for the performance of the feature based methods is presented in Table 5.18. When comparing with the error matrix of the methods without an explicit feature creation procedure, presented in Table 5.14, it is possible to verify a reduction of the misclassification caused by wake (after the sleep onset) for the A1 subtypes. Nonetheless, the classifier still retains the offset boundary detection issue. On the other hand, the issues identified for the methods without an explicit feature creation, regarding the A2 subtype classification, were mitigated by the feature based methods. A reduction of the misclassifications associated with the occurrence an A1 subtype was observed for the A3 subtype classification when using the feature based methods. The boundary issues related to this subtype were also significantly reduced with the tradeoff of an increase in the misclassifications associated with the occurrence of REM periods. It was conceptualized that these misclassifications could be linked to the REM sleep onset modulation which is associated with the A2 and A3 subtypes [19]. Conversely, this difficulty can possibly be diminished by using a method to isolate and remove the REM periods from the classification.

Table 5.18: Error matrix of viability assessment for the feature based for each A phase subtype [366].

	Miss-classified during an activation phase (%)			*Miss-classified during background activity (%)*		
CAP phase	*A1*	*A2*	*A3*	*NREM*	*REM*	*Wake*
A1	-	4.69	1.66	91.12	1.59	0.94
A2	7.75	-	3.03	79.92	6.26	3.04
A3	2.56	2.46	-	48.49	22.52	23.97

The best methods without an explicit feature creation procedure and feature based methods were further examined using the third CAPSD (15 control subjects and 4 subjects with sleep-disordered breathing), and LOOCV was employed for performance assessment (performing 50 repetitions of each iteration to achieve statistically significant results). The results for the method without an explicit feature creation procedure are presented in Table 5.19, while Table 5.20 presents the results for the feature based methods. These results are similar to the results attained by the viability assessment analysis and corroborate the prior conclusions.

Table 5.19: Performance (average ± standard deviation) of the methods without an explicit feature creation procedure for each A phase subtype [366].

CAP phase	*Acc (%)*	*Sen (%)*	*Spe (%)*	*AUC*
A1	88.30±2.29	61.90±14.20	89.56±2.68	0.881±0.032
A2	85.79±4.25	52.77±13.99	86.60±4.21	0.802±0.059
A3	80.15±12.26	57.65±14.20	80.51±13.90	0.765±0.063

Table 5.20: Performance (average ± standard deviation) of the feature based methods for each A phase subtype [366].

CAP phase	*Acc (%)*	*Sen (%)*	*Spe (%)*	*AUC*
A1	82.68±4.74	83.26±11.83	82.52±5.19	0.915±0.029
A2	83.25±5.82	73.83±12.84	83.40±6.00	0.874±0.044
A3	83.35±8.71	70.94±14.42	83.56±9.56	0.857±0.052

5.6. Sleep quality estimation

Most works presented in the state of the art performing the A phase examination use features created with domain-specific knowledge of the researchers. However, it was also observed that it is becoming considerably more difficult to find new sets of features which can be used for attaining a better performance than reported methods in the state of the art. Another relevant factor is that the combination of two or more features does not guarantee a performance improvement, and generally the feature based methods require a sorting procedure to identify the most relevant for the intended classification [370]. As it was previously discussed in section 5.3, these difficulties can be eliminated by deep learning models, which automatically learn the significant patterns directly from the data. On the other hand, the relevant patterns can only be identified if there is sufficient data to train the classifier. Such can be problematic for the CAP analysis as the classification is performed every second (using, in this work, 100 samples per second). For this reason, a different approach was followed for the classifier proposed in this section, evaluating consecutive overlapping windows which fed a 1D-CNN. This classifier was selected since it can exploit spatially local correlations in the input signal. A post-processing procedure was employed to correct the misclassified A phases and the output was fed to a FSM, to perform the CAP cycle scoring.

The block diagram of the developed model is presented in Figure 5.30. The first examination was a viability assessment using the subjects from the second CAPSD. The algorithms were implemented in Python 3 using Keras library. The NREM classifiers examine the same pre-processed signal employed for the A phase classification to provide a model which is more suitable for hardware implementation with few computational resources since it is less computational demanding.

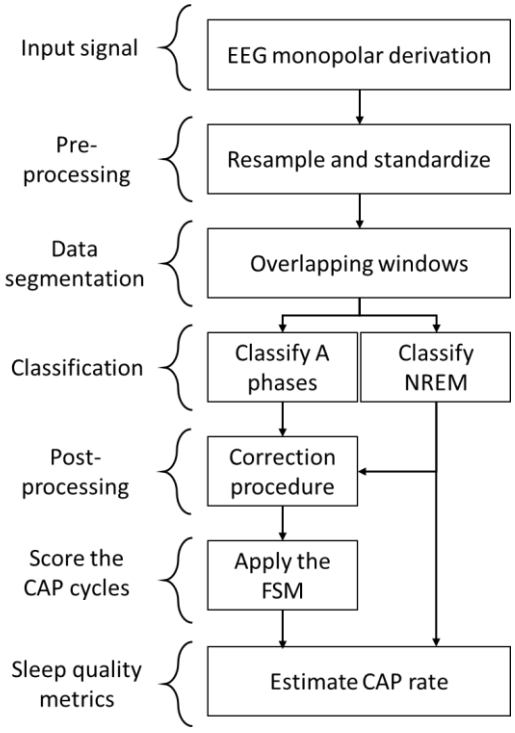


Figure 5.30. Block diagram of the developed algorithm for sleep quality estimation from the EEG signal analysis.

Three methodologies were evaluated to perform the overlapping, using either the first (first scenario overlaps on the right), the central (second scenario overlaps on the right and left), or the last (third scenario overlaps on the left) 100 samples of the overlapping window as the samples corresponding to the epoch's label. An example of the evaluated windows for each scenario is presented in Figure 5.31.

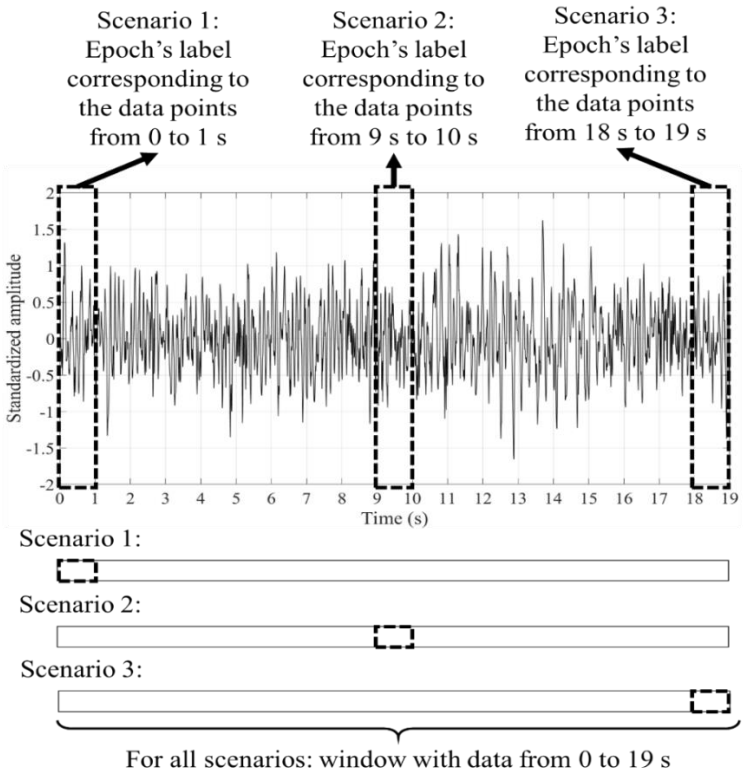


Figure 5.31. Example of the overlapping windows for each of the three examine scenarios.

The HOSA-C, a variation of the HOSA optimized for the 1D-CNN analysis, was developed to find the best classifier's structure for the optimal input configuration. For each evaluated combination of the most relevant hyper-parameters of the classifier (number of layers, kernel size, and number of kernels [371] [372]), the overlapping length, O, of the segmented windows was iteratively altered, testing all scenarios of overlapping, A_p, for the overlapping window W.

The required time for the algorithm's searching procedure to find the best solution was improved by using the group of GL concept, where (in the work presented in this section) each group was composed of fixed sequence of one convolution layer, followed by one subsampling layer (applying the max pooling operation with down sample by a factor of two), and 10% dropout was applied to the group's output. The stride and filter size of the GL was selected to be two since these values are frequently employed in the state of the art for 1D-CNN, allowing to reduce the data's dimensionality while preserving the highest excitations from the convolutional feature maps. ReLU was used as the activation function, and the network's error optimization and performance assessment was performed by the Adam algorithm [343] using cost-sensitive learning. The early stopping procedure presented in chapter 4 was used to avoid overfitting the classifier, and reduce the simulation time.

The HOSA-C was initiated without overlapping, and the overlapping's length was increased considering steps of two seconds, up to a maximum window length, O_{max}, of 17 seconds (upper limit empirically found to be in the saturation point for the A phase estimation performance). The algorithms started with one input layer (I_{pt}), one GL, and two output layers (F_C). The number of GL, G, was iteratively incremented until a maximum value, G_{max}, selected to be four, or until no relevant improvement was attained for the AUC (considering a threshold, t_h, of 1% to define the minimum relevant improvement) when the subsequent GL was introduced.

The number of kernels, K, used in the convolution layer of the first GL, was varied between eight and 128, using a step of 2^M where $K_{start} \leq M \leq K_{max}$ (K_{start} and K_{max} were four and seven, respectively). The subsequent GL were introduced using either the same or twice (representing an increment of two for the multiplier MUL_{max}) the number of kernels of the previous GL (providing a linear growth for the number of simulations). The network could include a FC layer between the last GL and the output layer, and the number of neurons, N, which composed this dense layer was varied from 50 (N_{start}) to the maximum value, N_{max}, of 150, in steps (N_{step}) of 50. The HOSA algorithm follows the subsequent pseudo code:

HOSA-C (*Data*, G_{max}, K_{start}, K_{max}, O_{max}, N_{start}, N_{step}, N_{max}, MUL_{max}, t_h)
$G = [1, 2, \ldots, G_{max}]$
$O = [0, 1, 3, 5, \ldots, O_{max}]$
$K = 2^M$ where $K_{start} \leq M \leq K_{max}$
$N = [N_{start}, N_{start} + N_{step}, \ldots, N_{max}]$
for g = 1 to length (G)
  for o = 1 to length (O)
    for k = 1 to length (K)
      for n = 1 to length (N)
        if O (o) > 0
          $W = [2 \times O(1) + 1, 2 \times O(2) + 1, \ldots, 2 \times O(\text{length}(O)) + 1]$
          $A_p = [W(1), W(floor(W / 2 + 1)), W(\text{length}(W))]$
        else
          $A_p = 1$
        for a = 1 to length (A_p)
          $Net \leftarrow I_{pt}(Data, O(o), A_p(a))$
          for z = 1 to g
            if z == 1
              $mul = 1$
              $Net_{g,o,k,n,a,z,mul:MULmax} \leftarrow Net + GL(K(k))$
              $k_{z,mul:MULmax} = K(k)$
            else
              for mul = 1 to $MULmax$
                $k_{z,mul} = mul \times k_{z-1,mul}$
                $Net_{g,o,k,n,a,z,mul} \leftarrow Net_{g,o,k,n,a,z-1,mul} + GL(k_{z,mul})$
          $Net_{g,o,k,n,a,z,mul} \leftarrow Net_{g,o,k,n,a,z,mul} + F_C(N(n)) + F_C(2)$
          $AUC_{g,o,k,n,a,z,mul} \leftarrow \text{test}(\text{train}(Net_{g,o,k,n,a,z,mul}))$
  $AUC_{g,o,k,n,a,z,mul,max} = max(AUC_{g,o,k,n,a,z,mul})|_{for\ all\ o,k,n,a,z,mul}$
  if g > 1
    if $AUC_{g,o,k,n,a,z,mul,max} - AUC_{g-1,o,k,n,a,z,mul,max} \leq t_h$
      if $AUC_{g,o,k,n,a,z,mul,max} > AUC_{g-1,o,k,n,a,z,mul,max}$

```
            BestNet = Net_{g,o,k,n,a,z,mul}|AUC_{g,o,k,n,a,z,mul,max}
        else
            BestNet = Net_{g-1,o,k,n,a,z,mul}|AUC_{g-1,o,k,n,a,z,mul,max}
        break
    else
        BestNet = Net_{g,o,k,n,a,z,mul}|AUC_{g,o,k,n,a,z,mul,max}
return BestNet
```

It was concluded that a window length of 31 s attained the best AUC for the second (overlapping on the right and left, and the epoch's label refers to the central 100 sample points) and third (overlapping on the left, and the epoch's label refers to the first 100 sample points) scenarios, while a window length of 19 s was the best for the first scenario. The second scenario attained the best results, while the first produced the lowest AUC. The optimal structure for the A phase classifier (found by the HOSA-C) was composed of 64 and 128 kernels in the first and second convolution layers, respectively, and employed 100 neurons in the first dense layer. The total number of examined networks was 2136, and each network was simulated 10 times. Thus, the total number of evaluated classifiers (using TFCV) was 21360. For the NREM assessment the optimized classifier's structure was composed of 32 and 64 kernels in the first and second convolution layers, respectively, followed by a dense layer with 150 neurons.

The variation of the AUC attained for the different window's length of the tested overlapping scenarios, regarding the A phase classification, is presented in Figure 5.32. It was verified that increasing the window length beyond 31 s was counterproductive, indicating that the added information led to misclassifications. Such occurrence can conceivably be connected to the findings reported by Terzano et al. [373], where it was concluded that the average A phase duration was about 13 s thus, extending the duration further than the 31 s can introduce too much information from the not-A epochs, leading to misclassifications.

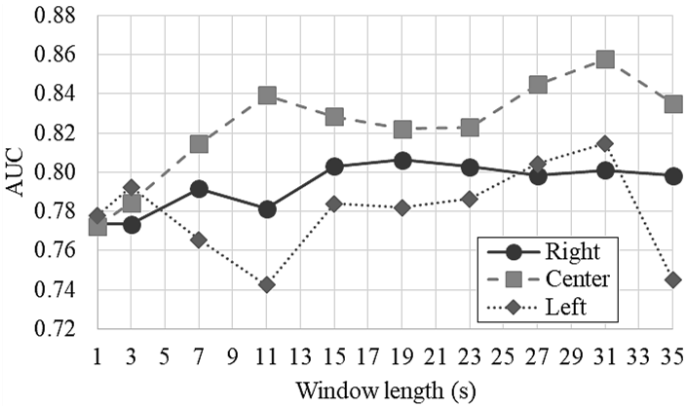


Figure 5.32. Variation of the AUC, according to the window's length of the tested overlapping scenarios, for the A phase classification.

The low performance produced by the first scenario was related to misclassifications of the A phase onset boundary, which occurred when the current epoch under classification (epoch related to the label) was not-A and the sampling points related to a subsequent A phase are present in the segmented window (due to the overlapping), leading the classifier to wrongly classify the current epoch as A. This effect was

diminished in the third scenario, although the reverse effect happened, associated with the A phase offset boundary detection, when the current epoch under classification was A, and the sampling points related to a subsequent not-A epochs are present in the segmented window, thus, leading the classifier to wrongly classify the current epoch as not-A. The second scenario was affected by both onset and offset misclassification although, these were lessened since the algorithm has contextual information, in the window, from both after and before epochs, and the current classification.

Nonetheless, it was verified that the proper offset boundary detection was problematic for all scenarios, particularly for the longer A phases, where the classification oscillated between not-A and A. This effect is in line with the indication presented by Terzano et al. [19], where it was specified that the A phases could exhibit limits that are ambiguous due to inconsistent voltage changes (in the EEG signal). However, this difficulty was lessened by the use of the post-processing procedure (the first part of this procedure was presented in section 5.2). The NREM classifier's output was also used to reduce the misclassifications by reclassifying (in the second part of the post-processing procedure) the epochs classified as A to not-A when the predicted epoch was not-NREM.

The performance of the best methods for A phase and NREM classification was assessed using TFCV, performing 50 repetitions of each iteration to achieve statistically significant results, using the classifiers found by the hyperparameter optimization procedure (HOSA-C). The Acc, Sen, Spe, and AUC for the A phase assessment were 81%, 69%, 84%, and 0.86, respectively. The A phase classification performance is highly relevant as it is similar to the results of the feature based methods, advocating the relevance of the posed method.

The performance of the method for sleep quality examination was assessed by LOOCV (performing 50 repetitions of each iteration to achieve statistically significant results), and using the third CAPSD (15 normal subjects and from the four subjects with sleep-disordered breathing). The attained results for each subject are presented in Table 5.21. The estimated average CAP rate error (given by *abs* ($CAP_1 - CAP_2$) / CAP_2), where *abs* is the absolute value, CAP_1 is the CAP rate predicted by the model, and CAP_2 is the CAP rate estimated from the database labels) was 32%, and the accuracy of the predicted sleep quality was 74%.

It was observed that the model had difficulties performing the assessment for the subjects suffering from sleep-disordered breathing. It was conceptualized that a classifier capable of exploring temporal information can possibly lead to a better performance than the use of the overlapping windows. Therefore, an approach based on LSTM was used in the subsequent examination for the A phase and NREM assessments, using the HOSA-L algorithm for optimization of the network. The block diagram of the developed model based on LSTM is the same as presented in Figure 5.30, without the overlapping windows.

The results for the LSTM optimization were similar to the results attained for the A phase subtype assessment (section 5.5), were it was observed that the use of BLSTM had an increase of less than 1% in the AUC when compared to the LSTM, and the cascade LSTM architecture attained a lower average AUC. Hence, the total number of examined network's architecture was 192. Each network was repeated 10 times, and the total

number of examined classifiers was 1920. The employed algorithm for sleep quality estimation follows the block diagram presented in Figure 5.30, but without the creation of overlapping windows. It was observed that the best network's configuration for the A phase classification was composed of one LSTM layer with 100 hidden units, followed by one dense layer with half of the LSTM's hidden units. The optimal number of time steps was 25.

Table 5.21: Performance of the method employed for sleep quality estimation based on the EEG signal examination using LOOCV, and classification performed by the 1D-CNN. *S* indicated if the predicted sleep quality was correct (C) or wrong (W).

	A phase				NREM				CAP			Sleep quality		
Subject	*Acc (%)*	*Sen (%)*	*Spe (%)*	*AUC*	*Acc (%)*	*Sen (%)*	*Spe (%)*	*AUC*	*Acc (%)*	*Sen (%)*	*Spe (%)*	*R*^ *(%)*	*N* (%)*	*S*
1	82.37	84.61	82.07	0.911	86.11	92.52	66.13	0.927	73.40	70.49	74.98	5.94	12.64	C
2	77.17	75.76	77.32	0.842	79.04	88.79	55.54	0.880	70.82	64.21	73.08	12.01	25.55	W
3	79.87	73.94	80.37	0.851	80.01	83.24	73.22	0.886	73.34	40.14	83.77	-1.68	3.57	C
4	79.41	85.77	78.91	0.900	85.46	87.57	82.63	0.933	77.46	76.08	77.82	22.45	47.77	W
5	82.25	84.93	81.93	0.912	85.57	82.36	94.79	0.926	80.09	77.11	81.87	14.82	31.53	C
6	83.53	81.49	83.86	0.900	80.12	71.98	98.22	0.937	75.29	49.47	91.34	-7.12	15.15	C
7	80.45	93.37	79.17	0.938	80.91	75.42	92.76	0.922	72.93	47.87	83.42	3.50	7.45	C
8	76.58	83.62	75.72	0.872	81.64	86.83	70.80	0.896	72.41	64.33	76.00	7.54	16.04	C
9	84.16	83.60	84.20	0.910	70.77	62.87	86.52	0.884	81.18	37.26	92.29	2.16	4.60	C
10	79.44	57.92	82.01	0.818	77.72	74.70	83.84	0.864	77.37	24.44	90.11	-4.72	10.04	C
11	80.62	72.01	81.56	0.858	68.94	68.94	68.93	0.769	80.20	48.96	91.76	-11.86	25.23	C
13	84.08	82.75	84.27	0.903	85.33	96.74	67.36	0.955	86.41	82.12	88.53	-5.47	11.64	C
14	85.77	76.31	87.06	0.889	76.14	96.23	35.76	0.903	73.05	66.55	75.33	2.35	5.00	C
15	83.56	87.82	82.94	0.924	85.23	86.48	81.31	0.920	78.65	74.28	80.94	8.60	18.30	C
16	84.44	73.36	86.24	0.872	74.71	97.03	17.20	0.842	81.10	70.71	88.22	-18.04	38.38	C
17	77.57	60.76	79.08	0.782	74.63	83.08	63.25	0.823	76.14	22.09	94.02	-26.28	55.91	C
18	78.07	61.17	84.09	0.824	73.71	73.54	74.47	0.822	44.42	10.07	99.59	-67.26	143.11	W
19	72.37	58.75	78.74	0.758	85.63	90.96	72.43	0.915	61.91	49.21	76.12	-22.98	48.89	W
20	74.57	55.71	83.52	0.789	53.62	48.47	78.53	0.707	43.90	25.59	87.96	-38.97	82.91	W
Mean	80.33	75.46	81.74	0.866	78.17	81.46	71.77	0.880	72.63	52.68	84.59	-	31.77	-
SD#	3.55	11.22	2.94	0.051	7.77	12.32	19.15	0.062	10.98	20.92	7.49	-	33.29	-

^ Difference between the CAP rate predicted by the model and the CAP rate estimated from the database labels

* Result from the division of the absolute value of *R* (given by the difference between the CAP rate predicted by the model and the CAP rate estimated from the database labels) by the CAP rate estimated from the database

Standard deviation

For the NREM classification the best network's architecture was composed of one LSTM layer with 300 hidden units, followed by one dense layer with 150 hidden units.

The performance of the classifiers was assessed using TFCV (performing 50 repetitions of each iteration to achieve statistically significant results). The Acc, Sen, Spe, and AUC for the A phase assessment were 84%, 73%, 85%, and 0.86, respectively. The reached performance was the best attained in this work for the methods without an explicit feature creation procedure, suggesting that the LSTM is more suitable for this analysis than the 1D-CNN, although a significant improvement was attained using the overlapping windows.

The learning curves of both 1D-CNN and LSTM based classifiers for the A phase classification were assessed to check if the inclusion of more data could improve the performance. The attained curves are presented in Figure 5.33, and it was possible to conclude that the LSTM could possibly benefit more, from additional data, than the 1D-CNN.

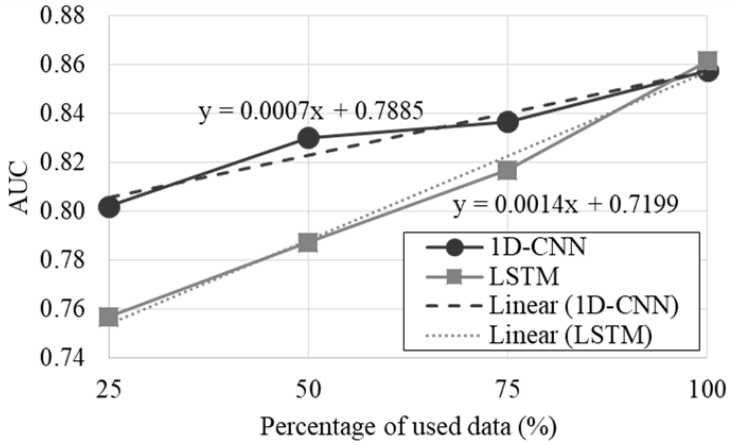


Figure 5.33. Learning curves for the best classifiers without an explicit feature creation procedure, presenting the linear tendency line and its respective equation.

The third CAPSD was employed for the performance assessment of the LSTM based classifiers for sleep quality assessment. LOOCV was used, and each simulation was repeated 50 times to achieve statistically significant results. The attained results for each subject are presented in Table 5.22. The assessed average CAP rate error was 17%, and the accuracy of the sleep quality prediction was 79%. The performance of the A phase estimation was lower when using the LSTM but for the NREM estimation performance was superior, leading to a more precise estimation of the CAP rate which, in turn, improved the sleep quality prediction's accuracy.

The last test was performed considering the hypotheses that the use of features can improve the classifier's capability to recognize the relevant patterns for the most challenging subjects (specially the patients suffering from sleep-disordered breathing). The optimization procedure was carried out using TFCV, evaluating the recordings from the third CAPSD, and the employed model is the same used for the LSTM without an explicit feature creation procedure but with a feature creation step before the classification. The examined features were the same as presented in section 5.5, and the classifiers' architectures was the same as the models without an explicit feature procedure for A phase and NREM assessment. The performance was then assessed by TFCV, performing 50 repetitions of each iteration to achieve statistically significant results, and the relevance of the features for the A phase classification was determined by the mRMR

algorithm. The order of the features, identifiable by the identification number presented in Table 5.15, was: 11; 10; 14; 12; 20; 13; 16; 19; 18; 15; 17; 3; 7; 2; 9; 8; 4; 1; 6; 5. Figure 5.34 presents the variation of the A phase classification performance metrics according to the number of employed features. It was observed that the best AUC was attained using the 12 most relevant features, reaching an Acc, Sen, Spe, and AUC of 86%, 81%, 86%, and 0.91, respectively.

Table 5.22: Performance of the method employed for sleep quality estimation based on the EEG signal examination using LOOCV, and classification performed by the LSTM without an explicit feature creation procedure. *S* indicated if the predicted sleep quality was correct (C) or wrong (W).

	A phase				NREM				CAP			Sleep quality		
Subject	*Acc (%)*	*Sen (%)*	*Spe (%)*	*AUC*	*Acc (%)*	*Sen (%)*	*Spe (%)*	*AUC*	*Acc (%)*	*Sen (%)*	*Spe (%)*	*R^ (%)*	*N* (%)*	*S*
1	79.96	75.25	80.60	0.852	85.79	92.52	64.79	0.924	73.61	79.65	70.31	13.60	28.94	C
2	83.00	73.15	84.44	0.863	88.22	94.31	78.61	0.957	83.53	84.52	83.04	5.54	11.79	W
3	83.17	72.39	84.74	0.862	88.52	94.11	79.71	0.957	83.16	84.55	82.48	6.60	14.04	C
4	82.41	72.01	83.93	0.853	88.24	93.73	79.58	0.953	82.57	81.09	83.31	4.14	8.81	W
5	85.40	61.16	89.08	0.856	87.16	88.06	84.30	0.927	81.73	67.49	90.60	-5.60	11.91	C
6	83.07	66.37	85.72	0.830	87.81	91.27	80.11	0.941	80.61	77.67	82.44	3.25	6.91	C
7	81.68	84.94	81.35	0.897	89.57	92.76	82.69	0.950	76.35	75.98	76.50	13.39	28.49	C
8	82.37	72.00	83.89	0.854	88.30	93.13	80.69	0.953	82.73	82.91	82.64	6.48	13.79	C
9	91.14	61.44	93.30	0.877	91.31	91.01	91.91	0.949	83.84	48.36	92.82	-5.87	12.49	C
10	82.41	73.19	83.76	0.859	88.56	93.99	79.99	0.958	82.43	83.81	81.75	7.09	15.09	W
11	84.03	55.87	87.10	0.812	80.02	76.26	85.33	0.881	81.90	59.89	90.05	-0.06	0.13	C
13	83.91	78.84	84.65	0.888	89.56	94.00	82.56	0.960	84.96	87.57	83.67	8.08	17.19	C
14	80.39	73.14	81.38	0.843	82.85	93.42	61.59	0.899	71.61	80.68	68.42	20.13	42.83	C
15	85.64	74.86	87.21	0.885	90.42	96.12	72.57	0.958	82.62	81.89	83.01	4.10	8.72	C
16	77.31	61.99	79.79	0.785	81.46	95.80	44.47	0.905	73.16	72.59	73.55	-2.98	6.34	C
17	78.81	52.75	81.16	0.749	72.46	81.56	60.20	0.825	73.07	45.64	82.14	-4.49	9.55	C
18	77.76	52.90	86.63	0.775	73.54	71.19	83.55	0.841	60.86	41.45	92.03	-29.35	62.45	W
19	62.02	51.49	66.94	0.628	78.41	86.46	58.43	0.839	67.94	71.68	63.74	0.71	1.51	C
20	69.23	57.01	75.04	0.703	79.60	86.38	46.75	0.763	69.43	70.16	67.65	-12.07	25.68	C
Mean	80.72	66.88	83.19	0.825	84.83	89.79	73.57	0.913	77.69	72.51	80.53	-	17.19	-
SD#	6.11	9.57	5.40	0.068	5.54	6.62	13.14	0.056	6.64	13.63	8.22	-	14.71	-

^ Difference between the CAP rate predicted by the model and the CAP rate estimated from the database labels

* Result from the division of the absolute value of *R* (given by the difference between the CAP rate predicted by the model and the CAP rate estimated from the database labels) by the CAP rate estimated from the database

Standard deviation

The sleep quality performance assessment was carried out using LOOCV (performing 50 repetitions of each iteration to achieve statistically significant results), and the attained results are presented in Table 5.23. By examining the table, it is possible to observe that

the average CAP rate error was 22%, and the accuracy of the predicted sleep quality was 90%. By comparing with the results of the method without an explicit feature creation procedure it is possible to observe a significant improvement in the A phase classification performance, with a relevant increment in the NREM classification performance. These improvements allowed the model to attain a significantly greater accuracy for the sleep quality estimation. It is also relevant to notice that the A phase estimation's performance is the best attained in this work, and is considerably better than all state of the art works, supporting the relevance of the proposed methodology.

Table 5.23: Performance of the method employed for sleep quality estimation based on the EEG signal examination using LOOCV, and classification performed by the LSTM evaluating selected features. *S* indicated if the predicted sleep quality was correct (C) or wrong (W).

	A phase				NREM				CAP			Sleep quality		
Subject	*Acc (%)*	*Sen (%)*	*Spe (%)*	*AUC*	*Acc (%)*	*Sen (%)*	*Spe (%)*	*AUC*	*Acc (%)*	*Sen (%)*	*Spe (%)*	*R^ (%)*	*N* (%)*	*S*
1	85.22	79.66	85.97	0.907	89.84	88.64	93.59	0.956	79.83	69.60	85.41	2.93	6.23	C
2	85.67	81.82	86.24	0.906	93.73	94.75	92.11	0.979	84.75	81.87	86.17	5.30	11.28	W
3	85.17	61.87	87.14	0.823	82.36	77.93	91.67	0.925	78.34	40.26	90.31	-4.56	9.70	C
4	84.90	82.67	85.22	0.898	91.71	94.55	87.22	0.975	84.47	85.50	83.96	8.29	17.64	W
5	88.07	76.81	89.77	0.923	89.24	86.89	96.70	0.964	85.88	72.02	94.52	-4.14	8.81	C
6	83.51	79.76	84.11	0.886	89.93	88.53	93.05	0.956	82.25	76.57	85.78	4.87	10.36	C
7	84.01	91.77	83.24	0.946	95.53	96.17	94.16	0.984	78.99	81.96	77.74	15.82	33.66	C
8	73.23	90.40	71.12	0.898	85.21	87.65	80.12	0.924	69.36	80.78	64.21	29.84	63.49	C
9	91.24	78.08	92.20	0.939	93.20	93.34	92.92	0.967	84.68	54.68	92.28	-3.67	7.81	C
10	83.80	48.49	88.01	0.816	95.48	94.73	97.03	0.986	75.91	50.89	81.93	8.94	19.02	C
11	83.58	66.57	85.43	0.851	89.80	86.86	93.95	0.952	77.34	66.64	81.30	13.00	27.66	C
13	85.36	84.20	85.53	0.919	90.71	93.12	86.91	0.970	82.83	78.04	85.20	3.52	7.49	C
14	84.87	79.22	85.65	0.897	89.67	96.52	75.86	0.951	75.15	78.88	73.83	16.09	34.23	C
15	86.78	87.70	86.65	0.938	95.05	96.43	90.69	0.982	84.71	84.50	84.83	6.39	13.60	C
16	78.84	73.82	79.65	0.838	82.34	94.42	51.12	0.923	80.73	79.27	81.74	-3.93	8.36	C
17	87.32	55.03	90.23	0.842	82.53	77.42	89.42	0.906	78.25	32.01	93.56	-17.54	37.32	C
18	83.58	70.06	88.40	0.873	73.59	68.82	93.89	0.921	67.52	50.83	94.35	-17.61	37.47	C
19	71.87	83.77	66.30	0.842	77.49	76.48	80.01	0.877	72.72	76.24	68.78	17.29	36.79	C
20	69.26	82.39	63.03	0.821	80.97	83.21	70.14	0.858	75.60	83.17	57.38	10.94	23.28	C
Mean	82.96	76.53	83.36	0.882	87.81	88.24	86.87	0.945	78.91	69.67	82.28	-	21.80	-
SD#	5.54	11.24	7.75	0.042	6.18	7.88	11.04	0.036	5.17	15.63	9.91	-	14.96	-

^ Difference between the CAP rate predicted by the model and the CAP rate estimated from the database labels

* Result from the division of the absolute value of *R* (given by the difference between the CAP rate predicted by the model and the CAP rate estimated from the database labels) by the CAP rate estimated from the database

Standard deviation

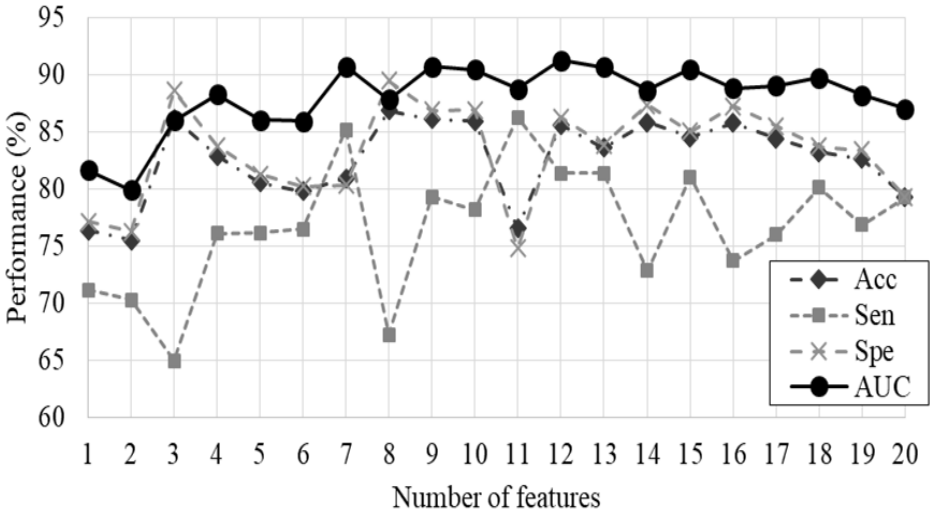


Figure 5.34. Variation of the A phase classification performance metrics according to the number of employed features which were fed to the LSTM.

The boxplots of the CAP rate error produced by the examined classifiers is presented in Figure 5.35. By examining the figure, it is possible to observe that the model based on the LSTM without an explicit feature creation procedure has the lowest variation, while the opposite occurred for the model based on the 1D-CNN. Nevertheless, it was observed that the feature based model (LSTM fed with features) attained a better performance for the sleep quality prediction. These results are likely to be related to the performance of the algorithms for the examination of the subjects suffering from sleep-disordered breathing, where the feature based method surpassed the methods without an explicit feature creation procedure, corroborating the hypothesis that the use of features can improve the performance of the classifier for the most challenging subjects. Figure 5.36 presents the normalized CAP rate error for all studied subjects, where it is possible to conclude that subject 18 was the most challenging for the classifiers without an explicit feature creation procedure, while the examination of subject 8 produced the highest error for the feature based model. These results can possibly be due to the relatively small number of subjects with sleep-disordered breathing available in the database (four subjects), which may not provide enough data for the method without an explicit feature creation procedure to recognize all the relevant patterns.

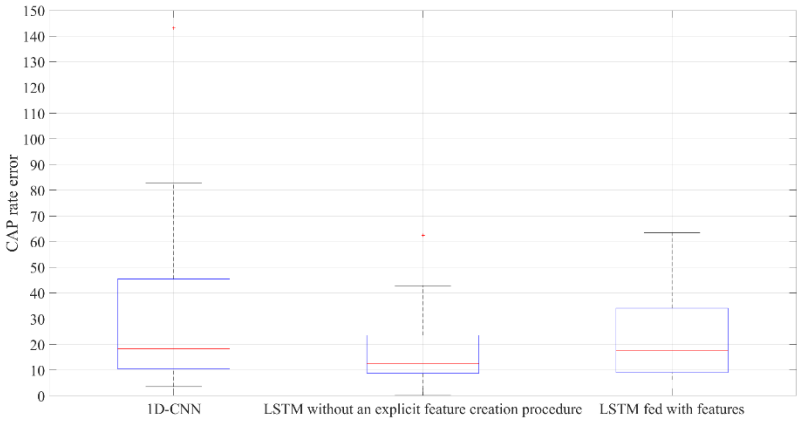


Figure 5.35. Boxplots of the estimated CAP rate error for each examined classifier.

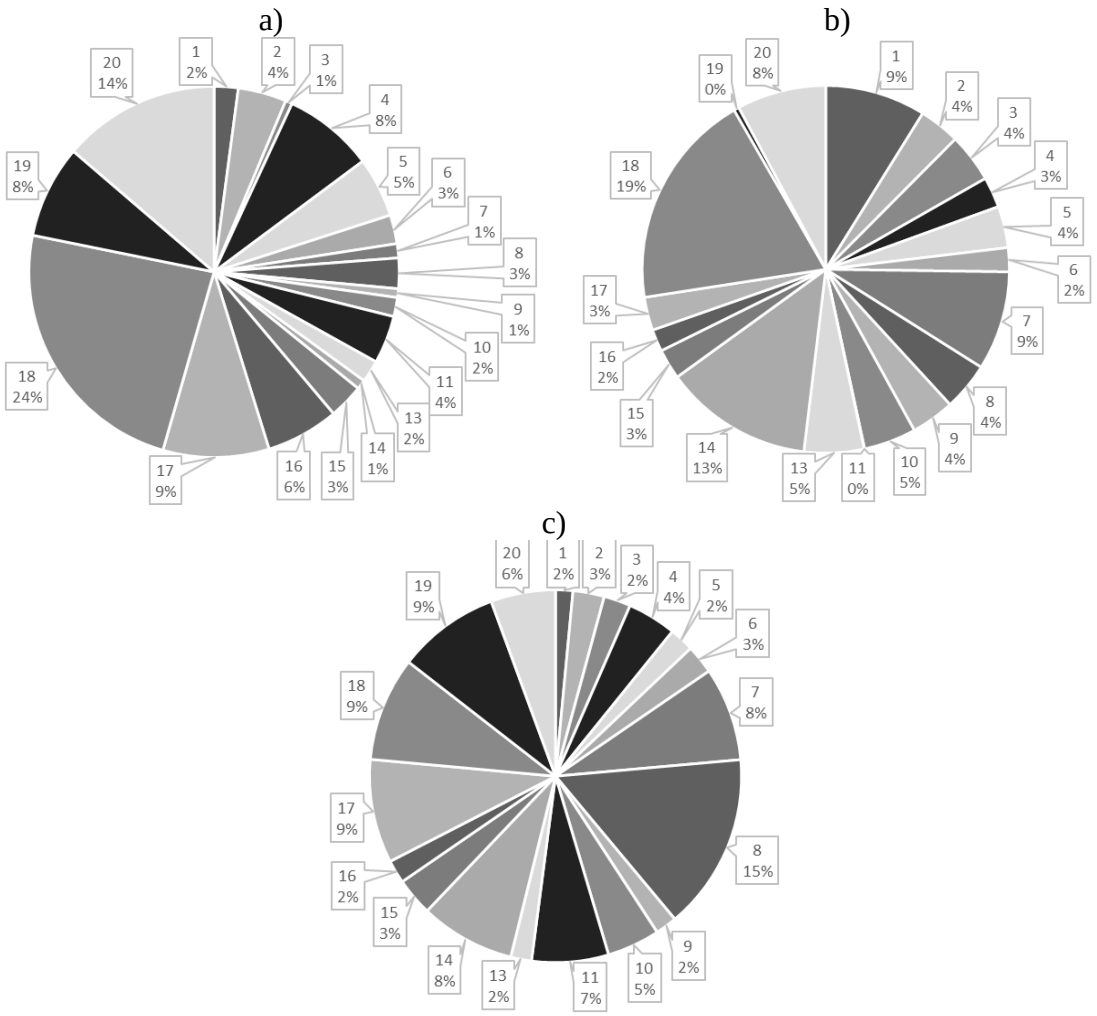


Figure 5.36. Normalized CAP rate error produced by a) the model based on the 1D-CNN, the model based on the LSTM without an explicit feature creation procedure, and c) the model based on the LSTM fed with features.

5.7. Discussion of the results

The comparative analysis between the results attained by the developed algorithms for the classification of the CAP A phases, and the results of the methods proposed in the state of the art is presented in Table 5.24. By analyzing the table, it is possible to verify that Barcaro et al. [324] achieved the highest Acc from the group of works that did not use a machine learning approach. Regarding the employment of LDA, Mariani et al. [329] reported the best Acc and Spe, but the lowest Sen. Thus, the results are unbalanced, and are possibly not suitable for clinical applications, while the results achieved in this work are balanced and have a higher Sen. Machado et al. [331] achieved a lower performance than the best results of this work.

Examining the results reported by Mariani et al. [328] and [330] for the SVM, it is possible to attest that a lower Acc and Spe were attained in this work, but the Sen is higher. These results could be related to the usual unbalanced data in a normal subject, with significant more not-A than A epochs, indicating that an increase in the Spe has a

larger impact in the Acc than a growth in Sen. Likewise, most of the state of the art works examined a lower number of subjects than this work, and have only examined subjects free of sleep related disorders, which is likely to lead to a better performance of the classifier. The same conclusions can be reached for the FFNN and AdaBoost.

Regarding the examination of the four classifiers evaluated for the methods without an explicit feature creation procedure, it was observed that LSTM attained the most balanced performance. A further examination was performed using overlapping windows to feed the 1D-CNN, leading to the best performance of all proposed methods without an explicit feature creation procedure. Nonetheless, a superior performance was reached when features were fed to the LSTM, suggesting that some of the relevant patterns were highlighted by these features. According to Rosa et al. [53], the specialist agreement ranges from 69% to 78%, examining the same EEG results. Therefore, the attained results with the best classifiers are considerably above the upper bound of the specialist agreement. These results advocate the feasibility of the developed algorithms for clinical diagnosis, and are indicative of the possibility for the developed algorithms to become useful tools for medical diagnosis.

Table 5.24: Comparative analysis between the state of the art methods for the A phase classification and the results attained in this work.

Work	*Classifier*	*Population*	*Acc (%)*	*Sen (%)*	*Spe (%)*	*CO (%)*
[331]	LDA	30*	68	-	-	-
[28]	Tuned threshold	8	69	59	71	66
[331]	kNN	30*	70	-	-	-
[331]	SVM	30*	71	-	-	-
[28]	Tuned threshold	8	72	52	76	67
[28]	Tuned threshold	8	72	70	72	71
[323]	Tuned threshold	10	77	84	90	84
[330]	AdaBoost	8	79	69	79	76
[326]	Tuned threshold	6*	81	76	81	79
[330]	SVM	8	82	70	84	79
[330]	FFNN	8	82	73	82	79
[79]	FFNN	4	82	76	83	80
[324]	Tuned threshold	10	84	-	-	-
[328]	SVM	4	84	74	86	81
[330]	LDA	8	85	73	87	82
[329]	LDA	16*	86	67	90	81
This work – feature based methods with features selected by SFS	LDA	13*	75	78	74	76
	LR	13*	76	80	75	77
	CT	13*	70	58	73	67
	ET	13*	70	64	71	68
	SVM	13*	72	80	70	74
	FFNN	**13***	**79**	**76**	**80**	**78**
	CFNN	13*	76	77	76	76
	kMC	13*	78	67	81	75
	kNN	13*	72	70	72	71
	SOM	13*	67	79	66	71
This work – methods without an explicit feature creation procedure	DSAE	13*	67	55	69	64
	LSTM	15*	76	75	77	76
	GRU	**15***	**81**	**66**	**83**	**77**
	1D-CNN	15*	74	71	75	73
This work – method for sleep quality examination	1D-CNN	19*	80	76	82	79
	LSTM without an explicit feature creation procedure	19*	81	67	83	77
	LSTM fed with feature	**19***	**83**	**77**	**84**	**81**

* The CAP Sleep Database – PhysioNet [19]

Regarding the A phase subtype detection, the comparative analysis between the results of the state of the art methods and the attained results is presented in Table 5.25. By examining the table, it is possible to conclude that the state of the art works which employed tuned thresholds for classification presented a poor classification performance for one of the A phase subtypes, which is not suitable for medical examination. On the other hand, the work presented by Mendez et al. [327] reported a performance which is similar to the results attained in this work, although only five subjects were examined. By comparing the methods without an explicit feature creation procedure with the feature based methods, it is possible to conclude that the employment of features allowed the classifier to achieve results which are more balanced (similar Sen and Spe), and are, therefore, more relevant for clinical diagnosis.

Table 5.25: Comparative analysis between the state of the art methods for the A phase subtype classification and the results attained in this work.

Work	*Classifier*	*Population*	*CL*[+]	*Acc (%)*	*Sen (%)*	*Spe (%)*	*CO (%)*
[322]	Tuned threshold	10	A1; A2; A3	95; 85; 60	-	-	-
[325]	Tuned threshold	6	A1; A2; A3	-	80; 77; 68	83; 73, 74	-
[327]	kNN	5	A1; A2; A3	90; 43; 80	-	-	-
This work – methods without an explicit feature creation procedure	LSTM	19[*]	A1; A2; A3	88; 86; 80	62; 53; 58	90; 87; 81	80; 75; 73
This work – feature based methods with features selected by SFS	**LSTM**	**19**[*]	**A1; A2; A3**	**83; 83; 83**	**83; 74; 71**	**83; 83; 84**	**83; 80; 79**

* The CAP Sleep Database – PhysioNet [19]
[+] Classified A phase subtype

Only Karimzadeh et al. [336] have performed the CAP cycle detection, and reported the results using the same performance metrics employed in this work (the other works reported the performance suing proposed metrics or employing correlation analysis). The comparative analysis with the developed algorithms for the classification of the CAP cycle detection is presented in Table 5.26. The reported results of the best developed methods are similar to the results Karimzadeh et al. [336]. However, the full EEG signal was used in this work for the performance assessment (instead of evaluating a selected part of the signal, as it was done by Karimzadeh et al. [336]), and more subjects were examined. The comparative analysis between the attained performance and the state of the art results for the sleep quality estimation is presented in section 8.2.

5.8. Key remarks

Several methods for CAP analysis, based on the examination of the EEG signal, were proposed and evaluated in this chapter. The first approach was the evaluation of the state of art to identify the most successful proposals and ascertain the gaps in the literature. The most relevant features were tested, and sets of new features were proposed for multiple classifiers. The second analysis was the development of methods without an explicit feature extraction process that are capable of learning high dimensional and abstract patterns from the input signal without requiring a feature selection method.

Both feature based and methods without an explicit feature extraction process attained an accuracy that is the same as the upper limit of the specialist agreement, advocating the

relevance of the developed models. A sleep model was subsequently proposed to test the hypothesis that the CAP cycles have a temporal dependency, which could possibly be identified by a stochastic model if the sleep process could be described with an approximation to continuous traces. It was identified that this model was feasible for CAP analysis and could provide a framework for future analysis in the sleep microstructure, taking into consideration that it is a new way of interpreting sleep. The subsequent examination was the A phase subtype assessment, which allowed to perform the characterization of CAP. The last examination was the proposal of methods for sleep quality assessment based on the CAP rate estimation. It was observed that the feature based methods can attain a better performance possibly because they can better highlight the relevant patterns for the classification.

Table 5.26: Comparative analysis between the state of the art methods for the CAP cycles classification and the results attained in this work.

Work	*Classifier*	*Population*	*Acc (%)*	*Sen (%)*	*Spe (%)*	*CO (%)*
[336]	kNN	8*[+]	77	-	-	
[336]	LDA	8*[+]	79	-	-	
[336]	SVM	8*[+]	79	-	-	
This work – feature based methods with features selected by SFS	LDA	13*	75	-	-	
	LR	13*	78	-	-	
	CT	13*	64	-	-	
	ET	13*	70	-	-	
	SVM	13*	75	-	-	
	FFNN	**13***	**79**	-	-	
	CFNN	13*	77	-	-	
	kMC	13*	78	-	-	
	kNN	13*	70	-	-	
	SOM	13*	68	-	-	
This work – methods without an explicit feature creation procedure	DSAE	13*	62	67	59	63
	LSTM	**15***	**76**	**71**	**84**	**77**
	GRU	15*	72	60	88	73
	1D-CNN	15*	73	66	83	74
This work – proposed sleep model	HMM	15*	72	66	75	71
This work – method for sleep quality examination	1D-CNN	19*	73	53	85	70
	LSTM without an explicit feature creation procedure	19*	78	73	81	77
	LSTM fed with feature	**19***	**79**	**70**	**82**	**77**

* The CAP Sleep Database – PhysioNet [19]

[+] Average of the reported results

The main limitation of the performed analysis was the relatively small number of subjects with sleep-disordered breathing, which may not have allowed the methods without a specific feature creation procedure to properly identify all the relevant patterns associated with the intended classification. Another limitation was the absence of an extensive validation (by comparing with an exhaustive grid search method) of the proposed heuristic oriented method for the classifiers optimization, and the absence of comparative analysis with other heuristic based approaches proposed in the state of the art. The last limitation was that the evaluation only considered one sleep related disorder. Hence, it is not possible to know if other disorders could significantly change the results.

6. Indirect measurement of CAP for sleep quality estimation

This chapter presents the development of methods to perform the indirect measurement of CAP from the ECG signal. The predicted metrics (based on CAP characteristics) allow to estimate the quality of sleep according to the first approach for sleep quality examination theorized in this work from the state of the art analysis. By seeing CAP a marker of sleep instability, the proposed approaches consider an extended concept for CAP, where the oscillation between the stable and unstable sleep are assessed by evaluating the indirect effect of CAP in other physiological signals, in this case the ECG signal [41].

For this purpose, a CPC technique was employed to perform a minute-by-minute estimation of the CAP epochs, a concept developed in this work that considers an optimal threshold to define each minute as either CAP or non-CAP. The NREM minutes were also estimated from the CPC, and this information was combined with the CAP epochs' classification to predict the CAP rate. The age-related CAP rate percentages in healthy subjects was then consider as a threshold to assess the sleep quality. A second sleep quality prediction method was also proposed by evaluating the average of the spectrographic CPC measure. A tool for time series analysis was developed at the end of the chapter to study the connection between the EDR and the N-N series with the goal of estimating the CAP epochs.

6.1. CAP estimation from ECG

6.1.1. Viability assessment

For healthy subjects, the HRV spectrum can be analyzed to assess the sleep stage by considering the ratio of the power on the 0.05 to 0.15 Hz band to the 0.15 to 0.4 Hz band. A large increase in the ratio is related to REM periods, while a characteristic pattern is associated with the presence of NREM sleep. A significant decrease in the ratio indicates the occurrence of more synchronized sleep (deep sleep) [374]. Unstable periods of sleep, associated with the existence of CAP, can be detected using CPC, between EDR and HRV, examining the power on the LF coupling. On the other hand, wake or REM sleep periods can be assessed by evaluating the CPC coupling in the VLF while periods of stable sleep (absence of CAP) can be identified by the power in the HF coupling [41] [52].

Therefore, the relation between the HRV and EDR was analyzed to estimate both CAP and NREM periods, considering minute-by-minute analysis (epochs with a duration of 60 s were employed to follow Thomas et al. [41] indication that his duration is suitable for scoring CAP), and this information was then employed to infer the CAP rate, for sleep quality estimation. The age-related CAP rate percentages in healthy subjects [10] were considered as a reference to define the quality of sleep. This analysis was validated by the fact that CAP rate is characterized by a low night-to-night intra-individual variability, despite the complex changes that this metric undertakes through the life period of a person [10]. Thus, the concept of CAP was interpreted in a broader context, and the CAP cycles occurrence was indirectly estimated.

Sleep stages, associated with the sleep macrostructure, are labeled every 30 s, thus the one minute epoch's label was defined as NREM if one or both of the 30 s annotations

(which compose the minute long epoch) correspond to NREM sleep. REM and Wake periods were labeled as non-NREM to allow binary classification (either NREM or non-NREM). The CAP cycle annotations were generated using the Terzano reference atlas [19] for one second epochs as presented in section 4.1.

Although a minute-by-minute classification was performed in this work, in healthy adults, the average CAP sequence lasts 2 min and 33 s and the average CAP cycle duration is 26.9 ± 4.1 s [38]. Hence, a CAP sequence is likely to occur over several one minute epochs. This is particularly problematic for the first and last CAP cycle as an ambiguous situation can be created if the CAP cycle occurred over two consecutive one minute epochs. For example, a CAP cycle which lasts 50 s and has 5 s in a one minute epoch and 45 s in the subsequent one minute epoch (it is important to bear in mind that contrarily to the macrostructure, where an epoch is score every 30 s, the CAP cycles can range from 4 s to 120 s, performing the scoring at every second). Another difficulty is associated with the isolated CAP cycles with a short duration or the short CAP sequences with few and short CAP cycles. These may represent short periods of instable sleep without leading to noticeable alterations in the ECG signal.

As an example, three isolated CAP cycle composed of phases lasting 2 s may be too short to produce a noticeable alteration in other physiological signals. In these situations, it is questionable if the minute should be considered CAP or non-CAP, denoting a minute of instable or stable sleep, respectively. An approach to address this issue was proposed in this work by formulating the CAP epoch concept, indicating that a period is related to instable sleep (CAP) only when more than a defined percentage of the 60 s of data (one minute epoch) was scored as a CAP cycle. Therefore, a threshold based methodology defined the one minute epoch as either unstable (CAP) or stable sleep (non-CAP). In the first test (viability assessment), an epoch was labeled as CAP if more than 30% of the minute corresponds to a CAP period (filtering the CAP related events that occur in the epoch but are shorted than 18 s). Otherwise, the minute was labeled as non-CAP.

The data was then segmented into one minute epochs, and a seven minutes window was employed, where the central minute was the one corresponding to the epoch label, and the first and last three minutes overlapped with the other windows. ECG records the electrical activity produced by the heart on the body surface. The characteristics of the measured waves are dependent upon the relative speed and direction of the activation wave front, and the amount of tissue that is activated at one time [375]. Several wave forms can be measured by ECG. Conversely, the QRS complex (large ventricular waveform) is particularly relevant since it is the simplest portion of the ECG tracing which allows to reliably estimate the heart rate. The QRS detection was implemented using an adaptation of the method developed by Pan and Tompkins [302], since it is capable of correctly detecting the QRS complexes in signals significantly contaminated with noise.

The ECG signal was band-pass filtered in the band with the maximum desired QRS energy (5 to 15 Hz band) with a filter composed of a combination of high-pass and low-pass filters. This approach allows the reduction of: the baseline wander; the influence of muscle noise; the T-wave interference. Afterwards, the QRS complex slope was assessed using a derivative filter, implementing an approximation of the signal derivative, and the

signal was squared to perform a nonlinear amplification that emphasizes the higher frequencies. Information from the waveform was attained using a Moving-Window Integration (MWI). However, if the window is too narrow, the QRS complex can be poorly detected, and if the window is too wide, other wave forms can merge with the QRS complex. In both cases, additional noise can be introduced in the signal. It was empirically verified, for the employed sampling rate, that a window with a length of 30 samples provides the best compromise [302].

The rising edge of the MWI marks the beginning of a QRS complex thus, a fiducial mark was created to identify the temporal location of the QRS complex. The maximum slope of the MWI corresponds to the R-wave peak, and the duration of the rising edge indicates the QRS complex width. A refractory period (0.2 s after a QRS detection) was employed to avoid the detection of the same QRS complex multiple times. The next QRS complex should occur within a period of 0.36 s from the last detection. Otherwise, the wave was classified as a T-wave if the mean slope of the waveform was lower than half of the previous QRS complex [302].

The band-pass filtered signal and the QRS complex detection algorithm output were compared to produce a time series with the amplitude and time occurrence of the R-waves. This signal was then used as a reference to determine the interbeat intervals, and all abnormal R-peaks (peaks with an amplitude higher or lower than two times the standard deviation from a running average) were rectified, considering the average of the next and previous peaks as a reference, ensuring the validity of data, producing the N-N series.

A modulation of the QRS morphology occurs during the respiration cycle due to changes in the transthoracic impedance, as the lungs fill and empty, and the movement of the ECG electrodes in relation to the heart, altering the heart-to-electrode distance. EDR is based on the analysis of this modulation [376], and an adaptation of the algorithm developed by Arunachalam and Brown [377] was employed to produce this signal. The respiratory amplitude modulation factor a was computed for all elements of the N-N series, considering the current n^{th} value of the R-peak amplitude, R, and a running average of the current and all previous R-peak amplitudes, S, by [377]

$$a_n = \frac{R_n}{S_n} \tag{6.1}$$

The EDR series was produced by running an interpolation over a. An example of a 30 s analysis of the EDR and squared QRS signals is presented in Figure 6.1.

The relation between the variability of respiratory volume and the heart rate can be assessed using a CPC technique. Cross-spectral coherence was selected to perform this analysis, since it can be implemented using only data from a single-lead ECG [41]. It consists of computation of the coherence and cross spectral power of the input (N-N series and EDR) for successive overlapping windows in the 0 to 0.5 Hz band. Therefore, this technique allows to describe the signals relationship, in the frequency domain, based on their second order statistics. The cross-spectrum between the discrete Fourier transform components, m, of N-N series, N, and EDR, E, is given by [378]

$$\Gamma_m(N,E) = A_{N,m}A_{E,m}e^{j(\Phi_{E,m}-\Phi_{N,m})} \tag{6.2}$$

where Φ and A are the phase and amplitude of the Fourier components, respectively. The regularity of the phase difference between the two signals was determined by the magnitude squared coherence [378]

$$\Delta_m(N,E) = \frac{\Gamma_m(N,E)^2}{\left(A_{N,m}e^{j\Phi_{N,m}}\right)^2\left(A_{E,m}e^{j\Phi_{E,m}}\right)^2} \quad 6.3$$

The Welch's averaged periodogram method was applied, since the number of samples in each window was finite, to obtain the estimates of the cross-correlation matrices [379].

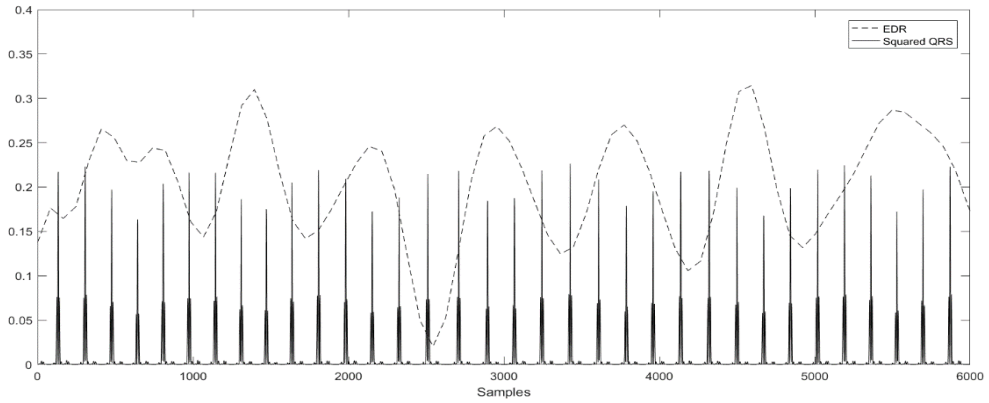


Figure 6.1. Example of the squared EDR and QRS signals [80].

The quantitative degree of CPC was assessed by [41]

$$\beta(f_m) = \Gamma_m(N,E)^2\Delta_m(N,E) \quad 6.4$$

considering the power of the heart rate and the variability of respiratory volume (evaluated by the cross-spectral power), and the consistency with which these signals track each other (assessed by the coherence) at each frequency. However, the used method requires that the heart rate is at least two times higher than the respiration rate in order to meet the Nyquist frequency requirements for a consistent recognition of the respiration cycle, and avoid aliasing [80]. Therefore, a spectrographic measure of the CPC was produced, and it was fed to the classifiers to estimate the CAP and NREM periods. An example of a CPC spectrographic measure, with movement noise and periods of stable and unstable sleep is presented in Figure 6.2.

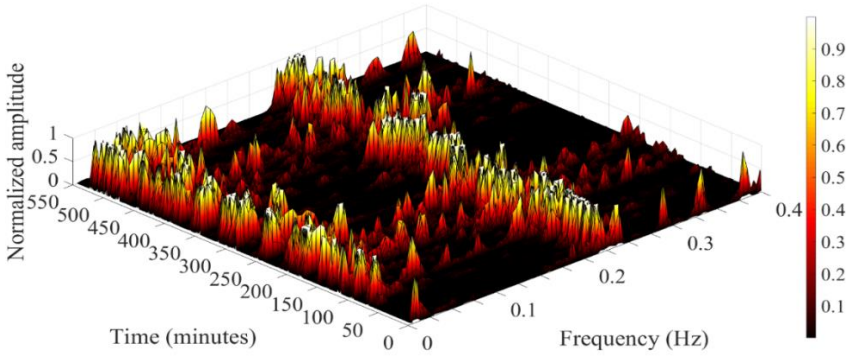


Figure 6.2. Example of a CPC spectrographic measure, with movement noise and periods of stable and unstable sleep [299].

Ten classifiers implementing binary classification were examined in the first test with the purpose of covering multiple solutions (including supervised and unsupervised learning methods), and identify the most suitable for the intended classification. Eight classifiers employ supervised learning: LR; LDA; QDA; CART; TreeBagger (bagging CART); kNN; SVM; FFNN. AdaBoost was used to produce an ensemble of weak learners generated by kMC, forming an unsupervised learning classifier. The last classifier was a DSAE (composed of two stacked autoencoders, followed by an output layer to perform the classification) that was trained using unsupervised learning to individually train the HLs, and then supervised learning was employed, at the end of the training, to fine tune the weights (process described in section 5.3).

Cost-sensitive learning was used to minimize the effect of the data unbalance in all supervised learning classifiers, while oversampling (of the minority class) was applied to the training dataset for the unsupervised learning methods. Cost-sensitive learning was also used during the fine tune of the DSAE weights, performed using supervised learning at the end of the training of the DSAE, considering the distribution of the training dataset's labels before oversampling. The block diagram of the developed model is presented in Figure 6.3.

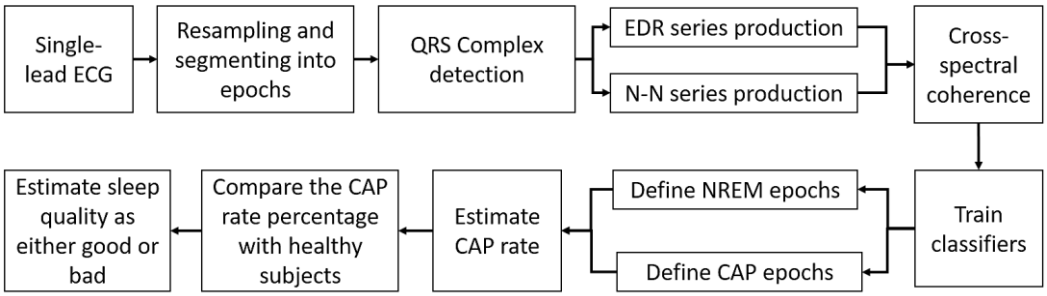


Figure 6.3. Block diagram of the implemented model for sleep quality estimation with cardiopulmonary coupling [81].

Hyperparameter optimization and performance examination was carried out using the recordings from the fourth CAPSD (nine normal subjects and four subjects with sleep-disordered breathing) and LOOCV (performing 50 repetitions of each iteration to achieve statistically significant results). The flowchart of the training and testing procedure is presented in Figure 6.4.

It was verified that regularization had no significant effect on the CAP or NREM classification for the LR while it considerably improved QDA performance. For LDA, only NREM detection was improved with regularization. For the CART, the maximum number of splits and size were varied between 1 and 100. The CAP classification achieved the best results using, respectively, 30 and 70, respectively. For NREM, classification was 60 and 90, respectively. The highest AUC for both NREM and CAP classification with the TreeBagger was attained using 29 trees (the number of trees was varied between 1 and 50).

The number of nearest neighbors considered for kNN was varied between 1 and 20, achieving the best results with 6 for NREM classification, and 3 for the CAP detection. A Gaussian kernel, and an outlier fraction of 5% attained the highest AUC for the classification with the SVM. Scaled conjugate gradient [308] was used as the training

algorithm for the FFNN, with the hyperbolic tangent sigmoid as transfer function. The number of neurons of the HL was varied between 1 and 30, in steps of 1. 17 neurons produced the best performance for NREM detection while 19 were used for CAP classification. The number of weak learners that were employed by the AdaBoost classifier was varied between 1 and 40. The highest AUC for both CAP and NREM classifications was achieved using 15 weak learners. The number of neurons in the autoencoders of the DSAE, was varied between 20 and 140, in steps of 20 neurons. The classifier was fine-tuned using the scaled conjugate gradient and, for both classifications, the best results were achieved using 120 and 60 neurons in the first and second autoencoders, respectively.

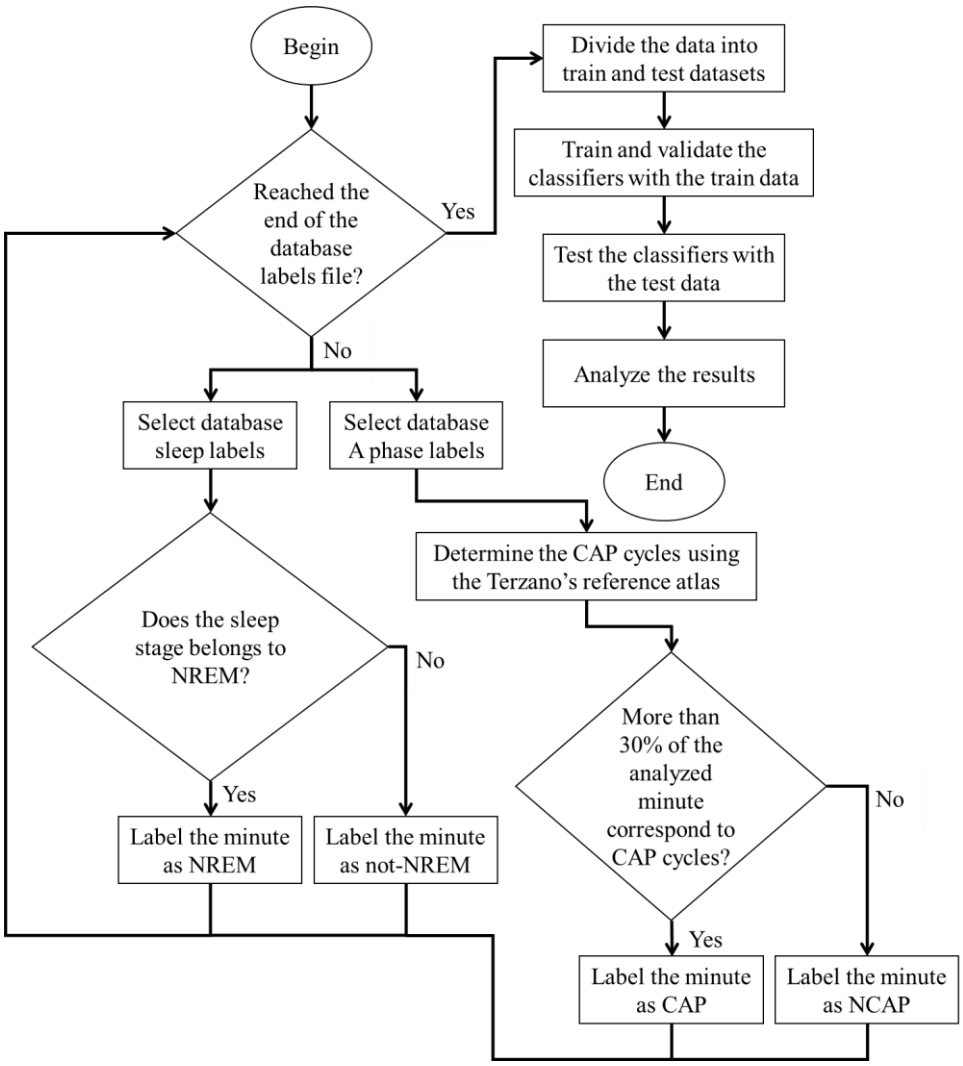


Figure 6.4. Flowchart of the employed procedure for training and testing the classifiers [80].

The results of each classifier for NREM classification are presented in Table 6.1, while Table 6.2 presents the results for the CAP classification. DSAE achieved the uppermost Sen and Acc for NREM detection while LDA provided the best Spe. By analyzing Table 6.2, it was observed that the DSAE attained the highest Acc and Spe while kNN reached the best Sen. Taking into consideration these results, it was concluded that DSAE is the best classifier to be further examined.

Table 6.1: Results of each classifier, fed with CPC, for NREM detection [80].

Classifier	*Acc (%)*	*Sen (%)*	*Spe (%)*
LR	58.15 ± 0.27	57.56 ± 0.57	60.01 ± 0.78
LDA	64.37 ± 0.13	62.40 ± 0.22	74.54 ± 0.18
QDA	60.39 ± 0.06	60.48 ± 0.11	60.13 ± 0.10
CART	61.85 ± 2.02	61.62 ± 3.05	62.59 ± 1.59
TreeBagger	66.31 ± 0.50	68.32 ± 0.93	60.01 ± 1.37
kNN	67.67 ± 0.96	70.35 ± 1.59	59.16 ± 2.13
SVM	67.76 ± 0.80	70.03 ± 1.58	60.65 ± 1.67
FFNN	62.82 ± 1.02	61.27 ± 1.77	68.68 ± 2.35
AdaBoost	61.91 ± 0.12	61.47 ± 0.39	63.17 ± 0.24
DSAE	69.51 ± 0.47	72.06 ± 1.11	60.20 ± 0.78

Table 6.2: Results of each classifier, fed with CPC, for CAP detection [80].

Classifier	*Acc (%)*	*Sen (%)*	*Spe (%)*
LR	57.11 ± 1.05	52.70 ± 1.54	58.33 ± 1.35
LDA	53.51 ± 2.37	51.87 ± 6.75	54.58 ± 4.34
QDA	58.38 ± 10.22	50.39 ± 15.85	60.61 ± 17.38
CART	56.55 ± 1.78	51.83 ± 2.47	58.15 ± 3.16
TreeBagger	59.52 ± 0.85	56.45 ± 1.69	60.21 ± 1.41
kNN	57.12 ± 1.42	61.07 ± 3.52	56.08 ± 2.39
SVM	55.86 ± 2.15	50.78 ± 4.00	57.27 ± 3.57
FFNN	61.75 ± 1.12	51.23 ± 1.43	64.64 ± 1.68
AdaBoost	53.41 ± 0.28	53.41 ± 0.14	53.48 ± 0.18
DSAE	61.89 ± 1.19	50.19 ± 1.33	65.85 1.66

6.1.2. Improving the minute-by-minute classification

A deeper study regarding the effectiveness of CPC for CAP estimation was performed testing the DSAE (classifier that achieved the best results in the viability assessment) and the FFNN (allowing to compare the performance of a deep and a shallow network) as classifiers.

Three thresholds (20%, 35% and 50%) were examined to establish the CAP epoch, and determine the predictive capability of the algorithms. The 20% threshold recognizes a minute as CAP if more than 20% of the total duration (60 s) corresponds to unstable sleep (CAP), filtering the CAP related events that are shorter than 12 s. This threshold was employed to verify if the algorithm could detect the "short" CAP related events, which last between 12 and 21 s. The 35% threshold filters the "short" CAP related events, in the analyzed minute, only considering the minute as unstable sleep if the CAP related events are longer than 21 s. The 50% threshold only considers the "longer" CAP related events that last more than 30 s to consider the minute as unstable sleep.

A threshold based approach, proposed by Thomas et al. [41] and Ibrahim et al. [52], was also analyzed as a possible alternative to the machine learning classifiers. This method considers three CPC frequency bands (VLF, LF, and HF) and examines two ratios. The first was defined by the sum of the two highest peaks in the LF band to the sum of the two highest peaks in the HF band. The second was given by the sum of the two highest peaks in the VLF band to the combined power of the two highest peaks in the LF and HF bands. Regarding the first ratio, a dominance of power in the HF band was associated with physiologic respiratory sinus arrhythmia, deep sleep, and non-CAP periods, while a preponderance of power in the LF was related to the presence of CAP periods, and can possibly indicate the occurrence of periodic respiration in SBD subjects. Concerning the second ratio, a dominance of power in the VLF was related to wake or

REM periods [41] [52]. Though this is a simple and easy to implement approach, it is likely to be dependent on the population that was employed to tune the thresholds.

The DSAE and FFNN hyperparameter optimization was done using BS and TFCV (performing 10 repetitions of each iteration for both algorithms) [318] for both CAP and NREM classification, using the configuration that maximized the AUC for each analysis. Afterwards, BS and TFCV were used for performance assessment using the tuned classifiers. The experiments were repeated 50 times to ensure the statistically significance of the results. LOOCV was used for the CAP rate estimation (repeating the simulation 50 times). The values specified by Thomas et al. [41] were employed for the thresholds based classification, testing the predictive capability of the classifier in each subject. Oversampling was used to balance the training dataset while cost-sensitive learning (considering the distribution of the training dataset's labels before oversampling) was used during the fine tune of the weights, performed using supervised learning, at the end of the training of the DSAE.

The number of neurons employed in the FFNN HL was varied between 1 and 150, in steps of 1 neuron. It was verified that 140 and 129 neurons achieved the best AUC for the NREM and CAP classification, respectively. The training algorithm and transfer function were the same as employed in the previous subsection (6.1.1).

A variation of the proposed heuristic search method for finding the structure of classifiers without an explicit feature creation procedure, named HOSA-A, was employed for optimizing the DSAE architecture. The number of encoders, *Enc*, which composed the network, *G*, was varied from 1 to the maximum number, G_{max}, chosen to be 4 or until no significant improvement was reached for the AUC (considering a threshold, t_h, of 1% to define the minimum significant improvement) when the following encoder was introduced. The number of neurons in the first encoder, *N*, was varied between 10 (N_{start}) and 150, which was chosen to be the maximum number, N_{max}, in steps, N_{step}, of 10. The number of hidden units used by the subsequent encoders was chosen to be either half (using the *floor* function to round the result of the division), the same, or twice the number of hidden units employed by the previous encoder. The output of all networks was a fully connected layer that used the soft-max function for classification. The HOSA-A follows the subsequent pseudo code:

```
HOSA-A (Data, G_max, N_start, N_step, N_max, t_h)
G = [1, 2, …, G_max]
N = [N_start, N_start + N_step, …, N_max]
for g = 1 to length (G)
   for n = 1 to length (N)
      for m = 1 to 3
         Net_0 ← I_pt (Data)
         N_prev = N (n)
         for z = 1 to g
            Net_z ← Net_z-1 + Enc (N_prev)
            if m == 1
               N_prev = floor (N (n) / 2 + 1 / 2)
            else
               if m == 2
```

```
                N_prev = N (n)
            else
                N_prev = N (n) × 2
        Net_g,n,m ← Net_g + F_C (2)
        AUC_g,n,m ← test (train (Net_g,n,m)
    AUC_g,n,m,max = max(AUC_g,n,m)|for all n,m
    if g > 1
        if AUC_g,n,m,max − AUC_g-1,n,m,max ≤ t_h
            if AUC_g,n,m,max > AUC_g-1,n,m,max
                BestNet = Net_g,n,m|AUC_g,n,m,max
            else
                BestNet = Net_g-1,n,m|AUC_g-1,n,m,max
            break
        else
            BestNet = Net_g,n,m|AUC_g,n,m,max
return BestNet
```

The DSAE was fine-tuned using the scaled conjugate gradient and, for both NREM and CAP classifications, the best results were achieved using 120 and 60 neurons in the first and second autoencoders, respectively. The number of examined networks was 135, and each simulation was repeated 10 times. Thus, the total number of simulated networks was 1350 for each examined validation method.

Three windows were tested (3 min, 5 min and 7 min) using the FFNN for the NREM and CAP classification. It was verified that the 7 min window, with six min overlapping (windows used in the viability test, presented in the previous subsection), achieved the highest AUC that was, on average, 7% and 3% higher when comparing with the 3 min and 5 min window, respectively. Therefore, increasing the window also improves the performance of the classifier, possibly due to the growth in the amount of data available. However, an increase to a 9 min window did not provide a notorious increase in the AUC. Consequently, the 7 min window was used in the tests, allowing to compare the new results with the ones previously attained.

Tables 6.3 and 6.4 present the performance of the classifiers regarding the NREM and CAP classification, respectively. By analyzing the tables, it is possible to verify that DSAE achieved the best results, followed by the FFNN, and the thresholds based classifier produced the worst results (possibly indicating that the thresholds, tuned using a different population, did not generalized well for the studied subjects). It is also possible to verify that BS improved the AUC, when comparing with the TFCV, and reduced the standard deviation. This could possibly indicate that the higher amount of data available using BS allowed to train the classifiers with more patterns, leading to a performance improvement.

Table 6.3: Results for NREM detection of the second test with CPC [81].

Classifier	*Method*	*Acc (%)*	*Sen (%)*	*Spe (%)*	*AUC*
FFNN	TFCV	62.82 ±1.02	61.27 ±2.36	68.68 ±1.77	0.646 ±0.007
	BS	73.92 ±0.75	76.67 ±1.13	66.36 ±1.72	0.701 ±0.007
DSAE	TFCV	70.69 ±4.56	74.06 ±5.95	61.06 ±5.83	0.661 ±0.035
	BS	73.67 ±0.67	76.65 ±1.14	66.52 ±1.09	0.710 ±0.004
Threshold	-	60.19 ±5.74	31.09 ±12.39	67.23 ±6.03	0.492 ±0.072

The entropy of the variability, causality, and connection between N-N series and EDR were analyzed in the third test for both classifiers (FFNN and DSAE), implementing the model presented in Figure 6.5, considering the 7 minute window with the 35% threshold for the CAP epoch definition. CPC analysis was based on the previously employed method. However, since the Fourier expansion is based on frequency resolution, it is not possible to know when the frequencies occur in a signal. To address this concern the wavelet analysis, a time–frequency joint representation [380], was also performed. Specifically, the Morlet wavelet (chosen because it is frequently employed in the state of the art to distinguish the abnormal heartbeat behavior in the ECG, which is non-stationary [380]) was used in the estimation of both cross-spectrum and magnitude squared coherence, having time-widths that are adapted to their frequencies.

Table 6.4: Results for CAP detection of the second test with CPC [81].

Classifier	*Method*	*CAP (%)*	*Acc (%)*	*Sen (%)*	*Spe (%)*	*AUC*
FFNN	TFCV	20	57.57 ±6.09	51.26 ±1.79	61.26 ±5.64	0.563 ±0.071
		35	61.75 ±1.12	51.23 ±1.42	64.64 ± 1.68	0.584 ±0.008
		50	61.03 ±7.01	48.44 ±7.70	64.89 ±1.68	0.567 ±0.101
	BS	20	63.82 ±0.56	47.40 ±1.10	71.31 ±1.00	0.594 ± 0.005
		35	72.73 ±0.62	52.92 ±1.25	73.36 ±0.96	0.616 ± 0.004
		50	74.43 ±0.80	67.03 ±1.71	74.99 ±0.93	0.710 ± 0.007
DSAE	TFCV	20	61.88 ±7.35	41.69 ±2.00	73.56 ±5.73	0.576 ±0.097
		35	72.23 ±4.18	58.37 ±3.17	75.94 ±6.24	0.688 ±0.074
		50	83.25 ±3.76	51.30 ±2.43	87.15 ±5.87	0.692 ±0.090
	BS	20	62.76 ±0.65	68.11 ±2.63	60.41 ±1.42	0.643 ±0.009
		35	72.97 ±0.89	59.65 ±1.11	75.46 ±1.18	0.691 ±0.013
		50	80.83 ±2.34	59.84 ±3.66	83.52 ±2.78	0.727 ±0.011
Threshold	-	20	54.02 ±9.51	36.86 ±10.05	66.38 ±8.81	0.516 ±0.019
		35	58.02 ±7.43	38.35 ±11.9	66.10 ±8.71	0.522 ±0.034
		50	62.79 ±6.50	46.07 ±20.17	65.92 ±8.43	0.559 ±0.061

A causality analysis was performed through the matrix of lags (a method for causality examination that was proposed in this work; detailed information about this method is presented in section 6.3) that was composed of the past values of the EDR signal, which are causal with the N-N series. The maximum number of lags was selected to be 70, thus corresponding to roughly 30 s of data, and this value was chosen since it is the standard scoring period for scoring sleep macrostructure and also because it is higher than the average CAP cycle duration (26.9 s) [38]. The causality between the EDR and the N-N series is expected to be a good marker for instable sleep (CAP) as in phenomena such as respiratory sinus arrhythmia, the heart rate is modulated by the breathing pattern, and this phenomena is also associated with the CPC HF coupling which, in turn, is linked to periods of stable sleep (non-CAP).

The residuals of the regression were used to estimate the Bayesian information in terms of a residual sum of squares, which was employed to select the relevant lags [381]. The output matrix, named Matrix of Lags (MoL), was created by moving the energy of the selected lags to the MoL. Information dynamics characterize the causal statistical structure of time series, concerning a driver and target systems that are causally connected [382]. Taking into consideration that both EDR and N-N series present significant variations during periods of sable (non-CAP) and unstable (CAP) sleep, and that entropy allows to qualify the degree of complexity in a signal. Hence, it was theorized that these oscillations could be identified by entropy variations. Therefore, the self-entropy (stored information) and transfer entropy (transferred information) were analyzed by considering the N-N as the target system and EDR as the driver system (this selection takes into

consideration that in phenomena such as respiratory sinus arrhythmia, the heart rate is modulated by the breathing pattern [383]).

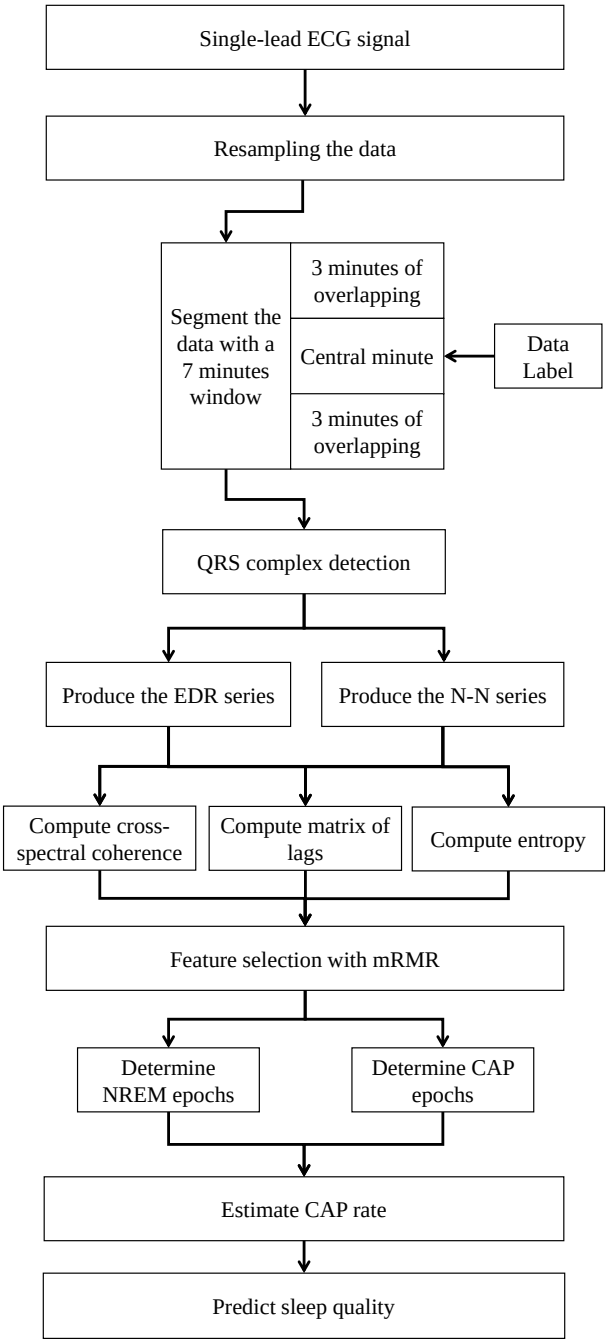


Figure 6.5. Flowchart of developed algorithm for the minute-by-minute detection of CAP considering characteristics of the N-N series and EDR [42].

A total of 672 features were examined (500 for CPC based on the Fourier analysis, 100 for CPC based on the wavelets, 70 for the MoL, and 2 for entropy analysis, specifically, self-entropy and transfer entropy), and their relevance was assessed by the mRMR algorithm. The optimal number of features was determined for both classifiers by applying the same training procedure of the second test and starting with 50 features identified as the most relevant. Afterwards, the number of features was incremented in steps of 50 features until no improvement was attained. At this point, the second phase

started to fine-tune the optimal number of features, varying 5 features each time. The optimal classifiers architecture (with the chosen number of features) and the attained results are presented in Tables 6.5 and 6.6, respectively. By analyzing Table 6.5 it is possible to determine that BS applied to the DSAE achieved the best results as it happened in the previous test. The improvement of 4% attained for both NREM and CAP detection by using more features, justified the need for introducing a higher complexity in the method.

Table 6.5: Architecture of the classifiers and selected number of features for the NREM or non-NREM, and CAP or non-CAP classification [42].

Classification	*Classifier*	*Method*	*Number of neurons*	*Number of features*
NREM or non-NREM	FFNN	CV	125	46
		BS	250	138
	DSAE	CV	250 , 125	201
		BS	250 , 125	200
CAP or non-CAP	FFNN	CV	200	154
		BS	225	146
	DSAE	CV	200 , 100	155
		BS	200 , 100	149

Table 6.6: Performance of the developed algorithms (mean ± standard deviation) for the NREM or non-NREM, and CAP or non-CAP classification [42].

Classification	*Classifier*	*Method*	*Acc (%)*	*Sen (%)*	*Spe (%)*	*AUC*
NREM or non-NREM	FFNN	CV	72.64 ± 3.74	72.77 ± 6.24	72.22 ± 7.61	0.725 ± 0.031
		BS	75.20 ± 0.97	76.48 ± 1.27	72.35 ± 1.26	0.749 ± 0.009
	DSAE	CV	73.41 ± 1.63	72.12 ± 2.71	74.71 ± 2.01	0.732 ± 0.035
		BS	76.65 ± 1.06	77.23 ± 0.96	75.86 ± 3.04	0.760 ± 0.016
CAP or non-CAP	FFNN	CV	75.84 ± 0.91	66.51 ± 0.98	79.18 ± 1.27	0.728 ± 0.007
		BS	73.02 ± 1.04	69.97 ± 1.94	74.72 ± 1.83	0.730 ± 0.003
	DSAE	CV	75.96 ± 4.67	69.33 ± 4.08	78.04 ± 7.80	0.739 ± 0.023
		BS	76.04 ± 1.71	70.67 ± 3.32	78.96 ± 1.15	0.741 ± 0.022

6.2. Global sleep quality assessment

It was verified in the previous section (6.1) that the best results for the minute-by-minute examination were attained using the 35% threshold for CAP detection. Therefore, this value was selected for the CAP rate analysis, used as the global sleep quality metric. The recordings of the fourth CAPSD were employed for this analysis. The performance of the classifiers analyzing only the CPC signal for the CAP rate estimation is presented in Table 6.7. The DSAE provided the best results for both mean absolute difference of the predicted and the true CAP rate, and the accuracy of the sleep quality prediction (considering the normal age-related CAP rate as threshold [10]).

Table 6.7: Results for CAP rate detection of the second test with CPC [81].

Classifier	*Mean absolute difference (%)*	*Acc sleep quality (%)*
FFNN	15.06	62
DSAE	9.35	77
Threshold	37.14	31

To further study the sleep quality predictions based on the CAP rate assessment, the predictions of the best model developed in the third test (features based on CPC, causality and entropy fed to the DSAE) were analyzed in the databases: UCDSAD; MrOSSS; DrNUH.

It is intended to calculate the agreement between the sleep quality prediction, based on the CAP rate, and other metrics used to indicate sleep quality. For the CAPSD, the

references metrics were: CAP rate; TST according to the normal values for the subject's age [384]; SE considering 85% as the threshold [385]. For the UCDSAD the considered sleep quality metrics were: TST; SE; AHI greater than 10, indicating the incidence of a mild sleep-related disorder [386]; ESS greater than 10, advises excessive daytime sleepiness, thus, assuming poor sleep quality [387]. For MrOSSS the sleep quality metrics were: TST; SE; AI according to the subject's age [388]; PSQI greater than 5, suggesting poor sleep quality [389]. For the DrNUH patients, the sleep quality metrics were: TST; SE; AHI; ESS. The summary of this analysis is presented in Table 6.8.

Table 6.8: Agreement of the sleep quality prediction with the sleep quality metrics [42].

Data source	*Metric*	*Agreement (%)*
CAPSD	TST	54
	SE	69
	CAP rate	77
UCDSAD	TST	48
	SE	69
	AHI	76
	ESS	60
MrOSSS	TST	60
	SE	70
	AI	72
	PSQI	58
DrNUH	TST	84
	SE	79
	AHI	73
	ESS	57

By analyzing Table 6.8, it is possible to determine that CAP rate, AHI, and AI are the metrics which attained the highest agreement with the developed model's predictions. These results are expected since the model is based on CAP analysis that is strongly related to these metrics. A good agreement was attained with the SE with a standard deviation of 4.2% among the analyzed databases. TST presented the highest standard deviation (13.7%), suggesting that the population, and the way that the study was conducted, can strongly affect this metric. These results agree with the conclusions presented Åkerstedt et al. [9] and Kaplan et al. [11] which suggested that continuity based metrics have a good correlation with subjective ratings of prior-night sleep quality, while the opposite occurs with duration based metrics. The low agreement achieved with both ESS and PSDI, about 58%, supports the findings of Rošťáková et al. [63], which have determined that there is a low correspondence between objective sleep metrics and the subject's assessment of the sleep quality.

Further examination regarding the agreement between the predicted sleep quality and the analyzed sleep quality metrics was performed by a regression study, using the Random Sample Consensus (RANSAC) algorithm [304] to diminish the effect of outliers. The results are presented in Figure 6.6, and the data was standardized to allow a better comparison (both TST and SE horizontal axis were inverted with the purpose of having all regression lines with a positive slope). The highest value for TST was chosen to be 10 hours, and the theoretical maximum values for both AI and AHI were considered. PSQI and ESS limits were, respectively, 21 and 25. Two groups were studied, specifically, the middle aged and the elderly population (young adults were not included since the number of recordings available in the databases for this age group was less than 2% of the total number of subjects). Two CAP rates were considered, 37.5% for middle aged and the 55.3% for the elderly, in agreement with the defined age-related percentages [10]. By

analyzing the regression lines, it is possible to conclude that the CAP rate predictions (considering, respectively, 37.5 and 55.3 as the threshold for the middle aged and elderly subjects) for sleep quality have a good agreement with the other studied sleep quality metrics (as they define the good or poor sleep quality according to the previously indicated definitions). Therefore, these results further validate the developed method.

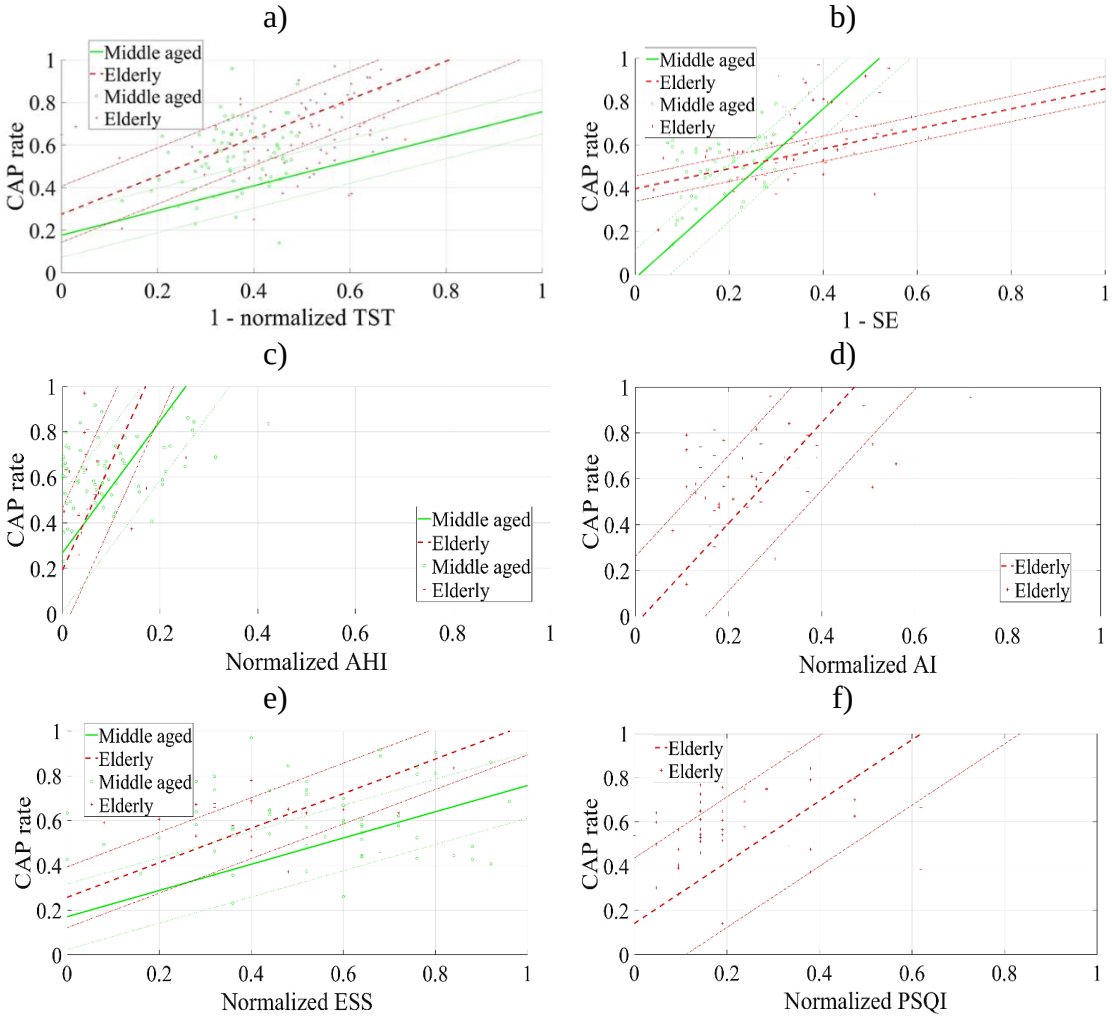


Figure 6.6. Data points (symbolized by '+' or 'o') and regression lines (dashed or solid lines, with margins denoted by doted-dashed lines), from RANSAC, for the predicted CAP rate (estimated sleep quality metric), and the sleep quality metric: a) TST; b) SE; c) AHI; d) AI; e) ESS; f) PSQI [42].

The possibility of providing a global estimation for the quality of sleep, based on the age-related CAP rate percentages [10], was tested by developing a method that produces an average of the CPC analysis for the all night recorded signal instead of performing the estimation for each epoch. Consequently, the CPC algorithm is the same as previously analyzed but a new way of training the classifier was developed, using the subjects of the fourth CAPSB. A comparative example of the globally estimated CPC for a subject with good and a subject with poor sleep quality is presented in Figure 6.7. By analyzing the figure, it is possible to attest that the subject with good sleep quality has a substantial amount of power in the HF band (0.1-0.4 Hz), while the subject with poor sleep quality has most of the power in the VLF and LF bands (0-0.01 Hz and 0.01-0.1 Hz, respectively). These observations agree with the conclusions reported in the state of the art where the

power in the HF coupling is associated with stable sleep (absence of CAP) while the power in the LF coupling is related with unstable sleep (suggesting the occurrence of CAP) [41].

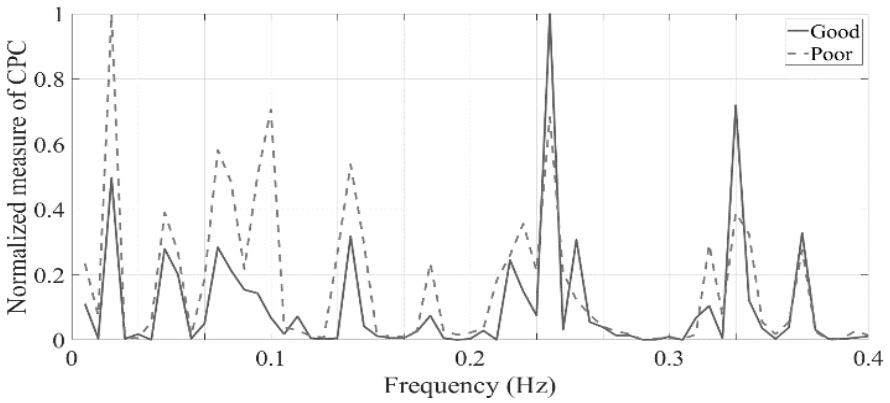


Figure 6.7. Example of a global spectrographic measure of a subject with good sleep quality and poor sleep quality.

The sleep quality labels were produced using the methodology presented in section 4.1. The spectrographic measure of CPC was feed to a 1D-CNN to perform the sleep quality classification. This neural network was chosen since it is capable of identifying complex patterns in image based inputs which are similar to the spectrographic measure created by CPC. The configuration of the CNN architecture was chosen by the heuristic oriented grid search method developed for the 1D-CNN which is presented in section 8.1. The ADAM algorithm [343] was used for the network's optimization, performing 10 runs, at each iteration, with LOOCV. The methods were developed using the recordings from the fourth CAPSD and were implemented in Python 3 using the Keras library. The optimized classifier was ten tested with LOOCV repeating each iteration 50 times to attain statistically significant results.

An accuracy of 74% was achieved (regarding the sleep quality estimation) by using two GL in the HLs, each composed of a convolution, followed by the normalization, and a pooling operation. The first convolution layer employed 64 filters with length 4, while the second used 128 filters with the same length. The pool size (of the pooling layers) was 2. A stride of 1 was employed for the convolution layers, and a stride of 2 was used for the pooling layers. It was verified that the developed method is considerably simpler than the previously presented methods for sleep quality estimation, requiring less computational resources. Therefore, this method attained a better performance to complexity ratio, and could possibly be suitable for hardware implementation.

6.3. Causality model for sleep quality analysis

A tool for time series analysis was developed to further study the connection between the EDR and the N-N series. The tested hypothesis considers that a model, based on matrix constructed with the measured energy of a fixed number of lags (previously named MoL), can be used to analyze the causality of dependent and independent variables. Therefore, the two time series, produced from the examination of the single-lead ECG signal, were used to perform the indirect detection of CAP, which was previously

described, by feeding a machine learning classifier (in this work a SVM was used and the model was implemented in MATLAB) with the energy of the lags that compose the MoL.

The proposed approach is based on the Granger causality concept, which considers that if a time series has a causal effect on another time series, then the information of the past of the first time series (independent series) can be employed to estimate the future values of the second time series (dependent series). In this work, it was hypothesized that the number (and energy) of lags considered by the information criterion as relevant (best compromise between complexity and information) would be different from stable to unstable sleep, as the EDR and the N-N series become less synchronized, and oscillate at a dissimilar frequency. The causality evaluation was performed by looking for dependencies within the measured time series (from the physiological system under investigation). Therefore, causality is defined in terms of predictability, and considers the directionality of time to assess a causal ordering of the dependent variables [390].

The same 7 minutes window with 6 minutes overlapping, which was evaluated in section 6.1, was used. Taking into consideration that the number of data points from both EDR and N-N series varies according to the HRV thus, both EDR and N-N series were resampled to 2.333 Hz, providing 980 points for the 7 minutes. Consequently, a uniform input size was created, and a fixed lag window of 30 s, with size *L*, was used. The dataflow of the developed method to generate each row of the MoL is presented in Figure 6.8.

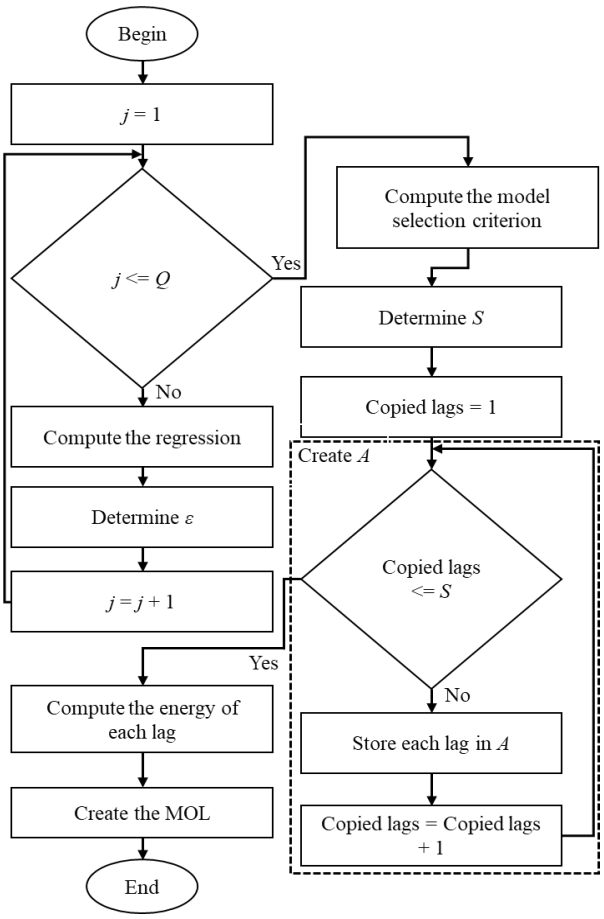


Figure 6.8. Dataflow of the algorithm that generates each row of the MoL [391].

A dataset $D = \{T1_i, T2_i\}$, where *T1* and *T2* are, correspondingly, the response and predictor time series of each epoch, *i*, with *m* points, was analyzed, and Figure 6.9 presents an example of an epoch for both time series.

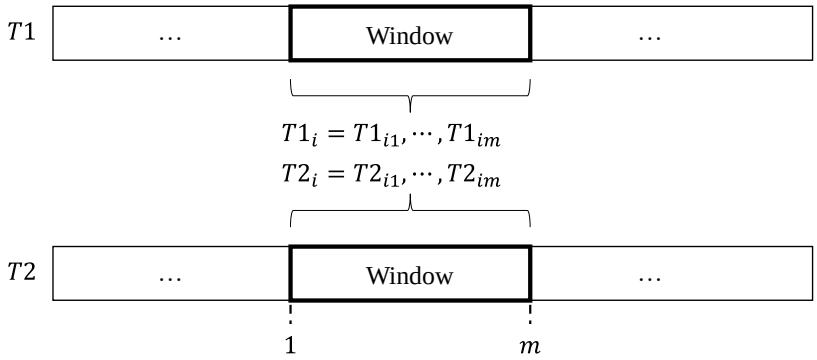


Figure 6.9. Example of an epoch for both time series of the MoL [391].

Each row of the MoL is composed of the energy of the lags that comprise a lag window, named lag, for each epoch. The lag runs over the two time series *Q* times without overlapping. *Q* is defined by the ratio *m*/*L*, and must be a positive integer, otherwise, zero padding was used in the last lag window. Therefore, *Q* time series are created with length *l* which changes in each iteration *j* as presented in Figure 6.10.

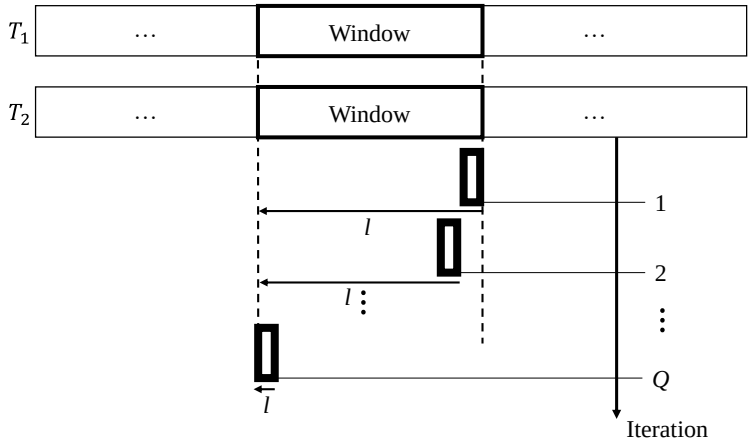


Figure 6.10. Example of the time series creation for each iteration [391].

At each iteration *L* regression models were fitted, and a model selection criterion was used to decide what is the number of lags, *S*, that produces the best compromise between complexity and information for each iteration. The selected data points, in each iteration, that are related to *T2* (from *l* until *l* – *S*) were copied to the j^{tm} row of the *Q*×*L* auxiliary matrix *A*.

However, if *S* was lower than *L* then the elements of the row that were between *S*+1 and *L* (are outside the selected number of lags) were left with a zero value. At the end of the final iteration, the energy of each lag was assessed and stored in the corresponding row of the MoL. An example that presents the creation of a MoL row can be followed by analyzing Figure 6.11. When the lag begins in the first point of the window, as displayed in figure a), the predictors, *x*, and the fitted values, *y*, are given by

$$\begin{cases} y_1 = \begin{bmatrix} T1_{im} \\ \vdots \\ T1_{i1} \end{bmatrix} \\ y_2 = \begin{bmatrix} T1_{i(m-1)} \\ \vdots \\ T1_{i1} \end{bmatrix} \\ \vdots \\ y_L = \begin{bmatrix} T1_{i(m-L)} \\ \vdots \\ T1_{i1} \end{bmatrix} \end{cases} \quad 6.5$$

$$\begin{cases} x_1 = \begin{bmatrix} T2_{im} \\ \vdots \\ T2_{i1} \end{bmatrix} \\ x_2 = \begin{bmatrix} T2_{i(m-1)} \\ \vdots \\ T2_{i1} \end{bmatrix} \\ \vdots \\ x_L = \begin{bmatrix} T2_{i(m-L)} \\ \vdots \\ T2_{i1} \end{bmatrix} \end{cases} \quad 6.6$$

Consequently, the *L* regression models can be describe as [340]

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_1 + \varepsilon_1 \\ y_2 = \beta_0 + \beta_1 x_1(2:m) + \beta_2 x_2 + \varepsilon_2 \\ \vdots \\ y_L = \beta_0 + \beta_1 x_1(L:m) + \beta_2 x_2(L-1:m) \cdots + \beta_L x_L + \varepsilon_L \end{cases} \quad 6.7$$

where ε is the residual of the regression, which designates the difference between the responses and the fitted values, and β are the fitting parameters. Second and third degree polynomial regression were also analyzed as an alternative for the regression models.

An information criterion was then calculated for each model, and the attained score, φ, was stored in an *L*-dimensional vector

$$\Phi = \begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_L \end{pmatrix} \quad 6.8$$

Therefore, the value of *S* (the number of lags that produced the best compromise between complexity and information) was specified, and the corresponding lags were copied to *A*. Following the example regarding the creation of a MoL row (Figure 6.11) for the case when *S*=3 the first row of the auxiliary matrix is

$$A = \begin{bmatrix} T2_{im} & T2_{i(m-1)} & T2_{i(m-2)} & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \quad 6.9$$

At the second iteration, the lag starts in the $m-L-1$ point of the window, as exhibited in Figure 6.11 b), hence,

$$\begin{cases} y_1 = \begin{bmatrix} T1_{i(m-L-1)} \\ \vdots \\ T1_{i1} \end{bmatrix} \\ y_2 = \begin{bmatrix} T1_{i(m-L-2)} \\ \vdots \\ T1_{i1} \end{bmatrix} \\ \vdots \\ y_L = \begin{bmatrix} T1_{i(m-2L)} \\ \vdots \\ T1_{i1} \end{bmatrix} \end{cases} \quad 6.10$$

$$\begin{cases} x_1 = \begin{bmatrix} T2_{i(m-L-1)} \\ \vdots \\ T2_{i1} \end{bmatrix} \\ x_2 = \begin{bmatrix} T2_{i(m-L-2)} \\ \vdots \\ T2_{i1} \end{bmatrix} \\ \vdots \\ x_L = \begin{bmatrix} T2_{i(m-2L)} \\ \vdots \\ T2_{i1} \end{bmatrix} \end{cases} \quad 6.11$$

and the regression models were defined by 6.12. *S* was calculated by applying the model selection criterion to the new scores and, for example, if *S*=*L* then

$$A = \begin{bmatrix} T2_{im} & T2_{i(m-1)} & T2_{i(m-2)} & \cdots & 0 \\ T2_{i(m-L-1)} & T2_{i(m-L-2)} & T2_{i(m-L-3)} & \cdots & T2_{i(m-2L)} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \quad 6.12$$

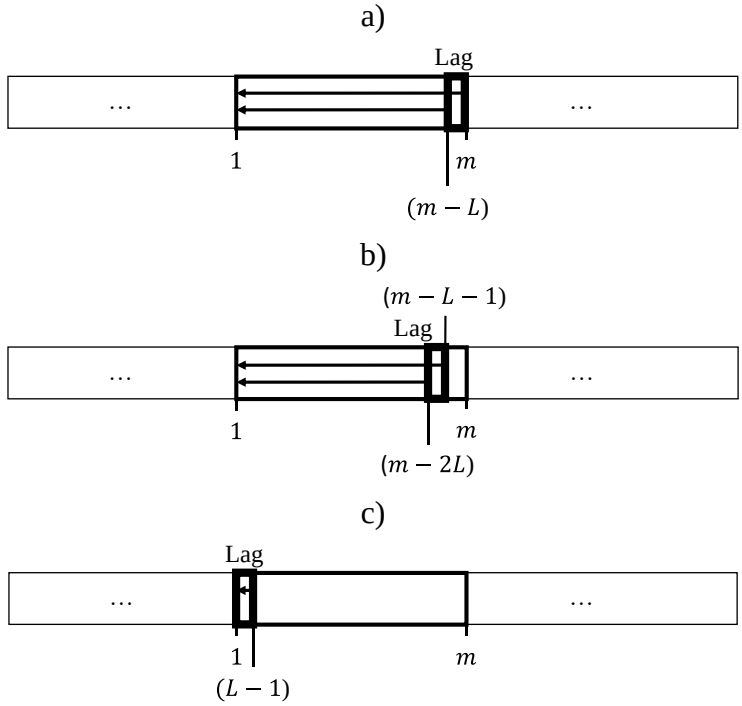


Figure 6.11. Length of the time series that will compose the lag when the lag starts: a) in last point of the window; b) in the $m-L-1$ point of the window; c) in the L-1 point of the window [391].

As presented in Figure 6.11 c), in the final iteration, the lag begins in the $L-1$ point of the window, and the same process was repeated to assess the value of *S*. As an example, if *S*=2 then

$$A = \begin{bmatrix} T2_{im} & T2_{i(m-1)} & T2_{i(m-2)} & \cdots & 0 \\ T2_{i(m-L-1)} & T2_{i(m-L-2)} & T2_{i(m-L-3)} & \cdots & T2_{i(m-2L)} \\ \vdots & \vdots & \vdots & & \vdots \\ T2_{i(L-1)} & T2_{i(L-2)} & 0 & \cdots & 0 \end{bmatrix} \quad 6.13$$

The energy of each column, *k*, of *A* was calculated by [392]

$$E_k = \sum_{e=1}^{Q} |A(e,k)|^2 \quad 6.14$$

and the value was copied to the i^{th} row of the MoL that is an $I \times L$ matrix defined as

$$MoL = \begin{bmatrix} \vdots & & \vdots \\ E_1 & \cdots & E_L \\ \vdots & & \vdots \end{bmatrix} \quad 6.15$$

where *I* is the total number of epochs.

The values of φ were assessed by the computation of an information criterion in terms of a residual sum of squares. The optimum number of lags was given by the index of the smallest element from Φ, since a penalization for the complexity of the model, given by the number of parameters required, *O*, was used to avoid overfitting. Three information criterion were tested, specifically the Bayesian Information Criterion (BIC) [381], the Akaike Information Criterion (AIC) [307], and the Hannan–Quinn Information Criterion (HQC) [307], calculated by

$$BIC = l \log\left(\frac{\varepsilon^T \varepsilon}{l}\right) + O \log(l) \quad 6.16$$

$$AIC = l \log\left(\frac{\varepsilon^T \varepsilon}{l}\right) + 2O \quad 6.17$$

$$HQC = l \log\left(\frac{\varepsilon^T \varepsilon}{l}\right) + 2O \log[\log(l)] \quad 6.18$$

These criteria were chosen because they diverge on how much the complexity is penalized, allowing to verify multiple possibilities for the model selection.

The relevance of the lags for the CAP classification (identification of unstable sleep) was assessed by a classifier dependent algorithm, specifically SBS, and a classifier independent developed method (developed in this work) based on the characteristics of a return map. SBS evaluated the AUC to predict the relevance of each lag, and the optimal number of lags to use was determined by analyzing in which iteration of the SBS algorithm the AUC was the highest.

The developed classifier independent feature selection method considers the analysis of two return maps, one created by epochs with unstable sleep (CAP) and the other by epochs without unstable sleep (non-CAP). These maps were produced by analyzing the variation of the selected number of lags from the current to the next row of the MoL

(indicating the variation in the number of chosen lags from the current epoch to the next epoch). An example of the two maps is presented in Figure 6.12.

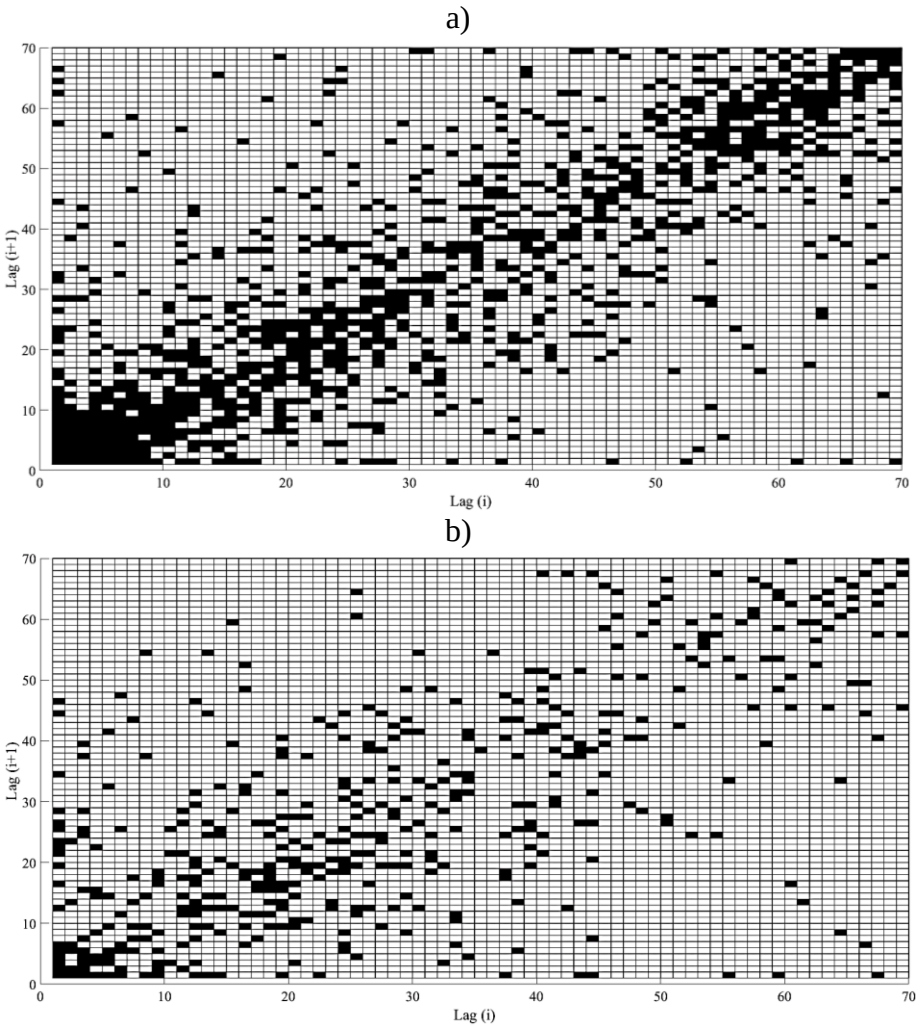


Figure 6.12. Return map presence of a) the CAP epochs; b) the non-CAP epochs [391].

Through the Figure 6.12 analysis, it is possible to verify that the variation in the lags transitions (changes from one lag to another) are significantly reduced in the CAP epochs. Therefore, a characteristic pattern is produced. It is also possible to attest the presence of a diagonal line pattern, designating that most of the time the selected number of lags in the next epoch is the same as the number of the lags of the current epoch. This behavior was anticipated since physiological signals are characterized by a continuous process. The selection of the relevant lags was performed by applying a column-wise analysis to the CAP epochs return map, and a lag was chosen when the number of transitions was 3 or lower (this value was defined as the optimal threshold for the selection by performing a grid search where the value was changed from 1 to 10, and the best average AUC was assessed by running the simulation 10 times).

The model's hyperparameter optimization (select the hyperparameters which maximize the AUC) and performance assessment was performed by TFCV using recordings from the fourth CAPSD and cost-sensitive learning. Each simulation was repeated 10 times for the hyperparameter optimization and 50 for the performance assessment (to achieve statistical significance), and the averaged value of the performance

metrics was considered. The problem related to the data unbalance was minimized by employing cost-sensitive learning. The Gaussian radial basis function was found to be the best kernel for all tests, and the value of L was selected to be 70 sampling points (corresponding to roughly 30 s of data).

A total of 6 thresholds (20%, 25%, 30%, 33%, 35%, and 40%) were analyzed in the first test to define the most suitable to define the CAP epoch, were the MoL (without the use of feature selection) was fed to the SVM. The same thresholds were used in the second tests were the performance of different regression methods were studied, using the model parameter, which achieved the highest AUC in the first test. The same model was then used in the final test, where the two feature selection methods were employed to select the lags that were fed to the SVM.

In the first test, the three model selection criteria were analyzed, and the classification performance is presented in Table 6.9. By analyzing the results, it is possible to conclude that the highest AUC was attained by using BIC for the model selection with a threshold of 35%. The lowest performance was produced by the 20% threshold, conceivably due to the presence of short CAP periods that did not significantly manifest in the ECG signal.

Table 6.9: Results of the first test regarding the models fed directly with the MoL [391].

Model selection criterion	*Method*	*CAP threshold (%)*	*Acc (%)*	*Sen (%)*	*Spe (%)*	*AUC*
BIC	N-N series as predictor	20	55.22	76.92	42.37	0.596
		25	60.29	62.43	57.93	0.602
		30	62.01	60.07	62.89	0.615
		33	62.47	60.10	63.74	0.619
		35	63.29	61.14	64.60	0.629
		40	67.63	58.66	69.84	**0.643**
	EDR as predictor	20	58.62	64.29	55.51	0.599
		25	59.03	62.60	56.70	0.601
		30	60.93	64.55	57.54	0.626
		33	62.59	67.72	61.50	0.636
		35	64.40	68.48	64.39	**0.659**
		40	76.20	50.47	82.35	0.659
AIC	N-N as predictor	20	51.93	77.43	38.21	0.578
		25	52.50	70.06	42.33	0.562
		30	58.03	62.74	60.09	0.579
		33	58.82	55.86	59.93	0.579
		35	73.88	38.59	85.42	**0.620**
		40	74.91	33.73	83.43	0.596
	EDR as predictor	20	65.29	80.20	48.73	0.645
		25	65.10	72.95	51.69	0.633
		30	65.44	71.56	54.30	**0.629**
		33	69.09	66.08	63.24	0.627
		35	71.53	60.92	67.25	0.622
		40	84.95	30.36	96.45	0.624
HQC	N-N as predictor	20	65.62	41.05	76.10	0.596
		25	65.86	48.53	75.79	0.622
		30	65.89	51.52	74.69	0.631
		33	65.12	59.98	68.21	**0.646**
		35	65.82	62.78	66.05	0.644
		40	66.76	68.89	53.10	0.610
	EDR as predictor	20	61.48	43.87	71.70	0.578
		25	62.76	49.34	69.45	0.594
		30	63.70	60.84	64.89	0.629
		33	63.89	61.84	62.92	0.634
		35	64.18	60.51	65.86	0.634
		40	74.90	44.80	82.11	**0.635**

Both BIC and HCQ attained balanced results, suggesting that the linear penalization of AIC is inadequate for this classification. It was also verified that, on average the models where EDR was used as the predictor series had a better performance, possibly due to the fact that in phenomena such as respiratory sinus arrhythmia, the heart rate is modulated

by the breathing pattern [383]. A representation of the average normalized output of the 3 selection criteria for each lag (for all CAP and non-CAP epochs) is presented in Figure 6.13. It was verified that after 26 lags the output variation is considerably low (less than 5%), possibly suggesting that the significant information is attained with less than 26 lags.

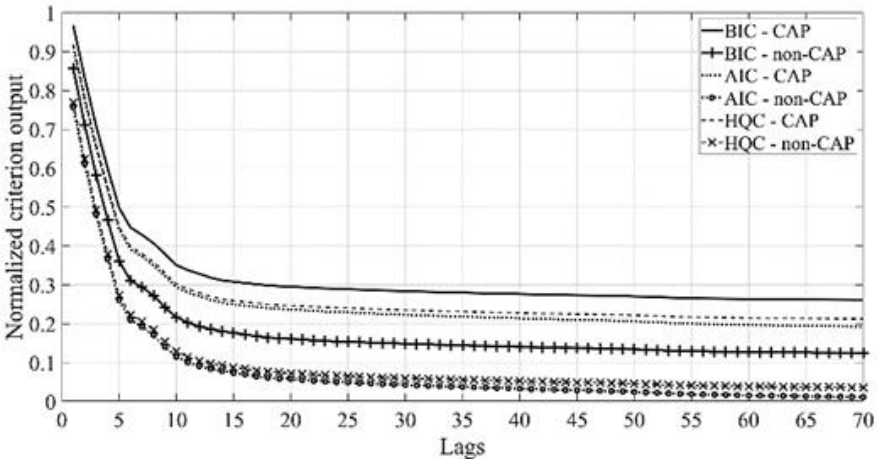


Figure 6.13. Average normalized output of the 3 selection criteria for all possible number of chosen lags [391].

The results of the second test are presented in Table 6.10, where Second Degree Polynomial Regression (SDPR) and Third Degree Polynomial Regression (TDPR) were tested. By analyzing the results, is it possible to conclude that SDPR attained a higher AUC for all CAP threshold while TDPR is possibly overfitted since, on average, the AUC is lower than the models based on SDPR. However, the 35% threshold was previously identified as the most relevant for CAP detection when considering the ECG signal and, for this threshold, the SDPR model does not significantly improve the results of the first test. Also, both SDPR and TDPR are considerably more computational demanding. Thus, the multiple linear regression models are a more suitable choice for this work.

Table 6.10: Results of the second test regarding the use of other regression models to create the MoL [391].

Degree of polynomial regression	*CAP threshold (%)*	*Acc (%)*	*Sen (%)*	*Spe (%)*	*AUC*
Second	20	56.02	84.19	39.83	0.620
	25	57.84	79.83	48.14	0.639
	30	60.59	79.19	51.21	0.652
	33	62.78	76.90	57.97	**0.672**
	35	66.25	61.75	68.30	0.663
	40	73.55	50.04	78.19	0.641
Third	20	57.40	79.78	44.66	0.622
	25	57.65	77.05	48.63	0.628
	30	58.82	70.85	54.68	0.629
	33	59.78	71.99	55.35	0.638
	35	65.76	63.14	66.77	**0.650**
	40	76.37	46.22	83.48	0.649

SBS was used in the third test, considering the model that attained the best results in the first test. The sequence of the chosen lags at each iteration of the algorithm is presented in Figure 6.14, while the variation of performance metrics according to the selected lags is presented Figure 6.15. The highest AUC was attained when the 35 less relevant lags were removed (iteration 1 to 35 of Figure 6.14). The average Acc, Sen, Spe, and AUC was, respectively, 76.45%, 71.37%, 81.55%, and 0.765. This analysis agrees

with the deduction attained from Figure 6.13 that less than half of the features are significant for the classification.

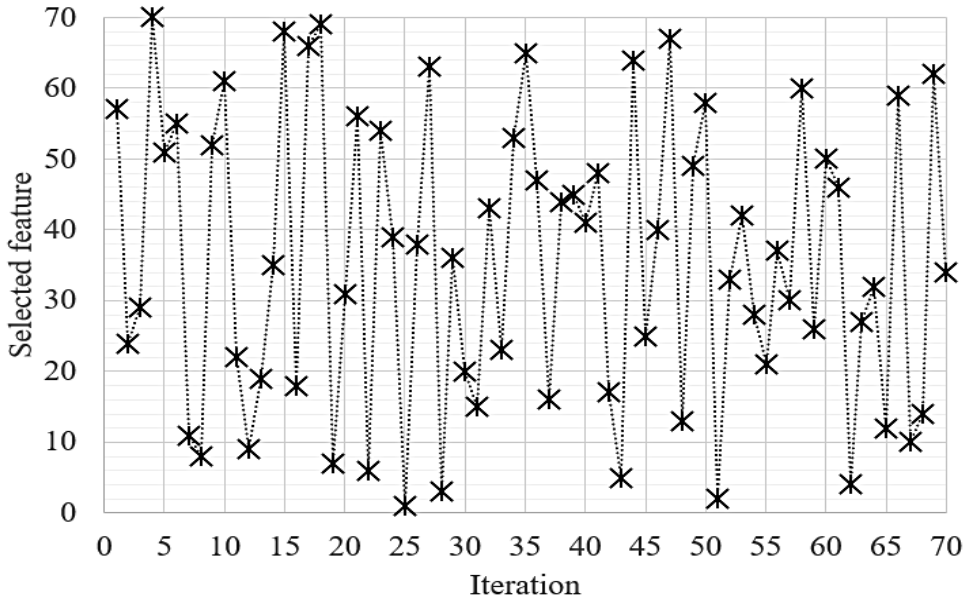


Figure 6.14. Sequence of features selected by SBS [391].

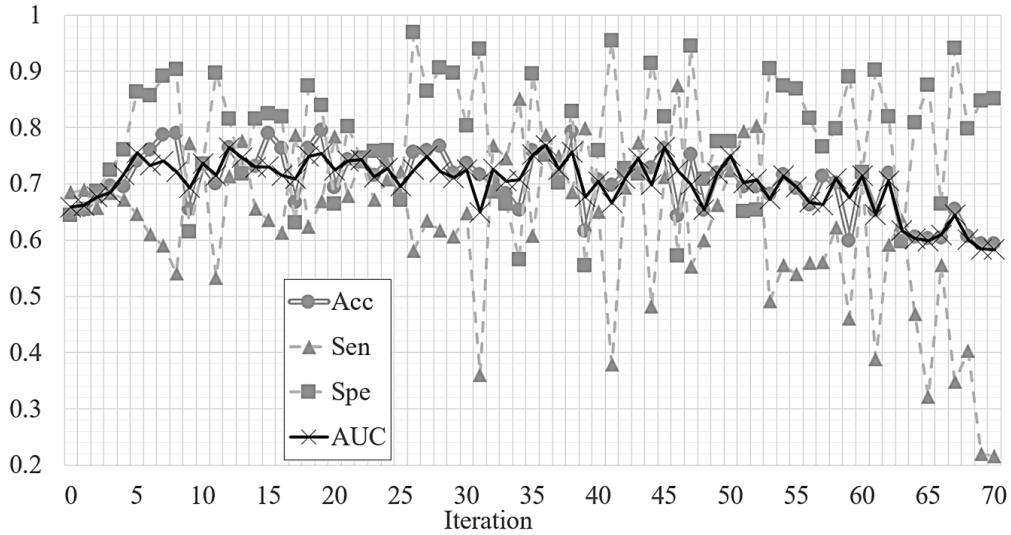


Figure 6.15. Performance metrics according to the selected lags [391].

A total of 12 features were selected by the developed classifier independent model, which achieved an average Acc, Sen, Spe, and AUC of, respectively, 70.13%, 65.12%, 74.33%, and 0.701. Although the AUC is 6.4% inferior, when comparing with the SBS model, the results produced with this method are still relevant with the advantage of taking considerably less time to select the relevant features (less than one minute while SFS took roughly one week).

6.4. Discussion of the results

Regarding the NREM classification, the best developed method achieved an average accuracy of 77% (DSAE fed with features selected by mRMR, and trained using BS),

which is in the range of the methods based on EEG (70% to 94% [393]) using only the signal from a single-lead ECG. Other works have also proposed methods for sleep classification using ECG signals. Xiao et al. [394] analyzed the HRV to detect REM, NREM, and wake periods using CARTs trained with the Random Forest (RF) algorithm. LDA and QDA were tested by Ebrahimi et al. [395] to perform the sleep analysis in two scenarios, using entropy measures, empirical mode decomposition, and discrete wavelet transform to generate features from the HRV. Four classes (wake, N2, SWS, and REM) were considered in the first scenario, while a binary classification was employed in the second scenario (REM or NREM, excluding the N1 stage). The best results were attained in the second scenario using the LDA to perform the classification of five minutes' epochs.

Mendez et al. [396] have also used binary classification for NREM and REM classification, using a time-varying Autoregressive Model (ARM) to produce features that were fed to a HMM. Aktaruzzaman et al. [397] employed a FFNN to perform the R-R series (interval between consecutive R peaks) classification with frequency and time domain measures, regularity features, and detrended fluctuation analysis. A comparison between the achieved results and the results of the analyzed works is presented in Table 6.11.

Table 6.11: Comparative analysis between the proposed NREM assessment methods and the methods proposed in the state of the art, using the ECG signal.

Work	*Method*	*Acc (%)*	*Sen (%)*	*Spe (%)*	*CO (%)*
[397]	R-R series and FFNN	72	74	69	72
[394]	HRV features and RF	73	-	-	-
[396]	ARM and HMM	79	85	70	78
[395]	HRV features and LDA	81	-	-	-
This work – first test	CPC and LR	58	58	60	59
	CPC and LDA	64	62	75	67
	CPC and QDA	60	61	60	60
	CPC and CART	62	62	63	62
	CPC and TreeBagger	66	68	60	65
	CPC and kNN	68	70	59	66
	CPC and SVM	68	70	61	66
	CPC and FFNN	63	61	69	64
	CPC and AdaBoost	62	62	63	62
	CPC and DSAE	**70**	**72**	**60**	**67**
This work – second test	CPC and FFNN	74	77	66	72
	CPC and DSAE	**74**	**77**	**66**	**72**
	Thresholds proposed by Thomas et al. [41]	60	31	67	53
This work – third test	mRMR selected features and FFNN	75	77	72	75
	mRMR selected features and DSAE	**77**	**77**	**76**	**77**

By analyzing Table 6.11, it was possible to assess that both Aktaruzzaman et al. [397] and Xiao et al. [394] reported a lower performance, while Ebrahimi et al. [395] achieved the highest Acc, but without reporting other performance metrics, such as AUC, it is not possible to verify if the results had a balanced Sen and Spe. The method developed by Mendez et al. [396] achieved better performance, but the approach is considerably more complex, requiring the estimation of both HMM and ARM parameters that could be significantly dependable upon the characteristics of the population employed for training. It is also possible to observe the significant improvement from the results attained in the first test to the results reached in the second test. The improved methods from the second test based on the FFNN and the DSE attained the same performance when fed with the CPC features. However, in the third test the DSAE attained better results than the FFNN.

The summary of the methods for indirect CAP is presented in Table 6.12. Only the method developed by Thomas et al. [41] was found, in the literature review, regarding the CAP estimation from ECG signal, presenting the results with the same performance metrics employed in this work. The reported Sen and Spe were, respectively, 40% and 84%. Thus, the developed algorithms (in the third test, where mRMR was employed for feature selection, and in the MoL method with features selected by SBS) have a lower Spe but a considerably higher Sen, suggesting that the developed method is more appropriate for a clinical diagnosis since the results are more balanced.

Table 6.12: Comparative analysis between the proposed CAP assessment methods and methods proposed in the state of the art, using the ECG signal.

Work	*Method*	*Acc (%)*	*Sen (%)*	*Spe (%)*	*CO (%)*
[41]	Tuned thresholds	-	40	84	-
This work – first test	CPC and LR	57	53	58	56
	CPC and LDA	54	52	55	54
	CPC and QDA	58	50	61	56
	CPC and CART	57	52	58	56
	CPC and TreeBagger	60	57	60	59
	CPC and kNN	57	61	56	58
	CPC and SVM	56	51	57	55
	CPC and FFNN	62	51	65	59
	CPC and AdaBoost	53	53	54	53
	CPC and DSAE	**62**	**50**	**66**	**59**
This work – second test	CPC and FFNN	73	53	73	66
	CPC and DSAE	**73**	**60**	**76**	**70**
	Thresholds proposed by Thomas et al. [41]	58	38	66	54
This work – third test	mRMR selected features and FFNN	73	70	75	73
	mRMR selected features and DSAE	**76**	**71**	**79**	**75**
This work – using the MoL	MoL and SVM	64	69	64	66
	MoL features selected by SBS and SVM	**77**	**71**	**82**	**77**
	MoL features selected by proposed method and SVM	70	65	74	70

It is also possible to notice the performance improvement from the first to the third test with CPC, although the method based on MoL attained the best performance. Another relevant factor it the Acc of the CAP detection from the best models, which is in the upper range of the specialist agreement range (69% to 78% [53]), and are also similar to the best methods for CAP assessment based on the EEG signal analysis, suggesting that the developed methods could be suitable for medical application.

6.5. Key remarks

Multiple approaches for indirect CAP analysis (seeing CAP as the marker of sleep instability, and using the CAP epoch concept which was created in this work) were evaluated in the first section of this chapter, with the goal of finding the feasibility of these methods. The research hypothesis that CAP can be interpreted in a broader context, denoting instability of sleep, and therefore, be indirectly estimated by other sensors (instead of the EEG) such as ECG, was confirmed. The CAP rate predicted by the proposed method was found to have a good correlation with the CAP rate estimated by PSG. A good correlation was also attained by the sleep quality predicted from the developed approach and the sleep quality predicted by other established sleep quality metrics.

A second sleep quality prediction method, based on the evaluation of the average CPC signal was developed, and found to be more suitable for hardware implementation. A tool for time series analysis was also proposed at the end of the chapter. This tool evaluated

the causality of a dependent and an independent variable by measuring the energy of a fixed number of lags. It was verified that the model created by this tool allowed to reliably estimate the CAP epochs. However, it is important to bear in mind that this tool can be used in other contexts involving dependent time series.

The limitation of the work developed in this chapter was the absence of subjects suffering from heart related disorders such as heart failure or cardiac arrhythmias (although the subjects suffering from OSA present cardiorespiratory dysautonomias). These disorders could significantly change the cardiac vagal activity [398], which could affect the performance of both CPC and MoL based algorithms.

7. Models for OSA examination

7.1. Based on the SpO2 signal

7.2. Based on the ECG signal

7.3. Discussion of the results

7.4. Key remarks

Models for both minute by minute and global OSA assessment are presented in this chapter as a way of predicting the quality of sleep by assessing the occurrence of sleep related disorders, according to the second approach for sleep quality examination theorized in this work (from the state of the art analysis). This disorder was selected for the examination as it is one of the most prevalent and frequently undiagnosed sleep related disorders, that it is also related to the occurrence of CAP in subjects suffering from SBD

From the OSA detection methods review, it was determined that ECG and SpO2 achieved the highest global classification. However, it was also verified that a combination of source sensors did not improve the performance. Therefore, two models were developed. The first was based on the SpO2 signal examination, and employed features which were identified in the state of the art as suitable for OSA detection based on SpO2 or other source signals (such as HRV). The second evaluated the ECG signal through CPC analysis by a deep learning classifier, comprising a new methodology proposed in this work.

7.1. Based on the SpO2 signal

The first version of the algorithm for sleep apnea detection was developed using data from 35 suspected OSA patients of the DrNUH dataset. The examination based on the SpO2 signal which was found to be suitable for OSA detection in the state of the art review. The employment of such analysis also has three relevant advantages: simplicity of self-assembly and utilization of a pulse oximeter by the subject; the direct correlation between apnea events and the reduction of the SpO2 level; the minimal invasiveness of the sensor. The signal was analyzed in both time and frequency domains, considering five-minute epochs with one-minute displacement between adjacent frames, producing 22 features that had been reported by Ravelo-García et al. [49] as relevant for apnea detection.

Two features were studied for the time domain analysis; specifically, the variance (Var) of one and five minutes, centered in the desired time segment, given by [399]

$$Var = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 \tag{7.1}$$

where N is the number of samples of the signal x and μ is the average of x. A filter bank with 20 equally spaced filters was employed in the frequency domain analysis. The implementation was made on the transformed domain as [49]

$$F_b{}^T = \frac{\sum_{k=b_m-\Delta_m}^{b_m+\Delta_m} S(k)U_{\Delta_m}(k)}{\sum_{k=0}^{\frac{N}{2}-1} S(k)}; 1 \le m \le M \tag{7.2}$$

considering U as the rectangular windowing process applied to each filter m, for the number of employed filters M, with a center frequency of the filter band b, and bandwidth Δ. U is the power spectrum of the signal, and was determined through the periodogram [49]

$$U(k) = \left| \frac{1}{N} \sum_{n=0}^{N-1} x_n e^{\frac{-j2\pi k}{N}} \right|^2 \quad 7.3$$

These features were selected since relevant information for the OSA classification can be attained by analyzing the SpO2 signal in the spectral domain, by detecting the oscillations in the signals. It was also observed that longer time periods allow to easily detect the SpO2 oscillations through the analysis of the signal's variance.

Feature selection was performed using the SFS algorithm, repeating 50 times each iteration of the algorithm, and the selected features were the two variance features and the energy of the filters: 2; 3; 8; 11; 12; 20. The Neperian logarithm was applied to all features before they were fed to the classifier, producing a dynamic compression, which creates a system less sensitive to dynamic changes of the features.

Discrimination of apnea and non-apnea periods was performed using LR. A threshold-based diagnostic rule was afterwards implemented to classify the epoch according to the resulting probability. All methods were developed in Python 3, and TFCV was employed for performance assessment (performing 50 repetitions of each iteration to achieve statistically significant results).

It was verified that the average Acc, Sen, Spe, and AUC of the algorithm were, respectively, 86.6%, 66.9%, 94.5%, and 0.90. A second version of the algorithm was created by introducing data from another 35 subjects (a total of 70). The SFS procedure was repeated, and five features were selected. Specifically, the variance of the central minute in the 5 min segment, and the energy of the filters: 2; 3; 8; 9. It was verified that the average Acc, Sen, Spe, and AUC of the improved algorithm were, respectively, 87.5%, 79.5%, 90.8%, and 0.92. Figure 7.1 presents an example of the variation of the selected features during OSA events.

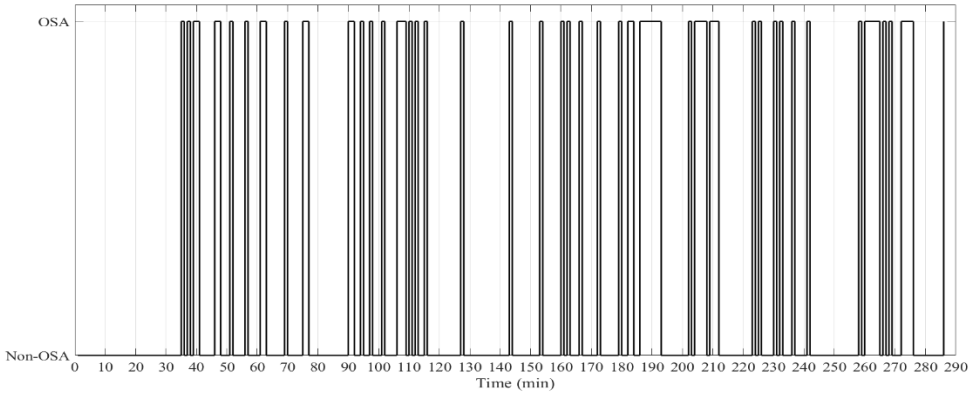
The number of OSA events was examined to perform the global OSA diagnosis by comparing the ratio

$$\mathrm{m-AHI-tib} = \frac{\text{minutes classified as OSA}}{\text{time in bed in minutes}} \quad 7.4$$

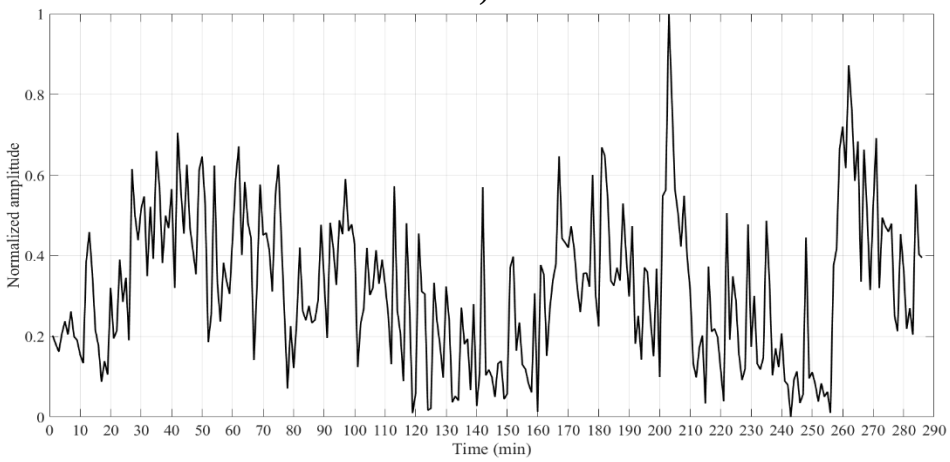
named m-AHI-tib, with a threshold. This ratio is highly correlated to the AHI [49] [400] [228] measured by PSG, and it has the benefit of not requiring a sleep or wake classification. According to the AASM, the reference AHI for OSA diagnosis is having an average of more than 5 events in 60 minutes [21]. Therefore, the selected threshold was 0.083 (≈ 5/60).

The regression plot of the AHI acquired by PSG and the predicted m-AHI-tib is presented in Figure 7.2. The regression R^2 was 0.87, further validating the technique for OSA diagnosis. The attained accuracy for the OSA diagnose was 95%.

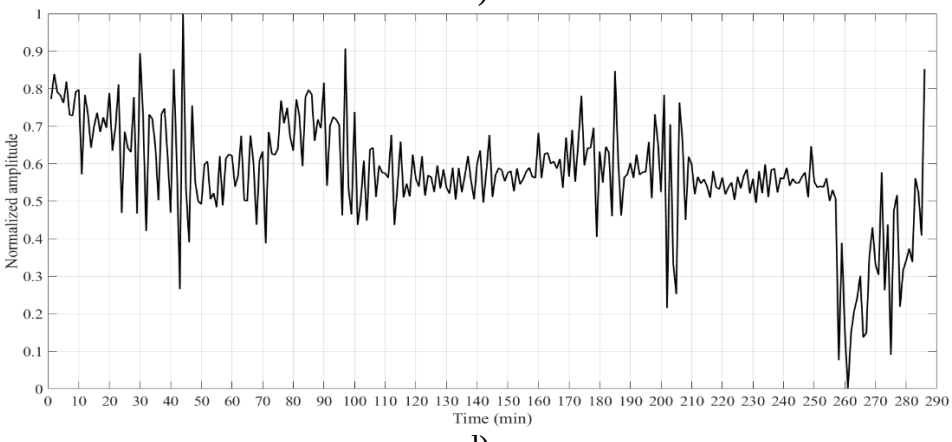
a)


OSA
Non-OSA
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250 260 270 280 290
Time (min)

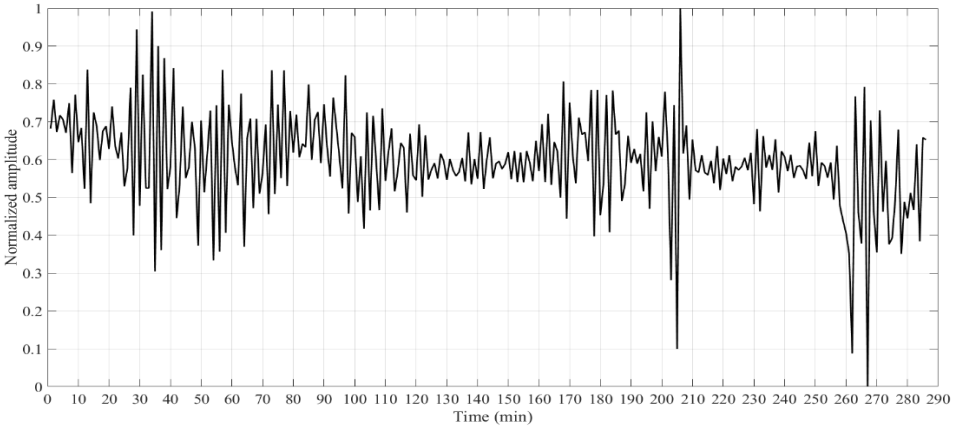
b)


Normalized amplitude
0 0.2 0.4 0.6 0.8 1
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250 260 270 280 290
Time (min)

c)


Normalized amplitude
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250 260 270 280 290
Time (min)

d)


Normalized amplitude
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250 260 270 280 290
Time (min)

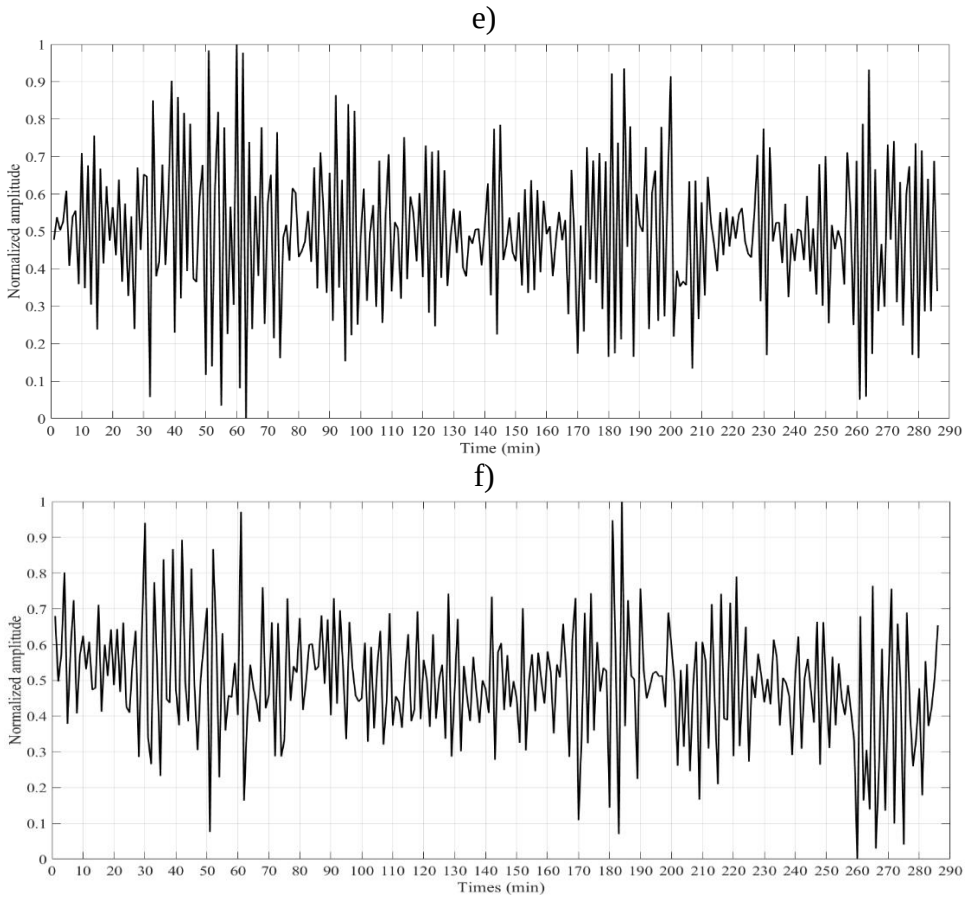


Figure 7.1. Variation of the selected futures during OSA events, identified in a). The features are the b) variance of the central minute, and the energy of the filters c) 2; d) 3; e) 8; f) 9.

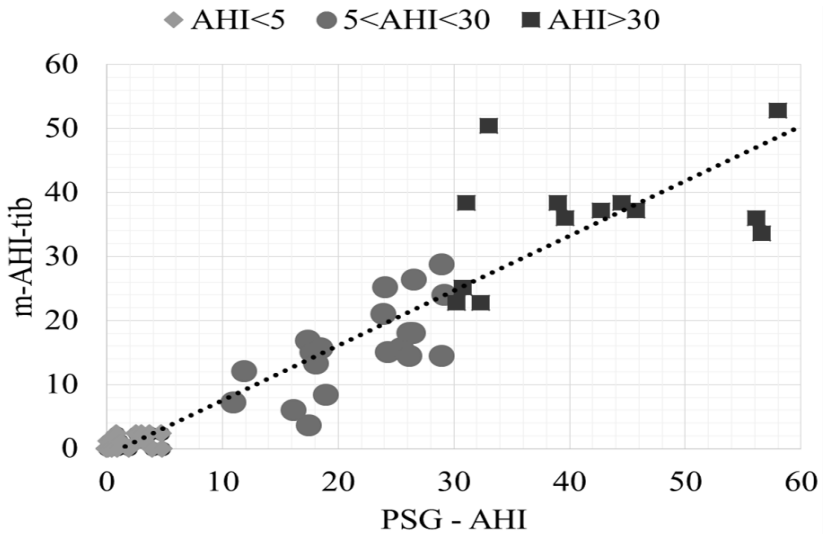


Figure 7.2. Regression plot of the AHI acquired by PSG and the estimated m-AHI-tib from the SpO2 based model [401].

7.2. Based on the ECG signal

The CPC signal was examined for the OSA detection based on the ECG signals from the DrNUH dataset. This approach was selected since ECG signal analysis was identified

in the state of the art methods review as the most suitable for OSA detection. The CPC signal was evaluated as it is intended to use only this signal for the proposed sleep quality model, presented in chapter 8. Therefore, the previously examined 7 minute window for the indirect measurement of CAP (based on the ECG signal) was used to create the CPC signal which was fed to a 1D-CNN. All methods were developed in Python 3 using the Keras library.

The hyperparameter optimization and performance assessment were performed by TFCV repeating each iteration 10 and 50 times (the optimized network was employed for performance assessment), respectively, using the ADAM algorithm [343] for the network's optimization. However, the configuration of the CNN architecture was chosen by the heuristic oriented grid search method developed for the 1D-CNN, which is presented in section 8.1. The best AUC was attained by using two GL, each composed of a convolution, followed by the normalization, and a pooling operation. A total of 128 filters, with length 5, were used in the first convolution layer, while 256 filters (with length 5) were employed in the second convolution layer. A stride of 2 was used for the pooling layers (with a pool size of 2), while a stride of 1 was used for the convolution layers. For the minute-by-minute examination, the attained Acc, Sen, Spe, and AUC were, respectively (average ± standard deviation), 76.4% ± 4.33%, 72.7% ± 6.49%, 75.3% ± 3.61%, and 0.77 ± 0.04. This performance is in the range of the reported metrics of the works available in the state of the art for OSA classification based on ECG [56].

The global classification procedure applied for the SpO2 signal was also employed for this model, and the attained global Acc was 94.3%. The regression plot of the AHI acquired by PSG and the predicted m-AHI-tib is presented in Figure 7.3, attaining an R^2 of 0.79.

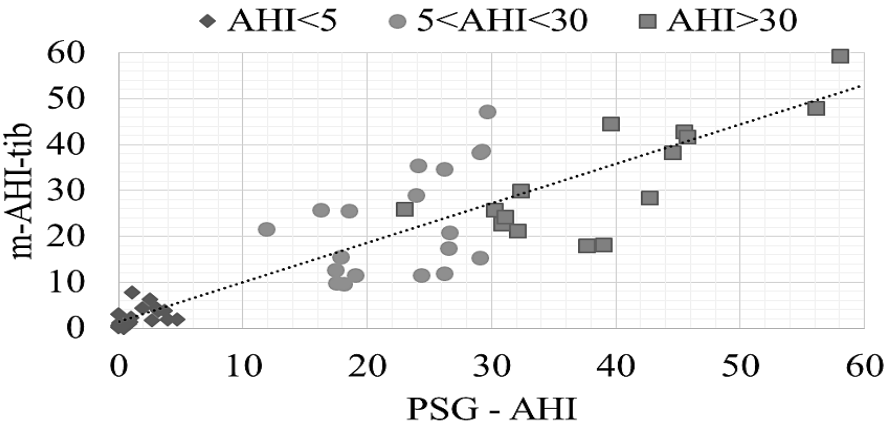


Figure 7.3. Regression plot of the AHI acquired by PSG and the estimated m-AHI-tib from the ECG based model [299].

7.3. Discussion of the results

A comparative analysis between the developed algorithms for the epoch based OSA classification, and the state of the art works reviewed in section 3.2 is presented in Table 7.1. By analyzing the table, it is possible to verify that the employment of machine learning methods achieved better and balanced results. Comparing the first and second versions of the developed SpO2 based algorithms, it is notorious that an improvement in

the performance was attained, with a significant increment (13%) in the Sen. Reaching a similar Acc with the bigger dataset is a good indicator of the robustness of the method.

Table 7.1: Comparison between the developed algorithms and the methods presented in the state of the art for epoch based OSA detection.

Work	*Population*	*Signal*	*Acc (%)*	*Sen (%)*	*Spe (%)*	*CO (%)*
Ravelo-Garcia et al. [181]	35*	ECG	-	74	86	
Lázaro et al. [157]	21	SpO2	70	82	69	74
MartÍnez-Vargas et al. [178]	35*	ECG	76	-	-	-
Kesper et al. [180]	35*	ECG	81	-	-	-
Hassan [188]	35*	ECG	84	-	-	-
Ravelo et al. [169]	35*	ECG	84	79	87	83
Mostafa et al. [161]	25	SpO2	85	60	92	79
Ravelo-García et al. [189]	35*	ECG	85	75	91	84
Martín-González et al. [196]	35*	ECG	85	82	87	85
Cheng et al. [191]	35*	ECG	85	83	82	83
Nguyen et al. [185]	35*	ECG	85	86	83	85
Song et al. [193]	35*	ECG	86	83	88	86
Mendez et al. [171]	25*	ECG	86	84	89	86
Hassan [192]	35*	ECG	87	82	91	87
Hassan and Haque [194]	35*	ECG	89	88	91	89
Chazal et al. [168]	35*	ECG	90	89	91	90
Jung et al. [164]	92	SpO2	91	83	89	88
Quiceno-Manrique, et al. [173]	35*	ECG	93	-	-	-
Almazaydeh et al. [156]	8	SpO2	93	88	100	94
Khandoker et al. [176]	30*	ECG	93	90	100	94
Rachim et al. [186]	35*	ECG	94	95	93	94
Pathinarupothi et al. [163]	8	SpO2	96	-	-	-
Mostafa et al. [165]	8	SpO2	98	97	99	98
Travieso et al. [182]	35*	ECG	99	-	-	-
Pathinarupothi et al. [199]	35*	ECG	100	-	-	-
This work – first version of the SpO2 algorithm	35	SpO2	87	67	95	83
This work – second version of the SpO2 algorithm	**70**	**SpO2**	**88**	**80**	**91**	**86**
This work – ECG algorithm	70	ECG	76	73	75	75

* recordings from PhysioNet apnea-ECG database [43]

When comparing with the state of the art works evaluating SPO2 signal, it is notorious that the works which have reported the best performance have only used 8 subjects in the analysis, while the best proposed method examined 70 subjects. The work presented by Jung et al. [164] was the only based on SpO2, which has used a higher population than the proposed method, and reached a better Acc (3% more) although with a lower specificity.

The state of the art methods based on ECG reported the best results. However, it is relevant to notice that all evaluated works which presented ECG methods were tested in the same public databases, which is likely to have cleaner signals that might contribute to increase the algorithm's diagnostic capability [56]. This observation is possibly related to the results attained in this work where the method based on ECG attained a lower performance than the developed methods based on SpO2.

The comparative analysis between the developed algorithms for the global OSA assessment, and the state of the art works (reviewed in section 3.2) is presented in Table 7.2. It is possible to conclude that the proposed methods attained a performance which is in the upper range of the state of the art works while using a higher population than the works which presented a higher Acc. It was also observed that although the model based on the SpO2 signal analysis reached a better performance for the minute-by-minute examination, the global Acc (OSA diagnosis) attained by the models developed for the two source signals (SpO2 and ECG) is almost the same.

Table 7.2: Comparison between the developed algorithms and the methods presented in the state of the art for global OSA assessment.

Work	*Population*	*Signal*	*Acc (%)*
Gutiérrez-Tobal et al. [187]	188	ECG	72
Garde et al. [159]	36	SpO2	85
Marcos et al. [151]	83	SpO2	86
Yılmaz et al. [177]	17	ECG	87
Lázaro et al. [157]	21	SpO2	87
Álvarez et al. [155]	144	SpO2	87
Álvarez et al. [150]	187	SpO2	87
Mendez et al. [172]	25	ECG	88
Marcos et al. [152]	113	SpO2	88
Chazal et al. [168]	35*	ECG	89
Álvarez et al. [154]	148	SpO2	90
Álvarez et al. [160]	127	SpO2	90
Roche et al. [166]	147	ECG	91
Ravelo-Garcia et al. [181]	35*	ECG	93
Khandoker et al. [175]	42	ECG	93
Chen et al. [190]	70*	ECG	93
Marcos et al. [153]	129	SpO2	93
Rachim et al. [186]	35*	ECG	94
Morales et al. [162]	79	SpO2	94
Morillo and Gross [158]	115	SpO2	94
Smruthy and Suchetha [195]	9+	ECG	95
Khandoker et al. [174]	16	ECG	95
Almazaydeh et al. [16]	32*	ECG	97
Song et al. [193]	35*	ECG	97
Martín-González et al. [196]	35*	ECG	97
Jung et al. [164]	92	SpO2	97
Cheng et al. [198]	10*	ECG	98
Chen and Zhang [197]	69*	ECG	98
Khandoker et al. [176]	30*	ECG	100
Yildiz et al. [179]	60*	ECG	100
This work – second version of the SpO2 algorithm	**70**	**SpO2**	**95**
This work – ECG algorithm	70	ECG	94

* recordings from PhysioNet apnea-ECG database [43]
\+ University college of Dublin sleep apnea Database [233]

7.4. Key remarks

Two source sensors were evaluated in this chapter for OSA analysis. It was verified that the model based on SpO2 reached a better performance for the minute-by-minute examination, conceivably because the oscillations in the SpO2 signal are more related to the OSA events. However, the attained global accuracy (OSA diagnosis) by the models developed for the two source sensors is essentially the same. However, the ECG based model is more suitable for the proposed sleep quality model (presented in chapter 8) since the CAP epochs can also be assessed from the ECG signal. Hence, only one sensor is required.

The main limitation of the methods based on the SpO2 signal is the inability of SpO2 to display distinctive patterns during short respiratory pauses. Another factor is that breathing pauses may be produced by pulmonary diseases instead of an OSA event, and the distinction between these occurrences cannot be properly evaluated by examining the SpO2 signal. Regarding the ECG signal examination, based on CPC, the main limitation was the possibility of subjects suffering from heart related disorders to affect the evaluation, and the possibility of occurrence of different influences on the dynamic of the HRV that are not related with OSA. Another relevant aspect was that the number of GL employed by the classifier was optimized for the proposed sleep quality model, which may have led to lower performance.

8. Proposed sleep quality model based on the ECG signal examination

8.1. Development of the model

8.2. Discussion of the results

8.3. Key remarks

A sleep quality model was developed in this chapter using the Keras library in Python 3. ECG was chosen as the source sensor since it is easier to self-assemble, when comparing with the EEG, and the developed models for the sleep quality estimation based on this sensor, presented in chapter 6, attained promising results. The CAP epochs' assessment was employed to define a new sleep quality metric, and the result of this estimation was combined with the OSA diagnosis (using the model presented in section 7.2), and the estimation performed by model which evaluates the average CPC signal (presented in section 6.2) to acquire the final sleep quality estimation. The proposed model predicts the quality of sleep by combining the information of sleep quality metrics and the sleep disorder detection to provide a better view of the global sleep quality, according to the third approach for sleep quality examination, theorized in this work (from the state of the art analysis).

8.1. Development of the model

The previous methods developed for sleep quality assessment, based on the ECG signal, required the assessments of the NREM periods to estimate the CAP rate. However, a simpler and possibly more accurate approach would be the employment of the same methodology used to evaluate the m-AHI-tib. Thus, a new sleep quality metric can be created by comparing the ratio [299]

$$\mathrm{m-CAP-tib} = \frac{\text{minutes classified as CAP}}{\text{time in bed in minutes}} \tag{8.1}$$

named m-CAP-tib, with a tuned threshold. To evaluate the performance of this metric, the CPC signal fed a 1D-CNN to estimate the CAP epochs.

It was conceptualized in this work that an improved sleep quality estimation can be attained by combining the assessment of both sleep disorders and sleep quality metrics. Therefore, two more methods were chosen to combine their assessment with the sleep quality estimation from the m-CAP-tib. Specifically, the global estimation of the sleep quality by considering the average of the CPC signal (presented in section 6.2; this method attained the best complexity to performance ratio), and the model based on CPC for OSA diagnosis (disorder that is strongly correlated to the occurrence of poor sleep quality, whose developed method was presented in section 7.2), were selected. The developed model is presented in Figure 8.1.

The algorithm has two main steps. The first creates the CPC signal from the N-N series and EDR signal by employing the method previously described, applied to the pre-processed (resampling and normalizing) single-lead ECG signal. The second step performs the examination of the CPC signal by the three 1D-CNN whose outputs were combined to produce the global estimation of the sleep quality (named SQ-g). The CNN was chosen for this model since it was acknowledged as one of the best networks for automatic feature extraction [342].

Two classifiers performed the minute-by-minute estimation, and were fed with the CPC epochs. Each epoch has a duration of 7 minutes with 6 minutes overlapping (identified in section 6.1 as the best window for CAP examination). The combination of all epochs composes a Spectrographic Image (SI), which is a time-frequency matrix

representation, where each line (row of the time-frequency matrix) contain the frequency based information of the examined epoch.

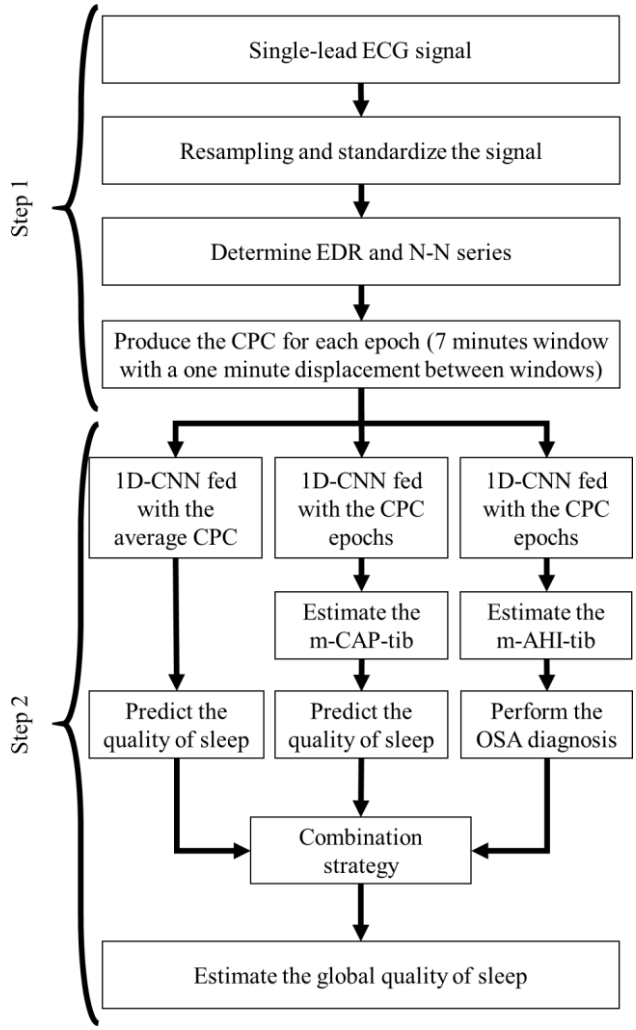


Figure 8.1. Block diagram of the sleep quality model based on the ECG signal examination [299].

One of the classifiers estimated the presence of OSA and another the occurrence of a CAP epoch. This information was then used to compute the m-AHI-tib, and m-CAP-tib, which were compared with thresholds for the OSA diagnosis (method presented in section 7.2) and sleep quality prediction (named SQ-m, using equation 8.1), respectively. The average of the CPC estimation fed the third 1D-CNN to perform the second sleep quality estimation (named SQ-ave). A representation of the SI creation is presented in Figure 8.2.

The sleep quality assessment was either “0” if the determined CAP rate was higher than the CAP rate percentage in healthy subject with the same age group as the patient, designating a poor sleep quality, or “1” otherwise. In the end, the SQ-g was calculated by a majority voting strategy, considering each input as a vote, and the system selects the output class with more votes (input class that was chosen by either two or three models, since three binary classification models were considered) to perform the classifiers ensemble. The developed process for feature creation, classification, and sleep quality assessment is presented in Figure 8.3.

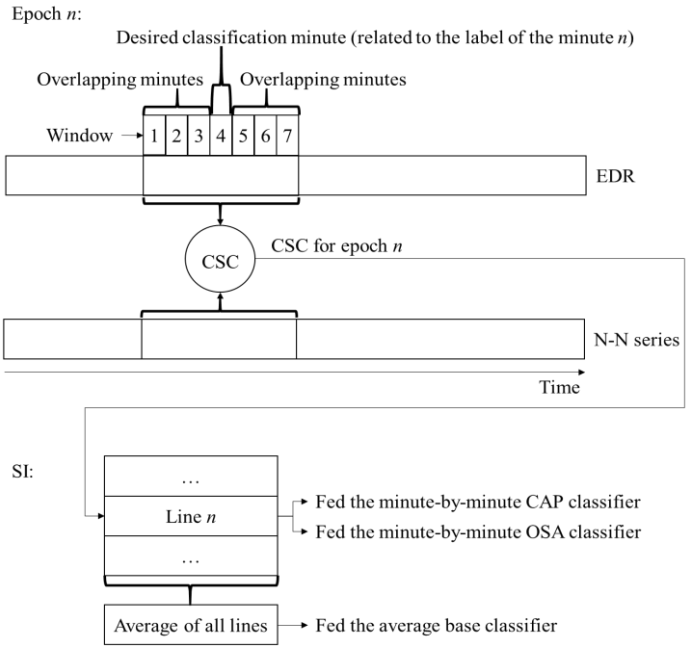


Figure 8.2. Representation of the SI creation [299].

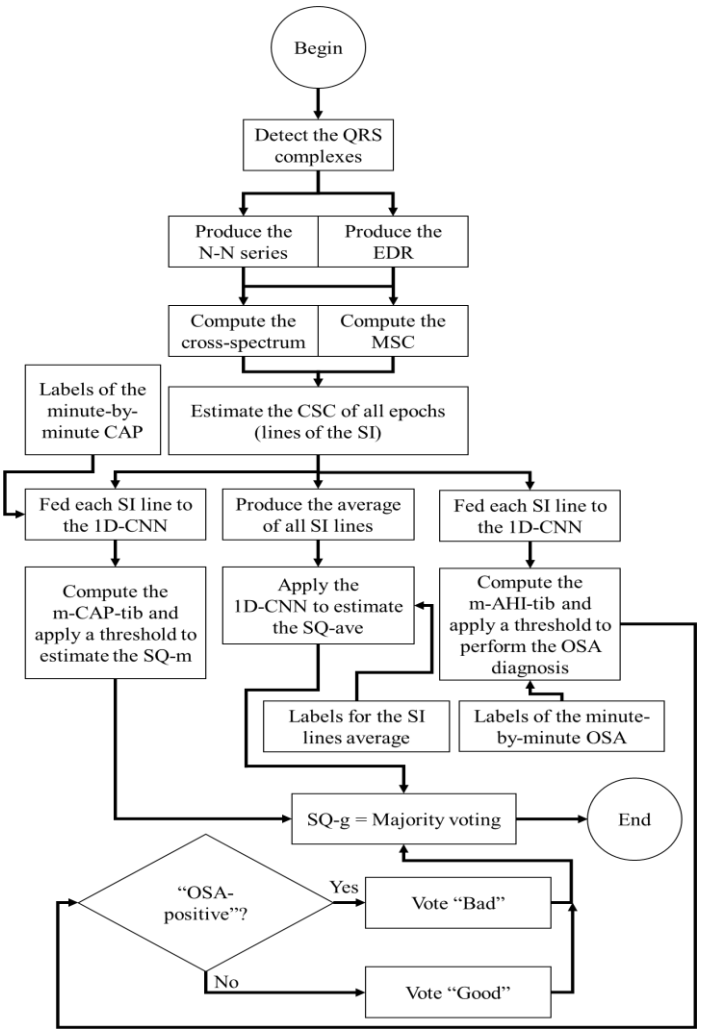


Figure 8.3. Flow diagram of the proposed method for sleep quality estimation from the ECG signal [299].

A grid search method was used to select the 1D-CNN hyperparameters and structure employed for the three classifiers, which compose the proposed sleep quality model. All classifiers started with one GL (composed of one convolution layer, followed by batch normalization, and a pooling layer), and the subsequent GLs were introduced in all classifiers if the Acc-G increased by a minimum value, t_h, of 1% or until the maximum number of GL, G_{max}, chosen to be 4. Consequently, all classifiers have an equal number of layers, leading to a considerable reduction of the simulation time, which would have been necessary to test all possible combinations. The number of filters (K) employed by the first convolution layer varied from 8 to 512. A step size with a power of two was used for optimization; hence, the starting (K_{start}) and maximum (K_{max}) value was 3 and 9, respectively. The filter length, F, varied from 1 (F_{start}) to F_{max}, selected to be 10, in steps (F_{step}) of 1. The examined activation functions were *ReLU* and *SELU*.

The number of filters, of the convolution layers, for the subsequent GL after the first was selected to be either half, the same, or twice the amount used in the previous GL, with the same F. For the batch normalization layer, the employed number of channels was chosen the same as the number of filters used by the convolution layer. Each pooling layer employed a pool size of 2 with equal stride. A stride of 1 was used by all convolution layers. ReLU was used as the activation function, and the network's error optimization was performed by the Adam algorithm [343]. A dense layer could be used between the last GL and the output layer, having a number of neurons which was either half, the same, or twice the F from the previous GL (the pooling layer of the last GL was chosen to be global. Thus, this GL was named GL2). The Output layer employed the soft-max function.

The optimization was performed using a variation of the HOSA algorithm (named HOSA-G) for the developed sleep quality model. The examined data was $Data_{OSA}$, $Data_{CAP}$, $Data_{SQ}$, for minute by minute OSA, minute by minute CAP, and sleep quality (based on the average of CPC window) examination, respectively. The predictions of the optimized classifiers were then combined by majority voting, *MajorityVoting*, to estimate the Acc of the global sleep model. The HOSA-G follows the subsequent pseudo code:

HOSA-G ($Data_{SQ}$, $Data_{CAP}$, $Data_{OSA}$, G_{max}, K_{start}, K_{max}, F_{start}, F_{step}, F_{max}, t_h)
$G = [1, 2, \ldots, G_{max}]$
$K = 2^N$ where $K_{start} \leq N \leq K_{max}$
$F = [F_{start}, F_{start} + F_{step}, \ldots, F_{max}]$
P = [*MaxP*, *AveP*]
A = [*ReLU*, *SELU*]
C = [*SQ*, *CAP*, *OSA*]
for g = 1 to length (G)
  for c = 1 to length (C)
    for k = 1 to length (K)
      for f = 1 to length (F)
        for p = 1 to length (P)
          for a = 1 to length (A)
            for m = 1 to 3
              if c == 1

$Net_{0,c,k,f,p,a,m,0} \leftarrow I_{pt}\ (Data_{SQ})$
else
if $c == 2$
$Net_{0,c,k,f,p,a,m,0} \leftarrow I_{pt}\ (Data_{CAP})$
else
$Net_{0,c,k,f,p,a,m,0} \leftarrow I_{pt}\ (Data_{OSA})$
$K_p = K\ (k)$
for $z = 1$ to g
if $z == g$
$Net_{z,c,k,f,p,a,m,0} \leftarrow Net_{z-1,c,k,f,p,a,m,0} + GL2\ (K_p, F\ (f), P\ (p), A\ (a))$
else
$Net_{z,c,k,f,p,a,m,0} \leftarrow Net_{z-1,c,k,f,p,a,m,0} + GL\ (K_p, F\ (f), P\ (p), A\ (a))$
if $m == 1$
$K_p = floor\ (K_p / 2 + 1 / 2)$
else
if $m == 2$
$K_p = K_p$
else
$K_p = K_p \times 2$
$Net_{g,c,k,f,p,a,m,1} \leftarrow Net_{g,c,k,f,p,a,m,0} + F_C\ (floor\ (K_p / 2 + 1 / 2)) + F_C\ (2)$
$Net_{g,c,k,f,p,a,m,2} \leftarrow Net_{g,c,k,f,p,a,m,0} + F_C\ (K_p) + F_C\ (2)$
$Net_{g,c,k,f,p,a,m,3} \leftarrow Net_{g,c,k,f,p,a,m,0} + F_C\ (K_p \times 2) + F_C\ (2)$
$Net_{g,c,k,f,p,a,m,4} \leftarrow Net_{g,c,k,f,p,a,m,0} + F_C\ (2)$
for $l = 1$ to 4
$\mathrm{AUC}_{g,c,k,f,p,a,m,l} \leftarrow \mathrm{test}\ (\mathrm{train}\ (Net_{g,c,k,f,p,a,m,l}))$
$\mathrm{AUC}_{g,c,k,f,p,a,m,l,max} = max(\mathrm{AUC}_{g,c,k,f,p,a,m,l})|_{for\ all\ k,f,p,a,m,l}$
$BestNet_{SQ} = Net_{g,1,k,f,p,a,m,l}|\mathrm{AUC}_{g,1,k,f,p,a,m,l,max}$
$BestNet_{CAP} = Net_{g,2,k,f,p,a,m,l}|\mathrm{AUC}_{g,2,k,f,p,a,m,l,max}$
$BestNet_{OSA} = Net_{g,3,k,f,p,a,m,l}|\mathrm{AUC}_{g,3,k,f,p,a,m,l,max}$
$Acc_g \leftarrow \mathrm{test}\ (MajorityVoting\ (BestNet_{SQ}, BestNet_{CAP}, BestNet_{OSA}))$
if $g > 1$
if $Acc_g - Acc_{g-1} \leq t_h$
break
return $BestNet_{SQ}$, $BestNet_{CAP}$, $BestNet_{OSA}$

The global sleep quality classifier structure selection and hyperparameter optimization was performed by either TFCV (for the CAP and OSA classifiers) or LOOCV (for the sleep quality prediction based on the average of the CPC signal), using cost-sensitive learning. The layer parameters of all classifiers is presented in Table 8.1. The total number of examined networks was 30240 and each was repeated 10 times. Therefore, the number of simulation was 302400.

Recordings from the fifth CAPSD (15 normal subjects and from the four subjects with sleep-disordered breathing) were examined for the minute-by-minute CAP cycle assessment using the classifier optimized by the HOSA-G, and TFCV was employed for the performance assessment. Each iteration was repeated 50 times to achieve statistically significant results. For the minute-by-minute CAP cycle assessment, the Acc, Sen, Spe, and AUC (average ± standard deviation), were 70.5% ± 2.17%, 73.20% ± 5.41%, 69.8% ± 1.80%, and 0.72 ± 0.04, respectively. An example of the line examination executed by

the classifier, which performed the minute-by-minute CAP assessment is presented in Figure 8.4.

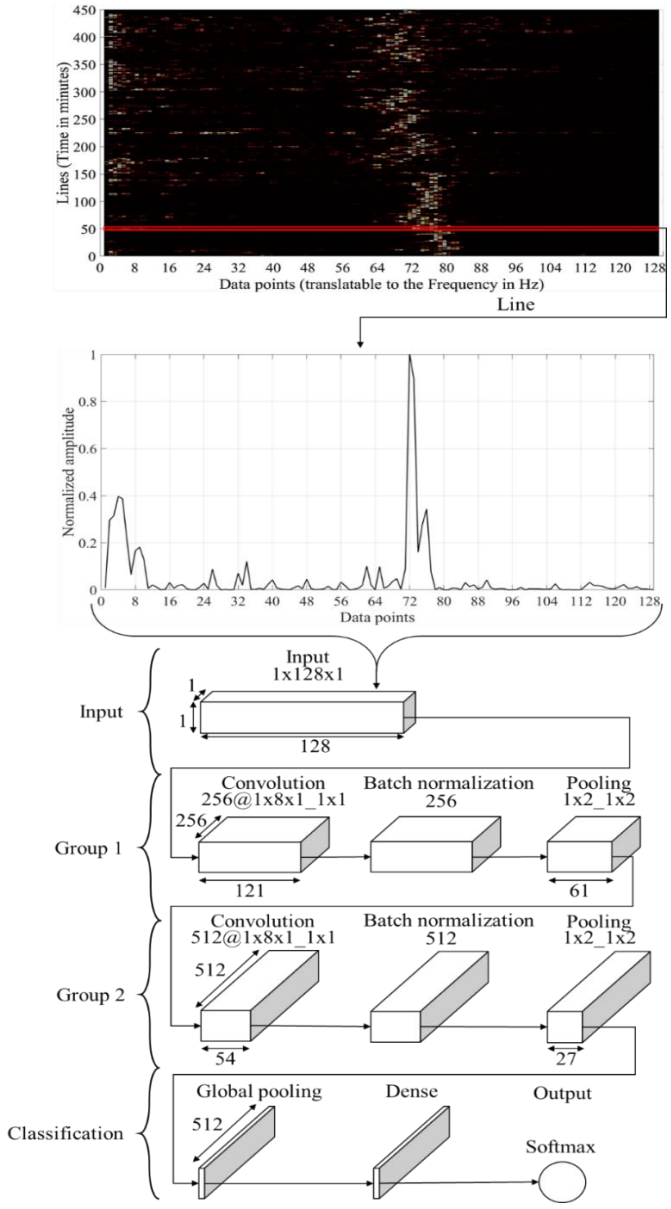


Figure 8.4. Example of the line examination executed by the classifier which performed the minute-by-minute CAP assessment [299].

Table 8.1: Layer parameters of all classifiers which composed the proposed sleep quality model [299].

Layer	***Average sleep quality examination***	***Minute-by-minute – CAP***	***Minute-by-minute – OSA***
Input[1]	1x128x1	1x128x1	1x128x1
Convolution[2]	64@1x4x1_1x1	256@1x8x1_1x1	128@1x5x1_1x1
Batch normalization[3]	64	256	128
Pooling[4]	1x2_1x2	1x2_1x2	1x2_1x2
Convolution[2]	128@1x4x1_1x1	512@1x8x1_1x1	256@1x5x1_1x1
Batch normalization[3]	128	512	256
Pooling[4]	1x2_1x2	1x2_1x2	1x2_1x2
Global pooling	128	512	256
Dense	128	512	256
Dense	2	2	2

[2]number of filters @ vertical width of the filters x horizontal width of the filters x number of channels _ vertical stride x horizontal stride
[3]number of channels
[4]vertical pool size x horizontal pool size _ vertical stride x horizontal stride

The minute-by-minute classifier's predictions (employed to estimate the m-CAP-tib) were then used to assess the SQ-m, considering the age-related CAP rate percentages in healthy subjects [10] as the reference to define the subject's quality of sleep, and using LOOCV for performance assessment (the value of m-CAP-tib of each subject was calculated by averaging the 50 repetitions of LOOCV). It was verified that the highest SQ-g was attained using a threshold of 0.22 to define the SQ-m. The regression plot of the CAP rate obtained by PSG and the predicted m-CAP-tib is presented in Figure 8.5. The regression R^2 was 0.87, further advocating the relevance of the proposed metric. The average performance of the global classification algorithms is presented in Table 8.2, where it is also presented the predicted sleep quality for each subject for each of the three classifiers (the OSA diagnosis indicated the quality of sleep as poor when the subject was classified as suffering from OSA, and as good otherwise).

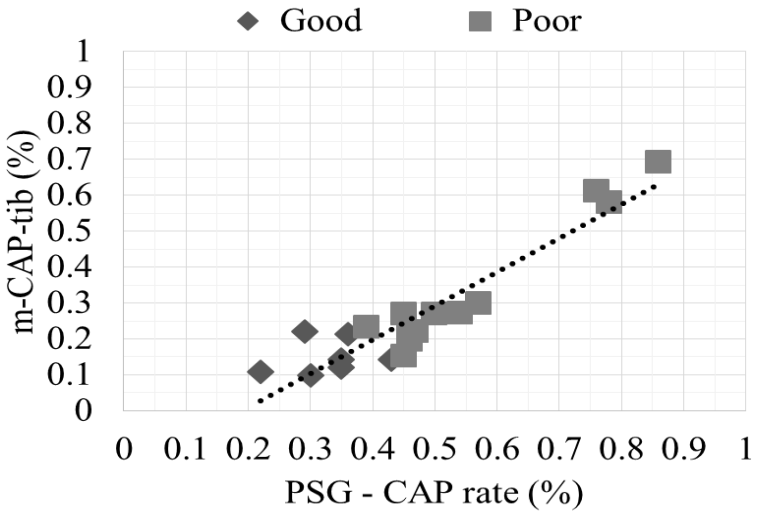


Figure 8.5. Regression plot of the CAP rate acquired by PSG and the estimated m-CAP-tib.

Table 8.2: Sleep quality predictions (good, G, or poor, P) and performance (agreement with the true sleep quality defined by the database labels) of the sleep quality model.

Subject	*SQ-ave*	*SQ-m*	*OSA diagnosis*	*SQ-g*	*m-CAP-tib*	*True CAP rate*	*True sleep quality*
1	G	P	P	P	0.23	0.47	P
2	G	G	G	G	0.21	0.36	G
3	G	G	G	G	0.12	0.35	G
4	G	G	G	P	0.14	0.35	G
5	G	P	P	P	0.27	0.50	P
6	P	P	P	P	0.23	0.57	P
7	P	G	P	P	0.15	0.45	P
8	P	G	P	P	0.20	0.46	P
9	G	G	G	G	0.10	0.30	G
10	G	P	G	G	0.22	0.29	G
11	P	P	G	P	0.23	0.46	P
12	P	G	G	G	0.11	0.22	G
13	G	P	P	P	0.27	0.54	P
14	P	P	G	P	0.23	0.39	P
15	P	P	G	P	0.27	0.45	P
17	P	G	G	G	0.14	0.43	G
18	P	P	P	P	0.61	0.76	P
19	P	P	P	P	0.58	0.78	P
20	P	P	P	P	0.69	0.86	P
Global Acc (%)	73.68	84.21	89.47	100.00	-	-	-

The average normalized measure of CPC for all subjects diagnosed as having a good or a poor sleep quality, by the classifier performing the SQ-ave estimation, is presented in Figure 8.6. It is possible to observe that the classifier indicated the sleep quality as poor when most of the power of the CPC estimation was in the LF band while the opposite occurred when most of the power was in the HF band. The classifier performing the OSA diagnosis detected a lower CAP rate as a negative diagnosis (indicative of good sleep quality) and the opposite as a positive diagnosis (indicative of poor sleep quality).

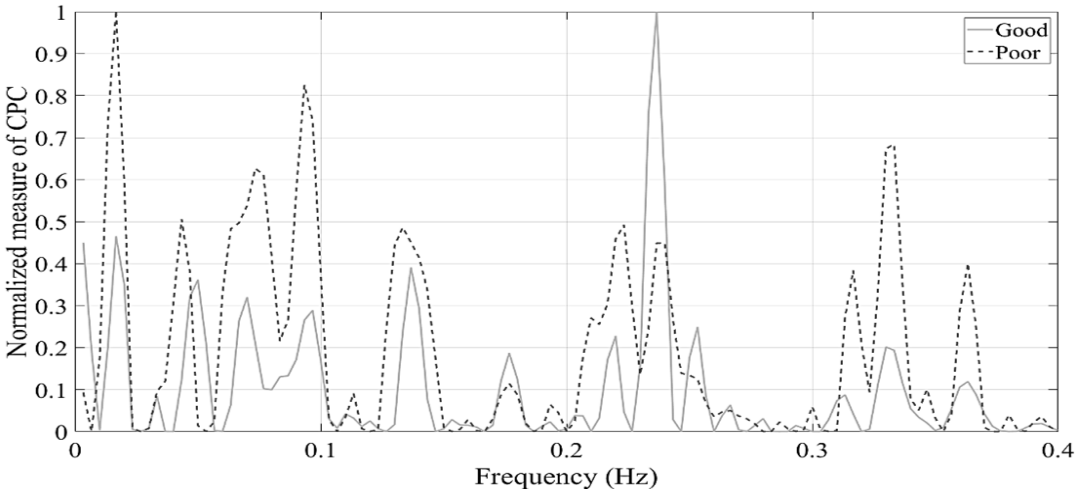


Figure 8.6. Average normalized measure of CPC for all subjects diagnosed as having a good or a poor sleep quality by the classifier performing the SQ-ave estimation.

8.2. Discussion of the results

Several methods were proposed in the state of the art for sleep quality examination, identified in the performed review [55]. A summary of the works that reported the global accuracy is presented in Table 8.3.

Table 8.3: Comparison of the result attained by the proposed sleep quality model with the works presented in the state of the art [299].

Work	***Description***	***Global Acc (%)***
[70]	Clustered sound events classified by an HMM	70
[89]	Set of rules that analyzed attributes	73
[82]	Features from EEG to fed a GELM	76
[71]	Features from EDR and HRV to fed a SVM	78
[76]	Actigraphy signal fed to a CNN	93
This work – method for sleep quality examination based on EEG	CAP rate estimation by a FSM, whose A phases of the CAP cycles were classified by a 1D-CNN	74
	CAP rate estimation by a FSM, whose A phases of the CAP cycles were classified by a LSTM	79
	CAP rate estimation by a FSM, whose A phases of the CAP cycles were classified by a LSTM, fed with features	**90**
This work – prediction from SQ-ave	Classification of the average CPC signal by a 1D-CNN	74
This work – prediction from SQ-m	Classification of the CAP epochs by a 1D-CNN to estimate the m-CAP-tib	84
This work – prediction from SQ-G	**Combinatory scheme that examined the output of three 1D-CNN**	**100**

Conventional sleep metrics were evaluated by Bsoul et al. [71], Sathyanarayana et al. [76], and Wang et al. [82]. The first work examined features from both time and frequency domains, extracted from HRV and EDR. The features were fed a SVM (with a Gaussian kernel) to estimate the deep sleep efficiency. The second work estimated the sleep

efficiency by feeding actigraphy signals to a CNN, while the third work examined features from PSD to feed a Graph regularized Extreme Learning Machine (GELM) with the goal of assessing the TST.

On the other hand, Wu et al. [70] and Choi et al. [89] proposed new methodologies for sleep quality estimation. The first work evaluated clustered sounds related to sleep events, using a SOM, and the output fed a multinomial HMM to predict the quality of sleep, while the second work proposed a set of rules which examined 19 attributes to quantify the quality of sleep.

By examining Table 8.3, it is possible to verify that the developed work attained the best performance, correctly predicting the sleep quality of all subjects, significantly improving the results attained individually by the sleep quality models (SQ-ave and SQ-m). Thus, validating the significance of the ensemble for this classification. Another relevant aspect was that the best proposed EEG based method's performance was 10% lower than the proposed ECG model, further emphasizing the relevance of this model.

8.3. Key remarks

It was verified that the sleep quality model developed in this chapter attained the best possible performance by correctly classifying the sleep quality of all tested subjects. This result was accomplished by combining three classifications from the CPC signal that considered the proposed sleep quality metric, the average CPC signal and the OSA diagnosis. Therefore, these results support the hypothesis considered for this work that an improved sleep quality estimation can be attained by combining the assessment of both sleep disorders and sleep quality metrics. The developed model is scalable as the combination strategy allows for more classifiers to be considered. Therefore, the proposed method can include the detection of more sleep disorders and/or other sleep quality metrics (estimated by new classification procedures). As a result, it can be a relevant tool for the future of sleep quality analysis at the patient's home. It is also relevant to notice that the proposed tools can be of support for physiological characterization if some specific and controlled databases are designed.

The main limitation of the proposed method is the possibility of subjects suffering from heart related disorders to affect the evaluation. Another limitation was that the threshold used to classify the sleep quality based on the m-CAP-tib was selected in a way that maximized the SQ-G. This approach may lead to difficulties in generalizing the method to other populations that are significantly different from the suited subjects. The last limitation was the fact that the examined subjects either suffer from sleep-disordered breathing or were free of sleep related disorders. However, other sleep related disorders may also significantly affect the sleep quality, although they were not considered for this work.

9. Development of the HMDs for sleep analysis

The proposed methods for sleep quality metrics and/or OSA assessment were implemented in HMDs, considering the three exanimated source sensors (EEG, ECG and SpO2). The goal of these devices is to evaluate the viability of hardware implementation of the proposed methods considering one of each of the theorized approaches for sleep quality examination (theorized in this work from the state of the art analysis). Specifically, the assessment based on sleep quality metrics was performed for the devices based on EEG or ECG, the evaluation based on the presence of sleep related disorders (OSA was examined in this work since it is one of the most prevalent and frequently undiagnosed sleep related disorder) was performed for the two devices based on SpO2, and the assessment based on the combination of estimated sleep quality metrics and the detection of sleep related disorders (to provide a better view of the global sleep quality) was performed for the second device based on ECG.

It is intended to develop HMDs as a proof of concept for the feasibility of hardware implementation of the developed methods. The first approach was the development of two devices that can perform the OSA analysis, using pulse oximetry (SCOPER characterization: O_1). The first can be used in multiple configurations, and was directed for clinical usage, while the second was simpler to operate, and was orientated for the personal consumer market. The second step was the development of two devices for sleep quality estimation, the first evaluated the EEG signal (SCOPER characterization: S_2) while the second examined the ECG signal (SCOPER characterization: C_3). The final device was developed to assess both sleep quality and OSA (minute-by-minute and global examination) using the single-lead ECG signal (SCOPER characterization: C_3).

9.1. HMD for OSA detection

9.1.1. First HMD for OSA detection

The first HMD implemented the first version of the algorithm for OSA detection (presented in section 7.1), and had two main units (sensing unit and a processing unit) that wirelessly communicate, as presented in Figure 9.1. The first unit was composed of the sensor (Contec cms50d+ pulse oximeter), the protocol transcription and error detection module, implemented in a Field-Programmable Gate Array (FPGA), and a Bluetooth transmitter. The second unit received the data in a Bluetooth receiver (in this case, a Bluetooth dongle) and employed a processing module to analyze the data. The block diagram of the employed hardware, and the HMD are presented in Figures 9.2 and 9.3, respectively.

The pulse oximeter acquired the SpO2 and heart rate signals with a sampling rate of 60 Hz, and a resolution of 8 bit. These signals were sent to the FPGA, Avnet Spartan-6 LX9 MicroBoard (3.5 cm by 7.5 cm) for checking errors and reshaping them to the Universal Asynchronous Receiver/Transmitter (UART) protocol that was used by the Bluetooth transmitter, the Bluetooth Bee. This device sent the data, using the Bluetooth 2.0 protocol in the 2.4 GHz frequency band with a baud rate of 19200 b/s, to the processing unit. The sensing unit was fed by a 5 V (nominal voltage) battery, with a nominal capacity of 2500 mAh. Taking into consideration that the average current consumption of the unit was 107 mA, the fully charged battery was enough to last at least 13 h.

The FPGA was programed to implement two FSM, each controlled by a state variable, in Very high speed integrated circuits Hardware Description Language (VHDL). Figure 9.4 presents the VHDL algorithm flowchart. The first FSM decoded the words received from the pulse oximeter (the protocol was composed of groups of five words that were sent 60 times per second) to extract the SpO2 and heart rate signals, checked the words for errors by analyzing the parity bit, and if no error was detected the words were stored in a buffer. The second FSM sent the information which was stored in the buffer, using the UART protocol, to the Bluetooth Bee.

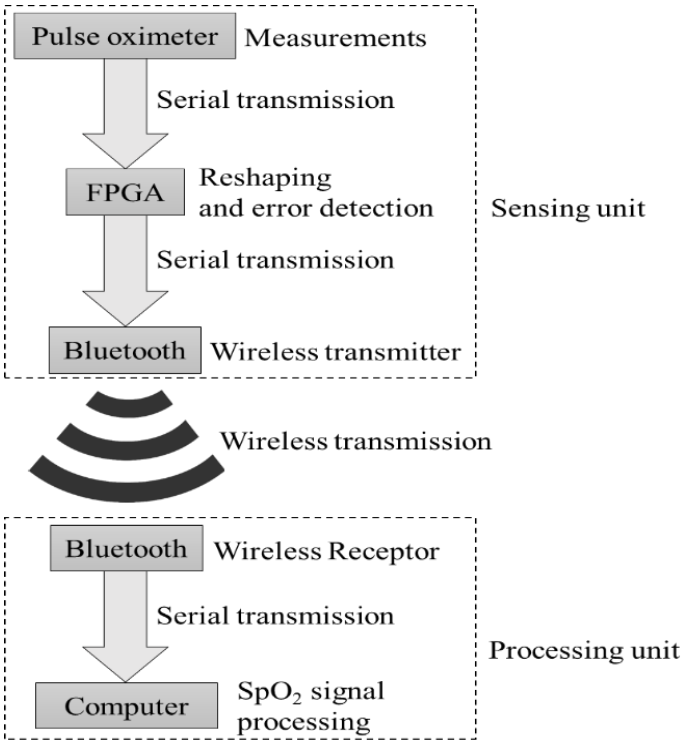


Figure 9.1. System architecture of the HMD for apnea detection (first version) [402].

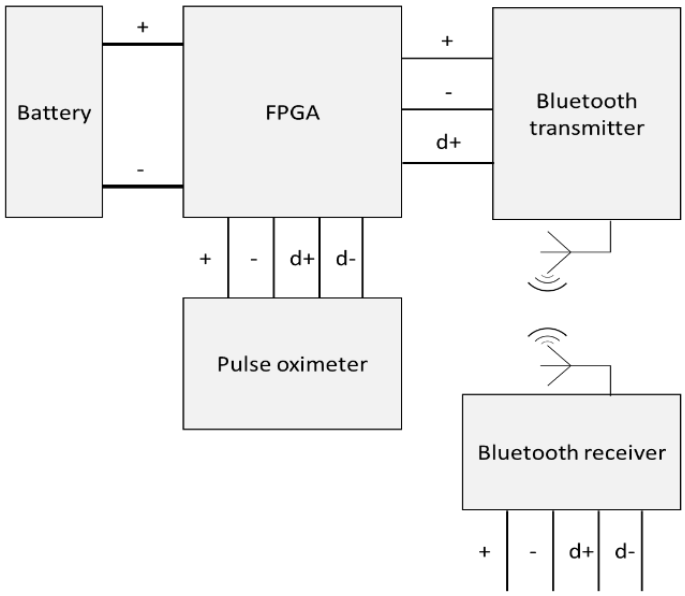


Figure 9.2. Block diagram of the HMD hardware for apnea detection (first version) [402].

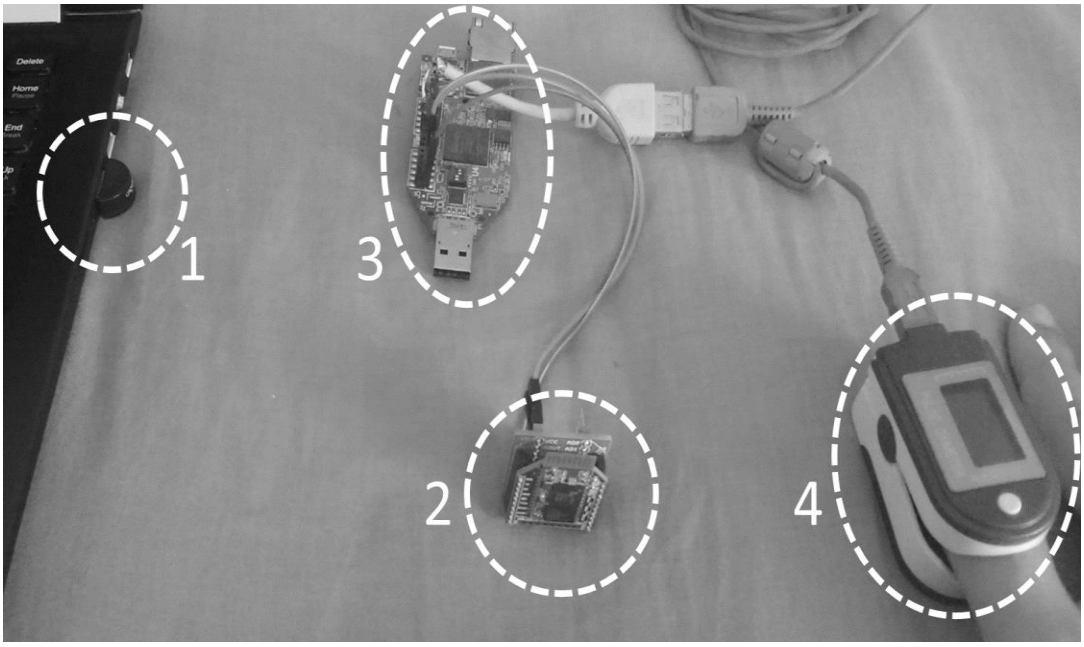


Figure 9.3. HMD hardware for apnea detection (first version). 1-bluetooth dongle; 2-bluetooth transmitter; 3-FPGA; 4-pulse oximeter [402].

It was verified that the average current consumption of the FPGA, 75 mA, was 25 mA higher than the Texas Instruments MSP430 (a commonly used microcontroller for low power consumption applications). However, the FPGA allows to easily implement the serial communication protocol of the pulse oximetry, a task that was considerably harder using a microcontroller. Therefore, the FPGA was chosen.

A Bluetooth dongle receives the data on the processing unit and sends it to a computer, using the universal serial bus protocol, to be analyzed by the developed algorithm previously examined (second version of the OSA detection based on SpO2). The flowchart of the implementation is presented in Figure 9.5.

The algorithm first stores each valid received word in a file to be analyzed. If the finger was removed from the sensor, all received words had the value 0; thus, considered not valid. If this happens for more than 30 s the test session is considered to be finished and the data analysis begins. Every five-minute epoch was analyzed to produce the feature vector that was combined with the weights vector to determine the OSA probability. At the end of each epoch analysis a threshold was applied to perform the binary classification, 0 for OSA negative and 1 for OSA positive.

A Graphical User Interface (GUI) was developed in Python to allow the user to: test and configure and the connection; start the exam; visualize the results in a graph, which displays the hearth rate and SpO2 variations as presented in Figure 9.6, and a text box with indication of the total number of detected apneas. The results were stored in a text file.

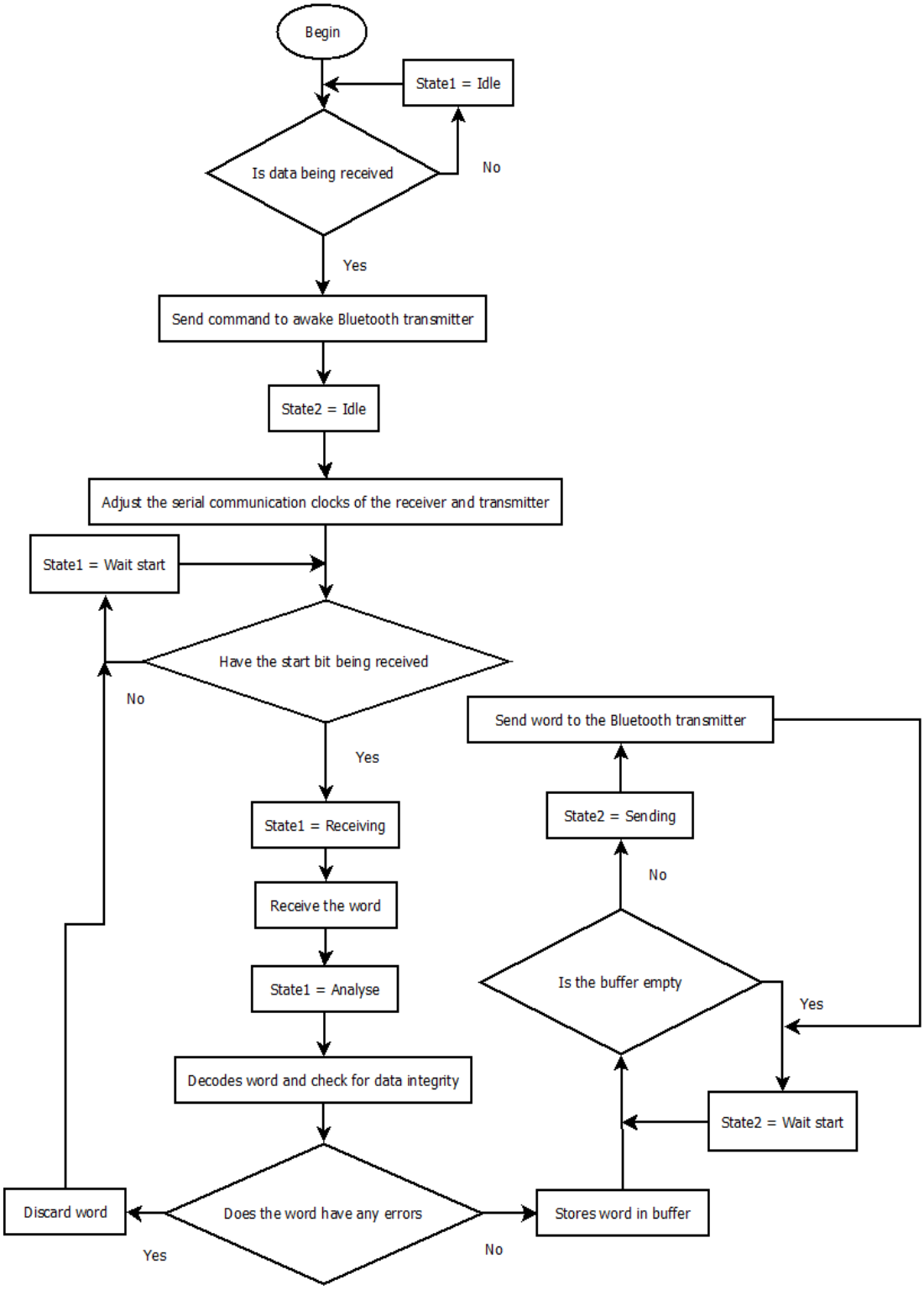


Figure 9.4. VHDL algorithm flowchart (first version) [402].

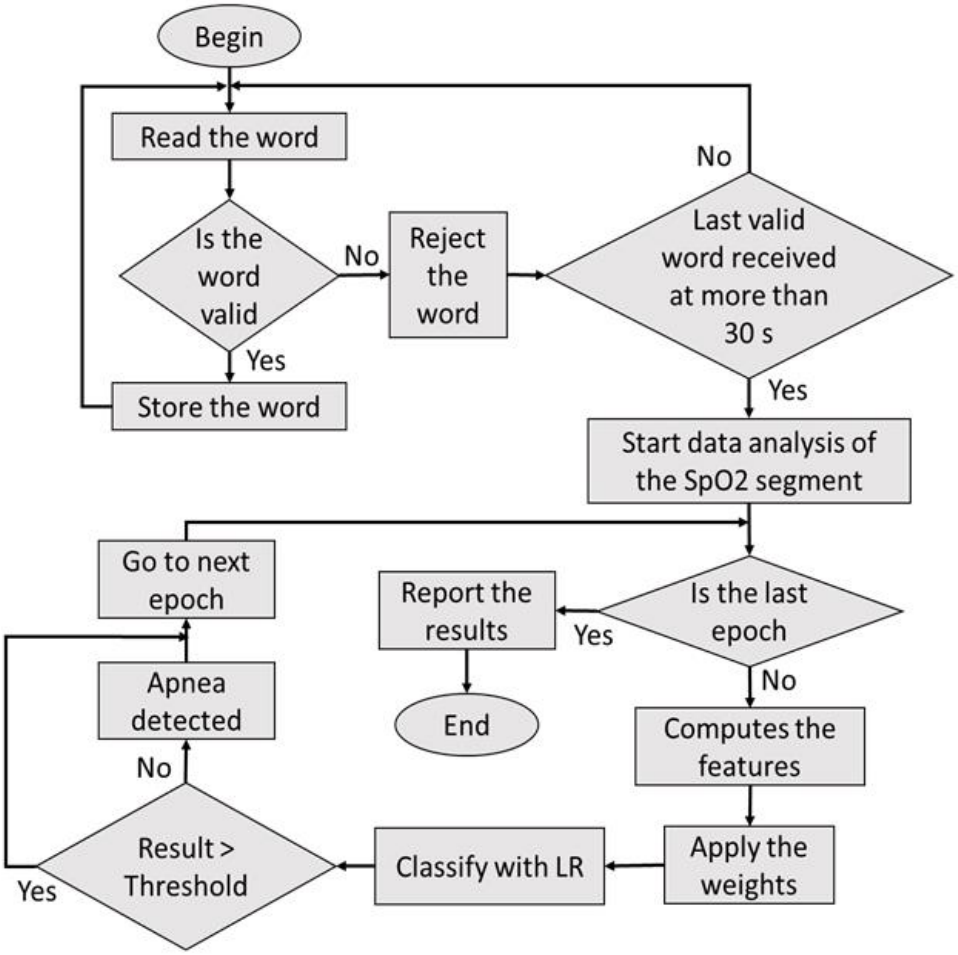


Figure 9.5. Flowchart of the Python algorithm [403].

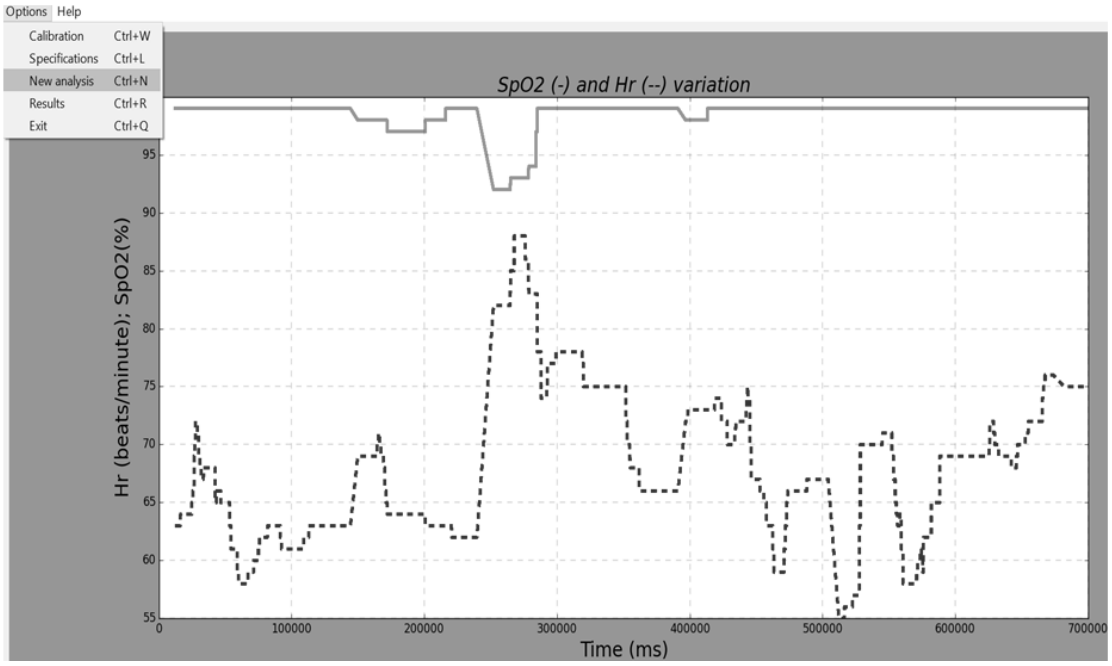


Figure 9.6. GUI of the HMD for OSA detection [402].

The HMD architecture was improved to include two new units (routing and portable processing), as presented in Figure 9.7. Four possible configurations were defined,

according to the utilization purpose, with the description presented in Table 9.1. The second wireless transmitter (ESP8266, operating in the station mode) of the sensing unit sends the information by Wi-Fi to a router in the routing unit, and the information can either go to the portable processing unit (in this case a smartphone) or to a server, through the internet, to be analyzed. Therefore, a third FSM was introduced in the VHDL algorithm, and the updated version of the flowchart is presented in Figure 9.8.

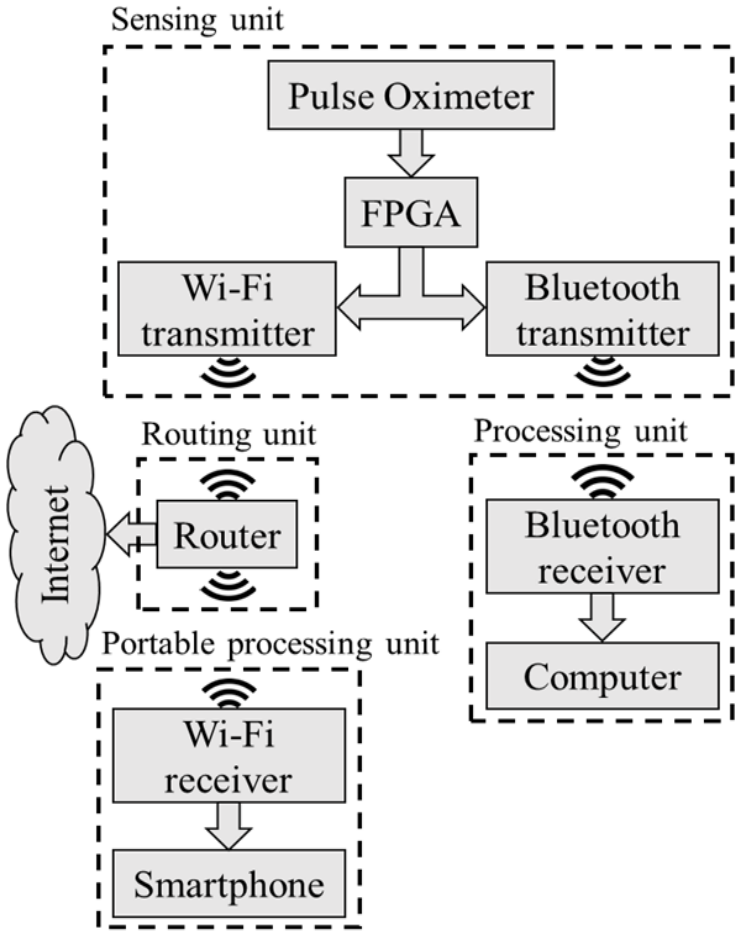


Figure 9.7. System architecture of the HMD for apnea detection (second version) [403].

Table 9.1: Configuration of the HMD for apnea detection (second version) [403].

Configuration number	*Description*	*Purpose*	*Architecture*	*Example of application*
1	Clinical diagnosis	Produce a report at the end of the test	Sensing unit sends data to the processing unit	Subject performing the test at home
2	Clinical diagnosis with real time monitoring	Display the signals and the number of apneas in real time, producing a report at the end of the test	Sensing unit sends data to the processing unit and to the portable processing unit, passing by the routing unit	Disabled subject performing the test, monitored by a person
3	Real time monitoring	Present, in real time basis, the measured signals and the number of apneas detected	Sensing unit sends data to the routing unit which in turn sends it to the portable processing unit	Self-analysis of the subject in a daily basis
4	Data acquisition	Send the measured signals to a server	Sensing unit sends data to the routing unit which in turn sends it to a server	Data sent to server to be stored and analyzed

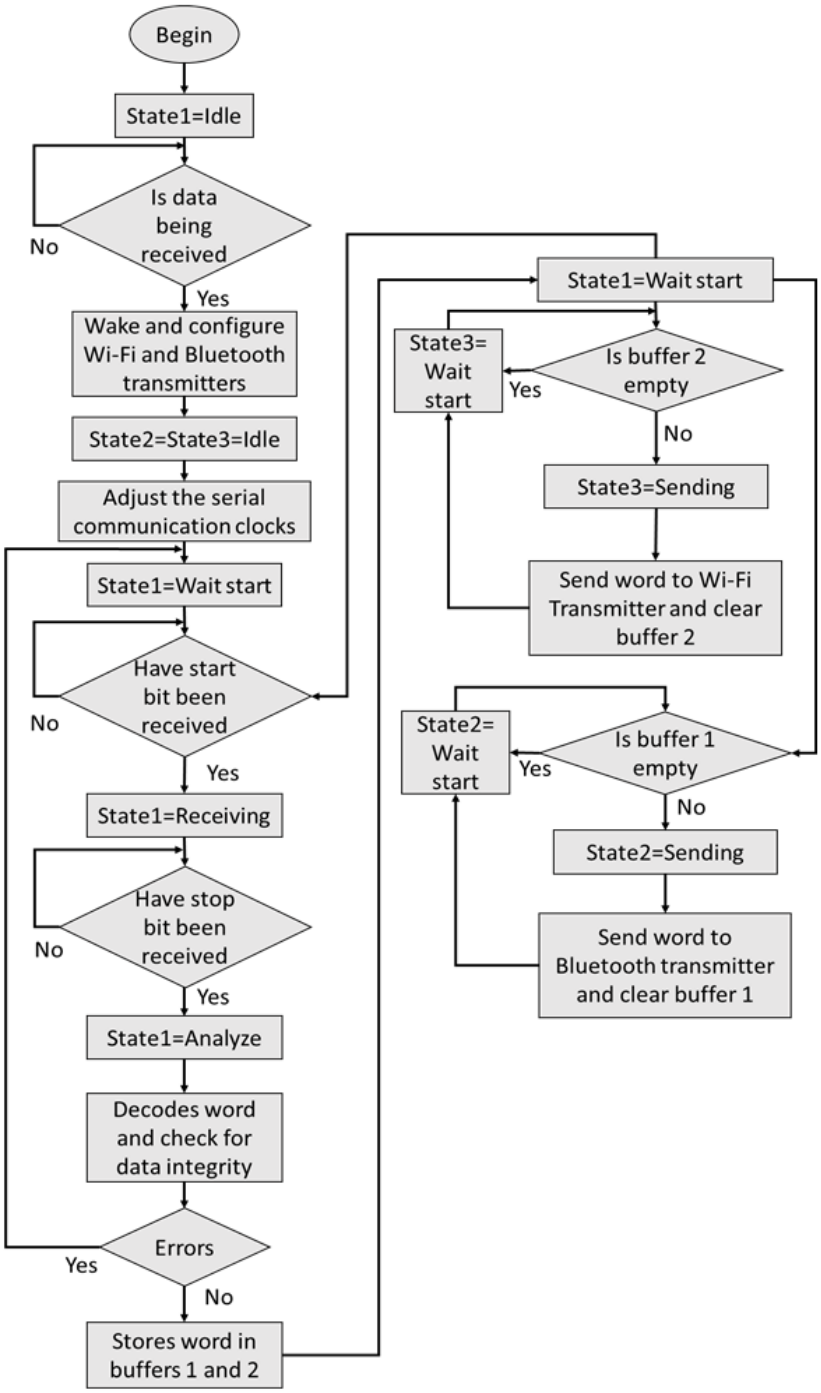


Figure 9.8. VHDL algorithm flowchart (second version) [403].

A Java application was developed to perform the analysis of the data in near-real time (classifies each minute as either apnea or non-apnea) and was implemented in the portable processing unit (smartphone with Android operating system). The algorithm flowchart is presented in Figure 9.9.

When the user opens the application, the smartphone's Wi-Fi is activated, and the user can connect to a configured router network. The GUI of the application presented the heart rate and SpO2 variation on a plot in real time, and a 3 s average of the signals in text boxes. Figure 9.10 displays the GUI in the absence and presence of apnea.

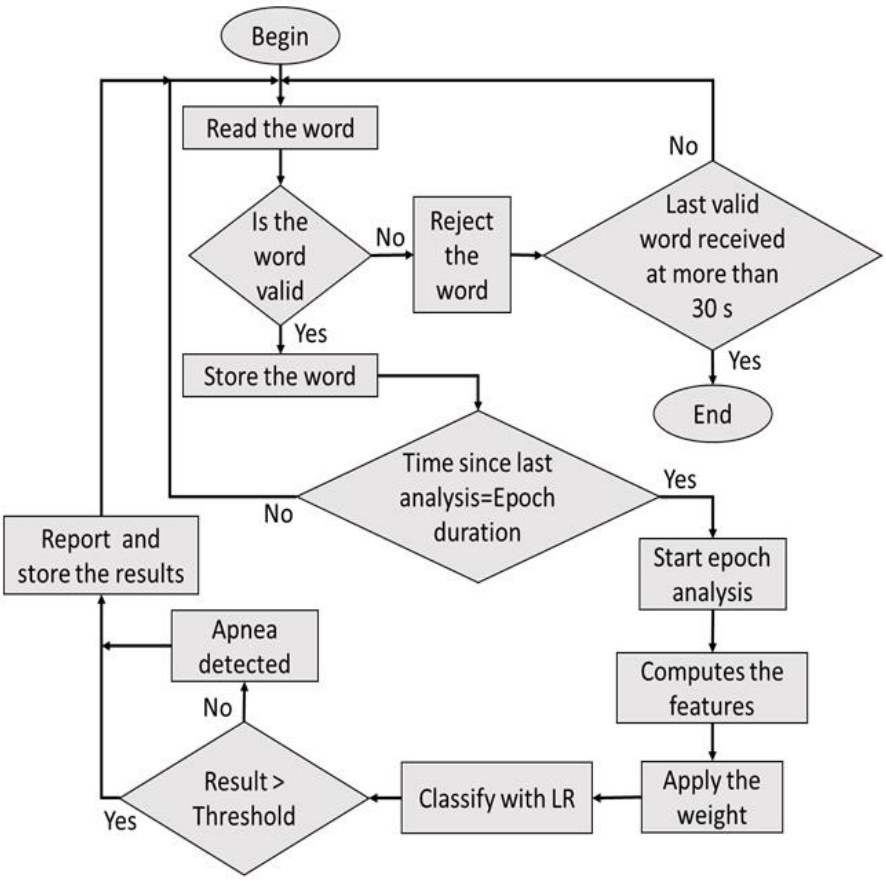


Figure 9.9. Flowchart of the Java algorithm [403].

a) b)

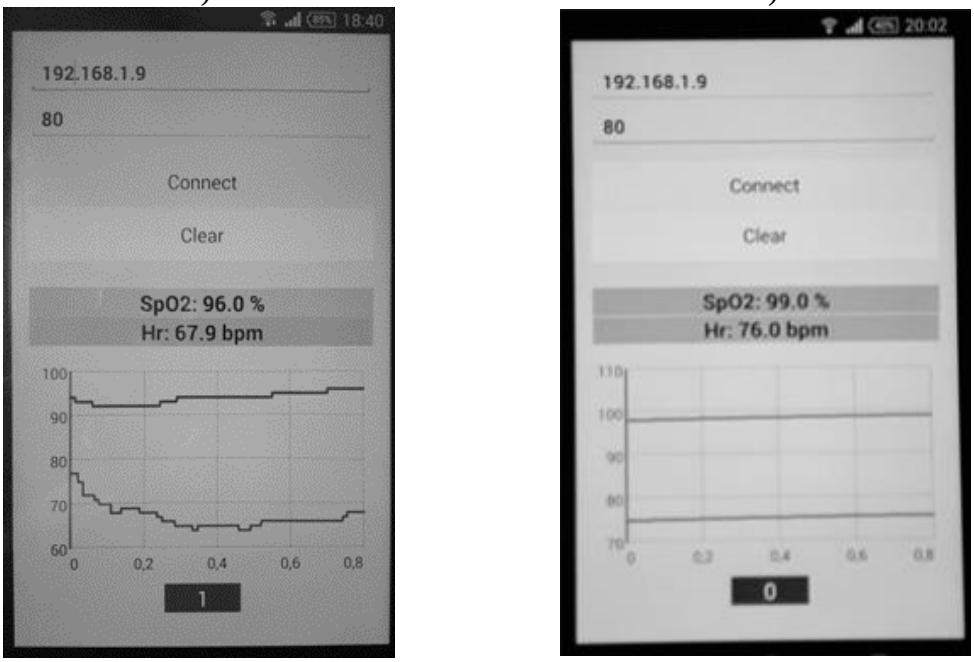


Figure 9.10. GUI of the Java application displaying the results of a test a) with an apnea event and b) with normal breathing [403].

It was verified that the average current consumption, of the sensing unit, is dependent upon the selected configuration, and is 0.22 A when both transmitters are used. When

only one of the wireless transmitter were used, the average current consumption was either 0.19 A (only Wi-Fi) or 0.11 A (only Bluetooth). The same 5 V, nominal voltage, battery was employed. Thus, the battery lasts at least 9 h (in the configuration where both transmitters were used) when fully charged. The average current consumption of the FPGA remained the same, and the implementation used roughly 10% of the accessible on-chip logic.

The smartphone's application allowed the development of a low cost solution (considering the cost of a basic smartphone), for personal analysis, while the computer analysis offers a more robust solution for medical diagnosis. The total cost of the device was 915 € (165 € for the pulse oximeter; 82 € for the FPGA; 7 € for the Wi-Fi transmitter; 36 € for the Bluetooth transmitter; 400 € for the processing unit; 25€ for the router; 200 € for the smartphone).

9.1.2. Second HMD for OSA detection

The HMD developed in the previous section was versatile and adjustable to multiple scenarios. However, an improved user friendly solution can be attained by employing a touch screen for user interaction, and simplifying the device's design to facilitate the manipulation and assembly. Such improvements can be significant for the older adults with physical disabilities.

Therefore, a second OSA detection device, based on the SpO2 signal examination, was developed to be easier to handle by the user, and implemented the second version of the algorithm for OSA detection (presented in section 7.1). The device can provide both a minute-by-minute and a global diagnosis (by applying a threshold to the estimated m-AHI-tib) of the disorder. The developed architecture is presented in Figure 9.11. The employed hardware is presented in Figure 9.12, and the device is composed of two units that wirelessly communicate via Bluetooth.

The BITalino Core BT [404] was used in the sensing unit to collect the signals measured by the pulse oximeter (Contec cms50d+), and it is composed of an ATmega328P microcontroller, a power management block, and a Bluetooth communication block. This unit was feed by a 3.7 V lithium ion battery, and the average load current was 50 mAh. Hence, the unit lasts 17 hours in real-time acquisition over Bluetooth [404]. The sensing unit cost was 240 € (75 € for the BITalino Core, and 165 € for the pulse oximeter). A representation of the assembled pulse oximeter and sensing unit is presented in Figure 9.13.

The processing unit is composed of a single-board computer (Raspberry Pi 3 B+ with a 64-bit, 1.4 GHz, ARM quad-core processor), fed by the Direct Current (DC) power supply, and a touch screen that displays the GUI. The two units automatically establish the Bluetooth connection when the GUI is opened with the default bit rate of 19200 bit/s. The processing unit cost was 60 €. When comparing with the first HMD (previous subsection, 9.1.1), a significant total cost reduction (from 915 € to 300 €) was attained.

The sensing rate of the sensing unit can be chosen through the GUI, and the device supports either 1, 10, 100, or 1 kHz. Nevertheless, the default value of 100 Hz was used for the measurements since it had fewer noise related artifacts than the measurements at

1 kHz. The signal's resolution can be either 6 or 10 bit, depending upon the Analog to Digital Conversion (ADC) port. Conversely, only the 10 bit ports were used.

The procedure for a normal examination can be summarized in the following steps:

1. Introduce the index finger in the pulse oximeter.
2. Attach the armband around the lower arm.
3. Attach the sensing unit to the armband.
4. Connect the sensing and processing units, and wait until the GUI is open.
5. Press "Start Test", and a new window will pop-up with the option of "Stop Test" (the sensing unit will begin the transmission of the SpO2 signal to the processing unit, which in turn will store the data in a text file with a timestamp).
6. Press "Stop Test" to finish the data recording.
7. Press "Analyze Results" and the application will perform the analysis and store the results in a text file.

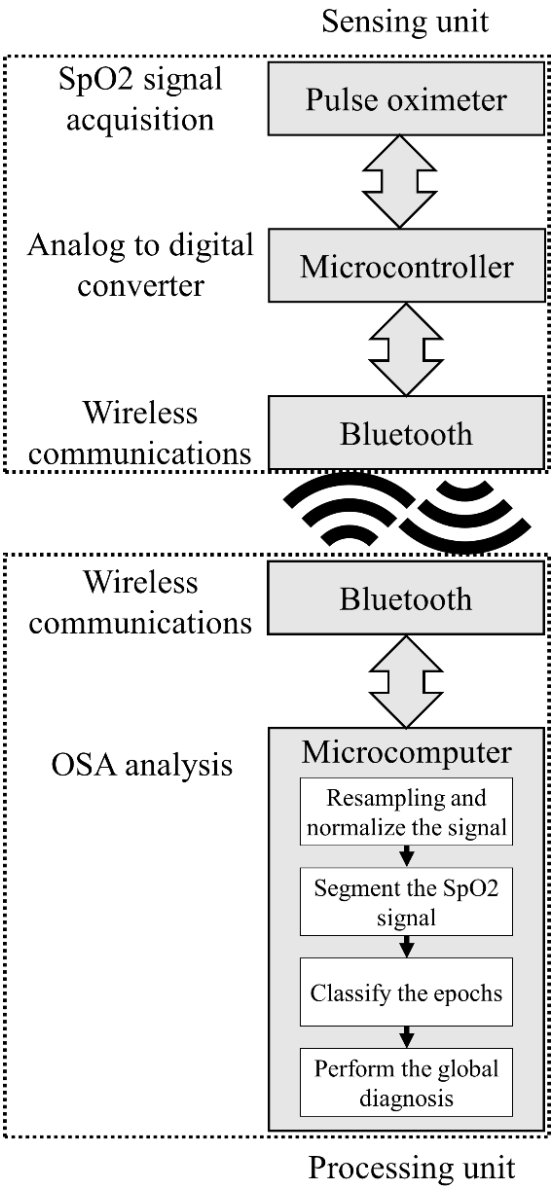


Figure 9.11. Architecture of the second developed HMD for OSA diagnosis [401].

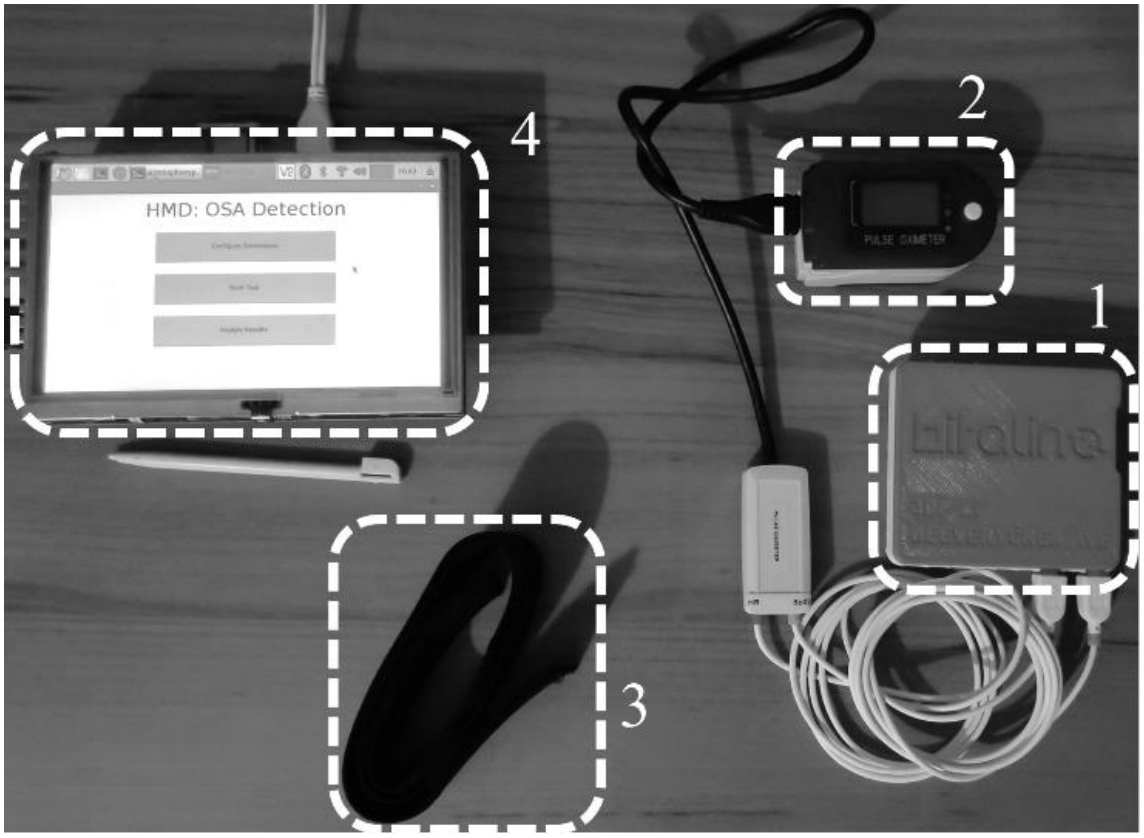


Figure 9.12. Hardware that composes the second HMD developed for OSA examination. 1- Sensing unit, 2- Pulse oximeter, 3- Armband, 4- Processing unit [401].

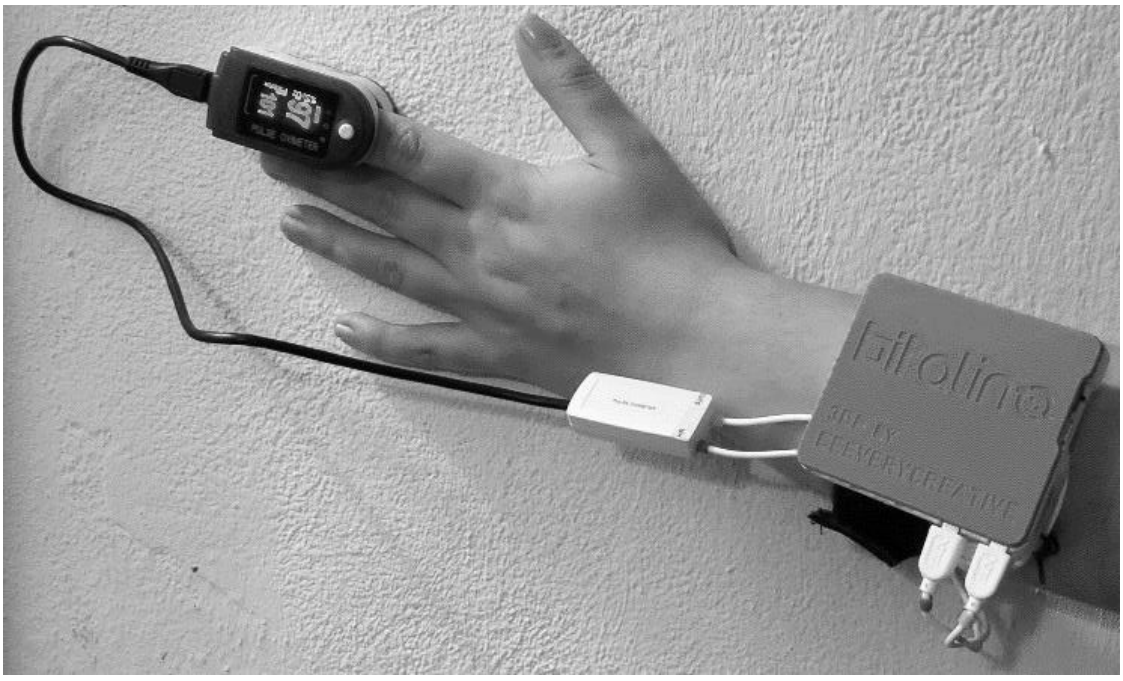


Figure 9.13. Pulse oximeter and sensing unit assembly of the second HMD for OSA examination [401].

9.2. HMD for sleep quality estimation based on the EEG signal

A low cost, and minimally invasive HMD that is capable of performing a minute-by-minute CAP recognition for sleep quality estimation, was developed using a model that examines the EEG signal (from one of the monopolar derivations, either C4-A1 or C3-A2) directly as input (the tested model was based on the improved implementation with the LSTM classifier, presented in section 5.3).

The device's architecture, presented in Figure 9.14, is composed of two units. The processing unit performs the analysis of the signals that were measured by the sensing unit, and the two units wirelessly communicate by Bluetooth. This approach allows the sensing unit to be small, since minimum processing capability is needed, and easily self-assembled, while the processing unit can be larger to accommodate a touchscreen that displays the GUI, presented in Figure 9.15. The user can configure the Bluetooth connection, the sensor sampling frequency, start or stop the exam, and examine the results. The implemented hardware is presented in Figure 9.16.

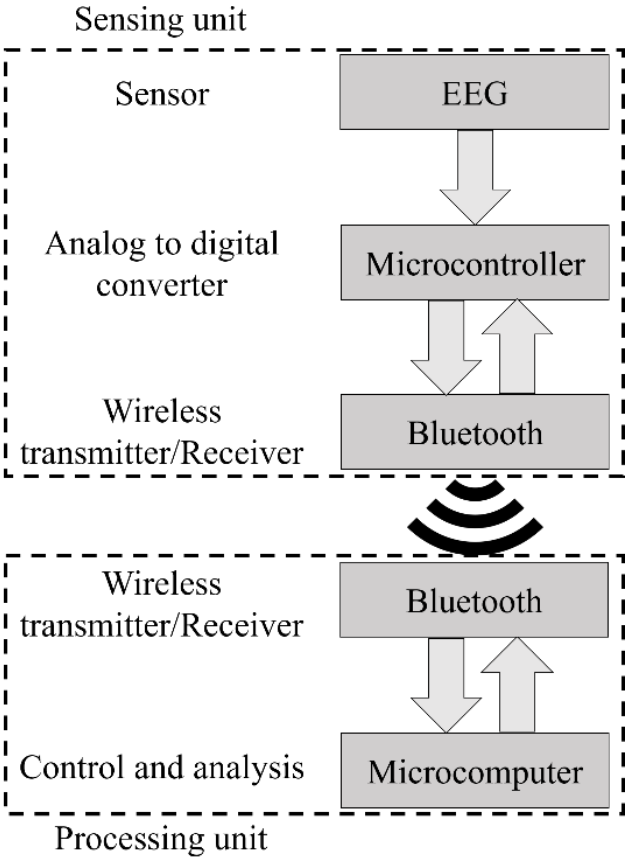


Figure 9.14. Developed architecture of the HMD that performs the sleep quality estimations based on the EEG signal analysis [26].

The BITalino Core BT [404] was used in the sensing unit to collect the signals measured by the EEG sensor (the sensor's cost was 145 €). The default sampling rate (100 Hz) was employed, and the sensor was connected to a 10 bit ADC port. The EEG sensor measures the electrical potentials over the electrode placed in the scalp with respect to a ground reference, measured by the ground cable. Through a comparison with an established gold standard device (BioPac MP35 Student Lab Pro), it was verified that the

average root mean squared error of the measurement was 0.013±0.005 [405]. The minor error substantiates the viability of the sensor for medical diagnosis.

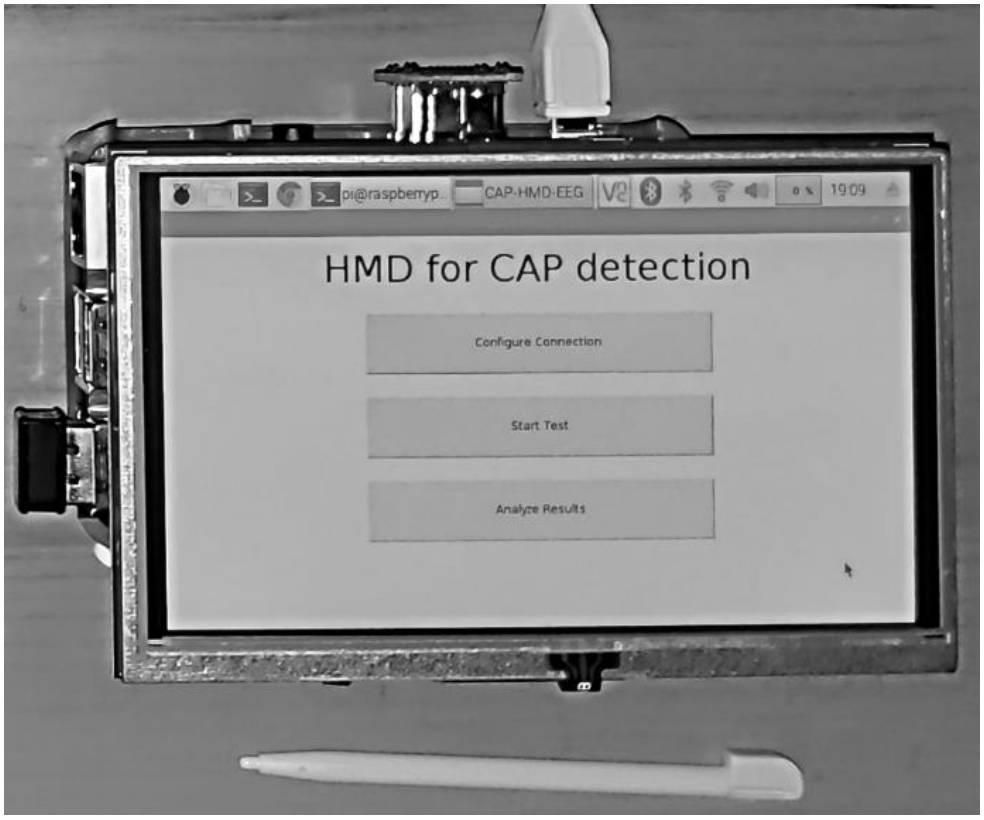


Figure 9.15. GUI of the HMD based on the EEG signal analysis [26].

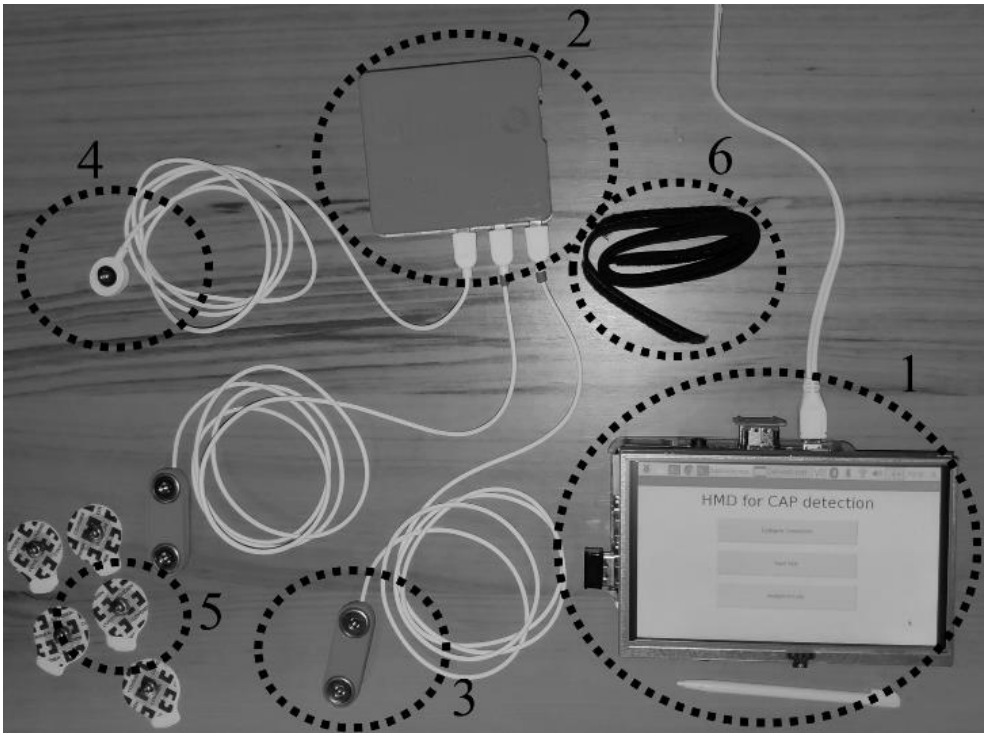


Figure 9.16. Implemented hardware: 1. Processing unit; 2. Sensing unit; 3. EEG sensor; 4. Ground cable; 5. Electrode; 6. Headband [26].

A single-board computer (Raspberry Pi 3 B+) was used on the processing unit, and it was fed by the DC power supply. A touch screen displays the GUI that allows the user to interact with the developed Python application, and it automatically connects with the sensing unit once the application is opened. The user can specify a new bit rate (the default value is 19200 bits/s) or change the ADCs that will be used on the GUI. Nevertheless, the usual examination only requires the user to follow the following steps:

1. Position the EEG sensor in the C4 or C3 positions according to the 10-20 international system (the second EEG sensor, that is connected to another ADC, can be positioned in another location of interest, such as Fp1, and the measured signal will be stored).
2. Tight the headband around the sensor to avoid the introduction of measurement noise.
3. Position the ground electrode in the mastoid region, such as A1 if the sensor was positioned on C4 or A2 otherwise.
4. Activate the sensing unit.
5. Activate the processing unit, wait for the application to open, and press on "Start Test" (a new window will be displayed with the "Stop Test" button).
6. Press on "Stop Test" when the test is finished, the measured signals were stored in a text file with a timestamp on the secure digital memory card of the processing unit.
7. Press on "Analyze Results" (the application performs the CAP classification) and wait until a message indicating that the examination was finished is displayed; hence, the classified CAP phases and cycles were stored, with a timestamp, in a text file.
8. The user can either examine the results file or deliver the device to an expert that can retrieve the file and perform the analysis of the results.

On a trial test, it was substantiated that the HMD can be effortlessly self-assembled, it is easy to operate, can properly record the EEG signal, and perform the minute-by-minute CAP classification.

9.3. HMD for sleep quality estimation based on the ECG signal

A sturdy, easy to operate, and fault resilient platform was employed to develop the HMD for sleep quality estimation based on the ECG signal evaluation (implementing the algorithm which performs the assessment from the average CPC signal, presented in section 6.2, since this method was found to have the best performance to complexity ratio). Therefore, an FPGA board with a processor was used to produce the HMD as it complies with the sturdy and fault resilient requisites [406], and has a good performance for a method with low memory requirements. The goal of this device is to be suitable to be used by the elderly population. Therefore, it is only composed of one unit, and operated by two switches (one to initiate the application, and the other start/stop the examination), providing an easy to operate platform.

The device is composed of an FPGA board (Xilinx PYNQ-Z2 board) with a processor, an ECG sensor, and a display. This board is based on the Zynq XC7Z020 system on a chip, with programmable logic comparable to an Artix-7 FPGA, and a processing system

composed of a 650MHz dual-core Cortex-A9 processor. In addition to that, the board implements a novel approach to the use of FPGAs where the device is programmed in Python, whose functions can be accelerated through hardware libraries (named overlays), which are developed in the programmable logic (programed in VHDL). Therefore, the applications developed for the processing system can use the programmable logic resources.

The single-lead ECG sensor's instrument amplifier was the Analog Devices AD8232 (band-width of 0.67 to 40 Hz, with impedance larger than 10 MΩ), chosen because it small, has a low power consumption, is inexpensive, and has a very small measurement error [407]. A liquid crystal display was used to present the message indicated by a low power microcontroller (Arduino Nano Every) that receives the command of what message should be displayed by reading the digital pins, which are connected to the FPGA board. A 3D case was designed (15×10.5×4.8 cm) and printed for enclosing all of the components. Figure 9. 17 presents the complete device.

Three designed components were implemented in the board: data acquisition through the analog sensor; denoising and processing of the data; display the information. Although the EEG sensor has an analog filtering circuit, it was observed that a substantial amount of motion artifacts remained in the signal. Thus, a lowpass finite impulse response filter, with 20 Hz cutoff frequency and order 20, was used for lessening the noise. However, the filtering operations at 200 Hz (employed sampling frequency) can considerably delay the examination process. Therefore, the filter was designed in Vivado, and exported (in tcl format) to be imported by the developed Python program (stored in the micro secure digital card), which is executed on the boot when the FPGA board is turned on.

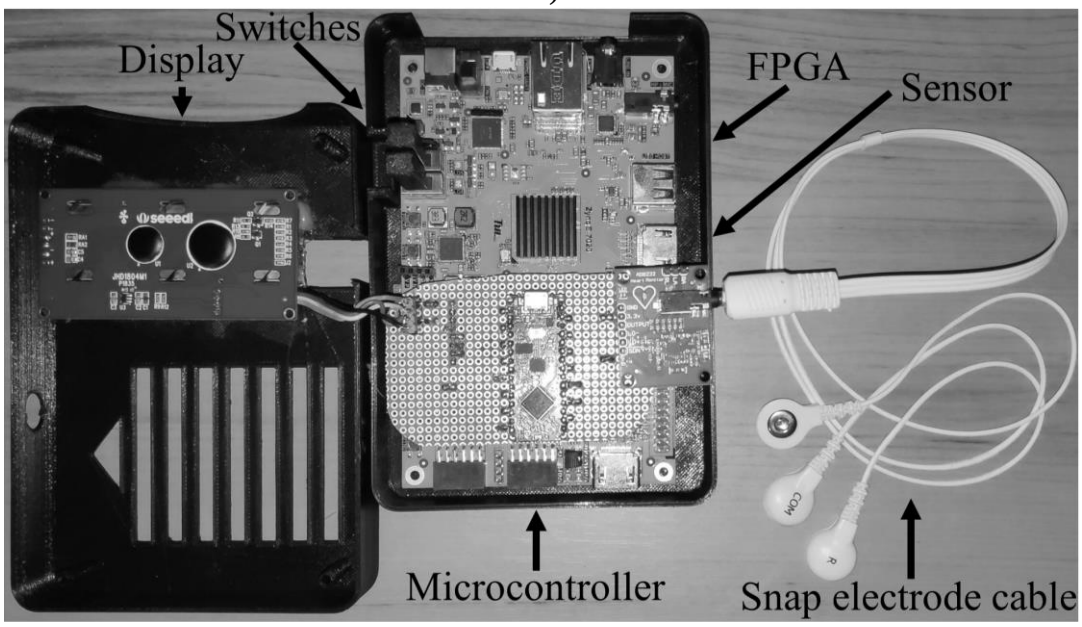
Therefore, the used approach has fast data acquisition and manipulation (implemented in the programmable logic), with a reliable classification process (implemented in the processing system). The interaction with the device is performed by operating the two slide switches, and the procedure for a standard examination requires the user to follow the following steps:

1. Move the first switch up to start the application, and wait until the message "Sleep Quality" is presented on the display.
2. Attach the ECG sensor's pads to the body, forming the single-lead position.
3. Move the second switch up (this action leads the device to start recording the information), and the message "Recording" is presented on the display.
4. Move the second switch down to stop the recording session, and the message "Analyzing" is presented on the display.
5. Wait until the message "Finished Exam" is displayed, indicating that the examination was finished (the result are stored in a text file, and the application closes).
6. The user can either retrieve the text file, to examine the results, or deliver the device to a specialist to perform the examination.

The total cost of the device was 152 € (110 € for the FPGA board, 24 € for the display and microcontroller, and 18 € for the ECG sensor). The device was fed by an external

power source (a 15 V, 3000 mAH, battery was used for the examination). The average current consumption was 305 mA; hence, the test can last at least 9 hours.

a)



b)

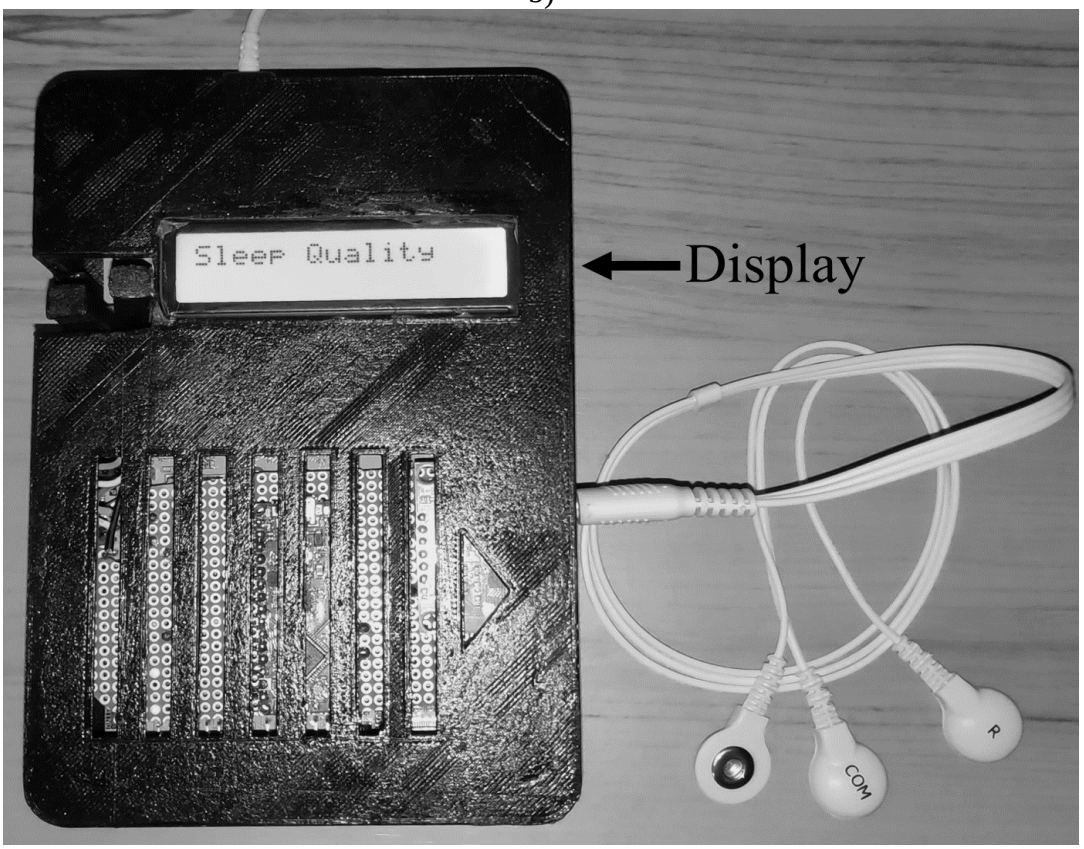


Figure 9.17. HMD for sleep quality estimation based on the ECG signal, with a) open box, and b) closed box [408].

9.4. HMD for sleep quality and OSA assessment

The developed model for sleep quality estimation was implemented on a developed HMD that is capable of predicting both the quality of sleep and the AHI (using the method presented in chapter 8). The conceived architecture of the device is the same as the one presented in Figure 9.14. It is composed of a sensing unit that acquires and wirelessly transmits (by Bluetooth) the signals to the processing unit, which performs the examination of the signals. The goal of this device is to be user friendly, and easy to operate with minimal invasiveness. Hence the approach of having two units (a small unit for data acquisition, and a larger unit with a touch screen for the user to interact) is more suitable.

The employed sensing unit is similar to the unit previously presented, using the BITalino Core BT [404] to collect the signals from a single-lead ECG sensor (the sensor's cost was 40 €). The average measurement root mean squared error of the sensor was 0.049±0.016, when comparing with an established gold standard device (BioPac MP35 Student Lab Pro) [405], advocating the viability of the sensor for clinical diagnostics. The processing unit is the same as previously described with a similar GUI, and the used hardware is presented in Figure 9.18.

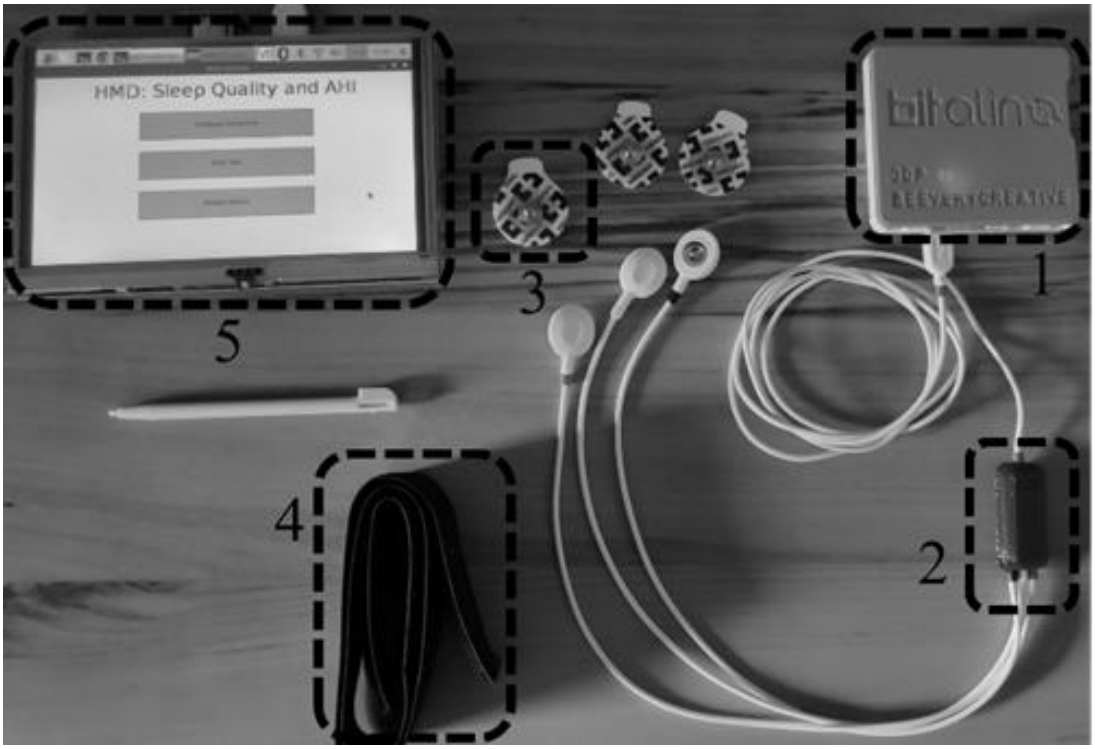


Figure 9.18. Hardware used on the HMD that predicts both the sleep quality and the AHI. 1- Sensing unit, 2- ECG sensor, 3- Electrode, 4- Armband, 5- Processing unit [299].

For the normal examination, the user needs to follow the steps:

1. Position the ECG electrodes to produce a single-lead ECG signal.
2. Tight the armband around the upper arm.
3. Attach the sensing unit to the armband using the s-shaped belt clip.

4. Activate the sensing unit.
5. Activate the processing unit, wait for the application to open, and press on "Start Test" (a new window will be displayed with the "Stop Test" button).
6. Press on "Stop Test" when the test is finished, the measured signals will be stored in a text file, with a timestamp, on the 32 GB secure digital memory card of the processing unit.
7. Press on "Analyze Results", and the algorithm developed in chapter 8 is employed to perform the minute-by-minute and global sleep quality and OSA examination (the results are stored in a text file).
8. The user can either examine the results file or deliver the device to an expert that can retrieve the file and perform the analysis of the results.

9.5. Key remarks

Two approaches were considered in the HMDs development. The first was the implementation of the HMD for each of the examined source sensor. Hence, the OSA evaluation was performed using the SpO2 signal (using the methods presented in sections 7.1 and 7.2), while the CAP estimation was based on either the EEG signal (employing the LSTM based method presented in section 5.3) or the ECG signal (using the method presented in sections 6.2) analysis. Therefore, these devices allow to estimate the quality of sleep according to the first two methodologies theorized in this work, by evaluating sleep quality metrics, and by assessing the occurrence of a sleep related disorder, respectively.

The second approach, applied for the development of the last HMD, was based on the third approach theorized in this work for sleep quality assessment (the combination of the information from sleep quality metrics and the sleep disorder detection to provide a better view of the global sleep quality, using the method presented in chapter 8).

The developed devices were intended to perform the examination at the patient's home without requiring the attendance of a specialist (to assemble the sensor or the device). Since the HMDs employed algorithms that attained a performance which is in the upper range of the specialist agreement; thus, they can possibly be used for clinical analysis.

The HMD were proposed as a proof of concept to demonstrate the implementation feasibility of the proposed methods. It was verified that the HMDs have enough computational resources, can record the physiological signals, and are energy efficient. However, the main limitation was that the devices were not validated in real world environment (performing recordings in parallel with a PSG, and comparing the results to verify if the performance of the proposed device is similar to the PSG estimations). Hence, it is not possible to know if they are sturdy enough to perform proper continuous monitoring of the subject in the difficulties of the sleep environment.

The proposed validation protocol follows the lines presented by Santos-Silva et al. [282], carrying out three nights of sleep evaluation (the sequence of these evaluations is randomly determined):

- One night performing recordings with the developed HMD at the patient's home.
- One night performing a parallel recording, in a sleep laboratory, of PSG, and the developed HMD.
- One night performing a recording, in a sleep laboratory, with only PSG.

All evaluations should take place within a two week period, and an expert physician must score the OSA events, the AHI, the CAP phases, and the CAP cycles. The examined population must have an equal number of subjects suffering from sleep-disordered breathing, and subjects free of sleep related disorders.

A proper validation could be attained using 80 participants, and all participants be free of cardiac pathologies. Ethical permission must be obtained from an ethics committee and all subjects must provide written informed consent for participating in the experiment. The examination must comply with the ten points required to achieve the highest EL and QR.

10. Conclusion

10.1. Overview of the work

10.2. Limitations of the work

10.3. Future work

10.1. Overview of the work

From the state of the art examination, it was observed that subjective methods for sleep quality assessment are fast and economic but have a high subjectivity. Thus, PSG based metrics are preferable to acquire objective information. Despite the simplicity associated with the estimation of metrics based on the sleep stage scoring, they can be seen as a rough indicator of the sleep quality since they are based on a synthetic segmentation of the continuous process that is sleep, drastically reducing the information provided by the PSG. As an example, a PSG that monitors 12 channels, sampled at 256 Hz with 16 bit resolution, produces about 1.5 Mb of information for each 30 s epoch, which is reduced to 3 bit of information to indicate the sleep stage.

However, there is a lack of definitional consensus regarding the proper method to define sleep quality. As a result, three main approaches were identified in this work as suitable paths to perform the sleep quality examination. The first considered the examination of sleep quality metrics, and CAP based examination was employed since these metrics examine the stability of sleep (found to possibly be the best metrics to describe the subjective experience of sleep), and are highly correlated to the occurrence of OSA in SBD patients. As a result, approaches based on the EEG sensor were evaluated as the CAP is composed of characteristic patterns captured by this sensor. The second approach was to assess the occurrence of sleep related disorders, theorizing that the presence of such disorders may be the main contributor to poor sleep quality. The fact that more than 60 sleep related disorders have been identified consubstantiate this hypothesis. Therefore, OSA was evaluated in this work as it was identified to be one of the most prevalent and frequently undiagnosed sleep related disorders, which can significantly affect the quality of sleep. From the reviewed literature, it was possible to conclude that either ECG or oximetry based approaches are likely to be the most relevant for the OSA research.

The last approach theorized that the combination of sleep quality metrics with the detection of sleep related disorders could provide a better estimation of global sleep quality. Although the CAP based analysis is conventionally performed by examining the EEG sensor, it was proposed in the state of the art that it can be indirectly evaluated by examining the ECG sensor (considering CAP in a broader context, where the occurrence of a CAP cycle designates the instability of sleep). Taking into consideration that the development of a method for OSA analysis was performed for the EEG sensor, hence, it was considered the development of a new sleep quality model (for the single-lead ECG) where the same source signal can be used to estimate sleep quality metrics (based on CAP), and assess the occurrence of OSA. This approach allowed the development of a less complex HMD, which is easier to be self-assembled, and addressed a gap in the state of the art regarding the proposal of approaches for sleep quality analysis based on the ECG signals.

Through the reviews of the state of the art, it was possible to provide an answer to the first research question (are the self-rating indexes of sleep quality the best way to estimate the quality of sleep or should objective measures, such as CAP rate, be used?). It was assessed that it exists a poor correlation between the self-rating indexes and the predicted quality of sleep; thus, advocating for the need to employ objective metrics. It was verified

in the state of the art that intensity, duration, and continuity metrics do not necessarily characterize the totality of the subject's sleep experience since some subjects present sleep complaints while having similar intensity, duration, and continuity metrics as those seen in non-complaining individuals [12]. Hence, the reasons for such sleep complaints may not be related to the architecture, timing, or amount of sleep but rather to variations in the experience of sleep itself. CAP based metrics were identified as having the highest capability of predicting the occurrence of instability during sleep that is caused by sleep-related disorders. It was also verified that OSA is the most prevalent sleep-related disorder that is commonly undiagnosed; hence, supporting the need to include this disorder in the analysis.

By taking into consideration that CAP is a characteristic EEG pattern, the first approach was the development of methods that can classify CAP with a minimally invasive sensor. Therefore, the signal from one EEG monopolar derivation was analyzed. Two approaches have been followed. The first one was based on features that fed a classifier (testing multiple features and classifiers), and the second approach employed methods without an explicit feature creation procedure. Both methods reached the highest bound of the specialist agreement; thus, becoming suitable for clinical analysis. Subsequently, a positive answer was attained for the second research question (can CAP be reliably assessed by analyzing the signal from one EEG monopolar derivation?). From the review analysis, it was also verified that the sleep microstructure provides significantly more information than the sleep macrostructure. However, most of the research focuses on the examination of the macrostructure with the conventional 30 s epoch. A sleep model was proposed to address this issue, using the EEG signal to estimate the CAP sequences through a statistical model. It was verified that this model can be used for the sleep quality estimation, and can possibly further improve the research in the sleep microstructure, either based on the CAP or on other microstructure metrics. A characterization analysis for CAP was also conducted, and methods for the A phase subtype detection were proposed as suitable future approaches for sleep quality assessment based on the first two methods for sleep quality analysis theorized in this work.

An indirect approach for CAP analysis was studied with the goal of having a validated metric, for sleep quality analysis, as the basis but measured by an easier to self-assemble sensor since the EEG sensor is potentially problematic for the elderly population. Therefore, the followed methodology considers the sleep instability concept associated with CAP, which can be extended to a broader context where the occurrence of sleep disturbances influence other physiological signals, especially the single-lead ECG [41]. Therefore, this sensor was chosen, and multiple models have been developed using a CPC technique as the foundation of the classification. This indirect approach achieved comparable performance to the models based on the EEG signal analysis. Thus, the third research question (can the sleep quality be assessed by considering an indirect estimation of the CAP rate?) also has a positive answer. Through the analysis of the ECG signals, it was verified the occurrence of a causal relationship between the heart rate and the respiration. Accordingly, a tool for the causality analysis, with the focus on physiological signals was proposed and tested for the indirect CAP estimation. The attained results are in the same range as the EEG based models; thus, further supporting the validity of the tool.

Two algorithms for OSA detection were proposed. The first examined features from the SpO2 signal, while the second analyzed the CPC signal produced from the single-lead ECG signal. It was verified that both approaches could provide a highly accurate AHI estimation by considering the threshold-based approach applied to the predicted m-AHI-tib. This methodology was followed to propose a sleep quality metric, the m-CAP-tib, for CAP analysis. It was verified that this metric is highly correlated to the CAP estimated by PSG, supporting the relevance of the proposed method. This metric was then employed by the developed model for sleep quality estimation, based on the appraisal of both predicted sleep quality and AHI, performed by a classifier ensemble.

The development of HMDs was the basis for the final stage of the work, and three source signals were evaluated. SpO2 was the first examined signal, and was evaluated by two HMD. The first performed the minute-by-minute OSA detection, while the second provided a global assessment of the disorder through the analysis of the SpO2 signal. The third HMD analyzed the signal from one EEG monopolar derivation to perform the sleep quality estimation (based on the CAP analysis) in a minute-by-minute approach that was used for the global sleep quality estimation. A similar approach was employed for the fourth HMD, which examined the single-lead ECG signal to perform the sleep quality assessment based on the indirect CAP assessment. The final device implemented the proposed sleep quality model by examining the single-lead ECG signal. The last device can also predict the AHI by examining the classified OSA minutes. All devices were cost-effective, and simple to use; therefore, providing a positive answer to the final research question (can a cost-effective home monitoring device be developed to perform both the estimation of the sleep quality and the presence of a sleep disorder?). Accordingly, all the established objectives have been accomplished in this research.

The novelties of this work can be summarized in the following points:

- Evaluation of the state of the art regarding the methods and devices for sleep quality and OSA analysis, assessing the research tendencies of these fields.
- Proposal, and examination of feature based and methods without an explicit feature extraction process for the classification of the CAP A phase and their subtypes, based on the EEG signal.
- Proposal of a sleep model that is suitable for CAP assessment.
- Characterization analysis of the CAP.
- Proposal of approaches for indirect CAP evaluation based on the ECG signal analysis.
- Proposal of an ECG based sleep quality metric (m-CAP-tib) that is correlated with the CAP rate predicted from the ECG signal.
- Development of a tool for time series analysis, capable of evaluating the causality of a dependent and an independent variable.
- Development of models for OSA analysis, based on SpO2 and ECG signals.
- Proposal of a new sleep quality model that evaluates both sleep quality metrics and the occurrence of a sleep related disorder (OSA estimation) to predict the quality of sleep, from the ECG signal.

- Implementation of HMDs for CAP, OSA, and sleep quality estimation, providing a proof of concept for each one of the approaches identified in this work as suitable paths to perform the sleep quality examination.

The proposed models for sleep quality analysis, based on the direct or indirect CAP estimation, achieved a performance that is equal to the highest agreement between two specialized physicians. The algorithms for OSA detection and diagnosis also attained a performance that in the range of the methods proposed in the state of the art. Therefore, the proposed method could possibly be used for medical analysis.

10.2. Limitations of the work

The main limitation of this work was the relatively low number of subjects involved in the CAP based analysis. A larger dataset is required to further validate the developed methods. Another limitation was that only one sleep related disorder was examined hence, it is not possible to know if the proposed methods will work properly for subjects suffering from other sleep related disorders.

Although not considered in this work, the evaluation of subjects suffering from specific pathologies, such as cardiac diseases, may significantly affect the evaluation of the proposed methods based on the ECG signal.

The last limitation is related to the proposed HMD, which were not validated in real world environment. Therefore, it is not possible to ensure that the devices are simple enough to be self-assembled and used by the elderly population, and if they are sturdy enough to perform an appropriate continuous monitoring of the subjects, in the difficulties of the sleep environment.

10.3. Future work

The possible next steps of this research are:

- Development of a larger dataset (with microstructure annotations) to allow a further validation of the proposed methods.
- Inclusion of other sleep related disorders in the examined population to assess the effect that his disorders can have in the performance of the algorithms.
- Evaluate subjects with cardiac pathologies to assess if the methods based on the ECG signal examination can work properly in the presence of such pathologies.
- Propose a mathematical approach for the definition of the CAP scoring rules.
- Validation of the developed HMD according to the proposed protocol (presented in section 9.5).
- Employ sensor fusion to assess if the combined information from different EEG deviations (such as F4-C4) can improve the performance for the CAP A phase classification.
- Perform a comparative analysis between the performance of a classifier attained by the proposed heuristic search algorithm (HOSA), a classifier produced by an exhaustive grid search, and a classifier attained by other heuristic based algorithms (such as genetic algorithms) frequently employed in the state of the

art, to assess the differences in the optimization time and in the attained performance.

- Examine other classifiers for the A phase assessment, such as the combination of 1D-CNN (for the automatic feature extraction) with an LSTM (to perform a classification with recurrent information).
- Study if a third hidden state, for the HMM employed by the proposed sleep model, related to the patterns where the conventional sleep model scores as either REM or wake after sleep onset periods, can improve the performance of the proposed sleep model for CAP estimation.

References

[1] H. Schulz and P. Salzarulo, “Forerunners of REM sleep,” *Sleep Med. Rev.*, vol. 16, no. 1, pp. 95–108, 2012.
[2] T. Young, P. Peppard, and D. Gottlieb, “Epidemiology of obstructive sleep apnea: a population health perspective,” *Am. J. Respir. Crit. Care Med.*, vol. 165, no. 9, pp. 1217–1239, 2002.
[3] I. Capellini, P. McNamara, B. Preston, C. Nunn, and R. Barton, “Does Sleep Play a Role in Memory Consolidation? A Comparative Test,” *PLOS ONE*, vol. 4, no. 2, p. e4609, 2009.
[4] M. Mahowald and C. Schenck, “Insights from studying human sleep disorders,” *Nature*, vol. 437, no. 7063, pp. 1279–1285, 2005.
[5] M. Hafner, M. Stepanek, J. Taylor, W. Troxel, and C. Stolk, “Why Sleep Matters-The Economic Costs of Insufficient Sleep: A Cross-Country Comparative Analysis,” *RAND Health Q.*, vol. 6, no. 4, p. 11, 2017.
[6] K. Waller *et al.*, “Subjective sleep quality and daytime sleepiness in late midlife and their association with age-related changes in cognition,” *Sleep Med.*, vol. 17, no. 1, pp. 165–173, 2016.
[7] B. Mander, J. Winer, and M. Walker, “Sleep and Human Aging,” *Neuron*, vol. 94, no. 1, pp. 19–36, 2017.
[8] T. Penzel, “Sleep Quality Assessment: Challenges and Opportunities,” *IEEE Pulse*, vol. 7, no. 4, pp. 3–3, 2016.
[9] T. Åkerstedt, K. Hume, D. Minors, and J. Waterhouse, “The meaning of good sleep: a longitudinal study of polysomnography and subjective sleep quality,” *J. Sleep Res.*, vol. 3, no. 3, pp. 152–158, 1994.
[10] L. Parrino, G. Milioli, A. Melpignano, and I. Trippi, “The Cyclic Alternating Pattern and the Brain-Body-Coupling During Sleep,” *Epileptologie*, vol. 33, no. 1, pp. 150–160, 2016.
[11] K. Kaplan *et al.*, “When a gold standard isn’t so golden: Lack of prediction of subjective sleep quality from sleep polysomnography,” *Biol. Psychol.*, vol. 123, no. 1, pp. 37–46, 2017.
[12] A. Krystal and J. Edinger, “Measuring sleep quality,” *Sleep Med.*, vol. 9, no. 1, pp. S10–S17, 2008.
[13] M. Sateia, “International classification of sleep disorders-third edition: highlights and modifications,” *Chest*, vol. 145, no. 5, pp. 1387–1394, 2014.
[14] P. Peppard, T. Young, J. Barnet, M. Palta, E. Hagen, and K. Hla, “Increased Prevalence of Sleep-Disordered Breathing in Adults,” *Am. J. Epidemiol.*, vol. 177, no. 9, pp. 1006–1014, 2013.
[15] T. Young, L. Evans, L. Finn, and M. Palta, “Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women,” *Sleep*, vol. 20, no. 9, pp. 705–706, 1997.
[16] L. Almazaydeh, K. Elleithy, and M. Faezipour, “Detection of Obstructive Sleep Apnea Through ECG Signal Features,” presented at the 2012 IEEE International Conference on Electro Information Technology, Indiana, USA, May 2012.
[17] S. Verhulst, N. Schrauwen, W. Backer, and K. Desager, “First night effect for polysomnographic data in children and adolescents with suspected sleep disordered breathing,” *Arch. Dis. Child.*, vol. 91, no. 3, pp. 233–237, 2006.
[18] L. Piwek, D. Ellis, S. Andrews, and A. Joinson, “The Rise of Consumer Health Wearables: Promises and Barriers,” *PLOS Med.*, vol. 13, no. 2, pp. 1–9, 2016.
[19] M. Terzano *et al.*, “Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep,” *Sleep Med.*, vol. 2, no. 6, pp. 537–553, 2001.
[20] D. Schomer and F. Silva, *Electroencephalography: Basic principles, clinical applications and related fields*. Philadelphia, USA: Lippincott Williams and Wilkins, 2010.
[21] R. Berry *et al.*, *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. Darien, Illinois, USA: American Academy of Sleep Medicine, 2017.
[22] F. Mendonça, A. Fred, S. Mostafa, F. Morgado-Dias, and A. Ravelo-García, “Automatic Detection of a Phases for CAP Classification,” presented at the 7th International Conference on Pattern Recognition Applications and Methods (ICPRAM), Funchal, Portugal, Jan. 2018.
[23] M. Peker, “An efficient sleep scoring system based on EEG signal using complex-valued machine learning algorithms,” *Neurocomputing*, vol. 207, no. C, pp. 165–177, 2016.
[24] T. Penzel and R. Conradt, “Computer based sleep recording and analysis,” *Sleep Med. Rev.*, vol. 4, no. 2, pp. 131–148, 2000.
[25] R. Boostani, F. Karimzadeh, and M. Nami, “A comparative review on sleep stage classification methods in patients and healthy individuals,” *Comput. Methods Programs Biomed.*, vol. 140, no. 1, pp. 77–91, 2017.
[26] F. Mendonça, S. Mostafa, F. Morgado-Dias, and A. Ravelo-García, “A Portable Wireless Device for Cyclic Alternating Pattern Estimation from an EEG Monopolar Derivation,” *Entropy*, vol. 21, no. 12, p. 1203, 2019.
[27] M. Terzano and L. Parrino, “Origin and Significance of the Cyclic Alternating Pattern (CAP),” *Sleep Med. Rev.*, vol. 4, no. 1, pp. 101–123, 2000.

[28] S. Mariani *et al.*, "Characterization of A phases during the Cyclic Alternating Pattern of sleep," *Clin. Neurophysiol.*, vol. 122, no. 10, pp. 2016–2024, 2011.
[29] M. Terzano, L. Parrino, M. Boselli, A. Smerieri, and M. Spaggiari, "CAP components and EEG synchronization in the first 3 sleep cycles," *Clin. Neurophysiol.*, vol. 111, no. 2, pp. 283–290, 2000.
[30] F. Mendonça, A. Fred, S. Mostafa, F. Morgado-Dias, and A. Ravelo-García, "Automatic detection of cyclic alternating pattern," *Neural Comput. Appl.*, pp. 1–11, 2018, doi: 10.1007/s00521-018-3474-5.
[31] M. Terzano and L. Parrino, "Clinical applications of cyclic alternating pattern," *Physiol. Behav.*, vol. 54, no. 4, pp. 807–813, 1993.
[32] L. Parrino, M. Boselli, P. Buccino, M. Spaggiari, G. Giovanni, and M. Terzano, "The cyclic alternating pattern plays a gate-control on periodic limb movements during non-rapid eye movement sleep," *J. Clin. Neurophysiol.*, vol. 13, no. 4, pp. 314–323, 1996.
[33] S. Korkmaz, N. Bilecenoglu, M. Aksu, and T. Yoldas, "Cyclic Alternating Pattern in Obstructive Sleep Apnea Patients with versus without Excessive Sleepiness," *Sleep Disord.*, vol. 2018, no. 1, pp. 1–7, 2018.
[34] M. Terzano, A. Smerieri, A. Felice, F. Giglia, V. Palomba, and L. Parrino, "Cyclic alternating pattern (CAP) alterations in narcolepsy," *Sleep Med.*, vol. 7, no. 8, pp. 619–626, 2006.
[35] L. Parrino, G. Milioli, F. Paolis, A. Grassi, and M. Terzano, "Paradoxical insomnia: the role of CAP and arousals in sleep misperception," *Sleep Med.*, vol. 10, no. 10, pp. 1139–1145, 2009.
[36] L. Parrino *et al.*, "Distinctive polysomnographic traits in nocturnal frontal lobe epilepsy," *Epilepsia*, vol. 53, no. 7, pp. 1178–1184, 2012.
[37] L. Parrino, R. Ferri, O. Bruni, and M. Terzano, "Cyclic alternating pattern (CAP): the marker of sleep instability," *Sleep Med. Rev.*, vol. 16, no. 1, pp. 27–45, 2012.
[38] A. Smerieri, L. Parrino, M. Agosti, R. Ferri, and M. Terzano, "Cyclic alternating pattern sequences and non-cyclic alternating pattern periods in human sleep," *Clin. Neurophysiol.*, vol. 118, no. 10, pp. 2305–2313, 2007.
[39] L. Parrino, M. Boselli, M. Spaggiari, A. Smerieri, and M. Terzano, "Cyclic alternating pattern (CAP) in normal sleep: polysomnographic parameters in different age groups," *Electroencephalogr. Clin. Neurophysiol.*, vol. 107, no. 6, pp. 439–450, 1998.
[40] R. Ferri *et al.*, "Cyclic alternating pattern and spectral analysis of heart rate variability during normal sleep," *J. Sleep Res.*, vol. 9, no. 1, pp. 13–18, 2000.
[41] R. Thomas, J. Mietus, C. Peng, and A. Goldberger, "An Electrocardiogram-Based Technique to Assess Cardiopulmonary Coupling During Sleep," *Sleep*, vol. 28, no. 9, pp. 1151–1161, 2005.
[42] F. Mendonça, S. Mostafa, F. Morgado-Dias, A. Ravelo-García, and T. Penzel, "Sleep quality of subjects with and without sleep-disordered breathing based on the cyclic alternating pattern rate estimation from single-lead ECG," *Physiol. Meas.*, vol. 40, pp. 1–14, 2019, doi: 10.1088/1361-6579/ab4f08.
[43] T. Penzel, G. Moody, R. Mark, A. Goldberger, and J. Peter, "The apnea-ecg database," in *Computers in Cardiology*, Massachusetts, USA, 2000, pp. 255–258.
[44] S. Martín-González, J. Navarro-Mesa, G. Juliá-Serdá, G. Ramírez-Ávila, and A. Ravelo-García, "Improving the understanding of sleep apnea characterization using Recurrence Quantification Analysis by defining overall acceptable values for the dimensionality of the system, the delay, and the distance threshold," *PLoS One*, vol. 13, no. 4, p. e0194462, 2018.
[45] A. Malhotra and R. Owens, "What Is Central Sleep Apnea?," *Respir. Care*, vol. 55, no. 9, pp. 1168–1178, 2010.
[46] D. Eckert, A. Jordan, P. Merchia, and A. Malhotra, "Central Sleep Apnea," *Chest*, vol. 131, no. 2, pp. 595–607, 2007.
[47] J. Wang, Y. Wang, J. Feng, B. Chen, and J. Cao, "Complex sleep apnea syndrome," *Patient Prefer. Adherence*, vol. 7, no. 1, pp. 633–641.
[48] T. Penzel, J. McNames, P. Chazal, B. Raymond, A. Murray, and G. Moody, "Systematic comparison of different algorithms for apnoea detection based on electrocardiogram recordings," *Med. Biol. Eng. Comput.*, vol. 40, no. 4, pp. 402–407, 2002.
[49] A. Ravelo-García *et al.*, "Oxygen Saturation and RR Intervals Feature Selection for Sleep Apnea Detection," *Entropy*, vol. 17, no. 5, pp. 2932–2957, 2015.
[50] C. Guilleminault, S. Connolly, R. Winkle, K. Melvin, and A. Tilkian, "Cyclical variation of the heart rate in sleep apnoea syndrome. Mechanisms, and usefulness of 24 h electrocardiography as a screening technique," *Lancet*, vol. 323, no. 8369, pp. 126–131, 1984.
[51] R. Thomas *et al.*, "Differentiating obstructive from central and complex sleep apnea using an automated electrocardiogram-based method," *Sleep*, vol. 30, no. 12, pp. 1756–1769, 2007.
[52] L. Ibrahim *et al.*, "Heritability of abnormalities in cardiopulmonary coupling in sleep apnea: use of an electrocardiogram-based technique," *Sleep*, vol. 33, no. 5, pp. 643–646, 2010.

[53] A. Rosa, G. Alves, M. Brito, M. Lopes, and S. Tufik, “Visual and automatic cyclic alternating pattern (CAP) scoring: inter-rater reliability study,” *Arq. Neuropsiquiatr.*, vol. 64, no. 3A, pp. 578–581, 2006.
[54] R. Largo, M. Lopes, K. Spruyt, C. Guilleminault, Y. Wang, and A. Rosa, “Visual and automatic classification of the cyclic alternating pattern in electroencephalography during sleep,” *Braz. J. Med. Biol. Res.*, vol. 52, no. 3, p. e8059, 2019.
[55] F. Mendonça, S. Mostafa, M. Morgado-Dias, A. Ravelo-García, and T. Penzel, “A Review of Approaches for Sleep Quality Analysis,” *IEEE Access*, vol. 7, no. 1, pp. 24527–24546, 2019.
[56] F. Mendonça, S. Mostafa, A. Ravelo-García, F. Morgado-Dias, and T. Penzel, “A Review of Obstructive Sleep Apnea Detection Approaches,” *IEEE J. Biomed. Health Inform.*, vol. 23, no. 2, pp. 825–837, 2019.
[57] F. Mendonça, S. Mostafa, A. Ravelo-García, F. Morgado-Dias, and T. Penzel, “Devices for home detection of obstructive sleep apnea: A review,” *Sleep Med. Rev.*, vol. 41, no. 1, pp. 149–160, 2018.
[58] G. Keklund and T. Akerstedt, “Objective components of individual differences in subjective sleep quality,” *J. Sleep Res.*, vol. 6, no. 4, pp. 217–220, 1997.
[59] M. Ohayon *et al.*, “National Sleep Foundation’s sleep quality recommendations: first report,” *Sleep Health*, vol. 3, no. 1, pp. 6–19, 2017.
[60] K. Knutson *et al.*, “The National Sleep Foundation’s Sleep Health Index,” *Sleep Health*, vol. 3, no. 4, pp. 234–240, 2017.
[61] B. Saletu, P. Wessely, J. Grünberger, and M. Schultes, “Erste klinische Erfahrungen mit einem neuen schlafanstoßenden Benzodiazepin, Cinolazepam mittels eines Selbstbeurteilungsbogens für Schlaf- und Aufwachqualität (SSA),” *Neuropsychiatrie*, vol. 1, no. 1, pp. 169–176, 1987.
[62] E. Snyder, B. Cai, C. DeMuro, M. Morrison, and W. Ball, “A New Single-Item Sleep Quality Scale: Results of Psychometric Evaluation in Patients With Chronic Primary Insomnia and Depression,” *J. Clin. Sleep Med.*, vol. 14, no. 11, pp. 1849–1857.
[63] Z. Rošt’áková, G. Dorffner, O. Aydemir, and R. Rosipal, “Estimation of Sleep Quality by Using Microstructure Profiles,” presented at the 16th Conference on Artificial Intelligence in Medicine (AIME), Vienna, Austria, Jun. 2017.
[64] A. Ong and M. Gillespie, “Overview of smartphone applications for sleep analysis,” *World J. Otorhinolaryngol. - Head Neck Surg.*, vol. 2, no. 1, pp. 45–49, 2016.
[65] B. Kemp, A. Zwinderman, B. Tuk, H. Kamphuisen, and J. Oberye, “Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG,” *IEEE Trans. Biomed. Eng.*, vol. 47, no. 9, pp. 1185–1194, 2000.
[66] I. Badreldin and A. Morsy, “Consistency of sleep restoration gain (SRG) as a measure for assessing sleep quality,” presented at the 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, California USA, Sep. 2012.
[67] Y. Wakuda, Y. Hasegawa, T. Fukuda, A. Noda, F. Arai, and M. Kawaguchi, “Estimation of sleep cycle and quality based on nonlinear analysis of heart rate variability,” presented at the The Fourth Symposium Micro-Nanomechatronics for Information-Based Society, Nagoya, Japan, Nov. 2004.
[68] E. Dafna, A. Tarasiuk, and Y. Zigel, “Sleep-Wake Evaluation from Whole-Night Non-Contact Audio Recordings of Breathing Sounds,” *PLoS One*, vol. 10, no. 2, p. e0117382, 2015.
[69] E. Dafna, A. Tarasiuk, and Y. Zigel, “Sleep-quality assessment from full night audio recordings of sleep apnea patients,” presented at the 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, California, USA, Sep. 2012.
[70] H. Wu, T. Kato, M. Numao, and K. Fukui, “Statistical sleep pattern modelling for sleep quality assessment based on sound events,” *Health Inf. Sci. Syst.*, vol. 5, no. 1, pp. 1–11, 2017.
[71] M. Bsoul, H. Minn, M. Nourani, G. Gupta, and L. Tamil, “Real-time sleep quality assessment using single-lead ECG and multi-stage SVM classifier,” presented at the 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, Buenos Aires, Argentina, Sep. 2010.
[72] S. Cheng and J. Huang, “An intelligent sleep quality detection system based on wearable device,” presented at the 2017 International Conference on Machine Learning and Cybernetics (ICMLC), Ningbo, China, Jul. 2017.
[73] H. Tseng, C. Huang, L. Yen, T. Lin, Y. Lee, and Y. Chen, “A method of measuring sleep quality by using PPG,” presented at the 2016 IEEE International Conference on Consumer Electronics - Taiwan, Nantou, Taiwan, 2016.
[74] P. Fonseca *et al.*, “Validation of Photoplethysmography-Based Sleep Staging Compared With Polysomnography in Healthy Middle-Aged Adults,” *Sleep*, vol. 40, no. 7, pp. 1–10, 2017.
[75] T. Blackwell, S. Ancoli-Israel, S. Redline, and K. Stone, “Factors that May Influence the Classification of Sleep-Wake by Wrist Actigraphy: The MrOS Sleep Study,” *J. Clin. Sleep Med.*, vol. 7, no. 4, pp. 357–367, 2011.

[76] A. Sathyanarayana *et al.*, “Sleep Quality Prediction From Wearable Data Using Deep Learning,” *J. Med. Internet Res. Mhealth Uhealth*, vol. 4, no. e125, p. 4, 2016.
[77] D. Jung, S. Hwang, G. Chung, Y. Lee, D. Jeong, and K. Park, “Estimation of Sleep Onset Latency Based on the Blood Pressure Regulatory Reflex Mechanism,” *IEEE J. Biomed. Health Inform.*, vol. 17, no. 3, pp. 539–544, 2013.
[78] K. Park, S. Hwang, D. Jung, H. Yoon, and W. Lee, “Ballistocardiography for nonintrusive sleep structure estimation,” presented at the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Illinois, USA, Aug. 2014.
[79] S. Mariani *et al.*, “Automatic detection of A phases of the Cyclic Alternating Pattern during sleep,” presented at the 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, Buenos Aires, Argentina, Sep. 2010.
[80] F. Mendonça, S. Mostafa, F. Morgado-Dias, and A. Ravelo-García, “Sleep Quality Analysis with Cardiopulmonary Coupling,” presented at the IEEE 2018 International Conference on Biomedical Engineering and Applications (ICBEA), Funchal, Portugal, Jul. 2018.
[81] F. Mendonça, S. Mostafa, F. Morgado-Dias, and A. Ravelo-Garcia, “Sleep quality estimation by cardiopulmonary coupling analysis,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 12, pp. 2233–2239, 2018.
[82] L. Wang, W. Zheng, H. Ma, and B. Lu, “Measuring sleep quality from EEG with machine learning approaches,” presented at the 2016 International Joint Conference on Neural Networks (IJCNN), British Columbia, Canada, Jul. 2016.
[83] B. Takhtsabzy and C. Thomsen, “Assessment of Sleep Quality in Powernapping,” presented at the 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Massachusetts, USA, Sep. 2011.
[84] Z. Rošťáková and R. Rosipal, “Differences in sleep microstate curves among healthy sleepers and patients after stroke,” *Clin. Neurophysiol.*, vol. 127, no. 3, p. e126, 2016.
[85] A. Lewandowski, R. Rosipal, and G. Dorffner, “Extracting more information from EEG recordings for a better description of sleep,” *Comput. Methods Programs Biomed.*, vol. 108, no. 3, pp. 961–972, 2012.
[86] R. Rosipal, A. Lewandowski, and G. Dorffner, “In search of objective components for sleep quality indexing in normal sleep,” *Biol. Psychol.*, vol. 94, no. 1, pp. 210–220, 2013.
[87] A. Flexer, G. Gruber, and G. Dorffner, “A reliable probabilistic sleep stager based on a single EEG signal,” *Artif. Intell. Med.*, vol. 33, no. 3, pp. 199–207, 2005.
[88] A. Ravelo-García, L. Lorenzo-García, and J. Navarro-Mesa, *Sleep Quality Differences According to a Statistical Continuous Sleep Model*. Berlin, Germany: Springer Nature, 2009.
[89] R. Choi, W. Kang, and C. Son, “Explainable sleep quality evaluation model using machine learning approach,” presented at the 2017 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), Daegu, South Korea, Nov. 2017.
[90] Y. Peng, C. Lin, M. Sun, and C. Landis, “Multimodality Sensor System for Long-Term Sleep Quality Monitoring,” *IEEE Trans. Biomed. Circuits Syst.*, vol. 1, no. 3, pp. 217–227, 2007.
[91] Y. Peng, C. Lin, M. Sun, and M. Feng, “Sleep condition inferencing using simple multimodality sensors,” presented at the 2006 IEEE International Symposium on Circuits and Systems, Island of Kos, Greece, May 2006.
[92] Y. Peng, C. Lin, and M. Sun, “A Distributed Multimodality Sensor System for Home-Used Sleep Condition Inference and Monitoring,” presented at the 1st Transdisciplinary Conference on Distributed Diagnosis and Home Healthcare, Virginia, USA, Apr. 2006.
[93] A. Alivar *et al.*, “Motion Detection in Bed-Based Ballistocardiogram to Quantify Sleep Quality,” presented at the 2017 IEEE Global Communications Conference, Singapore, Singapore, Dec. 2017.
[94] S. Cheng and H. Mei, “A Personalized Sleep Quality Assessment Mechanism Based on Sleep Pattern Analysis,” presented at the Innovations in Bio-Inspired Computing and Applications (IBICA), 2012 Third International Conference on, Kaohsiung, Taiwan, Sep. 2012.
[95] M. Morrell, L. Finn, H. Kim, P. Peppard, M. Badr, and T. Young, “Sleep fragmentation, awake blood pressure, and sleep-disordered breathing in a population-based study,” *Am. J. Respir. Crit. Care Med.*, vol. 162, no. 6, pp. 2091–2096, 2000.
[96] V. Swarnkar, U. Abeyratne, C. Hukins, and B. Duce, “A state transition-based method for quantifying EEG sleep fragmentation,” *Med. Biol. Eng. Comput.*, vol. 47, no. 10, pp. 1053–1061, 2009.
[97] J. Haba-Rubio, V. Ibanez, and E. Sforza, “An alternative measure of sleep fragmentation in clinical practice: the sleep fragmentation index,” *Sleep Med.*, vol. 5, no. 6, pp. 577–581, 2004.
[98] E. Bouazizi *et al.*, “Mathematical modelling of sleep fragmentation diagnosis,” *Biomed. Signal Process. Control*, vol. 24, no. 1, pp. 83–92, 2016.

[99] M. Kirsch, K. Monahan, J. Weng, S. Redline, and K. Loparo, “Entropy-Based Measures for Quantifying Sleep-Stage Transition Dynamics: Relationship to Sleep Fragmentation and Daytime Sleepiness,” *IEEE Trans. Biomed. Eng.*, vol. 59, no. 3, pp. 787–796, 2012.
[100] R. Naeck *et al.*, “Sleep diversity index for fragmentation analysis,” *J. Nonlinear Syst. Appl.*, vol. 1, no. 1, pp. 151–154, 2014.
[101] R. Naeck *et al.*, “Sleep fragmentation thresholds of sleep fragmentation indices,” presented at the 42nd Annual Conference of the IEEE Industrial Electronics Society, Florence, Italy, Oct. 2016.
[102] Y. Nam, Y. Kim, and J. Lee, “Sleep Monitoring Based on a Tri-Axial Accelerometer and a Pressure Sensor,” *Sensors*, vol. 16, no. 5, pp. 1–14, 2016.
[103] H. Miwa, S. Sasahara, and T. Matsui, “Roll-over Detection and Sleep Quality Measurement using a Wearable Sensor,” presented at the 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Lyon, France, Aug. 2007.
[104] T. Han, S. Min, and Y. Nam, “A Real-time Sleep Monitoring System with a Smartphone,” presented at the 9th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, Blumenau, Brazil, Jul. 2015.
[105] M. Pouliot, V. Joshi, R. Goubran, and F. Knoefel, “Bed occupancy monitoring: Data processing and clinician user interface design,” presented at the 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, California, USA, Sep. 2012.
[106] T. Guettari, D. Istrate, J. Boudy, B. Benkelfat, B. Fumel, and J. Daviet, “Design and First Evaluation of a Sleep Characterization Monitoring System Using a Remote Contactless Sensor,” *IEEE J. Biomed. Health Inform.*, vol. 21, no. 6, pp. 1511–1523, 2017.
[107] R. Norman, M. Scott, I. Ayappa, J. Walsleben, and D. Rapoport, “Sleep continuity measured by survival curve analysis,” *Sleep*, vol. 29, no. 12, pp. 1625–1631, 2006.
[108] H. Tseng, C. Huang, L. Yen, Y. Liao, and Y. Lee, “A Method of Measurement Sleep Quality by Using EEG,” presented at the 2016 IEEE International Conference on Consumer Electronics, Nantou, Taiwan, May 2016.
[109] Y. Yu, S. Chen, C. Huang, L. Ko, and C. Lin, “Portable and wireless EEG device used in sleep quality tracking,” presented at the 2015 IEEE International Conference on Consumer Electronics, Taipei, Taiwan, Jun. 2015.
[110] S. Hamdan, H. Osman, M. Eid, and A. Saddik, “A biofeedback system for sleep management,” presented at the 2012 IEEE International Symposium on Robotic and Sensors Environments Proceedings, Magdeburg, Germany, Nov. 2012.
[111] C. Kuo, Y. Liu, D. Chang, C. Young, F. Shaw, and S. Liang, “Development and Evaluation of a Wearable Device for Sleep Quality Assessment,” *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1547–1557, 2016.
[112] A. Gaddam, K. Kaur, G. Gupta, and S. Mukhopadhyay, “Determination of sleep quality of inhabitant in a smart home using an intelligent bed sensing system,” presented at the 2010 IEEE Instrumentation & Measurement Technology Conference Proceedings, Texas, USA, May 2010.
[113] E. Pino, A. Paz, and P. Aqueveque, “Noninvasive Monitoring Device to Evaluate Sleep Quality at Mining Facilities,” *IEEE Trans. Ind. Appl.*, vol. 51, no. 1, pp. 101–108, 2015.
[114] E. Pino, A. Morán, A. Paz, and P. Aqueveque, “Validation of non-invasive monitoring device to evaluate sleep quality,” presented at the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Milan, Italy, Aug. 2015.
[115] P. Prakash *et al.*, “Sensors and Instrumentation for Unobtrusive Sleep Quality Assessment in Autistic Children,” presented at the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Illinois, USA, Aug. 2014.
[116] J. Paalasmaa, M. Waris, H. Toivonen, L. Leppäkorpi, and M. Partinen, “Unobtrusive online monitoring of sleep at home,” presented at the 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, California, USA, Sep. 2012.
[117] T. Rahman *et al.*, “DoppleSleep: A contactless unobtrusive sleep sensing system using short-range doppler radar,” presented at the 3rd ACM International Joint Conference on Pervasive and Ubiquitous Computing, Osaka, Japan, Sep. 2015.
[118] X. Liu, J. Cao, S. Tang, and J. Wen, “Wi-Sleep: Contactless Sleep Monitoring via WiFi Signals,” presented at the 2014 IEEE Real-Time Systems Symposium, Rome, Italy, Dec. 2014.
[119] S. Milici, A. Lázaro, R. Villarino, D. Girbau, and M. Magnarosa, “Wireless Wearable Magnetometer-Based Sensor for Sleep Quality Monitoring,” *IEEE Sens. J.*, vol. 18, no. 5, pp. 2154–2152, 2018.
[120] A. Veiga, L. García, J. Lloret, and V. Augele, “An IoT-based Smart Pillow for Sleep Quality Monitoring in AAL Environments,” presented at the 2018 Third International Conference on Fog and Mobile Edge Computing, Barcelona, Spain, Apr. 2018.

[121] A. Weiss, N. Johnson, N. Berger, and S. Redline, “Validity of activity-based devices to estimate sleep,” *J. Clin. Sleep Med.*, vol. 6, no. 4, pp. 336–342, 2010.
[122] G. Jean-Louis, D. Kripke, R. Cole, J. Assmus, and R. Langer, “Sleep detection with an accelerometer actigraph: comparisons with polysomnography,” *Physiol. Behav.*, vol. 72, no. 1, pp. 21–28, 2001.
[123] F. Borges and F. Fischer, “Twelve-Hour Night Shifts of Healthcare Workers: A Risk to the Patients?,” *Chronobiol. Int.*, vol. 20, no. 2, pp. 351–360, 2003.
[124] A. Keill and J. Lee, “Validity of wearable fitness trackers on sleep measure,” *Med. Sci. Sports Exerc.*, vol. 48, no. 1, p. 10, 2016.
[125] M. Shin, P. Swan, and C. Chow, “The validity of Actiwatch2 and SenseWear armband compared against polysomnography at different ambient temperature conditions,” *Sleep Sci.*, vol. 8, no. 1, pp. 9–15, 2015.
[126] S. Tworoger, S. Davis, M. Vitiello, M. Lentz, and A. McTiernan, “Factors associated with objective (actigraphic) and subjective sleep quality in young adult women,” *J. Psychosom. Res.*, vol. 59, no. 1, pp. 11–19, 2005.
[127] N. Cellini, M. Buman, E. McDevitt, A. Ricker, and S. Mednick, “Direct comparison of two actigraphy devices with polysomnographically recorded naps in healthy young adults,” *Chronobiol. Int.*, vol. 30, no. 5, pp. 691–698, 2013.
[128] C. McCrae, M. Rowe, C. Tierney, N. Dautovich, A. DeFinis, and J. McNamara, “Sleep complaints, subjective and objective sleep patterns, health, psychological adjustment, and daytime functioning in community-dwelling older adults,” *J. Gerontol. Psychol. Sci.*, vol. 60, no. 4, pp. 182–189, 2005.
[129] J. Mantua, N. Gravel, and R. Spencer, “Reliability of Sleep Measures from Four Personal Health Monitoring Devices Compared to Research-Based Actigraphy and Polysomnography,” *Sensors*, vol. 6, no. 5, p. E646, 2016.
[130] C. Kushida, A. Chang, C. Gadkary, C. Guilleminault, O. Carrillo, and W. Dement, “Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients,” *Sleep Med.*, vol. 2, no. 5, pp. 389–396, 2001.
[131] H. Montgomery-Downs, S. Insana, and J. Bond, “Movement toward a novel activity monitoring device,” *Sleep Breath*, vol. 16, no. 3, pp. 913–917, 2012.
[132] A. Sheth *et al.*, “Feasibility of Recording Sleep Quality And Sleep Duration Using Fitbit in Children with Asthma,” presented at the 32nd Annual Meeting of the Associated Professional Sleep Societies, Maryland, USA, Jun. 2018.
[133] D. Dickinson, J. Cazier, and T. Cech, “A practical validation study of a commercial accelerometer using good and poor sleepers,” *Health Psychol. Open*, vol. 3, no. 2, pp. 1–10, 2016.
[134] J. Perez-Macias, H. Jimison, I. Korhonen, and M. Pavel, “Comparative assessment of sleep quality estimates using home monitoring technology,” presented at the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Illinois, USA, Aug. 2014.
[135] L. Meltzer, L. Hiruma, K. Avis, H. Montgomery-Downs, and J. Valentin, “Comparison of a Commercial Accelerometer with Polysomnography and Actigraphy in Children and Adolescents,” *Sleep*, vol. 38, no. 8, pp. 1323–1230, 2015.
[136] P. Paavilainen *et al.*, “Circadian activity rhythm in demented and non-demented nursing-home residents measured by telemetric actigraphy,” *J. Sleep Res.*, vol. 14, no. 1, pp. 61–68, 2005.
[137] M. Zambotti, F. Baker, and I. Colrain, “Validation of Sleep-Tracking Technology Compared with Polysomnography in Adolescents,” *Sleep*, vol. 38, no. 9, pp. 1461–1468, 2015.
[138] M. Souders *et al.*, “Sleep Behaviors and Sleep Quality in Children with Autism Spectrum Disorders,” *Sleep*, vol. 32, no. 12, pp. 1566–1578, 2009.
[139] L. Souza, A. Benedito-Silva, M. Pires, D. Poyares, S. Tufik, and H. Calil, “Further Validation of Actigraphy for Sleep Studies,” *Sleep*, vol. 26, no. 1, pp. 81–85, 2003.
[140] M. Sharif and A. BaHammam, “Sleep estimation using BodyMedia’s SenseWear™ armband in patients with obstructive sleep apnea,” *Ann. Thorac. Med.*, vol. 8, no. 1, pp. 53–57, 2013.
[141] J. Merilahti, J. Pärkkä, K. Antila, E. Mattila, and I. Korhonen, “Long-term subjective and objective sleep analysis of total sleep time and sleep quality in real life settings,” presented at the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Lyon, France, Aug. 2007.
[142] D. Leger, C. Guilleminault, C. Santos, and M. Paillard, “Sleep/wake cycles in the dark: sleep recorded by polysomnography in 26 totally blind subjects compared to controls,” *Clin. Neurophysiol.*, vol. 113, no. 10, pp. 1607–1614, 2002.
[143] M. Zambotti, L. Rosas, I. Colrain, and F. Baker, “The Sleep of the Ring: Comparison of the OURA Sleep Tracker Against Polysomnography,” *Behav. Sleep. Med.*, vol. 1, no. 1, pp. 1–15, 2017.

[144] J. Shambroom, S. Fábregas, and J. Johnstone, "Validation of an automated wireless system to monitor sleep in healthy adults," *J. Sleep Res.*, vol. 21, no. 2, pp. 221–230, 2012.
[145] Z. Liang and M. Martell, "Validity of Consumer Activity Wristbands and Wearable EEG for Measuring Overall Sleep Parameters and Sleep Structure in Free-Living Conditions," *J. Healthc. Inform. Res.*, vol. 2, no. 1, pp. 152–178, 2018.
[146] D. Visco, M. Niesley, A. Hughes, and M. Lewis, "Incidence of Sleep Disturbance and Evaluation of Sleep Quality in Patients Diagnosed with Cancer," *Clin. Res. Pulmonol.*, vol. 3, no. 1, p. 1030, 2015.
[147] R. Thomas, C. Wood, and M. Bianchi, "Cardiopulmonary coupling spectrogram as an ambulatory clinical biomarker of sleep stability and quality in health, sleep apnea, and insomnia," *Sleep*, vol. 41, no. 2, 2018.
[148] A. Tal, Z. Shinar, D. Shaki, S. Codish, and A. Goldbart, "Validation of Contact-Free Sleep Monitoring Device with Comparison to Polysomnography," *J. Clin. Sleep Med.*, vol. 13, no. 3, pp. 517–522, 2017.
[149] M. Smith *et al.*, "Use of Actigraphy for the Evaluation of Sleep Disorders and Circadian Rhythm Sleep-Wake Disorders: An American Academy of Sleep Medicine Systematic Review, Meta-Analysis, and GRADE Assessment," *J. Clin. Sleep Med.*, vol. 14, no. 7, pp. 1209–1230, 2018.
[150] D. Álvarez, R. Hornero, D. Abásolo, F. Campo, and C. Zamarrón, "Nonlinear characteristics of blood oxygen saturation from nocturnal oximetry for obstructive sleep apnoea detection," *Physiol. Meas.*, vol. 27, no. 4, pp. 399–412, 2006.
[151] J. Marcos, R. Hornero, D. Álvarez, F. Campo, C. Zamarrón, and M. López, "Utility of multilayer perceptron neural network classifiers in the diagnosis of the obstructive sleep apnoea syndrome from nocturnal oximetry," *Comput. Methods Programs Biomed.*, vol. 92, no. 1, pp. 79–89, 2008.
[152] J. Marcos, R. Hornero, D. Álvarez, F. Campo, and C. Zamarrón, "Assessment of four statistical pattern recognition techniques to assist in obstructive sleep apnoea diagnosis from nocturnal oximetry," *Med. Eng. Phys.*, vol. 31, no. 8, pp. 971–978, 2009.
[153] J. Marcos, R. Hornero, D. Álvarez, F. Campo, and M. Aboy, "Automated detection of obstructive sleep apnoea syndrome from oxygen saturation recordings using linear discriminant analysis," *Med. Biol. Eng. Comput.*, vol. 48, no. 9, pp. 895–902, 2010.
[154] D. Álvarez, R. Hornero, J. Marcos, and F. Campo, "Multivariate Analysis of Blood Oxygen Saturation Recordings in Obstructive Sleep Apnea Diagnosis," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 12, pp. 2816–2824, 2010.
[155] D. Álvarez, R. Hornero, J. Marcos, and F. Campo, "Feature selection from nocturnal oximetry using genetic algorithms to assist in obstructive sleep apnoea diagnosis," *Med. Eng. Phys.*, vol. 34, no. 8, pp. 1049–1057, 2012.
[156] L. Almazaydeh, M. Faezipour, and K. Elleithy, "A Neural Network System for Detection of Obstructive Sleep Apnea Through SpO2 Signal Features," *Int. J. Adv. Comput. Sci. Appl.*, vol. 3, no. 5, pp. 7–11, 2012.
[157] J. Lázaro, E. Gil, J. Vergara, and P. Laguna, "Pulse Rate Variability Analysis for Discrimination of Sleep-Apnea-Related Decreases in the Amplitude Fluctuations of Pulse Photoplethysmographic Signal in Children," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 1, pp. 240–246, 2013.
[158] D. Morillo and N. Gross, "Probabilistic neural network approach for the detection of SAHS from overnight pulse oximetry," *Med. Biol. Eng. Comput.*, vol. 51, no. 3, pp. 305–315, 2013.
[159] A. Garde, P. Dehkordi, W. Karlen, D. Wensley, J. Ansermino, and G. Dumont, "Development of a Screening Tool for Sleep Disordered Breathing in Children Using the Phone Oximeter," *PLOS ONE*, vol. 9, no. 11, p. e112959, 2014.
[160] D. Álvarez *et al.*, "Automated analysis of unattended portable oximetry by means of Bayesian neural networks to assist in the diagnosis of sleep apnea," presented at the 2016 Global Medical Engineering Physics Exchanges/Pan American Health Care Exchanges, Madrid, Spain, Apr. 2016.
[161] S. Mostafa, F. Mendonça, F. Morgado-Dias, and A. Ravelo-García, "SpO2 based sleep apnea detection using deep learning," presented at the 2017 21st International Conference on Intelligent Engineering Systems, Larnaca, Cyprus, Oct. 2017.
[162] J. Morales, C. Varon, M. Deviaene, P. Borzée, D. Testelmans, and B. Buyse, "Sleep Apnea Hypopnea Syndrome classification in SpO2 signals using wavelet decomposition and phase space reconstruction," presented at the IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks, Eindhoven, Netherlands, May 2017.
[163] R. Pathinarupothi, J. Dhara, E. Rangan, E. Gopalakrishnan, R. Vinaykumar, and P. Soman, "Single Sensor Techniques for Sleep Apnea Diagnosis using Deep Learning," presented at the 2017 IEEE International Conference on Healthcare Informatics, Utah, USA, Aug. 2017.
[164] D. Jung *et al.*, "Real-Time Automatic Apneic Event Detection Using Nocturnal Pulse Oximetry," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 3, pp. 706–712, 2018.

[165] S. Mostafa, J. Carvalho, F. Morgado-Dias, and A. Ravelo-García, "Optimization of sleep apnea detection using SpO2 and ANN," presented at the 2017 XXVI International Conference on Information, Communication and Automation Technologies, Sarajevo, Bosnia-Herzegovina, Oct. 2017.
[166] F. Roche *et al.*, "Predicting sleep apnoea syndrome from heart period: a time-frequency wavelet analysis," *Eur. Respir. J.*, vol. 22, no. 6, pp. 937–942, 2003.
[167] C. Zywietz, V. Einem, B. Widiger, and G. Joseph, "ECG analysis for sleep apnea detection," *Methods Inf. Med.*, vol. 43, no. 1, pp. 56–59, 2004.
[168] P. Chazal, T. Penzel, and C. Heneghan, "Automated Detection of Obstructive Sleep Apnoea at Different Time Scales Using the Electrocardiogram," *Physiol. Meas.*, vol. 25, no. 4, pp. 967–983, 2004.
[169] A. Ravelo *et al.*, "Application of Support Vector Machines and Gaussian Mixture Models for the Detection of Obstructive Sleep Apnoea based on the RR Series," presented at the 8th WSEAS International Conference on Applied Mathematics, Tenerife, Spain, Dec. 2005.
[170] R. Lin, R. Lee, C. Tseng, H. Zhou, C. Chao, and J. Jiang, "A New Approach for Identifying Sleep Apnea Syndrome Using Wavelet Transform and Neural Networks," *Biomed. Eng. Appl. Basis Commun.*, vol. 18, no. 3, pp. 138–143, 2006.
[171] M. Mendez *et al.*, "Detection of sleep apnea from surface ECG based on features extracted by an autoregressive model," presented at the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Lyon, France, Aug. 2007.
[172] M. Mendez, A. Bianchi, M. Matteucci, S. Cerutti, and T. Penzel, "Sleep Apnea Screening by Autoregressive Models From a Single ECG Lead," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 12, pp. 2838–2850, 2009.
[173] A. Quiceno-Manrique, J. Alonso-Hern´andez, C. Travieso-Gonz´alez, M. Ferrer-Ballester, and G. Castellanos-Dom´ınguez, "Detection of obstructive sleep apnea in ECG recordings using time-frequency distributions and dynamic features," presented at the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009, Minnesota, USA, Sep. 2009.
[174] A. Khandoker, J. Gubbi, and M. Palaniswami, "Automated Scoring of Obstructive Sleep Apnea and Hypopnea Events Using Short-Term Electrocardiogram Recordings," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 6, pp. 1057–1067, 2009.
[175] A. Khandoker, M. Palaniswami, and C. Karmakar, "Support Vector Machines for Automated Recognition of Obstructive Sleep Apnea Syndrome From ECG Recordings," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 1, pp. 34–48, 2009.
[176] A. Khandoker, C. Karmakar, and M. Palaniswami, "Automated recognition of patients with obstructive sleep apnoea using wavelet-based features of electrocardiogram recordings," *Comput. Biol. Med.*, vol. 39, no. 1, pp. 88–96, 2009.
[177] B. Yılmaz, M. Asyalı, E. Arıkan, S. Yetkin, and F. Özgen, "Sleep stage and obstructive apneaic epoch classification using single-lead ECG," *Biomed. Eng. OnLine*, vol. 9, no. 39, pp. 1–14, 2010.
[178] J. MartÍnez-Vargas, L. Seúlveda-Cano, and G. Castellanos-Dominguez, "On Determining Available Stochastic Features by Spectral Splitting in Obstructive Sleep Apnea Detection," presented at the 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Massachusetts, USA, Sep. 2011.
[179] A. Yildiz, M. Akın, and M. Poyraz, "An expert system for automated recognition of patients with obstructive sleep apnea using electrocardiogram recordings," *Expert Syst. Appl.*, vol. 38, no. 10, pp. 12880–12890, Sep. 2011.
[180] K. Kesper, S. Canisius, T. Penzel, T. Ploch, and W. Cassel, "ECG signal analysis for the assessment of sleep-disordered breathing and sleep pattern," *Med. Biol. Eng. Comput.*, vol. 50, no. 2, pp. 135–144, Feb. 2012.
[181] A. Ravelo-Garcia *et al.*, "Cepstrum feature selection for the classification of Sleep Apnea-Hypopnea Syndrome based on heart rate variability," presented at the 2013 Computing in Cardiology Conference, Zaragoza, Spain, Sep. 2013.
[182] C. Travieso, J. Alonso, M. Pozo-Banõs, J. Ticay-Rivas, and K. L. de-Ipinã, "Automatic Apnea Identification by Transformation of the Cepstral Domain," *Cogn. Comput.*, vol. 5, no. 4, pp. 1–8, 2013.
[183] A. Ravelo-García *et al.*, "Symbolic dynamics marker of heart rate variability combined with clinical variables enhance obstructive sleep apnea screening," *Chaos*, vol. 24, no. 2, pp. 1–8, 2014.
[184] C. Maier, H. Wenz, and H. Dickhaus, "Robust detection of sleep apnea from Holter ECGs," *Methods Inf. Med.*, vol. 53, no. 4, pp. 303–307, 2014.
[185] H. Nguyen, B. Wilkins, Q. Cheng, and B. Benjamin, "An online sleep apnea detection method based on recurrence quantification analysis," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 4, pp. 1285–1293, 2014.

[186] V. Rachim, G. Li, and W. Chung, “Sleep apnea classification using ECG-signal wavelet-PCA features,” *Biomed. Mater. Eng.*, vol. 24, no. 6, pp. 2875–2882, 2014.
[187] G. Gutiérrez-Tobal, D. Álvarez, J. Gomez-Pilar, F. Campo, and R. Hornero, “Assessment of Time and Frequency Domain Entropies to Detect Sleep Apnoea in Heart Rate Variability Recordings from Men and Women,” *Entropy*, vol. 17, no. 1, pp. 123–141, 2015.
[188] A. Hassan, “Automatic screening of Obstructive Sleep Apnea from single-lead Electrocardiogram,” presented at the 2015 International Conference on Electrical Engineering and Information Communication Technology, Dhaka, Bangladesh, May 2015.
[189] A. Ravelo-García *et al.*, “Application of the Permutation Entropy over the Heart Rate Variability for the Improvement of Electrocardiogram-based Sleep Breathing Pause Detection,” *Entropy*, vol. 17, no. 3, pp. 914–927, 2015.
[190] L. Chen, X. Zhang, and C. Song, “An Automatic Screening Approach for Obstructive Sleep Apnea Diagnosis Based on Single-Lead Electrocardiogram,” *IEEE Trans. Autom. Sci. Eng.*, vol. 12, no. 1, pp. 106–115, 2015.
[191] C. Cheng, C. Kan, and H. Yang, “Heterogeneous recurrence analysis of heart beat dynamics for the identification of sleep apnea events,” *Comput. Biol. Med.*, vol. 75, no. 1, pp. 10–18, 2016.
[192] A. Hassan, “Computer-aided obstructive sleep apnea detection using normal inverse Gaussian parameters and adaptive boosting,” *Biomed. Signal Process. Control*, vol. 29, no. 1, pp. 22–30, 2016.
[193] C. Song, K. Liu, X. Zhang, L. Chen, and X. Xian, “An Obstructive Sleep Apnea Detection Approach Using a Discriminative Hidden Markov Model From ECG Signals,” *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1532–1542, 2016.
[194] A. Hassan and A. Haque, “An expert system for automated identification of obstructive sleep apnea from single-lead ECG using random under sampling boosting,” *Neurocomputing*, vol. 235, pp. 122–130, Apr. 2017.
[195] A. Smruthy and M. Suchetha, “Real-Time Classification of Healthy and Apnea Subjects Using ECG Signals With Variational Mode Decomposition,” *IEEE Sens. J.*, vol. 17, no. 10, pp. 3092–3099, May 2017.
[196] S. Martín-González, J. Navarro-Mesa, G. Juliá-Serdá, J. Kraemer, N. Wessel, and A. Ravelo-García, “Heart rate variability feature selection in the presence of sleep apnea: An expert system for the characterization and detection of the disorder,” *Comput. Biol. Med.*, vol. 91, no. 1, pp. 47–58, 2017.
[197] L. Chen and X. Zhang, “State-Based General Gamma CUSUM for Modeling Heart Rate Variability Using Electrocardiography Signals,” *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 2, pp. 1160–1171, 2017.
[198] M. Cheng, W. Sori, F. Jiang, A. Khan, and S. Liu, “Recurrent Neural Network Based classification of ECG signal Features for Obstruction of Sleep Apnea Detection,” presented at the 2017 IEEE International Conference on Computational Science and Engineering and Embedded and Ubiquitous Computing, Guangzhou, China, Aug. 2017.
[199] R. Pathinarupothi, R. Vinaykumar, R. Ekanath, E. Gopalakrishnan, and P. Soman, “Instantaneous heart rate as a robust feature for sleep apnea severity detection using deep learning,” presented at the 2017 IEEE International Conference on Biomedical & Health Informatics, Florida, USA, Feb. 2017.
[200] P. Caseiro, R. Fonseca-Pinto, and A. Andrade, “Screening of obstructive sleep apnea using Hilbert-Huang decomposition of oronasal airway pressure recordings,” *Med. Eng. Phys.*, vol. 32, no. 6, pp. 561–568, 2010.
[201] Y. Maali and A. Al-Jumaily, “Automated detecting sleep apnea syndrome: A novel system based on genetic SVM,” presented at the 2011 11th International Conference on Hybrid Intelligent Systems, Melacca, Malaysia, Dec. 2011.
[202] G. Gutiérrez-Tobal, R. Hornero, D. Álvarez, J. Marcos, and F. Campo, “Linear and nonlinear analysis of airflow recordings to help in sleep apnoea–hypopnoea syndrome diagnosis,” *Physiol. Meas.*, vol. 33, no. 7, pp. 1261–1275, 2012.
[203] B. Koley and D. Dey, “Automated detection of apnea and hypopnea events,” presented at the 2012 Third International Conference on Emerging Applications of Information Technology, West Bengal, India, Dec. 2012.
[204] N. Selvaraj and R. Narasimhan, “Detection of Sleep Apnea on a Per-Second Basis Using Respiratory Signals,” presented at the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Osaka, Japan, Jul. 2013.
[205] A. Thommandram, J. Eklund, and C. McGregor, “Detection of apnoea from respiratory time series data using clinically recognizable features and kNN classification,” presented at the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Osaka, Japan, Jul. 2013.

[206] B. Koley and D. Dey, “Real-Time Adaptive Apnea and Hypopnea Event Detection Methodology for Portable Sleep Apnea Monitoring Devices,” *IEEE Trans. Biomed. Eng.*, vol. 60, no. 12, pp. 3354–3363, 2013.
[207] C. Avci and A. Akbas, “Sleep apnea classification based on respiration signals by using ensemble methods,” *Biomed. Mater. Eng.*, vol. 26, no. 1, pp. S1703–S1710, 2015.
[208] G. Ozdemir, H. Nasifoglu, and O. Erogul, “A Time-Series Approach to Predict Obstructive Sleep Apnea (OSA) Episodes,” presented at the The 2nd World Congress on Electrical Engineering and Computer Systems and Science, Budapest, Hungary, Jul. 2016.
[209] P. Minu and M. Amithab, “SAHS Detection Based on ANFIS Using Single Channel Airflow Signal,” *Int. J. Innov. Res. Sci. Eng. Technol.*, vol. 5, no. 7, pp. 13053–13061, 2016.
[210] R. Haidar, I. Koprinska, and B. Jeffries, “Sleep Apnea Event Detection from Nasal Airflow Using Convolutional Neural Networks,” presented at the 2017 International Conference on Neural Information Processing, Guangzhou, China, Oct. 2017.
[211] A. Ng, T. Koh, E. Baey, T. Lee, U. Abeyratne, and K. Puvanendran, “Could formant frequencies of snore signals be an alternative means for the diagnosis of obstructive sleep apnea?,” *Sleep Med.*, vol. 9, no. 8, pp. 894–898, 2008.
[212] R. Pozo, J. Murillo, L. Gómez, E. Gonzalo, J. Ramírez, and D. Toledano, “Assessment of Severe Apnoea through Voice Analysis, Automatic Speech, and Speaker Recognition Techniques,” *EURASIP J. Adv. Signal Process.*, vol. 2009, no. 1, pp. 1–11, 2009.
[213] O. Elisha, A. Tarasiuk, and Y. Zigel, “Automatic Detection of Obstructive Sleep Apnea Using Speech Signals,” *IEEE Trans. Biomed. Eng.*, vol. 58, no. 5, pp. 1373–1382, 2011.
[214] A. Karunajeewa, U. Abeyratne, and C. Hukins, “Multi-feature snore sound analysis in obstructive sleep apnea-hypopnea syndrome,” *Physiol. Meas.*, vol. 32, no. 1, pp. 83–97, 2011.
[215] L. Almazaydeh, K. Elleithy, M. Faezipour, and A. Abushakra, “Apnea Detection Based on Respiratory Signal Classification,” presented at the The 3rd International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare, Ontario, Canada, Oct. 2013.
[216] A. Benavides, R. Pozo, D. Toledano, J. Murillo, E. Gonzalo, and L. Gómez, “Analysis of voice features related to obstructive sleep apnoea andtheir application in diagnosis support,” *Comput. Speech Lang.*, vol. 28, no. 2, pp. 434–452, Mar. 2014.
[217] T. Rosenwein, E. Dafna, A. Tarasiuk, and Y. Zigel, “Breath-by-breath detection of apneic events for OSA severity estimation using non-contact audio recordings,” presented at the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Milan, Italy, Aug. 2015.
[218] T. Praydas, B. Wongkittisuksa, and S. Tanthanuch, “Obstructive Sleep Apnea Severity Multiclass Classification Using Analysis of Snoring Sounds,” presented at the 2nd World Congress on Electrical Engineering and Computer Systems and Science, Budapest, Hungary, Aug. 2016.
[219] C. Kalkbrenner, M. Eichenlaub, S. Rudiger, C. Kropf-Sanchen, W. Rottbauer, and R. Brucher, “Apnea and heart rate detection from tracheal body sounds for the diagnosis of sleep-related breathing disorders,” *Med. Biol. Eng. Comput.*, vol. 56, no. 4, pp. 671–681, 2018.
[220] C. Zamarrón, F. Gude, J. Barcala, J. Rodriguez, and P. Romero, “Utility of oxygen saturation and heart rate spectral analysis obtained from pulse oximetric recordings in the diagnosis of sleep apnea syndrome,” *Chest*, vol. 123, no. 5, pp. 1567–1576, 2003.
[221] J. Tian and J. Liu, “Apnea Detection Based on Time Delay Neural Network,” presented at the 27th Annual International Conference of the Engineering in Medicine and Biology Society, Shanghai, China, Sep. 2005.
[222] K. Polat, S. Yosunkaya, and S. Günes, “Comparison of Different Classifier Algorithms on the Automated Detection of Obstructive Sleep Apnea Syndrome,” *J. Med. Syst.*, vol. 32, no. 3, pp. 243–250, Jun. 2008.
[223] D. Álvarez, R. Hornero, V. Marcos, F. Campo, and M. López, “Spectral analysis of electroencephalogram and oximetric signals in obstructive sleep apnea diagnosis,” presented at the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009, New Jersey, USA, Sep. 2009.
[224] A. Yadollahi, E. Giannouli, and Z. Moussavi, “Sleep apnea monitoring and diagnosis based on pulse oximetry and tracheal sound signals,” *Med. Biol. Eng. Comput.*, vol. 48, no. 11, pp. 1087–1097, 2010.
[225] L. Poupard, C. Philippe, M. Goldman, R. Sartène, and M. Mathieu, “Novel mathematical processing method of nocturnal oximetry for screening patients with suspected sleep apnoea syndrome,” *Sleep Breath*, vol. 16, no. 2, pp. 419–425, 2012.

[226] D. Sommermeyer, D. Zou, L. Grote, and J. Hedner, “Detection of sleep disordered breathing and its central/obstructive character using nasal cannula and finger pulse oximeter,” *J. Clin. Sleep Med.*, vol. 8, no. 5, pp. 527–533, 2012.
[227] B. Xie and H. Minn, “Real-Time Sleep Apnea Detection by Classifier Combination,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 3, pp. 469–477, 2012.
[228] H. Al-Angari and A. Sahakian, “Automated recognition of obstructive sleep apnea syndrome using support vector machine classifier,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 3, pp. 463–468, 2012.
[229] F. Espinoza-Cuadros, R. Fernández-Pozo, D. Toledano, J. Alcázar-Ramírez, E. López-Gonzalo, and L. Hernández-Gómez, “Speech Signal and Facial Image Processing for Obstructive Sleep Apnea Assessment,” *Comput. Math. Methods Med.*, vol. 2015, no. 1, pp. 1–13, 2015.
[230] K. Madhav, E. Krishna, and K. Reddy, “Detection of sleep apnea from multiparameter monitor signals using empirical mode decomposition,” presented at the 2017 International Conference on Computer, Communication and Signal Processing, Tamilnadu, India, Jan. 2017.
[231] G. Memis and M. Sert, “Multimodal Classification of Obstructive Sleep Apnea Using Feature Level Fusion,” presented at the 2017 IEEE 11th International Conference on Semantic Computing, California, USA, Jan. 2017.
[232] Y. Ichimaru and G. Moody, “Development of the polysomnographic database on CD-ROM,” *Psychiatry Clin. Neurosci.*, vol. 53, no. 2, pp. 175–177, 1999.
[233] A. Goldberger *et al.*, “PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research,” *Circulation*, vol. 101, no. 23, pp. 215–220, 2000.
[234] S. Quan *et al.*, “The Sleep Heart Health Study: design, rationale, and methods,” *Sleep*, vol. 20, no. 12, pp. 1077–1085, 1997.
[235] D. Dean *et al.*, “Scaling Up Scientific Discovery in Sleep Medicine: The National Sleep Research Resource,” *Sleep*, vol. 29, no. 5, pp. 1151–1164, 2016.
[236] W. Flemons *et al.*, “Home Diagnosis of Sleep Apnea: A Systematic Review of the Literature,” *Chest*, vol. 124, no. 4, pp. 1543–1579, 2003.
[237] N. Collop *et al.*, “Obstructive Sleep apnea Devices for Out-Of- Center (OOC) testing: technology Evaluation,” *J. Clin. Sleep Med.*, vol. 7, no. 5, pp. 531–548, 2011.
[238] E. Kaniusas, L. Mehnen, H. Pfützner, B. Saletu, and R. Popovic, “EVALUATION OF ACOUSTIC SIGNALS FOR THE DETECTION OF SLEEP APNEA EVENTS,” *Proc. Int. Meas. Confed.*, vol. 7, pp. 177–182, 2002.
[239] H. Nakano, M. Hayashi, E. Ohshima, N. Nishikata, and T. Shinohara, “Validation of a new system of tracheal sound analysis for the diagnosis of sleep apnea-hypopnea syndrome,” *Sleep*, vol. 27, no. 5, pp. 951–957, 2004.
[240] D. Mack, M. Alwan, B. Turner, P. Suratt, and R. Felder, “A Passive and Portable System for Monitoring Heart Rate and Detecting Sleep Apnea and Arousals: Preliminary Validation,” Virginia, USA, Apr. 2006.
[241] C. Heneghan *et al.*, “A Portable Automated Assessment Tool for Sleep Apnea Using a Combined Holter-Oximeter,” *Sleep*, vol. 31, no. 10, pp. 1432–1439, 2008.
[242] A. Otero, S. Dapena, P. Félix, J. Presedo, and M. Tarascó, “A Low Cost Screening Test for Obstructive Sleep Apnea that can be Performed at the Patient’s Home,” Budapest, Hungary, Aug. 2009.
[243] A. Burgos, A. Gõni, A. Illarramendi, and J. Bermúdez, “Real-Time Detection of Apneas on a PDA,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 4, pp. 995–1002, 2010.
[244] J. Shin, Y. Chee, and D. Jeong, “Nonconstrained Sleep Monitoring System and Algorithms Using Air-Mattress With Balancing Tube Method,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 1, pp. 147–156, 2010.
[245] A. Yilmaz and T. Dündar, “Home recording for pre-phase sleep apnea diagnosis by Holter recorder using MMC memory,” presented at the 2010 IEEE International Conference on Virtual Environments Human-Computer Interfaces and Measurement Systems, Taranto, Italy, Sep. 2010.
[246] M. Bsoul, H. Minn, and L. Tamil, “Apnea MedAssist: Real-time Sleep Apnea Monitor Using Single-Lead ECG,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 3, pp. 416–427, 2011.
[247] Y. Zhao, H. Zhang, W. Liu, and S. Ding, “A snoring detector for OSAHS based on patient’s individual personality,” Dalian, China, Sep. 2011.
[248] T. Dündar, A. Yılmaz, and Ö. Çaglar, “Sleep apnea detection for prephase diagnosis using third level holter recording device,” Antalya, Turkey, Apr. 2011.
[249] J. Zhang, Q. Zhang, Y. Wang, and C. Qiu, “A real-time auto-adjustable smart pillow system for sleep apnea detection and treatment,” Pennsylvania, USA, Apr. 2013.

[250] A. Garde, W. Karlen, P. Dehkordi, D. Wensley, J. Ansermino, and G. Dumont, “Oxygen saturation in children with and without obstructive sleep apnea using the phone-oximeter,” Osaka, Japan, Jul. 2013.
[251] M. Al-Mardini, F. Aloul, A. Sagahyroon, and L. Al-Husseini, “On the Use of Smartphones for Detecting Obstructive Sleep Apnea,” Chania, Greece, Nov. 2013.
[252] S. Hwang *et al.*, “Unconstrained Sleep Apnea Monitoring Using Polyvinylidene Fluoride Film-Based Sensor,” *IEEE Trans. Biomed. Eng.*, vol. 61, no. 7, pp. 2125–2134, 2014.
[253] E. Rodriguez-Villegas, G. Chen, J. Radcliffe, and J. Duncan, “A pilot study of a wearable apnoea detection device,” *BMJ Open*, vol. 4, no. 10, p. e005299, 2014.
[254] O. Salem, Y. Liu, and A. Mehaoua, “Pervasive detection of sleep apnea using medical wireless sensor networks,” Rio Grande do Norte, Brazil, Oct. 2014.
[255] A. Garde, P. Dehkordi, D. Wensley, J. Ansermino, and G. Dumont, “Pulse oximetry recorded from the Phone Oximeter for detection of obstructive sleep apnea events with and without oxygen desaturation in children,” Milan, Italy, Aug. 2015.
[256] J. Jin and E. Sánchez-Sinencio, “A Home Sleep Apnea Screening Device With Time-Domain Signal Processing and Autonomous Scoring Capability,” *IEEE Trans. Biomed. Circuits Syst.*, vol. 9, no. 1, pp. 96–104, 2015.
[257] R. Crupi, T. Faetti, and R. Paradiso, “Preliminary evaluation of Wearable Wellness System for Obstructive Sleep Apnea Detection,” Milan, Italy, Aug. 2015.
[258] R. Nandakumar, S. Gollakota, and N. Watson, “Contactless Sleep Apnea Detection on Smartphones,” presented at the 13th Annual International Conference on Mobile Systems, Applications, and Services, Florence, Italy, May 2015.
[259] J. Behar *et al.*, “SleepAp: An Automated Obstructive Sleep Apnoea Screening Application for Smartphones,” *IEEE J. Biomed. Health Inform.*, vol. 19, no. 1, pp. 325–331, 2015.
[260] G. Surrel, F. Rincón, D. Atienza, and S. Murali, “Low-Power Wearable System for Real-Time Screening of Obstructive Sleep Apnea,” Pennsylvania, USA, Jul. 2016.
[261] R. Golpe, A. Jiménez, and R. Carpizo, “Home sleep studies in the assessment of sleep apnea/hypopnea syndrome,” *Chest*, vol. 122, no. 4, pp. 1156–1161, 2002.
[262] T. Shochat *et al.*, “The SleepStrip: an apnoea screener for the early detection of sleep apnoea syndrome,” *Eur. Respir. J.*, vol. 19, pp. 121–126, 2002.
[263] J. Reichert, D. Bloch, E. Cundiff, and B. Votteri, “Comparison of the NovaSom QSG, a new sleep apnea home-diagnostic system, and polysomnography,” *Sleep Med.*, vol. 4, no. 3, pp. 213–218, 2003.
[264] R. Núñez *et al.*, “Estudio de la validez de un equipo de poligrafía respiratoria (BREAS SC-20) para el diagnóstico del síndrome de apneas-hipopneas durante el sueño,” *Arch. Bronconeumol.*, vol. 39, no. 12, pp. 537–543, 2003.
[265] K. Dingli *et al.*, “Evaluation of a portable device for diagnosing the sleep apnoea/hypopnoea syndrome,” *Eur. Respir. J.*, vol. 21, no. 2, pp. 253–259, 2003.
[266] E. Quintana-Gallego *et al.*, “Home respiratory polygraphy for diagnosis of sleep-disordered breathing in heart failure,” *Eur. Respir. J.*, vol. 24, no. 3, pp. 443–448, 2004.
[267] S. Su, F. Baroody, M. Kohrman, and D. Suskind, “A comparison of polysomnography and a portable home sleep study in the diagnosis of obstructive sleep apnea syndrome,” *Otolaryngol. Head Neck Surg.*, vol. 131, no. 6, pp. 844–850, 2004.
[268] A. Candela *et al.*, “Validation of a Respiratory Polygraphy System in the Diagnosis of Sleep Apnea Syndrome,” *Arch. Bronconeumol.*, vol. 41, no. 2, pp. 71–77, 2005.
[269] W. Abraham *et al.*, “Validation and Clinical Utility of a Simple In-Home Testing Tool for Sleep-Disordered Breathing and Arrhythmias in Heart Failure: Results of the Sleep Events, Arrhythmias, and Respiratory Analysis in Congestive Heart Failure (SEARCH) Study,” *Congest Heart Fail*, vol. 12, no. 5, pp. 241–247, 2006.
[270] K. Pang, C. Gourin, and T. Gourin, “A comparison of polysomnography and the WatchPAT in the diagnosis of obstructive sleep apnea,” *Otolaryngol. Head Neck Surg.*, vol. 137, no. 4, pp. 665–668, 2007.
[271] V. Jobin, P. Mayer, and F. Bellemare, “Predictive value of automated oxygen saturation analysis for the diagnosis and treatment of obstructive sleep apnoea in a home-based setting,” *Thorax*, vol. 62, no. 5, pp. 422–427, 2007.
[272] M. Álvarez *et al.*, “Fiabilidad de la poligrafía respiratoria domiciliaria para el diagnóstico del síndrome de apneas-hipopneas durante el sueño. Análisis de costes,” *Arch. Bronconeumol.*, vol. 44, no. 1, pp. 22–28, 2008.
[273] S. Grover and S. Pittman, “Automated detection of sleep disordered breathing using a nasal pressure monitoring device,” *Sleep Breath*, vol. 12, no. 4, pp. 339–345, 2008.

[274] I. Ayappa, R. Norman, V. Seelall, and D. Rapoport, "Validation of a self-applied unattended monitor for sleep disordered breathing," *J. Clin. Sleep Med.*, vol. 4, no. 1, pp. 26–37, 2008.
[275] K. Wong *et al.*, "Diagnostic test evaluation of a nasal flow monitor for obstructive sleep apnea detection in sleep apnea research," *Behav. Res. Methods*, vol. 40, no. 1, pp. 360–366, 2008.
[276] J. Pépin, P. Defaye, E. Vincent, S. Christophle-Boulard, R. Tamisier, and P. Lévy, "Sleep apnea diagnosis using an ECG Holter device including a nasal pressure (NP) recording: validation of visual and automatic analysis of nasal pressure versus full polysomnography," *Sleep Med.*, vol. 10, no. 6, pp. 651–656, 2009.
[277] J. Gjevre, R. Taylor-Gjevre, R. Skomro, J. Reid, M. Fenton, and D. Cotton, "Comparison of polysomnographic and portable home monitoring assessments of obstructive sleep apnea in Saskatchewan women," *Can. Respir. J.*, vol. 18, no. 5, pp. 271–274, 2011.
[278] S. Ng *et al.*, "Validation of a portable recording device (ApneaLink) for identifying patients with suspected obstructive sleep apnoea syndrome," *Intern. Med. J.*, vol. 39, no. 11, pp. 757–762, 2009.
[279] S. Goodrich and W. Orr, "An investigation of the validity of the Lifeshirt in comparison to standard polysomnography in the detection of obstructive sleep apnea," *Sleep Med.*, vol. 10, no. 1, pp. 118–122, 2009.
[280] A. Oliveira *et al.*, "Diagnosis of obstructive sleep apnea syndrome and its outcomes with home portable monitoring," *Chest*, vol. 135, no. 2, pp. 330–336, 2009.
[281] C. Nigro, S. Aimaretti, S. Gonzalez, and E. Rhodius, "Validation of the WristOx 3100™ oximeter for the diagnosis of sleep apnea/hypopnea syndrome," *Sleep Breath*, vol. 13, no. 2, pp. 127–136, 2009.
[282] R. Santos-Silva *et al.*, "Validation of a Portable Monitoring System for the Diagnosis of Obstructive Sleep Apnea Syndrome," *Sleep*, vol. 32, no. 5, pp. 629–636, 2009.
[283] A. Zaffaroni, P. Chazal, C. Heneghan, P. Boyle, P. Ronayne, and W. McNicholas, "SleepMinder: An Innovative Contact-Free Device for the Estimation of the Apnoea-Hypopnoea Index," Minnesota, USA, Sep. 2009.
[284] N. Takama and M. Kurabayashi, "Effectiveness of a portable device and the need for treatment of mild-to-moderate obstructive sleep-disordered breathing in patients with cardiovascular disease," *J. Cardiol.*, vol. 56, no. 1, pp. 73–78, 2010.
[285] H. Driver *et al.*, "Validation of the MediByte type 3 portable monitor compared with polysomnography for screening of obstructive sleep apnea," *Can. Respir. J.*, vol. 18, no. 3, pp. 137–143, 2011.
[286] D. Lesser, G. Haddad, R. Bush, and M. Pian, "The utility of a portable recording device for screening of obstructive sleep apnea in obese adolescents," *J. Clin. Sleep Med.*, vol. 15, no. 8, pp. 271–277, 2012.
[287] C. Morales *et al.*, "In-home, self-assembled sleep studies are useful in diagnosing sleep apnea in the elderly," *Sleep*, vol. 35, no. 11, pp. 1491–1501, 2012.
[288] M. Kobayashi *et al.*, "Validity of sheet-type portable monitoring device for screening obstructive sleep apnea syndrome," *Sleep Breath*, vol. 17, no. 2, pp. 589–595, 2013.
[289] N. Garg, A. Rolle, T. Lee, and B. Prasad, "Home-based Diagnosis of Obstructive Sleep Apnea in an Urban Population," *J. Clin. Sleep Med.*, vol. 10, no. 8, pp. 879–885, 2014.
[290] A. Guerrero *et al.*, "Management of sleep apnea without high pretest probability or with comorbidities by three nights of portable sleep monitoring," *Sleep*, vol. 37, no. 8, pp. 1363–1373, 2014.
[291] A. Cairns, E. Wickwire, E. Schaefer, and D. Nyanjom, "A pilot validation study for the NOX T3 portable monitor for the detection of OSA," *Sleep Breath*, vol. 18, no. 3, pp. 609–614, 2014.
[292] N. Norman, S. Middleton, O. Erskine, P. Middleton, J. Wheatley, and C. Sullivan, "Validation of the Sonomat: a contactless monitoring system used for the diagnosis of sleep disordered breathing," *Sleep*, vol. 37, no. 9, pp. 1477–1487, 2014.
[293] J. Zou *et al.*, "Evaluation of a 2-channel portable device and a predictive model to screen for obstructive sleep apnea in a laboratory environment," *Respir. Care*, vol. 60, no. 3, pp. 356–362, 2015.
[294] L. Meng, H. Xu, J. Guan, H. Yi, H. Wu, and S. Yin, "Validation of a novel sleep-monitoring system for diagnosing obstructive sleep apnea: A comparison with polysomnography," *Exp. Ther. Med.*, vol. 12, no. 5, pp. 2937–2941, 2016.
[295] S. December *et al.*, "The Accuracy of Portable Monitoring in Diagnosing Significant Sleep Disordered Breathing in Hospitalized Patients," *PLoS One*, vol. 11, no. 12, p. e0168073, 2016.
[296] H. Hara, M. Tsutsumi, S. Tarumoto, T. Shiga, and H. Yamashita, "Validation of a new snoring detection device based on a hysteresis extraction algorithm," *Auris. Nasus. Larynx*, vol. 24, no. 5, Oct. 2017, doi: http://dx.doi.org/10.1016/j.anl.2016.12.009.
[297] J. Durán-Cantolla *et al.*, "Validation of a new domiciliary diagnosis device for automatic diagnosis of patients with clinical suspicion of OSA," *Respirology*, vol. 22, no. 2, pp. 378–385, 2017.

[298] J. Blank *et al.*, "Overview of recruitment for the osteoporotic fractures in men study (MrOS)," *Contemp. Clin. Trials*, vol. 26, no. 5, pp. 557–568, 2005.
[299] F. Mendonça, S. Mostafa, F. Morgado-Dias, G. Juliá-Serdá, and A. Ravelo-García, "A Method for Sleep Quality Analysis Based on CNN Ensemble With Implementation in a Portable Wireless Device," *IEEE Access*, vol. 8, no. 1, pp. 158523–158537, 2020.
[300] F. Chapotot and G. Becq, "Automated sleep–wake staging combining robust feature extraction, artificial neural network classification, and flexible decision rules," *Int. J. Adapt. Control Signal Process.*, vol. 24, no. 5, pp. 409–423, 2010.
[301] I. Digital Signal Processing Committee, *Programs for Digital Signal Processing*. New York, USA: IEEE press, 1979.
[302] J. Pan and W. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. 32, no. 3, pp. 230–236, 1985.
[303] K. Murphy, *Machine Learning: A Probabilistic Perspective*. Massachusetts: The MIT Press, 2012.
[304] C. Bishop, *Pattern Recognition and Machine Learning*. Berlin, Germany: Springer, 2006.
[305] R. Schapire and Y. Singer, "Improved boosting algorithms using confidence-rated predictions," *Mach. Learn.*, vol. 37, no. 3, pp. 297–336, 1999.
[306] M. Warmuth, J. Liao, and G. Rätsch, "Totally corrective boosting algorithms that maximize the margin," presented at the 23rd international conference on Machine learning, Pennsylvania, USA, Jun. 2006.
[307] S. Ayalew, M. Babu, and L. Rao, "Comparison of New Approach Criteria for Estimating the Order of Autoregressive Process," *IOSR J. Math.*, vol. 1, no. 3, pp. 10–20, 2012.
[308] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Massachusetts: The MIT Press, 2016.
[309] M. Moller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Netw.*, vol. 6, no. 4, pp. 525–533, 1993.
[310] S. Goyal and G. Goyal, "Cascade and Feedforward Backpropagation Artificial Neural Network Models For Prediction of Sensory Quality of Instant Coffee Flavoured Sterilized Drink," *Can. J. Artif. Intell. Mach. Learn. Pattern Recognit.*, vol. 2, no. 6, pp. 78–82, 2011.
[311] S. Hochreiter and J. Schmidhuber, "Long Short-term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[312] K. Cho *et al.*, "Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation," presented at the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, Oct. 2014.
[313] G. Klambauer, T. Unterthiner, and A. Mayr, "Self-Normalizing Neural Networks," presented at the 31st Conference on Neural Information Processing Systems (NIPS 2017), California, USA, Dec. 2017.
[314] B. Curry and P. Morgan, "Evaluating Kohonen's learning rule: An approach through genetic algorithms," *Eur. J. Oper. Res.*, vol. 154, no. 1, pp. 191–205, 2004.
[315] B. Olshausen and D. Field, "Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1," *Vision Res.*, vol. 37, no. 23, pp. 3311–3325, 1997.
[316] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and minredundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, 2005.
[317] S. Mostafa, F. Morgado-Dias, and A. Ravelo-García, "Comparison of SFS and mRMR for oximetry feature selection in obstructive sleep apnea detection," *Neural Comput. Appl.*, pp. 1–21, 2018, doi: 10.1007/s00521-018-3455-8.
[318] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," presented at the 14th international joint conference on Artificial intelligence, Quebec, Canada, Aug. 1995.
[319] B. Efron and R. Tibshirani, "Improvements on Cross-Validation: The .632 + Bootstrap Method," *J. Am. Stat. Assoc.*, vol. 92, no. 138, pp. 548–560, 1997.
[320] D. Sackett, R. Haynes, G. Guyatt, and P. Tugwell, *Clinical Epidemiology: A Basic Science for Clinical Medicine*, 2nd ed. Pennsylvania, USA: Lippincott Williams and Wilkins, 1991.
[321] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, 2006.
[322] J. Tenorio *et al.*, "A Novel Method to Assist the Detection of the Cyclic Alternating Pattern (CAP)," presented at the 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, California, USA, Aug. 2012.
[323] C. Navona, U. Barcaro, E. Bonanni, F. Martino, M. Maestri, and L. Murri, "An automatic method for the recognition and classification of the A-phases of the cyclic alternating pattern," *Clin. Neurophysiol.*, vol. 113, no. 11, pp. 1826–1831, 2002.

[324] U. Barcaro, E. Bonanni, M. Maestri, L. Murri, L. Parrino, and M. Terzano, “A general automatic method for the analysis of NREM sleep microstructure,” *Sleep Med.*, vol. 5, no. 6, pp. 567–576, 2004.
[325] F. Machado, F. Sales, C. Bento, A. Dourado, and C. Teixeira, “Automatic identification of Cyclic Alternating Pattern (CAP) sequences based on the Teager Energy Operator,” presented at the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Milan italy, Aug. 2015.
[326] H. Niknazar, S. Seifpour, M. Mikaili, A. Nasrabadi, and A. Banaraki, “A Novel Method to Detect the A Phases of Cyclic Alternating Pattern (CAP) Using Similarity Index,” presented at the 23rd Iranian Conference on Electrical Engineering, Tehran, Iran, May 2015.
[327] M. Mendez *et al.*, “On separability of A-phases during the cyclic alternating pattern,” presented at the 36th Annual International Conference of the IEEE engineering in Medicine and Biology Society, Chicago, USA, Aug. 2014.
[328] S. Mariani, A. Grassi, M. Mendez, L. Parrino, M. Terzano, and A. Bianchi, “Automatic detection of CAP on central and fronto-central EEG leads via Support Vector Machines,” presented at the 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Massachusetts, USA, Aug. 2011.
[329] S. Mariani *et al.*, “EEG segmentation for improving automatic CAP detection,” *Clin. Neurophysiol.*, vol. 124, no. 9, pp. 1815–1823, 2013.
[330] S. Mariani *et al.*, “Efficient automatic classifiers for the detection of A phases of the cyclic alternating pattern in sleep,” *Med. Biol. Eng. Comput.*, vol. 50, no. 4, pp. 359–372, 2012.
[331] F. Machado, C. Teixeira, C. Santos, C. Bento, F. Sales, and A. Dourado, “A-phases subtype detection using different classification methods,” presented at the IEEE 38th Annual International Conference of the Engineering in Medicine and Biology Society, Florida, USA, Aug. 2016.
[332] J. Lima and A. Rosa, “Classification of microstructure of human sleep using EEG modelling,” presented at the 9th European Signal Processing Conference, Rhodes, Greece, Sep. 1998.
[333] R. Largo, C. Munteanu, and A. Rosa, “Wavelet based CAP detector with GA tuning,” presented at the 6th wseas international conference on evolutionary computing, Lisbon, Portugal, Jun. 2005.
[334] R. Largo, C. Munteanu, and A. Rosa, “CAP Event Detection by Wavelets and GA Tuning,” presented at the 2005 IEEE International Workshop on Intelligent Signal Processing, Faro, Portugal, Sep. 2005.
[335] A. Rosa, L. Parrino, and M. Terzano, “Automatic detection of cyclic alternating pattern (CAP) sequences in sleep: preliminary results,” *Clin. Neurophysiol.*, vol. 110, no. 4, pp. 585–592, 1999.
[336] F. Karimzadeh, E. Seraj, R. Boostani, and M. Torabi-Nami, “Presenting Efficient Features for Automatic CAP Detection in Sleep EEG Signals,” presented at the 38th International Conference on Telecommunications and Signal Processing, Prague, Czech Republic, Jul. 2015.
[337] S. Aydin, H. Saraoglu, and S. Kara, “Log Energy Entropy-Based EEG Classification with Multilayer Neural Networks in Seizur,” *Ann. Biomed. Eng.*, vol. 37, no. 12, pp. 2626–2630, 2009.
[338] S. Sarraf, “EEG-Based Movement Imagery Classification Using Machine Learning Techniques and Welch’s Power Spectral Density Estimation,” vol. 33, no. 1, pp. 124–145, 2017.
[339] M. Ortigueira, *Processamento Digital de Sinais*, 1st ed. Lisboa, Portugal: Fundação Calouste Gulbenkian, 2005.
[340] R. Shumway and D. Stoffer, *Time Series Analysis and Its Applications*. Basingstoke, United Kingdom: Springer International Publishing, 2017.
[341] S. Mostafa, F. Mendonça, A. Ravelo-García, and F. Morgado-Dias, “Combination of Deep and Shallow Networks for Cyclic Alternating Patterns Detection,” presented at the 2018 13th APCA International Conference on Automatic Control and Soft Computing (CONTROLO), Ponta Delgada, Portugal, Jun. 2018.
[342] A. Craik, Y. He, and J. Contreras-Vidal, “Deep learning for electroencephalogram (EEG) classification tasks: a review,” *J. Neural Eng.*, vol. 16, no. 3, pp. 1–28, 2019.
[343] D. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” presented at the 3rd International Conference on Learning Representations, California, USA, May 2015.
[344] F. Mendonça, S. Mostafa, F. Morgado-Dias, and A. Ravelo-Garcia, “Cyclic Alternating Pattern Estimation from One EEG Monopolar Derivation Using a Long Short-Term Memory,” presented at the 2019 International Conference in Engineering Applications (ICEA), São Miguel, Portugal, Jul. 2019.
[345] T. Steenkiste, W. Groenendaal, D. Deschrijver, and T. Dhaene, “Automated Sleep Apnea Detection in Raw Respiratory Signals using Long Short-Term Memory Neural Networks,” *IEEE J. Biomed. Health Inform.*, doi: 10.1109/JBHI.2018.2886064.

[346] J. Nagi *et al.*, “Max-pooling convolutional neural networks for vision-based hand gesture recognition,” presented at the 2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), Kuala Lumpur, Malaysia, Nov. 2011.
[347] L. Kaifeng, J. Shunhua, and L. Jian, “Learning Gradient Descent: Better Generalization and Longer Horizons,” presented at the 34 th International Conference on Machine Learning, Sydney, Australia, Aug. 2017.
[348] J. Duchi, E. Hazan, and Y. Singer, “Adaptive Subgradient Methods for Online Learning and Stochastic Optimization,” *J. Mach. Learn. Res.*, vol. 12, no. 1, pp. 2121–2159, 2011.
[349] S. Himanen and J. Hasan, “Limitations of Rechtschaffen and Kales,” *Sleep Med. Rev.*, vol. 4, no. 2, pp. 149–167, 2000.
[350] H. Schulz, “Rethinking Sleep Analysis: Comment on the AASM Manual for the Scoring of Sleep and Associated Events,” *J. Clin. Sleep Med.*, vol. 4, no. 2, pp. 99–103, 2008.
[351] F. Mendonça, S. Mostafa, F. Morgado-Dias, and A. Ravelo-García, “Cyclic alternating pattern estimation based on a probabilistic model over an EEG signal,” *Biomed. Signal Process. Control*, vol. 62, no. 1, p. 102063, 2020.
[352] D. Arthur and S. Vassilvitskii, “k-means++: the advantages of careful seeding,” presented at the Eighteenth annual ACM-SIAM symposium on Discrete algorithms, Louisiana, USA, Jan. 2007.
[353] G. Lee and C. Scott, “EM algorithms for multivariate Gaussian mixture models with truncated and censored data,” *Comput. Stat. Data Anal.*, vol. 56, no. 9, pp. 2816–2829, 2012.
[354] R. Durbin, S. Eddy, A. Krogh, and G. Mitchison, *Biological Sequence Analysis*. Cambridge, UK: Cambridge University Press, 1998.
[355] N. Marwan and C. Webber, “Mathematical and Computational Foundations of Recurrence Quantifications,” in *Recurrence Quantification Analysis - Theory and Best Practices*, Cham, Switzerland: Springer Nature, 2015, pp. 3–43.
[356] K. Becker *et al.*, “Anaesthesia Monitoring by Recurrence Quantification Analysis of EEG Data,” *PLoS One*, vol. 5, no. 1, p. e8876, 2010.
[357] M. Thiel, M. Romano, P. Read, and J. Kurths, “Estimation of dynamical invariants without embedding by recurrence plots,” *Chaos*, vol. 14, no. 2, pp. 234–243, 2004.
[358] S. Wallot, “Recurrence Quantification Analysis of Processes and Products of Discourse: A Tutorial in R,” *Discourse Process.*, vol. 54, no. 5–6, pp. 382–405, 2017.
[359] M. Kennel, R. Brown, and H. Abarbanel, “Determining embedding dimension for phase-space reconstruction using a geometrical construction,” *Phys. Rev. A*, vol. 45, no. 1, pp. 3403–3411, 1992.
[360] N. Marwan, M. Romano, M. Thiel, and J. Kurths, “Recurrence plots for the analysis of complex systems,” *Phys. Rep.*, vol. 438, no. 5–6, pp. 237–329, 2007.
[361] M. Thiel, M. Romano, J. Kurths, R. Meucci, E. Allaria, and F. Arecchi, “Influence of observational noise on the recurrence quantification analysis,” *Phys. Nonlinear Phenom.*, vol. 171, no. 3, pp. 138–152, 2002.
[362] N. Marwan, “How to avoid potential pitfalls in recurrence plot based data analysis,” *Int. J. Bifurc. Chaos*, vol. 21, no. 4, pp. 1003–1017, 2011.
[363] G. Lancaster, D. Iatsenko, A. Pidde, V. Ticcinelli, and A. Stefanovska, “Surrogate data for hypothesis testing of physical systems,” *Phys. Rep.*, vol. 748, no. 1, pp. 1–60, 2018.
[364] C. Stam, J. Pijn, P. Suffczynski, and F. Silva, “Dynamics of the human alpha rhythm: evidence for non-linearity?,” *Clin. Neurophysiol.*, vol. 110, no. 10, pp. 1801–1813, 1999.
[365] H. Helakari *et al.*, “Spectral entropy indicates electrophysiological and hemodynamic changes in drug-resistant epilepsy – A multimodal MREG study,” *NeuroImage Clin.*, vol. 22, no. 1, p. 101763, 2019.
[366] F. Mendonça, S. Mostafa, F. Morgado-Dias, and A. Ravelo-García, “On the use of patterns obtained from LSTM and feature-based methods for time series analysis: application in automatic classification of the CAP A phase subtypes,” *J. Neural Eng.*, 2020.
[367] H. Niknazar, K. Maghooli, and A. Nasrabadi, “Epileptic Seizure Prediction using Statistical Behavior of Local Extrema and Fuzzy Logic System,” *Int. J. Comput. Appl.*, vol. 113, no. 2, pp. 24–30, 2015.
[368] S. Hartmann and M. Baumert, “Automatic A-Phase Detection of Cyclic Alternating Patterns in Sleep Using Dynamic Temporal Information,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 9, pp. 1695–1703, 2019.
[369] A. Yadav, C. Jha, and A. Sharan, “Optimizing LSTM for time series prediction in Indian stock market,” *Procedia Comput. Sci.*, vol. 167, no. 1, pp. 2091–2100, 2020.
[370] T. Cover, “The Best Two Independent Measurements Are Not the Two Best,” *IEEE Trans. Syst. Man Cybern.*, vol. SMC-4, no. 1, pp. 116–117, 1974.

[371] R. Yamashita, M. Nishio, R. Do, and K. Togashi, “Convolutional neural networks: an overview and application in radiology,” *Insights Imaging*, vol. 9, no. 4, pp. 611–629, 2018.
[372] D. Guidici and M. Clark, “One-Dimensional Convolutional Neural Network Land-Cover Classification of Multi-Seasonal Hyperspectral Imagery in the San Francisco Bay Area, California,” *Remote Sens.*, vol. 9, no. 6, p. 629, 2017.
[373] M. Terzano, D. Mancia, M. Salati, G. Costani, A. Decembrino, and L. Parrino, “The cyclic alternating pattern as a physiologic component of normal NREM sleep,” *Sleep*, vol. 8, no. 2, pp. 137–145, 1985.
[374] U. Scholz, A. Bianchi, S. Cerutti, and S. Kubicki, “Vegetative Background of Sleep: Spectral Analysis of the Heart Rate Variability,” *Physiol. Behav.*, vol. 62, no. 5, pp. 1037–1043, 1997.
[375] J. Bronzino, *The biomedical engineering handbook*, 3º. Connecticut: CRC Press, 2006.
[376] E. Helfenbein, R. Firoozabadi, S. Chien, E. Carlson, and S. Babaeizadeh, “Development of three methods for extracting respiration from the surface ECG: a review,” *J. Electrocardiol.*, vol. 47, no. 6, pp. 819–825, 2014.
[377] S. Arunachalam and L. Brown, “Real-Time Estimation of the ECG-Derived Respiration (EDR) Signal using a New Algorithm for Baseline Wander Noise Removal,” presented at the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009, Minnesota, USA, Sep. 2009.
[378] P. Stoica and R. Moses, *Spectral Analysis of Signals*. New Jersey, USA: Prentice Hall, 2005.
[379] I. Santamaria and J. Via, “Estimation of the Magnitude Squared Coherence Spectrum Based on Reduced-Rank Canonical Coordinates,” presented at the IEEE International Conference on Acoustics, Speech and Signal Processing, 2007, Honolulu, Hawaii, USA, Apr. 2007.
[380] P. Addison, “Wavelet transforms and the ECG: a review,” *Physiol. Meas.*, vol. 26, no. 5, pp. 155–199, 2005.
[381] M. Bogdan, J. Ghosh, and R. Doerge, “Modifying the Schwarz Bayesian information criterion to locate multiple interacting quantitative trait loci,” *Genetics*, vol. 167, no. 2, pp. 989–999, 2004.
[382] L. Faes, A. Porta, and G. Nollo, “Information Decomposition in Bivariate Systems: Theory and Application to Cardiorespiratory Dynamics,” *Entropy*, vol. 17, no. 1, pp. 277–303, 2015.
[383] F. Yasuma and J. Hayano, “Respiratory sinus arrhythmia: why does the heartbeat synchronize with respiratory rhythm?,” *Chest*, vol. 125, no. 2, pp. 683–690, 2004.
[384] M. Hirshkowitz *et al.*, “National Sleep Foundation’s sleep time duration recommendations: methodology and results summary,” *Sleep Health*, vol. 1, no. 1, pp. 40–43, 2015.
[385] P. Lacks and C. Morin, “Recent advances in the assessment and treatment of insomnia,” *J. Consult. Clin. Psychol.*, vol. 60, no. 4, pp. 586–594, 1992.
[386] N. Punjabi, “The Epidemiology of Adult Obstructive Sleep Apnea,” *Proc. Am. Thorac. Soc.*, vol. 5, no. 2, pp. 136–143, 2008.
[387] L. Rosenthal and D. Dolan, “The Epworth sleepiness scale in the identification of obstructive sleep apnea,” *J. Nerv. Ment. Dis.*, vol. 196, no. 5, pp. 429–431, 2008.
[388] M. Bonnet and D. Arand, “EEG Arousal Norms by Age,” *J. Clin. Sleep Med.*, vol. 3, no. 3, pp. 271–274, 2007.
[389] A. Hashmi, I. Khawaja, Z. Butt, M. Umair, S. Naqvi, and J. Haq, “The Pittsburgh Sleep Quality Index: validation of the Urdu translation,” *J. Coll. Physicians Surg. Pak.*, vol. 24, no. 2, pp. 123–126, 2014.
[390] S. Schulz *et al.*, “Cardiovascular and cardiorespiratory coupling analyses: A review,” *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.*, vol. 271, no. 1997, p. 20120191, 213AD.
[391] F. Mendonça, S. Mostafa, F. Morgado-Dias, and A. Ravelo-García, “Matrix of Lags: a tool for Analysis of Multiple Dependent Time Series applied for CAP scoring,” *Comput. Methods Programs Biomed.*, vol. 189, no. 1, p. 105314, 2020.
[392] C. Phillips, J. Parr, and E. Riskin, *Signals, Systems, and Transforms*. London, UK: Pearson, 2013.
[393] K. Aboalayon, M. Faezipour, W. Almuhammadi, and S. Moslehpour, “Sleep Stage Classification Using EEG Signal Analysis: A Comprehensive Survey and New Investigation,” *Entropy*, vol. 18, no. 9, pp. 1–31, 2016.
[394] M. Xiao, H. Yan, J. Song, Y. Yang, and X. Yang, “Sleep stages classification based on heart rate variability and random forest,” *Biomed. Signal Process. Control*, vol. 8, no. 6, pp. 624–633, 2013.
[395] F. Ebrahimi, S. Setarehdan, J. Ayala-Moyeda, and H. Nazeran, “Automatic sleep staging using empirical mode decomposition, discrete wavelet transform, time-domain, and nonlinear dynamics features of heart rate variability signals,” *Comput. Methods Programs Biomed.*, vol. 112, no. 1, pp. 47–57, 2013.

[396] M. Mendez, M. Matteucci, V. Castronovo, L. Ferini-Strambi, S. Cerutti, and A. Bianchi, “Sleep staging from Heart Rate Variability: time-varying spectral features and Hidden Markov Models,” *Int. J. Biomed. Eng. Technol.*, vol. 3, no. 3, pp. 246–263, 2010.
[397] M. Aktaruzzaman, M. Migliorini, M. Tenhunen, S. Himanen, A. Bianchi, and R. Sassi, “The addition of entropy-based regularity parameters improves sleep stage classification based on heart rate variability,” *Med. Biol. Eng. Comput.*, vol. 53, no. 5, pp. 415–425, 2015.
[398] A. Garcia, J. Koschnitzky, T. Dashevskiy, and J. Ramirez, “Cardiorespiratory Coupling in Health and Disease,” *Auton. Neurosci.*, vol. 175, no. 1–2, pp. 26–37, 2013.
[399] R. Shanmugam and R. Chattamvelli, *Statistics for Scientists and Engineers*. New Jersey, USA: John Wiley & Sons, 2016.
[400] A. Ravelo-García, J. Navarro-Mesa, M. Murillo-Díaz, and G. Julia-Serda, “Application of RR Series and Oximetry to a Statistical Classifier for the Detection of Sleep Apnoea/Hipopnoea,” presented at the IEEE Computers in Cardiology Conference 2004 (CinC), Illinois, USA, Sep. 2004.
[401] F. Mendonça, S. Mostafa, F. Morgado-Dias, and Ravelo-García, “An Oximetry Based Wireless Device for Sleep Apnea Detection,” *Sensors*, vol. 20, no. 3, p. 888, 2020.
[402] F. Mendonça, S. Mostafa, F. Morgado-Dias, J. Navarro-Mesa, G. Juliá-Serdá, and A. Ravelo-García, “A minimally invasive portable system for sleep apnea detection,” presented at the 2017 International Conference and Workshop on Bioinspired Intelligence (IWOBI), Funchal, Portugal, Jul. 2017.
[403] F. Mendonça, S. Mostafa, F. Morgado-Dias, J. Navarro-Mesa, G. Juliá-Serdá, and A. Ravelo-García, “A portable wireless device based on oximetry for sleep apnea detection,” *Computing*, vol. 100, no. 11, pp. 1203–1219, 2018.
[404] H. Silva, J. Guerreiro, A. Lourenço, A. Fred, and R. Martins, “BITalino: A Novel Hardware Framework for Physiological Computing,” presented at the International Conference on Physiological Computing Systems (PhyCS), Lisbon, Portugal, Jan. 2014.
[405] A. Batista, H. Silva, A. Fred, C. Moreira, M. Reis, and A. Ferreira, “Benchmarking of the BITalino Biomedical Toolkit Against an Established Gold Standard,” *Healthc. Technol. Lett.*, vol. 6, no. 2, pp. 32–36, 2019.
[406] C. Toner *et al.*, “Fault resilient FPGA design for 28 nm ZYNQ system-on-chip based radiation monitoring system at CERN,” *Microelectron. Reliab.*, vol. 100, no. 1, p. 113492, 2019.
[407] S. Liu, G. Cai, Y. Huang, and Y. Chen, “A wearable ECG apperatus for ubiquitous health care,” presented at the 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, Hungary, Oct. 2016.
[408] F. Mendonça, S. Mostafa, F. Morgado-Dias, and A. Ravelo-García, “A Method based on Cardiopulmonary Coupling Analysis for Sleep Quality Assessment with FPGA Implementation,” 2021.