



Universidade Estadual de Campinas
Instituto de Computação



Rebeca Padovani Ederli

Representações Visuais para Classificação de Estágios do Sono

CAMPINAS
2024

Rebeca Padovani Ederli

**Representações Visuais para
Classificação de Estágios do Sono**

Dissertação apresentada ao Instituto de Computação da Universidade Estadual de Campinas como parte dos requisitos para a obtenção do título de Mestra em Ciência da Computação.

Orientador: Prof. Dr. Zandoni Dias

Coorientador: Prof. Dr. Anderson de Rezende Rocha

Este exemplar corresponde à versão final da Dissertação defendida por Rebeca Padovani Ederli e orientada pelo Prof. Dr. Zandoni Dias.

CAMPINAS
2024

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Ana Regina Machado - CRB 8/5467

Ed28r Ederli, Rebeca Padovani, 1999-
Representações visuais para classificação de estágios do sono / Rebeca Padovani Ederli. – Campinas, SP : [s.n.], 2024.

Orientador: Zanoni Dias.
Coorientador: Anderson de Rezende Rocha.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Computação.

1. Inteligência artificial. 2. Sono. 3. Classificação de imagem. 4. Relógios inteligentes. I. Dias, Zanoni, 1975-. II. Rocha, Anderson de Rezende, 1980-. III. Universidade Estadual de Campinas. Instituto de Computação. IV. Título.

Informações Complementares

Título em outro idioma: Visual representations for sleep stages classification

Palavras-chave em inglês:

Artificial intelligence

Sleep

Image classification

Smartwatches

Área de concentração: Ciência da Computação

Titulação: Mestra em Ciência da Computação

Banca examinadora:

Zanoni Dias [Orientador]

George Darmiton da Cunha Cavalcanti

Marcos Medeiros Raimundo

Data de defesa: 02-05-2024

Programa de Pós-Graduação: Ciência da Computação

Identificação e informações acadêmicas do(a) aluno(a)

- ORCID do autor: <https://orcid.org/0009-0008-8283-0812>

- Currículo Lattes do autor: <http://lattes.cnpq.br/7219284839066009>



Universidade Estadual de Campinas
Instituto de Computação



Rebeca Padovani Ederli

Representações Visuais para Classificação de Estágios do Sono

Banca Examinadora:

- Prof. Dr. Zanoni Dias
Universidade Estadual de Campinas
- Prof. Dr. George Darmiton da Cunha Cavalcanti
Universidade Federal de Pernambuco
- Prof. Dr. Marcos Medeiros Raimundo
Universidade Estadual de Campinas

A ata da defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

Campinas, 02 de maio de 2024

Agradecimentos

Gostaria de agradecer ao Programa de Pós-Graduação do Instituto de Computação (IC) e à Universidade Estadual de Campinas (Unicamp), pela infraestrutura disponibilizada, essencial para a realização da pesquisa de mestrado.

Ao orientador Zanoni Dias, pela paciência, pelos ensinamentos e incentivo ao longo dos últimos anos, que foram fundamentais para o meu crescimento acadêmico e profissional. Ao coorientador Anderson Rocha, pelo apoio e pelos comentários que me auxiliaram durante o mestrado.

Aos amigos e colegas do projeto Viva Bem (IC – Unicamp), em especial ao grupo de Inteligência Artificial, pelo compartilhamento de conhecimento, e à Aurea Soriano-Vargas e ao Didier Vega-Oliveros, pelo auxílio na escrita do artigo.

Também agradeço aos professores que colaboraram para o meu desenvolvimento acadêmico de alguma forma, tanto os de graduação da Unesp de Presidente Prudente (FCT – Unesp) quanto os da pós-graduação do IC – Unicamp. Em especial, agradeço às professoras, pelas quais tenho profunda admiração. Expresso meus agradecimentos ao Prof. Marco Antônio Piteri, por me incentivar, desde a graduação, a prosseguir com os estudos acadêmicos no programa de pós-graduação da Unicamp.

Aos meus pais, Antonio e Maria Sueli, pelo apoio, e aos meus irmãos, Jonas, Natana, Daniel, Glória e Maria Angélica, por estarem sempre presentes, oferecendo ajuda e suporte. Aos meus amigos e à minha sobrinha, Helena, por proporcionarem momentos de descontração.

Agradeço também a todos que direta ou indiretamente contribuíram ao longo desses dois anos de mestrado.

Parte dos resultados apresentados neste trabalho foi obtida por meio do projeto “Hub de Inteligência Artificial em Saúde e Bem-estar – Viva Bem”, financiado pela Samsung Eletrônica da Amazônia Ltda., no âmbito da Lei de Informática 8.248/91.

Resumo

A presença de anomalias durante o sono, originalmente classificada por especialistas, é 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. Recentemente, a análise do sono por meio de dispositivos vestíveis, como *smartwatches*, tem sido investigada para a classificação automática dos estágios do sono, visando combinar o conforto dos *smartwatches* com a precisão da polissonografia. No entanto, os desafios enfrentados incluem a correta classificação dos estágios mais complexos e o desbalanceamento das classes nos conjuntos de dados. Utilizando dados de sensores de *smartwatches*, a literatura explora o desempenho da classificação de estágios do sono por meio de diferentes representações de dados, como dados brutos (séries temporais do sensor) e características extraídas deles (como média, desvio padrão, etc.). Nesse contexto, a representação de séries temporais por meio de imagens pode produzir características informativas e robustas ao ruído.

Assim, este estudo inova ao explorar diferentes representações visuais de séries temporais derivadas de dados de acelerômetro e frequência cardíaca, avaliando sua eficácia na classificação dos estágios do sono. Baseado nos experimentos realizados, embora *Gramian Angular Field* (GAF) tenha demonstrado os resultados mais promissores, as representações por Gráficos de Recorrência (RP) e *Markov Transition Field* (MTF) também se revelaram técnicas promissoras para a classificação.

Além disso, foi investigada a estratégia de dividir as imagens em *patches*, e empregar técnicas de *ensembles* com o objetivo de aprimorar os resultados na classificação dos estágios do sono. Os experimentos revelaram que essa abordagem contribui substancialmente para a melhoria da classificação. Adicionalmente, a pesquisa avaliou o potencial dos dados provenientes de sensores de acelerômetro e frequência cardíaca em *smartwatches*, explorando a combinação desses dois conjuntos de dados através da aplicação de *ensembles*. Os resultados indicam que, dependendo do contexto específico de classificação (seja em dois ou três estágios), os dados de acelerômetro ou de frequência cardíaca podem desempenhar um papel mais significativo.

Finalmente, a comparação com outras representações comumente utilizadas na literatura ressaltou a superioridade das abordagens visuais neste contexto. A avaliação por meio de diferentes métricas confirmou a eficácia das técnicas propostas, indicando uma abordagem competitiva para a classificação dos estágios do sono com base em dados de dispositivos vestíveis.

Abstract

The presence of anomalies during sleep, originally classified by specialists, is used for the diagnosis of sleep disorders. Polysomnography is the gold standard used in clinical settings for monitoring and classification of sleep stages, however, this method requires expensive and uncomfortable equipment. Recently, the analysis of sleep through wearable devices, such as smartwatches, has been investigated for the automatic classification of sleep stages, aiming to combine the comfort of smartwatches with the precision of polysomnography. Nevertheless, challenges faced include the correct classification of more complex stages and the imbalance of classes in datasets. Using data from smartwatch sensors, the literature explores the performance of sleep stage classification through different data representations, such as raw data (sensor time series) and extracted features from them (like mean, standard deviation, etc.). In this context, the representation of time series through images can produce informative and noise-robust features.

Thus, this study innovates by exploring different visual representations of time series derived from accelerometer and heart rate data, assessing their effectiveness in classifying sleep stages. Based on the experiments conducted, although Gramian Angular Field (GAF) showed the most promising results, the representations by Recurrence Plots (RP) and Markov Transition Field (MTF) also proved to be promising techniques for classification.

Furthermore, the strategy of dividing the images into patches, and employing ensemble techniques to enhance the results in the classification of sleep stages was investigated. The experiments revealed that this approach contributes substantially to the improvement of classification. Additionally, the research evaluated the potential of data from accelerometer and heart rate sensors in smartwatches, exploring the combination of these two data sets through the application of ensembles. The results indicate that, depending on the specific classification context (whether in two or three stages), the accelerometer or heart rate data may play a more significant role.

Finally, the comparison with other representations commonly used in the literature highlighted the superiority of visual approaches in this context. The evaluation through different assessment metrics confirmed the effectiveness of the proposed techniques, indicating a competitive approach for the classification of sleep stages based on data from wearable devices.

Lista de Figuras

2.1	Exemplo ilustrativo de diferentes comprimentos de janela utilizados para calcular as métricas de VFC da época n [31].	23
2.2	Arquitetura de rede CNN e RNN utilizada por Korkalainen <i>et al.</i> [44]. . . .	28
2.3	Estrutura do modelo <i>UltraNet</i> [37].	29
2.4	Visão geral da metodologia proposta por Lu e Tong [52] utilizando composição RGB de imagens RP.	29
2.5	Estrutura do modelo <i>ResNet</i> de fusão [64].	30
3.1	Esquema da metodologia: transformação de dados brutos de AC (cor verde) e de FC (cor azul) em representações visuais. A partir das representações, as imagens originais são utilizadas como entrada para o treinamento e validação e, então, o <i>ensemble</i> AC + FC e a validação são realizados (cor rosa). As imagens divididas em <i>patches</i> são utilizadas como entrada para o respectivo treinamento e a validação é feita a partir do <i>ensemble</i> dos resultados obtidos com todos os <i>patches</i> . O <i>ensemble</i> AC + FC, após a obtenção dos resultados do <i>ensemble</i> dos <i>patches</i> , é realizado e validado (cor rosa).	33
3.2	Classificação binária dos dados de acelerômetro de um indivíduo, cada imagem corresponde a 30 segundos.	33
3.3	Imagens de “vigília” geradas com GAF e dados de acelerômetro.	39
3.4	Estágios do sono de um indivíduo da base de dados <i>Sleep Accel</i> ao longo de 8 horas.	42
3.5	Dados da base <i>Sleep Accel</i> [81] de um indivíduo, coletados ao longo de 8 horas.	43
3.6	Quantidade de rótulos da base de dados <i>Sleep Accel</i> por intervalo de tempo ao longo de 8 horas (considerando três estágios do sono).	44
4.1	RP: matrizes de confusão geradas com os resultados obtidos por meio dos dados originais e dos melhores resultados de <i>ensembles</i> para a classificação de sono/vigília.	51
4.2	GAF: matrizes de confusão geradas com os resultados obtidos por meio dos dados originais e dos melhores resultados de <i>ensembles</i> para a classificação de sono/vigília.	52
4.3	MTF: matrizes de confusão geradas com os resultados obtidos por meio dos dados originais e dos melhores resultados de <i>ensembles</i> para a classificação de sono/vigília.	53
4.4	Espectrogramas: matrizes de confusão geradas com os resultados obtidos por meio dos dados originais e dos melhores resultados de <i>ensembles</i> para a classificação de sono/vigília.	54

4.5	Classificação de sono/vigília ao longo de uma noite de sono de um indivíduo utilizando a representação GAF.	54
4.6	RP: matrizes de confusão geradas com os resultados obtidos por meio dos dados originais e dos melhores resultados de <i>ensembles</i> para a classificação dos estágios do sono.	57
4.7	GAF: matrizes de confusão geradas com os resultados obtidos por meio dos dados originais e dos melhores resultados de <i>ensembles</i> para a classificação dos estágios do sono.	58
4.8	MTF: matrizes de confusão geradas com os resultados obtidos por meio dos dados originais e dos melhores resultados de <i>ensembles</i>	59
4.9	Espectrogramas: matrizes de confusão geradas com os resultados obtidos por meio dos dados originais e dos melhores resultados de <i>ensembles</i>	60
4.10	Classificação de estágios do sono ao longo de uma noite de sono de um indivíduo utilizando a representação GAF.	60

Lista de Tabelas

2.1	Agrupamentos dos estágios do sono para simplificar o problema observados em trabalhos relacionados.	18
2.2	Resumo dos trabalhos relacionados que classificam os estágios do sono. . .	22
2.3	Resumo dos trabalhos relacionados que envolvem outros problemas.	28
3.1	Exemplos de imagens geradas com dados de acelerômetro, com cada tipo de representação.	34
3.2	Exemplos de imagens geradas com dados de frequência cardíaca, com cada tipo de representação.	34
4.1	Acurácias balanceadas obtidas com cada representação para a classificação de sono/vigília.	49
4.2	Coeficientes κ de Cohen obtidos com cada representação para a classificação de sono/vigília.	50
4.3	Acurácias balanceadas obtidas com cada representação para a classificação dos estágios do sono.	55
4.4	Coeficientes κ de Cohen obtidos com cada representação para a classificação dos estágios do sono.	56
4.5	Acurácias balanceadas obtidas com dados brutos para a classificação de sono/vigília.	59
4.6	Acurácias balanceadas obtidas com dados brutos para a classificação dos estágios do sono.	61
4.7	Acurácias balanceadas obtidas com extração de <i>features</i> para as classificações de sono/vigília e de estágios do sono.	61
4.8	Comparação das melhores acurácias balanceadas obtidas com as diferentes representações de dados para a classificação de sono/vigília.	62
4.9	Comparação das melhores acurácias balanceadas obtidas com as diferentes representações de dados para a classificação dos estágios do sono.	62

Sumário

1	Introdução	13
1.1	Justificativa	14
1.2	Questões de Pesquisa	15
1.3	Objetivos	16
1.4	Organização do Texto	16
2	Fundamentação Teórica	17
2.1	Conceitos e Técnicas	17
2.1.1	Estágios do Sono	17
2.1.2	Polissonografia	18
2.1.3	Séries Temporais	19
2.1.4	Redes Neurais Convolucionais	19
2.1.5	Redes Neurais Recorrentes	20
2.1.6	Transformação de Séries Temporais em Imagens	20
2.1.7	Transferência de Aprendizado	21
2.2	Trabalhos Relacionados	21
2.2.1	Deteção de Sono/Vigília	22
2.2.2	Classificação em Três Estágios	24
2.2.3	Classificação em Quatro Estágios	25
2.2.4	Classificação em Cinco Estágios	27
2.2.5	Outros Problemas	27
2.2.6	Limitações dos Trabalhos Relacionados	30
3	Materiais e Métodos	32
3.1	Metodologia	32
3.1.1	Preparação dos Dados	33
3.1.2	Representações	34
3.1.3	<i>Patches</i>	38
3.1.4	<i>Ensembles</i>	38
3.2	Métricas de Avaliação	41
3.3	Conjunto de Dados	42
3.3.1	Artigo de Referência da Base de Dados	43
3.4	Ferramentas e Recursos Computacionais	44
4	Experimentos e Resultados	46
4.1	Experimentos	46
4.1.1	Treinamento e Validação	46
4.1.2	Comparação com Outras Representações	47
4.1.3	Descrição dos Experimentos	47

4.2	Resultados e Discussões	48
4.2.1	Classificação de Sono/Vigília	48
4.2.2	Classificação dos Estágios do Sono	52
4.2.3	Outras Representações	58
5	Conclusões e Trabalhos Futuros	63
	Referências Bibliográficas	67

Capítulo 1

Introdução

O sono é altamente preservado ao longo do processo evolutivo, indicando funções biológicas cruciais que permanecem em grande parte desconhecidas. Tanto animais quanto seres humanos submetidos a limitações parciais de sono frequentemente manifestam respostas fisiológicas adversas [58].

As mudanças do estilo de vida da população mundial nas últimas décadas têm levado a um aumento significativo na prevalência de distúrbios do sono [12], tendo em vista que fatores como a falta de atividade física, estresse diário e o excesso de mídia eletrônica estão relacionados ao aumento da proporção de pessoas que dormem menos do que o recomendado [30]. Estes distúrbios interferem negativamente em órgãos e sistemas do corpo humano, causando repercussões durante as vigílias, logo na qualidade de vida, e promovendo a origem de diversas patologias.

Uma variedade de problemas relacionados ao sono impacta milhões de indivíduos globalmente, abrangendo desde casos de insônia crônica e apneia do sono até narcolepsia e síndrome das pernas inquietas [9]. O primeiro posicionamento brasileiro sobre o impacto dos distúrbios do sono nas doenças cardiovasculares da Sociedade Brasileira de Cardiologia [19] destaca que esses distúrbios aumentam o risco de doenças cardiovasculares e, além disso, a baixa qualidade do sono tem recebido atenção em estudos geriátricos por estar relacionada ao declínio cognitivo em idosos [50]. Nesse sentido, a pesquisa médica busca automatizar procedimentos para diagnósticos mais rápidos e antecipados de distúrbios do sono [84].

A avaliação do comportamento do sono, incluindo parâmetros como frequência cardíaca, frequência respiratória, temperatura e movimento corporal, pode ser realizada para determinar aspectos como tempo, duração, latência, despertar após o início do sono e eficiência do sono [34, 67]. Embora a polissonografia (PSG) seja considerada o padrão-ouro em ambientes clínicos para extrair dados essenciais e determinar os estágios do sono, ela apresenta desafios, como a necessidade de equipamentos caros e desconfortáveis, compreendendo pelo menos onze canais de sinais fisiológicos coletados de diferentes sensores, além de monitoramento de curto prazo e possíveis efeitos no paciente após a primeira noite sob registros laboratoriais [56, 78].

A actigrafia utiliza um dispositivo semelhante a um relógio, equipado com sensores, como acelerômetro, e emprega algoritmos para mapear medições e identificar padrões de sono/vigília sem identificar com precisão os diferentes estágios do sono [6]. Essa aborda-

gem é recomendada pela Academia Americana de Medicina do Sono (AASM) e tem sido comparada à PSG em diversos estudos desde a década de 1990 [16,61,63,83], destacando-se pela confiabilidade e precisão em relação a questionários subjetivos e diários de sono.

Dispositivos vestíveis, como *smartwatches*, equipados com sensores, possibilitam o monitoramento do sono em casa, coletando dados sobre sua qualidade e duração [23]. Essas informações, baseadas nos movimentos ou sinais fisiológicos durante o sono, possibilitam o ajuste de hábitos e melhorarias na saúde do usuário [17,72]. As tecnologias vestíveis, ao automatizar a detecção e classificação de padrões de sono, podem oferecer diagnósticos rápidos de distúrbios, utilizando protocolos confortáveis, como actigrafia, e obtendo resultados precisos, semelhantes à PSG.

Diversas técnicas de Inteligência Artificial têm sido aplicadas para problemas relacionados ao sono (classificação dos estágios, reconhecimento da posição e investigação de distúrbios), desde métodos como *Random Forest*, *Support Vector Machine* (SVM) e *K-Nearest Neighbors* (KNN), utilizando extração de *features*, até a adaptação de Redes Neurais Recorrentes (RNN), como *Long Short-Term Memory* (LSTM) e *Gated Recurrent Unit* (GRU), e Redes Neurais Convolucionais (CNN) [23].

A conversão de dados brutos de sensores em imagens para classificação possibilita a utilização de CNNs 2D na extração de padrões complexos destas imagens, introduzindo uma diversidade de características inacessíveis a sinais 1D [38].

Neste estudo, levanta-se a hipótese de que a representação de dados de sono através de imagens, adquiridos por sensores de *smartwatches*, pode se revelar uma abordagem eficaz. Acredita-se que tal metodologia possui o potencial de aprimorar a precisão na avaliação de dados temporais originados de *smartwatches*, possibilitando uma classificação mais eficientes dos estágios do sono.

1.1 Justificativa

Os estágios do sono são diferentes estados fisiológicos que o corpo alterna durante o sono e contêm informações cruciais, uma vez que o entendimento desses padrões, juntamente com outras variáveis, constitui a base para diagnosticar vários distúrbios relacionados ao sono [71]. Inicialmente avaliados por Rechtschaffen e Kales (R&K) [66], esses estágios compreendiam seis categorias, incluindo vigília, sono de movimento rápido dos olhos (REM) e quatro níveis do sono não REM (NREM): S1, S2, S3 e S4. A AASM [35] posteriormente definiu cinco estágios: vigília, REM e NREM divididos em N1, N2 e N3. Os estágios R&K são equivalentes à definição da AASM, que agrupa S3 e S4 como N3. Uma análise ideal deve ser capaz de distinguir entre cinco ou seis estágios, dependendo do padrão escolhido. Muitos dos estudos realizados para a tomada de decisão automática do sono foram propostos apenas com três classes incluindo vigília, sono REM e sono não REM [27,81].

Os dados provenientes de sensores de *smartwatches* 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 fotopletismografia. Alguns estudos utilizam a extração de novas *features* e

obtem 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. A validação é feita em relação à classificação manual realizada por especialistas, que inclui variabilidade conhecida pelo fato de que a classificação de estágios específicos pode ser feita de forma diferente por diferentes profissionais [62].

Classificar os estágios do sono com dados de *smartwatches* e técnicas de Inteligência Artificial requer a construção de modelos capazes de lidar com a complexidade dos padrões presentes nesses dados. A identificação dos estágios do sono representa a primeira etapa para uma classificação mais precisa dos distúrbios do sono [73]. Tendo em vista que, embora existam outras causas possíveis, as modificações na arquitetura comum do sono potencialmente indicam a presença de distúrbios do sono. Por exemplo, anomalias no sono REM podem estar relacionadas à narcolepsia [57].

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 [2, 27].

Os trabalhos relacionados exploraram dados de diferentes sensores de dispositivos vestíveis para classificar os estágios do sono, como *smartwatches* [81], anéis [3] e outros [26]. Utilizando dados de *smartwatches* para diferentes problemas, a literatura está focada em utilizar os dados brutos do sensor, ou seja, séries temporais, em um modelo de *Deep Learning* (DL) [29]. No entanto, essa abordagem apresenta desvantagens relacionadas ao ruído nas séries temporais e à qualidade dos dados brutos.

Uma abordagem promissora investigada em estudos de problemas relacionados, como o Reconhecimento de Atividades Humanas (HAR) [52] e a identificação da doença de Parkinson [1], consiste em representar dados de sensores por meio de imagens, como os Gráficos de Recorrência (RP). Essa técnica, ainda não investigada na classificação de estágios do sono, demonstra robustez em relação ao ruído presente nos dados brutos e pode proporcionar a geração de características mais informativas capturados por modelos de *Deep Learning*.

Dessa forma, surge o interesse em explorar a utilização da representação de dados do sono em formato de imagens, especificamente provenientes de sensores integrados a *smartwatches*, e empregar algoritmos de *Deep Learning* na tarefa de classificação dos estágios do sono.

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 classificação dos estágios do sono?
- Quais são as representações de imagens mais adequadas para classificar estágios do sono?

- Quais técnicas de aprendizado de máquina podem tornar a classificação dos estágios do sono mais precisa?

1.3 Objetivos

O principal objetivo deste trabalho é investigar técnicas de Aprendizado Profundo (*Deep Learning*) para melhorar a classificação dos estágios do sono com dados provenientes de sensores de *smartwatches*, utilizando representações na forma de imagens. Para alcançar esse propósito, alguns objetivos específicos devem ser atingidos:

- Analisar as características dos dados provenientes de sensores de *smartwatches*;
- Selecionar parâmetros apropriados para representar os dados por meio de imagens;
- Comparar as diferentes representações geradas;
- Explorar a combinação de informações provenientes dos diversos sensores presentes nos *smartwatches*;
- Aplicar técnicas de aprendizado de máquina que permitam aprimorar os resultados originais;
- Desenvolver um classificador de sono/vigília;
- Desenvolver um classificador de estágios do sono;
- Comparar os resultados obtidos com os encontrados na literatura.

Esta dissertação visa contribuir com o desenvolvimento de uma metodologia aplicando Inteligência Artificial 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

Esta dissertação está organizada em cinco capítulos. No Capítulo 1, é descrito o problema de pesquisa a ser abordado neste mestrado e também os principais objetivos e contribuições esperadas. No Capítulo 2, são introduzidos conceitos importantes para o entendimento da dissertação e são apresentados alguns trabalhos relacionados relevantes existentes na literatura. No Capítulo 3, são apresentados os materiais e os métodos utilizados. No Capítulo 4, os experimentos e resultados são detalhados. Por fim, as conclusões e trabalhos futuros são abordados no Capítulo 5.

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, polissonografia 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, à transformação de séries temporais em imagens e à transferência de aprendizado. Esses conceitos e técnicas são apresentados nesta seção.

2.1.1 Estágios do Sono

A classificação dos estágios do sono ocorre em épocas de 30 segundos com base nas atuais diretrizes da AASM, que é o padrão mais recente, e os estágios do sono são definidos como [71]:

- **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 aparecem no EEG. É um momento típico para ocorrerem parassonias do NREM, como sonambulismo.

- **REM:** paralisia temporária dos músculos, exceto dos olhos e dos músculos que controlam a respiração. É conhecido como sono dos sonhos e a atividade cerebral aumenta, aproximando-se dos níveis observados durante a vigília.

Os ciclos do sono consistem em padrões distintos de NREM e REM, com cada um durando cerca de 90 a 120 minutos. O primeiro ciclo transita pelos estágios N1, N2 e N3 até o primeiro período REM. Em uma noite padrão de oito horas de sono, costumam ocorrer de quatro a cinco desses ciclos. Ao longo do sono, há um aumento frequente na porcentagem de sono REM em cada ciclo, enquanto a proporção do estágio N3 diminui, atingindo a maior quantidade na primeira metade da noite.

Em trabalhos relacionados, os autores podem agrupar os estágios do sono em apenas quatro, três ou até dois estágios, conforme é descrito na Tabela 2.1, para simplificar a tarefa de classificação.

5 estágios do sono (AASM)	4 estágios do sono	3 estágios do sono	2 estágios do sono
Vigília	Vigília	Vigília	Vigília
N1	Sono leve	NREM	Sono
N2			
N3	Sono profundo		
REM	REM	REM	

Tabela 2.1: Agrupamentos dos estágios do sono para simplificar o problema observados em trabalhos relacionados.

2.1.2 Polissonografia

A polissonografia (PSG) é um exame realizado para diagnosticar distúrbios do sono. Durante o exame, o paciente passa uma noite em um laboratório do sono, no qual diversos sensores são utilizados para monitorar parâmetros fisiológicos. Esses sensores incluem eletroencefalogramas (EEG) para registrar a atividade cerebral, eletrocardiogramas (ECG) para monitorar o ritmo cardíaco, eletrooculogramas (EOG) para monitorar movimentos oculares e sensores respiratórios para avaliar os padrões de respiração e níveis de oxigênio no sangue [46]. Esses sinais são divididos em épocas de 30 segundos, cada uma das quais é classificada por um especialista em diferentes estágios de sono (vigília, N1, N2, N3 ou REM). A PSG é considerada o padrão-ouro para análise do sono.

Como é necessária a utilização de equipamentos especializados e a presença de profissionais qualificados para monitorar e interpretar os dados coletados, a PSG apresenta um custo elevado. Além disso, a infraestrutura do laboratório do sono, que simula um ambiente de quarto inclui tecnologia avançada de monitoramento, contribuindo para os custos operacionais. A PSG é um exame desconfortável, pois diferentes sensores são aplicados ao corpo do paciente, incluindo o couro cabeludo, rosto, peito e pernas, que podem restringir movimentos e dificultar o sono natural. O ambiente desconhecido do laboratório e a presença de câmeras e sistemas de áudio para monitoramento contínuo também podem tornar a experiência menos confortável para muitos pacientes [55].

Como alternativa à PSG, os actígrafos são dispositivos portáteis utilizados para monitorar os ciclos de sono e vigília através da detecção de movimentos. Esses dispositivos

são menos invasivos e podem ser utilizados em casa, oferecendo uma solução mais conveniente para o monitoramento do sono em alguns casos como, por exemplo, pacientes que precisam ser monitorados por diversos dias. No entanto, os actígrafos fornecem dados menos detalhados em comparação com a PSG, sendo menos adequados para diagnósticos complexos [51].

2.1.3 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 definido 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 [22].

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 ou *Unix Timestamp*). 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” [78]. 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. Os dados coletados por sensores de *smartwatches*, como o acelerômetro, são séries temporais devido à natureza cronológica das observações registradas. O acelerômetro, ao medir a aceleração do dispositivo em diferentes direções ao longo do tempo, produz sequências ordenadas de dados sobre o movimento do usuário. 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.

2.1.4 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 consiste em aplicar um filtro que percorre a entrada, realizando multiplicações e produzindo um mapa de características que realça padrões importantes enquanto preserva a relação espacial dos dados.

A arquitetura CNN unidimensional (CNN 1D) é composta por múltiplas camadas, sendo que a camada de entrada pode receber sinais de séries temporais [74]. As CNNs

1D são comumente aplicadas a séries temporais para capturar características espaciais ao longo da dimensão temporal dos sinais de sensores [42].

As CNNs 2D aplicam operações para identificar padrões locais, aprendendo características como bordas, texturas e formas. Camadas de *pooling* são frequentemente empregadas para reduzir a dimensionalidade dos dados, preservando informações cruciais. Essas redes são notáveis por sua habilidade em realizar tarefas de visão computacional. A convolução 2D e a estrutura hierárquica das CNNs 2D possibilitam a captura eficiente de padrões visuais em dados bidimensionais, tornando-as essenciais para aplicações que demandam análise e compreensão de imagens [79].

2.1.5 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 utilizando 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 [33] e a GRU [14].

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.

2.1.6 Transformação de Séries Temporais em Imagens

Dados provenientes de sensores, organizados como séries temporais, têm sido amplamente explorados em diversas aplicações. A abordagem tradicional envolve a extração de *features* dessas séries temporais, muitas vezes utilizando algoritmos clássicos de aprendizado de máquina, como SVM, KNN e *Random Forest*. Com os avanços em *Deep Learning*, duas abordagens aplicadas diretamente com os dados brutos têm ganhado destaque. A primeira envolve a aplicação de RNNs, como LSTM ou GRU, que são projetadas para capturar padrões temporais complexos e dependências de longo prazo nos dados; e a segunda abordagem utiliza CNNs 1D, as quais são eficazes na aprendizagem de padrões locais em dados sequenciais [42].

Utilizando séries temporais transformadas em imagens, a classificação de dados de sensores vestíveis pode ser realizada utilizando técnicas de visão computacional para reconhe-

cimento de imagens. Isso envolve converter dados temporais em imagens bidimensionais, permitindo a aplicação direta de modelos de aprendizado profundo, como CNNs 2D. Essa abordagem captura informações temporais e padrões complexos em diferentes dimensões, facilitando a detecção de correlações e características ocultas. Assim, a representação dos dados é simplificada e a capacidade das CNNs 2D em extrair padrões visuais complexos é aproveitada [1]. Os Gráficos de Recorrência (RP) [20], *Gramian Angular Field* (GAF) [82], *Markov Transition Field* (MTF) [82] e espectrogramas são estratégias encontradas na literatura que codificam representações unidimensionais de dados coletados por sensores vestíveis em imagens [60, 64].

2.1.7 Transferência de Aprendizado

A transferência de aprendizado é uma técnica de Inteligência Artificial, notadamente empregada na utilização de modelos pré-treinados em conjuntos de dados abrangentes [65], exemplificado pelo conjunto de dados *ImageNet* [69], para aprimorar o desempenho em tarefas específicas.

A *ResNet* introduziu blocos residuais, proporcionando uma abordagem eficiente para o treinamento de redes de maior profundidade [32]. Por sua vez, a *EfficientNet* otimiza a arquitetura considerando simultaneamente a largura, profundidade e resolução da rede [76]. A *ImageNet*, como uma base de dados volumosa e categorizada, configura-se como um recurso significativo para treinar essas redes.

No processo de transferência de aprendizado, é prática comum congelar as camadas pré-treinadas da rede, impedindo a adaptação destas durante o treinamento na tarefa específica. Tal estratégia é especialmente pertinente quando o conjunto de dados alvo é limitado, preservando as representações aprendidas previamente. As camadas não congeladas, geralmente as finais, são ajustadas para se adequarem à nova tarefa, aproveitando os padrões gerais internalizados nas camadas congeladas [36]. Essa abordagem mitigadora acelera o treinamento e amplifica o desempenho nas tarefas específicas, utilizando o conhecimento prévio adquirido pela rede em grandes conjuntos de dados, como o *ImageNet*.

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 (vigília, NREM e REM) ou quatro (vigília, sono leve, sono profundo e REM) estágios. 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.

A Tabela 2.2 apresenta um resumo dos trabalhos de classificação de estágios do sono,

destacando os autores, estágios, tipo de representação dos dados e principais características.

Trabalho	Estágios	Representação	Principais Características
Paquet <i>et al.</i> [61] (2007)	2	<i>Features</i>	- Avaliação da actigrafia para detectar vigília/sono; - Indivíduos consumidores moderados de cafeína.
Sano e Picard [70] (2014)	2	<i>Features</i>	- Temperatura da pele (ST) desempenhou um papel importante na classificação de vigília/sono.
Haghighyegh <i>et al.</i> [31] (2020)	2	<i>Features</i>	- Comparação de diferentes comprimentos de janela para o cálculo da Variabilidade de Frequência Cardíaca (VFC); - <i>DeepConvLSTM</i> .
Walch <i>et al.</i> [81] (2019)	3	<i>Features</i>	- Base de dados <i>Sleep Accel</i> ; - <i>Feature proxy</i> de relógio, que representa o relógio circadiano; - <i>MESA Sleep</i> para verificar generalização.
Fonseca <i>et al.</i> [26] (2017)	3	<i>Features</i>	- O treinamento foi feito com dados da base SIESTA [43].
Beattie <i>et al.</i> [7] (2017)	4	<i>Features</i>	- Pós-processamento que penaliza padrões fisiológicos improváveis.
Altini e Kinnunen [3] (2021)	4	<i>Features</i>	- Funções cosseno [81], exponencial e linear para modelar características circadianas.
Radha <i>et al.</i> [65] (2021)	4	<i>Features</i>	- Transferência de aprendizado; - Treinamento com dados SIESTA [43].
Korkalainen <i>et al.</i> [44] (2020)	5	Dados brutos	- Não realiza a extração e seleção prévia de características; - Arquitetura CNN e GRU.

Tabela 2.2: Resumo dos trabalhos relacionados que classificam os estágios do sono.

2.2.1 Detecção de Sono/Vigília

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 sono/vigília foi abordada em diversos trabalhos [16,61,63,83].

O estudo de Paquet *et al.* [61] avalia a actigrafia para detectar sono/vigília 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 classificada 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.

Sano e Picard [70] 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 sono/vigília. 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 [16];

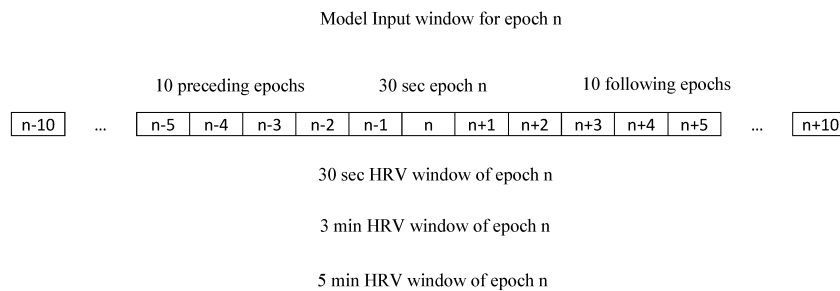


Figura 2.1: Exemplo ilustrativo de diferentes comprimentos de janela utilizados para calcular as métricas de VFC da época n [31].

- 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 dos sensores da 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.

Haghighat *et al.* [31] utilizaram os dados de PSG e actigrafia de punho do estudo MESA Sleep [13,86]. 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 [59], 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 [26]. 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 sono/vigília além da classificação de mais estágios do sono [3, 26, 81].

2.2.2 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 [81]. Os modelos treinados na base de dados criada, denominada *Sleep Accel*, foram também testados em um subconjunto dos dados disponibilizados publicamente, *MESA Sleep* [8], para avaliar a generalização dos algoritmos. Os autores utilizaram um contador de atividades calculado após aplicar um filtro passa-banda Butterworth de 5^a ordem entre 3 e 11Hz [77], 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 sono/vigília 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 sono/vigília e os três estágios, respectivamente. Ao utilizar o subconjunto dos dados *MESA Sleep*, obteve-se $\kappa = 0,52$ para a detecção de sono/vigília e $\kappa = 0,40$ para a classificação dos estágios.

Fonseca *et al.* [26] 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 [43] 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 [25], a acurácia atingiu 91,5% ($\kappa = 0,55$) para a detecção de sono/vigília, 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 estágios [44], como poderá ser observado nas seções seguintes.

2.2.3 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.* [7] 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 [77]. 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.* [7], 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 [41]. Neste estudo, são utilizadas as seguintes informações circadianas: uma função cosseno, também aplicada por Walch *et al.* [81], 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.

Altini e Kinnunen [3] utilizam um classificador *Light Gradient Boosting Machine* (LightGBM) [40], com *Dropouts meet Multiple Additive Regression Trees* (DART) [80]. 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.* [65], 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.* [26]. 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 [25]. 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.

2.2.4 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.* [44] 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 dedo (*Nonin Xpod 3011*). O estudo foi realizado com 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 2.2.

Os resultados gerais obtidos foram acurácias de 80,1% ($\kappa = 0,65$), 68,5% ($\kappa = 0,54$) e 64,1% ($\kappa = 0,51$) para os modelos de três, 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 [23].

2.2.5 Outros Problemas

Outros problemas envolvendo dados de sensores de *smartwatches*, dispositivos móveis ou dados de acelerômetros de pulso são abordados na literatura. Dentre eles, é importante destacar os trabalhos que desenvolveram redes neurais para HAR. Algumas abordagens recentes incluem a utilização direta dos dados brutos ou a transformação de séries temporais em imagens. A Tabela 2.3 mostra de forma sumarizada esses trabalhos, dando ênfase nos autores, no tipo de problema, na representação de dados utilizada e nas principais características.

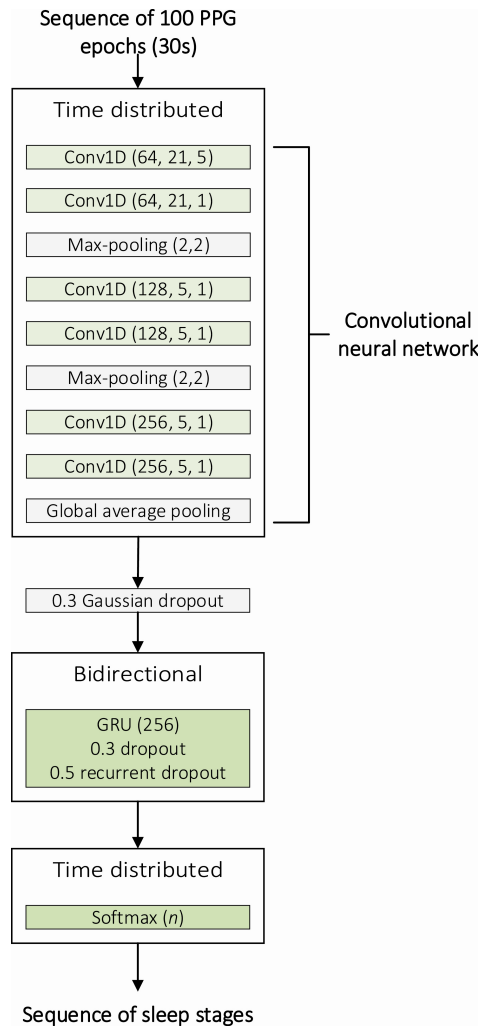


Figura 2.2: Arquitetura de rede CNN e RNN utilizada por Korkalainen *et al.* [44].

Trabalho	Problema	Representação	Principais Características
Sharen <i>et al.</i> [74] (2022)	HAR	Dados brutos	- Conjunto WISDM [47]; - CNN 1D.
Imran [37] (2022)	HAR	Dados brutos	- Conjunto WISDM [47]; - <i>UltraNet</i> [37] com BiGRU [11].
Lu e Tong [52] (2019)	HAR	Imagens (RP)	- <i>ResNet</i> com imagens RGB; - Eixos x , y e z como canais RGB.
Qin <i>et al.</i> [64] (2020)	HAR	Imagens (GAF)	- Conjuntos de dados: HHAR [75] e MHEALTH [5]; - <i>ResNet</i> de fusão.
Ortega <i>et al.</i> [60] (2021)	Reconhecimento de ingestão de alimentos e bebidas	Imagens (GAF, MTF, espectrogramas)	- Redes <i>multi-input</i> e CNNs.

Tabela 2.3: Resumo dos trabalhos relacionados que envolvem outros problemas.

Dados Brutos

A rede *WISNet* [74] 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 [47], atingindo uma acurácia de 96,4%. Outra proposta para HAR utilizando o conjunto de dados WISDM é descrita por Imran [37], que desenvolveu a rede *UltraNet* [37] com uma unidade recorrente fechada bidirecional (BiGRU) [11] com 64 unidades, como pode ser observada na

Figura 2.3. O modelo atingiu acurácia de 97,7% sob a base de dados utilizada.

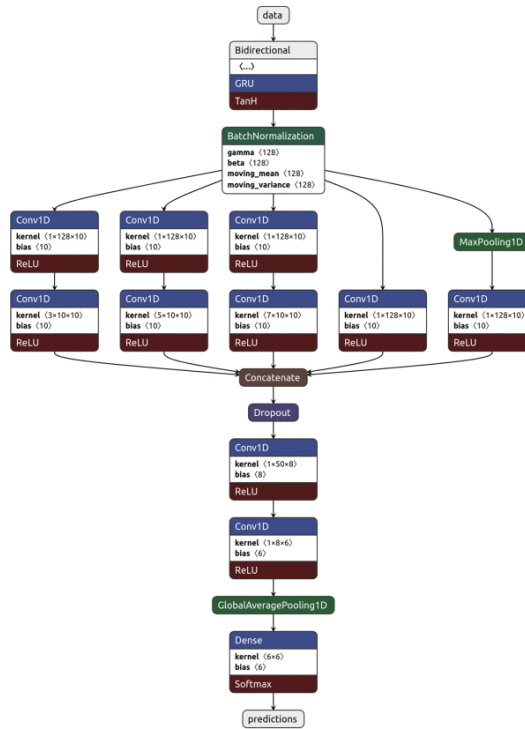


Figura 2.3: Estrutura do modelo *UtaNet* [37].

Transformação de Séries Temporais em Imagens

Considerando a representação por RP para HAR, Lu e Tong [52] codificaram os sinais dos três eixos do acelerômetro como imagens de três canais (RGB) usando RP modificado. Como pode ser observado na representação da metodologia na Figura 2.4, os autores treinaram uma *ResNet* para classificar atividades humanas através das imagens. O algoritmo foi avaliado em um novo banco de dados e um conjunto de dados público. Cada canal das imagens RGB corresponde a um eixo do acelerômetro (x , y e z). Os autores atingiram uma precisão média de 93.9/91.3% com o novo conjunto de dados e 90.9/88.7% com os dados públicos, sob configurações mistas/independentes do usuário. Outros trabalhos que abordam o uso de RP apresentam diferentes abordagens para compor uma imagem RGB [24].

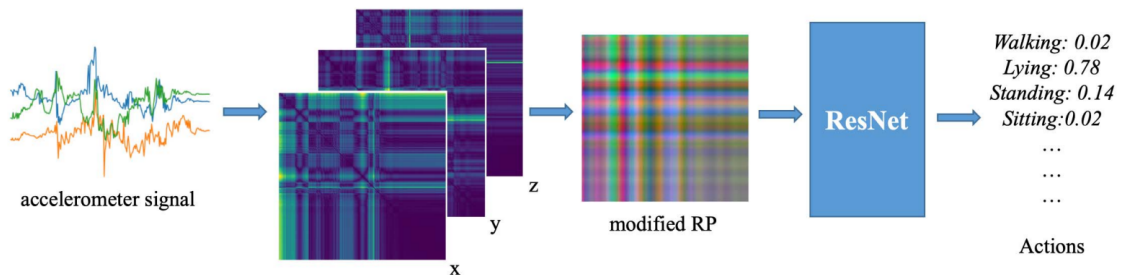


Figura 2.4: Visão geral da metodologia proposta por Lu e Tong [52] utilizando composição RGB de imagens RP.

Qin *et al.* [64] transformam dados de sensores de dispositivos móveis em imagens para realizar a classificação de seis atividade humanas e avaliam o método em dois conjuntos de dados: HHAR [75] e MHEALTH [5]. Os autores utilizam GAF para transformar os dados (aceleração, velocidade angular e orientação do campo magnético) em imagens e desenvolveram uma *ResNet* de fusão para resolver o problema de dados heterogêneos, como pode ser observada na Figura 2.5. GAF representa os dados de séries temporais em um sistema de coordenadas polares em vez de coordenadas cartesianas, na matriz de Gramian, cada elemento é o cosseno da soma ou diferença dos valores temporais em pares. Os autores alcançaram até 99,2% de acurácia ao utilizar todos os dados.

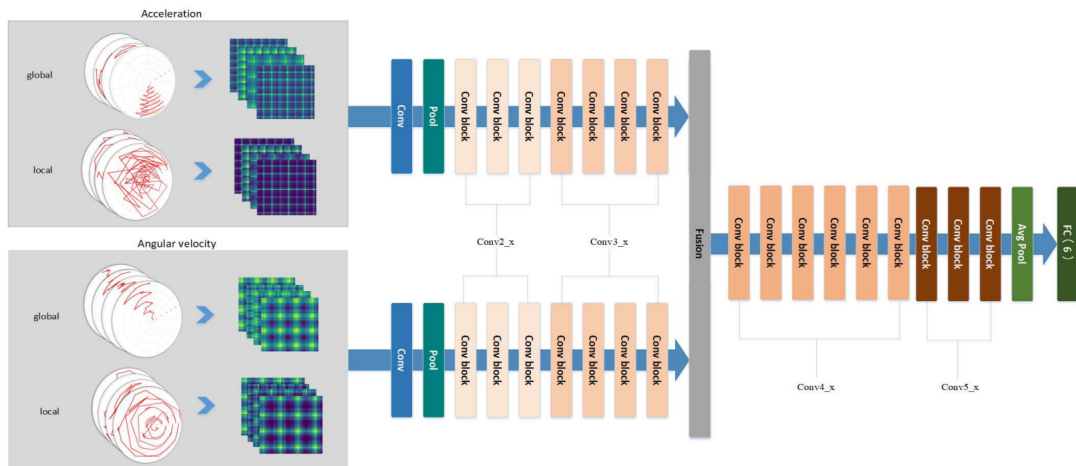


Figura 2.5: Estrutura do modelo *ResNet* de fusão [64].

Ortega *et al.* [60] desenvolveram um estudo sobre o reconhecimento de ingestão de alimentos e bebidas, tendo três classes: “comer”, “beber” e “nulo”, sendo a classe “nulo” todos os gestos irrelevantes. Para isso, os autores transformaram os dados de um acelerômetro triaxial de pulso em imagens utilizando GAF, MTF e espectrogramas; e compararam diversas rede *multi-input multi-domain*, incluindo um classificador baseado em CNNs para extrair *features* de cada representação de imagem. As saídas das CNNs 2D são unidas em um vetor que alimenta uma CNN 1D, com isso, os autores atingem uma acurácia de 97,1%.

É importante ressaltar que embora essas abordagens não busquem resolver o problema a ser explorado neste trabalho, os métodos desenvolvidos pelos autores são estudos recentes para lidar com dados de sensores dos *smartwatches* ou acelerômetros.

2.2.6 Limitações dos Trabalhos Relacionados

Observa-se uma carência de estudos atuais quanto à classificação dos estágios do sono utilizando dados oriundos de *smartwatches*. Além disso, embora exista uma prática consolidada de converter dados brutos de *smartwatches* em imagens para outros problemas de classificação, como o HAR, essa abordagem não foi previamente aplicada ao desafio de classificar os estágios do sono.

Diante disso, identifica-se uma lacuna de pesquisa no uso de séries temporais transformadas em imagens e técnicas de *Deep Learning* para a classificação dos estágios do

sono. A exploração dessa abordagem pode contribuir para um avanço no entendimento dos padrões de sono, representando um campo promissor para futuras investigações.

A transformação de dados de sono em imagens oferece uma estratégia diferente para a análise que pode capturar características complexas e sutis dos padrões de sono. Métodos como as transformações RP e GAF já mostraram eficácia em outros tipos de problemas, como para HAR [52, 64]. No entanto, a aplicação dessas técnicas no contexto do sono ainda é escassa, sugerindo um potencial significativo para a classificação dos estágios do sono.

Além disso, a utilização de CNNs 2D para a análise de imagens derivadas de dados de sono pode proporcionar o conhecimento de padrões relevantes. Pesquisas anteriores demonstraram que CNNs são eficazes na extração de características complexas de imagens [45, 49]. Integrar essas abordagens com dados de *smartwatches* pode levar a melhorias na precisão da classificação dos estágios do sono, preenchendo uma lacuna na literatura atual.

Capítulo 3

Materiais e Métodos

Neste capítulo, são descritas a metodologia, as métricas e a base de dados disponível publicamente utilizadas para a construção da dissertação.

3.1 Metodologia

A metodologia para classificar os estágios do sono consiste, principalmente, em explorar a representação de dados em imagens e utilizar métodos que consideram as informações espaciais e temporais dos dados, além de combinar os dados de acelerômetro (AC) e frequência cardíaca (FC) de *smartwatches*. A metodologia também inclui a comparação entre esse tipo de representação com a extração de *features* e com dados brutos, que são as representações de dados mais encontradas em trabalhos relacionados. A classificação dos estágios do sono é simplificada em sono/vigília (classificação binária) e vigília/NREM/REM (classificação de três estágios do sono).

Conforme ilustrado na Figura 3.1, a transformação de dados brutos em imagens foi aplicada utilizando técnicas encontradas em alguns trabalhos relacionados: RP, GAF, MTF e espectrogramas. As imagens foram geradas a partir dos dados de acelerômetro e dos dados de frequência cardíaca separadamente e, também, foram realizadas técnicas de *ensemble* para combinar as classificações obtidas com cada tipo de dados.

A utilização de imagens para realizar tarefas de classificação permite a aplicação de uma técnica que divide a imagem original em subimagens, ou *patches*. Como pode ser observado no esquema da Figura 3.1, as imagens das diferentes representações (RP, GAF, MTF e espectrogramas), além de serem processadas no formato que foram originalmente geradas, também são divididas em *patches*. Os *ensembles* são aplicados para combinar as previsões dos modelos treinados com as seguintes entradas: originais dos dados de AC e FC; *patches* dos dados de AC; *patches* dos dados de FC; e *patches* dos dados de AC e *patches* dos dados de FC.

Os detalhes da metodologia sobre a preparação dos dados, representações, *patches* e *ensembles* são apresentados nas próximas subseções. É importante destacar que as decisões tomadas para construir a metodologia foram baseadas em experimentos e/ou abordagens relatadas na literatura.

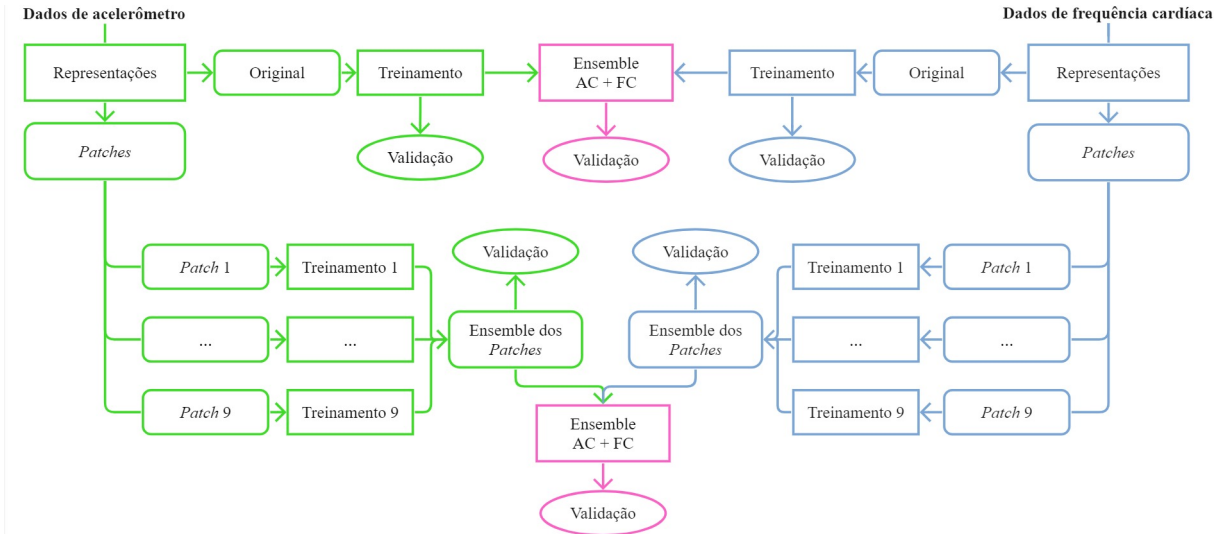


Figura 3.1: Esquema da metodologia: transformação de dados brutos de AC (cor verde) e de FC (cor azul) em representações visuais. A partir das representações, as imagens originais são utilizadas como entrada para o treinamento e validação e, então, o *ensemble* AC + FC e a validação são realizados (cor rosa). As imagens divididas em *patches* são utilizadas como entrada para o respectivo treinamento e a validação é feita a partir do *ensemble* dos resultados obtidos com todos os *patches*. O *ensemble* AC + FC, após a obtenção dos resultados do *ensemble* dos *patches*, é realizado e validado (cor rosa).

3.1.1 Preparação dos Dados

Antes de gerar as imagens, os dados brutos de movimento e de frequência cardíaca foram interpolados e cada imagem foi gerada considerando 600 medições [81], possuindo assim, dimensão 600×600 (exceto os espectrogramas, que possuem dimensões variadas). Para serem processadas como entrada em uma rede, as imagens são redimensionadas para 224×224 .

Cada imagem corresponde a uma janela de 30 segundos, conforme o padrão das gravações de PSG empregado pela literatura. Assim, cada imagem deve ser classificada como sono ou vigília, no problema de classificação binária, e como vigília, NREM ou REM, na classificação de três estágios. O cenário da classificação binária é ilustrado na Figura 3.2.

As diferentes representações de imagens foram obtidas para cada eixo do acelerômetro (x , y e z) e uma imagem RGB da composição de x , y e z é gerada para sustentar as informações dos três eixos [52]. As imagens resultantes dos dados de frequência cardíaca, por outro lado, são geradas em escala de cinza.

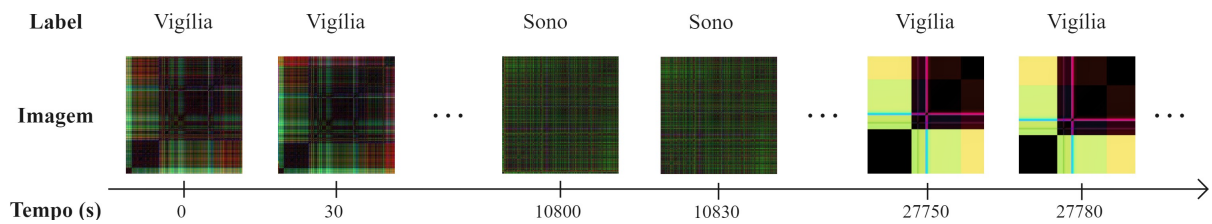


Figura 3.2: Classificação binária dos dados de acelerômetro de um indivíduo, cada imagem corresponde a 30 segundos.

3.1.2 Representações

As transformações de séries temporais para imagens foram realizadas utilizando quatro tipos de representações: RP, GAF, MTF e espectrogramas. Os detalhes sobre cada representação são apresentados nesta seção. Alguns exemplos das imagens obtidas com dados do acelerômetro e de frequência cardíaca podem ser visualizados nas tabelas 3.1 e 3.2, respectivamente.

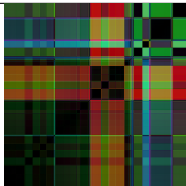
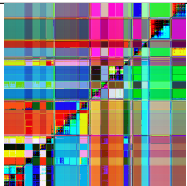
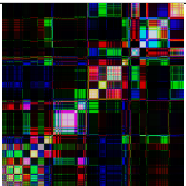
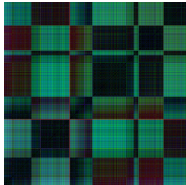
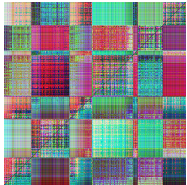
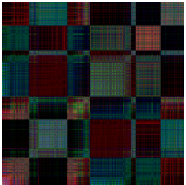
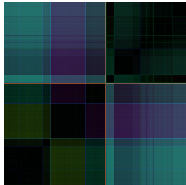
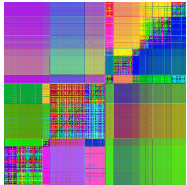
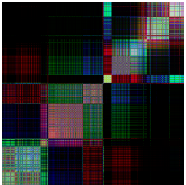
	RP	GAF	MTF	Espectrograma
Vigília				
Sono (NREM)				
Sono (REM)				

Tabela 3.1: Exemplos de imagens geradas com dados de acelerômetro, com cada tipo de representação.

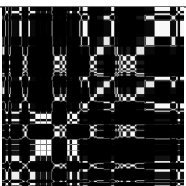
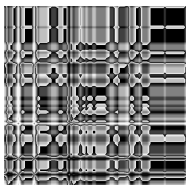
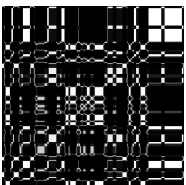
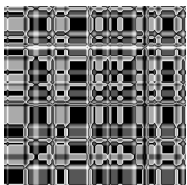
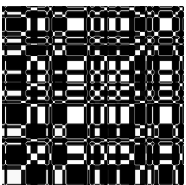
	RP	GAF	MTF	Espectrograma
Vigília				
Sono (NREM)				
Sono (REM)				

Tabela 3.2: Exemplos de imagens geradas com dados de frequência cardíaca, com cada tipo de representação.

Gráficos de Recorrência (RP)

As representações de RP foram propostas por Eckmann *et al.* [20] para análise não-linear de dados de séries temporais e possibilitam, por meio da visualização, adquirir informações sobre comportamento recorrente em séries temporais. O RP é uma matriz de pontos de tamanho $N \times N$, na qual N é o número de estados, e uma recorrência acontece quando uma trajetória visita a mesma vizinhança no espaço de fase em que estava em algum momento anterior.

A matriz de recorrência R pode ser descrita pela Equação 3.1 (função de Heaviside), sendo ε o limiar de recorrência; $d(x[i], x[j])$ representa a distância entre os valores correspondentes aos instantes i e j na série temporal x . Os valores da série temporal são normalizados antes da transformação para RP.

$$R[i, j] = \begin{cases} 1 & \text{se } d(x[i], x[j]) \leq \varepsilon \\ 0 & \text{caso contrário} \end{cases} \quad (3.1)$$

O limiar de recorrência (ε) é um parâmetro crucial nesta representação. Para determinar o valor ótimo de ε , foram exploradas diversas alternativas [21], incluindo:

- 10% da distância máxima;
- média das distâncias; e
- percentil 10.

Além do método tradicional de construção de RP, que utiliza a função de Heaviside para binarizar a matriz resultante, há diferentes variações do RP [53]. A abordagem denominada *unthresholded* RP [39] mapeia as distâncias entre pares de pontos numa série temporal para uma escala de cinza, proporcionando uma visualização mais rica. Para isso, as distâncias calculadas são normalizadas, e então os valores normalizados são invertidos, de forma que distâncias menores correspondam a tonalidades mais escuras na escala de cinza. Isso indica uma maior proximidade ou similaridade entre os pontos analisados.

Gramian Angular Field (GAF)

A representação com GAF [82] realiza a conversão dos valores da série temporal em ângulos e utiliza esses ângulos para gerar uma matriz que captura as relações temporais.

Dada uma série temporal $X = \{x_1, x_2, \dots, x_N\}$, a transformação para GAF envolve normalizar X para que todos os seus valores estejam no intervalo $[-1, 1]$. Em seguida é feita a transformação angular, que consiste em converter cada valor normalizado $\tilde{x}_i$ em um ângulo ϕ_i usando arco cosseno:

$$\phi_i = \arccos(\tilde{x}_i) \quad (3.2)$$

Há duas variantes principais do GAF:

- *Gramian Angular Summation Field* (GASF):

$$G_{i,j}^{(\text{GASF})} = \cos(\phi_i + \phi_j) \quad (3.3)$$

Tendo a matriz definida por 3.4

$$GASF = \begin{bmatrix} \cos(\phi_1 + \phi_1) & \cos(\phi_1 + \phi_2) & \cdots & \cos(\phi_1 + \phi_N) \\ \cos(\phi_2 + \phi_1) & \cos(\phi_2 + \phi_2) & \cdots & \cos(\phi_2 + \phi_N) \\ \vdots & \vdots & \ddots & \vdots \\ \cos(\phi_N + \phi_1) & \cos(\phi_N + \phi_2) & \cdots & \cos(\phi_N + \phi_N) \end{bmatrix} \quad (3.4)$$

- *Gramian Angular Difference Field* (GADF):

$$G_{i,j}^{(\text{GADF})} = \sin(\phi_i - \phi_j) \quad (3.5)$$

A matriz é definida por 3.6

$$GADF = \begin{bmatrix} \sin(\phi_1 - \phi_1) & \sin(\phi_1 - \phi_2) & \cdots & \sin(\phi_1 - \phi_N) \\ \sin(\phi_2 - \phi_1) & \sin(\phi_2 - \phi_2) & \cdots & \sin(\phi_2 - \phi_N) \\ \vdots & \vdots & \ddots & \vdots \\ \sin(\phi_N - \phi_1) & \sin(\phi_N - \phi_2) & \cdots & \sin(\phi_N - \phi_N) \end{bmatrix} \quad (3.6)$$

Estas matrizes, $G^{(\text{GASF})}$ e $G^{(\text{GADF})}$, codificam as relações angulares entre os pontos da série temporal. Elas podem ser visualizadas como imagens, nas quais padrões e estruturas podem ser identificados visualmente. Neste contexto, a representação com GADF mostrou-se mais adequada por destacar mudanças nas tendências de séries temporais, tendo em vista que o GADF enfatiza as diferenças angulares entre pontos de tempo consecutivos [64].

***Markov Transition Field* (MTF)**

Cada elemento da matriz MTF [82] reflete a probabilidade de transição entre dois estados em momentos diferentes da série temporal, capturando a dinâmica temporal da série em uma representação bidimensional.

Dada uma série temporal $X = \{x_1, x_2, \dots, x_N\}$, a construção do MTF é realizada a partir da discretização em Q quantis. Cada valor x_i de X é atribuído a um quantil correspondente, resultando em uma série discretizada $\tilde{X} = \{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_N\}$, na qual $\tilde{x}_i$ representa o quantil ao qual x_i pertence.

A matriz de transição de Markov W de dimensão $Q \times Q$ consiste em representar w_{ij} como a frequência com que um ponto no quantil q_j é seguido por um ponto no quantil q_i ($q_j \rightarrow q_i$). Após a normalização, na qual $\sum_{j=1}^Q w_{ij} = 1$ para cada i , W torna-se a matriz de transição de Markov. A matriz W antes da normalização é uma matriz de contagem de transições e pode ser visualizada como:

$$W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1Q} \\ w_{21} & w_{22} & \cdots & w_{2Q} \\ \vdots & \vdots & \ddots & \vdots \\ w_{Q1} & w_{Q2} & \cdots & w_{QQ} \end{bmatrix}, \quad (3.7)$$

sendo w_{ij} a contagem de ocorrências de transição na série temporal.

A MTF é uma matriz $Q \times Q$ e, na sua construção, cada elemento M_{ij} denota a probabilidade de transição do quantil q_i para o quantil q_j , considerando as posições temporais na série. MTF codifica, assim, as probabilidades de transição de multi-escala da série temporal. Por exemplo, M_{ij} com $|i - j| = k$ representa a probabilidade de transição entre pontos com intervalo temporal k . A diagonal principal M_{ii} captura as probabilidades de auto-transição.

Espectrogramas

Essa transformação é feita por meio da aplicação da *Short-Time Fourier Transform* (STFT), que basicamente decompõe o sinal em suas frequências componentes em diferentes instantes de tempo, permitindo visualizar como o espectro de frequência do sinal varia com o tempo [85].

Dada uma série temporal $x(t)$, a STFT é definida como:

$$STFT\{x(n)\}(m, k) = \sum_{n=-\infty}^{\infty} x(n)w(n - m)e^{-j\frac{2\pi k(n-m)}{N}}, \quad (3.8)$$

sendo:

- $x(n)$ é o valor da série temporal no instante n ;
- $w(n - m)$ é a função de janela aplicada ao sinal, deslocada pelo índice do quadro m ;
- N é o número total de pontos utilizado na transformação de Fourier, influenciando a resolução da frequência;
- $e^{-j\frac{2\pi k(n-m)}{N}}$ representa a base da transformada de Fourier;
- m indica a posição do quadro atual;
- k é o índice de frequência.

A função de janela de Hann é definida por:

$$w(n) = \begin{cases} 0.5 \left[1 - \cos \left(\frac{2\pi n}{M-1} \right) \right], & \text{se } 0 \leq n \leq M-1 \\ 0, & \text{caso contrário} \end{cases} \quad (3.9)$$

A série temporal é dividida em segmentos de tempo menores e sobrepostos, utilizando uma função de janela e a transformada de Fourier assume que o sinal é periódico, o que não é verdade para a maioria dos sinais reais. Então, a janela limita o sinal a um intervalo

de tempo finito, permitindo a análise frequencial local. Em seguida, a transformada de Fourier é aplicada a cada segmento de tempo, transformando os dados do domínio do tempo para o domínio da frequência. Isso fornece a amplitude e a fase das frequências presentes em cada segmento de tempo.

Os resultados da transformada de Fourier para cada segmento são organizados em uma matriz, na qual uma dimensão representa o tempo (os segmentos de tempo), e a outra representa a frequência. Os valores na matriz representam a magnitude das frequências em cada segmento de tempo.

A magnitude ao quadrado da STFT é frequentemente utilizada para construir o espectrograma do sinal, que é uma representação visual da intensidade das frequências em função do tempo:

$$\text{Espectrograma}(x(n)) = |STFT\{x(n)\}(m, k)|^2 \quad (3.10)$$

No espectrograma, o eixo x representa o tempo, o eixo y representa a frequência, e a intensidade das cores ou a escala de cinza em um ponto específico representa a magnitude da frequência naquele ponto do tempo.

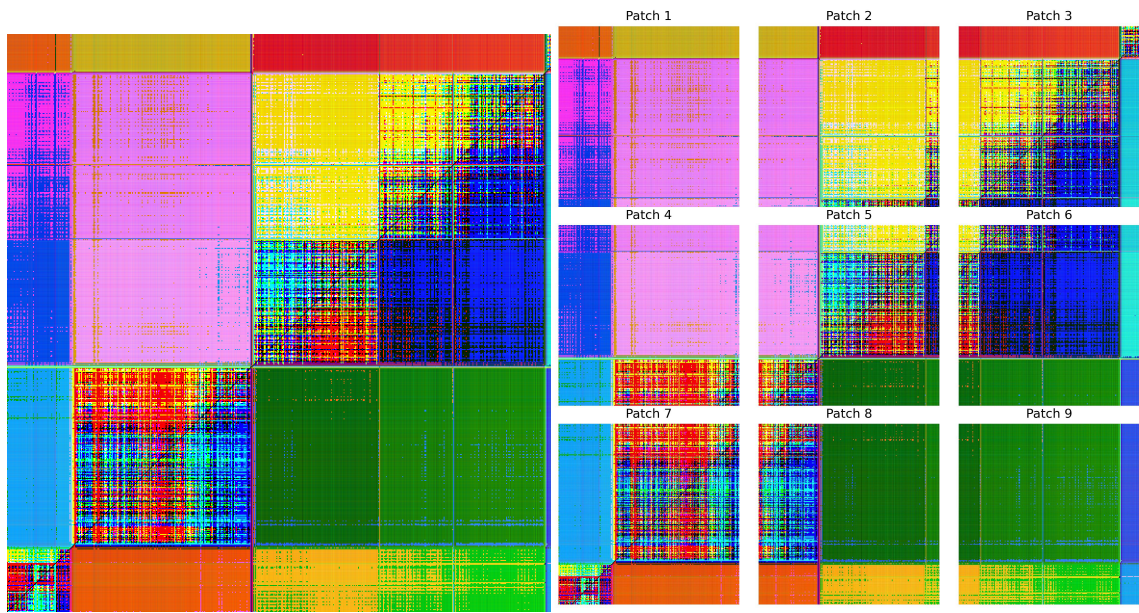
3.1.3 *Patches*

A técnica de dividir uma imagem em subimagens, ou *patches*, é uma abordagem comum em aplicações de visualização de imagens e aprendizado de máquina [48, 68], tendo como principal objetivo melhorar o foco em detalhes locais da imagem. Isso permite que o modelo aprenda características mais refinadas que podem ser perdidas quando se observa a imagem original.

Como pode ser observado no exemplo da Figura 3.3, cada imagem gerada pelas diferentes representações (600×600 ou outras dimensões, no caso dos espectrogramas) é dividida em 9 *patches* e cada *patch* (224×224 , ou seja, com uma pequena sobreposição entre os patches, para evitar a perda de informação local) é tratado como uma entrada independente para o processo de treinamento. Após o treinamento, os *patches* podem ser reagrupados para fazer previsões sobre a imagem inteira, ou seja, o modelo pode prever a classe de cada região do *patch* correspondente e, em seguida, combinar essas previsões para formar a imagem segmentada completa.

3.1.4 *Ensembles*

Técnicas de *ensemble* são utilizadas para combinar as previsões de múltiplos modelos de forma a melhorar a robustez e a precisão da predição final. Nesse sentido, um *ensemble* utiliza dois (*ensemble* de um modelo treinado com dados de acelerômetro e outro com dados de frequência cardíaca) ou vários modelos (*ensemble* envolvendo os modelos treinados com *patches*) para obter uma predição final, tentando explorar as decisões individuais e mitigar as desvantagens de cada modelo. Essas técnicas são úteis em problemas complexos, nos quais um único modelo pode não ser suficiente para capturar todas as nuances dos dados [28].



(a) Exemplo de imagem original (600×600). (b) Exemplos de *patches* (224×224 cada).

Figura 3.3: Imagens de “vigília” geradas com GAF e dados de acelerômetro.

Os *ensembles* de média simples, média ponderada, com rede simples e com *deep features* são descritas a seguir.

Média Simples

No *ensemble* de média simples, as previsões de cada classificador são combinadas calculando a média aritmética das previsões. Para o problema de classificação binária, isso significa calcular a média das probabilidades preditas para as classes “sono” e “vigília” pelos classificadores; e para a classificação de três classes, a média das previsões para cada uma das classes é calculada individualmente. Este tipo de *ensemble* foi aplicado para combinar as previsões dos modelos treinados com:

- imagens originais dos dados de acelerômetro + imagens originais dos dados de frequência cardíaca;
- *patches* dos dados de acelerômetro;
- *patches* dos dados de frequência cardíaca; e
- *patches* dos dados de acelerômetro + *patches* dos dados de frequência cardíaca.

Média Ponderada

A média ponderada é uma variação do método de média simples, na qual cada previsão do classificador contribui com um peso diferente para a previsão final. Os pesos são geralmente atribuídos com base no desempenho de cada classificador. Foram gerados 1000 conjuntos aleatórios de n pesos para encontrar a melhor combinação em cada caso. Este tipo de *ensemble* foi aplicado para combinar as previsões dos modelos treinados com:

- imagens originais dos dados de acelerômetro + imagens originais dos dados de frequência cardíaca ($n = 2$);
- *patches* dos dados de acelerômetro ($n = \text{número de patches} = 9$);
- *patches* dos dados de frequência cardíaca ($n = \text{número de patches} = 9$); e
- *patches* dos dados de acelerômetro + *patches* dos dados de frequência cardíaca ($n = 9 \times 2 = \text{número de patches} \times \text{número de tipos de dados}$).

Rede Simples

Neste método, as predições dos classificadores são utilizadas como entrada para uma rede neural simples, que aprende a melhor forma de combinar essas predições. As probabilidades preditas para as classes de interesse por um classificador treinado com os *patches* são as entradas para esta rede, que então fornece a previsão final. Com o intuito de manter apenas uma rede simples, este *ensemble* foi aplicado apenas com as predições dos modelos treinados com:

- *patches* dos dados de acelerômetro; e
- *patches* dos dados de frequência cardíaca.

Utilizando os modelos treinados apenas com *patches*, várias predições contribuem para uma entrada mais informativa, diferentemente do uso de acelerômetro e frequência cardíaca, na qual apenas dois modelos contribuiriam. Isso faz com que uma rede simples seja menos vantajosa para combinar somente duas predições, devido ao esforço adicional que não se justifica pela complexidade do problema.

Deep Features

Neste tipo de *ensemble*, vetores de características são extraídos das camadas profundas de cada classificador e combinadas para serem utilizadas como entrada para um modelo final. Este modelo final é treinado para fazer a previsão final usando essas *features* combinadas, aproveitando as representações de dados fornecidas pelos diferentes classificadores. Este tipo de *ensemble* foi aplicado para combinar as predições dos modelos treinados com:

- imagens originais dos dados de acelerômetro + imagens originais dos dados de frequência cardíaca.

Embora a diversidade de informações fornecidas pelos vários vetores de *deep features* de *patches* possa ser informativa, ela também introduz uma complexidade significativa no processo de modelagem. Essa complexidade se manifesta na dimensão dos dados a serem processados. Nesse sentido, trabalhar com apenas dois vetores de *deep features*, gerados a partir dos dados de acelerômetro e frequência cardíaca, simplifica o processo de modelagem.

3.2 Métricas de Avaliação

A fim de facilitar a comparação do desempenho do modelo proposto nesta dissertação com os encontrados na literatura, serão utilizadas as métricas comumente apresentadas nos trabalhos da área. Além da acurácia, essas medidas incluem sensibilidade, especificidade e Coeficiente Kappa (κ) de Cohen.

Embora os trabalhos relacionados apresentem a acurácia como principal medida de desempenho do modelo, nesta dissertação 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 [10]. Como pode ser observado na Equação 3.12, a acurácia balanceada é a média das sensibilidades (*recall*) de todas as c classes. Na Equação 3.11, VP_n é o número de verdadeiros positivos da classe n e FN_n é o número de falsos positivos da classe n .

$$sensibilidade_n = \frac{VP_n}{VP_n + FN_n} \quad (3.11)$$

$$\text{acurácia balanceada} = \frac{\sum_{n=1}^{n=c} sensibilidade_n}{c} \quad (3.12)$$

Na detecção sono/vigília, 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.

O coeficiente κ (kappa) de Cohen é a concordância entre diferentes avaliadores para problemas multiclasse 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 κ feita por Cohen [15] é descrita na Equação 3.13.

$$\kappa = \frac{\% \text{concordância observada} - \% \text{concordância por acaso}}{1 - \% \text{concordância por acaso}} \quad (3.13)$$

Os valores de κ 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 Conjunto de Dados

Neste trabalho foi utilizada a base de dados *Sleep Accel*, coletada com *Apple Watch*¹ (séries 2 e 3, *Apple Inc.*) [81], principalmente por ser a base de dados mais adequada disponível publicamente. 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. Para cada dado obtido, há uma informação de tempo (em segundos desde o início da PSG). A Figura 3.4 mostra os estágios do sono de um dos indivíduos deste conjunto.

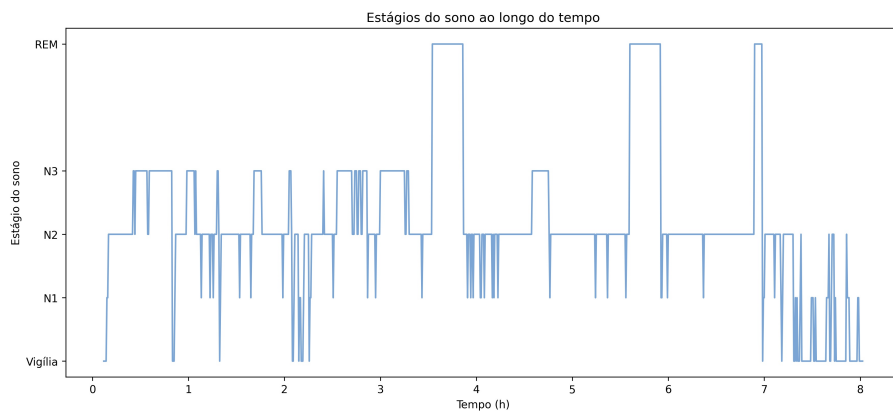
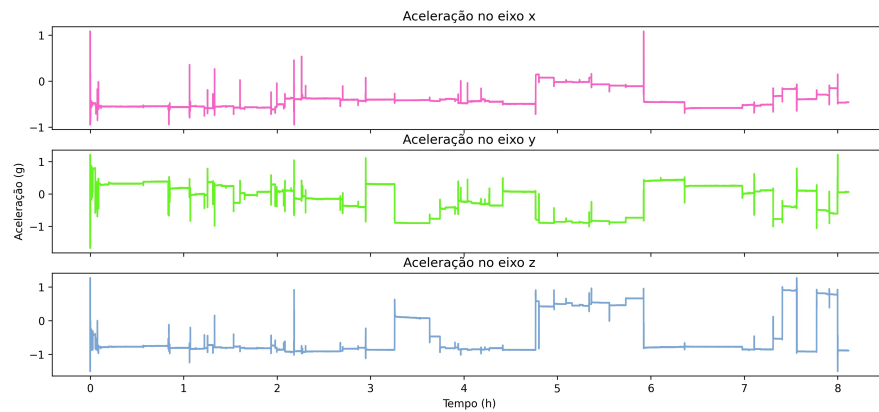


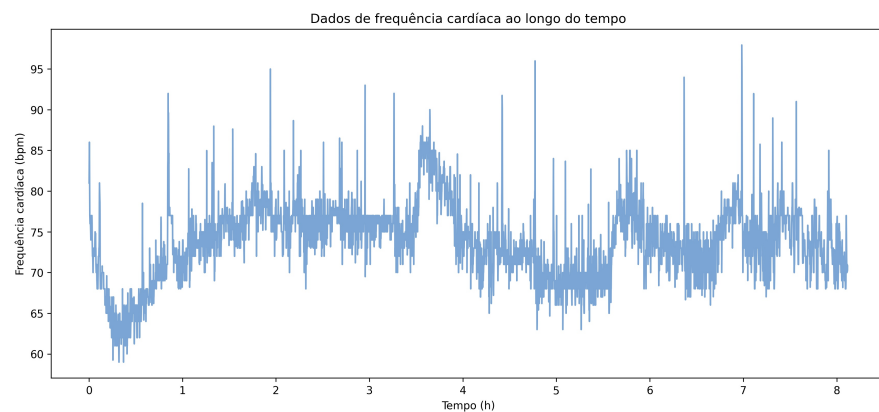
Figura 3.4: Estágios do sono de um indivíduo da base de dados *Sleep Accel* ao longo de 8 horas.

O *Apple Watch* utiliza 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). Os dados de acelerômetro e frequência cardíaca coletados ao longo de 8 horas de sono de um indivíduo da base de dados podem ser visualizados nas figuras 3.5a e 3.5b, respectivamente.

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



(a) Dados de acelerômetro.



(b) Dados de frequência cardíaca.

Figura 3.5: Dados da base *Sleep Accel* [81] de um indivíduo, coletados ao longo de 8 horas.

O *Apple Watch* foi utilizado durante um período de 7 a 14 dias e, 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*. Os participantes deste estudo não possuem diagnóstico conhecido de distúrbios do sono.

É importante destacar que este conjunto de dados é desbalanceado considerando a classificação de 2 estágios (sono e vigília) e, também, a classificação de três estágios (vigília, NREM e REM). A proporção das classes é 12 amostras “sono” para 1 amostra “vigília” (classificação binária); e 9 amostras “NREM” para 3 amostras “REM” para 1 amostra “vigília” (classificação de três estágios). A proporção e o desequilíbrio das classes ficam visíveis no gráfico da Figura 3.6, que apresenta a porcentagem de dados de cada classe a cada 1 hora.

3.3.1 Artigo de Referência da Base de Dados

Alguns dos resultados obtidos foram comparados com o artigo de referência para a base *Sleep Accel* [81], descrito na Seção 2.2 (Trabalhos Relacionados).

Esse estudo compara diferentes *features* para classificar sono/vigília e estágios do sono. Uma delas é o “tempo”, que é baseada no tempo desde o início da gravação de PSG.

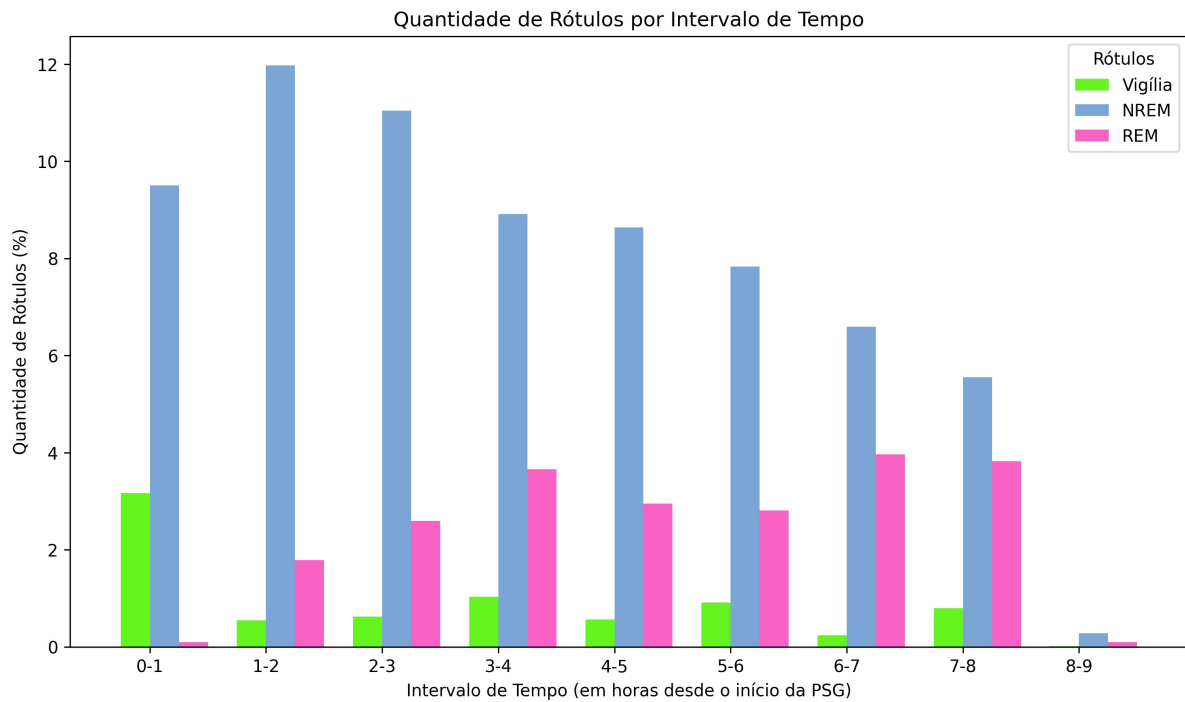


Figura 3.6: Quantidade de rótulos da base de dados *Sleep Accel* por intervalo de tempo ao longo de 8 horas (considerando três estágios do sono).

Idealmente, as *features* utilizadas para prever os estágios do sono devem ser independentes do tempo específico de uma sessão de PSG, tendo em vista que, em situações reais, apenas dados do *smartwatch* serão utilizados. Outros problemas encontrados no trabalho dos autores incluem modelos com *overfitting* e a ausência de um conjunto de validação. Além disso, uma das métricas utilizadas pelos autores é a acurácia, que quando utilizada em conjuntos de dados desequilibrados, como esse, pode apresentar resultados enganosos. A acurácia balanceada, por outro lado, considera a proporção de cada classe, oferecendo uma visão mais justa do desempenho do modelo e é a mais adequada para esse tipo de base.

3.4 Ferramentas e Recursos Computacionais

A implementação foi realizada na linguagem Python, tendo em vista que há um grande número de ferramentas desenvolvidas com suporte para essa linguagem e com boa documentação. Foram utilizadas bibliotecas científicas, numéricas, de redes neurais, de apresentação de gráficos, entre outras. Algumas bibliotecas que podem ser destacadas são: NumPy², pandas³, Matplotlib⁴, scikit-learn⁵ e TensorFlow⁶.

Os experimentos desta dissertação foram realizados em dois ambientes: Kubeflow,

²<https://numpy.org/>

³<https://pandas.pydata.org/>

⁴<https://matplotlib.org/>

⁵<https://scikit-learn.org/>

⁶<https://www.tensorflow.org/>

fornecido pelo projeto Viva Bem ⁷; e no Laboratório de Inteligência Artificial (Recod.ai) do Instituto de Computação (IC – Unicamp). Ambos os ambientes oferecem recursos de computação, como CPU, GPU, memória RAM e armazenamento.

⁷<https://vivabem.unicamp.br>

Capítulo 4

Experimentos e Resultados

Neste capítulo, os experimentos conduzidos são detalhados e os principais resultados obtidos, assim como a discussão, são apresentados.

4.1 Experimentos

Nesta seção, são abordadas as informações sobre o treinamento e validação, assim como a comparação com outras representações, e a descrição dos experimentos realizados.

4.1.1 Treinamento e Validação

Para o treinamento, foi empregada a abordagem de transferência de aprendizado utilizando a rede *EfficientNet* pré-treinada no extenso conjunto de dados *ImageNet*. A estratégia inclui o congelamento de algumas das camadas iniciais dessas redes para preservar os recursos genéricos aprendidos, enquanto as camadas mais profundas foram adaptadas ao conjunto de dados específico. Essa adaptação foi realizada por meio da inclusão de uma rede densa adicional no final da arquitetura, processo conhecido como *fine tuning*, permitindo um ajuste fino dos parâmetros da rede para melhor adequação às classes de interesse.

Por meio de testes de congelamento de camadas, foi estabelecido utilizar 90% das primeiras camadas da rede *EfficientNet* congeladas. Além disso, foram adicionadas uma camada densa de tamanho 512 com 50% de *dropout* e uma camada densa de tamanho 256 com 20% de *dropout*.

Para avaliar a robustez e a generalização dos modelos, foi aplicada a técnica de validação cruzada *k-fold*, com $k = 5$, dividindo o conjunto de dados em cinco partições distintas para garantir que cada amostra não fosse utilizada tanto para treinamento quanto para validação. É importante destacar que dados de um mesmo indivíduo não foram utilizados para treinamento e validação simultaneamente.

Reconhecendo o desafio imposto pelo desbalanceamento das classes no conjunto de dados, foi implementada a atribuição de pesos às classes (inversamente proporcionais aos tamanhos das classes), tentando mitigar o efeito do desequilíbrio e promover um treinamento mais equitativo e representativo.

4.1.2 Comparação com Outras Representações

A representação de dados de séries temporais por meio de imagens para classificar os estágios do sono foi comparada com duas outras abordagens populares: a extração de *features* e o uso de *Deep Learning* aplicado diretamente aos dados brutos. A motivação para essa comparação reside na necessidade de explorar e quantificar a eficácia de diferentes métodos de processamento de séries temporais, visando identificar qual abordagem maximiza o desempenho do modelo em termos das métricas de avaliação.

A extração de *features* busca resumir os dados brutos em um conjunto compacto de indicadores que capturam aspectos essenciais da série temporal, enquanto a aplicação direta de *Deep Learning* aos dados brutos visa aprender representações complexas sem a necessidade de pré-processamento manual. Comparando essas estratégias com a técnica proposta, é possível determinar se a nova abordagem oferece vantagens significativas, como a redução da perda de informações críticas ou a melhoria na classificação de padrões dos dados de sono.

4.1.3 Descrição dos Experimentos

Os experimentos foram realizados para dois cenários: classificação de sono/vigília e classificação dos estágios do sono. Os modelos possuem uma mesma arquitetura, baseada na rede *EfficientNet*, como descrita na Seção 3.1. Além disso, os experimentos realizados com os dados de acelerômetro e frequência cardíaca permitem comparar as seguintes configurações dos dados: original, que correspondem às imagens redimensionadas para 224×224 como entradas para o modelo; *patch n*, de dimensão 224×224 obtida da imagem de dimensão 600×600 na posição *n*; e diferentes *ensembles*:

- *Ensembles AC + FC* (imagens originais dos dados de acelerômetro + imagens originais dos dados de frequência cardíaca):
 - média simples;
 - média ponderada; e
 - *deep features*.
- *Ensembles dos Patches* dos dados de acelerômetro:
 - média simples;
 - média ponderada; e
 - rede simples.
- *Ensembles dos Patches* dos dados de frequência cardíaca:
 - média simples;
 - média ponderada; e
 - rede simples.

- *Ensembles patches AC + patches FC* (*patches* dos dados de acelerômetro + *patches* dos dados de frequência cardíaca):
 - média simples; e
 - média ponderada.

Os resultados de cada representação correspondem a uma cor nas tabelas e os melhores resultados para cada bloco de linhas estão em negrito. É importante destacar que todos os valores são médias de 5 *folds*, como descrito na seção de metodologia (Seção 3.1). Para os melhores resultados de acurácias balanceadas obtidos com cada configuração de dados, é exibida a matriz de confusão correspondente.

4.2 Resultados e Discussões

Nesta seção, serão apresentados os principais resultados obtidos com a base de dados *Sleep Accel* [81] descrita na Seção 3.3 para as classificações de sono/vigília e dos estágios do sono, bem como as discussões desses resultados. Em seguida, é realizada a comparação da representação visual com outras representações de dados tipicamente utilizadas em trabalhos relacionados.

4.2.1 Classificação de Sono/Vigília

Nesta subseção, são apresentadas as acurácias balanceadas, os coeficientes κ e as matrizes de confusão obtidas para a classificação de sono/vigília.

Acurácia Balanceada e Coeficiente κ de Cohen

A Tabela 4.1 apresenta as acurácias balanceadas obtidas com cada representação para a classificação de sono/vigília.

É possível observar que para este cenário, os dados de acelerômetro, comparados com os de frequência cardíaca, apresentam melhores resultados, com todas as representações, tanto para as imagens Originais quanto para os *Patches* individuais e para os *Ensembles* dos *Patches* individuais.

Observando os resultados com os *ensembles* que combinam os diferentes tipos de dados (AC e FC), nota-se que em nenhum caso esses apresentam melhores acurácias balanceadas em comparação com as obtidas com os dados de acelerômetro. Ou seja, os resultados obtidos com os dados de frequência cardíaca combinados com os de acelerômetro não contribuem para um ganho de acurácia balanceada neste cenário. Por outro lado, os resultados obtidos com os *Ensembles* dos *Patches* individuais, tanto para os dados de acelerômetro quanto para os de frequência cardíaca, apresentam um ganho significativo comparados com os resultados obtidos com as Originais (até 3,7 pontos percentuais para RP, até 2,7 para GAF, até 2,9 para MTF e até 0,7 para Espectrogramas).

Comparando as diferentes representações, a GAF alcançou 82,0% de acurácia balanceada com o *ensemble* dos *patches* por meio de média simples e dados de acelerômetro, indicando o melhor resultado para a classificação de sono/vigília. As representações RP e

		RP		GAF		MTF		Espectrogramas	
Rede	Config.	AC	FC	AC	FC	AC	FC	AC	FC
Eff.Net	Original	76.620	69.910	79.630	69.318	77.436	64.882	78.340	61.216
Ensembles AC + FC	M. Simples	71.382		69.038		71.269		75.937	
	M. Ponderada	76.541		77.569		74.946		78.190	
	Deep Features	75.592		77.610		75.430		77.124	
Eff.Net	Patch 1	76.796	66.364	76.790	65.686	74.298	63.212	72.044	51.074
	Patch 2	77.916	68.840	77.734	67.934	75.794	64.664	76.564	51.235
	Patch 3	73.486	64.176	73.166	64.736	71.452	61.166	71.658	51.250
	Patch 4	79.970	69.010	79.568	68.068	78.572	64.990	76.686	54.830
	Patch 5	76.852	66.056	78.704	69.044	76.818	65.122	76.432	52.832
	Patch 6	78.206	67.444	77.738	68.506	75.482	65.874	71.900	52.376
	Patch 7	74.058	62.856	72.366	62.610	72.376	60.772	77.718	53.138
	Patch 8	80.302	69.396	79.828	67.680	78.078	66.412	76.360	51.440
	Patch 9	76.320	64.990	77.736	63.938	74.792	63.210	71.938	53.070
Ensembles dos Patches	M. Simples	80.388	71.393	82.358	73.712	80.034	70.209	79.105	54.903
	M. Ponderada	80.382	71.282	82.035	72.148	80.203	68.777	79.007	54.907
	Rede Simples	80.259	70.812	81.544	72.874	79.853	69.410	79.032	53.302
Ensembles P. AC + P. FC	M. Simples	76.144		77.245		78.283		77.778	
	M. Ponderada	79.276		81.444		80.327		78.641	

Tabela 4.1: Acurácias balanceadas obtidas com cada representação para a classificação de sono/vigília.

MTF apresentam acurácias balanceadas maiores que 80,0% com *Ensembles* dos *Patches* e dados de acelerômetro (exceto MTF e *ensemble* dos *patches* com rede simples). Com GAF e MTF, o *Ensemble* dos *Patches* AC + *Patches* FC também apresenta resultados maiores que 80,0% e os Espectrogramas superam 78,0%.

Os desvios-padrão obtidos das acurácias balanceadas para a classificação de sono/vigília são de até 3,0% para RP; até 4,0% para GAF; até 6,0% para MTF; e até 4,0% para Espectrogramas.

Os melhores coeficientes κ , conforme indica a Tabela 4.2, estão concentrados nos resultados do *Ensemble* AC + FC com *deep features*, passando de 0,4 para a representação GAF (concordância moderada). Para a maioria das configurações (exceto *patches* 1 e 3 da MTF), os κ obtidos com dados de acelerômetro passam de 0,2, indicando uma concordância regular. Enquanto que com os dados de frequência cardíaca, κ ultrapassa 0,2 apenas com a aplicação de *ensembles* (e com o *patch* 6 da GAF) e são menores com os Espectrogramas. Mostrando novamente, assim, a importância dos dados de acelerômetro para a classificação de sono/vigília.

Matrizes de Confusão

As figuras 4.1, 4.2, 4.3 e 4.4 mostram, respectivamente, as matrizes de confusão geradas para a classificação de sono/vigília utilizando as representações RP, GAF, MTF e Espectrogramas com os dados originais de acelerômetro (figuras 4.1a, 4.2a, 4.3a e 4.4a) e frequência cardíaca (figuras 4.1b, 4.2b, 4.3b e 4.4b) e melhores *ensembles* acelerômetro + frequência cardíaca (figuras 4.1c, 4.2c, 4.3c e 4.4c). As figuras também incluem as matrizes de confusão dos melhores *ensembles* dos *patches* dos dados de acelerômetro (figuras 4.1d, 4.2d, 4.3d e 4.4d) e dos *patches* dos dados de frequência cardíaca (figuras 4.1e, 4.2e, 4.3e e 4.4e) e melhores *ensembles* dos *patches* dos dados de acelerômetro + *patches* dos dados de frequência cardíaca (figuras 4.1f, 4.2f, 4.3f e 4.4f).

Observando as matrizes de confusão correspondentes à representação RP, nota-se que,

Rede	Config.	RP		GAF		MTF		Espectrogramas	
		AC	FC	AC	FC	AC	FC	AC	FC
Eff.Net	Original	0.317	0.178	0.353	0.184	0.322	0.152	0.297	0.108
Ensembles AC + FC	M. Simples	0.237		0.213		0.320		0.290	
	M. Ponderada	0.320		0.319		0.334		0.314	
	Deep Features	0.339		0.418		0.351		0.389	
Eff.Net	Patch 1	0.256	0.141	0.267	0.170	0.181	0.125	0.241	0.105
	Patch 2	0.319	0.155	0.324	0.163	0.234	0.120	0.245	0.110
	Patch 3	0.199	0.097	0.225	0.128	0.180	0.114	0.259	0.104
	Patch 4	0.354	0.159	0.348	0.165	0.314	0.129	0.315	0.119
	Patch 5	0.335	0.124	0.328	0.147	0.289	0.099	0.300	0.099
	Patch 6	0.281	0.149	0.294	0.211	0.288	0.123	0.248	0.113
	Patch 7	0.215	0.110	0.269	0.097	0.228	0.082	0.342	0.082
	Patch 8	0.333	0.178	0.368	0.183	0.303	0.113	0.288	0.103
	Patch 9	0.277	0.131	0.286	0.115	0.247	0.099	0.232	0.099
Ensembles dos Patches	M. Simples	0.299	0.195	0.370	0.259	0.300	0.171	0.350	0.141
	M. Ponderada	0.306	0.219	0.372	0.279	0.327	0.187	0.339	0.137
	Rede Simples	0.304	0.183	0.355	0.255	0.303	0.174	0.340	0.124
Ensembles P. AC + P. FC	M. Simples	0.281		0.361		0.297		0.262	
	M. Ponderada	0.303		0.390		0.337		0.310	

Tabela 4.2: Coeficientes κ de Cohen obtidos com cada representação para a classificação de sono/vigília.

o *ensemble* de dados originais de acelerômetro + dados originais de frequência cardíaca apresenta maior sensibilidade que as demais. Com os dados originais de frequência cardíaca (Figura 4.1b), o modelo classificou mais “sono” como “vigília” do que o modelo correspondente aos dados originais de acelerômetro (Figura 4.1a), ou seja, utilizando dados originais de frequência cardíaca, o modelo apresentou mais erros comparados com dados originais de acelerômetro para classificar “sono” corretamente. Fazendo a análise das matrizes de confusão obtidas com dados originais versus *patches*, é possível observar a melhora da classificação de “vigília”, tendo em vista que tanto acelerômetro (Figura 4.1a versus Figura 4.1d) quanto para frequência cardíaca (Figura 4.1b versus Figura 4.1e) os acertos de “vigília” aumentaram com o uso de *patches*. A classificação de sono/vigília é mais equilibrada utilizando dados de acelerômetro e, conseqüentemente, o *ensemble* de *patches* dos dados de acelerômetro + *patches* dos dados de frequência cardíaca (Figura 4.1f).

As matrizes de confusão geradas para a representação GAF, exibidas na Figura 4.2, mostram que a classificação de sono/vigília utilizando dados originais de frequência cardíaca (Figura 4.2b) é mais balanceada que a com dados originais de acelerômetro (Figura 4.2a), confundindo, assim, as duas classes de forma semelhante. É possível notar que a classificação com dados originais de acelerômetro superestima a quantidade de “sono” e que, ao melhorar as predições para “vigília” utilizando *patches*, a classificação com dados de acelerômetro é balanceada (Figura 4.2d). Comparando os dois *ensembles* (Figura 4.2c versus Figura 4.2f), observa-se que ambos superestimam a quantidade de “sono”, ou seja, comentem mais erros e acertos de predição de “sono”. Assim, o *ensemble* dos *patches* de acelerômetro + *patches* de frequência cardíaca demonstra sensibilidade de 85,0%, sendo a maior para GAF.

Ao observar as matrizes de confusão geradas com a representação MTF, na Figura 4.3, nota-se que utilizando dados originais, tanto de acelerômetro quanto de frequência car-

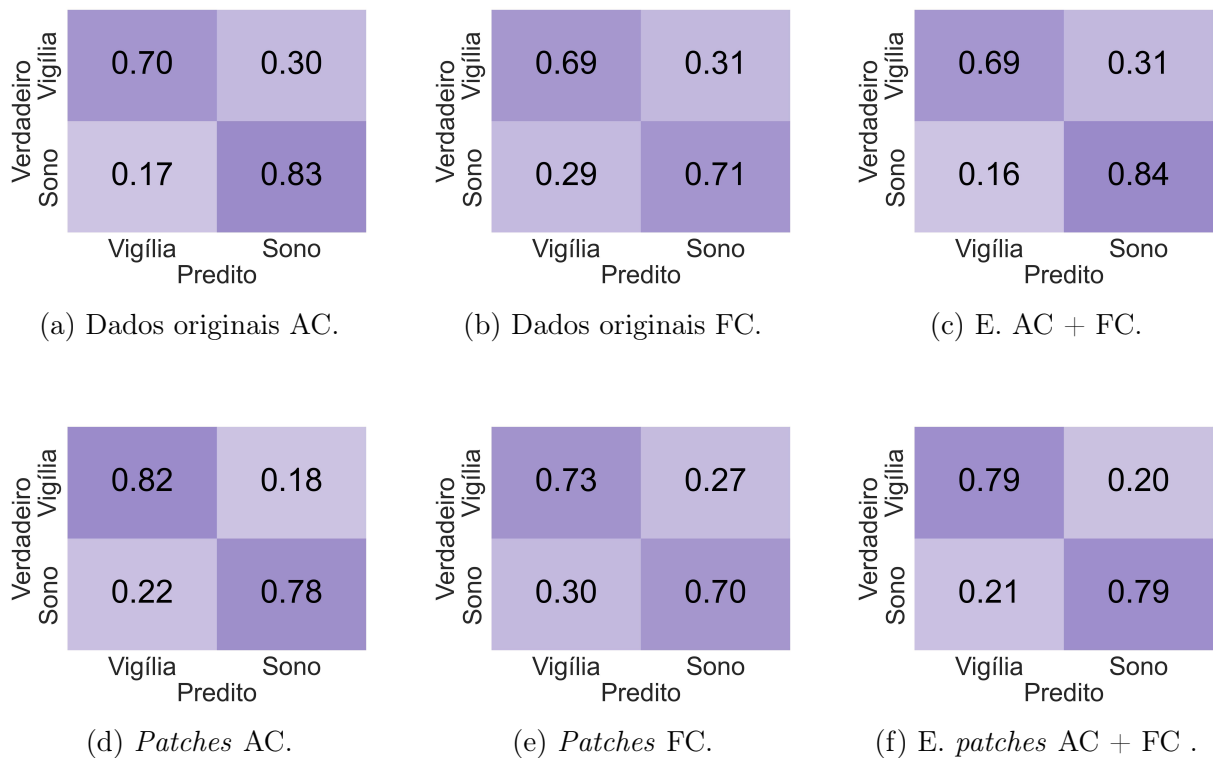


Figura 4.1: RP: matrizes de confusão geradas com os resultados obtidos por meio dos dados originais e dos melhores resultados de *ensembles* para a classificação de sono/vigília.

díaca (figuras 4.3a e 4.3b), a classificação correta de “sono” é melhor que a de “vigília”, e isso também reflete no *ensemble* com esses dados (Figura 4.3c). Além disso, destaca-se a baixa capacidade de classificação de “vigília” correta com dados originais de frequência cardíaca. Utilizando *patches* e dados de acelerômetro (Figura 4.3d), assim como nas representações RP e GAF, a classificação é mais balanceada que com os dados originais. Consequentemente, a matriz de confusão do *ensemble* dos *patches* dos dados de acelerômetro + *patches* dos dados de frequência cardíaca (Figura 4.3f) também é mais balanceada que a do *ensemble* dos dados originais, que chega a sensibilidade de 87,0%.

A representação por Espectrogramas apresenta as matrizes de confusão da Figura 4.4. Ao analisar as classificações obtidas com os dados originais, presentes nas figuras 4.4a e 4.4b, nota-se que ambas são balanceadas, enquanto que o *ensemble* com esses dados apresenta mais facilidade para classificar “sono” corretamente (Figura 4.4c). Ao utilizar *patches* dos dados de acelerômetro, a capacidade de classificar “vigília” corretamente melhora (Figura 4.4d). Com dados de frequência cardíaca, as “vigílias” são superestimadas, fazendo com que a classificação correta de “sono” seja drasticamente afetada (Figura 4.4e). Os dois tipos de *ensembles* (figuras 4.4c e 4.4f) apresentaram matriz de confusão similar, sendo ambos ligeiramente melhores para classificar “sono” corretamente, e apresentaram a maior sensibilidade obtida com essa representação, chegando a 80,0%.

As figura 4.5a e 4.5b ilustram as predições de uma noite de sono de um mesmo indivíduo para a classificação de sono/vigília utilizando dados originais de acelerômetro e dados do *ensemble* dos *patches* dos dados de acelerômetro. Observando essas figuras, é possível comparar a melhor noite de sono obtida para a classificação de sono/vigília (*ensemble*

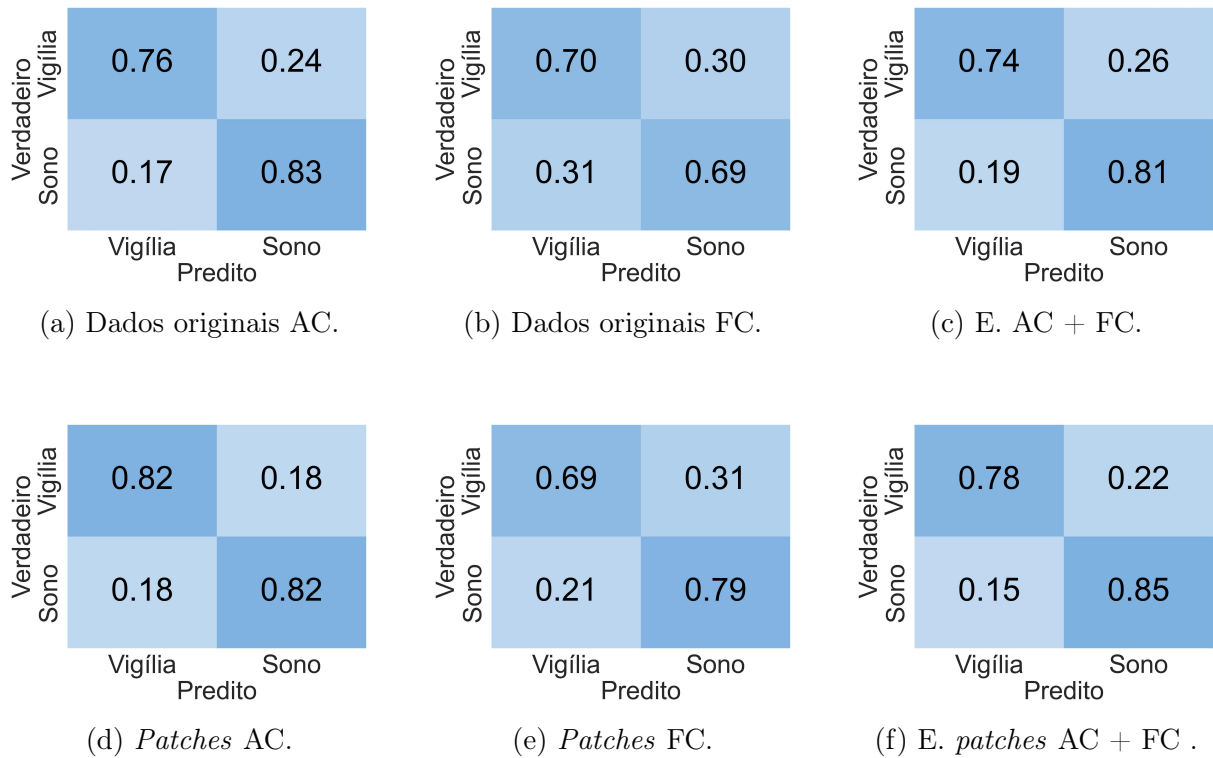


Figura 4.2: GAF: matrizes de confusão geradas com os resultados obtidos por meio dos dados originais e dos melhores resultados de *ensembles* para a classificação de sono/vigília.

dos *patches* dos dados de acelerômetro), que foi obtida por meio da GAF, com a noite de sono correspondente utilizando dados originais. Pode-se observar, principalmente, a diminuição de erros de predição “sono” quando a classe verdadeira é “vigília” ao utilizar *ensemble* de *patches*, esse fato também foi ilustrado na matriz de confusão correspondente (Figura 4.2d).

Com dados originais de acelerômetro, ocorrem mais erros “sono” para “vigília” no início da noite e poucas vezes ao redor das 6 horas de sono. Os erros de “vigília” para “sono” também ocorrem com mais frequência no início da noite.

Utilizando *ensemble* dos *patches* dos dados de acelerômetro, ocorrem alguns erros “sono” para “vigília” apenas no início da noite, enquanto que os erros “vigília” para “sono” ocorrem com maior frequência por volta das 2 e 6 horas de sono.

4.2.2 Classificação dos Estágios do Sono

Nesta subseção, são apresentadas as acurácias balanceadas, os coeficientes κ e as matrizes de confusão obtidas para a classificação dos estágios do sono.

Acurácias Balanceadas e Coeficiente κ de Cohen

As tabelas 4.3 e 4.4 apresentam, respectivamente, as acurácias balanceadas e coeficientes κ obtidos com cada representação para a classificação dos estágios do sono.

Ao analisar a Tabela 4.3, nota-se que, em contraste com os resultados obtidos para a classificação de sono/vigília, as acurácias balanceadas exibidas com dados de frequência

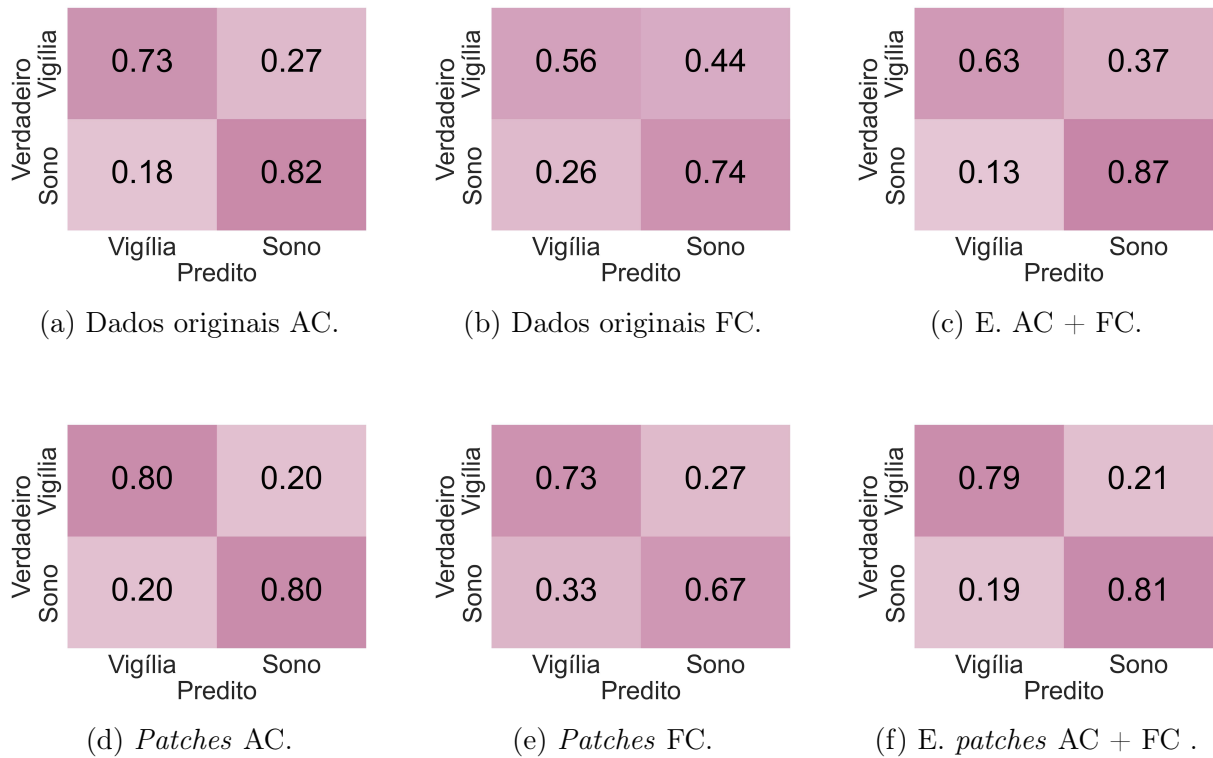


Figura 4.3: MTF: matrizes de confusão geradas com os resultados obtidos por meio dos dados originais e dos melhores resultados de *ensembles* para a classificação de sono/vigília.

cardíaca são maiores que as com dados de acelerômetro em muitos casos (exceto para os casos da representação por Espectrogramas). Para combinar AC + FC, a técnica que apresentou melhores acurácias balanceadas foi o *Ensemble* dos *Patches* AC + *Patches* FC por meio de média ponderada, obtendo acurácia balanceada superior a 60,0% para a maioria das representações (exceto Espectrogramas).

Utilizando Espectrogramas, o *Ensemble* AC + FC por média ponderada com os dados da configuração Original apresentou vantagem em comparação com os dados originais de acelerômetro e frequência cardíaca individuais, enquanto que as outras representações não mostraram melhorias na acurácia balanceada ao aplicar esse tipo de *ensemble*. Por outro lado, o *Ensemble* dos *Patches* AC + *Patches* FC com média ponderada apresentou melhores acurácias balanceadas, se comparado com o *Ensemble* dos *Patches* de dados de acelerômetro, para a maioria das representações. Mostrando, novamente, a importância dos *Patches* de dados de frequência cardíaca para este cenário.

Assim como a classificação de sono/vigília, os melhores resultados de todas as representações envolvem *ensembles* dos *patches*, passando de 62,0% com GAF, de 61,0% com RP, de 60,0% com MTF e de 57,0% com Espectrogramas. Ao comparar as melhores acurácias balanceadas obtidas e a configuração Original de cada representação, é possível observar ganhos de até 4,0 pontos percentuais para RP; 4,5 para GAF; 6,0 para MTF; e 1,4 para Espectrogramas.

Os desvios-padrão obtidos das acurácias balanceadas para a classificação dos estágios do sono são de até 6,0% para RP; até 3,0% para GAF; até 2,0% para MTF; e até 3,0% para Espectrogramas.

Verdadeiro	Vigília	0.78	0.22
	Sono	0.21	0.79
	Predito	Vigília	Sono

(a) Dados originais AC.

Verdadeiro	Vigília	0.62	0.38
	Sono	0.39	0.61
	Predito	Vigília	Sono

(b) Dados originais FC.

Verdadeiro	Vigília	0.76	0.24
	Sono	0.20	0.80
	Predito	Vigília	Sono

(c) E. AC + FC.

Verdadeiro	Vigília	0.83	0.17
	Sono	0.24	0.76
	Predito	Vigília	Sono

(d) *Patches* AC.

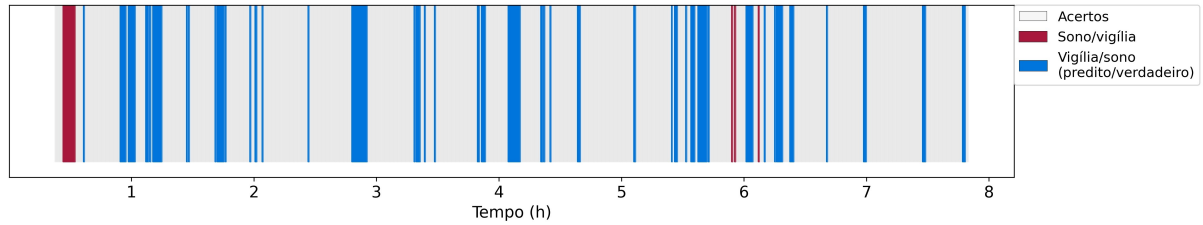
Verdadeiro	Vigília	0.92	0.08
	Sono	0.81	0.19
	Predito	Vigília	Sono

(e) *Patches* FC.

Verdadeiro	Vigília	0.77	0.23
	Sono	0.20	0.80
	Predito	Vigília	Sono

(f) E. *patches* AC + FC.

Figura 4.4: Espectrogramas: matrizes de confusão geradas com os resultados obtidos por meio dos dados originais e dos melhores resultados de *ensembles* para a classificação de sono/vigília.



(a) Dados originais de acelerômetro.

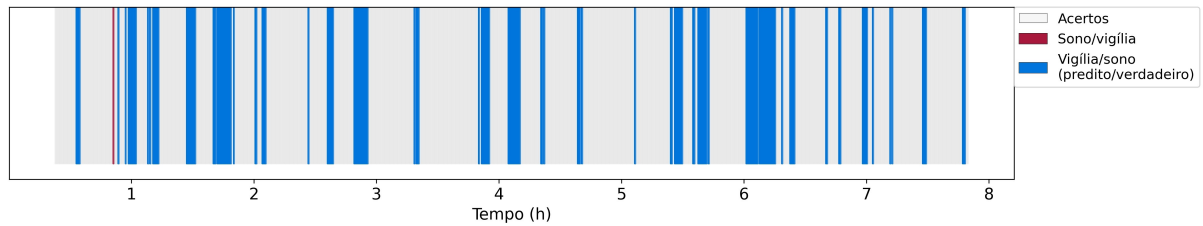
(b) *Ensemble* dos *patches* dos dados de acelerômetro.

Figura 4.5: Classificação de sono/vigília ao longo de uma noite de sono de um indivíduo utilizando a representação GAF.

Observando a Tabela 4.4, nota-se que os coeficientes κ obtidos com dados de acelerômetro, na maioria casos, apresentam-se inferiores a 0,2. Com os Espectrogramas, os κ obtidos com dados de frequência cardíaca não atingem 0,1 para quase todos os resultados. No entanto, ultrapassem 0,2 para RP, GAF e MTF (exceto para *patches* 3 e 7 de RP e *patch* 3 de MTF) mostrando uma concordância regular para a maioria dos casos. Indi-

		RP		GAF		MTF		Espectrogramas	
Rede	Config.	AC	FC	AC	FC	AC	FC	AC	FC
Eff.Net	Original	55.734	57.854	57.684	57.092	55.004	53.354	55.956	40.326
Ensembles AC + FC	M. Simples	46.009		46.135		46.354		52.410	
	M. Ponderada	50.244		50.104		50.366		56.002	
	Deep Features	53.598		53.392		54.818		51.006	
Eff.Net	Patch 1	53.676	54.080	54.956	53.926	51.584	51.476	52.722	35.330
	Patch 2	55.288	55.508	56.224	55.602	53.636	51.652	52.122	35.344
	Patch 3	44.114	45.854	52.060	51.468	45.788	47.920	43.610	35.628
	Patch 4	56.896	55.276	55.504	55.228	55.006	52.792	52.052	34.506
	Patch 5	54.624	54.364	56.178	54.722	53.544	50.616	52.038	36.252
	Patch 6	55.476	55.188	56.452	55.914	52.222	52.048	40.672	35.420
	Patch 7	49.712	50.064	51.236	50.284	46.450	48.034	53.518	35.342
	Patch 8	57.438	56.036	57.878	54.782	55.322	51.646	50.524	35.632
	Patch 9	54.212	53.930	55.880	53.246	52.538	50.924	48.286	35.330
Ensembles dos Patches	M. Simples	59.191	61.866	60.661	62.180	58.375	57.814	55.534	39.501
	M. Ponderada	59.405	61.459	60.164	61.568	58.300	57.340	57.362	39.277
	Rede Simples	49.960	52.435	51.186	51.247	48.293	51.293	48.689	39.331
Ensembles P. AC + P. FC	M. Simples	49.849		49.984		50.953		53.089	
	M. Ponderada	61.481		61.172		60.974		55.346	

Tabela 4.3: Acurácias balanceadas obtidas com cada representação para a classificação dos estágios do sono.

cando, novamente, a relevância desse tipo de dados para a classificação dos estágios do sono. Assim como observado com as acurácias balanceadas, os melhores κ foram obtidos com os experimentos que envolvem *ensembles* de *patches*.

Com exceção dos Espectrogramas, os melhores κ obtidos para a classificação dos estágios do sono foram alcançados com o *Ensemble* dos *Patches* de dados de frequência cardíaca (média ponderada para RP e GAF e rede simples para MTF), sendo $\kappa = 0,38$ para RP (concordância regular), $\kappa = 0,41$ para GAF (concordância moderada) e $\kappa = 0,35$ para MTF (concordância regular).

Matrizes de Confusão

As figuras 4.6, 4.7, 4.8 e 4.9 mostram, respectivamente, as matrizes de confusão geradas para a classificação dos estágios do sono utilizando as representações RP, GAF, MTF e Espectrogramas com os dados originais de acelerômetro (figuras 4.6a, 4.7a, 4.8a e 4.9a) e frequência cardíaca (figuras 4.6b, 4.7b, 4.8b e 4.9b) e melhores *ensembles* acelerômetro + frequência cardíaca (figuras 4.6c, 4.7c, 4.8c e 4.9c). As figuras também incluem as matrizes de confusão dos melhores *ensembles* dos *patches* dos dados de acelerômetro (figuras 4.6d, 4.7d e 4.8d, 4.9d) e dos *patches* dos dados de frequência cardíaca (figuras 4.6e, 4.7e, 4.8e e 4.9e) e melhores *ensembles* dos *patches* dos dados de acelerômetro + *patches* dos dados de frequência cardíaca (figuras 4.6f, 4.7f, 4.8f e 4.9f).

As matrizes de confusão geradas utilizando RP mostram que os dados originais de acelerômetro classificaram mais “vigília” corretamente do que os outros estágios e menos “NREM” corretamente (Figura 4.6a). Enquanto que, utilizando dados originais de frequência cardíaca (Figura 4.6b), a melhor classificação correta é de “NREM” e a pior é “vigília”. Utilizando *patches* de dados de acelerômetro, é possível observar que a classificação correta de “vigília” melhorou significativamente embora a classificação de “NREM” como “REM” aumentou (Figura 4.6d). Na matriz de confusão do *ensemble* que utiliza dados originais (Figura 4.6c) a confusão entre “NREM” e “REM” é maior, se comparada

Rede	Config.	RP		GAF		MTF		Espectrogramas	
		AC	FC	AC	FC	AC	FC	AC	FC
Eff.Net	Original	0.171	0.322	0.241	0.326	0.195	0.279	0.203	0.085
Ensembles AC + FC	M. Simples	0.131		0.134		0.151		0.191	
	M. Ponderada	0.172		0.172		0.173		0.189	
	Deep Features	0.216		0.254		0.215		0.112	
Eff.Net	Patch 1	0.150	0.252	0.183	0.277	0.167	0.237	0.153	0.051
	Patch 2	0.166	0.283	0.197	0.269	0.152	0.253	0.184	0.084
	Patch 3	0.087	0.157	0.142	0.232	0.106	0.175	0.103	0.092
	Patch 4	0.189	0.284	0.132	0.322	0.170	0.244	0.142	0.083
	Patch 5	0.175	0.261	0.171	0.272	0.141	0.205	0.162	0.081
	Patch 6	0.190	0.292	0.202	0.308	0.149	0.229	0.071	0.073
	Patch 7	0.149	0.189	0.140	0.252	0.092	0.200	0.138	0.081
	Patch 8	0.205	0.274	0.187	0.300	0.154	0.213	0.154	0.080
	Patch 9	0.167	0.254	0.175	0.272	0.181	0.207	0.154	0.086
Ensembles dos Patches	M. Simples	0.214	0.375	0.225	0.413	0.211	0.331	0.208	0.091
	M. Ponderada	0.207	0.380	0.230	0.414	0.205	0.335	0.211	0.101
	Rede Simples	0.192	0.370	0.209	0.352	0.171	0.347	0.176	0.102
Ensembles P. AC + P. FC	M. Simples	0.165		0.180		0.192		0.199	
	M. Ponderada	0.284		0.272		0.283		0.190	

Tabela 4.4: Coeficientes κ de Cohen obtidos com cada representação para a classificação dos estágios do sono.

com as outras matrizes que utilizam esse tipo de dados.

É importante destacar que as matrizes de confusão que envolvem dados de acelerômetro, incluindo as matrizes de confusão dos *ensembles*, classificaram mais “vigília” corretamente, enquanto que as matrizes de confusão geradas apenas com dados de frequência cardíaca mostraram classificar mais “NREM” e “REM” corretamente (Figura 4.6b e 4.6e). Além disso, nota-se que com dados de acelerômetro (originais e *patches*), a confusão mais frequente é classificar “NREM” como “REM”. Com dados de frequência cardíaca, a classificação incorreta mais frequente é “vigília” como “REM”.

Com a representação GAF, as matrizes de confusão obtidas para a classificação dos estágios do sono indicam que, utilizando dados originais de acelerômetro (Figura 4.7a), a classe mais corretamente classificada é “vigília” e a menos, “REM”. Já com dados originais de frequência cardíaca (Figura 4.7b), a classe mais corretamente classificada é “NREM” e a menos, “vigília”. Utilizando *patches*, nota-se que, com dados de acelerômetro (Figura 4.7d), as classificações corretas de “vigília” e de “REM” aumentam enquanto que a de “NREM” diminuem, se comparadas com a matriz de confusão obtida com os dados originais de acelerômetro.

A matriz de confusão com *patches* de frequência cardíaca mostra um aumento nas classificações corretas de “NREM” e “REM” (Figura 4.2e). Analisando as confusões mais comuns, nota-se que, com dados de acelerômetro, as classificações incorretas mais frequentes são “NREM” como “REM” e “REM” como “NREM”. Enquanto que, com dados de frequência cardíaca, a confusão mais frequente é classificar “vigília” como “REM”.

As matrizes de confusão correspondentes à representação MTF mostram que, com dados de acelerômetro (figuras 4.8a e 4.8d) e os dois tipos de *ensemble* (figuras 4.8c e 4.8f), a classificação correta mais frequente, assim como ocorre nas outras representações, é a da “vigília”. A confusão mais comum, utilizando dados de acelerômetro e os *ensembles*, é classificar “NREM” como “REM”. Analisando as matrizes de confusão dos dados

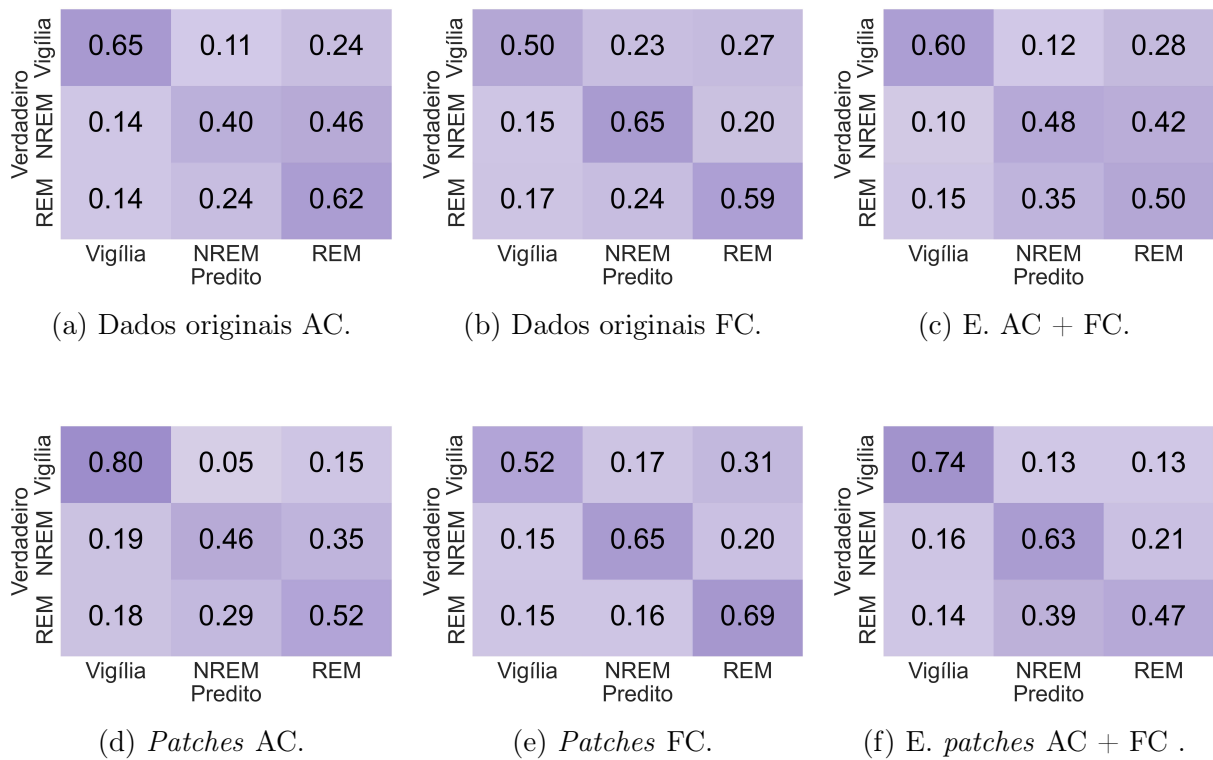


Figura 4.6: RP: matrizes de confusão geradas com os resultados obtidos por meio dos dados originais e dos melhores resultados de *ensembles* para a classificação dos estágios do sono.

de frequência cardíaca, é possível observar que a quantidade de classificações incorretas de “vigília” como “REM” apresentada com dados originais (Figura 4.8b) diminui com a utilização de *patches* (Figura 4.8e), tornando a classificação mais balanceada.

Com a representação Espectrogramas, as matrizes de confusão indicam principalmente que, com dados de frequência cardíaca, tanto na configuração original quanto em *patches* (figuras 4.9b e 4.9e, respectivamente) apresentam poucas classificações para “NREM” e “REM”, superestimando, assim, as “vigílias”. Com os dados de acelerômetro, em ambas as configurações (figuras 4.9a e 4.9d), a classe mais corretamente classificada foi “vigília” e a menos, “REM”. Ao utilizar a configuração em *patches*, a quantidade de classificação correta “REM” aumentou, assim como as predições de “REM” incorretas para as outras classes também aumentaram, mas as predições incorretas de “vigília” diminuíram. Ambos os tipos de *ensemble* (figuras 4.9c e 4.9f) apresentaram melhor classificação correta para “vigília” e maior confusão em “NREM” quando a classe verdadeira é “REM”.

As figuras 4.10a e 4.10b ilustram as predições de uma noite de sono de um mesmo indivíduo para a classificação dos estágios do sono utilizando dados originais de frequência cardíaca e *ensemble* dos *patches* de frequência cardíaca. Observando essas figuras, é possível comparar a melhor noite de sono obtida para a classificação dos estágios do sono (*ensemble* dos *patches* de frequência cardíaca), que foi obtida com GAF, com a noite de sono correspondente utilizando dados originais.

Pode-se observar que, com os dados originais, os erros mais frequentes são predições de “vigília” para “NREM” e predições de “REM” para “NREM”. Nota-se também que

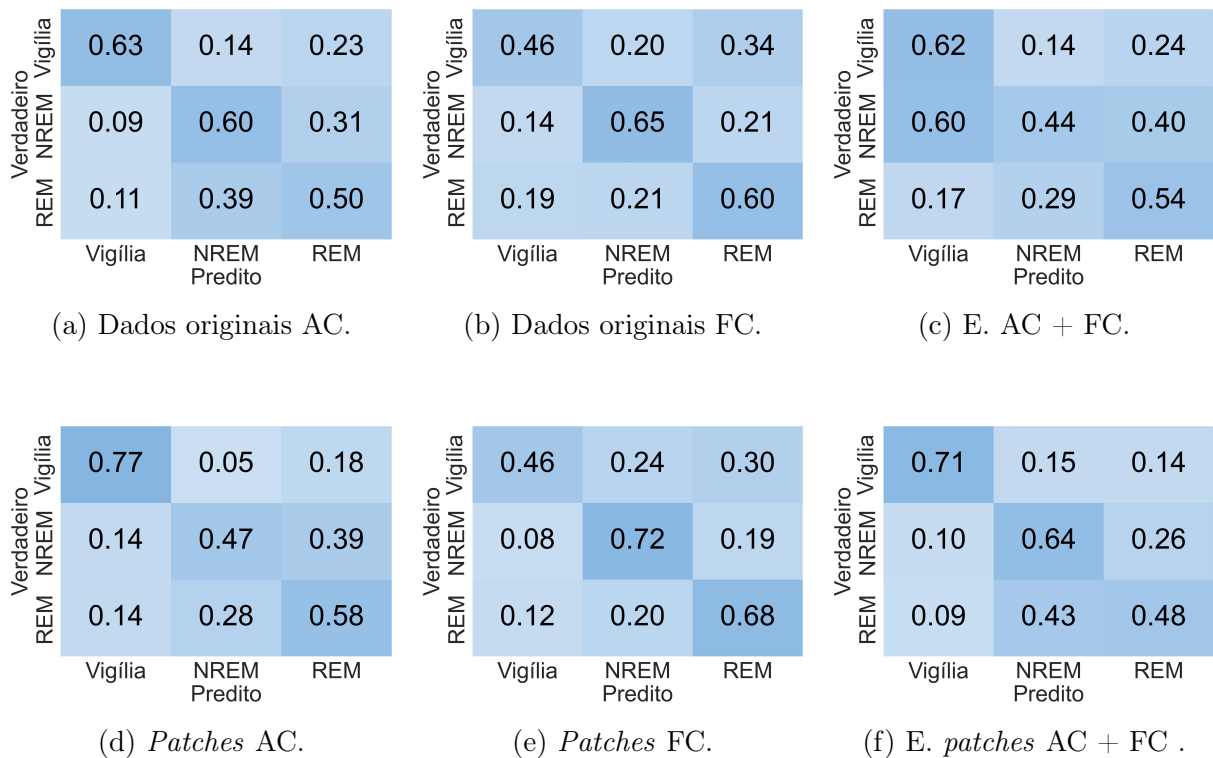


Figura 4.7: GAF: matrizes de confusão geradas com os resultados obtidos por meio dos dados originais e dos melhores resultados de *ensembles* para a classificação dos estágios do sono.

as predições de “vigília” são superestimadas para a classe “REM”. Analisando os erros relacionados às predições de “NREM”, o modelo confundiu mais essa classe com “REM” do que com “vigília”. Por fim, as classificações erradas menos comuns são as de “REM” para “vigília”.

Ao analisar os resultados do *ensemble* dos *patches* dos dados de frequência cardíaca, observa-se uma redução significativa nas predições incorretas de “vigília” para “NREM” ao longo do tempo. As predições equivocadas de “NREM” para “REM” também apresentaram uma diminuição. No entanto, os erros de classificação de “REM” para “vigília” no início da noite mantiveram-se constantes. Adicionalmente, destaca-se a ocorrência de predições incorretas de “NREM” para “vigília” próxima a marca de 8 horas do teste.

4.2.3 Outras Representações

Para comparar a abordagem da representação visual das séries temporais com as abordagens tipicamente utilizadas em trabalhos relacionados, obtivemos resultados utilizando os dados brutos como entrada para redes baseadas em CNN 1D, LSTM e GRU. Além disso, comparamos os resultados obtidos utilizando extração de *features* dos dados brutos como entrada para *Random Forest* e Regressão Logística.

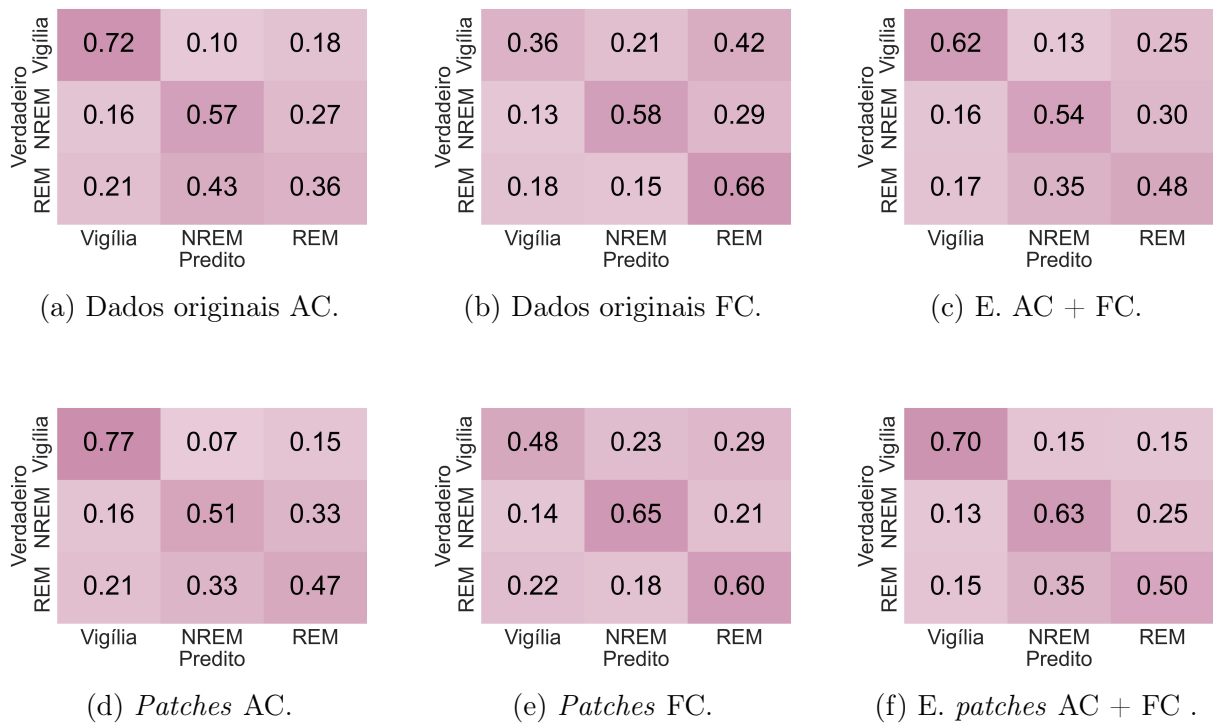


Figura 4.8: MTF: matrizes de confusão geradas com os resultados obtidos por meio dos dados originais e dos melhores resultados de *ensembles*.

Dados Brutos

Os modelos baseados em CNN 1D, LSTM e GRU foram implementados com base na arquitetura descrita por Mekruksavanich e Jitpattanakul [54] para HAR, sendo a primeira camada uma convolução 1D; a segunda, *MaxPooling* 1D; a terceira camada é outra convolução 1D, seguida de outra *MaxPooling*; a penúltima camada é LSTM, e a última, uma camada densa. Para comparar com modelos baseados em CNN 1D e GRU, a penúltima camada dessa arquitetura é substituída por uma CNN 1D e por uma GRU, respectivamente.

Assim como realizado para as representações visuais, as redes receberam como entrada dados de acelerômetro e dados de frequência cardíaca e, com os modelos treinados, foram realizados os *ensembles* utilizando média simples, média ponderada e *deep features* para combinar os resultados obtidos com dados de acelerômetro + dados de frequência cardíaca.

Os resultados obtidos para a classificação de sono/vigília e para a dos estágios do sono, com cada configuração e com cada rede, são apresentados nas tabelas 4.5 e 4.6.

		CNN		LSTM		GRU	
Config.		AC	FC	AC	FC	AC	FC
Original		71.596	52.000	71.554	56.290	72.542	63.884
Ensembles AC + FC	M. Simples	71.300		72.784		75.620	
	M. Ponderada	71.940		72.835		74.959	
	Deep Features	70.052		71.910		61.985	

Tabela 4.5: Acurácias balanceadas obtidas com dados brutos para a classificação de sono/vigília.

Verdadeiro	Vigília	0.71	0.15	0.14
	NREM	0.14	0.64	0.23
	REM	0.16	0.51	0.33
		Vigília	NREM Predito	REM

(a) Dados originais AC.

Verdadeiro	Vigília	0.85	0.07	0.08
	NREM	0.72	0.21	0.07
	REM	0.80	0.06	0.15
		Vigília	NREM Predito	REM

(b) Dados originais FC.

Verdadeiro	Vigília	0.72	0.15	0.13
	NREM	0.15	0.64	0.21
	REM	0.17	0.51	0.32
		Vigília	NREM Predito	REM

(c) E. AC + FC.

Verdadeiro	Vigília	0.72	0.10	0.18
	NREM	0.13	0.56	0.31
	REM	0.15	0.41	0.44
		Vigília	NREM Predito	REM

(d) *Patches* AC.

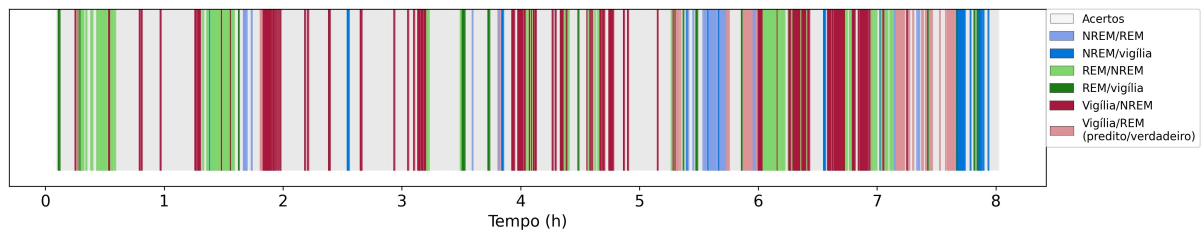
Verdadeiro	Vigília	0.87	0.12	0.01
	NREM	0.75	0.24	0.01
	REM	0.79	0.13	0.08
		Vigília	NREM Predito	REM

(e) *Patches* FC.

Verdadeiro	Vigília	0.75	0.12	0.13
	NREM	0.15	0.65	0.20
	REM	0.18	0.55	0.27
		Vigília	NREM Predito	REM

(f) E. *patches* AC + FC .

Figura 4.9: Espectrogramas: matrizes de confusão geradas com os resultados obtidos por meio dos dados originais e dos melhores resultados de *ensembles*.



(a) Dados originais de frequência cardíaca.

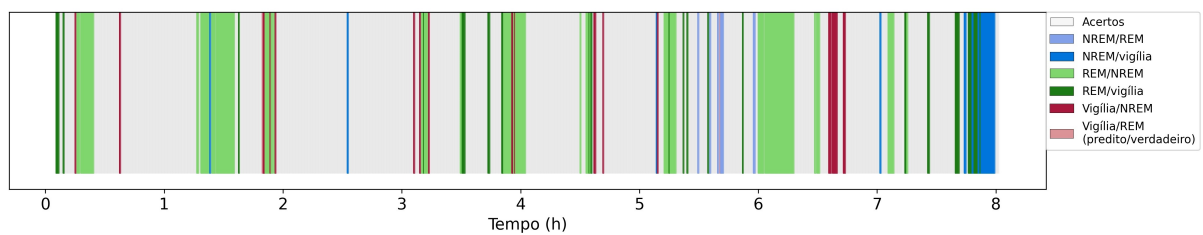
(b) *Ensemble* dos *patches* dos dados de frequência cardíaca.

Figura 4.10: Classificação de estágios do sono ao longo de uma noite de sono de um indivíduo utilizando a representação GAF.

Para a classificação de sono/vigília, nota-se que, assim como as representações visuais, os dados de acelerômetro apresentam maiores acurácias balanceadas comparadas com as obtidas por meio dos dados de frequência cardíaca. Os melhores resultados envolvem *ensemble* dos dois tipos de dados para todas as redes. A maior acurácia balanceada supera 75,0% com a rede GRU e o *ensemble* média simples. Os melhores resultados apresentados apenas com dados de acelerômetro e apenas frequência cardíaca também foram alcançados

com GRU, superando 72,0% e 63,0%, respectivamente.

Config.		CNN		LSTM		GRU	
		AC	FC	AC	FC	AC	FC
Original		46.656	46.700	47.874	57.546	46.844	56.030
Ensembles AC + FC	M. Simples	45.660		48.715		45.078	
	M. Ponderada	46.267		48.419		45.040	
	Deep Features	45.521		51.754		36.770	

Tabela 4.6: Acurácias balanceadas obtidas com dados brutos para a classificação dos estágios do sono.

Para a classificação dos estágios do sono, é possível observar, que assim como as representações visuais, as melhores acurácias balanceadas foram obtidas com dados de frequência cardíaca, ultrapassando 57,0% utilizando a rede LSTM. Os resultados obtidos com *ensembles* não apresentaram melhorias para as redes CNN 1D e GRU, enquanto que os *ensembles* realizados com a rede LSTM melhoram os resultados apresentados apenas com dados de acelerômetro.

Extração de *Features*

A extração de *features* dos dados de acelerômetro e de frequência cardíaca, assim como o *ensemble* realizado, foram baseados no trabalho de Walch *et al.* [81], que utiliza contadores de atividade como *feature* extraída dos dados de acelerômetro e desvios padrão locais extraídos dos dados de frequência cardíaca. O *ensemble* consistiu em utilizar a *feature* dos dados de acelerômetro e a *feature* dos dados de frequência cardíaca como entradas para os modelos. A Tabela 4.7 apresenta as acurácias balanceadas obtidas para a classificação de sono/vigília e para a classificação dos estágios do sono.

Config.	Classificação de Sono/Vigília				Classificação dos Estágios do Sono			
	Random Forest		Regressão Logística		Random Forest		Regressão Logística	
	AC	FC	AC	FC	AC	FC	AC	FC
Original	76.553	64.807	73.854	63.592	51.550	48.732	44.501	40.502
Ensembles AC + FC	76.654		73.982		58.510		46.510	

Tabela 4.7: Acurácias balanceadas obtidas com extração de *features* para as classificações de sono/vigília e de estágios do sono.

É possível observar que, assim como a representação visual, os dados de acelerômetro, comparados aos dados de frequência cardíaca, apresentam as melhores acurácias balanceadas para a classificação de sono/vigília, alcançando 76,0% com *Random Forest*. Para a classificação dos estágios do sono, o melhor resultado ultrapassa 58,0% utilizando *Random Forest* e o *ensemble* das *features*.

Comparação entre as Representações

A Tabela 4.8 apresenta a comparação das acurácias balanceadas obtidas com cada representação de dados utilizando acelerômetro, frequência cardíaca e *ensemble* para a classificação de sono/vigília. É possível observar que a representação visual atinge os melhores

	AC			FC			Ensemble AC + FC		
	Imagens	D. Brutos	Features	Imagens	D. Brutos	Features	Imagens	D. Brutos	Features
Ac. bal.	82.358	72.542	76.553	73.712	63.884	64.807	81.444	75.620	76.654
Config.	GAF Patches	GRU	RF	GAF Patches	GRU	RF	GAF Patches	GRU	RF

Tabela 4.8: Comparação das melhores acurácias balanceadas obtidas com as diferentes representações de dados para a classificação de sono/vigília.

	AC			FC			Ensemble AC + FC		
	Imagens	D. Brutos	Features	Imagens	D. Brutos	Features	Imagens	D. Brutos	Features
Ac. bal.	60.661	47.874	51.550	62.180	57.546	48.732	61.481	51.754	58.510
Config.	GAF Patches	LSTM	RF	GAF Patches	LSTM	RF	RP Patches	LSTM	RF

Tabela 4.9: Comparação das melhores acurácias balanceadas obtidas com as diferentes representações de dados para a classificação dos estágios do sono.

resultados em todos os casos, tendo uma vantagem de até 5,8 pontos percentuais com dados de acelerômetro; 8,9 com dados de frequência cardíaca e 4,8 com *ensemble*.

A Tabela 4.9 apresenta as melhores acurácias balanceadas obtidas com cada tipo de representação utilizando dados de acelerômetro, frequência cardíaca e *ensemble* e realiza uma comparação para a classificação dos estágios do sono. Assim como observado na classificação de sono/vigília, a representação visual se destaca por apresentar os melhores resultados em todos os casos. A diferença de acurácia balanceada utilizando imagens chega a pelo menos 9,1 pontos percentuais com dados de acelerômetro; 4,6 com dados de frequência cardíaca; e 3,0 com *ensemble*.

Comparando os coeficiente κ obtidos com as representações visuais e com os relatados no trabalho de referência para a base de dados [81], observa-se:

- Para a classificação de sono/vigília, Walch *et al.* [81] relataram: $\kappa = 0,36$ (AC), $\kappa = 0,19$ (FC) e $\kappa = 0,37$ (*ensemble* AC + FC). Todos os melhores κ foram obtidos com Regressão Logística. Utilizando as representações visuais, os seguintes melhores κ obtidos foram: $\kappa = 0,37$ (AC), $\kappa = 0,28$ (FC) e $\kappa = 0,42$ (*ensemble* AC + FC).
- Para a classificação dos estágios do sono, os autores relataram: $\kappa = 0,08$ (AC), $\kappa = 0,16$ (FC) e $\kappa = 0,29$ (*ensemble* AC + FC). Utilizando representações visuais, os melhores κ foram: $\kappa = 0,24$ (AC) $\kappa = 0,41$ (FC) e $\kappa = 0,28$ (*ensemble* AC + FC).

É importante destacar que as acurácias balanceadas não podem ser comparadas, tendo em vista que os autores apresentam apenas os valores de acurácia simples.

Pode-se afirmar que, embora a utilização de imagens para representação de dados mostre uma redução na vantagem ao empregar técnicas de *ensemble* comparativamente a outras formas de representação, ela destaca-se ao apresentar ganhos significativos nos dados isolados de acelerômetro (levando a obtenção do melhor resultado para classificação de sono e vigília) e de frequência cardíaca (levando a obtenção do melhor resultado para classificação de estágios do sono).

Capítulo 5

Conclusões e Trabalhos Futuros

Com a realização dos experimentos apresentados, é possível concluir que os dados de acelerômetro são mais eficazes que os de frequência cardíaca na classificação de sono/vigília, demonstrando melhores resultados em todas as representações analisadas. Embora os *ensembles* que combinam dados de acelerômetro e frequência cardíaca não tenham apresentado um aumento de acurácia balanceada, os *Ensembles* dos *Patches* individuais, tanto para dados de acelerômetro quanto de frequência cardíaca, mostraram um ganho significativo de acurácia em comparação com os resultados das imagens Originais. Especificamente, a representação GAF com dados de acelerômetro e *Ensemble* dos *Patches* alcançou a maior acurácia balanceada para a classificação de sono/vigília. Além disso, os coeficientes κ reforçam a superioridade dos dados de acelerômetro, indicando uma concordância regular para a maioria das configurações, em contraste com os resultados mais limitados obtidos com os dados de frequência cardíaca. Portanto, os dados de acelerômetro, especialmente quando utilizados em *Ensembles* de *Patches* individuais, provaram ser adequados para a classificação de sono/vigília.

Parte dos resultados obtidos para a classificação de sono/vigília utilizando dados de acelerômetro e a representação de Gráficos de Recorrência (RP) foram explorados no artigo “*Sleep-Wake Classification using Recurrence Plots from Smartwatch Accelerometer Data*” (Rebeca Padovani Ederli, Didier Vega-Oliveros, Aurea Soriano-Vargas, Anderson Rocha e Zanoni Dias) [21], apresentado no *9th IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, realizado em Recife (PE), em outubro de 2023.

A análise dos resultados relacionados à classificação dos estágios do sono revela um cenário distinto da classificação de sono/vigília, no qual os dados de frequência cardíaca demonstram ser mais eficazes do que os dados de acelerômetro em muitos aspectos, alcançando acurácias balanceadas superiores. Notavelmente, os *Ensemble* dos *Patches* que combinam ambos os tipos de dados (AC + FC) mostraram melhorias em relação aos resultados obtidos apenas com dados de acelerômetro, especialmente, indicando a eficácia dessa abordagem na melhora das acurácias balanceadas para as diferentes representações. Os melhores resultados foram observados com os *Ensembles* dos *Patches*, que não só superaram os dados Originais de acelerômetro e frequência cardíaca, mas também proporcionaram ganhos substanciais de acurácia balanceada, reforçando a importância dos *patches*.

Além disso, os coeficientes κ foram predominantemente melhores com os dados de

frequência cardíaca, superando a marca de 0,2 em várias configurações, exceto em casos específicos. Isso sugere uma concordância regular na classificação dos estágios do sono, destacando novamente a relevância dos dados de frequência cardíaca neste contexto.

Portanto, em ambos os cenários, a abordagem dos *patches* foram importantes. E, enquanto os dados de acelerômetro predominam na classificação de sono/vigília, para a classificação mais detalhada dos estágios do sono, os dados de frequência cardíaca demonstram ser mais efetivos. A importância dos dados de acelerômetro para a classificação de sono/vigília e a importância dos dados de frequência cardíaca para a classificação dos estágios do sono também foram observadas em trabalhos relacionados [23, 81].

Representar séries temporais com Espectrogramas levaram a comportamentos diferentes quando comparados com as outras representações visuais, isso pode ter ocorrido devido à natureza intrínseca de cada método de representação. Espectrogramas enfatizam a frequência ou o espectro de potências das séries ao longo do tempo, capturando assim a variação de frequência em diferentes janelas temporais. Isso os torna particularmente úteis para análises envolvendo principalmente sons ou sinais onde a frequência é um componente crucial. Por outro lado, RP visualizam a periodicidade e as dinâmicas temporais das séries, destacando os momentos em que a série temporal visita o mesmo ponto ou estado no espaço de fase. GAF transforma a série temporal em uma imagem codificando as relações angulares entre os pontos, o que permite capturar informações de tempo e amplitude simultaneamente, preservando assim a correlação temporal. MTF, entretanto, representa a série temporal como uma matriz que descreve as probabilidades de transição entre diferentes estados, capturando a dinâmica de transição de estados ao longo do tempo. Visualmente é possível notar que as representações RP, GAF e MTF são parecidas, enquanto os Espectrogramas são diferenciados.

Comparando as representações visuais com dados brutos e extração de *features*, é possível concluir a eficácia dessa nova abordagem para representar dados de sono. Isso ocorre pois as imagens permitem que o modelo *Deep Learning* explore esses padrões de uma forma que não é possível com dados brutos linearmente considerados. Além disso, ao contrário dos métodos tradicionais, como *Random Forest* e Regressão Logística, que exigem conhecimento específico para a seleção e extração manual de *features* relevantes, as técnicas de *Deep Learning* podem automaticamente identificar e extrair características significativas das imagens geradas a partir das séries temporais dos dados de acelerômetro e frequência cardíaca. No entanto, a representação visual também podem ser menos eficientes em termos de recursos computacionais e tempo de treinamento, devido aos requisitos computacionais elevados e tempo prolongado de treinamento.

Assim, as questões de pesquisa levantadas no Capítulo 1 podem ser respondidas da seguinte forma:

- Como combinar as informações dos dados de diferentes sensores de *smartwatches* para obter um método competitivo de classificação dos estágios do sono?

Comparando os dois tipos de *ensembles* que combinam os resultados dos dados de acelerômetro + dados de frequência cardíaca, o *ensemble* utilizando *patches* permitiu combinar os resultados de modo a obter melhor acurácia balanceada para a maioria das representações (exceto com Espectrogramas para a classificação dos estágios do

sono).

- Quais são as representações de imagens mais adequadas para classificar estágios do sono?

Para a classificação de sono/vigília e dos estágios do sono, a representação GAF demonstrou os melhores resultados, destacando-se como a mais eficaz. No entanto, no contexto da classificação de sono/vigília, as representações RP, MTF e Espectrogramas também exibiram desempenhos superiores em comparação às abordagens baseadas em dados brutos e na extração de *features*, indicando que também são adequadas para esta classificação. Por outro lado, para classificar os estágios do sono, a representação por Espectrogramas não superou os resultados alcançados com o uso de dados brutos ou com técnicas de extração de *features*, mostrando que não é a mais adequada para essa tarefa.

- Quais técnicas de aprendizado de máquina podem tornar a classificação dos estágios do sono mais precisa?

A implementação de *ensembles* de *patches* individuais para dados de acelerômetro e de frequência cardíaca assim como o *ensemble* dos *patches* dos dados de acelerômetro + *patches* dos dados de frequência cardíaca tornaram a classificação de sono/vigília e dos estágios do sono mais precisas para a maioria das representações (exceto com Espectrogramas para a classificação dos estágios do sono).

Para trabalhos futuros, é evidente que a utilização de *patches* apresentou resultados promissores tanto na classificação de sono/vigília quanto nos estágios do sono, indicando um caminho interessante para a exploração de redes baseadas em *Transformers*, como a *Vision Transformer* (ViT) [18]. Tal abordagem pode potencializar ainda mais a capacidade de modelagem dos dados ao capturar relações complexas dentro das imagens, eventualmente, eliminando a necessidade de dividir as imagens originais em *patches*. Além disso, a investigação de métodos de normalização para os dados brutos de acelerômetro e frequência cardíaca surge como uma medida crucial para aprimorar a qualidade dos dados antes da aplicação de modelos de aprendizado de máquina, possibilitando uma representação mais estável e comparável dos sinais.

A exploração de técnicas de filtragem, tais como filtros passa-faixa e passa-baixa, nos dados brutos dos sensores, para a remoção de ruídos podem melhorar a sinalização de características relevantes para as classificações. Técnicas de pós-processamento, como correção de classificações baseadas em regras de contexto, podem ser investigadas para aumentar a precisão e a confiabilidade das classificações finais, melhorando assim a utilidade prática dos modelos em aplicações de monitoramento de sono.

Uma limitação importante surge na implementação dessas tecnologias em dispositivos vestíveis, que possuem capacidades computacionais mais restritas. A execução de modelos de *Deep Learning*, somada à geração e ao processamento das imagens, exige elevados recursos computacionais, tornando sua aplicação direta em dispositivos com poder computacional limitado um desafio significativo. Essa questão destaca a necessidade de pesquisas futuras focadas na otimização de representações visuais para torná-las mais

eficientes do ponto de vista computacional. Investigar métodos que reduzam o custo computacional sem comprometer significativamente a precisão dos modelos pode viabilizar a implementação dessa abordagem em dispositivos vestíveis.

Por fim, é indispensável empregar técnicas de explicabilidade, como métodos de visualização e localização de ativação relevante, a fim de tornar os modelos de *Deep Learning* mais compreensíveis, permitindo que especialistas identifiquem partes importantes nos dados, validem previsões e compreendam padrões.

Referências Bibliográficas

- [1] L. C. S. Afonso, G. Rosa, C. R. Pereira, S. Weber, C. Hook, V. H. C. Albuquerque, and J. P. Papa. A recurrence plot-based approach for Parkinson’s disease identification. *Future Generation Computer Systems*, 94:282–292, 2019.
- [2] 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.
- [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] O. Banos, R. Garcia, J. A. Holgado-Terriza, M. Damas, H. Pomares, I. Rojas, A. Saez, and C. Villalonga. mHealthDroid: a novel framework for agile development of mobile health applications. In *Ambient Assisted Living and Daily Activities*, pages 91–98. Springer, 2014.
- [6] 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, 2017.
- [7] 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.
- [8] 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.
- [9] M. B. Bin Heyat, F. Akhtar, M. Ansari, A. Khan, F. Alkahtani, H. Khan, and D. Lai. Progress in detection of insomnia sleep disorder: a comprehensive review. *Current Drug Targets*, 22(6):672–684, 2021.

- [10] 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.
- [11] A. Chakrabarty, O. A. Pandit, and U. Garain. Context sensitive lemmatization using two successive bidirectional gated recurrent networks. In *Annual Meeting of the Association for Computational Linguistics*, pages 1481–1491, 2017.
- [12] J.-P. Chaput. Sleep patterns, diet quality and energy balance. *Physiology & Behavior*, 134:86–91, 2014.
- [13] 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.
- [14] 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.
- [15] J. Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46, 1960.
- [16] 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.
- [17] F. de Arriba-Pérez, M. Caeiro-Rodríguez, and J. M. Santos-Gago. How do you sleep? Using off the shelf wrist wearables to estimate sleep quality, sleepiness level, chronotype and sleep regularity indicators. *Journal of Ambient Intelligence and Humanized Computing*, 9:897–917, 2018.
- [18] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
- [19] 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^o 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.

- [20] J.-P. Eckmann, S. O. Kamphorst, and D. Ruelle. Recurrence plots of dynamical systems. *World Scientific Series on Nonlinear Science Series A*, 16:441–446, 1995.
- [21] R. P. Ederli, D. Vega-Oliveros, A. Soriano-Vargas, A. Rocha, and Z. Dias. Sleep-wake classification using recurrence plots from smartwatch accelerometer data. In *IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, pages 1–6. IEEE, 2023.
- [22] P. Esling and C. Agon. Time-series data mining. *ACM Computing Surveys*, 45, 12 2012.
- [23] S. Fallmann and L. Chen. Computational sleep behavior analysis: A survey. *IEEE Access*, 7:142421–142440, 2019.
- [24] F. A. Faria, J. Almeida, B. Alberton, L. P. C. Morellato, and R. d. S. Torres. Fusion of time series representations for plant recognition in phenology studies. *Pattern Recognition Letters*, 83:205–214, 2016.
- [25] 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.
- [26] 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.
- [27] 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.
- [28] M. A. Ganaie, M. Hu, A. Malik, M. Tanveer, and P. Suganthan. Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*, 115:105151, 2022.
- [29] F. Gu, M.-H. Chung, M. Chignell, S. Valaee, B. Zhou, and X. Liu. A survey on deep learning for human activity recognition. *ACM Computing Surveys*, 54(8):1–34, 2021.
- [30] 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.
- [31] 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.
- [32] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

- [33] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.
- [34] V. Ibáñez, J. Silva, and O. Cauli. A survey on sleep assessment methods. *PeerJ*, 6:e4849, 2018.
- [35] C. Iber. The AASM manual for the scoring of sleep and associated events: Rules. *Terminology and Technical Specification*, 2007.
- [36] M. Iman, H. R. Arabnia, and K. Rasheed. A review of deep transfer learning and recent advancements. *Technologies*, 11(2):40, 2023.
- [37] H. A. Imran. Ultanet: An antithesis neural network for recognizing human activity using inertial sensors signals. *IEEE Sensors Letters*, 6, 1 2022.
- [38] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, 33(4):917–963, 2019.
- [39] J. S. Iwanski and E. Bradley. Recurrence plots of experimental data: To embed or not to embed? *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 8(4):861–871, 1998.
- [40] 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.
- [41] H. Kinnunen, A. Rantanen, T. Kenttö, and H. Koskimäki. 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(4):04NT01, 2020.
- [42] 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.
- [43] G. Klish, 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 (EMBC)*, 20(3):51–57, 2001.
- [44] 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.
- [45] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

- [46] C. A. Kushida, M. R. Littner, T. Morgenthaler, C. A. Alessi, D. Bailey, J. Coleman Jr, L. Friedman, M. Hirshkowitz, S. Kapen, M. Kramer, T. Lee-Chiong, D. L. Loube, J. Owens, J. P. Pancer, and M. Wise. Practice parameters for the indications for polysomnography and related procedures: an update for 2005. *Sleep*, 28(4):499–523, 2005.
- [47] 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.
- [48] C. Lam, C. Yu, L. Huang, and D. Rubin. Retinal lesion detection with deep learning using image patches. *Investigative Ophthalmology & Visual Science*, 59(1):590–596, 2018.
- [49] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.
- [50] J. Li, M. V. Vitiello, and N. S. Gooneratne. Sleep in normal aging. *Sleep Medicine Clinics*, 13(1):1–11, 2018.
- [51] K. L. Lichstein, K. C. Stone, J. Donaldson, S. D. Nau, J. P. Soeffing, D. Murray, K. W. Lester, and R. N. Aguillard. Actigraphy validation with insomnia. *Sleep*, 29(2):232–239, 2006.
- [52] J. Lu and K.-Y. Tong. Robust single accelerometer-based activity recognition using modified recurrence plot. *IEEE Sensors Journal*, 19(15):6317–6324, 2019.
- [53] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths. Recurrence plots for the analysis of complex systems. *Physics Reports*, 438(5-6):237–329, 2007.
- [54] S. Mekruksavanich and A. Jitpattanakul. Biometric user identification based on human activity recognition using wearable sensors: An experiment using deep learning models. *Electronics*, 10(3):308, 2021.
- [55] T. Morgenthaler, C. Alessi, L. Friedman, J. Owens, V. Kapur, B. Boehlecke, T. Brown, A. Chesson Jr, J. Coleman, T. Lee-Chiong, J. Pancer, and T. J. Swick. Practice parameters for the use of actigraphy in the assessment of sleep and sleep disorders: an update for 2007. *Sleep*, 30(4):519–529, 2007.
- [56] 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.
- [57] S. Nightingale, J. Orgill, I. Ebrahim, S. De Lacy, S. Agrawal, and A. Williams. The association between narcolepsy and REM behavior disorder (RBD). *Sleep Medicine*, 6(3):253–258, 2005.
- [58] M. Nollet, W. Wisden, and N. P. Franks. Sleep deprivation and stress: a reciprocal relationship. *Interface Focus*, 10(3):20190092, 2020.

- [59] F. J. Ordóñez and D. Roggen. Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. *Sensors*, 16(1):115, 2016.
- [60] D. Ortega Anderez, A. Lotfi, and A. Pourabdollah. A deep learning based wearable system for food and drink intake recognition. *Journal of Ambient Intelligence and Humanized Computing*, 12:9435–9447, 2021.
- [61] J. Paquet, A. Kawinska, and J. Carrier. Wake detection capacity of actigraphy during sleep. *Sleep*, 30(10):1362–1369, 2007.
- [62] 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.
- [63] 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.
- [64] Z. Qin, Y. Zhang, S. Meng, Z. Qin, and K.-K. R. Choo. Imaging and fusing time series for wearable sensor-based human activity recognition. *Information Fusion*, 53:80–87, 2020.
- [65] 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.
- [66] A. Rechtschaffen. A manual for standardized terminology, techniques and scoring system for sleep stages in human subjects. *Brain Information Service*, 1968.
- [67] 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.
- [68] K. Roy, D. Banik, D. Bhattacharjee, and M. Nasipuri. Patch-based system for classification of breast histology images using deep learning. *Computerized Medical Imaging and Graphics*, 71:90–103, 2019.
- [69] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, 115:211–252, 2015.
- [70] A. Sano and R. W. Picard. Comparison of sleep-wake classification using electroencephalogram and wrist-worn multi-modal sensor data. In *International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 930–933. IEEE, 2014.
- [71] M. J. Sateia. International classification of sleep disorders. *Chest*, 146(5):1387–1394, 2014.

- [72] S. Schutte-Rodin, M. C. Deak, S. Khosla, C. A. Goldstein, M. Yurcheshen, A. Chiang, D. Gault, J. Kern, D. O’Hearn, S. Ryals, N. Verma, D. B. Kirsch, K. Baron, S. Holfinger, J. Miller, P. Patel, S. Bhargava, and R. Kannan. Evaluating consumer and clinical sleep technologies: an American Academy of Sleep Medicine update. *Journal of Clinical Sleep Medicine*, 17(11):2275–2282, 2021.
- [73] R. N. Sekkal, F. Bereksi-Reguig, D. Ruiz-Fernandez, N. Dib, and S. Sekkal. Automatic sleep stage classification: From classical machine learning methods to deep learning. *Biomedical Signal Processing and Control*, 77:103751, 2022.
- [74] 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.
- [75] A. Stisen, H. Blunck, S. Bhattacharya, T. S. Prentow, M. B. Kjærgaard, A. Dey, T. Sonne, and M. M. Jensen. Smart devices are different: Assessing and mitigating mobile sensing heterogeneities for activity recognition. In *ACM Conference on Embedded Networked Sensor Systems (SenSys)*, pages 127–140, 2015.
- [76] M. Tan and Q. Le. EfficientNet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning (ICML)*, pages 6105–6114, 2019.
- [77] B. H. Te Lindert and E. J. Van Someren. Sleep estimates using Microelectromechanical Systems (MEMS). *Sleep*, 36(5):781–789, 2013.
- [78] S. M. G. P. Togeiro and A. K. Smith. Métodos diagnósticos nos distúrbios do sono. *Brazilian Journal of Psychiatry*, 27:8–15, 2005.
- [79] M. Vakalopoulou, S. Christodoulidis, N. Burgos, O. Colliot, and V. Lepetit. Deep learning: basics and convolutional neural networks (CNNs). *Machine Learning for Brain Disorders*, pages 77–115, 2023.
- [80] 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.
- [81] 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.
- [82] Z. Wang and T. Oates. Imaging time-series to improve classification and imputation. *arXiv preprint arXiv:1506.00327*, 2015.
- [83] 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.
- [84] S. Xu, O. Faust, S. Seoni, S. Chakraborty, P. D. Barua, H. W. Loh, H. Elphick, F. Molinari, and U. R. Acharya. A review of automated sleep disorder detection. *Computers in Biology and Medicine*, 150:106100, 2022.

- [85] Z. Zeng, M. G. Amin, and T. Shan. Arm motion classification using time-series analysis of the spectrogram frequency envelopes. *Remote Sensing*, 12(3):454, 2020.
- [86] 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.