

UNIVERSIDADE FEDERAL DE SANTA MARIA  
CENTRO DE TECNOLOGIA  
CURSO DE ENGENHARIA DE COMPUTAÇÃO

Nelson Roberto Weirich Junior

**ESTUDO DE ALGORITMOS DE APRENDIZADO DE MÁQUINA NA  
PREDIÇÃO DE APNEIA DO SONO POR SINAIS DE EEG**

Santa Maria, RS  
2022

**Nelson Roberto Weirich Junior**

**ESTUDO DE ALGORITMOS DE APRENDIZADO DE MÁQUINA NA PREDIÇÃO DE  
APNEIA DO SONO POR SINAIS DE EEG**

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia de Computação da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Bacharel em Engenharia de Computação**. Defesa realizada por videoconferência.

Orientador: Cesar Augusto Prior

Santa Maria, RS  
2022

**Nelson Roberto Weirich Junior**

**ESTUDO DE ALGORITMOS DE APRENDIZADO DE MÁQUINA NA PREDIÇÃO DE APNEIA DO SONO POR SINAIS DE EEG**

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia de Computação da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Bacharel em Engenharia de Computação**. Defesa realizada por videoconferência.

**Aprovado em 17 de fevereiro de 2022:**

---

**Cesar Augusto Prior, Dr. (UFSM)**  
(Presidente/Orientador)



Documento assinado digitalmente

Cesar Ramos Rodrigues

Data: 21/02/2022 17:58:23-0300

CPF: 562.557.990-34

Verifique as assinaturas em <https://v.ufsc.br>

---

**Cesar Ramos Rodrigues, Dr. (UFSC)**

---

**César Abascal Machado, Me. Eng.**

Santa Maria, RS  
2022

*À minha família.*

## **AGRADECIMENTOS**

Meus sinceros agradecimentos ao prof. Cesar Augusto Prior, que orientou este trabalho e manteve suas esperanças, foi paciente e confiante em me ajudar e persistiu nessa trajetória até o fim; à minha família, que me apoiou incondicionalmente na esperança de me ver alcançando meus objetivos; e às psicólogas da CAED, que ajudaram-me no período mais crítico da minha jornada.

Agradecimentos especiais são direcionados à Coordenadoria de Ações Educacionais da UFSM, por oferecer o serviço de apoio psicopedagógico aos estudantes; à própria UFSM, que manteve toda a estrutura necessária à realização do curso de graduação; às amigas que construí durante esses anos, aprendi muito com elas e foram fundamentais para o meu bem-estar; e à todas as pessoas que, de alguma forma, me apoiaram direta ou indiretamente.

## RESUMO

### ESTUDO DE ALGORITMOS DE APRENDIZADO DE MÁQUINA NA PREDIÇÃO DE APNEIA DO SONO POR SINAIS DE EEG

AUTOR: Nelson Roberto Weirich Junior

ORIENTADOR: Cesar Augusto Prior

Apneia do sono é um distúrbio do sono caracterizado pela obstrução parcial ou total do fluxo de ar no trato respiratório. Isso leva a redução da oxigenação sanguínea, induzindo ao despertar. A longo prazo, leva a pessoa a desenvolver outras doenças que afetam sua saúde e bem-estar social. Contudo, o processo de diagnóstico é conhecido e realizado através de exame de polissonografia, onde são coletados sinais bioelétricos e fisiológicos do paciente em uma noite de sono. Analisar sinais bioelétricos é uma tarefa complexa que envolve processamento de sinais; atualmente existem, porém, ferramentas computacionais no campo da inteligência artificial que visam auxiliar tal análise. À vista disso, esta pesquisa teve por objetivo estudar uma nova técnica para predição de apneia do sono, através do processamento de, apenas, sinais de eletroencefalograma (EEG). Para alcançar tal objetivo, foram utilizados sinais de EEG do banco de dados de polissonografia, de livre acesso, do Hospital Universitário St. Vincent da Universidade de Dublin. Esse banco é composto por anotações de exames de polissonografia de 25 participantes — 21 masculinos e 4 femininos; média de idade 50 ( $\pm$  10) anos totalizando 25 polissonogramas (PSGs) de noite completa com dois canais de EEG (C3–A2 e C4–A1). O banco apresenta também anotações de eventos respiratórios de apneia ocorridos durante o exame. O banco de dados foi organizado, agrupando, em um só arquivo, os dados de sinais de EEG e as anotações dos eventos respiratórios. Com esses dados preparados, os sinais foram divididos em bandas de frequência e extraídos os atributos que os caracterizam, de modo a treinar os modelos de aprendizado de máquina. O estudo apresentou resultados que revelam a possibilidade de conseguir identificar apneia do sono apenas com sinais de EEG, facilitando, assim, o processo de análise e diagnóstico da síndrome. O algoritmo *Gradient Boosting*, que obteve o melhor desempenho neste trabalho, apresentou 79,56% de sensibilidade, 90,59% de especificidade e 84,96% de acurácia. Pensando na continuidade do trabalho, sugere-se aprimorar os modelos gerados, com foco na seleção de atributos e estudo de outros algoritmos. Por fim, esses estudos podem ser fonte de recursos para aplicações práticas, como ferramentas de apoio a profissionais em exames e estudos de apneias do sono.

**Palavras-chave:** Engenharia Biomédica. Apneia do Sono. Eletroencefalografia. Inteligência Artificial. Aprendizado de Máquina. Processamento de sinais.

## ABSTRACT

### STUDY OF MACHINE LEARNING ALGORITHMS IN THE PREDICTION OF SLEEP APNEA FROM EEG SIGNALS

AUTHOR: Nelson Roberto Weirich Junior

ADVISOR: Cesar Augusto Prior

Sleep apnea is a sleep disorder characterized by partial or total obstruction of airflow in the respiratory tract. This leads to reduced blood oxygenation, inducing awakening. In the long term, it leads the person to develop other diseases that affect their health and social well-being. However, the diagnostic process is known and performed through the polysomnography, where bioelectrical and physiological signals are collected from the patient during a night's sleep. Analysing bioelectrical signals is a complex task that involves signal processing; currently, however, there are computational tools in the field of artificial intelligence that aim to assist such analysis. In view of this, this research aimed to study a new technique for predicting sleep apnea, through the processing of only electroencephalogram (EEG) signals. To achieve this goal, EEG signals from the open access polysomnography database were used from the University Hospital St. Vincent, University of Dublin. This database is composed of polysomnography exam notes from 25 participants – 21 males and 4 females; mean age 50 ( $\pm$  10) years, totalling 25 full-night polysomnograms (PSGs) with two EEG channels (C3-A2 and C4-A1). The database presents also notes of apnea respiratory events that occurred during the exam. The database was organized, grouping, in a single file, the EEG signal data and the annotations of respiratory events. With this data prepared, the signals were divided into frequency bands and the features that characterize them were extracted, in order to train the machine learning models. The study presented results that reveal the possibility of being able to identify sleep apnea only with EEG signals, thus facilitating the process of analysis and diagnosis of the syndrome. The Gradient Boosting algorithm, which had the best performance in this work, presented 79.56% of sensitivity, 90.59% of specificity and 84.96% of accuracy. Thinking about the continuity of the work, it is suggested to improve the generated models, focusing on the selection of attributes and the study of other algorithms. Finally, these studies can be a source of resources for practical applications, such as tools to support professionals in sleep apnea exams and studies.

**Keywords:** Biomedical Engineering. Sleep Apnea. Electroencephalography. Artificial Intelligence. Machine Learning. Signal Processing.

## LISTA DE ILUSTRAÇÕES

Figura 2.2.1–Exame de polissonografia em laboratório . . . . .	16
Figura 2.3.1–Projeção da cabeça com o posicionamento dos eletrodos no sistema 10-20 . . . . .	17
Figura 2.5.1–Matriz de confusão . . . . .	19
Figura 3.3.1–Projeção da cabeça com a marcação dos eletrodos do banco de dados conforme o sistema 10-20 . . . . .	25
Figura 3.4.1–Diagrama de fluxo do projeto executado . . . . .	26
Figura 3.4.2–Diagrama do fluxo de dados no processo de ML . . . . .	27
Figura 3.4.3–Indicação da faixa de marcação no sinal de EEG . . . . .	28
Figura 4.1.1–Curva de um recorte do sinal de EEG canal C3-A2 . . . . .	30
Figura 4.1.2–Curva de um recorte do sinal filtrado de EEG . . . . .	31
Figura 4.1.3–Histograma da quantidade de épocas por classe . . . . .	31
Figura 4.2.1–Curva ROC dos modelos gerados para cada classificador. . . . .	33



## LISTA DE QUADROS

Quadro 3.1 – Especificações do ambiente pessoal de trabalho . . . . .	23
Quadro 3.2 – Especificações do servidor utilizado nos experimentos . . . . .	24
Quadro A.1 – Relação de resultados por plataforma e termo de busca utilizado . . . . .	39
Quadro B.1 – Principais informações de cada indivíduo registrado no banco de dados do estudo . . . . .	47

## LISTA DE TABELAS

Tabela 4.2.1–Resultados das métricas atingidas neste estudo . . . . .	32
Tabela 4.3.1–Comparação entre as métricas atingidas neste estudo contra de outros estudos da literatura científica . . . . .	34

## LISTA DE ABREVIATURAS E SIGLAS

AUC	<i>Area under the ROC curve</i>
EEG	Eletroencefalograma
ECG	Eletrocardiograma
CNN	<i>Convolutional neural network</i>
CWT	<i>Continuous wavelet transform</i>
DFA	<i>Detrended fluctuation analysis</i>
DL	<i>Deep learning</i>
DVNN	<i>Deep vanilla neural network</i>
HUSM	Hospital Universitário de Santa Maria
IA	Inteligência artificial
IMC	Índice de massa corpórea
ML	<i>Machine learning</i>
PSI	<i>Power spectral intensity</i>
PSG	Polissonograma
RIR	<i>Relative intensity ratio</i>
RNN	<i>Recurrent neural network</i>
ROC	<i>Receiver operating characteristic</i>
SAOS	Síndrome da apneia obstrutiva do sono
SVM	<i>Support vector machine</i>

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>13</b>
1.1	OBJETIVOS	13
1.2	JUSTIFICATIVA	14
1.3	ESTRUTURA DO TEXTO	14
<b>2</b>	<b>REVISÃO BIBLIOGRÁFICA</b>	<b>15</b>
2.1	APNEIA DO SONO	15
2.2	POLISSONOGRAMA	16
2.3	SINAIS DE ELETROENCEFALOGRAMA	17
2.4	INTELIGÊNCIA ARTIFICIAL	17
<b>2.4.1</b>	<b>Aprendizado de máquina</b>	<b>18</b>
<b>2.4.2</b>	<b>Classificadores</b>	<b>18</b>
2.5	MÉTRICAS DE AVALIAÇÃO	19
<b>2.5.1</b>	<b>Métricas para classificação</b>	<b>19</b>
2.5.1.1	Acurácia	19
2.5.1.2	Precisão	20
2.5.1.3	Sensibilidade	20
2.5.1.4	Especificidade	20
2.5.1.5	F1-score	20
2.6	TRABALHOS RELACIONADOS	20
<b>3</b>	<b>MATERIAIS E MÉTODOS</b>	<b>23</b>
3.1	AMBIENTE DE DESENVOLVIMENTO	23
3.2	FERRAMENTAS E CÓDIGOS	24
3.3	BANCO DE DADOS	24
3.4	PROCEDIMENTOS INICIAIS	26
<b>3.4.1</b>	<b>Exploração do banco de dados</b>	<b>27</b>
<b>3.4.2</b>	<b>Criação do sistema de trabalho</b>	<b>27</b>
<b>3.4.3</b>	<b>Pré-processamento</b>	<b>27</b>
<b>3.4.4</b>	<b>Extração de atributos</b>	<b>28</b>
<b>3.4.5</b>	<b>Treinamentos e testes</b>	<b>29</b>
<b>3.4.6</b>	<b>Análise qualitativa</b>	<b>29</b>
<b>4</b>	<b>RESULTADOS E DISCUSSÃO</b>	<b>30</b>
4.1	PROCEDIMENTOS INICIAIS	30
<b>4.1.1</b>	<b>Pré-processamento</b>	<b>30</b>

4.2	FLUXO DE CLASSIFICAÇÃO DOS SINAIS . . . . .	32
4.2.1	Treinamentos e testes . . . . .	32
4.2.2	Validação . . . . .	32
4.2.3	Análise qualitativa . . . . .	33
4.3	DISCUSSÃO . . . . .	34
4.3.1	Potencial para aplicação . . . . .	35
5	CONCLUSÃO . . . . .	36
5.1	SUGESTÕES DE TRABALHOS FUTUROS . . . . .	36
	<b>REFERÊNCIAS . . . . .</b>	<b>37</b>
	<b>APÊNDICE A – PESQUISA DE TRABALHOS RELACIONADOS .</b>	<b>39</b>
A.1	SUMÁRIO . . . . .	39
A.2	PLATAFORMA: PUBMED . . . . .	39
A.3	PLATAFORMA: NATURE . . . . .	41
A.4	PLATAFORMA: GOOGLE SCHOLAR . . . . .	42
A.5	PLATAFORMA: PHYSIONET . . . . .	43
A.6	PLATAFORMA: GITHUB . . . . .	43
A.7	PLATAFORMA: KAGGLE . . . . .	44
A.8	PLATAFORMA: ZENODO . . . . .	45
	<b>APÊNDICE B – INFORMAÇÕES DOS INDIVÍDUOS NO BANCO DE DADOS . . . . .</b>	<b>47</b>

## 1 INTRODUÇÃO

Doenças relacionadas ao sono são problemas recorrentes na sociedade e levam ao desenvolvimento de comorbidades a seus portadores. Nos EUA, 31,6% dos adultos com 18 anos ou mais tem sono insuficiente, enquanto 72,7% das crianças da faixa de 9–12 anos apresentam essa característica (CENTER FOR DISEASE CONTROL AND PREVENTION, 2017). O distúrbio mais comum é a apneia do sono, junto à insônia e síndrome das pernas inquietas.

Síndrome da apneia do sono é um distúrbio do sono caracterizado pela obstrução parcial ou total do fluxo de ar no trato respiratório. O portador desenvolve, a longo prazo, outras doenças relacionadas, que causam prejuízos em sua saúde e bem-estar. Algumas doenças são menos danosas, enquanto outras podem levar a sérios riscos se não tratadas.

Contudo, existem exames que servem para diagnosticar a síndrome e permitem que o paciente inicie um tratamento adequado. Dentre eles, destaca-se o exame de polissonografia — padrão atual para seu diagnóstico — que extrai sinais fisiológicos e bioelétricos do paciente durante uma noite de sono completa. Alguns desses sinais são os sinais de eletroencefalograma (EEG), eletrocardiograma (ECG), taxa de oxigenação sanguínea.

A análise desses sinais é, há muito tempo, conduzida exclusivamente por profissionais qualificados nessa área. Com o passar do tempo, novas tecnologias surgiram e serviram de auxílio ao processo. Mais ainda, com o avanço de tecnologias na área da inteligência artificial (IA), começaram-se pesquisas com os sinais de polissonografia digitalizados. Existem universidades e instituições de pesquisas que possuem centros especializados na análise do sono, onde realizam o exame de polissonografia. Algumas dessas disponibilizam publicamente seus bancos de dados de coletas dos sinais de polissonograma (PSG); este trabalho utilizará como objeto de estudos um desses bancos.

O desenvolvimento deste trabalho foi orientado por duas questões principais:

- a) que outras abordagens podem ser utilizadas para identificar apneia do sono, e suas variações, através dos dados de polissonografia?
- b) é possível utilizar apenas sinais de EEG nesse processo?

Baseando-se nesses questionamentos, considera-se possível uma nova abordagem para classificação de sinais de EEG para predição de apneia do sono.

### 1.1 OBJETIVOS

Este trabalho tem por objetivo analisar uma técnica para classificação de apneia do sono, unicamente, por sinais de EEG, utilizando algoritmos de aprendizado de máquina e compará-los com técnicas já utilizadas na literatura.

Em via de alcançar o objetivo geral, destacam-se os seguintes objetivos específicos:

- a) pesquisar trabalhos relacionados;

- b) estudar bancos de dados públicos disponíveis;
- c) estudar técnicas utilizadas em classificação de sinais de EEG;
- d) implementar os métodos computacionais para classificação de sinais;
- e) estudar e implementar algoritmos otimizados para detecção de apneia do sono;
- f) auferir métricas de desempenho computacional dos métodos;
- g) analisar os resultados.

## 1.2 JUSTIFICATIVA

Pretende-se com este estudo explorar novas formas de analisar sinais de polissonografia usando recursos computacionais aplicados com técnicas de IA. Com tais estudos, pretende-se futuramente aplicá-los junto a laboratórios que estudam os distúrbios do sono, como no Hospital Universitário de Santa Maria (HUSM), por exemplo, auxiliando na formação de base de dados com anotações dos profissionais.

Essas análises servem como estudos iniciais para melhorar os processos de tratamento de dados de polissonografia. Isso auxilia o processo de diagnóstico e acompanhamento médico do portador de apneia.

## 1.3 ESTRUTURA DO TEXTO

A obra está dividida em cinco capítulos estruturais, incluindo este.

O [Capítulo 2](#) apresenta uma revisão da literatura acerca do tema. São analisadas obras que embasam o referencial teórico do estudo sobre apneia do sono.

O [Capítulo 3](#) apresenta as ferramentas, o ambiente de trabalho, o objeto de estudo e a metodologia utilizada para o desenvolvimento deste trabalho.

No [Capítulo 4](#), são mostrados os resultados da aplicação dos procedimentos descritos no [Capítulo 3](#).

O texto encerra-se com o [Capítulo 5](#), que apresenta uma conclusão para os objetivos deste trabalho e deixa exposto caminhos para futuros estudos que partam deste.

## 2 REVISÃO BIBLIOGRÁFICA

Para montar o aporte teórico deste trabalho, pesquisaram-se materiais em livros, artigos publicados, portais de saúde e portais virtuais de bases de dados e de códigos-fonte. As publicações foram pesquisadas nas seguintes bases: PubMed, Nature e IEEE Xplore; banco de dados nas plataformas PhysioNet, Kaggle, Zenodo e Google Dataset Search; e códigos-fonte em GitHub. Ainda assim, utilizaram-se as bases não-convencionais Google Search e DuckDuckGo Search, de modo a encontrar trabalhos não técnicos ou com ligação perdida em meio a outras bases de pesquisa. No [Apêndice A](#) encontra-se uma lista dos principais resultados refinados da pesquisa conforme a base de dados, assim como o termo utilizado como chave de busca.

### 2.1 APNEIA DO SONO

Apneia do sono é um distúrbio do sono caracterizado pela obstrução total ou parcial do fluxo de ar. Isso resulta na diminuição da oxigenação do sangue, levando a pessoa ao despertar para recuperar o controle da respiração. Pelos eventos que se desencadeiam durante um episódio de apneia, esse distúrbio é caracterizado como grave. Fatores que podem iniciar ou agravar um quadro de apneia são: obesidade, características fisiológicas favoráveis, uso de álcool, tabagismo, cansaço excessivo, idade avançada ([INSTITUTO DO SONO DA ASA NORTE, 2021](#)).

A síndrome é categorizada segundo os seguintes fatores: nível de obstrução do fluxo de ar e fonte causadora da obstrução. Quando há obstrução total do fluxo de ar, denomina-se por síndrome da apneia do sono<sup>1</sup>; quando há obstrução parcial do fluxo de ar, denomina-se por síndrome da hipopneia do sono. Por outro lado, quando a causa da obstrução for decorrente das vias aéreas, classifica-se como apneia/hipopneia obstrutiva do sono; quando a causa da obstrução for decorrente da falha na comunicação do sistema nervoso central, classifica-se como apneia/hipopneia do sono tipo central; nos casos onde ambas causas são presentes, o evento classifica-se como apneia/hipopneia mista do sono. A síndrome da apneia obstrutiva do sono (SAOS) é o mais comum de todos os casos ([THEORELL-HAGLÖW \*et al.\*, 2018](#)).

Esse quadro desencadeia uma série de prejuízos ao indivíduo: cansaço diurno, respiração anormal e ruidosa, dor de cabeça, fadiga, boca seca, irritabilidade, bruxismo, privação do sono e insônia. Se não tratado, no decorrer do tempo, as consequências tendem a ser mais graves e tornarem-se crônicas, necessitando de intervenção mais severa ([INSTITUTO DO SONO DA ASA NORTE, 2021](#)).

Para diagnóstico e análise do quadro do paciente, existe um exame clínico padrão atual, denominado exame de polissonografia. Pessoas que apresentam alguns dos problemas relatados, são avaliadas nesse exame para serem posteriormente direcionadas ao tratamento adequado ([INSTITUTO DO SONO DA ASA NORTE, 2021](#)).

<sup>1</sup> Termo grafado segundo Descritores em Ciências da Saúde DeCS/MeSH.

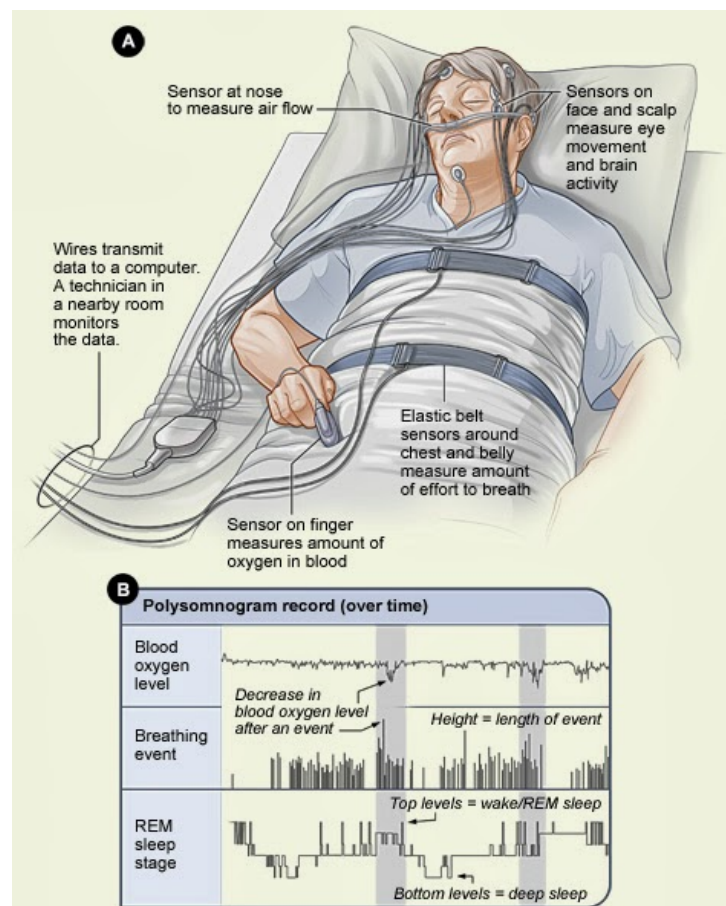


## 2.2 POLISSONOGRAFIA

A polissonografia é um teste clínico não invasivo realizado para investigar distúrbios do sono. O teste é caracterizado por registrar diferentes sinais fisiológicos do paciente; alguns exemplos são: sinais de EEG, nível de oxigênio no sangue, frequência cardíaca, movimentos do tórax e posição corporal. Há exames que extraem mais sinais além dos mencionados (MAYO CLINIC, 2021).

O teste é feito em um paciente, durante uma noite completa de sono, pelo aparelho de polissonografia. O aparelho tem sensor nasal, para medir o fluxo de ar; sensor na face e no escalpo, para medir os movimentos dos olhos e a atividade cerebral; sensor no dedo, para medir a oxigenação sanguínea; e faixas no tronco, para medir a movimentação da caixa torácica. O aparelho é ligado a um computador, que registra os sinais, através de cabos. A Figura 2.2.1 ilustra um paciente com o aparelho instalado e alguns sinais do polissonograma.

Figura 2.2.1 – Exame de polissonografia em laboratório



Fonte – Cardiorespira Cardiopulmonar (2013).

### 2.3 SINAIS DE ELETROENCEFALOGRAMA

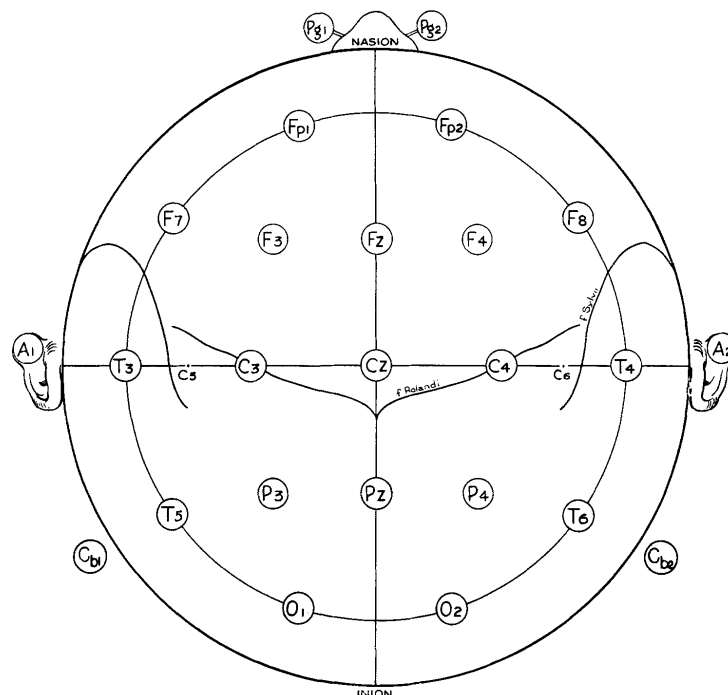
Eletroencefalografia é um método de monitoramento não invasivo da atividade bioelétrica do cérebro. As medições são feitas por eletrodos dispostos no escalpo segundo o sistema 10-20 (NIEDERMEYER; SCHOMER; HENRIQUE, 2011).

Esse sinal é utilizado para análise e diagnóstico de doenças, mas também pode ser utilizado para estudo dos padrões de pensamento.

Atualmente, existem dispositivos que facilitam a medição e registro dos sinais de EEG. Esses dispositivos tornam-se úteis para a criação de aplicações que utilizem processamento de sinais em sistemas inteligentes.

Os eletrodos são posicionados no escalpo conforme o sistema internacional 10–20. Esse sistema é um método para localizar os eletrodos a partir de distâncias padronizadas. Sendo assim, padroniza o local de medição dos sinais de EEG em diferentes cenários de uso. A Figura 2.3.1 ilustra um plano de projeção do topo da cabeça com a localização dos eletrodos, conforme foi apresentado por (JASPER, 1958).

Figura 2.3.1 – Projeção da cabeça com o posicionamento dos eletrodos no sistema 10-20



Fonte – Jasper (1958).

### 2.4 INTELIGÊNCIA ARTIFICIAL

Inteligência artificial (IA) é definida como “a ciência e engenharia de produzir sistemas inteligentes” (MCCARTY, 2007). É, hoje, uma área de pesquisa da computação que se dedica em estudar sistemas com capacidade de resolver problemas a semelhança do pensamento hu-

mano. Diferente das aplicações tradicionais na computação, que resolviam problemas a partir de um algoritmo específico para um conjunto de problemas escrito por um humano, os sistemas inteligentes utilizam apenas dados de entrada para “aprender” algum padrão de comportamento e gerar um modelo, ou algoritmo, para resolver problemas com dados semelhantes (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012). Essa capacidade de adaptar-se aos dados é o que torna os sistemas, de fato, inteligentes e próximos à capacidade de pensamento humano, embora não cheguem perto da complexidade do cérebro biológico.

Esse ramo da Ciência da Computação fornece conteúdo para diversas outras áreas científicas para pesquisa e desenvolvimento de produtos e serviços. Na área médica, em especial no campo da medicina clínica, têm-se desenvolvido estudos desde os anos 70, quando a Engenharia Biomédica começou a ser explorada no Brasil (INSTITUTO DE ENGENHARIA BIOMÉDICA DA UFSC, 2011).

### 2.4.1 Aprendizado de máquina

Aprendizado de máquina, ou *machine learning* (ML), é o campo de estudo científico, dentro da área de IA, focado nos algoritmos de indução (KONAVI; FOSTER, 1998). Pode ser descrito, ainda, como o conjunto de ferramentas e métodos computacionais focados em melhorar o desempenho, ou fazer previsões precisas, a partir da própria experiência desses métodos (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012).

Originou-se de estudos que se preocupavam “com a programação de um computador digital para se comportar de uma maneira que, se feito por seres humanos ou animais, seria descrito como envolvendo o processo de aprendizagem” (SAMUEL, 1959).

Os algoritmos de ML são utilizados para a criação de modelos que descrevem o comportamento de um conjunto de dados. Esses modelos são utilizados para classificação, regressão ou outras formas de identificação de padrões. O grande desafio do ML é projetar, ou encontrar, algoritmos de previsão eficientes e precisos (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012). O objetivo principal é obter o algoritmo que atenda às especificações de qualidade baseadas no problema a ser resolvido.

### 2.4.2 Classificadores

Dentre muitos classificadores desenvolvidos ao longo dos anos, existe um grupo seletivo que apresentam ótima eficiência em sistemas de aprendizado profundo, ou *deep learning* (DL); esses foram os mais utilizados para classificação de sinais de polissonograma nos últimos dez anos: *convolutional neural network* (CNN), *deep vanilla neural network* (DVNN) e *recurrent neural network* (RNN) (MOSTAFA *et al.*, 2019).

## 2.5 MÉTRICAS DE AVALIAÇÃO

### 2.5.1 Métricas para classificação

Dos métodos de ML que realizam a classificação de dados separados em categorias, tem-se definidos valores de desempenho medidos de tal forma que seja possível avaliá-los. Tendo um banco de dados com um conjunto de classes de duas dimensões, então o classificador irá estimar com base nos atributos fornecidos qual classe se encaixa esse vetor de atributos. Disso, tem-se os dados classificados, podendo ser dos seguintes tipos: verdadeiro positivo, falso positivo, verdadeiro negativo ou falso negativo.

- a) Verdadeiro positivo (TP – true positive): predição correta da classe positiva.
- b) Falso positivo (FP – false positive): predição incorreta da classe positiva.
- c) Verdadeiro negativo (TN – true negative): predição correta da classe negativa.
- d) Falso negativo (FN – false negative): predição incorreta da classe negativa.

Dispondo os valores esperados (ou verdadeiros) de cada classe em colunas e os valores previstos de cada classe em linhas baseado nos valores verdadeiros, gera-se um matriz, denominada matriz de confusão, como apresentada na [Figura 2.5.1](#).

Figura 2.5.1 – Matriz de confusão

		Classe Predita	
		0	1
Classe Original	0	<b>TN</b>	<b>FP</b>
	1	<b>FN</b>	<b>TP</b>

Fonte – Bezerra (2018)

A combinação desses valores gera alguns indicadores que servem para avaliar comparativamente os modelos. Elencam-se abaixo os três indicadores principais de desempenho.

#### 2.5.1.1 Acurácia

Representa a taxa de acertos sobre os resultados totais de um modelo de ML. Do inglês *accuracy* (ACC).

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.1)$$

### 2.5.1.2 Precisão

Representa a taxa de acertos sobre os resultados positivos. É a fração de resultados positivos que, de fato, são verdadeiros positivos. Também conhecida como os valores preditivos verdadeiros, do inglês *positive predictive value* (PPV).

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2.2)$$

### 2.5.1.3 Sensibilidade

Descreve a probabilidade de um teste positivo ocorrer com sucesso. Também conhecido como a taxa de verdadeiros positivos, do inglês *true positive rate* (TPR).

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2.3)$$

### 2.5.1.4 Especificidade

Descreve a probabilidade de um teste negativo ocorre com sucesso. Também conhecida como a taxa de verdadeiros negativos, do inglês *true negative rate* (TNR).

$$\text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (2.4)$$

### 2.5.1.5 F1-score

É o valor da média harmônica da precisão e sensibilidade. Esse valor representa a acurácia de um classificador binário.

$$\text{F1} = \frac{2 \times \text{precisão} \times \text{sensibilidade}}{\text{precisao} + \text{sensibilidade}} \quad (2.5)$$

## 2.6 TRABALHOS RELACIONADOS

Diversos trabalhos publicados abordam a detecção de apneia do sono utilizando ML. Foi realizada uma busca por trabalhos nas plataformas PubMed, Nature e IEEE Xplore; banco de dados nas plataformas PhysioNet, Kaggle, Zenodo e Google Dataset Search; e códigos-fonte em GitHub.

Para selecionar os trabalhos publicados, utilizaram-se os seguintes critérios de inclusão: trabalhos que utilizam técnicas de ML e sinais de polissonograma para predição e diagnóstico de síndromes de apneia e hipopneia; e trabalhos que trazem informações sobre variações do uso de EEG e técnicas utilizadas com esse tipo de sinal. Por outro lado, utilizaram-se os seguintes critérios de exclusão: trabalhos que utilizam outros métodos de diagnóstico de síndromes de apneia e hipopneia e trabalhos que utilizam EEG para diagnóstico de outras doenças não relacionadas às síndromes.

Em [Keshavarz et al. \(2020\)](#) foi investigado e avaliado os métodos de aprendizado de máquina supervisionado para prever a SAOS. Neste trabalho, utilizaram uma base de dados com 231 registros de polissonografia. Sobre a base de dados, aplicaram algoritmos de ML populares com o método de validação cruzada. Os classificadores com maior valor de área sobre a curva ROC (*Receiver operating characteristic*), ou *area under the ROC curve* (AUC), foram *Naïve Bayes* e *Support Vector Machine* (SVM), com AUC de 0,768 e 0,761, respectivamente.

Em [Gutierrez-Tobal et al. \(2019\)](#) utilizaram-se registros de SpO<sub>2</sub> de canal único coletados de 320 indivíduos para obter informações estatísticas, espectrais, não lineares e clínicas relacionadas às síndromes de apneia e hipopneia do sono. Utilizaram quatro modelos de ML para classificar a gravidade em quatro graus: sem síndrome, leve, moderada e grave. O modelo de *AdaBoost* com discriminantes lineares foi o melhor classificador; alcançou acurácias de 92,9%, 87,4% e 78,7% em tarefas de classificação binária com limiares de gravidade de índice 5, 15 e 30 eventos por hora, respectivamente. Os autores concluíram que os “resultados sugerem que o ML pode ser usado com as informações de SpO<sub>2</sub> adquiridas na residência de um paciente para ajudar no diagnóstico”.

[Mostafa et al. \(2019\)](#) apresenta, em seu estudo de revisão sobre os trabalhos relacionados à apneia do sono com uso de aprendizado de máquina da última década, uma visão das diversas técnicas utilizadas para análise dos sinais; e tenta responder “que tipo de pré-processamento ou extração de recursos são necessários” e as “vantagens e desvantagens dos diferentes tipos de redes neurais”. Consideraram, no estudo, a diferença dos sinais empregados, sensores, banco de dados e os desafios de implementação. Para a pesquisa, realizaram uma busca sistemática em cinco serviços de indexação de 2008 a 2018; encontraram, no final, 255 artigos e selecionaram 21 conforme os critérios de inclusão e exclusão.

[Zhou, Wu e Zeng \(2015\)](#) apresenta um estudo visando “encontrar um método novo e mais simples para detectar pacientes com apneia e quantificar as características não lineares da apneia do sono”. Utilizando dados de EEG de 12 indivíduos (6 saudáveis e 6 com síndrome), calcularam os expoentes de escala de EEG de 30 min utilizando a *destrended fluctuation analysis* (DFA). Com o cálculo dos expoentes de escala média, calculados a cada 30 segundos, encontraram 360 valores de controle e 360 valores de apneia. Utilizaram SVM para classificar os pacientes, tal algoritmo obteve acurácia média de 95,1% com sensibilidade de 93,2% e especificidade de 98,6%. Com isso, concluíram que a DFA em sinais de EEG é “um método eficiente e praticável, e é útil clinicamente no diagnóstico de apneia do sono”.

Em [Zhao et al. \(2021\)](#) é apresentado um método para categorizar eventos de apneia do sono através de características de sinais de EEG. No trabalho, utilizou-se filtro passa-banda *Butterworth* de resposta ao impulso infinito nos sinais dos canais C3-A2 e C4-A1. As características extraídas foram entropia e variância da amostra de cada sub-banda; e o método para selecionar as características foi o método de análise de composição de vizinhos. Para a classificação, utilizaram os seguintes classificadores: *Random Forest*, *k-Nearest Neighbors*, e SVM com validação cruzada de 10 vezes. Ao final, obtiveram uma acurácia média de 88,99%. Por fim, concluíram

que o método tem potencial para classificação automática de eventos respiratórios.

Biswal *et al.* (2018) utilizou espectrograma de sinal para calcular a densidade espectral de potência, ou *power spectral density* (PSD), pelo método *multitaper* de Thompson. No sinal de EEG, utilizaram o tamanho da janela de PSD de 2 segundos.

Cen *et al.* (2018) utilizou o banco de dados público do Hospital Universitário St. Vincent, disponível na plataforma PhysioNet. Com uso de CNN alcançou o valor médio absoluto de 79,61% de acurácia para detecção de hipopneia e apneia em sinais de oxigenação sanguínea, fluxo de ar e movimento do abdômen.

Almuhammadi, Aboalayon e Faezipour (2015) introduz uma metodologia eficiente que pode ser implementada em hardware para diferenciar pacientes com SAOS. A partir de um banco de dados público, extraíram as componentes delta, theta, alpha e beta, com filtro passa-banda de resposta ao impulso infinita. Dessas, obtiveram as features: energia e variância. Para treinar os modelos, utilizaram os algoritmos SVM, *Artificial Neural Network*, *Linear Discriminant Analysis* e *Naïve Bayes*. O melhor resultado obtido foi para o SVM com acurácia de 97,14%, utilizando 10% das amostras como conjunto de teste.

### 3 MATERIAIS E MÉTODOS

O desenvolvimento da metodologia deste trabalho envolveu as atividades de busca, análise e organização de dados; avaliação, desenvolvimento e teste de algoritmos de processamento de sinais e de ML; e visualização de resultados.

Inicialmente, após a pesquisa, escolheu-se o banco de dados da Universidade de Dublin, descrito com detalhes na [seção 3.3](#). O diretório principal continha arquivos com anotações de eventos respiratórios, em texto simples, separados dos arquivos de dados brutos, no formato EDF. A primeira tarefa foi identificar os dados de interesse e organizá-los para, em seguida, iniciar o processamento dos sinais.

#### 3.1 AMBIENTE DE DESENVOLVIMENTO

A maior parte do desenvolvimento do projeto foi feita em um computador pessoal com as especificações mostradas no [Quadro 3.1](#). O processo de escrita de códigos era feito em um emulador de terminal no ambiente VIM<sup>1</sup>. Por outro lado, para escrever as rotinas e registrar os resultados, eram utilizados *notebooks* Python. Nesses *notebooks*, chamavam-se os métodos para criar o fluxo de dados e gravavam-se os resultados, que ficavam registrados no arquivo.

Quadro 3.1 – Especificações do ambiente pessoal de trabalho

Recurso	Especificação
Processador	Intel(R) 2.5GHz 64bits
Memória	8GB
Armazenamento	1TB
Sistema Operacional	Ubuntu 18.04.5
Kernel	Linux 4.15.0-154-generic
Shell	zsh 5.4.2
Versão Conda	4.9.2
Versão Python	3.7.10

Fonte – autor.

Os treinamentos e testes ocorreram em um servidor hospedado na Universidade Federal de Santa Maria (UFSM), o qual era constituído dos recursos apresentados no [Quadro 3.2](#).

<sup>1</sup> VIM é um programa de terminal para edição de arquivos de texto. Sendo assim, é uma poderosa ferramenta para criação e edição de todo tipo de arquivo de texto bruto.



Quadro 3.2 – Especificações do servidor utilizado nos experimentos

Recurso	Especificação
Processador	Intel(R) Xeon(R) 2GHz 64bits
Memória	8GB
Armazenamento	323GB
Sistema Operacional	Ubuntu 18.04.1
Kernel	Linux 5.4.0-70-generic
Shell	bash 4.4.20
Versão Conda	4.9.2
Versão Python	3.7.10

Fonte – autor.

### 3.2 FERRAMENTAS E CÓDIGOS

Todo o processo inicial de exploração dos dados foi realizado com o uso de *scripts* em shell<sup>2</sup>, incluindo *scripts* em AWK<sup>3</sup>, tendo em vista que são ferramentas mais robustas e rápidas para manipular grandes volumes de dados. Nas etapas posteriores do fluxo de trabalho, utilizaram-se algoritmos e *scripts* em Python, já que há uma gama de recursos prontos para trabalhar em projetos de IA.

Dentre os pacotes em Python utilizados, destacam-se: conda<sup>4</sup>, scikit-learn<sup>5</sup>, NumPy<sup>6</sup> e pandas<sup>7</sup>.

### 3.3 BANCO DE DADOS

Para treinar e avaliar o desempenho do método proposto, utilizaram-se sinais de EEG retirados do banco de dados do Hospital Universitário St. Vincent da Universidade de Dublin. Os dados vieram de estudos realizados com indivíduos randomicamente selecionados em um período de seis meses de pacientes ligados ao Sleep Clinic Disorders da St Vincent’s University Hospital, Dublin, para possível diagnóstico de apneia obstrutiva do sono, apneia central do sono ou roncos primários (HENEGHAN, 2021).

O banco de dados era composto por anotação de exames de polissonografia de 25 participantes (21M, 4F; idade: 50 anos ( $\pm 10$ ), intervalo 28–68 anos; índice de massa corpórea (IMC): 31,6 ( $\pm 4,0$ ) kg/m<sup>2</sup>, intervalo); totalizando 25 polissonogramas (PSGs), de aproximadamente

<sup>2</sup> Shell é um programa de computador de sistemas Unix, escrito por Ken Thompson, que permite acesso aos serviços do núcleo sistema operacional, funcionando como uma interface de usuário e acessado através de uma interface de linha de comando (CLI – do inglês *command line interface*). Por isso, também é conhecido como um interpretador de comandos.

<sup>3</sup> AWK é uma linguagem de programação interpretada que serve para manipulação de texto e facilita a criação de *scripts* shell. Criada por Alfred Aho, Peter J. Weinberger e Brian Kernighan no laboratório Bell Labs, seu nome é um acrônimo dos nomes de seus criadores.

<sup>4</sup> conda versão 4.9.2.

<sup>5</sup> scikit-learn versão 0.24.1.

<sup>6</sup> NumPy versão 1.19.2.

<sup>7</sup> pandas versão 1.2.3.

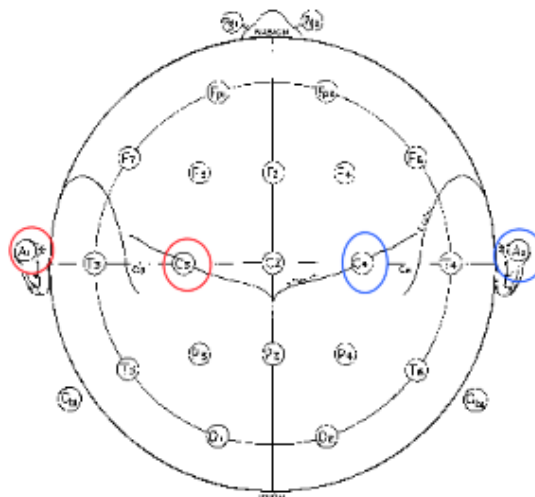
7h, de noite de sono completa com três canais simultâneos de Holter ECGs. O [Quadro B.1](#), apresenta algumas informações características dos indivíduos e do estudo realizado.

Os seguintes sinais fisiológicos estavam registrados neste banco:

- a) EEG (C3-A2);
- b) EEG (C4-A1);
- c) EOG esquerdo;
- d) EOG submental;
- e) ECG (derivação V2 modificada);
- f) fluxo de ar oro-nasal (termistor);
- g) movimentos da caixa torácica;
- h) movimentos do abdômen (medidores de tensão não calibrados);
- i) saturação de oxigênio (oxímetro de pulso);
- j) ronco (microfone traqueal);
- k) posição corporal.

A [Figura 3.3.1](#) apresenta o posicionamento dos eletrodos utilizados neste banco de dados: à esquerda o canal C3-A2 (vermelho) e à direita o canal C4-A1 (azul).

Figura 3.3.1 – Projeção da cabeça com a marcação dos eletrodos do banco de dados conforme o sistema 10-20



Fonte – Adaptação de [Jasper \(1958\)](#).

Junto ao banco de dados, haviam arquivos de anotações de estágios do sono e arquivos de anotações de eventos respiratórios. Os eventos respiratórios estavam descritos pelo tempo inicial, tipo de evento e duração; estavam anotados os seguintes tipos de eventos:

- a) O: apneias obstrutivas;
- b) C: apneias centrais;

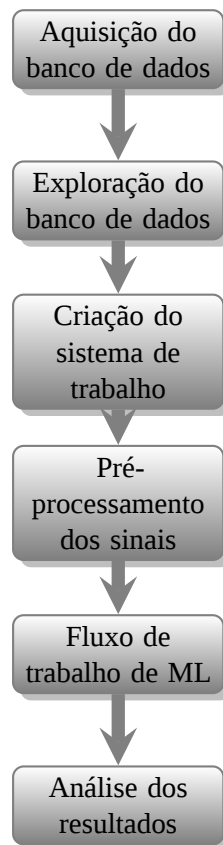
- c) M: apneias mistas;
- d) HYP: hipoapneias;
- e) PB: episódios respiratórios periódicos;
- f) CS: Cheynes-Stokes
- g) Bradicardia/Taquicardia.

### 3.4 PROCEDIMENTOS INICIAIS

Inicialmente, foi realizada a exploração do banco de dados para conhecer a organização e a estrutura dos dados. Após esse processo, foi desenvolvido um sistema de trabalho com a criação de módulos de funções e *scripts* Python, AWK e shell. Iniciou-se, então, o pré-processamento dos sinais armazenados no banco de dados. Com isso, partiu-se para a modelagem dos classificadores com os algoritmos de ML usando os atributos extraídos manualmente.

A [Figura 3.4.1](#) ilustra o fluxo de processos realizados neste trabalho.

Figura 3.4.1 – Diagrama de fluxo do projeto executado



Fonte – autor.

### 3.4.1 Exploração do banco de dados

Com os *scripts* em shell, AWK e Python, realizou-se a exploração dos dados, efetuando as seguintes tarefas:

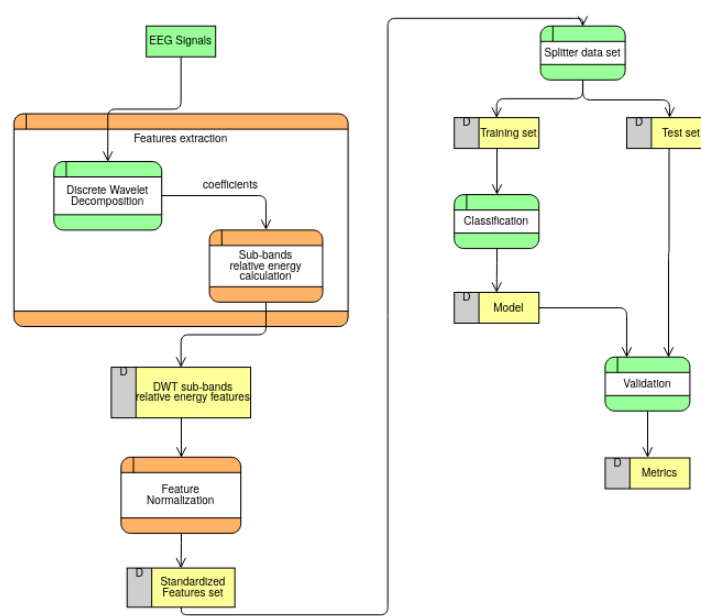
- baixar o banco do servidor da PhysioNet;
- extrair as amostras dos arquivos EDF para arquivos CSV;
- extrair os sinais dos canais de EEG;
- agregar os dados de sinais com suas anotações;
- concaternar todos os arquivos e criar o conjunto de dados de treino, teste e validação.

### 3.4.2 Criação do sistema de trabalho

Tendo em vista o fluxo de execução e o encadeamento das tarefas a serem executadas, criou-se uma organização de códigos fonte em Python, na forma de módulos de funções. Essa organização modular auxiliou nas chamadas de métodos, posteriormente, nos *notebooks*.

A [Figura 3.4.2](#) apresenta do fluxo de processo de ML, desde o pré-processamento até a geração de métricas de desempenho.

Figura 3.4.2 – Diagrama do fluxo de dados no processo de ML



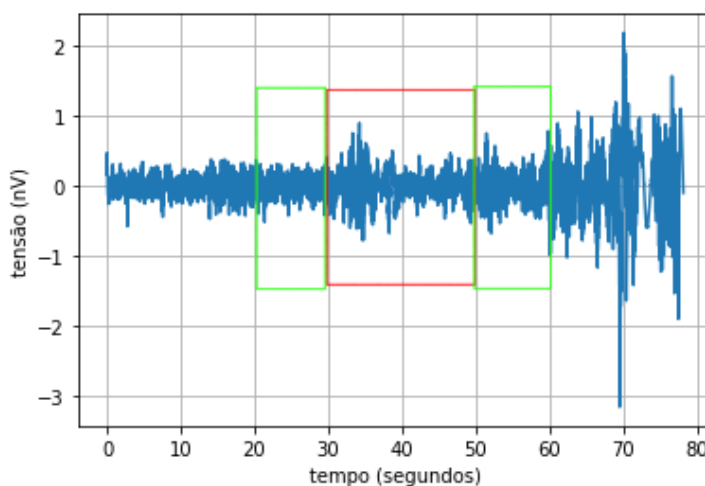
Fonte – autor.

### 3.4.3 Pré-processamento

Primeiramente, os sinais de cada sujeito do banco de dados passou por um processo de filtro, com objetivo de remover ruídos de alta e baixa frequência. Para isso, foi utilizado um filtro *Butterworth* passa-banda de 0,5 a 70 Hz, que preserva a faixa de frequências características dos

sinais de EEG, conforme descrito em (VANHATALO; VOIPIO; KAILA, 2005). A partir disso, os sinais foram divididos em épocas<sup>8</sup> de 10 segundos. Cada época representa uma instância da base de dados final, ou seja, cada época possui determinados atributos e está ligada a uma classe. Então, para cada uma, separaram-se as bandas de frequência características do sinal de EEG: delta, theta, alpha, beta, gamma; extrairam-se os atributos presentes no sinal em cada intervalo e atribuiu-se uma classe que identificasse o sinal. Foram marcados sinais no intervalo de até 10 segundos antes do começo e após o término de um evento de apneia como pertencentes à classe 'APNEA', como mostra a Figura 3.4.3. Todas as três categorias de apneia (obstrutiva, mista e central) foram classificadas com uma classe indistinta. O resto foi definido como ausência de apneia.

Figura 3.4.3 – Indicação da faixa de marcação no sinal de EEG



Fonte – autor.

Identificou-se que as classes estavam desbalanceadas, então foi necessário balanceá-las, de modo a não comprometer o processo de treino dos classificadores. Foi utilizada a função *NearMiss*<sup>9</sup> em Python, que realiza um processo de subamostragem dos dados da classe majoritária, reduzindo as amostras de forma aleatória e baseando-se na distância entre as instâncias. Após a aplicação da função, ambas as classes apresentam a mesma quantidade de amostrar da classe minoritária.

#### 3.4.4 Extração de atributos

Foram extraídos os seguintes atributos dos sinais de EEG: média, desvio padrão, variância, zero crossing, entropia, PSI<sup>10</sup> de cada banda e RIR<sup>11</sup> de cada banda.

Após, dividiu-se o conjunto de dados em conjuntos de treino e teste, com 80% e 20% das amostras, respectivamente.

<sup>8</sup> Uma época é um intervalo de tempo que contém as amostras desse intervalo.

<sup>9</sup> *NearMiss* é um método implementado pela biblioteca *imblearn*.

<sup>10</sup> *Power spectral intensity*.

<sup>11</sup> *Relative intensity ratio*.

### 3.4.5 Treinamentos e testes

Em um *notebook* Python foram escritos os procedimentos para o fluxo de treino e testes dos modelos de classificadores. Nesse processo, foram utilizados os seguintes algoritmos: *Ada Boost*, *Gradient Boosting*, *Logistic regression*, *K-nearest neighbors*, *Naïve Bayes*, *Linear SVM*, *Random Forest*. Para cada um, foram configurados os parâmetros da função em Python.

### 3.4.6 Análise qualitativa

O conjunto de teste e validação consistiu de 20% do total de dados resultantes do balanceamento por reamostragem. Esse conjunto foi utilizado para testar os modelos gerados pelo treinamento dos classificadores. Para isso, extraíram-se as seguintes métricas estatísticas: sensibilidade, especificidade, precisão, acurácia e *F1-score*.

## 4 RESULTADOS E DISCUSSÃO

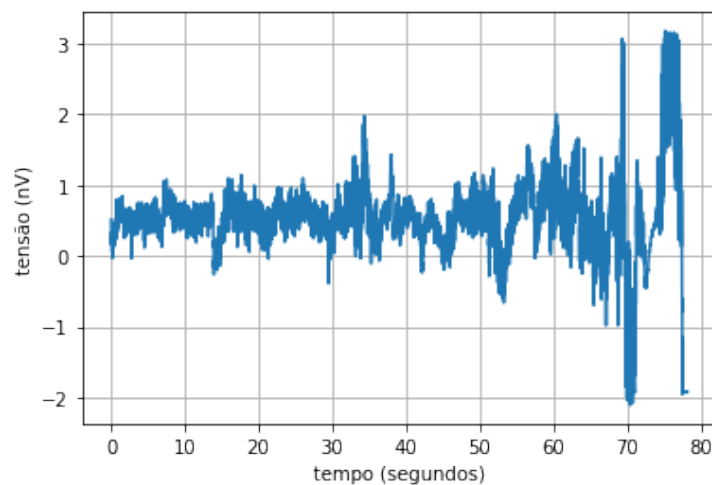
Neste capítulo, são apresentados os resultados advindos do sistema desenvolvido. Após apresentar os subprodutos das etapas prévias ao treino, encontram-se os resultados dos modelos de ML em que se extraíram os atributos manualmente.

### 4.1 PROCEDIMENTOS INICIAIS

A organização sistêmica dos códigos, permitiu trabalhar com os algoritmos em forma de pacotes em Python. Sendo assim, nos arquivos de *notebook*, eram chamadas as funções para executar tarefas pré-definidas. O sistema modular ficou dividido em: módulo de tratamento de dados, módulo de treino, módulo de validação, módulo de visualização.

Seguindo o fluxo de tarefas, após a aquisição e extração dos sinais, resultou da exploração a forma de onda ilustrada na [Figura 4.1.1](#).

Figura 4.1.1 – Curva de um recorte do sinal de EEG canal C3-A2



Fonte – autor.

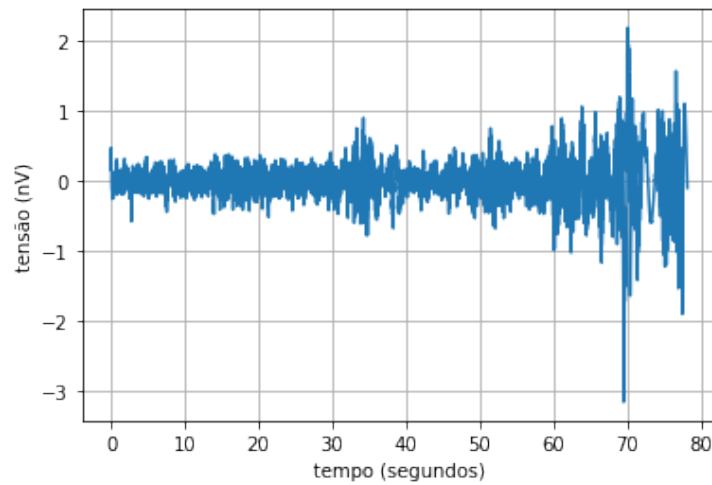
Nota – intervalo do sinal de EEG bruto, retirado do banco de dados, com os valores em nV e o tempo em segundos.

#### 4.1.1 Pré-processamento

Na fase de pré-processamento, como se tratava de dados brutos, foram removidos qualquer ruído do sinal, após passar pelo filtro *Butterworth* passa banda de 0,5 a 70 Hz (veja [subseção 3.4.3](#)). A [Figura 4.1.2](#) é um recorte do sinal de um indivíduo após a filtragem.

Com isso, reduziu-se qualquer componente de ruído branco e de leituras erradas do equipamento que não pertencem à faixa de frequência do EEG.

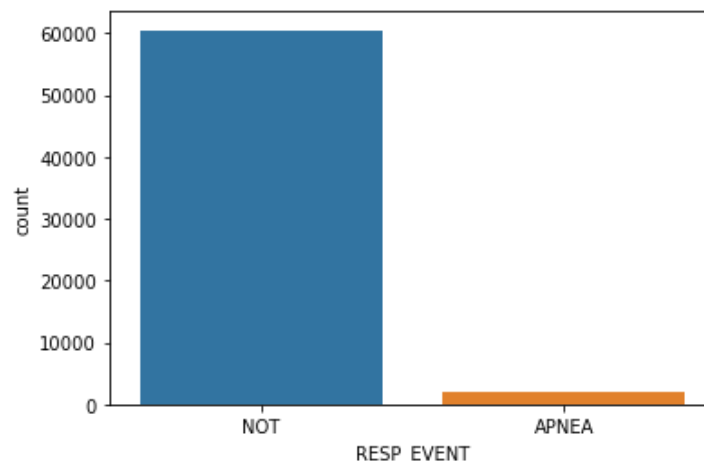
Figura 4.1.2 – Curva de um recorte do sinal filtrado de EEG



Fonte – autor.

Nesta etapa, verificou-se que as classes de eventos estavam desbalanceadas: haviam mais registros para a classe “NOT”, que representa a ausência de apneia, como apresentado na [Figura 4.1.3](#), totalizando em 60.473 épocas, contra 1.945 épocas da classe “APNEA”.

Figura 4.1.3 – Histograma da quantidade de épocas por classe



Fonte – autor.

O desbalanceamento de classes, neste caso, leva aos classificadores criarem modelos mais eficientes para a classe majoritária. Para resolver esse problema, foi realizado o processo de reamostragem dos dados, diminuindo, portanto, a classe majoritária através de uma função que remove as instâncias de forma aleatória com base na distância. O resultado deste processo foi um banco equilibrado, com o mesmo número de instâncias para ambas as classes: 1945 instâncias de cada classe.

Apesar de auxiliar no processo, a remoção de parte dos dados da classe majoritária, pode resultar em um modelo menos acurado, ou seja, que não é tão assertivo para identificar os da-



dos de ambas as classes. Mas, por outro lado, há um ganho em sensibilidade e especificidade, garantindo que um dado tido como verdadeiro terá maior probabilidade de ser verdadeiro e um dado tido como falso terá maior probabilidade de ser, de fato, falso. Essas duas métricas são mais adequadas para este tipo de aplicação, já que se espera resultados confiáveis.

## 4.2 FLUXO DE CLASSIFICAÇÃO DOS SINAIS

Elencam-se os resultados para as etapas de treino e testes, e análise dos classificadores.

### 4.2.1 Treinamentos e testes

Como as ferramentas de ML estavam prontas em bibliotecas escritas em Python, as chamadas de funções foram feitas diretamente nos *notebooks* Python, facilitando o fluxo de treino e testes.

### 4.2.2 Validação

Analisando-se a capacidade do modelo em identificar eventos de apneia obstrutiva do sono, têm-se os resultados descritos na [Tabela 4.2.1](#).

Tabela 4.2.1 – Resultados das métricas atingidas neste estudo

Classificador	Sensibilidade	Especificidade	Precisão	Acurácia	F1-score
<i>AdaBoost</i>	0,7857	0,9167	0,9114	0,8483	0,8439
<i>Gradient Boosting</i>	0,7956	0,9059	0,9048	0,8496	0,8467
<i>K-nearest Neighbors</i>	0,7167	0,9382	0,9268	0,8226	0,8083
<i>Linear SVM</i>	0,7167	0,9543	0,9448	0,8303	0,8151
<i>Logistic Regression</i>	0,7512	0,9220	0,9132	0,8329	0,8243
<i>Naïve Bayes</i>	0,7044	0,9731	0,9662	0,8329	0,8148
<i>Random Forest</i>	0,8177	0,9032	0,8973	0,8560	0,8557

Fonte – autor.

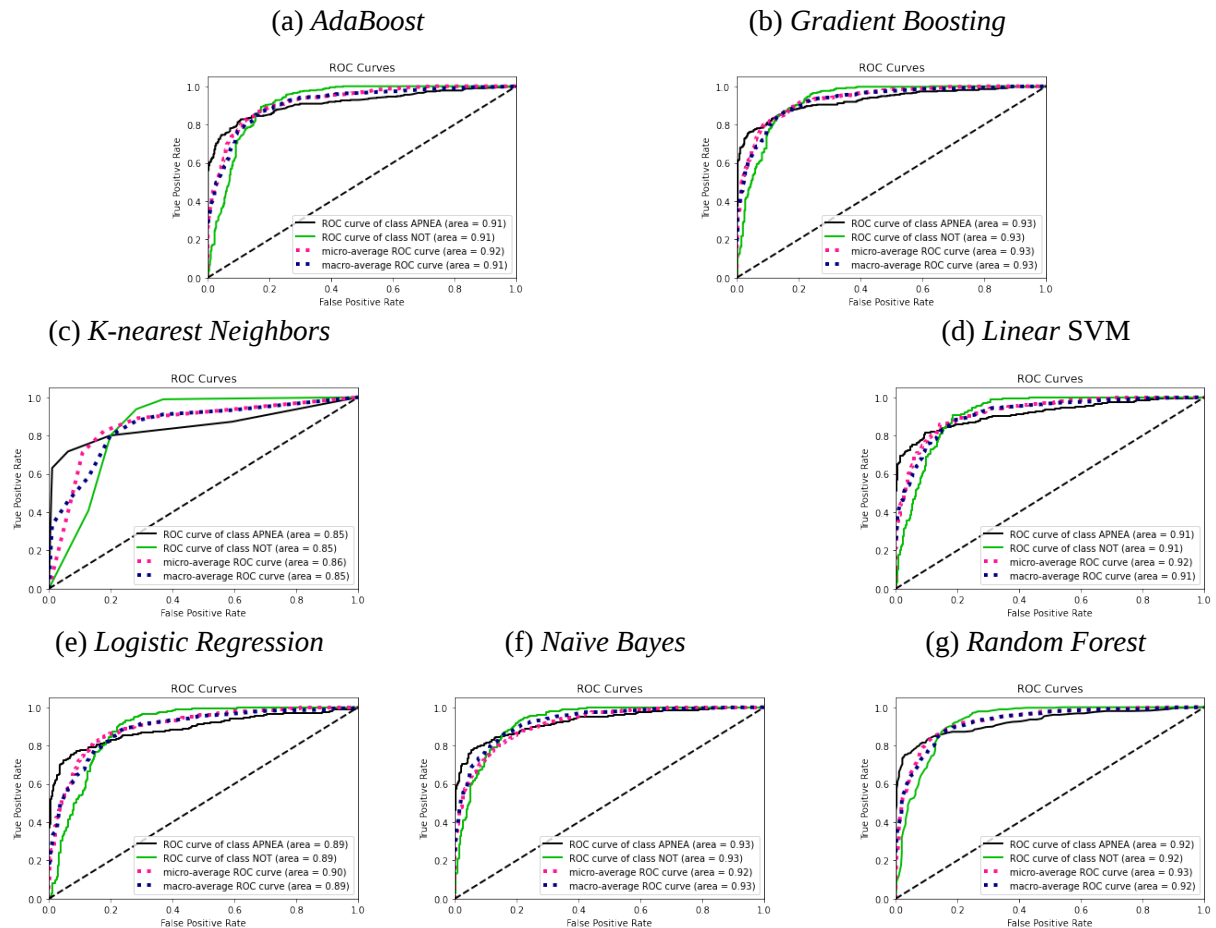
Os classificadores apresentam valores próximos a uma média para cada parâmetro. Por esse aspecto, todos esses algoritmos são consistentes na predição das classes em sinais de apenas dois canais de EEG, conforme os atributos selecionados.

Verifica-se que o modelo *Random Forest* apresentou maior relação entre sensibilidade e especificidade, reforçado pelo valor de *F1-score*, o que é um bom indicativo para avaliar o melhor classificador para a predição das classes de dados. Além disso, obteve a maior sensibilidade, indicando que consegue identificar melhor as classes negativas dentre os algoritmos analisados; e maior acurácia, o que significa que os resultados da predição se aproximam mais do valor verdadeiro, mas isso precisa estar equilibrado com a precisão para garantir resultados com menor variabilidade. No entanto, o modelo *Naïve Bayes* identificou melhor as classes verdadeiras, indicado pelo valor da especificidade; e foi o mais preciso dos classificadores, ou seja, apresentou menor variabilidade dos resultados.

### 4.2.3 Análise qualitativa

Após o registro dos resultados, foi executada a análise qualitativa dos dados de treino e teste. Para cada classificador, montou-se, a curva ROC<sup>1</sup> que se vê na Figura 4.2.1.

Figura 4.2.1 – Curva ROC dos modelos gerados para cada classificador.



Fonte – autor.

Ao diferenciar classificadores, observando seu desempenho pela curva ROC, é preciso dar destaque àqueles que possuem o maior valor AUC e uma curva com uma inclinação inicial mais acentuada, aproximando-se do valor máximo. Quanto mais achatada e com uma inclinação mais próxima à 45°, menor é o desempenho do classificador.

Visualiza-se nos gráficos que o maior valor de AUC é apresentado pelo classificador *Gradient Boosting* e *Naïve Bayes*, seguidos por *Random Forest*, *Linear SVM* e *AdaBoost*, com desempenho semelhante, e por *Logistic Regression*ont. O classificador *K-nearest Neighbors* foi o que apresentou menor desempenho, como é percebido pela sua curva característica.

O tipo de sinal, uma série temporal, e os atributos escolhidos para a geração dos modelos influenciam no desempenho final de cada classificador. Ainda assim, a quantidade de amostras

<sup>1</sup> A curva ROC (do inglês: *Receiver Operating Characteristic*) apresenta graficamente a relação entre a taxa de verdadeiros positivos e a taxa de falsos positivos. Ela é utilizada para analisar o desempenho de classificadores binários.

é significativa para obter melhores resultados, mais ainda, a variabilidade das amostras é também importante para a geração de modelos que representam maior variação dos casos de apneia.

### 4.3 DISCUSSÃO

Os resultados estavam condizentes com a base retirada de estudos da literatura científica, como mostra a

A [Tabela 4.3.1](#) apresenta a comparação de estudos da literatura científica com este trabalho, através das métricas sobre alguns classificadores que apresentaram melhor desempenho.

Tabela 4.3.1 – Comparação entre as métricas atingidas neste estudo contra de outros estudos da literatura científica

Estudo	Classificador	Acurácia	Especificidade	Sensibilidade
A	<i>Naïve Bayes</i>	0,7230	0,5949	0,7895
B	<i>AdaBoost</i> com LDA	0,9290	0,5000	0,9660
C	SVM	0,9510	0,9860	0,9320
D	<i>Random Forest</i>	0,8600	0,8900	–
E	SVM	0,9714	0,9726	0,9701
F	<i>Gradient Boosting</i>	0,8496	0,9059	0,7956
G	<i>Random Forest</i>	0,8560	0,9032	0,8177

Fonte – autor.

Nota – A: [Keshavarz et al. \(2020\)](#). B: [Gutierrez-Tobal et al. \(2019\)](#). C: [Zhou, Wu e Zeng \(2015\)](#). D: [Zhao et al. \(2021\)](#). E: [Almuhammadi, Aboalayon e Faezipour \(2015\)](#). F, G: este trabalho.

Mesmo trabalhando apenas com dois canais de EEG, este trabalho alcançou resultados próximos a trabalhos que utilizavam mais sinais de PSGs. O estudo com o classificador *Random Forest* (G) aproximou-se dos resultados de [Zhao et al. \(2021\)](#) (D) com o mesmo classificador, apesar de não utilizar o método de validação cruzada. Os estudos de [Zhou, Wu e Zeng \(2015\)](#) (C) e [Almuhammadi, Aboalayon e Faezipour \(2015\)](#) (E), com classificador SVM, atingiram os melhores resultados. [Almuhammadi, Aboalayon e Faezipour \(2015\)](#) utilizou, da mesma forma, somente sinais de EEG, porém com canais diferentes e trabalhou mais na seleção de atributos para caracterizar melhor o sinal. O estudo de [Gutierrez-Tobal et al. \(2019\)](#) (B), no que lhe concerne, trabalhou com um arranjo de classificadores e utilizou o sinal de SpO2 de 320 sujeitos. Esses são estudos com resultados a serem alcançados em futuras investigações. Mesmo assim, com as ferramentas e técnicas implementadas neste trabalho, percebe-se que os modelos gerados conseguem, de fato, classificar os sinais quanto a ocorrência de apneia do sono.

Nas condições atuais, no entanto, nenhum modelo é ideal para ser utilizado em análises clínicas. Para isso, é necessário garantir que o classificador seja mais acurado e preciso e, ainda assim, atingir maiores valores de especificidade e sensibilidade.

### 4.3.1 Potencial para aplicação

Uma abordagem diferente do padrão foi utilizada neste trabalho, o uso de somente o sinal de eletroencefalograma para classificação de síndromes de apneia do sono. Para esse diagnóstico, é comum o uso de sinais de ECG ou outros sinais derivados do sistema respiratório e circulatório. Apesar disso, usar unicamente sinais eletrofisiológicos de EEG, coletados em exame de polissonografia, é possível. Necessita-se, porém, o aprimoramento desse estudo partindo para uma investigação mais detalhada.

Para melhorar o processo de classificação, utilizando ainda apenas sinais de EEG, faz-se necessário estudar os sinais de mais canais além dos utilizados aqui. Os canais C3-A2 C4-A1 são algumas das referências que registram sinais no momento do sono; no entanto, há outros canais que podem ser utilizados para refinar o processo e, possivelmente, adicionar atributos mais relevantes de outros canais. Utilizar mais sinais no processo, seria outra alternativa, visto que o exame de polissonografia já abrange uma gama considerável deles. Apesar disso, pensar em uma aplicação mais acessível e rápida requer pensar em utilizar apenas sinais necessários, que representem melhor os eventos e tenham boa qualidade para processá-los. Portanto, mantendo o estudo focado somente em sinais de EEG é importante para alcançar esses requisitos.

Ainda assim, novos algoritmos podem ser testados. Saindo da área de ML, há diversos algoritmos de DL que realizam o trabalho de predição em menor tempo e, no processo de treino, extraem os atributos de forma implícita, evitando a seleção manual dos mesmos. Outra ideia é a extração do espectrograma do sinal de EEG, técnica conhecida como *scaleogram*. A imagem espectrográfica seria, então, a entrada para uma rede neural. Dessa forma, os atributos do sinal de EEG estariam representados pelo *scaleogram* implicitamente, já que ele extrai as características de tempo e frequência do sinal. Embora sejam aplicações que demandam mais recursos computacionais, essas técnicas de DL podem ser aplicadas em dispositivos dedicados ao DL. Atualmente, há uma gama de dispositivos embarcados para desenvolvimento e prototipagem de modelos de IA.

Os achados desse trabalho podem ser utilizados para conduzir estudos com outros bancos de dados, como do HUSM, por exemplo. Dessa forma, dar-se-a continuidade à implementação e descoberta de métodos em IA que auxiliem no diagnóstico das síndromes de apneia do sono. Ainda assim, pensando em uma aplicação prática, sugere-se que se estude implantações do sistema em dispositivos de sistemas embarcados, para inovação em análises clínicas. A título de exemplo, pode ser testada a implantação de modelos para processamento *offline* e aplicação da coleta de dados por telemedicina.

Ainda, algumas hipóteses para serem investigadas em estudos futuros são: quais seriam os melhores atributos para maximizar o desempenho do modelo? Quais outros eletrodos caracterizam melhor o evento de apneia do sono?

## 5 CONCLUSÃO

Este trabalho apresentou o estudo de diferentes abordagens para classificação de apneia do sono com sinais de EEG, utilizando ML. O estudo mostrou que prever os eventos de apneia do sono com algoritmos de ML, apenas em sinais de EEG, é possível. Assim sendo, a metodologia aplicada trouxe desempenhos próximos aos trabalhos que utilizavam diferentes sinais fisiológicos do paciente. No entanto, exige-se mais investigações sobre este método, promovendo um estudo para sua implantação em sistemas de monitoramento do sono.

### 5.1 SUGESTÕES DE TRABALHOS FUTUROS

Sugere-se, como continuidade deste trabalho, o aprofundamento do estudo dessa técnica para, futuramente, planejar a implantação em dispositivos embarcados dedicados a ML. Algumas tarefas que podem ser executadas em novos estudos são: trabalhar com outros bancos de dados, testar outros algoritmos de amostragem dos dados, testar outros algoritmos de ML ou DL, e investigar os sinais de EEG de outros eletrodos.

Dessa forma, abre-se a possibilidade do uso deste trabalho como base para desenvolvimento de soluções que auxiliem profissionais da área médica em exames e estudos de apneias e outros distúrbios do sono.

## REFERÊNCIAS

- ALMUHAMMADI, W. S.; ABOALAYON, K. A. I.; FAEZIPOUR, M. Efficient obstructive sleep apnea classification based on eeg signals. In: **2015 IEEE Long Island Systems, Applications and Technology Conference (LISAT)**. IEEE, 2015. Acessado em 17 set 2021. Disponível em: <<http://dx.doi.org/10.1109/lisat.2015.7160186>>.
- BEZERRA, C. S. **Uma abordagem de segmentação semântica de íris para fins biométricos usando aprendizagem profunda**. 80 f. Dissertação (Mestrado em Ciências) — Universidade Federal do Paraná, Curitiba, 2018.
- BISWAL, S. *et al.* Expert-level sleep scoring with deep neural networks. **Journal of the American Medical Informatics Association**, v. 25, n. 12, p. 1643–1650, 11 2018. ISSN 1527-974X. Acessado em 17 set 2021. Disponível em: <<https://doi.org/10.1093/jamia/ocy131>>.
- CARDIORESPIRA CARDIOPULMONAR. **Dicionário Cardiopulmonar para todos: Polissonografia**. 2013. Acessado em: 21 set 2021. Disponível em: <<https://cardiorespira.blogspot.com/2013/09/dicionario-cardiopulmonar-para-todos.html>>.
- CEN, L. *et al.* Automatic system for obstructive sleep apnea events detection using convolutional neural network. In: **2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)**. IEEE, 2018. Acessado em 17 set 2021. Disponível em: <<http://dx.doi.org/10.1109/embc.2018.8513363>>.
- CENTER FOR DISEASE CONTROL AND PREVENTION. **CDC - Data and Statistics - Sleep and Sleep Disorders**. 2017. Acesso em: 21 set 2021. Disponível em: <[https://www.cdc.gov/sleep/data\\_statistics.html](https://www.cdc.gov/sleep/data_statistics.html)>.
- GUTIERREZ-TOBAL, G. C. *et al.* Evaluation of machine-learning approaches to estimate sleep apnea severity from at-home oximetry recordings. **IEEE Journal of Biomedical and Health Informatics**, Institute of Electrical and Electronics Engineers (IEEE), v. 23, p. 882–892, 3 2019. Acessado em 17 set 2021. Disponível em: <<http://dx.doi.org/10.1109/jbhi.2018.2823384>>.
- HENEGHAN, C. **St. Vincent’s University Hospital / University College Dublin Sleep Apnea**. PhysioNet, 2021. Acessado em 17 set 2021. Disponível em: <<https://archive.physionet.org/pn3/ucddb/>>.
- INSTITUTO DE ENGENHARIA BIOMÉDICA DA UFSC. **História**: Ieb-ufsc. 2011. Acesso em: 17 set. 2021. Disponível em: <[http://www.ieb.ufsc.br/?page\\_id=52](http://www.ieb.ufsc.br/?page_id=52)>.
- INSTITUTO DO SONO DA ASA NORTE. **Apneia e hipopneia obstrutiva do sono**. 2021. Acesso em 08 set. 2021. Disponível em: <<https://isan.med.br/web/doencas-do-sono/apneia-e-hipopneia-obstrutiva-do-sono/>>.
- JASPER, H. H. Report of the committee on methods of clinical examination in electroencephalography: 1957. **Electroencephalography and Clinical Neurophysiology**, v. 10, n. 2, p. 370–375, 1958. ISSN 0013-4694. Acesso em: 17 set. 2021. Disponível em: <<https://www.sciencedirect.com/science/article/pii/0013469458900531>>.

- KESHAVARZ, Z. *et al.* Obstructive sleep apnea: A prediction model using supervised machine learning method. **Studies in health technology and informatics**, v. 272, p. 387–390, 2020. ISSN 0926-9630. Acessado em 17 set 2021. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/32604683/>>.
- KONAVI, R.; FOSTER, P. **Glossary of terms**: Special issue on application of machine learning and the knowledge discovery process. Kluwer Academic Publishers, 1998. Acesso em: 24 fev. 2022. Disponível em: <<https://ai.stanford.edu/~ronnyk/glossary.html>>.
- MAYO CLINIC. **Sleep apnea**: Symptoms and causes. 2021. Acesso em 08 set. 2021. Disponível em: <<https://www.mayoclinic.org/diseases-conditions/sleep-apnea/symptoms-causes/syc-20377631>>.
- MCCARTY, J. What is artificial intelligence? Stanford University, 2007. Acesso em: 17 set. 2021. Disponível em: <<http://jmc.stanford.edu/articles/whatisai.html>>.
- MOHRI, M.; ROSTAMIZADEH, A.; TALWALKAR, A. **Foundations of Machine Learning**. MIT press, 2012. 414 p. Acesso em: 24 fev. 2022. Disponível em: <<https://cs.nyu.edu/~mohri/mlbook/>>.
- MOSTAFA, S. S. *et al.* A systematic review of detecting sleep apnea using deep learning. **Sensors**, MDPI AG, v. 19, p. 4934, 11 2019. Acesso em: 17 set. 2021. Disponível em: <<http://dx.doi.org/10.3390/s19224934>>.
- NIEDERMEYER, E.; SCHOMER, D. L.; HENRIQUE, L. d. S. F. *Niedermeyers electroencephalography: basic principles, clinical applications, and related fields*. Lippincott William & Wilkins, Reino Unido, p. 1309, 2011.
- SAMUEL, A. L. Some studies in machine learning using the game of checkers. **IBM Journal of Research and Development**, v. 3, n. 3, p. 210–229, 1959. Acesso em: 17 set. 2021. Disponível em: <<https://ieeexplore.ieee.org/document/5392560>>.
- THEORELL-HAGLÖW, J. *et al.* Gender differences in obstructive sleep apnoea, insomnia and restless legs syndrome in adults - what do we know? a clinical update. **Sleep Med Rev**, p. 28–38, 2018. Acesso em: 17 set. 2021. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/28495359/>>.
- VANHATALO, S.; VOIPIO, J.; KAILA, K. Full-band eeg (fbeeeg): an emerging standard in electroencephalography. **Clinical Neurophysiology**, Elsevier, v. 116, n. 1, p. 1–8, 2005. Acesso em: 17 fev. 2022. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/15589176/>>.
- ZHAO, X. *et al.* Classification of sleep apnea based on eeg sub-band signal characteristics. **Scientific Reports**, Springer Science and Business Media LLC, v. 11, 12 2021. Acesso em: 17 set. 2021. Disponível em: <<http://dx.doi.org/10.1038/s41598-021-85138-0>>.
- ZHOU, J.; WU, X.; ZENG, W. Automatic detection of sleep apnea based on eeg detrended fluctuation analysis and support vector machine. **Journal of Clinical Monitoring and Computing**, Springer Science and Business Media LLC, v. 29, p. 767–772, 12 2015. Acesso em: 17 set. 2021. Disponível em: <<http://dx.doi.org/10.1007/s10877-015-9664-0>>.

## APÊNDICE A – PESQUISA DE TRABALHOS RELACIONADOS

Pesquisa de trabalhos relacionados publicados em artigos, bases de dados e como trabalhos amadores. Foram selecionados os trabalhos mais relevantes e compatíveis com o tema do projeto. O [Quadro A.1](#) apresenta os dados bibliométricos de pesquisa de termos chave.

### A.1 SUMÁRIO

Quadro A.1 – Relação de resultados por plataforma e termo de busca utilizado

Plataforma	Termo de busca	Quantidade
Pubmed	(apnea) AND (machine learning)	222
Pubmed	((apnea) AND (machine learning)) AND (eeg)	23
Nature	((apnea) AND (machine learning)) AND (eeg)	48
Google Scholar	apnea "machine learning"eeg	3680
Physionet	apnea eeg	17
GitHub	apnea eeg	2
GitHub	apnea machine learning	10
Kaggle	apnea	53
Zenodo	apnea AND machine learning AND eeg	1
Zenodo	apnea eeg	16

Fonte – autor.

### A.2 PLATAFORMA: PUBMED

**Termo de busca:** (apnea) AND (machine learning)

222 resultados

1. [Application of machine learning to predict obstructive sleep apnea syndrome severity](#)
2. [Obstructive Sleep Apnea: A Prediction Model Using Supervised Machine Learning Method](#)
3. [The faces of sleep apnea in the age of machine learning](#)
4. [Predicting Nondiagnostic Home Sleep Apnea Tests Using Machine Learning](#)
5. [Objective Relationship Between Sleep Apnea and Frequency of Snoring Assessed by Machine Learning](#)
6. [Phenotypes in obstructive sleep apnea: A definition, examples and evolution of approaches](#)
7. [Evaluation of Machine Learning Approaches to Estimate Sleep Apnea Severity From At-Home Oximetry Recordings](#)



8. Obstructive sleep apnea syndrome detection based on ballistocardiogram via machine learning approach
9. Diagnosis of Obstructive Sleep Apnea during Wakefulness Using Upper Airway Negative Pressure and Machine Learning
10. Can Statistical Machine Learning Algorithms Help for Classification of Obstructive Sleep Apnea Severity to Optimal Utilization of Polysomnography Resources?
11. Machine Learning for automatic identification of thoracoabdominal asynchrony in children
12. A Systematic Review of Detecting Sleep Apnea Using Machine Learning

**Termo de busca:** ((apnea) AND (machine learning)) AND (eeg)

23 resultados

1. Automatic detection of sleep apnea based on EEG detrended fluctuation analysis and support vector machine
2. An Intelligent Sleep Apnea Classification System Based on EEG Signals
3. Multi-channel EEG-based sleep stage classification with joint collaborative representation and multiple kernel learning
4. Sleep-wake stage detection with single channel ECG and hybrid machine learning model in patients with obstructive sleep apnea
5. A deep learning-based decision support system for diagnosis of OSAS using PTT signals
6. Accurate Deep Learning-Based Sleep Staging in a Clinical Population With Suspected Obstructive Sleep Apnea
7. A human amygdala site that inhibits respiration and elicits apnea in pediatric epilepsy
8. Expert-level sleep scoring with deep neural network
9. Sleep staging from electrocardiography and respiration with deep learning
10. Large-Scale Automated Sleep Staging
11. A Novel Multi-Class EEG-Based Sleep Stage Classification System
12. Deep learning enables sleep staging from photoplethysmogram for patients with suspected sleep apnea
13. Orthogonal convolutional neural network for automatic sleep stage classification based on single-channel EEG

14. Feature relevance in physiological networks for classification of obstructive sleep apnea
15. Automatic Detection of Respiratory Effort Related Arousals With Deep Neural Networks From Polysomnographic Recordings
16. Single-channel oximetry monitor versus in-lab polysomnography oximetry analysis: does it make a difference?
17. Data-Driven Multimodal Sleep Apnea Events Detection: Synchrosqueezing Transform Processing and Riemannian Geometry Classification Approaches
18. Automatic, electrocardiographic-based detection of autonomic arousals and their association with cortical arousals, leg movements, and respiratory events in sleep
19. A two-step automatic sleep stage classification method with dubious range detection
20. Automatic sleep staging: from young adults to elderly patients using multi-class support vector machine
21. Validation of overnight oximetry to diagnose patients with moderate to severe obstructive sleep apnea
22. Analysis and visualization of sleep stages based on deep neural networks

Relacionados:

1. Classification of sleep apnea based on EEG sub-band signal characteristics

### A.3 PLATAFORMA: NATURE

**Termo de busca:** ((apnea) AND (machine learning)) AND (eeg)

48 resultados

1. Classification of sleep apnea based on EEG sub-band signal characteristics
2. Deep learning for automated sleep staging using instantaneous heart rate
3. Contactless recording of sleep apnea and periodic leg movements by nocturnal 3-D-video and subsequent visual perceptive computing
4. Neural network analysis of sleep stages enables efficient diagnosis of narcolepsy
5. DeepSleep convolutional neural network allows accurate and fast detection of sleep arousal
6. Accurate whole-night sleep monitoring with dry-contact ear-EEG

7. Sleep stage classification from heart-rate variability using long short-term memory neural networks
8. Brain wave classification using long short-term memory network based OPTICAL predictor
9. Developing and adopting safe and effective digital biomarkers to improve patient outcomes

#### A.4 PLATAFORMA: GOOGLE SCHOLAR

**Termo de busca:** apnea "machine learning" eeg  
3680 resultados

1. Efficient obstructive sleep apnea classification based on EEG signals
2. An intelligent sleep apnea classification system based on EEG signals
3. Model comparison for the detection of EEG arousals in sleep apnea patients
4. Classification of sleep apnea based on EEG sub-band signal characteristics
5. An efficient method to detect sleep hypopnea-apnea events based on EEG signals
6. Automatic detection of sleep apnea based on EEG detrended fluctuation analysis and support vector machine
7. A comparison of different machine learning algorithms using single channel EEG signal for classifying human sleep stages
8. Detection of Sleep Apnea Based on the Analysis of Sleep Stages Data Using Channel EEG
9. Detection of Excessive Daytime Sleepiness in Resting-Stage EEG Recordings..
10. Combining machine learning models for the automatic detection of EEG arousals
11. Feature Extraction and Detection of Obstructive Sleep Apnea from Raw EEG Signal
12. Automatic Classification of Sleep Apnea Type and Severity using EEG Signals
13. An Automated System for Diagnosis of Sleep Apnea Syndrome Using Single-Channel EEG Signal
14. Feature relevance in physiological networks for classification of obstructive sleep apnea
15. An intelligent sleep apnea detection system

16. [A multi-scale parallel convolutional neural network for automatic sleep apnea detection using single-channel EEG signals](#)
17. [Estimating daytime sleepiness with previous night EEG, EOG and EMG spectrograms in patients with suspected sleep apnea using a convolutional neural network](#)
18. [A reliable algorithm based on combination of EMG, ECG and EEG signals for sleep apnea detection](#)
19. [Development of a support vector machine learning and smart phone Internet of Things-based architecture for real-time sleep apnea diagnosis](#)
20. [EEG signal analysis of patients with obstructive sleep apnea syndrome using spectrum and fuzzy entropy](#)
21. [Automatic Detection of Sleep Apnea Using Sub-Band Features from EEG Signals](#)
22. [Detection of sleep apnea from single-channel EEG using an explainable CNN](#)

#### A.5 PLATAFORMA: PHYSIONET

**Termo de busca:** apnea eeg  
17 resultados

1. [CAP Sleep Database](#)
2. [Haaglanden Medisch Centrum sleep staging datase](#)
3. [Sleep-EDF Databse Expanded](#)
4. [Sleep-EDF Database](#)
5. [NCH Sleep DataBank](#)
6. [MIT-BIH Polysomnographic Database](#)
7. [Siena Scalp EEG Database](#)

#### A.6 PATAFORMA: GITHUB

**Termo de busca:** apnea eeg  
2 resultados

1. [luciaquirke/sleep-apnea](#)
2. [chamikasudusinghe/apnea](#)

**Termo de busca:** [apnea machine learning](#)

10 resultados

1. [lakshmanmakkidi/SleepApneaDetection](#)
2. [jfrjunio/OSAML](#)
3. [Naman-18/SleepApnea\\_Application](#)
4. [berkott/SciFairPPG](#)
5. [jacobstac/Bachelors-Thesis-Detection-of-Sleep-Apnea](#)
6. [haniyacoub/Prediciting-Apnea-](#)
7. [jgarnierUQ/Thesis](#)
8. [jrinconcol/Machine-learning-model-to-predict-OSA](#)

#### A.7 PLATAFORMA: KAGGLE

**Termo de busca:** [apnea](#)

53 resultados

*Notebooks:*

1. [Read Apnea Dat Files](#)
2. [Sleep Apnea Notebook](#)
3. [Apnea ECG](#)
4. [Read Apnea Dat Files - Rachael](#)
5. [Sleep Apnea cases](#)
6. [kernel3aa2b96a53](#)

*Databases:*

1. [The Apnea-ECG database](#)
2. [Sleep EDF and Apnea](#)

## A.8 PLATAFORMA: ZENODO

**Termo de busca:** apnea AND machine learning AND eeg  
1 resultado

1. [Apnea-ECG Database](#)
2. [The Apnea-ECG database](#)
3. [Obstructive Sleep Apnea Prediction Data](#)
4. [Nonlinear dynamics forecasting of obstructive sleep apnea onsets](#)
5. [Detecting and Quantifying Apnea Based on the ECG](#)
6. [APNEA HRV DATASET](#)
7. [Apnea Detection from the ECG](#)
8. [St. Vincent's University Hospital/UCB Sleep Apnea Database](#)
9. [Sleep EDF and Apnea](#)
10. [MIT-BIH Polysomnography Database](#)
11. [Obstructive sleep apnea in adults with Down syndrome](#)
12. [Evaluation of Potential Biomarkers for Obstructive Sleep Apnea ..](#)
13. [Sleep Apnea](#)
14. [Heart Biomarker Evaluatoin in Apnea Treatment](#)

**Termo de busca:** apnea eeg  
16 resultados

1. [Obstructive Sleep Apnea Prediction Data](#)
2. [The DREAMS Databases and Assessment Algorithm](#)
3. [Nonlinear dynamics forecasting of obstructive sleep apnea onsets](#)
4. [MIT-BIH Polysomnography Database](#)
5. [Apnea Recognition with Wavelet Neural Networks](#)
6. [Role of Spectral Power and Sample Entropy of the EEG in Restoration of Stable Breathing from Respiratory Events in Sleep Apnea Hypopnea Syndrome](#)
7. [St. Vincent's University Hospital/UCB Sleep Apnea Database](#)

8. APNEA HRV DATASET
9. Sleep-EDF Database
10. Sleep Heart Health Study PSG Database

## APÊNDICE B – INFORMAÇÕES DOS INDIVÍDUOS NO BANCO DE DADOS

Quadro B.1 – Principais informações de cada indivíduo registrado no banco de dados do estudo

Número do estudo	Peso (kg)	Sexo	IMC (kg/m <sup>2</sup> )	Idade (anos)	Duração do estudo (horas)	Eficiência do sono (%)	Duração da PSG (segundos)
ucddb002	100,3	M	33,9	54	6,2	84	22470
ucddb003	102,0	M	31,8	48	7,3	81	26478
ucddb005	100,4	M	32,4	65	6,9	63	24798
ucddb006	103,5	M	30,2	52	6,7	89	24267
ucddb007	84,0	M	25,1	47	6,8	90	24405
ucddb008	59,8	F	28,4	63	6,4	64	23041
ucddb009	101,5	M	31,3	52	7,7	80	27759
ucddb010	119,0	M	39,3	38	7,6	92	27211
ucddb011	101,0	M	28,6	51	7,5	60	27030
ucddb012	97,5	M	30,4	51	7,2	85	25941
ucddb013	80,0	F	34,2	62	6,8	61	24333
ucddb014	91,0	M	29,0	56	6,4	79	23239
ucddb015	83,9	M	29,0	28	7,6	77	27488
ucddb017	117,0	M	37,8	53	6,6	87	23684
ucddb018	77,0	M	26,3	35	6,8	60	24685
ucddb019	97,8	M	30,9	49	7,1	92	25573
ucddb020	108,8	M	34,0	52	6,3	78	22586
ucddb021	87,0	F	33,6	41	7,6	82	27409
ucddb022	80,7	M	29,3	34	6,6	58	23640
ucddb023	89,0	F	32,7	68	7,2	67	25850
ucddb024	99,9	M	33,8	54	7,6	83	27250
ucddb025	128,6	M	42,5	52	5,9	77	21350
ucddb026	84,0	M	27,4	49	7,0	87	25160
ucddb027	93,0	M	28,1	45	7,4	86	26791
ucddb028	88,9	M	30,1	50	6,0	68	21660

Fonte – Adaptação de [Heneghan \(2021\)](#).