



Universidade Estadual de Campinas  
Instituto de Computação



Rebeca Padovani Ederli

## Detecção e Classificação de Estágios do Sono

Orientador: Prof. Dr. Zanoni Dias

Coorientador: Prof. Dr. Anderson Rocha

CAMPINAS  
2022

# Resumo

A presença de anomalias durante o sono foi originalmente classificada por especialistas e utilizada para o diagnóstico de distúrbios do sono. A polissonografia é o padrão-ouro utilizado em ambientes clínicos para monitoramento e classificação dos estágios do sono, entretanto, esse método exige equipamentos caros e desconfortáveis. A análise do comportamento do sono por meio de *smartwatches* tem sido estudada para a detecção e classificação automática de estágios do sono a fim de possibilitar a utilização de tecnologias vestíveis que são confortáveis como os actígrafos e precisos como a polissonografia. Os principais desafios encontrados ao utilizar métodos de tomada de decisão automática para o sono são a incapacidade de classificar corretamente os estágios mais complexos além do desequilíbrio das classes no conjunto de dados. Então, muitos autores simplificam o problema em classificar o sono em apenas três ou quatro estágios, enquanto a análise ideal deveria ser capaz de distinguir cinco ou seis estágios, conforme o padrão escolhido (AASM ou R&K). Neste trabalho, modelos de aprendizado de máquina e diferentes combinações dos dados de diferentes sensores de *smartwatches* serão exploradas para obter um método competitivo de detecção e classificação dos estágios do sono.

# Abstract

The presence of abnormalities during sleep was originally classified by specialists and used for the diagnosis of sleep disorders. Polysomnography is the gold standard used in clinical settings for monitoring and classifying sleep stages, however, this method requires expensive and uncomfortable equipment. The analysis of sleep behavior through smartwatches has been studied for the automatic detection and classification of sleep stages in order to enable the use of wearable technologies that are comfortable like actigraphs and accurate like polysomnography. The main challenges encountered when using automatic decision-making methods for sleep are the inability to correctly classify the most complex stages in addition to the imbalance of classes in the data set. Therefore, many authors simplify the problem by classifying sleep into only three or four stages, while the ideal analysis should be able to distinguish five or six stages, depending on the chosen standard (AASM or R&K). In this work, machine learning models and different combinations of data from different smartwatches sensors will be explored to obtain a competitive method of detecting and classifying sleep stages.

# Lista de Figuras

2.1	Exemplo ilustrativo de diferentes comprimentos de janela utilizados para calcular as métricas de VFC da época $n$ [19]. . . . .	7
2.2	Funções cosseno, exponencial e linear utilizadas para modelar características circadianas [3]. . . . .	10
2.3	Arquitetura de rede CNN e RNN utilizada por Korkalainen <i>et al.</i> [28]. . . . .	12
2.4	Estrutura do modelo <i>1D-ResNet</i> [31]. . . . .	13
2.5	Estrutura do modelo <i>WISNet</i> [44]. . . . .	13
2.6	Estrutura do modelo <i>UltraNet</i> [23]. . . . .	14
3.1	Esquema de metodologia a ser utilizada. . . . .	16
A.1	Quantidade de dados de vigília e sono da base de dados <i>Sleep Accel</i> a cada hora após o início das gravações de PSG. . . . .	28
A.2	Importância relativa de cada <i>feature</i> de movimento determinada pelo classificador Random Forest. . . . .	31
A.3	Matrizes de confusão para as classificações sem e com aumento de dados e adição de dados passados com Regressão Logística considerando o conjunto com todas as <i>features</i> (movimento, FC e tempo). . . . .	33
A.4	Matrizes de confusão para as classificações sem e com aumento de dados e adição de dados passados com Random Forest considerando o conjunto com todas as <i>features</i> (movimento, FC e tempo). . . . .	34

# Lista de Tabelas

4.1	Cronograma das atividades propostas dividido em meses. . . . .	20
A.1	Exemplo de <i>features</i> de movimento, sem considerar períodos passados. . . .	28
A.2	Exemplo de <i>features</i> de movimento, considerando $n = 2$ períodos passados.	28
A.3	Comparação dos resultados obtidos com $n = 80$ períodos passados utilizando Regressão Logística considerando apenas <i>features</i> de movimento com os demais resultados, que não consideram dados passados. . . . .	29
A.4	Comparação dos resultados obtidos com e sem a aumentação de dados, com o classificador Regressão Logística. . . . .	30
A.5	Comparação dos resultados obtidos com $n = 80$ períodos passados utilizando Random Forest considerando apenas <i>features</i> de movimento com os demais resultados, que não consideram dados passados. . . . .	30
A.6	Comparação dos resultados obtidos com e sem a aumentação de dados, com o classificador Random Forest. . . . .	30
A.7	Comparação das acurácias obtidas em diferentes experimentos realizados utilizando Regressão Logística e apenas movimento, com $n = 30$ dados passados (quando aplicável). . . . .	32
A.8	Comparação das acurácias obtidas em diferentes experimentos realizados utilizando Random Forest e apenas movimento, com $n = 30$ dados passados (quando aplicável). . . . .	32
A.9	Comparação dos resultados obtidos pela Regressão Logística com a configuração original e com a utilização de aumentação de dados e $n = 30$ dados passados. . . . .	32
A.10	Comparação dos resultados obtidos pelo Random Forest com a configuração original e com a utilização de aumentação de dados e $n = 30$ dados passados.	33

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Descrição do Problema . . . . .	2
1.2	Questões de Pesquisa . . . . .	3
1.3	Objetivos . . . . .	3
1.4	Organização do Texto . . . . .	3
<b>2</b>	<b>Fundamentação Teórica</b>	<b>4</b>
2.1	Conceitos e Técnicas . . . . .	4
2.1.1	Estágios do sono . . . . .	4
2.1.2	Séries temporais . . . . .	5
2.1.3	Redes Neurais Convolucionais . . . . .	5
2.1.4	Redes Neurais Recorrentes . . . . .	5
2.2	Trabalhos Relacionados . . . . .	6
<b>3</b>	<b>Materiais e Métodos</b>	<b>15</b>
3.1	Metodologia . . . . .	15
3.2	Métricas de Avaliação . . . . .	16
3.3	Conjuntos de Dados . . . . .	17
3.4	Ferramentas e Recursos Computacionais . . . . .	18
<b>4</b>	<b>Plano de Trabalho e Cronograma de Execução</b>	<b>19</b>
	<b>Referências Bibliográficas</b>	<b>21</b>
	<b>Apêndice Appendices</b>	<b>25</b>
	<b>Apêndice A Resultados Preliminares</b>	<b>26</b>
A.1	Descrição dos experimentos . . . . .	27
A.1.1	Novos conjuntos de <i>features</i> . . . . .	27
A.1.2	Inclusão de <i>features</i> de períodos passados . . . . .	27
A.1.3	Aumentação de dados . . . . .	28
A.1.4	<i>Features</i> de movimento com os eixos x, y e z . . . . .	28
A.2	Resultados e discussões . . . . .	29

# Capítulo 1

## Introdução

As mudanças do estilo de vida da população mundial nas últimas décadas, como falta de atividade física, estresse diário e o excesso de mídia eletrônica, estão relacionadas ao aumento da proporção de pessoas que dormem menos do que o recomendado [18]. O sono insuficiente e toda a vasta gama de distúrbios do sono interferem negativamente em órgãos e sistemas do corpo humano, causando repercussões também durante as vigílias, logo na qualidade de vida, e promovendo a origem de muitas patologias. O primeiro posicionamento brasileiro sobre o impacto dos distúrbios do sono nas doenças cardiovasculares da Sociedade Brasileira de Cardiologia [12] destaca que esses distúrbios aumentam o risco de diversas doenças cardiovasculares. Pelo fato da crescente quantidade de evidências que ligam distúrbios do sono a resultados desfavoráveis à saúde, principalmente o declínio cognitivo em adultos mais velhos, o sono tem recebido atenção também em estudos geriátricos [30]. Nesse sentido, a pesquisa médica tem interesse em automatizar processos para oferecer diagnósticos mais rápidos e adiantados de distúrbios do sono e explorar o comportamento de indivíduos enquanto dormem.

A partir de parâmetros como frequência cardíaca, frequência respiratória, temperatura e movimento corporal [40], é possível determinar o comportamento do sono em termos de tempo, duração, latência, despertar após o início do sono e eficiência do sono [21]. A polissonografia (PSG) é o padrão-ouro utilizado em ambientes clínicos para extrair dados que possibilitam calcular alguns desses parâmetros e determinar os estágios do sono. Entretanto, esse método exige equipamentos caros e desconfortáveis (pelo menos onze canais de vários sinais fisiológicos coletados de diferentes sensores) e o monitoramento é de curto prazo [46], além do efeito no paciente causado pela primeira noite sob registros laboratoriais [33].

Embora a actigrafia seja aplicada apenas para aproximar automaticamente os padrões de sono-vigília e não distingue de forma confiável entre os diferentes estágios do sono, essa é outra ferramenta utilizada em domínio médico, que coleta principalmente dados de movimento e fica posicionada apenas no pulso [46]. Um método que possibilita a detecção automática dos estágios do sono em ambiente domiciliar pode ser capaz de avançar ainda mais na avaliação do sono.

O monitoramento do sono em domínio domiciliar pode ser realizado por tecnologias vestíveis e não vestíveis. Dentre elas destaca-se o uso de *smartwatches*, dos quais vários dispositivos são disponíveis comercialmente e já foram analisados de acordo com o desem-

penho, precisão e percepção do usuário [14]. Nesse sentido, a automatização da detecção e classificação de distúrbios do sono pode proporcionar a utilização de tecnologias vestíveis que são confortáveis como os actígrafos e precisos como a PSG, além de fornecer diagnósticos rápidos e precoces desses distúrbios.

Diversas técnicas de Inteligência Artificial têm sido aplicadas para problemas sobre sono, desde métodos como *Random Forests*, *Support Vector Machine* (SVM) e *K-Nearest Neighbors* (KNN) até a adaptação de Redes Neurais Recorrentes (RNN), como *Long Short-Term Memory* (LSTM) e *Gated Recurrent Unit* (GRU), e Redes Neurais Convolucionais (CNN) [14].

## 1.1 Descrição do Problema

A presença de anomalias durante o sono foi originalmente classificada por especialistas e utilizada para o diagnóstico de distúrbios do sono. Atualmente, a análise do comportamento do sono tem sido estudada na classificação automática de distúrbios do sono, como apneia e insônia [2, 32].

A detecção e classificação de distúrbios do sono envolve a detecção de estágios do sono e, frequentemente, a detecção de vigília e sono é realizada como a primeira etapa para uma classificação mais refinada desses estágios [14]. A validação é feita em relação à classificação manual realizada por especialistas, que inclui variabilidade conhecida pelo fato de que a classificação de estágios específicas pode ser feita de forma diferente pelos profissionais [36].

Os estágios são diferentes estados fisiológicos que o corpo alterna durante o sono e contêm informações importantes, tendo em vista que o conhecimento desses padrões juntamente com outras variáveis são a base para o diagnóstico de diversos distúrbios relacionados ao sono [42]. Rechtschaffen e Kales (R&K) [39] avaliaram, originalmente, seis estágios do sono, sendo vigília, sono de movimento rápido dos olhos (REM) e os quatro níveis do sono não REM (NREM): S1, S2, S3 e S4. Já a Academia Americana de Medicina do Sono (AASM) [22] define cinco estágios diferentes do sono: vigília, REM e NREM dividida em N1, N2 e N3. Os estágios R&K são equivalentes à definição da AASM, que combina S3 e S4 como N3. A análise ideal deveria ser capaz de distinguir cinco ou seis estágios, conforme o padrão escolhido. Muitos dos estudos realizados nos últimos anos para a tomada de decisão automática do sono foram propostos apenas com três classes incluindo vigília, sono REM e sono não REM [17, 48]. Alguns métodos foram avaliados em trabalhos utilizando a classificação de quatro e cinco estágios por meio de modelos tradicionais de aprendizado de máquina [50].

Os principais desafios encontrados ao utilizar métodos de classificação automática para o sono são a incapacidade de classificar corretamente os estágios mais complexos (difíceis de distinção) e o desequilíbrio das classes no conjunto de dados [1, 17]. Os dados utilizados são caracterizados como de varredura, isso é, as variáveis são capturadas continuamente ao longo do tempo [4]. No geral, há informações de movimento e batimentos cardíacos provenientes de sensores como acelerômetro e fotopletiografia. Alguns estudos utilizam a extração de novas características e obtêm mais informações sobre esses e outros dados [3].



Além disso, duas formas de utilizar os dados são encontradas na literatura: *online*, no qual apenas os dados conhecidos até o momento atual são analisados; e *offline*, sendo possível utilizar dados em torno do momento atual e aplicação de pós-processamentos. Nesse sentido, é de interesse explorar os resultados que se pode gerar na combinação de características calculadas e modelos de aprendizado de máquina para este tipo de problema.

## 1.2 Questões de Pesquisa

Este trabalho busca investigar e responder as seguintes questões de pesquisa:

- Como combinar as informações dos dados de diferentes sensores de *smartwatches* para obter um método competitivo de detecção de sono/vigília?
- Como combinar as informações dos dados de diferentes sensores de *smartwatches* para obter um método competitivo de classificação dos estágios do sono?
- Qual o tipo de abordagem (*online* ou *offline*) é mais adequada para a detecção e classificação dos estágios do sono?

## 1.3 Objetivos

O principal objetivo deste trabalho é investigar técnicas de Inteligência Artificial para melhorar a precisão da detecção e classificação de distúrbios do sono. Para atingir esse propósito, alguns objetivos específicos devem ser alcançados:

- Explorar a combinação de dados dos diferentes sensores presentes em *smartwatches*;
- Desenvolver um detector e classificador de estágios do sono.

Este projeto visa contribuir com o desenvolvimento de uma metodologia aplicando Inteligência Artificial ligada aos desafios presentes na classificação de estágios do sono, a fim de obter um modelo para a tomada de decisão automática com dados disponíveis, e limitados, para esse tipo de problema.

## 1.4 Organização do Texto

Este projeto está organizado em 5 capítulos. No Capítulo 1, é descrito o problema de pesquisa a ser abordado neste projeto e também os principais objetivos e contribuições esperadas. No Capítulo 2, são introduzidos conceitos importantes para o entendimento do projeto e são apresentados alguns trabalhos relacionados relevantes existentes na literatura. No Capítulo 3, é abordada a metodologia, além das bases de dados e as métricas que serão utilizadas para avaliação dos modelos. No Capítulo 4, são mostrados o plano de trabalho e o cronograma de execução. O Apêndice A contém alguns resultados preliminares e discussões.

# Capítulo 2

## Fundamentação Teórica

Neste capítulo será apresentada a revisão bibliográfica feita para este trabalho. A Seção 2.1 contém os principais conceitos e técnicas relacionadas à definição de sono e à metodologia a ser abordada e, na Seção 2.2, serão analisados os trabalhos relacionados ao tema.

### 2.1 Conceitos e Técnicas

Os principais conceitos e técnicas empregados neste trabalho estão relacionados aos estágios do sono, às séries temporais e às redes neurais convolucionais e recorrentes. Esses conceitos e técnicas são apresentados nesta seção.

#### 2.1.1 Estágios do sono

Seguindo as diretrizes da AASM, que é o padrão mais recente, os estágios do sono são definidos como:

- **Vigília:** alerta de vigília para sonolência, sentir-se relaxado, adormecer rapidamente ou em poucos minutos;
- **NREM N1:** indica o início do sono, transição rápida e superficial. Ao ouvir um ruído não sentir vontade de responder a ele;
- **NREM N2:** menor frequência cardíaca e temperatura corporal. Os sinais de eletroencefalograma (EEG) continuam a diminuir em frequência enquanto aumentam em amplitude durante este período de sono leve;
- **NREM N3:** sono profundo, o tônus muscular, o pulso e a frequência respiratória diminuem. As ondas delta ocorrem no EEG;
- **REM:** paralisia temporária dos músculos, exceto dos olhos e os músculos que controlam a respiração. É conhecido como sono dos sonhos e a atividade cerebral aumenta, aproximando-se dos níveis observados durante a vigília.

### 2.1.2 Séries temporais

Uma série temporal  $T$  é definida como uma sequência ordenada de  $n$  variáveis de valor real

$$T = (t_1, \dots, t_n), t_i \in \mathbb{R}.$$

A série temporal é um conjunto de observações coletadas de forma sequencial, ao longo do tempo, e conforme uma determinada taxa de amostragem. Nesse sentido,  $T$  pode ser definida como um conjunto de instantes de tempo contíguos e pode abranger todo o conjunto de dados fornecidos pela observação de um processo e possuir comprimento considerável [13].

No contexto deste trabalho, as séries temporais são obtidas para toda a duração do registro de sono e as taxas de amostragem variam de acordo com o tipo de sensor utilizado para a coleta. Esses dados podem estar acompanhados com *timestamps* em diferente formatos (por exemplo: horário da coleta, tempo desde a PSG e formato *Unix*). A classificação de estágios do sono é baseada no padrão de ondas cerebrais, na atividade muscular no mento e no oculograma analisados a cada período de 30 segundos, que é denominado "época" [46]. As séries temporais de gravações de PSG são padronizadas com épocas de 30s e cada uma delas possui uma *label* correspondente a um estágio do sono, que será utilizada como validação. É importante destacar que os tempos de início e fim das séries temporais dos diferentes tipos de dados (de sensores e PSG) devem estar sincronizados.

### 2.1.3 Redes Neurais Convolucionais

Redes Neurais Convolucionais (CNNs) é uma técnica de aprendizado profundo que pertence ao grupo das Redes Neurais Artificiais (RNAs). As CNNs são redes neurais para processamento de dados que possuem uma topologia em grade como, por exemplo, séries temporais (que podem ser consideradas como uma grade 1D de amostras coletadas em intervalos de tempo regulares) e imagens (que podem ser vistas como uma grade 2D de pixels). A convolução é um tipo especializado de operação linear e é utilizada no lugar da multiplicação de matrizes, feita pelas redes neurais, em pelo menos uma de suas camadas. A arquitetura CNN unidimensional (CNN 1D) é composta por múltiplas camadas, tendo as camadas de entrada compostas por sinais de séries temporais [44]. As CNNs 1D são comumente aplicadas a séries temporais para capturar características espaciais ao longo da dimensão temporal dos sinais de sensores [26].

### 2.1.4 Redes Neurais Recorrentes

Redes Neurais Recorrentes (RNNs) possuem conexões entre os neurônios artificiais que podem gerar um ciclo, fazendo com que a saída de alguns neurônios artificiais influencie na entrada subsequente para os mesmos neurônios, isso permite que a rede exiba um comportamento dinâmico temporal. As RNNs conseguem utilizar seu estado interno para processar sequências de comprimento variável de entradas e geram uma única camada oculta ao fazerem um *loop* sobre ela quantas vezes forem necessárias, em vez de criar

várias camadas ocultas. Assim como uma CNN, que é especializada em processar uma grade de valores  $X$  (como, por exemplo, uma imagem), uma RNN é uma rede neural especializada em processar uma sequência de valores  $x(1), \dots, x(\tau)$ . Dentre as variantes deste tipo de rede, destacam-se a LSTM [20] e GRU [9].

As redes LSTM podem aprender dependências de longo prazo, pois possuem um componente denominado célula de memória, que retém o estado da célula ao longo do tempo. As GRUs são RNNs mais sofisticadas já que evitam o problema do gradiente de fuga, que é a incapacidade de propagar informações úteis de gradiente da extremidade de saída do modelo de volta para as camadas próximas à extremidade de entrada. Esse problema é evitado pelas GRUs por meio da atualização e redefinição de *gates*. A atualização de um *gate* ajuda o modelo a determinar quanta informação de etapas de tempo anteriores deve ser transferida para etapas de tempo futuras.

## 2.2 Trabalhos Relacionados

Nesta seção serão apresentados trabalhos que buscam detectar sono e vigília e outros que classificam os estágios do sono. Na classificação, alguns autores simplificam o problema em distinguir apenas três estágios (vigília, NREM e REM) e quatro (vigília, sono leve, sono profundo e REM). A classificação que segue a diretriz da AASM, em cinco estágios (vigília, N1, N2, N3 e REM), também é estudada em trabalhos relacionados. Como poderá ser observado na discussão dos estudos, é possível que mais de uma variação do problema seja abordada e comparada. Todos os trabalhos citados a seguir utilizam classificação a cada 30 segundos e, quando não mencionado, a forma de validação são gravações PSG. Outras abordagens também serão apresentadas para mostrar métodos que lidam com o mesmo tipo de dados a ser utilizado neste trabalho mas, com aplicações em problemas diferentes.

### Detecção de vigília/sono

O problema de detecção em dois estágios simplifica o sono NREM e REM em uma única fase e os trabalhos buscam diferenciar o sono da vigília. A actigrafia de pulso comparada com a PSG para detecção de vigília/sono foi abordada em diversos trabalhos [11,35,37,49].

O estudo de Paquet *et al.* [35] avalia a actigrafia para detectar vigília/sono utilizando dados de 23 indivíduos consumidores moderados de cafeína (um a três copos de café por dia). Foi utilizado o *Actiwatch-L* (*Mini Mitter, Respironics Inc.*), que contém um acelerômetro 3D. O movimento foi convertido em contagem de atividades e as medidas em torno da época pontuada foram obtidas de modo *offline*: atividade média em uma janela de 7 épocas; desvio padrão em uma janela de 8 épocas; número de contagens de atividade acima de 10 em uma janela de 11 épocas; e o logaritmo natural na época marcada. Com um classificador de Regressão Logística, os autores atingiram uma acurácia de 90,3% (sensibilidade de 94,6% e especificidade de 47,3%) e sob os dados de sono em efeito do uso de cafeína, a acurácia obtida foi de 76,8% (sensibilidade de 95,3% e especificidade de 48,5%). Com a baixa especificidade, os autores apontam que o algoritmo encontra dificuldades em detectar vigílias.

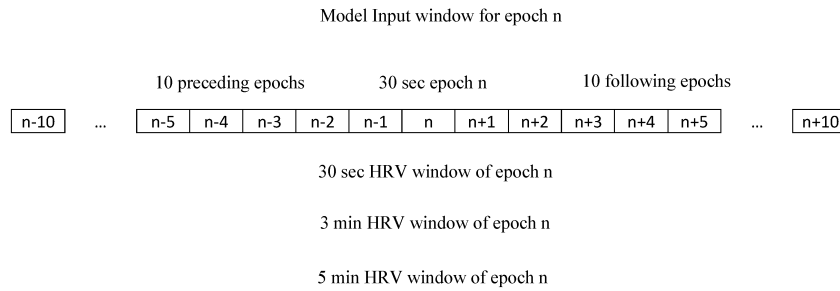


Figura 2.1: Exemplo ilustrativo de diferentes comprimentos de janela utilizados para calcular as métricas de VFC da época  $n$  [19].

Sano e Picard [41] coletaram dados de aceleração, condutância da pele (SC), temperatura da pele (ST) por meio de um sensor de pulseira e dados de EEG, de gravações de PSG, de 15 indivíduos para realizar uma comparação da classificação de vigília/sono. Foram computadas seis características do acelerômetro após aplicação de um filtro passa-banda de 2 a 3Hz, sendo elas:

- Número de cruzamentos de zero;
- Função "D" de Cole [11];
- Média e desvio padrão;
- Máximo e mínimo da raiz quadrada média dos dados de aceleração dos três eixos.

As características foram calculadas de modo *online* para cada época de 30s. Além das características do acelerômetro, o modelo teve como entrada 16 características de EEG; 7 de SC e 5 de ST. Os autores obtiveram melhores resultados com o classificador SVM de kernel gaussiano e com todas as características extraídas (EEG+SC+ST+acelerômetro), atingindo 85,0% de acurácia. Ao utilizarem características apenas do sensor de pulseira, o melhor resultado obtido foi uma acurácia de 74,0% (SC+ST+acelerômetro e ST+acelerômetro). Um dos pontos mais interessantes desse estudo é que, além do movimento, a ST desempenhou um papel importante na classificação de sono/vigília quando os sinais EEG são desconsiderados.

Haghighy *et al.* [19] utilizaram os dados de PSG e actigrafia de punho do estudo MESA *Sleep* [8,51]. O sinal de eletrocardiografia (ECG) da PSG foi utilizado para extrair as métricas de variabilidade da frequência cardíaca (VFC), totalizando 17 características (9 no domínio do tempo e 8 no domínio da frequência) e a actigrafia forneceu um contador de atividades. Neste estudo, foram comparados diferentes comprimentos de janela para o cálculo da VFC: janelas de 30 segundos, 3 minutos e 5 minutos. Uma ilustração do cálculo realizado utilizando diferentes tamanhos de janelas é apresentada na Figura 2.1, nota-se que as *features* foram obtidas de modo *offline*. O modelo que obteve melhores resultados foi um DeepConvLSTM, composto por nove camadas convolucionais e quatro camadas LSTM, apresentando acurácia de 84,5% (sensibilidade de 89,5% e especificidade de 70,0%). É importante destacar que a contagem de atividades combinada com as métricas de VFC calculadas por janelas de 5 minutos produziu a maior concordância com a PSG.

O uso de actigrafia para a coleta de dados é o mais presente em trabalhos que estudam apenas a detecção de vigília e sono. A AASM indicou a actigrafia como um método adequado para auxiliar na avaliação de pacientes com distúrbios do sono [16]. Como a actigrafia coleta essencialmente dados de movimento, essa informação é importante para atingir bons resultados nesse tipo de classificação. Outros trabalhos, que serão apresentados, realizam a detecção de vigília/sono além da classificação de mais estágios do sono [3, 16, 48].

## Classificação em três estágios

Neste tipo de problema, o objetivo é diferenciar vigília, sono NREM e sono REM.

Um estudo realizado com os dispositivos *Apple Watch* coletou dados de aceleração e frequência cardíaca do acelerômetro MEMS e do fotopletismógrafo de pulso óptico (PPG) de 31 indivíduos [48]. Os modelos treinados na base de dados criada, denominada *Sleep Accel*, foram também testados em um subconjunto dos dados disponibilizados publicamente, *MESA Sleep* [6], para avaliar a generalização dos algoritmos. Os autores utilizaram um contador de atividades calculado após aplicar um filtro passa-banda Butterworth de 5ª ordem entre 3 e 11Hz [45], e o desvio padrão dos dados de frequência cardíaca da janela acima (10 minutos) da época válida como características de entrada do modelo. Nesse sentido, a extração de *features* é feita de modo *online*. Além disso, o estudo investiga o uso de um recurso denominado *proxy* de relógio, que representa o relógio circadiano. Foi realizada uma comparação com os algoritmos de Regressão Logística, *K-Nearest Neighbors* (KNN), *Random Forest* e uma rede neural para os problemas de classificação de vigília/sono e de três estágios. Os melhores resultados obtidos com os dados *Sleep Accel* foram acurácias de 90,0% ( $\kappa = 0,45$ ) e 72,0% ( $\kappa = 0,30$ ) para a identificação de vigília/sono e os três estágios, respectivamente. Ao utilizar o subconjunto dos dados *MESA Sleep*, obteve-se  $\kappa = 0,52$  para a detecção de vigília/sono e  $\kappa = 0,40$  para a classificação dos estágios.

Fonseca *et al.* [16] utilizaram o sistema *Philips Respironics Alice PDx* (*Philips Respironics Inc.*) para gravações PSG de 101 indivíduos e o *Actiwatch Spectrum* (*Philips Respironics Inc.*) para validar um dispositivo de registro que contém PPG e acelerômetro de três eixos (*Royal Philips*). Para isso, o treinamento foi feito com as características extraídas do eletrocardiograma (ECG) de um subconjunto dos dados SIESTA [27] e na validação e teste foram utilizadas as informações do PPG e acelerômetro. O conjunto de características de VFC inclui características no domínio de tempo e no domínio de frequência e são extraídas de forma *offline* dos sinais de ECG. Essas características foram combinadas com as relacionadas ao movimento, que não são descritas no trabalho. Utilizando um classificador linear Bayesiano desenvolvido pelos autores em um trabalho anterior [15], a acurácia atingiu 91,5% ( $\kappa = 0,55$ ) para a detecção de vigília/sono, e para três e quatro estágios, o modelo obteve acurácias de 72,9% ( $\kappa = 0,46$ ) e 59,3% ( $\kappa = 0,42$ ), respectivamente.

No geral, os autores que estudam a classificação em três estágios apontam que as diferenças dos classificadores não são particularmente pronunciadas, mas a inclusão de características afeta significativamente o desempenho. Também é chamada a atenção para

que, embora apenas a frequência cardíaca não melhore muito o desempenho em relação ao movimento sozinho na classificação sono/vigília, ela desempenha um papel significativo na classificação vigília/NREM/REM. Alguns autores que classificam em quatro ou cinco estágios do sono também apresentam resultados utilizando três [28], como poderá ser observado nas seções seguintes.

## Classificação em quatro estágios

Por ser considerado difícil distinguir N1 e N2, a classificação em quatro estágios do sono busca diferenciar vigília, sono leve (N1 e N2), sono profundo (N3) e REM.

Beattie *et al.* [5] relataram a precisão da estimativa dos estágios do sono utilizando o relógio *Fitbit* (*Fitbit Surge*, *Fitbit*), que mede o movimento por meio de acelerômetro 3D e possui PPG, em 60 indivíduos (36 homens e 24 mulheres). As características foram extraídas de forma *offline*, sendo as de movimento:

- Contagem de atividade;
- Máximo e mínimo de cada eixo;
- Tempo desde o último movimento significativo;
- Tempo até o próximo movimento significativo.

Os sinais de PPG foram pré-processados por meio de um algoritmo detector de picos. As características de potências HF (*High Frequency*), LF (*Low Frequency*) e VLF (*Very Low Frequency*) foram calculadas para as frequências cardíaca e respiratória. Com a frequência cardíaca também foram extraídas as características de RMSSD, pNN50, delta RR, média, percentil 10 e percentil 90. Ao considerar essas e variantes dessas informações, que foram obtidas em diferentes janelas de tempo, os autores definiram 180 características. Um classificador discriminante linear (LDA) alcançou melhores resultados, tendo uma precisão geral de 69,0% e  $\kappa = 0,52$ . Uma questão interessante é que os autores realizam um pós-processamento que penaliza padrões fisiológicos improváveis como, por exemplo, ver uma única época de vigília isolada durante um período de épocas de sono profundo.

No estudo de Altini e Kinnunen [3] foram coletados dados fisiológicos de 106 indivíduos (Singapura, Finlândia e Estados Unidos) por meio do anel vestível *Oura* (primeira geração, *Oura Health*), obtendo dados do acelerômetro, temperatura, frequência cardíaca e VFC, além do cálculo de informações circadianas independentes de sensor. Sob os dados do acelerômetro foram calculadas estatísticas descritivas em cada eixo, após aplicar o filtro Butterworth [45]. As características foram extraídas de modo *offline*, usando janelas deslizantes de diferentes comprimentos e incluem:

- Média aparada (após retirar 10% de cada lado);
- Intervalo interquartil de cada eixo (janelas de 30s);
- Desvio de amplitude (janelas de 5s) e
- Diferença no ângulo do braço (janelas de 5s).

Com os dados de temperatura, foi aplicado uma etapa de rejeição de artefatos (valores fora da faixa de 31 a 40 graus Celsius não são marcados); e foram calculados os valores de média, mínimo, máximo e desvio padrão. As características extraídas dos sinais do PPG são semelhantes às obtidas por Beattie *et al.* [5], após aplicar um filtro de média móvel em tempo real para localização dos valores máximos e mínimos locais para denotar o tempo de cada batimento cardíaco. O algoritmo implementado no anel *Oura* rotula cada intervalo entre batimentos como normal ou anormal usando filtros medianos [25]. Neste estudo, são utilizadas as seguintes informações circadianas: uma função cosseno, também aplicada por Walch *et al.* [48], um decaimento exponencial e uma função linear. A função cosseno representa o componente de acionamento circadiano; a exponencial é a queda da pressão homeostática do sono ao longo da noite; e a linear indica o tempo decorrido desde o início da noite, variando de 0 a 1. Um exemplo dessas três características pode ser visualizado na Figura 2.2.

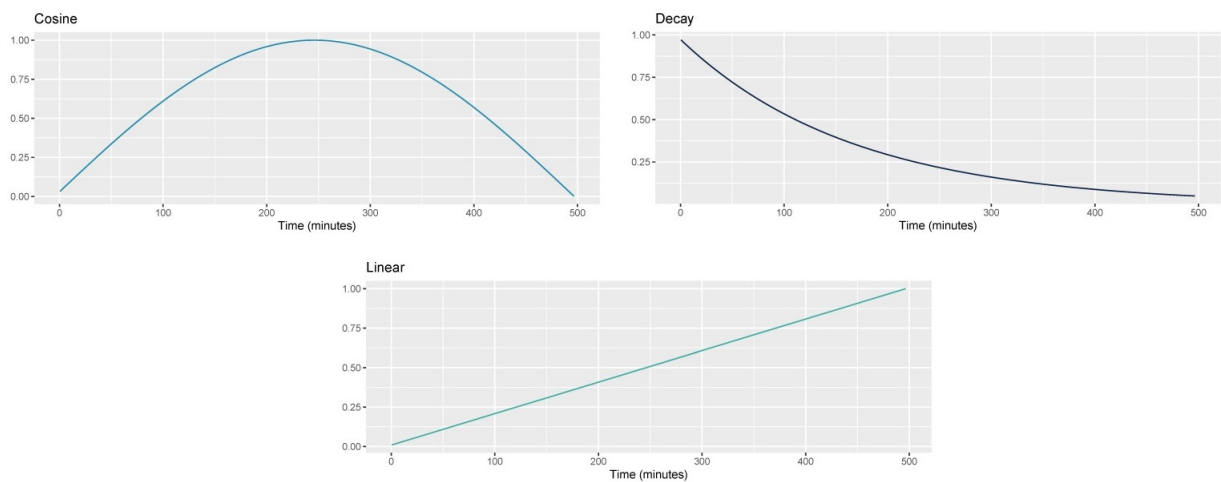


Figura 2.2: Funções cosseno, exponencial e linear utilizadas para modelar características circadianas [3].

Os autores [3] utilizam um classificador *Light Gradient Boosting Machine* (LightGBM) [24], com *Dropouts meet Multiple Additive Regression Trees* (DART) [47]. A precisão para detecção de sono e vigília aumentou de 94,0% ( $F_1 = 0,67$ ) para 96,0% ( $F_1 = 0,78$ ) ao utilizar características do PPG e circadianas, além das informações do acelerômetro. A precisão para detectar quatro estágios foi de 57,0% ( $F_1 = 0,68$ ) para o modelo com informações apenas do acelerômetro e 79,0% ( $F_1 = 0,78$ ) ao incluir os demais recursos.

Uma abordagem diferente é apresentada por Radha *et al.* [38], na qual é utilizada transferência de aprendizado de um modelo LSTM. Primeiro, o modelo é treinado com os dados SIESTA, que é um grande conjunto de dados de ECG disponível publicamente, e depois adaptado usando um pequeno conjunto de dados de ECG e PPG, descrito por Fonseca *et al.* [16]. A VFC é uma substituição fortemente estudada para a PSG. Nesse sentido, a transferência de aprendizado é capacitada pelo fato de que a VFC pode ser obtida por meio do PPG incorporado em relógios vestíveis. Para realizar a transferência de aprendizado é necessário um grande conjunto de dados, mas os dados disponíveis estão nas diretrizes R&K (mais antiga) e, por isso, os autores também utilizam um conjunto



de dados apresentado no trabalho anterior, que estão nas diretrizes AASM. A partir de dados de VFC calculados de sinais de ECG e PPG, um total de 127 características foram extraídas de modo *offline*, tendo em vista que foi utilizada uma janela deslizante centrada em cada época de 30s e as épocas no início e no final de cada gravação, por estarem nas bordas, foram marcadas como inválidas; os valores para essas épocas foram interpolados [15]. As *features* definidas incluem:

- 37 de domínio do tempo;
- 14 de domínio de frequência;
- 28 de entropia e auto-semelhança;
- 22 de energia *Teager*;
- 13 de gráficos de visibilidade;
- 8 de coordenação de fase e
- 5 de estatísticas descritivas.

É importante destacar que os autores não justificam a escolha dessas *features*. O modelo foi pré-treinado em todo o conjunto de dados SIESTA e avaliado nos sinais de ECG e PPG do conjunto de dados menor. Com isso, os autores alcançaram uma acurácia de 76,4% e  $\kappa = 0,65$ .

Um dos pontos mais importantes dos estudos de classificação em quatro estágios é que a actigrafia pode fornecer algumas informações sobre os estágios do sono (como vigília e sono), mas não pode prever com precisão todos os estágios. Ao adicionar os dados fisiológicos (frequência cardíaca e frequência respiratória, por exemplo) capturados pelo PPG e explorar ainda mais as características fornecidas por esse sensor, os estágios do sono podem ser detectados com precisão razoável.

## Classificação em cinco estágios

Seguindo as diretrizes da AASM, a classificação em cinco estágios do sono é definida em diferenciar vigília, N1, N2, N3 e REM.

Korkalainen *et al.* [28] desenvolveram um método baseado em CNN e RNN para identificar os estágios do sono a partir do sinal de PPG obtido com um oxímetro de pulso de dedo (*Nonin Xpod 3011*). O estudo foi realizado sob 894 indivíduos com suspeita de apneia obstrutiva do sono (AOS) e o desempenho foi demonstrado com três, quatro e cinco estágios. Um dos objetivos secundários do trabalho é não realizar a extração e seleção prévia de características. Nesse sentido, o modelo desenvolvido utiliza CNN para aprender as características de cada estágio do sono, enquanto a RNN foi utilizada para considerar a distribuição temporal dos estágios durante a noite. A arquitetura de rede foi a mesma para todos os modelos de classificação (três, quatro e cinco estágios), compreendendo em seis convoluções 1D (Conv1D), normalização em lote e função de ativação de unidade linear retificada (ReLU). Após a CNN, segue-se uma camada *dropout* gaussiana, unidade

recorrente fechada bidirecional (GRU) e uma camada densa com uma função de ativação *softmax*, sendo  $n$  o número de estágios do sono a serem classificados. A arquitetura com mais detalhes é ilustrada na Figura 2.3.

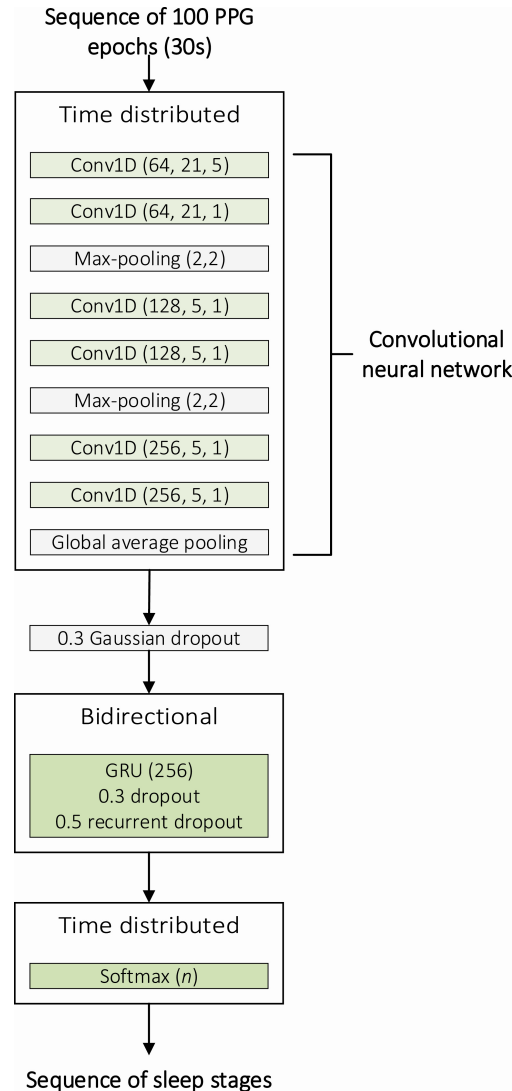


Figura 2.3: Arquitetura de rede CNN e RNN utilizada por Korkalainen *et al.* [28].

Os resultados gerais obtidos foram acurácias de 80,1% ( $\kappa = 0,65$ ) para a classificação de três estágios, e 68,5% ( $\kappa = 0,54$ ) e 64,1% ( $\kappa = 0,51$ ) para os modelos de quatro e cinco estágios, respectivamente. Os autores destacam que a baixa concordância com a pontuação manual de PSG de N1 é uma limitação para a classificação das fases do sono baseado em PPG. Outros trabalhos também chegaram a conclusão de que no geral, o estágio N1 é difícil de distinguir [14].

## Outros problemas

Outros problemas envolvendo dados de sensores de *smartwatches* são abordados na literatura. Dentre eles, é importante destacar os trabalhos que desenvolveram redes neurais para aprender informações dos dados provenientes dos acelerômetros e giroscópio para re-

alizer o reconhecimento e classificação da atividade humana (HAR). Algumas abordagens recentes para esse problema serão apresentados a seguir.

A rede neural residual profunda unidimensional (*1D-ResNet*) [31] atingiu acurácia geral de 94,15%. Essa rede possui camadas convolucionais e conexões residuais para obter informações por conta própria e também camadas de *batch normalization* e ReLU, como mostrado na Figura 2.4.

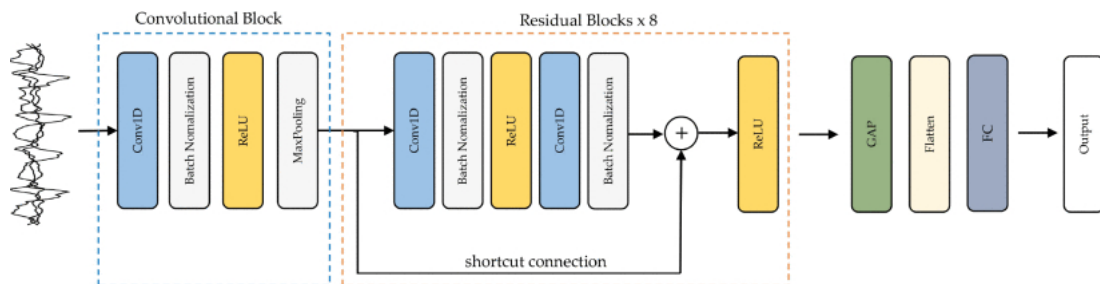


Figura 2.4: Estrutura do modelo *1D-ResNet* [31].

A rede *WISNet* [44] é uma abordagem personalizada de aprendizado profundo (CNN 1D) para reconhecer atividades humanas complexas, esse modelo foi comparado com as redes GRU (95,3% de acurácia), LSTM (95,2% de acurácia) e RNN (91,8% de acurácia) para classificar e detectar o conjunto de dados WISDM de seis classes [29], atingindo uma acurácia de 96,4%. A representação do modelo *WISNet* pode ser observada na Figura 2.5. Outra proposta para HAR utilizando o conjunto de dados WISDM é descrita por Imran [23], que desenvolveu a rede *UltraNet* com uma unidade recorrente fechada bidirecional (BiGRU) com 64 unidades, como pode ser observada na Figura 2.6. O modelo atingiu acurácia de 97,7% sob a base de dados utilizada.

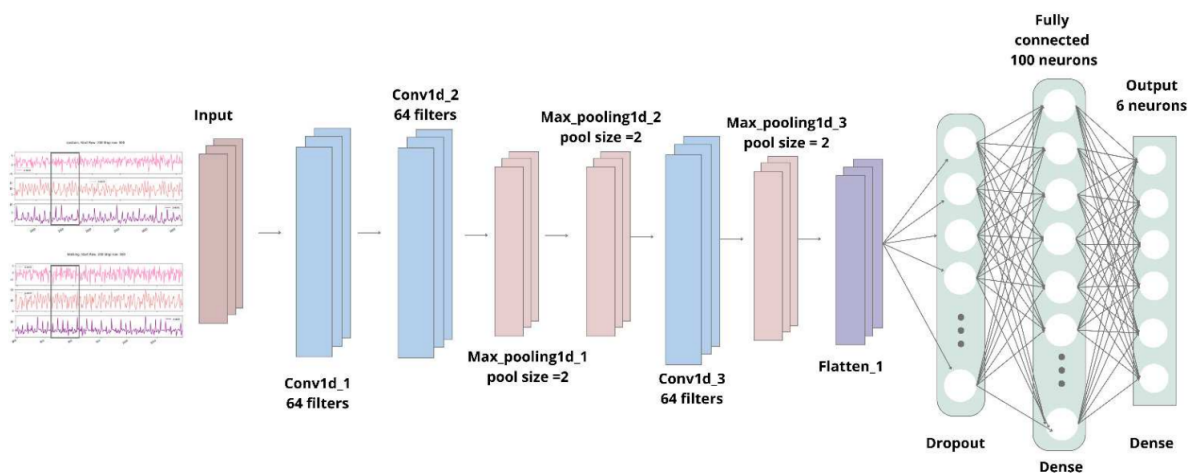


Figura 2.5: Estrutura do modelo *WISNet* [44].

É importante ressaltar que embora essas abordagens não buscam resolver o problema a ser explorado neste trabalho, os métodos desenvolvidos pelos autores são estudos recentes para lidar com dados de sensores dos *smartwatches*, destacando o acelerômetro. Nota-se que as abordagens mencionadas utilizam CNNs 1D na modelagem e os trabalhos fazem comparações com RNNs.

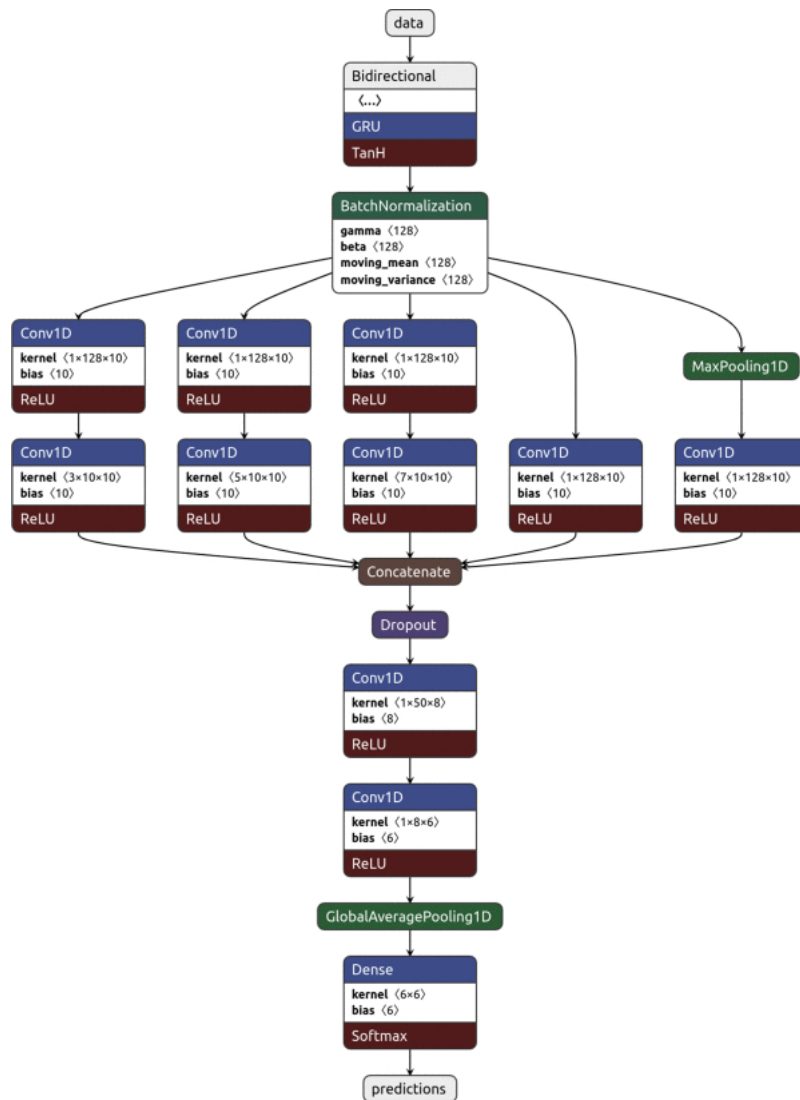


Figura 2.6: Estrutura do modelo *UltraNet* [23].

# Capítulo 3

## Materiais e Métodos

Neste capítulo são descritas a metodologia, as métricas e as bases de dados disponíveis publicamente que serão utilizadas para avaliação e validação do modelo.

### 3.1 Metodologia

A metodologia para detectar e classificar os estágios do sono consiste principalmente em explorar métodos que consideram as informações espaciais e temporais dos dados e também em combinar os diferentes dados dos sensores de *smartwatches* (como acelerômetro e PPG).

A exploração direta dos dados brutos (abordagem direcionada a dados), ou seja, sem a extração de características, permite que um modelo aprenda propriedades distinguíveis dos dados e resolva tarefas específicas. O aprendizado profundo extrai automaticamente características dos dados de entrada sem a intervenção humana, fazendo com que o modelo capture aspectos que podem ser ignorados pela seleção de característica realizada por humanos [43]. Como os dados utilizados neste trabalho estão relacionados a diversas variáveis fisiológicas, em especial os dados de PPG [28], escolhemos utilizar a metodologia de aprendizado profundo e abordagem direcionada a dados para analisar os sinais dos sensores de *smartwatches* sem seleção prévia de características, tendo em vista que padrões muito sutis obtidos diretamente dos dados podem não ser extraídos manualmente.

Como esquematizado na Figura 3.1, após a aquisição dos dados (*Sleep Accel* e *MESA*), será realizado um pré-processamento, com filtragem e normalização dos dados brutos. As amostras serão divididas em treino, validação e teste. Para obter as características espaciais de cada estágio do sono, será explorada a utilização de CNNs 1D, como, por exemplo, a *WaveNet* [34]. Já para analisar a distribuição temporal dos estágios do sono serão empregadas RNNs, como por exemplos arquiteturas LSTM e GRU, que já foram utilizadas em trabalhos relacionados [19,28]. Além da exploração de métodos e combinação de dados dos diferentes sensores, a metodologia inclui:

- Aquisição e preparação das bases de dados disponíveis publicamente;
- Combinar e comparar a utilização de dados dos diferentes sensores de *smartwatches* (como acelerômetro e PPG);

- Estudo de geração de novas *features* a partir dos dados disponibilizados;
- Realização de experimentos que possam mitigar o desequilíbrio de classes;
- Exploração de soluções baseadas nas técnicas do estado da arte e nos resultados das investigações realizadas;
- Comparação entre as abordagens *online* e *offline*;
- Comparação do método proposto com resultados obtidos em trabalhos relacionados.

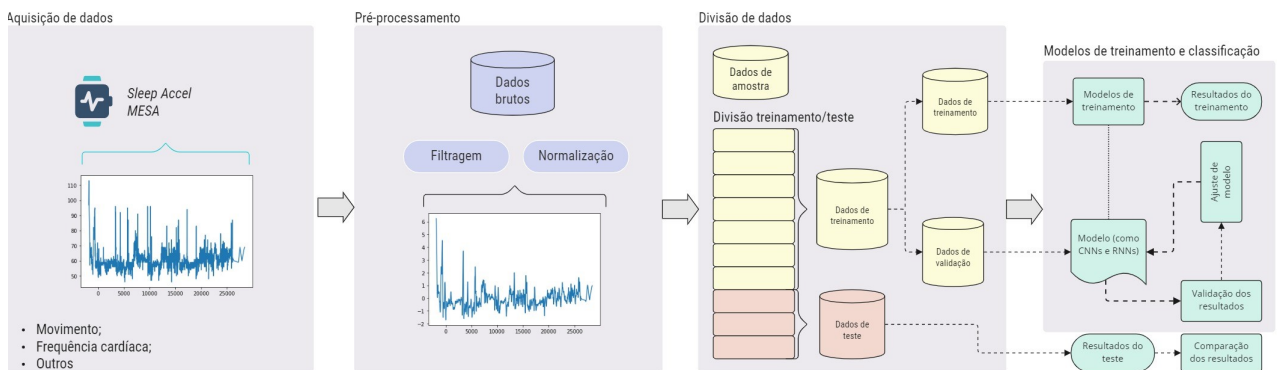


Figura 3.1: Esquema de metodologia a ser utilizada.

## 3.2 Métricas de Avaliação

A fim de facilitar a comparação do desempenho do modelo proposto neste projeto com os encontrados na literatura, serão utilizadas as métricas comumente apresentadas nos trabalhos da área. Além da acurácia, essas medidas incluem sensibilidade, especificidade, precisão,  $F_1$  e  $\kappa$  (kappa) de Cohen.

Embora os trabalhos relacionados apresentem a acurácia como principal medida de desempenho do modelo, neste projeto será utilizada a acurácia balanceada, tendo em vista que os dados contêm classes desbalanceadas e que a acurácia apresenta uma estimativa otimista quando um classificador é testado em um conjunto de dados desequilibrado [7]. Como pode ser observado na Equação 3.2, a acurácia balanceada é a média das sensibilidades (*recall*) de todas as  $c$  classes. Na Equação 3.1,  $VP_n$  é o número de verdadeiros positivos da classe  $n$  e  $FN_n$  é o número de falsos positivos da classe  $n$ .

$$sensibilidade_n = \frac{VP_n}{VP_n + FN_n} \quad (3.1)$$

$$acurácia\ balanceada = \frac{\sum_{n=1}^{n=c} sensibilidade_n}{c} \quad (3.2)$$

A precisão de  $c$  classes é definida conforme a Equação 3.3, sendo  $FP_n$  os falsos positivos da classe  $n$ .

$$precisão = \frac{\sum_{n=1}^{n=c} VP_n}{\sum_{n=1}^{n=c} (VP_n + FP_n)} \quad (3.3)$$

A Equação 3.4 descreve a métrica  $F_1$  de uma classe  $n$ .

$$F_{1_n} = 2 \times \frac{\text{precisão}_n \times \text{sensibilidade}_n}{\text{precisão}_n + \text{sensibilidade}_n} \quad (3.4)$$

Na detecção vigília/sono, a sensibilidade é definida como a porcentagem de períodos de sono corretamente rotulados como sono. Já a especificidade é a porcentagem de épocas de vigília corretamente rotuladas como vigília.

A medida  $\kappa$  (kappa) de Cohen é a concordância entre diferentes avaliadores para problemas multiclases e classes desequilibradas. Neste contexto, os avaliadores são as rotulagens de PSG e o algoritmo de classificação automática dos estágios do sono. A definição de  $\kappa$  feita por Cohen [10] é descrita conforme a Equação 3.5.

$$\kappa = \frac{\% \text{Concordância observada} - \% \text{Concordância por acaso}}{1 - \% \text{Concordância por acaso}} \quad (3.5)$$

Os valores de  $\kappa$  são interpretados como:

- $\kappa \leq 0,00$ : nenhuma concordância;
- $0,00 < \kappa \leq 0,20$ : concordância leve;
- $0,20 < \kappa \leq 0,40$ : concordância regular;
- $0,40 < \kappa \leq 0,60$ : concordância moderada;
- $0,60 < \kappa \leq 0,80$ : concordância substancial; e
- $0,80 < \kappa \leq 1,00$ : concordância quase perfeita.

### 3.3 Conjuntos de Dados

Este projeto utilizará duas bases de dados disponíveis publicamente: *Sleep Accel*, coletada com *Apple Watch*<sup>1</sup> (séries 2 e 3, *Apple Inc.*) [48], e *MESA Sleep*<sup>2</sup>, que contém dados de actigrafia de pulso provenientes do *Actiwatch Spectrum* (*Philips Respironics*) [6]. Ambos os conjuntos serão descritos a seguir.

O conjunto de dados *Sleep Accel* foi coletado na Universidade de Michigan entre junho de 2017 e março de 2019, e há dados de 31 indivíduos. Os dados são contagem de passos, aceleração e frequência cardíaca registradas com o *Apple Watch*, além dos rótulos dos estágios do sono pontuados com gravações de PSG conforme o padrão da AASM. Os participantes deste estudo não possuem diagnóstico conhecido de distúrbios do sono. Para cada dado obtido, há uma informação de tempo (em segundos desde o início da PSG). O *Apple Watch* usa um acelerômetro MEMS triaxial, que mede a aceleração nas direções  $x$ ,  $y$  e  $z$  (em  $g$ ), e um PPG que fica na face dorsal do pulso, que obtém a frequência cardíaca em batimentos por minuto (bpm). O *Apple Watch* foi utilizado durante 7 a 14 dias e,

<sup>1</sup><https://physionet.org/content/sleep-accel/1.0.0/>

<sup>2</sup><https://sleepdata.org/datasets/mesa>

no último dia, os participantes passaram a noite em um laboratório de sono para realizar uma gravação de PSG, com 8 horas de duração, enquanto utilizavam o *smartwatch*.

O *Multi-Ethnic Study of Atherosclerosis* (MESA) é um estudo apoiado pelo *National Heart, Lung, and Blood Institute* (NHLBI), que investiga fatores associados ao desenvolvimento de doença cardiovasculares. *MESA Sleep* é um projeto publicado pelo *National Sleep Research Resource* (NSRR) que recrutou 2237 participantes, de diferentes etnias e regiões dos Estados Unidos, para usarem o *Actiwatch Spectrum* entre 2010 e 2013. Os participantes não possuem doenças cardiovasculares e nem utilizam oxigênio ou dispositivo de pressão positiva nas vias aéreas (PAP) durante a noite. Os dados são, basicamente, a contagem de atividades coletada pela actigrafia e gravações de PSG, que foram pontuadas de acordo com as diretrizes da AASM. As gravações de PSG incluem uma noite completa de sono de cada participante e a actigrafia de pulso foi utilizada por 7 dias.

### 3.4 Ferramentas e Recursos Computacionais

A implementação deste projeto será realizada na linguagem Python, tendo em vista que há um grande número de ferramentas desenvolvidas com suporte para essa linguagem e com boa documentação. O projeto utilizará bibliotecas científicas, numéricas, de redes neurais, de apresentação de gráficos, entre outras. Algumas bibliotecas que podem ser destacadas são: NumPy<sup>3</sup>, pandas<sup>4</sup>, Matplotlib<sup>5</sup>, scikit-learn<sup>6</sup> e TensorFlow<sup>7</sup>.

Os experimentos deste projeto serão realizados em dois ambientes: no Laboratório de Inteligência Artificial (Recod.ai) do Instituto de Computação (IC – Unicamp); e na plataforma Google Colab<sup>8</sup>, que fornece recursos de computação em nuvem gratuitos, tais como CPU, GPU, TPU, memória RAM e armazenamento.

---

<sup>3</sup><https://numpy.org/>

<sup>4</sup><https://pandas.pydata.org/>

<sup>5</sup><https://matplotlib.org/>

<sup>6</sup><https://scikit-learn.org/>

<sup>7</sup><https://www.tensorflow.org/>

<sup>8</sup><https://colab.research.google.com/>



## Capítulo 4

# Plano de Trabalho e Cronograma de Execução

O plano de trabalho é composto pelas seguintes atividades:

1. Obtenção dos créditos obrigatórios em disciplinas do programa de mestrado.
2. Revisão bibliográfica.
3. Estudo e seleção de bases de dados.
4. Exame de Qualificação do Mestrado (EQM).
5. Preparação das bases de dados.
6. Seleção de modelos como ponto de partida.
7. Definição de um novo modelo.
8. Avaliação e aprimoramento do modelo.
9. Realização de testes e análise dos resultados.
10. Participação no Programa de Estágio Didático (PED).
11. Documentação e publicação dos resultados.
12. Escrita do documento da dissertação.
13. Defesa da Dissertação de Mestrado.



## Referências Bibliográficas

- [1] K. Aggarwal, S. Khadanga, S. Joty, L. Kazaglis, and J. Srivastava. A structured learning approach with neural conditional random fields for sleep staging. In *IEEE International Conference on Big Data (Big Data)*, pages 1318–1327. IEEE, 2018.
- [2] H. M. Al-Angari and A. V. Sahakian. Automated recognition of obstructive sleep apnea syndrome using support vector machine classifier. *IEEE Transactions on Information Technology in Biomedicine*, 16(3):463–468, 2012.
- [3] M. Altini and H. Kinnunen. The promise of sleep: A multi-sensor approach for accurate sleep stage detection using the oura ring. *Sensors*, 21(13):4302, 2021.
- [4] G. Atluri, A. Karpatne, and V. Kumar. Spatio-temporal data mining: A survey of problems and methods. *ACM Computing Surveys*, 51(4):1–41, 2018.
- [5] Z. Beattie, Y. Oyang, A. Statan, A. Ghoreyshi, A. Pantelopoulos, A. Russell, and C. Heneghan. Estimation of sleep stages in a healthy adult population from optical plethysmography and accelerometer signals. *Physiological Measurement*, 38(11):1968–1979, 2017.
- [6] D. E. Bild, D. A. Bluemke, G. L. Burke, R. Detrano, A. V. Diez Roux, A. R. Folsom, P. Greenland, D. R. Jacobs Jr, R. Kronmal, K. Liu, J. C. Nelson, D. O. Leary, M. F. Saad, S. Shea, M. Szklo, and R. P. Tracy. Multi-ethnic study of atherosclerosis: objectives and design. *American Journal of Epidemiology*, 156(9):871–881, 2002.
- [7] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann. The balanced accuracy and its posterior distribution. In *International Conference on Pattern Recognition*, pages 3121–3124. IEEE, 2010.
- [8] X. Chen, R. Wang, P. Zee, P. L. Lutsey, S. Javaheri, C. Alcántara, C. L. Jackson, M. A. Williams, and S. Redline. Racial/ethnic differences in sleep disturbances: the Multi-Ethnic Study of Atherosclerosis (MESA). *Sleep*, 38(6):877–888, 2015.
- [9] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.
- [10] J. Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46, 1960.

- [11] R. J. Cole, D. F. Kripke, W. Gruen, D. J. Mullaney, and J. C. Gillin. Automatic sleep/wake identification from wrist activity. *Sleep*, 15(5):461–469, 1992.
- [12] L. F. Drager, G. Lorenzi-Filho, F. D. Cintra, R. P. Pedrosa, L. R. Bittencourt, D. Poyares, C. G. Carvalho, S. M. G. P. T. Moura, R. Santos-Silva, P. F. De Bruin, G. R. Geovanini, F. N. Albuquerque, W. A. A. De Oliveira, G. A. Moreira, L. M. Ueno, F. B. Nerbass, M. U. P. B. Rondon, E. R. F. Barbosa, A. Bertolami, A. A. V. De Paola, B. B. S. Marques, C. F. Rizzi, C. E. Negrão, C. H. G. Uchôa, C. Maki-Nunes, D. Martinez, E. A. Fernández, F. U. Maroja, F. R. Almeida, I. C. Trombetta, L. J. Storti, L. A. Bortolotto, M. T. de Mello, M. A. Borges, M. L. Andersen, N. d. P. Portilho, P. Macedo, R. Alves, S. Tufik, S. C. Fagondes, and T. T. Risso. 1<sup>o</sup> posicionamento brasileiro sobre o impacto dos distúrbios de sono nas doenças cardiovasculares da sociedade brasileira de cardiologia. *Arquivos Brasileiros de Cardiologia*, 111:290–340, 2018.
- [13] P. Esling and C. Agon. Time-series data mining. *ACM Computing Surveys (CSUR)*, 45, 12 2012.
- [14] S. Fallmann and L. Chen. Computational sleep behavior analysis: A survey. *IEEE Access*, 7:142421–142440, 2019.
- [15] P. Fonseca, X. Long, M. Radha, R. Haakma, R. M. Aarts, and J. Rolink. Sleep stage classification with ECG and respiratory effort. *Physiological Measurement*, 36:2027, 8 2015.
- [16] P. Fonseca, T. Weysen, M. S. Goelema, E. I. Møst, M. Radha, C. Lunsingh Scheurleer, L. van den Heuvel, and R. M. Aarts. Validation of photoplethysmography-based sleep staging compared with polysomnography in healthy middle-aged adults. *Sleep*, 40(7), 2017.
- [17] K. Fujimoto, Y. Ding, and E. Takahashi. Sleep stage detection using a wristwatch-type physiological sensing device. *Sleep and Biological Rhythms*, 16(4):449–456, 2018.
- [18] M. Hafner, M. Stepanek, J. Taylor, W. M. Troxel, and C. Van Stolk. Why sleep matters — the economic costs of insufficient sleep: a cross-country comparative analysis. *RAND Health Quarterly*, 6(4), 2017.
- [19] S. Haghayegh, S. Khoshnevis, M. H. Smolensky, K. R. Diller, and R. J. Castriotta. Deep neural network sleep scoring using combined motion and heart rate variability data. *Sensors*, 21(1):25, 2020.
- [20] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.
- [21] V. Ibáñez, J. Silva, and O. Cauli. A survey on sleep assessment methods. *PeerJ*, 6:e4849, 2018.
- [22] C. Iber. The AASM manual for the scoring of sleep and associated events: Rules. *Terminology and Technical Specification*, 2007.

- [23] H. A. Imran. Ultanet: An antithesis neural network for recognizing human activity using inertial sensors signals. *IEEE Sensors Letters*, 6, 1 2022.
- [24] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu. LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30, 2017.
- [25] H. Kinnunen, A. Rantanen, T. Kentt, and H. K. ki. Feasible assessment of recovery and cardiovascular health: accuracy of nocturnal HR and HRV assessed via ring PPG in comparison to medical grade ECG. *Physiological Measurement*, 41:04NT01, 5 2020.
- [26] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman. 1d convolutional neural networks and applications: A survey. *Mechanical Systems and Signal Processing*, 151:107398, 4 2021.
- [27] G. Klash, B. Kemp, T. Penzel, A. Schlogl, P. Rappelsberger, E. Trenker, G. Gruber, J. Zeithofer, B. Saletu, W. Herrmann, S. Himanen, D. Kunz, M. Barbanoj, J. Roschke, A. Varri, and G. Dorffner. The SIESTA project polygraphic and clinical database. *IEEE Engineering in Medicine and Biology Magazine*, 20(3):51–57, 2001.
- [28] H. Korkalainen, J. Aakko, B. Duce, S. Kainulainen, A. Leino, S. Nikkonen, I. O. Afara, S. Myllymaa, J. Töyräs, and T. Leppänen. Deep learning enables sleep staging from photoplethysmogram for patients with suspected sleep apnea. *Sleep*, 43(11):zsaa098, 2020.
- [29] J. R. Kwapisz, G. M. Weiss, and S. A. Moore. Activity recognition using cell phone accelerometers. *ACM SIGKDD Explorations Newsletter*, 12:74–82, 3 2011.
- [30] J. Li, M. V. Vitiello, and N. S. Gooneratne. Sleep in normal aging. *Sleep Medicine Clinics*, 13(1):1–11, 2018.
- [31] S. Mekruksavanich, P. Jantawong, and A. Jitpattanakul. Recognition of complex human activities for wellness management from smartwatch using deep residual neural network. In *International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON)*, pages 350–353. IEEE, 2022.
- [32] S. S. Mostafa, F. Mendonça, F. Morgado-Dias, and A. Ravelo-García. SpO2 based sleep apnea detection using deep learning. In *IEEE International Conference on Intelligent engineering Systems (INES)*, pages 91–96. IEEE, 2017.
- [33] J. Newell, O. Mairesse, P. Verbanck, and D. Neu. Is a one-night stay in the lab really enough to conclude? First-night effect and night-to-night variability in polysomnographic recordings among different clinical population samples. *Psychiatry Research*, 200(2-3):795–801, 2012.

- [34] A. v. d. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu. Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*, 2016.
- [35] J. Paquet, A. Kawinska, and J. Carrier. Wake detection capacity of actigraphy during sleep. *Sleep*, 30(10):1362–1369, 2007.
- [36] A. Patanaik, J. L. Ong, J. J. Gooley, S. Ancoli-Israel, and M. W. Chee. An end-to-end framework for real-time automatic sleep stage classification. *Sleep*, 41(5):zsy041, 2018.
- [37] C. P. Pollak, W. W. Tryon, H. Nagaraja, and R. Dzwonczyk. How accurately does wrist actigraphy identify the states of sleep and wakefulness? *Sleep*, 24(8):957–965, 2001.
- [38] M. Radha, P. Fonseca, A. Moreau, M. Ross, A. Cerny, P. Anderer, X. Long, and R. M. Aarts. A deep transfer learning approach for wearable sleep stage classification with photoplethysmography. *npj Digital Medicine*, 4(1):1–11, 2021.
- [39] A. Rechtschaffen. A manual for standardized terminology, techniques and scoring system for sleep stages in human subjects. *Brain Information Service*, 1968.
- [40] A. Roebuck, V. Monasterio, E. Geder, M. Osipov, J. Behar, A. Malhotra, T. Penzel, and G. Clifford. A review of signals used in sleep analysis. *Physiological Measurement*, 35(1):R1, 2013.
- [41] A. Sano and R. W. Picard. Comparison of sleep-wake classification using electroencephalogram and wrist-worn multi-modal sensor data. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 930–933. IEEE, 2014.
- [42] M. J. Sateia. International classification of sleep disorders. *Chest*, 146(5):1387–1394, 2014.
- [43] H. T. Sencar, L. Verdoliva, and N. Memon. What’s in this book and why? In *Multimedia Forensics*, pages 3–6. Springer, Singapore, 2022.
- [44] H. Sharen, P. Rukmani, L. J. Anbarasi, and A. H. Gandomi. WISNet: A deep neural network based human activity recognition system. *Research Square (preprint)*, 2022.
- [45] B. H. Te Lindert and E. J. Van Someren. Sleep estimates using Microelectromechanical Systems (MEMS). *Sleep*, 36(5):781–789, 2013.
- [46] S. M. G. P. Togeiro and A. K. Smith. Métodos diagnósticos nos distúrbios do sono. *Brazilian Journal of Psychiatry*, 27:8–15, 2005.
- [47] R. K. Vinayak and R. Gilad-Bachrach. DART: Dropouts meet multiple additive regression trees. In *Artificial Intelligence and Statistics (AISTATS)*, pages 489–497. PMLR, 2015.

- [48] O. Walch, Y. Huang, D. Forger, and C. Goldstein. Sleep stage prediction with raw acceleration and photoplethysmography heart rate data derived from a consumer wearable device. *Sleep*, 42(12):zsz180, 2019.
- [49] J. B. Webster, D. F. Kripke, S. Messin, D. J. Mullaney, and G. Wyborney. An activity-based sleep monitor system for ambulatory use. *Sleep*, 5(4):389–399, 1982.
- [50] X. Wu, J. Yang, Y. Pan, X. Zhang, and Y. Luo. Automatic sleep-stage scoring based on photoplethysmographic signals. *Physiological Measurement*, 41(6):065008, 2020.
- [51] G.-Q. Zhang, L. Cui, R. Mueller, S. Tao, M. Kim, M. Rueschman, S. Mariani, D. Mogley, and S. Redline. The national sleep research resource: towards a sleep data commons. *Journal of the American Medical Informatics Association*, 25(10):1351–1358, 2018.

# Apêndice A

## Resultados Preliminares

Neste capítulo serão apresentados os principais resultados obtidos utilizando a base de dados *Sleep Accel*, descrita na Seção 3.3.

O trabalho de Walch *et al.* [48], citado na seção de Trabalhos Relacionados (Seção 2.2), foi a principal referência para a realização dos experimentos, tendo em vista que os autores disponibilizam a base de dados e o código da implementação, além de apresentar um estudo de classificação em três estágios do sono. Os testes foram feitos com o objetivo de estudar o conjunto de dados e entender algumas dificuldades que podem ser encontradas neste trabalho. Os experimentos consistiram em, primeiro reproduzir os resultados obtidos pelos autores [48] e testar novas estratégias para melhorar o modelo de detecção de vigília/sono e, em seguida, realizar o mesmo para o problema de classificação em três estágios do sono. É importante destacar que os experimentos realizados têm como intuito estudar a modelagem do problema, sem se preocupar em aplicar, por enquanto, classificadores sofisticados de aprendizado de máquina.

Os autores [48] utilizam Regressão Logística, Random Forest, KNN e uma rede neural como classificadores e os seguintes conjuntos de *features* para realizar a detecção de vigília/sono e classificação em três estágios do sono:

1. Apenas movimento;
2. Apenas frequência cardíaca (FC);
3. Movimento e frequência cardíaca (FC);
4. Movimento, frequência cardíaca (FC) e tempo.

Durante a reprodução dos resultados, notou-se que os algoritmos Random Forest e KNN apresentaram *overfitting* com as configurações utilizadas pelos autores [48]. Como o KNN e a rede neural exigiram mais tempo de processamento, decidimos utilizar os classificadores Regressão Logística, sem alterar os parâmetros escolhidos pelos os autores [48], e Random Forest, após ajustar os parâmetros a fim de reduzir o efeito de *overfitting*, para a realização dos experimentos.

É importante destacar que os autores [48], assim como os demais trabalhos desta área, apresentam os resultados apenas por meio da acurácia mas, como a base de dados é desbalanceada, calculamos também a acurácia balanceada, que é a métrica mais adequada



para esse tipo de base. A validação foi feita com os dados de PSG e épocas de 30s e os conjuntos de treinamento e teste foram divididos em aproximadamente 70% e 30% do total dos dados, respectivamente, no trabalho de referência. Mesmo não tendo um conjunto de validação, decidimos utilizar a mesma divisão dos autores [48] para comparação. A divisão foi feita aleatoriamente 20 vezes para obter a média dos resultados e as amostras de um único indivíduo não foram usadas simultaneamente nos conjuntos de treinamento e teste.

## A.1 Descrição dos experimentos

Nesta seção será apresentada uma descrição geral de cada experimento feito, sendo eles: novos conjuntos de *features*, inclusão de *features* de períodos passados, aumento de dados e estudo de *features* de movimento com os três eixos do acelerômetro. É importante ressaltar que os experimentos foram realizados primeiramente no contexto da detecção de vigília/sono e, em seguida, na classificação em três estágios do sono.

### A.1.1 Novos conjuntos de *features*

Além dos resultados obtidos com os conjuntos de *features* apresentados pelos autores no artigo de referência [48], foram incluídos os seguintes conjuntos de *features*, totalizando 7 conjuntos diferentes:

1. Apenas tempo;
2. Movimento e tempo;
3. Frequência cardíaca (FC) e tempo.

### A.1.2 Inclusão de *features* de períodos passados

Esse experimento utiliza *features* de períodos passados para classificar o período atual, simulando uma memória no classificador. Os testes foram feitos de modo *online*, então as informações extras são de  $n$  períodos anteriores ao atual (que será classificado), de modo que as *features* de cada período  $c$  sejam compostas pelas *features* indicadas na Equação A.1. Caso não exista uma  $Feature_x$  ( $x < 0$ ), a  $Feature_0$  (a primeira *feature* existente) é utilizada no seu lugar. Um exemplo com  $n = 2$  pode ser observado nas Tabelas A.1 e A.2.

$$Feature_c = [Feature_{c-n}, Feature_{c-n+1}, \dots, Feature_{c-1}, Feature_c] \quad (A.1)$$

Foram realizados testes com os classificadores Regressão Logística e Random Forest considerando apenas *features* de movimento, com  $n$  entre 0 e 128, em saltos de potência de 2, e com  $n$  entre 10 e 100, em saltos de 10.

Período	<i>Features</i>
0	31.640606
1	22.570999
2	19.341522
3	21.960207

Tabela A.1: Exemplo de *features* de movimento, sem considerar períodos passados.

Período	<i>Features</i>		
0	31.640606	31.640606	<b>31.640606</b>
1	31.640606	31.640606	<b>22.570999</b>
2	31.640606	22.570999	<b>19.341522</b>
3	22.570999	19.341522	<b>21.960207</b>

Tabela A.2: Exemplo de *features* de movimento, considerando  $n = 2$  períodos passados.

### A.1.3 Aumentação de dados

Tendo em vista que a base de dados *Sleep Accel* é desbalanceada, tanto em relação às classes do problema de detecção de vigília/sono quanto às de classificação em três estágios do sono, foi realizada uma aumento de dados partindo do pressuposto que 10 horas antes do início das gravações de PSG, os indivíduos estavam acordados (vigília). O gráfico da Figura A.1 mostra a relação, em porcentagem, a cada hora de dados das classes de vigília/sono antes da aumento, que apresenta proporção de 12 (sono) para 1 (vigília). Aplicando a técnica de aumento, os dados apresentam proporção de 7 (sono) para 1 (vigília).

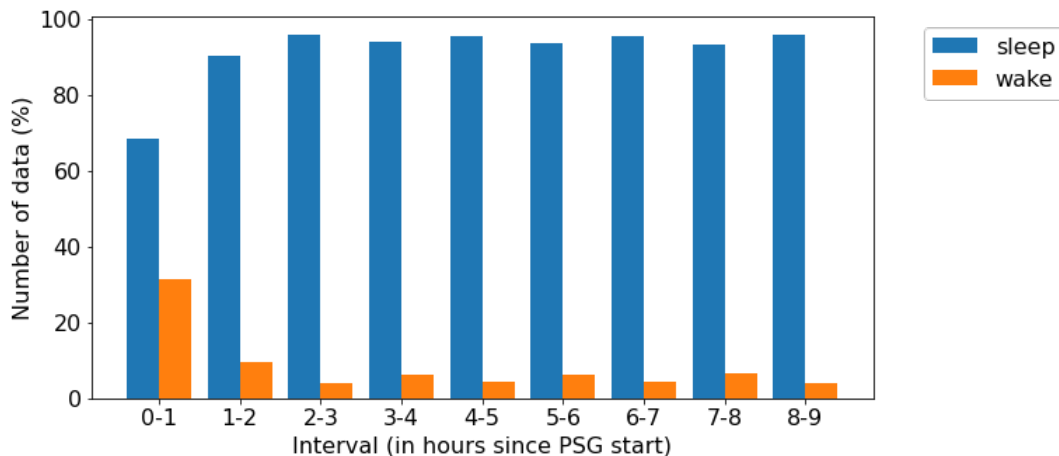


Figura A.1: Quantidade de dados de vigília e sono da base de dados *Sleep Accel* a cada hora após o início das gravações de PSG.

### A.1.4 *Features* de movimento com os eixos x, y e z

Walch *et al.* [48] converteram o movimento em um contador de atividades para a compatibilidade com os actígrafos. Entretanto, o cálculo de conversão desses sinais utilizam

apenas os dados do eixo z coletados pelo acelerômetro. Nesse sentido, foi realizado um estudo que consiste em considerar as informações de movimento nos três eixos, que foram denominadas como contador x, contador y e contador z. As três *features* geradas foram testadas individualmente e em conjunto. Também foram geradas *features* de movimento considerando diversos valores estatísticos, como média, desvio padrão, entropia, entre outros considerando os dados brutos para cada intervalo de 30 segundos.

## A.2 Resultados e discussões

Nesta seção são apresentados alguns resultados e discussões, primeiro no contexto de detecção de vigília/sono e, depois, na classificação em três estágios do sono.

### Detecção de vigília/sono

A Tabela A.3 mostra a comparação dos resultados das médias de 20 iterações da acurácia e acurácia balanceada na detecção de vigília/sono utilizando o classificador Regressão Logística e diferentes conjuntos de *features* e com o melhor resultado obtido no experimento com *features* de períodos passados ( $n = 80$ ).

Regressão Logística			
	Conjunto de <i>features</i>	Acurácia	Ac. balanceada
Sem períodos passados	Apenas movimento	0.84	0.73
	Apenas FC	0.67	0.64
	Apenas tempo	0.80	0.65
	Movimento e FC	0.85	0.74
	Movimento e tempo	<b>0.87</b>	0.74
	FC e tempo	0.77	0.74
	Movimento, FC e tempo	<b>0.87</b>	0.77
Com 80 períodos passados	Apenas movimento	0.85	<b>0.79</b>

Tabela A.3: Comparação dos resultados obtidos com  $n = 80$  períodos passados utilizando Regressão Logística considerando apenas *features* de movimento com os demais resultados, que não consideram dados passados.

Nota-se que Regressão Logística obteve melhor acurácia balanceada quando foram considerados dados passados ( $n = 80$ ), melhorando o resultado em 6 pontos percentuais quando se considera apenas o movimento e em 2 pontos percentuais quando se considera todas as *features*. O mesmo resultado de 79% de acurácia balanceada é obtido com a aumento de dados utilizando apenas o movimento, como pode ser observado na Tabela A.4, que compara os experimentos feitos utilizando aumento de dados e o realizado apenas com os dados originais.

Assim como a Regressão Logística, os melhores resultados obtidos utilizando o classificador Random Forest foram com o conjunto que contém todas as *features*, como pode ser observado na Tabela A.5. O melhor resultado obtido com o experimento de dados passados também foi com  $n = 80$ , mostrando que apenas as informações de movimento podem ser suficientes para uma detecção razoável de vigília/sono. No caso dos resultados obtidos

Conjunto de <i>features</i>	Dados originais		Com aumento de dados	
	Acurácia	Ac. balanceada	Acurácia	Ac. balanceada
Apenas movimento	0.84	0.73	<b>0.88</b>	<b>0.79</b>
Apenas FC	0.67	0.64	0.68	0.67
Apenas tempo	0.80	0.65	0.45	0.54
Movimento e FC	0.85	0.74	<b>0.88</b>	<b>0.79</b>
Movimento e tempo	<b>0.87</b>	0.74	<b>0.88</b>	<b>0.79</b>
FC e tempo	0.77	0.74	0.68	0.66
Movimento, FC e tempo	<b>0.87</b>	<b>0.77</b>	<b>0.88</b>	<b>0.79</b>

Tabela A.4: Comparação dos resultados obtidos com e sem a aumento de dados, com o classificador Regressão Logística.

com o Random Forest, as *features* que consideram 80 períodos passados melhoraram a acurácia balanceada, quando se considera apenas o movimento, em 4 pontos percentuais e atingiram a acurácia balanceada do conjunto que inclui todas as *features*. O experimento realizado com a aumento de dados superou a acurácia balanceada desses resultados com todas as *features*, como pode ser observado na Tabela A.6.

Random Forest			
Sem períodos passados	Conjunto de <i>features</i>	Acurácia	Ac. balanceada
	Apenas movimento	0.76	0.77
	Apenas FC	0.54	0.65
	Apenas tempo	<b>0.87</b>	0.64
	Movimento e FC	0.75	0.77
	Movimento e tempo	0.80	0.80
	FC e tempo	0.82	0.74
	Movimento, FC e tempo	0.81	<b>0.81</b>
	Com 80 períodos passados	Apenas movimento	0.77

Tabela A.5: Comparação dos resultados obtidos com  $n = 80$  períodos passados utilizando Random Forest considerando apenas *features* de movimento com os demais resultados, que não consideram dados passados.

Conjunto de <i>features</i>	Dados originais		Com aumento de dados	
	Acurácia	Ac. balanceada	Acurácia	Ac. balanceada
Apenas movimento	0.76	0.77	0.81	0.81
Apenas FC	0.54	0.65	0.67	0.68
Apenas tempo	<b>0.87</b>	0.64	0.73	0.70
Movimento e FC	0.75	0.77	0.82	0.82
Movimento e tempo	0.80	0.80	0.82	0.82
FC e tempo	0.82	0.74	0.76	0.71
Movimento, FC e tempo	0.81	<b>0.81</b>	<b>0.83</b>	<b>0.83</b>

Tabela A.6: Comparação dos resultados obtidos com e sem a aumento de dados, com o classificador Random Forest.

É possível observar que quando consideramos o conjunto mais completo de *features*

(movimento, FC e tempo), ambos os classificadores apresentaram melhor acurácia balanceada, mas nota-se que a principal *feature* para detectar vigília/sono é o movimento, tendo em vista que esse é o conjunto que atingiu as melhores acurácias balanceadas, dentre os que não combinam mais de uma *feature* (apenas frequência cardíaca e apenas tempo).

Antes de apresentar os resultados obtidos com as técnicas de adição de dados passados e de aumento de dados, é importante exibir alguns dos resultados alcançados com o estudo dos três eixos do acelerômetro, tendo em vista que os resultados apresentados até aqui utilizam um contador de atividades como *features* de movimento, obtido apenas com os dados do eixo z.

A Figura A.2 mostra a importância de cada *feature* determinada pelo classificador Random Forest. Além do contador, outras *features* de movimento, a maioria baseada em valores estatísticos dos dados dos acelerômetros, foram testadas. Nota-se que o contador nos três eixos são importantes para o classificador e que as outras *features* não apresentam importância significativa. Nesse sentido, os demais experimentos consideram as *features* de movimento como três números: contadores de movimentos nos eixos x, y e z.

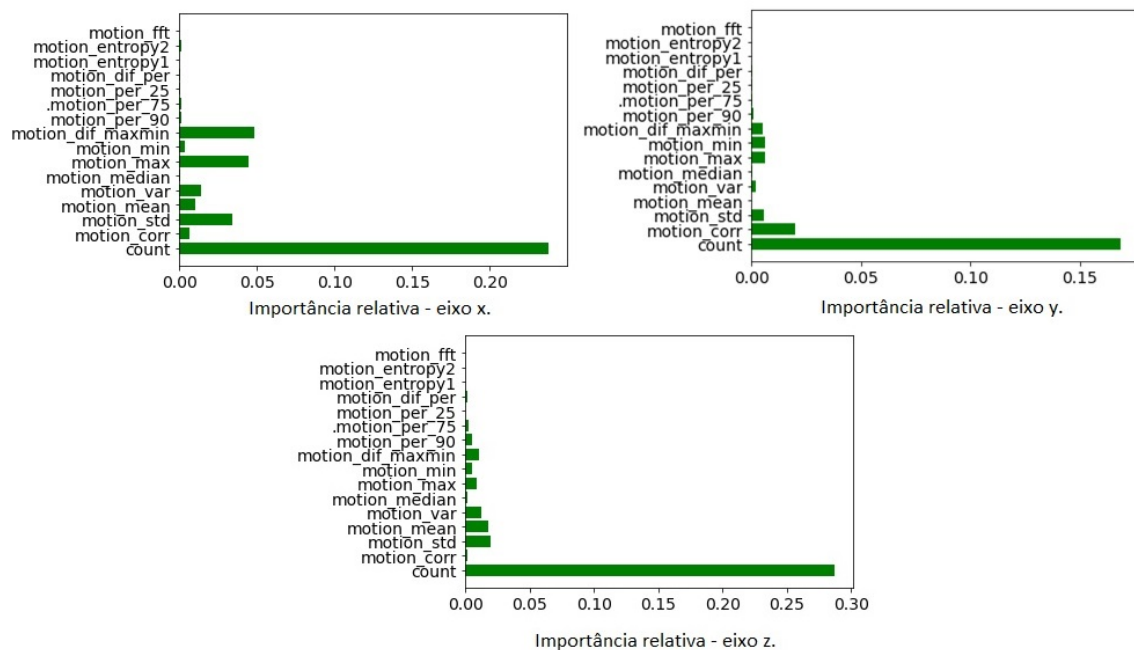


Figura A.2: Importância relativa de cada *feature* de movimento determinada pelo classificador Random Forest.

Os resumos da comparação das acurácias balanceadas dos experimentos utilizando apenas movimento com os classificadores Regressão Logística e Random Forest são apresentados nas Tabelas A.7 e A.8, respectivamente. Ao adicionar os dados que foram gerados com a técnica de aumento de dados, a acurácia balanceada já atinge os melhores resultados com  $n = 30$  dados passados com ambos os classificadores.

As principais conclusões obtidas por meio dos resultados apresentados são a importância do papel da *feature* de movimento na detecção de vigília/sono e que a aumento de dados e a utilização de dados passados, feitas de forma separada, apresentaram melhoria de desempenho de forma semelhante e, quando em conjunto, melhoram significativamente

Configuração	Acurácia		Ac. balanceada	
	Média	Desvio	Média	Desvio
Original	0.87	0.019	0.73	0.025
Com dados passados	0.87	0.023	0.78	0.034
Com aumento de dados	0.89	0.013	0.78	0.018
Com dados passados e aumento de dados	<b>0.91</b>	0.009	<b>0.85</b>	0.023

Tabela A.7: Comparação das acurácias obtidas em diferentes experimentos realizados utilizando Regressão Logística e apenas movimento, com  $n = 30$  dados passados (quando aplicável).

Configuração	Acurácia		Ac. balanceada	
	Média	Desvio	Média	Desvio
Original	0.75	0.042	0.77	0.025
Com dados passados	0.80	0.031	0.80	0.022
Com aumento de dados	0.85	0.018	0.80	0.015
Com dados passados e aumento de dados	<b>0.86</b>	0.028	<b>0.86</b>	0.015

Tabela A.8: Comparação das acurácias obtidas em diferentes experimentos realizados utilizando Random Forest e apenas movimento, com  $n = 30$  dados passados (quando aplicável).

o modelo. É importante ressaltar a diferença de resultados exibidos pelos dois tipos de acurácia, enquanto a acurácia balanceada geralmente apresenta os menores valores, a acurácia utilizada pelos autores [48] encontra valores otimistas para determinar o desempenho dos modelos.

## Classificação em três estágios

Tendo como base os testes realizados no problema de detecção sono/vigília, as melhores configurações encontradas foram utilizadas para a classificação em três estágios do sono, ou seja, a aumento de dados em conjunto com dados passados ( $n = 30$ ) e as *features* de movimento baseadas nos contadores dos três eixos do acelerômetro foram aplicados. A Tabela A.9 compara os resultados obtidos com e sem essas configurações utilizando Regressão Logística.

Conjunto de <i>features</i>	Original		Com dados passados e aumento de dados	
	Acurácia	Ac. balanceada	Acurácia	Ac. balanceada
Apenas movimento	<b>0.70</b>	0.45	<b>0.72</b>	0.54
Apenas FC	0.67	0.41	0.66	0.46
Apenas tempo	<b>0.70</b>	0.34	0.66	0.44
Movimento e FC	0.68	0.47	0.71	0.59
Movimento e tempo	<b>0.70</b>	0.45	<b>0.72</b>	0.55
FC e tempo	0.67	0.45	0.66	0.49
Movimento, FC e tempo	0.68	<b>0.50</b>	0.71	<b>0.61</b>

Tabela A.9: Comparação dos resultados obtidos pela Regressão Logística com a configuração original e com a utilização de aumento de dados e  $n = 30$  dados passados.

Pode-se observar que para o problema de classificação em três estágios, a frequência cardíaca apresentou um papel importante já que ao adicionar esse conjunto com o movimento, a acurácia balanceada aumentou significativamente, diferente do analisado na detecção de vigília/sono. Walch *et al.* [48] também apontaram sobre isso e destacaram que esse comportamento traz sentido para as variáveis biológicas e natureza de cada problema. As matrizes de confusão para os dois casos, comparados na Tabela A.9, com todas as *features* são apresentadas na Figura A.3.

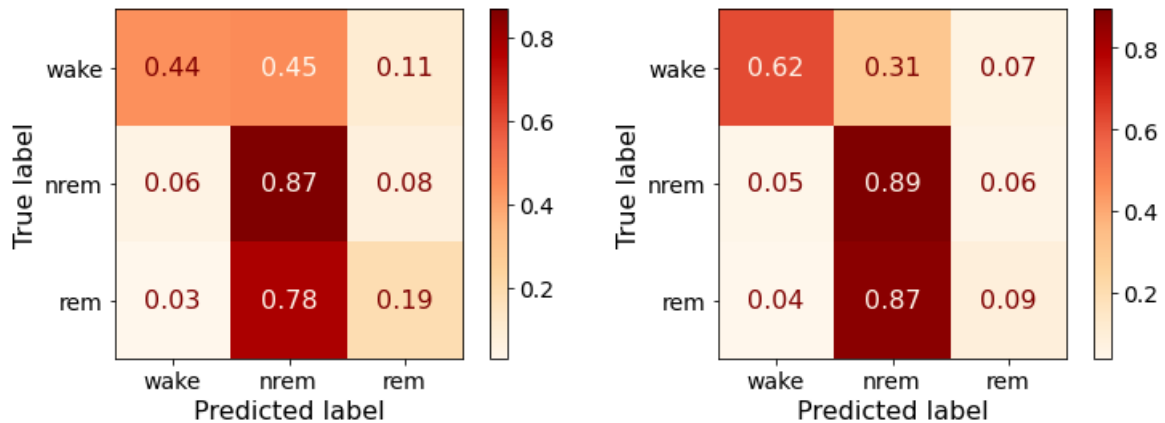


Figura A.3: Matrizes de confusão para as classificações sem e com aumento de dados e adição de dados passados com Regressão Logística considerando o conjunto com todas as *features* (movimento, FC e tempo).

Os resultados obtidos com o classificador Random Forest podem ser visualizados na Tabela A.10. As matrizes de confusão geradas para o conjunto de todas as *features* podem ser visualizadas na Figura A.4.

Conjunto de <i>features</i>	Original		Com dados passados e aumento de dados	
	Acurácia	Acurácia balanceada	Acurácia	Acurácia balanceada
Apenas movimento	0.55	0.52	<b>0.56</b>	0.58
Apenas FC	0.53	0.49	0.51	0.56
Apenas tempo	0.44	0.47	0.48	0.50
Movimento e FC	0.55	0.59	0.55	<b>0.67</b>
Movimento e tempo	0.47	0.55	0.53	0.56
FC e tempo	0.56	0.57	0.53	0.59
Movimento, FC e tempo	<b>0.59</b>	<b>0.63</b>	0.55	<b>0.67</b>

Tabela A.10: Comparação dos resultados obtidos pelo Random Forest com a configuração original e com a utilização de aumento de dados e  $n = 30$  dados passados.

É possível concluir que embora os resultados de classificação em três estágios não apresente resultados satisfatórios em nenhum dos cenários, já pode-se notar algumas melhorias com a aplicação dos experimentos apresentados. Ao analisar as matrizes de confusão, infere-se que a Regressão Logística melhorou a classificação do estágio NREM e que ambos os classificadores apresentaram melhorias na classificação das vigílias, isso pode ter ocorrido por conta, principalmente, da aumento de dados realizada.

Dessa maneira, os experimentos realizados permitem observar algumas dificuldades e possibilidades de soluções que podem ser encontradas com a abordagem de tratamento

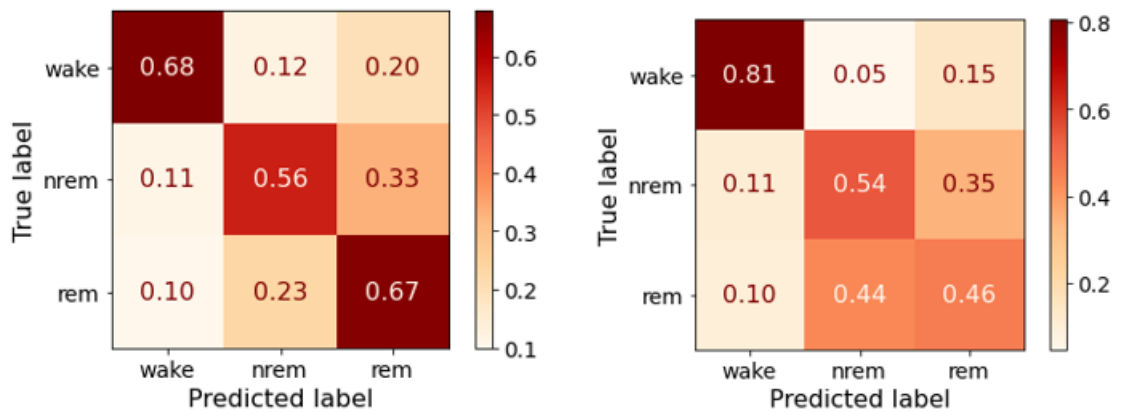


Figura A.4: Matrizes de confusão para as classificações sem e com aumento de dados e adição de dados passados com Random Forest considerando o conjunto com todas as *features* (movimento, FC e tempo).

de dados de forma *online* e uma visão geral de como são as bases de dados e do problema como um todo. Isso traz um direcionamento à investigação de novas metodologias que contribuam para uma melhor resolução dos problemas de classificação a serem tratados e à busca de métodos que permitam a superação das limitações que serão encontradas em cada abordagem.