

**Avaliação de Algoritmos Computacionais no  
contexto de Sistema de Monitorização de  
Apneia baseado no processamento de  
Bio-sinais**

(versão final após defesa)

**Eduardo Frango Paulos**

Dissertação para obtenção do Grau de Mestre em  
**Engenharia Informática**  
(2<sup>o</sup> ciclo de estudos)

Orientador: Prof. Doutor Nuno Gonçalo Coelho Costa Pombo  
Co-orientador: Prof. Doutor Sandeep Pirbhulal

**Covilhã, Julho de 2022**



# **Dedicatória**

”Não se deve ir atrás de objetivos fáceis, é preciso buscar o que só pode ser alcançado por meio dos maiores esforços.”

— Albert Einstein

Dedicado a todos aqueles que, de uma forma ou outra, me apoiaram e continuarão a apoiar. Dedicado ainda, a todos os que lutam por uma sociedade melhor.



# Agradecimentos

Agradeço em primeiro à minha Família pelo apoio incondicional que me deram ao longo da esta jornada que agora acaba. Pai, Mãe, Irmão, Avós, Tios e Primos um obrigado especial!

Em segundo lugar, ao meu orientador Nuno Pombo e co-orientador Sandeep que sem a sua ajuda esta caminhada não seria possível. A vocês, obrigado!

Em seguida, um obrigado a uma pessoal especial que tornou esta caminhada mais fácil e gratificante. Um grande obrigado Patrícia Fernandes!

Por último e não menos importante, um obrigado a todos os meus amigos que a vida académica me foi dando.

Obrigado a todos!



# Resumo

A doença da apneia do sono afeta cerca de 1 milhão de pessoas no mundo. Ela distingue-se em três tipos: Obstrutiva, Central e Mista. A investigação por uma alternativa à Polissonografia como meio de diagnóstico à Apneia Obstrutiva do Sono é uma necessidade, visto os custos e meios que ela envolve. Nesse sentido a presente dissertação descreve a envolvimento do problema e propõe modelo de detecção de episódios de apneia, com recurso ao aprendizado de máquina (em inglês, *Machine Learning*).

O modelo foi construído com o uso de diferentes classificadores (SVM, MLP, Adaboost e RandomForest). Como dados de entrada, foram processados e extraídas características do base de dados *ApneaECG* da *Physionet*. O melhor resultado alcançado foi uma exatidão (*accuracy*) de 78,45%, sensibilidade (*sensitivity*) de 70,26% e especificidade *specificity* de 83,49%.

Adicionalmente, foi construída uma aplicação de recolha de dados que utiliza um dispositivo capacitado com elétrodos que permitem a recolha de sinal de eletrocardiograma. Com estes dados, pretende-se no futuro criar uma base de dados e um sistema que ajude a detecção da apneia.

# Palavras-chave

Eletrocardiograma (ECG); Apneia Obstrutiva do Sono (AOS); Diagnóstico; Detecção; Machine Learning; Heart Rate Variability; ECG-Derived Respiration (EDR); Classificadores; Support Machine Vetor (SVM); Multi-layer Perceptron (MLP); AdaBoost; RandomForest; Recolha de dados; Aplicação Móvel.



# Abstract

Sleep apnea disease affects about 1 million people worldwide. It is distinguished into three types: Obstructive, Central and Mixed. The investigation for an alternative to Polysomnography as a means of diagnosing Obstructive Sleep Apnea is a necessity, given the costs and means that it involves. In this sense, the present dissertation describes the surroundings of the problem and proposes a model for the detection of apnea episodes, using machine learning (in English, *Machine Learning*).

The model was built using different classifiers (SVM, MLP, Adaboost and RandomForest). As input data, features were processed and extracted from the *ApneaECG* database of *Physionet*. The best result achieved was an accuracy (*accuracy*) of 78.45%, sensitivity (*sensitivity*) of 70.26% and specificity (*specificity*) of 83.49%.

Additionally, a data collection application was built that uses a device equipped with electrodes that allow the collection of an electrocardiogram signal. With these data, it is intended in the future to create a database and a system to help detect apnea.

# Keywords

Electrocardiogram (ECG); Obstructive Sleep Apnea (OSA); Diagnosis; Detection; Machine Learning; Heart Rate Variability; ECG-Derived Respiration (EDR); Classifiers; Support Machine Vector (SVM); Multi-layer Perceptron (MLP); AdaBoost; RandomForest; Data collection; Mobile Application.



# Índice

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Enquadramento e Motivação . . . . .	1
1.2	Organização do Documento . . . . .	2
<b>2</b>	<b>Enquadramento Teórico</b>	<b>5</b>
2.1	Visão Geral . . . . .	5
2.2	Epidemiologia em Portugal . . . . .	5
2.3	Diagnóstico . . . . .	5
2.4	Classificação . . . . .	6
2.5	Problemas associados . . . . .	7
2.6	Eletrocardiograma (ECG) e Apneia . . . . .	7
2.6.1	ECG . . . . .	7
2.6.2	ECG como ferramenta de diagnóstico . . . . .	8
2.6.3	Nota adicional . . . . .	9
2.7	<i>Machine Learning</i> (ML) e Apneia . . . . .	9
2.8	Conclusão . . . . .	10
<b>3</b>	<b>Estado da Arte</b>	<b>11</b>
3.1	Métodos . . . . .	11
3.1.1	Questões de pesquisa . . . . .	11
3.1.2	Critérios de inclusão . . . . .	11
3.1.3	Estratégia de pesquisa . . . . .	12
3.1.4	Extração das características do estudo . . . . .	12
3.2	Resultados . . . . .	12
3.2.1	Elegibilidade dos estudos . . . . .	13
3.2.2	Informação recolhida . . . . .	14
3.2.3	Estatísticas da informação retiradas . . . . .	17
3.3	Discussão . . . . .	18
3.3.1	RQ1 - Porque são necessários modelos de deteção para a Apneia Obstrutiva do Sono (AOS)? . . . . .	18
3.3.2	RQ2 - Como o ECG é aplicado para deteção de AOS? . . . . .	18
3.3.3	RQ3 - Quais são as Base de Dados para deteção de AOS? . . . . .	19
3.3.4	RQ4 - Quais são os métodos/técnicas/modelos para deteção de AOS baseada em ECG em registo ou em tempo real? . . . . .	20
3.3.5	RQ5 - Quais são os parâmetros usados para avaliar o desempenho da deteção de AOS? . . . . .	22
3.4	Conclusão . . . . .	23

<b>4</b>	<b>Ferramentas e Tecnologias Usadas</b>	<b>25</b>
4.1	Ferramentas . . . . .	25
4.1.1	Jupyter . . . . .	25
4.2	Tecnologias . . . . .	26
4.2.1	<i>Flutter</i> . . . . .	26
4.2.2	Bibliotecas e pacotes <i>Flutter</i> . . . . .	26
4.2.3	<i>Firebase</i> . . . . .	27
4.2.4	<i>Python</i> . . . . .	27
4.2.5	Bibliotecas <i>Python</i> . . . . .	27
4.3	Conclusão . . . . .	28
<b>5</b>	<b>Dispositivo e Aplicação Móvel</b>	<b>29</b>
5.1	Dispositivo . . . . .	29
5.1.1	CrITÉRIOS de seleção . . . . .	29
5.1.2	<i>Movesense Medical Developer Kit (MMDK)</i> . . . . .	29
5.1.3	Discussão . . . . .	30
5.2	Aplicação Móvel - <i>ApneaECG</i> . . . . .	30
5.2.1	Descrição da aplicação . . . . .	30
5.2.2	Engenharia de software . . . . .	32
5.3	Protocolo de Recolha de Dados . . . . .	33
5.4	Conclusão . . . . .	33
<b>6</b>	<b>Implementação e Resultados</b>	<b>35</b>
6.1	Implementação . . . . .	35
6.1.1	Pré-processamento . . . . .	35
6.1.2	Extração e Seleção de características . . . . .	36
6.1.3	Classificadores . . . . .	39
6.2	Resultados . . . . .	40
6.2.1	Normalização . . . . .	40
6.2.2	Resultados por número de características . . . . .	41
6.3	Discussão . . . . .	41
6.4	Conclusão . . . . .	44
<b>7</b>	<b>Conclusão e Trabalho Futuro</b>	<b>45</b>
7.1	Conclusões Principais . . . . .	45
7.2	Trabalho Futuro . . . . .	45
	<b>Bibliografia</b>	<b>47</b>
<b>A</b>	<b>Protocolo de recolha de Dados ECG</b>	<b>55</b>

# Lista de Figuras

2.1	Utente em pleno exame de Polissonografia. . . . .	6
2.2	Sinal ECG, as diferentes ondas e seus segmentos de onda. Fonte: [1] . . . .	7
2.3	Segmento ECG com episódios de apneia. . . . .	8
2.4	segmento ECG normal. . . . .	8
3.1	Gráfico da distribuição por data dos estudos. . . . .	17
3.2	Diagrama de bloco genérico do método utilizado para deteção/predição da AOS . . . . .	20
4.1	Ferramentas e Tecnologias utilizadas em contexto de desenvolvimento. . .	25
5.1	<i>Movesense Medical Developer Kit</i> . Fonte:[2] . . . . .	30
5.2	Esquema geral da aplicação ApneaECG. Adaptado de [3] . . . . .	31
5.3	<i>Layouts</i> da aplicação ApneaECG. . . . .	31
5.4	Diagrama de Caso de Uso - <i>Efetuar recolha de dados ECG</i> . . . . .	33
6.1	<i>Workflow</i> de desenvolvimento do Algoritmo. . . . .	35
6.2	<i>Heart Rate Variability</i> . Fonte: [4] . . . . .	37
6.3	<i>Features Domain</i> . . . . .	37
6.4	Estrutura do um <i>hidden layer Multi-layer Perceptron Classifier</i> (MLPC). .	39
6.5	Estrutura do um <i>hidden layer</i> MLPC. . . . .	41
6.6	Distribuição dos resultados ao longo da seleção de caraterísticas. . . . .	43
6.7	Média dos valores de <i>ACC</i> , <i>SENS</i> e <i>SPEC</i> para os diferentes classificadores. .	43
7.1	Arquitetura Modelo Futuro. . . . .	46



# Lista de Tabelas

2.1	<i>Apnea Hypopnea Index</i> . . . . .	6
3.1	Informação extraída 1 - Legenda: NR: <i>Number of Records</i> ; NF: <i>Number of Features</i> . . . . .	14
3.2	Informação extraída (continuação) - Legenda: Acc: <i>Accuracy</i> , Sens: <i>specificity</i> , Spec: <i>specificity</i> . . . . .	15
3.3	Pré-processamento, Extração de características e Seleção de características. .	16
5.1	Requisitos Funcionais. . . . .	32
5.2	Requisitos não Funcionais. . . . .	32
5.3	Requisitos de Domínio. . . . .	32
6.1	Caraterísticas/ <i>Feactures</i> extraídas do <i>Heart Rate Variability</i> (HRV). . .	38
6.2	Caraterísticas/ <i>Feactures</i> extraídas do <i>ECG-derived respiration</i> (EDR). . .	38
6.3	Resultados obtidos não normalizados vs normalizados. (N): Normalizados.	40
6.4	Resultados obtidos. . . . .	42



# Lista de Acrónimos

<b>AHI</b>	<i>Apnea Hypopnea Index</i>
<b>AI</b>	<i>Artificial Intelligence</i>
<b>AOS</b>	Apneia Obstrutiva do Sono
<b>API</b>	Application Programming Interface
<b>BD</b>	Base de Dados
<b>DL</b>	<i>Deep Learning</i>
<b>DGS</b>	Direção-Geral de Saúde
<b>ECG</b>	Eletrocardiograma
<b>EDR</b>	<i>ECG-derived respiration</i>
<b>EEG</b>	Eletroencefalograma
<b>FIR</b>	<i>Finite Impulse Response</i>
<b>HR</b>	<i>Heart Rate</i>
<b>HRV</b>	<i>Heart Rate Variability</i>
<b>IA</b>	Inteligência Artificial
<b>IHR</b>	<i>Instantaneous Heart Rate</i>
<b>IO</b>	<i>Input-Output</i>
<b>IoMT</b>	<i>Internet of Medical Things</i>
<b>IoT</b>	<i>Internet of Things</i>
<b>ML</b>	<i>Machine Learning</i>
<b>MLPC</b>	<i>Multi-layer Perceptron Classifier</i>
<b>MMDK</b>	<i>Movesense Medical Developer Kit</i>
<b>NN</b>	<i>Neural network</i>
<b>POO</b>	Programação Orientada a Objetos
<b>PRISMA</b>	<i>Preferred Reporting Items for Systematic Reviews and Meta-Analyses</i>
<b>PSG</b>	Polissonografia
<b>RBF</b>	<i>Radial Basis Function</i>
<b>RRI</b>	<i>RR interval</i>
<b>SAS</b>	Síndrome da Apneia do Sono
<b>SDK</b>	<i>Software Development Kit</i>
<b>SLR</b>	<i>Systematic Literature Review</i>
<b>SVM</b>	<i>Support-vector machine</i>
<b>ULS</b>	Unidade de Saúde Local



# Capítulo 1

## Introdução

Neste capítulo será feito um enquadramento do problema, assim como a motivação que leva à busca de uma solução para o problema. Por fim, será apresentada a organização do documento.

### 1.1 Enquadramento e Motivação

A criação de sistemas/modelos que facilitem o diagnóstico das mais diversas doenças, é hoje um “*must*” para os sistemas de saúde na sociedade. Se fizermos uma retrospectiva no sentido de percepção da evolução desses modelos, podemos constatar que existe um esforço da comunidade científica em torná-los de mais fácil utilização e cada vez mais eficientes. O desenvolvimento científico nos últimos anos, trouxe avanços significativos que hoje, sem qualquer dúvida, ajudam os serviços de saúde nas mais diversas áreas e permitem a sobrevivência de muitos pacientes. Remetendo para atualidade pandémica onde são utilizados dispositivos de ventilação assistida, este é um claro exemplo onde convergem várias áreas científicas, inclusive a informática, no desenvolvimento destes dispositivos. Refletindo neste exemplo, é possível constatar que a união entre os vários campos científicos é essencial, em última instância, para o desenvolvimento da sociedade, dos sistemas de saúde e da medicina.

É neste sentido que a exploração de novos métodos de tratamento, diagnóstico e até prevenção de doenças são essenciais. É também, através desta evolução, que doenças que anteriormente eram consideradas severas, hoje deixaram de o ser, dando a possibilidade aos pacientes de terem uma vida menos condicionada.

A popularização de dispositivos de *fitness* e *wearables*, veio disponibilizar ao utilizador comum o acesso a tecnologias de monitorização, como, por exemplo, sensores de batimento cardíaco, oxigenação sanguínea (SpO<sub>2</sub>) e sensores de ECG. Estes dispositivos podem ser considerados parte integrante na *Internet of Medical Things* (IoMT).

IoMT trata-se de um tópico emergente na Informática e uma vertente da conhecida *Internet of Things* (IoT), onde à semelhança desta última, são utilizados dispositivos como os citados anteriormente, que coletam e enviam os dados através da rede Internet para aplicações de domínio específico [5]. Através desses dispositivos, pretende-se chegar às necessidades individuais dos pacientes [6] e a uma monitorização mais personalizada das patologias, na medida que, em casos de emergência a reação médica seja antecipada [7]. De um modo geral, o IoMT pretende e permitirá a aproximação dos sistemas de saúde aos pacientes.

Em adição, os sistemas de IoMT vieram oferecer aos utilizadores a possibilidade de serem eles próprios a monitorizar a sua doença e, de antemão, aperceber-se de algo errado no

seu sistema fisiológico. Este propósito não se aplica estritamente a pessoas já diagnosticadas, mas também a pessoas saudáveis que apenas pretendem monitorizar o seu estado de saúde. Assim, de forma quase despropositada, estes dispositivos podem ajudar ao diagnóstico de doenças.

É evidente que, aliado a estes dispositivos existem *softwares* ou conjuntos destes que permitam o seu devido funcionamento, seja ele de monitorização ou diagnóstico. À medida que empresas de tecnologia disponibilizam estes dispositivos ao consumidor, em paralelo são desenvolvidos e aprimorados os sistemas que integram esses dispositivos. Completando, com o aumento da necessidade destes dispositivos é expectável que a necessidade de aplicações e/ou programas de cuidados de saúde aumente significativamente nos próximos anos [8].

A *Machine Learning* (ML) pode desempenhar um papel vital na busca de métodos e/ou sistemas que ajudem os profissionais de saúde. ML, parte da Inteligência Artificial (IA), caracteriza-se pelo desenvolvimento de algoritmos, que de acordo com dados empíricos, fazem evoluir os comportamentos computacionais dos computadores [9]. A evolução dos algoritmos de ML tem vindo a melhorar métodos de captura de dados, redes de computadores, levou ao aparecimento de novos sensores *Input-Output* (IO) e ao auto-ajuste de comportamento dependendo do utilizador [10].

Todos estes campos da tecnologia, a IoMT, o *software*, a ML e o contínuo desenvolvimento nestes campos, em união, podem melhorar a qualidade dos serviços de saúde, dando a possibilidade de obter diagnósticos mais precisos e até de fazer prognósticos mais assíduos de doenças em estágios iniciais.

Com este estudo pretende-se o desenvolvimento de um sistema de deteção focado na doença da Apneia Obstrutiva do Sono (AOS). Através da recolha de sinais vitais como o ECG e do processamento das suas características, pretende-se chegar a um modelo que consiga identificar episódios e até, se possível, de os prever. Por fim, será feito um esforço para que este modelo seja implementado num dispositivo real, que de alguma forma consiga processar em tempo real os dados recolhidos dos sensores e assim ajudar no diagnóstico da AOS.

## 1.2 Organização do Documento

Este documento encontra-se dividido em capítulos da seguinte forma:

- Primeiro capítulo – **Introdução** – Contém a motivação para o desenvolvimento do tema proposto, como também um enquadramento inicial ao problema;
- Segundo capítulo – **Enquadramento Teórico** – são abrangidos, no capítulo, conceitos que envolvem a doença da apneia e as soluções possíveis para o problema do seu diagnóstico;
- Terceiro capítulo – **Estado da Arte** – Neste capítulo será apresentada uma revisão sistemática descrevendo os diferentes desenvolvimentos na área.

- Quarto capítulo – **Dispositivo e Aplicação** – Serão apresentados vários dispositivos de recolha de bio-sinais ECG candidatos à sua utilização para o método proposto.
- Quinto capítulo – **Implementação e Resultados** – São descritas, neste capítulo, as decisões tomadas para o desenvolvimento do algoritmo, assim como os resultados obtidos a partir do mesmo;
- Sexto capítulo – **Conclusão e Trabalho Futuro** – apresenta uma reflexão do trabalho desenvolvido e do trabalho futuro a ser desenvolvido.

**Nota:** Ao longo do documento será, por várias vezes, utilizado o termo apneia. Sempre que é mencionado o termo, este refere-se exclusivamente à AOS.



# Capítulo 2

## Enquadramento Teórico

Neste capítulo, encontra-se um enquadramento à Síndrome da Apneia do Sono (SAS), onde serão distinguidas as diferentes variantes da doença, a sua epidemiologia em Portugal, o método de diagnóstico, o método de classificação e os problemas associados. Será feita uma breve descrição do sinal ECG, algumas das suas características e como ele pode ajudar ao diagnóstico. Por fim, também uma breve descrição do que é ML e o seu enquadramento na apneia.

### 2.1 Visão Geral

A SAS é a doença mais comum ligada ao sono que afeta quase 1 bilhão de pessoas no mundo [11]. A sua prevalência é maior no sexo masculino com cerca de 34% e 17% para o sexo feminino [12]. Esta caracteriza-se pela ocorrência de episódios de interrupção do fluxo aéreo durante o período de sono, sejam parciais (Hipoapneia) ou totais (Apneia). Considera-se um episódio de apneia quando a interrupção é de cerca de 10 segundos. Dependendo da etiologia da doença, distinguem-se 3 tipos de SAS: a obstrutiva, caracteriza-se pela obstrução parcial ou total do fluxo aéreo; a central, que corresponde a quando de falha de comunicação entre o cérebro e o sistema respiratório [13]; por fim a mista, que combina os sintomas tanto da obstrutiva como da central [14]. Este estudo, será focado essencialmente na AOS.

### 2.2 Epidemiologia em Portugal

Segundo a Direção-Geral de Saúde (DGS), pelo estudo efetuado em 2014, estima-se que a prevalência da doença seja de 2,2% na população portuguesa, ainda assim um valor menor comparado com países como os Estados Unidos da América [15]. Assim como outros estudos conduzidos em outros países, o número de pacientes masculinos é superior ao feminino. Em Portugal, pelo mesmo estudo, foi encontrado um razão entre o sexo masculino e feminino de 3,8:1. Além destes dados, é de salientar que da amostra do estudo cerca de 48% eram casos graves. Levanta-se assim a hipótese de subdiagnóstico, tendo em conta que existe uma maior probabilidade de que casos com diagnóstico sejam AOS mais agravados.

### 2.3 Diagnóstico

Como exame para o diagnóstico é comumente usada a Polissonografia (PSG). Trata-se de um exame que avalia o sono de um paciente através de vários parâmetros como moni-

torização da atividade cerebral, atividade muscular e o movimento dos olhos [16]. Desta forma, o exame requer vários sensores que são ligados ao paciente durante todo o seu período de sono, tornando-se incomodativo para o paciente. Adicionalmente, é necessário pessoal dedicado que monitorize continuamente o evoluir do exame identificando episódios de apneia [16]. Por consequência, este exame apenas pode ser realizado em clínicas especializadas, hospitais ou clínicas de sono [13]. Contudo, por vezes, de modo a reduzir os custos, certas clínicas realizam um teste de monitorização não tão detalhado, o que pode dificultar o diagnóstico. Todos esses fatores tornam a PSG e, consequentemente, o diagnóstico da Apneia caro. Neste sentido, é necessário desenvolver métodos igualmente eficazes, mas mais simples, que ajudem na redução deste custo.

Como alternativa à PSG e de modo a automatizar a deteção e monitorização da apneia, têm surgido várias propostas que englobam a monitorização do oxigénio no sangue (SpO<sub>2</sub>) [17, 18], ECG, Eletroencefalograma (EEG) [19, 20], entre outros tipos de *inputs*.

O ECG é uma das alternativas mais utilizadas pelas suas características morfológicas [21], tendo como vantagem não ser invasivo em comparação com os restantes. Um *milestone* para a sua utilização para a deteção da AOS foi o *Challenge* em 2000 [22], no qual foi disponibilizada a Base de Dados *Apnea-ECG*. Um dos pioneiros da sua utilização foi *Raymond et al.* [23], extraindo o EDR e o intervalo entre picos *R* do sinal ECG.



Fig. 2.1: Utente em pleno exame de Polissonografia.

## 2.4 Classificação

Através do *Apnea Hypopnea Index* (AHI), são distinguidos vários grupos da doença. Nesta qualificação é considerado o número de episódios de apneia por hora. Na tabela 2.1 são apresentados os diferentes grupos [24]:

Episódios/hora	Grupo
<5	Mínima ou inexistente
[5,15[	Ligeiro
[15,30[	Moderado
>30	Grave

Table 2.1: *Apnea Hypopnea Index*

## 2.5 Problemas associados

Como se trata de uma doença ligada ao sono, muitas vezes é difícil para o paciente aperceber-se de que sofre da mesma. Consequentemente, muitos pacientes não são diagnosticados com a doença, o que por sua vez leva a que estes não sejam tratados. Como resultado, a doença considera-se subdiagnosticada, ou seja, o número de casos expectáveis é maior que o número de casos diagnosticados. A curto prazo, um “não diagnóstico” pode levar a complicações como tensão arterial alta, batimento cardíaco acelerado, hipoxemia, e a longo termo problemas vasculares cerebrais e cardíacos, comprometimento do sistema imunitário, fadiga extrema, fraca concentração, tempos de reação baixos e em casos extremos até à própria morte [25].

## 2.6 ECG e Apneia

### 2.6.1 ECG

O Eletrocardiograma (ECG) é um dos bio-sinais mais usados na medicina, caracteriza-se pela assinatura elétrica produzida pelo coração [26]. Através deste sinal é possível observar o comportamento do coração e assim fazer o diagnóstico de doenças associadas ao mesmo. É subdividido em diferentes ondas e diferentes segmentos das mesmas. Estas são conhecidas por *P*, *Q*, *R*, *S*, *J*, *T* e *U*. A duração da onda *P*, o complexo *QRS* e o intervalo entre a onda *S* e *T* são alguns dos exemplos de características que podem ser observadas através do ECG. A representação das ondas e características citadas e outras, segue-se na Figura 2.2:

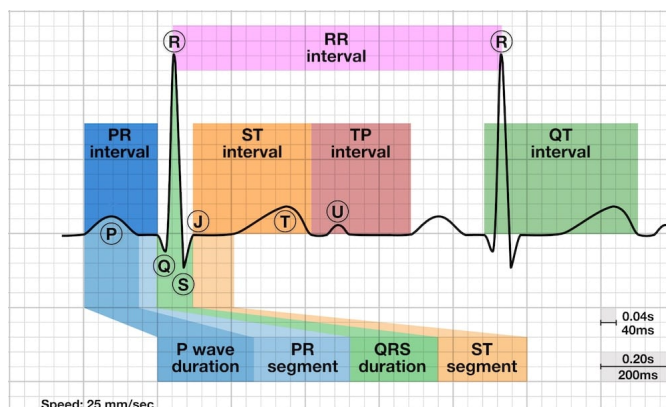


Fig. 2.2: Sinal ECG, as diferentes ondas e seus segmentos de onda. Fonte: [1]

Variações no comportamento das ondas, segmentos e intervalos pode indicar doenças como arritmias, ataques cardíacos e entre outras doenças cardíacas [27]. No entanto, a visualização de alterações nestes segmentos é um processo difícil até para os profissionais de saúde, visto a sua complexidade. Num estudo conclui-se que apenas um 1/4 de 800 médicos conseguiu classificar o intervalo *QT* como 'longo' ou 'normal' [28].

## 2.6.2 ECG como ferramenta de diagnóstico

Além das características mencionadas anteriormente, existem outras características que se podem retirar a partir do ECG. O EDR é uma das mais importantes no estudo da apneia. Representa o comportamento respiratório do paciente através do ECG. Em 1986, George B. Moody et al. realizaram um estudo de análise de técnicas aplicadas ao EDR [29]. Chegaram à conclusão que, através da utilização do EDR era possível destacar episódios de AOS, constituindo uma alternativa viável à PSG.

Adicionalmente, características como a amplitude e a frequência do ECG também são alterados em episódios de apneia [30].

Através da análise da Figura 2.4 podemos confirmar a afirmação anterior. O segmento com apneia contém um sinal que atinge maiores amplitudes e na generalidade um batimento cardíaco menos constante. Os segmentos de sinal forma extraídos de Base de Dados (BD) *Apnea-ECG* [31].

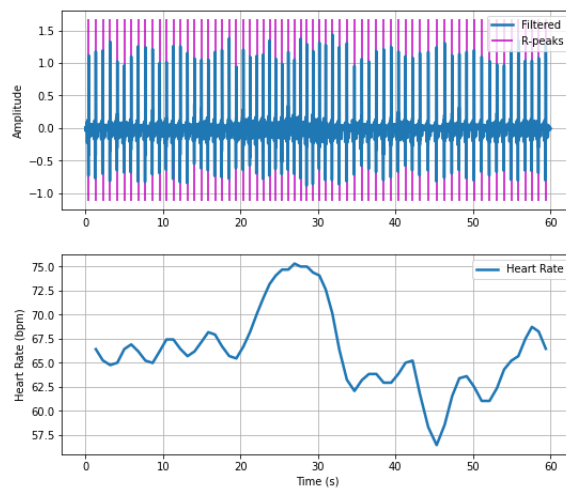


Fig. 2.3: Segmento ECG com episódios de apneia.

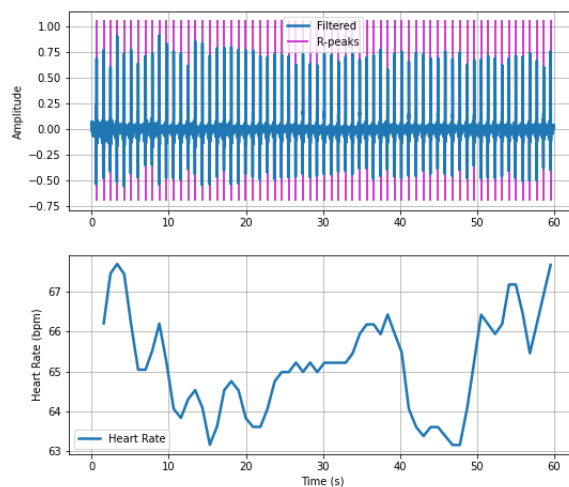


Fig. 2.4: segmento ECG normal.

Seguindo o problema da complexidade na interpretação de ECG, a necessidade de alterna-

tiva ao PSG e a certeza de que é possível observar episódios de apneia através do mesmo, torna-se ainda mais evidente a necessidade de criar modelos de classificação para a doença da apneia.

### 2.6.3 Nota adicional

O exame comum de ECG é composto por diferentes derivações de sinal (em inglês, *leads*), que são recolhidas pelo posicionamento de vários eléctrodos em zonas específicas do corpo. Neste estudo, apenas uma derivações será utilizada, contornando a complexidade da recolha das demais. Esta decisão deve-se tanto ao facto mencionado anteriormente, como também ao facto de que a base de dados *Apnea-ECG* [31] proposta pelo *challenge* [22] apenas conter uma *lead*, neste caso a *lead I*.

## 2.7 *Machine Learning* (ML) e Apneia

*Machine Learning* (ML) é uma das áreas mais exploradas no mundo da informática, atualmente. A introdução do termo remonta à data de 1959 por Arthur Samuel no seu artigo de “Some Studies in Machine Learning Using the Game of Checkers” [32], porém as bases do desenvolvimento nesta área surgiram anos antes por Walter Pitts e Warren McCulloch que desenvolveram o primeiro modelo de uma rede neuronal [33].

Apesar do conceito ter surgido há vários anos, apenas agora conseguimos começar a tirar o proveito real do seu potencial. Isto pode-se dever tanto ao poder computacional que anteriormente não existia, como também ao nível de complexidade da área.

A ML faz parte de uma esfera maior, a Inteligência Artificial. Por norma, *Artificial Intelligence* (AI) toma uma das duas abordagens, tal como nós humanos [34]:

1. Pelo raciocínio e lógica
2. Pela experiência

A ML apenas se foca no segundo ponto, por outras palavras um computador toma uma decisão pela experiência adquirida. A partir desta ideia é possível “ensinar” um computador a desempenhar uma determinada tarefa, tal como nos é ensinado a ler e a escrever. Por outras palavras, são recolhidas experiências anteriores, no caso dos computadores, dados que o ajudam a descrever e a perceber como realizar a tarefa. Através desta aprendizagem, não apenas é possível aprender, como também capacita o computador a reagir de forma correta mesmo com novos *inputs*.

A partir da observação de características específicas resultantes desses *inputs*, são formados os modelos de ML. Para exemplificar e seguindo o contexto da Apneia Obstrutiva do sono, existem certas características presente no ECG, como as mencionadas na Secção anterior 2.6, que indicam a presença de episódios de apneia. Cabe ao modelos de ML através da observação das características determinar se existe um episódio de apneia ou não. Aprendendo a observar as características de um determinado segmento ECG *A*, é expectável que o mesmo modelo ML consiga classificar um segmento *B*. Passamos assim a

ter um algoritmo que aprendeu com a experiência/observação. Porém um dos desafios no desenvolvimento destes algoritmos reside em encontrar a melhor forma de escolher e expor essas características ao computador.

Assim, ao longo dos últimos anos, os desenvolvedores têm vindo a desenvolver modelos cada vez mais capazes de deteção de eventos de apneia, tentando excluir a necessidade de exame de PSG e facilitando o processo de diagnóstico. Fazendo uma analogia com o que foi dito anteriormente, os modelos "aprendem" através das características dos dados ECG (experiência). Após este treino, estes serão capazes de identificar episódios de apneia através de novos dados ECG. Uma abordagem mais detalhada sobre o desenvolvimento na construção de modelos de deteção da AOS encontra-se no Capítulo 3 – Estado da Arte.

## **2.8 Conclusão**

A complexidade do exame de Polissonografia torna o diagnóstico da AOS pouco confortável e dispendioso. Adicionalmente, a análise dos sinais recolhidos por este exame não é deveras fácil, até mesmo para médicos. Tendo em conta estes fatores e, principalmente, o erro humano, é fundamental a busca por alternativas. Nesta esfera, a computação moderna, mais concretamente, a ML, poderá dar grandes contributos e, assim, ajudar a melhorar a qualidade de vida de muitas pessoas.

# Capítulo 3

## Estado da Arte

Para o estado da arte foi realizada uma revisão sistemática da literatura, em inglês *Systematic Literature Review* (SLR), de modo a otimizar a busca e obter mais conhecimento na área da predição da AOS através de biosinais ECG.

Neste capítulo serão apresentados os métodos utilizados por outros autores. Será, também exposta a forma como foi recolhida e filtrada a literatura, os seus resultados obtidos da extração da informação da literatura, uma discussão com os pontos mais importantes e uma conclusão.

### 3.1 Métodos

A SLR teve como auxílio e base o método de revisão sistemática de declaração de *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) [35].

#### 3.1.1 Questões de pesquisa

As questões de pesquisa, em inglês *Research Questions*, foram originalmente pensadas em inglês, daí a sua representação:

- (RQ1) - Porque são necessários modelos de deteção para a Apneia Obstrutiva do Sono (AOS)?
- (RQ2) - Como o ECG é aplicado para deteção de AOS?
- (RQ3) - Quais são as Base de Dados para deteção de AOS?
- (RQ4) - Quais são os métodos/técnicas/modelos para deteção de AOS baseada em ECG em registo ou em tempo real?
- (RQ5) - Quais são os parâmetros usados para avaliar o desempenho da deteção de AOS?

#### 3.1.2 Critérios de inclusão

Como método de inclusão, foram incluídos os artigos que preenchessem os critérios seguintes: (1) Não ter um público-alvo geograficamente localizado ou de alguma forma específico; (2) Basear-se em sistemas de computação; (3) O artigo terá que ser escrito na língua inglesa; (4) Os participantes do estudo em causa, não podem estar sob efeito de medicamentos e/ou drogas durante a recolhas dos dados do ECG; (5) Os participantes podem pertencer a uma faixa etária, tendo obrigatoriamente diagnóstico para apneia obstrutiva do sono;

(6) Deverá estar explícito ou o valor precisão ou sensibilidade ou especificidade ao qual chegaram; (7) Deverá ser do tipo *Journal* ou *Conference*.

Para a ajuda na seleção dos artigos perante os critérios apresentados, foi utilizada a ferramenta *Web* colaborativa de ajuda à SLR, *Rayyan*[36].

### 3.1.3 Estratégia de pesquisa

Para a extração da literatura relacionada a métodos de predição ou/e deteção da apneia do sono utilizando ECG, foram utilizadas as seguintes base de dados: *ScienceDirect*<sup>1</sup>, *Springer*<sup>2</sup>, *IEEE Xplore*<sup>3</sup>, *Scopus*<sup>4</sup>, *WebOfScience*<sup>5</sup> e *ACM Digital Library*<sup>6</sup>. Adicionalmente, apenas os estudos dentro uma janela temporal entre 1 de Janeiro de 2011 até 27 de Junho de 2020 e que cumprissem o critérios de inclusão (3.1.2), foram incluídos para este estudo. A última *query* a estas bases de dados foi feita a 10 de Novembro de 2020.

A palavras-chave utilizadas nesta pesquisa foram:

("ECG" OR "electrocardio" or "electrocardiogram") AND ("OSA" OR "obstructive sleep apnea" OR "apnea" OR "apnoea") AND ("prediction" OR "predictive" OR "foresee" or "detect" or "diagnosis" OR "diagnostic")

### 3.1.4 Extração das características do estudo

Na extração das características do estudo foram definidas categorias, para extrair dados importantes para análise dos modelos propostos. Essa categorias seguem-se:

- Informações do estudo: ano de publicação;
- *Input*: O tipo de sinal *ECG* utilizado no estudo, assim o seu *dataset* e número de registos utilizados;
- Métodos: o método de pré-processamento do sinal *ECG*, método de *features extraction* e *features selection* e o respetivo classificador para a AOS;
- Performance: qual o método utilizado para avaliar o *output* que valores de performance os autores;
- *Output*: o tipo de output que os utilizadores chegaram.

## 3.2 Resultados

De todas as bases de dados, a que resultou em um maior número de publicações pelas *queries* foi a *ScienceDirect* com 1096, seguindo-se a *Springer* com 607, a *ACM Digital*

<sup>1</sup>ScienceDirect - Science, health and medical journals, full-text articles and books, [Online], Available: <https://www.sciencedirect.com/>

<sup>2</sup>Springer - International Publisher Science Technology Medicine, [Online], Available: <https://www.springer.com/>

<sup>3</sup>IEEE Xplore Digital Library, [Online], Available: <https://ieeexplore.ieee.org/Xplore>

<sup>4</sup>Scopus, [Online], Available: <https://www.scopus.com/>

<sup>5</sup>WebOfScience - Web of Science Core Collection, [Online], Available: [www.webofknowledge.com](http://www.webofknowledge.com)

<sup>6</sup>ACM Digital Library, [Online], Available: <https://dl.acm.org/>

*Library* com 111. Finalmente, 39, 20 e 11 com a *Scopus*, *WebOfScience* e *IEEE Xplore*, respetivamente. A soma destes resultados resultou em 1884 publicações, que após retirar os duplicados converteram-se em 1837 registos únicos.

Após a análise do título e resumo, e seguindo os critérios de inclusão descritos anteriormente, foram incluídos para elegibilidade 41 publicações de texto completo. Por revisão do texto completo destas publicações foram excluídos 20 por não cumprirem os critérios inicialmente propostos. Foram incluídos, por fim, 23 publicações, duas delas por referência a uma das publicações originais 21.

Como razões de exclusão de 20 dos 41 para elegibilidade, encontram-se 4 excluídos por não utilizarem ECG para a predição/deteção da AOS, 9 por não fazer predição/deteção da AOS, 2 por se revelarem *background articles*, 1 por apresentar pouca informação ou informação incompleta, 1 por ter data de publicação anterior a 2011, 1 por ter um público específico, 1 por não apresentar métricas de acordo com o critérios e 1 por não especificar o classificador utilizado.

O diagrama, baseado no PRISMA, que acompanha a informação descrita nesta secção encontra-se esquematizada na figura 2.1.

### 3.2.1 Elegibilidade dos estudos

Perante os critérios de inclusão, existem alguns artigos que foram incluídos que não os cumprem na sua totalidade. Segue-se uma clarificação do porquê da sua inclusão.

No estudo [25] é apresentado um algoritmo de extracção de regras personalizadas *IF-THEN* por paciente. O algoritmo e posteriormente as regras extraídas são utilizadas para construir um perfil de paciente. Este perfil tem o intuito de tentar descobrir as características morfológicas do ECG que fazem antever a ocorrência de um episódio de apneia. O estudo em causa, não segue o normal fluxo normal de outros estudos, onde são extraídas e seleccionadas *features*. No entanto, revela-se ser uma abordagem promissora para a previsão de eventos de apneia, mesmo que esta seja feita caso a caso. A sua inclusão deve-se a estes factos, assim com os estudos [37] e [38] que foram adicionados por referência do estudo [25] por se revelarem precedentes deste.

O foco do estudo [21] é determinar o grau de AOS para ajustar adequadamente o tratamento mais eficiente e específico da apneia, o *continuous positive airway pressure* (CPAP). Por consequente, o propósito não é a predição ou deteção de apneia. Contudo, o estudo segue o normal fluxo de preparação de dados, extracção e seleção de *features* e a utilização de um classificador. Perante esta característica o estudo foi incluído para revisão.

No estudo [39] além da utilização de sinais ECG, também são utilizados sinais de eletromiografia e de eletroencefalografia. Ainda que o foco não esteja a utilização do ECG e seja uma combinação dos três sinais, estes são testados individualmente, em combinação de dois-a-dois e só por fim os três em conjunto. A sua inclusão deve-se ao teste individualizado de cada sinal.

A mesma situação acontece no estudo [40], onde além do sinal ECG é também utilizado o um sensor de aceleração de 3 eixos.

### 3.2.2 Informação recolhida

Nesta secção estão expressas as informações que foram sendo retiradas ao longo da análise dos estudos. Na Tabela 3.1 encontra-se os dados acerca da ano do estudo, o número de registos de ECG, o número de *features* utilizado. Nas Tabelas 3.2 estão representados os dados referentes à *accuracy*, *sensitivity* e *specifity*, por segmento ou registo, ou para ambos dependendo do estudo. Apenas estão contemplados os melhores resultados para cada estudo, mesmo que os autores tenham apresentados os resultado dos restantes classificadores utilizados. Ainda, na Tabela 3.1, para os estudos que compararam vários classificadores, o que teve maior performance encontra-se a negrito. Por fim, na Tabela 3.3 constam as soluções escolhidas pelos diferentes autores para as diferentes fases do processo.

Autor	Ano	NR	NF	Classificadores
[41]	2018	70	-	Decision fusion (ANN, ANN-HMM, SVM and SVM-HMM)
[42]	2011	30	64	LS-SVM: <b>Linear</b> , Polynomial and RBF
[43]	2019	30	23	KNN, RUSBoost, <b>GentleBoost</b> , Sub-spaceKNN, ANN e SVM
[44]	2020	70	14	Random Forest
[45]	2019	35	12	Gaussian SVM classifier
[46]	2016	35	8	Bagged Decision Trees
[47]	2019	76	-	kNN, <b>SVM</b> , bagging e LogitBoost
[48]	2018	70	15	LD, QD, KNN, SVM e <b>ANN</b>
[49]	2018	35	12	LS-SVM
[21]	2017	30	45-15	ANN
[50]	2020	35	8	SVM-RBF
[51]	2019	35	Auto	CNN using a pre-trained (AlexNet)
[52]	2019	35	-	Naive Bayes
[39]	2019	25	8	Multilayer perceptron (MLP) neural network
[40]	2015	20	6	AdaBoost
[53]	2011	30	8	NARX type ANN
[54]	2011	35	130	SVM
[55]	2019	-	17	Boosted Tree
[25]	2015	35	13	DEREx
[56]	2019	66	-	SVM
[57]	2019	53	10	Residual Neural Network
[37]	2013	35	12	DEREx

Table 3.1: Informação extraída 1 - Legenda: NR: *Number of Records*; NF: *Number of Features*.

Autor	Por segmento			Por registo		
	Acc (%)	Sens (%)	Spec (%)	Acc (%)	Sens (%)	Spec (%)
[41]	84.7	88.9	82.1	100	100	100
[42]	83.3	90	70	-	-	-
[43]	93.26	91.52	94.36	100	100	100
[44]	93.9	92.26	94.92	97.14	-	-
[45]	-	-	-	90.87	92.43	92.61
[46]	85.97	84.14	86.83	-	-	-
[47]	92.59	89.7	94.67	-	-	-
[48]	90.9	89.6	91.9	-	-	-
[49]	90.11	90.87	88.88	-	-	-
[21]	-	-	-	97.2	-	-
[50]	81.06	82.45	79.72	-	-	-
[51]	86.22	90	83.82	86.22	90	83.22
[52]	-	-	-	71	-	-
[39]	98,09	97,14	96,87	-	-	-
[40]	85	80	90	-	-	-
[53]	Learning: 83.33 Test: 82.58	-	-	Learning: 96.66 Test: 100	-	-
[54]	-	-	-	≈90	-	-
[55]	94.7	-	-	-	-	-
[25]	-	-	-	84.26	84.7	81.17
[56]	78.07	78.01	78.13	96.31	-	-
[57]	94.38	95.08	93.24	-	-	-
[37, 38]	-	-	-	88,57	90,32	90,25

Table 3.2: Informação extraída (continuação) - Legenda: Acc: *Accuracy*, Sens: *specificity*, Spec: *specificity*.

Autor	Pré-processamento	Extração de caraterísticas	Seleção de caraterísticas
[41]	Pan-Tompkins algorithm	DNN	-
[42]	DWT	FFT	Hill climbing algorithm
[43]	Autocorrelation fuction (ACF) Pan-Tompkins	Alphabet Entropy (AlphaEn)	SFS algorithm
[44]	Filtering, Denoising and Segmentation	Autoregressive model and Spectral Autocorrelection Function	SFFS Algorithm
[45]	WFB	FUEN and LOEN	Student t-test
[46]	Segmented into 1 min. basis	Statistical and spectral features	Statistical analysis
[47]	FDM (decompose ECG signal into FIBFs)	p-values from Kruskal-Wallis test	-
[48]	Divided into 1 minute segments and RR-peaks founded using Pan-Tompkins	Time Domain and PSD	-
[49]	ECG segmentation and normalization	FUEN and LOEN	Design of optimal JTFS Biorthogonal WFBs and Five-level Wavelet Decomposition
[21]	Pacient division by AHI	Pan-Tompkins	SFFS
[50]	Segmentation of ECG into 1 minute and extraction of EDR and HBI signals from ECG	Extraction of Modes or RCs fromboth HBi and EDR Signals using SM-SSA based data-driven Approach	Evaluation of Instataneous Frequency(IF) and Instatinous Amplitude (IA) from each RC using HilbertTransform.
[51]	Bandpass filter and Pan-Tomkins	Continuous Wavelet Transform (CWT)	-
[52]	Ative Learning	-	-
[39]	Noise cancelation	Discrete Wavelet Transform (DWT)	Principal Component Analysis (PCA)
[40]	Bandpass	ECG: Pan-Tompkins and short-time Fourier transform (STFT)	-
[53]	IHR calculation and FFT-interpolation	Daubechies order 3 (db3) wavelet	-
[54]	Automated Wavelet	-	-
[55]	Removal of unwanted signal & removal signal wandering	Extraction of QRS features	Calcualtion statistic of time and non linear and non linear analysis
[56]	Filtering and Segmentation of ECG Signals into one minute duration frames	Atomic decomposition using orthogonal matching pursuit (OMP) algorithm and varioous dictionaries	Sparse residual entropy (SRE) features
[57]	ECG segmented into 1-min	Neural Network	-

Table 3.3: Pré-processamento, Extração de caraterísticas e Seleção de caraterísticas.

### 3.2.3 Estatísticas da informação retiradas

Nesta subsecção serão apresentados alguns dados estatísticos resultantes da recolha de informação dos estudos incluídos.

Assim como é demonstrado pela tabela 3.1 a maioria do estudos incluídos têm data de 2019, o que corresponde a 39.1% do total, correspondente a 9 estudos. Seguem-se os anos de 2011 e 2018 com 13%, equivalente 3 estudos cada. As restantes datas apresentam valores abaixo de 10% que equivale entre 1 a 2 estudos. A informação encontra-se graficamente representada na Figura 3.1.

A soma do número de registos utilizados pelos autores dá um total de 815 registos, no entanto muitos deles têm a mesma origem. A maioria dos estudos utiliza entre 20 a 35 registos, ou seja, cerca 61% da totalidade dos estudos.

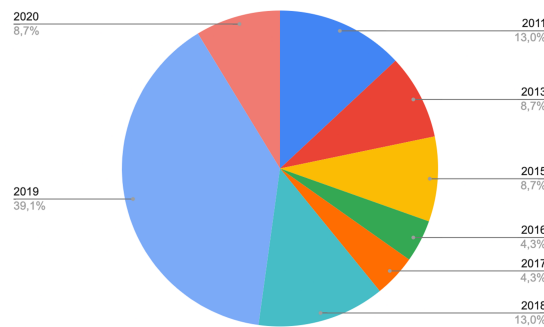


Fig. 3.1: Gráfico da distribuição por data dos estudos.

Existem várias abordagens para tentar encontrar uma resolução ao problema. Dentro dessa abordagens podem-se observar pontos de distinção, com o modo como os sinais são utilizados, ou seja, existem autores que preferem fazer uma análise segmento-a-segmento avaliando a possibilidade de existir evento de AOS nesse segmento de sinal. Porém o registo pode ser avaliado como um todo, ou seja, como *output* teremos se um sujeito têm probabilidade de ter AOS ou não. No entanto a maioria dos estudos recorre à análise segmentada, cerca de  $\approx 70\%$ . Desta percentagem, 37,5% faz duas abordagens em simultâneo. A restante parte utiliza o registo como um todo.

Existe uma prevalência da utilização de *Support-vector machine* (SVM) como classificador pelo autores. Aproximadamente 26% dos estudos selecionados utiliza de alguma forma o SVM [42, 45, 49, 50, 54, 56]. As *Neural network* (NN) são o segundo tipo de classificador mais utilizado com 21% [21, 51, 39, 53, 57]. Alguns autores optam por utilizar vários classificadores ( $\approx 17\%$ ) para encontra o melhor deles, com o melhor resultado [41, 43, 47, 48]. Nestes casos ou o SVM ou as NN são utilizadas ou até ambas, em conjunto com outros algoritmos. Existem ainda 13% que utilizaram uma forma de *Decision Tree*, que correspondem a 3 estudos [44, 46, 55]. Os restantes, utilizam outras formas de ML como classificadores.

### 3.3 Discussão

Como o objetivo que perceber o estado-da-arte dentro do tema de detecção de apnea com a utilização de ECG, foi realizada uma SLR. Nesta secção será apresentada uma discussão a cerca dos resultados apresentados anteriormente, resultantes da SLR, tentando desta forma responder às *Research Questions* propostas.

#### 3.3.1 RQ1 - Porque são necessários modelos de detecção para a AOS?

Existe um consenso entre os autores para o problema a predição e/ou detecção da apneia. Os autores referem que a doença da apneia é uma doença que afeta cerca de 4% da população mundial [54, 25], que afeta mais homens que mulheres [42, 44, 40]. Todos os autores referem pelo menos um problema associado à doença, desde problemas cardiovasculares até à própria morte do paciente, caso este não seja tratado.

Perante este problema a necessidade de diagnóstico é bastante valorizada. O modo de diagnóstico da doença mais utilizado é a PSG, um exame que análise o sono do paciente através de vários sensores, entre eles o ECG. Este é um exame dispendioso, que envolve a monitorização dos sinais ao longo de várias horas e por vezes vários dias. Além do número significativo necessário de equipamentos para o exame, o desconforto do paciente é outro fator tido em conta pelos autores. Sendo o exame realizado num ambiente desconhecido para o paciente e o grande número sensores ligados ao seu corpo durante o seu sono, tornam o exame desconfortável.

No sentido de contornar estes problemas, os autores propõem um forma que seja de algum modo menos dispendiosa em custo, mais confortável e que possivelmente possa ser feita em casa sem o auxílio de pessoal especializado.

#### 3.3.2 RQ2 - Como o ECG é aplicado para detecção de AOS?

O uso de sinais ECG na detecção de apnea computacionalmente, é hoje bastante utilizado [46]. Isto deve-se ao facto do ECG preencher uma das grandes desvantagens da PSG. O ECG, neste contexto, é um método não intrusivo de análise [45, 47, 48] que permite a recolha de informação que podem servir vários fins.

Através da análise morfológica do ECG é possível a extração de características (*features*), como HRV e EDR. Esta última, destacada por vários autores, descreve episódios respiratórios do paciente. Através dela é possível revelar comportamentos respiratórios e, assim sendo, perceber padrões de apneia [42, 48, 21]. Adicionalmente, o HRV, relativo aos intervalos entre os picos R-R do ECG, pode ser utilizado para a detecção da apneia devido às alterações que esta provoca nestes intervalos [42].

O autor [53] destaca ainda uma terceira característica, a *Instantaneous Heart Rate* (IHR). O autor evidencia que existem variações cíclicas nesta característica que descrevem AOS.

### 3.3.3 RQ3 - Quais são as Base de Dados para detecção de AOS?

A maioria dos autores recorre a bases de dados proveniente de um plataforma *web* bastante conhecida por disponibilizar conjuntos de dados clínicos e fisiológicos assim com *software open-source* [58]. De entre as bases de dados existentes na plataforma, distinguem-se três dentro do contexto da AOS:

- *Apnea-ECG Database* [31]: É uma base de dados resultante de uma desafio lançado em 2000 pela *Physionet* para a detecção da AOS utilizando um ECG de variação única [41]. É composta por 70 registos, que forma divididos em 3 grupos (A,B,C) dependendo do tempo de duração do registos em conjunto com o AHI do individuo e o tempo total de episódio de apneia. Destes três grupos resultam 35 registos. A restante metade dos registo é um *set* de dados de teste. Esta base de dados, com algumas exclusões pontuais de alguns registos por alguns autores, foi utilizada em todos os estudos, com a excepção dos estudos [40, 39]. A sua larga utilização deve-se ao facto de existirem anotações no sinal, minuto-a-minuto anotadas por especialistas, que identificam os episódios de AOS, assim como do complexo QRS.
- *St. Vincent's University Hospital / University College Dublin Sleep Apnea Database*: Utilizada pelos autores [47, 39], contém dados de 25 PSG. Em contraste com a base de dados anterior, o sinal ECG resultada 3 variações. Esta base de dados não contém apenas dados de sinal ECG. É possível encontrar registos de sinal electroencefalografia, eletromiografia, entre outros. Perante esta coleta de sinais, são também presentes um maior número de anotações como o estágio do sono e distinção entre os diferentes tipos de apneia do sono (obstrutiva, central, misto e hipopnéia);
- *MIT-BIH Polysomnographic Database* [59]: base de dados que contém registo de PSG's de um laboratório do sono em *Boston*. Foram gravadas com o intuito de ajudar no diagnóstico da AOS. Contém anotações de batimento cardíaco, estágios de sono e episódios de apneia. É utilizado pelo autor do estudo [47].

Todas as bases de dados mencionadas anteriormente então disponíveis em livre acesso pelo *website* da *Physionet* (<https://www.physionet.org>).

No estudo [43] é utilizada uma outra base de dados, também disponibilizada pela plataforma *Physionet*. A base de dados é composta por registos de sinal ECG, com a particularidade de serem obtidos enquanto os sujeitos assistiam o filme *Fantasia* da *Disney* [60].

Casos particulares de base de dados utilizadas ocorrem nos estudos [49] e [40]. No primeiro é utilizado um pequeno *set* de 9 pacientes do laboratório hospital universitário de *Necmettin Erbakan*. No segundo estudo, o autores desenvolveram um dispositivo capaz de obter sinais ECG e dados de um acelerómetro. Com esse dispositivo criou uma *dataset* com 20 registo.

### 3.3.4 RQ4 - Quais são os métodos/técnicas/modelos para detecção de AOS baseada em ECG em registo ou em tempo real?

É possível desde já destacar que a maioria dos autores utiliza um fluxo constituído por pré-processamento, extracção e selecção de características e por fim a utilização de um ou mais classificadores. O genérico desse fluxo encontra-se representado pela Figura 3.2.



Fig. 3.2: Diagrama de bloco genérico do método utilizado para detecção/predição da AOS

Para responder a esta questão, a subsecção será dividida pelas diferentes etapas: pré-processamento, extracção e selecção de características e por fim os classificadores utilizados pelos autores para os seus estudos.

#### 3.3.4.1 Pré-processamento

Um passo comum a todos os estudos incluídos é segmentação do sinal. A maioria dos estudos opta pela segmentação em 1 minuto, com excepção do [53] que opta por uma segmentação de 6 minutos e do [51] que não faz segmentação.

Como qualquer outro sinal, seja ele de carácter biológico ou não, são sempre susceptíveis a ruído que torna impróprios para análise no seu estado bruto. O ECG bruto não é excepção à regra, sofrendo de ruído muscular, movimento basal [51] e até dos próprios equipamentos electrónicos que fazem a recolha do sinal [56]. De forma a colmatar este problema os autores recorrem a diferentes métodos de eliminação do ruído.

Uma das formas recorrentes é a utilização de *high-pass filter* e *low-pass filter*. Neste filtros são utilizados para rejeitar frequências de sinal baixo de um determinado valor (*high-pass filter*) ou frequências de sinal acima de um determinado valor (*low-pass filter*). A utilização do conjunto das duas dá-se o nome de *bandpass filter*. Esta técnica é utilizada pelos autores [44, 51, 39, 40, 56].

Outros autores optam pela utilização de *wavelet transform* [42, 45, 49, 39, 53, 54, 55]. Este tipo de abordagem permite a decomposição do sinal, e assim remover ruído. Contudo, além da remoção de ruído, é também possível a possível a extracção de características através deste método [39, 56], criando-se assim uma ponte entre o pré-processamento e o extracção de características. Um caso particular ocorre no estudo [51], onde um *wavelet* específica (*continuous wavelet transform (CWT)*) é utilizada para transformar o sinal em representações de escalograma de frequência de tempo 2D.

No estudo [47] diferente dos demais, os sinais são decompostos utilizando o método de decomposição de *Fourier* (FMD), que gera as chamadas *Fourier intrinsic band functions* (FIBFs). As FIBFs que contém ruído são registadas para os próximos passos.

Adicionalmente, dependendo os requisitos do estudo, continuando na linha de preparação do sinal é também comum os autores utilizam um algoritmo bem conhecido, o *Pan-Tompkin's* [41, 43, 48, 21, 50, 51, 56]. Através dele é possível identificar os picos R do sinal. Assim

como acontece com as *wavelet transform*, existem autores que consideram a sua utilização como parte do pré-processamento e outros com parte da extração.

Foram ainda encontrados dois autores [45, 49] que recorrem à normalização do sinal através de *Z-score*.

### 3.3.4.2 Extração de características

Assim como está refletido na subsecção anterior, alguns autores consideram a *Wavelet* nesta fase do *workflow*, mais especificamente os estudos [39, 53]. Em particular no estudo [39] é utilizada a *Discrete Wavelet Transform (DWT)* em detrimento das Transformações de *Fourier*, afirmando que é uma forma mais eficiente para este passo. O estudo [42], contrariamente utiliza a Transformação Rápida de *Fourier*.

Para extrair as características necessárias ao cálculo do HRV, os estudos [21, 40, 55] utilizam ao algoritmo de *Pan-Tompkin's*. Tal como dito na subsecção 3.3.2, o uso do HRV é bastante recorrente devido as características que deles se podem tirar, como por exemplo o intervalo entre ondas do sinal ECG.

No estudo [44] são utilizados dois métodos de extração, o autoregressivo e função de autocorrelação espectral. Na primeira é extraída o seu coeficiente que servirá com característica e na segunda é utilizada para extrair características do domínio de frequência.

Os estudos [45, 49] utilizam as características *Fuzzy entropy (FUEN)* e *Log-energy (LOEN)*, entanto que no estudo [43] foi utilizado o *Alphabet Entropy (AlphaEn)*.

No estudo [46] é recorrido a características estatísticas (média, variância e *kurtosis*) e espectrais, estas últimas com a ajuda da transformação discreta de *Fourier*.

No estudo [47] com a utilização da já mencionada FDM são extraídas, por exemplo a máxima e mínima dos diferentes ondas do ECG.

Os autores do estudo [50] através da decomposição do sinal EDR conseguiram extrair características como frequência e amplitude instantânea.

Casos particulares são os estudos [41, 57] em que são utilizadas NN para este processo.

Existem alguns estudos que não mencionam como fizeram as extração de características [52, 54, 25, 37, 38].

### 3.3.4.3 Seleção de características

Assim como a anterior, esta etapa não é executada por todos os autores. Contudo para aqueles que utilizam, o método mais utilizado foi o algoritmo de *Sequential Feature Selector* pelos [43, 44, 21]. Os últimos dois utilizam em particular os o *Sequential Forward Feature Selection*. Esta método é utilizado para encontrar o melhor *set* de características possíveis, através da escolha sequencial da melhor característica de um subconjunto ideal. Os autores do estudo [42] optaram pela utilização do algoritmo *hill-climbing* para encontrar o melhor *set* de características.

No estudo [39] é usado um método chamado *Principal Component Analysis*, para diminuir a dimensão das características para mais tarde o classificador não cometa o erro de não conseguir distinguir entre dois grupos de características.

### 3.3.4.4 Classificador

Assim como foi dito na secção 3.2.3, o classificador mais escolhido pelos autores é o SVM. Neste, por norma esta associado a um *kernel* que pode ser por exemplo polinomial ou uma função racional [51]. Contudo, nem todos os autores referem o *kernel* que estão a utilizar. Sendo um algoritmo com bastante eficiência, esta pode ser a razão para a sua elevada escolha.

Segue-se a utilização de NN, como sendo a segunda alternativa mais escolhida [21, 51, 39, 53, 57]. Através de uma NN tenta-se fazer uma abstração do cérebro Humano para a computação. Em particular no estudo [53] é utilizada um NN com apelido de *NARX*, que contém *input layer* com 6 neurónios, 3 no *hidden layer* e o *output layout* com 1.

As *decision trees* foram as terceiras mais utilizadas pelos autores [44, 46, 55]. Neste tipo de abordagem é formada uma tabela de decisão em forma de árvore.

Existem ainda alguns autores que fazem uma seleção de classificados, tendendo perceber qual atinge maior performance [43, 44, 48, 50, 55]. Em particular, os estudos [41, 51] utilizam a técnica de *decision fusion* para fazer o teste dos vários classificadores selecionados.

Diferente dos demais, os estudos [25, 37, 38] utilizam um algoritmo de extração de regras do tipo *IF-THEN* do sinal ECG. Estas regras são posteriormente inseridas num *Rules Engine Module* para fazer a previsão de episódios de AOS em tempo real num sistema móvel desenvolvido pelos autores.

### 3.3.5 RQ5 - Quais são os parâmetros usados para avaliar o desempenho da detecção de AOS?

Tendo em conta que quase todos os estudos utilizam a base de dados *Apnea-ECG*, os autores recorrem à medir dos *True Positives*, *False Positives*, *True Negative* e *False Negative*, alguns chegam a representar as suas *Confusion Matrix* [41, 51]. Com a sua utilização é medida posteriormente a precisão (ACC), sensibilidade (SENS) e especificidade (SPEC) do método proposto, que são dadas pelas seguintes equações:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.1)$$

$$SENS = \frac{TP}{TP + FN} \quad (3.2)$$

$$SPEC = \frac{TN}{TN + FP} \quad (3.3)$$

onde, TP, TN, FP e FN correspondem a *True Positive*, *True Negative*, *False Positive* e *False Negative*, respetivamente.

### 3.4 Conclusão

Graças ao desafio proposto pela *Physionet* em 2000 [22], com a disponibilização de dados de sinais ECG à comunidade científica, abriu portas ao desenvolvimento de modelos de deteção da AOS com o uso do biosinal. Essa valência ainda se reflete, hoje em dia, pela grande utilização da base de dados no contexto da AOS.

É evidente que a demanda da criação modelos de predição ou deteção da AOS é cada vez maior na comunidade científica. Este facto é refletido pela quantidade de artigos encontrados nos últimos três anos, em especificamente em 2019 onde residem a maioria dos estudos incluídos.

É também possível observar que não existe uma abordagem única para o problema. Diferentes autores, apesar de seguirem maioritariamente o mesmo fluxo de trabalho (Fig. 3.2), optam por utilizar diferentes algoritmos para as diferentes fases, na tentativa de alcançarem a melhor performance possível. Dentre os algoritmos utilizados destaca-se a utilização de ML, com a exceção dos estudos [41, 50, 51, 57] que utilizam *Deep Learning* (DL). Destaca-se a utilização de SVM não só pela popularidade, com também pela sua performance nos resultados.



# Capítulo 4

## Ferramentas e Tecnologias Usadas

A utilização de várias tecnologias e ferramentas foram indispensáveis para o desenvolvimento da dissertação. Neste capítulo, serão expostas e explicadas, sucintamente, as tecnologias e ferramentas utilizadas.

Na Figura 4.1 encontram-se esquematizadas as ferramentas e tecnologias utilizadas.

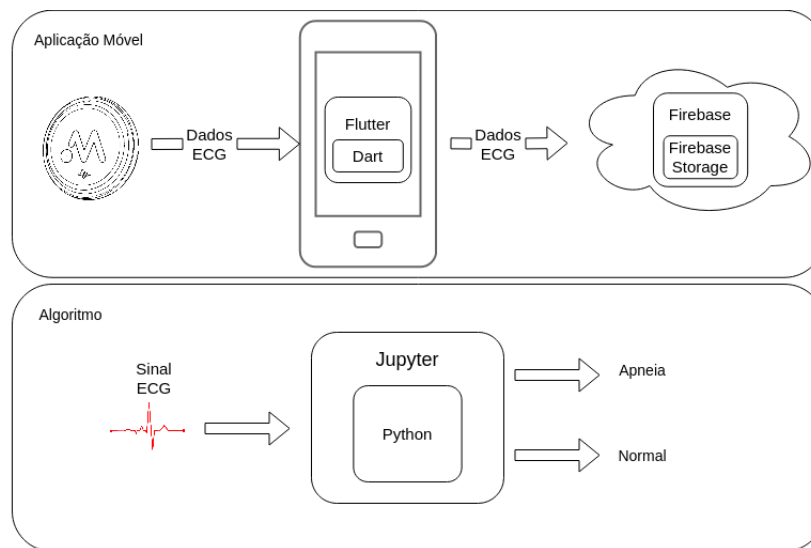


Fig. 4.1: Ferramentas e Tecnologias utilizadas em contexto de desenvolvimento.

### 4.1 Ferramentas

#### 4.1.1 Jupyter

*Jupyter* – também conhecido por *Jupyter Notebook* – é uma aplicação que permite criação e edição de documentos que contenham *live code*, equações, representações gráficas, tabelas, texto, entre outros. A aplicação resulta de um projeto intitulado de *Jupyter Project*, uma *spin-off* do *IPython*, sem fins lucrativos, criada em 2014.

É uma ferramenta utilizada pela comunidade científica devido a sua versatilidade, nomeadamente devido ao facto de funcionar tanto como editor de texto, como interpretador de código.

Para base de suporte a esta diversidade de informação num só ficheiro, o Jupyter recorre aos chamados *Notebook* ou *notebook documents*, assim como acontece com o *Google Colab*. Ficheiros que podem deter tanto código e *rich text*, como figuras, links, equações, etc. Apresentam uma extensão própria denominada *.ipynb*. Além destes ficheiros, a aplicação suporta outro tipo de ficheiros como CVS, imagens, etc, assim como editores de texto mais avançados.

Além do seu *kernel* base – *IPython* – que fornece suporte à interpretação de código *Python*, são suportadas outras linguagens de programação, como *R*, *Haskell* e *Ruby*, mediante a instalação dos respetivos *kernels* [61].

A aplicação dispõe de duas interfaces, o *JupyterLab* e o *Jupyter Notebook*. A primeira é uma interface *user friendly*, com mais recursos e funcionalidades em comparação com a última que apresenta uma interface mais simplista.

## 4.2 Tecnologias

### 4.2.1 *Flutter*

*Framework open source* focado no desenvolvimento de interfaces gráficas para multi-plataforma, suportada e desenvolvida pela *Google*. Tem como linguagem base a *Dart*, desenvolvida pela mesma empresa e lançada em 2011. A linguagem tem uma sintaxe bastante próxima do *JavaScript* e o seu principal paradigma é a Programação Orientada a Objetos (POO). Atualmente, a *framework*, encontra-se na versão 2.5, lançada em outubro de 2021. O *Flutter* teve o seu lançamento mais tardio em maio de 2017. Dá suporte ao desenvolvimento para plataformas como *Android*, *iOS*, sistemas embutidos, *Web* e *desktop* (*Windows*, *Mac* e *linux*).

### 4.2.2 Bibliotecas e pacotes *Flutter*

À semelhança do que acontece com outras linguagens de programação ou *frameworks*, existe uma vasta quantidade de bibliotecas e pacotes que auxiliam no desenvolvimento. Não fugindo a esta regra, segue-se a descrição das bibliotecas e pacotes que foram utilizados na aplicação criada:

- **csv:** biblioteca simples que permite a conversão de listas em arquivos do tipo *csv*;
- **ext\_storage:** providencia o diretório interno do dispositivo onde a aplicação está instalada. Na aplicação desenvolvida foi utilizado para conseguir gravar um *backup* dos dados recolhidos no num directório personalizado do dispositivo;
- **firebase\_core:** pacote desenvolvido pela *Google* para a integração base dos serviços *cloud Firebase*. Este é indispensável para utilização de qualquer serviço dessa *cloud*;
- **firebase\_storage:** permite a utilização do serviço de armazenamento (*storage*) da *Firebase*. No âmbito da aplicação foi utilizada para o armazenamento em *cloud* dos dados recolhidos;
- **mdsflutter:** pacote desenvolvido por terceiro para a integração da Application Programming Interface (API) *Movesense*, empresa responsável pela produção do dispositivo;
- **syncfusion\_flutter\_charts:** pacote que permite a visualização de dados. Dispõem de diferentes tipos de gráficos para a integração nas aplicações. Foi utilizado para representar o sinal ECG, de modo a dar um *feedback* ao utilizador.

### 4.2.3 *Firestore*

Inicialmente criada por uma empresa com o mesmo nome em 2011, foi posteriormente adquirida em 2014 pela *Google*, que é atualmente a responsável pelo seu desenvolvimento e suporte. É uma plataforma *cloud* com foco a entregar um *back-end* completo através das suas ferramentas. O seu objetivo principal é entregar soluções prontas que facilitem a escalção de uma aplicação para a *cloud*. Maioritariamente pensada para aplicações móveis (*Android* e *iOS*), neste momento permite a integração com desenvolvimento *web*, *Unity* e *C++*. Como o *Firestore* pertence ao universo *Google*, permite ainda a escalção dos seus projetos para a *Google Cloud Platform* que oferece mais ferramentas.

O principal serviço utilizado pela aplicação desenvolvida foi o *Firestore Storage*, que permite o armazenamento de ficheiros em ambiente *cloud*, quem seguidamente poderão ser utilizados por outros serviços do *Firestore*.

### 4.2.4 *Python*

*Python* é uma linguagem de programação de alto nível, de uso geral, *open-source*, lançada em 1991, pelo holandês *Guido van Rossum*. Segundo as estatísticas do site *TIOBE*, *Python* é a linguagem de programação mais popular a nível mundial [62]. É utilizada para diversos fins, desde o desenvolvimento de aplicações *GUI* a plataformas *web*. A linguagem foi pensada de forma a ter uma sintaxe simples, que facilitasse a manutenção do código, tornando-o mais legível e intuitivo. *Python* assenta em vários paradigmas da programação, tais como programação estruturada, POO, funcional e orientada a aspetos. Além destas características, o *Python* possui um sistema do tipo dinâmico e gestão automática de memória. Em consequência da sua essência *open-source*, o *Python* possui uma vasta gama de *frameworks*, bibliotecas e ferramentas de desenvolvimento de terceiros, que ajudam a uma programação menos demorada e mais eficiente. A linguagem encontra-se na versão 3.10.1, lançada em dezembro de 2021. O seu *core* é escrito em linguagem *C*, designado por *CPython* e pode ser encontrado na plataforma *Github*.

### 4.2.5 Bibliotecas *Python*

Ao longo do desenvolvimento, foram utilizadas várias bibliotecas *Python*, que serviram de veículo para alcançar o objetivo do projeto. Seguem-se abaixo, a descrição das mesmas:

- ***biosppy***: “*toolbox*” que ajuda no processamento de bio-sinais em *Python*. No âmbito desta tese foi utilizada para o pré-processamento do sinal ECG da base de dados *Apnea-ECG* [31];
- ***HRVAnalysis***: permite a extracção e análise do HRV e das *features* que o compõem. Foi desenvolvido em cima das outras bibliotecas, como *SciPy* e *NumPy*;
- ***Pandas***: uma das bibliotecas mais utilizadas na estruturação e análise de dados em *Python*. No âmbito do projeto é utilizado para a leitura e manipulação de dados contidos em ficheiros *.csv*;

- **Pickle:** módulo que permite a implementação de protocolos binários para a serialização e vice-versa de estruturas de objetos *Python* [63];
  - **Matplotlib:** a biblioteca mais comumente utilizada para a representação de gráficos em *Python*. Esta permite a representação de uma ampla variedade de gráficos, como estatísticos, com imagens, em *3D*, entre outros;
  - **Numpy:** pacote para a computação científica em *Python*. Tem foco principal em matrizes multidimensionais e *arrays*. Esta faz parte de uma biblioteca mãe, apelidada de *SciPy*;
  - **Sklearn:** biblioteca focada em entregar soluções para o desenvolvimento em ML. Disponibiliza implementação rápida e simples de diferentes classificadores, modelos de seleção, funções de extração de *features*, entre outros recursos. A biblioteca foi essencialmente utilizada para a criação de um modelo de classificação para a apneia através dos classificadores disponibilizados;
  - **tqdm:** pequena biblioteca que permite a criação de barras de progresso;
- wfdb-python:** disponibilizada pela *PhysioNet*, é uma biblioteca que permite a leitura, escrita e processamento dos ficheiros da mesma plataforma. Foi utilizado na extração do sinal ECG e dos respetivos *labels*.

### 4.3 Conclusão

O uso de ferramentas e tecnologias como as descritas anteriormente formam um veículo de auxílio no desenvolvimento de projetos nas mais diversas áreas. É possível afirmar que sem a sua existência, a execução da presente dissertação não seria possível. Tomando essa linha de pensamento, a aplicabilidade e versatilidade das mesmas também são dois dos fatores que permitem a continuidade do progresso tecnológico.

# Capítulo 5

## Dispositivo e Aplicação Móvel

Após a análise da Literatura feita no Capítulo 3, constatou-se que a aplicação em contexto real dos modelos propostos não era considerada pelos autores. Perante este pormenor, neste capítulo é proposta uma pequena aplicação móvel que recolhe dados ECG de pacientes com suspeitas ou já diagnosticados para a AOS. Aliado à aplicação foi estudado e utilizado um dispositivo de fácil utilização e que permita a recolha de dados ECG com qualidade médica. Complementando a aplicação e o dispositivo, é apresentado um protocolo de recolha com a Unidade de Saúde Local (ULS) da Guarda, a fim de encontrar um público-alvo elegível para a recolha dos dados.

Neste capítulo, segue-se uma descrição detalhada dos critérios de seleção para a escolha do dispositivo, assim como uma breve descrição do mesmo. Por fim, encontra-se uma descrição da aplicação desenvolvida e ainda o protocolo, fruto da parceria criada com a ULS da Guarda.

### 5.1 Dispositivo

#### 5.1.1 Critérios de seleção

De modo a encontrar o melhor dispositivo possível e aquele que melhor se adequar às necessidades do presente estudo, foram considerados alguns requisitos para esse dispositivo: (1) o equipamento terá de ser disponibilizado através de fontes seguras que assegurem a sua garantia; (2) Terá de ser certificado por entidades responsáveis que garantam a qualidade dos dados recolhidos; (3) Disponibilizar ferramentas de desenvolvimento com *Software Development Kit* (SDK) e/ou API ou código aberto que permitam perceber o funcionamento do dispositivo e construção de soluções personalizadas; (4) Terá de contar com documentação detalhada que facilite a compreensão do funcionamento do dispositivo, como também ajude no desenvolvimento de *software* de terceiros; (5) Conexão via *Wi-Fi* ou *Bluetooth*.

#### 5.1.2 *Movesense Medical Developer Kit* (MMDK)

Fabricado na Finlândia pela *Movesense*, subsidiária da empresa *Suunto*, este kit é constituído por um dispositivo com a capacidade de recolher dados ECG e de movimento para uso medicinal. Para isso este conta com um sensor de canal único (*one lead*) ECG. Além da recolha de dados, também estão disponíveis funcionalidades de medição de frequência cardíaca e intervalos R-R [64].

O objetivo principal deste dispositivo é disponibilizar o acesso a biosinais e outros de nível medicinal, ao mesmo tempo que permite a fácil integração e desenvolvimento de

aplicações para os programadores. Assim, os programadores dispõem de uma abstração sólida, capaz de recolher dados de alta qualidade, retirando a necessidade de desenvolver as suas próprias soluções de *hardware* [2].

Conta com uma API aberta e detalhada, que facilita a sua utilização, nomeadamente no desenvolvimento de aplicações. Para facilitar ainda mais o desenvolvimento à volta do dispositivo, estão também disponíveis bibliotecas desenvolvidas por terceiros, para a integração com diferentes ferramentas/*frameworks*.

Na conectividade, o dispositivo utiliza uma antena *Bluetooth* incorporada, pela qual é feito o envio e receção de dados e disponibilizado funcionalidades como a atualização do *firmware* e agregar código personalizado.

Mais concretamente, no sinal ECG é possível realizar recolha em frequências de 128, 256 e 512Hz. Ademais, a resolução é de 15 *bits*.



Fig. 5.1: *Movesense Medical Developer Kit*. Fonte:[2]

### 5.1.3 Discussão

Tratando-se da recolha de dados sensíveis, a necessidade do dispositivo ter certificados que garantam a sua qualidade é essencial para a viabilidade do estudo em questão. O MMDK é certificado pela *Eurofins Expert Services*, entidade certificadora de diversos produtos, incluindo dispositivos médicos. Estão presentes dois certificados, um com a norma *ISO 13485:2016* e o último com a diretiva europeia médica *93/42/EEC*, dando-lhe a classe *Ila*.

De entre todos os dispositivos encontrados para o estudo, o MMDK é o único que cumpre com os critérios de seleção.

O MMDK funciona no mesmo ambiente de desenvolvimento da *Suunto*, onde é facultado o código fonte, assim como uma API e até a arquitetura na qual o sistema está montado.

## 5.2 Aplicação Móvel - *ApneaECG*

No âmbito de recolha de dados ECG, foi desenvolvida uma aplicação móvel com recurso à *Framework Flutter*, descrita na secção 4.2.3. Nesta secção encontra-se uma descrição detalhada da aplicação, os requisitos propostos

### 5.2.1 Descrição da aplicação

A aplicação é constituída por dois *layouts* simples, onde o utilizador inicialmente se pode conectar com um dispositivo *Movesense*. Posteriormente, é possível iniciar a recolha de

dados como está representado na Figura 5.3. A aplicação é responsável pela recolha, armazenamento e envio dos dados recolhidos para ambiente *cloud*. Está desenhada para funcionar no sistema operativo *Android*.

Além da *framework* utilizada, a aplicação é suportada por uma biblioteca que faz ponte de ligação com API da *Movesense* para o ambiente *Flutter*. Em concreto, a API *Movesense* é construída no formato *REST* com o métodos *HTTP* mais comuns *PUT*, *GET*, *POST* e *DELETE*. É através deles que é possível obter os dados e assim interagir com o dispositivo. As respostas aos métodos, são formatadas em tipo *YAML*, que permite uma serialização dos dados.

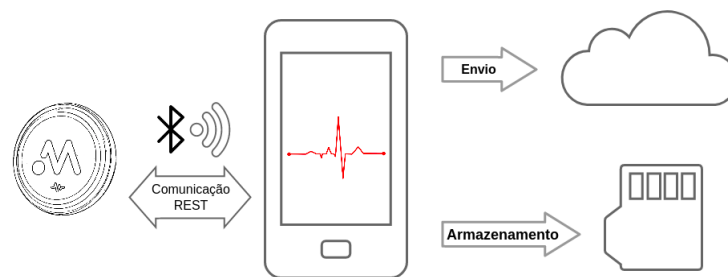


Fig. 5.2: Esquema geral da aplicação ApneaECG. Adaptado de [3]

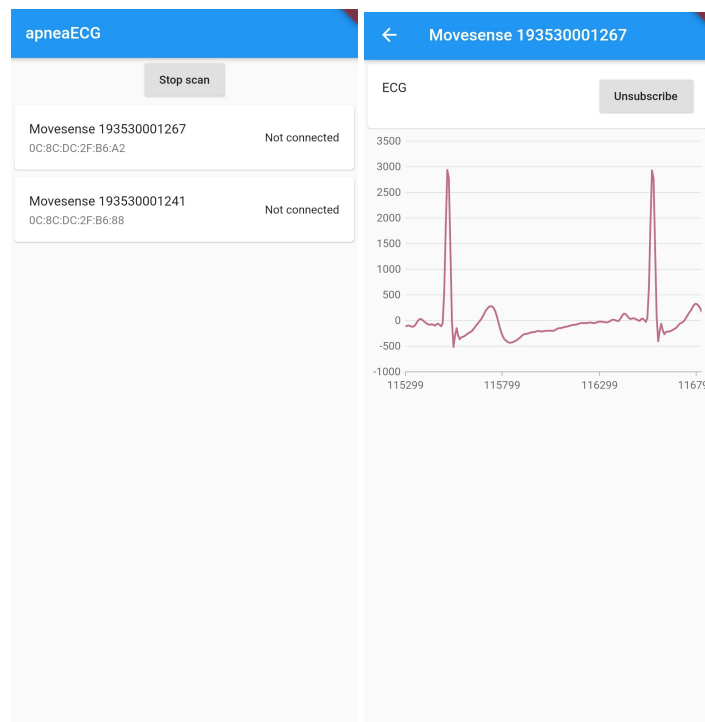


Fig. 5.3: Layouts da aplicação ApneaECG.

## 5.2.2 Engenharia de software

Apesar desta ser uma aplicação bastante simples com o único propósito da recolha de dados ECG através do dispositivo, é natural que faça um estudo da engenharia de software para que determinados requisitos sejam cumpridos, de modo que a recolha de dados não seja comprometida. Seguintes subseções seguintes os requisitos funcionais, não funcionais e de domínio, assim com os casos de uso.

### 5.2.2.1 Requisitos

Nas tabelas seguintes, respetivamente, os requisitos funcionais, não funcionais e de domínio para a aplicação.

RF01	A aplicação permitirá a mostrar os dispositivos <i>Movesense</i> disponíveis;
RF02	A aplicação permitirá a escolha de um dispositivo <i>Movesense</i> disponíveis;
RF03	A aplicação permitirá que iniciação e cessão da recolha de dados ECG;
RF04	A aplicação permitirá a visualização do sinal ECG em tempo-real;
RF05	A aplicação permitirá saber a qual dispositivo está ligado;
RF05	A aplicação permitirá fazer a escolha voltar à página inicial;

Table 5.1: Requisitos Funcionais.

RNF01	A aplicação deverá fazer a busca dos dispositivos <i>Movesense</i> disponíveis para conexão;
RNF02	A aplicação deverá fazer um <i>backup</i> interno dos dados recolhidos;
RNF03	Os dados recolhidos deverão ser enviados para a plataforma <i>cloud</i> ;
RNF04	Os dados recolhidos deverão ser gravados com a data em que são recolhidos e com a de extensão <i>.csv</i> ;
RNF05	Os dados recolhidos deverão ser gravados no tipo de extensão <i>.csv</i> ;

Table 5.2: Requisitos não Funcionais.

RD01	O dispositivo móvel (smartphone ou tablet) deverá estar ligado a uma rede <i>Wi-Fi</i> ;
RD02	A recolha de dados apenas será bem sucedida se o dispositivo <i>Movesense</i> estiver bem colocado;
RD03	A recolha de dados apenas será bem sucedida se não existir uma interrupção do sinal <i>Bluetooth</i> ;

Table 5.3: Requisitos de Domínio.

### 5.2.2.2 Casos de uso

Assim como foi referido anterior, a aplicação proposta é um sistema simples que apenas tem como intuito a recolha de respetivos dados. Nesse sentido o caso de uso possível reflete-se nesse mesmo processo. Na Figura 5.4 encontra-se representado esse mesmo caso.

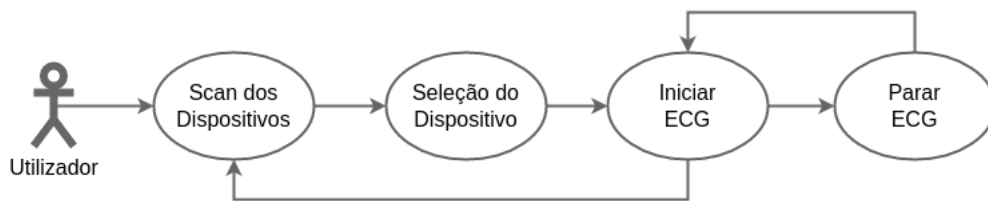


Fig. 5.4: Diagrama de Caso de Uso - *Efetuar recolha de dados ECG*

### 5.3 Protocolo de Recolha de Dados

No âmbito da recolha de dados foi criado um protocolo com a ULS da Guarda, mais concretamente com o Serviço de Pneumologia da mesma. No protocolo é possível encontrar a Descrição do problema, o objectivo do protocolo e do estudo em questão, a Metodologia na qual são apresentados os critérios de elegibilidade, a seleção da amostra, os materiais a serem usados, breves instruções de utilização, descrição do dispositivo, procedimento de recolha, entre outros pontos fulcrais para o bom funcionamento da recolha de dados. É ainda possível encontrar no documento os Aspectos Éticos e Confidencialidade dos Dados, onde é detalhada a anonimização dos dados recolhidos.

Este teve parecer positivo por parte da Comissão de Ética da ULS, porém, foi tardio e tornou-se impossível colocá-lo em prática dentro do tempo de execução da presente dissertação. O Protocolo encontra-se em anexo no Apêndice A.

### 5.4 Conclusão

O conjunto dos três componentes descritos ao longo deste Capítulo, aproximam o contexto científico ao mundo real. A recolha de dados de pacientes com suspeita ou com diagnóstico para a AOS, para o teste do algoritmo/modelo proposto é o componente que mais reflete esta questão. Deste modo, pretendeu-se testar o modelo com dados diferentes para além daqueles que são sugeridos pela Literatura, e assim perceber o seu comportamento em condições reais.

No entanto, este é apenas um passo inicial para a aproximação dos dois contextos. Existem ainda componentes a serem desenvolvidos, como a aplicação do modelo de classificação de AOS em *Cloud*, para que se crie uma ferramenta que ajude equipas médicas no diagnóstico da AOS.



# Capítulo 6

## Implementação e Resultados

Este capítulo descreve todo o processo de desenvolvimento do algoritmo (modelo) de detecção de apneia com recurso a sinais ECG. Posteriormente, são apresentados os resultados obtidos a partir do mesmo.

### 6.1 Implementação

Durante o processo de criação do modelo foram utilizados exclusivamente os dados da base de dados *ApneaECG* [31], disponível em acesso livre na plataforma *Physionet* [58]. A sua escolha deve-se essencialmente à sua viabilidade, contendo dados bem estruturados que facilitam na construção de novos modelos de detecção. É também de salientar, que esta é a base de dados mais utilizada por outros autores, assim como revelou a Revisão Sistemática.

À semelhança da Literatura revista pelo Estado da Arte no Capítulo 3 o processo de desenvolvimento do algoritmo seguiu um *workflow* semelhante aos demais. A sua representação encontra Figura seguinte:

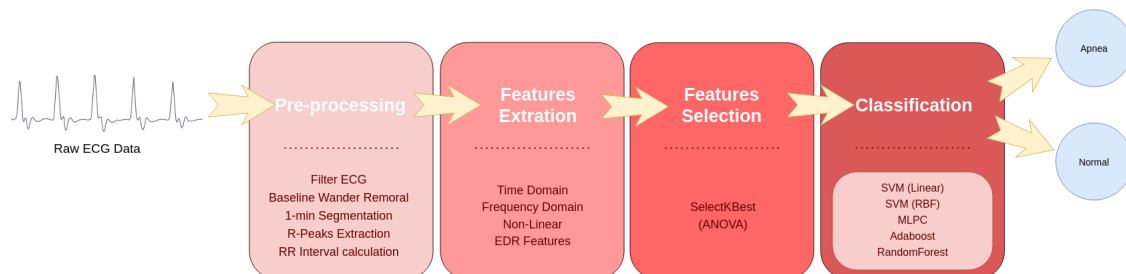


Fig. 6.1: *Workflow* de desenvolvimento do Algoritmo.

#### 6.1.1 Pré-processamento

Em primeira instância, são extraídos os sinais ECG da BD da *Physionet*. Complementado o que foi descrito na Seção 3.3.3, os dados estão a uma frequência de 100Hz e cada registo tem cerca de 8 horas de gravação. Agregado a estes dados, seguem as anotações feitas por especialistas do sono. Indicam a presença de apneia ou não no segmento respetivo, minuto a minuto. As anotações correspondem a "A" para um segmento com apneia e "N" para um segmento normal, que no pré-processamento são transformados em notação binária, ou seja, 1 e 0, respetivamente.

Como é natural para qualquer sinal, estes são suscetíveis a ruído que alteram a sua forma. Perante este problema é comumente aplicado ao sinal um ou mais filtros que eliminam

o ruído. No caso do ECG extraído foi aplicado um filtro do tipo *Finite Impulse Response* (FIR) com *bandpass* de 3 ~ 45Hz e por fim uma *baseline correction* através de um *median filter*.

Para concluir esta etapa, foram detetados e extraídos os picos *R* do sinal ECG por meio do algoritmo de detecção do complexo *QRS* desenvolvido por *P. Hamilton* [65]. Este algoritmo tem vários pontos de similaridade com um dos mais conhecidos algoritmos de detecção do complexo *QRS*, o *Pan-Tompkins*. De entre os algoritmos disponibilizados pela biblioteca *Biosppy*, o algoritmo *Hamilton* é um dos que oferece uma implementação mais simplificada e rápida. Perante a possibilidade de existirem picos fisiologicamente impossíveis, foi aplicado um novo *median filter* para os excluir. Este filtro define um intervalo de tempo, representado por:

$$[R - tol, R + tol[ \quad (6.1)$$

onde, *R* representa o valor numérico de um pico *R* e *tol* representa a tolerância de correção previamente definida.

Por fim é calculado o *RR interval* (RRI), que se compreende por o intervalo entre um pico *R* e o seu subsequente. O próprio intervalo já constitui uma característica do sinal ECG, porém, existem mais características que podem ser extraídas a partir dele. Na Secção seguinte serão mostradas algumas dessas características.

## 6.1.2 Extração e Seleção de características

### 6.1.2.1 Extração

Sumarizando o que foi descrito até agora ao longo do documento, a partir do sinal ECG é possível extrair características, que podem servir para a identificação de uma determinada patologia. No plano desta dissertação essas características serão utilizadas para a identificação de episódios de Apneia Obstrutiva do Sono. Na totalidade foram extraídas 34 características/*features* diferentes do sinal ECG.

Diferente de *Heart Rate* (HR) que conta o número de batimentos por minuto, o HRV é a variação do batimento cardíaco, ou seja, a variação nos intervalos de tempo entre batimentos adjacentes [66]. Em norma o HRV corresponde diretamente ao RRI, que define o intervalo entre picos *R*, como anteriormente visto.

As características resultantes da observação do HRV são divididas em vários grupos, como *time-domain*, *frequency-domain*, *non-linear* [67]. Para o desenvolvimento do algoritmo foram extraídas características dos três grupos mencionados. Seguem-se nas tabelas seguintes as mesmas.

Adicionalmente, foi extraído o sinal EDR através do método proposto por *Sarkar et al.* [68], que faz uso do HRV. Através deste sinal foram extraídas novas características que se seguem na Tabela 6.2. Na Figura subsequente, encontram-se esquematizadas os diferentes campos de origem das características.

É possível verificar que na Tabela 6.2 surge a nomenclatura "*NNI*" e "*NN*". Estes termos

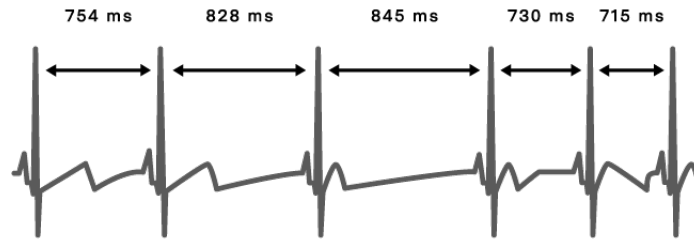


Fig. 6.2: Heart Rate Variability. Fonte: [4]

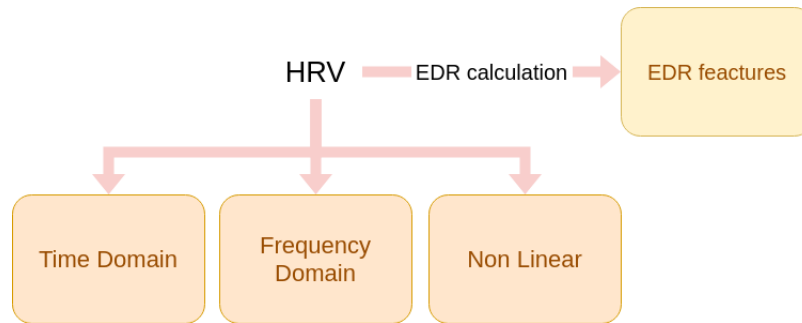


Fig. 6.3: Features Domain.

referenciam-se a "Normal-to-Normal" e são utilizados para enfatizar que foram utilizados batimentos "normais", ou seja, batimentos anormais foram removidos a priori [67]. Todavia, estes não deixam de se referir aos picos *R*, anteriormente já mencionados. As características tabeladas em 6.1 têm base as descrições apresentadas na documentação da biblioteca *HRVAnalysis* [69, 70], utilizada para a extração, e do estudo feito por *Shaffer, F et al.* [67].

Por último, nesta etapa do processo foi efetuada uma normalização dos dados através do *StandardScaler*. Esta é uma função disponibilizada pela *Sklearn*, definida pela equação de *Standardization* seguinte, que também é denominada por *Z-Score*:

$$z = \frac{x - \mu}{\sigma} \quad (6.2)$$

onde,  $x$  é a amostra,  $\mu$  a média da amostra de treino e  $\sigma$  o desvio padrão da amostra.

### 6.1.2.2 Seleção

A Seleção de Características permite a limpeza do ruído (em inglês, *noise*) que por vezes é criado pelas mesmas. Apesar das características serem essenciais para a classificação por um algoritmo, estas, por vezes, podem "baralhar" o classificador e diminuírem a eficiência do próprio. Nesta sequência, a seleção das características mais relevantes é, por norma, também uma etapa fundamental para a otimização dos algoritmos desenvolvidos. Por norma, a mesma traz benefícios com aumento da eficácia, redução do tempo de treino e, assim como dito anteriormente, remove o ruído.

Table 6.1: Caraterísticas/*Feactures* extraídas do HRV.

Caraterística	Descrição
Time Domain	
Mean_NNI	Média do intervalo RR
SDNN	Desvio padrão do intervalo de tempo entre batimentos cardíacos normais sucessivos
SDSD	Desvio padrão das diferenças entre os intervalos RR adjacentes
RMSSD	Raiz quadrada da média da soma dos quadrados das diferenças entre intervalos NN adjacentes
Median_NNI	Valor médio do intervalo RR
NNI_50	Número de diferenças de intervalo de intervalos RR sucessivos maiores que 50ms
PNNI_50	Proporção derivada da divisão de NNI_50 pelo número total de intervalos RR
NNI_20	Número de diferenças de intervalo de intervalos RR sucessivos maiores que 20ms
PNNI_20	Proporção derivada da divisão de NNI_20 pelo número total de intervalos RR
Range_NNI	Diferença entre o máximo e mínimo do intervalo RR
CVSD	Coefficiente de variação de diferenças sucessivas iguais ao rmissd dividido por Mean_NNI
CVNNI	Coefficiente de variação igual à razão de sdnm dividido por Mean_NNI
Max_HR	Valor máximo do HR
Min_HR	Valor mínimo do HR
STD_HR	Desvio padrão do HR
Mean_HR	Média do HR
Frequency Domain	
VLF	Potência na banda de <i>Very Low Frequência</i>
LF	Potência na banda de <i>Low Frequency</i>
HF	Potência na banda de <i>High Frequency</i>
LF_HF_Ratio	Razão da potência entre LF e HF
LFNU	Potência LF normalizada
HFNU	Potência HF normalizada
Total_Power	Espectral de densidade de potência total
Non-Linear	
SD1	Desvio padrão do Poincaré perpendicular à linha identidade
SD2	Desvio padrão do Poincaré ao longo da linha identidade
SD1SD2	Razão entre S1 e S2
CSI	<i>Cardiac Sympathetic Index</i>
CVI	<i>Cadiac Vagal Index</i>
Modified_CSI	CSI modificado

Table 6.2: Caraterísticas/*Feactures* extraídas do EDR.

Feacture	Descrição
Mean_Resp	Média da frequência respiratória
Mean_Period	Período respiratório médio
Resp_RMS	Raiz quadrada da média da respiração
Resp_STD	Desvio padrão da respiração
Max_freq	Frequência máxima

O campo da seleção de caraterísticas divide-se em dois grupos: supervisionado e não supervisionado. Este pode ser ainda subdividido. No caso do supervisionado existe o método *filter*, *wrapper* e *embedded* [71]. Estes diferenciam-se pela forma como fazem a seleção. No método de *filter* são utilizadas propriedades estatísticas dos dados de *input*. No segundo método correspondendo ao *wrapper*, é utilizado um classificador para a escolha, em outros termos, são classificadas as melhores caraterísticas por intermédio de um determinado algoritmo de classificação. Finalmente, o terceiro e último, o método *embedded*, faz uma fusão entre os dois primeiros métodos [72].

Tendo como base estes fundamentos foi efetuada uma seleção das caraterísticas descritas

na Secção anterior, utilizando o método *filter*. Inicialmente, foi testado o algoritmo de *Sequential Feature Selector*, porém, com este requer algum poder computacional, aumentando assim o tempo de execução. Desta forma, foi escolhida a função *SelectKBest* da *Sklearn*. Assim, como o nome sugere, são selecionadas  $k$  número de caraterísticas com melhor pontuação. Para o cálculo de pontuação é utilizado uma função (*scoring function*). Para esta função foi utilizada a *ANOVA*.

### 6.1.3 Classificadores

A etapa dos classificadores, por outras palavras, a classificação é a última dentro do processo já referido. Nesta fase, foram considerados 5 diferentes algoritmos como classificadores: MLPC, dois SVM, *Decision Tree* e *AdaBoost*.

No total, foram submetidos a treino 16363 segmentos, cada um de 1 minuto. Junto com os registos, foram também submetidas as anotações de extraídas no pré-processamento. É de salientar que, neste contexto, se trata de uma classificação binária, ou seja, existem apenas dois valores possíveis de *output*, Apneia e Não-apneia, que se podem traduzir em duas categorias.

#### 6.1.3.1 Support-vector machine (SVM)

*Support-vector machine* foram utilizados dois tipos de *kernel* diferentes: Linear e *Radial Basis Function* (RBF). Mais concretamente, para ambas as variantes foi utilizado o *C-Support Vector Classification* que suporta o parâmetro  $C$ . Este parâmetro foi introduzido por Cortes e Vapnik em 1995, com o objetivo de combater o problema da *SVM standard*, onde a margem pode fazer maus classificações. O parâmetro  $C$  cria uma *soft margin*, permitindo ignorar alguns valores, melhorando o *fit* do classificador [73]. O valor  $C$  utilizado foi de 1 assim como utilizado em [42].

#### 6.1.3.2 Multi-layer Perceptron Classifier (MLPC)

O classificador *MLP* é uma *Neural network* (NN) que mapeia dados de entrada fornecidos para um *output* esperado [74]. Os parâmetros utilizados foram baseados no estudo conduzido por Wang *et al.* [75], onde foi utilizado um *hidden layer* com 69 neurónios calculados através de:  $NF * 2 + 1$ , onde  $NF$  é o número de caraterísticas. A função de ativação utilizada foi a *ru*. Na Figura 6.5 representa a estrutura genérica da NN.

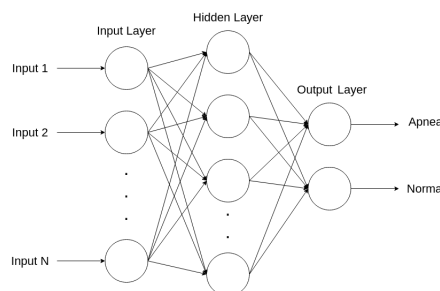


Fig. 6.4: Estrutura do um *hidden layer* MLPC.

## 6.2 Resultados

À semelhança da Literatura revista, os resultados foram medidos através da *accuracy*, *specificity* e *sensibility*, com recurso às equações já apresentadas na Secção 3.3.5.

Fazendo uma tradução das variáveis das referidas equações para o contexto dos classificadores, a variável *True Positive* refere-se à classificação de um segmento como "Apneia" e na variável *True Negative* como segmento "Normal". A classificação como segmento "Normal" ou "Apneia" é feita por cada um dos classificadores escolhidos.

Nesta etapa de resultados, que corresponde diretamente ao teste do modelo proposto, foi utilizado um *subset* da base de dados da *Apnea-ECG* da *Physionet*. Este *subset* tem 35 registos, que foram pré-processados e dos quais foram extraídas as respetivas características em conjunto com os dados de treino. Deste processo resultaram 16733 segmentos de 1 minuto para teste.

### 6.2.1 Normalização

Inicialmente e para efeitos de comparação, foram submetidas a treino a totalidade das características extraídas, ou seja, 34 não normalizadas e 34 normalizadas. Na Tabela 6.3 encontra-se os resultados obtidos.

Table 6.3: Resultados obtidos não normalizados vs normalizados. (N): Normalizados.

Classificador		Resultados		
Nº Features		ACC(%)	SENS(%)	SPEC(%)
34	SVM-Linear	71,29	52,17	83,05
	SVM-RBF	76,57	61,71	85,72
	MLPC	69,30	<b>89,15</b>	57,08
	Adaboost	76,74	67,58	82,37
	RandomForest	<b>77,78</b>	63,83	<b>86,37</b>
34 (N)	SVM-Linear	76,72	67,53	84,83
	SVM-RBF	<b>78,44</b>	67,09	85,43
	MLPC	78,31	<b>68,36</b>	84,43
	Adaboost	76,47	67,58	82,37
	RandomForest	77,79	63,83	<b>86,38</b>

Para a *accuracy*, o melhor resultado, de **78,44%**, foi obtido pelo *SVM-RBF* com a normalização. Na *sensibility*, foi atingido **89,15%** com o *MLPC*, neste caso sem normalização. A especificação não obteve resultados substanciais entre as características com e sem normalização. No entanto, em ambos os casos o *RandomForest* foi o que teve melhor resultado com aproximadamente **86,4%**. Contudo, o mesmo classificador, no mesmo ambiente apresentou valores mais baixos de *accuracy* e *specificity*.

Na generalidade, a normalização dos dados resultou numa melhoria dos resultados, principalmente nos algoritmos de SVM. Contrariamente, a normalização não surtiu tanto efeito nos algoritmos de *AdaBoost* e *RandomForest*, que aproximadamente mantiveram os seus resultados. Diante destes resultados, os restantes resultados apresentados, nomeadamente na Secção seguinte, são com dados normalizados.

### 6.2.2 Resultados por número de características

Assim como aconteceu no teste da normalização, primeiramente foram submetidas a testes todas as características a fim de perceber a performance dos classificadores.

Com recurso ao método de seleção apresentado na Secção 6.1.2.2, foram submetidos a testes diferentes números de características para os diferentes classificadores escolhidos. Os resultados encontram-se na Tabela 6.4. O melhor valor de *accuracy*, de 79,56%, foi encontrado com 22 características com o classificador MLPC. Para *sensibility* o melhor resultado foi de 70,26% pelo mesmo classificador, com 33 características. Por último, na *specificity* o melhor valor foi de 94,03% alcançado pela *SVM-Linear* com 2 características, porém, apesar deste resultado para os restantes dois valores de *accuracy* e *sensibility*, o mesmo classificador apresentou os valores mais baixos de toda a fase de teste.

Destacam-se os resultados do *Adaboost* para a *sensibility* que apesar de não ter obtido o melhor resultado, obteve os melhores valores para 30 características e para os sucessivos números decrescentes de características.

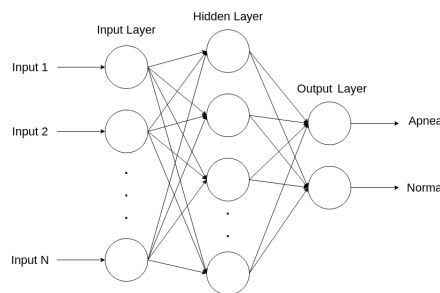


Fig. 6.5: Estrutura de um *hidden layer* MLPC.

## 6.3 Discussão

Perante os resultados apresentados anteriormente, verifica-se que os diferentes classificadores tiveram valores similares, à exceção de alguns valores. A título de exemplo, o SVM-Linear, foi o que apresentou valores mais dispersos, o que resultou, maioritariamente, em resultados piores que os restantes classificadores. Isto pode ser verificado na Figura 6.6, onde é visível que o classificador em questão, tem o piores valores de *accuracy* (*ACC*) e *sensibility* (*SENS*) mesmo que tenha alcançado 94,03% em *specificity* (*SPEC*).

O MLPC foi o classificador que teve melhor *accuracy* com 22 características, assim como já foi referenciado. Para o mesmo número de características, o classificador não teve os melhores valores de *sensibility* e especificação, contudo, não se distanciou muito dos valores de topo.

Em contraste, o mesmo classificador com *kernel RBF*, apesar de não ter sido o que atingiu o melhor valor, foi o que obteve melhor valor de *accuracy* em termos médios, com cerca de 77,6%, tal como consta na Figura 6.7.

Em relação ao *AdaBoost*, em relação a *sensibility* foi o que teve um comportamento mais constante, conseguindo 67% na média. Ainda assim, foi o que apresentou o pior resultado médio em *specificity*.

Table 6.4: Resultados obtidos.

N° Features	Classificador	Resultados		
		ACC(%)	SENS(%)	SPEC(%)
34	SVM-Linear	76,72	67,53	84,83
	SVM-RBF	<b>78,44</b>	67,09	85,43
	MLPC	78,31	<b>68,36</b>	84,43
	Adaboost	76,47	67,58	82,37
	RandomForest	77,79	63,83	<b>86,38</b>
33	SVM-Linear	76,44	63,31	84,51
	SVM-RBF	<b>78,52</b>	66,81	85,72
	MLPC	78,45	<b>70,26</b>	83,49
	Adaboost	76,65	67,47	82,30
	RandomForest	77,63	64,36	<b>85,79</b>
30	SVM-Linear	76,40	61,54	85,50
	SVM-RBF	<b>78,25</b>	66,60	85,41
	MLPC	78,09	67,34	84,70
	Adaboost	77,25	<b>67,88</b>	83,01
	RandomForest	77,41	63,42	<b>86,02</b>
25	SVM-Linear	76,48	58,95	<b>87,27</b>
	SVM-RBF	<b>78,24</b>	66,43	85,51
	MLPC	77,80	63,87	86,37
	Adaboost	76,94	<b>67,81</b>	82,56
	RandomForest	77,38	63,12	86,14
22	SVM-Linear	74,27	56,37	85,27
	SVM-RBF	78,18	64,96	86,31
	MLPC	<b>79,56</b>	66,79	85,45
	Adaboost	76,58	<b>66,86</b>	82,56
	RandomForest	77,13	62,21	<b>86,91</b>
18	SVM-Linear	74,32	55,81	85,71
	SVM-RBF	<b>78,11</b>	64,05	<b>86,76</b>
	MLPC	78,03	64,94	86,08
	Adaboost	76,25	<b>66,37</b>	82,32
	RandomForest	76,94	62,82	85,63
15	SVM-Linear	72,93	51,95	85,83
	SVM-RBF	<b>77,40</b>	61,11	<b>87,42</b>
	MLPC	77,26	65,44	84,72
	Adaboost	76,74	<b>66,59</b>	82,99
	RandomForest	76,54	61,58	85,75
10	SVM-Linear	73,36	52,23	<b>86,36</b>
	SVM-RBF	76,31	60,22	86,21
	MLPC	<b>76,50</b>	61,94	85,46
	Adaboost	76,17	<b>66,65</b>	82,02
	RandomForest	74,98	61,57	83,24
5	SVM-Linear	73,54	46,60	<b>90,11</b>
	SVM-RBF	<b>76,81</b>	62,49	85,61
	MLPC	76,42	62,90	84,74
	Adaboost	76,31	<b>66,40</b>	82,41
	RandomForest	74,80	63,03	82,04
2	SVM-Linear	70,84	33,15	<b>94,03</b>
	SVM-RBF	75,74	58,26	86,49
	MLPC	75,56	59,57	85,40
	Adaboost	<b>75,93</b>	<b>66,51</b>	81,72
	RandomForest	72,06	60,22	79,35

Por fim, o *RandomForest* aproximou-se dos restantes classificadores e apresentou alguns valores de topo para *specificity* com determinados números de características. Em termos

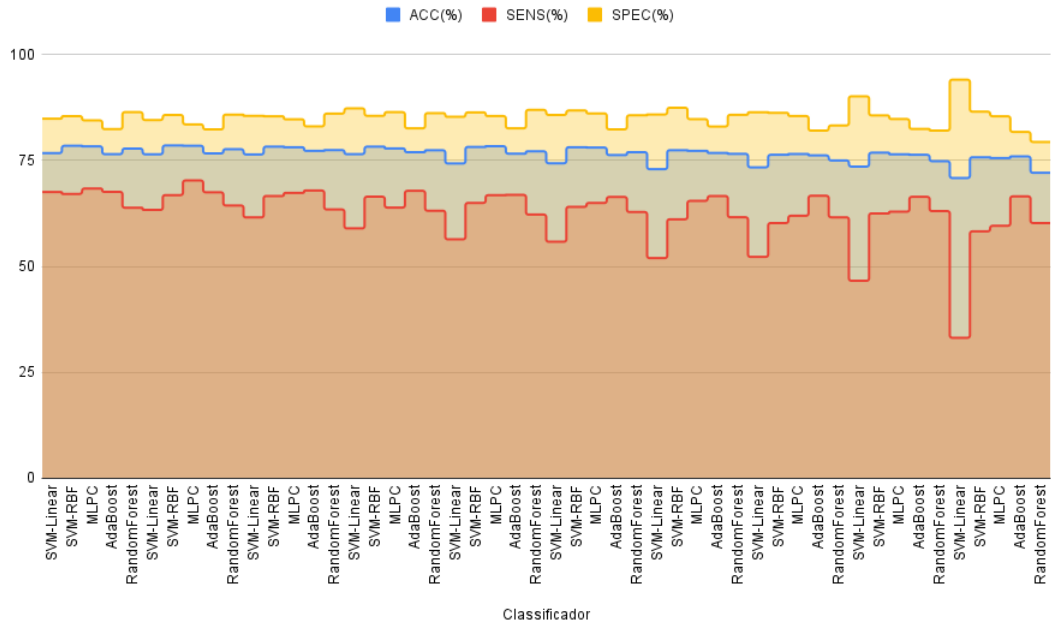


Fig. 6.6: Distribuição dos resultados ao longo da seleção de características.

médios, nenhum dos seus valores (*ACC*, *SENS* ou *SPEC*) se destacou.

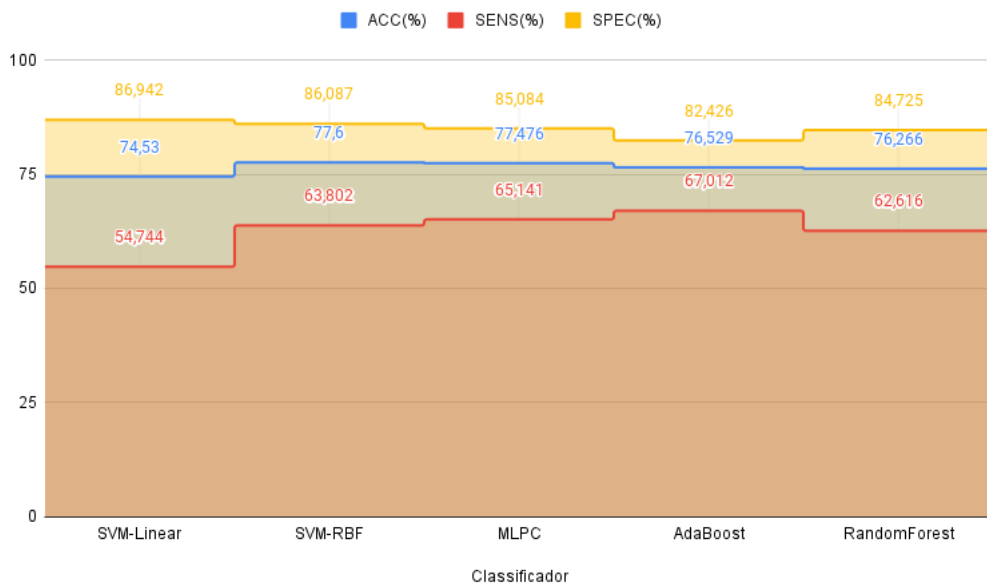


Fig. 6.7: Média dos valores de *ACC*, *SENS* e *SPEC* para os diferentes classificadores.

Em comparação, o melhor classificador encontrado na Literatura foi o MLP em [39], com uma *accuracy* de 98,09%. Apesar disso, este estudo não só fez uso de sinal ECG, como também utilizou sinais de EEG e eletromiograma (EMG). Apenas utilizando como *input* ECG, o mesmo estudo obteve 86,43% de *accuracy*.

Assim, o melhor resultado utilizando apenas ECG foi alcançado pelo estudo [55] com 94,7% como recurso *Boosted Tree* como classificador. Os restantes valores não são referi-

dos pelo estudo.

Para o classificador *SVM*, o mais utilizado pelos estudos encontrados, os melhores valores foram encontrados pelo estudo [43], com 93,14%, 94,39% e de 90,81% de *ACC*, *SENS* e *SPEC*, respetivamente. Porém este não foi o seu melhor classificador, sendo neste caso o *GentleBoost* com 93,26% em *accuracy*. O segundo melhor resultado foi conseguido pelo estudo [47], com 92,59%.

De todos os estudos, o [57] é o que apresenta resultados mais relevantes e o que utiliza apenas ECG como *input*. Este alcançou, aproximadamente, 94%, 95% e 93% para *ACC*, *SENS* e *SPEC*, respetivamente. Todavia, o estudo utilizou uma *Residual NN*, que pertence à esfera da DL, campo que não foi explorado no modelo desenvolvido.

É de notar que, a maioria dos autores revelaram resultados mais significativos que os apresentados pelo modelo proposto, no entanto, não significa que o modelo proposto não consiga fazer deteção de episódios de apneia. Alguns dos autores utilizaram uma abordagem em DL, isto pode-se traduzir num fator para a melhor performance nesses modelos. Não obstante, o modelo proposto apresentou alguns resultados significativos, mais concretamente na *specificity*, onde foram atingidos valores superiores a 85%.

A seleção de características, foi um etapa que não revelou grandes benefícios, sendo também por vezes uma etapa não mencionada ou considerada pelos autores. Contrariamente, a normalização das características revelou-se benéfica, melhorando substancialmente os resultados.

**Nota:** As comparações feitas acima com os estudos explorados na SLR, são apenas referentes aos resultados por segmento, visto que não foi feito um teste por registo.

## 6.4 Conclusão

O modelo proposto visa desenvolver uma alternativa à PSG, exame *golden standard* para o diagnóstico AOS, com o objetivo principal de a descomplicar. Nesse sentido, apenas foi utilizado um tipo de *input*, o sinal ECG. Através da base de dados *ApneaECG*, foram extraídos dados ECG e anotações dos minutos de apneia. Seguiu-se a construção de um *workflow*, semelhante aos autores encontrados na Literatura. O resultado foi o referido modelo de classificação de eventos apneia.

É possível concluir que o modelo construído evidencia que o sinal ECG é uma alternativa viável para o diagnóstico à apneia. Ainda é possível concluir que o este modelo consegue detetar eventos de apneia com uma *accuracy* de 78,45%, *sensibility* de 70,26% e *specificity* de 83,49%. Estes resultados foram conseguidos com o classificador MLPC e com 32 das 33 características extraídas.

Poderão ainda existir características, não extraídas no desenvolvimento e uma melhor utilização da seleção de características, que possam melhorar o desempenho do modelo.

# Capítulo 7

## Conclusão e Trabalho Futuro

Neste capítulo serão apresentadas as conclusões alcançadas e o trabalho futuro que poderá ter seguimento do presente trabalho.

### 7.1 Conclusões Principais

Através da *Systematic Literature Review*, foi possível perceber que os avanços na construção de soluções para o diagnóstico à AOS nos últimos 10 anos têm sido notáveis. Existem, cada vez mais, modelos capazes de ajudar médicos e utentes a melhorar os serviços de saúde, neste caso concreto para a doença da apneia. É também de salientar que a *Deep Learning* é uma área emergente neste contexto e onde foram encontrados resultados bastante significativos, concluindo-se assim que a exploração desta área para a construção de novos modelos poderá ser frutuoso.

A criação da aplicação móvel, utilizando o dispositivo MMDK, como o propósito de recolha de dados ECG é também um passo no progresso para contornar os problemas que envolvem a PSG. Isto, na medida em que, os dados recolhidos poderão testar o modelo proposto e outros em ambiente real, e logo à partida ajudar equipas médicas através desses resultados.

Finalmente, o modelo/ algoritmo de deteção da apneia proposto vem solidificar que a utilização de ECG é uma alternativa viável e que pode vir a ser utilizada com meio de diagnóstico primário. O modelo proposto obteve um valores de *accuracy* de 78,45%, *sensibility* de 70,26% e *specificity* de 83,49% como o uso do classificador MLPC, como o uso de 33 caraterísticas. A escolha das caraterísticas extraídas são resultados da análise feita à Literatura e as bibliotecas desenvolvidas pela comunidade em *Python*. Isto pode significar que ainda existiam caraterísticas não extraídas que melhorem os resultados do modelo. A título pessoal, o aprofundamento do conhecimento nas áreas da *Machine Learning*, na informática médica e um tanto na área da medicina, foi uma das maiores valências ao longo do desenvolvimento da dissertação.

Desta dissertação, resulta uma *paper*, ainda em desenvolvimento, com o nome " *ECG-based Detection of Obstructive Sleep Apnea using Intelligent Systems: A Systematic Literature Review*", que posteriormente será submetido a conferências.

### 7.2 Trabalho Futuro

Assim como já foi mencionado, o trabalho desenvolvido ao longo desta dissertação é um passo para encontrar soluções que melhorem o diagnóstico a AOS, no entanto, não é o passo final.

Nesse sentido, o melhoramento do modelo através da extração de novas características, utilização de novos classificadores com diferentes parâmetros, o teste de novos parâmetros nos classificadores já utilizados e a análise por registo, são alguns dos trabalhos que podem melhorar o seu desempenho.

Adicionalmente, a implementação de um sistema *cloud* com o modelo e a sucessiva integração da aplicação móvel poderão ajudar a criar um serviço de ajuda ao diagnóstico. Na Figura 7.1 segue uma proposta da arquitetura do modelo que poderá ser desenvolvido a partir daqui.

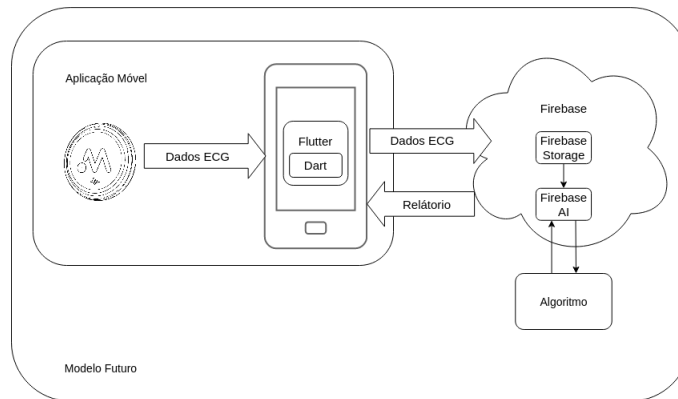


Fig. 7.1: Arquitetura Modelo Futuro.

Por último, a criação de uma base de dados com os futuros dados recolhidos a partir da aplicação e dispositivo, poderá adicionar à comunidade científica uma base de testes adicional à *ApneaECG*.

# Bibliografia

- [1] Ed Burns. (2021) U wave. [Online]. Available: <https://litfl.com/u-wave-ecg-library/>
- [2] Movesense. (2020) Movesense medical spec sheet. Create new solutions for health monitoring and remote healthcare. [Online]. Available: <https://www.movesense.com/movesense-medical/>
- [3] Movesense. *Movesense System Overview*. [Online]. Available: [https://www.movesense.com/docs/system/system\\_overview/](https://www.movesense.com/docs/system/system_overview/)
- [4] Oura Team. (2020) What is hrv? [Online]. Available: <https://ouraring.com/blog/what-is-heart-rate-variability/>
- [5] G. J. Joyia, R. M. Liaqat, A. Farooq, and S. Rehman, "Internet of medical things (iomt): applications, benefits and future challenges in healthcare domain," *J Commun*, vol. 12, no. 4, pp. 240–247, 2017.
- [6] B. Dong, J. Yang, Y. Ma, and X. Zhang, "Medical monitoring model of internet of things based on the adaptive threshold difference algorithm," *International Journal of Multimedia and Ubiquitous Engineering*, vol. 11, no. 5, pp. 75–82, 2016.
- [7] F. Andriopoulou, T. Dagiuklas, and T. Orphanoudakis, "Integrating iot and fog computing for healthcare service delivery," in *Components and services for IoT platforms*. Springer, 2017, pp. 213–232.
- [8] G. J. Joyia, R. M. Liaqat, A. Farooq, and S. Rehman, "Internet of medical things (iomt): applications, benefits and future challenges in healthcare domain," *J Commun*, vol. 12, no. 4, pp. 240–247, 2017.
- [9] Z. Rayan, M. Alfonse, and A.-B. M. Salem, "Machine learning approaches in smart health," *Procedia Computer Science*, vol. 154, pp. 361 – 368, 2019, proceedings of the 9th International Conference of Information and Communication Technology [ICICT-2019] Nanning, Guangxi, China January 11-13, 2019. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S187705091930821X>
- [10] J. D. Kelleher, B. Mac Namee, and A. D'arcy, *Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies*. MIT press, 2020.
- [11] A. Benjafield, K. Valentine, N. Ayas, P. Eastwood, R. Heinzer, M. Ip, S. Patel, P. Peppard, S. Sinha, S. Tufik *et al.*, "Global prevalence of obstructive sleep apnea in adults: estimation using currently available data," in *B67. Risk and prevalence of sleep disordered breathing*. American Thoracic Society, 2018, pp. A3962–A3962.
- [12] S. Javaheri, F. Barbe, F. Campos-Rodriguez, J. A. Dempsey, R. Khayat, S. Javaheri, A. Malhotra, M. A. Martinez-Garcia, R. Mehra, A. I. Pack *et al.*, "Sleep apnea: types,

- mechanisms, and clinical cardiovascular consequences,” *Journal of the American College of Cardiology*, vol. 69, no. 7, pp. 841–858, 2017.
- [13] P. D. Hung, “Detection of central sleep apnea based on a single-lead eeg,” in *Proceedings of the 2018 5th International Conference on Bioinformatics Research and Applications*, ser. ICBRA ’18. New York, NY, USA: Association for Computing Machinery, 2018, p. 78–83. [Online]. Available: <https://doi.org/10.1145/3309129.3309132>
- [14] N. Pombo, N. Garcia, and K. Bousson, “Classification techniques on computerized systems to predict and/or to detect apnea: A systematic review,” *Computer methods and programs in biomedicine*, vol. 140, pp. 265–274, 2017.
- [15] A. P. Rodrigues, P. Pinto, B. Nunes, and C. Bárbara, “Síndrome de apneia obstrutiva do sono: epidemiologia, diagnóstico e tratamento. um estudo da rede médicos-sentinela,” 2014.
- [16] B. Fatimah, P. Singh, A. Singhal, and R. B. Pachori, “Detection of apnea events from eeg segments using fourier decomposition method,” *Biomedical Signal Processing and Control*, vol. 61, p. 102005, 2020.
- [17] S. Warapongmanupong and A. Preutthipan, “Can standard deviation of overnight pulse oximetry be used to screen childhood obstructive sleep apnea,” *International Journal of Pediatric Otorhinolaryngology*, vol. 119, pp. 27 – 31, 2019. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0165587619300047>
- [18] B. Xie, W. Qiu, H. Minn, L. Tamil, and M. Nourani, “An improved approach for real-time detection of sleep apnea.” in *BIOSIGNALS*, 2011, pp. 169–175.
- [19] T. Sugi, F. Kawana, and M. Nakamura, “Automatic eeg arousal detection for sleep apnea syndrome,” *Biomedical Signal Processing and Control*, vol. 4, no. 4, pp. 329–337, 2009.
- [20] C.-C. Hsu and P.-T. Shih, “A novel sleep apnea detection system in electroencephalogram using frequency variation,” *Expert Systems with Applications*, vol. 38, no. 5, pp. 6014–6024, 2011.
- [21] Ş. Yücelbaş, C. Yücelbaş, G. Tezel, S. Özşen, S. Küçüktürk, and Ş. Yosunkaya, “Pre-determination of OSA degree using morphological features of the ECG signal,” *Expert Systems with Applications*, vol. 81, pp. 79–87, 2017.
- [22] G. Moody, R. Mark, A. Goldberger, and T. Penzel, “Stimulating rapid research advances via focused competition: The computers in cardiology challenge 2000,” in *Computers in Cardiology 2000. Vol. 27 (Cat. 00CH37163)*. IEEE, 2000, pp. 207–210.
- [23] B. Raymond, R. Cayton, R. Bates, and M. Chappell, “Screening for obstructive sleep apnoea based on the electrocardiogram-the computers in cardiology challenge,” in

- Computers in Cardiology 2000. Vol. 27 (Cat. 00CH37163)*. IEEE, 2000, pp. 267–270.
- [24] W. R. Ruehland, P. D. Rochford, F. J. O’Donoghue, R. J. Pierce, P. Singh, and A. T. Thornton, “The new aasm criteria for scoring hypopneas: impact on the apnea hypopnea index,” *sleep*, vol. 32, no. 2, pp. 150–157, 2009.
- [25] I. De Falco, G. De Pietro, and G. Sannino, “On Finding Explicit Rules for Personalized Forecasting of Obstructive Sleep Apnea Episodes,” *Proceedings - 2015 IEEE 16th International Conference on Information Reuse and Integration, IRI 2015*, pp. 326–333, 2015.
- [26] NHS. (2021) Electrocardiogram (ecg). [Online]. Available: <https://www.nhs.uk/conditions/electrocardiogram/>
- [27] J.-L. Vincent, “Understanding cardiac output,” *Critical care*, vol. 12, no. 4, pp. 1–3, 2008.
- [28] S. Viskin, U. Rosovski, A. J. Sands, E. Chen, P. M. Kistler, J. M. Kalman, L. R. Chavez, P. I. Torres, O. A. Centuri3n, A. Fujiki *et al.*, “Inaccurate electrocardiographic interpretation of long qt: the majority of physicians cannot recognize a long qt when they see one,” *Heart Rhythm*, vol. 2, no. 6, pp. 569–574, 2005.
- [29] G. B. Moody, R. G. Mark, M. A. Bump, J. S. Weinstein, A. D. Berman, J. E. Mietus, and A. L. Goldberger, “Clinical validation of the ecg-derived respiration (edr) technique,” *Computers in cardiology*, vol. 13, no. 1, pp. 507–510, 1986.
- [30] J. McNames and A. Fraser, “Obstructive sleep apnea classification based on spectrogram patterns in the electrocardiogram,” in *Computers in Cardiology 2000. Vol. 27 (Cat. 00CH37163)*. IEEE, 2000, pp. 749–752.
- [31] T. Penzel, G. B. Moody, R. G. Mark, A. L. Goldberger, and J. H. Peter, “The apnea-ecg database,” in *Computers in Cardiology 2000. Vol.27 (Cat. 00CH37163)*, Sep. 2000, pp. 255–258.
- [32] A. L. Samuel, “Some studies in machine learning using the game of checkers,” *IBM Journal of Research and Development*, vol. 3, no. 3, pp. 210–229, 1959.
- [33] W. S. McCulloch and W. Pitts, *A Logical Calculus of the Ideas Immanent in Nervous Activity*. Cambridge, MA, USA: MIT Press, 1988, p. 15–27.
- [34] L. G. Serrano, *Exploring Machine Learning Basics*, ser. ISBN: 9781617298127. Manning Publications, 2020. [Online]. Available: <https://www.manning.com/books/exploring-machine-learning-basics?query=machine>
- [35] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and T. P. Group, “Preferred reporting items for systematic reviews and meta-analyses: The prisma statement,” *PLOS Medicine*, vol. 6, no. 7, pp. 1–6, 07 2009. [Online]. Available: <https://doi.org/10.1371/journal.pmed.1000097>

- [36] M. Ouzzani, H. Hammady, Z. Fedorowicz, and A. Elmagarmid, “Rayyan—a web and mobile app for systematic reviews,” *Systematic Reviews*, vol. 5, no. 1, p. 210, 2016. [Online]. Available: <http://dx.doi.org/10.1186/s13643-016-0384-4>
- [37] G. Sannino, I. De Falco, and G. De Pietro, “Monitoring Obstructive Sleep Apnea by means of a real-time mobile system based on the automatic extraction of sets of rules through Differential Evolution,” *Journal of Biomedical Informatics*, vol. 49, pp. 84–100, 2014. [Online]. Available: <http://dx.doi.org/10.1016/j.jbi.2014.02.015>
- [38] G. Sannino, I. De Falco, and G. De Pietro, “An automatic rules extraction approach to support osa events detection in an mhealth system,” *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 5, pp. 1518–1524, 2014.
- [39] M. K. Moridani, M. Heydar, S. Sina, and J. Behnam, “A Reliable Algorithm Based on Combination of EMG, ECG and EEG Signals for Sleep Apnea Detection,” *2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI)*, pp. 256–262, 2019.
- [40] J. H. Lee, H. J. Park, and Y. N. Kim, “Monitoring obstructive sleep apnea with electrocardiography and 3-axis acceleration sensor,” *8th ACM International Conference on Pervasive Technologies Related to Assistive Environments, PETRA 2015 - Proceedings*, 2015.
- [41] K. Li, W. Pan, Y. Li, Q. Jiang, and G. Liu, “A method to detect sleep apnea based on deep neural network and hidden Markov model using single-lead ECG signal,” *Neurocomputing*, vol. 294, pp. 94–101, 2018. [Online]. Available: <https://doi.org/10.1016/j.neucom.2018.03.011>
- [42] A. Yildiz, M. Akin, and M. Poyraz, “An expert system for automated recognition of patients with obstructive sleep apnea using electrocardiogram recordings,” *Expert Systems with Applications*, vol. 38, no. 10, pp. 12 880–12 890, 2011. [Online]. Available: <http://dx.doi.org/10.1016/j.eswa.2011.04.080>
- [43] A. Zarei and B. M. Asl, “Automatic classification of apnea and normal subjects using new features extracted from HRV and ECG-derived respiration signals,” *Biomedical Signal Processing and Control*, vol. 59, p. 101927, 2020. [Online]. Available: <https://doi.org/10.1016/j.bspc.2020.101927>
- [44] A. Zarei and B. Mohammadzadeh Asl, “Performance evaluation of the spectral auto-correlation function and autoregressive models for automated sleep apnea detection using single-lead ECG signal,” *Computer Methods and Programs in Biomedicine*, vol. 195, 2020.
- [45] M. Sharma, M. Raval, and U. R. Acharya, “A new approach to identify obstructive sleep apnea using an optimal orthogonal wavelet filter bank with ECG signals,” *Informatics in Medicine Unlocked*, vol. 16, no. February, p. 100170, 2019. [Online]. Available: <https://doi.org/10.1016/j.imu.2019.100170>

- [46] A. R. Hassan and M. A. Haque, "Computer-aided obstructive sleep apnea screening from single-lead electrocardiogram using statistical and spectral features and bootstrap aggregating," *Biocybernetics and Biomedical Engineering*, vol. 36, no. 1, pp. 256–266, 2016. [Online]. Available: <http://dx.doi.org/10.1016/j.bbe.2015.11.003>
- [47] B. Fatimah, P. Singh, A. Singhal, and R. B. Pachori, "Detection of apnea events from ECG segments using Fourier decomposition method," *Biomedical Signal Processing and Control*, vol. 61, p. 102005, 2020. [Online]. Available: <https://doi.org/10.1016/j.bspc.2020.102005>
- [48] P. Janbakhshi and M. B. Shamsollahi, "Sleep Apnea Detection from Single-Lead ECG Using Features Based on ECG-Derived Respiration (EDR) Signals," *Irbm*, vol. 39, no. 3, pp. 206–218, 2018. [Online]. Available: <https://doi.org/10.1016/j.irbm.2018.03.002>
- [49] M. Sharma, S. Agarwal, and U. R. Acharya, "Application of an optimal class of antisymmetric wavelet filter banks for obstructive sleep apnea diagnosis using ECG signals," *Computers in Biology and Medicine*, vol. 100, no. June, pp. 100–113, 2018. [Online]. Available: <https://doi.org/10.1016/j.compbimed.2018.06.011>
- [50] H. Singh, R. K. Tripathy, and R. B. Pachori, "Detection of sleep apnea from heart beat interval and ECG derived respiration signals using sliding mode singular spectrum analysis," *Digital Signal Processing: A Review Journal*, vol. 104, p. 102796, 2020. [Online]. Available: <https://doi.org/10.1016/j.dsp.2020.102796>
- [51] S. A. Singh and S. Majumder, "a Novel Approach Osa Detection Using Single-Lead Ecg Scalogram Based on Deep Neural Network," *Journal of Mechanics in Medicine and Biology*, vol. 19, no. 4, pp. 1–18, 2019.
- [52] L. Cai, R. Datta, J. Huang, S. Dong, and M. Du, "Sleep Disorder Data Stream Classification Based on Classifiers Ensemble and Active Learning," *Proceedings - 2019 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2019*, pp. 1432–1435, 2019.
- [53] I. Delibaşoğlu, C. Avcı, and A. Akbaş, "ECG based sleep apnea detection using wavelet analysis of instantaneous heart rates," *ACM International Conference Proceeding Series*, pp. 1–5, 2011.
- [54] C. Shi, M. Nourani, G. Gupta, and L. Tamil, "Demo: A virtual sleep laboratory," *mHealthSys 2011 - Proceedings of the 1st ACM Workshop on Mobile Systems, Applications, and Services for HealthCare - Co-held with ACM SenSys 2011*, 2011.
- [55] I. Fahrudi, I. K. Eddy Purnama, H. Takahashi, and M. H. Purnomo, "Classification of Sleep Disorder from Single Lead Non-overlapping of ECG-apnea based Non-Linear Analysis using Ensemble Approach," *2019 IEEE 10th International Conference on Awareness Science and Technology, iCAST 2019 - Proceedings*, 2019.

- [56] C. S. Viswabhargav, R. K. Tripathy, and U. R. Acharya, “Automated detection of sleep apnea using sparse residual entropy features with various dictionaries extracted from heart rate and EDR signals,” *Computers in Biology and Medicine*, vol. 108, no. March, pp. 20–30, 2019. [Online]. Available: <https://doi.org/10.1016/j.compbimed.2019.03.016>
- [57] L. Wang, Y. Lin, and J. Wang, “A RR interval based automated apnea detection approach using residual network,” *Computer Methods and Programs in Biomedicine*, vol. 176, pp. 93–104, 2019. [Online]. Available: <https://doi.org/10.1016/j.cmpb.2019.05.002>
- [58] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, “Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals,” *circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [59] Y. Ichimaru and G. Moody, “Development of the polysomnographic database on cd-rom,” *Psychiatry and clinical neurosciences*, vol. 53, no. 2, pp. 175–177, 1999.
- [60] N. Iyengar, C. Peng, R. Morin, A. L. Goldberger, and L. A. Lipsitz, “Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics,” *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 271, no. 4, pp. R1078–R1084, 1996.
- [61] D. Oak, “Jupyter kernels,” 2019, [Online] <https://github.com/jupyter/jupyter/wiki/Jupyter-kernels>. Último acesso a 29 de dezembro de 2021.
- [62] Tiobe. (2021) Tiobe index for november 2021. [Online]. Available: <https://www.tiobe.com/tiobe-index/>
- [63] Pickle. (2021) pickle — serialização de objetos python. [Online]. Available: <https://docs.python.org/pt-br/3/library/pickle.html>
- [64] Movesense, *MOVESENSE MEDICAL SENSOR*, 2020.
- [65] P. Hamilton, “Open source ecg analysis,” in *Computers in Cardiology*, 2002, pp. 101–104.
- [66] R. McCraty and F. Shaffer, “Heart rate variability: new perspectives on physiological mechanisms, assessment of self-regulatory capacity, and health risk,” *Global advances in health and medicine*, vol. 4, no. 1, pp. 46–61, 2015.
- [67] F. Shaffer and J. P. Ginsberg, “An overview of heart rate variability metrics and norms,” *Frontiers in public health*, p. 258, 2017.
- [68] S. Sarkar, S. Bhattacharjee, and S. Pal, “Extraction of respiration signal from ecg for respiratory rate estimation,” 2015.

- [69] Robin Champseix. (2010) Welcome to hrv-analysis documentation! [Online]. Available: <https://aura-healthcare.github.io/hrv-analysis/index.html>
- [70] R. Champseix, L. Ribiere, and C. Le Couedic, “A python package for heart rate variability analysis and signal preprocessing,” *Journal of Open Research Software*, vol. 9, no. 1, 2021.
- [71] H. Liu and H. Motoda, *Computational methods of feature selection*. CRC Press, 2007.
- [72] M. d. S. OLIVEIRA, “Metodologia de seleção de features não supervisionada para clustering em conjunto de dados de alta dimensionalidade,” Master’s thesis, Universidade Federal de Pernambuco, 2018.
- [73] A. M. d. S. Pinho, “Classification models for sleep apnea detection,” Ph.D. dissertation.
- [74] Y. Zhang, Y. Sun, P. Phillips, G. Liu, X. Zhou, and S. Wang, “A multilayer perceptron based smart pathological brain detection system by fractional fourier entropy,” *Journal of medical systems*, vol. 40, no. 7, pp. 1–11, 2016.
- [75] T. Wang, C. Lu, and G. Shen, “Detection of sleep apnea from single-lead ecg signal using a time window artificial neural network,” *BioMed research international*, vol. 2019, 2019.



# **Appendix A**

## **Protocolo de recolha de Dados ECG**



## **Recolha de dados ECG de indivíduos que apresentem sintomatologia associada à apneia obstrutiva do sono**

### **Acrónimos:**

- Índice de Apneia-Hipopneia (IAH)
- Apneia Obstrutiva do Sono (AOS)
- Eletrocardiograma (ECG)
- ECG-Derived Respiration (EDR)
- Polissonografia (PSG)
- Síndrome da Apneia do Sono (SAS)
- Universidade da Beira Interior (UBI)

### **Equipa responsável:**

Lic. Eduardo Paulos, Estudante de Mestrado, UBI

(Contacto: 964245718, [edupaulos1@gmail.com](mailto:edupaulos1@gmail.com))

Prof. Nuno Pombo, PhD, Universidade da Beira Interior (UBI)

(Contactos: 965683580, [ngpombo@di.ubi.pt](mailto:ngpombo@di.ubi.pt))

Lic. Rui Charrinho, Estudante de Mestrado, UBI

Prof. Sandeep Pirbhulal, PhD, Norwegian University of Science and Technology (Top 2% mundial em investigação)

Dr<sup>a</sup> Filipa Jesus, MsC, Serviço de Pneumologia da U.L.S da Guarda

### **1) Introdução**

Este documento apresenta o protocolo experimental para a recolha de dados eletrocardiogramas (ECG) de indivíduos que tenham sido diagnosticados com Apneia Obstrutiva do Sono (AOS) e que manifestem os sintomas associados à doença, ou que tenham uma grande probabilidade de

vir a ser diagnosticado. O documento está organizado da seguinte forma: 1) Introdução, 2) Background e Relevância, 3) Objectivo(s) Geral(ais) e Específico(s), 4) Metodologia, 5) Aspectos Éticos e Confidencialidade dos Dados, 6) Bibliografia e Anexos.

Este documento tem como objectivo fundamentar o pedido de aprovação do protocolo experimental de recolha de dados médicos pela Comissão de Ética da Unidade de Saúde Local (ULS) da Guarda.

Este procedimento experimental enquadra-se nos trabalhos no âmbito da dissertação de mestrado dos alunos Eduardo Frango Paulos e Rui Pedro Belona Charrinho, da Universidade da Beira Interior, cujo orientador e representante da UBI é o Prof. Doutor Nuno Pombo, em coorientação com o Prof. Doutor Sandeep Pirbhulal. Por parte do Serviço de Pneumologia da U.L.S da Guarda, a Mestre Filipa Jesus será a responsável pelo acompanhamento do processo de recolha.

## **2) Background e Relevância**

A Síndrome da Apneia do Sono (SAS) é a doença mais comum ligada ao sono, afeta quase 1 bilhão de pessoas no mundo, segundo a *International Sleep Experts*. Mais especificamente, caracteriza-se pela paragem na respiração durante o período de sono. Esta paragem pode-se tratar de uma cessação total (apneia) ou parcial (hipopneia) (Pombo, 2017). Dependendo da etiologia da doença, distinguem-se 3 tipos de SAS: a obstrutiva, caracteriza-se pela obstrução parcial ou total do fluxo do canal aéreo; a central, que corresponde a uma descoordenação da comunicação entre o cérebro e o sistema respiratório (Phan Duy Hung, 2018); por fim a mista, que combina os sintomas tanto da obstrutiva como da central (Pombo, 2017). A recolha de dados terá focado indivíduos com na AOS.

Através do Índice de Apneia-Hipopneia (IAH), são distinguidos vários patamares da doença. Nesta qualificação são contados o número de episódios de apneia por hora. Na tabela 1 encontram-se os diferentes grupos do IAH.

Tabela 1 - Classificação de severidade através do IAH

<b>Episódios/hora</b>	<b>Grupo</b>
<5	Mínima ou inexistente
[5,15]	Ligeiro
[15,30]	Moderado
>30	Grave

Como se trata de uma doença ligada ao sono, muitas vezes é difícil para o paciente aperceber-se dos sintomas associados. Isto leva a que muitas pessoas não sejam diagnosticadas e por consequência não sejam tratadas, levando a que a doença seja subdiagnosticada. A curto prazo, um “não diagnóstico” pode levar a tensão arterial alta, batimento cardíaco acelerado, hipoxemia, e a longo termo problemas vasculares cerebrais e cardíacos, comprometimento do sistema imunitário, fadiga extrema, fraca concentração, tempos de reação baixos e em casos extremos até a própria morte (De Falco, 2015).

Como exame para o diagnóstico é comumente usada a Polissonografia (PSG). Trata-se de um exame que avalia o sono de um paciente através de vários parâmetros como monitorização da atividade cerebral, atividade muscular e o movimento dos olhos (Fatimah, 2020). Desta forma, o exame requer vários sensores que irão estar ligados ao paciente durante todo o seu período de sono, tornando-se incomodativo para o paciente. Adicionalmente, é necessário pessoal dedicado que monitorize continuamente a evolução do exame identificando e anotando episódios de apneia (Fatimah, 2020). Por consequência, este exame só pode ser realizado em clínicas especializadas, como hospitais ou clínicas de sono (Phan Duy Hung, 2018). Contudo, por vezes, de modo a reduzir os custos certos centros realizam um teste de diagnóstico mais

simplista que diminui os custos. Por consequência, diminui também a exatidão do exame, dificultando o diagnóstico. Todos esses fatores tornam a PSG e consequentemente o diagnóstico da Apneia caro. Neste sentido é necessário desenvolver métodos eficazes que ajudem na redução deste custo e que tornem o diagnóstico mais acessível.

Como alternativa a PSG e de modo a automatizar a deteção e monitorização da apneia, têm surgidas várias propostas que englobam a monitorização do oxigênio no sangue (SpO<sub>2</sub>) (Xie, 2011; Warapongmanupong, 2019), eletrocardiograma (ECG), Eletroencefalograma (EEG), entre outros tipos de *inputs* de dados.

O ECG é uma das alternativas mais utilizadas pelas suas características morfológicas (Yücelbaş, 2017) e por se tratar de um método não invasivo. Um *milestone* para a sua utilização para a deteção da AOS foi o *Challenge* em 2000 (Goldberger, 2000), no qual foi disponibilizada a Base de dados *Apnea-ECG*. Um dos pioneiros na sua utilização foi *Raymond et al.*, extraíndo o ECG-Derived Respiration (EDR) e o intervalo entre picos R do sinal ECG (Raymond, 2000).

Perante o descrito anteriormente, pretende-se a recolha de dados ECG de indivíduos com AOS para a criação de uma base de dados, que servirá posteriormente para o teste de algoritmos de deteção e/ou predição de episódios de apneia. Em última instância pertence facilitar o diagnóstico da AOS com a ajuda desta base de dados, criando alternativas à PSG.

### **3) Objetivo(s) geral(ais) e específico(s)**

O objetivo geral deste estudo experimental é a criação de uma base de dados de eletrocardiogramas (ECG) de indivíduos diagnosticados para a AOS.

A testagem do algoritmo em desenvolvimento por outras bases de dados que não a mencionada anteriormente, trará mais viabilidade aos algoritmos. Neste sentido, a criação de uma base de dados para a realização dos testes é também um dos objetivos deste protocolo.

É também objetivo deste protocolo a disponibilização de acesso livre e gratuito para a comunidade científica internacional, nos servidores da UBI, em <http://allab.it.ubi.pt/mediawiki>. Compreende-se como objetivos específicos a aquisição dos dados, sua validação, categorização e publicação.

#### 4) Metodologia

Tipo de estudo: O estudo é do tipo transversal.

Descrição e caracterização da amostra experimental: A amostra experimental será constituída por 4 grupos: Grupo I (controlo) – indivíduos sem AOS ; Grupo II - indivíduos com AOS e grau ligeira; Grupo III - indivíduos com AOS e grau moderado; Grupo IV - indivíduos com AOS e grave;

Critérios de elegibilidade:

Tabela 2 – Critérios de inclusão e exclusão para os grupos da amostra experimental

Inclusão	Exclusão
<ul style="list-style-type: none"><li>• 30 - 60 anos</li><li>• Diagnóstico AOS</li></ul>	<ul style="list-style-type: none"><li>• Medicação que altere o comportamento do traçado ECG</li><li>• Não se sente confortável a manusear um dispositivo móvel</li></ul>

Seleção da amostra: Os indivíduos candidatos serão selecionados pela equipa médica do Serviço de Pneumologia da U.L.S, mediante os critérios de elegibilidade, aquando da consulta de sono. Posteriormente, a equipa médica comunicará à equipa investigadora os candidatos.

Material e Instrumentação: sensor ECG - Movesense (equipamento do projeto), dispositivo móvel (*smartphone ou tablet*) , aplicação móvel pré-instalada, conexão à rede Internet (opcional), manual/panfleto de instruções.



Entrega e recolha do material: A entrega do material necessário à recolha de dados será feita ou pela equipa médica na consulta do utente ou por um dos investigadores deste estudo que se deslocará posteriormente ao domicílio do utente. Assim como a entrega, a recolha do material será feita em consulta ou por um dos investigadores.

Instruções de utilização: Aquando da entrega do material, será feito um esclarecimento do processo de recolha de dados, dos procedimentos a serem tomados e como devem ser utilizados os equipamentos. Adicionalmente será disponibilizado um manual/panfleto onde constarão instruções de utilização de modo para facilitar o processo.

Infraestrutura e Recursos Necessários: A recolha dos dados ECG será feita através de um dispositivo (mais detalhes na seção seguinte), que fica ligado a um dispositivo móvel através de Bluetooth. A recolha deve ser feita no período de sono do indivíduo.

Para que o processo seja descomplicado é necessário que o indivíduo esteja confortável com a utilização de um dispositivo móvel.

Descrição do dispositivo:

O processo de recolha será efetuado pelo dispositivo *Movesense Medical* que se encontra representado na figura abaixo:



**Figura 1 - Modelo do sensor *Movesense Medical*.**



Este dispositivo é composto por dois pinos de eletrodo na parte traseira, pelo qual são recolhidos os dados ECG de único canal. Além do sensor ECG, o dispositivo dispõe de sensores de movimento.

O dispositivo é certificado pela *Eurofins Expert Services*, entidade certificadora de vários produtos, incluindo dispositivos médicos. Estão presentes dois certificados, um com a norma de Sistemas de gestão da qualidade ISO 13485:2016 e o último com a diretiva europeia de dispositivo médico UE 93/42/EEC, dando-lhe a classe IIa.

A comunicação do dispositivo é feita através de Bluetooth com um dispositivo móvel que sucessivamente se conecta com uma aplicação móvel que se encontra em desenvolvimento.

Além do que já foi referido anteriormente, é ligado aos pinos um eletrodo adesivo que contém uma cola Ag/AgCl que facilita tanto a aderência à pele como também facilita na recolha dos sinais elétricos. Exemplos deste adesivo encontram-se na figura seguinte, assim como a sua aplicação na pele:



**Figura 2 - Eletrodo adesivo e colocação do mesmo em conjunto com o sensor.**

O sítio de aplicação do eletrodo não necessita de ser no local representado na Figura 2, no entanto é necessário que este seja colocado no peito ou no início do abdômen do indivíduo.

#### Procedimento de recolha:

Em primeiro lugar, o dispositivo é colocado tal como explicado anteriormente. Para ligar o dispositivo basta colocar o dispositivo em contacto com a pele. Quando ele liga, este mostra uma luz vermelha.

Assim que o dispositivo estiver colocado, basta aceder à aplicação através do dispositivo móvel. É necessário ter o *Bluetooth* do dispositivo móvel ligado para que a aplicação detecte o dispositivo e inicie a recolha de dados. Durante o processo é também necessário que o *bluetooth* se mantenha ligado.

Após a recolha dos dados, estes serão armazenados no dispositivo móvel e enviados para a *Cloud* (Opcional). Neste sentido é também necessário que aquando do início do processo de recolha o Wi-Fi do dispositivo móvel esteja ativo e ligado a uma rede Internet.

#### Prazo e Calendarização:

O protocolo será aplicado assim que for aprovado pela Comissão de Ética da U.L.S da Guarda, durante um período de cerca de três meses.

Aspetos a considerar para a situação pandémica: Atendendo a situação atual os objetos utilizados para o processo são descartáveis ou higienizáveis, nomeadamente, o adesivo eletrodo é descartável e tanto o dispositivo *movesense* como o dispositivo móvel são higienizáveis.

### **5) Aspetos Éticos e Confidencialidade dos Dados**

Riscos: Este procedimento não apresenta quaisquer riscos para os voluntários.

Benefícios: Este estudo permite um método de diagnóstico complementar e alternativo.

Confidencialidade dos dados: Os dados a publicar serão anonimizados por forma a impedir que seja realizada uma associação entre os dados e o indivíduo a quem foram recolhidos, ou a

identificar quais os indivíduos que efetivamente participaram no estudo, de acordo com as normas internacionalmente aceites e a legislação portuguesa em vigor, em particular, em conformidade com a Lei nº 58/2019 de 8 de agosto.

## 6) Bibliografia

N. Pombo, N. Garcia, and K. Bousson, "Classification techniques on computerized systems to predict and/or to detect apnea: A systematic review, "Computer methods and programs in biomedicine, vol. 140, pp. 265–274, 2017

Phan Duy Hung. 2018. Detection of Central Sleep Apnea Based on a Single-Lead ECG. In Proceedings of the 2018 5th International Conference on Bioinformatics Research and Applications (ICBRA '18). Association for Computing Machinery, New York, NY, USA, 78–83. DOI:<https://doi.org/10.1145/3309129.3309132>

Raymond, B., et al. "Screening for obstructive sleep apnoea based on the electrocardiogram-the computers in cardiology challenge." *Computers in Cardiology 2000. Vol. 27 (Cat. 00CH37163)*. IEEE, 2000.

I. De Falco, G. De Pietro and G. Sannino, "On Finding Explicit Rules for Personalized Forecasting of Obstructive Sleep Apnea Episodes," 2015 IEEE International Conference on Information Reuse and Integration, 2015, pp. 326-333, doi: 10.1109/IRI.2015.57.

Fatimah, Binish, et al. "Detection of apnea events from ecg segments using fourier decomposition method." *Biomedical Signal Processing and Control* 61 (2020): 102005.

Warapongmanupong, Siriporn, and Aroonwan Preutthipan. "Can standard deviation of overnight pulse oximetry be used to screen childhood obstructive sleep apnea." *International journal of pediatric otorhinolaryngology* 119 (2019): 27-31.

Xie, Baile, et al. "An improved approach for real-time detection of sleep apnea." *BIOSIGNALS*. 2011.

Yücelbaş, Şule, et al. "Pre-determination of OSA degree using morphological features of the ECG signal." *Expert Systems with Applications* 81 (2017): 79-87.

Goldberger, A., et al. "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* [Online]. 101 (23), pp. e215–e220." (2000).

## TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

**Título da Pesquisa:** Recolha de dados ECG de indivíduos que apresentem sintomatologia associada à apneia obstrutiva do sono

**Nome do Investigador Principal ou Orientador e filiação:** Prof. Doutor Nuno Pombo, Universidade da Beira Interior; Prof. Doutor. Sandeep Pirbhulal, Norwegian University of Science and Technology

**Nome do(s) Investigadores assistentes/alunos e filiação:** Licenciado Eduardo Paulos, Estudante de Mestrado, UBI; Licenciado Rui Charrinho, Estudante de Mestrado, UBI; Mestre Filipa Jesus, Serviço de Pneumologia da U.L.S da Guarda

1. **Natureza da pesquisa:** este estudo tem como finalidade fazer a recolha de dados eletrocardiograma (ECG) aquando do seu período de sono.
2. **Participantes da pesquisa:** um grupo de 10 a 15 indivíduos diagnosticados com Apneia Obstrutiva do Sono.
3. **Envolvimento na pesquisa:** ao participar neste estudo, o voluntário permitirá a recolha de dados ECG através do dispositivo fornecido. Deste modo, será necessário que o voluntário esteja confortável com a utilização do dispositivo e de um *smartphone* ou *tablet*, este último necessário para a comunicação com o dispositivo. Para que o processo de recolha tenha início, o voluntário terá de colocar o dispositivo fornecido no seu peito e posteriormente terá de ligá-lo ao *smartphone* ou *tablet* através da aplicação móvel fornecida. O *smartphone* ou *tablet* também serão fornecidos pela equipa investigadora. Opcionalmente, para que os dados sejam descarregados carregados para a *Cloud*, o voluntário poderá contactar o *smartphone* ou *tablet* à sua rede Wi-Fi local.
4. **Riscos e desconforto:** a participação nesta recolha não traz quaisquer riscos ou desconforto.
5. **Confidencialidade:** de acordo com a Lei nº 58/2019 de 8 de agosto em vigor, todas as informações pessoais recolhidas são estritamente confidenciais. Somente os investigadores terão conhecimento dos dados. Os dados ECG, anonimizados, serão usados em

investigação por outros investigadores, ao serem disponibilizados numa base de dados pública.

6. **Benefícios:** ao participar neste estudo, o voluntário não terá qualquer benefício direto. Entretanto, espera-se que os dados recolhidos tragam informações importantes e adicionais sobre a doença da Apneia Obstrutiva do Sono, de forma que este conhecimento possa contribuir para desenvolvimento e melhoramento dos métodos de avaliação deste efeito.
7. **Pagamento:** o voluntário não terá nenhum tipo de despesa para participar neste estudo, bem como não lhe serão efetuados quaisquer pagamentos.
8. **Disponibilização dos dados:** Os dados recolhidos estarão disponíveis para consulta até a conclusão de dissertação dos alunos de mestrado do estudo em causa. Após este período, serão disponibilizados em acesso livre e gratuito para a comunidade científica internacional, nos servidores da UBI, em <http://allab.it.ubi.pt/mediawiki>.
9. **Contatos:** Eduardo Paulos (Email: [edupaulos1@gmail.com](mailto:edupaulos1@gmail.com), Tel.: 964245718);  
Nuno Pombo (Email: [ngpombo@di.ubi.pt](mailto:ngpombo@di.ubi.pt), Tel.:965683580)

Após estes esclarecimentos, solicitamos o seu consentimento de forma livre para participar nesta pesquisa. Portanto, preencha, por favor, os itens que se seguem.

### **Consentimento Livre e Esclarecido**

Tendo em vistas os itens acima apresentados, eu, \_\_\_\_\_  
\_\_\_\_\_, fui informado(a) dos objetivos desta pesquisa e, de forma livre e esclarecida, manifesto o meu consentimento em participar como voluntário.

Covilhã, \_\_\_\_\_ de \_\_\_\_\_ de \_\_\_\_\_.

**O(A) voluntário(a)**

**O(A) investigador(a)**