PODER EXECUTIVO
MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO AMAZONAS
INSTITUTO DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO
EM INFORMÁTICA

DANIEL MITSUAKI DA SILVA UTYIAMA

RECONHECIMENTO DE EMOÇÕES BASEADO EM APRENDIZADO AUTOSSUPERVISIONADO

Manaus – Amazonas
Julho – 2024

DANIEL MITSUAKI DA SILVA UTYIAMA

RECONHECIMENTO DE EMOÇÕES BASEADO EM APRENDIZADO AUTOSSUPERVISIONADO

Dissertação apresentada ao Programa de Pós-Graduação em Informática do Instituto de Computação da Universidade Federal do Amazonas como requisito para a obtenção do grau de Mestre em Informática.

Orientador: Prof. Dr. Eduardo James Pereira Souto

Manaus – Amazonas
Julho – 2024

Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

Utyiama, Daniel Mitsuaki da Silva
U93r Reconhecimento de Emoções Baseado em Aprendizado Autossupervisionado / Daniel Mitsuaki da Silva Utyiama . 2024
90 f.: il. color; 31 cm.

Orientador: Eduardo James Pereira Souto
Dissertação (Mestrado em Informática) - Universidade Federal do Amazonas.

1. Reconhecimento de Emoções. 2. Aprendizado Autossupervisionado. 3. Bases de Dados Rotuladas. 4. Tarefas Auxiliares. 5. Redes Neurais Profundas. I. Souto, Eduardo James Pereira. II. Universidade Federal do Amazonas III. Título

Ministério da Educação
Universidade Federal do Amazonas
Coordenação do Programa de Pós-Graduação em Informática

FOLHA DE APROVAÇÃO

"RECONHECIMENTO DE EMOÇÕES BASEADO EM APRENDIZADO AUTO-SUPERVISIONADO"

DANIEL MITSUAKI DA SILVA UTYIAMA

DISSERTAÇÃO DE MESTRADO DEFENDIDA E APROVADA PELA BANCA EXAMINADORA CONSTITUÍDA PELOS PROFESSORES:

Prof. Dr. Eduardo James Pereira Souto - PRESIDENTE

Prof. Dr. Rafael Giusti - MEMBRO INTERNO

Profa. Dra. Patrícia Takako Endo - MEMBRO EXTERNO

MANAUS, 12 de julho de 2024.

Documento assinado eletronicamente por **Eduardo James Pereira Souto**, **Professor do Magistério Superior**, em 10/08/2024, às 10:37, conforme horário oficial de Manaus, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Rafael Giusti**, **Professor do Magistério Superior**, em 12/08/2024, às 17:09, conforme horário oficial de Manaus, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Patricia Takako Endo**, **Usuário Externo**, em 01/09/2024, às 09:13, conforme horário oficial de Manaus, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site https://sei.ufam.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2132934** e o código CRC **CEB570A9**.

Avenida General Rodrigo Octávio, 6200 - Bairro Coroado I Campus Universitário Senador Arthur Virgílio Filho, Setor Norte - Telefone: (92) 3305-1181 / Ramal 1193
CEP 69080-900, Manaus/AM, coordenadorppgi@icomp.ufam.edu.br

Referência: Processo nº 23105.029768/2024-78 SEI nº 2132934

AGRADECIMENTOS

Gostaria de agradecer ao professor Dr. Eduardo James Pereira Souto, cuja convivência ao longo dos anos muito me ensinou e contribuiu significativamente para meu crescimento científico e intelectual. Agradeço também ao Kevin Quispe pelo contínuo apoio, incentivo e atenção durante todo o processo. À professora Dr. Odette Mestrinho Passos, que durante minha jornada sempre me orientou, direcionou e esteve presente. Expresso minha gratidão à minha mãe, Maria de Fátima da Silva Utyiama, e à minha avó, Olinda Gonçalves (*in memorian*), pela educação e valores que me proporcionaram. Agradeço especialmente à minha namorada, Ana Carolina Queiroz, pelo carinho, apoio constante e motivação para seguir em frente. Este trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES-PROEX) - Código de Financiamento 001.

RESUMO

O reconhecimento de emoções é uma aplicação do aprendizado de máquina que envolve a análise de sinais fisiológicos, de áudio e/ou vídeo para identificar emoções expressas pelos indivíduos. A obtenção de bases de dados rotuladas para essa tarefa é desafiadora e onerosa, muitas vezes apresentando problemas estruturais como desequilíbrio de classes, dados faltantes e vieses de rotulagem.

Uma abordagem promissora para contornar esses problemas é desenvolver soluções de reconhecimento de padrões baseadas no aprendizado autossupervisionado. Essa abordagem permite treinar modelos utilizando dados não rotulados, transferindo o conhecimento adquirido para um modelo especializado no reconhecimento de emoções. Dessa forma, é possível superar a dependência de bases de dados rotuladas, tornando o processo mais eficiente e menos custoso.

A escolha de tarefas auxiliares no aprendizado autossupervisionado é crucial, pois possibilita o treinamento eficiente de modelos em grandes bases de dados não rotulados e contribui para a aprendizagem de representações robustas e generalizáveis. Isso permite que o modelo se adapte melhor a diferentes tarefas e cenários.

Nesse contexto, este trabalho apresenta uma arquitetura de rede neural que utiliza uma abordagem auto-supervisionada para o reconhecimento de emoções a partir de sinais de eletrocardiograma. Para avaliar o desempenho da arquitetura neural proposta, implementamos e avaliamos diferentes combinações de tarefas auxiliares, analisando como cada uma contribui para a eficácia e precisão do modelo. Identificamos as tarefas auxiliares mais significativas para a classificação de emoções e realizamos análises detalhadas dos parâmetros associados a essas tarefas.

Experimentos conduzidos em quatro bases de dados públicas demonstraram consistentemente o desempenho superior do método proposto em comparação com a mesma arquitetura treinada de forma supervisionada. Na base de dados SWELL, o método alcançou uma acurácia de 93,64% na classificação de excitação, que é uma dimensão da emoção, utilizando apenas 25% dos dados rotulados, comparado a 78,20% do método supervisionado.

Palavras-Chave: Reconhecimento de Emoções. Aprendizado Autossupervisionado. Bases de Dados Rotuladas. Atividades Auxiliares. Redes Neurais Profundas.

ABSTRACT

Emotion recognition is a machine learning application that involves analyzing physiological, audio, and/or video signals to identify emotions expressed by individuals. Obtaining labeled datasets for this task is challenging and costly, often presenting structural problems such as class imbalance, missing data, and labeling biases.

A promising approach to circumvent these problems is to develop pattern recognition solutions based on self-supervised learning. This approach allows training models using unlabeled data, transferring the acquired knowledge to a model specialized in emotion recognition. In this way, it is possible to overcome the dependency on labeled datasets, making the process more efficient and less costly.

The choice of auxiliary tasks in self-supervised learning is crucial, as it enables efficient training of models on large unlabeled datasets and contributes to learning robust and generalizable representations. This allows the model to better adapt to different tasks and scenarios.

In this context, this work presents a neural network architecture that uses a self-supervised approach for emotion recognition from electrocardiogram signals. To evaluate the performance of the proposed neural architecture, we implemented and evaluated different combinations of auxiliary tasks, analyzing how each one contributes to the model's effectiveness and accuracy. We identified the most significant auxiliary tasks for emotion classification and conducted detailed analyses of the parameters associated with these tasks.

Experiments conducted on four public datasets consistently demonstrated the superior performance of the proposed method compared to the same architecture trained in a supervised manner. In the SWELL dataset, the method achieved an accuracy of 93.64% in arousal classification, which is the degree of alertness of the emotion,, using only 25% of the labeled data, compared to 78.20% for the supervised method.

Keywords: Emotion Recognition. Self-Supervised Learning. Labeled Datasets. Auxiliary Tasks. Deep Neural Networks.

LISTA DE TABELAS

Tabela 3.1 - Trabalhos relacionados agrupados por método, base de dados, tipo de aprendizado autossupervisionado e desempenho obtido.....36
Tabela 4.1 - Descrição das tarefas auxiliares selecionadas para o modelo autossupervisionado.....43
Tabela 4.2 - Especificação da rede neural multitarefa convolucional profunda que foi implementada para pré-treinamento autossupervisionado. 44
Tabela 4.3 Estrutura da camada neural convolucional da rede e parâmetros para um treinamento totalmente supervisionado bem como a transferência de ajustes de aprendizagem/fine tuning 45
Tabela 5.1 - Exemplo de matriz de confusão para um problema de n classes.....55
Tabela 5.2 - Tarefas auxiliares combinadas no aprendizado autossupervisionado. 58
Tabela 5.3 - Resultado da classificação de emoções em excitação e valência para a base de dados AMIGOS. 59
Tabela 5.4 - Resultado da classificação de emoções em excitação e valência para a base de dados DREAMER. 60
Tabela 5.5 - Resultado da classificação de emoções em excitação e valência para a base de dados SWELL.....60
Tabela 5.6 - Resultado da classificação de emoções em excitação e valência para a base de dados SWELL.....61
Tabela 5.7 - Comparação entre os trabalhos da literatura e o nosso modelo para a base de dados AMIGOS. 69
Tabela 5.8 - Comparação entre os trabalhos da literatura e o nosso modelo para a base de dados DREAMER. 70
Tabela 5.9 - Comparação entre os trabalhos da literatura e o nosso modelo para a base de dados..70
Tabela 5.10 - Comparação entre os trabalhos da literatura e o nosso modelo para a base de dados WESAD. 70

LISTA DE FIGURAS

Figura 2.1 - Roda de emoções de Plutchik (1980). 17
Figura 2.2 - Modelo circunflexo de Russel (1980). 18
Figura 2.3 - Questionário SAM. 19
Figura 2.4 - Metodologia para o reconhecimento de emoções. 24
Figura 2.5 - Visão geral da abordagem autossupervisionada. 27
Figura 4.1 - Visão geral da arquitetura proposta para aprendizado autossupervisionado nos sinais de ECG para reconhecimento de emoções. 41
Figura 5.1 - Instâncias por participante na base de dados Amigos. 50
Figura 5.2 - Distribuição de instâncias por escalas na dimensão de excitação na base de dados Amigos. 50
Figura 5.3 - Distribuição de instâncias por escalas na dimensão de valência na base de dados Amigos 50
Figura 5.4 - Instâncias por participante na base de dados Dreamer 51
Figura 5.5 - Distribuição de instâncias por escalas na dimensão de excitação e valência na base de dados Dreamer 51
Figura 5.6 - Instâncias por participante na base de dados Swell. 52
Figura 5.7 - Distribuição de instâncias por escalas na dimensão de excitação na base de dados Swell. 53
Figura 5.8 - Distribuição de instâncias por escalas na dimensão de valência na base de dados Swell. 53
Figura 5.9 - Instâncias por participante na base de dados Wesad 54
Figura 5.10 - Distribuição de instâncias por níveis do estresse na base de dados Wesad 54
Figura 5.11 - Arquitetura implementada para o treinamento autossupervisionado e para classificação de emoções 62
Figura 5.12 - Resultados para a classificação de emoções na base de dado AMIGOS para os grupos de experimentos. 63
Figura 5.13 - Resultados para a classificação de emoções na base de dado DREAMER para os grupos de experimentos. 64
Figura 5.14 - Resultados para a classificação de emoções na base de dado SWELL para os grupos de experimentos. 65
Figura 5.15 - Resultados para a classificação de emoções na base de dado WESAD para os grupos de experimentos. 65
Figura 5.16 - Estrutura do treinamento do modelo com pré-treinamento autossupervisionado e totalmente supervisionado. 67
Figura 5.17- Resultados para a classificação de emoções na base de dado WESAD com a variação da quantidade de dados rotulados utilizados no treinamento. 68
Figura 5.18 - Resultados para a classificação de emoções na base de dado SWELL com a variação da quantidade de dados rotulados utilizados no treinamento. 68

LISTA DE EQUAÇÕES

Equação 5.1 - Quantidade de amostras de teste da classe i 55
Equação 5.2 - Quantidade de amostras de teste classificadas como j 55
Equação 5.3 - Acurácia 56
Equação 5.4 - Precisão 56
Equação 5.5 - Revocação 56
Equação 5.6 - Medida-F 57

SUMÁRIO

1. **INTRODUÇÃO** 10

1.1 **Contexto e Descrição do Problema** 10

1.2 **Justificativa** 10

1.3 **Objetivos** 13

1.4 **Organização do Documento** 14

2. **RECONHECIMENTO DE EMOÇÕES HUMANAS** 15

2.1 **Emoções Humanas** 15

2.2 **Representação das Emoções** 16

2.3 **Medição das Emoções** 18

2.4 **Relação entre Sinais Fisiológicos e Emoções** 19

2.4.1 Eletroencefalografia 20
2.4.2 Eletrocardiografia 21
2.4.3 Atividade eletrodérmica 21
2.4.4 Eletromiografia 22
2.4.5 Variação da frequência cardíaca 23

2.5 **Metodologia para o Reconhecimento de Emoções Humanas** 24

2.6 **Aprendizado Autossupervisionado** 25

2.7 **Considerações Finais** 27

3. **TRABALHOS RELACIONADOS** 28

3.1 **Aprendizado autossupervisionado baseado em Aprendizado Contrastivo** 28

3.2 **Aprendizado Autossupervisionado baseado em Transformações de Sinais** 32

3.3 **Considerações Finais** 39

4. **ARQUITETURA PARA O APRENDIZADO AUTOSSUPERVISIONADO NOS SINAIS DE ECG PARA O RECONHECIMENTO DE EMOÇÕES** 40

4.1 **Pré-Processamento dos Dados** 42

4.2 **Arquitetura CNN Auto-Supervisionada** 43

4.3 **Arquitetura CNN para o Reconhecimento de Emoções** 45

4.4 **Considerações Finais** 46

5. **EXPERIMENTOS E RESULTADOS** 47

5.1 **Protocolo Experimental** 47

5.2 **Resultados** 57

5.2.1 Análise do Treinamento Autossupervisionado Utilizando Combinação de Tarefas Auxiliares 57
5.2.2 Análise da Variação do Número de Filtros de Kernel no Treinamento Autossupervisionado 61
5.2.3 Análise da Variação da Quantidade dos Dados Rotulados utilizados para o Reconhecimento de Emoções 66

6. **CONCLUSÕES** 72

6.1 **Contribuições** 73

6.2 **Trabalhos Futuros** 73

REFERÊNCIAS BIBLIOGRÁFICAS 75

1. INTRODUÇÃO

1.1 Contexto e Descrição do Problema

A área de reconhecimento de emoções tem sido objeto de crescente interesse de pesquisa em inteligência artificial. Com o avanço da tecnologia e o aumento da capacidade computacional, tornou-se possível desenvolver sistemas capazes de identificar e compreender as emoções humanas, permitindo a criação de aplicações em diversas áreas. Por exemplo, na indústria de videogames, o reconhecimento de emoções contribui para melhorar a experiência do jogador, ajustando o nível de dificuldade do jogo com base na emoção detectada (YANG *et al*., 2018). Na educação, o reconhecimento de emoções pode ser usado para identificar alunos que estejam com dificuldades de aprendizagem ou com baixo engajamento, permitindo que o professor intervenha de maneira mais eficaz (WAN; GUO, 2020). Na área de saúde, o reconhecimento de emoções pode ser utilizado para monitorar sinais de depressão, ansiedade ou outros transtornos emocionais, permitindo intervenções precoces e prevenindo agravamentos (MARKOVA; GANCHEV; KALINKOV, 2019) (MONTESINOS *et al*., 2019).

Apesar dos avanços, o reconhecimento de emoções enfrenta desafios significativos. O progresso nas tecnologias de sensores permitiu a utilização de diversos modos de dados para identificar emoções em indivíduos. Entre esses modos, destacam-se dados extraídos de expressões faciais (CAI; DONG; WEI, 2020), de movimentos corporais e passos (AHMED; BARI; GAVRILOVA, 2020), da fala (KHALIL et al., 2019), obtidos a partir de eletrocardiogramas (HSU et al., 2020), eletroencefalogramas (YASEMIN; SARIKAYA; INCE, 2019) e os obtidos por meio da resposta galvânica da pele (JOY et al., 2021). Assim, a combinação de diferentes modalidades de dados e o uso de algoritmos de aprendizado de máquina são essenciais para uma detecção precisa e confiável das emoções.

Atualmente, a maioria das soluções de aprendizado de máquina usadas para o reconhecimento de emoções é baseada em técnicas supervisionadas, que exigem dados rotulados para treinar modelos. No entanto, essa abordagem enfrenta limitações, como a incapacidade de lidar com novos dados ou emoções não previamente catalogadas, o que pode restringir a aplicabilidade dos modelos em situações reais (MEHARI; STRODTHOFF, 2021). Além disso, a rotulagem de dados pode ser um processo caro e demorado, exigindo a participação de especialistas em emoções. Mesmo quando os dados estão rotulados, a qualidade dos rótulos depende da precisão e consistência dos rotuladores humanos, que podem cometer

erros devido a preconceitos inconscientes ou interpretações subjetivas das emoções. Além do mais, certas emoções podem ser difíceis de rotular e podem ser interpretadas de maneiras diferentes por diferentes rotuladores. Quando o método de rotulagem é inadequado para o tipo de dados em questão, podem ocorrer erros e inconsistências nos rótulos atribuídos, prejudicando a precisão do modelo de aprendizado de máquina e sua capacidade de generalização (SCHMIDT et al., 2018b).

Uma outra limitação também presente na abordagem de aprendizado de máquina supervisionada é o alto consumo de recursos computacionais e tempo, já que o modelo precisa ser treinado do zero para cada tarefa de classificação ou regressão (SARKAR, 2020). Além disso, as representações dos modelos treinados por meio dessa abordagem são frequentemente muito específicas para determinadas tarefas e pouco úteis para outras (SARKAR; ETEMAD, 2020). Em outras palavras, quando um modelo é treinado especificamente para uma tarefa usando uma grande base de dados rotulada, ele tende a aprender características que são altamente otimizadas para essa tarefa específica. Por exemplo, um modelo treinado para reconhecer emoções a partir de sinais de eletrocardiograma (ECG) pode aprender representações que capturam nuances únicas desses sinais relacionadas às emoções, mas essas representações podem não ser aplicáveis ou úteis para outras tarefas, como diagnóstico médico de condições cardíacas

Esse fenômeno é frequentemente referido como falta de generalização. Os modelos tornam-se tão afinados com as características específicas dos dados de treinamento e da tarefa que perdem a flexibilidade para lidar de maneira eficaz com problemas ou dados ligeiramente diferentes. Isso limita a utilidade desses modelos em aplicações mais amplas que requerem capacidade de generalização, fazendo com que novos modelos precisem ser treinados para cada nova tarefa, muitas vezes do zero, o que é tanto tempo quanto recurso intensivo.

1.2 Justificativa

O avanço da inteligência artificial tem permitido que diversas áreas da ciência adotem métodos de aprendizado de máquina para inovar e expandir o conhecimento. Contudo, a escassez de bases de dados rotulados é um obstáculo significativo para o progresso nessa área (LIU, G. et al., 2020). Diversas técnicas têm sido desenvolvidas para mitigar a limitação de bases de dados não rotuladas, incluindo o aumento de dados (LIU, P. et al., 2020), o uso de redes adversárias generativas (HOSSAIN et al., 2021), transferência de aprendizado (SAEED; OZCELEBI; LUKKIEN, 2019) e aprendizado autossupervisionado (SARKAR, 2020).

O aumento de dados, por exemplo, cria variações nos dados existentes para ampliar a quantidade e diversidade das amostras, melhorando assim a robustez e generalização dos modelos quando os conjuntos de dados de treinamento são limitados (SHORTEN; KHOSHGOFTAAR, 2019). As redes adversárias generativas, que envolvem uma rede geradora que cria dados sintéticos e uma rede discriminadora que aprende a diferenciar entre dados reais e sintéticos, têm se mostrado úteis para expandir conjuntos de treinamento e melhorar a eficácia dos modelos de aprendizado de máquina (HOSSAIN et al., 2021).

No entanto, essas técnicas enfrentam desafios como alto custo computacional, necessidade de memória adicional, treinamento instável e dificuldades na avaliação de qualidade dos dados gerados, que podem introduzir vieses (SHORTEN; KHOSHGOFTAAR, 2019; ALQAHTANI; KAVAKLI-THORNE; KUMAR, 2021).

A transferência de aprendizado aproveita conhecimento prévio adquirido em tarefas relacionadas para reduzir o tempo e o custo de treinamento de novos modelos, melhorando a generalização e o desempenho em tarefas específicas (SARKAR, 2020). Quando combinada com o aprendizado autossupervisionado, onde os próprios dados funcionam como fonte de supervisão, essa abordagem pode ainda mais diminuir a dependência de dados rotulados manualmente e aumentar a precisão dos modelos (JING et al., 2018).

Apesar dos avanços, a aplicação do aprendizado autossupervisionado, especialmente no reconhecimento de emoções a partir de sinais de ECG, ainda enfrenta desafios. Os sinais de ECG contêm nuances específicas que são difíceis de capturar sem supervisão adequada, tornando complexa a tarefa de aprender representações generalizáveis que possam ser efetivamente utilizadas em reconhecimento de emoções (ZHANG et al., 2016; JING et al., 2018).

Para abordar essas limitações, propomos uma abordagem auto-supervisionada para o aprendizado de sinais de ECG, utilizando uma combinação de atividades auxiliares em um domínio genérico. Esta abordagem permite que cada tarefa traga diferentes perspectivas ou restrições sobre as representações aprendidas, aumentando a probabilidade de que um conjunto dessas tarefas auxilie no reconhecimento efetivo de emoções. Ao diminuir a dependência de rotulagem manual e melhorar a capacidade de generalização, esta abordagem tem potencial para ajudar a superar as barreiras existentes e promover avanços significativos no reconhecimento de emoções em aplicações práticas.

1.3 Objetivos

O objetivo deste trabalho é desenvolver uma arquitetura de rede neural que utilize uma abordagem autossupervisionada para reconhecimento de emoções a partir de sinais de eletrocardiograma. Esta abordagem pretende superar as limitações dos métodos supervisionados tradicionais, que geralmente dependem de grandes volumes de dados rotulados, por meio da implementação de tarefas auxiliares que visam aprimorar a generalização das representações do modelo.

Para alcançar o objetivo geral, delineamos os seguintes objetivos específicos:

1. Avaliação de tarefas auxiliares: Avaliar o impacto de diferentes combinações de tarefas auxiliares na classificação de emoções, para determinar como cada tarefa contribui para a eficácia e precisão do modelo.
2. Análise de contribuição das tarefas: Identificar quais tarefas auxiliares são mais significativas para o desempenho na classificação de emoções, investigando a relevância de cada uma no resultado final.
3. Desenvolvimento de rede neural auto-supervisionada: Implementar uma arquitetura de rede neural auto-supervisionada para o aprendizado de representações de ECG focadas no reconhecimento de emoções, incluindo análises sobre o impacto do ajuste dos filtros de kernel nas convoluções no desempenho do modelo.
4. Análises detalhadas dos parâmetros: Realizar análises aprofundadas dos parâmetros associados às tarefas de aprendizado autossupervisionado para entender como essas tarefas contribuem para a formação de representações afetivas eficazes de ECG.

1.4 Organização do Documento

Este documento estruturado da seguinte maneira:

- **Capítulo 1:** Introduz o contexto e a problemática abordada, delineando a motivação e a justificativa para a realização deste estudo, além de definir os objetivos da pesquisa**.**
- **Capítulo 2:** Discorre sobre os conceitos fundamentais relacionados à pesquisa, incluindo reconhecimento de emoções, sinais fisiológicos, e as diversas metodologias de aprendizado de máquina, como aprendizado supervisionado, não supervisionado, semi-supervisionado, transferência de aprendizado e aprendizado autossupervisionado.
- **Capítulo 3:** Oferece uma revisão da literatura, organizada de acordo com o tipo de aprendizado e as tarefas de reconhecimento de emoções, visando facilitar o entendimento dos métodos e abordagens existente.
- **Capítulo 4:** Descreve a arquitetura de rede neural proposta, destacando os principais conceitos e mecanismos envolvidos em sua implementação.
- **Capítulo 5:** Detalha os experimentos realizados para testar o método proposto de reconhecimento de emoções, incluindo a descrição das tarefas auxiliares de aprendizado autossupervisionado, apresentando os resultados alcançados e uma discussão sobre esses.
- **Capítulo 6:** Apresenta as conclusões do estudo, resumindo os principais resultados e contribuições, além de sugerir direções para trabalhos futuros.

2. RECONHECIMENTO DE EMOÇÕES HUMANAS

Neste capítulo são apresentados os principais conceitos relacionados à definição de emoções, modelos de representação, as principais características dos sinais fisiológicos utilizados para medir os estados afetivos, metodologia para o reconhecimento de emoções, bem como uma introdução ao aprendizado autossupervisionado e às tarefas auxiliares. Esses conceitos contribuem para a compreensão da abordagem e dos tipos de emoções que são identificados nesta pesquisa.

2.1 Emoções Humanas

As emoções humanas são fenômenos complexos e multifacetados. Há inúmeras teorias e perspectivas a partir das quais as emoções são estudadas, o que as torna ainda mais desafiadoras de compreender (MOORS, 2010). As emoções são fundamentais na experiência humana, desempenhando um papel importante em nossas interações sociais, tomada de decisões e bem-estar emocional. Em geral, as emoções são experiências subjetivas que envolvem respostas fisiológicas, cognitivas e comportamentais a estímulos externos ou internos (MAUSS; ROBINSON, 2009) (ALARCAO; FONSECA, 2017).

As emoções iniciam a partir de um estímulo que produz uma experiência subjetiva, na qual uma grande variedade de elementos (e.g. cultura, educação, experiências anteriores, personalidade) podem determinar a percepção e as respostas de uma pessoa. As experiências subjetivas podem variar em intensidade de pessoa para pessoa e podem provocar diferentes respostas comportamentais em um único indivíduo (MAUSS; ROBINSON, 2009).

As respostas comportamentais são as expressões externas da emoção, como um sorriso, uma gargalhada, um grito e outras reações visíveis. Por exemplo, o susto em resposta a um estímulo inesperado e intenso é um reflexo universal que envolve múltiplas ações motoras, que incluem tensão nos músculos do pescoço, das costas e o piscar dos olhos (GANGULY; SINGLA, 2019). Já as respostas fisiológicas são as alterações internas do corpo em resposta às experiências emocionais, resultantes das reações do sistema nervoso autônomo.

O sistema nervoso autônomo é responsável por controlar as respostas involuntárias do corpo, tais como a respiração, o batimento cardíaco e o movimento da pupila. Por exemplo, em resposta a um estímulo estressante, substâncias como adrenalina e cortisol são rapidamente liberadas no corpo, preparando o indivíduo para uma reação de “lutar ou fugir”.

Para criar representações para as emoções, alguns pesquisadores desenvolveram trabalhos no campo da psicologia que tentam relacionar as características físicas e fisiológicas com os estados afetivos. Essas teorias se baseiam na ideia de que diferentes emoções têm padrões distintos de respostas corporais associados a elas. Por exemplo, a emoção de medo pode estar associada a um aumento na frequência cardíaca, sudorese e dilatação das pupilas (EKMAN, 1992). Contudo, é crucial reconhecer que a relação entre estados emocionais e respostas corporais é complexa, variando significativamente entre indivíduos e situações (SUBRAMANIAN *et al.*, 2018)

2.2 Representação das Emoções

Para um observador leigo, pode parecer fácil identificar se alguém está experimentando ou expressando emoções específicas, como felicidade ou medo. No entanto, definir ou medir o estado emocional de uma pessoa é um dos problemas mais discutidos na ciência afetiva (MAHESH, 2020), devido às diferentes perspectivas de representação das emoções.

Existem duas perspectivas comuns de representação emocional: a dimensional e a discreta. Na perspectiva discreta, cada emoção corresponde a um perfil único e universal em termos de experiência, fisiologia e comportamento. Essa abordagem categoriza as emoções em um conjunto limitado de emoções básicas e distintas. Por exemplo, Ekman (1992) argumenta que todas as pessoas podem expressar e reconhecer seis emoções básicas: tristeza, felicidade, surpresa, medo, raiva e desgosto. Embora muitos psicólogos tenham aceitado a teoria das emoções básicas, não há consenso sobre o número preciso de emoções básicas. Robert Plutchik (1980), por exemplo, propôs oito emoções primárias: raiva, medo, tristeza, nojo, surpresa, antecipação, confiança e alegria, organizadas em uma roda de cores, conforme exibido na Figura 2.1.

Enquanto a perspectiva discreta assume que cada emoção corresponde a uma categoria distinta, outras pesquisas apontam para a possibilidade de considerar outras emoções a partir da intensidade ou da combinação das emoções básicas. Zenonos et al. (2016) propuseram uma abordagem para distinguir oito emoções e humores diferentes, incluindo animação, felicidade, calma, cansaço, tédio, tristeza, estresse e irritação. Apesar de sua simplicidade e alto grau de interpretabilidade, alguns pesquisadores argumentam que os modelos discretos são incapazes de capturar a complexidade e a variabilidade individual na experiência emocional (COWIE; CORNELIUS, 2003). Portanto, é importante considerar diferentes perspectivas para compreender e medir as emoções humanas de forma mais precisa e abrangente.

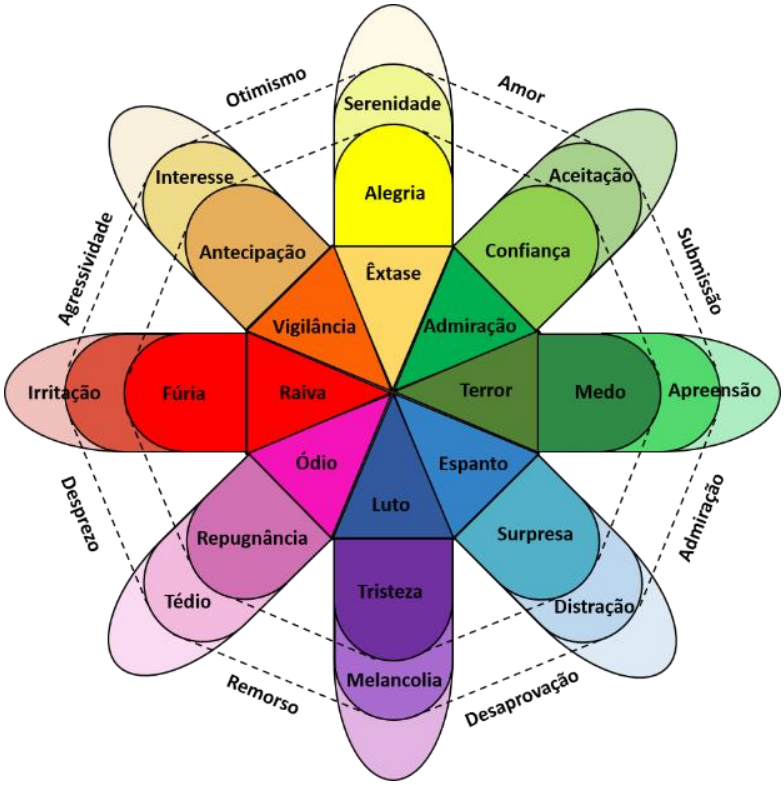

Figura 2.1 - Roda de emoções de Plutchik (1980).

A perspectiva dimensional baseia-se em uma abordagem contínua, onde as emoções são representadas ao longo de um espectro com extremos opostos (RUSSELL, 1980). Os estudos dimensionais das emoções remontam a Wundt (1913) que propôs que as emoções são definidas por três dimensões independentes: agradável-desagradável, tensão-relaxamento e excitação-calma. Posteriormente, Russell (1980) refinou essa abordagem ao introduzir um modelo circunflexo, distribuindo as emoções em um espaço circular baseado em duas dimensões principais: excitação e valência.

No modelo circunflexo de Russell, as emoções sãoagrupadas com base nas dimensões de excitação (ativação) e valência (positiva ou negativa) representando os eixos vertical e horizontal, respectivamente. O centro do círculo indica um estado de excitação média e valência neutra, enquanto os diferentes quadrantes do círculo ilustram combinações específicas de excitação e valência, como mostrado na Figura 2.2. Nesse modelo, as expressões emocionais podem ser ilustradas em qualquer nível de excitação e valência ou definidas a partir de quatro regiões (quadrantes).

As expressões emocionais neste modelo são ilustradas em qualquer ponto ao longo destes eixos ou definidas a partir dos quatro quadrantes, cada um representando diferentes combinações de ativação e valência – como animado (alta valência positiva e alta ativação) ou

depressivo (alta valência negativa e baixa ativação). A perspectiva dimensional permite representar emoções complexas em um espaço contínuo e permite uma análise mais refinada das relações entre diferentes emoções.

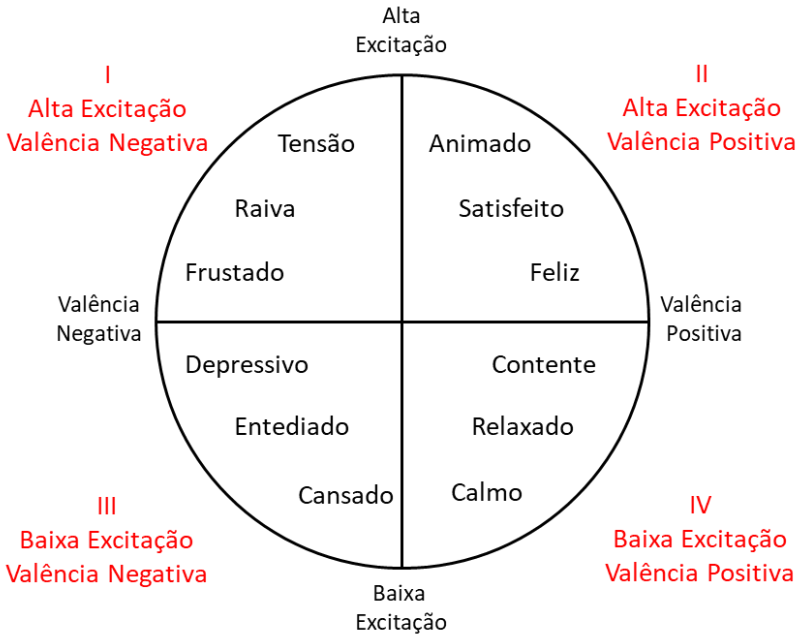

Figura 2.2 - Modelo circunflexo de Russel (1980).

2.3 Medição das Emoções

A mensuração das emoções pode ser feita por meio da observação dos três componentes de uma resposta emocional: o cognitivo, o comportamental e o fisiológico. O componente cognitivo está relacionado à avaliação subjetiva da emoção, que inclui a interpretação cognitiva da situação e a avaliação do significado pessoal atribuído a ela. Essa experiência pode ser capturada por meio de questionários de autoavaliação. A Figura 2.3 apresenta um exemplo de um questionário chamado *Self-Assessment Manikin* (SAM), projetado para capturar as experiências emocionais a partir de uma perspectiva dimensional, considerando três dimensões principais: excitação, valência e dominância.

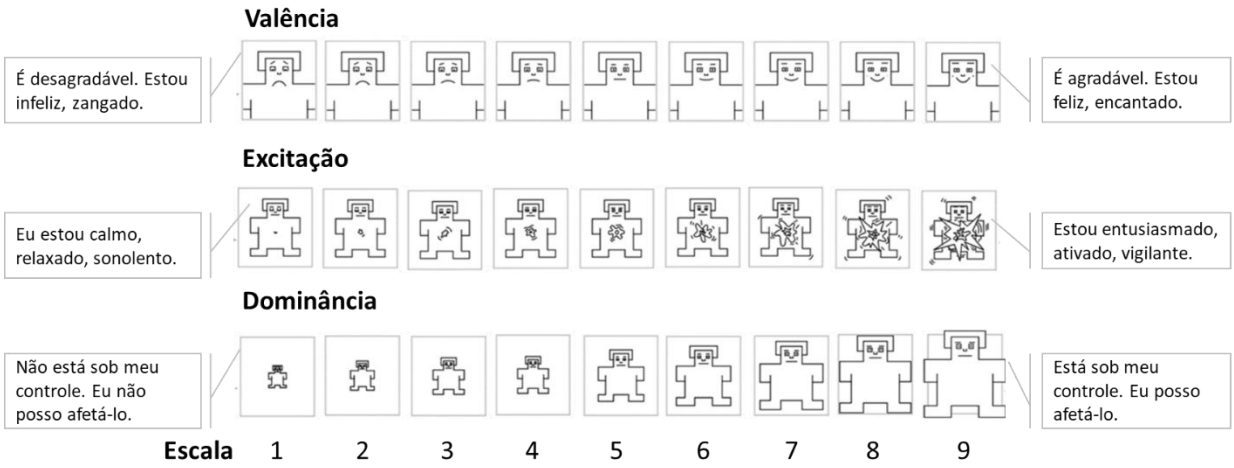

Figura 2.3 - Questionário SAM.

O componente comportamental refere-se às expressões comportamentais da emoção, como expressões faciais, gestos, postura, movimentos corporais e padrões de fala. A fala humana é uma das principais formas de expressão humana (ZENG *et al.*, 2008). Além de transmitir informações desejadas por meio de palavras sonoras, o falante também compartilha informações sobre o tom da voz, energia, velocidade e outras propriedades acústicas, que ajudam o receptor a medir as intenções e emoções da comunicação. Por outro lado, as expressões faciais e gestos são as formas mais comuns de identificação de emoções como indicado por Joy (2021), que apoia a existência de um conjunto de expressões faciais universalmente reconhecidas para emoções, como felicidade, surpresa, medo, tristeza, raiva e desgosto. Além disso, as pesquisas baseadas no movimento corporal, postura e gestos têm crescido nos últimos anos, dadas as possibilidades de reconhecimento de emoções a distância.

O componente fisiológico das emoções refere-se às mudanças físicas que ocorrem no corpo em resposta a uma emoção, como alterações na frequência cardíaca, pressão arterial, condutância da pele e atividade cerebral (KREIBIG, 2010). Para medir as emoções, os sinais fisiológicos mais comuns incluem o eletroencefalograma, o eletrocardiograma, e medidas como a atividade eletrodérmica e a eletromiografia. A próxima seção detalhará como esses sinais são utilizados para inferir os estados emocionais.

2.4 Relação entre Sinais Fisiológicos e Emoções

A relação entre sinais fisiológicos e emoções é objeto de estudos científicos na área da psicologia, que apontam uma forte conexão entre eles (SHU *et al.*, 2018) (YAN *et al.*, 2022). As emoções podem desencadear mudanças fisiológicas em todo o corpo, que podem ser medidas e utilizadas para entender a experiência emocional de uma pessoa. Por exemplo, alguns

estados emocionais negativos, como medo e ansiedade, podem levar a alterações fisiológicas fortes, tais como sudorese, boca seca e mal-estar (YADAV *et al.*, 2021) (KHATTAK *et al.*, 2021) (NGAI *et al.*, 2022) (XU *et al.*, 2022). Já o estado de felicidade é caracterizado pelo aumento da atividade cardíaca, vasodilatação, aumento da atividade eletrodérmica e aumento da atividade respiratória (KREIBIG, 2010).

A expressão das emoções por meio de respostas fisiológicas é um processo natural, geralmente inconsciente e controlado pelo sistema nervoso central. Isso dificulta a falsificação ou a mascaramento das reações emocionais pelo sujeito. Assim, a inferência de emoções a partir de sinais fisiológicos apresenta vantagens em comparação com inferência por experiências subjetivas ou respostas comportamentais (LI, J. *et al.*, 2022) (SUN; LIN, 2022).

Dentre os sinais fisiológicos existentes, os principais sinais e técnicas frequentemente aplicadas para inferir estados afetivos são:

2.4.1 Eletroencefalografia

A eletroencefalografia é uma técnica que mede a atividade elétrica do cérebro e tem sido indicada para identificar alterações neurológicas (LI, X. *et al.*, 2022). Características específicas dos sinais EEG, como as bandas alfa e beta, ajudam a identificar emoções autoavaliadas positivas, incluindo gratidão, inspiração e orgulho, e as bandas teta e gama são associadas a emoções prazerosas, tais como diversão, interesse e alegria (HU *et al.*, 2017). Por essa razão, os sinais EEG têm sido amplamente utilizados em estudos para detectar respostas emocionais do indivíduo a diferentes estímulos (CHAO *et al.*, 2019). Por exemplo, Krishna et al. (2019) propõem a utilização do sinal EEG para identificar as expressões de emoção de portadores de deficiência física ou imobilizados, enquanto Zhang et al. (2016) propõem um método para selecionar os melhores canais do sinal EEG para identificar as emoções de alegria, medo, tristeza e relaxamento. Além disso, outros estudos buscam avaliar diferentes emoções e discutem qual tipo de estímulo (visual, áudio ou audiovisual) é melhor para estabelecer as emoções a partir do EEG (SARMA; BARMA, 2020) (ZHUANG *et al.*, 2018).

No entanto, é importante destacar que a eletroencefalografia pode apresentar limitações, como a necessidade de o indivíduo estar em repouso para obtenção de dados precisos, o que pode ser difícil em algumas situações, e a interferência de outros sinais elétricos corporais que podem afetar a qualidade dos resultados (LI, X. *et al.*, 2022). Portanto, é necessário considerar essas limitações ao interpretar os resultados obtidos a partir da eletroencefalografia.

2.4.2 Eletrocardiografia

A eletrocardiografia é um método não invasivo para registrar a atividade elétrica do coração durante um determinado intervalo de tempo (BERKAYA *et al.*, 2018). Além de fornecer informações importantes para diagnóstico de anomalias cardíacas, a ECG pode ser utilizada para identificar estados emocionais em indivíduos (PAWEŁJEMIOŁO *et al.*, 2022), já que as emoções podem produzir variações no sinal do ECG (SUN; LIN, 2022).

Os principais parâmetros do sinal de eletrocardiograma, incluindo as ondas P, Q e T, os complexos QRS e o intervalo QT/QTc, são frequentemente utilizados para análise da atividade cardíaca de um indivíduo. A maioria dos estudos relacionados ao reconhecimento de emoções com base em ECG se concentra na avaliação da duração e amplitude do complexo QRS (DZEDZICKIS; KAKLAUSKAS; BUCINSKAS, 2020). Por exemplo, Cai, Liu e Hao (2009) analisaram características extraídas do complexo QRS e mostraram que a tristeza pode ser reconhecida com mais facilidade e precisão do que a emoção da alegria. Uyarel et al. (2006) analisaram a dispersão do parâmetro QT/QTc e comprovaram que essa medição fisiológica pode ser usada como marcador para reconhecer raiva intensa.

No entanto, uma desvantagem do uso do sinal eletrocardiogramaé que ele é bastante sensível a ruídos e normalmente é obtido em ambientes clínicos quando o paciente está em estado calmo.

2.4.3 Atividade eletrodérmica

Atividade eletrodérmica (EDA) é uma medida da mudança nas propriedades elétricas da pele em relação à excreção do suor, obtida pela variação contínua das características elétricas da pele humana (DZEDZICKIS; KAKLAUSKAS; BUCINSKAS, 2020). A variação da condutância da pele (SC) pode ser medida de forma não invasiva ao aplicar uma pequena corrente elétrica. Além disso, a resposta galvânica da pele (GSR) é a medida da variação da SC em resposta à atividade de excreção do suor. O GSR é frequentemente referido como EDA ou SC (DESAI; SHETTY, 2021). Essa medida não pode ser controlada voluntariamente e é uma variável importante para medir a excitação emocional (SHUKLA *et al.*, 2021).

As mudanças emocionais induzem reações de suor, que são principalmente perceptíveis na superfície dos dedos das mãos e nas solas dos pés. A reação ao suor provoca uma variação na quantidade de sal na pele humana, o que leva à alteração da resistência elétrica da pele (AYATA; YASLAN; KAMAŞAK, 2017). A condutância da pele está principalmente

relacionada ao nível de excitação: se o nível de excitação aumenta, a condutância da pele também aumenta. Por essa razão, algumas pesquisas buscam utilizar o sinal EDA para identificar doenças e alterações nos estados afetivos, como estresse, excitação, frustração, raiva e dor (LIU; DU, 2018) (ZONTONE *et al.*, 2019) (LANG *et al.*, 1993) (AQAJARI *et al.*, 2021).

Comparado ao eletroencefalograma e eletrocardiograma, o GSR requer uma quantidade menor de eletrodos de medição, o que facilita o uso de dispositivos vestíveis e a definição de estados emocionais quando uma pessoa está envolvida em atividades normais (DZEDZICKIS; KAKLAUSKAS; BUCINSKAS, 2020). Entretanto, semelhante a outras técnicas de medição, a precisão do GSR pode ser comprometida por artefatos decorrentes de movimentos.

2.4.4 Eletromiografia

A eletromiografia (EMG) é uma técnica que mede a atividade elétrica muscular em resposta a estímulos nervosos ou musculares (MAAOUI; PRUSKI; ABDAT, 2008). É amplamente utilizada em muitas áreas da ciência, incluindo avaliação da saúde neuromuscular (HOGREL, 2005), análise da ativação muscular em esportes (SAISHO *et al.*, 2019), análise da marcha (PAPAGIANNIS *et al.*, 2019), avaliação da fadiga muscular (SUBASI; KIYMIK, 2009), controle de próteses e exoesqueletos (FARINA *et al.*, 2014) e no campo da psicologia (GIANNAKAKIS *et al.*, 2022).

No reconhecimento de emoções, o EMG é usado para investigar a relação entre as emoções cognitivas e as reações fisiológicas (ZONG; CHETOUANI, 2009). A maioria dos estudos que utiliza o EMG para reconhecer as reações emocionais se concentra na análise das expressões faciais. Por exemplo, Kim et al. (2022) exploram o uso dos sinais EMG facial e EEG para classificar emoções como felicidade, surpresa, medo, raiva, tristeza e desgosto. Mithbavkar e Shah (2021) desenvolveram um conjunto de dados para o reconhecimento de emoções baseado em dados coletados por meio de eletromiogramas durante a dança, estimulando respostas emocionais como espanto, temor, humor e tranquilidade. Já Joesph et al. (2020) propõem a extração de características das medidas de EMG, pressão sanguínea e GSR para detectar estágios emocionais de felicidade, tristeza, raiva, ódio e respeito.

Como outras técnicas de medição por contato, incluindo EEG e ECG, o EMG pode comprometer o conforto e apresentar limitações para uso contínuo. No entanto, é uma técnica excelente para detectar emoções fortes, pois mudanças drásticas na valência e na intensidade de excitação produzem alterações nas expressões faciais (WIOLETA, 2013).

2.4.5 Variação da frequência cardíaca

A variação da frequência cardíaca (HRV) representa a variação no intervalo de tempo entre os batimentos cardíacos consecutivos (BENEZETH *et al.*, 2018). A variabilidade da frequência cardíaca é regulada pelo sistema nervoso autônomo, mais especificamente pelos nervos simpáticos que aceleram a frequência cardíaca e pelos nervos parassimpáticos que diminuem a frequência cardíaca. Alterações na frequência cardíaca são influenciadas pelas emoções, estresse e exercícios físicos (HUANG *et al.*, 2019) (BENEZETH *et al.*, 2018). As medições de HRV são usadas para monitorar estados afetivos como ansiedade, raiva, medo, estresse e relaxamento (GUO *et al.*, 2016) ou para auxiliar na detecção ou tratamento de doenças psiquiátricas, como depressão (KIM, E. Y. *et al.*, 2017), ansiedade (NA; CHO; CHO, 2021) e dependência química (YUAN; HUANG; YAN, 2019). Por exemplo, Thanapattheerakul et al. (2018) mostraram que o sentimento de tristeza, quando induzido pelo choro, tende a aumentar a HRV. Essa característica do HRV mostra que a intensidade e o contexto em que os estímulos são apresentados podem afetar a detecção dos estágios emocionais.

As medições de HRV são comumente obtidas a partir do sinal ECG, que fornece informações sobre a variação do intervalo RR em relação ao tempo. Entretanto, o uso do ECG apresenta problemas de sensibilidade e ruído, conforme mencionado anteriormente. Uma alternativa bastante utilizada, principalmente por causa da proliferação de relógios inteligentes, é a fotopletismografia (PPG). Essa técnica é usada para detectar mudanças no volume de sangue nos tecidos microvasculares. Seu funcionamento se dá por um fotodetector e uma fonte de luz que ilumina o tecido, e o fotodetector mede as pequenas variações da luz refletida (BENEZETH *et al.*, 2018). Há uma variedade de estudos que comprovam as vantagens em se utilizar essa técnica para a extração do HRV, quando comparada ao ECG (JEYHANI *et al.*, 2015) (CHOI *et al.*, 2017). Além da abordagem usual do PPG mencionada acima, também existe a abordagem remota, pela qual é possível recuperar a onda de pulso cardiovascular medindo as variações da luz emitida remotamente no ambiente, por meio de sistemas de visão computacional (BENEZETH *et al.*, 2018). Essa abordagem aumenta o nível de conforto da pessoa durante o procedimento de medição, mas aumenta o ruído no sinal, o que exige sistemas avançados de processamento e análise de sinais.

Sistemas de reconhecimento de emoções, como os apresentados neste trabalho, podem ser usados para inferir os estados emocionais dos seres humanos. No entanto, a análise de

padrões e correlações de alta dimensão dos sinais fisiológicos acima seria praticamente impossível sem o uso de computadores e métodos computacionais, como aprendizado de máquina (KOTOWSKI; STAPOR, 2021).

2.5 Metodologia para o Reconhecimento de Emoções Humanas

Diversas metodologias são descritas na literatura para o reconhecimento de emoções por meio de dispositivos vestíveis, abrangendo principalmente três estágios: aquisição do sinal, processamento de dados e modelo de aprendizado e avaliação (MONTESINOS et al., 2019), conforme demonstrado na Figura 2.4.

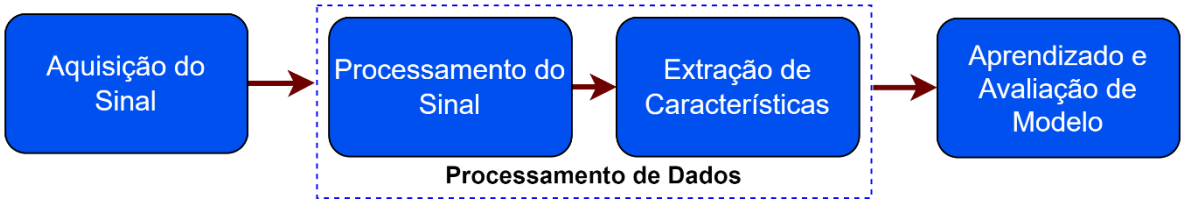

Figura 2.4 - Metodologia para o reconhecimento de emoções.

Aquisição do sinal: os dados podem ser coletados de várias maneiras, incluindo a gravação de expressões faciais, sinais fisiológicos, voz e texto. No contexto dos sinais fisiológicos, os dados podem ser obtidos por meio de sensores clínicos e dispositivos vestíveis de baixo custo. Normalmente, são capturados enquanto o indivíduo executa atividades físicas, intelectuais ou em exames clínicos de rotina. Considerando as dificuldades na aquisição desses sinais, diversos estudos disponibilizam bases de dados com registros de alterações fisiológicas em contextos emocionais, como DREAMER (KATSIGIANNIS; RAMZAN, 2017), AMIGOS (MIRANDA-CORREA et al., 2018), WESAD (SCHMIDT et al., 2018a), CLAS (MARKOVA; GANCHEV; KALINKOV, 2019) e K-EmoCon (PARK et al., 2020).

Processamento de dados: os dados coletados precisam ser processados para remover qualquer ruído e padronizá-los. Esse processo pode incluir segmentação de dados, normalização e redução de dimensionalidade. Esta etapa pode ser dividida em duas fases: processamento do sinal e extração de características. Na fase de processamento do sinal, os dados brutos são limpos para obter os sinais fisiológicos. Dessa forma, podem ser removidas as interferências causadas pela respiração do indivíduo, bem como filtradas as amplitudes causadas pelo uso de diferentes dispositivos na captura dos sinais. Na fase de extração de características, os dados processados na fase anterior, caso necessáriopodem ser segmentados em janelas, a partir das quais diferentes características numéricas são extraídas, como o tempo e características do

domínio de frequência (HSU et al., 2020). Além disso, técnicas de seleção de características, como análise de relevância, filtros e *wrappers*, podem ser empregadas para selecionar as características que são mais importantes para o problema, bem como para reduzir a sobrecarga e complexidade do modelo e o tempo de execução (MONTESINOS et al., 2019).

Aprendizado e avaliação de modelos: nesta etapa um ou mais modelos de aprendizado de máquina são treinados, seguido de avaliação do desempenho dos modelos usando um conjunto de teste. Várias técnicas podem ser utilizadas para avaliação, incluindo *cross-validation*, *leave-one-subject-out* (SCHMIDT et al., 2018a) e *leave-one-out* (MARKOVA; GANCHEV; KALINKOV, 2019). Métricas como acurácia, matriz de confusão, especificidade e sensibilidade são usadas para informar o desempenho do modelo nesta etapa. É importante lembrar que a avaliação deve ser feita com um conjunto de teste separado do conjunto de treinamento, a fim de evitar vazamento de dados.

Além da metodologia de reconhecimento de emoções supervisionada, descrita anteriormente, outra abordagem que pode ser utilizada para a tarefa é o aprendizado autossupervisionado. Nessa nova abordagem, grandes quantidades de dados não rotulados podem ser utilizadas para pré-treinar um modelo e reduzir a necessidade de ter grandes quantidades de dados rotulados para treinamento. A próxima seção descreve melhor os conceitos e a aplicação de aprendizado de representação autossupervisionado, com o objetivo de demonstrar as vantagens de desenvolver métodos aplicando essa nova abordagem.

2.6 Aprendizado Autossupervisionado

O desenvolvimento de novas abordagens de aprendizagem que não dependam de grandes quantidades de dados rotulados tem impulsionado a pesquisa em áreas como o aprendizado autossupervisionado. Essa metodologia tem a capacidade de aproveitar dados não rotulados para produzir representações generalistas que podem ser facilmente transferidas para problemas com limitações de dados rotulados. Ao contrário do aprendizado não supervisionado e semi-supervisionado, os métodos autossupervisionados criam sinais supervisionados artificialmente a partir de tarefas auxiliares e dados não-rotulados, e os utilizam para pré-treinar um modelo de rede neural profunda com o objetivo de criar representações generalizadas e independentes da tarefa alvo.

O aprendizado autossupervisionado é uma inovação do aprendizado não supervisionado, que tem sido aplicado com o objetivo de aprender representações de alto nível a partir de dados não rotulados por humanos. Nessa abordagem, o problema não supervisionado

é transformado em um problema supervisionado pela geração automática de rótulos. A geração de rótulos artificiais é baseada na criação de tarefas auxiliares, também chamadas de tarefas de pretexto. Essas tarefas auxiliares são utilizadas para aprender representações sobre os dados, mas tipicamente não são o verdadeiro propósito da aprendizagem. Por exemplo, prever a rotação de uma imagem (GIDARIS; SINGH; KOMODAKIS, 2018) ou prever uma palavra considerando as palavras ao redor (MIKOLOV *et al.*, 2013) são algumas tarefas comumente usadas em campos de visão computacional e processamento de linguagem natural, respectivamente.

Na área de sinais fisiológicos, por exemplo, tarefas auxiliares para aplicações podem incluir a predição de sequência, a diferenciação do sinal original da sua versão perturbada ou transformada. Transformações que podem ser realizadas sobre os sinais incluem a adição de ruídos, aplicação de um fator de escala ou inversão do sinal. Essas tarefas auxiliares são importantes para o aprendizado autossupervisionado, pois ajudam a criar representações generalizadas dos dados e a melhorar o desempenho do modelo em tarefas específicas.

O processo de aprendizado autossupervisionado pode ser dividido em duas etapas, como ilustrado na Figura 2.5. Na primeira etapa (**i**), o modelo é pré-treinado utilizando tarefas auxiliares que envolvem reconhecer transformações nos sinais. Um processo automatizado de geração de rótulos é definido com base nessas tarefas e a rede neural profunda é treinada para classificar qual transformação foi aplicada ao sinal original. Na segunda etapa (**ii**), o conhecimento (ou representação) adquirido na primeira etapa é utilizado para especializar o modelo para uma tarefa específica. No exemplo a seguir, a tarefa é a classificação de animais.

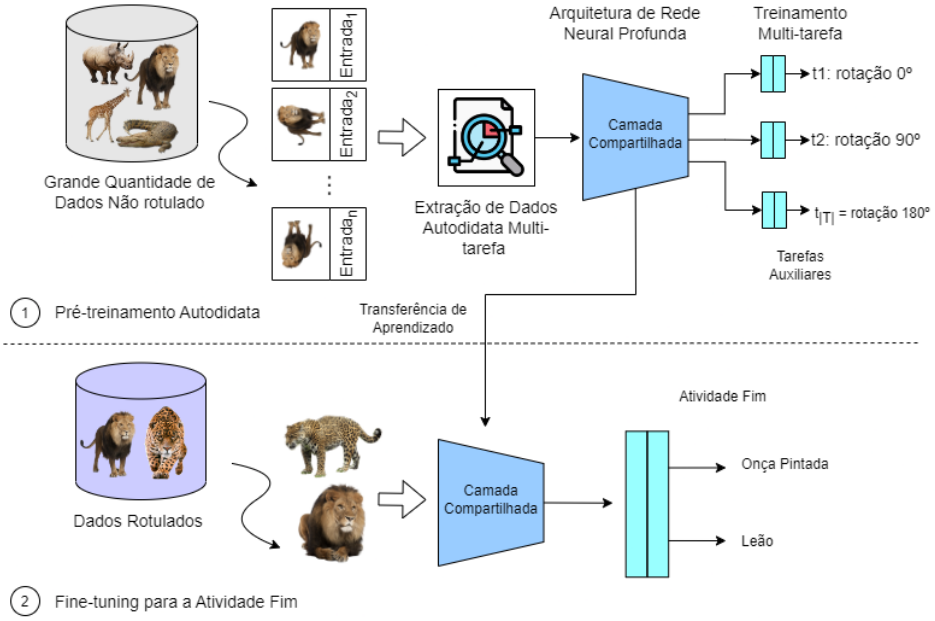

Figura 2.5 - Visão geral da abordagem autossupervisionada.

2.7 Considerações Finais

Neste capítulo abordamos a complexidade das emoções humanas e sua relação com as respostas físicas e fisiológicas. Discutimos duas perspectivas comuns de representação emocional: a discreta e a dimensional, e como a mensuração das emoções pode ser realizada ao observar seus componentes cognitivo, comportamental e fisiológico. Destacamos também as técnicas frequentemente utilizadas para inferir estados afetivos, suas vantagens e desvantagens, enfatizando que a escolha de uma técnica deve basear-se nas características específicas do estudo e do indivíduo.

3. TRABALHOS RELACIONADOS

Para a elaborar uma revisão bibliográfica, atributos comuns encontrados em trabalhos relacionados de aprendizado autossupervisionado foram identificados e utilizados para organizar este capítulo. Esses atributos incluem o tipo de sinal utilizado para o pré-treinamento ou treinamento do método (e.g., eletrocardiograma, eletroencefalograma) e a metodologia aplicada no pré-treinamento autossupervisionado para reconhecimento de emoções por meio de sinais fisiológicos (e.g., aprendizado contrastivo e transformações de sinais). Com base nesses atributos, os trabalhos relacionados foram categorizados em dois grupos: aqueles que usam aprendizado contrastivo e aqueles que utilizam transformações de sinais. Além disso, um resumo dos métodos propostos, resultados e contribuições desses trabalhos são apresentados nas seções seguintes.

3.1 Aprendizado autossupervisionado baseado em Aprendizado Contrastivo

O aprendizado contrastivo, em inglês, *contrastive learning*, é um tipo comum de aprendizado autossupervisionado que envolve treinar um modelo para aprender representações de alto nível dos dados comparando pares de amostras. Por exemplo, um modelo pode aprender a reconhecer diferentes visualizações de um objeto como sendo similares, mesmo que elas não sejam idênticas. Ao comparar pares de amostras similares e não-similares, o modelo aprende a reconhecer padrões, similaridades e outras características nos dados que podem ser úteis para tarefas de classificação e agrupamento. Na seção seguinte, apresentaremos trabalhos que utilizaram o aprendizado contrastivo com sinais fisiológicos, como eletrocardiogramas e eletroencefalogramas.

Zhang et al. (2022) propuseram uma abordagem para aprender representações do sinal do ECG com base em detecção temporal-espacial reversa (MTSRD - *Manipulated Temporal-Spatial Reverse Detection*). O framework MTSRD proposto consiste em dois módulos: um módulo de manipulação temporal e um módulo de manipulação espacial. O módulo de manipulação temporal envolve a inversão horizontal (reversão temporal) do sinal original e o módulo de manipulação espacial envolve a inversão vertical (reversão espacial) do sinal. O aprendizado é realizado pela classificação de quatro tipos de sinais, incluindo o sinal original. Os autores avaliaram a eficácia de sua abordagem proposta em várias bases de dados de referência e a compararam com outros métodos. Os resultados experimentais mostraram que o framework MTSRD proposto supera os métodos existentes e alcança um bom desempenho em

várias tarefas de classificação de ECG, incluindo a detecção de fibrilação atrial. Os autores também exploraram as representações do espaço de características e concluíram que a reversão temporal é mais eficaz para aprender representações de ECG do que a reversão espacial.

Lan et al. (2022) propuseram um modelo de aprendizado intra-inter-supervisionado para aprender representações gerais e personalizadas a partir de sinais cardíacos multivariados não rotulados. A complexa dinâmica temporal das arritmias cardíacas apresenta um desafio para esse tipo de aprendizado, que podem variar igualmente para o mesmo paciente (diferença intra-sujeitos) e entre diferentes pacientes (diferença entre-sujeitos). Experimentos em três bases de dados do mundo real demonstraram que o modelo intra-inter-supervisionado pré-treinado leva a uma melhoria de cerca de 10% em relação ao treinamento supervisionado, mesmo com apenas 1% dos dados rotulados disponíveis. Isso sugere uma forte generalização e robustez do modelo.

Wei et al. (2022) propuseram um novo framework de aprendizado de representações auto-supervisionada para ECG chamado CT-HB (*Contrastive Heartbeats*). O CT-HB utiliza uma técnica de aprendizado contrastivo, que consiste em aprender a diferença entre dois exemplos (um positivo e outro negativo) para obter representações gerais e robustas do sinal ECG. Para isso, os autores empregam uma técnica de amostragem para definir pares positivos e negativos de batimentos cardíacos para o aprendizado contrastivo, utilizando os padrões periódicos e significativos dos sinais de ECG. O CT-HB alcançou resultados superiores em comparação com outros métodos de aprendizado não supervisionado e autossupervisionado em diversas tarefas de predição de doenças cardíacas e conseguiu superar até mesmo métodos de aprendizado supervisionado em cenários com pouca quantidade de dados. O estudo mostra que essa abordagem pode ser útil para aplicações de ECG no mundo real, pois pode reduzir a necessidade de coletar grandes quantidades de dados rotulados.

Cheng et al. (2020) apresentaram um método baseado em aprendizado contrastivo para modelagem de sinais EEG e ECG. Esses tipos de sinais geralmente têm rótulos com ruído e um número limitado de sujeitos (menor que 100), o que torna o uso de aprendizado supervisionado ainda mais desafiador. Para lidar com esses desafios, a ideia de aprendizado específico do sujeito foi desenvolvida por meio de uma função de perda específica para cada sujeito e um treinamento adversarial para promover a invariância do sujeito durante o aprendizado autossupervisionado. Além disso, foram desenvolvidas técnicas de aumento de dados para séries temporais que podem ser usadas com a perda contrastiva. Os resultados mostram que a abordagem proposta é competitiva com os métodos supervisionados em relação à classificação, mesmo em condições de rotulagem limitada e número limitado de sujeitos. Os

experimentos também mostram que a invariância de sujeito melhora a qualidade das representações apreendidas e que a função de perda específica do sujeito melhora o desempenho do modelo quando os dados rotulados são usados para ajuste fino.

Alguns trabalhos utilizam o aprendizado contrastivo no campo da detecção dos estágios do sono. Xiao et al. (2021) apresentaram o método SleepDPC, um arcabouço de aprendizado autossupervisionado para classificação de estágios de sono a partir de sinais de EEG. O método utiliza técnicas de codificação preditiva contrastiva e codificação contrastiva discriminativa para extrair representações de alto nível dos sinais brutos de EEG. Na codificação preditiva contrastiva, o modelo é treinado para prever representações futuras dos dados em uma sequência, enquanto na codificação contrastiva discriminativa, o modelo é treinado para distinguir entre diferentes amostras de dados e maximizar a similaridade entre amostras pertencentes à mesma classe. Essas técnicas permitem extrair características relevantes dos dados em uma etapa de pré-treinamento autossupervisionado. Os autores avaliaram as representações aprendidas em duas bases de dados públicas e demonstraram que o método proposto aprendeu significativamente as representações, obtendo um desempenho superior em relação a vários métodos concorrentes, apesar do acesso limitado aos dados rotulados. Os resultados experimentais validam a eficácia do método SleepDPC em extrair características úteis dos sinais de EEG e destacam sua promissora aplicabilidade em outras áreas de séries temporais.

Jiang et al. (2021) propuseram um método de aprendizado de representação auto-supervisionada usando aprendizado contrastivo para classificação de estágios do sono a partir de sinais de EEG. Durante o treinamento, a rede foi estabelecida com uma tarefa auxiliar (*pretext task*) para combinar pares de transformações corretas geradas a partir dos sinais de EEG, melhorando assim a capacidade de representação dos sinais de EEG. Os experimentos foram conduzidos utilizando bases de dados com mais de 300.000 exemplos, incluindo Sleep-edf, Sleep-edfx, Dod-O e Dod-H. Os resultados empíricos no conjunto de dados Sleep-edf demonstraram que o método proposto apresenta desempenho competitivo no estágio de sono (88,16% de precisão e 81,96% de F1-Score) e verificaram a eficácia da estratégia autossupervisionadapara análise de sinais de EEG com poucos dados rotulados. Os autores sugerem que o método proposto tem potencial para ser aplicado em outras aplicações de EEG e em outros domínios.

Ye et al. (2022) propuseram um framework chamado CoSleep que tem a capacidade de aprender representações robustas e generalizáveis de sinais fisiológicos por meio de

aprendizagem auto-supervisionada para classificar as fases do sono. O CoSleep usa uma abordagem de aprendizado de máquina que utiliza múltiplos domínios dos dados para melhorar o desempenho do modelo. Cada domínio (tempo e frequência) dos dados é utilizado para extrair informações complementares a fim de identificar informações relevantes em um nível semântico de classe. Essas informações são combinadas para treinar o modelo. Adicionalmente, foi incluído um mecanismo de armazenamento em fila que captura uma sequência de amostras negativas e as utiliza como exemplos negativos na próxima iteração de treinamento, com o objetivo de melhorar a qualidade das representações apreendidas. Os experimentos realizados mostraram que o CoSleep obteve um desempenho superior em relação aos trabalhos na literatura relacionados à auto-aprendizagem, alcançando uma precisão de 71,6% e 57,9% em duas bases de dados de sono, SleepEDF e ISRUC, respectivamente.

Outros trabalhos utilizam o aprendizado contrastivo no campo de séries temporais e apresentam resultados promissores de forma geral, sem uma aplicação específica. Zang, X. et al. (2022) propuseram uma estratégia de pré-treinamento autossupervisionado em séries temporais, que utiliza a modelagem da Consistência de Tempo-Frequência (TF-C) para gerar representações temporais por meio de aprendizado contrastivo na dimensão temporal. A função de perda de consistência dedicada é utilizada para diminuir a distância entre as representações baseadas no tempo e na frequência, gerando sinais autossupervisionados para otimizar o modelo de pré-treinamento e capturar as relações latentes entre os domínios de tempo e frequência. As informações aprendidas são codificadas nos parâmetros do modelo e utilizadas para inicializar o modelo de ajuste fino e melhorar o desempenho em bases de dados de interesse. Os autores avaliaram o desempenho do modelo pré-treinado em oito bases de dados de séries temporais e os resultados mostram que a abordagem TF-C apresenta uma transferência de representação positiva, superando todos os baselines em uma média de 15,4% na pontuação F1 score. Essa técnica pode ser útil para melhorar o desempenho de modelos em outras aplicações de séries temporais, permitindo a captura de relações latentes entre diferentes domínios de sinais.

Zhang, Wenrui, et al. (2022) propuseram uma abordagem chamada CRT (*Cross Reconstruction Transformer*) para o aprendizado de representação não-supervisionado/autossupervisionado em séries temporais. Nessa abordagem usa-se uma transformação da série temporal para o domínio da frequência e a eliminação aleatória de certas partes na série temporal em ambos os domínios (tempo e frequência), maximizando a preservação do contexto global. Em seguida, os dados são reconstruídos em ambos os domínios usando uma arquitetura Transformer, chamada de *Dropped Temporal-Spectral Modeling*. Para

discriminar as representações no espaço de características global, os autores propuseram a constante chamada IDC (*Instance Discrimination Constraint*) para reduzir a informação mútua entre diferentes séries temporais e aprimorar as fronteiras de decisão. Além disso, uma estratégia de aprendizado que aumenta progressivamente a taxa de eliminação durante o processo de treinamento foi proposta para otimizar o CRT. Os resultados experimentais demonstram que o CRT alcança o melhor desempenho em comparação com os métodos existentes em várias bases de dados do mundo real, com melhorias de desempenho variando de 2% a 9%.

Kan et al. (2022) apresentaram o framework de SSL chamado SGMC (*Self-supervised Group Meiosis Contrastive learning*), baseado em estímulos consistentes dos sinais de EEG. Para aumentar a amostra de um grupo, os autores propuseram um método de aumento de dados inspirado em genética, chamado Meiose. Este método utiliza o alinhamento dos estímulos para aumentar as amostras de um grupo sem alterar as características dos estímulos, fornecendo amostras de grupo aumentadas para aprimorar o aprendizado construtivo. O SGMC utiliza um extrator de características de grupo capaz de examinar as relações entre as amostras de um grupo e identificar características compartilhadas a partir de amostras de EEG acionadas pelos mesmos estímulos emocionais em vídeos. A técnica de aprendizado contrastivo é empregada para maximizar a similaridade das representações em nível de grupo de grupos aumentados que possuem os mesmos estímulos. O SGMC alcançou excelentes resultados de reconhecimento de emoções na base de dados pública DEAP, com precisões de 94,72% e 95,68% nas dimensões de valência e excitação, respectivamente, além de um desempenho competitivo na base de dados pública SEED, com precisão de 94,04%. Os experimentos também mostraram que o SGMC manteve um bom desempenho mesmo com amostras de rótulos limitados.

3.2 SSL baseado em Transformações de Sinais

Nesta seção, descrevem-se os estudos que empregaram técnicas de transformação de sinais para aprender representações de sinais fisiológicos. Essa abordagem utiliza a transformação como uma forma de aumentar os dados de sinais rotulados. Os estudos são apresentados e categorizados de acordo com sua aplicação.

Xu et al. (2020) propuseram um método de detecção de anomalias em sinais de EEG, utilizando aprendizado autossupervisionado com uma rede neural convolucional (CNN). O método é baseado em transformações de escala aplicadas às sequências de EEG ao longo da dimensão temporal para criar dados auto-rotulados, permitindo treinar um classificador CNN

com os dados normais auto-rotulados. O objetivo é prever com precisão as transformações de escala em novos dados de EEG normais, mas não em dados anormais (epilépticos). Essa inconsistência entre as transformações de escala previstas e as transformações de escala verdadeiras é usada para detectar epilepsia. Este método é diferente de outros métodos de detecção de anomalias que exigem dados de EEG epilépticos, pois é baseado apenas em EEGs normais. Avaliações experimentais abrangentes mostraram que o método proposto superou métodos clássicos de detecção de anomalias, incluindo SVM e autocodificadores, e a robustez do método também foi comprovada empiricamente com diferentes estruturas de classificador e hiperparâmetros relevantes.

Banville et al. (2021) examinaram a auto-supervisão como uma abordagem geral para aprender representações de dados EEG, além de avaliar o desempenho do modelo autossupervisionado em comparação com outras abordagens encontradas na literatura. Os autores exploraram duas tarefas auxiliares baseadas em previsão de contexto temporal e codificação preditiva contrastiva. Eles realizaram experimentos em duas bases de dados e compararam com outras abordagens supervisionadas elaboradas. Os experimentos mostraram que os classificadores lineares, treinados utilizando os pesos obtidos pela rede auto-supervisionada superaram consistentemente as redes neurais profundas supervisionadas em cenários que há uma quantidade limitada de dados rotulados disponíveis para o treinamento do modelo, ao mesmo tempo que alcançaram desempenho competitivo quando todos os rótulos estavam disponíveis.

Soltanieh e Etemad (2022) investigaram a eficácia de várias técnicas de aumento de dados para o aprendizado autossupervisionado contrastivo de sinais de eletrocardiograma (ECG). O framework de auto-supervisão proposto consiste em duas etapas: o aprendizado contrastivo e a tarefa alvo. Na primeira etapa, um codificador é treinado usando várias técnicas de aumento de dados para extrair representações generalizáveis dos sinais de ECG. Em seguida, o codificador é congelado e algumas camadas lineares são ajustadas com quantidades diferentes de dados rotulados para a detecção de arritmias. Os experimentos foram realizados na base de dados PTB-XL, que consiste em um grande conjunto de dados de ECG de 12 derivações disponível publicamente. Os resultados indicam que a aplicação de técnicas de aumento de dados em uma faixa específica de complexidades é mais eficaz para o aprendizado autossupervisionado contrastivo.

Spathis et al. (2020) propuseram uma abordagem de aprendizagem de representação auto-supervisionada que utiliza sinais não rotulados de atividades físicas (acelerômetro) e

frequência cardíaca. Eles desenvolvem uma rede neural profunda que usa os dados de acelerômetro para prever o valor da frequência cardíaca. Após o pré-treinamento com as tarefas auxiliares, o modelo é avaliado em uma base de dados contendo sinais de acelerômetro e ECG. Os resultados dos experimentos indicam que as características aprendidas podem ser aplicadas em diversas tarefas, através da transferência de aprendizado com classificadores lineares. O modelo é capaz de capturar informações inerciais, que foram usadas para prever variáveis relacionadas à saúde, como aptidão física e características demográficas dos indivíduos (IMC, sexo, idade, aptidão física e gasto de energia), superando o desempenho de autocodificadores não supervisionados e o uso de características físicas e fisiológicas comuns.

Vazquez-Rodriguez et al. propuseram um modelo de reconhecimento de emoções baseado em transformador que processa o sinal de ECG. O modelo utiliza os mecanismos de atenção do transformador para construir representações contextualizadas, dando mais importância às partes relevantes do sinal. Essas representações são então alimentadas em uma rede totalmente conectada para prever as emoções. Para superar o problema de bases de dados com rótulos emocionais relativamente pequenos, os autores usaram a abordagem auto-supervisionada. Para isso, várias bases de dados de ECG não rotulados foram usadas para pré-treinar o modelo, que foi refinado para reconhecer emoções no conjunto de dados AMIGOS. Os experimentos mostraram que o modelo foi capaz de construir características robustas para prever a excitação e a valência no conjunto de dados AMIGOS, e apresentou resultados promissores em comparação com os métodos mais recentes.

Sarkar e Etemad (2020a) propuseram um *framework* de aprendizado profundo autossupervisionado e multitarefa para o reconhecimento de emoções a partir de eletrocardiograma (ECG). O processo de aprendizado é dividido em duas etapas: a) aprendizado de representações do ECG e b) aprendizado de classificação de emoções. Para aprender as representações do ECG, é utilizada uma rede de reconhecimento de transformação de sinais que aprende representações abstratas de alto nível a partir de dados não rotulados. Seis diferentes transformações de sinais são aplicadas aos sinais do ECG (adição de ruído, escalonamento, negação, inversão horizontal, permutação e deformação temporal), e o reconhecimento dessas transformações é usado como tarefas auxiliares. O treinamento do modelo com as tarefas auxiliares ajuda a rede a aprender representações espaço-temporais que generalizam bem em diferentes bases de dados e categorias de emoções. Em seguida, os pesos da rede auto-supervisionada são transferidos para uma rede de reconhecimento de emoções, onde as camadas convolucionais são mantidas congeladas e as camadas densas são treinadas com dados rotulados

do ECG. Os resultados experimentais, utilizando duas bases de dados públicas (SWELL e AMIGOS), mostram que a solução proposta apresenta um desempenho consideravelmente superior em comparação com uma rede treinada usando aprendizado totalmente supervisionado.

Em um outro trabalho, Sarkar e Etemad (2020b) propuseram a utilização de aprendizado autossupervisionado de várias tarefas para reconhecimento de emoções baseado em ECG. Foram utilizadas duas bases de dados públicas, o SWELL e o AMIGOS. Inicialmente, foram realizadas seis diferentes tarefas de transformação de sinal para treinar a rede com rótulos gerados automaticamente. Em seguida, os seis sinais transformados juntamente com os sinais originais foram utilizados para treinar uma rede neural convolucional de múltiplas tarefas de forma auto-supervisionada. A arquitetura da rede proposta consiste em três blocos convolucionais como camadas compartilhadas seguidas de duas camadas densas específicas para cada tarefa. O modelo pré-treinado foi então utilizado para classificação de emoções. Para isso, os pesos da rede pré-treinada foram transferidos para uma nova rede e treinada uma camada totalmente conectada, e o framework foi testado em duas bases de dados. A análise mostrou que o modelo autossupervisionado é melhor ou competitivo em comparação com a mesma rede treinada de forma totalmente supervisionada.

A Tabela 3.1 apresenta os trabalhos relacionados agrupados de acordo com o método, base de dados e tipo de aprendizado autossupervisionado utilizado e desempenho obtido.

Tabela 3.1 - Trabalhos relacionados agrupados por método, base de dados, tipo de aprendizado autossupervisionado e desempenho obtido

Ano	Referência	Métodos	Base de Dados	Tipo de self-learning	Métricas (%)				
2020	Cheng et al.	Encoder	MIT-BIH Arrhythmia	Contrastive learning	**Acurácia**				
					Intersubject		**Intrasubject**		
					Subj. specific	2 class.: 81.10	2 class.: 79.30		
						4 class.: 53.90	4 class.: 50.50		
			PhysioNet Motor Imagery		Subj. invariant	2 class.: 81.20	2 class.: 79.60		
						4 class.: 52.80	4 class.: 49.80		
	Sarkar e Etemad	CNN	AMIGOS	Signal Transformation	**Acurácia**				
					Arousal		**Valence**		
			SWELL		CNN com self-sup.	0.85	0.84		
	Xu et al.	CNN	The UPenn and Mayo Clinic's Seizure Detection Challenge	Signal Transformation	**AUC**				
					ResNet34	80.60			
	Spathis et al.	CNN	Privado	Não informado	**AUC**				
					Step2Heart (A/R/T)				
					PCA	**90%**	**95%**	**99%**	**99.9%**
					VO2max	68.30	67.80	68.00	68.20
					PAEE	78.20	79.20	80.60	79.70
					Height	70.30	74.00	80.50	81.30
					Weight	69.90	70.70	77.40	76.90
					Sex	76.20	81.50	91.10	93.40
					Age	61.10	63.80	67.30	67.60
					BMI	64.70	66.10	67.80	69.40
		CNN	AMIGOS		**Acurácia**				

<table>
<tr><td rowspan="5"></td><td rowspan="5">Sarkar e Etemad</td><td rowspan="5"></td><td></td><td rowspan="5">Signal Transformation</td><td>Dataset</td><td>Arousal</td><td>Valence</td><td>Estados afetivos</td><td>Estresse</td></tr>
<tr><td>WESAD</td><td>AMIGOS</td><td>88.00</td><td>87.00</td><td>-</td><td>-</td></tr>
<tr><td>DREAMER</td><td>WESAD</td><td>-</td><td>-</td><td>96.00</td><td>-</td></tr>
<tr><td rowspan="2">SWELL</td><td>DREAMER</td><td>85.00</td><td>85.00</td><td>-</td><td>-</td></tr>
<tr><td>SWELL</td><td>96.00</td><td>97.00</td><td>-</td><td>93.00</td></tr>
<tr><td rowspan="6">2021</td><td rowspan="3">Banville et al.</td><td rowspan="3">CNN</td><td rowspan="2">PC18</td><td rowspan="3">Signal Transformation</td><td colspan="5">Acurácia</td></tr>
<tr><td>Método</td><td colspan="2">2 classes</td><td colspan="2">32 classes</td></tr>
<tr><td>TUHab</td><td>PR e BT</td><td colspan="2">98.00</td><td colspan="2">-</td></tr>
<tr><td rowspan="3">Xiao et al.</td><td rowspan="3">Encoder</td><td rowspan="2">Sleep-EDF</td><td rowspan="3">Contrastive learning</td><td>CPC</td><td colspan="2">-</td><td colspan="2">95.40</td></tr>
<tr><td colspan="5">Acurácia</td></tr>
<tr><td>ISRUC</td><td>Método</td><td colspan="2">Sleep-EDF</td><td colspan="2">ISRUC</td></tr>
<tr><td rowspan="4">Jiang et al.</td><td rowspan="4">CNN</td><td>Sleep-edf</td><td rowspan="4">Signal Transformation</td><td>SleepDPC</td><td colspan="2">70.00</td><td colspan="2">53.00</td></tr>
<tr><td colspan="5">Acurácia</td></tr>
<tr><td>Sleep-edfx</td><td>Método</td><td>Dataset</td><td>Sujeitos</td><td colspan="2">Métricas gerais</td></tr>
<tr><td>Dod-O</td><td rowspan="2">Abordagem SSL</td><td>Sleep-edf</td><td>20 SC</td><td colspan="2">88.16</td></tr>
<tr><td>Dod-H</td><td>Sleep-edfx</td><td>78 SC + 23 ST</td><td colspan="2">84.42</td></tr>
<tr><td rowspan="6">2022</td><td rowspan="3">Ye et al.</td><td rowspan="3">Encoder</td><td rowspan="2">Sleep-edf</td><td rowspan="3">Contrastive learning</td><td colspan="5">Acurácia</td></tr>
<tr><td>Método</td><td colspan="2">Sleep-edf</td><td colspan="2">ISRUC</td></tr>
<tr><td>ISRUC</td><td>CoSleep</td><td colspan="2">71.00</td><td colspan="2">57.00</td></tr>
<tr><td rowspan="3">Kan et al.</td><td rowspan="3">Encoder</td><td rowspan="2">DEAP</td><td rowspan="3">Contrastive learning</td><td colspan="5">Acurácia</td></tr>
<tr><td rowspan="2">Método</td><td colspan="2">DEAP</td><td colspan="2">SEED</td></tr>
<tr><td rowspan="2">SEED</td><td>Valence</td><td>Arousal</td><td colspan="2">Porcentagem de rótulos</td></tr>
<tr><td>Totalmente Supervisionado</td><td>91.23</td><td>92.36</td><td>50%</td><td>100%</td></tr>
<tr><td>Zhang, X et al.</td><td>Encoder</td><td>Privado</td><td>Contrastive learning</td><td>Com fine-tuned</td><td>94.72</td><td>95.68</td><td>93.71</td><td>94.04</td></tr>
<tr><td colspan="5"></td><td>Modelo</td><td>Acurácia</td><td>Precisão</td><td>Recall</td><td>AUROC</td></tr>
<tr><td colspan="5"></td><td>TF-C</td><td>78.24</td><td>79.82</td><td>80.11</td><td>90.52</td></tr>
</table>

	Zhang, W[1]. et al.	Encoder	PTB-XL	Contrastive learning	**Método**	**Dataset**	**Acurácia**	**F1 - Score**	**ROC - AUC**
			HAR		CRT	PTB-XL	87.81	68.43	89.22
						HAR	90.09	90.51	98.94
			Sleep-EDF			Sleep-EDF	90.12	44.38	94.74
	Vazquez-Rodriguez et al.	Encoder	ASCERTAIN	Signal Transformation	**Modelo**	**F1 - Score**		**Acurácia**	
			DREAMER			**Arousal**	**Valence**	**Arousal**	**Valence**
			PsPM-FR		Transformer pré-treinado	87.00	83.00	88.00	83.00
			PsPM-RRM1-2						
			PsPM-VIS						
			AMIGOS						
	Zhang, W[2]. et al.	Encoder	CinC 2017	Contrastive learning	**Representações**	**Acurácia**	**Sensibilidade**	**AUC**	**Especificidade**
		FCNN			Espacial	70.00	63.00	72.10	70.90
					Temporal	83.40	81.00	90.04	83.80
					Espacial-temporal	83.10	83.00	90.08	83.10
	Xiang et al.	RNN	PTB-XL	Contrastive learning	**Dataset**	**Tamanho do embedding**		**AUROC**	
			Chapman		PTB-XL	128		91.50	
					Chapman			83.30	
			CPSC		CPSC			63.30	
	Wei et al.	CNN	MIT-BIH	Contrastive learning	**Método**	**Dataset**	**Acurácia**	**Macro AUROC**	
			LVH		CT-HB	MIT-BIH	95.74	95.01	
			Próprio			LVH	-	81.04	
	Soltanieh, S.; Etemad, A. e Hashemi, J.	CNN	PTB-XL	Signal Transformation	**Transformações**		**Acurácia**		
					Ruído Gaussiano		80.31		
					Escala		65.62		
					Permutação		84.73		
					Giro vertical		65.83		
					Giro horizontal		65.94		
					Máscara de zero		67.05		
					Time warping		81.18		

3.3 Considerações Finais

Neste capítulo discutiu-se o uso do aprendizado autossupervisionado na análise de sinais fisiológicos para a detecção de estados emocionais, área significativa de pesquisa na ciência afetiva. A organização dos estudos revisados por tipo de aprendizado permitiu a identificação de abordagens e métodos específicos adotados pelos pesquisadores, destacando o aprendizado contrastivo e o baseado em transformações de sinais como técnicas promissoras. Essas técnicas têm possibilitado uma análise objetiva dos estados emocionais, representando uma importante contribuição para a ciência afetiva.

A Tabela 1 apresenta algumas bases de dados recorrentes no treinamento e avaliação de modelos para reconhecimento de emoções utilizando o sinal de ECG. Alguns exemplos de bases de dados comuns incluem DREAMER, WESAD, SWELL e PTB-XL. Também é interessante notar que alguns trabalhos desenvolveram suas próprias bases de dados para o reconhecimento de emoções.

Quanto aos resultados, as redes neurais apresentaram resultados promissores nas tarefas de reconhecimento de emoções com base no sinal de ECG. No entanto, a dependência de dados rotulados para a execução dos experimentos é um problema comum nos estudos revisados. Em todos os trabalhos, a abordagem autosupervisionada é empregada para superar essa limitação e ajudar a melhorar a eficiência e a generalização dos modelos, tornando-os mais robustos a variações nos dados. Embora a abordagem autossupervisionada seja promissora, ainda há desafios a serem superados em sua implementação. Por exemplo, a escolha de uma representação adequada dos dados é crítica para o sucesso do modelo.

Além disso, a compreensão das emoções humanas ainda é um tema complexo e a relação entre estados emocionais e respostas corporais continua sendo um desafio para os pesquisadores. É necessário continuar a desenvolver novas técnicas e abordagens para a análise de sinais fisiológicos, a fim de avançar na compreensão dos estados emocionais e de seus efeitos sobre a mente e o comportamento humano. Portanto, este capítulo conclui que a utilização do aprendizado autossupervisionado na análise de sinais fisiológicos para a detecção de emoções constitui uma área de pesquisa crucial e em constante evolução, com potencial para trazer novas descobertas e insights para a ciência afetiva.

4. ARQUITETURA PARA O APRENDIZADO AUTOSSUPERVISIONADO NOS SINAIS DE ECG PARA O RECONHECIMENTO DE EMOÇÕES

Este capítulo apresenta a proposta de uma arquitetura baseada no aprendizado autossupervisionado para o reconhecimento de emoções a partir de sinais de eletrocardiograma (ECG). A abordagem visa minimizar a dependência de dados rotulados, aproveitando técnicas de pré-processamento, geração de modelos neurais profundos e classificação. A Figura 4.1 fornece uma visão geral do método, destacando cada etapa do processo. O método inclui:

1. Aquisição de Dados: O sinal de ECG é coletado a partir de sensores em indivíduos.
2. Pré-processamento: Os dados são submetidos a filtros para atenuação de ruídos, normalização e segmentação. Além disso, são aplicadas transformações no sinal, como adição de ruído, escala, inversão temporal, negação e permutação para criar novos dados e aumentar o tamanho da base de treinamento.
3. Geração do Modelo Neural Profundo: Um modelo neural profundo é treinado com os dados pré-processados para identificar padrões e classificar as transformações realizadas nos segmentos. O modelo consiste em três blocos de convolução, e as séries temporais são segmentos de tamanho fixo para lidar com entradas de comprimento variável.
4. Geração do Modelo Neural Profundo: Com os dados pré-processados, procede-se à geração do modelo neural profundo. Esta etapa utiliza uma arquitetura de rede neural, como *convolutional neural networks* (CNN), para aprender representações profundas dos dados de ECG. A rede é estruturada em camadas que incluem convoluções, ativações não-lineares, pooling e, finalmente, camadas densas que culminam em uma camada de saída projetada para a classificação de tarefas auxiliares/emoções.
5. Tarefa Alvo (Downstream Task): Consiste na classificação das emoções, identificando a que classe de emoções correspondem os valores preditos pelo modelo.

Em resumo, este processo envolve a aquisição de dados, pré-processamento, criação de um modelo neural profundo e classificação para o treinamento da rede neural com as tarefas auxiliares escolhidas utilizando dados não rotulados e ajuste fino da rede neural pré-treinada para a tarefa específica em questão utilizando dados rotulados. Nas próximas seções, serão

apresentados com mais detalhes o pré-processamento, as tarefas auxiliares utilizadas e a CNN auto-supervisionada que foram utilizadas neste trabalho. A fim de explicar melhor a arquitetura, ela foi dividida em duas seções distintas: a Seção 4.2 e a Seção 4.3.

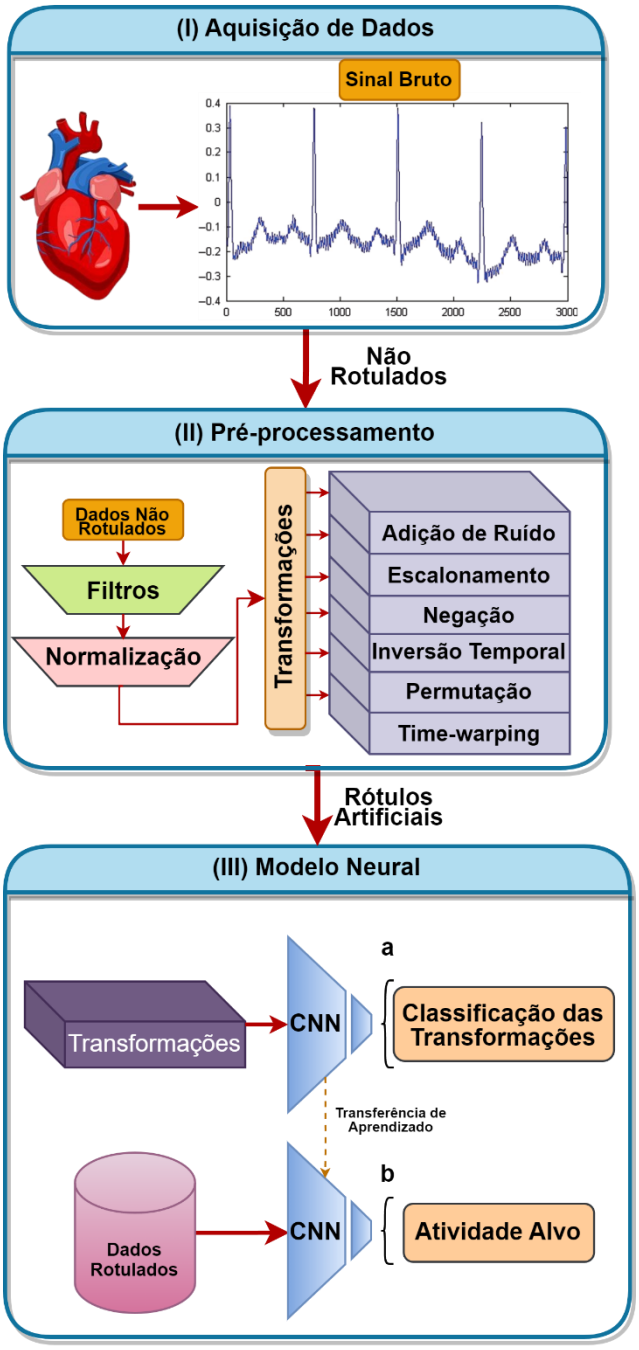

Figura 4.1 - Visão geral da arquitetura proposta para aprendizado autossupervisionado nos sinais de ECG para reconhecimento de emoções.

4.1 Pré-Processamento dos Dados

A coleta de dados pode ser um processo complexo, frequentemente envolvendo uma variedade de equipamentos e métodos. Em alguns casos, a utilização de diferentes hardwares pode resultar em sinais com propriedades espaço-temporais distintas, gerando variações e discrepâncias entre as amostras. Para mitigar esses efeitos, é comum realizar procedimentos de pré-processamento nos dados antes da análise.

Neste trabalho foram adotados dois procedimentos de pré-processamentos para reduzir as discrepâncias entre as amostras. Inicialmente, optou-se por remover as variações lentas ou constantes nos sinais que não estão diretamente relacionadas às informações de interesse. Isso pode incluir, por exemplo, flutuações devido a movimentos ou interferências externas que não refletem mudanças reais no sinal de interesse, como as oscilações naturais do corpo humano. Isso foi alcançado aplicando um filtro IIR (Infinite Impulse Response) de alta passagem, que remove os componentes de baixa frequência dos sinais, com uma frequência de passagem de banda de 0,8 Hz, eliminando assim frequências abaixo desse limiar.

O segundo procedimento adotado foi a normalização dos dados específica por usuário utilizando o z-score. Esse procedimento é útil ao trabalhar com dados de diferentes indivíduos, pois permite a comparação dos sinais levando em conta as características individuais de cada sujeito. A normalização z-score ajusta os dados para que tenham média zero e desvio padrão igual a um, permitindo a comparação dos valores de diferentes amostras com base em uma referência comum.

Após o pré-processamento, os sinais foram segmentados em janelas de 10 segundos, sem sobreposição. Essa medida foi tomada para garantir que cada segmento fosse independente e não houvesse vazamento de informações entre os dados de treinamento e de teste. Dessa forma, cada segmento de sinal pôde ser analisado individualmente, sem influência de informações de outros segmentos.

Para cada segmento de sinal, foram geradas transformações, conforme descrito na Tabela 4.1.

Tabela 4.1 - Descrição das tarefas auxiliares selecionadas para o modelo autossupervisionado.

Tarefas Auxiliares	Descrição	Parâmetros
Tarefas Genéricas		
Ruído	Adição de ruído em séries temporais é uma técnica comum de aumento de dados que pode ser usada para melhorar o desempenho do modelo de aprendizado de máquina. Essa técnica consiste em adicionar uma pequena quantidade de ruído aleatório ao sinal original para criar uma nova série temporal.	Relação sinal-ruído de 15
Escalonamento	Esta transformação aplica uma mudança nas magnitudes das amostras de sinal a partir da multiplicação de um valor escalar.	Fator de escala 1.1
Negação	A negação da série temporal é uma técnica simples de transformação de dados que envolve a inversão do sinal em uma série temporal. Essa transformação é realizada multiplicando-se todos os valores da série temporal por -1. Isso tem o efeito de inverter a direção do sinal, tornando os picos positivos em negativos e vice-versa.	-
Inversão Horizontal	Esta transformação aplica uma mudança na ordem temporal das amostras a partir de uma função de inversão a tempo.	-
Permutação	Esta transformação perturba aleatoriamente as amostras dentro de uma série temporal, cortando e trocando diferentes segmentos da série cronológica para gerar um novo um.	Vinte segmentos de permutação
Time Warping	Esta transformação estende ou deforma localmente as séries temporais, distorcendo suavemente os intervalos de tempo entre valores.	O fator de alongamento dos 20 segmentos time-warped é de 1,05.

4.2 Arquitetura CNN Auto-Supervisionada

A arquitetura da rede CNN auto-supervisionada, destinada a classificar emoções com base em sinais de ECG, é inspirada em um trabalho anterior de Pritam e Sarkar (2021). Ela é composta por duas etapas de aprendizado distintas. Na primeira etapa, a rede aprende a representação dos sinais de ECG, enquanto na segunda etapa, os pesos obtidos na etapa anterior são utilizados para treinar outra rede capaz de classificar as emoções. A estrutura da rede inclui três blocos convolucionais compartilhados e sete saídas, cada uma com duas camadas densas específicas para a tarefa.

Cada bloco de convolução consiste em duas camadas de convolução 1D com função de ativação ReLU, seguidas por uma camada de pooling máximo de tamanho 8. O número de

filtros aumenta de 32 para 64 e 128, à medida que a rede progride, enquanto o tamanho do núcleo é reduzido de 32 para 16 e 8, respectivamente. Após o último bloco de convolução, é realizada uma operação global de pooling máximo. Em seguida, a saída é conectada a duas camadas totalmente conectadas com 128 nós ocultos e função de ativação ReLU, com uma camada de dropout de 60% para evitar superajuste. A camada de saída do modelo é gerada por meio de uma função de ativação sigmoidal. A Tabela 4.2 apresenta as configurações da rede auto-supervisionada.

Tabela 4.2 - Especificação da rede neural multitarefa convolucional profunda que foi implementada para pré-treinamento autossupervisionado.

Camada	Especificação		Dimensões
Entrada	-		2560 x 1
Camadas Compartilhadas	Bloco Conv 1	2 x (Conv1D, 1 x 32, 32, ReLu)	2560 x 32
		Maxpool, 1 x 8, Stride 2	1277 x 32
	Bloco Conv 2	2 x (Conv1D, 1 x 16, 64, ReLu)	1277 x 64
		Maxpool, 1 x 8, Stride 2	635 x 64
	Bloco Conv 3	2 x (Conv1D, 1x 8, 128, ReLu)	635 x 128
		Global Max pooling	1 x 128
Camada de Tarefas Específicas	2 x (Dense, 128 units)		128
	7 x tarefas paralelas		
Saída	Sigmoid		2
	7 x saídas paralelas		

A saída da rede é uma função sigmoidal (0 ou 1), permitindo a identificação da classe de tarefa à qual o segmento de entrada pertence. Os pesos obtidos no treinamento são então transferidos para uma segunda rede dentro do mesmo domínio, acelerando o processo de treinamento e melhorando a capacidade de generalização.

A transferência de pesos é uma estratégia promissora para otimizar o treinamento de redes neurais. No caso da rede proposta para o reconhecimento de emoções a partir de sinais de ECG, a aprendizagem prévia realizada pela CNN autodidata na primeira etapa foi integralmente aproveitada no processo de transferência de pesos para a segunda rede. Isso permitiu que a segunda rede aprendesse a tarefa específica de reconhecimento de emoções baseada em uma

base de conhecimento prévia. O resultado final foi uma rede capaz de classificar emoções a partir de sinais de ECG, como será detalhado na próxima seção.

4.3 Arquitetura CNN para o Reconhecimento de Emoções

Durante a fase de transferência, as camadas convolucionais foram mantidas congeladas, enquanto as camadas densas foram treinadas com dados de eletrocardiograma previamente rotulados. Ao manter as camadas convolucionais congeladas durante a transferência, os padrões aprendidos foram preservados, permitindo à segunda rede adaptar-se a novos dados de entrada. Com essa abordagem de transferência de aprendizado, a rede será capaz de identificar emoções com base nos padrões de atividade elétrica do coração. A Tabela 4.3 apresenta as configurações da rede para o reconhecimento de emoções.

Tabela 4.3 Estrutura da camada neural convolucional da rede e parâmetros para um treinamento totalmente supervisionado bem como a transferência de ajustes de aprendizagem/fine tuning

Camada	Especificação	Dimensões
Entrada	-	2560 x 1
Camadas Compartilhadas	Bloco Conv 1	-
	Bloco Conv 2	-
	Bloco Conv 3	-
Camada Densa Reconhecimento de Emoções	2 x (Dense, 512 units)	512
Saída Reconhecimento de Emoções	Softmax	

A arquitetura da rede para o reconhecimento de emoções é composta por camadas convolucionais semelhantes às utilizadas pela rede que aprendeu as representações do ECG, seguidas por camadas totalmente conectadas com 512 nós ocultos. Durante a transferência de aprendizado, os pesos das camadas compartilhadas dessa rede multitarefa pré-treinada são congelados e transferidos para a rede de reconhecimento de emoções. No entanto, as camadas totalmente conectadas da rede de reconhecimento de emoções não são transferidas da rede multitarefa; ao invés disso, são treinadas usando o conjunto de dados rotulados para realizar o reconhecimento de emoções. Essa abordagem permite que a rede de reconhecimento de emoções se beneficie das representações aprendidas pela CNN multitarefa, enquanto se concentra especificamente na tarefa de reconhecimento de emoções.

4.4 Considerações Finais

Este capítulo apresentou uma visão detalhada da abordagem auto-supervisionada proposta para o reconhecimento de emoções, destacando suas diversas etapas bem definidas. Iniciamos com o pré-processamento dos dados, abrangendo desde a aplicação de filtros para atenuação de ruídos até a extração de características e transformações no sinal. Em seguida, exploramos a arquitetura da CNN auto-supervisionada, constituída por seus três blocos de convolução, responsáveis por extrair características e aprender representações espaciais do sinal de eletrocardiograma. Por fim, detalhamos a arquitetura da CNN para o reconhecimento de emoções, dedicada à identificação da classe de emoção correspondente aos valores preditos pelo modelo.

Como será demonstrado no próximo capítulo, combinação de tarefas auxiliares não apenas aprimora o aprendizado de representações eficazes, mas também facilita o treinamento subsequente do modelo para a classificação de emoções. Além disso, reconhecemos a importância do ajuste dos hiperparâmetros para a otimização dos resultados do modelo.

5. EXPERIMENTOS E RESULTADOS

Este capítulo detalha os experimentos conduzidos e os resultados alcançados pelo método proposto para o reconhecimento de emoções, bem como examina as tarefas auxiliares propostas para o aprendizado autossupervisionado. A seção inicial é dedicada à exposição do protocolo experimental, abrangendo a descrição da base de dados empregada, o método de validação utilizado e as métricas de desempenho escolhidas. Os resultados são divididos em três cenários: **(i)** análise do treinamento autossupervisionado com tarefas auxiliares combinadas; **(ii)** análise da variação do tamanho de filtros de kernel no treinamento autossupervisionado; **(iii)** análise da variação da quantidade dos dados rotulados utilizados para o reconhecimento de emoções. Conclusões sobre os resultados são discutidas, assim como a indicação de direções para trabalhos futuros nesta linha de pesquisa.

5.1 Protocolo Experimental

Esta seção oferece uma visão abrangente de como os experimentos foram concebidos, organizados e conduzidos para alcançar os objetivos estabelecidos. Os experimentos foram agrupados em três cenários distintos:

1. **Análise do treinamento autossupervisionado com tarefas auxiliares combinadas:** Neste cenário, investigamos o treinamento autossupervisionado utilizando diferentes combinações de transformações de sinal. Essas transformações incluem adição de ruído, escalonamento, negação, inversão horizontal, permutação e time-warping. O objetivo é comparar o impacto dessas tarefas auxiliares combinadas na classificação de emoções e identificar a combinação mais eficaz para maximizar o desempenho em métricas como acurácia e F1-Score.
2. **Análise da variação do número de filtros de kernel no treinamento autossupervisionado**:

   Neste cenário, investigamos o efeito das variações no número de filtros de kernel nos blocos de convolução durante o treinamento autossupervisionado. Inicialmente, analisamos o desempenho do modelo com diferentes configurações de números de filtros (16, 32 e 64) para os blocos de convolução. Em seguida, dobramos o tamanho dos filtros em cada configuração para avaliar o impacto dessas alterações na capacidade do modelo de aprender representações eficazes do sinal de ECG. Este experimento visa compreender como as configurações dos filtros de kernel influenciam o treinamento

autossupervisionado e, consequentemente, a capacidade do modelo de classificar emoções com precisão.

3. **Análise da variação da quantidade dos dados rotulados utilizados para o reconhecimento de emoções:**

   Este cenário avalia o impacto da variação da quantidade de dados rotulados no treinamento autossupervisionado para a tarefa final de reconhecimento de emoções. Exploramos diferentes proporções de dados rotulados para compreender como essas variações afetam os resultados do modelo na classificação de emoções.

As bases de dados AMIGOS, DREAMER, WESAD e SWELL foram utilizadas na construção de modelos para todos os cenários. Tais conjuntos de dados contêm sinais fisiológicos de eletrocardiograma capturados por sensores vestíveis. Os sinais foram ajustados uniformemente para uma frequência de 256 Hz e, em seguida, segmentados em janelas de 10 segundos para preparação das transformações subsequentes.

Após o janelamento, cada segmento de sinal das bases de dados anteriores foi submetido a seis transformação, visando o treinamento autossupervisionado dos modelos. As transformações aplicadas incluíram a adição de ruído, escalonamento, negação, inversão horizontal, permutação e time-warping. Como resultado, foi gerada uma base de dados ampliada, que incorporou tanto os sinais originais quanto suas variantes transformadas.

Com a base de dados ampliada, o particionamento foi realizado utilizando a estratégia de validação cruzada em 10 vias. Esta técnica envolve dividir a base de dados em dez partes iguais, utilizando nove delas para treinamento e uma para teste, repetindo esse processo dez vezes para garantir a robustez e a confiabilidade dos resultados. As métricas de desempenho escolhidas para avaliar os modelos incluíram acurácia e F1-Score, cujos detalhes são explorados nas seções subsequentes.

Em seguida, com a base de dados preparada e particionada, o treinamento autossupervisionado foi realizado. Esse treinamento consistiu na classificação dos segmentos de sinal conforme a transformação a que cada um havia sido submetido, com o intuito de ensinar o modelo a identificar e diferenciar as características de cada tipo de transformação aplicada.

Após a conclusão treinamento autossupervisionado, iniciou-se o segundo estágio, no qual os parâmetros – ou pesos – aprendidos anteriormente foram transferidos para um modelo de classificação de emoções. Este segundo modelo aproveitou o conhecimento prévio adquirido

para melhorar a precisão na identificação das emoções presentes nos sinais de eletrocardiograma.

Nas próximas seções deste capítulo, cada cenário de treinamento será detalhado, abordando as técnicas e processos utilizados e destacando como esses cenários contribuem para a robustez do modelo na tarefa de classificação de emoções.

5.1.1 Bases de Dados

Antes de detalharmos os experimentos conduzidos e os resultados obtidos, é essencial compreendermos as bases de dados que sustentaram este estudo. Cada conjunto de dados oferece informações únicas sobre as respostas emocionais humanas, capturadas através de uma variedade de estímulos e contextos. Nesta seção, examinamos de perto quatro bases de dados - AMIGOS, DREAMER, SWELL e WESAD - destacando suas características, metodologias de coleta de dados e os protocolos empregados para induzir diferentes estados emocionais nos participantes.

Base de Dados AMIGOS: A base de dados AMIGOS (Affect, Personality, and Mood in Groups and IndividualS) (MIRANDA-CORREA *et al.*, 2018) tem como objetivo investigar a personalidade, o humor e as respostas afetivas individuais por meio de sinais neurológicos e fisiológicos. O conjunto de dados é composto por dois experimentos distintos: um com vídeos curtos e outro com vídeos longos. No primeiro, 40 participantes assistiram a 16 clipes de vídeo curtos. No segundo, 37 participantes visualizaram quatro vídeos longos. Durante esses experimentos, foram capturados sinais fisiológicos como EEG, ECG e GSR, sendo os sinais de ECG registrados com o dispositivo Shimmer a uma frequência de amostragem de 256 Hz. Ao final de cada clipe de vídeo, os participantes responderam a questionários para avaliar seus estados afetivos em termos de excitação e valência, ambas em uma escala de 1 a 9. Dos 40 participantes apenas um foi removido, uma vez que os dados se encontravam incompletos, a Figura 5.1 apresenta os participantes e a quantidade instâncias de gravações coletadas.

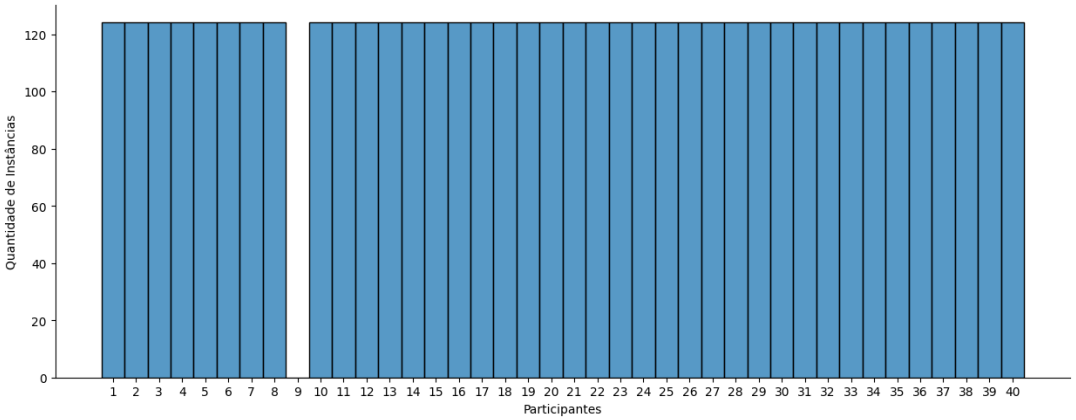

Figura 5.1 - Instâncias por participante na base de dados Amigos.

Quanto a quantidade de instâncias por escala na dimensão de excitação há a seguinte distribuição apresentada pela Figura 5.2.

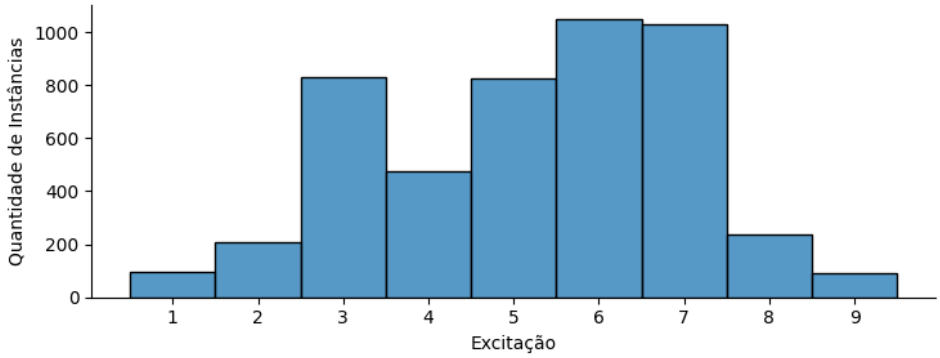

Figura 5.2 - Distribuição de instâncias por escalas na dimensão de excitação na base de dados Amigos.

A quantidade de instâncias por escala na dimensão de valência está distribuída conforme apresentado na Figura 5.3.

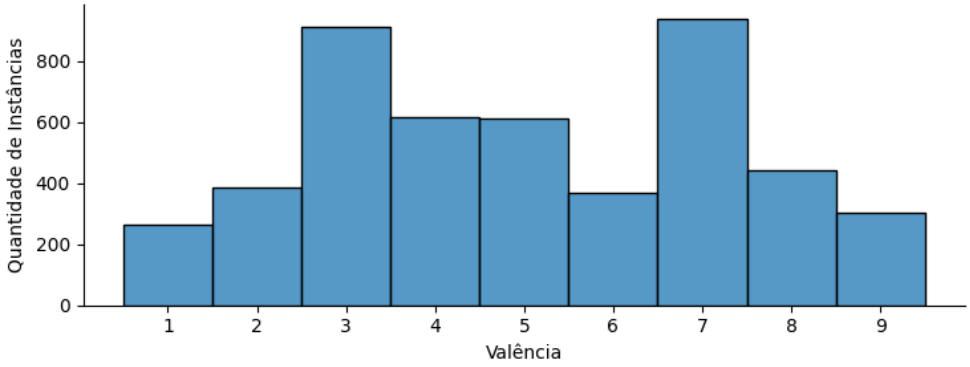

Figura 5.3 - Distribuição de instâncias por escalas na dimensão de valência na base de dados Amigos

Base de Dados DREAMER: A base de dados DREAMER (KATSIGIANNIS; RAMZAN, 2017) foi utilizada para estudar respostas emocionais induzidas por estímulos audiovisuais em 23 participantes, como apresentado na Figura 5.4. Estes foram expostos a diversos cortes de filmes selecionados para evocar emoções variadas como divertimento, excitação, alegria, calma, raiva, nojo, medo, tristeza e surpresa. Antes de cada sessão de filme, vídeos neutros eram apresentados para restabelecer um estado afetivo neutro nos participantes. Cada participante assistiu a 18 cortes de filmes, cada um com 60 segundos de duração, garantindo assim a exposição a um espectro amplo de emoções. Os sinais de ECG foram coletados utilizando o sensor Shimmer, com uma taxa de amostragem de 256 Hz. Após cada sessão de filme, os participantes avaliaram suas respostas emocionais em termos de excitação e valência, ambas em uma escala de 1 a 5, como apresentado na Figura 5.5.

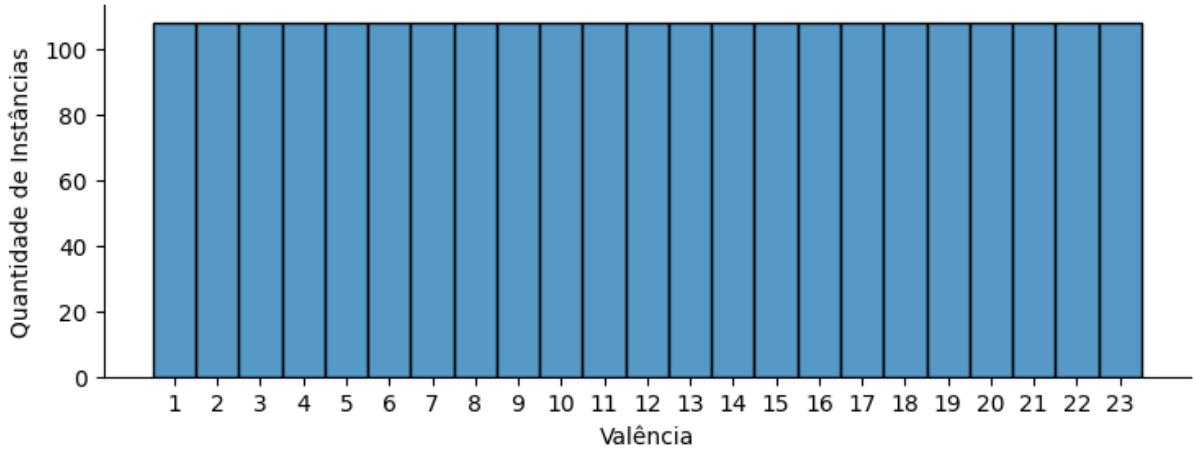

Figura 5.4 - Instâncias por participante na base de dados Dreamer

Quanto a quantidade de instâncias por escala na dimensão de excitação e valência há a seguinte distribuição apresentada pela Figura 5.5.

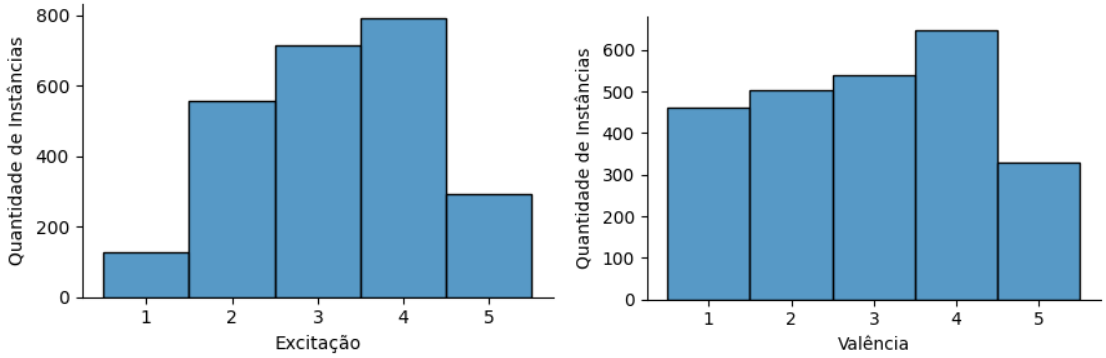

Figura 5.5 - Distribuição de instâncias por escalas na dimensão de excitação e valência na base de dados Dreamer

Base de Dados SWELL: A base de dados SWELL (KOLDIJK, 2018) foi desenvolvida para investigar o estresse em contextos de trabalho. Os 25 participantes foram submetidos a tarefas cotidianas do trabalho (redação de relatórios, realização de apresentações, leitura de e-mails e buscas na internet) sob diferentes condições de estresse, enquanto seus sinais de ECG eram registrados, a Figura 5.6 apresenta a quantidade instâncias coletadas por cada participante. Durante a execução dessas atividades, o ambiente de trabalho foi intencionalmente modificado para introduzir elementos estressantes, como interrupções frequentes por e-mails e pressões relacionadas ao tempo de conclusão das tarefas. O estudo focou em avaliar três estados afetivos distintos: um estado neutro, sem interrupções ou restrições de tempo; um estado de estresse induzido por limitações temporais (atividades a serem concluídas em 30 minutos); e um estado de estresse provocado por interrupções (recebimento de vários e-mails, alguns importantes e outros não). Os participantes avaliaram suas respostas emocionais após cada sessão de trabalho em termos de excitação e valência, ambas em uma escala de 1 a 9. Os sinais de ECG foram coletados usando o dispositivo Mobi TMSI a uma taxa de amostragem de 2048 Hz.

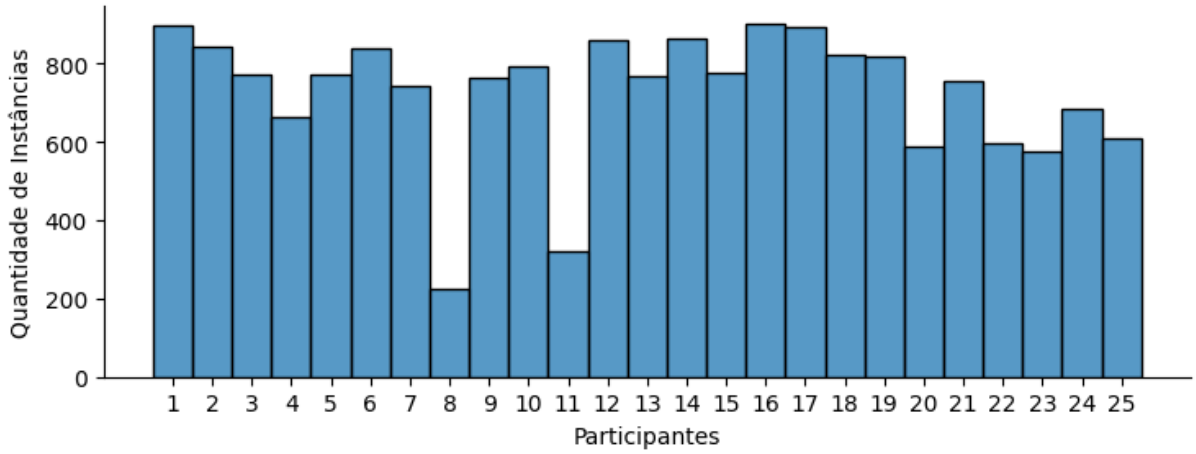

Figura 5.6 - Instâncias por participante na base de dados Swell.

Quanto a quantidade de instâncias por escala na dimensão de excitação há a seguinte distribuição apresentada pela Figura 5.7.

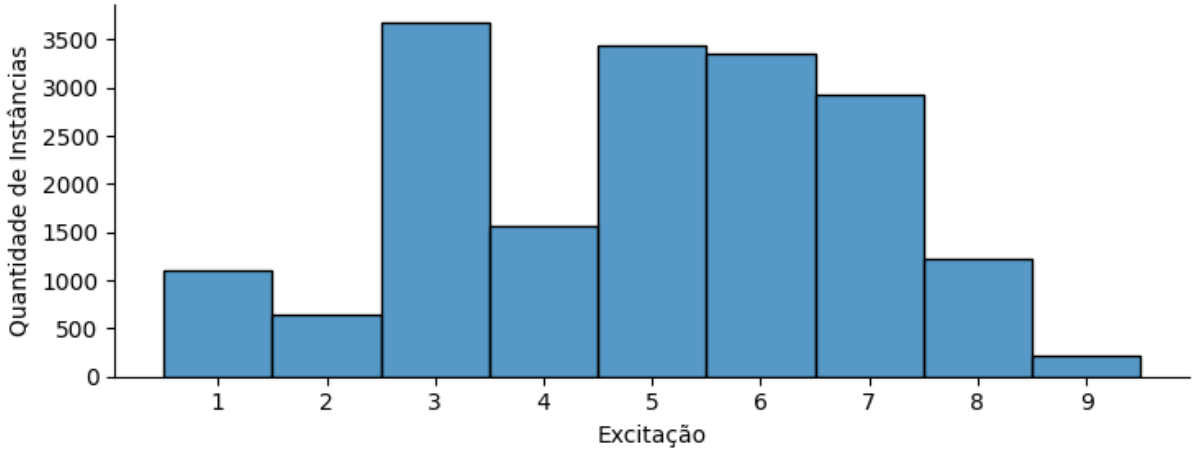

Figura 5.7 - Distribuição de instâncias por escalas na dimensão de excitação na base de dados Swell.

Quanto a quantidade de instâncias por escala na dimensão de valência há a seguinte distribuição apresentada pela Figura 5.8.

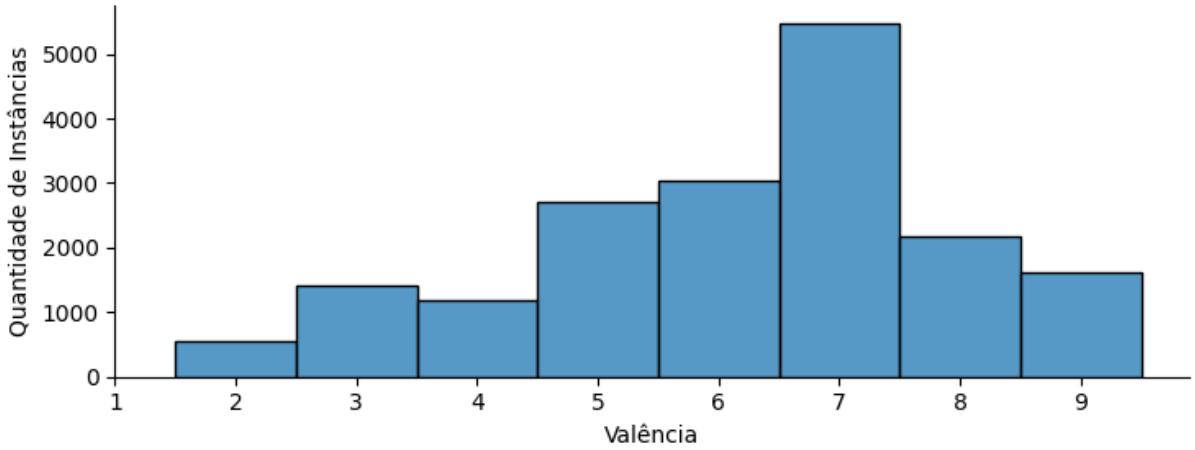

Figura 5.8 - Distribuição de instâncias por escalas na dimensão de valência na base de dados Swell.

Base de Dados WESAD: Na base de dados WESAD (SCHMIDT et al., 2018a), sinais de ECG foram coletados de 15 participantes, como apresentado pela Figura 5.9, submetidos a um protocolo que induzia estados de estresse, diversão e neutralidade. Técnicas comuns como exposição a conteúdo de vídeos, sons ou tarefas sob pressão foram utilizadas. A avaliação das emoções foi feita por aplicação de um questionário, utilizando conceitos de estresse (em uma escala de 1 a 4). Para capturar as reações fisiológicas, os sinais de ECG foram coletados pelo dispositivo RespiBAN com uma taxa de amostragem de 700 Hz.

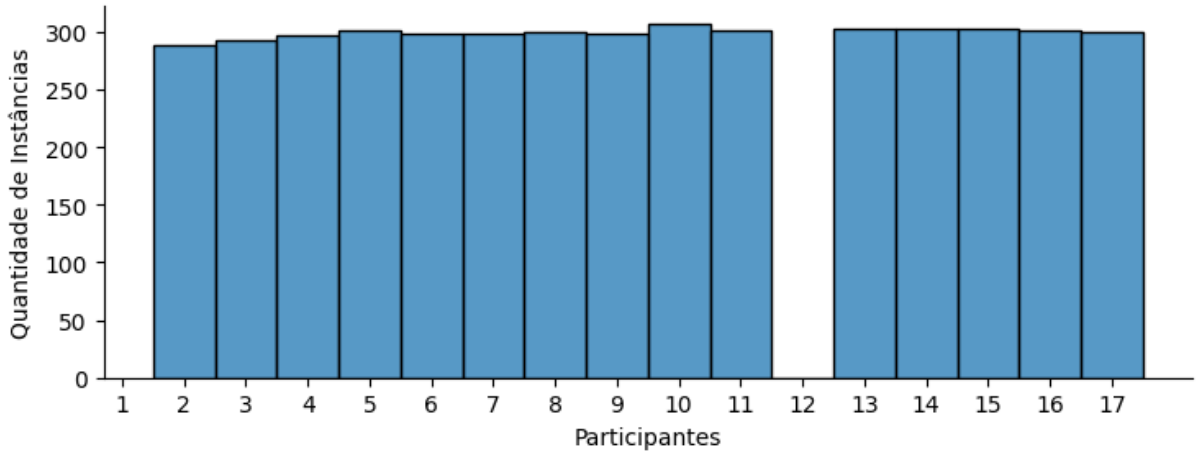

Figura 5.9 - Instâncias por participante na base de dados Wesad

Quanto a quantidade de instâncias por níveis de estresse há a seguinte distribuição apresentada pela Figura 5.10.

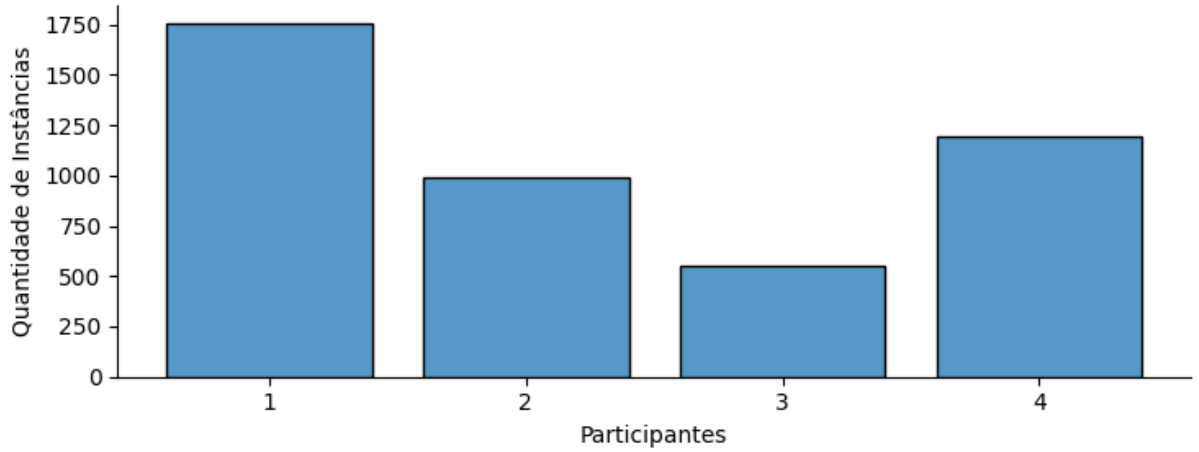

Figura 5.10 - Distribuição de instâncias por níveis do estresse na base de dados Wesad

5.1.2 Estratégia de Particionamento

Nos experimentos, foi utilizada a técnica de divisão dos dados em dez conjuntos de treino e teste, conhecida como *10-fold cross-validation*. Essa técnica consiste em dividir o conjunto de dados em *k* partes iguais, chamadas de *'folds'*. O modelo é treinado em *k-1 folds* e avaliado no *fold* restante, usado como conjunto de teste. Esse processo é repetido *k* vezes, com cada uma das *k folds* servindo alternadamente como conjunto de teste.

O resultado de cada *fold* fornece uma amostra do desempenho do modelo. A média desses resultados oferece uma estimativa geral do desempenho do modelo. Essa abordagem apresenta várias vantagens em relação à divisão tradicional de treino/teste, pois reduz a variação da avaliação e fornece uma estimativa mais confiável do desempenho do modelo. Além disso,

a técnica de *10-fold cross-validation* faz uso mais eficiente dos dados disponíveis, uma vez que cada exemplo é utilizado tanto para treinamento quanto para teste.

5.1.3 Métricas de Avaliação

As métricas de acurácia e F1-Score foram utilizadas para avaliar o desempenho do método proposto. Essas métricas são derivadas da matriz de confusão, que é comumente usada para apresentar os valores corretamente classificados pelo modelo. Os verdadeiros positivos (VP) ocorrem quando a classe verdadeira é classificada corretamente. Além disso, a matriz de confusão indica onde o modelo pode estar se confundindo, como nos casos de falsos positivos (FP), quando um caso falso é classificado como verdadeiro, conforme mostrado na Tabela 5.1.

Tabela 5.1 - Exemplo de matriz de confusão para um problema de n classes.

		Classe Predita			
		1	**2**	**N**	**Soma**
Classe Real	**1**	VP_{11}	FP_{12}	FP_{1N}	SI_1
	2	FP_{21}	VP_{22}	FP_{2N}	SI_2
	N	FP_{N1}	FP_{N2}	VP_{NN}	SI_N
	Soma	ST_1	ST_2	ST_N	total

Para facilitar a compreensão dos cálculos, algumas variáveis são apresentadas previamente. C representa a quantidade de classes existentes no problema em questão; VPii indica o número de exemplos de teste pertencentes à classe i que foram corretamente classificadas com o rótulo i; FNij representa o total de amostras de teste da classe i que foram erroneamente classificadas com o rótulo j; SIi refere-se à quantidade de amostras de teste pertencentes à classe i; e STj denota o número total de amostras de teste que foram classificadas com o rótulo j. As duas últimas variáveis podem ser determinadas pelas Equação 5.1 e 5.2 a seguir:

Equação 5.1 - Quantidade de amostras de teste da classe i

$$\boldsymbol{SI_i} = VP_{ii} + \sum_{j=1, j\neq i}^{C} FP_{ij}$$

Equação 5.2 - Quantidade de amostras de teste classificadas como j

$$\boldsymbol{ST_j} = VP_{jj} + \sum_{i=1, i\neq j}^{C} FP_{ij}$$

Os resultados foram obtidos extraindo as seguintes métricas de desempenho:

Acurácia: avalia a taxa de acerto do classificador, que corresponde à probabilidade de uma instância de teste ser classificada corretamente. É calculada pela razão entre o número de instâncias corretamente classificadas e o total de instâncias classificadas, conforme Equação 5.3 a seguir:

Equação 5.3 - Acurácia

$$\text{Acurácia} = \frac{\sum_{i=1}^{C} VP_{ii}}{total} = \frac{\sum_{i=1}^{C} VP_{ii}}{\sum_{i=1}^{C} SI_i} = \frac{\sum_{j=1}^{C} VP_{jj}}{\sum_{j=1}^{C} ST_j}$$

Precisão: é definida como a proporção de exemplos positivos classificados corretamente em relação ao total de exemplos classificados como positivos. Conforme a Equação 5.4:

Equação 5.4 - Precisão

$$\text{Precisão}_i = \frac{VP_{ii}}{VP_{ii} + FP_{ij}}$$

Revocação: também conhecida como *recall*, é uma métrica de avaliação de modelos de classificação que mede a capacidade do modelo em detectar corretamente todas as amostras positivas de uma determinada classe. Em outras palavras, ela avalia a proporção de instâncias positivas que foram corretamente identificadas pelo modelo. A revocação é calculada pela divisão do número de verdadeiros positivos pela soma do número de verdadeiros positivos e falsos negativos. É uma medida importante para problemas em que os falsos negativos são mais críticos do que os falsos positivos, como em diagnósticos médicos.

De forma mais técnica, a revocação pode ser definida como a média ponderada da fração dos rótulos verdadeiros que foram corretamente classificados para cada classe de atividade. Matematicamente, a revocação é dada pela seguinte Equação 5.5:

Equação 5.5 - Revocação

$$\text{Revocação}_i = \frac{VP_{ii}}{SI_i}$$

Medida-F (do inglês F1 Score)**:** é uma métrica que combina a precisão e a revocação para avaliar o desempenho geral do modelo de classificação. É particularmente útil em casos de desbalanceamento de dados, pois pondera tanto a precisão quanto a revocação de maneira equilibrada. A Medida-F é calculada pela média harmônica entre a precisão e a revocação, conforme a

Equação 5.*6*:

Equação 5.6 - Medida-F

$$\text{Medida-F} = 2 * \frac{Precisão_i * Revocação_i}{Precisão_i + Revocação_i}$$

onde, n é o número total de amostras, $\mathbf{y}_j$ é a previsão do modelo para uma amostra e $\hat{\mathbf{y}}_j$ é o valor real correspondente a essa amostra.

5.2 Resultados

A seguir serão apresentadas a análise do treinamento autossupervisionado com tarefas auxiliares combinadas (5.2.1), análise da variação de neurônios no treinamento autossupervisionado (5.2.2) e análise da variação da dimensionalidade dos dados rotulados no treinamento autossupervisionado no reconhecimento de emoções (5.2.3). As informações detalhadas acerca do desempenho desses modelos serão fornecidas a fim de avaliar o progresso de cada um deles. Tabelas contendo informações sobre a acurácia e F1-score dos modelos serão apresentadas, o que permitirá a identificação de possíveis melhorias e otimizações do desempenho dos modelos na tarefa em questão.

5.2.1 Análise do Treinamento Autossupervisionado Utilizando Combinação de Tarefas Auxiliares

Neste cenário foram conduzidos treinamentos autossupervisionados, empregando uma variedade de transformações como tarefas auxiliares. O objetivo principal dessa abordagem foi explorar os efeitos potenciais resultantes do uso de diferentes conjuntos de tarefas.

A Tabela 5.2 oferece uma visão geral do impacto de cada combinação na acurácia. O Grupo 6, que inclui sete das tarefas auxiliares, demonstrou o melhor desempenho entre as configurações testadas para o treinamento autossupervisionado. É importante destacar que o treinamento de cada modelo autossupervisionado utilizou a base de dados ampliada, mencionada na Seção 5.1 do protocolo experimental, composta por amostras do sinal original e dos sinais transformados.

Tabela 5.2 - Tarefas auxiliares combinadas no aprendizado autossupervisionado.

Experimento	Tarefas Auxiliares — Acurácia						
	Sinal Original	**Adição de Ruído**	**Escalonamento**	**Negação**	**Inversão Horizontal**	**Permutação**	**Time Warping**
Grupo 1	96,06% ± 2,40	99,52% ± 0,25	-	-	-	-	-
Grupo 2	96,93% ± 1,95	99,55% ± 0,22	97,29% ± 1,90	-	-	-	-
Grupo 3	96,94% ± 1,90	99,56% ± 0,20	97,98% ± 1,81	99,99% ± 0,01	-	-	-
Grupo 4	97,00% ± 1,78	99,56% ± 0,25	97,32% ± 1,76	99,99% ± 0,01	99,95% ± 0,05	-	-
Grupo 5	96,81% ± 2,17	99,57% ± 0,43	97,17% ± 2,08	100% ± 0,03	99,92% ± 0,47	99,66% ± 0,68	-
Grupo 6	97,04% ± 2,28	96,64% ± 0,21	97,27% ± 2,20	100% ± 0,01	99,95% ± 0,25	99,74% ± 0,61	99,52% ± 1,56

Acurácia Média: 98,55 ± 1,39

Os diferentes conjuntos de dados de tarefas auxiliares também foram avaliados no domínio alvo (classificação de emoções). Para isso, foram transferidos os pesos de cada modelo e empregados no treinamento dos modelos utilizando diferentes bases de dados, incluindo AMIGOS (9 escalas), DREAMER (5 escalas), SWELL (9 escalas) e WESAD (5 níveis). As tabelas 5.3, 5.4, 5.5 e 5.6 apresentam uma análise detalhada dos resultados obtidos.

Na base de dados AMIGOS, que categoriza emoções em nove classes distintas, para a dimensão de Excitação: o Grupo 2 obteve o melhor desempenho, alcançando uma acurácia de 72,58% e um F1 Score de 72,65%, seguido pelo Grupo 5 com 71,65% de acurácia e 71,69% de F1 Score. A média geral de acurácia para a dimensão de excitação foi de 71,26%, com um F1 Score médio de 71,49%, conforme mostrado na Tabela 5.3. Para a dimensão de Valência, o Grupo 4 apresentou o melhor desempenho com uma acurácia 64,14% de acurácia e 63,02% de F1 Score. A média geral de acurácia para a dimensão de valência foi de 63,32%, com um F1 Score médio de 62,35%.

Tabela 5.3 - Resultado da classificação de emoções em excitação e valência para a base de dados AMIGOS.

AMIGOS (9 Escalas)					
Excitação			**Valência**		
Grupo	**Acurácia**	**F1 Score**	**Grupo**	**Acurácia**	**F1 Score**
1	70,50%	70,62%	**1**	63,17%	61,86%
2	72,58%	72,65%	**2**	62,93%	61,91%
3	71,06%	71,13%	**3**	63,48%	62,57%
4	70,64%	70,71%	**4**	64,14%	63,02%
5	71,65%	71,69%	**5**	63,79%	63,07%
6	71,12%	71,16%	**6**	62,42%	61,68%
Média	71,26%	71,49%	**Média**	63,32%	62,35%

Para a base de dados DREAMER, que foca na classificação de emoções em cinco categorias distintas, refletem variações no desempenho entre os grupos tanto para excitação quanto para valência. Observa-se que, para a dimensão de excitação, o Grupo 5 apresentou o melhor desempenho com uma acurácia de 77,06% e um F1 Score de 77,00%, enquanto o Grupo 1 registrou o desempenho mais baixo, com 72,87% de acurácia e 72,63% de F1 Score. A média de desempenho para excitação ficou em 75,39% de acurácia e 75,33% de F1 Score, indicando um resultado razoável como apresentada pela Tabela 5.4.

Tabela 5.4 - Resultado da classificação de emoções em excitação e valência para a base de dados DREAMER.

DREAMER (5 Escalas)					
Excitação			**Valência**		
Grupo	**Acurácia**	**F1 Score**	**Grupo**	**Acurácia**	**F1 Score**
1	72,87%	72,63%	**1**	71,42%	69,65%
2	75,56%	75,41%	**2**	73,43%	71,99%
3	75,97%	75,69%	**3**	72,70%	71,57%
4	76,82%	76,56%	**4**	74,36%	73,17%
5	77,06%	77,00%	**5**	73,15%	72,11%
6	74,03%	73,70%	**6**	72,94%	71,63%
Média	75,39%	75,33%	**Média**	73,00%	71,69%

Na dimensão de valência, os resultados foram mais uniformes. O Grupo 4 alcançou o melhor desempenho com uma acurácia de 74,36% e um F1 Score de 73,17%, enquanto o Grupo 1 novamente mostrou o menor desempenho, com acurácia de 71,42% e F1 Score de 69,65%. A média para valência foi de 73,00% de acurácia e 71,69% de F1 Score. Esses resultados são consistentemente inferiores aos da excitação, sugerindo que a dimensão de valência pode ser mais desafiadora para classificar com precisão nesta base de dados.

Os resultados para a base de dados SWELL, que analisa a classificação de emoções em nove categorias, mostram um alto nível de desempenho nas dimensões de excitação e valência, com algumas variações notáveis entre os grupos. Para excitação, o Grupo 5 obteve o melhor desempenho, alcançando 95,68% de acurácia e 95,60% de F1 Score. Por outro lado, o Grupo 1 apresentou o menor desempenho, com 92,82% de acurácia e 92,74% de F1 Score, como apresentado na Tabela 5.5.

Tabela 5.5 - Resultado da classificação de emoções em excitação e valência para a base de dados SWELL.

SWELL (9 Escalas)					
Excitação			**Valência**		
Grupo	**Acurácia**	**F1 Score**	**Grupo**	**Acurácia**	**F1 Score**
1	92,82%	92,74%	**1**	89,56%	89,61%
2	94,39%	94,30%	**2**	92,01%	92,02%
3	94,49%	94,39%	**3**	91,96%	92,00%
4	95,29%	95,20%	**4**	92,68%	92,70%
5	95,68%	95,60%	**5**	93,18%	93,20%
6	95,47%	95,38%	**6**	93,15%	93,16%
Média	94,69%	94,60%	**Média**	92,09%	92,12%

Quanto na dimensão de valência, os resultados foram menores quanto comparado ao de excitação. O Grupo 5 alcançou o melhor desempenho com uma acurácia de 93,18% e um F1 Score de 93,20%, enquanto o Grupo 1 novamente mostrou o menor desempenho, com acurácia

de 89,56% e F1 Score de 89,61%. A média para valência foi de 92,09% de acurácia e 92,12% de F1 Score.

A análise dos resultados de classificação de estresse na base de dados WESAD para cinco classes revela um bom desempenho, quando comparada às bases de dados AMIGOS e DREAMER. Os Grupos 5 e 4 mostraram os melhores desempenhos, com o Grupo 5 alcançando a maior acurácia de 93,96% e o maior F1 Score de 92,75%. Essa performance elevada indica uma eficácia na classificação de estresse usando o modelo treinado. Em contraste, o Grupo 1 apresentou a menor acurácia e F1 Score, com 91,43% e 89,94%, respectivamente. A média de acurácia entre todos os grupos foi de 92,87%, e a média do F1 Score foi de 91,50%, refletindo um desempenho geral elevado do modelo na tarefa de classificação de estresse, como apresentado pela Tabela 5.6.

Tabela 5.6 - **Resultado da classificação de emoções em excitação e valência para a base de dados SWELL.**

WESAD (4 Níveis)		
Estresse		
Grupo	**Acurácia**	**F1 Score**
1	91,43%	89,94%
2	92,60%	91,27%
3	92,83%	91,33%
4	93,05%	91,76%
5	93,96%	92,75%
6	93,36%	91,97%
Média	92,87%	91,50%

A diversidade de transformações incorporadas nas tarefas auxiliares do Grupo 5, como adição de ruído, negação e permutação, foi determinante para a contribuição observada nos resultados. Essas transformações induzem o modelo a aprender a partir das variações nos dados, promovendo uma generalização aprimorada ao reconhecer as características do sinal de eletrocardiograma (ECG) associadas a diferentes tipos de emoções. Os resultados obtidos pelo Grupo 5 sugerem que a eficácia do aprendizado autossupervisionado em tarefas de classificação de emoções pode ser atribuída à capacidade do modelo de generalizar a partir de representações mais robustas, desenvolvidas através dos desafios impostos pelas tarefas auxiliares.

5.2.2 Análise da Variação do Número de Filtros de Kernel no Treinamento Autossupervisionado

Neste estudo, investigamos o impacto da variação do número de filtros de kernel nos blocos de convolução sobre o aprendizado do modelo, alterando assim sua capacidade de

capturar características dos sinais de ECG. Para isso, foram conduzidos três experimentos com diferentes configurações de filtros.

- Configuração de Filtros 1: 16, 32 e 64 filtros nos blocos 1, 2 e 3, respectivamente.
- Configuração de Filtros 2: 32, 64 e 128 filtros nos blocos 1, 2 e 3, respectivamente.
- Configuração de Filtros 3: 64, 128 e 256 filtros nos blocos 1, 2 e 3, respectivamente.

Para cada experimento, realizamos um treinamento autossupervisionado utilizando o melhor grupo de tarefas auxiliares identificado na seção anterior, o Grupo 5. Em seguida, cada modelo autossupervisionado foi utilizado para treinar a classificação de emoções, reutilizando os pesos do modelo autossupervisionado, conforme mostrado na Figura 5.11.

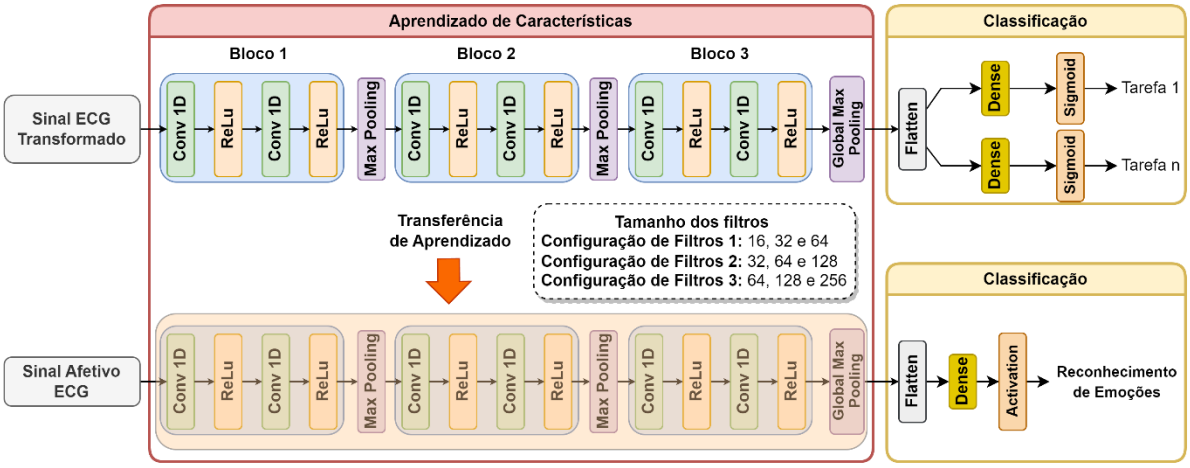

Figura 5.11 - Arquitetura implementada para o treinamento autossupervisionado e para classificação de emoções.

A variação do número de filtros de kernel revela como mudanças nos hiperparâmetros da arquitetura podem impactar a capacidade do modelo em desenvolver representações de dados generalizáveis e robustas. Esse comportamento é explorado mais a fundo ao se analisar o desempenho do modelo nas variações dos três experimentos.

Os resultados obtidos desta investigação derivam de um conjunto diversificado de bases de dados, incluindo AMIGOS, DREAMER, SWELL e WESAD, e são validados por meio de validação cruzada de 10 folds. Essa abordagem permite uma compreensão mais aprofundada sobre como a escolha do número de filtros de kernel para os blocos de convolução pode ser crucial para melhorar a aprendizagem auto-supervisionada e sua aplicação na análise de emoções.

Ao analisar os resultados para a base de dados AMIGOS, observou-se que na atividade de classificar emoções contendo 9 classes, a configuração de filtros 1 obteve acurácia de

62,60%. Este valor aumentou significativamente para 71,65%, com um F1 Score de 71,69%, além do menor desvio padrão (1,64% e 1,61%, respectivamente) quando utilizada a configuração de filtros 2. Contudo, ao utilizar a configuração de filtros 3, houve uma tendência de diminuição da performance e aumento do desvio padrão, sugerindo que números maiores de filtros não trazem benefícios adicionais e podem até prejudicar o desempenho do modelo, como apresentado na Figura 5.12.

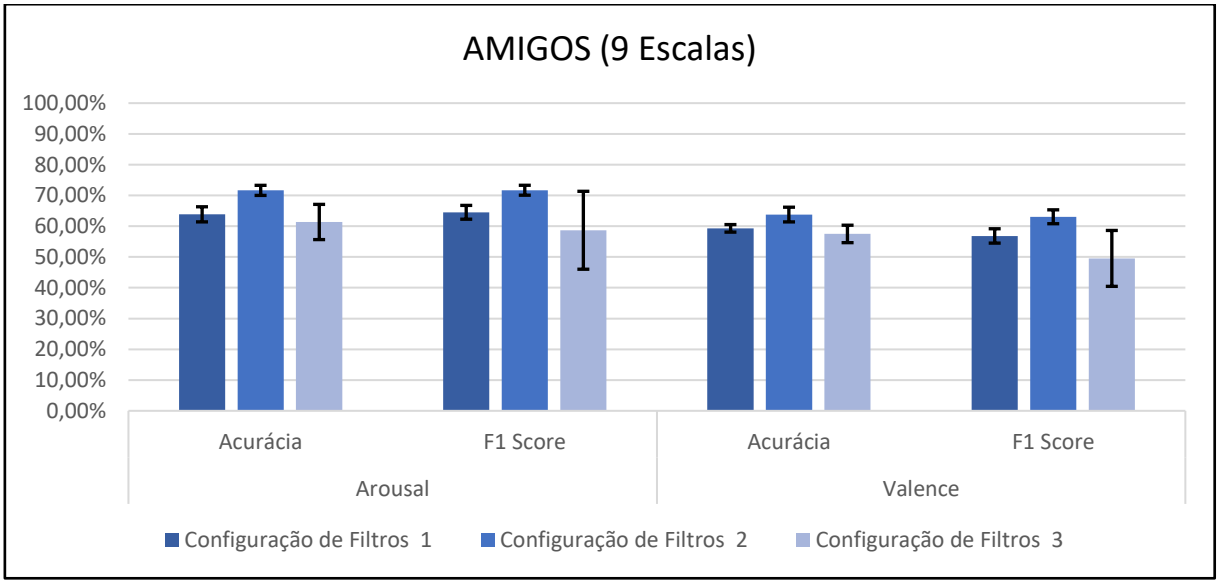

Figura 5.12 - Resultados para a classificação de emoções na base de dado AMIGOS para os grupos de experimentos.

Os resultados para excitação e valência na base de dados DREAMER para cinco classes foram melhores quando comparados aos da AMIGOS. A análise dos resultados na classificação de emoções em excitação iniciou com a configuração de filtros 1, alcançando uma acurácia de 59,95% e um F1 Score de 58,68%. Este desempenho foi significativamente superado ao utilizar a configuração de filtros 2, que impulsionou a acurácia para 77,06% e o F1 Score para 77,00%. No entanto, ao utilizar a configuração de filtros 3, essa tendência foi revertida, reduzindo a acurácia para 61,27% e o F1 Score para 58,08%, com um notável aumento no desvio padrão. Em termos de valência, os resultados foram semelhantes: partindo de 61,92% de acurácia e 57,02% de F1 Score para a configuração de filtros 1, houve uma melhora significativa para 73,15% e 72,11%, respectivamente, com a configuração de filtros 2. Já para a configuração de filtros 3, houve uma queda na acurácia para 60,91% e no F1 Score para 55,52%, acompanhada por um aumento na variação do desvio padrão, conforme apresentado na Figura 5.13.

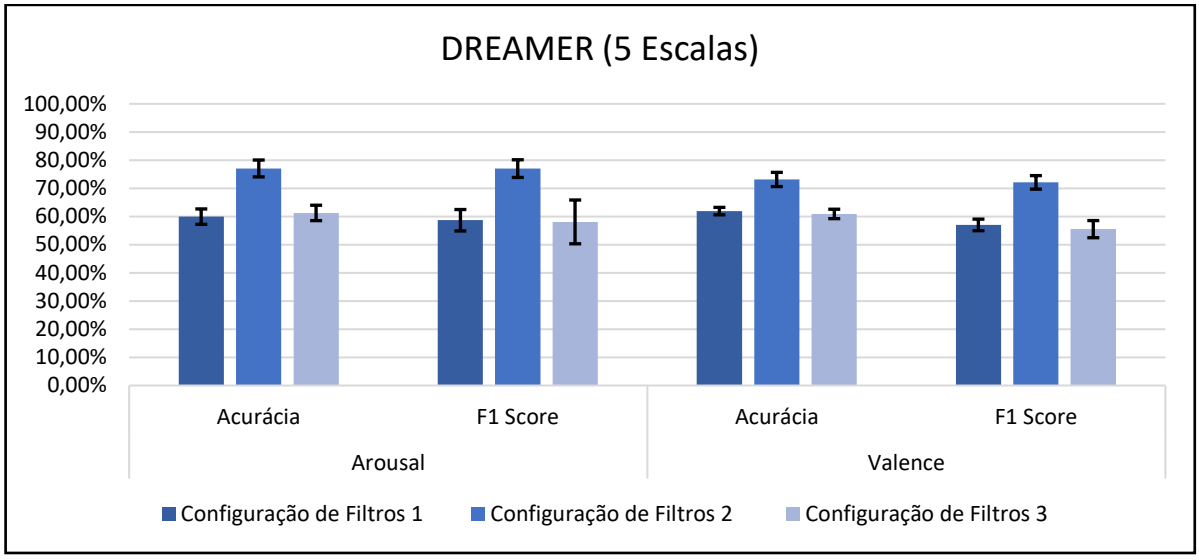

Figura 5.13 - Resultados para a classificação de emoções na base de dado DREAMER para os grupos de experimentos.

Quanto ao desempenho do modelo aplicado à base de dados SWELL, foi possível observar uma evolução nas métricas, quando comparado ao desempenho das bases de dados AMIGOS e DREAMER. Inicialmente, com a configuração de filtros 1, o modelo demonstrou uma performance promissora, atingindo 77,32% de acurácia e 77,70% de F1 Score para a classificação de excitação. Ao utilizar a configuração de filtros 2, houve um salto significativo no desempenho, elevando a acurácia para 95,68% e o F1 Score para 95,60%, sugerindo que um maior número de filtros de kernel permite uma captura mais eficiente das características dos dados. No entanto, ampliar ainda mais o número de filtros, como na configuração de filtros 3, resultou em uma queda acentuada no desempenho, com a acurácia e o F1 Score descendo para 64,20% e 67,96%, respectivamente, e um incremento notável no desvio padrão, indicativo de sobreajuste do modelo. A configuração de filtros 2 se mostrou ideal, aumentando a acurácia para 93,18% e o F1 Score para 93,20%, conforme apresentado na Figura 5.14.

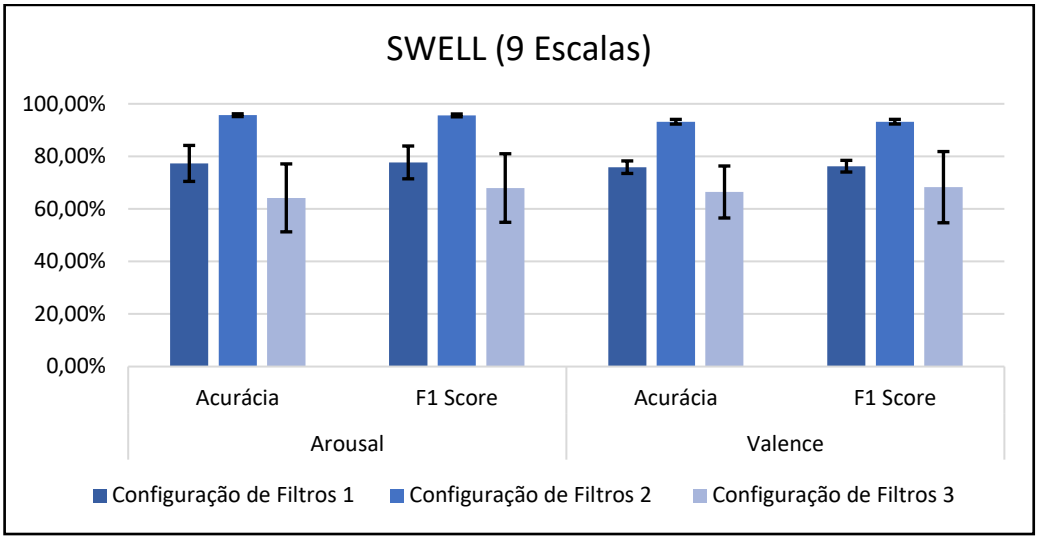

Figura 5.14 - Resultados para a classificação de emoções na base de dado SWELL para os grupos de experimentos.

Na base de dados WESAD, dedicada à classificação de estresse, o número de filtros de kernel demonstrou ter um papel importante no desempenho do modelo. Iniciando com a configuração de filtros 1, o modelo apresentou uma acurácia de 72,24% e um F1 Score de 68,61%. O aumento do número de filtros, como na configuração de filtros 2, gerou uma melhora significativa, elevando a acurácia para 93,96% e o F1 Score para 92,75%, indicando que essa configuração de hiperparâmetros captura de maneira eficiente as características do sinal para a tarefa de classificação de emoções. No entanto, ao utilizar a configuração de filtros 3, houve um declínio na acurácia para 69,62% e no F1 Score para 59,44%, com um incremento no desvio padrão, sinalizando um possível sobreajuste e uma perda de capacidade de generalização, conforme apresentado na Figura 5.15.

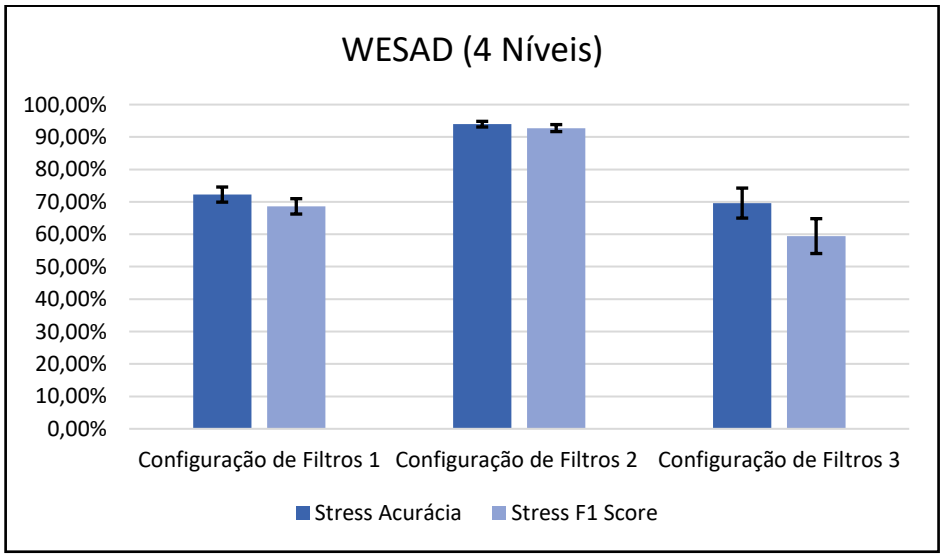

Figura 5.15 - Resultados para a classificação de emoções na base de dado WESAD para os grupos de experimentos.

Por meio da análise de desempenho de modelos nas bases de dados AMIGOS, DREAMER, SWELL e WESAD, com foco na classificação de emoções, foram identificadas tendências que realçam a importância do número de filtros de kernel no aprendizado autossupervisionado. Em todas as quatro bases de dados, observou-se que um tamanho intermediário de filtros, especificamente o utilizado na configuração de filtros 2, proporciona um equilíbrio ideal entre a generalização dos dados e o desempenho do modelo. Esta configuração resultou em melhorias significativas tanto na acurácia quanto no F1 Score em comparação com a configuração de filtros 1.

Entretanto, configurações maiores, como as utilizadas na configuração de filtros 3, frequentemente resultaram em uma diminuição do desempenho. Isso sugere que existe um limite na configuração do número de filtros, além do qual o aumento no número de filtros não apenas deixa de oferecer benefícios adicionais, mas também pode prejudicar a capacidade do modelo de generalizar devido ao sobreajuste. Esses achados indicam que uma abordagem cuidadosa na seleção do número de filtros de kernel é crucial para otimizar o aprendizado autossupervisionado e a classificação de emoções, garantindo que o modelo capture as características essenciais dos dados sem comprometer sua capacidade de generalização.

5.2.3 Análise da Variação da Quantidade dos Dados Rotulados utilizados para o Reconhecimento de Emoções

Neste cenário, avaliamos o efeito da quantidade de dados rotulados na classificação de emoções, considerando três proporções específicas: 5%, 25% e 50% de dados rotulados. Para estabelecer uma comparação efetiva, inicialmente, um pré-treinamento autossupervisionado foi implementado utilizando a Combinação 5 de tarefas auxiliares, conforme descrito na seção 5.1.1.

Posteriormente, foram conduzidos treinamentos supervisionados com o objetivo de comparar os resultados com o cenário autossupervisionado. A escolha desse método foi motivada pelo desafio de classificação de emoções, que é frequentemente limitado pela disponibilidade de dados rotulados. Ao separar frações específicas de dados rotulados, buscou-se entender a partir de qual proporção de dados rotulados os modelos que utilizam aprendizado autossupervisionado superam os totalmente supervisionados. A Figura 5.16 apresenta a estrutura do treinamento executado neste cenário.

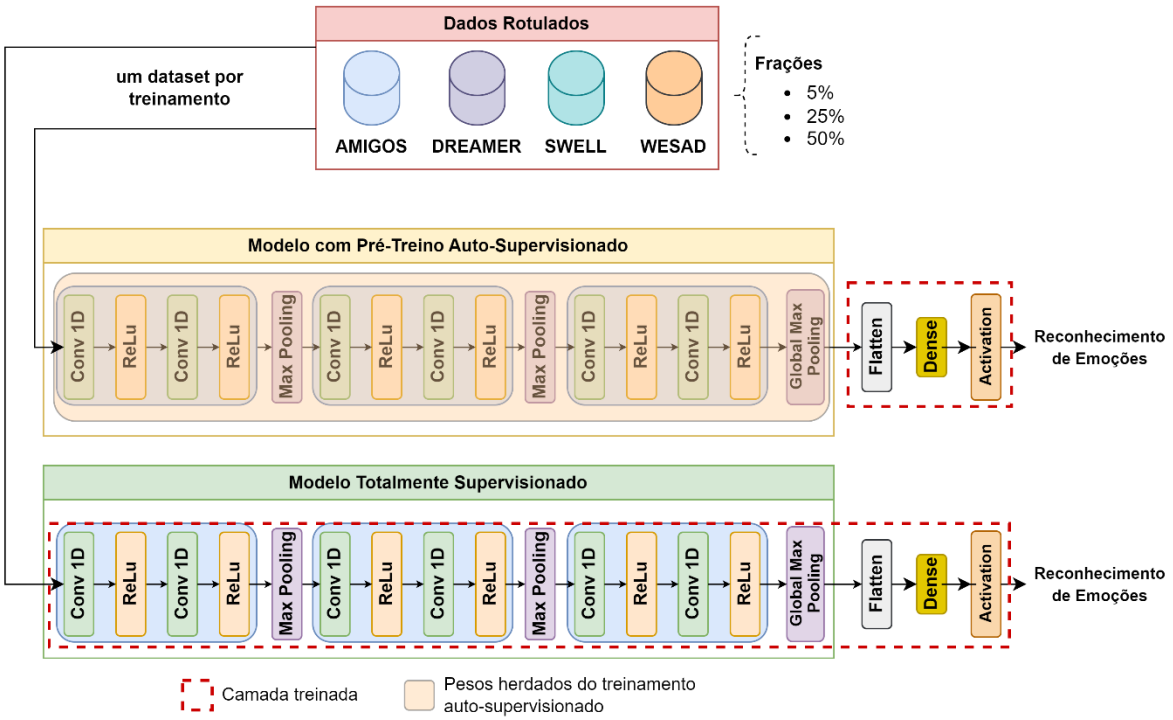

Figura 5.16 - Estrutura do treinamento do modelo com pré-treinamento autossupervisionado e totalmente supervisionado.

Nos experimentos realizados utilizando o modelo com pré-treino autossupervisionado, observou-se que os resultados na classificação de emoções foram influenciados tanto pela quantidade de dados rotulados disponíveis quanto pela base de dados utilizada. Notou-se que, com apenas 5% de dados rotulados, houve variações significativas nos resultados de acurácia entre as bases de dados WESAD e SWELL.

A análise dos resultados da base de dados WESAD na classificação de emoções revela um aumento no desempenho conforme a proporção de dados rotulados cresce. Com apenas 5% dos dados rotulados, o modelo alcançou uma acurácia de 70,01%. Este valor aumentou significativamente para 89,57% quando a proporção de dados rotulados foi elevada para 25%. Essa tendência de melhoria se manteve, com a acurácia chegando a 91,48% ao utilizar 50% dos dados rotulados, conforme demonstrado na Figura 5.17.

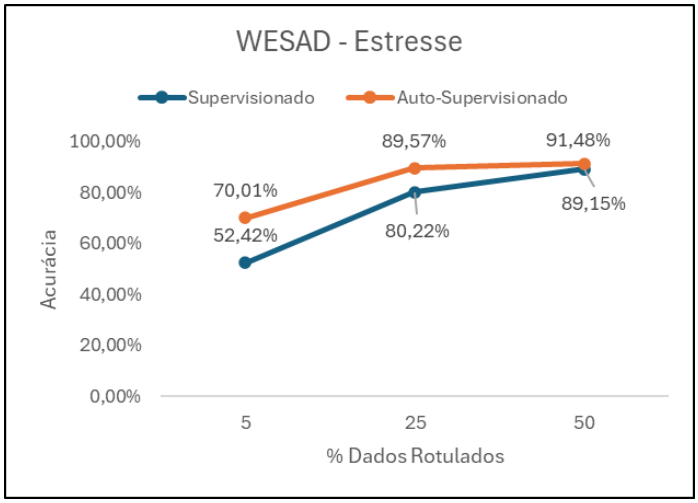

Figura 5.17- Resultados para a classificação de emoções na base de dado WESAD com a variação da quantidade de dados rotulados utilizados no treinamento.

Na base de dados SWELL, a análise de classificação de emoções em um cenário de 9 classes também revelou uma correlação positiva entre o aumento do volume de dados rotulados e a melhoria das métricas de desempenho. Com apenas 5% dos dados rotulados, a acurácia do modelo foi de 88,84%. Este valor aumentou para 93,64% com 25% dos dados rotulados e atingiu 94,18% quando 50% dos dados foram utilizados. Além disso, a valência também mostrou melhorias significativas; a acurácia subiu de 84,94% para 91,29% com o mesmo incremento nos dados rotulados, como demonstrado na Figura 5.18.

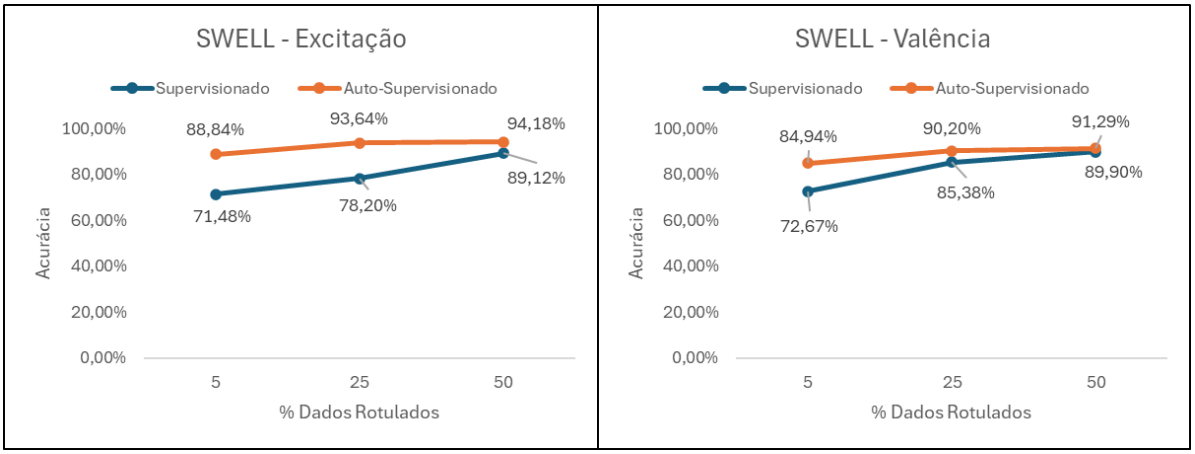

Figura 5.18 - Resultados para a classificação de emoções na base de dado SWELL com a variação da quantidade de dados rotulados utilizados no treinamento.

Os resultados obtidos das bases de dados WESAD e SWELL mostram que o aprendizado autossupervisionado oferece vantagens significativas, especialmente em cenários com dados rotulados limitados.

Na base de dados WESAD, mesmo com apenas 5% dos dados rotulados, o modelo autossupervisionado conseguiu atingir uma acurácia de 70,01%, que aumentou para 91,48% com 50% dos dados rotulados. Este crescimento na acurácia demonstra a capacidade do modelo autossupervisionado de aproveitar eficientemente os dados rotulados disponíveis, adaptando-se e aprendendo de variações nos dados para melhorar progressivamente seu desempenho.

Na base de dados SWELL, o desempenho também foi satisfatório. Com 5% dos dados rotulados, o modelo alcançou uma acurácia inicial de 88,84%, que aumentou para 94,18% com o incremento para 50% dos dados rotulados. Esses resultados confirmam a eficácia do modelo autossupervisionado em lidar com diferentes proporções de dados rotulados, destacando sua capacidade de generalização em um cenário de classificação mais complexo de 9 classes.

5.3 Comparação com outros trabalhos na literatura

Os resultados dos diversos trabalhos de reconhecimento de emoções são comparados de forma estruturada, agrupando-os conforme as bases de dados utilizadas: AMIGOS, DREAMER, SWELL e WESAD. Essa abordagem facilita a análise comparativa ao destacar como diferentes técnicas de reconhecimento de emoções se desempenham quando aplicadas a conjuntos de dados distintos, cada um com suas próprias características e desafios. Tal organização permite uma avaliação mais clara das metodologias em relação à sua eficácia e aplicabilidade, considerando as especificidades de cada base de dados.

A Tabela 5.7 exibe as médias de acurácia na classificação de emoções utilizando a base de dados AMIGOS. O modelo proposto, que empregou a Combinação 5 de tarefas auxiliares, alcançou uma acurácia de 71,65% para Excitação e 63,79% para Valência. Esses resultados foram inferiores aos apresentados pelos estudos de Sarkar e Etemad (2020) e Rodriguez et al. (2022). A discrepância nos resultados pode estar associada ao método de validação empregado no modelo proposto, que utilizou a técnica de validação cruzada leave-one-out. Esta técnica garante que, em cada iteração do treinamento, os dados de um mesmo participante não sejam utilizados simultaneamente para teste, evitando assim o vazamento de dados.

Tabela 5.7 - Comparação entre os trabalhos da literatura e o nosso modelo para a base de dados AMIGOS.

Abordagem	Acurácia		F1-Score	
	Excitação	Valência	Excitação	Valência
Rodriguez et al., 2022	88,00 %	83,00 %	87,00 %	83,00 %
Sarkar e Etemad, 2020	79,60 %	78,30 %	77,70 %	76,50 %
Modelo proposto	71,65 %	63,79 %	71,69 %	63,07 %

A Tabela 5.8 mostra as médias de acurácia na classificação de emoções com o uso da base de dados DREAMER. O modelo desenvolvido atingiu uma acurácia de 77,06% para Excitação e 73,15% para Valência, superando os resultados do estudo de Quispe et al., 2022, que empregou a mesma técnica de validação.

Tabela 5.8 - Comparação entre os trabalhos da literatura e o nosso modelo para a base de dados DREAMER.

Abordagem	Acurácia		F1-Score	
	Excitação	Valência	Excitação	Valência
Sarkar e Etemad, 2020	77,10 %	74,90 %	74,00 %	74,70 %
Modelo proposto	77,06 %	73,15 %	77,00 %	72,11 %
Quispe et al, 2022	71,27 %	70,24 %	70,86 %	68,49 %

A Tabela 5.9 apresenta as médias de acurácia na classificação de emoções usando a base de dados SWELL. O modelo desenvolvido alcançou uma acurácia de 95,68% para Excitação e 93,18% para Valência. Esses resultados são superiores aos reportados por Sarkar e Etemad (2020), mas não alcançam os de Quispe et al. (2022). As variações nos resultados podem ser atribuídas às diferentes tarefas auxiliares utilizadas e às especificidades da arquitetura dos modelos implementados.

Tabela 5.9 - Comparação entre os trabalhos da literatura e o nosso modelo para a base de dados SWELL.

Abordagem	Acurácia		F1-Score	
	Excitação	Valência	Excitação	Valência
Quispe et al, 2022	96,94 %	95,58 %	96,87 %	95,58 %
Modelo proposto	95,68 %	93,18 %	95,60 %	93,20 %
Sarkar e Etemad, 2020	92,60 %	93,80 %	93,00 %	94,30 %

A Tabela 5.10 apresenta as médias de acurácia na classificação de emoções usando a base de dados WESAD. O modelo desenvolvido alcançou uma acurácia de 93,96% para a classificação de estresse, superior ao trabalho de Rabbani e Khan (2022), que utilizou uma abordagem de aprendizado contrastivo.

Tabela 5.10 - Comparação entre os trabalhos da literatura e o nosso modelo para a base de dados WESAD.

Abordagem	Acurácia	F1-Score
Sarkar e Etemad, 2020	95,00 %	94,00 %
Modelo proposto	93,96 %	92,75 %
Rabbani, 2022	73,80 %	-

5.4 Considerações Finais

Como considerações finais, alguns pontos centrais emergem dos cenários explorados. A seção que aborda a análise do treinamento autossupervisionado com tarefas auxiliares combinadas revelou que a combinação entre diferentes tarefas pode desempenhar um papel crucial na capacidade de generalização do modelo de classificação de emoções. O emprego cuidadoso de tarefas auxiliares resultou em melhores resultados na aprendizagem de características, o que é essencial para a classificação de emoções. Esta abordagem, que integra tarefas de transformações, demonstrou ser especialmente eficiente em comparação com métodos que utilizam tarefas de maneira isolada, apontando para a eficácia de estratégias de treinamento mais abrangentes no domínio do reconhecimento de emoções.

Além disso, a análise da variação de neurônios e da quantidade dos dados rotulados no treinamento autossupervisionado ofereceu perspectivas fundamentais sobre a otimização de redes neurais. A variação do tamanho dos filtros de kernel teve um impacto direto na qualidade da representação aprendida, com um tamanho excessivo ou insuficiente de podendo levar a overfitting ou underfitting. Da mesma forma, a análise da variação da quantidade dos dados rotulados destacou como a quantidade de dados rotulados influencia a eficácia do treinamento na etapa de reconhecimento de emoções. À medida que a quantidade de dados rotulados aumentou, houve uma melhoria correspondente no desempenho do modelo, especialmente evidente quando a transição ocorreu de poucos dados rotulados para uma quantidade moderada.

6. CONCLUSÕES

Este trabalho abordou o desenvolvimento e a avaliação de modelos de aprendizado autossupervisionado para o reconhecimento de emoções utilizando sinais de eletrocardiograma, analisando diversas combinações de tarefas auxiliares em quatro bases de dados distintas: AMIGOS, DREAMER, SWELL e WESAD. A pesquisa focou na capacidade dessas combinações de transformações de sinal em melhorar a generalização do modelo para diferentes tipos de emoções, utilizando uma abordagem que minimiza a dependência de grandes volumes de dados rotulados.

Os experimentos demonstraram que as transformações de sinal, como adição de ruído, escalonamento, negação, inversão horizontal, permutação e time-warping, aplicadas no contexto do aprendizado autossupervisionado, foram eficazes em extrair características dos dados, facilitando a classificação de emoções. Por exemplo, na base de dados SWELL, o modelo alcançou acurácias de até 95,68%, mostrando pouca perda de desempenho mesmo com reduções na quantidade de dados rotulados, especialmente na classificação em nove classes de emoções.

Os resultados foram menos consistentes na base de dados WESAD, onde o desempenho variou mais significativamente entre os grupos, indicando desafios particulares em contextos de estresse. No entanto, a abordagem auto-supervisionada provou ser superior aos métodos tradicionalmente supervisionados, especialmente em cenários com disponibilidade limitada de dados rotulados. Nas bases de dados WESAD e SWELL, mesmo com apenas 25% dos dados rotulados, o desempenho manteve-se competitivo.

Além disso, investigamos o impacto da variação do número de filtros de kernel nos blocos de convolução no aprendizado do modelo. Os experimentos realizados com três configurações revelaram que um número intermediário de filtros, especificamente aquele utilizado na Configuração de Filtros 2, ofereceu um equilíbrio ideal, melhorando significativamente tanto a acurácia quanto o F1 Score em comparação com a Configuração de Filtros 1. No entanto, configurações mais altas, como as da Configuração de Filtros 3, frequentemente resultaram em uma diminuição do desempenho e em um aumento do desvio padrão, indicativo de potencial sobreajuste.

6.1 Contribuições

Além de contribuir com novas metodologias para o reconhecimento de emoções, este estudo também destaca a importância de desenvolver técnicas de aprendizado que se adaptam a restrições de dados. Conclui-se que as técnicas de aprendizado autossupervisionado são não apenas viáveis, mas também eficazes para a tarefa de reconhecimento de emoções, oferecendo uma abordagem flexível e robusta que pode ser ajustada conforme as necessidades do cenário de aplicação.

Essa capacidade de adaptação e eficácia foi comprovada pelos resultados obtidos, o que não só demonstra o potencial do modelo, mas também seu valor científico. A relevância e o impacto deste trabalho foram reconhecidos por meio de sua publicação do artigo Applying Self-Supervised Representation Learning for Emotion Recognition Using Physiological Signals (SENSORS, v. 22, p. 9102-9126, 2022), destacando sua contribuição para a comunidade acadêmica.

Além disso, outro artigo está em elaboração para a publicação dos resultados das análises apresentadas neste trabalho. Este reconhecimento valida os esforços e os resultados apresentados, e encoraja o desenvolvimento contínuo de pesquisas que abordam desafios semelhantes na área de reconhecimento de emoções.

6.2 Trabalhos Futuros

Neste trabalho, explorou-se a eficácia do aprendizado autossupervisionado utilizando várias combinações de transformações de sinal para reconhecer emoções a partir de sinais de eletrocardiograma. Os experimentos indicaram a viabilidade do método proposto em diversas bases de dados, tais como AMIGOS, DREAMER, SWELL e WESAD. Considerando esses resultados, vários caminhos se abrem para pesquisas futuras.

Primeiramente, é essencial avaliar a implementação e o desempenho desses modelos em ambientes online, particularmente em dispositivos móveis e sistemas em tempo real, onde a classificação de emoções pode ser aplicada para intervenções imediatas em contextos de saúde mental ou monitoramento contínuo do bem estar.

Além disso, a extensão do método proposto para outros tipos de sinais fisiológicos, como eletroencefalograma ou sinais de atividade eletrodérmica, pode facilitar o diagnóstico de condições médicas ou aprimorar sistemas de interface cérebro-computador. Essa expansão do trabalho atual poderia também incluir a investigação de tarefas auxiliares específicas que sejam

particularmente eficazes para cada tipo de sinal, como transformações que considerem as características únicas de sinais eletroencefalograma em comparação com eletrocardiograma.

Uma área de grande interesse envolve a exploração de diferentes métodos de fusão de características e de decisão. A implementação de técnicas como ensembles de modelos ou a introdução de novos métodos de extração de características, como a aplicação de transformadores, poderiam melhorar a discriminação entre diferentes tipos de emoções, especialmente em cenários onde a quantidade de dados rotulados é limitada.

Por fim, considerando as variações significativas de desempenho entre diferentes grupos de dados, futuros trabalhos poderiam focar em métodos adaptativos que ajustem dinamicamente as características aprendidas ou as transformações aplicadas com base nas especificidades dos dados em análise. Investigar tarefas auxiliares mais específicas, como normalização adaptativa ou geração de características sintéticas que possam melhor refletir estados emocionais sutis, também são caminhos promissores para trabalhos futuras.

REFERÊNCIAS BIBLIOGRÁFICAS

AHMED, Ferdous; BARI, A. S.M.Hossain; GAVRILOVA, Marina L. Emotion Recognition from Body Movement. IEEE Access, [*s. l.*], v. 8, p. 11761–11781, 2020. Disponível em: https://doi.org/10.1109/ACCESS.2019.2963113

ALARCAO, Soraia M; FONSECA, Manuel J. Emotions recognition using EEG signals: A survey. IEEE Transactions on Affective Computing, [s. l.], v. 10, n. 3, p. 374–393, 2017.

ALQAHTANI, Hamed; KAVAKLI-THORNE, Manolya; KUMAR, Gulshan. Applications of Generative Adversarial Networks (GANs): An Updated Review. Archives of Computational Methods in Engineering, [*s. l.*], v. 28, n. 2, p. 525–552, 2021. Disponível em: https://doi.org/10.1007/s11831-019-09388-y

AQAJARI, Seyed Amir Hossein et al. Pain Assessment Tool With Electrodermal Activity for Postoperative Patients: Method Validation Study, [s. l.], v. 9, n. 5, p. e25258, 2021. Disponível em: https://doi.org/10.2196/25258

AYATA, Deger; YASLAN, Yusuf; KAMAŞAK, Mustafa. Emotion Recognition via Galvanic Skin Response: Comparison of Machine Learning Algorithms and Feature Extraction Methods. IU-Journal of Electrical Electronics Engineering, [s. l.], v. 17, n. 1, p. 3147–3156, 2017.

BAZGIR, Omid; MOHAMMADI, Zeynab; HABIBI, Seyed Amir Hassan. Emotion Recognition with Machine Learning Using EEG Signals. 2018 25th Iranian Conference on Biomedical Engineering and 2018 3rd International Iranian Conference on Biomedical Engineering, ICBME 2018, [*s. l.*], p. 1–5, 2018. Disponível em: https://doi.org/10.1109/ICBME.2018.8703559

BENEZETH, Yannick et al. Remote heart rate variability for emotional state monitoring. In: , 2018. 2018 IEEE EMBS International Conference on Biomedical Health Informatics (BHI). [S. l.: s. n.], 2018. p. 153–156. Disponível em: https://doi.org/10.1109/BHI.2018.8333392

BERKAYA, Selcan Kaplan et al. A survey on ECG analysis. Biomedical Signal Processing and Control, [s. l.], v. 43, p. 216–235, 2018. Disponível em: https://doi.org/10.1016/j.bspc.2018.03.003

BOBADE, Pramod; VANI, M. Stress Detection with Machine Learning and Deep Learning using Multimodal Physiological Data. Proceedings of the 2nd International Conference on Inventive Research in Computing Applications, ICIRCA 2020, [*s. l.*], p. 51–57, 2020. Disponível em: https://doi.org/10.1109/ICIRCA48905.2020.9183244

CAI, Linqin; DONG, Jiangong; WEI, Min. Multi-Modal Emotion Recognition from Speech and Facial Expression Based on Deep Learning. Proceedings - 2020 Chinese Automation Congress, CAC 2020, [*s. l.*], p. 5726–5729, 2020. Disponível em: https://doi.org/10.1109/CAC51589.2020.9327178

CHAO, Hao et al. Emotion Recognition from Multiband EEG Signals Using CapsNet. Sensors, [s. l.], v. 19, n. 9, p. 2212, 2019. Disponível em: https://doi.org/10.3390/s19092212

CHEN, Guijun *et al.* Emotion Feature Analysis and Recognition Based on Reconstructed EEG Sources. IEEE Access, [*s. l.*], v. 8, p. 11907–11916, 2020. Disponível em: https://doi.org/10.1109/ACCESS.2020.2966144

CHOI, Kwang-Ho et al. Is heart rate variability (HRV) an adequate tool for evaluating human emotions? A focus on the use of the International Affective Picture System (IAPS). Psychiatry Research, [s. l.], v. 251, p. 192–196, 2017. Disponível em: https://doi.org/10.1016/j.psychres.2017.02.025

CORREA, Juan Abdon Miranda et al. Amigos: A dataset for affect, personality and mood research on individuals and groups. IEEE Transactions on Affective Computing, [s. l.], 2018.

COWIE, Roddy; CORNELIUS, Randolph R. Describing the emotional states that are expressed in speech. Speech communication, [s. l.], v. 40, n. 1–2, p. 5–32, 2003.

DESAI, Usha; SHETTY, Akshaya D. Electrodermal Activity (EDA) for Treatment of Neurological and Psychiatric Disorder Patients: A Review. In: , 2021. 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS). [S. l.: s. n.], 2021. p. 1424–1430.

DOHENY, Emer P. *et al.* Feature-Based Evaluation of a Wearable Surface EMG Sensor against Laboratory Standard EMG during Force-Varying and Fatiguing Contractions. IEEE Sensors Journal, [*s. l.*], v. 20, n. 5, p. 2757–2765, 2020. Disponível em: https://doi.org/10.1109/JSEN.2019.2953354

DZEDZICKIS, Andrius; KAKLAUSKAS, Art\=uras; BUCINSKAS, Vytautas. Human Emotion Recognition: Review of Sensors and Methods. Sensors, [s. l.], v. 20, n. 3, p. 592, 2020. Disponível em: https://doi.org/10.3390/s20030592

EKMAN, Paul. An Argument for Basic Emotions. Cognition and Emotion, [s. l.], v. 6, n. 3–4, p. 169–200, 1992. Disponível em: https://doi.org/10.1080/02699939208411068

FARINA, Dario et al. The extraction of neural information from the surface EMG for the

control of upper-limb prostheses: emerging avenues and challenges. IEEE Transactions on Neural Systems and Rehabilitation Engineering, [s. l.], v. 22, n. 4, p. 797–809, 2014.

FILATOVA, Natalya Nikolaevna; TEREKHIN, Sergey Alexeevich. Bioengineering system for research on human emotional response to external stimuli. Proceedings - 2015 International Conference on Biomedical Engineering and Computational Technologies, SIBIRCON 2015, [*s. l.*], p. 13–17, 2015. Disponível em: https://doi.org/10.1109/SIBIRCON.2015.7361841

GANGULY, Suranjita; SINGLA, Rajesh. Electrode Channel Selection for Emotion Recognition based on EEG Signal. 2019 IEEE 5th International Conference for Convergence in Technology, I2CT 2019, [s. l.], p. 1–4, 2019. Disponível em: https://doi.org/10.1109/I2CT45611.2019.9033929

GIANNAKAKIS, Giorgos et al. Review on Psychological Stress Detection Using Biosignals. IEEE Transactions on Affective Computing, [s. l.], v. 13, n. 1, p. 440–460, 2022. Disponível em: https://doi.org/10.1109/taffc.2019.2927337

GIDARIS, Spyros; SINGH, Praveer; KOMODAKIS, Nikos. Unsupervised Representation Learning by Predicting Image Rotations. [S. l.]: arXiv, 2018. Disponível em: https://doi.org/10.48550/ARXIV.1803.07728

GONG, Chao *et al.* Amygdala-inspired affective computing: To realize personalized intracranial emotions with accurately observed external emotions. China Communications, [*s. l.*], v. 16, n. 8, p. 115–129, 2019. Disponível em: https://doi.org/10.23919/JCC.2019.08.011

GUANGHUI, Chen; XIAOPING, Zeng. Multi-Modal Emotion Recognition by Fusing Correlation Features of Speech-Visual. IEEE Signal Processing Letters, [*s. l.*], v. 28, p. 533–537, 2021. Disponível em: https://doi.org/10.1109/LSP.2021.3055755

GUO, Han-Wen et al. Heart Rate Variability Signal Features for Emotion Recognition by Using Principal Component Analysis and Support Vectors Machine. In: , 2016. 2016 IEEE 16th International Conference on Bioinformatics and Bioengineering (BIBE). [S. l.]: IEEE, 2016. Disponível em: https://doi.org/10.1109/bibe.2016.40

HOGREL, Jean-Yves. Clinical applications of surface electromyography in neuromuscular disorders. Neurophysiologie Clinique/Clinical Neurophysiology, [s. l.], v. 35, n. 2–3, p. 59–71, 2005. Disponível em: https://doi.org/10.1016/j.neucli.2005.03.001

HOSSAIN, Khondker Fariha *et al.* ECG-ATK-GAN: Robustness against Adversarial Attacks on ECG using Conditional Generative Adversarial Networks. [*s. l.*], n. 1, p. 1–5, 2021.

Disponível em: http://arxiv.org/abs/2110.09983

HSU, Yu Liang *et al.* Automatic ECG-Based Emotion Recognition in Music Listening. IEEE Transactions on Affective Computing, [*s. l.*], v. 11, n. 1, p. 85–99, 2020. Disponível em: https://doi.org/10.1109/TAFFC.2017.2781732

HU, Ping *et al.* Learning supervised scoring ensemble for emotion recognition in the wild. ICMI 2017 - Proceedings of the 19th ACM International Conference on Multimodal Interaction, [*s. l.*], v. 2017-Janua, p. 553–560, 2017. Disponível em: https://doi.org/10.1145/3136755.3143009

HU, Xin et al. EEG Correlates of Ten Positive Emotions. Frontiers in Human Neuroscience, [s. l.], v. 11, 2017. Disponível em: https://doi.org/10.3389/fnhum.2017.00026

HUANG, Chuhui et al. Discovery of Irreversible Inhibitors Targeting Histone Methyltransferase, SMYD. ACS Medicinal Chemistry Letters, [s. l.], v. 10, n. 6, p. 978–984, 2019. Disponível em: https://doi.org/10.1021/acsmedchemlett.9b00170

HUSSEIN, Sarfaraz *et al.* Lung and Pancreatic Tumor Characterization in the Deep Learning Era: Novel Supervised and Unsupervised Learning Approaches. IEEE Transactions on Medical Imaging, [*s. l.*], v. 38, n. 8, p. 1777–1787, 2019. Disponível em: https://doi.org/10.1109/TMI.2019.2894349

JEYHANI, Vala et al. Comparison of HRV parameters derived from photoplethysmography and electrocardiography signals. In: , 2015. 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). [S. l.]: IEEE, 2015. Disponível em: https://doi.org/10.1109/embc.2015.7319747

JING, Cai; LIU, Guangyuan; HAO, Min. The Research on Emotion Recognition from ECG Signal. In: , 2009. 2009 International Conference on Information Technology and Computer Science. [S. l.]: IEEE, 2009. Disponível em: https://doi.org/10.1109/itcs.2009.108

JING, Longlong *et al.* Self-Supervised Spatiotemporal Feature Learning via Video Rotation Prediction. [*s. l.*], 2018. Disponível em: http://arxiv.org/abs/1811.11387

JOESPH, Cynthia et al. Implementation of physiological signal based emotion recognition algorithm. In: , 2020. 2020 IEEE 36th International Conference on Data Engineering (ICDE). [S. l.: s. n.], 2020. p. 2075–2079.

JOY, Emily et al. Recent Survey on Emotion Recognition Using Physiological Signals. 2021 7th International Conference on Advanced Computing and Communication Systems, ICACCS 2021, [s. l.], p. 1858–1863, 2021. Disponível em:

https://doi.org/10.1109/ICACCS51430.2021.9441999

JOY, Emily *et al.* Recent Survey on Emotion Recognition Using Physiological Signals. 2021 7th International Conference on Advanced Computing and Communication Systems, ICACCS 2021, [*s. l.*], p. 1858–1863, 2021. Disponível em: https://doi.org/10.1109/ICACCS51430.2021.9441999

KATSIGIANNIS, Stamos; RAMZAN, Naeem. DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals from Wireless Low-cost Off-the-Shelf Devices. IEEE Journal of Biomedical and Health Informatics, [*s. l.*], v. 22, n. 1, p. 98–107, 2017. Disponível em: https://doi.org/10.1109/JBHI.2017.2688239

KHALIL, Ruhul Amin *et al.* Speech Emotion Recognition Using Deep Learning Techniques: A Review. IEEE Access, [*s. l.*], v. 7, p. 117327–117345, 2019. Disponível em: https://doi.org/10.1109/ACCESS.2019.2936124

KHATTAK, Asad et al. An efficient deep learning technique for facial emotion recognition. Multimedia Tools and Applications, [s. l.], v. 81, n. 2, p. 1649–1683, 2021. Disponível em: https://doi.org/10.1007/s11042-021-11298-w

KIM, Eun Young et al. Diagnosis of major depressive disorder by combining multimodal information from heart rate dynamics and serum proteomics using machine-learning algorithm. Progress in Neuro-Psychopharmacology and Biological Psychiatry, [s. l.], v. 76, p. 65–71, 2017. Disponível em: https://doi.org/10.1016/j.pnpbp.2017.02.014

KIM, Hodam et al. Classification of Individual's discrete emotions reflected in facial microexpressions using electroencephalogram and facial electromyogram. Expert Systems with Applications, [s. l.], v. 188, p. 116101, 2022. Disponível em: https://doi.org/10.1016/j.eswa.2021.116101

KOELSTRA, S. *et al.* DEAP: A Database for Emotion Analysis using Physiological Signals. IEEE TRANS. AFFECTIVE COMPUTING, [*s. l.*], v. 3, n. 1, p. 18–31, 2011.

KOLDIJK, Saskia *et al.* The Swell knowledge work dataset for stress and user modeling research. ICMI 2014 - Proceedings of the 2014 International Conference on Multimodal Interaction, [*s. l.*], p. 291–298, 2014. Disponível em: https://doi.org/10.1145/2663204.2663257

KOLDIJK, Saskia *et al.* The Swell knowledge work dataset for stress and user modeling research. IEEE Transactions on Affective Computing, [*s. l.*], v. 22, n. 1, p. 291–298, 2018. Disponível em: https://doi.org/10.1145/3242969.3242985

KOTOWSKI, Krzysztof; STAPOR, Katarzyna. Machine Learning and EEG for Emotional State Estimation. The Science of Emotional Intelligence, [s. l.], p. 75, 2021.

KREIBIG, Sylvia D. Autonomic nervous system activity in emotion: A review. Biological psychology, [s. l.], v. 84, n. 3, p. 394–421, 2010.

KRISHNA, N Murali et al. An Efficient Mixture Model Approach in Brain-Machine Interface Systems for Extracting the Psychological Status of Mentally Impaired Persons Using EEG Signals. IEEE Access, [s. l.], v. 7, p. 77905–77914, 2019. Disponível em: https://doi.org/10.1109/access.2019.2922047

LANG, PETER J et al. Looking at pictures: Affective, facial, visceral, and behavioral reactions. Psychophysiology, [s. l.], v. 30, n. 3, p. 261–273, 1993. Disponível em: https://doi.org/10.1111/j.1469-8986.1993.tb03352.x

LI, Jingjing et al. DRS-Net: A spatial\textendashtemporal affective computing model based on multichannel EEG data. Biomedical Signal Processing and Control, [s. l.], v. 76, p. 103660, 2022. Disponível em: https://doi.org/10.1016/j.bspc.2022.103660

LI, Xiang et al. EEG based Emotion Recognition: A Tutorial and Review. ACM Computing Surveys, [s. l.], 2022. Disponível em: https://doi.org/10.1145/3524499

LI, Yu Feng; GUO, Lan Zhe; ZHOU, Zhi Hua. Towards Safe Weakly Supervised Learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, [*s. l.*], v. 43, n. 1, p. 334–346, 2021. Disponível em: https://doi.org/10.1109/TPAMI.2019.2922396

LIU, Guohang *et al.* Easy Data Augmentation Method for Classification Tasks. 2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing, ICCWAMTIP 2020, [*s. l.*], p. 166–169, 2020. Disponível em: https://doi.org/10.1109/ICCWAMTIP51612.2020.9317525

LIU, Pei *et al.* A Survey of Text Data Augmentation. Proceedings - 2020 International Conference on Computer Communication and Network Security, CCNS 2020, [*s. l.*], p. 191–195, 2020. Disponível em: https://doi.org/10.1109/CCNS50731.2020.00049

LIU, Yun; DU, Siqing. Psychological stress level detection based on electrodermal activity. Behavioural Brain Research, [s. l.], v. 341, p. 50–53, 2018. Disponível em: https://doi.org/10.1016/j.bbr.2017.12.021

MAAOUI, Choubeila; PRUSKI, Alain; ABDAT, Faiza. Emotion Recognition for hHman-Machine Communication. In: , 2008. 2008 IEEE/RSJ International Conference on Intelligent

Robots and Systems. [S. l.]: IEEE, 2008. Disponível em: https://doi.org/10.1109/iros.2008.4650870

MAHESH, Batta. Machine learning algorithms-a review. International Journal of Science and Research (IJSR). [Internet], [s. l.], v. 9, p. 381–386, 2020.

MARKOVA, Valentina; GANCHEV, Todor; KALINKOV, Kalin. CLAS: A Database for Cognitive Load, Affect and Stress Recognition. Proceedings of the International Conference on Biomedical Innovations and Applications, BIA 2019, [*s. l.*], p. 2019–2022, 2019. Disponível em: https://doi.org/10.1109/BIA48344.2019.8967457

MAUSS, Iris B; ROBINSON, Michael D. Measures of emotion: A review. Cognition and Emotion, [s. l.], v. 23, n. 2, p. 209–237, 2009. Disponível em: https://doi.org/10.1080/02699930802204677

MEHARI, Temesgen; STRODTHOFF, Nils. Self-supervised representation learning from 12-lead ECG data. [*s. l.*], p. 1–11, 2021. Disponível em: http://arxiv.org/abs/2103.12676

MIKOLOV, Tomas et al. Efficient Estimation of Word Representations in Vector Space. [S. l.]: arXiv, 2013. Disponível em: https://doi.org/10.48550/ARXIV.1301.3781

MIRANDA-CORREA, Juan Abdon *et al.* AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups. IEEE Transactions on Affective Computing, [*s. l.*], v. 12, n. 2, p. 479–493, 2018. Disponível em: https://doi.org/10.1109/TAFFC.2018.2884461

MITHBAVKAR, Shraddha A; SHAH, Milind S. Analysis of EMG Based Emotion Recognition for Multiple People and Emotions. In: , 2021. 2021 IEEE 3rd Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS). [S. l.: s. n.], 2021. p. 1–4.

MONTERO QUISPE, Kevin G et al. Applying Self-Supervised Representation Learning for Emotion Recognition Using Physiological Signals. Sensors, [s. l.], v. 22, n. 23, 2022. Disponível em: https://doi.org/10.3390/s22239102

MONTESINOS, Victoriano *et al.* Multi-Modal Acute Stress Recognition Using Off-the-Shelf Wearable Devices. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, [*s. l.*], p. 2196–2201, 2019. Disponível em: https://doi.org/10.1109/EMBC.2019.8857130

MOORS, Agnes. Theories of emotion causation: A review. [S. l.]: Psychology Press, 2010.

MOSCHONA, Danai Styliani. An Affective Service based on Multi-Modal Emotion

Recognition, using EEG enabled Emotion Tracking and Speech Emotion Recognition. 2020 IEEE International Conference on Consumer Electronics - Asia, ICCE-Asia 2020, [*s. l.*], 2020. Disponível em: https://doi.org/10.1109/ICCE-Asia49877.2020.9277291

NA, Kyoung-Sae; CHO, Seo-Eun; CHO, Seong-Jin. Machine learning-based discrimination of panic disorder from other anxiety disorders. Journal of Affective Disorders, [s. l.], v. 278, p. 1–4, 2021. Disponível em: https://doi.org/10.1016/j.jad.2020.09.027

NGAI, Wang Kay et al. Emotion recognition based on convolutional neural networks and heterogeneous bio-signal data sources. Information Fusion, [s. l.], v. 77, p. 107–117, 2022. Disponível em: https://doi.org/10.1016/j.inffus.2021.07.007

NIJHAWAN, Rahul; SRIVASTAVA, Ishita; SHUKLA, Pushkar. Land Cover Classification Using Supervised and Unsupervised Learning Techniques. Nonlinear System Identification, [*s. l.*], p. 137–155, 2017. Disponível em: https://doi.org/10.1007/978-3-662-04323-3_6

PAPAGIANNIS, Georgios I et al. Methodology of surface electromyography in gait analysis: review of the literature. Journal of Medical Engineering and Technology, [s. l.], v. 43, n. 1, p. 59–65, 2019. Disponível em: https://doi.org/10.1080/03091902.2019.1609610

PARK, Cheul Young *et al.* K-EmoCon, a multimodal sensor dataset for continuous emotion recognition in naturalistic conversations. Scientific Data, [*s. l.*], v. 7, n. 1, p. 1–16, 2020. Disponível em: https://doi.org/10.1038/s41597-020-00630-y

PAWEŁJEMIOŁO et al. Datasets for Automated Affect and Emotion Recognition from Cardiovascular Signals Using Artificial Intelligence A Systematic Review. Sensors, [s. l.], v. 22, n. 7, p. 2538, 2022. Disponível em: https://doi.org/10.3390/s22072538

PLUTCHIK, Robert. A general psychoevolutionary theory of emotion. In: THEORIES OF EMOTION. [S. l.]: Elsevier, 1980. p. 3–33.

RABBANI, Suha; KHAN, Naimul. Contrastive self-supervised learning for stress detection from ecg data. Bioengineering, v. 9, n. 8, p. 374, 2022.

RASYADA, Ihda *et al.* Sentiment Analysis of BPJS Kesehatan's Services Based on Affective Models. IES 2020 - International Electronics Symposium: The Role of Autonomous and Intelligent Systems for Human Life and Comfort, [*s. l.*], n. January 2019, p. 549–556, 2020. Disponível em: https://doi.org/10.1109/IES50839.2020.9231940

RUSSELL, James A. A circumplex model of affect. Journal of personality and social psychology, [s. l.], v. 39, n. 6, p. 1161, 1980.

SAEED, Aaqib; OZCELEBI, Tanir; LUKKIEN, Johan. Multi-task Self-Supervised Learning for Human Activity Detection. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, [*s. l.*], v. 3, n. 2, p. 1–30, 2019. Disponível em: https://doi.org/10.1145/3328932

SAISHO, Osamu et al. Enhancing support for optimal muscle usage in sports. In: , 2019. Proceedings of the 23rd International Symposium on Wearable Computers. [S. l.]: ACM, 2019. Disponível em: https://doi.org/10.1145/3341163.3347722

SAITO, Shunsuke *et al.* SCANimate: Weakly Supervised Learning of Skinned Clothed Avatar Networks. [*s. l.*], v. 1, n. 2, p. 2886–2897, 2021. Disponível em: http://arxiv.org/abs/2104.03313

SARKAR, Pritam. Self-supervised ECG Representation Learning for Affective Computing. [*s. l.*], 2020.

SARKAR, Pritam; ETEMAD, Ali. Self-supervised ECG Representation Learning for Emotion Recognition. IEEE Transactions on Affective Computing, [*s. l.*], p. 1–13, 2020. Disponível em: https://doi.org/10.1109/TAFFC.2020.3014842

SARMA, Parthana; BARMA, Shovan. Review on Stimuli Presentation for Affect Analysis Based on EEG. IEEE Access, [s. l.], v. 8, p. 51991–52009, 2020. Disponível em: https://doi.org/10.1109/access.2020.2980893

SARROUTI, Mourad; ABACHA, Asma Ben; DEMNER-FUSHMAN, Dina. Multi-Task Transfer Learning with Data Augmentation for Recognizing Question Entailment in the Medical Domain. 2021 IEEE 9th International Conference on Healthcare Informatics (ICHI), [*s. l.*], p. 339–346, 2021. Disponível em: https://doi.org/10.1109/ichi52183.2021.00058

SCHMIDT, Philip *et al.* Introducing WeSAD, a multimodal dataset for wearable stress and affect detection. ICMI 2018 - Proceedings of the 2018 International Conference on Multimodal Interaction, [*s. l.*], p. 400–408, 2018a. Disponível em: https://doi.org/10.1145/3242969.3242985

SCHMIDT, Philip *et al.* Labelling affective states “in the wild”: Practical guidelines and lessons learned. UbiComp/ISWC 2018 - Adjunct Proceedings of the 2018 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2018 ACM International Symposium on Wearable Computers, [*s. l.*], p. 654–659, 2018b. Disponível em: https://doi.org/10.1145/3267305.3267551

SHAO, Ling; ZHU, Fan; LI, Xuelong. Transfer learning for visual categorization: A survey.

IEEE Transactions on Neural Networks and Learning Systems, [*s. l.*], v. 26, n. 5, p. 1019–1034, 2015. Disponível em: https://doi.org/10.1109/TNNLS.2014.2330900

SHORTEN, Connor; KHOSHGOFTAAR, Taghi M. A survey on Image Data Augmentation for Deep Learning. Journal of Big Data, [*s. l.*], v. 6, n. 1, 2019. Disponível em: https://doi.org/10.1186/s40537-019-0197-0

SHU, Lin et al. A Review of Emotion Recognition Using Physiological Signals. Sensors, [s. l.], v. 18, n. 7, p. 2074, 2018. Disponível em: https://doi.org/10.3390/s18072074

SHUKLA, Jainendra et al. Feature Extraction and Selection for Emotion Recognition from Electrodermal Activity. IEEE Transactions on Affective Computing, [s. l.], v. 12, n. 4, p. 857–869, 2021. Disponível em: https://doi.org/10.1109/taffc.2019.2901673

SUBASI, Abdulhamit; KIYMIK, M Kemal. Muscle Fatigue Detection in EMG Using Time\textendashFrequency Methods, ICA and Neural Networks. Journal of Medical Systems, [s. l.], v. 34, n. 4, p. 777–785, 2009. Disponível em: https://doi.org/10.1007/s10916-009-9292-7

SUBRAMANIAN, Ramanathan et al. Ascertain: Emotion and personality recognition using commercial sensors. IEEE Transactions on Affective Computing, [s. l.], v. 9, n. 2, p. 147–160, 2018. Disponível em: https://doi.org/10.1109/TAFFC.2016.2625250

SUN, Bo; LIN, Zihuai. Emotion Recognition using Machine Learning and ECG signals. [S. l.]: arXiv, 2022. Disponível em: https://doi.org/10.48550/ARXIV.2203.08477

THANAPATTHEERAKUL, Thanyathorn et al. Emotion in a Century. In: , 2018. Proceedings of the 10th International Conference on Advances in Information Technology - IAIT 2018. [S. l.]: ACM Press, 2018. Disponível em: https://doi.org/10.1145/3291280.3291788

TSAI, Chia Wei *et al.* Anomaly Detection Mechanism for Solar Generation using Semi-supervision Learning Model. Indo - Taiwan 2nd International Conference on Computing, Analytics and Networks, Indo-Taiwan ICAN 2020 - Proceedings, [*s. l.*], p. 9–13, 2020. Disponível em: https://doi.org/10.1109/Indo-TaiwanICAN48429.2020.9181310

UYAREL, Huseyin et al. Effects of anxiety on QT dispersion in healthy young men. Acta cardiologica, [s. l.], v. 61, n. 1, p. 83–87, 2006.

VAZQUEZ-RODRIGUEZ, Juan et al. Transformer-based self-supervised learning for emotion recognition. In: 2022 26th International Conference on Pattern Recognition (ICPR). IEEE, 2022. p. 2605-2612.

VIEGAS, Carla. Two Stage Emotion Recognition using Frame-level and Video-level Features. Proceedings - 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2020, [*s. l.*], v. 1, p. 912–915, 2020. Disponível em: https://doi.org/10.1109/FG47880.2020.00143

WAN, Boxin; GUO, Junqi. Learning Immersion Assessment Model Based on Multi-dimensional Physiological Characteristics. Proceedings of 2020 IEEE International Conference on Power, Intelligent Computing and Systems, ICPICS 2020, [*s. l.*], p. 87–90, 2020. Disponível em: https://doi.org/10.1109/ICPICS50287.2020.9202208

WANG, Xiaolong; HE, Kaiming; GUPTA, Abhinav. Transitive Invariance for Self-Supervised Visual Representation Learning. Proceedings of the IEEE International Conference on Computer Vision, [*s. l.*], v. 2017-Octob, p. 1338–1347, 2017. Disponível em: https://doi.org/10.1109/ICCV.2017.149

WANG, Zhengwei; SHE, Qi; WARD, Tomás E. Generative Adversarial Networks in Computer Vision: A Survey and Taxonomy. ACM Computing Surveys, [*s. l.*], v. 54, n. 2, 2021. Disponível em: https://doi.org/10.1145/3439723

WIJASENA, Hamidan Z.; FERDIANA, Ridi; WIBIRAMA, Sunu. A Survey of Emotion Recognition using Physiological Signal in Wearable Devices. AIMS 2021 - International Conference on Artificial Intelligence and Mechatronics Systems, [*s. l.*], 2021. Disponível em: https://doi.org/10.1109/AIMS52415.2021.9466092

WIOLETA, Szwoch. Using physiological signals for emotion recognition. In: , 2013. 2013 6th International Conference on Human System Interactions (HSI). [S. l.]: IEEE, 2013. Disponível em: https://doi.org/10.1109/hsi.2013.6577880

WU, Yifu; WEI, Jin; ROCHE, Rigoberto. A domain knowledge - Enabled hybrid semi-supervision learning method. GlobalSIP 2019 - 7th IEEE Global Conference on Signal and Information Processing, Proceedings, [*s. l.*], 2019. Disponível em: https://doi.org/10.1109/GlobalSIP45357.2019.8969462

WUNDT, Wilhelm Max. Grundriss der psychologie. [S. l.]: A. Kröner, 1913.

XU, Cuiting et al. A novel facial emotion recognition method for stress inference of facial nerve paralysis patients. Expert Systems with Applications, [s. l.], v. 197, p. 116705, 2022. Disponível em: https://doi.org/10.1016/j.eswa.2022.116705

YADAV, Satya Prakash et al. Survey on Machine Learning in Speech Emotion Recognition

and Vision Systems Using a Recurrent Neural Network (RNN). Archives of Computational Methods in Engineering, [s. l.], v. 29, n. 3, p. 1753–1770, 2021. Disponível em: https://doi.org/10.1007/s11831-021-09647-x

YAN, MaoSong et al. Emotion classification with multichannel physiological signals using hybrid feature and adaptive decision fusion. Biomedical Signal Processing and Control, [s. l.], v. 71, p. 103235, 2022. Disponível em: https://doi.org/10.1016/j.bspc.2021.103235

YANG, Wenlu *et al.* Physiological-Based Emotion Detection and Recognition in a Video Game Context. Proceedings of the International Joint Conference on Neural Networks, [*s. l.*], v. 2018-July, n. i, p. 1–8, 2018. Disponível em: https://doi.org/10.1109/IJCNN.2018.8489125

YASEMIN, Mine; SARIKAYA, Mehmet Ali; INCE, Gokhan. Emotional State Estimation using Sensor Fusion of EEG and EDA. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, [*s. l.*], p. 5609–5612, 2019. Disponível em: https://doi.org/10.1109/EMBC.2019.8856895

YUAN, Yue; HUANG, Jing; YAN, Ke. Virtual Reality Therapy and Machine Learning Techniques in Drug Addiction Treatment. In: , 2019. 2019 10th International Conference on Information Technology in Medicine and Education (ITME). [S. l.]: IEEE, 2019. Disponível em: https://doi.org/10.1109/itme.2019.00062

ZENG, Zhihong et al. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE transactions on pattern analysis and machine intelligence, [s. l.], v. 31, n. 1, p. 39–58, 2008.

ZENONOS, Alexandros et al. HealthyOffice: Mood recognition at work using smartphones and wearable sensors. In: , 2016. 2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops). [S. l.: s. n.], 2016. p. 1–6.

ZHANG, Jianhai et al. ReliefF-Based EEG Sensor Selection Methods for Emotion Recognition. Sensors, [s. l.], v. 16, n. 10, p. 1558, 2016. Disponível em: https://doi.org/10.3390/s16101558

ZHANG, Shujian; GONG, Chengyue; CHOI, Eunsol. Learning with Different Amounts of Annotation: From Zero to Many Labels. [*s. l.*], 2021. Disponível em: http://arxiv.org/abs/2109.04408

ZHENG, Wei Long; LU, Bao Liang. A multimodal approach to estimating vigilance using EEG and forehead EOG. Journal of Neural Engineering, [*s. l.*], v. 14, n. 2, 2017. Disponível em: https://doi.org/10.1088/1741-2552/aa5a98

ZHUANG, Ning et al. Investigating Patterns for Self-Induced Emotion Recognition from EEG Signals. Sensors, [s. l.], v. 18, n. 3, p. 841, 2018. Disponível em: https://doi.org/10.3390/s18030841

ZONG, Cong; CHETOUANI, Mohamed. Hilbert-Huang transform based physiological signals analysis for emotion recognition. In: , 2009. 2009 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT). [S. l.]: IEEE, 2009. Disponível em: https://doi.org/10.1109/isspit.2009.5407547

ZONTONE, Pamela et al. Stress Detection Through Electrodermal Activity (EDA) and Electrocardiogram (ECG) Analysis in Car Drivers. In: , 2019. 2019 27th European Signal Processing Conference (EUSIPCO). [S. l.]: IEEE, 2019. Disponível em: https://doi.org/10.23919/eusipco.2019.8902631