

DETECÇÃO DE DISTÚRBIOS DO SONO EM DADOS DE POLISSONOGRAFIA
ATRAVÉS DE APRENDIZADO DE MÁQUINA

Thiago Barral Fernandes Reis

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Ciência da Computação, Centro Federal de Educação Tecnológica Celso Suckow da Fonseca CEFET/RJ, como parte dos requisitos necessários à obtenção do grau de mestre.

Orientadores:
Felipe da Rocha Henriques
Michel Pompeu Tcheou

Detecção de Distúrbios do Sono em Dados de Polissonografia Através de Aprendizado de Máquina

Dissertação de Mestrado em Ciência da Computação, Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, CEFET/RJ.

Thiago Barral Fernandes Reis

Aprovada por:

Presidente, Professor D.Sc. Felipe da Rocha Henriques (CEFET/RJ)

Professor D.Sc. Michel Pompeu Tcheou (UERJ)

Professora D.Sc. Laura Silva de Assis (CEFET/RJ)

Professor D.Sc. Tadeu Nagashima Ferreira (UFF)

Rio de Janeiro,
6 de Junho de 2024

Ficha catalográfica elaborada pela Biblioteca Central do CEFET/RJ

R375 Reis, Thiago Barral Fernandes
Detecção de distúrbios do sono em dados de polissonografia
através de aprendizado de máquina / Thiago Barral Fernandes
Reis. — 2024.
52f. : il. color. , enc.

Dissertação (Mestrado) Centro Federal de Educação
Tecnológica Celso Suckow da Fonseca, 2024.

Bibliografia : f. 48-52

Orientador: Felipe da Rocha Henriques

Coorientador: Michel Pompeu Tcheou

1. Polissonografia. 2. Distúrbio do sono. 3. Internet das coisas.
4. Aprendizado do computador. I. Henriques, Felipe da Rocha.
(Orient.). II. Tcheou, Michel Pompeu. (Coorient.). III. Título.

CDD 616.849

Elaborada pela bibliotecária Tania Mello – CRB/7 nº 5507/04

DEDICATÓRIA

Dedico esse trabalho aos meus pais, pois hoje com dois lindos filhos vejo o tamanho da responsabilidade e da entrega. Dedico também aos meus filhos que possam ver nesse trabalho uma motivação para sempre buscarem educação e conhecimento. E por fim, mas não menos importante, dedico também a minha esposa que me apoiou e sempre esteve ao meu lado.

AGRADECIMENTOS

Agradeço aos meus orientadores Felipe da Rocha Henriques e Michel Pompeu Tcheou, e a todo corpo docente e discente. Por fim, agradeço ao órgão de fomento desta pesquisa.

RESUMO

Detecção de Distúrbios do Sono em Dados de Polissonografia Através de Aprendizado de Máquina

Thiago Barral Fernandes Reis

Orientadores:

Felipe da Rocha Henriques

Resumo da Dissertação submetida ao Programa de Pós-graduação em Ciência da Computação do Centro Federal de Educação Tecnológica Celso Suckow da Fonseca CEFET/RJ como parte dos requisitos necessários à obtenção do grau de mestre.

A qualidade do sono é crucial para a recuperação das funções biológicas, neurológicas e psicológicas. Indivíduos privados de sono podem sofrer de diversos impactos negativos na saúde, sendo a apneia do sono um dos distúrbios mais comuns a perturbar o sono. A polissonografia é o principal método de diagnósticos de distúrbios do sono e é realizado em clínicas especializadas. Com a evolução Internet das Coisas, dispositivos vestíveis tornaram-se capazes de coletar dados para a análise do sono, fomentando novas perspectivas no campo de saúde e da ciência da computação. Este estudo se propõe a então a explorar a aplicação de aprendizado de máquina na detecção de distúrbios do sono, auxiliando no diagnóstico da polissonografia. Dessa forma quatro algoritmos de aprendizado de máquina foram avaliados com os dados provenientes de uma base de dados reais de polissonografia, advindos de treze sinais fisiológicos. Além disso, também investigou-se o impacto da redução do número de sinais no desempenho dos algoritmos, onde o objetivo era avaliar a possível utilização de dispositivos vestíveis para captação desses sinais. Os resultados apontaram o algoritmo *Random Forest* como o mais promissor, apresentando performance satisfatória mesmo com o uso de um número reduzido de sinais.

Palavras-chave:

Polissonografia; Distúrbio do sono; Análise de dados; Internet das Coisas; Aprendizado de Máquina.

Rio de Janeiro,

6 de Junho de 2024

ABSTRACT

Detecting Sleep Disorders in Polysomnography Data Through Machine Learning

Thiago Barral Fernandes Reis

Advisors:

Felipe da Rocha Henriques

Abstract of dissertation submitted to Programa de Pós-graduação em Ciência da Computação - Centro Federal de Educação Tecnológica Celso Suckow da Fonseca CEFET/RJ as partial fulfillment of the requirements for the degree of master.

Sleep quality is crucial for the recovery of physiological, neurological and biological functions. Individuals deprived of sleep may suffer from various negative health impacts, with sleep apnea being one of the most common disorders that disturb the sleep. Polysomnography is the primary diagnostic method for sleep disorders and is performed in a specialized clinics. With the evolution of the Internet of Things, wearable devices have become capable of collecting data for sleep analysis, creating new perspectives in the field of health and computer science. This study aims to explore the application of machine learning in the detection of sleep disorders, assisting in the polysomnography diagnosis. In this way, four machine learning algorithms were evaluated using data from a real polysomnography database, derived from thirteen physiological signals. Furthermore, the impact of reducing the number of signals on the algorithms performance was also analyzed, with the purpose of assessing the possible of use wearable devices for capturing these signals. The results pointed to the Random Forest algorithm as the most promising algorithm, showing satisfactory performance even with a reduced number of signals.

Key-words:

Polysomnography; Sleep disturb; Data analytics; Internet of Things; Machine learning.

Rio de Janeiro,

6 de Junho de 2024

Sumário

I	Introdução	12
I.1	Ciclos do Sono	12
I.2	Privação do Sono	14
I.3	Objetivos e contribuições	17
I.4	Estrutura	18
II	Fundamentação Teórica e Trabalhos Relacionados	19
II.1	Polissonografia	19
II.2	Algoritmos de Aprendizado de Máquina	21
II.2.1	<i>K-Nearest Neighbors</i> (KNN)	22
II.2.2	<i>Support Vector Machine</i> (SVM)	24
II.2.3	Árvore de Decisão e <i>Random Forest</i>	27
II.3	<i>Pipeline</i> para Implementação do Modelo	27
II.3.1	Internet das Coisas	28
II.3.2	Armazenamento dos dados	29
II.3.3	Processamento dos dados	30
II.3.4	Formulação dos Resultados	31
II.4	Trabalhos Relacionados	31
III	Metodologia	34
III.1	Base de Dados e Pré-Processamento	34
III.2	Métricas para Avaliação dos Modelos	37
IV	Resultados	40
IV.1	Avaliação dos modelos de aprendizado de máquina com todos os atributos	40
IV.2	Avaliação do <i>Random Forest</i> após a otimização dos hiperparâmetros	43
IV.3	Avaliação dos modelos de aprendizado de máquina com menos atributos	43
IV.4	Avaliação do <i>Random Forest</i> após a otimização dos hiperparâmetros, considerando menos atributos	45

V Conclusões	46
V.1 Futuras Direções	46
Referências	48

Lista de Figuras

I.1	A arquitetura do sono e suas fases. Fonte: Walker, 2018	13
I.2	Relação entre a média de sono e a chance de lesão em esportes. Fonte: Walker, 2018	14
I.3	Relação entre a perda de sono e o aumento do risco de colisão de veículos. Fonte: Walker, 2018	15
I.4	Relação entre a perda de sono e obesidade. Fonte: Walker, 2018	15
II.1	Esquema do aprendizado supervisionado.	21
II.2	Esquema do aprendizado não supervisionado.	22
II.3	Ilustração do método <i>K-Nearest Neighbors</i> (KNN). Fonte: Adaptado de Wang et al., 2020.	23
II.4	Ilustração do hiperplano gerado pelo algoritmo <i>Support Vector Machine</i> (SVM). Fonte: Adaptado de Souza, 2011.	25
II.5	Arquitetura do fluxo dos dados na <i>Microsoft Azure</i> .	29
II.6	Exemplo de arquitetura de Internet das Coisas para comunicação de dados biométricos de pacientes.	30
III.1	Visualização de exemplo dos sinais fisiológicos da base de dados de estudo.	35
III.2	Fluxo da metodologia proposta neste trabalho.	38
IV.1	Gráficos do teste estatístico <i>Kolmogorov-Smirnov</i> (KS) para o algoritmo KNN, a esquerda, e para o algoritmo <i>Random Forest</i> , a direita, utilizando todos os atributos dos dados (todos os sinais do Polissonografia (PSG)).	41
IV.2	Curva Receiver Operating Characteristic (ROC) para o modelo <i>Random Forest</i> utilizando todos os atributos.	42
IV.3	Curva ROC para o modelo <i>Random Forest</i> com menos <i>features</i> .	44
IV.4	Gráficos do teste estatístico KS para o algoritmo KNN, a esquerda, e para o algoritmo <i>Random Forest</i> , a direita, utilizando menos atributos.	45

Lista de Tabelas

II.1 Níveis de realização da PSG e os parâmetros monitorados.	19
II.2 Métricas extraídas da leitura dos dados oriundos da PSG.	20
II.3 Formulação matemática das funções <i>kernel</i> .	26
III.1 Valores Mínimo, Máximo, Médias e Desvios Padrão dos sinais do exame de PSG.	36
IV.1 Hiperparâmetros dos modelos.	41
IV.2 Avaliação dos modelos em relação aos dados de teste.	41
IV.3 Resultados numéricos do teste estatístico KS para os quatro algoritmos avaliados, considerando todos os atributos.	42
IV.4 Resultados do algoritmo <i>Random Forest</i> , após o processo de <i>tuning</i> dos hiperparâmetros.	43
IV.5 Avaliação dos modelos, treinados com menos atributos.	44
IV.6 Resultados numéricos do teste estatístico KS para os quatro algoritmos avaliados, considerando menos atributos.	45

Lista de Abreviações

API	<i>Application Programming Interface</i>	30
AUPRC	<i>Area Under the Precision-Recall Curve</i>	31, 32, 37, 39
AUROC	Area Under The Receiver Operating Characteristic Curve	39, 42, 44
CCNL	Clinical And Cognitive Neuroscience Laboratory	34, 37
CDAC	Laboratório Da Clínica De <i>Data Animation</i>	34
CITS	Classificação Internacional Dos Transtornos Do Sono	16
CNN	<i>Convolutional Neural Network</i>	32
DFA	Detrended Fluctuation Analysis	32
ECG	Eletrocardiograma	17, 19, 31, 33, 34, 43
EEG	Eletroencefalograma	17, 19, 31, 32, 33, 34
ELM	<i>Extreme Learning Machine</i>	33
EMG	Eletromiograma	17, 19, 31, 32, 34
EOG	Eletrooculograma	17, 19, 31, 34
FC	Frequência Cardíaca	19, 20
FN	Falso Negativo	38, 39, 43
FP	Falso Positivo	38, 39, 43
HMM	<i>Hidden Markov Model</i>	32
IOT	<i>Internet of Things</i>	20, 28
KNN	<i>K-Nearest Neighbors</i>	22, 23, 24, 26, 37, 40, 41, 42, 44, 45, 46
KS	<i>Kolmogorov–Smirnov</i>	40, 41, 42, 44, 45
LCP	<i>Laboratory for Computational Physiology</i>	34
LS	Latência Do Sono	20
LSTM	<i>Long Short Term Memory</i>	32
MGH	<i>Massachusetts General Hospital</i>	34, 37
MIT	<i>Massachusetts Institute of Technology</i>	34
MP	Movimento De Pernas	19
NFE	National Sleep Foundation	13
NREM	<i>Not Rapid Eyes Movement</i>	12
OMS	Organização Mundial Da Saúde	14

PSG	Polissonografia	16, 17, 18, 19, 20, 21, 28, 30, 32, 33, 34, 36, 37, 41, 46, 47
RBF	<i>Radial Basis Function</i>	25, 26
REM	<i>Rapid Eyes Movement</i>	12, 13, 20
RESNETS	<i>Residuals Networks</i>	32
ROC	Receiver Operating Characteristic	39, 42, 44
SAS	Sleep Apnea Syndrome	32, 33
SE	<i>Sleep efficiency</i>	13
SVM	<i>Support Vector Machine</i>	24, 25, 26, 32, 33, 37, 40, 42, 44, 45, 46
TNC	Tempo Na Cama	20
TTS	Total Tempo Sono	20
VA	<i>Viterbi algorithm</i>	32
VAIS	Vigília Após Início Sono	20
VP	Verdadeiro Positivo	38, 39

Capítulo I Introdução

O sono pode ser caracterizado como um estado oposto ao estado de vigília, onde há um desligamento temporário das percepções. O sono também é caracterizado por ser um estado inconsciente, onde há diminuição da atividade cerebral, relaxamento muscular e diminuição da resposta a fatores ambientais, mas que pode ser revertido por estímulos sensoriais ou sonoros [Santana et al., 2021]. Porém, é nesse estado físico e mental que o corpo restaura várias de suas funções neurológicas, biológicas e psicológicas. Da mesma forma, o sono também mantém processos cognitivos em bom funcionamento, tais como o humor e a memória, e também é necessário para o desempenho do sistema imunológico e endócrino.

Desordens no sono estão associadas a alguns problemas de saúde, como obesidade, diabetes, hipertensão e doenças cardiovasculares. Outra doença que está associada com as desordens no sono é a depressão. Um estudo recente mostrou que noites de sono mais curtas aumentam a probabilidade de início ou a recorrência da doença para pessoas com mais de 60 anos [Hutka et al., 2021].

I.1 Ciclos do Sono

O sono não é apenas um estado inconsciente. Ele também pode ser subdividido em quatro estágios, sendo um estágio conhecido como *Rapid Eyes Movement* (REM) e outros três estágios conhecidos como não REM, ou *Not Rapid Eyes Movement* (NREM). Durante uma noite de sono de um paciente adulto, observa-se entre 4 e 6 ciclos dependendo do número de horas dormidas ao todo. Os ciclos de sono variam durante a noite, sendo o primeiro ciclo com duração média de 90 minutos, e diminuindo ao decorrer da noite, e 80% do tempo de sono está relacionado aos estágios NREM [Hutka et al., 2021].

O sono inicia-se pelo estágio N1 (ou primeiro estágio NREM), seguindo para os estágios N2 e N3 de acordo com o aprofundamento do sono. O estágio N3 apresenta ondas cerebrais mais lentas e é tido como a etapa mais restauradora e mais profunda do sono. Ao final da fase N3 o ciclo do sono começa a se encerrar, passando novamente pelas fases N2 e N1 até atingir a fase REM onde é delimitado o início de um novo ciclo. A fase mais profunda do sono, a fase N3, apresenta maior duração nos primeiros ciclos; e de forma oposta, a fase REM vai aumentando sua duração durante a noite, podendo chegar a 45 minutos de duração [Laranjeira et al., 2017]. Ao todo, um indivíduo

jovem adulto permanece por volta de 20% a 25% do tempo de sono na fase N3 e o mesmo padrão se repete para a fase REM. Além disso, a eficiência do sono média de uma pessoa é de 90% (*Sleep efficiency* (SE)), isso significa que em 90% do tempo total que a pessoa passa na cama ela estará dormindo [Hutka et al., 2021]. No estágio N1, onde o sono ainda é mais leve, há maior chance do indivíduo despertar, assim como de perturbações no sono causadas por distúrbios como a apneia. Essas características também estão presentes no estágio REM [Hutka et al., 2021]. A Figura I.1 apresenta a arquitetura e as fases do sono durante uma noite de sono de um indivíduo adulto.

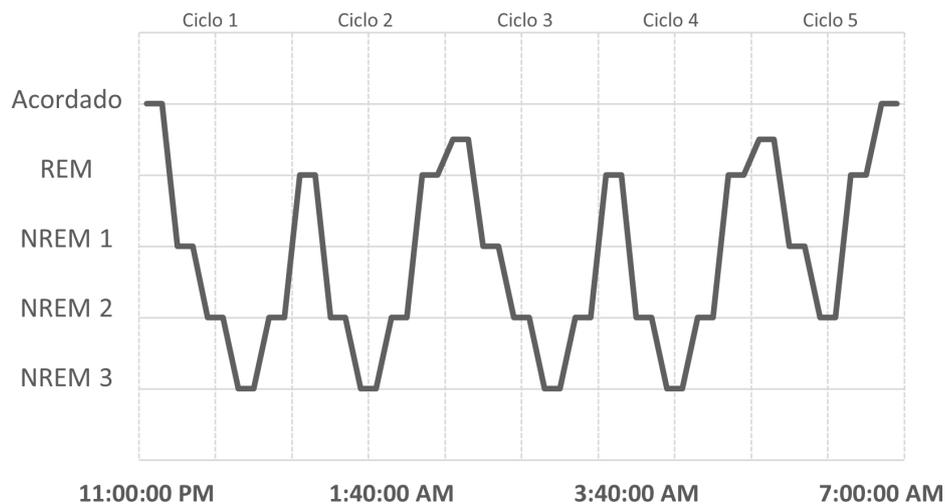


Figura I.1: A arquitetura do sono e suas fases. Fonte: Walker, 2018

Tendo em vista essa arquitetura do sono e as suas transições de fases, uma noite de sono na qual o indivíduo consegue passar por todas as fases com tempo suficiente tem grande influência positiva em sua saúde. Por outro lado, uma noite de sono mal dormida e, portanto, não restauradora, acometida por distúrbios do sono, pode acarretar em impactos negativos na qualidade de vida do indivíduo [Silva Júnior et al., 2015]. A depressão, por exemplo, é associada às anormalidades observadas na arquitetura do sono como a sua continuidade, a sua profundidade e a distribuição dos seus estágios [Armitage, 2000].

Da mesma forma que são observadas transições entre as fases do sono, a duração total de sono de um indivíduo também sofre alterações em função da idade. Recém nascidos costumam dormir 16 horas por dia, em média. Já para um indivíduo adulto, recomenda-se dormir entre 7 a 9 horas, segundo a National Sleep Foundation (NFE) [Hirshkowitz et al., 2015].

Durante as 24 horas do dia, há um ciclo entre se manter desperto e o ciclo do sono, que todo indivíduo passa por toda sua vida, denominado ciclo de vigília. Porém somente em 1982, Alexander Boerbély, propôs um modelo que evidenciou a regulação entre esses dois ciclos. Segundo Alexander, essa regulação se dá através da ação recíproca dos sistemas circadianos e homeostáticos. O sistema

circadiano é responsável pela manutenção da vigília (fase em que está acordado), já no sistema homeostático, quando atinge um limiar superior, o indivíduo se encontra com maior propensão ao sono [Santana et al., 2021].

I.2 Privação do Sono

Um indivíduo acometido por alterações da qualidade do seu sono, ou privação de sono, observa as consequências na fase de vigília, ao apresentar comprometimentos relacionados à cognição, dificuldade de concentração e memória [Santana et al., 2021].

Ainda existem inúmeras perguntas a serem respondidas acerca do sono, e uma delas diz respeito a quanto tempo seria possível ficar em privação do sono. Foi com o objetivo de responder a essa pergunta que *Randy Gardner* e um grupo de jovens resolveram bater o recorde de privação de sono e monitoraram todo o processo. *Randy Gardner*, que foi o escolhido para ficar sem dormir, alcançou a marca de 11 dias e 25 minutos sem pegar no sono. No decorrer desse período, foi relatado por *Randy* variações no olfato, na audição e no paladar, bem como em tarefas cognitivas. Além disso, ele também observou grandes variações de humor [Schuler and Pacheco, 2022].

Outras experiências de cunho científico foram realizadas contabilizando-se o risco de lesões no esporte profissional (vide Figura I.2), como um maior risco de acidentes de carro e um aumento da obesidade devido a diminuição de horas dormidas. Apesar desses experimentos terem sido feitos em americanos, a Organização Mundial da Saúde (OMS) já declarou que há uma epidemia de privação de sono em todos os países industrializados [Walker, 2018].

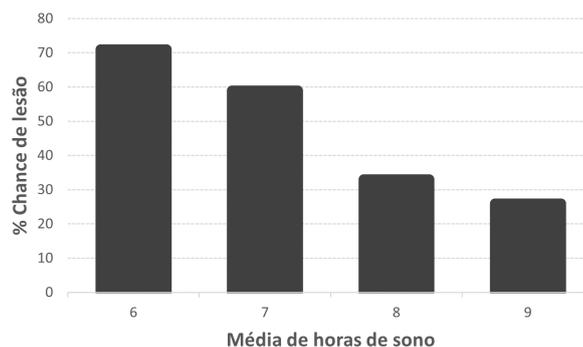


Figura I.2: Relação entre a média de sono e a chance de lesão em esportes. Fonte: Walker, 2018

Muitos estudos são realizados na área da medicina esportiva, com o objetivo de minimizar a probabilidade de os esportistas sofrerem lesões. Contudo, um dos grandes fatores para a reabilitação motora, mental e muscular do atleta é o sono. A qualidade do sono não é uma variável levada muito em questão, pois ainda existe grande desconhecimento na área. Porém, fica claro através do gráfico apresentado na Figura I.2 que um sono de boa qualidade e restaurador diminui consideravelmente as chances de lesões desses atletas.

Uma outra estatística alarmante, apresentada na Figura I.3, ilustra a relação entre o aumento do risco de acidentes de veículos e quantidade de horas de sono do motorista. Pode-se perceber, a partir dos exemplos apresentados, que o sono tem um papel restaurador de todas as condições motoras e mentais, afetando assim a resposta imediata a um perigo. Esse reflexo, quando atrasado, mesmo que em frações de segundos, pode significar entre bater ou não com o carro.

Essas duas relações evidenciam o papel indispensável do sono, mesmo que a diferença entre a quantidade de horas dormidas e a quantidade de horas necessárias para que um indivíduo possua um sono restaurador seja pequena. A privação do sono causa uma resposta aos estímulos ambientais atrasada em função da quantidade deficitária de horas dormidas. Embora muitas vezes imperceptível para um indivíduo, o risco de colisão entre veículos é alto.

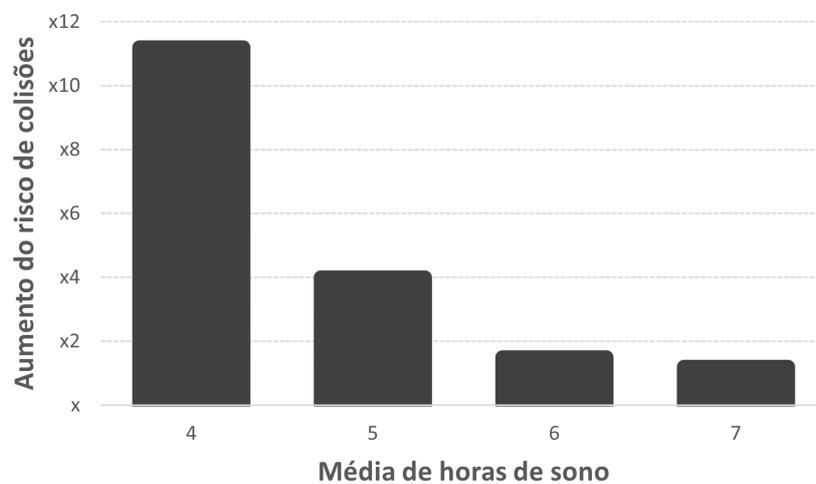


Figura I.3: Relação entre a perda de sono e o aumento do risco de colisão de veículos. Fonte: Walker, 2018

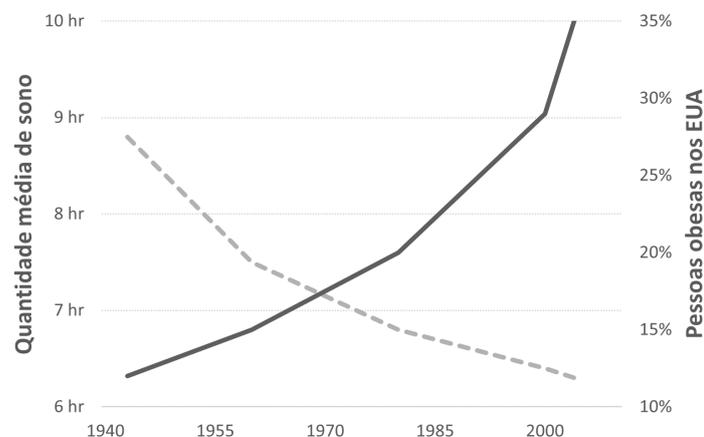


Figura I.4: Relação entre a perda de sono e obesidade. Fonte: Walker, 2018

Os dados apresentados na Figura I.4 mostram que a perda do sono pode ser uma das causas da obesidade. Esta hipótese está relacionada com o fato de que a insuficiência de sono aumenta

a concentração de um hormônio que nos faz sentir fome e, por outro lado, freia a produção do hormônio de saciedade. Esse é o motivo pelo qual há vontade de se comer alimentos doces quando o indivíduo teve uma noite de sono de má qualidade [Walker, 2018].

Na Figura I.4, também é possível observar que a quantidade média de sono da população está em declínio, e essa queda é relacionada a alteração comportamental da sociedade nas últimas décadas. A exposição à luz artificial e as pressões socioeconômicas são tidas como os principais responsáveis pela insônia. Segundo a Classificação Internacional dos Transtornos do Sono (CITS), a insônia é caracterizada pela dificuldade em se iniciar e manter o sono [Santana et al., 2021].

Nos dados dos estudos que geraram as Figuras I.3, I.2 e I.4, pode-se observar a importância do sono para saúde, tanto para o cidadão comum, quanto para um atleta profissional. No gráfico da Figura I.4, pode-se ainda identificar que nas últimas décadas, com o cotidiano cada vez mais corrido e demandante, a quantidade média de sono do cidadão americano vem diminuindo na mesma proporção que o número de indivíduos americanos obesos vêm crescendo. Não seria de se espantar se pudéssemos extrapolar esse resultado para cidadãos de outros países mundo afora.

Em um outro estudo realizado na Austrália [Walker, 2018], recrutara-se dois grupos de adultos saudáveis: um deles foi embriagado até o limite jurídico para dirigir (0,08% de álcool no sangue); enquanto que o outro foi privado de sono por uma única noite. Os dois grupos executaram o teste de concentração para avaliar o desempenho da atenção, medido através do número de lapsos que os integrantes do estudo tiveram. Após passarem 19 horas acordadas, as pessoas privadas de sono tiveram a cognição tão prejudicada quanto a daquelas que tinham sido embriagadas.

Curiosamente, os participantes dos estudos supracitados não percebem que aplicam menos esforço ao enfrentar os desafios do trabalho ou que são menos eficientes quando privados de sono, apesar de ambos os fatos serem verdadeiros. Observou-se que os tempos gastos em rotinas diárias mais simples, como arrumar com cuidado ou elegância para ir trabalhar, aumentam após uma noite mal dormida [Walker, 2018].

Dentre os transtornos mais citados pelos pacientes acometidos por distúrbios de sono, a depressão está entre o principal transtorno psiquiátrico. Estima-se que entre 50-80% dos pacientes com transtornos mentais têm distúrbios do sono e pelo menos 50% dos pacientes com distúrbios do sono têm algum transtorno mental.

Para a avaliação desses distúrbios do sono, a polissonografia (PSG) é até hoje o principal exame para diagnosticar grande parte desses transtornos [Laranjeira et al., 2017]. O exame de polissonografia tradicional é realizado em uma clínica especializada, onde o paciente dorme uma noite sob monitoramento. Esse exame registra vários sinais neurológicos e fisiológicos através de sensores ligados a pontos específicos do corpo do paciente. Esses sensores são acoplados a aparelhos para a leitura dos sinais biométricos. O exame de PSG registra os sinais correspondentes a um Eletroence-

falograma (EEG), Eletrooculograma (EOG), Eletromiograma (EMG), Eletrocardiograma (ECG), fluxo aéreo (nasal e oral), esforço respiratório (torácico e abdominal) e gases sanguíneos (saturação de oxigênio; concentração de dióxido de carbono) [Guimarães, 2010].

Apesar de ser o principal exame realizado para o diagnóstico de transtornos do sono, a PSG apresenta dois pontos negativos a serem considerados. O primeiro ponto é o fato de ser um exame caro e realizado apenas em clínicas especializadas, acarretando em possíveis grandes períodos de espera para sua realização. O segundo ponto é o desconforto gerado pela realização do exame, conhecido como “efeito de primeira noite em laboratório do sono”, o que pode gerar sinais inverossímeis [Azman et al., 2019].

I.3 Objetivos e contribuições

Tendo em vista os pontos abordados anteriormente, principalmente a relevância do sono para a fisiologia humana e a dificuldade para a realização do exame de PSG, o presente estudo tem como principal objetivo detectar distúrbios de sono através da utilização de algoritmos de aprendizado de máquina [Reis et al., 2024]. Nesse sentido, pretende-se não só realizar a detecção de distúrbios do sono com os dados de entrada provenientes de exames reais de polissonografia, mas também conseguir um bom desempenho dos modelos de aprendizado de máquina a partir de menos dados de entrada para ter a flexibilidade da realização do exame de forma domiciliar, por exemplo, a partir de dispositivos vestíveis [Scott et al., 2020; Rentz et al., 2021].

Nesse cenário, elencamos os seguintes objetivos e contribuições deste trabalho, cujos resultados principais foram publicados em [Reis et al., 2024]:

- Avaliar algoritmos de aprendizado de máquina para a detecção de distúrbios de sono, a partir dos dados provenientes do exame de PSG;
- Avaliar o desempenho da classificação dos algoritmos de aprendizado de máquina quando disponibilizados menos dados de entrada, sendo esses possíveis de serem captados através de dispositivos vestíveis.

A base de dados referenciada em [Ghassemi et al., 2018] foi considerada neste trabalho. Essa base de dados contém 1985 noites monitoradas, onde os sinais fisiológicos registrado incluíram ECG, EOG, EMG e saturação de oxigênio. Além disso, o *arousal* será considerado como distúrbio a ser detectado pelos algoritmos. *Arousal* nesse contexto de sono é definido como qualquer distúrbio do sono diferente de apneia [Shahrbabaki et al., 2015; Qian et al., 2021].

I.4 Estrutura

Este trabalho está estruturado da seguinte forma: O Capítulo II apresenta a fundamentação teórica estudada, que é composta pela conceitualização do exame de PSG e dos sinais obtidos a partir desse exame, também por uma explicação teórica dos algoritmos de aprendizado de máquina que foram avaliados nesse estudo, além de uma revisão da literária sobre o tema proposto. O Capítulo III apresenta a base de dados utilizada neste estudo, o pré-processamento dos dados e os principais desafios encontrados. Em seguida, o Capítulo IV apresenta os resultados obtidos através dos teste realizados com todos os dados de entrada e também após apenas alguns dados de entrada serem selecionados. Por último, o Capítulo V apresenta as conclusões obtidas através dos resultados, assim como algumas sugestões de futuras direções para o tema aqui investigado.

Capítulo II Fundamentação Teórica e Trabalhos Relacionados

O Capítulo II tem como base estabelecer o embasamento teórico e a contextualização necessária para compreender os conceitos e argumentos apresentados ao longo do trabalho. Neste Capítulo são abordados em detalhes o exame de PSG, os algoritmos de aprendizado de máquina avaliados nesse trabalho, o *pipeline* para a futura implementação do modelo e a análise dos trabalhos relacionados.

II.1 Polissonografia

Existem alguns métodos para monitorar a qualidade do sono. Porém, a PSG ainda é tida como o padrão para avaliação do sono e diagnóstico de distúrbios do sono, porque a PSG proporciona informações mais detalhadas e mais precisas. A PSG é um exame não invasivo que tem como principal objetivo diagnosticar distúrbios do sono. O exame consiste basicamente no monitoramento e registro dos sinais fisiológicos captados por eletrodos e sensores que estão conectados ao corpo do paciente.

O exame de PSG pode monitorar muitos parâmetros. Dessa forma, foram criados níveis de realização desse exame, e em cada nível, diferentes parâmetros são monitorados. A Tabela II.1 apresenta os parâmetros monitorados, agrupados por nível [Guimarães, 2010].

Tabela II.1: Níveis de realização da PSG e os parâmetros monitorados.

	Descrição	Parâmetros monitorados
Nível 1	PSG completa em laboratório especializado	EEG; EMG; EOG; ECG; Fluxo aéreo; Esforço respiratório; Oximetria; Movimento de pernas (MP).
Nível 2	PSG completa domiciliar	EEG; EMG; EOG; ECG; Fluxo aéreo; Esforço respiratório; Oximetria.
Nível 3	Teste modificado portátil	Fluxo aéreo; Esforço respiratório; Oximetria; Frequência Cardíaca (FC); ECG.

A PSG de nível 1 consiste no registro simultâneo de todos os possíveis parâmetros fisiológicos de um indivíduo durante uma noite de sono. Porém, para a execução da PSG nesse nível, há a necessidade da sua realização em uma clínica especializada, com a colocação de eletrodos e sensores ao longo de todo o corpo do paciente para captação dos sinais.

Já para o nível 2, o exame tem como objetivo realizar a PSG em pacientes que possuem alguma dificuldade de acessibilidade a clínica de sono. Então, nesse nível, o exame é adaptado para ser realizado em casa. Já a PSG de nível 3 é um exame portátil, muitas vezes preliminar com o objetivo de descartar algum possível distúrbio facilmente identificável.

Cabe destacar que um dos objetivos deste trabalho tem a ver com a PSG de nível 3, quando se deseja realizar um exame com um equipamento portátil, com menos sinais monitorados. Este tipo de exame poderia ser realizado a partir de dispositivos vestíveis, em um contexto de *Internet of Things* (IoT). Desta forma, deseja-se avaliar o impacto da redução do número de sinais no desempenho dos algoritmos de aprendizado de máquina avaliados para a detecção de distúrbios do sono.

A partir da leitura dos dados recebidos dos sensores e eletrodos no exame de PSG durante uma noite de sono, é possível calcular algumas métricas para classificação da qualidade do sono do paciente, conforme apresentado na Tabela II.2 [Rodrigues et al., 2012].

Tabela II.2: **Métricas extraídas da leitura dos dados oriundos da PSG.**

Categoria	Descrição
Métricas do sono	Total tempo sono (TTS) Tempo na cama (TNC) Latência do sono (LS): tempo do apagar das luzes até a fase REM do primeiro ciclo. Vigília após início sono (VAIS): Minutos de vigília após início sono e antes do final sono. Eficiência sono (%): $\frac{TTS \times 100}{TNC}$ Tempo total em cada estágio Percentual de TTS em cada estágio: $(\text{Tempo em cada estágio} / TTS) \times 100$
Eventos de despertar	Quantidade de despertares
Eventos respiratórios	Número de apneias e hipoapneias Esforço respiratório relativo aos eventos de despertar. Quantidade de eventos de dessaturação de oxigênio Valor médio da saturação de oxigênio Ocorrência de hipoventilação
Eventos cardíacos	FC média Arritmias Taquicardia
Eventos de movimentos	Quantidade de movimentos dos membros no período.

II.2 Algoritmos de Aprendizado de Máquina

Os algoritmos de aprendizado de máquina são programados para detectarem padrões nos dados de entradas e modelarem uma resposta generalista. Desta forma, o algoritmo consegue aprender e prever respostas quando confrontados com novos dados de entrada.

Há dois grupos de algoritmos de aprendizado de máquina. São eles: os supervisionados e os não supervisionados. Os algoritmos supervisionados são aqueles nos quais os dados de entrada também apresentam os valores da variável alvo, ou resposta, como esquematizado na Figura II.1. Esses algoritmos podem ser de classificação ou de regressão. Os algoritmos de classificação basicamente tentam identificar a categoria, ou seja, uma variável qualitativa, de cada instância a partir das relações encontradas entre a variável alvo e as demais variáveis nos dados de treinamento. Para os algoritmos de regressão, o conceito que visa relacionar as variáveis é o mesmo, porém nesse caso a variável alvo não é uma categoria e sim um número contínuo, ou seja, uma variável quantitativa [Bishop and Nasrabadi, 2006].

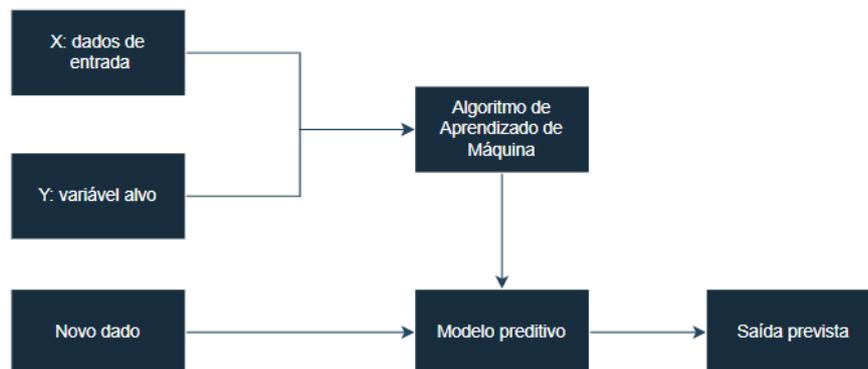


Figura II.1: Esquema do aprendizado supervisionado.

Já os algoritmos não supervisionados são aqueles nos quais os dados de entrada não estão rotulados, ou seja, não apresentam a variável alvo, permitindo apenas uma *clusterização*, ou seja, uma categorização dos dados em grupos com maior relação entre si, como esquematizado na Figura II.2.

O objetivo proposto nesse estudo é detectar distúrbios do sono a partir dos dados provenientes da PSG, que classificam distúrbios do sono durante a noite de sono do paciente, logo esse objetivo se encaixa com os algoritmos supervisionados de classificação.

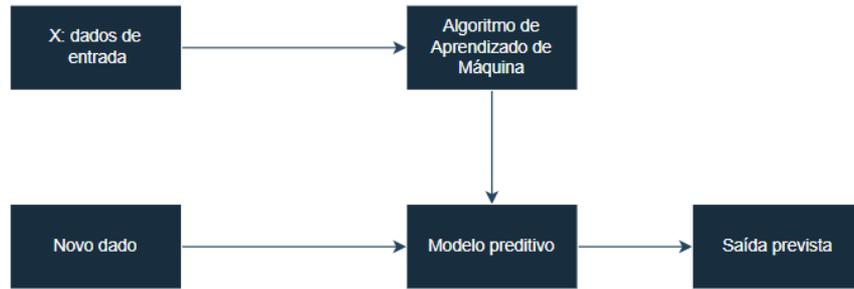


Figura II.2: Esquema do aprendizado não supervisionado.

II.2.1 *K-Nearest Neighbors* (KNN)

O KNN é um algoritmo de aprendizado de máquina de classificação de simples entendimento. O algoritmo compara k vizinhos mais próximos do dado ainda não classificado e os utiliza para a classificação [Zhang et al., 2017].

Em amostras de tamanho pequenas, os métodos paramétricos, como o *Naive Bayes*, leva a bons resultados com baixa variância. Entretanto, quando o tamanho amostral é grande, muitas vezes é vantajoso aumentar o número de parâmetros resultando em um viés menor às custas de uma variância maior. É por essa razão que os métodos não paramétricos ganham relevância [Izbicki and dos Santos, 2020], sendo o algoritmo KNN um exemplo de método não paramétrico.

A formalização matemática do algoritmo KNN é expressa pela Eq. (II.1), a seguir:

$$g(x) = \frac{1}{k} \sum_{i \in N_x} y_i, \quad (\text{II.1})$$

onde a resposta se baseia nas distâncias y dos k -vizinhos mais próximos da instância a ser classificada x . Além disso, N_x é o conjunto de k observações.

Um valor alto de k resulta em um modelo muito simples, gerando assim alto viés e baixa variância. Por outro lado, um valor baixo de k resulta em um modelo com uma variância elevada, mas com um viés baixo [Izbicki and dos Santos, 2020].

A Figura II.3 ilustra o emprego do método KNN com diferentes valores de K , onde, são apresentadas duas classes, A e B, que são representadas pelos quadrados pretos e círculos verdes, respectivamente. Há um círculo menor e um círculo maior envolvendo alguns pontos e determinando uma borda. O círculo menor representa o modelo para $k = 5$, e o círculo maior representa o modelo para um $k = 12$, ou seja, nesse último há 12 pontos categorizados incluídos nesse círculo maior.

O modelo, ao tentar prever um dado novo, que está representado como X na Figura II.3, avalia a categoria dos vizinhos dentro do limite do círculo definido pelo k , ou seja, se o valor de k for igual

a 5, como representado pelo círculo menor, o dado não classificado seria identificado como sendo da classe dos quadrados; já se o valor de k escolhido for k igual a 12, como representado pelo círculo maior, o dado não classificado seria identificado como sendo da classe dos círculos verdes.

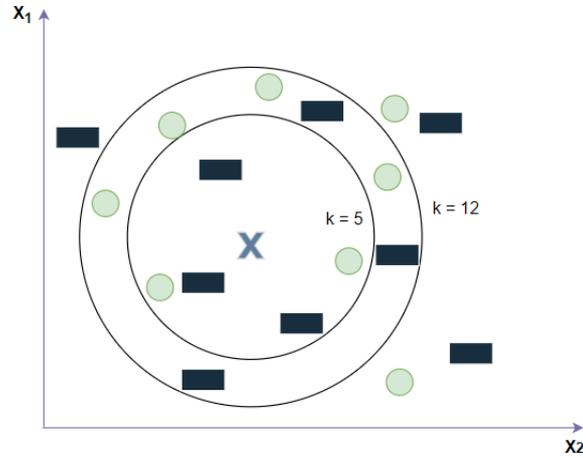


Figura II.3: Ilustração do método KNN. Fonte: Adaptado de Wang et al., 2020.

As etapas de um algoritmo KNN são:

1. Recebe um dado não classificado;
2. Mede a distância (*Euclidiana, Manhattan, Minkowski, etc*) entre o novo dado e todos os outros dados que já estão classificados;
3. Obtém as X ($X = K$) menores distâncias;
4. Verifica a classe de cada um dos dados que tiveram a menor distância e conta a quantidade de cada classe que aparece;
5. Toma como resultado a classe que mais apareceu dentre os dados que tiveram as X menores distâncias;
6. Classifica o novo dado com a classe tomada como resultado da classificação.

As funções de distância comumente usadas são definidas como:

- Função da distância de *Manhattan*:

$$\sum_{k=0}^{n-1} |a_i - b_i| \quad (\text{II.2})$$

- Função da distância *Euclidiana*:

$$\sqrt{\sum_{k=0}^{n-1} (a_i - b_i)^2} \quad (\text{II.3})$$

- Função da distância de *Minkowski*:

$$\left(\sum_{k=0}^{n-1} |a_i - b_i|^{1/p}\right)^p \quad (\text{II.4})$$

onde a e b são as instâncias sendo $k = 1, \dots, n$, e p a ordem ou o grau da distância.

Diferente de algoritmos que classificam as instâncias de acordo com probabilidades, o KNN classifica os dados de acordo com a distância entre a instância e os outros pontos já categorizados mais próximos. O fato desse algoritmo se basear em uma distância significa que é indispensável que os dados estejam normalizados, pois não influenciariam de forma exagerada no modelo [Hastie et al., 2009].

II.2.2 *Support Vector Machine (SVM)*

SVM é um modelo de aprendizado de máquina supervisionado que pode ser utilizado tanto para regressão quanto para classificação. Porém, ele costuma apresentar melhores resultados quando utilizado em modelos de classificação [Mammone et al., 2009].

O SVM utiliza um método, chamado de truque do *kernel*, para aumentar a dimensão dos dados que não podem ser inicialmente separados linearmente, e conseguir realizar essa separação linear em uma dimensão n , maior do que a original. Nesse sentido, o objetivo do SVM é encontrar um hiperplano em um espaço de n dimensões que classifica distintamente os pontos de dados. Esse hiperplano deve ter a margem máxima entre os pontos de dados das duas classes. Os vetores de suporte são pontos de dados que estão mais próximos ao hiperplano e influenciam diretamente a posição do hiperplano [Noble, 2006].

Para melhor entendimento do algoritmo SVM, por exemplo, supõe-se um conjunto de dados de entrada onde y pode ser categorizado por 2 parâmetros em função de x . O SVM constrói classificadores lineares, que separam os dados através de um hiperplano com maior margem [Kalinowski et al., 2023].

Em duas dimensões, o hiperplano é uma reta e sua formulação matemática se dá pela Eq. (II.5).

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 = 0. \quad (\text{II.5})$$

Dessa forma, se $\beta_0 + \beta_1 x_1 + \beta_2 x_2 > 0$, o dado ainda não classificado é de uma classe; e se $\beta_0 + \beta_1 x_1 + \beta_2 x_2 < 0$, o dado ainda não classificado é de outra classe. A Figura II.4 ilustra o hiperplano gerado pelo SVM.

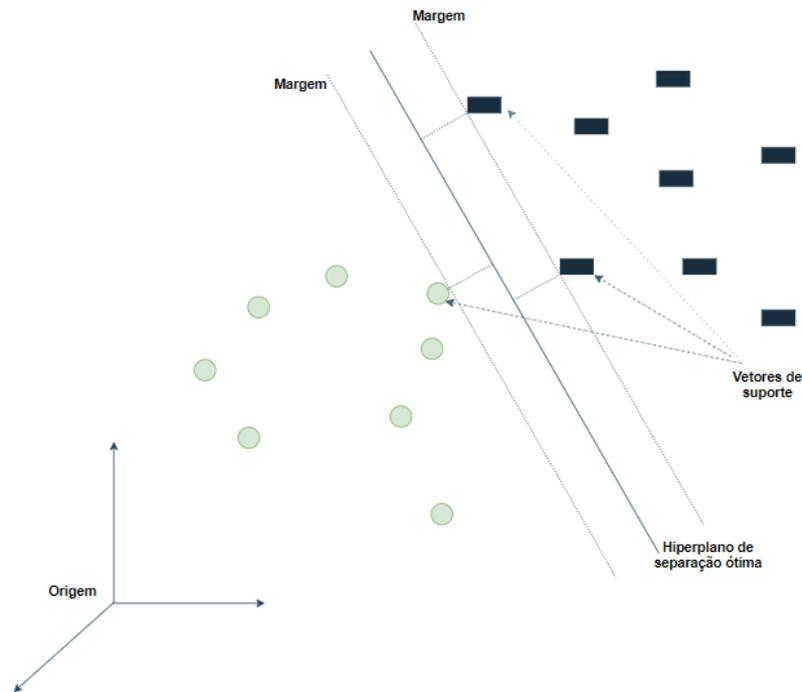


Figura II.4: Ilustração do hiperplano gerado pelo algoritmo SVM. Fonte: Adaptado de Souza, 2011.

Para o algoritmo SVM utilizando o *kernel Radial Basis Function* (RBF), como será apresentado na Tabela II.3, existem dois hiperparâmetros que podem ser variados em busca de um melhor desempenho do modelo: γ e C . O parâmetro C representa uma constante, chamada de parâmetro de regularização, que determina o compromisso entre a maior distância das margens e o menor erro de classificação do modelo. Já o parâmetro γ define até que ponto a influência de uma única instância alcança.

As etapas do SVM são:

1. Recebe um dado não classificado;
2. Definir o par de dados de classes diferente que serão os vetores de suporte;
3. Define o hiperplano de separação ótimo;
4. Classifica o novo dado de acordo com o posicionamento espacial em relação ao hiperplano de separação.

O processo de busca do hiperplano que melhor se ajusta ao modelo possui complexidade computacional quadrática, e que pode ser resolvida pela função de *Lagrange* a seguir, onde se deseja maximizar:

$$L(\alpha) = \sum_{i=1}^L \alpha_i - \frac{1}{2} \sum_{i=1}^L \sum_{j=1}^L y_i y_j \alpha_i \alpha_j x_i^D x_j \quad (\text{II.6})$$

com as seguintes restrições:

$$\sum_{i=1}^L y_i \alpha_i = 0 \quad e \quad 0 \leq \alpha_i \leq C \quad \forall_i \quad (\text{II.7})$$

sendo α um multiplicador da função de *Lagrange*, onde a instância x_i corresponde a $\alpha_i > 0$. Além disso, y_i representam as respostas das instâncias x_i dos dados de treino.

Dado um conjunto de treinamento:

$$D = (x_i, y_i) : i = 1, \dots, L \subset \mathbb{R}^n \times \{-1, +1\}, \quad (\text{II.8})$$

os pontos de dados com $\alpha_i > 0$ que ficam ao longo das margens do limite de decisão são chamados de vetores de suporte.

Na Eq. (II.6), o termo multidimensional $x_i^D x_j$ é linearizado usando-se o “truque de *kernel*” de acordo com a Eq. (II.9).

$$x_i^D x_j = K(x_i, x_j). \quad (\text{II.9})$$

Logo, para solucionar a Eq. (II.9), há a necessidade de se escolher o tipo de *kernel* a ser usado para gerar o modelo de SVM. Nenhum estudo até hoje demonstrou a utilização preferencial de um tipo de *kernel* em relação a outro. Portanto, a escolha do *kernel* se dá através de uma escolha arbitrária e avaliação do desempenho do modelo [Khandoker et al., 2008]. Para o caso de estudo desse trabalho, o *kernel* utilizado foi o padrão, já implementado pelo pacote da linguagem *Python* que possibilitou testar o algoritmo, ou seja, o *kernel* RBF. A Tabela II.3 apresenta três tipos comuns de *kernels*, onde d representa o grau do polinômio, σ representa a largura do RBF e $\gamma = \frac{1}{2\sigma^2}$ [Kecman, 2001].

Tabela II.3: **Formulação matemática das funções *kernel*.**

Função <i>kernel</i>	Fórmula matemática
Linear	$K(x_i, x_j) = x_i \cdot x_j$
Polinomial	$K(x_i, x_j) = (x_i \cdot x_j + 1)^d$
RBF	$K(x_i, x_j) = \exp\left(-\frac{\ x_i - x_j\ ^2}{2\sigma^2}\right)$

Da mesma forma como no KNN, o SVM é um algoritmo que se baseia na distância das instâncias para realizar a classificação. Logo, há também a necessidade de que os dados estejam normalizados

para que a resposta seja coerente [Zhou et al., 2015].

II.2.3 Árvore de Decisão e *Random Forest*

Um modelo *ensemble* se baseia em agrupar modelos de aprendizado de máquina mais simples para produzir um modelo agrupado mais complexo com objetivo de obter melhores resultados. No caso do *Random Forest*, esse é formado por um grupo de modelos de Árvore de Decisão [Cutler et al., 2012]. Dessa forma para entender o modelo *Random Forest*, inicialmente é necessário compreender a Árvore de Decisão.

O modelo de aprendizado de máquina do tipo Árvore de Decisão é bastante conhecido por possuir um bom desempenho e é conceitualmente simples se comparado com outros tipos de algoritmos de aprendizado de máquina. Esse modelo faz o cálculo do ganho de informação para avaliação da resposta em cada nó da folha até a raiz.

Em geral, as funções resultantes dos modelos de Árvore de Decisão apresentam alta correlação. Desta forma, o algoritmo *Random Forest* tem como um dos objetivos diminuir essa alta correlação ao alterar o mecanismo de criação das árvores. Diferente do processo da modelagem da Árvore de Decisão, onde cada parâmetro é avaliado distintamente em um nó específico, na modelagem do *Random Forest*, apenas n parâmetros são escolhidos aleatoriamente para compor cada árvore.

As etapas do *Random Forest* são:

1. Recebe um dado não classificado;
2. O dado não classificado atravessa a *Random Forest* percorrendo os nós das árvores e alterando o caminho de acordo com a condição do nó aplicada ao dado;
3. Ao final do percurso é atribuído uma classificação ao dado.

A formulação matemática para o agrupamento das diferentes árvores de decisão no algoritmo *Random Forest* se dá através da Eq. (II.10), onde B é o número total de árvores de decisão, g é a resposta de uma única árvore e G é a resposta do *Random Forest*, que nada mais é do que uma medida estatística central, a moda, da resposta de todas as árvores [Izbicki and dos Santos, 2020].

$$G(x) = \text{moda}\{g^b(x), b = 1, \dots, B\}. \quad (\text{II.10})$$

II.3 Pipeline para Implementação do Modelo

Nesta seção será abordado uma opção de arquitetura para implementação do modelo, desde a coleta dos dados até a obtenção dos resultados.

Os dados provenientes do exame de PSG, através dos sensores e eletrodos conectados ao longo do corpo dos pacientes, são tradicionalmente enviados ao aparelho receptor através de fios. Porém, com o avanço da tecnologia, principalmente das áreas de transmissão de dados e armazenamento de energia, houve o desenvolvimento acelerado da área de IoT, o que acarretou na possibilidade de obter alguns desses sinais fisiológicos através de dispositivos vestíveis [Ding et al., 2020].

A possibilidade da informação ser transmitida sem a necessidade de aparelhos receptores conectados a fios, possibilitou a criação de aparelhos vestíveis que se tornaram uma realidade no cotidiano moderno. Da mesma forma, a realização do exame de PSG completo de forma assíncrona e sem a necessidade de estar em uma clínica especializada passou a ser viável [Ding et al., 2020].

Dessa forma, pacientes poderiam utilizar esses aparelhos vestíveis adaptáveis para o exame e realizar a PSG em outros locais, assim reduzindo a influência do chamado “efeito de primeira noite em laboratório do sono”. Contudo, vale notar que esse tipo de exame (PSG de nível 3) precisa de uma infraestrutura menos complexa do que a PSG completa, o que implica em usar menos sensores (captando menos sinais). Para tanto, faz-se necessário investigar o impacto da redução no número de sinais de entrada dos modelos de aprendizado de máquina no desempenho dos classificadores, que é um dos objetivos desse trabalho [Reis et al., 2024].

A Figura II.5 representa essa arquitetura em nuvem para implementação dos modelos desse estudo. O *Blob Storage* é um recurso da *Microsoft Azure* para armazenamento de grande quantidade de dados, logo serviria como local de armazenamento dos dados provenientes da PSG, em seguida os dados seriam processados no *Synapse Analytics* e consumidos pelo modelo de aprendizado de máquina na etapa de *Machine Learning*, e como uma última etapa do fluxo o *Logic Apps*, que serviria para automatizar o restante do trabalho e entregar um relatório com os resultados do exame.

II.3.1 Internet das Coisas

A IoT não tem uma descrição única de seus conceitos. Porém, de maneira geral, a Internet das Coisas se caracteriza como um ecossistema interconectado com a Internet entre computadores e sensores, com o objetivo de captação e processamento dos dados [Magrani, 2021].

Para implementação desse ecossistema de IoT para exame de PSG domiciliar, cuja infraestrutura de comunicação se utiliza de uma rede sem fio, um dos pontos importantes é o fornecimento de energia para os sensores de monitoramento de sinais fisiológicos. Para isso, uma possibilidade seria o uso de dispositivos vestíveis com suprimento energético por meio de baterias, que têm limitação de tempo de vida. Desse modo, o estudo de técnicas de economia de energia para esse tipo de rede, chamada de Rede de Sensores Corporais Sem Fio, é importante [Yang, 2008; Verma and Rai, 2015;

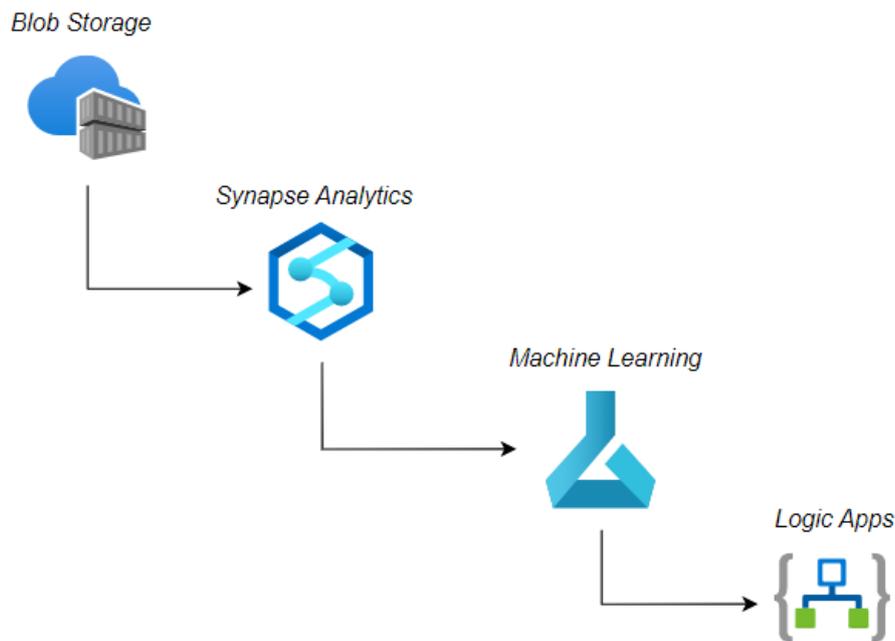


Figura II.5: Arquitetura do fluxo dos dados na *Microsoft Azure*.

Santos and Motoyam, 2018; Ferreira et al., 2019a,b].

A Figura II.6 apresenta um exemplo de arquitetura da captação e comunicação de dados biométricos de pacientes, em uma Rede de Sensores Corporais Sem Fio. O processo exemplificado por essa arquitetura é composto das seguintes etapas: captação de dados, amplificação do sinal, filtro dos ruídos e conversão do sinal analógico em digital. Em seguida, o sinal é transmitido para um computador via *Bluetooth* e via Internet para armazenamento em computação em nuvem. Apesar desse exemplo considerar parte da comunicação usando o padrão de rede *Bluetooth*, outras tecnologias de rede sem fio podem ser usadas, tais como o Wi-Fi (IEEE 802.11), o ZigBee (IEEE 802.15.4) ou a WBAN (IEEE 802.15.6).

II.3.2 Armazenamento dos dados

Após a coleta dos dados através dos sensores e a sua transmissão para a Internet, faz-se necessário armazenar os dados antes de serem processados. A última parte da Figura II.6 apresenta essa etapa, chamada aqui de “migração dos dados para a nuvem”. Hoje em dia, há diversas arquiteturas que dispõem de recursos para realizar esse armazenamento. Neste trabalho, os recursos de computação em nuvem da *Microsoft Azure* foram considerados.

O recurso de armazenamento de dados da *Microsoft Azure* considerado aqui é chamado de Armazenamento de *Blobs* da *Azure*. O armazenamento de dados no *Blob* é altamente escalável,

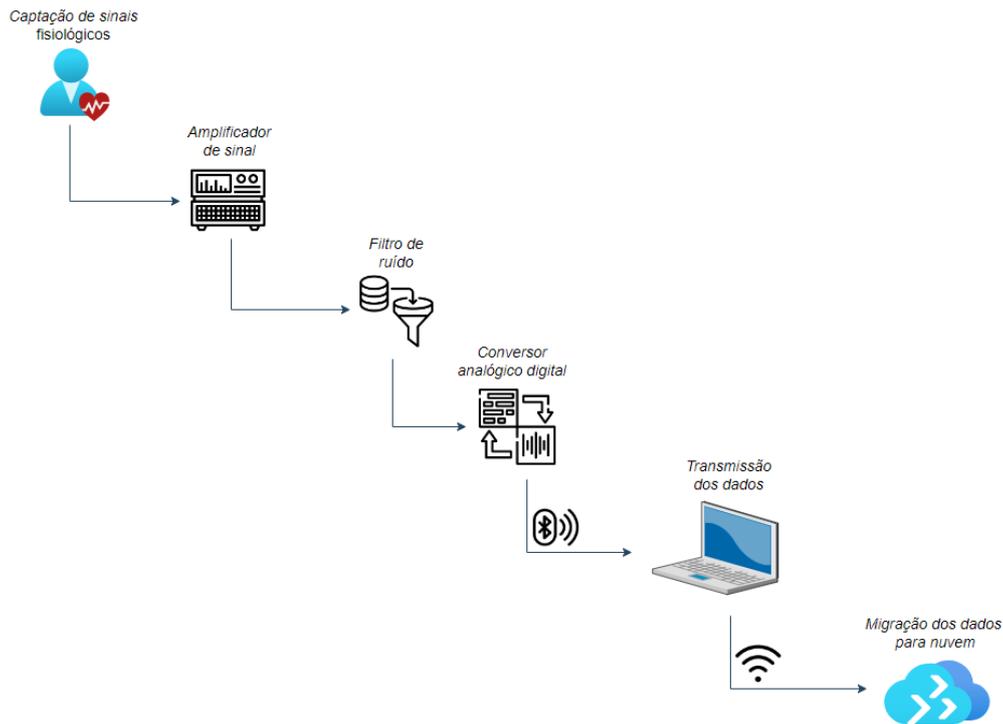


Figura II.6: Exemplo de arquitetura de Internet das Coisas para comunicação de dados biométricos de pacientes.

possibilitando armazenar grandes quantidades de dados, não sendo necessário que esses dados sejam estruturados. Além disso, os dados são armazenados de forma eficiente. O acesso aos arquivos armazenados nos *blobs* se dá através de *Application Programming Interface* (API) [Collier and Shahan, 2015].

II.3.3 Processamento dos dados

Com a grande quantidade de dados gerados pelos exames de PSG realizados já armazenados na nuvem, mais precisamente, já armazenados nos *blobs*, o próximo passo da arquitetura é o processamento desses dados e o recurso necessário, dentro da plataforma de computação em nuvem desse exemplo, é o *Azure Synapse Analytics*.

O *Synapse Analytics* é capaz de criar uma instância de serviço para gerenciar os recursos nesse *workspace*. E para dar seguimento ao fluxo de dados é necessário a criação dos seguintes passos: coleta e inserção dos dados dos exames de PSG, e a etapa de transformação dos dados, adaptando os dados para servirem como insumos para o modelo de aprendizado de máquina.

Para implementação dos modelos de aprendizado de máquina, o recurso utilizado foi o *Azure Machine Learning* sendo executado dentro do próprio *workspace* do *Synapse Analytics* [Collier and Shahan, 2015].

II.3.4 Formulação dos Resultados

Ao final desse fluxo de dados há, a possibilidade de ser elaborado um relatório baseado nos resultados e enviados de forma automatizada com as conclusões realizadas pelo modelo. Essa automatização da estruturação e envio do relatório é realizado pelo *Azure Logic Apps*.

O *Azure Logic Apps* tem como objetivo automatizar fluxos de negócio repetitivo, no modelo *drag and drop*, de uso bem intuitivo e com ações já pré definidas, e esses fluxos são criados com pouco ou nenhum código [Collier and Shahan, 2015].

II.4 Trabalhos Relacionados

Segundo a Associação Americana de Distúrbios do Sono, *arousal* pode ser traduzido como uma excitação ou distúrbio do sono. Ele é caracterizado pela alteração do padrão da frequência do EEG em um valor mínimo de 16 Hz por mais de 3 segundos, tendo em vista que o padrão normal do EEG para uma pessoa dormindo não ultrapassa 8 Hz. Desta forma, Plesinger et al. [2018] consideraram apenas os sinais do EEG para identificação dessas excitações de forma automatizada. Para esse referido estudo foi utilizado um modelo de aprendizado de máquina, do tipo *bagged tree ensemble*, que teve um desempenho em termos de *Area Under the Precision-Recall Curve* (AUPRC) de 20%, quando confrontado com os dados de teste. O modelo tinha o objetivo identificar as áreas onde o paciente sofreu alguma excitação do sono através apenas da leitura dos sinais do EEG, e a base de dados utilizada considerou dados de polissonografia fornecidos pela *PhysioNet* para o desafio *Computing in Cardiology Challenge 2018* [Ghassemi et al., 2018], a mesma base de dados utilizada nesta Dissertação.

Para o estudo de [Shen, 2018], foram utilizados todos os 13 sinais obtidos na polissonografia, que incluem os sinais fisiológicos do EEG, EOG, EMG, ECG, saturação de oxigênio, fluxo de ar e esforço respiratório, como dados de entrada para o modelo. O modelo utilizado foi uma rede neural convolucional profunda, que obteve um desempenho em termos de AUPRC de 0,50 na identificação de excitações do sono. AUPRC é uma métrica para avaliar modelos de aprendizado de máquina do tipo classificação, e o seu valor máximo, onde o modelo seria perfeito nas classificações, seria 1.

A rede convolucional do estudo de Shen [2018] foi constituída de 33 camadas e uma camada final totalmente conectada seguida por uma camada *softmax*, para a classificação da resposta entre uma das duas classes. Entre cada uma das 33 camadas convolucionais também foram adicionadas camadas de agrupamento, conhecidas como *pooling*. Nesse caso, a técnica de *max pooling* foi considerada pelos autores. Essas camadas de agrupamento têm como objetivo extrair os valores máximos de cada unidade de *pooling* do neurônio anterior. Essa técnica ajuda a reduzir a redundância e o custo computacional do modelo. Para evitar um possível *overfitting*, foram utilizadas as técnicas de

normalização em lote, *dropout* e de parada antecipada. O principal fator considerado pelos autores para o desempenho ter sido abaixo do esperado foi o desbalanceamento da base de dados. A base de dados em Shen [2018] é a mesma utilizada neste trabalho.

Já em Li et al. [2018], foram escolhidos 6 sinais de EEG e mais 3 sinais de EMG como sinais de entrada para o modelo utilizado. Nesse estudo, uma rede neural convolucional - *Convolutional Neural Network* (CNN) - com 35 camadas foi considerada, obtendo-se um desempenho de AUPRC de 0,31.

No caso do estudo de Li et al. [2018] a rede neural convolucional profunda - CNN - foi constituída de 3 partes: 1 camada *linear spatial filtering*, 33 camadas *Residuals Networks* (ResNets) (sendo 16 camadas residuais), e finalmente uma última camada totalmente conectada com a utilização da função *softmax*, para classificação da instância ainda não classificada. A camada *linear spatial filtering*, tem como objetivo de aumentar a razão entre sinal e ruído, sem alterar o formato do dado de entrada.

Esse estudo ([Li et al., 2018]) partiu da premissa que o algoritmo de CNN atingiu o estado da arte ao se tratar de visão computação, e que CNNs com módulos residuais tem um desempenho melhor do que CNNs normais. Tendo essa motivação, [Li et al., 2018] utilizou ResNets e considerou cada sinal como se fosse uma imagem unidimensional para se adequar ao modelo proposto. Mais uma vez, o desbalanceamento da mesma base de dados foi levantada como o principal problema para o estudo.

Uma outra estratégia foi considerada por Szalma et al. [2018] para a detecção das áreas de excitação do sono. Nesse caso, um modelo do tipo *Hidden Markov Model* (HMM) foi considerado, em conjunto com o algoritmo de *Viterbi*. Três algoritmos do tipo *Random Forest* foram treinados para gerarem um modelo de probabilidade. Usando os resultados dos modelos *Random Forest* o *Viterbi algorithm* (VA) é capaz de calcular as camadas ocultas mais prováveis do HMM. O HMM então, conseguiu chegar em um desempenho, em termos de AUPRC, de 0,29, tendo como base a mesma fonte de dados de PSG utilizado nesta dissertação.

Ainda utilizando a mesma fonte de dados da base *PhysioNet*, o estudo [Schellenberger et al., 2018] se propôs a detectar distúrbios de sono diferentes de apnéia utilizando um modelo do tipo *Long Short Term Memory* (LSTM). Após testar a combinação de vários hiperparâmetros, a rede que obteve o melhor resultado consistiu em duas células LSTM com 256 camadas ocultas, e obteve um desempenho de AUPRC de 0,14.

Já para o estudo de Zhou et al. [2015], diferente dos estudos anteriores citados até aqui, o objetivo foi investigar a capacidade do modelo SVM para identificar pacientes com Sleep Apnea Syndrome (SAS) tendo o EEG como dado de entrada e utilizando Detrended Fluctuation Analysis (DFA). A DFA é utilizada para analisar séries temporais em que há um componente de tendência

não linear ou para detectar a presença de correlações de longo alcance.

Nesse mesmo estudo, foram analisados dados de seis pacientes com SAS, que tiveram seus exames de PSG realizados pelo *Boston Beth Israel Hospital Sleep Laboratory*, e mais dados de seis pacientes saudáveis que tiveram seus exames de PSG realizados pelo *Westeinde Hospital Sleep Center*. Os dados de entrada utilizados foram apenas aqueles relativos ao sinal de EEG para o mesmo modelo de aprendizado de máquina, o SVM, obtendo 95,1% de acurácia. Há de se considerar como ponto de atenção que foram utilizados apenas 12 pacientes para esse estudo, entretanto, mesmo com poucos dados, conseguiu-se um bom desempenho, em termos da acurácia da classificação.

No trabalho de Sadr and De Chazal [2014], foi utilizado apenas o sinal de ECG para gerar os dados de entrada para o *Extreme Learning Machine* (ELM). A base de dados utilizada foi fornecida pelo *MIT PhysioNet Apnea-ECG* [Penzel et al., 2000] com 35 noites de treino e 35 noites de teste. A acurácia obtida por esse modelo foi de 87,7%.

No estudo de Khandoker et al. [2008], foram observados pacientes em 125 noites, divididos em 83 noites de treino e 42 noites de teste. A base de dados foi uma contribuição da *University College Dublin e St. Vincent's University Hospital*¹. Os dados de entrada utilizadas foram sinais de ECG e os dados respiratórios. Esses dados serviram como entrada para o modelo SVM, que apresentou uma acurácia de 92,85% para identificação de distúrbios do sono.

Nesse presente trabalho, além de avaliar quatro algoritmos de aprendizado de máquina para a detecção de *arousals*, ou distúrbios do sono diferentes de apneia, também há um segundo objetivo que é o de avaliar a quantidade de sinais de entrada a ser considerada pelos algoritmos na detecção dos distúrbios do sono. Com isso, pretendemos investigar o compromisso entre a quantidade de sinais usados como dados de entrada na dos distúrbios e o desempenho dos algoritmos de aprendizado de máquina, de forma a possibilitar aplicações em dispositivos com menor poder computacional e armazenamento de energia [Dhruba et al., 2021; MK et al., 2021].

¹<http://physionet.org/physiobank/database/ucddb/>

Capítulo III Metodologia

O objetivo deste capítulo é descrever a metodologia adotada nesta pesquisa para investigar se é viável a identificação de distúrbios do sono através de técnicas de aprendizado de máquina. Este capítulo aborda detalhadamente a base de dados, todos os passos do pré-processamento e as métricas utilizadas para avaliar os modelos. A apresentação detalhada da metodologia utilizada tem como objetivo garantir a transparência, a replicabilidade e a validade dos resultados obtidos neste estudo.

III.1 Base de Dados e Pré-Processamento

A base de dados considerada neste trabalho foi a mesma que deu origem ao desafio *Computing in Cardiology* referenciado em [Ghassemi et al., 2018]. A plataforma *PhysioNet* é um repositório de dados de pesquisa médica disponíveis gratuitamente, administrado pelo *Laboratory for Computational Physiology* (LCP) do *Massachusetts Institute of Technology* (MIT). A base de dados utilizada nesse estudo foi uma contribuição feita pelo Clinical and Cognitive Neuroscience Laboratory (CCNL) do *Massachusetts General Hospital* (MGH), junto com o Laboratório da Clínica de *Data Annotation* (CDAC). Esse banco de dados inclui 1985 pacientes monitorados no Hospital Geral de *Massachusetts* [Ghassemi et al., 2018].

Ao longo da PSG todos os paciente tiveram os seguintes sinais fisiológicos monitorados: EEG, EOG, EMG, ECG e saturação de oxigênio. Excluindo a saturação de oxigênio, todos os demais sinais foram monitorados em uma frequência de amostragem de 200 Hz. Por conveniência, o sinal de saturação de oxigênio foi reamostrado para se adaptar à forma de monitoramento dos outros sinais.

Na Figura III.1, são apresentados exemplos dos sinais fisiológicos gerados a partir da manipulação dessa base de dados. Os sinais representados por $F_3 - M_2$, $F_4 - M_1$ e $O_1 - M_2$ compõem o EEG; e o ECG representa a atividade cardiovascular. O *Arousal* é a variável alvo que desejamos detectar, ou variável independente, onde o objetivo é identificar como as alterações das variáveis dependentes afetam o comportamento dessa variável independente.

Todas as instâncias foram categorizadas como sendo “-1”, “0” ou “1”, onde “1” significa que o paciente está tendo algum tipo de excitação noturna, ou seja, algum distúrbio no sono; “0” significa um sono normal, sem nenhum tipo de distúrbio; e “-1” identifica uma parte do exame onde não foi

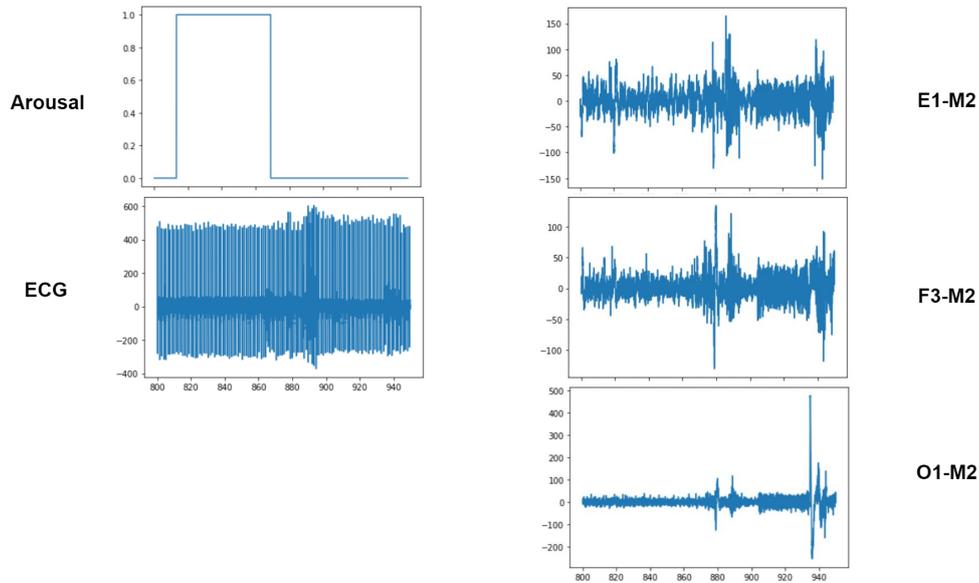


Figura III.1: Visualização de exemplo dos sinais fisiológicos da base de dados de estudo.

possível a realização da categorização.

O período do distúrbio do sono descrito pela base da *PhysioNet* é considerado como sendo 2 segundos antes do primeiro dado ser classificado como “1”, até 2 segundos após o último dado, ou seja, daquele distúrbio específico ser classificado como “1”.

Ao longo de uma noite de um paciente que sofre com algum distúrbio do sono, por mais severo e repetitivo que seja esse distúrbio, espera-se que o paciente passe muito mais tempo dormindo do que sofrendo com o distúrbio, ou seja, nesse caso, haverá muito mais dados categorizados como “0” do que dados categorizados como “1”.

Para refutar a hipótese oposta, uma amostragem aleatória simples foi considerada, com 105 pacientes do total, o que corresponde a uma margem de erro de 1%. Os resultados observados indicam que a média da quantidade de tempo que o paciente passa sofrendo alguma perturbação no sono comparado com o tempo que ele passa dormindo, é de 5%. Logo, a partir desses dados, pode-se afirmar que essa base de dados é classificada como sendo desbalanceada, o que pode gerar problemas com os modelos de classificação.

Oversampling e *undersampling* são duas técnicas para tratar bases de dados desbalanceadas. O *oversampling* trata-se de um processo de geração de dados sintéticos da classe com menor número de registros, já o *undersampling* é o processo oposto, ou seja, trata-se de eliminar dados da classe que tem mais registros [Bishop and Nasrabadi, 2006].

O método de *oversampling* é bastante indicado quando se tem uma amostra de dados pequena, onde a eliminação de dados pode trazer grande prejuízo. Nesse trabalho, há o registro de 1985 pacientes monitorados durante uma noite de sono, totalizando 135 *gigabytes* de dados, ou seja, há uma quantidade significativa de dados à disposição, acarretando até em uma dificuldade computacional

para manipulação de toda essa quantidade de dados. Uma forma de se contornar o problema com a dificuldade computacional é trabalhar com amostras de dados menores e, por consequência, neste trabalho, foi utilizado o método de *undersampling*, que não gerou nenhum tipo de prejuízo com perda de informação.

Para a execução do *undersampling* foi feita uma técnica descrita como janelamento. O janelamento dos dados consiste na realização de um recorte temporal de uma região de distúrbio do sono, ou seja, onde os dados são categorizados como “1”. Essa tarefa foi automatizada através de uma função criada em *python*, da seguinte forma: onde houver uma região que tenha algumas instâncias categorizados como “1”, então essa função realiza um recorte temporal de um valor arbitrário antes do primeiro valor “1” e depois do último valor “1” mais esse valor arbitrário.

Conforme citado anteriormente, para se adequar à descrição do *PhysioNet*, esse valor arbitrário não pode ser menor do que 2 segundos. Como os dados são obtidos numa frequência de 200 *Hz*, logo isso corresponde a 400 instâncias de informação dos sinais monitorados. Após alguns testes preliminares, foi escolhido o valor de 3000 instâncias para esse valor arbitrário com o objetivo de se obter uma amostra de dados já balanceada.

Para leitura dos dados brutos, o próprio repositório de dados do *PhysioNet* fornece uma biblioteca em código *python*. Com esses dados brutos disponíveis, a segunda etapa do processo foi o pré-processamento, que englobou *undersampling* e a normalização.

Os sinais monitorados são muitos discrepantes em número absoluto. Para evidenciar essa situação, a Tabela III.1 apresenta os valores máximos, mínimos, as médias e os desvios padrões de cada sinal considerando os sinais dos exames de PSG de 105 pacientes escolhidos aleatoriamente.

Tabela III.1: Valores Mínimo, Máximo, Médias e Desvios Padrão dos sinais do exame de PSG.

Sinal	Mínimo	Máximo	Média	Desvio Padrão
F3-M2	-7135	6294	0,001608	51,4064
F4-M1	-6644	7018	0,000730	54,0166
C3-M2	-7390	7509	0,000125	51,6248
C4-M1	-6849	7089	-0,000755	61,73256
O1-M2	-7062	6282	0,003117	48,2796
O2-M1	-6641	5252	0,001814	47,00
E1-M2	-7094	7558	0,001854	53,052
Chin1-Chin2	-4741	5278	-0,000056	36,3156
ABD	-8416	8328	-0,2352	632,5017
CHEST	-8382	8281	-0,006390	435,7844
AIRFLOW	-7795	7927	-0,063039	614,4929
SaO2	-32443	32443	28022,884251	7660,8344
ECG	-4596	4538	-0,003362	204,6287

A falta de normalização dos dados poderia gerar pesos diferentes para algoritmos que levam em consideração distâncias para a formalização do modelo, e em contrapartida não geram prejuízos

para os demais algoritmos. Para normalização dos dados foi utilizado o procedimento “min-max”, definido pela Equação (III.1).

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}. \quad (III.1)$$

Em seguida, com os dados processados, houve a necessidade de dividi-los em dados de treinamento, validação e teste. Os dados de treinamento foram utilizados para treinar os quatro modelos de aprendizado de máquina apresentados Capítulo II: KNN, SVM, *Decision Tree* e *Random Forest*. Esses quatro modelos selecionados foram os mais encontrados na literatura tendo como referência a mesma base de dados. O processo de treinamento de um algoritmo de aprendizado de máquina supervisionado permite que o algoritmo tenha acesso a variável resposta e consiga identificar uma relação estatística entre as demais variáveis e a variável alvo, tendo como resultado uma formulação matemática que relaciona essas variáveis.

Os dados de teste foram utilizados para avaliar os modelos. Esses dados nunca tinham sido vistos pelos modelos, e esses têm o objetivo de avaliar o desempenho de cada algoritmo. Já com a identificação do modelo de aprendizado de máquina com melhor resultado para esse caso de estudo, os dados de validação foram utilizados para otimização dos seus hiperparâmetros.

Após essas etapas, ainda houve uma ponderação da quantidade de sinais usados para a detecção dos distúrbios do sono, e a reavaliação dos métodos de aprendizado de máquina considerando menos parâmetros de entrada.

Na Figura III.2 está apresentado um fluxograma da metodologia que foi implementada nesse estudo. Os dados são oriundos de exames de PSG realizados pelo CCNL do MGH, onde foram monitorados alguns sinais fisiológicos. Houve a necessidade de manipulação desses dados, caracterizado na Figura III.2 como Mineração dos dados, e em seguida foram realizados os testes com os modelos de aprendizado de máquina utilizando todas as variáveis disponíveis como também utilizando apenas algumas variáveis selecionadas. O pacote *scikit-learn* da linguagem de programação *Python* teve a finalidade de servir como fonte para utilização dos algoritmos. Finalmente, a última etapa do fluxo foi a realização do *tuning* através de iterações para encontrar a melhor combinação dos hiperparâmetros do modelo mais promissor segundo os resultados da etapa anterior.

III.2 Métricas para Avaliação dos Modelos

A AUPRC serve como uma métrica para avaliar modelos binários. Ela é calculada a partir da área sob a curva de precisão-*recall*, que representa essa relação em diferentes limiares.

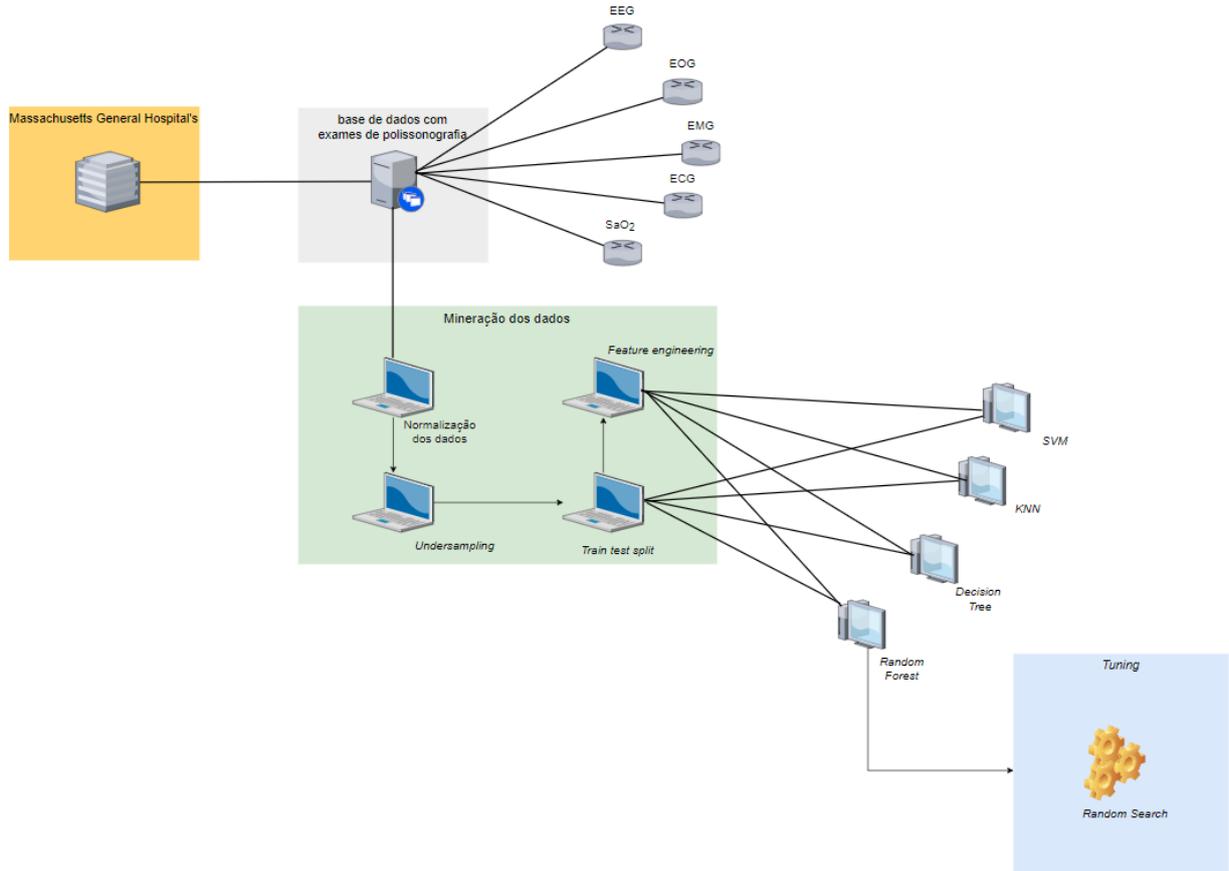


Figura III.2: Fluxo da metodologia proposta neste trabalho.

A precisão indica a proporção do Verdadeiro Positivo (VP) em relação ao total das instâncias classificadas como positivas pelo modelo, ou seja, para calcular essa métrica basta dividir o número de VP pela soma de VP e Falso Positivo (FP), definida pela Equação (III.2):

$$Precisão = \frac{VP}{VP + FP}. \quad (III.2)$$

Recall, também conhecido como sensibilidade ou taxa de VP, indica a proporção de casos positivos identificados corretamente. Para calcular o *recall*, a contagem de observações classificadas corretamente como positivas pelo modelo é dividida pela soma de VP e Falso Negativo (FN), resultando na Equação (III.3):

$$Recall = \frac{VP}{VP + FN}. \quad (III.3)$$

O *F1-score* é uma métrica que estabelece uma conexão entre precisão e o *recall*, e é calculado como uma média harmônica entre essas duas métricas.

$$F1-score = \frac{2 \times Precisao \times Recall}{Precisao + Recall}. \quad (III.4)$$

Assim como o AUPRC, Area Under the Receiver Operating Characteristic Curve (AUROC), é outra métrica empregada para avaliar modelos binários. Ele quantifica a área sob a curva ROC, que ilustra a relação entre a taxa de VP e a FP em vários limiares, e fornece informações sobre a capacidade do modelo de discriminar entre instâncias positivas e negativas.

Por outro lado, a acurácia quantifica a correção geral do modelo, considerando a soma de verdadeiros positivos e verdadeiros negativos. Em outras palavras, ela representa a proporção de instâncias corretamente previstas (positivas e negativas) para o número total de instâncias classificadas pelo modelo, abrangendo ambas as classificações corretas e incorretas [Sofaer et al., 2019]. A acurácia é definida pela Equação (III.5).

$$Acurácia = \frac{VP + VN}{VP + VN + FP + FN}. \quad (III.5)$$

Capítulo IV Resultados

Este capítulo apresenta os resultados obtidos a partir da análise dos dados coletados e dos experimentos realizados durante a pesquisa. O principal objetivo desta pesquisa foi investigar a viabilidade da utilização de aparelhos vestíveis para identificação de distúrbios do sono e os modelos utilizados foram o KNN, o SVM, Árvore de Decisão e *Random Forest*.

Antes de apresentar os resultados, é importante ressaltar que todas as análises foram realizadas cumprindo a metodologia estabelecida no capítulo anterior, garantindo a confiabilidade dos resultados.

Os resultados estão organizados da seguinte forma: a Seção IV.1 descreve os resultados obtidos a partir dos modelos treinados com todos os parâmetros. Já a Seção IV.3 apresenta os resultados obtidos quando os modelos foram treinados apenas com alguns sinais fisiológicos específicos.

IV.1 Avaliação dos modelos de aprendizado de máquina com todos os atributos

Já com os modelos devidamente treinados, a Tabela IV.2 apresenta os resultados finais de cada modelo em relação a cada métrica dividido por classe. Nesse caso, os dados de entrada são os dados de teste e assim como na base de dados, foi mantido o mesmo padrão e os resultados da classe “0” correspondem às métricas ao se avaliar períodos onde não houve distúrbios do sono; e ao contrário, os resultados da classe “1” correspondem aos períodos onde houve distúrbios do sono. Os valores dos hiperparâmetros utilizados nos modelos estão apresentados na Tabela IV.1. Para este experimento inicial, foram considerados os valores padrão presentes no pacote da linguagem *Python*, chamado *skit-learn*, na sua versão 1.3.2., para os hiperparâmetros de todos os modelos avaliados. O modelo que obteve os melhores resultados nesse caso foi o *Random Forest*, como apresentado na Tabela IV.2.

Outra medida que corroborou com a escolha do melhor modelo foi o teste estatístico KS. Este é um teste estatístico não paramétrico utilizado para comparar duas distribuições, e que se baseia na diferença máxima absoluta entre as funções de distribuição.

Na Figura IV.1 a linha tracejada corresponde ao valor de KS, que é definido como a diferença entre a taxa de verdadeiro positivo e a taxa de falso positivo. Quanto maior o valor de KS, melhor é considerado o modelo [Dodge, 2008]. O gráfico da esquerda apresenta os resultados do teste KS para o algoritmo KNN, e o gráfico da direita da Figura IV.1 mostra os resultados para o algoritmo

Tabela IV.1: Hiperparâmetros dos modelos.

Modelo	Hiperparâmetros
SVM	$C = 1$ $kernel = \text{“RBF”}$
KNN	$k\text{-vizinhos} = 2$
Árvore de Decisão	profundidade máxima da árvore = ilimitado número máximo de nós nas folhas = ilimitado
<i>Random Forest</i>	número total de árvores = 150 profundidade máxima de cada árvore = 50

Tabela IV.2: Avaliação dos modelos em relação aos dados de teste.

Modelo	Classe	<i>Precision</i>	<i>Recall</i>	<i>F1 score</i>
SVM	classe 0	0,44	0,68	0,53
	classe 1	0,82	0,63	0,72
KNN	classe 0	0,86	0,85	0,85
	classe 1	0,87	0,88	0,88
Árvore de Decisão	classe 0	0,84	0,85	0,85
	classe 1	0,87	0,87	0,87
<i>Random Forest</i>	classe 0	0,94	0,90	0,92
	classe 1	0,94	0,95	0,94

Random Forest.

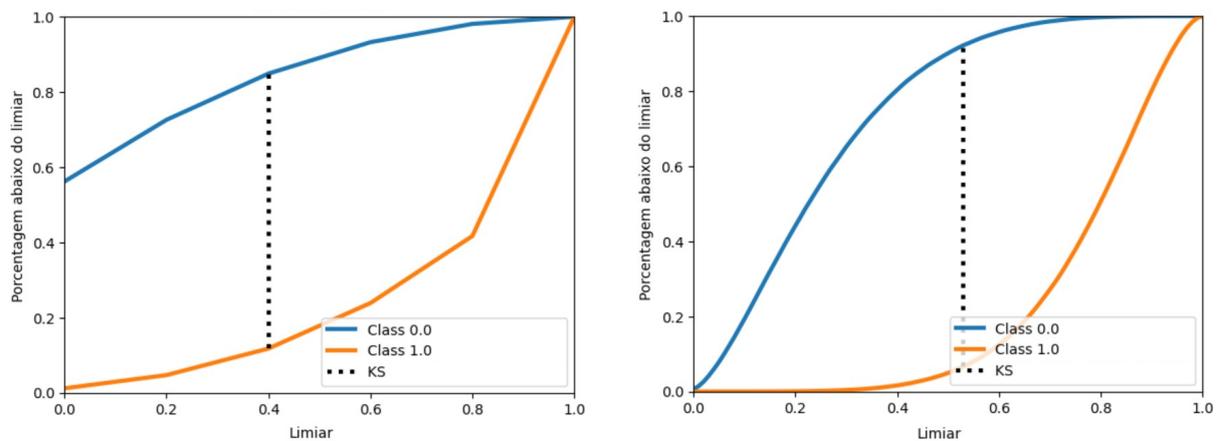


Figura IV.1: Gráficos do teste estatístico KS para o algoritmo KNN, a esquerda, e para o algoritmo *Random Forest*, a direita, utilizando todos os atributos dos dados (todos os sinais do PSG).

A Figura IV.1 só apresentou os gráficos do teste para os algoritmos KNN e *Random Forest*,

pois os demais modelos obtiveram valores inferiores. O modelo da árvore de decisão obteve um valor de KS de 0,716; e o SVM obteve um valor bem mais baixo, igual a 0,127. Os resultados numéricos do teste estatístico KS para os quatro algoritmos de aprendizado de máquina avaliados são apresentados na Tabela IV.3, onde se observa a superioridade do algoritmo *Random Forest*.

Tabela IV.3: **Resultados numéricos do teste estatístico KS para os quatro algoritmos avaliados, considerando todos os atributos.**

Modelo	KS
SVM	0,127
Árvore de Decisão	0,716
KNN	0,731
<i>Random Forest</i>	0,856

Logo, baseado nos resultados apresentados nas Tabelas IV.2 e IV.3, o *Random Forest* foi o algoritmo mais promissor, quando todos os atributos são considerados como dados de entrada. Esse modelo ainda obteve uma acurácia de 93% e uma AUROC igual a 0,92, como observado na Figura IV.2. Na Figura IV.2 está representado pela linha tracejada um classificador binário aleatório, que teria uma AUROC igual a 0,5 e representado pela linha azul é a curva do modelo *Random Forest*.

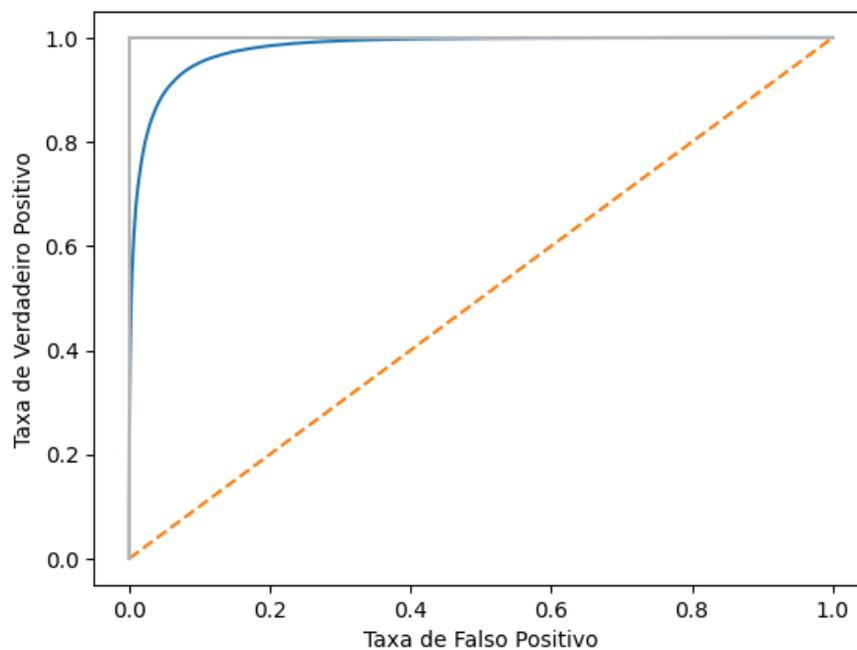


Figura IV.2: Curva ROC para o modelo *Random Forest* utilizando todos os atributos.

IV.2 Avaliação do *Random Forest* após a otimização dos hiperparâmetros

Após essa análise inicial, e considerando que o algoritmo *Random Forest* obteve os melhores resultados, a função *Random Search* foi utilizada, juntamente com os dados de validação, para otimização dos hiperparâmetros (*tuning*) desse algoritmo. Os hiperparâmetros avaliados nesse modelo foram: o número total de árvores (n) da árvore aleatória e a profundidade máxima (d) de cada árvore. Após a busca a partir do *Random Search*, a configuração otimizada de parâmetros desse modelo passou a ser: $n = 150$ e $d = 46$.

A Tabela IV.4 apresenta os resultados das métricas avaliadas para o algoritmo *Random Forest*, após o processo de *tuning* dos seus hiperparâmetros.

Tabela IV.4: Resultados do algoritmo *Random Forest*, após o processo de *tuning* dos hiperparâmetros.

Modelo	Classe	<i>Precision</i>	<i>Recall</i>	<i>F1 score</i>
<i>Random Forest</i>	classe 0	0,90	0,94	0,92
	classe 1	0,95	0,92	0,94

Apesar do *F1 score* não ter apresentado melhora após a otimização, o *Recall* teve uma melhora significativa para a classe 0, e o aumento do *Recall* indica uma diminuição dos resultados FN, o que para o objetivo desse trabalho é a métrica mais interessante. Considere o seguinte exemplo, em que um dado paciente seja caracterizado como tendo distúrbios do sono mas que na verdade não os tenha. Nesse caso, esse paciente seguiria para uma avaliação mais apurada de um especialista, o qual descartaria o diagnóstico errado. Esse seria um caso de FP. Já um paciente caracterizado como não tendo distúrbios do sono, o que seria um FN (situação avaliada pela métrica *Recall* da classe 0), ele permaneceria com o diagnóstico errado, sem ir ao médico. Exatamente por esse motivo, consideramos que a métrica *Recall* da classe 0 apresenta maior relevância em relação às demais.

IV.3 Avaliação dos modelos de aprendizado de máquina com menos atributos

Além da avaliação dos resultados das métricas para os modelos, conforme apresentado nas Subseções IV.1 e IV.3, outro importante objetivo deste trabalho é a análise dos modelos de aprendizado de máquina para detecção de distúrbios do sono quando menos sinais de entrada (ou atributos) são considerados. Uma justificativa para tal análise tem a ver com a premissa de que esses dados podem ser provenientes de dispositivos vestíveis, como em uma Rede de Sensores Corporais Sem Fio. A literatura indica que tais aparelhos vestíveis conseguem medir pelo menos a saturação de oxigênio do usuário, ECG, movimentos do abdômen e peito [Ding et al., 2020].

Com apenas esses quatro sinais de entrada supracitados, os mesmos algoritmos foram avali-

ados, com a mesma configuração de hiperparâmetros que geraram os resultados apresentados na Tabela IV.2. Os resultados estão apresentados na Tabela IV.5. Mais uma vez os resultados finais são apresentados para cada modelo em relação a cada métrica dividido por classe, e o modelo *Random Forest* também se destaca como tendo os melhores resultados.

Tabela IV.5: Avaliação dos modelos, treinados com menos atributos.

Modelo	Classe	<i>Precision</i>	<i>Recall</i>	<i>F1 score</i>
SVM	classe 0	0,63	0,33	0,43
	classe 1	0,60	0,84	0,70
KNN	classe 0	0,78	0,76	0,77
	classe 1	0,80	0,82	0,81
Árvore de Decisão	classe 0	0,76	0,76	0,76
	classe 1	0,80	0,79	0,80
<i>Random Forest</i>	classe 0	0,81	0,79	0,80
	classe 1	0,83	0,85	0,84

Os resultados obtidos com os dados de teste e com menos sinais de entrada mostram que o *Random Forest* foi novamente o modelo mais promissor, obtendo, ainda, uma acurácia de 82% e AUROC também de 0,82, com a curva ROC sendo apresentada na Figura IV.3.

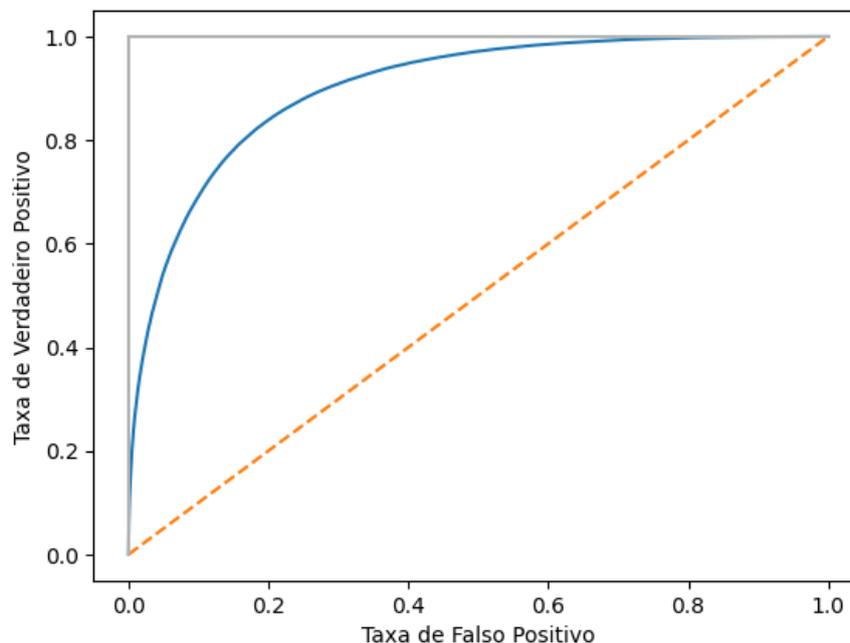


Figura IV.3: Curva ROC para o modelo *Random Forest* com menos *features*.

Ademais, mais uma vez, os testes estatístico de KS corroboraram com os dados obtidos na Tabela IV.5 e o modelo *Random Forest* apresentou a maior diferença entre a taxa de verdadeiro positivo e a taxa de falso positivo, sendo assim considerado o melhor modelo, para esse caso de

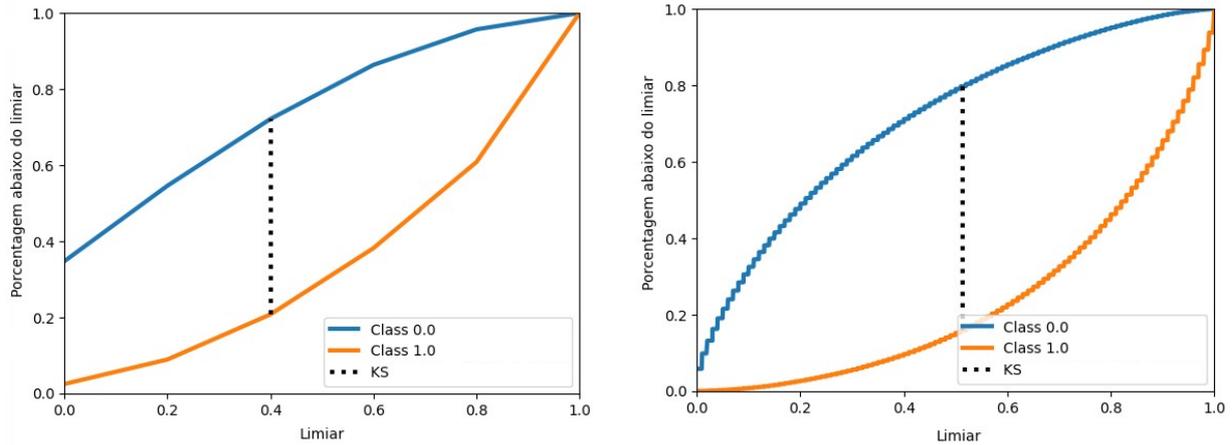


Figura IV.4: Gráficos do teste estatístico KS para o algoritmo KNN, a esquerda, e para o algoritmo *Random Forest*, a direita, utilizando menos atributos.

estudo. A Figura IV.4 apresenta os resultados do teste estatístico KS para os algoritmos KNN (esquerda) e *Random Forest* (direita), quando menos atributos (dados de entrada) são considerados. Além disso, a Tabela IV.6 mostra os resultados numéricos desse mesmo estudo.

Tabela IV.6: **Resultados numéricos do teste estatístico KS para os quatro algoritmos avaliados, considerando menos atributos.**

Modelo	KS
SVM	0,138
KNN	0,514
Árvore de Decisão	0,555
<i>Random Forest</i>	0,639

IV.4 Avaliação do *Random Forest* após a otimização dos hiperparâmetros, considerando menos atributos

Após o experimento, cujos resultados foram apresentados anteriormente, novamente considerou-se o *Random Search* para otimização dos hiperparâmetros do algoritmo *Random Forest*, visto que ele foi o algoritmo com melhor desempenho na detecção de distúrbios do sono.

Os hiperparâmetros avaliados nesse estudo foram os mesmos da etapa de *tuning* do modelo anterior (Subseção IV.3). Após a busca, não houve nenhuma contribuição para os resultados, e a melhor configuração desse modelo foi: $n = 170$ e $d = 46$, com os mesmos resultados encontrados na subseção anterior, apresentados na Tabela IV.5.

Capítulo V Conclusões

Este trabalho teve como sua principal proposta avaliar algoritmos de aprendizado de máquina para detectar distúrbios do sono a partir de sinais que pudessem ser monitorados através de aparelhos vestíveis.

Inicialmente, foi analisado o melhor modelo, dentre os algoritmos SVM, KNN, árvore de decisão e *Random Forest*, utilizando a base de dados de PSG disponibilizada pela *PhysioNet*. Nesse primeiro caso, considerou-se o uso de todos os atributos (sinais disponíveis), e o *Random Forest* gerou o modelo com os melhores resultados. Em seguida, foi realizado o *tuning* (otimização dos hiperparâmetros) desse modelo através da função de otimização *Random Search*. Ao final desses passos os seguintes resultados numéricos foram obtidos: *F1 Score* de 0,92 para a classe 0 e de 0,94 para classe 1; *Recall* de 0,94 para a classe 0 e de 0,92 para classe 1; e uma acurácia de 0,82. A configuração obtida para o algoritmo *Random Forest* com seus hiperparâmetros otimizados foi a seguinte: número total de árvores de 150; e profundidade máxima de 46 para cada árvore.

Tendo como referência a metodologia de obtenção de modelos com todas as variáveis independentes disponíveis, o mesmo padrão de teste foi realizado para calcular os modelos de aprendizado de máquina, para o caso onde as variáveis de entrada seriam obtidos através aparelhos vestíveis, ou seja, considerando menos dados de entrada. Os resultados novamente mostraram que o algoritmo *Random Forest* obteve o melhor desempenho, alcançando um valor de 0,80 para a classe 0 e de 0,84 para classe 1 para o *F1 Score*; além de uma acurácia de 82%.

Embora tenha sido observado uma diminuição na acurácia de 92% para 82% para a detecção de distúrbios do sono, o modelo proposto nesse caso, com menos dados de entrada, ainda se mostrou promissor, tendo em vista que o propósito foi possibilitar um exame mais simples e menos oneroso, com isso diminuindo consideravelmente o número de exames em clínicas especializadas e a possibilidade da realização do exame de forma domiciliar em pessoas com alguma dificuldade de acessar a clínica.

V.1 Futuras Direções

Como futuras direções, a implementação da estrutura de *cloud computing* descrita no Capítulo II com a possibilidade de escalar o processo computacional para processar a quantidade de dados

obtida nos exames de PSG traria muitos benefícios para esse estudo.

Da mesma forma, a possibilidade da aplicação prática da captação dos sinais fisiológicos a partir de dispositivos vestíveis em pacientes durante o sono, também seria mais um passo importante para viabilizar a aplicação prática da proposta desse estudo.

Ainda destacamos a possibilidade de explorar outros algoritmos de aprendizado de máquina, como por exemplo, as redes neurais [Schellenberger et al., 2018], que ainda foram pouco explorados para identificação de distúrbios do sono como visto na Seção II.4. Dessa forma, as Redes Neurais seriam uma boa opção a serem analisadas e comparadas aos demais modelos.

Por fim, destacamos o uso de técnicas específicas de redução de dimensionalidade, a serem aplicadas nos sinais de PSG, como a Análise de Componentes Principais.

Referências

- Armitage, R. The effects of antidepressants on sleep in patients with depression. *The Canadian Journal of Psychiatry*, 45(9):803–809, 2000.
- Azman, N., Wicaksono, S. R., and Sinaga, E. Offline-first sleep assessment (offsa): Internet of things polysomnography sleep assessment framework for sleep monitoring in rural environment. *International Journal of Recent Technology and Engineering*, 8(4):6112–6118, 2019.
- Bishop, C. M. and Nasrabadi, N. M. *Pattern recognition and machine learning*, volume 4. Springer, 2006.
- Collier, M. and Shahan, R. *Microsoft Azure Essentials-Fundamentals of Azure*. Microsoft Press, 2015.
- Cutler, A., Cutler, D. R., and Stevens, J. R. Random forests. In *Ensemble machine learning*, pages 157–175. Springer, 2012.
- Dhruba, A. R., Alam, K. N., Khan, M. S., Bourouis, S., and Khan, M. M. Development of an iot-based sleep apnea monitoring system for healthcare applications. *Computational and Mathematical Methods in Medicine*, 2021, 2021.
- Ding, X., Clifton, D., Ji, N., Lovell, N. H., Bonato, P., Chen, W., Yu, X., Xue, Z., Xiang, T., Long, X., et al. Wearable sensing and telehealth technology with potential applications in the coronavirus pandemic. *IEEE reviews in biomedical engineering*, 14:48–70, 2020.
- Dodge, Y. *The concise encyclopedia of statistics*. Springer Science & Business Media, 2008.
- Ferreira, C. B. M., Guedes, R. M., and Henriques, F. R. Cluster-Head Switching Algorithm Based on Node Temperature in Wireless Body Sensors Networks. In *Anais do Simpósio Brasileiro de Sistemas Multimídia e Web (WEBMEDIA)*, Rio de Janeiro-RJ, 2019a.
- Ferreira, C. B. M., Guedes, R. M., and Henriques, F. R. Algoritmo DTN com Controle de Energia Aplicado em Redes de Sensores Corporais Sem-Fio. In *Anais do XXXVII SIMPÓSIO BRASILEIRO DE TELECOMUNICAÇÕES E PROCESSAMENTO DE SINAIS - SBrT2019*, Petrópolis-RJ, 2019b.

- Ghassemi, M. M., Moody, B. E., Lehman, L.-W. H., Song, C., Li, Q., Sun, H., Mark, R. G., Westover, M. B., and Clifford, G. D. You snooze, you win: the physionet/computing in cardiology challenge 2018. In *2018 Computing in Cardiology Conference (CinC)*, volume 45, pages 1–4. IEEE, 2018.
- Guimarães, G. M. Diagnóstico polissonográfico. *Pulmão.(Rio de Janeiro)*, 9(3-4):88–92, 2010.
- Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.
- Hirshkowitz, M., Whiton, K., Albert, S. M., Alessi, C., Bruni, O., DonCarlos, L., Hazen, N., Herman, J., Hillard, P. J. A., Katz, E. S., et al. National sleep foundation’s updated sleep duration recommendations. *Sleep health*, 1(4):233–243, 2015.
- Hutka, P., Krivosova, M., Muchova, Z., Tonhajzerova, I., Hamrakova, A., Mlynckova, Z., Mokry, J., and Ondrejka, I. Association of sleep architecture and physiology with depressive disorder and antidepressants treatment. *International journal of molecular sciences*, 22(3):1333, 2021.
- Izbicki, R. and dos Santos, T. M. *Aprendizado de máquina: uma abordagem estatística*. Rafael Izbicki, 2020.
- Kalinowski, M., Escovedo, T., Villamizar, H., and Lopes, H. *Engenharia de Software para Ciência de Dados: Um guia de boas práticas com ênfase na construção de sistemas de Machine Learning em Python*. Casa do Código, 2023.
- Kecman, V. *Learning and soft computing: support vector machines, neural networks, and fuzzy logic models*. MIT press, 2001.
- Khandoker, A. H., Palaniswami, M., and Karmakar, C. K. Support vector machines for automated recognition of obstructive sleep apnea syndrome from ecg recordings. *IEEE transactions on information technology in biomedicine*, 13(1):37–48, 2008.
- Laranjeira, C. d. M. et al. Percepção do sono em pacientes submetidos à polissonografia, 2017.
- Li, H., Cao, Q., Zhong, Y., and Pan, Y. Sleep arousal detection using end-to-end deep learning method based on multi-physiological signals. In *2018 Computing in Cardiology Conference (CinC)*, volume 45, pages 1–4. IEEE, 2018.
- Magrani, E. *A internet das coisas*. BOD GmbH DE, 2021.
- Mammone, A., Turchi, M., and Cristianini, N. Support vector machines. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(3):283–289, 2009.

- MK, S. et al. Detection of obstructive sleep apnea using internet of things: A review. *International Journal of Sensors Wireless Communications and Control*, 11(3):263–270, 2021.
- Noble, W. S. What is a support vector machine? *Nature biotechnology*, 24(12):1565–1567, 2006.
- Penzel, T., Moody, G. B., Mark, R. G., Goldberger, A. L., and Peter, J. H. The apnea-ecg database. In *Computers in Cardiology 2000. Vol. 27 (Cat. 00CH37163)*, pages 255–258. IEEE, 2000.
- Plesinger, F., Viscor, I., Nejedly, P., Andrla, P., Halamek, J., and Jurak, P. Automated sleep arousal detection based on eeg envelopgrams. In *2018 Computing in Cardiology Conference (CinC)*, volume 45, pages 1–4. IEEE, 2018.
- Qian, X., Qiu, Y., He, Q., Lu, Y., Lin, H., Xu, F., Zhu, F., Liu, Z., Li, X., Cao, Y., et al. A review of methods for sleep arousal detection using polysomnographic signals. *Brain Sciences*, 11(10):1274, 2021.
- Reis, T. B. F., Tcheou, M. P., and Henriques, F. d. R. Detecting sleep disorders in polysomnography data. In *Latin American Symposium on Circuits and Systems (LASCAS 2024)*, pages 1–5. IEEE, 2024.
- Rentz, L. E., Ulman, H. K., and Galster, S. M. Deconstructing commercial wearable technology: Contributions toward accurate and free-living monitoring of sleep. *Sensors*, 21(15):5071, 2021.
- Rodrigues, M. H., Ramalheira, J., Lopes, J., Gomes, M. d. M., and Silva, A. M. d. Polissonografia: aspectos técnicos e clínicos. *Rev. bras. neurol*, pages 7–22, 2012.
- Sadr, N. and De Chazal, P. Automated detection of obstructive sleep apnoea by single-lead ecg through elm classification. In *Computing in Cardiology 2014*, pages 909–912. IEEE, 2014.
- Santana, T. P., Muniz, J. A., Lopes, M. C., Lúcio, M. F., dos Santos, A. B. B., Ahumada, D. A. R., and Silva, A. C. Sono e imunidade: papel do sistema imune, distúrbios do sono e terapêuticas/sleep and immunity: role of the immune system, sleep disorders and treatment. *Brazilian Journal of Development*, 7(6):55769–55784, 2021.
- Santos, N. S. N. and Motoyam, S. M. S. A large scale platform using wban technology for patient monitoring. *IEEE Latin America Transactions*, 16(3):705–711, 2018.
- Schellenberger, S., Shi, K., Mai, M., Wiedemann, J. P., Steigleder, T., Eskofier, B., Weigel, R., and Kölpin, A. Detecting respiratory effort-related arousals in polysomnographic data using lstm networks. In *2018 Computing in Cardiology Conference (CinC)*, volume 45, pages 1–4. IEEE, 2018.

- Schuler, L. and Pacheco, P. A influência dos aparelhos eletrônicos nos episódios de paralisia do sono. In *Anais do II Congresso de Psicologia do Unilasalle-Rj*, page 27, 2022.
- Scott, H., Lack, L., and Lovato, N. A systematic review of the accuracy of sleep wearable devices for estimating sleep onset. *Sleep Medicine Reviews*, 49:101227, 2020.
- Shahrbabaki, S. S., Dissanayaka, C., Patti, C. R., and Cvetkovic, D. Automatic detection of sleep arousal events from polysomnographic biosignals. In *2015 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pages 1–4. IEEE, 2015.
- Shen, Y. Effectiveness of a convolutional neural network in sleep arousal classification using multiple physiological signals. In *2018 Computing in Cardiology Conference (CinC)*, volume 45, pages 1–4. IEEE, 2018.
- Silva Júnior, J. L. R. d. et al. Depressão e distúrbios respiratórios do sono em pacientes com doença pulmonar obstrutiva crônica, 2015.
- Sofaer, H. R., Hoeting, J. A., and Jarnevich, C. S. The area under the precision-recall curve as a performance metric for rare binary events. *Methods in Ecology and Evolution*, 10(4):565–577, 2019.
- Souza, F. C. M. Técnicas de buscas heurísticas para otimização de parâmetros de máquinas de vetores suportes. Master’s thesis, Universidade Federal de Pernambuco, 2011.
- Szalma, J., Bánhalmi, A., and Bilicki, V. Detection of respiratory effort-related arousals using a hidden markov model and random decision forest. In *2018 Computing in Cardiology Conference (CinC)*, volume 45, pages 1–5. IEEE, 2018.
- Verma, M. and Rai, R. Energy-efficient cluster-based mechanism for wban communications for healthcare applications. *International Journal of Computer Applications*, 120(19), 2015.
- Walker, M. *Por que nós dormimos: A nova ciência do sono e do sonho*. Editora Intrínseca, 2018.
- Wang, X., Huang, L., Huang, H., Li, B., Xia, Z., and Li, J. An ensemble learning model for short-term passenger flow prediction. *Complexity*, 2020:1–13, 2020.
- Yang, G. *Body Sensor Networks*. Springer, 2008.
- Zhang, S., Li, X., Zong, M., Zhu, X., and Wang, R. Efficient knn classification with different numbers of nearest neighbors. *IEEE transactions on neural networks and learning systems*, 29(5):1774–1785, 2017.

Zhou, J., Wu, X.-m., and Zeng, W.-j. Automatic detection of sleep apnea based on eeg detrended fluctuation analysis and support vector machine. *Journal of clinical monitoring and computing*, 29(6):767–772, 2015.