

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE

CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA

Coordenação de Pós-Graduação em Ciência da Computação

TESE

**Reconhecimento de emoções em sinais de EEG
via *deep learning* e reconstrução
do espaço de fase**

Moisés Roberto de Araújo Mota

Campina Grande - Paraíba - Brasil

30/03/2020

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE

CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA

Coordenação de Pós-Graduação em Ciência da Computação

**Reconhecimento de emoções em sinais de EEG
via *deep learning* e reconstrução
do espaço de fase**

Moisés Roberto de Araújo Mota

Tese submetida à Coordenação do Curso de Pós-Graduação em Ciência da Computação da Universidade Federal de Campina Grande - Campus I como parte dos requisitos necessários para obtenção do grau de Doutor em Ciência da Computação.

Área de concentração: Ciência da Computação

Linha de pesquisa: Modelos Computacionais e Cognitivos

Herman M. Gomes

Eanes T. Pereira

Orientadores

Campina Grande - Paraíba - Brasil

30/03/2020

M917r

Mota, Moisés Roberto de Araújo.

Reconhecimento de emoções em sinais de EEG via *deep learning* e reconstrução do espaço de fase / Moisés Roberto de Araújo Mota. - Campina Grande, 2020.

115 f. : il. Color.

Tese (Doutorado em Ciência da Computação) - Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática, 2020.

"Orientação: Prof. Dr. Herman Martins Gomes, Prof. Dr. Eanes Torres Pereira.

Referências.

1. Reconhecimento de Emoções. 2. Eletroencefalograma. 3. Imagens de Espaço de Fase Empilhadas. 4. Rede Convolutacional 3D. I. Gomes, Herman Martins. II. Pereira, Eanes Torres. III. Título.

CDU 004.932(043)

Reconhecimento de emoções em sinais de EEG via *deep learning* e reconstrução do espaço de fase

Moisés Roberto de Araújo Mota

Tese submetida à Coordenação do Curso de Pós-Graduação em Ciência da Computação da Universidade Federal de Campina Grande - Campus I como parte dos requisitos necessários para obtenção do grau de Doutor em Ciência da Computação.

Área de concentração: Ciência da Computação

Linha de pesquisa: Modelos Computacionais e Cognitivos

**RECONHECIMENTO DE EMOÇÕES EM SINAIS DE EEG VIA DEEP LEARNING E
RECONSTRUÇÃO DO ESPAÇO DE FASE**

MOISÉS ROBERTO DE ARAUJO MOTA

TESE APROVADA EM 30/03/2020

**HERMAN MARTINS GOMES, Ph.D, UFCG
Orientador(a)**

**EANES TORRES PEREIRA, Dr., UFCG
Orientador(a)**

**JOSEANA MACÊDO FECHINE RÉGIS DE ARAÚJO, Dra., UFCG
Examinador(a)**

**CARLOS WILSON DANTAS ALMEIDA, Dr., UFCG
Examinador(a)**

**FRANCISCO MARCOS DE ASSIS, Dr., UFCG
Examinador(a)**

**FRANCISCO MADEIRO BERNARDINO JUNIOR, Dr., UNICAP
Examinador(a)**

CAMPINA GRANDE - PB

Agradecimentos

Agradeço, principalmente, à minha esposa, por todo o apoio afetivo, financeiro e psicológico que me ofereceu, suportando sozinha todas as dificuldades que surgiram durante esta minha etapa acadêmica. Sem ela, definitivamente, eu não teria chegado ao final de todo esse processo. Não posso esquecer da minha cunhada, que revisou agilmente esta tese e fez um grande trabalho de correção. Agradeço, também, aos meus orientadores, por acreditarem em mim, pela enorme paciência que tiveram e por todo o suporte instrumental e intelectual. São os melhores orientadores que eu poderia ter. Não posso esquecer também dos meus amigos, que entenderam todas as recusas aos convites para lazer ao longo de quase sete anos. Ainda, é necessário agradecer ao programa de pós-graduação da UFCG, por todo o suporte, e ao CNPQ, pela bolsa recebida durante quase um ano. Por fim, agradeço aos meus pais, pelo incentivo e também pelo suporte financeiro. Desejo conseguir retribuir de alguma forma todas essas pessoas que me proporcionaram chegar até aqui e realizar esse sonho.

Resumo

Emoções são importantes para os seres humanos, influencia nosso comportamento social, memória e cognição. A área da computação afetiva visa a auxiliar na compreensão dos fenômenos emocionais humanos utilizando sistemas computacionais para isso. Neste contexto, definir, caracterizar e classificar emoções e suas partes constituintes a partir de sinais fisiológicos, como sinais de eletroencefalograma (EEG), expressões faciais, etc., é uma tarefa desafiadora. No entanto, ainda não se conhece um conjunto de características ótimas para o reconhecimento de emoções a partir de sinais de EEG, apesar de existirem características como PSD (*power spectral density*) e HOC (*higher-order crossings*), amplamente utilizadas pela comunidade científica. Novas pesquisas têm conseguido avançar substancialmente o estado da arte no que diz respeito ao reconhecimento de emoções a partir de EEG ao considerarem o emprego de características extraídas do espaço de fase desses sinais e classificadores baseados em técnicas de *deep learning*. Neste contexto, a presente tese investigou o emprego de imagens de reconstrução do espaço de fase de sinais de EEG, que são submetidas à uma rede convolucional 3D (uma das técnicas de *deep learning*), para aprender automaticamente características do espaço de fase de sinais de EEG para reconhecer diferentes estados emocionais de modo independente do indivíduo. Os resultados dos experimentos demonstraram a viabilidade e competitividade da abordagem proposta utilizando imagens de reconstrução do espaço de fase empilhadas de, apenas um canal de EEG FP1, alcançando acurácias de $0,84 \pm 0,07$, para quatro classes de valência-excitação, e $0,88 \pm 0,05$ para duas classes de valência e $0,94 \pm 0,01$ para duas classes de excitação, ambos utilizando a metodologia de treinamento e classificação do tipo LOSO (*leave-one-subject-out*). Esses resultados, além de reduzir a quantidade de canais necessários para a classificação ao mínimo possível, contribui para avançar o estado da arte ao apresentar uma nova abordagem para a classificação de emoções a partir de sinais de EEG.

Palavras-chave: Emoções, eletroencefalograma, imagens de espaço de fase empilhadas, rede convolucional 3D.

Abstract

Emotions are important for human beings, it influences our social behavior, memory and cognition. The affective computing area aims to assist in the understanding of human emotional phenomena using computer systems for this. In this context, defining, characterizing and classifying emotions and their constituent parts based on physiological signals, such as electroencephalogram signals (EEG), facial expressions, etc, is a challenging task. However, an optimal characteristics set for recognizing emotions from EEG signals is not yet known, although there are characteristics such as PSD (power spectral density) and HOC (higher-order crossings), widely used by the scientific community. New researches have substantially advanced the state of the art with regard to the recognition of emotions from EEG when considering the use of characteristics extracted from the phase space of these signals and classifiers based on deep learning techniques. In this context, the present thesis investigated the use of reconstruction of the EEG signals images of phase space, which are subjected to a 3D convolutional network (one of the deep learning techniques), to automatically learn the EEG signal space phase characteristics to recognize different emotional states independently of the individual. The results of the experiments demonstrated the feasibility and competitiveness of the proposed approach using images of the stacked phase space reconstruction of only one EEG FP1 channel, reaching accuracy of $0,84 \pm 0,07$, for four valence-arousal classes, and $0,88 \pm 0,05$ for two valence classes and $0,94 \pm 0,01$ for two arousal classes, both using the training and classification methodology of the LOSO type (leave-one-subject-out). These results, in addition to reducing the number of channels required for classification to the minimum possible, contribute to advancing the state of the art by presenting a new approach to the emotions classification based on EEG signals.

Keywords: Emotions, electroencephalogram, stacked phase space images, 3D convolutional network.

Lista de ilustrações

Figura 1 – Esquema 10-20 de posicionamento de eletrodos	31
Figura 2 – Espaço de fase de um pêndulo	33
Figura 3 – Modelo climático simplificado de Lorenz.	34
Figura 4 – Curva da informação mútua média da subbanda do sinal	37
Figura 5 – Dimensão de imersão da sub-banda gama	38
Figura 6 – Imagens de PSR com o intervalo de 60 segundos.	39
Figura 7 – Filtro de suavização aplicado para retirar ruídos	40
Figura 8 – Arquitetura típica de uma rede neural convolucional	41
Figura 9 – Exemplo de Max pooling	43
Figura 10 – Filtros extratores de características	44
Figura 11 – Classificação de emoções de acordo com Excitação e Valência	46
Figura 12 – Classificação das emoções da base DEAP	51
Figura 13 – Sinais fisiológicos captados pela base DEAP	52
Figura 14 – Interface web utilizada pelo DEAP para classificar os vídeos	53
Figura 15 – Distribuição de amostras da base DEAP para quatros classes	54
Figura 16 – Distribuição das amostras da base DEAP para duas classes.	55
Figura 17 – Etapas da abordagem proposta nesta pesquisa	66
Figura 18 – Exemplo de sinal de EEG dos canais frontais da base DEAP	68
Figura 19 – Sinal de EEG do canal FP1 e suas sub-bandas	69
Figura 20 – Emoções baseadas em valência e excitação	70
Figura 21 – Algoritmo de aumento de dados do Keras	72
Figura 22 – Diagrama dos experimentos iniciais	77
Figura 23 – Técnica de contagem de caixas implementada	79
Figura 24 – Resultado da compactação da imagem de PSR	80
Figura 25 – Imagem de PSR antes e após converter para tons de cinza	80
Figura 26 – Imagens de PSR da sub-banda gama em diferentes intervalos.	82
Figura 27 – Imagem de PSR empilhada.	84
Figura 28 – Evolução do treinamento do classificador	86
Figura 29 – Evolução do treinamento do classificador para duas classes de excitação	86
Figura 30 – Evolução do treinamento do classificador para duas classes de valência	87
Figura 31 – Arquitetura da rede neural convolucional utilizada para classificação. .	90
Figura 32 – Comparação dos intervalos das médias dos resultados para duas classes	93
Figura 33 – Comparação dos intervalos das médias dos resultados para quatro classes	94
Figura 34 – Possíveis regiões de ativação dos pesos da rede convolucional.	98
Figura 35 – Área de destaque dos mapas de características do classificador.	99

Lista de tabelas

Tabela 1 – Subsistemas orgânicos e funções componentes da emoção	24
Tabela 2 – Resumo do experimento para construção da base DEAP	52
Tabela 3 – Resumo dos trabalhos correlatos	62
Tabela 4 – Dados da base DEAP com o aumento de dados	73
Tabela 5 – Quantidade de amostras por classe da base DEAP	74
Tabela 6 – Configurações dos experimentos iniciais	78
Tabela 7 – Acurácia Média de diferentes tipos de balanceamento.	78
Tabela 8 – Parâmetros avaliados para configurar o classificador mais adequado. . .	81
Tabela 9 – Acurácia Média com ausência e presença da técnica WESML.	83
Tabela 10 – Resultados dos experimentos	83
Tabela 11 – Acurácia Média dos diferentes classificadores testados	84
Tabela 12 – Dados da base DEAP com o aumento de dados	85
Tabela 13 – Média das demais métricas do classificador	88
Tabela 14 – Média das demais métricas do classificador	88
Tabela 15 – Parâmetros selecionados para o classificador	89
Tabela 16 – Comparação de resultados para duas classes	90
Tabela 17 – Comparação de resultados para quatro classes	92
Tabela 18 – Comparação quanto a quantidade canais.	94
Tabela 19 – Sequência de quadros de imagens de PSR de um sinal de EEG.	95

Lista de abreviaturas e siglas

EEG	<i>Eletroencefalograma</i>
PSD	<i>Power Spectral Density</i>
HOC	<i>Higher-Order Crossings</i>
PSR	<i>Phase-Space Reconstruction</i>
DEAP	<i>Database for Emotion Analysis using Physiological Signals</i>
SNC	<i>Sistema Nervoso Central</i>
SNE	<i>Sistema Neuro-Endócrino</i>
SNA	<i>Sistema Nervoso Autônomo</i>
SNS	<i>Sistema Nervoso Somático</i>
BCI	<i>Brain-Computer Interface</i>
PCA	<i>Principal Component Analysis</i>
ICA	<i>Independent Component Analysis</i>
CNN	<i>Convolutional Neural Networks</i>
SVM	<i>Support Vector Machine</i>
RNA	<i>Redes Neurais Artificiais</i>
MLP	<i>multilayer perceptron</i>
NB	<i>Naïve Bayes</i>
KNN	<i>K-Nearest Neighbor</i>
RT	<i>Regression Tree)</i>
DBN	<i>Deep Belief Networks</i>
RBM	<i>Restricted Boltzmann Machine</i>
HMM	<i>Hidden Markov Model</i>
CLRNN	<i>Convolutional Networks and Long-Short-Term-Memory Recurrent Neural Networks</i>

PCC	<i>Pearson Correlation Coefficient</i>
MFI	<i>Multidimensional Feature Image</i>
AE	<i>Autoencoder</i>
DLN	<i>Deep Learning Network</i>
DM	<i>Density Matrix</i>
GLCM	<i>Gray Level Cooccurrence Matrix</i>
VPP	<i>Valor Preditivo Positivo</i>
VP	<i>Verdadeiros Positivos</i>
VN	<i>Verdadeiros Negativos</i>
FP	<i>Falsos Positivos</i>
FN	<i>Falsos Negativos</i>
ADASYN	<i>Adaptive Synthetic Sampling Approach for Learning from Imbalanced Datasets</i>
SGD	<i>Stochastic Gradient Descent</i>
ReLU	<i>Rectified Linear Unit</i>
LOSO	<i>Leave-One-Subject-Out</i>
LOTO	<i>Leave-One-Trial-Out</i>
BCE	<i>Binary Cross-Entropy</i>
CCE	<i>Categorical Cross-Entropy</i>
GSR	<i>Galvanic Skin Response</i>
PLV	<i>Phase Locking Value</i>
PLI	<i>Phase Lag Index</i>
DNN	<i>Deep Neural Networks</i>
FFT	<i>Fast Fourier Transform</i>
TM	<i>Topological Map</i>
PCRNN	<i>Parallel Convolutional Recurrent Neural Network</i>

LIME	<i>Local Interpretable Model-agnostic Explanations</i>
HAHV	<i>High arousal-high valence</i>
LAHV	<i>Low arousal-high valence</i>
LALV	<i>Low arousal-low valence</i>
HALV	<i>High arousal-low valence</i>
WESML	<i>Weighted exogenous sampling maximum-likelihood estimator</i>
MSVM	<i>Multi-class support vector machine</i>
MC-LS-SVM	<i>Multi-class least squares support vector machine</i>
PP	<i>Poincaré points</i>
DE	<i>Differential entropy</i>
s	Segundos

Lista de símbolos

$\vec{\xi}_t$	<i>Vetores m-dimensionais da série temporal m</i>
m	<i>Dimensão de imersão</i>
τ	<i>Passo de reconstrução ou tempo de defasagem</i>
$x(t)$	<i>Sinal no domínio do tempo</i>
$x(t+\tau)$	<i>Versão de $x(t)$ defasada de τ</i>
$I(\tau)$	<i>Função de informação mútua</i>
d	<i>Dimensão do espaço de fase real dos vetores</i>
w_i	<i>Pesos</i>
Y_i	<i>Amostra aleatória da população</i>
β	<i>Valor de maximização da função</i>
x_i	<i>Variável explanatória.</i>

Sumário

1	INTRODUÇÃO	23
2	FUNDAMENTAÇÃO	27
2.1	Sinais de eletroencefalograma (EEG)	27
2.1.1	Aquisição e processamento de sinais de EEG	30
2.1.2	Reconstrução do espaço de fase	32
2.2	Redes neurais convolucionais	39
2.3	Classificação de emoções	44
2.3.1	Reconhecimento de emoções dependente e independente	47
2.3.2	Emoções puras ou mistas	48
2.3.3	DEAP	50
2.3.4	Considerações Finais	56
3	REVISÃO BIBLIOGRÁFICA	57
3.1	Pesquisas correlatas	57
3.1.1	O uso de redes convolucionais para classificação de emoções	58
3.1.2	O uso de características da reconstrução do espaço de fase	60
3.1.3	Considerações Finais	63
4	METODOLOGIA E EXPERIMENTOS	65
4.1	Materiais e métodos	65
4.1.1	Métricas de avaliação de desempenho da classificação	69
4.1.2	Balanceamento e aumento de dados	71
4.1.3	Canais de EEG selecionados	75
4.2	Experimentos	76
4.3	Resultados e discussão	89
5	CONCLUSÕES E PESQUISAS FUTURAS	101
5.1	Conclusões	101
5.1.1	Principais contribuições	103
5.1.2	Limitações desta pesquisa	103
5.1.3	Sugestões para Pesquisas Futuras	104
	REFERÊNCIAS BIBLIOGRÁFICAS	107

1 Introdução

Nesse capítulo, são introduzidos os elementos fundamentais para a realização desta pesquisa. Estes elementos são apresentados por meio do problema em estudo - o reconhecimento de emoções a partir de sinais de EEG, bem como do contexto e das principais motivações que justificam a importância deste trabalho para a área da Computação Afetiva.

Um dos principais desafios enfrentados por quem procura abordar cientificamente o conceito de emoção, ou simplesmente usar as teorias da emoção em uma estrutura particular, é a variedade de teorias disponíveis. As teorias da emoção diferem extremamente no que diz respeito aos componentes que as representam e aos níveis de processamento que as endereçam. Essa paisagem bastante diversa pode gerar confusão, pois nem todas as teorias são boas para todos os fins. Um erro comum ao descrever o domínio das emoções como um todo é se referir à "emoção" e suas formas afins ("emocional", "emotivo", etc) de uma maneira generalizada (PICARD, 2000).

Traduzir a intuição em palavras e conceitos bem definidos é difícil, muitas palavras são naturalmente ambíguas, por exemplo: emoção, sentimento e afeto. Todas elas representam o mesmo tipo de problema, mas nas circunstâncias certas, elas podem ser usadas para se referir a emoções. No entanto, cada um delas também tem pelo menos um outro sentido, correspondente a uma parte específica do domínio (PETTA; PELACHAUD; COWIE, 2011).

No entanto, adotaremos a definição de emoção da base DEAP (KOELSTRA et al., 2012), segundo Koelstra et al. (2012), a emoção pode ser definida como um processo psicofisiológico desencadeado pela percepção consciente, ou não, de um objeto ou situação e é frequentemente associado a humor, temperamento, personalidade, disposição e motivação. Embora essa definição tenha sido adotada nesta tese, é possível verificar a discussão sobre definições e taxonomias a respeito da emoção em trabalhos como os de Frijda (2016) e Keltner (2019). Também, de acordo com Scherer (2005) e Scherer e Moors (2019), é possível associar componentes emocionais a subsistemas fisiológicos.

A Tabela 1 contém a relação entre componentes emocionais e subsistemas fisiológicos, bem como suas funções presumidas. Três dos componentes têm status de longa data como modalidades de emoção, expressão, sintomas corporais e excitação, e experiência subjetiva. A elicitación das tendências de ação e a preparação da ação também foram implicitamente associadas à excitação emocional, mas somente após a inclusão explícita dessas consequências motivacionais, essas características importantes dos episódios de emoção adquiriram o status de um componente principal. A inclusão de um componente

Tabela 1 – Subsistemas orgânicos e funções componentes da emoção.

Função Emocional	Subsistema orgânico	Componente Emocional
Avaliação de objetos e eventos	Processamento de informações (SNC)	Componente cognitivo (apreciação)
Regulação do sistema	Suporte e Preparação (SNC, SNE, SNA)	Componente Neurofisiológico (sintomas corporais)
Preparação e direção de ações	Sistema Executivo (SNC)	Componente motivacional (tendências de ações)
Comunicação de reações e intenções	Ação (SNS)	Componente da expressão motora (expressões faciais e vocais)
Monitoramento do estado interno do organismo	Monitor (SNC)	Componente do sentimento subjetivo (experiência emocional)

Notas: SNC = Sistema Nervoso Central; SNE = Sistema Neuro-Endócrino; SNA = Sistema Nervoso Autônomo; SNS = Sistema Nervoso Somático.

Fonte: Tabela traduzida de [Scherer \(2005, p. 698\)](#)

cognitivo de processamento da informação, não é consensual, pois muitos teóricos ainda preferem ver a emoção e a cognição como dois sistemas independentes que interagem entre si ([SCHERER, 2005](#); [SCHERER; MOORS, 2019](#)).

Como podemos observar, há dificuldade de diferenciar a emoção de outros tipos de fenômenos afetivos ([FRIJDA, 2016](#)). Isso é análogo à tentativa de definir a especificidade da linguagem em comparação com outros tipos de sistemas de comunicação, humanos ou animais. Mas há algumas definições bem aceitas na área, como no trabalho de [Dzedzickis, Kaklauskas e Bucinskas \(2020\)](#), que podem ser usadas para diminuir essa confusão, por exemplo, é sugerido que "os sentimentos são sempre experimentados em relação a um objeto específico de que a pessoa está ciente. Sua duração depende do tempo em que a representação do objeto permanece ativa na mente da pessoa". Portanto, o sentimento poderia ser uma reação à emoção experimentada pelo ser humano, que refletiria o processo de experiência subjetiva, servindo de gatilho para todo o fenômeno multimodal que é o episódio emocional ([SCHERER, 2005](#); [SCHERER; MOORS, 2019](#)).

Também é possível dizer que os episódios emocionais compreendem componentes diferentes, como avaliação da situação, preparação da ação, respostas fisiológicas, comportamento expressivo e sentimentos subjetivos ([SCHERER; MOORS, 2019](#)), o que torna esses eventos notoriamente difíceis de caracterizar. Apesar de existirem diferentes formas de medir e categorizar as emoções ([DZEDZICKIS; KAKLAUSKAS; BUCINSKAS, 2020](#)), mas a categorização de [Russell e Barret \(1999\)](#), eles sugerem que a emoção é caracterizada principalmente pela valência, posteriormente pela excitação, esse tipo de caracterização

é a mais utilizada nas pesquisas de reconhecimento de emoções na revisão de [Dzedzickis, Kaklauskas e Bucinskas \(2020\)](#), e também é utilizada na base DEAP ([KOELSTRA et al., 2012](#)).

Uma vez que definir, caracterizar e classificar emoções e suas partes componentes é uma atividade complexa para o ser humano, o reconhecimento da emoção baseado em sinais fisiológicos, como sinais de eletroencefalograma (EEG), expressões faciais, etc., é uma tarefa desafiadora para algoritmos e técnicas computacionais ([CHRISTENSEN; ABDULLAH, 2018](#)). Essa atividade requer a extração de um conjunto de características dos sinais de um episódio emocional. O reconhecimento de emoção aqui se refere à classificação de características em uma das várias classes de emoções. Geralmente, um classificador supervisionado pré-treinado com características emocionais como entrada, e classe de emoção como saída é usado para determinar a classe de uma manifestação emocional ([KONAR; CHAKRABORTY, 2015](#)).

Nesse contexto, surge a computação afetiva, uma área multidisciplinar que abrange ciência da computação, psicologia e ciência cognitiva ([KOELSTRA et al., 2012](#)). Ela é focada no estudo e desenvolvimento de sistemas e dispositivos que podem reconhecer, interpretar, processar e simular emoções humanas ([D'MELLO; KAPPAS; GRATCH, 2018](#)), além de utilizar avanços em neurociência para ajudar a expandir a experiência interativa entre humanos e máquinas e, efetivamente criar máquinas capazes de gerar empatia ([BASTOS-FILHO et al., 2012](#)).

Inspirado pelo elo indissolúvel entre emoções e cognição, a computação afetiva visa a reduzir o hiato de comunicação entre o humano emocional e o computador, emocionalmente desafiado pela necessidade de reconhecer e responder a estados afetivos (por exemplo, estados de ânimo e emoções). Como exemplo, um resultado da Computação Afetiva pode ser um ambiente de aprendizagem sensível às emoções, que detecta e responde à frustração dos alunos, a fim de aumentar a motivação e melhorar os ganhos de aprendizagem em oposição a um sistema que ignora o estado afetivo do aluno ([CALVO; MELLO, 2010](#); [DZEDZICKIS; KAKLAUSKAS; BUCINSKAS, 2020](#)).

De acordo com [Picard \(2000\)](#), a computação afetiva é guiada por três questões:

- Como as emoções podem ser geradas em computadores?
- Como as emoções podem ser reconhecidas pelos computadores?
- Como as emoções podem ser expressas pelos computadores?

A presente pesquisa de doutorado se insere no contexto da segunda questão. Ou seja, a tese tem como objetivo investigar a classificação de emoções, especificamente atuando na melhoria da extração de características de sinais de EEG, atuando na etapa de pré-processamento dos sinais, como também na posterior classificação.

Várias abordagens têm sido empregadas para reconhecer estados emocionais humanos. De acordo com [Meiselman \(2016\)](#) e [Robinson e Mauss \(2009\)](#), identificaram métricas que podem ser usadas no reconhecimento de emoções. Entre as métricas mencionadas, podemos destacar as seguintes:

- **Autorrelatos:** de acordo com os autores, a precisão dos resultados obtidos por meio dos autorrelatos depende da proximidade temporal entre o evento emocional e o momento do relato. Quanto mais próximo for o momento da experiência emocional, mais preciso será o relato. Além disso, pode haver uma tendência a evadir relatos de experiências emocionais negativas entre pessoas de certas classes sociais. Os estudiosos afirmam que os autorrelatos são muito importantes para capturar informações dimensionais (valência, excitação, abordagem e partida);
- **Medidas autonômicas:** medidas de ação do sistema nervoso autônomo (SNA) podem ser obtidas a partir de sinais cardiovasculares ou nível de condutibilidade da pele, devido à presença de suor. Segundo os autores, alguns estudos indicaram inconsistências nas classificações de estados emocionais usando medidas autonômicas;
- **Estados cerebrais:** avaliam o fator de assimetria dos sinais das regiões cerebrais frontais, uma vez que há evidências na literatura de que essa assimetria está relacionada à valência emocional e seu equilíbrio. Entre as respostas comportamentais evocadas pelas emoções, Mauss e Robinson analisaram em seu artigo: características vocais (relacionadas à medida de excitação), expressões faciais (relacionadas à medida de valência) e comportamento integral (relacionadas ao orgulho e embaraço).

Em consonância com as pesquisas de [Meiselman \(2016\)](#) e [Robinson e Mauss \(2009\)](#), esta tese lida apenas com o reconhecimento de emoções usando EEG, especificamente inserido nas seguintes categorias: autorrelato e estados cerebrais, que tem como objetivo discutir o processamento desses sinais, a fim de permitir a pesquisas futuras avaliar e desenvolver novos estudos no campo de reconhecimento de emoções.

Para que esse trabalho consiga avançar o estado da arte no reconhecimento de emoções, foi realizada uma revisão bibliográfica sistemática, que pode ser vista no Capítulo 2 dessa tese, para que fosse possível identificar subáreas de pesquisa que permitissem o desenvolvimento de um trabalho relevante.

2 Fundamentação

Este capítulo apresenta uma fundamentação da área de reconhecimento de emoções via sinais de eletroencefalograma (EEG), mas especificamente os tópicos referentes à base DEAP (KOELSTRA et al., 2012) utilizada nessa pesquisa, ao processamento do sinal, ao uso da reconstrução do espaço de fase (*phase-space reconstruction*, PSR) juntamente com redes neurais convolucionais e detalhes sobre o problema de reconhecimento de emoções.

2.1 Sinais de eletroencefalograma (EEG)

A emoção pode ser vista como uma expressão do estado psicológico da mente humana e dos processos de pensamento e, portanto, um tipo de evento que poderia ser captado por sinais fisiológicos como o EEG (KONAR; CHAKRABORTY, 2015). Apesar do evento emocional poder ser mapeado em todo o sistema nervoso autônomo (SNA), evidências apontam que o lobo frontal do cérebro é uma região importante para o processamento das emoções (FONSECA; ALARCAO, 2017; HISER; KOENIGS, 2018; DZEDZICKIS; KAKLAUSKAS; BUCINSKAS, 2020).

Dada a expressão de uma emoção, a tarefa de reconhecê-la é um típico problema de reconhecimento de padrões. Por exemplo, o reconhecimento da emoção baseado em sinais de EEG requer a extração de conjuntos de características do sinal de um determinado indivíduo (MEISELMAN, 2016). Reconhecimento de emoção aqui refere-se à classificação de características do sinal de EEG em uma das várias classes de emoção (KONAR; CHAKRABORTY, 2015). Geralmente, um classificador supervisionado pré-treinado com características emocionais como entrada e classe de emoção como saída é usado para determinar a classe de uma expressão emocional desconhecida (DZEDZICKIS; KAKLAUSKAS; BUCINSKAS, 2020).

As várias formas de onda do sinal de eletroencefalograma (EEG), provenientes dos múltiplos canais de aquisição, capturam informações clinicamente valiosas. Uma forma de onda pode representar um evento isolado, ou várias formas de onda podem constituir um padrão de sinal composto. Em ambos os casos, é importante desenvolver métodos para a detecção e quantificação objetiva de características do sinal com o propósito de facilitar a sua interpretação (KONAR; CHAKRABORTY, 2015). A extração de características relevantes do sinal é particularmente crucial quando se pretende conceber um sistema de classificação a partir do EEG. A filtragem do ruído, como sinais provenientes da musculatura do crânio ou da face, é outra questão importante no processamento de sinais de EEG e um pré-requisito para a análise de sinal subsequente ser confiável (NORMAN; KIRAKOWSKI; WILEY, 2018).

Conforme apontado em pesquisas anteriores, como aquelas de [Puce e Hämäläinen \(2017\)](#), [Kuppens e Verduyn \(2017\)](#), [Prochazkova e Kret \(2017\)](#), cada modalidade utilizada para o reconhecimento de emoções tem suas limitações. São exemplos de desvantagens:

- Autorrelato de estado emocional pode ser errôneo ou incompleto devido a vários fatores psicológicos;
- Padrões de fala e de expressões faciais estão fortemente ligados à interação social e, como tal, tendem a fornecer pouca informação no caso de uma pessoa que não está interagindo socialmente;
- Expressões faciais ou corporais podem ser silenciadas ou alteradas em uma tentativa de esconder o estado emocional;
- Os correlatos fisiológicos da emoção (mesmo dentro de um indivíduo) são dependentes do contexto;

Além dos argumentos baseados nos fundamentos neuro-fisiológicos da emoção e o reconhecimento da emoção com base no estado fisiológico na interação cotidiana, há ainda outra forte razão para estudar os sinais fisiológicos na análise do estado emocional ([SCHE-RER; MOORS, 2019](#)). Às vezes, as características dos sinais diretamente relacionadas as emoções não são suficientemente expressos ou conhecidos e, naturalmente, a utilização de uma característica menos expressiva pode levar à uma classificação errada. Este problema pode ser evitado tentando-se reconhecer uma emoção a partir da combinação de características de diferentes sinais fisiológicos. Tal processo é muitas vezes referido como reconhecimento de emoção multimodal ([KONAR; CHAKRABORTY, 2015](#)).

A combinação de múltiplas modalidades, em que qualquer modalidade única pode ser "ruidosa" ou ambígua (incluindo sinais fisiológicos), tem o potencial de melhorar o reconhecimento de padrões que podem fornecer pistas para os estados emocionais de interesse ([PETTA; PELACHAUD; COWIE, 2011](#)). Porém, conforme relatado nas pesquisas de [Zheng et al. \(2019\)](#), [Li et al. \(2019\)](#), [Shin, Maeng e Kim \(2018\)](#), apesar de bons resultados utilizando sinais multimodais entende-se que é possível avançar em modalidades monomodais ([CHRISTENSEN; ABDULLAH, 2018](#)). Portanto, esta tese se dedica ao reconhecimento de emoções a partir apenas do sinal de EEG.

Um dos problemas pendentes na tentativa de reconhecer emoções é fazê-lo para diferentes indivíduos, que podem expressar a mesma emoção de forma diferente ([SCHE-RER; MOORS, 2019](#)). Padrões de expressão variam de muitas maneiras - por exemplo, as mãos de uma pessoa quando ela está nervosa podem transpirar, enquanto as mãos de outra pessoa podem não ter nenhuma reação aparente.

Temperamento e personalidade dão pistas sobre esses padrões de expressão, pessoas extrovertidas tendem a ser mais expressivas do que introvertidas. Extrovertidos po-

dem falar com uma voz mais alta e com menos pausas de hesitação do que os introvertidos. Adultos e crianças que exibem a maioria das expressões faciais têm respostas de condutividade da pele mais baixas (PICARD, 2000). Recém-nascidos expressam suas emoções mais em medidas de frequência cardíaca do que expressões faciais. Os padrões emocionais também podem depender de gênero, idade, contexto e expectativas (CALVO et al., 2014; WHITTLE; SIMMONS; ALLEN, 2017).

Crianças podem ser mais expressivas para determinadas emoções de acordo com a idade. Elas também podem ser menos expressivas, de acordo com a aceitação social para determinadas emoções (LINDSEY, 2019). Em outras palavras, dado que uma emoção particular é sentida, uma variedade de fatores influencia a forma como a emoção será exibida (PICARD, 2000).

Essa situação é paralela a outro problema clássico de processamento de sinais, o da construção de sistemas de reconhecimento de fala. Um dos objetivos desses tipos sistemas pode ser reconhecer o que foi dito independentemente de quem o disse (KHAN et al., 2019). De modo que, fatores como, sotaques, timbres e intensidades de fala são importantes para a identificação correta das palavras e podem confundir o classificador (KHAN et al., 2019). De maneira semelhante, reconhecimento de emoções independente do indivíduo é uma tarefa complexa (PANDEY; SEEJA, 2019), uma vez que o sinal de EEG varia entre indivíduos e esse tipo de sinal é sensível a contexto (WHITTLE; SIMMONS; ALLEN, 2017).

Um outro problema em aberto na área de reconhecimento de emoções via sinais de EEG é a quantidade mínima de canais necessária que contenha informações emocionais suficientes para a atividade de classificação (CHRISTENSEN; ABDULLAH, 2018). Na revisão de Dzedzickis, Kaklauskas e Bucinskas (2020), por exemplo, foram encontrados diferentes trabalhos nos quais a quantidade de canais de EEG varia de quatro até quatorze. Na revisão bibliográfica realizada para esta tese identificamos pesquisas utilizando uma quantidade menor de canais como, por exemplo, o trabalho de Jalilifard, Pizzolato e Islam (2016) no qual é utilizado apenas um canal para o reconhecimento de emoções via sinal de EEG.

Também foram encontrados estudos como os de Meng, Yan e Xu (2017), Tong, Zhao e Fu (2018) e Wang, Hu e Song (2019) que adotam a estratégia de selecionar os canais de EEG com informações emocionais latentes, uma vez que utilizar todos os canais aumenta a complexidade da atividade de classificação além de inserir ruídos de canais irrelevantes, reduzindo a robustez do sistema. Wang, Hu e Song (2019) selecionam oito canais para valência e dez para excitação, e Tong, Zhao e Fu (2018) encontraram no mínimo seis canais e no máximo treze, com informações emocionais latentes.

Outras pesquisas como as de Jalilifard, Pizzolato e Islam (2016), Xu et al. (2019) e Ogino et al. (2019), conseguem classificar estados emocionais com apenas um canal

de EEG. A partir dessas observações em um problema correlato, é possível afirmar que o reconhecimento independente do indivíduo, a partir de imagens de PSR de sinais de EEG, oriundos de apenas um canal é uma contribuição para a área de reconhecimento de emoções (PANDEY; SEEJA, 2019). É importante salientar que a presente tese pode contribuir para motivar o uso de equipamentos de registro de sinais de EEG vestíveis como o do trabalho de Athavipach, Pan-Ngum e Israsena (2019), ou o uso de dispositivos mais simples como o utilizado no trabalho de Ogino et al. (2019).

Tendo em vista que o processamento de sinais de EEG é uma atividade complexa, bem como seu reconhecimento, iremos detalhar nas próximas seções os passos necessários desde a aquisição até sua classificação, discutindo as técnicas utilizadas nessa proposta.

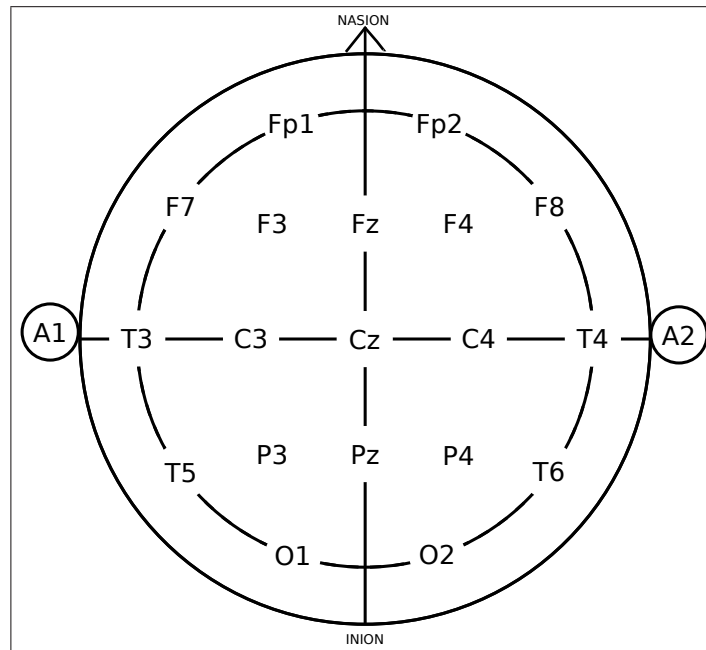
2.1.1 Aquisição e processamento de sinais de EEG

O eletroencefalograma (EEG) é uma modalidade viável para o reconhecimento da emoção. Sob um ambiente hostil, as pessoas às vezes tentam ocultar a manifestação de seus estados emocionais na expressão facial e na voz, além de algumas formas de aquisição de sinais serem invasivas, requererem contatos físicos etc (AHIRWAL; KOSE, 2018). O EEG, por outro lado, dá uma modalidade mais realista de reconhecimento da emoção, particularmente, devido a suas mudanças temporais durante a excitação da emoção, e assim, a ocultação da emoção no EEG também pode ser difícil senão impossível (KONAR; CHAKRABORTY, 2015).

Para registrar os sinais de EEG, é preciso seguir um dos esquemas de disposição de eletrodos. O sistema 10 – 20 é um dos esquemas internacionalmente aceito para colocação de eletrodos no couro cabeludo (KOELSTRA et al., 2012). Tal disposição de eletrodos busca assegurar que a maioria das funções cerebrais, tais como sistema motor, processamento de emoção, raciocínio, etc., podem ser recuperadas corretamente a partir dos sinais de EEG obtidos a partir desses canais (AHIRWAL; KOSE, 2018). No sistema 10 – 20 (mostrado na Figura 1) que consiste na vista do topo do crânio da disposição dos eletrodos.

De acordo com Mecarelli (2009), a nomenclatura dos canais do sistema 10 – 20 é dada de acordo com a região em que estão localizados, fronto-polar (FP), frontal (F), temporal (T), central (C), parietal (P) e occipital (O). E os pontos localizados sobre as linhas média do crânio são indexados pela letra "z", de "zero", por exemplo, os canais Fz e Pz. Para determinação das posições são usados dois pontos de referência, o *nasion* localizado no topo do nariz entre as sobrancelhas e o *inion* que fica na base do crânio atrás da cabeça. A distância desses dois pontos é medida e a partir deles são marcados os pontos sobre a linha mediana do crânio, Fpz e Oz com 10% da distância entre o *nasion* e o *inion* e os pontos Fz, Cz e Pz com 20% desta distância.

Figura 1 – Esquema 10-20 de posicionamento de eletrodos para aquisição de sinais de EEG



Fonte: Autoria Própria

Os eletrodos com números pares ficam localizados no hemisfério esquerdo do crânio, por exemplo Fp2, F4, F8, C4, P4, T4, T6, enquanto os números ímpares ficam no hemisfério direito, por exemplo Fp1, F3, F7, C3, P3, T3, T5. E de acordo com [Tong, Zhao e Fu \(2018\)](#) e [Konar e Chakraborty \(2015\)](#), os canais F3, F4, Fp1, e Fp2 são comumente usados para reconhecimento de emoção.

De acordo com revisões como as de [Dzedzickis, Kaklauskas e Bucinskas \(2020\)](#) e [Christensen e Abdullah \(2018\)](#) características no domínio do tempo, frequência e tempo-frequência estão entre as mais utilizadas para a classificação de emoções. Características como PSD (*power spectral density*, no domínio da , densidade do espectro de potência) ([MOON; JANG; LEE, 2018](#)) e HOC (*higher order crossings*) ([PETRANTONAKIS; HADJILEONTIADIS, 2010](#)) estão entre as características com informações robustas para a classificação de emoções ([YU et al., 2019](#)). No entanto, trabalhos como os de [García-Martínez et al. \(2017\)](#), [Zangeneh et al. \(2019b\)](#), [Zangeneh et al. \(2019a\)](#) demonstram que é possível utilizar características do espaço de fase bem como a imagem da sua reconstrução ([DAWID, 2019](#)) para o reconhecimento de estados emocionais.

Como na pesquisa de [Zangeneh et al. \(2019a\)](#) em que são utilizadas as características do espaço de fase de sinais de EEG para o reconhecimento de estados emocionais. Diferente da presente tese, que utiliza diretamente as imagens do espaço de fase da sub-banda gama do sinal de EEG.

2.1.2 Reconstrução do espaço de fase

Segundo [Zangeneh et al. \(2019b\)](#), a análise de características não-lineares tem sido utilizada para atividades de reconhecimento de emoções em sinais de EEG. De acordo com [Watt e Hameroff \(1988\)](#), imagens do espaço de fase de sinais de EEG são visualmente reconhecíveis e os padrões das trajetórias e atratores podem conter informações visuais sobre os estados do sinal. Portanto, como o sinal de EEG exibe um comportamento não-linear, é possível utilizar a reconstrução do espaço de fase para visualizar a dinâmica do sinal de EEG ([DUTTA; SINGH; KUMAR, 2018](#)).

Trabalhos como os de [Zangeneh et al. \(2019b\)](#) e [Sun et al. \(2018\)](#), [Dong, Zhou e Du \(2019\)](#), demonstram que o espaço de fase pode conter informações úteis para atividades de classificação de sinais fisiológicos como EEG e ECG, assim como nos trabalhos de [Marinus, Araújo e Gomes \(2019\)](#) e [Costa \(2012\)](#), com imagens do espaço de fase a partir de sinais de voz para classificação de patologias em pregas vocais. Há também o trabalho de [Dawid \(2019\)](#) que utiliza imagens do espaço de fase geradas a partir de sinais de EEG para a classificação de três diferentes níveis de estresse, demonstrando uma possível viabilidade da abordagem proposta nesta tese.

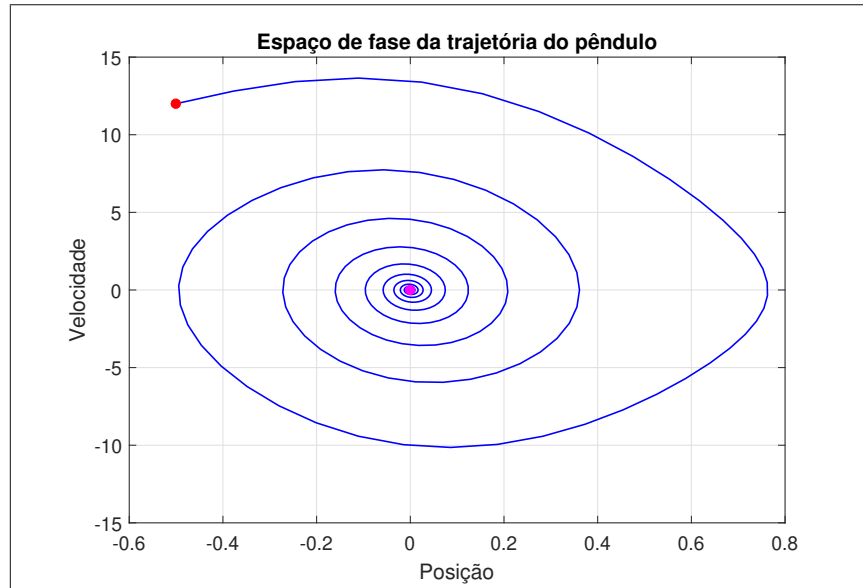
Sinais fisiológicos como os de EEG ([DAWID, 2019](#)), sinais de voz ([MARINUS; ARAÚJO; GOMES, 2019](#)) e ECG ([SUN et al., 2018](#)) são séries temporais experimentais. Em sinais experimentais é inevitável a presença de ruídos, e esses podem afetar os resultados de atividades de classificação ([COSTA, 2012](#)), desse modo é importante selecionar técnicas adequadas para análise de séries temporais.

Uma série temporal consiste em um conjunto de valores discreto e numerável de uma variável de estudo de um sistema dinâmico ([COSTA, 2012](#)), por exemplo uma série de medidas de variações de tensão ao longo do tempo em um eletrodo no lobo frontal do crânio de um paciente (EEG) ([COHEN, 2014](#)). Segundo [Marinus, Araújo e Gomes \(2019\)](#), um sistema dinâmico é um modelo matemático para a representação de diversos sistemas, que consistem em uma sequência de dados que são amostrados através do tempo ([MOON, 2004](#)). Conforme [Santos \(2019\)](#), sistemas dinâmicos, como séries temporais experimentais, podem ser representadas no espaço de fase.

De acordo com [Santos \(2019\)](#), os estudos de sistemas dinâmicos têm como um dos seus pilares [Poincaré \(1893\)](#), cujo trabalho sobre a estabilidade do sistema solar usou o pêndulo simples para exemplificar alguns conceitos e ideias. Por exemplo, na Figura 2 pode ser visto o espaço de fase de um pêndulo sob o efeito de atrito com posição x e velocidade y . Ou seja, em um sistema de duas variáveis (x, y) , o espaço de fase consiste em um gráfico cartesiano em que cada ponto representa um estado do sistema em cada instante. Desta forma, é possível afirmar que o espaço de fase permite investigar a evolução de um sistema dinâmico ao longo do tempo a partir das suas variáveis de estado ([MARINUS; ARAÚJO;](#)

GOMES, 2019).

Figura 2 – Espaço de fase de um pêndulo sob efeito de atrito



Fonte: Autoria própria

Apesar do trabalho de [Poincaré \(1893\)](#) ser seminal para o estudo do espaço de fase de sistemas dinâmicos, percebendo a utilidade do estudo das estruturas topológicas no espaço de fases de trajetórias dinâmicas, Figura 2, foi com [Lorenz \(1963\)](#) no domínio dos sistemas dissipativos, que demonstrou-se a imprevisibilidade de determinados sistemas para tempos longos, como sistemas climáticos, devido a sua sensibilidade às condições iniciais ([SANTOS, 2019](#)).

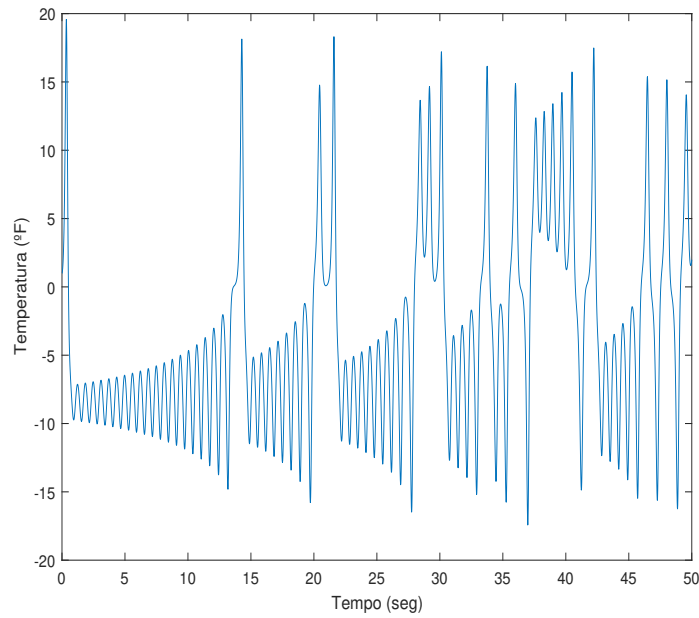
Na Figura 3a, por exemplo, vemos o modelo climático simplificado de Lorenz, enquanto a Figura 3b representa a sua reconstrução do espaço de fase em três dimensões ([LORENZ, 1963](#)). De acordo com [Lorenz \(1963\)](#), é possível visualizar no espaço de fase a impraticabilidade da previsão meteorológica para tempos longos devido às imprecisões nas determinações das condições iniciais do sistema ([SANTOS, 2019](#)).

É possível notar na Figura 3b que apesar da trajetória periódica das órbitas em torno de dois atratores, a posição dos pontos do sistema nestas órbitas é aleatória se alterando com o passar do tempo, alternando entre os dois atratores em instantes aleatórios. Essas trajetórias não utilizam o mesmo caminho novamente, mesmo visitando um estado anterior aparentemente numa região próxima no espaço.

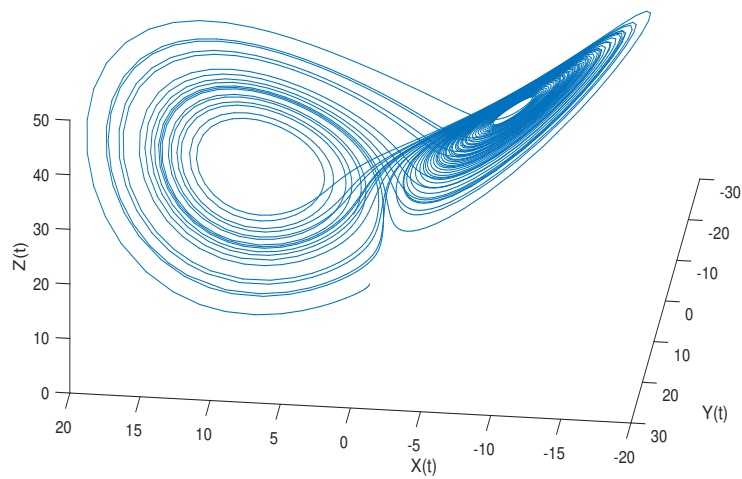
Sistemas não-lineares, como EEG ([COHEN, 2014](#)), evoluem com um comportamento instável e aperiódico, mas seu comportamento pode ser classificado como determinístico uma vez que seu estado futuro pode ser conhecido, desde que seja sabido seu estado atual ([SANTOS, 2019](#)). Porém, o seu estado futuro pode ser radicalmente alterado a par-

Figura 3 – Modelo climático simplificado de Lorenz e seu retrato no espaço de fase. (a) Modelo climático simplificado de Lorenz. (b) Retrato do espaço de fase em 3D do modelo climático.

(a) Modelo climático de Lorenz



(b) PSR do modelo climático de Lorenz



Fonte: Autoria própria

tir de pequenas alterações no estado atual. Dada a dificuldade de se conhecer o estado presente com exatidão, se faz necessário modelar o sistema não-linear como aleatório.

Desse modo, a teoria dos sistemas dinâmicos tem fornecido novas ferramentas para se analisar séries temporais obtidas em experimentos (SANTOS, 2019). De acordo com Marinus, Araújo e Gomes (2019), na teoria dos sistemas dinâmicos, existem técnicas que permitem modelar uma série temporal qualquer via a reconstrução do espaço de fase, como o teorema de imersão de Takens (*Takens embedding theorem*).

Segundo o teorema de imersão de Takens (1981), uma vez que sinais experimentais como sinais vocais (MARINUS; ARAÚJO; GOMES, 2019), e de EEG (COHEN, 2014), podem apresentar sua evolução no tempo a partir de uma única variável de estado, a variável medida (COSTA, 2012), $x(t)$, é possível converter uma sequência de medidas escalares $x(t)$ em vetores de estado (SANTOS, 2019). Assim, vetores $\vec{\xi}_t$ m -dimensionais são reconstruídos a partir da série temporal x_i , em que $x_i = x(t_i)$, $i = 1, \dots, T$.

$$\vec{\xi}_t = x(t_i), x(t_i - \tau), \dots, x(N - (m - 1)\tau) \quad (2.1)$$

Em que, o parâmetro τ representa o tempo de atraso, N é o total de amostras do sistema e m é a dimensão de imersão.

Segundo Santos (2019), o teorema de Takens (TAKENS, 1981) é um teorema sobre existência do tempo de atraso τ da reconstrução do espaço de fase, bem como a dimensão de imersão m . A dimensão de imersão pode ser obtida por vários métodos, por exemplo, o método dos falsos vizinhos próximos. Nesta tese, utilizaremos o método dos falsos vizinhos próximos conforme adotado nas pesquisas de Watt e Hameroff (1988), Costa (2012), Marinus, Araújo e Gomes (2019) e Dawid (2019).

Uma escolha adequada do parâmetro τ é importante para uma reconstrução do espaço de fase de modo que as características topológicas do sinal original sejam mantidas (MARINUS; ARAÚJO; GOMES, 2019). Um tempo de atraso τ muito curto não capturará a dinâmica dos dados, enquanto um atraso de reconstrução muito grande tornará os vetores reconstruídos completamente não relacionados em um sentido estatístico (COSTA, 2012). De acordo com Santos (2019), Fraser e Swinney (1986) propuseram um dos métodos conhecidos e populares para determinar o atraso de reconstrução adequado, o método da informação mútua média.

O método da informação mútua média, proposto por Fraser e Swinney (1986), consiste em escolher o passo τ , de modo que uma medida realizada entre um determinado instante de tempo t e um tempo posterior $t + \tau$, que minimize a informação mútua contida em vetores vizinhos ao longo de sua evolução. O valor de τ é determinado pelo ponto em que ocorre o primeiro mínimo na curva de informação mútua. Desse modo, medindo a dependência do sinal consigo mesmo, garante-se assim a reconstrução de vetores com

o menor nível de informação redundante (estatisticamente independentes), mas ainda correlacionados (COSTA, 2012; MARINUS; ARAÚJO; GOMES, 2019; SANTOS, 2019).

De acordo com Fraser e Swinney (1986) a informação mútua média entre $x(t)$ e sua versão defasada $x(t + \tau)$ pode ser obtida a partir de um histograma de b intervalos de classe (*bins*), a partir da distribuição de probabilidade de um sinal ($x(t)$) (MARINUS; ARAÚJO; GOMES, 2019; SANTOS, 2019), conforme a Equação 2.2:

$$I(\tau) = \sum_{i=1}^b \sum_{j=1}^b P_{(i,j)}(x(t), x(t + \tau)) \log_2 \left[\frac{P_{(i,j)}(x(t), x(t + \tau))}{P_{(i)}(x(t))P_{(j)}(x(t + \tau))} \right] \quad (2.2)$$

Em que $P_{(i)}(x(t))$ é a probabilidade do sinal $x(t)$ assumir um valor dentro do i -ésimo intervalo do histograma, $P_{(j)}(x(t + \tau))$ é a probabilidade de $x(t + \tau)$ estar no j -ésimo intervalo e $P_{(i,j)}(x(t), x(t + \tau))$ é a probabilidade de que, simultaneamente, $x(t)$ esteja no i -ésimo intervalo e $x(t + \tau)$ no j -ésimo intervalo (COSTA, 2012).

Então, o tempo de defasagem adequado corresponde ao valor de τ para o qual a função de informação mútua média atinge o seu primeiro mínimo local (FRASER; SWINNEY, 1986; COSTA, 2012; MARINUS; ARAÚJO; GOMES, 2019; SANTOS, 2019). Com isso, garante-se que o espaço de estados reconstruído seja topologicamente equivalente ao espaço de estados do sistema dinâmico original. Na Figura 4 é apresentada a curva de informação mútua para obtenção do valor de τ ótimo. A Figura 4a mostra o primeiro mínimo da função de informação mútua, e o espaço de fase reconstruído do sinal 4b.

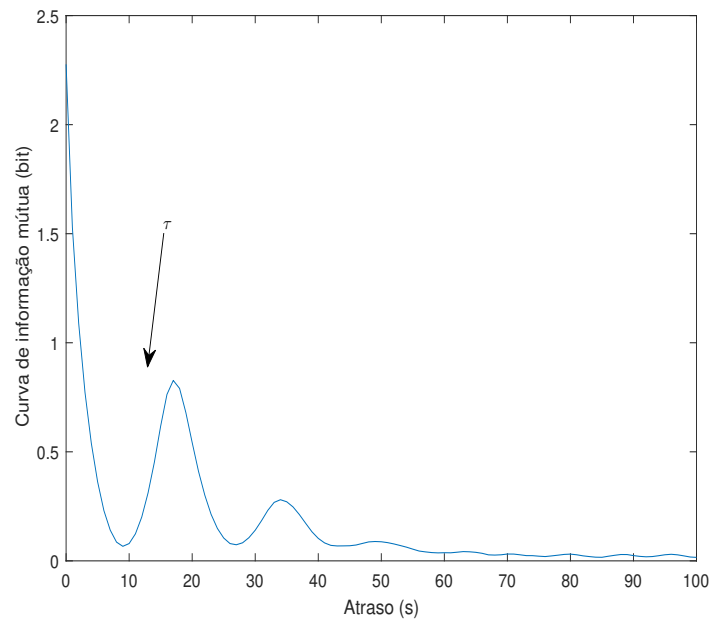
De acordo com Cohen (2014), o sinal de EEG compreende, pelo menos, quatro dimensões (tempo, espaço, frequência, potência e fase), sendo que potência e fase são elementos discretos de uma dimensão e fornecem informação amplamente independente. Apesar da multidimensionalidade do sinal de EEG, é possível analisá-lo no espaço de fase a partir da imagem da reconstrução do espaço de fase em duas dimensões como nos trabalhos de Watt e Hameroff (1988) e Dawid (2019), uma vez que o sinal base está também em duas dimensões (tempo e frequência). Portanto, mesmo calculando o parâmetro m adequado, nesta pesquisa utilizaremos a projeção do espaço de fase em duas dimensões.

Ainda assim, calculamos a dimensão de imersão m , que segundo Gao et al. (2007), pode ser obtida a partir do método dos falsos vizinhos mais próximos. De acordo com Costa (2012), quando o número de falsos vizinhos mais próximos cai a zero, o atrator terá sido suficientemente desdobrado e assim é possível identificar a menor dimensão de imersão m capaz de representá-lo adequadamente.

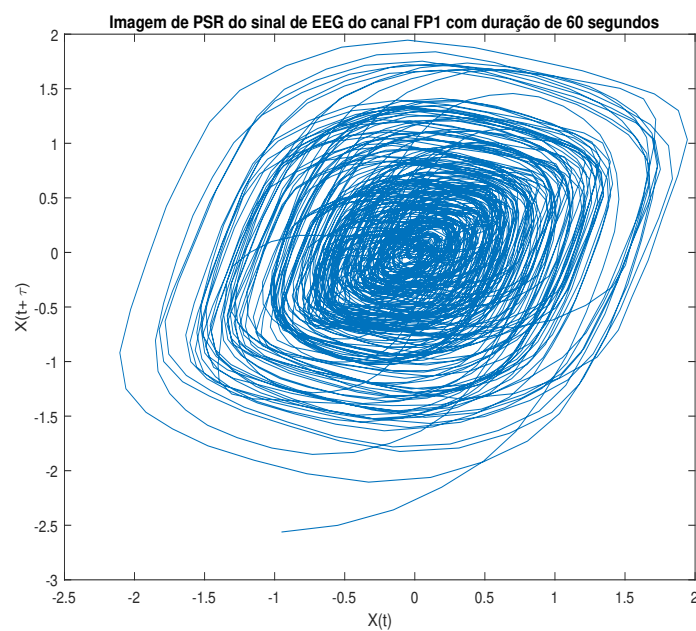
Para isso, projeta-se o sinal a uma determinada dimensão d do sistema comparando-a com uma projeção em uma dimensão maior $d+1$, até que não haja falsos vizinhos (COSTA, 2012). É possível calcular o falso vizinho mais próximo a partir da Equação 2.3 (KENNEL;

Figura 4 – Curva da informação mútua média da subbanda do sinal de EEG. (a) Curva da informação mútua média do sinal, em que a seta aponta para o τ . (b) Retrato do espaço de fase do sinal.

(a) Informação Mútua Média do sinal de EEG do canal FP1

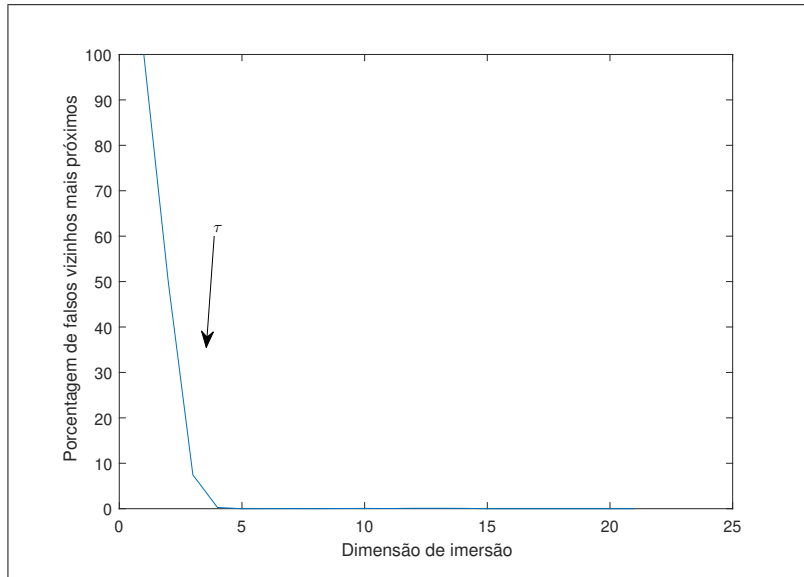


(b) Imagem de PSR do sinal de EEG do canal FP1



Fonte: Autoria própria

Figura 5 – Dimensão de imersão da sub-banda gama utilizada na reconstrução do espaço de fase, em que a seta aponta para o τ .



Fonte: Autoria própria

BROWN; ABARBANEL, 1992).

$$\left[\frac{R_{d+1}^2(n, r) - R_d^2(n, r)}{R_d^2(n, r)} \right]^{\frac{1}{2}} = \frac{|x(n + d\tau) - x^{(r)}(n + d\tau)|}{R_d(n, r)} > R_{tol} \quad (2.3)$$

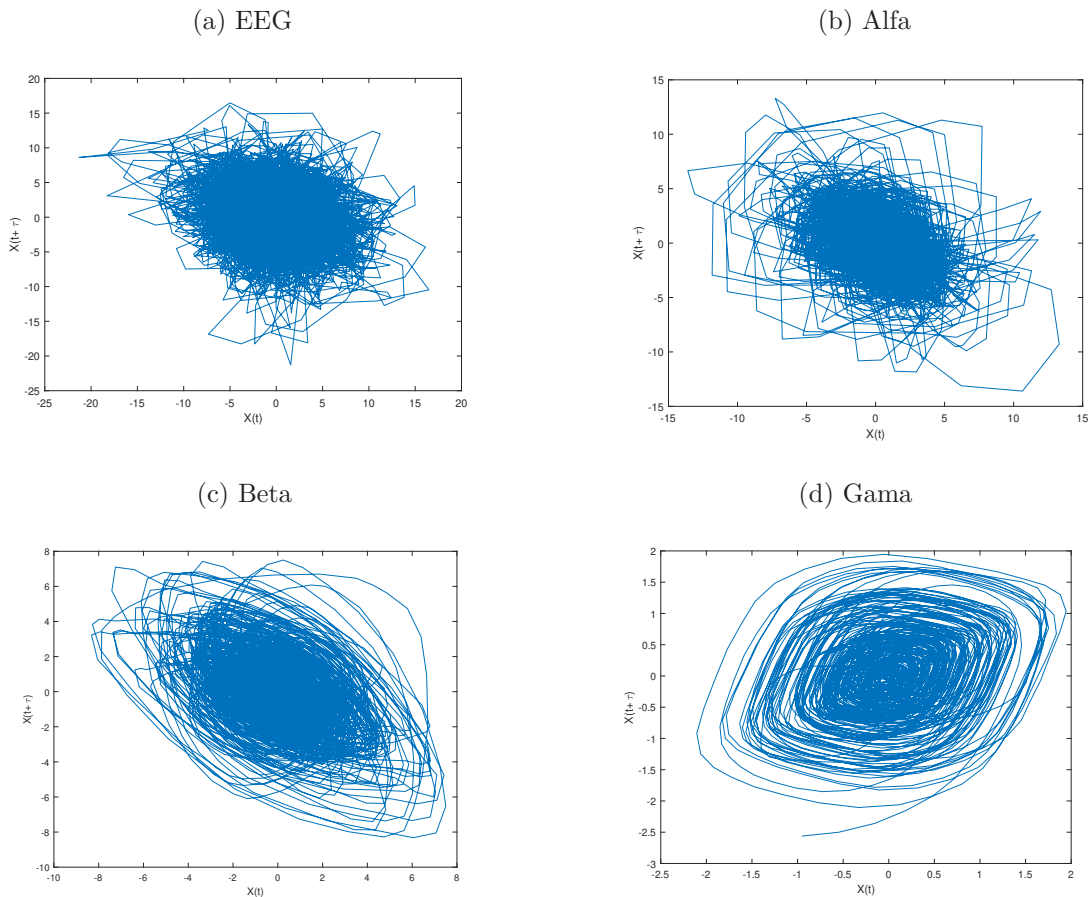
Em que R_d representa a distância no espaço de fase com a dimensão de imersão d , R_{d+1} a distância com a dimensão $d + 1$, n é o ponto da série temporal no espaço de fase, r é o vizinho mais próximo, x é a série temporal, τ é o tempo de atraso. R_{tol} é um limiar de tolerância que pode ser definido empiricamente (KENNEL; BROWN; ABARBANEL, 1992).

Portanto, a partir da Equação 2.3 é possível gerar a Figura 5, na qual observa-se a quantidade de vizinhos mais próximos se reduzindo até que a quantidade de dimensões seja $m = 4$. Diante do exposto, demonstra-se que, para representar o espaço de fase do sinal de EEG apropriadamente seriam necessárias, no mínimo, quatro dimensões.

Uma vez definidas as formas de calcular o atraso temporal τ e a dimensão de imersão m , podemos gerar as imagens de reconstrução de espaço de fase para o sinal de EEG e suas sub-bandas, vide a Figura 6. Nas Figuras 6a, 6b, 6c, 6d é possível perceber que as trajetórias em torno do atrator formam uma estrutura visualmente densa, embora seja possível diferenciar as sub-bandas do sinal original.

Uma vez geradas as imagens de PSR do sinal de EEG e das suas sub-bandas, Figura 6, o passo seguinte consistiu em encontrar o classificador adequado para operar

Figura 6 – Imagem de PSR a partir do sinal de EEG com o intervalo de 60 segundos do sinal de EEG e suas sub-bandas. (a) PSR do sinal de EEG sem extração de sub-bandas. (b) PSR da sub-banda alfa do sinal de EEG. (c) PSR da sub-banda beta do sinal de EEG. (d) PSR da sub-banda gama do sinal de EEG.



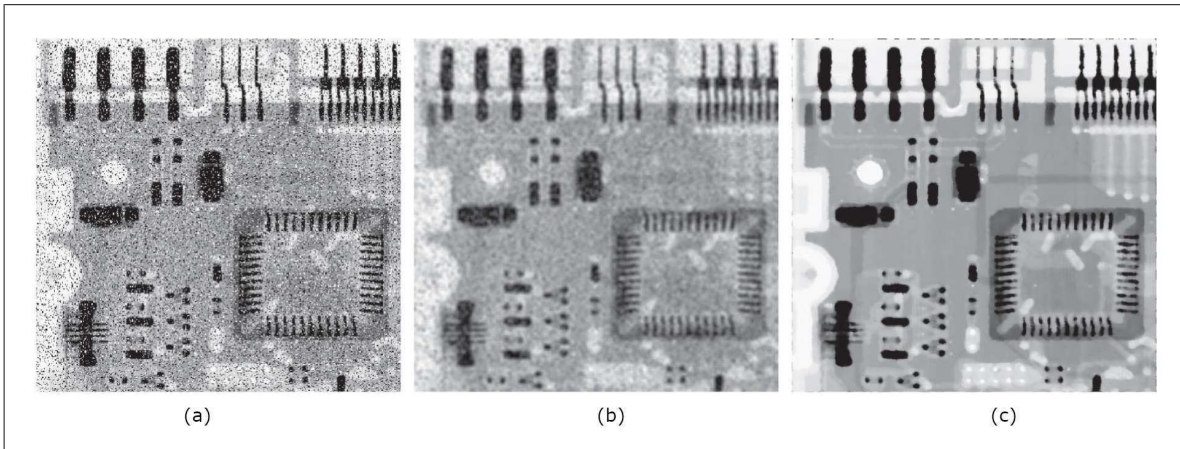
Fonte: Autoria própria

com este tipo de sinal. A próxima seção, portanto, apresenta o algoritmo das redes neurais convolucionais modelos adequados à extração de características e classificação de imagens, exatamente o mesmo tipo de problema investigado nesta tese, que trata do reconhecimento de emoções a partir das imagens de PSR do sinal de EEG.

2.2 Redes neurais convolucionais

Convolutional Networks ou *Convolutional Neural Networks* (CNN), também conhecidas como redes neurais convolucionais, são técnicas que têm despertado bastante interesse nos campos da inteligência artificial nos últimos anos (GOODFELLOW; BENGIO; COURVILLE, 2016). Dentre as abordagens existentes, as CNN têm sido bem sucedidas em se adaptar a diferentes tipos de problemas e extrair características de sinais como EEG (TRIPATHI et al., 2017; MOON; JANG; LEE, 2018; WANG et al., 2018).

Figura 7 – Filtro de suavização aplicado para retirar ruídos. (a) imagem original, (b) imagem com ruído e (c) imagem com o ruído suavizado.



Fonte: (GONZALEZ; WOODS, 2018, p. 252).

Uma rede convolucional se refere a uma rede neural especializada para o processamento de dados com uma topologia do tipo grade (*grid-like topology*). Como séries temporais podem ser descritas em grades unidimensionais (1D) ou em imagens que podem ser organizadas bidimensionalmente (2D), por exemplo, esse tipo de rede é adequada para o processamento de séries temporais (GOODFELLOW; BENGIO; COURVILLE, 2016).

A rede convolucional é assim denominada devido ao emprego da operação matemática conhecida como convolução, que diz respeito a um tipo especializado de operação linear. De maneira geral, a convolução é uma operação espacial linear que executa uma soma de produtos entre uma imagem f e um kernel de filtro, w (GONZALEZ; WOODS, 2018). Para exemplificar, suponha que temos uma imagem com ruído e precisamos suavizá-lo, vide Figura 7. Para tanto, é aplicado um filtro de suavização com o objetivo de reduzir o ruído. Para aplicar um filtro de suavização na imagem é preciso executar uma operação de convolução entre a imagem f e o filtro w (GONZALEZ; WOODS, 2018).

A aplicação de um filtro a uma imagem funciona a partir da soma dos produtos do pixel central e sua vizinhança para qualquer ponto (x, y) resultando na saída $g(x, y)$ da imagem. Então o filtro é movido por toda extensão da imagem de modo que o centro do filtro esteja sempre alinhado com o pixel referencial. Sendo assim, a operação de convolução é tipicamente denotada por um asterisco (GOODFELLOW; BENGIO; COURVILLE, 2016), como pode ser visto na Equação (2.4).

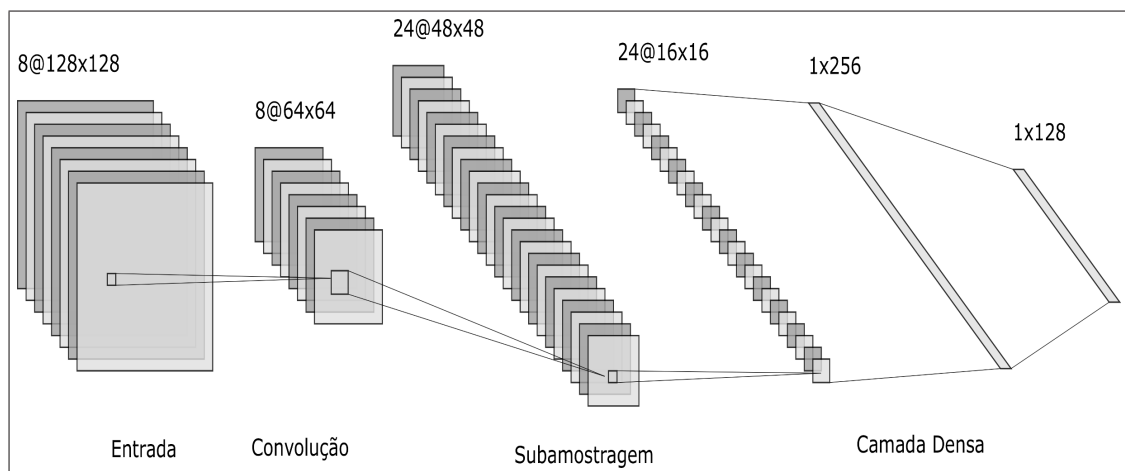
$$S(i, j) = (I * K)(i, j) = \sum_{m=1}^M \sum_{n=1}^N I(m, n)K(i - m, j - n) \quad (2.4)$$

Em que $I(m, n)$ representa a imagem nas coordenadas $m \times n$, K é o *kernel* do filtro que

irá convoluir sobre ela e M , N são a largura e altura da imagem, respectivamente.

Uma das motivações para utilizar esse tipo de rede neural se encontra em trabalhos como os de Rouast, Adam e Chiong (2018), Alaku e Türkoğlu (2019) e Rim et al. (2020). Técnicas de *deep learning* como redes convolucionais são promissoras para a classificação de sinais fisiológicos, como EEG, tendo propiciado avanços no estado da arte em reconhecimento de emoções (RIM et al., 2020).

Figura 8 – Arquitetura típica de uma rede neural convolucional



Fonte: Autoria própria.

Os dados de entrada de uma rede convolucional geralmente são matrizes multidimensionais, essas matrizes chamam-se tensores. Um tensor pode ser definido, no geral, como uma matriz de números dispostos em uma grade regular com um número variável de eixos Goodfellow, Bengio e Courville (2016). A aplicação de filtros de convolução nos dados de entrada da rede convolucional cria um mapa de características, também conhecido como mapa de ativações, com a mesma profundidade dos dados de entrada. Ao usar mais filtros (ou *kernels*), a rede é capaz de preservar melhor as dimensões espaciais e as convoluções em mais de um eixo por vez (GOODFELLOW; BENGIO; COURVILLE, 2016). Por exemplo, se usarmos uma imagem bidimensional I como entrada, provavelmente também queremos usar um *kernel* bidimensional K , conforme Equação 2.4.

Para ilustrar, ao processar uma imagem de entrada com milhares ou milhões de pixels, é possível detectar pequenas características significativas, como bordas com *kernels* que ocupam apenas dezenas ou centenas de pixels. Isso significa que precisamos armazenar menos parâmetros. No caso de convolução, a forma particular de compartilhamento de características faz com que ela seja equivariante à translação. Dizer que uma função é equivariante significa que, se a entrada for alterada, a saída será alterada da mesma maneira. Especificamente, uma função $f(x)$ é equivariante para uma função $g(x)$ se $f(g(x)) = g(f(x))$. De modo que, seja g qualquer função que desloque a entrada, então

a função de convolução é equivariante para g (GOODFELLOW; BENGIO; COURVILLE, 2016).

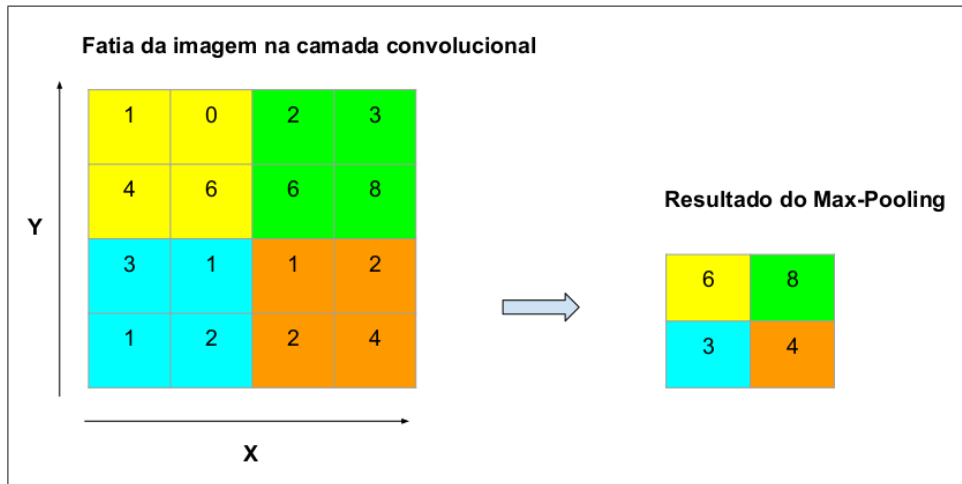
A título de exemplificação, seja uma função de mapeamento de uma função de imagem para outra função de imagem, tal que $I' = g(I)$ é a função de imagem e $I'(x, y) = I(x - 1, y)$, é a imagem I com cada pixel deslocado em uma unidade para a direita. Se essa transformação for aplicada a I e, em seguida, a convolução for aplicada, o resultado será o mesmo que se aplicássemos convolução a I , e depois aplicarmos a transformação à saída. Ao processar dados de séries temporais, isso significa que a convolução produz um tipo de linha de tempo que mostra quando diferentes características aparecem na entrada. Se movermos um evento mais tarde na entrada, a mesma representação dela aparecerá na saída, apenas no final do tempo. Da mesma forma, com imagens, a convolução cria um mapa 2D de onde determinados recursos aparecem na entrada. No entanto, a convolução não é naturalmente equivariante a algumas outras transformações, como alterações na escala ou rotação de uma imagem. Outros mecanismos são necessários para lidar com esses tipos de transformações (GOODFELLOW; BENGIO; COURVILLE, 2016).

De acordo com Goodfellow, Bengio e Courville (2016), produzir inferências com redes convolucionais, no geral, requer três etapas, a saber:

- Na primeira etapa, a camada executa várias convoluções em paralelo para produzir um conjunto de ativações lineares (mapa de ativações ou de características);
- Na segunda etapa, cada ativação linear é processada através de uma função de ativação não linear, como a função de ativação linear retificada (relu). Este estágio também pode ser chamado de estágio de detecção;
- Na terceira etapa, utiliza-se um processo de *pooling*, que tem o propósito de reduzir a dimensionalidade da saída da camada.

A Figura 9 demonstra como funciona o processo de *pooling*, que consiste em um filtro convolucional que desliza sobre a imagem, com um deslocamento (*stride*) igual a 2, agrupando os pixels da imagem por meio do uso de uma função especial de re-amostragem. O *max pooling*, por exemplo, agrupa os pixels selecionando o valor máximo dentro do filtro na imagem e, em seguida, se desloca, se o passo for 2, 2 pixels, para selecionar novos valores. O resultado dessa operação pode ser definido como mapa de características, ou ativações, de uma rede convolucional. A camada de *pooling*, é também referida como uma camada de *downsampling*. Nesta categoria, existem opções desse tipo de função, como *maxpooling*, *L2-norm pooling* e o *average pooling*.

O *pooling* auxilia a tornar a representação aproximadamente invariante para pequenas translações do sinal de entrada. Invariância à translação significa que, se trasladarmos a entrada, os valores da maioria das saídas agrupadas não serão alterados. Essa

Figura 9 – Exemplo de *max pooling* com um tamanho 2×2 e passo 2

Fonte: Autoria Própria

propriedade pode ser muito útil em situações em que alguma característica está presente na imagem mas não é necessário informar para o algoritmo sua localização exata (GOODFELLOW; BENGIO; COURVILLE, 2016). Citando caso análogo, ao determinar se uma imagem contém um rosto, não é necessário saber a localização dos olhos com precisão de pixels, só é necessário saber que há um olho no lado esquerdo do rosto e um olho no lado direito. Em outros contextos, é mais importante preservar a localização de uma característica. Assim, se o objetivo for encontrar um canto definido por duas arestas reunidas em uma orientação específica, precisamos preservar a localização das arestas o suficiente para testar se elas se encontram (GOODFELLOW; BENGIO; COURVILLE, 2016).

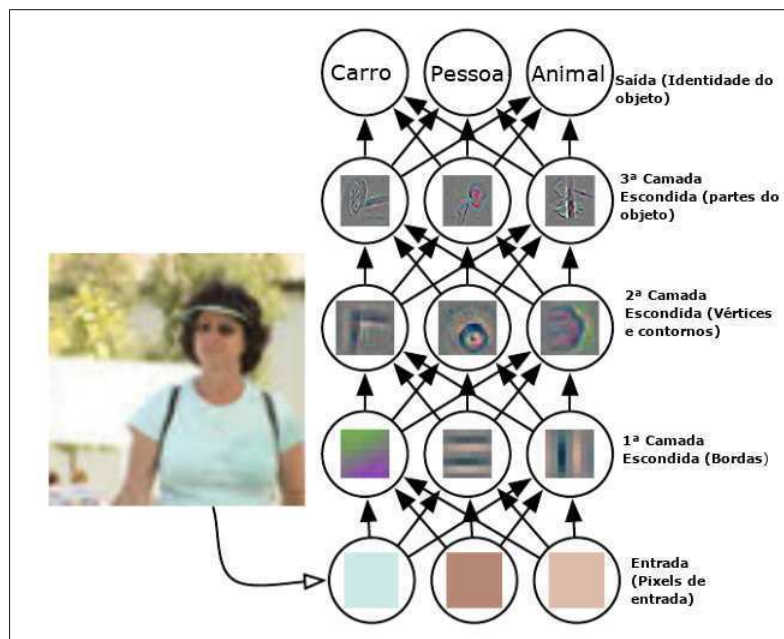
Para muitas tarefas, o *pooling* é essencial para manipular entradas de tamanho variável. Para ilustrar, se o objetivo é classificar imagens de tamanho variável, a entrada para a camada de classificação deve ter um tamanho fixo. Isso geralmente é feito variando-se o tamanho de um deslocamento entre as regiões de *pooling* para que a camada de classificação sempre receba o mesmo número de características abstraídas, independentemente do tamanho da entrada externa. Desse modo, a camada final de *pooling* da rede pode ser definida para gerar quatro conjuntos de características abstraídas, uma para cada quadrante de uma imagem, independentemente do tamanho desta (GOODFELLOW; BENGIO; COURVILLE, 2016).

Outro parâmetro importante para a rede convolucional é conhecido como *padding*. É uma variável que adiciona linhas e colunas a entrada com valor zero (*zero padding*) com o objetivo de evitar a perda de informação na borda da imagem para preservar as características espaciais nas primeiras camadas da rede (YANG et al., 2018).

O raciocínio intuitivo por trás da rede convolucional é que, uma vez que sabemos

que uma característica específica está no volume de entrada original, sua localização exata não é tão importante pois a rede consegue aprendê-la, de modo que cada camada da rede convolucional se especializa em um tipo de característica da imagem de entrada, como mostra Figura 10. Dessa forma, as camadas posteriores à primeira recebem um mapa de ativações da camada anterior, e a cada camada da rede aprende características de mais baixo nível à medida que a proximidade com a última camada aumenta (GOODFELLOW; BENGIO; COURVILLE, 2016).

Figura 10 – Filtros extratores de características de uma rede neural convolucional



Fonte: Goodfellow, Bengio e Courville (2016, p. 6)

Conforme mencionado anteriormente, redes convolucionais são algoritmos indicados para tarefas que envolvam séries temporais (*time-series*) e imagens (GOODFELLOW; BENGIO; COURVILLE, 2016). Como, no presente trabalho, foram geradas imagens de PSR a partir de sinais de EEG, optamos por utilizar redes convolucionais para o reconhecimento de emoções fazendo uso dessas imagens.

Uma vez definido e apresentado o extrator de características, será possível passar para a etapa do reconhecimento de emoções. Na próxima seção, são definidos os grupos de emoções investigados nesta pesquisa, assim como a teoria subjacente.

2.3 Classificação de emoções

Os teóricos discutiram há muito tempo um conjunto de categorias para descrever estados emocionais, foi sugerido por, por exemplo, que existem oito emoções básicas: medo,

raiva, angústia, alegria, nojo, surpresa, interesse e vergonha (PICARD, 2000). Posteriormente se distinguiu entre oito diferentes emoções básicas: medo, raiva, tristeza, alegria, nojo, surpresa, aceitação e antecipação. Atualmente existem diversas listas com diferentes emoções consideradas básicas (FRIJDA, 2016; KELTNER, 2019).

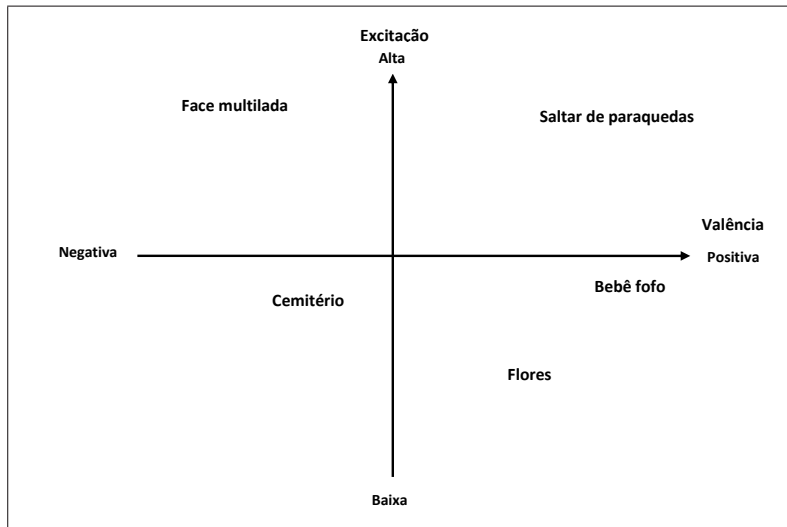
Segundo Ekman (1999), foram identificadas em seus experimentos seis emoções mais comuns, a saber, medo, raiva, tristeza, alegria (sinônimos próximos, como alegria e felicidade, podem ser combinados), nojo e surpresa. Depois destas seis, a quantidade de emoções divergem entre pesquisadores (FRIJDA, 2016; KELTNER, 2019). Estas emoções universalmente identificadas por Ekman (1999), foram denominadas emoções básicas.

As "emoções básicas" podem ser definidas de muitas maneiras, talvez a definição mais conhecida tenha sido dada por Paul Ekman, que ligou emoções básicas àquelas que têm expressões faciais universais distintas associadas, bem como outras oito propriedades (EKMAN, 1992). Por estes critérios, Ekman identificou seis emoções básicas: medo, raiva, tristeza, felicidade, nojo e surpresa. As emoções básicas também podem ser deduzidas pela análise de palavras associadas à emoção, uma abordagem tomada por Johnson-laird e Oatley (1989), em que foram elencados 590 termos em inglês que descrevem emoções. Os estudiosos concluíram que as palavras poderiam ser baseadas em uma ou mais das cinco emoções básicas: medo, raiva, tristeza, felicidade e desgosto. A partir dessa análise é possível inferir que algumas emoções aparecem universalmente, e outras parecem envolver especificidades culturais (PICARD, 2000).

Alguns autores têm se preocupado menos com a enumeração de oito ou mais emoções básicas e, ao invés disso, se referem a dimensões contínuas da emoção. Três dimensões aparecem mais comumente na literatura, embora apenas os nomes de duas sejam amplamente acordadas. As duas dimensões mais comuns são "excitação" (calma / excitada) e "valência" (negativa / positiva), *arousal* e *valence*, respectivamente, da língua inglesa. Esses foram os eixos ilustrados anteriormente na Figura 11, juntamente com títulos de imagens classificadas nesse espaço contínuo, de acordo com o trabalho de (BRADLEY; LANG, 1994). Bradley e Lang (1994) reuniram um arquivo internacional de imagens classificadas por excitação e valência, para construir o gráfico da Figura 11. Numerosos pesquisadores têm trabalhado com dimensões de emoção ao invés de emoções básicas ou categorias discretas de emoções. Bradley e Lang (1994) escreve que os autorrelatos entre os sujeitos são mais confiáveis em relação às dimensões do que com categorias discretas, como raiva, medo, etc. (PICARD, 2000).

A falta de uma definição única para o conceito de emoção e a falta de concordância sobre se existem emoções básicas ou espaços contínuos de emoções são obstáculos aos objetivos de reconhecimento de estados emocionais. No entanto, esses obstáculos não são intransponíveis. Similarmente, obstáculos ocorrem em campos como análise de conteúdo de imagem em que, apesar das dificuldades, análise de padrões e ferramentas de

Figura 11 – Classificação de emoções de acordo com excitação e valência



Fonte: Figura traduzida de [Picard \(2000, p. 100\)](#).

aprendizagem têm se mostrado úteis. Portanto, é razoável esperar um sucesso semelhante na modelagem de padrões afetivos. Além disso, a questão de tentar representar emoções fazendo uso de categorias discretas ou dimensões contínuas pode ser considerada uma escolha, pois cada representação tem vantagens em diferentes aplicações ([PICARD, 2000](#)).

Categorias discretas de emoções podem ser tratadas em um espaço contínuo, por exemplo, um sentimento de tristeza pode ocorrer em "tristeza" e "melancolia". Pesquisadores como Paul Ekman definiram fenômenos afetivos que não são emoções básicas, como "dor", para que não sejam emoções, mas "complicações emocionais". Alternativamente, pode-se considerar a dor como sendo uma emoção cognitivamente gerada, ou talvez uma mistura de emoções mais básicas. Desse modo é possível concluir que as emoções básicas são mais fáceis de retratar (felicidade, raiva, tristeza), e as emoções como o sofrimento envolvem mais esforço (depressão, dor, frustração) ([PICARD, 2000](#)).

Em resumo, os problemas de reconhecimento e modelagem de emoções são simplificados por meio da suposição de um pequeno conjunto de emoções "básicas" discretas, ou pela suposição de um pequeno número de dimensões contínuas. Ambos produzem uma representação concisa e possuem suas vantagens e desvantagens, no entanto, mesmo sendo simplificações da realidade, tais representações constituem um bom ponto de partida para possibilidades de aplicação de diferentes abordagens para a classificação de emoções. É interessante notar também que considerar um pequeno repertório de emoções tem respaldo nas diferentes etapas do desenvolvimento humano, por exemplo: um bebê tem um repertório de emoções menor do que uma criança, e a criança, por sua vez, possui um repertório menor do que um adulto ([PICARD, 2000](#)).

2.3.1 Reconhecimento de emoções dependente e independente do indivíduo

O estudo da classificação de emoções torna-se complexo devido aos muitos questionamentos acerca da teoria da emoção, tais como: de que modo a emoção é definida, expressa, e comunicada? Diferentes línguas não necessariamente usam as mesmas palavras para descrever fenômenos emotivos, o que dificulta ainda mais as tentativas de demonstrar sua universalidade. Um dos benefícios potenciais da computação afetiva reside na sua capacidade de fazer medições e analisar padrões de sinais afetivos, condicionados aos indivíduos e às circunstâncias que os afetam. Dadas as condições, medidas e padrões de respostas semelhantes, pode-se chegar a conclusões sobre a universalidade de vários tipos de expressões afetivas. Portanto, a solução proposta pela computação afetiva é, em primeiro lugar, a medição de dados pessoais para indivíduos de todos os tipos e, em segundo lugar, a análise universal dos dados individuais para ver quais padrões comuns estão presentes (PICARD, 2000).

Padrões comuns são esperados para emoções universais, e podem diferir ligeiramente entre diferentes indivíduos. Por exemplo, no contexto de faces, mais de 60 expressões de raiva foram encontradas (EKMAN, 1992), mas todos os membros da família raiva incluem três características comuns: as sobrancelhas são abaixadas e arqueadas em conjunto, a parte superior da pálpebra é levantada e o músculo nos lábios é apertado. Variações nesta expressão básica levantam hipóteses para refletir se a raiva é controlada, espontânea, simulada e assim por diante (EKMAN, 1992). Como essas variações tendem a ocorrer com frequências diferentes em indivíduos diferentes, elas podem invocar várias outras respostas individuais, talvez reforçadas pelo ambiente em que a pessoa se localiza, reforçando, assim, a tese da variação cultural das emoções, sendo influenciadas, inclusive, por dialetos. Fatores individuais como temperamento afetam os limiares da expressão, bem como outras características fisiológicas (EKMAN, 1992).

No presente trabalho, optamos por abordar os dois tipos de reconhecimento de emoções, tanto o dependente do indivíduo, quanto o independente. Entendemos que ambos ainda possuem muitas oportunidades de pesquisa como deixam claro as revisões de Santhiya e Chitrakala (2019) e Fonseca e Alarcao (2017). Essas pesquisas destacam, entre outras oportunidades, o reconhecimento de emoções entre diferentes indivíduos como um problema ainda a ser resolvido, bem como o reconhecimento em uma única pessoa ainda precisa ser refinado, principalmente no que diz respeito a definir um conjunto de características ótimas para esse tipo de tarefa de reconhecimento de padrões. Outro tópico ainda em aberto na área de reconhecimento de emoções, como destaca Fonseca e Alarcao (2017), e corroborado posteriormente por Santhiya e Chitrakala (2019), é a complexidade das emoções no mundo real, e sua provável falta de pureza, uma vez que podemos sentir mais de uma emoção ao mesmo tempo, a depender do estímulo, conforme discutido na próxima subseção.

2.3.2 Emoções puras ou mistas

[Manfred \(1977\)](#) descobriu em medições da pressão do dedo que uma expressão de melancolia começa com uma forma que se parece com o amor e termina com uma forma que parece tristeza. Em outras palavras, a mistura emoção da melancolia é descrita como uma justaposição de duas formas no tempo. Por outro lado, a maioria dos teóricos propôs cenários mais simples e determinísticos, por exemplo: sentir-se "cauteloso", segundo essa hipótese, é como sentir uma mistura de interesse e medo.

Os relacionamentos "amor-ódio" são um exemplo onde sentimentos se confundem em um determinado ciclo no tempo. O resultado não é uma simples soma das duas emoções, ou um sentimento que está entre o amor e o ódio, mas uma rápida mudança entre os dois em curtos espaços de tempo. De fato, para certos pares de emoções como amor e ódio, ou talvez tristeza e alegria, pode não ser possível para eles realmente co-ocorrerem ([PFELFER, 1994](#)).

Essa hipótese define que as diferentes emoções não precisam se misturar da mesma maneira. Na verdade, trata-se de uma previsão baseada na maneira como as emoções coincidem com diferentes padrões corporais, e surgem com diferentes mecanismos, na medida em que duas emoções tenham mecanismos generativos não-sobrepostos, e seus mecanismos podem não se misturar, podendo então coexistir no tempo. Mas se eles exigirem o mesmo mecanismo generativo, então somente um deles pode ser gerado de maneira instantânea ([PICARD, 2000](#)). Alternativamente, duas emoções geradas pelo mesmo mecanismo também podem ter diferentes durações de decaimento. Se o segundo é iniciado antes do primeiro e decai mais rapidamente, isso pode produzir um tipo diferente de sobreposição no tempo. Contudo, dado que as emoções são eventos curtos, essa sobreposição não deve ser significativa ([PICARD, 2000](#)).

Com este raciocínio, talvez seja possível predizer que uma emoção primária como o medo, gerada inicialmente na região do cérebro denominada amígdala, poderia coexistir com um estado gerado cognitivamente como nojo, muito embora seja provável que o medo extremo consiga substituir temporariamente quaisquer emoções cognitivas ([PICARD, 2000](#)). Eventos cognitivos podem interferir na pureza das emoções. Por exemplo, se uma pessoa está profundamente envolvida enquanto escuta uma música triste, essa pessoa pode alcançar e expressar um puro estado de tristeza. Entretanto, se a mente vagueia em um evento feliz e se o indivíduo está ansioso enquanto escuta essa música, então a tristeza pode ainda surgir, mas não será tão pura. Este tipo de mistura dilui a expressão de uma emoção ([PICARD, 2000](#)).

Emoções e cognições podem inibir outras emoções. Uma intrigante experiência de mentira e expressão emocional ilustra essa inibição [Picard \(2000\)](#). Trinta e um sujeitos

foram solicitados a expressar raiva ou amor, usando um sentógrafo¹. O dispositivo diferencia significativamente as formas essenciais dessas duas emoções. Em seguida, os sujeitos participaram de vários ensaios em que eles deveriam mentir em vários pontos sobre os cartões que estavam segurando em suas mãos. Quando mentindo ao expressar a raiva, nenhuma mudança significativa foi encontrada tampouco nos relatos desses indivíduos, ou em suas expressões gravadas. No entanto, quando lhes foi pedido para mentir ao expressar amor, não só foram seus autorrelatos de amor significativamente menores, mas as formas de onda para o amor foram significativamente alteradas. Isso sugere que certos eventos cognitivos, como mentir, podem inibir certas expressões emocionais (amor) e não outras (raiva) (PICARD, 2000).

Imagine um ator que se sente irritado na noite de um show, mas tem que expressar o papel de um personagem alegre. Para expressar deliberadamente alegria, ele suprime sua raiva, ou a sobrepõe com alegria (PICARD, 2000). Se ele é bem-sucedido no palco ao comunicar a alegria para a plateia, ele simplesmente "esqueceu" sua raiva, de modo que ela retornará depois que terminar seu tempo no palco? Ou há um efeito terapêutico e a raiva é realmente suprimida? Medições da emoção do ator antes, durante e depois do desempenho poderiam ser estudadas tanto para entender a pureza das emoções quanto para compreender os seus efeitos terapêuticos (PICARD, 2000). As medições poderiam ser combinadas com relatórios do ator e do público, para reunir suas avaliações subjetivas (cognitivas/perceptuais) a fim de buscar observar a sincronização do corpo com essas medições (PICARD, 2000).

Se o ator apenas "esqueceu" sua raiva, então isso sugere que um ato, tem de ocorrer conscientemente e subconscientemente para que a resposta corporal àquela emoção desapareça. Caso contrário o público, ainda que o ator se esforce mas em vez de alegria expresse outra emoção, talvez pense que ele é um mau ator (PICARD, 2000). Quando o ator está irritado e tenso o seu corpo não pode simplesmente, ao pensar em "sorrir" e expressar-se de maneira despreocupada e leve. O corpo também possui uma memória de curto prazo, os músculos armazenam tensão, a postura pode ser alterada. A intensidade da comunicação afetiva não é apenas uma função dos pensamentos, mas uma função da modulação corporal - voz, rosto, postura e muito mais (PICARD, 2000). À medida que o ator adota deliberadamente todas essas expressões de maneira coerente, não só sua comunicação é mais eficaz, mas move-se mais perto de um estado puro de emoção. Quanto mais puro o estado emocional, mais poderosa será a capacidade do ator para se comunicar com a audiência e levá-la para um estado semelhante (PICARD, 2000). Teorias que examinam a pureza das emoções através de seu poder de expressão verificam que quando há coerência corporal durante o registro daquele estado, os sinais tornam-se empiricamente mais testáveis pelo computador, e podem facilitar a modelagem dos

¹ Equipamento para medir a pressão do dedo ao pressionar um botão (MANFRED, 1977).

estados emocionais para síntese e reconhecimento (PICARD, 2000).

Os modelos de representação de emoções tendem a variar entre "nível baixo" e "nível médio", mapeando emoções para sinalizar padrões de geração, expressão ou síntese. Alguns dos modelos assumem a emoção discreta enquanto outras assumem dimensões contínuas. Nenhum deles é "de alto nível" e nenhum considera a semântica da situação que pode gerar uma resposta emocional (conforme necessário na representação cognitiva da emoção) (VERDUYN et al., 2015).

O termo "modelo" refere-se a uma fórmula que seja capaz de explicar completamente um fenômeno, analisando-o e sintetizando-o. No sentido mais rico, um modelo pode reconhecer e sintetizar o padrão. Mas podemos estender esse conceito e considerar que conjuntos de características que discriminam estados emocionais, mas que não podem sintetizá-los podem ser considerados modelos, assim como outros que podem sintetizar certas expressões afetivas, mas não fornecem parâmetros para o reconhecimento. Desse modo, o termo "modelo", neste trabalho é usado para descrever um conjunto de parâmetros e procedimentos úteis para a análise, síntese dos padrões ou ambos, como define Picard (2000).

É importante deixar claro que existe um equívoco comum de que haja um único modelo correto para representação ou reconhecimento de emoções. Pelo contrário, a experiência demonstrou que a melhor escolha do modelo depende da aplicação. Cada modelo tem seus pontos fortes e fracos, e, às vezes, uma combinação de modelos pode conseguir melhores resultados do que um único (PICARD, 2000).

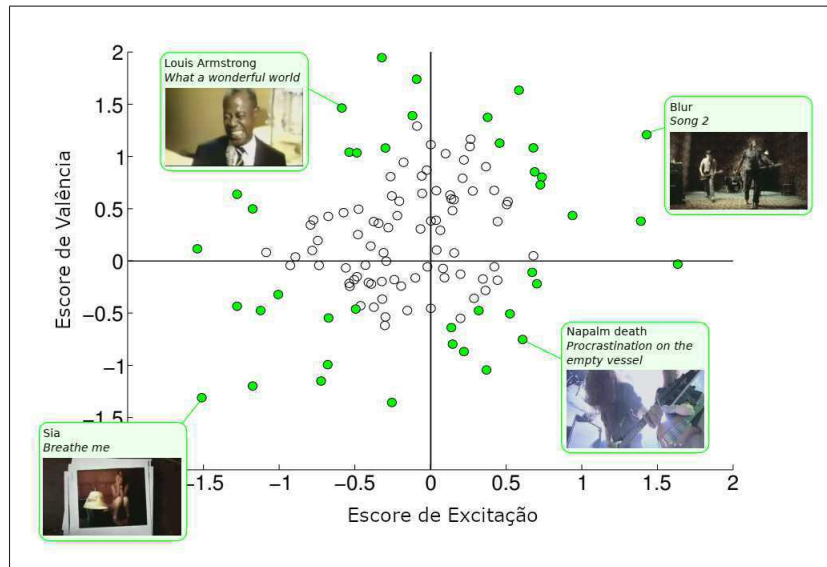
Para essa pesquisa, como consequência da base de dados selecionada para os experimentos, adotamos o modelo dimensional baseado em valência (*valence*) e excitação (*arousal*) e sua representação no espaço amostral apresentadas pela pesquisa de (KOELSTRA et al., 2012). Na próxima seção, iremos detalhar a base de dados utilizada.

2.3.3 DEAP: *A Database for Emotion Analysis using Physiological signals*

A base de sinais de EEG DEAP (*A Database for Emotion Analysis using Physiological signals*) explora a possibilidade de classificar as dimensões emocionais induzidas por música e vídeos para diferentes usuários. A base inova ao obter as respostas a estes estímulos (clipes de música), que anteriormente se resumia a bases focadas principalmente em imagens, músicas ou segmentos de vídeo. Várias categorizações discretas de emoções foram propostas, como as seis emoções básicas introduzidas por Ekman (1992) e a estrutura em árvore das emoções proposta por Parrott (2001). As escalas de emoção também foram propostas, como a *Plutchik's emotion wheel* e a escala de valência-excitação por Russell (KOELSTRA et al., 2012).

Nessa base de dados, foi utilizada a escala de valência-excitação de Russell, am-

Figura 12 – Classificação das emoções da base DEAP



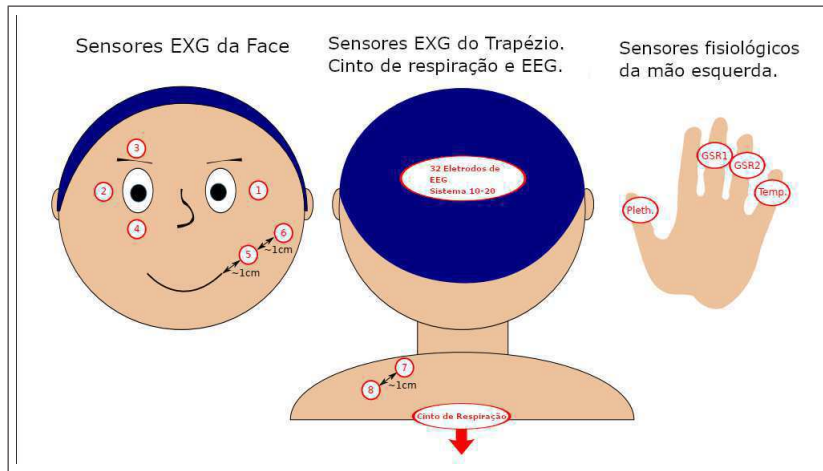
Fonte: Figura traduzida de [Koelstra et al. \(2012, p. 5\)](#).

plamente aceita na pesquisa sobre afeto, para descrever quantitativamente as emoções. Nessa escala, cada estado emocional pode ser colocado em um plano bidimensional com excitação e valência como os eixos horizontal e vertical. Embora a excitação (*arousal*) e a valência (*valence*) expliquem a maior parte da variação nos estados emocionais, uma terceira dimensão de dominância (*dominance*) também pode ser incluída no modelo. A excitação pode variar de inativo (por exemplo, desinteressado, entediado) a ativo (por exemplo, alerta, excitado), a valência varia de desagradável (por exemplo, triste, estressado) para agradável (por exemplo, feliz, eufórico), a dominância varia de um sentimento indefeso e fraco (sem controle) a um sentimento fortalecido (no controle de tudo) (KOELSTRA et al., 2012), como é possível se observar na Figura 12.

Para estimular as emoções nos usuários, um conjunto de 120 clipes de música foi coletado e submetido a um teste subjetivo para selecionar o material mais apropriado. Para cada vídeo, um destaque de um minuto foi selecionado automaticamente. Foram selecionados 32 participantes para participar do experimento e ter seu sinal de EEG e sinais fisiológicos registrados pelos sensores periféricos, como os sensores EXG que podem ser utilizados para captar sinais de EEG ou musculares, conforme pode ser visto na Figura 13.

Após serem gravados assistindo aos 40 vídeos de músicas, os participantes avaliaram cada vídeo em termos de excitação, valência, agradável/desagradável, dominância (quão forte é o sentimento experimentado?) e familiaridade (você já conhecia esse vídeo?). Além disso, para 22 participantes, também foi gravado vídeo frontal visando a captar suas expressões faciais durante o experimento. Um resumo da base pode ser visto na Tabela 2.

Figura 13 – Sinais fisiológicos captados pela base DEAP



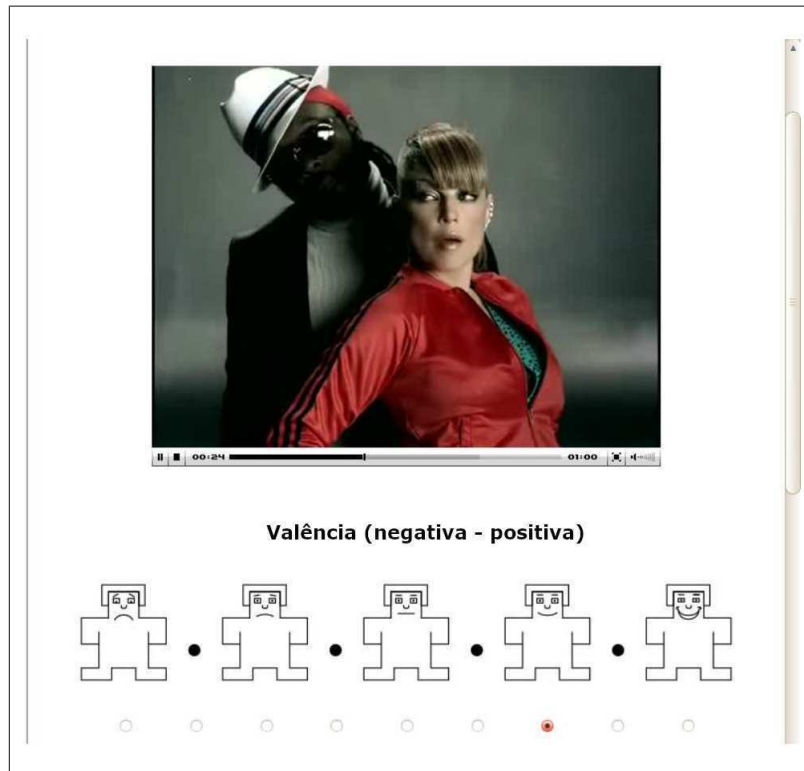
Fonte: Figura traduzida de Koelstra et al. (2012, p. 6).

Tabela 2 – Resumo do experimento para construção da base DEAP

Características	Configurações
Quantidade de participantes	32
Quantidade de vídeos	40
Método de seleção	Subconjunto de vídeos rotulados a partir da experiência dos participantes
Escala de avaliação	Excitação, Valência, Dominância (quão forte é o sentimento experimentado?), Gostar (quanto você gostou do vídeo?), Familiaridade (você já conhecia esse vídeo?)
Valores de classificação	Familiaridade: escala discreta de 1 até 5, outros: escala contínua de 1 até 9
Sinais captados	32-canais no sistema 10-20, EEG e sinais fisiológicos em 512Hz, vídeos das expressões faciais (para 22 participantes)

Os criadores da base dados disponibilizam todo seu conteúdo online, que consiste em todos os dados de sinal gravados, vídeo da face frontal para um subconjunto dos participantes e avaliações subjetivas dos participantes. Também estão incluídas as classificações subjetivas da lista inicial dos 120 vídeos utilizados. Devido a questões de licenciamento, não foram capazes de incluir os vídeos, mas os links dos mesmos também foram disponibilizados. Também está disponível uma análise estatística das avaliações dos participantes e suas correlações entre os sinais EEG e as classificações dos usuários (*ratings*). São apresentados os resultados preliminares com os sinais e EEG, bem como os sinais fisiológicos periféricos (KOELSTRA et al., 2012).

Figura 14 – Interface Web utilizada pelo DEAP para classificar os vídeos

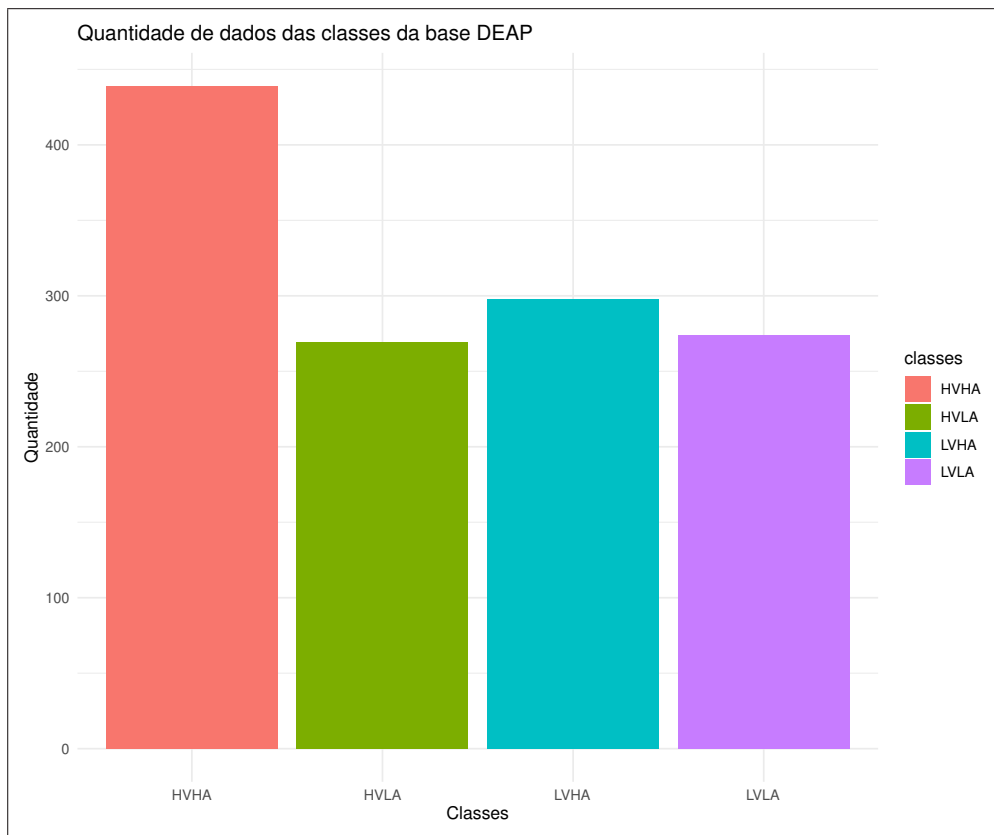


Fonte: Figura traduzida de Koelstra et al. (2012, p. 5).

A partir da coleção inicial de 120 vídeos para estimular os participantes, 40 vídeos de teste foram escolhidos por intermédio de uma interface de avaliação emocional subjetiva baseada na web, como pode ser visto na Figura 14. Participantes assistiram aos vídeos e os avaliaram em uma escala de 9 pontos para valência, excitação e dominância. Uma imagem da interface é mostrada na Figura 14. A ordem dos cliques foi randomizada, mas dando preferência aos cliques classificados pelo menor número de participantes. Esse critério assegurou um número similar de avaliações para cada vídeo (14 a 16 avaliações por vídeo foram coletadas). Foi assegurado também que os participantes nunca viram o mesmo vídeo duas vezes (KOELSTRA et al., 2012).

Depois de todos os 120 vídeos terem sido classificados por pelo menos 14 voluntários cada, os 40 vídeos finais para uso na base de dados foram selecionados. Para maximizar a força das emoções, foram selecionados aqueles vídeos que tiveram as maiores notas dadas pelos voluntários e, ao mesmo tempo, foram incluídos vídeos com classificações menores como uma espécie de ruído na base. Para cada vídeo, foi calculada uma escala de validade e valorização normalizada. Para cada quadrante no espaço de valência-excitação normalizado, foram selecionados os 10 vídeos que estão mais próximos do canto extremo do quadrante. A Figura 12 mostra a pontuação para as avaliações de cada vídeo. Dos 40 vídeos selecionados, 17 foram selecionados através de *tags* afetivas do Last.fm, indicando

Figura 15 – Distribuição de amostras da base DEAP para quatro classes com base em valência e excitação.



Fonte: Autoria própria.

que os estímulos podem ser selecionados através desse método (KOELSTRA et al., 2012).

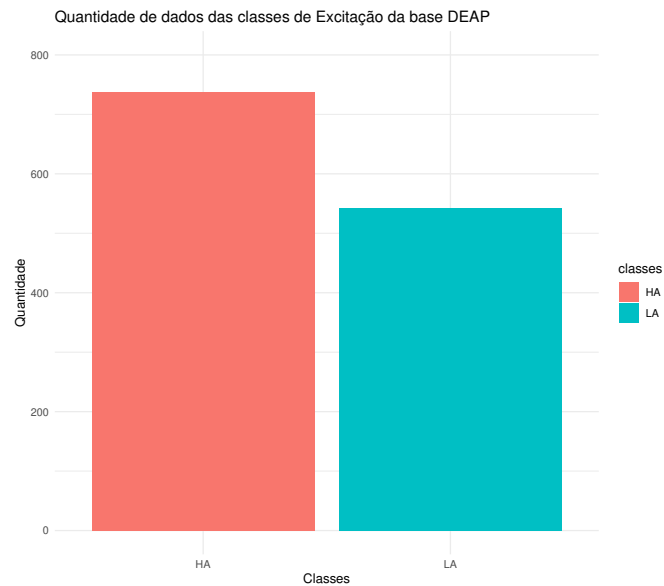
Os estímulos foram selecionados para induzir emoções nos quatro quadrantes do espaço de valência-excitação (*low-Arousal-Low-Valence*, *High-Arousal-High-Valence*, *Low-Arousal-High-Valence*, *High-Arousal-High-Valence*, respectivamente LALV, HALV, LAHV, HAHV). Nessas quatro condições de elicitación de afeto, são asseguradas que as grandes partes do plano de excitação-valência estão cobertas (KOELSTRA et al., 2012).

Existe também a possibilidade de executar a classificação de emoções utilizando apenas a valência ou a excitação, ou seja em duas classes, como nos trabalhos de Wang et al. (2018) e Yang et al. (2018), nesse caso as classes são, alta-excitação (HA) e baixa-excitação (LA), alta-valência (HV) e baixa-valência (LV). A Figura 16 mostra a distribuição das amostras nas classes baseadas em excitação na Figura 16a e em valência Figura 16a.

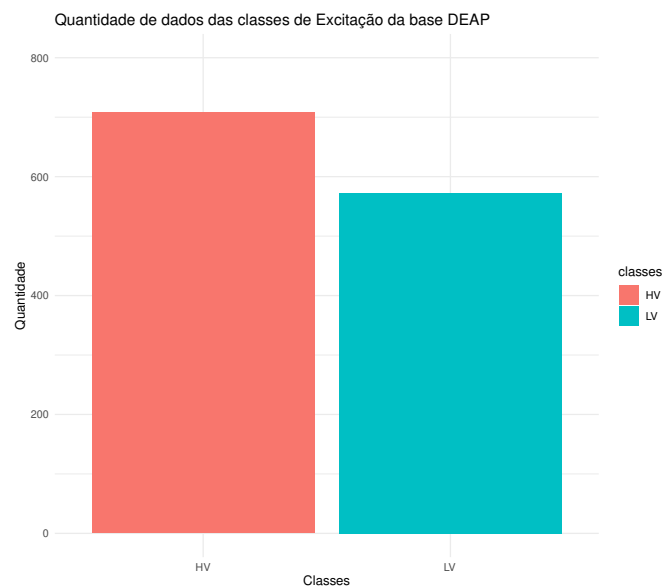
Tanto na Figura 15 quanto na Figura 16 mostram a proporção de amostras para quatro e duas classes, respectivamente, baseadas nos eixos valência-excitação. Dessa forma é visível que existe um desbalanceamento entre as classes da base e que é possível que seja necessário utilizar uma técnica de balanceamento para não prejudicar o

Figura 16 – Distribuição das amostras da base DEAP para duas classes.. (a) Distribuição das amostras da classe com base em excitação. (b) Distribuição das amostras de classe com base em valência.

(a) Distribuição das amostras de classes com base em excitação



(b) [Distribuição das amostras de classes com base em valência



Fonte: Autoria própria

classificador durante o aprendizado (DONG; GONG; ZHU, 2019).

É importante salientar que há dois conjuntos de disponíveis da base DEAP, a base original que foi descrita na Tabela 2 e a base filtrada (utilizada para os experimentos dessa proposta) em que as amostras foram filtradas para um frequência de $128Hz$, os artefatos

EOG (Eletrooculograma) ² foram removidos, foi aplicado um filtro de frequência passa-faixa de 4,0 – 45,0 Hz (a sub-banda gamma que pode ir acima dos 50 Hz é truncada para 45, perdendo informação) (KOELSTRA et al., 2012).

Os dados foram segmentados em ensaios de 60 segundos e uma linha de base pré-teste de 3 segundos, totalizando 63 segundos de sinal, esses 3 segundos iniciais são considerados como período de adaptação do indivíduo no início do experimento (KOELSTRA et al., 2012). Por isso é prudente retirá-los do sinal que será utilizado para a classificação uma vez que podem prejudicar o resultado do reconhecimento já que é um sinal ruidoso (XIN et al., 2016; MEI et al., 2017; YANG et al., 2018). ³

2.3.4 Considerações Finais

Neste capítulo, foi explanado sobre como é executado o registro de sinais de EEG e sua categorização em estados emocionais. Assim como a categorização desses sinais é realizada de diferentes formas, tanto dependente como independente do indivíduo, além do detalhamento da base de dados de sinais de EEG que foi utilizada. Também foi apresentado um estudo sobre os métodos de processamento dos sinais de EEG e o classificador utilizado nesta pesquisa, reconstrução do espaço de fase, redes neurais convolucionais, respectivamente.

No próximo capítulo serão apresentados os trabalhos relacionados ao objeto desta pesquisa de doutorado, esses estudos foram selecionados a partir de uma revisão bibliográfica sistemática sobre o estado da arte do reconhecimento de emoções a partir de técnicas de *deep learning*.

² Frequência das piscadas dos olhos, energia, média e variância do sinal.

³ Essas informações estão disponíveis online <http://www.eecs.qmul.ac.uk/mmv/datasets/deap/readme.html>, acessada em: 31/12/2019.

3 Revisão bibliográfica

Neste capítulo são detalhados os trabalhos relacionados ao objeto de pesquisa. Para tanto, este capítulo apresenta uma síntese da revisão sistemática realizada no início desta pesquisa, abordando o escopo do reconhecimento de emoções a partir de sinais de EEG, extração de características e algoritmos de classificação.

3.1 Pesquisas correlatas

Segundo [Kitchenham e Charters \(2007\)](#), uma revisão da literatura ¹ decorre da necessidade de resumir toda a informação existente sobre uma área de pesquisa de forma completa e imparcial. Portanto, a definição de um protocolo para orientar o desenvolvimento deste tipo de pesquisa é necessária, e permite que essa pesquisa seja auditada e replicada por outros. Na primeira etapa do estudo, foram selecionados 977 trabalhos. Na segunda etapa, após uma seleção mais detalhada baseada em critérios bem definidos de inclusão e exclusão, 61 trabalhos permaneceram como relevantes para esta revisão da literatura. Antes da última etapa, foram também eliminados 3 trabalhos. Finalmente, 58 trabalhos foram selecionados para esta revisão.

Para a seleção de estudos ser mais precisa e eficiente, foram adotados os seguintes critérios de inclusão:

- Artigos publicados em revistas e conferências;
- Documentos diretamente relacionados ao reconhecimento de emoções de dados EEG;
- Artigos publicados entre 2010 e 2020;
- Documentos teóricos ou empíricos.

Com a definição desses critérios, foi possível selecionar os artigos, analisar e sintetizar os mesmos e tirar conclusões sobre a área que está sob investigação. De modo a direcionar essa introdução, a revisão não será totalmente detalhada, porém as próximas seções irão tratar sobre os artigos que serviram como base para essa proposta de tese. Para facilitar a compreensão dessa revisão, separamos a área de reconhecimento de emoções via sinais de EEG em categorias, que visam refletir o estado da arte. É importante observar que as categorias relacionadas refletem inclusive possíveis abordagens para o problema do reconhecimento de emoções. Que vão desde a investigação por características ótimas e

¹ Se encontra no Apêndice A desta tese

discriminatórias, até a exploração de novas abordagens, melhoria de classificadores existentes e etc, que justificam as decisões tomadas para a elaboração desta tese. Os principais tópicos de pesquisa que nortearam esta tese foram:

- Reconhecimento de emoções usando sinais de EEG (CHRISTENSEN; ABDULLAH, 2018);
- Características do espaço de fase para o reconhecimento de emoções em sinais de EEG (ZANGENEH SOROUSH et al., 2018);
- O uso de imagens de reconstrução do espaço de fase para o reconhecimento de emoções via sinais de EEG (ZANGENEH et al., 2019a);
- O uso de *deep learning* para extração de características em sinais de EEG (RIM et al., 2020).

Como o foco desta tese é a utilização de uma técnica de aprendizagem profunda *deep learning* chamada rede convolucional como classificador de emoções a partir de imagens de PSR (*phase space reconstruction*) a partir de sinais de EEG, a próxima seção consiste em um recorte da revisão bibliográfica focada nesse objetivo.

3.1.1 O uso de redes convolucionais para classificação de emoções a partir de características de sinais de EEG

Li et al. (2017) utilizaram redes neurais convolucionais e redes neurais recorrentes do tipo *Long-Short-Term-Memory* (LSTM) para classificar estados emocionais com sinais de EEG da base DEAP em quatro classes distintas a partir da valência e excitação.

As características utilizadas para a classificação foram a EEG MFI (*EEG Multi-dimensional Feature Image*), que podem ser extraídas da seguinte forma: a densidade do espectro de potência (PSD) é estimada a partir do sinal de EEG, utilizando o método de Welch com uma janela de Hamming e diferentes tamanhos de janela de tempo em segundos (1, 2, 3, 4, 5, 6, 8, 10, 12, 15, 20, 30 e 60) sem sobreposição de parâmetros. A quantidade de amostras varia de acordo com a janela de tempo, em que Tl é o tamanho da janela de tempo, então o método de janelamento também funciona como um aumento de dados. Os dados utilizados são dos 32 canais, resultando em uma imagem por vídeo.

A classificação foi executada de forma independente do indivíduo e o conjunto final de dados foi aumentado em 20 vezes a partir da adição do ruído sal e pimenta (*salt & pepper*) de modo que houvesse pelo menos 20 mil amostras de cada conjunto (PSD de 1 segundo até 60). Cada imagem foi submetida a uma rede convolucional 3D, com duas camadas convolucionais e uma camada densa, e em seguida passava por uma

rede recorrente do tipo LSTM que, segundo [Li et al. \(2017\)](#), são ótimas para reconhecer características espaciais de imagens. Dessa forma, foi obtido um resultado de, em média, 75,21% de acurácia, com as quatro classes balanceadas, que em comparação a KNN, SVM, CNN+RNN, obtém uma melhora entre 9 e 13% na acurácia.

Na Pesquisa de [Wen, Xu e Du \(2017\)](#) é utilizada uma entropia calculada através das sub-bandas de um sinal de EEG da base DEAP. Essa característica é extraída a partir de cada um dos canais dos 32 da base. Outro parâmetro utilizado naquela pesquisa foi calcular o coeficiente de correlação de Pearson entre os canais, ordenando-os em pares a partir desse valor. Uma vez reordenados, o vetor de características é montado. Essas características são extraídas a partir do sinal de EEG recortado com uma janela de 1 segundo com o passo de 0,5 segundo. Dessa forma o número de amostras foi aumentado de 40 para 4760 amostras por indivíduo.

De acordo com os autores, os experimentos dependentes do indivíduo, foram realizados com os dados com e sem o aumento via janelamento. Os resultados demonstram que, em média, o conjunto com os dados aumentados teve acurácias mais altas, tanto para valência, como para excitação. As melhores acurácias médias são 77,98% e 72,98%, respectivamente, para valência e excitação.

No trabalho desenvolvido por [Yang, Han e Tang \(2017\)](#) a entropia diferencial é calculada para cada sub-banda de um sinal de EEG. A base utilizada (SEED) possui 62 canais, sendo os sinais rotulados em três classes (Feliz, Neutro e Triste), com três mil amostras cada uma. O vetor de características foi submetido à uma rede neural profunda de três camadas, que aproveita os pesos de um autocodificador. O experimento foi executado para cada uma das sub-bandas (alfa, beta, gama, delta e theta), a cada um dos 14 usuários selecionados.

Os resultados são de 59,6%, 66,27%, 71,97%, 78,48%, 82,56%, 85,5% de acurácia para as bandas delta, theta, alfa, beta e gama, respectivamente. Estes resultados indicam que, de acordo com a faixa de frequência das bandas, as porcentagens de acurácia aumentam, uma vez que gama corresponde a maior faixa de frequência do sinal.

[Tripathi et al. \(2017\)](#) realizaram o reconhecimento de emoções independente do indivíduo com os dados da DEAP utilizando duas técnicas distintas, uma rede neural profunda e uma rede neural convolucional com quatro camadas. Para alimentar as redes os autores extraíram 101 estatísticas dos sinais dos 32 indivíduos da base DEAP, com 40 canais por indivíduo.

Os resultados se mostram promissores, pois os autores comparam a acurácia deles com a de outros trabalhos, assim como o de [Wen, Xu e Du \(2017\)](#), cujos os experimentos realizados foram para duas classes baseadas nos níveis de valência e excitação, apresentando como resultados, respectivamente, 75,78% e 73,125% (Rede Neural Profunda),

81,406% e 73,36% (Rede Convolutacional). Também foram realizados experimentos com 3 classes baseados em valência e excitação, respectivamente, 58,44% e 55,70% (Rede Neural Profunda), 66,79% e 57,58% (Rede Convolutacional).

Vale salientar que [Tripathi et al. \(2017\)](#) não realizaram nenhum tipo de aumento de dados, nem balanceamento, uma vez que utilizar os dados de todos os indivíduos da base DEAP amenize esse problema, porém é possível que esse resultado esteja enviesado, já que a acurácia é sensível a dados desbalanceados ([KONAR; CHAKRABORTY, 2015](#)).

Os autores [Mei et al. \(2017\)](#) propõem utilizar uma matriz de correlação de Pearson gerada via sinais de EEG da base DEAP, e submetê-la a uma rede neural convolutacional. Para tanto o dados da base DEAP foram aumentados a partir do seu janelamento em segmentos de quatro segundos com uma superposição de dois segundos. As matrizes submetidas à rede convolutacional foram geradas através da banda gama e, separadamente, das bandas alfa, beta e theta.

Os experimentos foram executados com duas, três e quatro classes baseadas em valência e excitação. A classificação foi feita dependente do indivíduo e os resultados são as médias entre todos os 32 experimentos

Para todos os casos apresentados pelos autores, a rede convolutacional foi melhor do que outras técnicas, como o SVM. Os resultados foram, de valência e excitação respectivamente, para duas classes na banda gama: 82,7% e 81,2%; para as três bandas: 83,6% e 83%. Para três classes na banda gama: 85,7% e 85,9%, valência e excitação; para três bandas: 85,8% e 86,7%. Para quatro classes na banda gama: 72,2%; 73,1% para as três bandas ([MEI et al., 2017](#)).

Dessa forma, fica evidente que ao se fazer uso de matrizes de correlação de Pearson geradas a partir das três bandas, obteve-se melhores resultados ao se fazer uso de apenas a banda gama. Dessa forma, fica evidenciado que redes neurais convolucionais, são melhores do que classificadores baseados em SVM para esse tipo de atividade ([MEI et al., 2017](#)).

Além dos resultados citados, revisões como as de [Lotte et al. \(2018\)](#) e [Rim et al. \(2020\)](#) apresentam evidências de que técnicas de *deep learning*, como redes neurais convolucionais, são promissoras para classificação de estados emocionais a partir de sinais de EEG.

3.1.2 O uso de características da reconstrução do espaço de fase para classificação de estados emocionais

A pesquisa de [Dawid \(2019\)](#), utilizou imagens de reconstrução do espaço de fase gerados a partir do sinal de EEG para classificar sinais de EEG contaminados com movimentos dos músculos faciais, como aperto e afrouxamento dos músculos dos lábios, esta

classe é *lip*, a terceira classe é de aperto e afrouxamento das mandíbulas se chama *jaw* e classe neutra, categorizada a partir da ausência desses movimentos.

O objetivo desse estudo é a partir do sinal de EEG reconhecer diferentes expressões faciais. Foi criada uma base de dados de sinais de EEG para o estudo, os dados foram submetidos à uma rede convolucional em diferentes níveis de pré-processamento, como imagens de PSR em preto e branco ou coloridas, alcançando os melhores resultados, de $90,54\% \pm 3,98$, com as imagens coloridas.

Apesar de Dawid (2019) utilizar imagens de PSR para classificação de níveis de expressões faciais baseadas em estresse, há certas limitações nesse trabalho. A base de dados só possui registros de sinais do autor. São 400 sinais de EEG no total mas não há variabilidade, então é possível que o classificador tenha aprendido apenas os sinais relacionados a esse indivíduo. Dessa forma, o experimento teria que ser replicado com uma base de dados com diferentes indivíduos para testar sua viabilidade para o reconhecimento independente do indivíduo.

O trabalho de Elamir, Al-Atabany e Aldosouky (2018), utilizou a matriz de densidade do espaço de fase do sinal de EEG como entrada para o reconhecimento de emoções. O trabalho usou as sub-bandas beta e gama dos canais Fp1-Fp2-Cz-Fz-F3-F4, levando em consideração a assimetria do lobo frontal do cérebro. Como classificadores foram testados o SVM, KNN e RT (*regression tree*), a fim de categorizar as emoções em dois grupos de 3 classes: feliz, triste, neutro, sonolento, calmo e animado.

Após gerar as imagens de PSR (*phase space reconstruction*), foram extraídas características da matriz de densidade como contraste, correlação, energia e homogeneidade, contendo o vetor de características, 22 características. Ainda foi aplicado no vetor um teste ANOVA, com o objetivo de filtrar as características mais importantes, assim, a quantidade de características foi reduzida para 18 na banda gama e 14 na beta.

Os melhores resultados são para a banda gama, com a acurácia de 95,8% para valência utilizando RT e 95,8% para excitação utilizando SVM (ELAMIR; AL-ATABANY; ALDOSOUKY, 2018). No entanto, há questões que ficam em aberto, como a escolha da quantidade de dimensões do espaço de fase para a extração das características do sinal, além da não divulgação do desvio padrão dos resultados, dificultando sua comparação com diferentes trabalhos.

Também temos o trabalho de Zangeneh et al. (2019b), em que são extraídas características da trajetória do sinal de EEG no espaço de fase para em seguida transformá-lo no *angle space* (AS), para então estudar a dinâmica do sistema. Nesse trabalho foram extraídas características estatisticamente significativas do AS são submetidos à dois classificadores, perceptron multicamadas (MLP) e *Naïve Bayes* (NB), eles são combinados utilizando a teoria de Dempster-Shafer, e os resultados são avaliados.

Utilizando a base DEAP (KOELSTRA et al., 2012; ZANGENEH et al., 2019b) categorizaram as emoções em quatro classes distintas HAHV, LAHV, LALV e HALV, que consiste nos quadrantes de estados emocionais de Russell e Barret (1999). Foram selecionadas 20 características estatisticamente relevantes que proporcionaram o resultado de 91,37%, também esse estudo demonstrou que os canais Fp1 e Fp2, possuem um papel relevante para sinais emocionais. De maneira semelhante ao trabalho de Elamir, Al-Atabany e Aldosouky (2018), também não é divulgado o desvio padrão o que torna a comparação difícil.

É interessante notar que os trabalhos de Zangeneh et al. (2019b) e Dawid (2019) demonstram que é possível que haja informações emocionais nesse tipo de sinal e apontando a viabilidade desta tese. No entanto, a classificação aparentemente pode ter sido feita de modo independente do indivíduo, mas é provável que o treinamento do classificador tenha sido utilizado a metodologia *leave-one-trial-out* o que torna a atividade de classificação mais simples (PEREIRA; GOMES, 2016).

Tabela 3 – Resumo dos trabalhos correlatos

Referência	Extrator	Classificador	Tipo	Base	Classes	Resultado
Li et al., (2017)	EEG MFI	CLRNN	I	DEAP	4	75,21%
Wen, Xu e Du, (2017)	PCCF	CNN	D	DEAP	4	77,98%(V) e 72,98%(A)
Yang, Han e Tang, (2017)	AE	DLN	D	SEED	3	85,50% ± 3,02
Tripathi et al., (2017)	CNN	CNN	I	DEAP	2	81,40%(V) e 73,36%(A)
Mei et al., (2017)	CNN	CNN	I	DEAP	4	73,10% ± 14,2
Elamir, Al-Atabany e Aldosouky, (2018)	GLCM	RT e SVM	I	DEAP	2	95,80%
Dawid, (2019)	PSR	CNN	N.I.	Proprietária	3	90,54% ± 3,98
Zangeneh et al., (2019b)	AS	MLP e NB	N.I.	DEAP	4	95,8%

¹I = Independente do indivíduo, D = Dependente do indivíduo, N.I. = Não identificado.

Os resultados da Tabela 3 são gerados a partir de bases de dados de EEG diferentes que, por não terem uma configuração experimental padronizada, dificulta a comparação direta de desempenho de diferentes abordagens (RIM et al., 2020). Fatores que também contribuem com esse fato são, a falta de divulgação de métricas estatísticas como o desvio padrão dos resultados e não deixar claro a metodologia de treinamento do classificador (RIM et al., 2020).

No entanto, mesmo com essas dificuldades, a partir da Tabela 3 que as técnicas de *deep learning*, em destaque as redes convolucionais, têm apresentado resultados promissores tanto como extratores de características quanto como classificadores (RIM et al., 2020), assim como características do espaço de fase, corroborando com o trabalho de Zangeneh Soroush et al. (2018). Esse tipo de informação pode ser benéfico para o estudo do processamento e classificação de emoções via sinais de EEG, uma vez que pode revelar novas abordagens possibilitando o avanço no estado da arte nesse escopo.

3.1.3 Considerações Finais

Neste capítulo, foi apresentado uma revisão sistemática com ênfase nos classificadores baseados em técnicas de *deep learning*, demonstrando que são promissoras quando comparadas com outras técnicas de reconhecimento de padrões. Além disto, também foram discutidas pesquisas como as de [Zangeneh et al. \(2019b\)](#) e [Dawid \(2019\)](#) que possuem evidências que o espaço de fase do sinal de EEG pode conter informações emocionais. Tais resultados estão alinhados com a revisão de [Zangeneh Soroush et al. \(2018\)](#).

No próximo capítulo serão apresentadas as técnicas utilizadas e os resultados dos experimentos realizados com a abordagem desta tese, além das principais hipóteses e comparações com o estado da arte sobre o reconhecimento de emoções a partir de sinais de EEG.

4 Metodologia e experimentos

Neste capítulo, a metodologia de pesquisa é apresentada, juntamente com experimentos que demonstraram indícios da viabilidade desta tese, ancorada na utilização de imagens de reconstrução do espaço de fase e redes convolucionais, um algoritmo de *deep learning*, para descobrir se imagens de PSR geradas a partir de sinais de EEG são adequadas para o reconhecimento de emoções de maneira independente do indivíduo, bem como se é possível utilizar uma quantidade mínima de canais para esse tipo de tarefa.

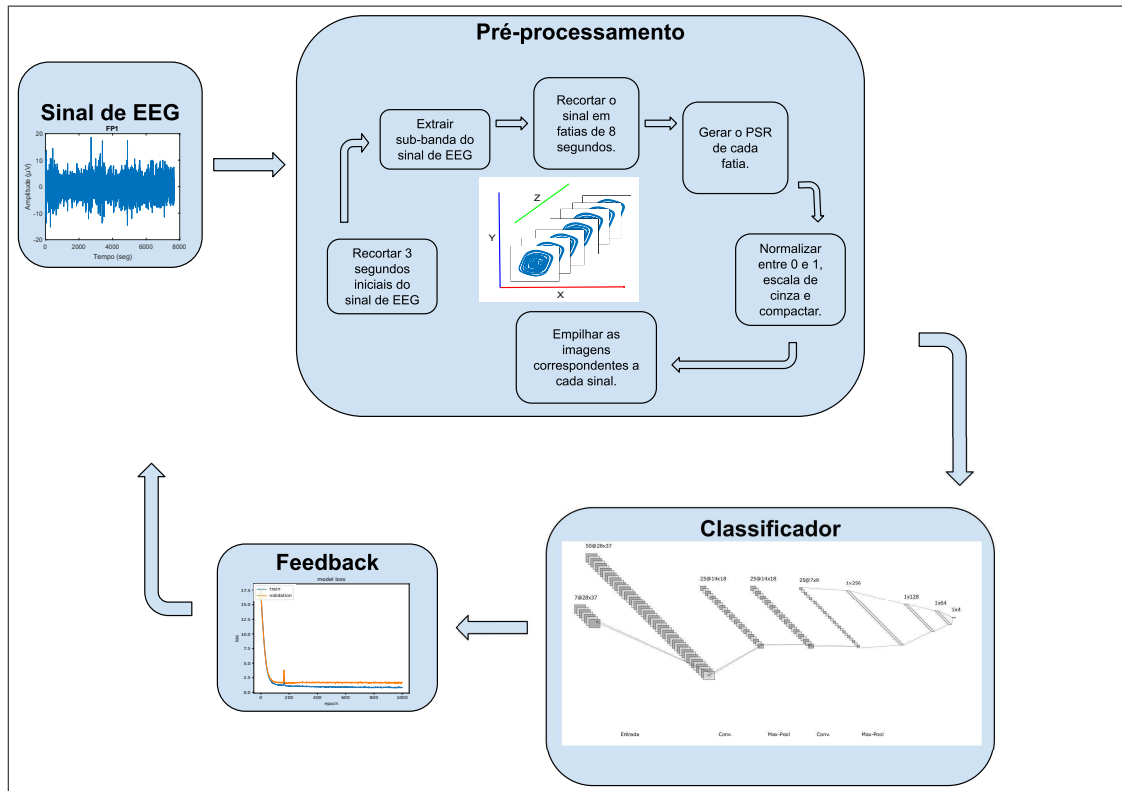
4.1 Materiais e métodos

Para a execução dos experimentos nesta tese, utilizamos a ferramenta de software Matlab, que foi empregada com frequência em pesquisas relacionadas a essa proposta (TORRES; OROZCO; ALVAREZ, 2013; ZHENG; ZHU; LU, 2014; KUMAR; KUMAR, 2016), devido à existência de ferramentas (*toolboxes*) específicas para processamento de sinais. Além disso, utilizamos algoritmos implementados na linguagem de programação Python (BROWNLIE, 2017), a biblioteca *Tensorflow* juntamente com a API Keras (CHOLLET, 2015) e o software estatístico R (COHEN, 2014), para a execução de procedimentos mais específicos, a execução de algoritmos de *deep learning*, entre outros. No que concerne à execução desses algoritmos houve dificuldade com as redes convolucionais, uma vez que demandavam um alto poder de processamento. Houve a necessidade de executar essa tarefa em um computador com maior capacidade (processador Intel i7 com 8GB de RAM e uma GPU NVidia Tesla C2705 com 448 núcleos e 6GB de memória DDR5) para o pré-processamento dos dados, e uma outra máquina (processador Intel i7 com 16GB de RAM e uma GPU NVidia GTX Titan X com 2880 núcleos e 12GB de RAM) para executar o treinamento e validação do classificador de maneira mais rápida, uma configuração de computador inicialmente utilizada consistiu de um processador Intel i5, com 8GB de RAM.

Tal configuração mostrou-se satisfatória para os experimentos clássicos (como reconhecimento de emoções utilizando características como PSD e HOC), mas proibitiva quando tentava-se executar algoritmos de *deep learning*. Uma vez definido o hardware e base de dados para o desenvolvimento da pesquisa, foram planejados experimentos de reconhecimento de emoções com o objetivo de melhorar o estado da arte visto em trabalhos como Torres, Orozco e Alvarez (2013), Verma e Tiwary (2014), Chen e Wang (2016), Li et al. (2017), Wen, Xu e Du (2017), Tarnowski et al. (2018) e Moon, Jang e Lee (2018).

Como já citado no Capítulo 2, optamos por utilizar a base de sinais de EEG DEAP (KOELSTRA et al., 2012), por ser de fácil acesso, e bastante utilizada na área con-

Figura 17 – Etapas da abordagem proposta nesta pesquisa



Fonte: Autoria própria

forme pode ser visto em pesquisas correlatas em [Chen e Wang \(2016\)](#), [Li et al. \(2017\)](#), [Wen, Xu e Du \(2017\)](#), [Mei et al. \(2017\)](#), [Wang et al. \(2018\)](#), [Tarnowski et al. \(2018\)](#) e [Moon, Jang e Lee \(2018\)](#).

A Figura 17 ilustra as etapas da abordagem apresentada nesta pesquisa, como também reflete a ordem na qual serão abordadas as técnicas utilizadas para o reconhecimento de emoções neste capítulo. A primeira etapa, a de pré-processamento, consiste em utilizar o sinal de EEG da base DEAP para gerar a pilha de imagens de PSR.

Para tanto, inicialmente recorta-se os 3 segundos iniciais do sinal de EEG da base DEAP ([KOELSTRA et al., 2012](#)), para retirar o período de adaptação do indivíduo durante o registro dos sinais emocionais, uma vez que esse período pode atrapalhar a etapa de classificação ([XIN et al., 2016](#); [MEI et al., 2017](#); [YANG et al., 2018](#)). Em seguida, os sinais de EEG são recortados em fatias de 8 segundos ([KUPPENS; VERDUYN, 2017](#)) com sobreposição de 2 segundos ([PIHO; TJAHJADI, 2018](#); [WANG; HU; SONG, 2019](#); [YANG et al., 2018](#)).

Posteriormente, são geradas as imagens de PSR de cada recorte do sinal de EEG por indivíduo, para então normalizar suas intensidades para o intervalo entre 0 e 1 entre 1 e 0 ([BROWNLIEE, 2017](#)). As dimensões da imagem são reduzidas utilizando a téc-

nica de contagem de caixas (MARINUS, 2019), transforma-as em imagens na escala de cinza (GONZALEZ; WOODS, 2018). Por último, realiza-se a conversão das imagens para escala de cinza e faz-se o empilhamento das imagens de todos os recortes. Sendo assim, cada sinal da base DEAP (KOELSTRA et al., 2012) passa a ser representado doravante por uma pilha de 7 imagens de PSR.

A segunda etapa é a de classificação, em que as pilhas de imagens de PSR alimentam uma rede convolucional (GONZALEZ; WOODS, 2018) em 3 dimensões (3D), e em seguida avança-se para a última etapa. Na terceira etapa, o *feedback*, os resultados da classificação são analisados para então, se necessário, retornar à uma das etapas anteriores para reajustar os parâmetros de pré-processamento, ou para calibrar o classificador, a fim de melhorar os resultados.

As decisões quanto aos parâmetros escolhidos na execução das etapas da abordagem proposta nesta tese, serão justificados experimentalmente nesta seção, desde a escolha dos canais, o tamanho da janela de recorte dos sinais, a redução do tamanho das imagens, etc. Já os parâmetros da rede convolucional foram definidos empiricamente, ajustando-os de acordo com os resultados dos experimentos.

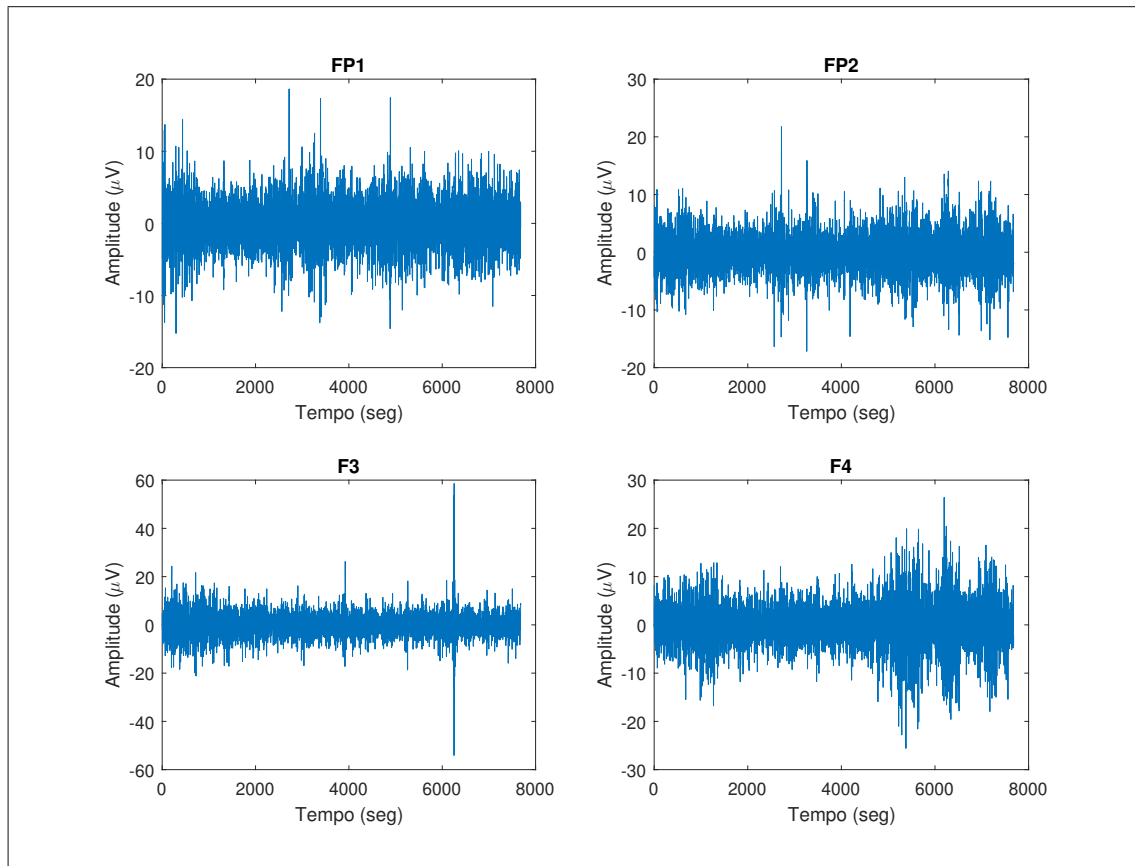
Um dos parâmetros importantes para a abordagem proposta é a seleção dos canais de EEG, uma vez que, de acordo com trabalhos como os de Wang, Hu e Song (2019), Xu et al. (2019) e Wang e Xu (2016) há canais de EEG que possuem informações emocionais latentes.

Na Figura 18, por exemplo, vemos os sinais da base DEAP dos canais FP1, FP2, F3 e F4. São os canais frontais, ou seja os canais que estão localizados no lobo frontal, que segundo a literatura (FONSECA; ALARCAO, 2017; WU et al., 2017; TONG; ZHAO; FU, 2018; TARAN; BAJAJ, 2019), é uma das regiões em que está presente parte do processamento emocional humano como já demonstrava o trabalho de Damasio (1995). Os sinais da Figura 18 são de um mesmo quadrante emocional, ou seja representam uma mesma emoção mapeada em diferentes canais.

Na Figura 19 mostra-se o sinal de EEG do canal frontal FP1 e suas sub-bandas alfa, beta e gama. Estas sub-bandas são frequentemente associadas a sinais emocionais (FONSECA; ALARCAO, 2017; CHRISTENSEN; ABDULLAH, 2018; SANTHIYA; CHITRAKALA, 2019; DZEDZICKIS; KAKLAUSKAS; BUCINSKAS, 2020). Principalmente a banda gama, que é uma das mais utilizadas para o reconhecimento de emoções (FONSECA; ALARCAO, 2017).

Assim como na pesquisa de Xu et al. (2019), optamos por utilizar o sinal de EEG na sub-banda gama, uma vez que é uma faixa de frequência que contém informação emocional (LI; LU, 2009; WANG; NIE; LU, 2014) latente como indica a pesquisa de Tibor et al. (2017). Apesar de na base DEAP, conforme já detalhado no Capítulo 2, ter sido

Figura 18 – Exemplo de sinal de EEG dos canais frontais da base DEAP usados nos experimentos

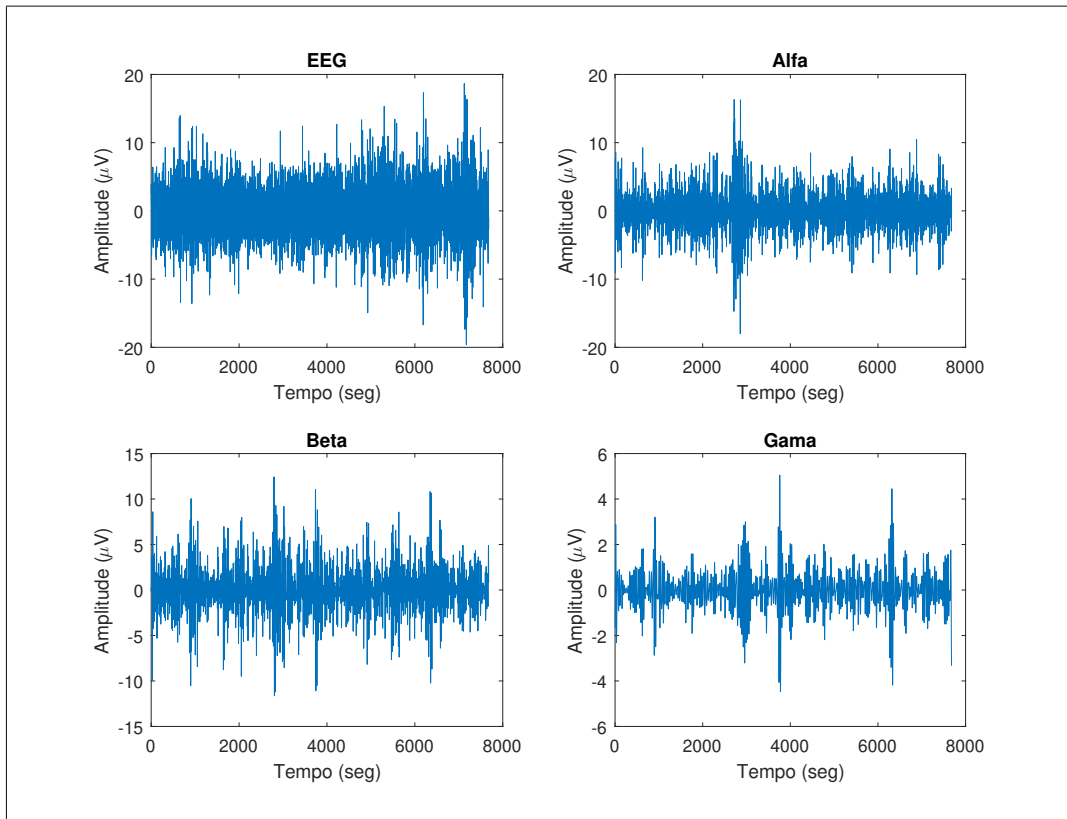


Fonte: Autoria própria

aplicado um filtro de frequência passa-faixa de $4 - 45 \text{ Hz}$, ou seja na banda gama, que pode variar entre 30 e 60 Hz (podendo chegar a 150 Hz) (KONAR; CHAKRABORTY, 2015), sendo truncada e possivelmente perdendo parte da sua informação, pesquisas como aquelas de Tripathi et al. (2017), Yin et al. (2017), Elamir, Al-Atabany e Aldosouky (2018), Rouast, Adam e Chiong (2018), Tarnowski et al. (2018) e Wang et al. (2018), em seus resultados apontam que ainda há informação emocional relevante para a classificação utilizando a banda gama, apesar de ter sido filtrada.

Definimos também nesta pesquisa que o espaço de classificação fosse dividido em quatro classes distintas, sendo elas: *Low-Arousal-Low-Valence*, *High-Arousal-High-Valence*, *Low-Arousal-High-Valence*, *High-Arousal-High-Valence*, respectivamente LALV, HALV, LAHV, HAHV. Tais classes representam os quadrantes, que cobrem as emoções básicas definidas por Ekman (EKMAN, 1992; EKMAN, 1999) e o plano criado por Russell e Barret (1999) em que são distribuídas as emoções baseadas em excitação e valência, conforme a Figura 20. Também foram realizados experimentos utilizando apenas duas classes para valência e excitação, respectivamente baixa valência e alta valência, baixa excitação

Figura 19 – Sinal de EEG do canal FP1 e suas sub-bandas alfa, beta e gama.



Fonte: Autoria própria

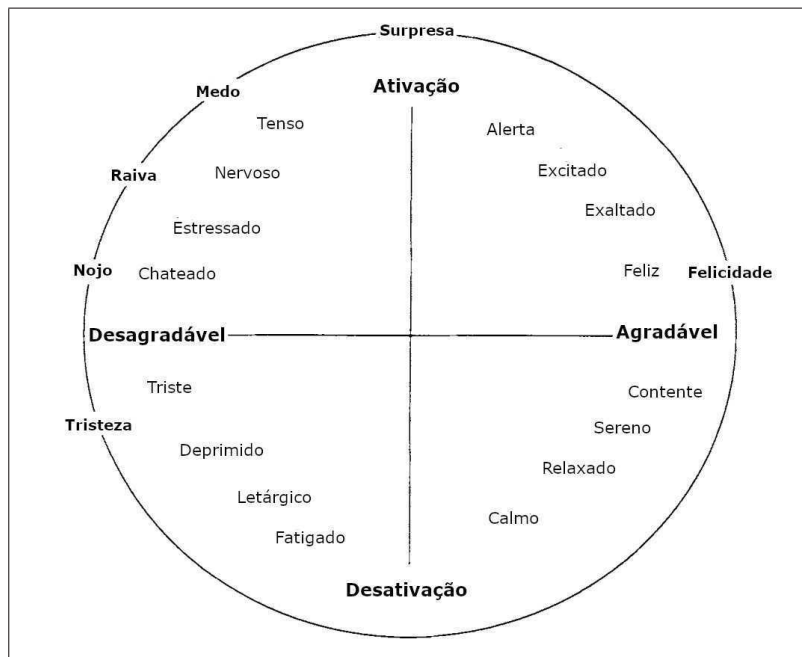
e alta excitação, com o objetivo de verificar a viabilidade da técnica para posteriormente aplicá-la em complexidades maiores como a classificação de emoções em quatro classes.

Para executar as etapas de pré-processamento e aumento de dados da Figura 17 recorreremos a experimentos e analisamos os resultados de maneira empírica. Para tanto, utilizamos métricas descritas na literatura (ZHENG, 2015; CASH; STANKOVIĆ, 2016; BERGER; MAURER, 2018) e em pesquisas que serviram de referência para esta tese como os de Tripathi et al. (2017), Tarnowski et al. (2018), Moon, Jang e Lee (2018), Teo e Chia (2018), Wang et al. (2018) e Yang et al. (2018). As métricas utilizadas estão descritas na próxima seção.

4.1.1 Métricas de avaliação de desempenho da classificação

Segundo Zheng (2015) a natureza dos dados pode afetar as métricas. Em particular, dados enviesados, ou seja dados em que um dos seus "tipos" é mais raro do que outros e classes desequilibradas podem ser um problema para a acurácia. Este problema existe para todas as métricas que dão peso igual a cada tipo de dados. Por exemplo, um conjunto de dados em que a classe positiva seja uma pequena parte dos dados observados, digamos

Figura 20 – Emoções baseadas em valência e excitação definidas por Russel e Feldman.



Fonte: [Russell e Barret \(1999, p. 808\)](#)

1%, isso significa que um classificador que sempre classifique os dados recebidos como negativos alcançaria 99% de acurácia.

Desse modo, qualquer métrica que dê peso igual a cada instância de uma classe tem dificuldade em lidar com classes desequilibradas, porque, por definição, a métrica será dominada pelas classes com mais dados. Além disso, tais métricas são problemáticas não apenas para a fase de avaliação, mas também ao treinar o modelo. O efeito de grandes valores discrepantes durante a avaliação pode ser mitigado por meio de métricas robustas. Dentre as métricas de avaliação existentes, selecionamos as seguintes: acurácia (Acc), precisão (Pre), F1, sensibilidade (Sen) e especificidade (Esp). Utilizar mais de uma métrica foi necessário para verificar o aprendizado do classificador e mitigar o problema do desequilíbrio dos dados no momento em que a atividade de classificação era executada com quatro classes.

De acordo com [Zheng \(2015\)](#) a acurácia, Equação 4.1, leva em conta a frequência com que a previsão é feita pelo classificador. É a razão entre a quantidade de estimativas corretas (o número de amostras¹ no conjunto de testes) e a quantidade total de dados.

$$Acc = \frac{VP + VN}{Total} \quad (4.1)$$

¹ Nesta tese, o termo "amostra" se refere a cada um dos exemplos de dados de entrada utilizado para treinamento e teste do classificador de emoções.

Em que, VP significa verdadeiros positivos, VN significa verdadeiros negativos, FP significa falsos positivos e FN significa falsos negativos.

Precisão (*precision*) e revocação (*recall*) são duas métricas que são frequentemente usadas juntas (ZHENG, 2015). Precisão (Prec) corresponde ao valor preditivo positivo (VPP), Equação 4.2, enquanto a revocação (Rec, também conhecida como sensibilidade), neste contexto, é também conhecida como a taxa de verdadeiros positivos, Equação 4.3. No entanto, utilizamos essas duas métricas para calcular uma métrica mais robusta com relação a acurácia (ZHENG, 2015). Combinando a precisão e a revocação em uma média harmônica de precisão e revocação, chega-se a tradicional *F-measure*, Equação 4.4, ou *F-score* balanceada:

$$Prec = \frac{VP}{VP + FP} \quad (4.2)$$

$$Rec = \frac{VP}{VP + FN} \quad (4.3)$$

$$F\text{-measure} = 2 \cdot \frac{Prec \cdot Rec}{Prec + Rec} \quad (4.4)$$

Nessa perspectiva, uma vez que o problema de classificação para quatro classes nesta tese continha grupos discrepantes quanto à quantidade de exemplares, a métrica *F-measure* se mostrou mais adequada para calibração do classificador, enquanto para o problema de duas classes a acurácia foi suficiente.

Após definir as métricas de avaliação, foi necessário estabelecer uma etapa importante do pré-processamento de dados desequilibrados, a metodologia de balanceamento (DONG; GONG; ZHU, 2019). A próxima seção é dedicada ao balanceamento e aumento de dados utilizados nesta tese.

4.1.2 Balanceamento e aumento de dados

O desbalanceamento entre as classes foi outro fator considerado antes da realização dos experimentos, uma vez que, de um modo geral, um conjunto de dados desbalanceado pode afetar o desempenho do aprendizado (DONG; GONG; ZHU, 2019). A base DEAP (KOELSTRA et al., 2012) possui um desbalanceamento tanto para quatro classes, quanto para duas classes, respectivamente, Figura 15 e Figura 16.

Para quatro classes o problema de desbalanceamento é mais severo, pois a classe majoritária possui praticamente o dobro de amostras das classes minoritárias, de modo que, o classificador pode não aprender a classificar corretamente uma das classes minoritárias e manter a acurácia acima de 70%. Para duas classes, tanto para valência, quanto para excitação, o problema do desbalanceamento não se torna uma preocupação visto que

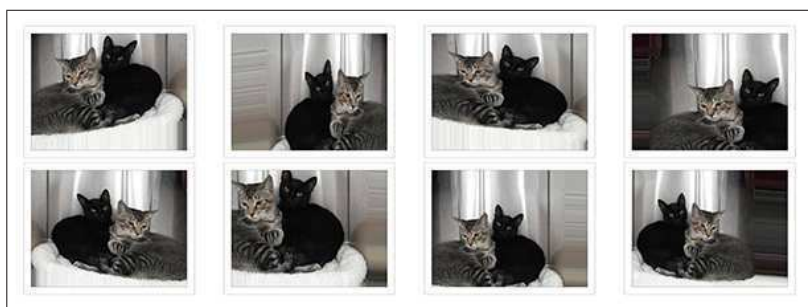
as proporções de amostras de cada classe são próximas (DONG; GONG; ZHU, 2019). No trabalho de Dong, Gong e Zhu (2019) são discutidas as possíveis soluções para o problema do balanceamento.

Dentre as soluções apresentadas, foram verificados aumento de dados baseado no algoritmo ADASYN (*adaptive synthetic sampling approach for learning from imbalanced data sets*) (HE et al., 2008), assim como o balanceamento a partir do aumento de dados (GUENNEC; MALINOWSKI; TAVENARD, 2016), com a técnica da janela deslizante *window sliding* (TARNOWSKI et al., 2018; MOON; JANG; LEE, 2018; WANG et al., 2018) e a partir da atribuição de pesos à classes com menor quantidade de amostras (BROWNLIE, 2017).

Foi necessário também, resolver o problema da pequena quantidade de dados da base DEAP, visto que, conforme a documentação da base (KOELSTRA et al., 2012), cada usuário é representado por um conjunto de dados de 40 sinais de EEG com comprimento de 7680, ou seja, com as dimensões de 40×7680 . E de acordo com (SALAMON; BELLO, 2017), algoritmos de *deep learning* possuem uma elevada capacidade para construir modelos de dados, porém são particularmente dependentes da disponibilidade de grandes quantidades de dados de treinamento.

Uma solução para esse problema é o aumento de dados, que consiste em adicionar deformações, ou ruído, na coleção de dados de treinamento originais para gerar novas amostras. A chave desse processo é não perder a a informação semântica dos dados. Por exemplo, na visão computacional é possível criar novas amostras de imagens a partir de operações como rotação, translação, espelhamento, dentre outras. Porém, essas técnicas devem ser aplicadas de modo que a informação de uma imagem permaneça coerente e reconhecível (SALAMON; BELLO, 2017; HE et al., 2008), método de aumento que já está implementado na API Keras (CHOLLET, 2015) como pode ser visto no resultado da sua aplicação na Figura 21.

Figura 21 – Resultado do algoritmo de aumento de dados do Keras quando aplicado em uma imagem



Fonte: Página do Keras (CHOLLET, 2015)

Tabela 4 – Dados da base DEAP com o aumento de dados (Amostras X Comprimento do sinal)

Dados	Dados da DEAP sem aumento	Dados da DEAP com aumento
Sinais	1280 × 7680	9575 × 7680
Conjunto de treino	896 × 7680	7732 × 7680
Conjunto de validação	344 × 7680	1843 × 7680
Conjunto de teste	40 × 7680	40 × 7680

Existem diversos métodos de *data augmentation* (HE et al., 2008). Na pesquisa de (GUENNEC; MALINOWSKI; TAVENARD, 2016), uma das abordagens utilizadas foi a técnica da janela deslizante *window sliding*. Esse tipo de aumento também foi utilizado nos trabalhos de (TARNOWSKI et al., 2018; MOON; JANG; LEE, 2018) e Wang et al. (2018), com janelas de tamanhos diferentes, cinco, três e quatro segundos respectivamente. Consiste em extrair as características do sinal e executar a classificação nesse nível. O tamanho da fatia é um parâmetro desse método e a quantidade de dados novos gerados depende da relação entre o tamanho, o passo da janela e se há ou não superposição.

Um dos problemas desse tipo de aumento é que cada fatia extraída de uma série temporal pertencente a uma classe "y" é rotulada na mesma classe. De acordo Kuppens e Verduyn (2017), em um minuto de duração é possível que o indivíduo experimente diferentes emoções com diferentes durações. Desse modo é possível que cada recorte do sinal tenha informação de mais de uma emoção.

Devido aos resultados dos trabalhos de Tarnowski et al. (2018), Moon, Jang e Lee (2018) e Wang et al. (2018), há evidências de que uma emoção pode se destacar entre as outras emoções num mesmo sinal Kuppens e Verduyn (2017). Sendo assim, partir desse resultados, há suporte para o uso do aumento de dados por janelamento no contexto de sinais de EEG.

A Tabela 4 mostra a quantidade de amostras na base DEAP antes e após o aumento de dados utilizando a API Keras (CHOLLET, 2015). O aumento foi de aproximadamente 7 vezes a quantidade original de dados. Chegou-se nessa quantidade amostras de maneira empírica, como será detalhado na seção de experimentos.

Na Tabela 5, vemos a quantidade de amostras por classe e como ficou após o aumento de dados. É importante deixar claro que embora exista desbalanceamento quando o problema é reduzido para duas classes, Figura 16, de excitação e valência, não tão acentuado quanto para quatro classes, Tabela 5.

No entanto, o problema do balanceamento permanece, ver Tabela 5, uma vez que as classes majoritárias continuam com uma maior quantidade de dados e as minoritárias não terem seus dados aumentados o suficiente.

Tabela 5 – Quantidade de amostras por classe da base DEAP com o aumento de dados

Classe	Sem aumento	Com aumento
HVHA	439	3286
HVLA	269	2011
LVHA	298	2229
LVLA	274	2049
Total	1280	9575

Uma das estratégias experimentadas no presente trabalho que proporcionou resultados considerados satisfatórios, consistiu em atribuir pesos diferentes aos dados das classes majoritárias e minoritárias de modo que quanto menor a quantidade de dados maior o peso atribuído àquele tipo de dado (KING; ZENG, 2001). A API Keras (CHOLLET, 2015), utilizada para implementar os algoritmos de *deep learning*, disponibiliza o balanceamento de dados via atribuição de pesos sugerido no trabalho de (KING; ZENG, 2001).

O procedimento consiste em ponderar os dados para compensar as diferenças nas frações da amostra (\bar{y}) e da população (τ). O estimador de probabilidade máxima de amostragem exógena ponderada (*weighted exogenous sampling maximum-likelihood estimator*, WESML), maximiza-se os pesos da função logarítmica King e Zeng (2001), como pode ser visto na Equação 4.5:

$$\ln L_w(\beta|y) = - \sum_{i=1}^n w_i \ln(1 + e^{(1-2Y_i)x_i\beta}) \quad (4.5)$$

Em que os pesos são $w_1 = \frac{\tau}{\bar{y}}$, $w_0 = \frac{(1-\tau)}{(1-\bar{y})}$ e $w_i = w_1 Y_i + w_0 (1 - Y_i)$. β é o valor desconhecido que maximizará o valor desta função. Y_i , ($i = 1, \dots, n$) é uma amostra aleatória da população. x_i consiste em um vetor de variáveis explanatórias.

A partir da Equação 4.5, maximizam-se os pesos à medida que as classes minoritárias sejam pequenas frações das classes majoritárias ou da população total. Esse tipo de estratégia permite uma espécie de balanceamento via estimadores de peso, de modo que o classificador passa a ter como foco aprender as características mais discriminatórias das classes minoritárias de maneira mais eficiente.

Na seção posterior a Tabela 7 mostra os experimentos realizados com diferentes tipos de aumento de dados e qual se mostrou mais adequado para a abordagem proposta nesta tese. Após explicar qual técnica de aumento de dados foi selecionada para o problema apresentado nesta tese, serão apresentados os canais selecionados bem como as demais técnicas utilizadas no pré-processamento do sinal de EEG.

4.1.3 Canais de EEG selecionados

Os canais do sistema 10-20 também foram um tópico específico para projetar os experimentos (KONAR; CHAKRABORTY, 2015). Com base em pesquisas como as de Konar e Chakraborty (2015), Zheng e Lu (2015) e Tong, Zhao e Fu (2018), os canais FP1, FP2, F3, F4 estão entre os canais com informações emocionais mais perceptíveis e, portanto, foram selecionados para os experimentos. Baseados também em trabalhos como os de Wu et al. (2017), Taran e Bajaj (2019) e Ogino et al. (2019), optou-se por selecionar a menor quantidade de canais de modo que o reconhecimento de emoções não fosse prejudicado.

Como os sinais de EEG captados a partir de dispositivos não invasivos, também conhecidos como *scalp* EEG, estão sujeitos a diversos tipos de ruídos provenientes da musculatura do usuário, certos canais, como os da região occipital, podem captar sinais elétricos dos músculos do pescoço mesmo que o usuário permaneça imóvel, mesmo a musculatura tensa pode enviar sinais elétricos que não estejam relacionados com emoções (FONSECA; ALARCAO, 2017). Esse sinais indesejados atuam como ruído no sinal de EEG, prejudicando o reconhecimento das emoções, bem como dificultando a extração de características discriminatórias relacionadas a estados emocionais.

Não apenas os canais localizados na região occipital, canais O1 e O2, sofrem com o ruído oriundo da musculatura, os canais frontais também podem ser afetados por esse tipo de ruído, sendo um dos mais comuns é o do movimentos das pálpebras. Porém, de acordo com Koelstra et al. (2012) os criadores da base DEAP removeram esse tipo de ruído dos seus dados, uma vez que os sinais da musculatura da face foram igualmente registrados durante o desenvolvimento da base, como mostra a Figura 13 para assim facilitar a atividade de filtrar esse tipo de ruído.

É bem verdade que mesmo com esse tipo de filtragem de ruído, é possível que ainda haja sinais da musculatura que envolvem a caixa craniana misturado ao sinal de EEG, por isso é uma atividade complexa remover esse tipo de interferência por completo. Portanto, estudos sobre seleção de canais como os de Wang, Hu e Song (2019), Tong, Zhao e Fu (2018) e Zheng e Lu (2015) são importantes para área, pois ajudam a tornar mais simples a escolha por diferentes configurações de sistemas de reconhecimentos de emoções, possibilitando uma extração de característica bem como um reconhecimento de emoções mais preciso.

Outro fator que pode influenciar os resultados de uma atividade de reconhecimento de emoções utilizando a base DEAP é a escala de tempo em que o sinal foi gravado, (OSTWALD et al., 2012; SCHACHT et al., 2012; STEINBERG et al., 2013; SCHULREICH et al., 2013), pois os eventos emocionais ocorrem rapidamente. O cérebro geralmente responde ao estímulo entre 30 – 60 *ms* pós-estímulo auditório e 45 – 60 *ms*

pós estímulo visual, chegando ao seu pico entre 100 e 130 *ms* e continua reverberando até entre 300 e 400 *ms* (STEINBERG et al., 2013). Ou seja, todo o evento emocional acontece ao longo de 1 segundo, mais especificamente até 500 *ms*, obviamente é provável que existam variações que dependem do indivíduo e do estímulo (STEINBERG et al., 2013).

No entanto, Gunes et al. (2011) critica as bases de dados de sinais de EEG afirmando em sua revisão que sinais menores que 8 segundos são curtos para os eventos emocionais se manifestarem completamente. O tamanho da janela a ser usada para alcançar um reconhecimento ótimo de emoções é uma das questões da literatura existente que não tem uma resposta única. Usar o autorrelato como a verdade absoluta é arriscado (GUNES et al., 2011), já que para obter uma precisão e uma previsão confiável dessa característica requer monitoramento de longo prazo dos indivíduos.

Pesquisas selecionadas pela revisão bibliográfica realizada nesta tese, como Gunes et al. (2011), relataram baixa classificação baseada em excitação e valência, além de diferenças entre o autorrelato e características derivadas de dados infravermelhos e térmicos. Porém os autores também deixam claro que, mesmo considerando 8 segundos um sinal curto para uma ativação completa de sinais como os de EEG, este poderia ser o suficiente para identificar as assinaturas de três estados emocionais: estresse, medo e prazer.

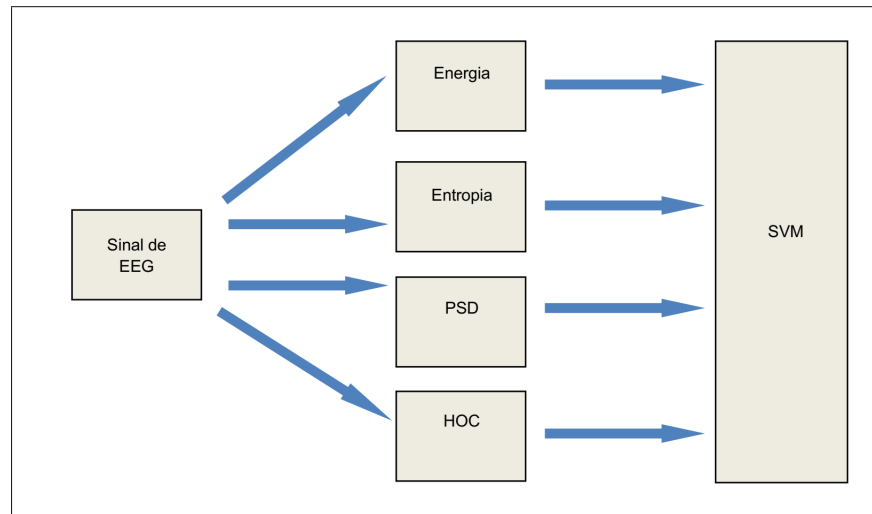
Então, é possível perceber que o reconhecimento de emoções é uma atividade complexa que depende de diversos fatores, além da diferença entre indivíduos (PICARD, 2000). Sendo assim, possuir características eficientes que consigam resultados ótimos, mesmo com essas dificuldades inerentes à área, é um desafio que pode avançar o estado da arte resultando em um trabalho relevante. Após considerar esses fatores, optamos por utilizar o sinal de EEG conforme o disponibilizado pela base DEAP (KOELSTRA et al., 2012) de acordo com a escala em que os sinais foram captados, analisando toda a extensão do sinal.

Na seção a seguir, detalhamos os experimentos que proporcionaram verificar quais variáveis são importantes para o desempenho do classificador (GONZALEZ; WOODS, 2018; WANG et al., 2018; MOON; JANG; LEE, 2018; WANG et al., 2019; CAO et al., 2019; WANG; TONG; HENG, 2019), a definição da configuração do classificador e, posteriormente, relatamos a análise dos resultados, comparações com trabalhos da área e as evidências documentadas nesta tese.

4.2 Experimentos

Conforme discutido no Capítulo 1, A principal hipótese desta tese consiste em verificar se imagens do espaço de fase (PSR) geradas a partir de sinais de EEG quando aplicadas em redes neurais convolucionais profundas, fornecem resultados elevados de acurácias, para o reconhecimento de emoções. Um outra hipótese formulada é sobre a

Figura 22 – Diagrama dos experimentos iniciais com características clássicas extraídas do sinal de EEG



Fonte: Autoria própria

possibilidade de executar esta tarefa com apenas um canal entre os quatro selecionados, Fp1, Fp2, F3 e F4. Sendo assim, os experimentos executados nesta seção foram desenvolvidos com o objetivo de comprovar essas hipóteses. Para verificar o desempenho do classificador, os resultados alcançados nesta tese foram comparados aos de outras pesquisas, para tal, a métrica utilizada foi a acurácia, uma vez que, as pesquisas selecionadas apresentam essa métrica como principal forma de avaliação de desempenho.

Para tanto, inicialmente, focamos em reproduzir experimentos do reconhecimento de emoções com características do sinal de EEG que obtiveram resultados promissores na área (DZEDZICKIS; KAKLAUSKAS; BUCINSKAS, 2020), como PSD e HOC (CHRISTENSEN; ABDULLAH, 2018), energia e entropia (YU et al., 2019), da seguinte forma: foram extraídas as características de cada um dos canais Fp1, Fp2, F3, F4, para em seguida, submeter ao SVM (Máquina de Vetores de Suporte), o classificador mais utilizado na área (CHRISTENSEN; ABDULLAH, 2018).

A Figura 22, contém um diagrama ilustrando como foram executados os experimentos com as características do sinal de EEG citadas, sendo cada experimento executado para os 32 usuários da base DEAP (KOELSTRA et al., 2012), de maneira dependente e independente do indivíduo. Esses experimentos foram importantes para instrumentalização quanto à base de dados, API de técnicas de reconhecimento de padrões Keras (CHOLLET, 2015), que foram utilizadas para o desenvolvimento desta tese, conforme a Tabela 6.

Como base para os parâmetros do SVM, inspiramo-nos em trabalhos como o de Candra et al. (2015), no qual foram utilizados o kernel RBF (*Radial Basis Function*), o algoritmo de treinamento por validação cruzada (*K-Fold Cross-Validation*, em que $K =$

Tabela 6 – Configurações dos experimentos iniciais

Características	Configurações
Canais	Fp1, Fp2, F3 e F4
Extratores de características	HOC, PSD, Energia e Entropia
Classificador	SVM
Tipo de Reconhecimento	Dependente e Independente do indivíduo

10), entre outras técnicas de otimização de parâmetros.

Após a etapa de instrumentalização quanto a tecnologias e base de dados, buscou-se alcançar e, posteriormente, avançar no estado da arte. Para tanto, foi necessário executar por volta de 250 experimentos com a finalidade de encontrar os parâmetros mais adequados, assim como diferentes configurações dos dados entrada, e o classificador com os melhores resultados.

Estudos como os de [Du et al. \(2018\)](#) demonstram que classificadores como redes convolucionais possuem uma relação com a quantidade de dados utilizadas para o aprendizado. De modo que, a medida que a quantidade de dados aumenta, o aprendizado da rede melhora.

Devido às dimensões das imagens de PSR geradas (579×440 pixels), que fazem com que a rede tenha que aprender características entre os 254.760 pixels de apenas 1280 imagens, considerou-se utilizar técnicas de aumento de dados para melhorar o desempenho do classificador.

A Tabela 7 mostra os experimentos realizados com diferentes técnicas de aumento de dados, de modo que é perceptível que a técnica com melhores resultados tanto para o treinamento quanto para o teste é a implementada pela Keras API ([CHOLLET, 2015](#)). É importante deixar claro que todos os experimentos da Tabela 7 foram executados com classificadores com as mesmas configurações.

Tabela 7 – Acurácia Média de diferentes tipos de balanceamento.

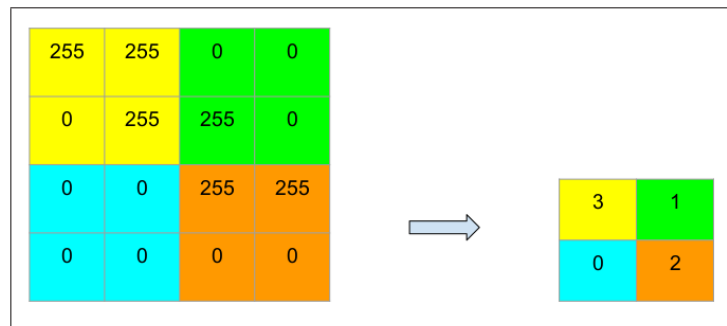
Tipo de Aumento	Acurácia Média Treino	Acurácia Média Teste
Sem aumento	$0,55 \pm 0,20$	$0,55 \pm 0,30$
Adasyn	$0,99 \pm 0,01$	$0,41 \pm 0,13$
Janela Deslizante	$0,62 \pm 0,01$	$0,34 \pm 0,20$
Keras API	$0,89 \pm 0,02$	$0,60 \pm 0,15$

Apesar da utilização do aumento de dados, evitou-se também a praga da dimensionalidade ([ALBON, 2018](#); [BROWNLIE, 2017](#); [ROUAST](#); [ADAM](#); [CHIONG, 2018](#)). Uma vez que, à medida que a complexidade aumenta, as características que o classificador pre-

cisa aprender e a quantidade de amostras de entrada também aumentam (ALBON, 2018; BROWNLEE, 2017; ROUAST; ADAM; CHIONG, 2018).

Portanto, para evitar a praga da dimensionalidade, implementamos a técnica apresentada por Marinus, Araújo e Gomes (2019), de contagem de caixas. Que consiste em dividir a imagem em caixas de tamanho $N \times N$ pixels sem sobreposição, contar quantos pixels com valores mais altos existem na caixa, e numa nova imagem atribuir a um pixel o valor correspondente à quantidade de pixels altos da imagem original, como mostra a Figura 23.

Figura 23 – Técnica de contagem de caixas implementada para reduzir o tamanho da imagem de PSR



Fonte: Autoria própria

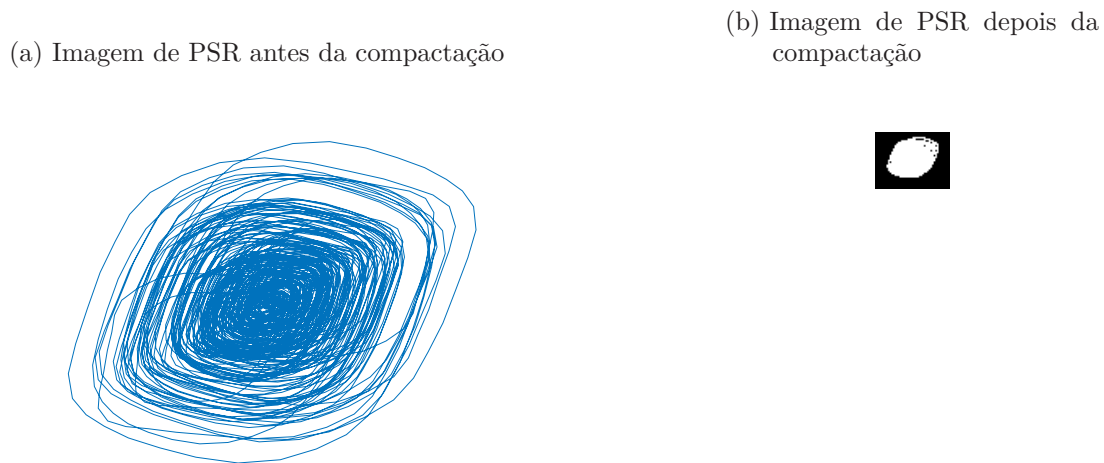
A saída do algoritmo de contagem de caixas consiste em uma imagem em escala de cinza em que as intensidades de pixel representam os pixels com valores mais altos em cada uma das caixas (MARINUS; ARAÚJO; GOMES, 2019). Na Figura 34, é possível verificar o resultado da compactação da imagem.

Para alcançar esse tipo de compactação foi escolhida uma caixa de tamanho 15, reduzindo assim as dimensões da Figura 24a de 579×440 para 28×37 pixels, conforme se pode observar na Figura 24b. Uma redução de aproximadamente 40% na área da imagem, diminuindo a possibilidade da praga da dimensionalidade.

Na Figura 25, é possível verificar a imagem de PSR após compactação da imagem utilizando a técnica da contagem de caixas (MARINUS; ARAÚJO; GOMES, 2019). Antes de ser convertida para tons de cinza 25a, e depois de convertida para a escala de cinza 25b.

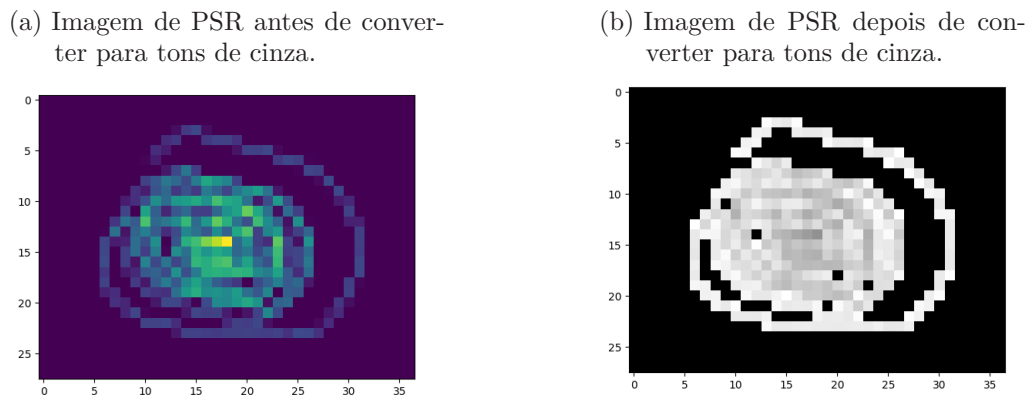
A partir dos experimentos com as imagens de PSR compactadas, foram mapeados 44 parâmetros que poderiam influenciar nos resultados do classificador, e foi necessário investigar de maneira empírica (EASTERLING, 2015; CASH; STANKOVIĆ, 2016; BERGER; MAURER, 2018) quais seriam os parâmetros mais importantes para a classificação de emoções. A Tabela 8 mostra os parâmetros e os tipos de redes convolucionais testadas,

Figura 24 – Resultado da compactação da imagem de PSR via técnica de contagem de caixas. (a) Imagem de PSR antes da compactação. (b) Imagem de PSR depois da compactação.



Fonte: Autoria própria

Figura 25 – Imagem de PSR antes e após converter para tons de cinza. (a) Imagem de PSR antes de converter para tons de cinza. (b) Imagem de PSR depois de converter para tons de cinza.



Fonte: Autoria própria

cada configuração das redes convolucionais foram submetidas a 32 experimentos baseados em técnicas de *design* experimental descritas nas pesquisas de [Easterling \(2015\)](#), [Zheng \(2015\)](#), [Cash e Stanković \(2016\)](#) e [Berger e Maurer \(2018\)](#).

As configurações, bem como os parâmetros das redes neurais convolucionais foram fundamentadas em pesquisas como as de [Tripathi et al. \(2017\)](#), [Tarnowski et al. \(2018\)](#), [Moon, Jang e Lee \(2018\)](#), [Teo e Chia \(2018\)](#), [Wang et al. \(2018\)](#) e [Yang et al.](#)

Tabela 8 – Parâmetros avaliados para configurar o classificador mais adequado.

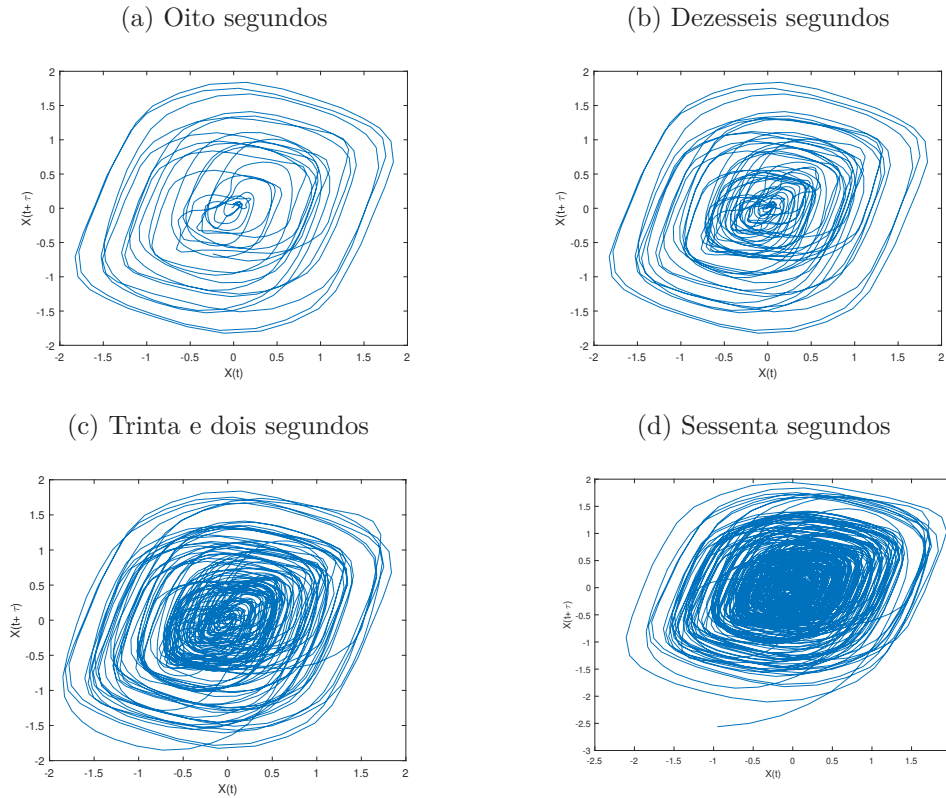
Parâmetros	Variações
Quantidade de Camadas convolucionais	1 – 6
Dimensão dos filtros	$3 \times 3 - 5 \times 5$
Otimizador	<i>SGD, Adadelta, Adam</i>
Dropout	0, 1 – 0, 5
Regularizador	L1, L2 e L1L2
Função de ativação	ReLU, Sigmoid
Pool	<i>Average, Max</i>
Dimensão do pool	$1 \times 1 - 3 \times 2 \times 4$
Épocas	40 – 1000
Normalização do batch	Sim, Não
Quantidade de camadas densas	1 – 4
Nós da camada densa	2 – 256
Tamanho do sinal de EEG	8 – 60 segundos
Função de perda	<i>binary cross-entropy, categorical cross-entropy</i>
Taxa de aprendizado	0, 1 – 0, 0001
Decaimento do fator de aprendizado	0, 1 – 0, 0001
Canais de EEG	1 – 4
Configuração da entrada da rede	<i>2D, 3D</i>
Quantidade de amostras de treino	140 – 35960
Quantidade de amostras de validação	40 – 1843

(2018). Os parâmetros que sofreram maior alteração foram a quantidade de amostras tanto para treino quanto para validação, que variou sua quantidade a partir da técnica de janela deslizante e também com o auxílio de dados sintéticos gerados pela metodologia de aumento de dados da API Keras (CHOLLET, 2015).

Além dos parâmetros selecionados para avaliação, como a dimensão dos filtros, quantidade de camadas, dimensão dos *pools*, função de regularização, quantidade de amostras de treinamento, quantidade de nós na camada densa, taxa de aprendizado e o *dropout*, sendo apontadas pela literatura base para esta pesquisa (GOODFELLOW; BENGIO; COURVILLE, 2016; GONZALEZ; WOODS, 2018) como principais parâmetros capazes de influenciar no aprendizado da rede, outros parâmetros se mostraram relevantes para a classificação de emoções nos experimentos executados.

Foi observado que o intervalo de tempo do sinal a partir do qual a imagem de PSR é gerada, a presença do aumento de dados, da técnica WESML (KING; ZENG, 2001), o formato da entrada e a quantidade de canais de EEG, foram parâmetros relevantes para a classificação de emoções. De acordo com Kuppens e Verduyn (2017), apesar da possibilidade de um sinal de EEG ter várias emoções ao mesmo tempo com durações e intensidades diferentes, uma única emoção pode ser encontrada em toda a duração de um

Figura 26 – Imagens de PSR da sub-banda gama em diferentes intervalos. (a) PSR no intervalo de oito segundos. (b) PSR no intervalo de dezesseis segundos. (c) PSR no intervalo de trinta e dois segundos. (d) PSR no intervalo de sessenta segundos.



Fonte: Autoria própria

sinal fisiológico (KUPPENS; VERDUYN, 2015).

Os experimentos do trabalho de Pereira et al. (2018) demonstram que é possível classificar emoções com sinais de duração menores de 60 segundos. E trabalhos como os de Yang et al. (2018), demonstram que é possível encontrar informações emocionais em intervalos de tempo em toda a duração de um sinal de EEG da base DEAP (KOELSTRA et al., 2012).

Portanto, decidiu-se que o sinal de EEG que daria origem às imagens de PSR seria testado em diferentes intervalos de tempo, com o objetivo de investigar qual intervalo seria mais adequado para a técnica proposta nesta tese. Então foi definido que as imagens de PSR dos sinais de EEG seriam geradas, conforme a Tabela 8, nos seguintes intervalos 8, 16, 32 e 60, conforme a Figura 26, respectivamente, 26a 26b 26c 26d.

Como pode ser visto na Figura 26, as imagens geradas a partir do PSR possuem uma aparência mais confusa à medida que o intervalo de tempo aumenta, enquanto que de maneira oposta as imagens possuem padrões mais distinguíveis de acordo com a redução

Tabela 9 – Acurácia Média com ausência e presença da técnica WESML.

Presença da técnica WESML	Acurácia Média Treino	Acurácia Média Teste
dados aumentados sem WESML	$0,89 \pm 0,02$	$0,60 \pm 0,15$
sem aumento de dados com WESML	$0,82 \pm 0,10$	$0,58 \pm 0,30$
dados aumentados com WESML	$0,97 \pm 0,03$	$0,88 \pm 0,05$

do intervalo. Uma vez que também empiricamente os resultados foram melhores com este intervalo optamos por trabalhar com o intervalo de oito segundos.

Conforme citado anteriormente, outro parâmetro que se mostrou relevante para a classificação de emoções com a técnica proposta foi a utilização da da técnica WESML (KING; ZENG, 2001). A partir dos resultados de experimentos da Tabela 9, o uso da técnica de WESML (KING; ZENG, 2001), em conjunto com o aumento de dados, proporciona os melhores resultados. No entanto é importante lembrar que essa técnica foi necessária apenas para o reconhecimento de emoções em quatro classes.

Combinando as técnicas de aumento de dados, com o balanceamento das classes e imagens de PSR no intervalo de oito segundos, foram executados os experimentos cujos resultados estão presentes na Tabela 10. De modo que é perceptível que reduzir os intervalos do sinal de EEG melhorou os resultados de classificação. Como os resultados ainda estão distantes dos de trabalhos como os de Taran e Bajaj (2019), acurácia de 0,90, alteramos outro parâmetros experimental com o objetivo de melhorar os resultados.

Tabela 10 – Resultados dos experimentos de reconhecimento de emoções com variações da janela de PSR

Intervalo PSR	Acurácia treinamento	Acurácia Teste
60 segundos	$0,79 \pm 0,06$	$0,55 \pm 0,13$
32 segundos	$0,66 \pm 0,06$	$0,11 \pm 0,30$
16 segundos	$0,64 \pm 0,13$	$0,45 \pm 0,18$
8 segundos	$0,75 \pm 0,06$	$0,74 \pm 0,17$

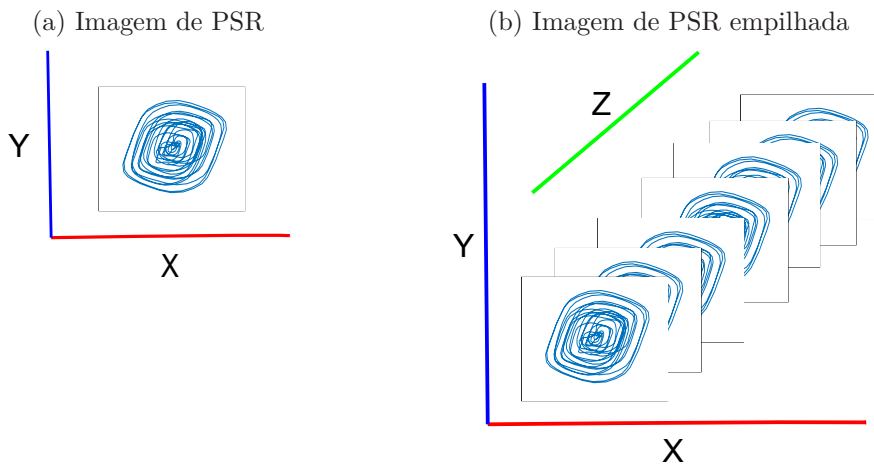
Outra característica que se destacou foi o formato da entrada para a rede convolucional. De acordo com os resultados dos experimentos presentes na Tabela 11, o melhor formato de entrada foi uma pilha de imagens formada pelas diversas imagens de PSR proveniente do fatiamento do sinal original.

Na Tabela 11, há diferentes tipos de redes convolucionais e diferentes formatos de entradas a partir do sinal de PSR, e é importante salientar que, exceto pela dimensão das camadas convolucionais e dos *pools*, possuem as mesmas configurações. A entrada em forma de imagem significa que o sinal foi recortado pela técnica da janela deslizante e

Tabela 11 – Acurácia Média dos diferentes classificadores e formatos de entrada testados

Classificador	Entrada	Canais	Acurácia Média Treino	Acurácia Média Teste
CNN2D	Imagem	4	$0,89 \pm 0,02$	$0,53 \pm 0,03$
CNN2D	Imagem	2	$0,90 \pm 0,30$	$0,50 \pm 0,30$
CNN2D	Pilha	1	$0,89 \pm 0,20$	$0,60 \pm 0,10$
CNN3D	Pilha	4	$0,99 \pm 0,06$	$0,59 \pm 0,01$
CNN3D	Pilha	2	$0,99 \pm 0,04$	$0,63 \pm 0,01$
CNN3D	Pilha	1	$0,97 \pm 0,03$	$0,88 \pm 0,05$

Figura 27 – Imagem de PSR empilhada. (a) Imagem de PSR antes da compactação. (b) Imagem de PSR depois da compactação.



Fonte: Autoria própria

submetidas à rede individualmente e fazendo isso para os quatro canais frontais de EEG do sistema 10-20 da base DEAP (KOELSTRA et al., 2012) ao mesmo tempo.

Conforme os resultados reportados na Tabela 11, utilizar apenas 1 canal produziu melhores resultados do que utilizar mais canais, contrastando com as pesquisas de Wang et al. (2019), Xu et al. (2019) e Wang, Hu e Song (2019), em que vários canais foram utilizados.

É perceptível também que, apenas ao utilizar uma rede convolucional em três dimensões, Figura 27, a pilha de imagens de PSR, Figura 27b, teve um desempenho melhor do que as outras abordagens experimentadas, sendo possível afirmar que a informação emocional pode estar disposta nesta nova dimensão adicionada no experimento. Portanto, a pilha de imagens de PSR possivelmente fornece informações do sinal de EEG no espaço de fase, e que nessa dimensão as características das emoções estão latentes.

Também é preciso deixar claro que foram utilizadas, em todos os experimentos executados, duas variações da validação cruzada (*cross validation*) (BROWNEE, 2017;

ALBON, 2018), a saber, *leave-one-subject-out* (LOSO) e *leave-one-trial-out* (LOTO), de maneira semelhante aos experimentos realizados no trabalho de Pereira et al. (2018). Os resultados experimentais da Tabela 7, Tabela 9, Tabela 10, Tabela 11, são da validação cruzada do tipo LOSO.

Dessa forma, aplicando a metodologia LOSO, os conjuntos de treinamento, validação e teste foram organizados do seguinte modo, para o treinamento foram separadas as amostras de trinta e um indivíduos apenas com dados aumentados. Para a validação foram separadas as amostras originais dos mesmos indivíduos do treino e uma pequena proporção dos dados aumentados. Por fim, para o teste, foram separadas apenas as amostras de um indivíduo, que ficou fora dos conjuntos anteriores.

De maneira semelhante, com a metodologia LOTO, os conjuntos de treinamento, validação e teste, foram divididos de modo que as amostras aumentadas dos trinta e dois indivíduos ficassem restritas ao treinamento. E as amostras originais ficassem divididas entre validação e teste, de modo que os três conjuntos fossem totalmente disjuntos.

A Tabela 12 mostra a configuração dos conjuntos com as diferentes metodologias, LOTO e LOSO. Também é preciso deixar claro que a validação cruzada (*k-fold cross validation*) foi executada com $k = 10$, e que cada experimento foi executado trinta e duas vezes.

Tabela 12 – Dados da base DEAP com o aumento de dados

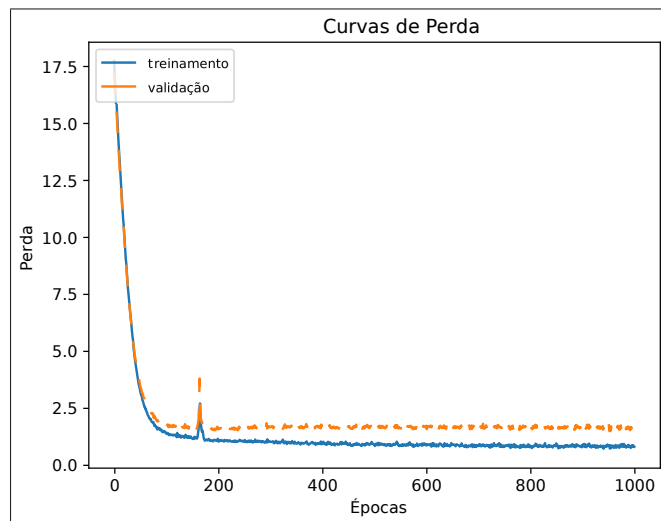
Conjunto	LOTO	LOSO
Conjunto de treino	$7732 \times 7 \times 28 \times 37$	$7732 \times 7 \times 28 \times 37$
Conjunto de validação	$1843 \times 7 \times 28 \times 37$	$1843 \times 7 \times 28 \times 37$
Conjunto de teste	$40 \times 7 \times 28 \times 37$	$40 \times 7 \times 28 \times 37$

Na Figura 28 vê-se o resultado so experimentos utilizando a metodologia LOTO, no qual é possível identificar a redução do erro do classificador durante o treinamento e validação. É perceptível que o erro chega em um intervalo de valores constante entre 200 e 400 épocas, sendo verificado que o treinamento do classificador em até 250 épocas evita sua super-especialização e, conseqüentemente, a perda de capacidade de generalização.

É interessante acrescentar que a evolução do classificador para duas classes também chega a seu menor erro no mesmo intervalo de épocas, tanto para excitação, a Figura 29, quanto para valência, Figura 30. É possível notar que utilizando a metodologia LOSO, as etapas de treino e validação possuem valores semelhantes de redução de erro, mesmo com conjuntos disjuntos, o que indica que o classificador consegue reconhecer a informação emocional nas pilhas de imagens de PSR tanto as aumentadas quanto as originais.

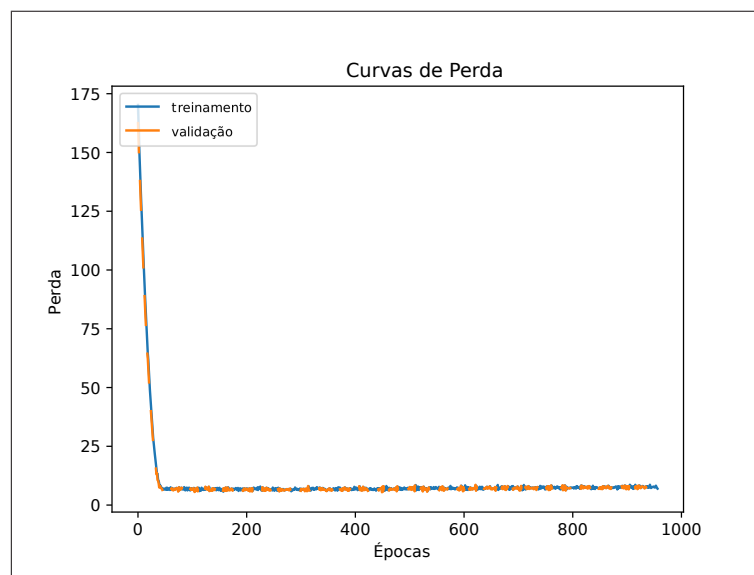
Também é notável que o classificador da Figura 28 tem mais dificuldade em cate-

Figura 28 – Evolução do treinamento do classificador a partir das imagens de PSR para quatro classes



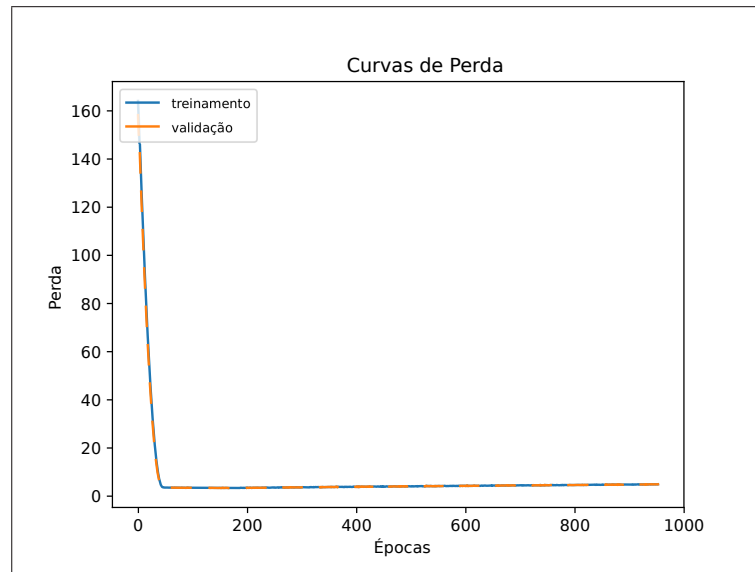
Fonte: Autoria própria

Figura 29 – Evolução do treinamento do classificador a partir das imagens de PSR para duas classes baseadas em excitação



Fonte: Autoria própria

Figura 30 – Evolução do treinamento do classificador a partir das imagens de PSR para duas classes baseadas em valência



Fonte: Autoria própria

gorizar as imagens de PSR em quatro classes, do que em duas classes, Figura 29 e Figura 30. As Figuras 28, Figura 29 e Figura 30 também indicam que é possível desenvolver um classificador de emoções a partir das imagens de PSR. Portanto, a partir da combinação de tipo de rede, tamanho do intervalo, compactação da imagem (MARINUS; ARAÚJO; GOMES, 2019), Figura 25, e formato da entrada, foi possível a melhora nos resultados vistos nas Tabela 13 e Tabela 14.

Além da evidência de que as imagens de PSR empilhadas contêm informação emocional, o formato da entrada, Figura 27, levanta a possibilidade de que a informação emocional esteja disposta também na dimensão temporal. Por fim, é importante deixar claro que nos ativemos a diversas métricas para avaliar o classificador, embora na Tabela 11 tenhamos utilizado a acurácia média dos experimentos, algo necessário devido a trabalhos que não divulgam todas as métricas de avaliação ou trabalham com o foco apenas na acurácia, especificidade e sensibilidade (MOON; JANG; LEE, 2018; ELAMIR; AL-ATABANY; ALDOSOUKY, 2018).

Como podemos visualizar na Tabela 13 as métricas Precisão, Sensibilidade e *F1-Score* estão alinhadas com a acurácia da Tabela 16 para duas classes com o treinamento e teste do tipo LOSO (*leave-one-subject-out*), assim como a Tabela 14 demonstra o mesmo para quatro classes. Deste modo, é possível atestar empiricamente a viabilidade do método confirmando que as estratégias selecionadas para o treinamento do classificador de fato foram bem sucedidas.

Tabela 13 – Média das demais métricas do classificador selecionado para duas classes de valência e excitação

Excitação			
Classes	Precisão	Sensibilidade	<i>F1-Score</i>
0	0,85 ± 0,01	0,82 ± 0,03	0,84 ± 0,04
1	0,87 ± 0,04	0,90 ± 0,02	0,88 ± 0,01
Valência			
Classes	Precisão	Sensibilidade	<i>F1-Score</i>
0	0,87 ± 0,04	0,88 ± 0,17	0,88 ± 0,01
1	0,90 ± 0,01	0,91 ± 0,02	0,90 ± 0,03

Tabela 14 – Média das demais métricas do classificador selecionado para quatro classes

Classes	Precisão	Sensibilidade	<i>F1-Score</i>
0	0,85 ± 0,09	0,94 ± 0,04	0,88 ± 0,04
1	0,97 ± 0,03	0,82 ± 0,77	0,91 ± 0,04
2	0,84 ± 0,05	0,96 ± 0,01	0,93 ± 0,06
3	0,98 ± 0,01	0,90 ± 0,07	0,89 ± 0,02

Com base nos resultados evidenciados nas Tabelas 13 e 14, a configuração do classificador selecionada pode ser vista na Tabela 15. Apesar de utilizar uma técnica de aprendizagem profunda (*deep learning*) a rede convolucional é rasa, com apenas duas camadas convolucionais e três densas. Outro aspecto que merece destaque é que a mesma rede foi utilizada para classificar as imagens de PSR em duas e quatro classes independentes do indivíduo, alterando apenas a função de perda, conforme a Tabela 15.

A Figura 31 mostra a arquitetura final do classificador utilizado tanto para quatro quanto para duas classes. Para a classificação de duas classes basta alterar a quantidade de nós de saída da camada *softmax* da rede neural convolucional para dois nós, e a função de perda para *binary cross-entropy*. Para quatro classes é preciso alterar os nós de saída para quatro e a função de perda para *categorical cross-entropy*.

Foi verificado também, a partir dos experimentos desta tese que a função de perda *binary cross-entropy* funciona melhor para a classificação binária, tanto para excitação quanto para valência, e que a função *categorical cross-entropy* funciona melhor para quatro classes. Ou seja, que as métricas dos experimentos foram superiores quando utilizamos as funções conforme explicado.

Sendo assim, após a execução dos experimentos é possível afirmar que os resultados foram promissores tendo em vista a Tabela 14 e Tabela 13, ou seja, foi possível executar o reconhecimento de emoções utilizando as imagens de PSR dos sinais de EEG

Tabela 15 – Parâmetros selecionados para o classificador

Parâmetros	Dimensões
Quantidade de camadas convolucionais	2
Tamanhos das camadas convolucionais	50 – 12
Dimensão dos filtros	$5 \times 5 \times 7$
Otimizador	<i>SGD</i>
Regularizador	L1
Função de ativação	ReLU
Pool	<i>Max</i>
Quantidade de pools	2
Dimensão do pool	$2 \times 2 \times 1$ e $1 \times 1 \times 1$
Épocas	250
Normalização do batch	Sim
Quantidade de camadas densas	3
Nós da camada densa	256 – 128 – 64
Tamanho do sinal de EEG	8 segundos
Função de perda	bce*, cce**
Taxa de aprendizado	0,001
Decaimento do fator de aprendizado	0,0001
Canais de EEG	1 (FP1)
Configuração da entrada da rede	<i>3D</i>
Quantidade de amostras de treino	7732
Quantidade de amostras de validação	1843
Quantidade de amostras de teste	40

¹Binary cross-entropy (bce), categorical cross-entropy(cce).

²*Resultado para duas classes, **Resultado para quatro classes.

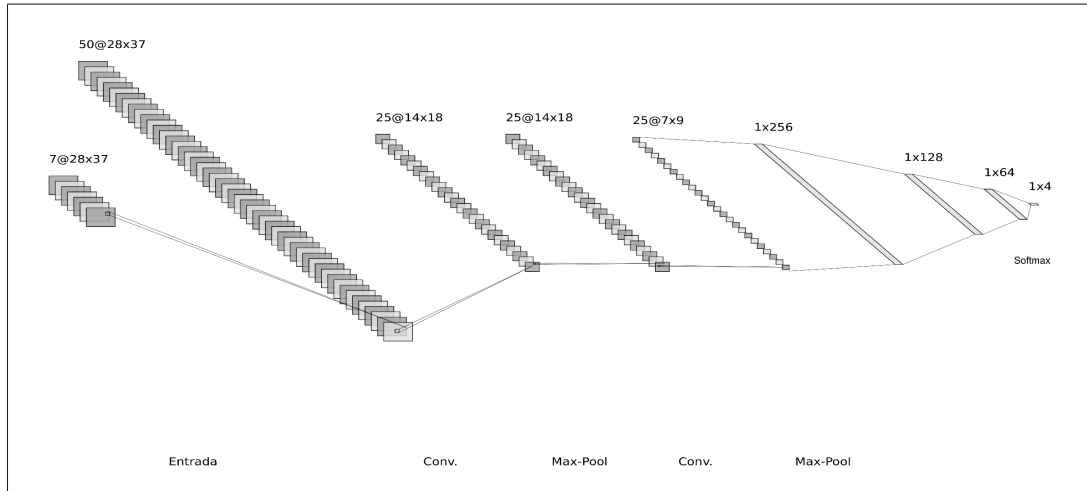
da base DEAP. Posteriormente discutimos as possíveis implicações destes resultados e comparando-os com outros trabalhos do estado da arte.

4.3 Resultados e discussão

Os resultados dos experimentos executados nesta tese alcançaram resultados similares ao estado da arte, que estão resumidos em duas tabelas, as pesquisas selecionadas para comparação compreenderam os anos de 2017 até o primeiro semestre de 2020, a Tabela 16, na qual constam experimentos de outros autores para comparação entre métodos e resultados alcançados para duas classes da valência (V) e excitação (*arousal*, A). A Tabela 17, na qual são comparados os resultados de trabalhos no estado da arte para classificação de emoções em quatro classes.

Conforme apresentado anteriormente, os experimentos foram executados com duas metodologias diferentes, LOSO e LOTO, utilizando como entrada para o classificador,

Figura 31 – Rede neural convolucional utilizada para classificação.



Fonte: Autoria própria

Tabela 16 – Comparação de resultados para duas classes

ID	Característ.	Dataset	Classif.	Metodologia	Acurácia Teste
Tripathi et al. (2017)	CNN 2D	DEAP	CNN 2D	LOTO	0,81 (V) 0,73 (A)
Metodologia proposta	PSR	DEAP	CNN 3D	LOSO	0,88(V) ± 0,05
Metodologia proposta	PSR	DEAP	CNN 3D	LOTO	0,88 (A) ± 0,05
Tarnowski et al. (2018)	EEG e GSR	Própria	SVM	LOTO	0,96 ± 0,05 (A)
Moon, Jang e Lee (2018)	PLV	DEAP	CNN	LOTO	0,99 (V)
Teo e Chia (2018)	FFT	Própria	DNN	LOTO	0,88 (A)
Wang et al. (2018)	TM	DEAP	CNN 3D	LOTO	0,72 (V) 0,73 (A)
Yang et al. (2018)	2D EEG Frames	DEAP	PCRNN	LOTO	0,90 ± 3,08 (V)
					0,91 ± 2,99 (A)

¹ Power Spectral Density (PSD), Higher-Order Crossings (HOC), Phase Space Reconstruction (PSR), Convolutional Neural Network (CNN), Galvanic Skin Response (GSR), Phase Locking Value (PLV), Deep Neural Networks (DNN), Topological Map (TM), Parallel Convolutional Recurrent Neural Network (PCRNN).

²*Resultado para quatro classes.

cujas configurações podem ser vistas na Figura 31 e na Tabela 15, as imagens de PSR a partir dos sinais de EEG da base DEAP (KOELSTRA et al., 2012).

Também é importante salientar que, há algumas dificuldades em comparar trabalhos da área de reconhecimento de emoções, principalmente quando são bases de dados diferentes, uma vez que essas bases não seguem uma configuração experimental padronizada (RIM et al., 2020). Entre outras dificuldades, também vale destacar que, como pode ser visto na Tabela 16 e na Tabela 17, métricas como o desvio padrão ou tipo de validação cruzada utilizada não são divulgados.

Como é possível ver na Tabela 16 quando comparamos a acurácia dos experimentos executados no presente trabalho, a característica do PSR utilizada para alimentar a rede convolucional (CNN) é comparável aos resultados Rim et al. (2020), superior aos de Li et al. (2013) e Tripathi et al. (2017). Porém, segundo Tripathi et al. (2017) nos experimentos com a metodologia LOSO seus resultados ficaram entre 60 e 70% de acurácia tanto para valência quanto para excitação.

Consideramos que a metodologia LOSO impõe uma maior dificuldade aos classificadores uma vez que a etapa de teste é executada com dados de um indivíduo desconhecido para o algoritmo, porém os resultados do algoritmo LOTO não devem ser descartados uma vez que demonstra que o classificador consegue reconhecer os padrões de indivíduos que participaram da etapa de treino mesmo que posteriormente esses dados não sejam conhecidos.

Na pesquisa de Tarnowski et al. (2018), os autores utilizam os sinais de GSR para acompanhar o evento emocional até seu declínio. A partir da baixa do sinal de GSR, os estudiosos recortam o sinal de EEG em intervalos de cinco segundos, tentando isolar o momento do fenômeno emocional. De acordo com os resultados do artigo, esse método se mostrou bem sucedido uma vez que o alcançou a acurácia de 0,96 para excitação. Porém, vale salientar outras características que podem ter melhorado a acurácia do classificador, por exemplo, o sinal da base construída para o experimento, os autores Tarnowski et al. (2018)(TARNOWSKI et al., 2018) utilizaram a resolução de $512Hz$ do sinal de EEG captado, aumentando o nível de detalhe do sinal mesmo em fatias de cinco segundos (enquanto a base DEAP capta o sinal de EEG em $128Hz$).

Assim como na pesquisa de Tarnowski et al. (2018), o artigo de Moon, Jang e Lee (2018) também utiliza o recorte dos sinais de EEG com um intervalo de tempo menor (três segundos) na base DEAP. A característica com os melhores resultados é a PLV (*phase locking value*), embora tenham sido avaliadas outras como PLI (*phase lag index*) e PCC (*Pearson correlation coefficient*). Novamente, a partir do janelamento do sinal de EEG, o trabalho de Wang et al. (2018) utiliza janelas de 4 segundos subdivididas em 2 segundos para criar uma sequência temporal que evidencie o evento emocional no tempo e no espaço através do mapa topológico de ativações do sinal, conseguindo assim uma representação espaço-temporal do mesmo. Yang et al. (2018) segue uma metodologia semelhante a Wang et al. (2018), transformando os dados da DEAP em uma sequência de quadros a partir do recorte do sinal original em intervalos de 1 (um) segundo e sem superposição, dessa forma cada sinal teria 60 (sessenta) quadros.

Com esse tipo de abordagem, tanto Wang et al. (2018) quanto Yang et al. (2018), como pode ser visto na Tabela 16, conseguiram os resultados com maiores acurácias quando comparados com outras pesquisas. Os autores recortaram o sinal em intervalos de 8 (oito) segundos com superposição de 1 (um) segundo, resultando em 7 quadros por sinal.

Além disso também utilizamos apenas um canal como entrada da CNN, o FP1, uma vez que os canais frontais são os que possuem informação emocional mais proeminente (Zheng e Lu, 2015; Picard, 2000; Damasio, 1995), diferente de Wang et al. (2018) e Yang et al. (2018) que utilizam os 32 (trinta e dois) canais da base DEAP (Yang et al., 2018).

Tabela 17 – Comparação de resultados para quatro classes

ID	Característ.	Dataset	Classif.	Metodologia	Acurácia Teste
Munoz et al. (2018)	PSD e DE	MAHNOB HCI	SVM	N.I.	$0,92 \pm 0,00$
Taran e Bajaj (2019)	FF	Particular	MC-LS-SVM	LOTO	0,90
Zangeneh et al. (2019a)	PP	DEAP	MSVM	N.I.	0,81
Zheng et al. (2019)	PSD e DE	Particular	DNN	N.I.	$0,85 \pm 0,12$
Metodologia proposta	PSR	DEAP	CNN 3D	LOSO	$0,84 \pm 0,07$
Metodologia proposta	PSR	DEAP	CNN 3D	LOTO	$0,92 \pm 0,08$

– *multi-class least squares support vector machine* (MC-LS-SVM), *fused features* (FF), *multi-class support vector machine* (MSVM), *poincare points* (PP), *differential entropy* (DE)

– N.I. = Não identificado.

A Tabela 17 mostra que a acurácia da pesquisa de Munoz et al. (2018) está entre as mais altas dos trabalhos selecionados e equivalente à classificação da metodologia proposta nesta tese. No entanto, vale salientar que apesar de utilizar a validação cruzada não fica claro se a classificação é independente do indivíduo, ou se está aplicando LOTO ou LOSO para alcançar esses resultados. Outra observação pertinente é quanto a base, como nesta tese não utilizamos a base MAHNOB HCI (MUNOZ et al., 2018), não é possível afirmar como a técnica aqui proposta se comportaria.

Também é possível considerar o trabalho de Taran e Bajaj (2019) equivalente, mesmo que os resultados apresentados no estudo não sejam acompanhados dos desvios-padrão. Outra observação é que a base de dados utilizada foi criada pelos pesquisadores, e para uma comparação mais precisa seria necessário replicar o experimento do estudo em outras bases.

Um detalhe importante é que os estudos de Munoz et al. (2018) e de Taran e Bajaj (2019), utilizam poucos canais para alcançar os resultados de 0,92 e 0,90, respectivamente. Enquanto os resultados da metodologia proposta alcança resultados equivalentes com apenas um canal de EEG.

O estudo de Zangeneh et al. (2019a) é uma comparação importante pois utiliza características do espaço de fase extraídas a partir das imagens de PSR do sinal de EEG da base DEAP. A metodologia proposta nesta tese pode ser superior ao resultado de 0,81 de Zangeneh et al. (2019a), na metodologia LOTO com $0,92 \pm 0,08$, e que pode ser considerada competitiva na metodologia LOSO com $0,84 \pm 0,07$. Sem o desvio-padrão do trabalho de Zangeneh et al. (2019a), ou a metodologia de treinamento, a comparação não é precisa.

Os resultados do trabalho de Zheng et al. (2019) são equivalentes ao da metodologia aqui proposta, alcançando $0,85 \pm 0,12$ contra $0,92 \pm 0,08$ com o método LOTO, e $0,84 \pm 0,07$ com o método LOSO. Ou seja, ambos métodos, os resultados da presente tese são superiores ao de Zheng et al. (2019).

Então, a Tabela 16 e a Tabela 17 mostram que os resultados alcançados com a técnica proposta na presente tese são competitivos com o estado da arte. Para uma comparação mais precisa seria necessário replicar os experimentos com as mesmas condições, no entanto, a comparação é válida uma vez que serve para demonstrar a viabilidade da técnica proposta.

A partir da Figura 32 e da Figura 33, temos um gráfico de intervalos gerado a partir das médias e desvios-padrão dos resultados comparados, no qual é possível verificar que a técnica proposta é competitiva e equivalente aos resultados com acurácia mais alta identificados na revisão da literatura desta tese. Uma vez que os resultados da técnica proposta possuem uma grande interseção com os intervalos dos resultados das pesquisas na Tabela 16 e Tabela 17, conforme pode-se notar nas Figura 32 e Figura 33, é possível afirmar que a abordagem desta tese é equivalente ao estado da arte apresentado.

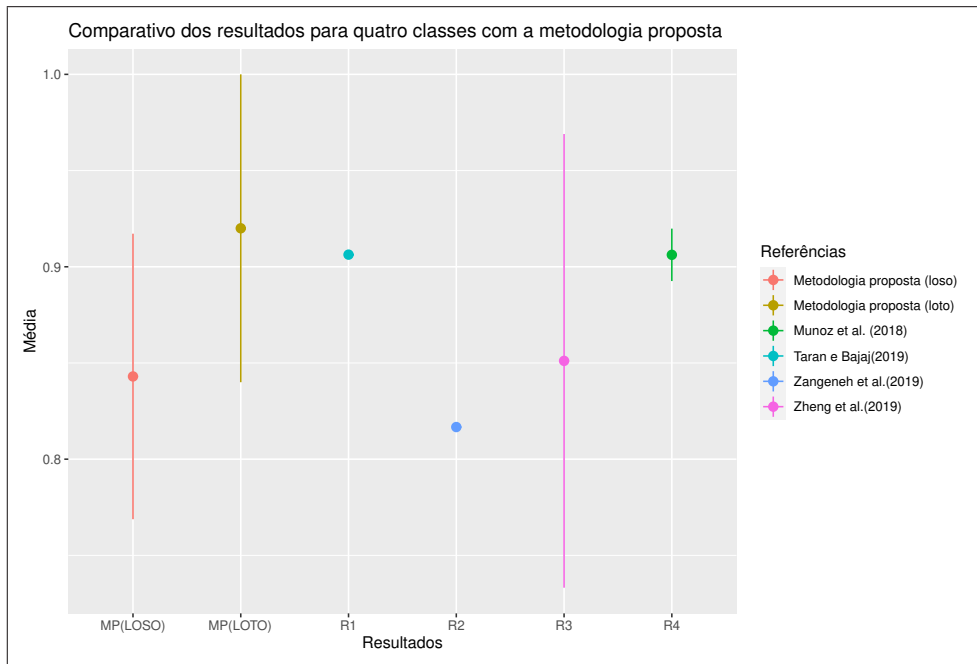
Figura 32 – Comparação dos intervalos das médias dos resultados para duas classes da metodologia proposta



Fonte: Autoria própria

Portanto, os resultados confirmam a hipótese de que as imagens de PSR podem ser utilizadas para o reconhecimento de emoções. Além disso, quando utilizadas como uma sequência de quadros como mostrado na Tabela 11, os resultados melhoraram signi-

Figura 33 – Comparação dos intervalos das médias dos resultados para quatro classes da metodologia proposta



Fonte: Autoria própria

ficativamente, por volta de 40% após organizar as imagens desta forma. De modo que, é possível que a informação emocional esteja mais identificável na dimensão edicional das imagens de PSR.

Tabela 18 – Comparação quanto a quantidade canais utilizados para classificação.

Referência	Quantidade de Canais
Tarnowski et al. (2018)	16
Moon, Jang e Lee (2018)	32
Teo e Chia (2018)	5
Wang et al. (2018)	10 (A) e 8 (V)
Munoz et al. (2018)	2
Taran e Bajaj (2019)	1
Zangeneh et al. (2019a)	32
Zheng et al. (2019)	6
Metodologia proposta	1

























































Um resultado também relevante é o uso de apenas um canal de EEG para o reconhecimento de emoções, como mostra a Tabela 18, comparando com os estudos selecionados a partir da revisão de literatura, temos que a metodologia proposta é o que utiliza a menor quantidade de canais, assim como Taran e Bajaj (2019). Esse tipo de resultado é impor-



































































































tante para a viabilização de sensores mais simples, ou execução de experimentos [Ogino et al. \(2019\)](#).



































































































A Tabela 19 mostra as imagens de PSR geradas a partir da técnica da janela deslizante ([TARNOWSKI et al., 2018](#); [MOON](#); [JANG](#); [LEE, 2018](#); [WANG et al., 2018](#)) para obter as sequências de quadros, sendo a nomenclatura V1-C1 significa vídeo 1 da classe 1. Esses padrões são submetidos à CNN como entrada, após a compactação via a técnica de contagem de caixas ([MARINUS](#); [ARAÚJO](#); [GOMES, 2019](#)). Na Tabela 19 as classes discriminadas na primeira coluna são os dois níveis de excitação.





























Conforme a Tabela 11 mostra, foi a sequência de imagens que permitiu os resultados competitivos da Figura 32 e Figura 33, de modo que é possível que as informações emocionais dispostas no espaço de fase sejam mais identificáveis durante a evolução das imagens no tempo.

Tabela 19 – Sequência de quadros de imagens de PSR de um sinal de EEG.

Base DEAP, Indivíduo 1, Canal FP1, banda Gama, 8 s com superposição de 1 s.							
Amostra	Fatia 1	Fatia 2	Fatia 3	Fatia 4	Fatia 5	Fatia 6	Fatia 7
V1-C1							
V2-C1							
V3-C1							
V4-C1							
V5-C0							
V6-C0							
V7-C0							
V8-C0							

Continuação da Tabela 19							
Amostra	Fatia 1	Fatia 2	Fatia 3	Fatia 4	Fatia 5	Fatia 6	Fatia 7
V9-C0							
V10-C0							
V11-C0							
V12-C0							
V13-C0							
V14-C0							
V15-C1							
V16-C1							
V17-C1							
V18-C1							
V19-C1							
V20-C1							
V21-C1							
V22-C0							

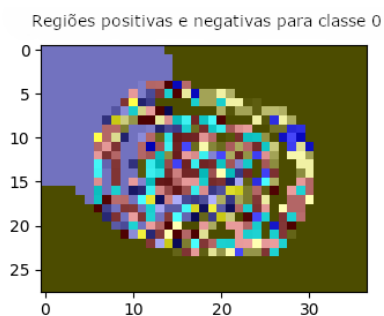
Continuação da Tabela 19							
Amostra	Fatia 1	Fatia 2	Fatia 3	Fatia 4	Fatia 5	Fatia 6	Fatia 7
V23-C0							
V24-C1							
V25-C1							
V26-C1							
V27-C1							
V28-C0							
V29-C0							
V30-C0							
V31-C1							
V32-C1							
V33-C0							
V34-C1							
V35-C1							
V36-C1							

Continuação da Tabela 19							
Amostra	Fatia 1	Fatia 2	Fatia 3	Fatia 4	Fatia 5	Fatia 6	Fatia 7
V37-C1							
V38-C1							
V39-C1							
V40-C1							
Fim da Tabela							

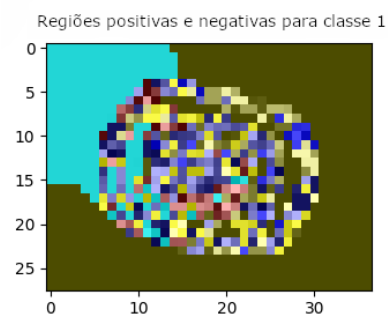
Como resultado adicional, tentou-se encontrar um auxílio visual que ajudasse a explicar como o classificador aprendeu os padrões emocionais nas pilhas de imagem de PSR, recorreremos ao LIME (*Local Interpretable Model-agnostic Explanations*) (RIBEIRO; SINGH; GUESTRIN, 2016; ELSHAWI et al., 2019; ZHANG et al., 2018), uma técnica que consiste em analisar as regiões em que os pesos dos mapas de características (*feature maps*) da rede convolucional são mais ativados, e a partir disso e da imagem de entrada, destacar essas regiões. A saída do algoritmo LIME está presente na Figura 34.

Figura 34 – Imagem de PSR empilhada. (a) Regiões da classe 0 baseada em excitação. (b) Regiões da classe 1 baseada em excitação.

(a) Regiões de interesse da rede convolucional para a classe 0



(b) Regiões de interesse da rede convolucional para a classe 1

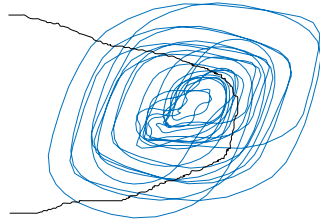


Fonte: Autoria própria

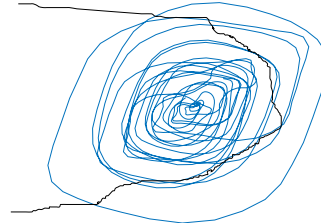
Observando a Figura 34, as regiões em que os pesos dos mapas de características (*feature maps*) são mais ativados encontram-se na região do centróide da imagem do PSR e também em regiões específicas da trajetória do espaço de fase. É interessante notar que

Figura 35 – Área de destaque dos mapas de características do classificador. (a) Área de destaque da classe 0. (b) Área de destaque da classe 1. (c) Área de destaque da classe 2. (d) Área de destaque da classe 3.

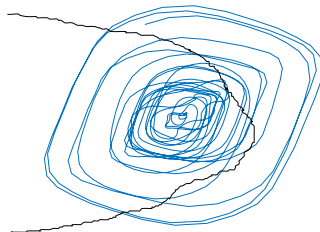
(a) Área de destaque da classe 0



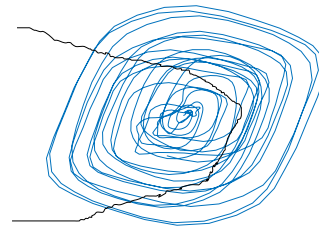
(b) Área de destaque da classe 1



(c) Área de destaque da classe 2



(d) Área de destaque da classe 3



Fonte: Autoria própria

esse padrão se repete para todas as imagens de PSR para as duas classes de valência (*valence*, classe 1) e excitação (*arousal*, classe 0). Dessa forma, é possível afirmar que essas regiões das imagens de PSR possuem informação emocional do sinal de EEG que são úteis para classificação de estados emocionais.

Então, a partir da Figura 34a e da Figura 34b destacaram-se as áreas de interesse do classificador na Figura 35 para as quatro classes baseadas em valência e excitação e estes padrões se repetem nas demais fatias do PSR. Sendo assim, é possível inferir que as informações emocionais estejam concentradas em torno do atrator da imagem do PSR do sinal de EEG e que a evolução das órbitas em torno dele que definam as classes. No entanto, estes resultados precisam de uma avaliação mais detalhada.

Por fim, neste capítulo foi possível verificar a viabilidade e o sucesso da metodologia proposta a partir dos resultados da Tabela 16 e da Tabela 17. Considerando a comparação com os resultados de outros trabalhos do estado da arte, vê-se que utilizar as imagens de PSR para classificar emoções é uma estratégia promissora e que evidencia a existência de informações emocionais na dimensão da fase do sinal de EEG. Também obteve-se um resultado extra, da Figura 34 e da Figura 35, nas quais é perceptível uma área específica das imagens de PSR que podem conter informações emocionais.

5 Conclusões e pesquisas futuras

Este capítulo relata as conclusões do presente trabalho, bem como as contribuições e avanços para o estado da arte na classificação de emoções a partir de sinais de EEG. Além disso, há uma discussão de como se pretende aprimorar e utilizar a abordagem proposta em trabalhos futuros.

5.1 Conclusões

O objetivo principal da presente tese foi desenvolver um método para classificar emoções a partir de imagens de PSR de sinais de EEG. Desta forma, considerando os resultados alcançados reportados na Tabela 16 e na Tabela 17, é possível afirmar que este objetivo foi alcançado.

Desenvolvemos uma abordagem de classificação que consiste em alimentar uma rede convolucional com a arquitetura definida na Tabela 15, com imagens de PSR pré-processadas conforme a Figura 17, é importante salientar que conforme os resultados demonstraram, Tabela 16 e Tabela 17, só foi possível após o empilhamento das imagens de PSR.

Estudos como os de Zangeneh et al. (2019a) mostram que a dinâmica do espaço de fase do sinal de EEG possui informações emocionais, de modo que é possível que as informações da dinâmica do espaço não esteja contida em uma imagem de PSR mas que possa ser aprendida pelo classificador a partir de uma pilha de imagens de PSR. Os resultados da Tabela 11 corroboram com esta afirmação, de modo que a acurácia dos resultados melhoram cerca de 40% após utilizar como entrada da rede convolucional a pilha de imagens de PSR. Outro detalhe importante é que apenas a rede convolucional 3D conseguiu aprender a dinâmica do sistema a partir da pilha de imagens, com uma rede convolucional 2D não foi possível alcançar os mesmos resultados.

No que diz respeito aos objetivos específicos, verificou-se experimentalmente que as imagens de PSR possuem informações emocionais, embora tenha-se conseguido alcançar essas informações apenas após empilhar as imagens de PSR. Também foi identificada a necessidade de utilizar metodologias para aumento de balanceamento de dados, conforme evidenciam a Tabela 7 e a Tabela 9. No entanto, o balanceamento via WESML (KING; ZENG, 2001) só foi necessário devido ao desbalanceamento da base DEAP (KOELSTRA et al., 2012), conforme pode ser visto na Figura 15. Porém, o procedimento de balanceamento não foi necessário quando a classificação era realizada para apenas duas classes de excitação e valência, uma vez que o desbalanceamento não é significativo, conforme

Figura 16.

Revisões como as de [Christensen e Abdullah \(2018\)](#), [D’Mello, Kappas e Gratch \(2018\)](#), [Lotte et al. \(2018\)](#), [Yu et al. \(2019\)](#) e [Dzedzickis, Kaklauskas e Bucinskas \(2020\)](#), nos permitem concluir que não foram encontradas, entre as pesquisas selecionados para esta tese, características adequadas para o reconhecimento de emoções via sinais de EEG, bem como que há a necessidade de se selecionar os canais de EEG mais relevantes para a classificação de emoções. Trabalhos como os de [Yu et al. \(2019\)](#) e [Taran e Bajaj \(2019\)](#), respectivamente, seguem nessa direção.

A metodologia proposta nesta tese, se alinha com o estado da arte ([GARCÍA-MARTÍNEZ et al., 2017](#); [CHRISTENSEN; ABDULLAH, 2018](#); [DAWID, 2019](#)), fornecendo um novo recursos tecnológico como alternativa para o reconhecimento de emoções a partir do espaço de fase do sinal de EEG, bem como, apresenta a nova abordagem para o reconhecimento de emoções via imagens de PSR.

Também é possível afirmar que a presente tese fornece uma alternativa para o reconhecimento de emoções com apenas um canal de EEG, conforme pode ser visto na Tabela 18, esta característica da metodologia proposta pode contribuir com o desenvolvimento de equipamentos mais simples, ou seja, com menos canais, e vestíveis para o registro de sinais de EEG, como os do estudo de [Athavipach, Pan-Ngum e Israsena \(2019\)](#), [Suzuki et al. \(2019\)](#) e [Kosmyna et al. \(2019\)](#). De modo que, uma vez seja possível efetuar o reconhecimento de emoções com apenas um canal ([TARAN; BAJAJ, 2019](#)), dispositivos vestíveis se tornar mais viáveis [Ogino et al. \(2019\)](#).

Como resultado extra também é possível citar as regiões das imagens de PSR destacadas pelo LIME ([RIBEIRO; SINGH; GUESTRIN, 2016](#); [ZHANG et al., 2018](#); [ELSHAWI et al., 2019](#)), a partir dos mapas de características (*feature maps*) da rede convolucional, Figura 34. No entanto, é preciso uma análise mais detalhada destas imagens, bem como testar outras técnicas de interpretabilidade do aprendizado de máquina ([ELSHAWI et al., 2019](#)), de modo a confirmar que nas regiões apontadas pelo LIME ([RIBEIRO; SINGH; GUESTRIN, 2016](#)) na Figura 34 e Figura 35, também são destacadas utilizando outras técnicas como SHAP e Anchors ([ELSHAWI et al., 2019](#)).

Portanto, é possível concluir que a abordagem desenvolvida no presente trabalho permite efetuar o reconhecimento de emoções a partir de imagens de PSR de sinais de EEG, juntamente com um classificador do tipo rede neural convolucional 3D. Os resultados demonstraram que a técnica desenvolvida é robusta e serve tanto para a classificação em duas quanto em quatro classes, precisando apenas alterar a saída da camada softmax da rede e a função de perda, conforme Tabela 15. Sendo assim, foi evidenciado experimentalmente que as imagens do espaços de fase empilhadas podem fornecer informações emocionais do sinal de EEG nessa dimensão.

Por fim, é possível pontuar também que a metodologia apresentada nesta tese contribuiu com o estado da arte do reconhecimento de emoções a partir de sinais de EEG, quando comparada às pesquisas desenvolvidas por estudiosos, mencionadas no trabalho, cujos resultados obtidos se encontram presentes na Tabela 16 e na Tabela 17.

5.1.1 Principais contribuições

Os resultados presentes na tese foram alcançados utilizando apenas um canal de EEG (FP1) do esquema 10–20 (KOELSTRA et al., 2012), fazendo com que este trabalho se destaque com relação aos demais contidos na Tabela 16 e Tabela 17, os quais utilizaram uma quantidade de canais maior. Defendemos que a utilização de apenas um canal torna mais prática a coleta dos sinais de EEG e pode auxiliar no uso de uma quantidade menor de sensores em experimentos voltados para atividades diárias ou em ambientes fora do laboratório (KAWANA et al., 2019).

Diante do exposto, é possível listar as seguintes contribuições deste trabalho:

- Utilização de imagens de PSR para a classificação de emoções;
- Empilhamento de imagens de PSR para fornecer informações emocionais do sinal de EEG;
- Utilização de apenas um canal (FP1) para reconhecimento de emoções;
- O método de classificação empregado funciona tanto para a divisão dos rótulos emocionais em duas quanto em quatro classes;

No entanto, também é preciso reconhecer as principais limitações da abordagem apresenta nesta tese. Demonstrando como a abordagem proposta pode ser aprimorada e o que ainda pode ser investigado a partir dela.

5.1.2 Limitações desta pesquisa

Assim como foram identificadas as principais contribuição desta tese, também é preciso apresentar quais são as limitações da solução apresentada. É preciso salientar que os resultados alcançados foram verificados para a base DEAP (KOELSTRA et al., 2012), de modo que que é necessário realizar experimentos com a abordagem desenvolvida em, ou entre, outras bases de sinais de EEG para verificar a robustez do procedimento (RAYAT-DOOST; SOLEYMANI, 2018).

Também, se faz necessária uma investigação sobre os canais de EEG utilizados para a classificação de emoções a partir de imagens de PSR. Tendo em vista que, a seleção foi realizada nos canais frontais Fp1, Fp2, F3 e F4, devido a evidências de que

são canais com informações emocionais latentes (RAYATDOOST; SOLEYMANI, 2018; CHRISTENSEN; ABDULLAH, 2018; YU et al., 2019; DZEDZICKIS; KAKLAUSKAS; BUCINSKAS, 2020), é possível que outros canais também possam ser utilizados e talvez alcançar resultados melhores.

Além disso, é possível que a pilha de imagens, bem como, as imagens de PSR individualmente possam conter características que podem ser extraídas e utilizadas para a classificação de emoções. As imagens de PSR não foram analisadas de maneira detalhada, quanto a possíveis características relacionadas ao reconhecimento de emoções (ZANGENEH SOROUSH et al., 2018) de modo que podem revelar mais informações sobre os estados emocionais registrados no sinal de EEG.

A presente tese também não avaliou outros classificadores além da rede neural convolucional e existe a possibilidade de que outras técnicas de *deep learning* (RIM et al., 2020) sejam mais adequadas para as imagens de PSR. De modo que, não é possível concluir que não há informações emocionais nas imagens de PSR ou que apenas quando empilhadas que possibilita a classificação de emoções.

Por fim, mesmo com as limitações apresentadas, é possível afirmar que apresentamos resultados positivos, de modo a colaborar com o estado da arte no que diz respeito ao reconhecimento de emoções utilizando a reconstrução do espaço de fase de sinais de EEG na sub-banda gama de apenas um canal de EEG (Fp1), com classificação independente do indivíduo, podendo auxiliar na adoção de sensores mais simples em pesquisas da área e apresentar uma nova ferramenta tecnológica para a identificação de estados emocionais.

5.1.3 Sugestões para Pesquisas Futuras

Devido à importância das emoções para os seres humanos (KAHNEMAN, 2011), espera-se que os resultados alcançados na presente tese possam auxiliar numa melhor compreensão dos estados emocionais, principalmente no que concerne à identificação em atividades de classificação (LOTTE et al., 2018; YU et al., 2019; DZEDZICKIS; KAKLAUSKAS; BUCINSKAS, 2020). Também é esperado que as imagens de PSR revelem um novo recurso tecnológico para a identificação de informações emocionais em sinais de EEG (GARCÍA-MARTÍNEZ et al., 2017), assim como possibilite a adoção de sensores mais simples e vestíveis (SUZUKI et al., 2019; KOSMYNA et al., 2019; OGINO et al., 2019).

No entanto, levando em consideração as limitações da abordagem desenvolvida nesta tese apresentadas na seção anterior, sugere-se as seguintes linhas de pesquisas futuras:

- Análise de características das imagens de PSR de sinais de EEG geradas com os parâmetros da abordagem proposta que possibilitem o reconhecimento de emoções

e melhorar os resultados apresentados na Tabela 16 e Tabela 17;

- Análise de características da pilha de imagens de PSR relacionando-a com informações da dinâmica do espaço de fase do sinal de EEG, com o objetivo de identificar características adequadas para o reconhecimento de emoções;
- Investigar se imagens de PSR geradas a partir de sinais de outros canais de EEG resultam em resultados melhores dos que apresentados nesta tese;
- Investigar a adoção de outras técnicas *deep learning* (RIM et al., 2020) que sejam adequadas para imagens de PSR;
- Verificar a possibilidade de efetuar a classificação de emoções entre diferentes bases de sinais de EEG (RAYATDOOST; SOLEYMANI, 2018), para atestar a robustez do método proposto;
- Investigar as áreas de ativação dos mapas de características do classificador geradas a partir de diferentes técnicas de interpretabilidade do aprendizado de máquina (ELSHAWI et al., 2019), de modo a identificar padrões visuais nas imagens de PSR que auxiliem na classificação de emoções.

Desse modo, espera-se que as sugestões de pesquisa possam contribuir para o diagnóstico de estados emocionais, refinando a metodologia proposta nesta tese e identificando novas características adequadas para o reconhecimento de emoções independente do indivíduo e entre diferentes bases.

Também há a possibilidade de contribuir com sistemas de acompanhamento e registro em tempo real de sinais emocionais a partir de dispositivos vestíveis, já que sensores mais simples podem possibilitar o desenvolvimento de equipamentos que serão utilizados diariamente (KOSMYNA et al., 2019).

Por fim, desenvolver um auxílio visual para a classificação de emoções que possa auxiliar na identificação e caracterização de durações, picos, entre outros fenômenos que podem ocorrer durante eventos emocionais (KUPPENS; VERDUYN, 2017), auxiliando assim estudos de dinâmica emocional (KRONE et al., 2018).

Referências Bibliográficas

- AHIRWAL, M. K.; KOSE, M. R. Emotion Recognition System based on EEG signal: A Comparative Study of Different Features and Classifiers. In: *Proceedings of the 2nd International Conference on Computing Methodologies and Communication (ICCMC 2018)*. [S.l.]: IEEE, 2018. p. 472–476. Citado na página 30.
- ALAKU, T. B.; TÜRKOĞLU, İ. Eeg-based emotion estimation with different deep learning models. In: *2019 4th International Conference on Computer Science and Engineering (UBMK)*. [S.l.: s.n.], 2019. p. 33–37. Citado na página 41.
- ALBON, C. *Python Machine Learning Cookbook*. 1. ed. Sebastopol, CA: O'Reilly Media, Inc, 2018. 366 p. Citado 4 vezes nas páginas 78, 79, 84 e 85.
- ATHAVIPACH, C.; PAN-NGUM, S.; ISRASENA, P. A wearable in-ear EEG device for emotion monitoring. *Sensors (Switzerland)*, v. 19, n. 18, p. 1–16, 2019. Citado 2 vezes nas páginas 30 e 102.
- BASTOS-FILHO, T. F. et al. Evaluation of feature extraction techniques in emotional state recognition. In: *International Conference on Intelligent Human Computer Interaction*. Kharagpur, India: IEEE, 2012. p. 1–6. Citado na página 25.
- BERGER, P. D.; MAURER, R. E. *Experimental Design - With Applications in Management, Engineering, and the Sciences*. Second. Cham, Switzerland: Springer International Publishing, 2018. 644 p. Citado 3 vezes nas páginas 69, 79 e 80.
- BRADLEY, M. M.; LANG, P. J. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, v. 25, n. 1, p. 49 – 59, 1994. Citado na página 45.
- BROWNLEE, J. *Machine Learning Mastery With Python - Understand Your Data, Create Accurate Models and Work Projects End-To-End*. Melbourne, Australia: Jason Brownlee, 2017. v. 1.4. 179 p. Citado 7 vezes nas páginas 65, 66, 72, 78, 79, 84 e 85.
- CALVO, R. et al. *The Oxford Handbook of Affective Computing*. [S.l.]: Oxford University Press, 2014. 923 p. Citado na página 29.
- CALVO, R. A.; MELLO, S. D. Affect detection : An interdisciplinary review of models , methods , and their applications. *IEEE Transactions on Affective Computing*, v. 1, n. 1, p. 18–37, 2010. Citado na página 25.
- CANDRA, H. et al. Investigation of window size in classification of eeg-emotion signal with wavelet entropy and support vector machine. In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. [S.l.: s.n.], 2015. p. 7250–7253. Citado na página 77.
- CAO, G. et al. Emotion recognition based on CNN. *Chinese Control Conference, CCC*, Technical Committee on Control Theory, Chinese Association of Automation, v. 2019-July, n. 61372023, p. 8627–8630, 2019. Citado na página 76.

- CASH, P.; STANKOVIĆ, T. *Experimental Design Research - Approaches, Perspectives, Applications*. First. Zurich: Springer International Publishing Switzerland, 2016. 273 p. Citado 3 vezes nas páginas 69, 79 e 80.
- CHEN, Z. G. S.; WANG, S. Emotion recognition from peripheral physiological signals enhanced by eeg. In: *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [S.l.: s.n.], 2016. p. 2827–2831. Citado 2 vezes nas páginas 65 e 66.
- CHOLLET, F. *Keras*. [S.l.]: GitHub, 2015. <<https://github.com/fchollet/keras>>. Citado 7 vezes nas páginas 65, 72, 73, 74, 77, 78 e 81.
- CHRISTENSEN, L. R.; ABDULLAH, M. A. EEG emotion detection review. *2018 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB 2018*, IEEE, p. 1–7, 2018. Citado 9 vezes nas páginas 25, 28, 29, 31, 58, 67, 77, 102 e 104.
- COHEN, X. M. *Analyzing Neural Time Series Data*. [S.l.]: MIT Press, 2014. 1–5 p. (1, 1). Citado 5 vezes nas páginas 32, 33, 35, 36 e 65.
- COSTA, W. C. D. A. *Análise Dinâmica Não Linear de Sinais de Voz para Detecção de Patologias Laríngeas*. Tese (Doutorado) — Universidade Federal de Campina Grande, Campina Grande, PB, Novembro 2012. Citado 3 vezes nas páginas 32, 35 e 36.
- DAMASIO, A. R. *Descartes' Error - Emotion, Reason and Human Brain*. 1. ed. New York, NY, USA: Avon Books, 1995. 3–331 p. Citado 2 vezes nas páginas 67 e 92.
- DAWID, A. PSR-based research of feature extraction from one-second EEG signals: a neural network study. *SN Applied Sciences*, Springer International Publishing, v. 1, n. 12, p. 1–12, 2019. Citado 9 vezes nas páginas 31, 32, 35, 36, 60, 61, 62, 63 e 102.
- D'MELLO, S.; KAPPAS, A.; GRATCH, J. The Affective Computing Approach to Affect Measurement. *Emotion Review*, v. 10, n. 2, p. 174–183, 2018. Citado 2 vezes nas páginas 25 e 102.
- DONG, E.; ZHOU, K.; DU, S. Phase Space Reconstruction Based Multi-Task Classification for Motor Imagery EEG. *Proceedings of 2019 IEEE International Conference on Mechatronics and Automation, ICMA 2019*, IEEE, p. 1260–1264, 2019. Citado na página 32.
- DONG, Q.; GONG, S.; ZHU, X. Imbalanced deep learning by minority class incremental rectification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 41, n. 6, p. 1367–1381, June 2019. Citado 3 vezes nas páginas 55, 71 e 72.
- DU, S. S. et al. How many samples are needed to estimate a convolutional neural network? *Advances in Neural Information Processing Systems*, v. 2018-December, n. NeurIPS 2018, p. 373–383, 2018. Citado na página 78.
- DUTTA, S.; SINGH, M.; KUMAR, A. Classification of non-motor cognitive task in EEG based brain-computer interface using phase space features in multivariate empirical mode decomposition domain. *Biomedical Signal Processing and Control*, Elsevier Ltd, v. 39, p. 378–389, 2018. Citado na página 32.

DZEDZICKIS, A.; KAKLAUSKAS, A.; BUCINSKAS, V. Human emotion recognition: Review of sensors and methods. *Sensors (Switzerland)*, v. 20, n. 3, p. 1–41, 2020. Citado 9 vezes nas páginas 24, 25, 27, 29, 31, 67, 77, 102 e 104.

EASTERLING, R. G. *Fundamentals of Statistical Experimental Design and Analysis*. First. Cedar Crest: Wiley, 2015. 269 p. Citado 2 vezes nas páginas 79 e 80.

EKMAN, P. An argument for basic emotions. *Cognition & Emotion*, v. 6, n. 3-4, p. 169–200, 1992. Citado 4 vezes nas páginas 45, 47, 50 e 68.

EKMAN, P. Basic emotions. In: DALGLEISH, T.; POWER, M. (Ed.). *Handbook of cognition and emotion*. U.K.: John Wiley Sons, Ltd, 1999. cap. 3, p. 45–60. Citado 2 vezes nas páginas 45 e 68.

ELAMIR, M. M.; AL-ATABANY, W.; ALDOSOUKY, M. A. Phase Space Density Matrix for Emotion Recognition. *International Journal of Computer Applications*, v. 179, n. 49, p. 37–42, 2018. Citado 4 vezes nas páginas 61, 62, 68 e 87.

ELSHAWI, R. et al. Interpretability in healthcare a comparative study of local machine learning interpretability techniques. *Proceedings - IEEE Symposium on Computer-Based Medical Systems*, v. 2019-June, p. 275–280, 2019. Citado 3 vezes nas páginas 98, 102 e 105.

FONSECA, M. J.; ALARCAO, S. M. Emotions recognition using eeg signals : A survey. *IEEE Transactions on Affective Computing*, IEEE, PP, n. 99, p. 1–20, 2017. Citado 4 vezes nas páginas 27, 47, 67 e 75.

FRASER, A. M.; SWINNEY, H. L. Independent coordinates for strange attractors from mutual information. *Physical Review A*, American Physical Society, v. 33, p. 1134–1140, Feb 1986. Citado 2 vezes nas páginas 35 e 36.

FRIJDA, N. H. The evolutionary emergence of what we call “emotions”. *Cognition and Emotion*, v. 30, n. 4, p. 609–620, 2016. Citado 3 vezes nas páginas 23, 24 e 45.

GAO, J. et al. *Multiscale Analysis of Complex Time Series: Integration of Chaos and Random Fractal Theory, and Beyond*. [S.l.]: Wiley, 2007. 1–352 p. Citado na página 36.

GARCÍA-MARTÍNEZ, B. et al. Nonlinear Methodologies Applied to Automatic Recognition of Emotions: An EEG Review. *Lecture Notes in Computer Science*, v. 10586, p. 754–765, 2017. Citado 3 vezes nas páginas 31, 102 e 104.

GONZALEZ, R. C.; WOODS, R. E. *Digital image processing*. 4. ed. New York, NY: Pearson, 2018. 1306 p. Citado 4 vezes nas páginas 40, 67, 76 e 81.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. <<http://www.deeplearningbook.org>>. Citado 7 vezes nas páginas 39, 40, 41, 42, 43, 44 e 81.

GUENNEC, A. L.; MALINOWSKI, S.; TAVENARD, R. Data Augmentation for Time Series Classification using Convolutional Neural Networks. In: *ECML/PKDD Workshop on Advanced Analytics and Learning on Temporal Data*. Riva Del Garda, Italy: [s.n.], 2016. Citado 2 vezes nas páginas 72 e 73.

- GUNES, H. et al. Emotion representation, analysis and synthesis in continuous space: A survey. In: *IEEE International Conference on Automatic Face and Gesture Recognition*. Santa Barbara, CA, USA: Ieee, 2011. p. 827–834. ISBN 978-1-4244-9140-7. Citado na página 76.
- HE, H. et al. Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In: *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*. [S.l.: s.n.], 2008. p. 1322–1328. Citado 2 vezes nas páginas 72 e 73.
- HISER, J.; KOENIGS, M. The Multifaceted Role of the Ventromedial Prefrontal Cortex in Emotion, Decision Making, Social Cognition, and Psychopathology. *Biological Psychiatry*, Society of Biological Psychiatry, v. 83, n. 8, p. 638–647, 2018. Citado na página 27.
- JALILIFARD, A.; PIZZOLATO, E. B.; ISLAM, M. K. Emotion classification using single-channel scalp-eeeg recording. In: *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. [S.l.: s.n.], 2016. p. 845–849. Citado na página 29.
- JOHNSON-LAIRD, P. N.; OATLEY, K. The language of emotions: An analysis of a semantic field. *Cognition & Emotion*, v. 3, n. 2, p. 81–123, 04 1989. Citado na página 45.
- KAHNEMAN, D. *Thinking, fast and slow*. New York: Farrar, Straus and Giroux, 2011. Citado na página 104.
- KAWANA, T. et al. Eeg-hat with candle-like microneedle electrode. In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. [S.l.: s.n.], 2019. p. 1111–1114. Citado na página 103.
- KELTNER, D. Toward a consensual taxonomy of emotions. *Cognition and Emotion*, Taylor & Francis, v. 33, n. 1, p. 14–19, 2019. Citado 2 vezes nas páginas 23 e 45.
- KENNEL, M. B.; BROWN, R.; ABARBANEL, H. D. I. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A*, American Physical Society, v. 45, p. 3403–3411, Mar 1992. Citado na página 38.
- KHAN, U. et al. Feature extraction and modeling techniques in speech recognition: A review. In: *2019 4th International Conference on Information Systems Engineering (ICISE)*. Shanghai, China: IEEE, 2019. v. 4, p. 63–67. Citado na página 29.
- KING, G.; ZENG, L. Logistic Regression in Rare Events Data. *Political Analysis*, v. 9, n. 2, p. 137–163, 2001. Citado 4 vezes nas páginas 74, 81, 83 e 101.
- KITCHENHAM, B.; CHARTERS, S. *Guidelines for performing Systematic Literature Reviews in Software Engineering*. [S.l.], 2007. v. 2.3, n. EBSE 2007-001. Citado na página 57.
- KOELSTRA, S. et al. Deap : A database for emotion analysis using physiological signals. *IEEE Transactions on Affective Computing*, v. 3, n. 1, p. 1–15, 2012. Citado 24 vezes nas páginas 23, 25, 27, 30, 50, 51, 52, 53, 54, 56, 62, 65, 66, 67, 71, 72, 75, 76, 77, 82, 84, 90, 101 e 103.

KONAR, A.; CHAKRABORTY, A. *Emotion Recognition - A Pattern Analysis Approach*. Hoboken, New Jersey: John Wiley & Sons, 2015. 1–543 p. (1, September). Citado 8 vezes nas páginas [25](#), [27](#), [28](#), [30](#), [31](#), [60](#), [68](#) e [75](#).

KOSMYNA, N. et al. Attentivu: A wearable pair of eeg and eog glasses for real-time physiological processing. In: *2019 IEEE 16th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. [S.l.: s.n.], 2019. p. 1–4. Citado 3 vezes nas páginas [102](#), [104](#) e [105](#).

KRONE, T. et al. A multivariate statistical model for emotion dynamics. *Emotion*, v. 18, n. 5, p. 739–754, 2018. Citado na página [105](#).

KUMAR, J.; KUMAR, J. Affective modelling of users in hci using eeg. *Procedia Computer Science*, v. 84, p. 107–114, 2016. Proceeding of the Seventh International Conference on Intelligent Human Computer Interaction (IHCI 2015). Citado na página [65](#).

KUPPENS, P.; VERDUYN, P. Looking at emotion regulation through the window of emotion dynamics. *Psychological Inquiry*, Routledge, v. 26, n. 1, p. 72–79, 2015. Citado na página [82](#).

KUPPENS, P.; VERDUYN, P. Emotion dynamics. *Current Opinion in Psychology*, v. 17, p. 22 – 26, 2017. *Emotion*. Citado 5 vezes nas páginas [28](#), [66](#), [73](#), [81](#) e [105](#).

LI, D. et al. The fusion of electroencephalography and facial expression for continuous emotion recognition. *IEEE Access*, IEEE, v. 7, p. 155724–155736, 2019. Citado na página [28](#).

LI, K. et al. Affective state recognition from eeg with deep belief networks. In: *IEEE International Conference on Bioinformatics and Biomedicine*. Shanghai, China: IEEE, 2013. p. 305–310. Citado na página [91](#).

LI, M.; LU, B. L. Emotion classification based on gamma-band eeg. In: *Proc. of Annual International Conference of the IEEE EMBS*. Minneapolis, Minnesota, USA: IEEE, 2009. v. 2009, p. 1323–1326. Citado na página [67](#).

LI, Y. et al. Human emotion recognition with electroencephalographic multidimensional features by hybrid deep neural networks. *Applied Sciences*, v. 7, n. 10, p. 1–20, 2017. Citado 5 vezes nas páginas [58](#), [59](#), [62](#), [65](#) e [66](#).

LINDSEY, E. W. Frequency and Intensity of Emotional Expressiveness and Preschool Children’s Peer Competence. *Journal of Genetic Psychology*, Routledge, v. 180, n. 1, p. 45–61, 2019. Citado na página [29](#).

LORENZ, E. N. Deterministic nonperiodic flow. *Journal of the atmospheric sciences*, v. 20(2), p. 130–141, 1963. Citado na página [33](#).

LOTTE, F. et al. A review of classification algorithms for EEG-based brain-computer interfaces: A 10 year update. *Journal of Neural Engineering*, IOP Publishing, v. 15, n. 3, 2018. Citado 3 vezes nas páginas [60](#), [102](#) e [104](#).

MANFRED, C. *Sentics: The Touch of Emotions*. [S.l.]: Anchor Press, 1977. 1–273 p. (1, 1). Citado 2 vezes nas páginas [48](#) e [49](#).

- MARINUS, J.; ARAÚJO, J.; GOMES, H. Reconstructed phase space and convolutional neural networks for classifying voice pathologies. In: *23rd Iberoamerican Congress on Pattern Recognition, CIARP 2018, Proceedings*. [S.l.: s.n.], 2019. p. 792–802. Citado 7 vezes nas páginas [32](#), [33](#), [35](#), [36](#), [79](#), [87](#) e [95](#).
- MARINUS, J. V. d. M. L. *Classificação de Sinais de Voz Afetada por Patologia nas Pregas Vocais Utilizando Reconstrução do Espaço de Fases*. 154 p. Tese (Doutorado) — Universidade Federal de Campina Grande, 2019. Citado na página [67](#).
- MECARELLI, O. *Clinical Electroencephalography*. Cham, Switzerland: Springer International Publishing Switzerland, 2009. 809 p. Citado na página [30](#).
- MEI, H. et al. Eeg-based emotion classification using convolutional neural network. In: *2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)*. [S.l.: s.n.], 2017. p. 130–135. Citado 4 vezes nas páginas [56](#), [60](#), [62](#) e [66](#).
- MEISELMAN, H. L. *Emotion Measurement*. 1. ed. Cambridge: Elsevier, 2016. v. 1. 1–752 p. Citado 2 vezes nas páginas [26](#) e [27](#).
- MENG, Q.; YAN, J.; XU, H. Research on EEG signal recognition based on channel selection. *Proceedings - 2017 Chinese Automation Congress, CAC 2017*, v. 2017-Janua, p. 6413–6417, 2017. Citado na página [29](#).
- MOON, F. C. *Chaotic and Fractal Dynamics: An Introduction for Applied Scientists and Engineers*. 1. ed. Weinheim, Germany: Wiley-VCH, 2004. v. 3. 527 p. Citado na página [32](#).
- MOON, S. E.; JANG, S.; LEE, J. S. Convolutional Neural Network Approach for EEG-Based Emotion Recognition Using Brain Connectivity and its Spatial Information. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, IEEE, v. 2018-April, p. 2556–2560, 2018. Citado 14 vezes nas páginas [31](#), [39](#), [65](#), [66](#), [69](#), [72](#), [73](#), [76](#), [80](#), [87](#), [90](#), [91](#), [94](#) e [95](#).
- MUNOZ, R. et al. Using Black Hole Algorithm to Improve EEG-Based Emotion Recognition. *Computational Intelligence and Neuroscience*, v. 2018-December, n. 1, p. 1–21, 2018. Citado 2 vezes nas páginas [92](#) e [94](#).
- NORMAN, K. L.; KIRAKOWSKI, J.; WILEY, J. *The Wiley Handbook of Human Computer Interaction Volume 1*. 1. ed. [S.l.]: Wiley, 2018. 995 p. Citado na página [27](#).
- OGINO, M. et al. Analysis of Prefrontal Single-Channel EEG Data for Portable Auditory ERP-Based Brain – Computer Interfaces. *Frontiers in Human Neuroscience*, v. 13, n. July, p. 1–14, 2019. Citado 6 vezes nas páginas [29](#), [30](#), [75](#), [95](#), [102](#) e [104](#).
- OSTWALD, D. et al. Evidence for neural encoding of bayesian surprise in human somatosensation. *NeuroImage*, v. 62, n. 1, p. 177–88, 2012. Citado na página [75](#).
- PANDEY, P.; SEEJA, K. *Emotional State Recognition with EEG Signals Using Subject Independent Approach*. Singapore: Springer, 2019. v. 16. Citado 2 vezes nas páginas [29](#) e [30](#).
- PARROTT, W. G. *Emotions in social psychology : essential readings*. [S.l.]: Philadelphia (Pa.): Psychology Press, 2001. Citado na página [50](#).

- PEREIRA, E. T.; GOMES, H. M. The role of data balancing for emotion classification using eeg signals. In: *IEEE International Conference on Digital Signal Processing (DSP)*. Beijing, China: IEEE, 2016. p. 555–559. Citado na página 62.
- Pereira, E. T. et al. Empirical evidence relating eeg signal duration to emotion classification performance. *IEEE Transactions on Affective Computing*, p. 1–1, 2018. ISSN 2371-9850. Citado 2 vezes nas páginas 82 e 85.
- PETRANTONAKIS, P. C.; HADJILEONTIADIS, L. J. Emotion recognition from eeg using higher order crossings. *IEEE Transactions on information technology in biomedicine*, v. 14, n. 2, p. 186–197, 2010. Citado na página 31.
- PETTA, P.; PELACHAUD, C.; COWIE, R. *Emotion-Oriented Systems: The Humaine Handbook*. 1st. ed. Berlin: Springer Publishing Company, Incorporated, 2011. 794 p. Citado 2 vezes nas páginas 23 e 28.
- PFEIFER, R. The "fungus eater approach" to emotion: A view from artificial intelligence. In: *Cognitive Studies, The Japanese Society for Cognitive Science*. [S.l.: s.n.], 1994. p. 42–57. Citado na página 48.
- PICARD, R. W. *Affective computing*. 1. ed. Cambridge, MA, USA: The MIT Press, 2000. 304 p. Citado 11 vezes nas páginas 23, 25, 29, 45, 46, 47, 48, 49, 50, 76 e 92.
- PIHO, L.; TJAHJADI, T. A mutual information based adaptive windowing of informative eeg for emotion recognition. *IEEE Transactions on Affective Computing*, p. 1–1, 2018. Citado na página 66.
- POINCARÉ, H. *Les méthodes nouvelles de la mécanique céleste*. Paris, França: Gauthier-Villars et fils, Imprimeurs-Libraires, 1893. v. 2. 408 p. Citado 2 vezes nas páginas 32 e 33.
- PROCHAZKOVA, E.; KRET, M. E. Connecting minds and sharing emotions through mimicry: A neurocognitive model of emotional contagion. *Neuroscience Biobehavioral Reviews*, v. 80, p. 99 – 114, 2017. ISSN 0149-7634. Citado na página 28.
- PUCE, A.; HÄMÄLÄINEN, M. S. A review of issues related to data acquisition and analysis in eeg/meg studies. *Brain Sciences*, v. 7, n. 6, 2017. Citado na página 28.
- RAYATDOOST, S.; SOLEYMANI, M. Cross-corpus eeg-based emotion recognition. In: *2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP)*. [S.l.: s.n.], 2018. p. 1–6. Citado 3 vezes nas páginas 103, 104 e 105.
- RIBEIRO, M. T.; SINGH, S.; GUESTRIN, C. "why should I trust you?": Explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*. [S.l.: s.n.], 2016. p. 1135–1144. Citado 2 vezes nas páginas 98 e 102.
- RIM, B. et al. Deep learning in physiological signal data: A survey. *Sensors (Switzerland)*, v. 20, n. 4, 2020. Citado 8 vezes nas páginas 41, 58, 60, 62, 90, 91, 104 e 105.
- ROBINSON, M. D.; MAUSS, I. B. Measures of emotion : A review. *Cognition & Emotion*, Routledge, v. 23, n. 2, p. 209–237, 2009. Citado na página 26.

ROUAST, P. V.; ADAM, M.; CHIONG, R. Deep Learning for Human Affect Recognition: Insights and New Developments. *IEEE Transactions on Affective Computing*, v. 14, n. 8, p. 1–20, 2018. Citado 4 vezes nas páginas 41, 68, 78 e 79.

RUSSELL, J. A.; BARRET, L. F. Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, American Psychological Association, v. 76, n. 5, p. 805–819, 1999. Citado 4 vezes nas páginas 24, 62, 68 e 70.

SALAMON, J.; BELLO, J. P. Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal Processing Letters*, IEEE, v. 24, n. 3, p. 279–283, March 2017. Citado na página 72.

SANTHIYA, P.; CHITRAKALA, S. A survey on emotion recognition from eeg signals: Approaches, techniques challenges. In: *2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN)*. [S.l.: s.n.], 2019. p. 1–6. Citado 2 vezes nas páginas 47 e 67.

SANTOS, M. O. *Emprego do Estimador KSG na Classificação de Patologias Laríngeas por Intermédio da Análise Dinâmica Não-Linear*. 100 p. Tese (Doutorado) — Universidade Federal de Campina Grande, 2019. Citado 4 vezes nas páginas 32, 33, 35 e 36.

SCHACHT, A. et al. Association with positive outcome induces early effects in event-related brain potentials. *Biological psychology*, v. 89, n. 1, p. 130–6, 2012. Citado na página 75.

SCHERER, K. R. What are emotions? and how can they be measured? *Social Science Information*, v. 44, n. 4, p. 695–729, dec 2005. Citado 2 vezes nas páginas 23 e 24.

SCHERER, K. R.; MOORS, A. The Emotion Process: Event Appraisal and Component Differentiation. *Annual Review of Psychology*, v. 70, n. 1, p. 719–745, 2019. Citado 3 vezes nas páginas 23, 24 e 28.

SCHULREICH, S. et al. *Fearless dominance and reduced feedback-related negativity amplitudes in a time-estimation task - Further neuroscientific evidence for dual-process models of psychopathy*. [S.l.]: Elsevier B.V., 2013. 352–363 p. Citado na página 75.

SHIN, J.; MAENG, J.; KIM, D. H. Inner Emotion Recognition Using Multi Bio-Signals. *2018 IEEE International Conference on Consumer Electronics - Asia, ICCE-Asia 2018*, IEEE, p. 206–212, 2018. Citado na página 28.

STEINBERG, C. et al. Rapid and highly resolving associative affective learning: convergent electro- and magnetoencephalographic evidence from vision and audition. *Biological psychology*, v. 92, n. 3, p. 526–40, 2013. Citado 2 vezes nas páginas 75 e 76.

SUN, Z. et al. Data prediction of eeg based on phase space reconstruction and neural network. In: *2018 8th International Conference on Electronics Information and Emergency Communication (ICEIEC)*. [S.l.: s.n.], 2018. p. 162–165. Citado na página 32.

SUZUKI, A. et al. Emotional recognition with wearable eeg device. In: *2019 IEEE 1st Global Conference on Life Sciences and Technologies (LifeTech)*. [S.l.: s.n.], 2019. p. 214–215. Citado 2 vezes nas páginas 102 e 104.

- TAKENS, F. Detecting strange attractors in turbulence. In: _____. *Lecture Notes in Mathematics, Berlin Springer Verlag*. [S.l.]: Springer, 1981. v. 898, p. 366. Citado na página 35.
- TARAN, S.; BAJAJ, V. Emotion recognition from single-channel EEG signals using a two-stage correlation and instantaneous frequency-based filtering method. *Computer Methods and Programs in Biomedicine*, Elsevier B.V., v. 173, p. 157–165, 2019. Citado 6 vezes nas páginas 67, 75, 83, 92, 94 e 102.
- TARNOWSKI, P. et al. Combined analysis of GSR and EEG signals for emotion recognition. *2018 International Interdisciplinary PhD Workshop, IIPhDW 2018*, p. 137–141, 2018. Citado 11 vezes nas páginas 65, 66, 68, 69, 72, 73, 80, 90, 91, 94 e 95.
- TEO, J.; CHIA, J. T. Deep Neural Classifiers for Eeg-Based Emotion Recognition in Immersive Environments. *2018 International Conference on Smart Computing and Electronic Enterprise, ICSCCE 2018*, IEEE, p. 1–6, 2018. Citado 4 vezes nas páginas 69, 80, 90 e 94.
- TIBOR, S. R. et al. Deep learning with convolutional neural networks for eeg decoding and visualization. *Human Brain Mapping*, v. 38, n. 11, p. 5391–5420, 2017. Citado na página 67.
- TONG, L.; ZHAO, J.; FU, W. Emotion Recognition and Channel Selection Based on EEG Signal. *Proceedings - 11th International Conference on Intelligent Computation Technology and Automation, ICICTA 2018*, p. 101–105, 2018. Citado 4 vezes nas páginas 29, 31, 67 e 75.
- TORRES, C.; OROZCO, A.; ALVAREZ, M. Feature selection for multimodal emotion recognition in the arousal-valence space. In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. [S.l.]: IEEE, 2013. (35), p. 4330–4333. Citado na página 65.
- TRIPATHI, S. et al. Using deep and convolutional neural networks for accurate emotion classification on deap dataset. In: *Innovative Applications of Artificial Intelligence*. [S.l.: s.n.], 2017. p. 4746–4752. Citado 9 vezes nas páginas 39, 59, 60, 62, 68, 69, 80, 90 e 91.
- VERDUYN, P. et al. Determinants of emotion duration and underlying psychological and neural mechanisms. *Emotion Review*, v. 7, n. 4, p. 330–335, 2015. Citado na página 50.
- VERMA, G. K.; TIWARY, U. S. Multimodal fusion framework: A multiresolution approach for emotion classification and recognition from physiological signals. *NeuroImage*, Elsevier Inc., v. 102, p. 162–172, 2014. Citado na página 65.
- WANG, J.; XU, W. Research of video advertisements effect based on eeg: Ersp and emotion for commercial effect. In: *2016 13th International Conference on Service Systems and Service Management (ICSSSM)*. Kunming, China: IEEE, 2016. p. 0–4. Citado na página 67.
- WANG, K. Y. et al. Design of Intelligent EEG System for Human Emotion Recognition with Convolutional Neural Network. *Proceedings 2019 IEEE International Conference on Artificial Intelligence Circuits and Systems, AICAS 2019*, IEEE, p. 142–145, 2019. Citado 2 vezes nas páginas 76 e 84.

- WANG, X.-W.; NIE, D.; LU, B.-L. Emotional state classification from eeg data using machine learning approach. *Neurocomputing*, Elsevier, v. 129, p. 94–106, 2014. Citado na página 67.
- WANG, Y. et al. EmotioNet: A 3-D Convolutional Neural Network for EEG-based Emotion Recognition. *Proceedings of the International Joint Conference on Neural Networks*, IEEE, v. 2018-July, p. 1–7, 2018. Citado 14 vezes nas páginas 39, 54, 66, 68, 69, 72, 73, 76, 80, 90, 91, 92, 94 e 95.
- WANG, Z.; TONG, Y.; HENG, X. Phase-Locking Value Based Graph Convolutional Neural Networks for Emotion Recognition. *IEEE Access*, v. 7, p. 93711–93722, 2019. Citado na página 76.
- WANG, Z. M.; HU, S. Y.; SONG, H. Channel Selection Method for EEG Emotion Recognition Using Normalized Mutual Information. *IEEE Access*, v. 7, p. 143303–143311, 2019. Citado 5 vezes nas páginas 29, 66, 67, 75 e 84.
- WATT, R. C.; HAMEROFF, S. R. Phase space electroencephalography (EEG): A new mode of intraoperative EEG analysis. *International Journal of Clinical Monitoring and Computing*, v. 5, n. 1, p. 3–13, 1988. Citado 3 vezes nas páginas 32, 35 e 36.
- WEN, Z.; XU, R.; DU, J. A novel convolutional neural networks for emotion recognition based on eeg signal. In: *2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)*. [S.l.: s.n.], 2017. p. 672–677. Citado 4 vezes nas páginas 59, 62, 65 e 66.
- WHITTLE, S.; SIMMONS, J. G.; ALLEN, N. B. *Emotion and Gender-Specific Neural Processing in Men and Women*. [S.l.]: Elsevier Inc., 2017. 183–201 p. Citado na página 29.
- WU, S. et al. Estimation of valence of emotion using two frontal EEG channels. *Proceedings - 2017 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2017*, v. 2017-Janua, p. 1127–1130, 2017. Citado 2 vezes nas páginas 67 e 75.
- XIN, L. et al. Relevance vector machine based eeg emotion recognition. In: *2016 Sixth International Conference on Instrumentation Measurement, Computer, Communication and Control (IMCCC)*. [S.l.: s.n.], 2016. p. 293–297. Citado 2 vezes nas páginas 56 e 66.
- XU, T. et al. Emotion Recognition Using Frontal EEG in VR Affective Scenes. *IEEE MTT-S 2019 International Microwave Biomedical Conference, IMBioC 2019 - Proceedings*, IEEE, v. 1, p. 1–4, 2019. Citado 3 vezes nas páginas 29, 67 e 84.
- YANG, B.; HAN, X.; TANG, J. Three class emotions recognition based on deep learning using staked autoencoder. In: *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. [S.l.: s.n.], 2017. p. 1–5. Citado 2 vezes nas páginas 59 e 62.
- YANG, Y. et al. Emotion Recognition from Multi-Channel EEG through Parallel Convolutional Recurrent Neural Network. *Proceedings of the International Joint Conference on Neural Networks*, IEEE, July, p. 1–7, 2018. Citado 10 vezes nas páginas 43, 54, 56, 66, 69, 81, 82, 90, 91 e 92.

YIN, Z. et al. Cross-subject EEG feature selection for emotion recognition using transfer recursive feature elimination. *Frontiers in Neurorobotics*, v. 11, n. APR, p. 1–16, 2017. Citado na página 68.

YU, M. et al. A review of EEG features for emotion recognition. *SCIENTIA SINICA Informationis*, v. 49, n. 9, p. 1097–1118, 2019. Citado 4 vezes nas páginas 31, 77, 102 e 104.

ZANGENEH, S. M. et al. A novel EEG-based approach to classify emotions through phase space dynamics. *Signal, Image and Video Processing*, Springer, Cham, v. 13, p. 1149–1156, 2019. Citado 5 vezes nas páginas 31, 58, 92, 94 e 101.

ZANGENEH, S. M. et al. Emotion recognition through EEG phase space dynamics and Dempster-Shafer theory. *Medical Hypotheses*, Elsevier, v. 127, n. March, p. 34–45, 2019. Citado 5 vezes nas páginas 31, 32, 61, 62 e 63.

ZANGENEH SOROUSH, M. et al. Emotion Classification through Nonlinear EEG Analysis Using Machine Learning Methods. *International Clinical Neuroscience Journal*, v. 5, n. 4, p. 135–149, 2018. Citado 4 vezes nas páginas 58, 62, 63 e 104.

ZHANG, W. et al. Radar signal recognition based on tspot and lime. In: *2018 37th Chinese Control Conference (CCC)*. [S.l.: s.n.], 2018. p. 4158–4163. Citado 2 vezes nas páginas 98 e 102.

ZHENG, A. *Evaluating Machine Learning Models*. First edit. [S.l.]: O’Reilly Media, Inc, 2015. 1–58 p. Citado 4 vezes nas páginas 69, 70, 71 e 80.

ZHENG, W. L. et al. EmotionMeter: A Multimodal Framework for Recognizing Human Emotions. *IEEE Transactions on Cybernetics*, v. 49, n. 3, p. 1110–1122, 2019. Citado 4 vezes nas páginas 28, 92, 93 e 94.

ZHENG, W.-L.; LU, B.-L. Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks. *IEEE Transactions on Autonomous Mental Development*, v. 7, n. 3, p. 162–175, Sept 2015. Citado 2 vezes nas páginas 75 e 92.

ZHENG, W.-L.; ZHU, J.-Y.; LU, B.-L. Multimodel emotion analysis in response to multimedia. In: *IEEE International Conference on Multimedia and Expo Workshops*. Chengdu: IEEE, 2014. p. 1–2. Citado na página 65.