



**UNIVERSIDADE FEDERAL DO VALE DO SÃO FRANCISCO
PROGRAMA DE PÓS-GRADUAÇÃO CIÊNCIAS DA SAÚDE E
BIOLÓGICAS**

ANA BEATRIZ RODRIGUES MASSARANDUBA

**AVALIAÇÃO DA LATERALIZAÇÃO CORTICAL PARA
IDENTIFICAÇÃO DE PACIENTES COM DOENÇA DE PARKINSON
UTILIZANDO SINAIS DE ELETROENCEFALOGRAMA**

**PETROLINA-PE
2023**

ANA BEATRIZ RODRIGUES MASSARANDUBA

**AVALIAÇÃO DA LATERALIZAÇÃO CORTICAL PARA
IDENTIFICAÇÃO DE PACIENTES COM DOENÇA DE PARKINSON
UTILIZANDO SINAIS DE ELETROENCEFALOGRAMA**

Trabalho apresentado à Universidade Federal do Vale do São Francisco - UNIVASF, Campus Petrolina, como requisito da obtenção do grau de mestre em Ciências da Saúde e Biológicas.

Orientador: Prof. Dr. Rodrigo Pereira Ramos
Coorientadora: Prof^a. Dra. Ivani Brys

**PETROLINA-PE
2023**

Massaranduba, Ana Beatriz Rodrigues
M414a Avaliação da lateralização cortical para identificação de
pacientes com doença de Parkinson utilizando sinais de
eletroencefalograma / Ana Beatriz Rodrigues Massaranduba. –
Petrolina - PE, 2023.
xiv, 65 f. : il. ; 29 cm.

Dissertação de mestrado (Mestrado em Ciências da Saúde e
Biológicas) - Universidade Federal do Vale do São Francisco,
Campus Petrolina - PE, 2023.

Orientador: Prof. Dr. Rodrigo Pereira Ramos.
Coorientadora: Profa. Dra. Ivani Brys.

Inclui referências.

1. Doença de Parkinson. 2. Biomarcadores. 3. EEG. 4.
Lateralização. 5. Diagnóstico. I. Título. II. Ramos, Rodrigo Pereira. III.
Universidade Federal do Vale do São Francisco.

CDD 616.833

**UNIVERSIDADE FEDERAL DO VALE DO SÃO FRANCISCO
PÓS-GRADUAÇÃO CIÊNCIAS DA SAÚDE E BIOLÓGICAS**

FOLHA DE APROVAÇÃO

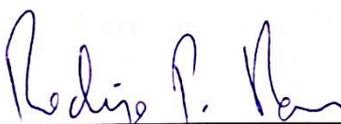
ANA BEATRIZ RODRIGUES MASSARANDUBA

**AVALIAÇÃO DA LATERALIZAÇÃO CORTICAL PARA IDENTIFICAÇÃO DE
PACIENTES COM DOENÇA DE PARKINSON UTILIZANDO SINAIS DE
ELETROENCEFALOGRAMA**

Dissertação apresentada como requisito para obtenção do título de Mestre em Ciências com ênfase na linha de pesquisa: Fundamentação Conceitual e Metodologias Inovadoras Integradoras em Ambiente, Tecnologia e Saúde, pela Universidade Federal do Vale do São Francisco.

Aprovada em: 16 de fevereiro de 2023

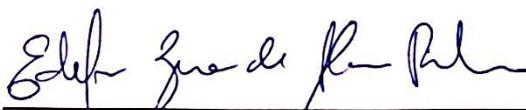
Banca Examinadora



Rodrigo Pereira Ramos, Doutor
Universidade Federal do Vale do São Francisco – Univasf



Rosalvo Ferreira de Oliveira Neto, Doutor
Universidade Federal do Vale do São Francisco – Univasf



Edilson Beserra de Alencar Filho, Doutor
Universidade Federal do Vale do São Francisco – Univasf

DEDICATÓRIA

Dedico este trabalho a todas as pessoas que amo.

AGRADECIMENTOS

Início agradecendo a Deus pela vida da minha família e por ter me dado saúde física e mental para realizar grandes coisas ao longo de uma pandemia.

Agradeço aos meus pais, que me dão apoio, amor e forças todos os dias para seguir lutando. Agradeço também à minha irmã Mariana pela amizade, companheirismo, amor e por me motivar sempre.

Agradeço aos meus avós pelo amor, carinho, cuidado e orgulho.

Agradeço aos meus tios(as) por cuidarem de mim como se fossem meus pais, e por sempre me darem apoio e carinho. Quero agradecer especialmente a minha tia Elisabeth por sempre me incentivar, estimular e se preocupar tanto com a minha educação, o tempo todo cuidando para que eu tivesse tudo que fosse necessário.

Agradeço aos meus primos(as), que são como irmãos para mim, por sempre serem tão amorosos e companheiros. Em especial à minha prima Juliana, colega de formação, por todos os conselhos, pelo incentivo e parceria.

Agradeço a Thiago, meu namorado, por tanto apoio, cuidado, incentivo, amor e companheirismo.

Quero agradecer às minhas amigas Anna Luisa, Bianca, Isabella, Sarah e Vitória por estarem comigo há tantos anos, sempre me apoiando e me incentivando. Agradeço também aos meus amigos do mestrado pelo companheirismo, em especial ao meu amigo Bruno, que sempre esteve presente, desde a graduação, me dando todo apoio.

Agradeço ao professor Rodrigo Ramos, que é meu orientador há 5 anos e se tornou família, sempre ensinando, apoiando e inspirando. Agradeço à minha coorientadora, professora Ivani Brys, pela confiança, apoio e por ser inspiração.

Agradeço à Universidade Federal do Vale do São Francisco, ao Programa de Pós-Graduação Ciências da Saúde e Biológicas e a todos os professores, técnicos e funcionários que contribuíram para a minha formação.

Agradeço também aos participantes da banca examinadora.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

"O otimismo é a cura para muitas coisas."

Michael J. Fox

RESUMO

A doença de Parkinson (DP) é uma das doenças neurodegenerativas mais comuns que afetam a população de meia-idade e idosa em todo o mundo. A DP ocorre devido à morte dos neurônios dopaminérgicos na via nigroestriatal, causando uma deficiência de dopamina, o neurotransmissor responsável pelo controle dos movimentos do corpo. A doença de Parkinson traz sintomas de origem motora, como tremores, perda de equilíbrio e bradicinesia, e de origem não motora, como problemas no sono e distúrbios sensoriais. Entretanto, o diagnóstico depende do aparecimento dos sintomas motores, que ocorrem quando a doença já está em um estágio avançado. Diante disso, diversos trabalhos vêm sendo desenvolvidos com o intuito de obter biomarcadores capazes de possibilitar o diagnóstico da DP através de algoritmos de inteligência artificial. Nesse contexto, sinais de eletroencefalograma (EEG) podem ser utilizados na busca de características que possam auxiliar na detecção da doença. Para isso, é necessário analisar diferentes métodos de caracterização e classificação desses sinais. Além disso, sabe-se que pacientes com a DP possuem um lado do corpo mais afetado pelos sintomas motores, o que significa que o hemisfério oposto do cérebro é mais afetado pela doença. Assim, este projeto visou investigar o uso da lateralização cortical no desenvolvimento de um sistema para auxílio do diagnóstico da doença de Parkinson por meio de sinais de EEG. Foi utilizada uma base de dados com sinais de EEG coletados enquanto os participantes, com e sem diagnóstico de Parkinson, permaneciam em repouso ou realizavam a tarefa do *Oddball Paradigm*. O trabalho foi dividido em duas partes: na primeira, foi utilizada a diferença entre os canais opostos Fp1-Fp2, F3-F4, C3-C4 e P3-P4 para analisar o desempenho do estudo da lateralização da doença na busca por biomarcadores. Na segunda parte, foram utilizados os 60 canais e diferentes métodos de lateralização, como a subtração entre os pares de canais opostos, o hemisfério do cérebro mais afetado, entre outras. Foram utilizadas técnicas de análise estatística e espectral, com separação de frequências alfa, beta e gama para extrair características dos sinais que pudessem servir como biomarcadores. O desempenho dessas características foi investigado por meio de ferramentas de aprendizado de máquina, como máquinas de vetores de suporte (SVM, do inglês Support-vector Machine), k vizinhos mais próximos (k NN, k-nearest Neighbor) e florestas aleatórias (RF, Random Forest). Na primeira parte do trabalho, foi obtida uma acurácia de 74% para a frequência alfa de Fp1-Fp2. Já na segunda etapa, foram obtidas acurácias de mais de 70 %, sendo o melhor classificador para esses dados o RF. Foi observado que a lateralização cortical melhora o desempenho do modelo, sendo a técnica mais eficiente a subtração de canais de hemisférios opostos.

Palavras-chave: Parkinson. Biomarcadores. EEG. Lateralização. Diagnóstico. Aprendizado de Máquina.

ABSTRACT

Parkinson's disease (PD) is one of the most common neurodegenerative diseases affecting middle-aged and elderly people worldwide. PD occurs due to the death of dopaminergic neurons, causing a deficiency of dopamine, the neurotransmitter responsible for controlling the body's movements. Parkinson's disease brings on symptoms of motor origin, such as tremors, loss of balance and bradykinesia, and of non-motor origin, such as sleep problems and sensory disturbances. However, the diagnosis depends on the onset of motor symptoms, which occur when the disease is already in an advanced stage. Therefore, several studies have been developed with the purpose of obtaining biomarkers capable of enabling the diagnosis of PD through artificial intelligence algorithms. In this context, electroencephalogram (EEG) signals can be used in the search for characteristics that can help detect the disease. For this, it is necessary to analyze different methods of characterization and classification of these signals. In addition, it is known that patients with PD have one side of the body more affected by motor symptoms, which means that the opposite side of the brain is more affected by the disease. Thus, this project aimed to investigate the use of cortical lateralization in the development of a system to aid the diagnosis of Parkinson's disease using EEG signals. A database was used with EEG signals collected while participants, with and without Parkinson's diagnosis, remained at rest or performed the *Oddball Paradigm* task. The work was divided into two parts: in the first part, the difference between the opposite channels Fp1-Fp2, F3-F4, C3-C4 and P3-P4 was used to analyze the performance of the lateralization study of the disease in the search for biomarkers. In the second part, the 60 channels and different lateralization methods were used, such as subtraction between pairs of opposing channels, the hemisphere of the brain most affected, and others. Statistical and spectral analysis techniques were used, with separation of alpha, beta and gamma frequencies to extract features from the signals that could serve as biomarkers. The performance of these features was investigated using machine learning tools, such as support-vector machines (SVM), k-nearest neighbors (*k*NN) and random forests (RF). In the first part of the work, an accuracy of 74% was obtained for the alpha frequency of Fp1-Fp2. In the second stage, accuracies of more than 70% were obtained, and the best classifier for these data was the RF. It was observed that cortical lateralization improves model performance, and the most efficient technique is the subtraction of channels from opposite hemispheres.

Keywords: Parkinson. Biomarkers. EEG. Lateralization. Diagnosis. Machine Learning.

LISTA DE FIGURAS

Figura 1	O sistema internacional 10/20.	23
Figura 2	Sistema com 64 eletrodos.	24
Figura 3	Exemplos de imagens obtidas pelo HOS	29
Figura 4	Árvore de decisão.	30
Figura 5	Etapas da metodologia do projeto	34
Figura 6	Etapas do pré-processamento dos sinais	37
Figura 7	Passos da etapa de caracterização dos sinais	40
Figura 8	Matriz de confusão para a frequência beta dos canais C3-C4 no re- posso.	45
Figura 9	Matriz de confusão para a frequência beta dos canais F3-F4 na tarefa do Oddball para CTRL x ON.	48
Figura 10	Matriz de confusão para a frequência alfa dos canais Fp1-Fp2 na tarefa do Oddball para CTRL x OFF.	48
Figura 11	Matriz de confusão para a frequência gama CTRL x ON e RF consi- derando todos os canais	52
Figura 12	Matriz de confusão para a frequência gama CTRL x ON e RF consi- derando a subtração dos canais	52
Figura 13	Matriz de confusão para a frequência alfa CTRL x ON e SVM consi- derando todos os canais	55
Figura 14	Matriz de confusão para a frequência alfa CTRL x ON e SVM consi- derando a subtração dos canais	56

LISTA DE TABELAS

Tabela 1	Matriz de Confusão	32
Tabela 2	Informações dos pacientes com DP (ON e OFF) e CTRL (média +/- desvio padrão)	35
Tabela 3	Etapas de cada um dos 4 métodos de extração de características na ordem cronológica.	39
Tabela 4	Aprendizado de máquina para CTRL x ON e para a primeira abordagem da análise espectral do repouso.	43
Tabela 5	Aprendizado de máquina para CTRL x OFF e para a primeira abordagem da análise espectral do repouso.	44
Tabela 6	Aprendizado de máquina para CTRL x ON e para a análise espectral do Oddball Paradigm.	46
Tabela 7	Aprendizado de máquina para CTRL x OFF e para a análise espectral do Oddball Paradigm.	47
Tabela 8	Acurácias (%) e melhores classificadores para todos os canais, subtração de canais opostos e hemisfério mais afetado do cérebro.	49
Tabela 9	Acurácias (%) e melhores classificadores para o hemisfério menos afetado do cérebro, hemisfério direito e hemisfério esquerdo.	50
Tabela 10	Acurácias (%) do bootstrapping para a análise do Oddball Paradigm.	51
Tabela 11	bootstrapping para a análise espectral do Oddball Paradigm.	51
Tabela 12	P-valor para todos os canais x subtração entre canais opostos.	53
Tabela 13	Acurácias (%) e melhores classificadores para os coeficientes da bi-coerência como características.	54
Tabela 14	Acurácias (%) e melhores classificadores para os coeficientes da bi-coerência como características.	54
Tabela 15	P-valor para todos os canais x subtração entre canais opostos utilizando o HOS.	57
Tabela 16	P-valor para as acurácias de todos os canais utilizando medidas estatísticas x todos os canais utilizando HOS.	57
Tabela 17	P-valor para as acurácias da subtração de canais opostos utilizando medidas estatísticas x subtração de canais opostos utilizando HOS.	58

LISTA DE ABREVIATURAS E SIGLAS

kNN	K-nearest neighbor
ACC	Acurácia
CNN	Convolutional Neural Network
CTRL	Controle
DA	Discriminant Analysis
DP	Doença de Parkinson
DT	Decision Tree
EEG	Eletroencefalografia
EN	Ensemble Bagged Trees
FFT	Fast Fourier Transform
fMRI	Functional Magnetic Resonance
FN	False Negative
FP	False Positive
HOC	Higher Order Crossings
HOS	Higher Order Spectra
IA	Inteligência Artificial
ICA	Independent Component Analysis
LEDD	L-Dopa equivalent daily dose
LR	Logistic Regression
MEG	Magnetoencefalografia
OMS	Organização Mundial de Saúde

RF	Random Forest
SVM	Support Vector Machine
TN	True Negative
TP	True Positive

SUMÁRIO

1	INTRODUÇÃO	16
2	OBJETIVOS	20
2.1	OBJETIVO GERAL	20
2.2	OBJETIVOS ESPECÍFICOS	20
3	REFERENCIAL TEÓRICO	21
3.1	A DOENÇA DE PARKINSON	21
3.2	SINAIS DE EEG E SEU PROCESSAMENTO	22
3.3	TÉCNICAS DE ANÁLISE E PROCESSAMENTO DE SINAIS	25
3.3.1	Análise Temporal e Estatística	25
3.3.2	Análise Espectral	27
3.4	APRENDIZADO DE MÁQUINA	29
3.5	ESTIMATIVA DO MODELO	31
3.6	ESTUDOS CORRELATOS	32
4	MÉTODO	34
4.1	AQUISIÇÃO DOS DADOS	34
4.2	PRÉ-PROCESSAMENTO DOS SINAIS	36
4.3	ETAPA DE CARACTERIZAÇÃO	37
4.3.1	Utilizando quatro pares de canais	37
4.3.1.1	Repouso	38
4.3.1.2	Tarefa do Oddball	39
4.3.2	Utilizando todos os canais	39
4.4	ETAPA DE CLASSIFICAÇÃO	40
4.4.1	Quatro pares de canais	40
4.4.2	Todos os canais	41
5	RESULTADOS E DISCUSSÃO	42
5.1	UTILIZANDO QUATRO PARES DE CANAIS	42
5.1.1	Repouso	42
5.1.2	Oddball Paradigm	45

5.2	TODOS OS CANAIS	49
6	CONCLUSÃO E TRABALHOS FUTUROS	60
	REFERÊNCIAS	62

1 INTRODUÇÃO

A doença de Parkinson (DP) é a segunda doença neurodegenerativa mais comum no mundo, ficando atrás apenas da doença de Alzheimer, afeta de 2 a 3% da população acima de 65 anos de idade, e não possui cura conhecida (1). Essa doença é caracterizada pela morte dos neurônios dopaminérgicos localizados na região cerebral conhecida como substância nigra, causando deficiência da dopamina, neurotransmissor responsável por controlar os movimentos do corpo. Os principais sintomas da doença de Parkinson são rigidez muscular, postura instável, tremor, perda de equilíbrio, bradicinesia e habilidade motora fina danificada (2).

Embora a DP esteja também associada a muitos sintomas não motores que contribuem para a incapacidade geral, o diagnóstico clínico dessa doença depende das alterações nas características motoras do paciente, que, na maioria dos casos, ocorre quando o indivíduo já perdeu mais de 60% dos neurônios dopaminérgicos (3). Dessa forma, nos estágios iniciais da doença, enquanto os sintomas motores não podem ser observados, o diagnóstico da DP fica inviabilizado, impossibilitando o tratamento precoce capaz de retardar a progressão da doença (4). Além disso, não existem estratégias preventivas para a doença, então, a sua identificação precoce é fundamental para que esses tratamentos possam ser estudados e desenvolvidos, além de melhorar o prognóstico do paciente.

Algoritmos computacionais para aprendizado de máquina vêm sendo aplicados em diversas áreas de pesquisa nos últimos anos e vários trabalhos no âmbito da engenharia são desenvolvidos atualmente em busca de se obter biomarcadores capazes de realizar o diagnóstico da doença de Parkinson (5). Dentro desse contexto, parâmetros extraídos de gravações de voz, do modo de caminhar e de sinais de eletroencefalograma (EEG) são comumente usados como biomarcadores, em conjunto com algoritmos de inteligência artificial (IA), na busca pela detecção automática de Parkinson. O primeiro desses parâmetros é utilizado devido aos distúrbios vocais causados pela doença, já que pacientes com DP possuem dificuldades na pronúncia de sons vocálicos, causando marcações na fala (6). O modo de caminhar é analisado como um biomarcador para DP por causa das limitações motoras acarretadas pela degeneração neuronal, pois indivíduos portadores de DP possuem uma forma de caminhar característica (7).

Já os sinais de EEG possuem uma natureza elétrica, que está associada à atividade cerebral, e são capazes de refletir o estado das camadas corticais e estruturas subcorticais correspondentes, áreas que são afetadas pela DP. Além disso, esses sinais são frequentemente utilizados na literatura para diagnóstico de doenças neurológicas como epilepsia, esquizofrenia e Alzheimer (8). A sua captação é feita por meio de eletrodos posicionados no couro cabeludo (9), que permitem observar o comportamento das redes neurais biológicas associadas às disfunções cognitivas e complicações motoras presentes em indivíduos com DP (10).

Os sinais de eletroencefalograma são coletados enquanto os pacientes se encontram em um estado de repouso ou enquanto estão desenvolvendo alguma tarefa, que tem como objetivo induzir diferentes respostas cerebrais. Dentre essas tarefas, destaca-se o Oddball Paradigm, que consiste em apresentar aos voluntários, periodicamente, dois tipos de estímulos de natureza visual ou auditiva: estímulos padrão e estímulos alvo. Na presença de um estímulo padrão, nada deve ser feito. Porém, quando há um estímulo alvo, o participante deve computar esse estímulo, seja mentalmente ou por meio de uma ferramenta auxiliar. Ao final do processo, o participante informa a contagem final de estímulos alvo. Esse experimento pode ser utilizado para avaliar a cognição e a atenção (11).

Pacientes diagnosticados com a doença de Parkinson possuem um lado do corpo mais afetado pelos sintomas motores, que corresponde ao hemisfério oposto do cérebro mais afetado pela morte dos neurônios dopaminérgicos (12). Por exemplo, se o indivíduo apresenta mais tremores do lado direito do corpo, o hemisfério esquerdo do cérebro é mais afetado pela doença. Dessa forma, espera-se que haja mais diferenciação entre os sinais de EEG de hemisférios opostos em pessoas com a DP, do que em indivíduos de controle. Por isso, o estudo da lateralização da doença no cérebro pode ser de grande valia no seu processo de diagnóstico.

Vários domínios de análise têm sido usados na extração de características de sinais de EEG, incluindo tempo, frequência e análise estatística (13). Se o sinal é não estacionário, métodos de mudança do domínio temporal para o domínio espectral podem trazer informações adicionais considerando mudanças dinâmicas (14). Além disso, sinais de EEG contêm informações relevantes em várias bandas de frequência, como alfa (8 a 13 Hz), beta (14 a 30 Hz) e gama (acima de 30 Hz) (15). Desse modo, a

caracterização espectral do sinal é fundamental para extrair informações específicas, que não são visíveis quando o sinal está representado no domínio do tempo. Assim, a decomposição do sinal em frequências, obtida por meio do uso da transformada de Fourier, pode ter um bom desempenho na busca de biomarcadores em sinais de EEG (16). Além disso, uma ferramenta de análise espectral comumente utilizada na literatura para caracterização de sinais de EEG é o *Higher Order Spectra* (HOS), que é a representação das cumulantes de ordem superior de um processo aleatório no domínio da frequência (17).

De posse das características dos sinais, é necessário utilizar ferramentas capazes de realizar a sua classificação. Mais recentemente, algoritmos de inteligência artificial vêm sendo empregados para avaliação de padrões e auxílio ao diagnóstico da DP, como as máquinas de vetores de suporte (SVM, do inglês *support-vector machine*), os k vizinhos mais próximos (k NN, *k-nearest neighbor*), entre outros (18) (19).

Assim, este trabalho buscou inicialmente extrair características estatísticas da diferença de canais opostos do cérebro no domínio da frequência. Para isso, foi utilizada a transformada de Fourier e as frequências alfa, beta e gama foram analisadas separadamente. Em seguida, foram empregados classificadores de aprendizado de máquina, como SVM e k NN, para identificar o desempenho dessas características como biomarcadores da DP. Posteriormente, essas características estatísticas foram utilizadas para comparar o desempenho de diferentes formas de lateralização cortical, como a subtração entre canais opostos, canais referentes apenas ao hemisfério mais afetado do cérebro do indivíduo, entre outras, por meio da utilização de classificadores de aprendizado de máquina. Por último, foram empregadas como características as frequências do HOS a fim de visualizar se essa técnica mais complexa traria melhores resultados na diferenciação de indivíduos com e sem Parkinson, também fazendo uso da assimetria do cérebro.

Neste trabalho, foi utilizada a base de dados disponibilizada por Cavanagh *et al.* (2017, p. 27) na plataforma aberta e gratuita *Openneuro*. Ela contém sinais de EEG coletados durante a execução do teste do Oddball Paradigm e com os pacientes em repouso (20).

A seguir, serão apresentados os Capítulos 2 - 6 que compõem esta dissertação e estão organizados da seguinte forma: o Capítulo 2 mostra os objetivos geral e es-

pecíficos do trabalho. No Capítulo 3, constam alguns conceitos necessários para o entendimento do modelo que foi desenvolvido. Já o Capítulo 4 diz respeito à metodologia utilizada, que contém as bases de dados, o pré-processamento dos sinais, as abordagens usadas na etapa de caracterização dos sinais de EEG e, por fim, a etapa de classificação. Os resultados e discussão encontram-se no Capítulo 5 e a conclusão do trabalho bem como as sugestões para trabalhos futuros podem ser vistas no Capítulo 6.

2 OBJETIVOS

2.1 OBJETIVO GERAL

Investigar a lateralização cortical para detecção da doença de Parkinson utilizando como características biomarcadores obtidos de sinais de EEG e usando a aprendizagem de máquina para classificação.

2.2 OBJETIVOS ESPECÍFICOS

- Verificar se a lateralização cortical melhora a diferenciação entre indivíduos com e sem Parkinson;
- Comparar diferentes formas de lateralização cortical, como subtração entre hemisférios, utilização apenas do hemisfério mais afetado, entre outras;
- Comparar o HOS (*Higher Order Spectra*) com medidas estatísticas simples como características dos sinais;
- Avaliar o desempenho dos classificadores empregados por meio das métricas matrizes de confusão e acurácias e verificar que tipo de classificador apresenta melhores resultados para os biomarcadores utilizados.

3 REFERENCIAL TEÓRICO

3.1 A DOENÇA DE PARKINSON

No nascimento, o cérebro humano contém o número máximo de neurônios que poderá conter ao longo da vida, pois, ao contrário de outras células do nosso corpo, como as da pele ou ossos, as células nervosas não podem se autorreparar. Portanto, com o passar do tempo, os neurônios morrem devido ao envelhecimento e não são substituídos. A doença de Parkinson (DP) é uma condição neurológica avançada, de processo lento, que destrói os neurônios dopaminérgicos na substância nigra do cérebro, afetando as suas vias de comunicação. A DP geralmente atinge pessoas acima de 50 anos e, até hoje, acometeu entre 7 e 10 milhões de pessoas em todo o mundo, de acordo com as estatísticas do último registro realizado pela Organização Mundial de Saúde (OMS) até a publicação do presente trabalho (21).

A causa da maioria dos casos da doença de Parkinson não é conhecida até hoje. No entanto, existem pesquisas que indicam que, em raros casos, a combinação entre fatores ambientais e genéticos desempenha um papel importante como causa de DP. Para entendimento geral, a DP é tratada como um distúrbio do sistema nervoso, que é resultado da perda de células de várias partes do cérebro. Dentre essas células, as mais afetadas são os neurônios responsáveis por produzir dopamina. A dopamina é o principal neurotransmissor modulatório do estriado e de outros núcleos da base, que desempenha um papel vital na coordenação do movimento. Então, em decorrência da perda dessas células, os pacientes sofrem de distúrbios motores (22).

Os sintomas primários podem ser classificados em sintomas motores e sintomas não motores. Os sintomas motores manifestam-se como tremor, bradicinesia, que é a lentificação dos movimentos voluntários, rigidez muscular e marcha anormal, enquanto os sintomas não motores apresentam-se como transtornos mentais, problemas de sono, disfunção autonômica e distúrbio sensorial (23).

O diagnóstico da DP é baseado no histórico do paciente e em exame físico. Também podem ser usadas tomografias computadorizadas para complementação. Entretanto, a observação dos sintomas motores é o que determina o diagnóstico clínico da doença (24).

A DP tem como principal tratamento a reposição dos níveis de dopamina no sis-

tema nervoso central. Para isso, é utilizado o fármaco levodopa (também chamada de L-Dopa), abreviação de ácido (S)-2-amino-3-(3,4-dihidroxifenil) propanóico, sintetizado pela primeira vez em 1910 e utilizado para tratar a doença de Parkinson desde o início dos anos 1960. A levodopa é uma substância que, quando chega ao sistema nervoso central, é convertida em dopamina pelos neurônios dopaminérgicos remanescentes (25).

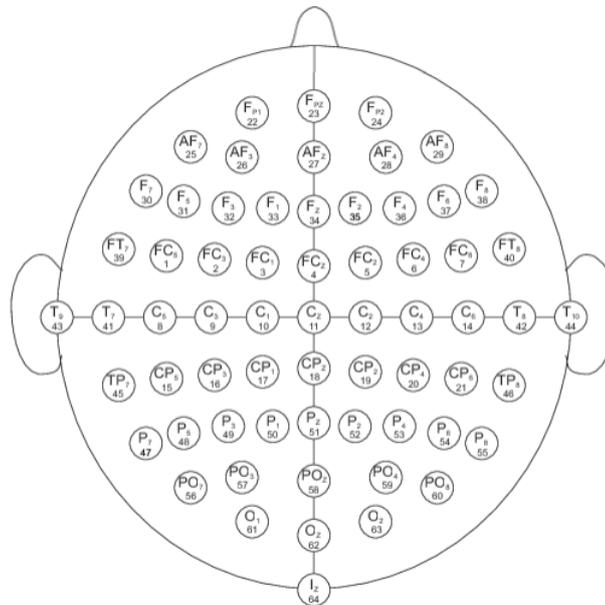
A levodopa funciona bem nos primeiros estágios da doença, enquanto ainda há neurônios dopaminérgicos remanescentes para convertê-la em dopamina. No entanto, com a progressão do processo degenerativo, ela deixa de ser efetiva e resulta em discinesias e outras complicações motoras que limitam o manejo farmacológico dos sintomas da DP (26).

3.2 SINAIS DE EEG E SEU PROCESSAMENTO

Os sinais de eletroencefalografia (EEG) são obtidos por meio do monitoramento eletrofisiológico com a finalidade de registrar graficamente as correntes elétricas produzidas no encéfalo, ou seja, a atividade elétrica cerebral (27). Para coletar os sinais, eletrodos em posições pré-definidas são dispostos na superfície do couro cabeludo do paciente ou participante da pesquisa (27). O EEG coleta o sinal elétrico proveniente da diferença de potencial entre os eletrodos de registro sobre o couro cabeludo e o tecido cortical localizado abaixo (28). Em seguida, um amplificador aumenta as intensidades dos potenciais elétricos obtidos e, ao final, o sinal é submetido à análise gráfica (27). O EEG é uma técnica não invasiva que possui alta resolução temporal e baixo custo financeiro em relação a técnicas de imagem como tomografia, imagem por ressonância magnética funcional (fMRI, do inglês *functional magnetic resonance imaging*) e a magnetoencefalografia (MEG), por isso é amplamente utilizada em pesquisas envolvendo engenharia biomédica, engenharia neural e neurociência (29).

Existem padrões de disposições pré-definidas dos eletrodos no crânio, produzindo configurações replicáveis. Um deles é o sistema internacional 10/20, mostrado na Figura 1, que se baseia na relação entre a localização de um eletrodo e a área subjacente do córtex cerebral (30).

Figura 2 – Sistema com 64 eletrodos.



Fonte: (31)

Uma das principais formas de análise de sinais de EEG é a análise espectral, que ocorre por meio da avaliação da presença de diferentes faixas de frequência no sinal, sendo que as 5 bandas de frequência mais conhecidas são (32) (33):

- Delta: menor que 4 Hz, é usualmente encontrada durante o estágio de sono profundo. Sua presença em adultos acordados pode ser um indício de doença ou lesão cerebral.
- Teta: de 4 a 7 Hz, é mais comum na infância e está associada ao estado de sonolência.
- Alfa: de 8 a 13 Hz, aparece com mais frequência em indivíduos normais no estado de relaxamento, despertos, mas com os olhos fechados.
- Beta: de 14 a 30 Hz, pode indicar estímulo intenso do sistema nervoso central ou estado de tensão e sua ausência está relacionada ao esforço mental e atenção.
- Gama: maior que 30 Hz, relaciona-se ao estado de processamento ativo de informações do córtex, responsável pelas funções motoras.

Os sinais de EEG se contaminam com diversos ruídos de origem mecânica, eletromagnética e/ou eletrostática, além de ruídos conhecidos como artefatos, que têm

origem fisiológica, como batimentos cardíacos, movimentos musculares, piscadas de olhos, entre outros (34). Esses ruídos podem atrapalhar a visualização de informações relevantes para a interpretação do sinal. Por essa razão, e também para selecionar faixas de frequência específicas, é necessário utilizar filtros digitais no pré-processamento dos sinais (35). O filtro passa-faixa é muito usado, pois seleciona uma faixa de frequência do sinal e elimina frequências acima e abaixo do intervalo escolhido (36).

Para analisar e processar os sinais de EEG, uma das ferramentas mais utilizadas é o EEGLAB, disponibilizada gratuitamente para uso no *software* comercial MATLAB®. O EEGLAB contém algoritmos de pré-processamento padrão disponíveis, como re-referenciamento e filtros. Possui ainda ferramentas de visualização úteis para quando o usuário necessita realizar, de forma manual, a rejeição de frequências indesejadas ou ruídos. Além disso, por meio desse ambiente, é possível analisar tanto características temporais quanto frequenciais do sinal e ele é capaz de executar a análise por componentes independentes (ICA, do inglês *independent component analysis*) (37).

3.3 TÉCNICAS DE ANÁLISE E PROCESSAMENTO DE SINAIS

Vários domínios de análise têm sido usados na extração de características de sinais de EEG, incluindo tempo e frequência (13). Dentro desses domínios, existem diversas ferramentas relevantes para caracterização de sinais de EEG (14) e algumas delas serão brevemente explicadas abaixo.

3.3.1 Análise Temporal e Estatística

Média É a medida central mais popular de um conjunto de dados amostrais. Se x_1, x_2, \dots, x_n representam as n observações de uma amostra da variável x , então, sua média aritmética simples é dada pela equação (38):

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (3.1)$$

Mediana A mediana é o valor que ocupa a posição central em um conjunto de dados ordenados. Portanto, ela tem a propriedade de dividir um conjunto de observações em duas partes iguais quanto ao número de seus elementos: os valores que são menores

ou iguais à mediana e os valores que são maiores ou iguais a ela (38). Seja n o número de observações num conjunto de dados:

- Se n for ímpar, a mediana é dada por:

$$M_d = x_{\frac{n+1}{2}} \quad (3.2)$$

- Se n for par, a mediana é dada pela equação:

$$M_d = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} \quad (3.3)$$

Desvio Padrão O desvio padrão é uma medida que expressa o nível de dispersão de um conjunto de dados, indicando o quanto o conjunto é uniforme. Quanto mais homogêneos os dados, mais próximo de zero o desvio (38). O desvio padrão é expresso por:

$$D = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \quad (3.4)$$

Potência Uma das características mais simples que pode ser extraída, e que é bastante utilizada para sinais de EEG (14), é a potência média, que para um sinal $x(t)$ é dada pela equação:

$$P = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=-N}^N |x_k|^2 \quad (3.5)$$

onde $x_k, k = -N, N + 1, \dots, N$, são amostras do sinal $x(t)$.

Assimetria A assimetria indica o quanto a curva de frequência de uma distribuição estatística desvia ou se afasta da posição simétrica. Se a assimetria for negativa, os dados espalham-se mais para a esquerda da média do que para a direita. Se a assimetria for positiva, os dados se espalham mais para a direita. A assimetria da distribuição normal (ou qualquer distribuição perfeitamente simétrica) é zero. A equação que apresenta a definição da assimetria de uma distribuição é dada por:

$$s = \frac{E(x - \bar{x})^3}{D^3} \quad (3.6)$$

em que \bar{x} é a média de x , D é o desvio padrão de x , e $E(y)$ representa o operador de esperança estatística da quantidade y .

Curtose A curtose mede a concentração ou dispersão dos valores de um conjunto de dados em relação às medidas de tendência central em uma distribuição de frequências conhecida (a distribuição Normal). A curtose da distribuição normal é 3. As distribuições que têm a curva mais aberta do que a distribuição normal têm uma curtose superior a 3; já as distribuições que têm uma curva mais fechada do que a distribuição normal têm uma curtose inferior a 3. A curtose de uma distribuição é definida por (39):

$$k = \frac{E(x - \bar{x})^4}{D^4} \quad (3.7)$$

3.3.2 Análise Espectral

A análise espectral dos sinais diz respeito à obtenção de características dos sinais no domínio da frequência em contraste com a análise no domínio do tempo. Um sinal no domínio do tempo pode ser convertido para o domínio da frequência através de operadores matemáticos chamados genericamente de transformadas (40). Uma das transformadas mais utilizadas é a de Fourier (a mais utilizada para análise de sinais de EEG), que decompõe um sinal periódico em uma soma de senos e cossenos (41). Matematicamente, tem-se que a transformada de Fourier de um sinal contínuo $x(t)$ é dada pela equação (40):

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft} dt \quad (3.8)$$

Para o cálculo da transformada de Fourier para sinais discretos, utiliza-se o algoritmo FFT (do inglês *Fast Fourier Transform*), que corresponde a amostras da transformada discreta da equação:

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x_n e^{-j\omega n} \quad (3.9)$$

em que x_n são as amostras do sinal e ω é a frequência discreta.

Os espectros HOS (*Higher Order Spectra*) são representações espectrais de momentos de ordem mais alta ou cumulantes do sinal. Uma das características do HOS é a detecção de acoplamento de fase quadrática, em que seu espectro de segunda

ordem (ou ordens mais altas) é capaz de reconhecer acoplamentos não lineares entre fases. Ou seja, se um elemento de frequência é composto por dois outros elementos de frequência, é possível diferenciá-los. Além disso, o HOS elimina ruídos gaussianos (42). Dentro do HOS, existem diversos métodos para caracterização do sinal, dois dos mais usados na literatura serão brevemente explicados abaixo.

O biespectro é a transformada de Fourier da correlação de terceira ordem do sinal. Se o biespectro de um sinal é zero, as componentes de onda não estão acopladas e/ou relacionadas umas com as outras. A Equação (3.10) descreve o biespectro (43).

$$Bis(f_1, f_2) = E[X(f_1) \cdot X(f_2) \cdot X^*(f_1 + f_2)] \quad (3.10)$$

Na Equação (3.10), X^* representa o complexo conjugado de X , $X(f)$ é a transformada de Fourier do sinal $x(t)$ e $E[\cdot]$ representa a operação de esperança estatística.

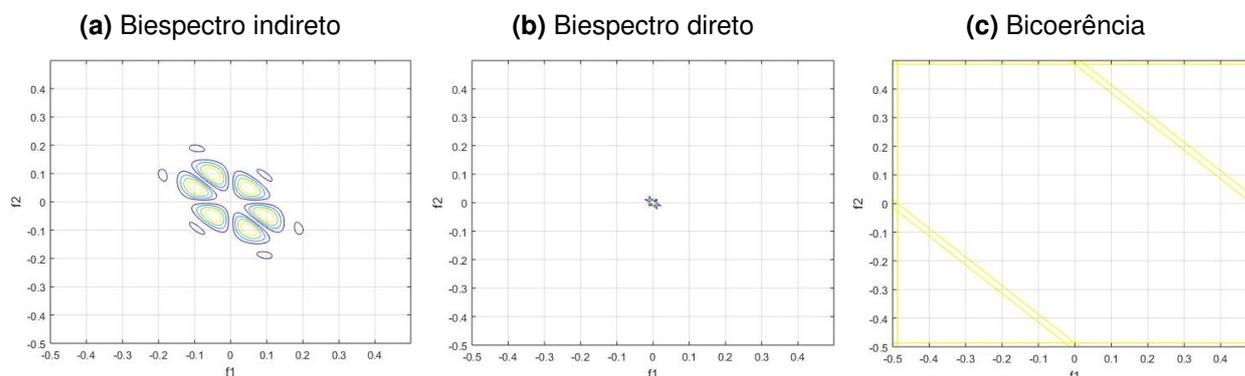
A bicoerência é o biespectro normalizado e é dada pela equação (3.11):

$$Bic(f_1, f_2) = \frac{Bis(f_1, f_2)}{\sqrt{P(f_1) \cdot P(f_2) \cdot P(f_1 + f_2)}} \quad (3.11)$$

sendo $P(f)$ a densidade espectral de potência (43). Como os dados de entrada são discretos no tempo, as ferramentas utilizam o processo de discretização para implementar o HOS.

Para se obter o biespectro de um sinal, existem abordagens diferentes, sendo duas delas o método indireto e o método direto. Tanto para a bicoerência quanto para os biespectros direto e indireto, os parâmetros de saída são uma matriz quadrática com comprimento igual ao da FFT, contendo os biespectros ou a bicoerência, e um vetor de frequências normalizadas associadas com as linhas e colunas dessa matriz (44).

É possível ainda obter a representação gráfica dos parâmetros de saída: uma imagem em duas dimensões associando pares de frequências através de uma espécie de curva de nível. Para fins de exemplificação, a Figura 3 mostra gráficos de biespectro indireto, direto e bicoerência plotados para a função seno com frequência de $1Hz$ e amplitude unitária.

Figura 3 – Exemplos de imagens obtidas pelo HOS

Fonte: A Autora

3.4 APRENDIZADO DE MÁQUINA

O aprendizado de máquina é um ramo em evolução de algoritmos computacionais que são desenvolvidos para simular a inteligência humana, aprendendo com o ambiente. Técnicas baseadas em aprendizado de máquina foram aplicadas com sucesso em diferentes campos, como o reconhecimento de padrões, visão computacional, engenharia espacial, direito, finanças, entretenimento, biologia, biomedicina, medicina, psicologia e muitos outros (45).

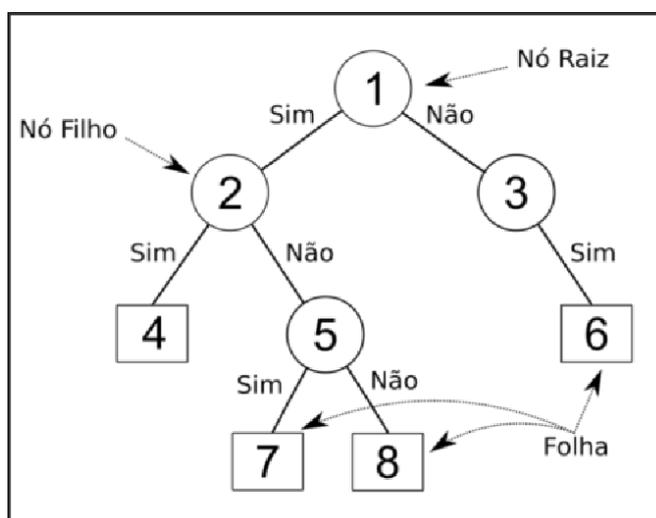
Um algoritmo de aprendizado de máquina é um processo computacional que usa dados de entrada para realizar uma tarefa desejada de maneira automática, para produzir um resultado específico. Esses algoritmos alteram ou adaptam automaticamente sua arquitetura por meio da experiência, para que se tornem cada vez melhores na realização da tarefa desejada. No aprendizado de máquina supervisionado, o processo de adaptação é chamado de treinamento, no qual são fornecidas amostras de dados de entrada e o resultado esperado para aquelas entradas. O algoritmo aprende para que ele possa não apenas produzir o resultado desejado com as entradas de treinamento, mas também generalizar o aprendizado, produzindo o resultado desejado a partir de dados novos e nunca antes vistos (45). Uma maneira de utilizar algoritmos de aprendizado de máquina é através de uma aplicação do Matlab chamada *Classification Learner*. Ela possui diferentes técnicas em seu escopo, sendo possível utilizar todas elas ao mesmo tempo com os mesmos dados e verificar qual delas possui o melhor desempenho. Algumas dessas técnicas serão citadas a seguir.

Máquina de vetores de suporte O SVM é um classificador binário não probabilístico e supervisionado, capaz de analisar um conjunto de dados rotulados e, a partir dessa análise, fazer previsões em dados não rotulados. Ele funciona por meio da busca por hiperplanos que melhor separam as informações (7).

K vizinhos mais próximos O método supervisionado dos k vizinhos mais próximos (k NN) é um dos representantes de algoritmos baseados em instâncias mais conhecidos e utilizados. No aprendizado baseado em instância, os exemplos são armazenados sem modificações ou generalizações. Dessa forma, uma função de distância determina quais instâncias do conjunto de dados são próximas ao novo exemplo a ser predito. O algoritmo k NN é baseado na ideia de que a classe da nova instância é a classe majoritária das k instâncias mais simples (46).

Árvore de decisão Uma técnica de aprendizado de máquina que é muito conhecida e funciona de maneira simples é a árvore de decisão. A árvore de decisão possui a mesma estrutura de uma árvore normal, contendo o nó de raiz, ramos e nós folhas. Um nó de raiz é pai de todos os nós e é o nó mais alto na árvore. Cada nó mostra uma característica (atributo), cada ligação (ramo) mostra uma decisão (regra) e cada folha mostra um resultado (categórico ou contínuo) (47). A Figura 4 ilustra a estrutura de uma árvore de decisão.

Figura 4 – Árvore de decisão.



Fonte: (48)

Florestas aleatórias As florestas aleatórias são uma combinação de preditores de árvore, de modo que cada árvore depende dos valores de um vetor aleatório amostrado independentemente e com a mesma distribuição para todas as árvores da floresta. O erro para esse tipo de classificador converge até um limite à medida que o número de árvores na floresta aumenta. Além disso, o erro depende da força das árvores individuais na floresta e da correlação entre elas (49).

Regressão logística A técnica de regressão logística funciona por meio da modelagem de previsão. Essa corresponde à tarefa de construir um modelo para a variável dependente em função das variáveis independentes. A principal característica da regressão logística é que sua variável dependente é categórica e geralmente binária (50).

Análise discriminante A análise discriminante constrói um modelo preditivo para fazer parte de um grupo. O modelo é composto por uma função discriminante (ou, para mais de dois grupos, um conjunto de funções discriminantes) com base em combinações lineares das variáveis preditoras que proporcionam a melhor discriminação entre os grupos. As funções são geradas a partir de uma amostra de casos que fazem parte de um grupo conhecido e podem então ser aplicadas a novos casos que tenham medidas para as variáveis preditoras, mas que fazem parte de um grupo desconhecido (51).

3.5 ESTIMATIVA DO MODELO

Uma forma eficaz de se avaliar o resultado de um modelo balanceado é por meio da discriminação dos erros (ou acertos) cometidos para cada classe (52), o que pode ser obtido com o uso da matriz de confusão. A matriz de confusão é uma tabela que mostra uma comparação entre as predições do modelo e o padrão verdadeiro. A Tabela 1 mostra a representação de uma matriz de confusão de dimensão 2x2.

Tabela 1 – Matriz de Confusão

		Classe Predita	
		Positivo	Negativo
Classe Real	Positivo	TP	FN
	Negativo	FP	TN

Fonte: A Autora.

A diagonal principal da matriz indica instâncias das classes que foram previstas corretamente (do inglês, True Negative - TN e True Positive - TP), enquanto os elementos da diagonal secundária representam as instâncias que foram classificadas erroneamente (do inglês, False Negative - FN e False Positive - FP). A partir desses valores, é possível calcular importantes parâmetros de avaliação, como acurácia (ACC), dada por:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.12)$$

A acurácia é o número de previsões acertadas pelo modelo dividido pela soma de todos os tipos de previsões realizadas.

3.6 ESTUDOS CORRELATOS

Há poucos trabalhos na literatura que estudaram a assimetria cortical utilizando sinais de EEG para diagnóstico da doença de Parkinson. Em um deles (53), foram selecionados 34 indivíduos com doença de Parkinson que ainda não haviam sido submetidos ao tratamento com medicação dopaminérgica e 18 indivíduos controles pareados por idade, sexo e dominância manual. Os sinais de EEG foram coletados por meio do sistema 10/20 enquanto os voluntários estavam em repouso com olhos abertos e olhos fechados. Posteriormente, uma dose aguda padronizada de medicação dopaminérgica foi administrada em todos os indivíduos com DP para calcular a magnitude da resposta de curta duração. Foram analisados especificamente os sinais dos eletrodos médio-frontal (F3 e F4), lateral-frontal (F7 e F8), anterior-temporal (T3 e T4), médio-parietal (P3 e P4) e occipital (O1 e O2). Após pré-processamento, foi realizada a análise espectral dos sinais por meio do periodograma obtido pelo método de

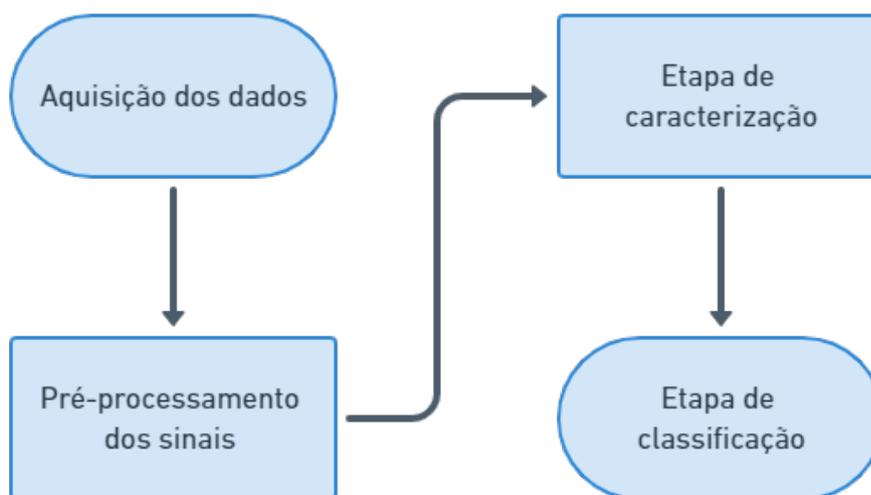
Welch (54). Para avaliar a assimetria cortical, foi utilizado um índice de lateralização (IL), que era obtido da subtração do logaritmo da densidade espectral de potência dos canais opostos. Tudo isso para as bandas de frequência delta (de 0 a 4 Hz), teta (4 a 8 Hz), alfa (8 a 13 Hz) e beta (13 a 30 Hz). Como resultado, foi encontrado que nas regiões frontal média/lateral havia um maior índice de lateralização para a banda beta e menor índice de lateralização para a banda teta em indivíduos com DP em comparação com controles. Já na região occipital, um menor índice de lateralização para a banda alfa foi encontrado na DP, correlacionando-se com a magnitude da resposta de curta duração à medicação dopaminérgica. Dessa forma, o trabalho citado teve como conclusão que a lateralização da atividade eletroencefalográfica beta do córtex frontal está associada à incapacidade clínica, e a atividade alfa do córtex occipital pode estar relacionada à capacidade de resposta à L-dopa na DP não tratada.

Um outro trabalho (55) utilizou sinais de EEG de 14 eletrodos no couro cabeludo de 20 pacientes com DP e 30 indivíduos controle de mesma idade. Durante a coleta dos sinais, os indivíduos foram submetidos a estímulos multimodais (audiovisuais) para evocar estados emocionais direcionados específicos, como felicidade, tristeza, medo, raiva, surpresa e nojo. Medidas derivadas da análise espectral dos sinais (potência absoluta, relativa, frequência e assimetria) foram submetidas ao teste ANOVA para comparação entre os grupos. Além disso, avaliações subjetivas foram obtidas para os estímulos emocionais usados. Como resultado, foi encontrado que os pacientes com DP apresentaram potências nas frequências delta, teta, alfa e beta menores em relação aos indivíduos controle em diferentes estados emocionais. A lateralização cortical foi observada na assimetria inter-hemisférica nas frequências teta, alfa e beta diferenças de potência, com os indivíduos controle exibindo maior ativação do hemisfério direito do que esquerdo. Além disso, a análise discriminante classificou corretamente 95,0% dos pacientes e indivíduos controle durante estímulos emocionais. Assim, o trabalho concluiu que essas potências espectrais distribuídas em diferentes bandas de frequência podem fornecer informações significativas sobre o processamento emocional em pacientes com DP.

4 MÉTODO

A Figura 5 mostra o fluxograma com as etapas da metodologia utilizada, que serão descritas a seguir.

Figura 5 – Etapas da metodologia do projeto



Fonte: A Autora.

4.1 AQUISIÇÃO DOS DADOS

A base de dados que foi utilizada está disponibilizada na plataforma aberta e gratuita *Openneuro* (56). Essa plataforma hospeda e disponibiliza 539 bases de dados públicas de diversos tipos de exames fisiológicos, como imagens de ressonância magnética, ecocardiograma e EEG. Foi usada a base compartilhada por Cavanagh *et al.* (2017, p. 27), grupo de pesquisa do departamento de psicologia da Universidade do Novo México, nos Estados Unidos, que contém sinais de EEG destinados ao estudo da doença de Parkinson, bem como ferramentas de análise para esses dados. Esse grupo disponibiliza 4 bases de dados para o estudo da DP, sendo que, em cada uma delas, os sinais de EEG foram gravados enquanto os participantes realizavam diferentes tarefas: repouso, Oddball Paradigm, teste de Simon e tarefa de aprendizado por reforço. Este trabalho se dedicou a estudar a base referente ao repouso e ao Oddball Paradigm, como será descrito a seguir.

O experimento ocorreu em 2015 na Universidade do Novo México e nas tarefas

Oddball Paradigm e repouso, participaram 25 pessoas diagnosticadas clinicamente com a doença de Parkinson, denominadas grupo DP, e um grupo controle, denominado grupo CTRL, formado por um número equivalente de indivíduos saudáveis (57). Os indivíduos com a doença de Parkinson realizaram o teste duas vezes com intervalo de uma semana, uma vez sob efeito da medicação dopaminérgica (grupo ON) e outra sem a utilização do medicamento nas últimas 15 horas (grupo OFF).

Todos os participantes foram submetidos ao mini exame de teste mental (MMSE, do inglês *Mini Mental State Exam*), que não identificou casos severos de demência. Além disso, os pacientes com DP e os indivíduos de controle foram pareados em idade e sexo e não possuíam diferenças significativas em nível de escolaridade. A Tabela 2 mostra as informações referentes aos participantes, em que LEDD (*L-Dopa equivalent daily dose*) refere-se à quantidade de medicação que o paciente toma e que tem efeito sobre os sintomas de forma equivalente a 100 mg de levodopa de liberação imediata. Ou seja, como os pacientes geralmente tomam várias medicações que têm efeitos semelhantes à levodopa, o LEDD é uma medida (em levodopas) para estimar o quanto de medicação o paciente usa por dia (58).

Tabela 2 – Informações dos pacientes com DP (ON e OFF) e CTRL (média +/- desvio padrão)

Informações	ON/OFF	CTRL
Sexo	16 H, 9 M	16 H, 9 M
Idade	69,68 (8,73)	69,32 (9,58)
LEDD	685 (462)	-
Anos desde o diagnóstico	5,4 (4,09)	-

Fonte: A Autora.

Para o repouso, os sinais de EEG dos participantes foram coletados enquanto eles ficaram sentados, sem realizar nenhuma atividade, por um minuto com os olhos abertos e um minuto com os olhos fechados.

A tarefa do Oddball Paradigm, nesse caso, consistiu em submeter os participantes ao *3-Stimulus Auditory Oddball Task*. Essa atividade corresponde à utilização de estímulos auditivos para investigar disfunções no processamento sensorial e cognitivo. Os estímulos são compostos por um sinal padrão (tom de 440 Hz), com uma taxa de

ocorrência de 70%, um sinal alvo, (tom de 660 Hz) com taxa de ocorrência de 15%, e um sinal divergente, representado por um som de natureza, ocorrendo 15% das vezes. Dessa forma, ao fim do experimento, os participantes deveriam contabilizar apenas os sinais alvo. Foram realizados dois blocos com 100 estímulos cada e toda a atividade durou em média 12 min por participante.

Os sinais de EEG foram gravados continuamente por um período de repouso seguido da tarefa do Oddball, por meio de 64 eletrodos de cloreto de prata com referência no canal AFz, uma taxa de amostragem de 500 Hz e uma banda de passagem entre 0,1 Hz e 100 Hz.

4.2 PRÉ-PROCESSAMENTO DOS SINAIS

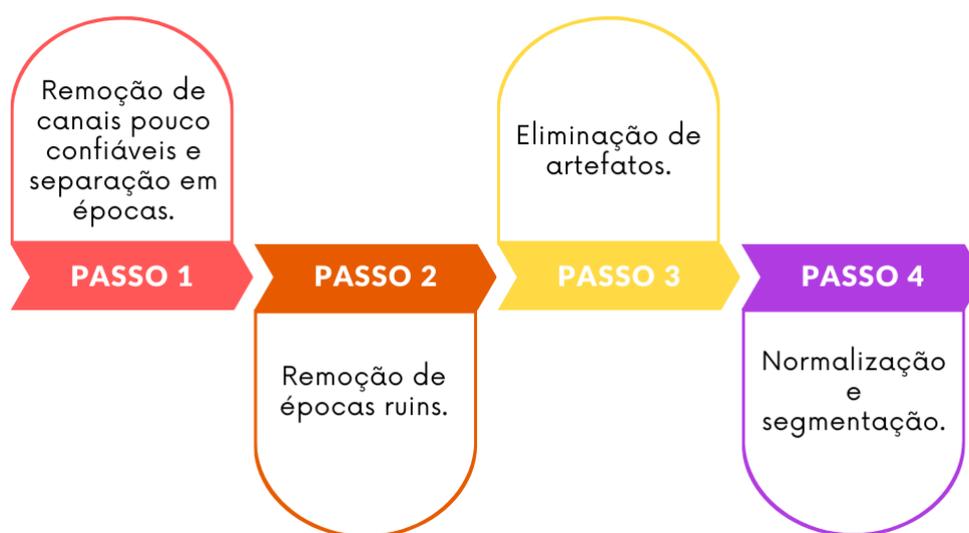
O pré-processamento foi realizado por meio do *software* MATLAB em conjunto com a *toolbox* EEGLab (59). Primeiro, foram removidos os quatro canais localizados na região temporal ventral (FT9, FT10, TP9 e TP10), pois eles apresentaram problemas na coleta do sinal (57), restando, assim, 60 eletrodos. Posteriormente, os sinais foram separados em épocas de 4 segundos de duração, de forma que cada época continha os 2 segundos anteriores a um estímulo sonoro e os 2 segundos subsequentes a esse mesmo estímulo. Dessa forma, os sinais foram compostos apenas pelos trechos correspondentes às respostas do EEG aos estímulos sonoros. Em seguida, canais e épocas ruins foram identificados e removidos por meio de uma função interna do EEGLab (*pop_rejchan*). Essa ferramenta compara o histograma do sinal com uma curva Gaussiana. Como espera-se que um sinal de EEG tenha o comportamento semelhante a uma Gaussiana, quando o seu histograma diferia consideravelmente dela, esse canal ou época era marcado como ruim (59).

Também foram removidos artefatos por meio do algoritmo FASTER (*Fully Automated Statistical Thresholding for EEG artifact Rejection*), que realiza essa remoção utilizando a técnica de ICA (60). O ICA consiste na busca por uma transformação linear que, ao ser aplicada a um conjunto de vetores aleatórios, minimize a dependência estatística das componentes geradas pela transformação (61). Assim, como se assume que os sinais captados pelos eletrodos são provenientes de fontes linearmente independentes, como os sinais cerebrais e os artefatos, é possível separar essas fontes com o uso do ICA. Os dados foram então re-referenciados com base na

média de todos os canais, ficando com suas amplitudes normalizadas.

Por último, foi necessário realizar a segmentação dos sinais de EEG de forma a separar a parte em que os indivíduos estavam em repouso da parte em que estavam submetidos à tarefa do Oddball. A Figura 6 resume os passos do pré-processamento citados acima.

Figura 6 – Etapas do pré-processamento dos sinais



Fonte: A Autora.

4.3 ETAPA DE CARACTERIZAÇÃO

4.3.1 Utilizando quatro pares de canais

Nessa etapa, foram utilizadas ferramentas para extrair características dos sinais de EEG no domínio da frequência, de forma a avaliar quais delas seriam mais promissoras para identificar biomarcadores para a doença de Parkinson. Com o objetivo de explorar a lateralização da doença de Parkinson para seu diagnóstico, foram escolhidos quatro pares de canais opostos para serem subtraídos.

Buscando analisar os sinais de EEG referentes aos participantes em repouso (antes do início da tarefa do Oddball), foram selecionados canais que, de acordo com a literatura, se mostraram relevantes para pacientes com DP nesse estado (1), (62), sendo eles: F3-F4, C3-C4, Fp1-Fp2 e P3-P4. Além disso, cada par corresponde a uma região diferente do cérebro (frontal, central, fronto-polar e parietal), permitindo uma

comparação entre lobos distintos. As características dos sinais utilizadas inicialmente foram as seis medidas estatísticas citadas na seção 3.3, selecionadas com base em um trabalho que buscava estudar a lateralização cortical para estudo de emoções (63), sendo elas: mediana, média, potência, desvio padrão, curtose e assimetria.

4.3.1.1 Repouso

Primeira abordagem Para obter características frequenciais, foi extraído o módulo ao quadrado dos coeficientes da transformada de Fourier (FFT) dos 85 primeiros segundos dos sinais e esses valores foram subtraídos para canais opostos. O módulo ao quadrado corresponde a uma estimativa da densidade espectral de potência, que é uma medida de força do sinal. Em seguida, foram separadas as frequências alfa (8 a 13 Hz), beta (13 a 30 Hz) e gama (30 a 70 Hz), e, por último, as medidas estatísticas foram extraídas. Todos esses procedimentos foram realizados para os grupos CTRL, ON e OFF.

Segunda abordagem Para uma segunda técnica de extração de características, a abordagem acima foi repetida, mas aplicando-se a normalização dos coeficientes da FFT, deixando todos os seus valores entre 0 e 1.

Terceira abordagem É a repetição da segunda, mas com o módulo dos coeficientes da FFT sem ser elevado ao quadrado, o que representa os valores absolutos dos coeficientes.

Quarta abordagem Modifica a etapa de subtração entre canais opostos. Nesse caso, os sinais desses canais foram primeiramente subtraídos e, em seguida, foi calculada a FFT. O módulo do seu resultado foi obtido e elevado ao quadrado, os resultados foram normalizados, as frequências alfa, beta e gama separadas e as medidas estatísticas extraídas. A Tabela 3 esquematiza as etapas de cada uma das abordagens explicadas acima.

Tabela 3 – Etapas de cada um dos 4 métodos de extração de características na ordem cronológica.

Método	Etapa 1	Etapa 2	Etapa 3	Etapa 4	Etapa 5	Etapa 6	Etapa 7
1	FFT	Módulo	Elevar ao quadrado	Subtrair	Separar freq.	Medidas estatíst.	
2	FFT	Módulo	Elevar ao quadrado	Normaliz.	Subtrair	Separar freq.	Medidas estatíst.
3	FFT	Módulo	Normaliz.	Subtrair	Separar freq.	Medidas estatíst.	
4	Subtrair	FFT	Módulo	Elevar ao quadrado	Normaliz.	Separar freq.	Medidas estatíst.

Fonte: A Autora.

4.3.1.2 Tarefa do Oddball

Nesse caso, em vez de utilizar os primeiros 85 segundos do sinal, utilizou-se um tempo intermediário, de em média 100 segundos, que se refere ao momento da realização da tarefa do Oddball, contemplando épocas referentes a vários tipos de estímulos. A técnica empregada para obter características do sinal foi a mesma usada na última abordagem para o repouso, descrita na última linha da Tabela 3.

4.3.2 Utilizando todos os canais

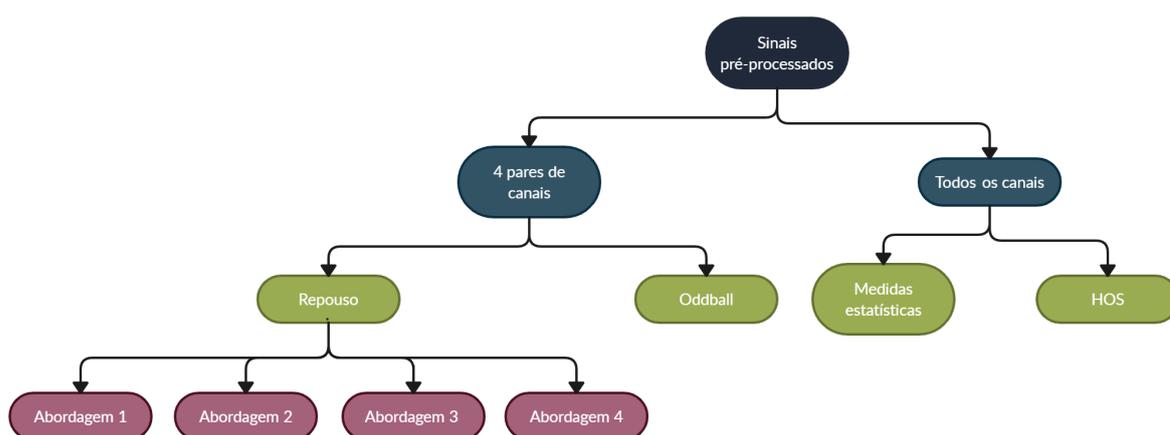
Numa análise posterior, foram utilizados todos os 60 canais com toda a extensão do sinal referente à tarefa do Oddball em comparação com diferentes formas de lateralização. O intuito foi comparar as acurácias do modelo confrontando CTRL e DP (ON e OFF) nas seguintes situações: todos os canais, subtração dos canais opostos, canais do hemisfério mais afetado do cérebro do indivíduo, canais do hemisfério menos afetado do cérebro, canais do hemisfério direito e canais do hemisfério esquerdo. Dessa forma, seria possível verificar se a assimetria cortical melhora o desempenho do modelo e qual tipo de assimetria apresenta o melhor resultado. Nessa abordagem, foram utilizadas duas técnicas de extração de características.

Medidas estatísticas A técnica de extração de características utilizada nessa abordagem, inicialmente, foi a mesma empregada por último na subtração dos quatro pa-

res de canais: foi utilizado o módulo ao quadrado normalizado da FFT do sinal, as frequências alfa, beta e gama separadas e as medidas estatísticas extraídas (mediana, média, potência, desvio padrão, curtose e assimetria).

HOS Como as medidas utilizadas anteriormente são características obtidas de estatísticas mais simples dos sinais, optou-se por utilizar uma técnica mais robusta de extrair esses possíveis biomarcadores. Por isso, posteriormente, foram empregadas como características dos sinais as frequências da bicoerência vetorizadas. O HOS é utilizado na literatura e foi apontado como um método eficaz para análise de sinais de EEG, sendo a bicoerência um dos métodos mais utilizados (17). A Figura 7 apresenta de forma esquematizada a etapa de extração de características dos sinais.

Figura 7 – Passos da etapa de caracterização dos sinais



Fonte: A Autora.

4.4 ETAPA DE CLASSIFICAÇÃO

4.4.1 Quatro pares de canais

Tanto para o repouso quanto para a tarefa do Oddball, foram usadas ferramentas de aprendizado de máquina para a classificação. Para isso, foi empregada a aplicação do Matlab *Classification Learner*, em que é possível inserir os dados e testar a classificação com 5 tipos de ferramentas de aprendizado de máquina diferentes. Dessa forma, pode-se identificar qual delas possui o melhor desempenho para os dados utilizados. Nesse caso, os classificadores utilizam as medidas estatísticas como ca-

racterísticas das classes CTRL x ON e CTRL x OFF, usaram o *5-fold cross-validation* como técnica de reamostragem e retornam a acurácia obtida pelos modelos.

4.4.2 Todos os canais

Para a etapa em que todos os canais foram utilizados, inicialmente foi empregado o *Classification Learner* com o *5-fold cross-validation*, para identificar quais classificadores tinham melhor desempenho nesse caso. Uma vez que foram escolhidos os classificadores (*k*NN, SVM e RF), foi aplicada uma técnica de reamostragem na classificação chamada *bootstrapping*, em que as características usadas para teste e treinamento variam. Nesse caso, o algoritmo fez 50 iterações com o *bootstrapping*, cada uma referente a uma configuração diferente de teste e treinamento e ,a cada iteração, os três classificadores obtiveram uma acurácia. Ao final das 50 iterações, o algoritmo extraiu a média dessas 50 acurácias de cada classificador.

5 RESULTADOS E DISCUSSÃO

5.1 UTILIZANDO QUATRO PARES DE CANAIS

5.1.1 Repouso

As ferramentas de aprendizado de máquina foram utilizadas em conjunto com a técnica *cross validation*. Optou-se por esta abordagem por estes algoritmos se basearem em critérios não lineares para encontrar padrões, o que poderia levar a resultados mais assertivos. Nas Tabelas 4 e 5, estão os resultados para as comparações CTRL x ON e CTRL x OFF, respectivamente. Os valores apresentados são as acurácias dos melhores classificadores para cada caso, sendo que DT se refere à árvore de decisão (do inglês, *decision tree*), DA à análise discriminante (*discriminant analysis*), kNN é referente ao método dos k vizinhos mais próximos (*K-nearest neighbor*) e SVM à máquina de vetores de suporte (*support vector machine*). Quanto mais próximo o resultado é de 50%, pior o desempenho do classificador, pois indica que ele realizou as predições das classes ao acaso. Dessa forma, os melhores resultados encontrados (em negrito na tabela) foram para a frequência gama do par P3-P4, em que as classes eram CTRL x ON, e a frequência beta de C3-C4 para as classes CTRL x OFF, ambos com uma acurácia de 64%.

Tabela 4 – Aprendizado de máquina para CTRL x ON e para a primeira abordagem da análise espectral do repouso.

Canais	Frequência	Acurácia (%)	Classificador
Fp1-Fp2	Alfa	58	DT
	Beta	62	kNN
	Gama	58	DT
F3-F4	Alfa	54	DA
	Beta	62	DA
	Gama	54	DT
C3-C4	Alfa	62	SVM
	Beta	52	SVM
	Gama	58	SVM
P3-P4	Alfa	58	DT
	Beta	54	SVM
	Gama	64	DT

Fonte: A Autora.

Tabela 5 – Aprendizado de máquina para CTRL x OFF e para a primeira abordagem da análise espectral do repouso.

Canais	Frequência	Acurácia (%)	Classificador
Fp1-Fp2	Alfa	58	kNN
	Beta	56	kNN
	Gama	54	kNN
F3-F4	Alfa	52	DA
	Beta	58	DA
	Gama	54	DT
C3-C4	Alfa	56	kNN
	Beta	64	kNN
	Gama	58	kNN
P3-P4	Alfa	56	kNN
	Beta	54	SVM
	Gama	60	DT

Fonte: A Autora.

Com o objetivo de melhorar a acurácia, um dos melhores casos foi escolhido para ser testado em novas abordagens. Então, foi realizada a segunda abordagem, em que os coeficientes da transformada de Fourier com a frequência na faixa beta foram normalizados e apenas os canais C3-C4 foram utilizados para discriminar as classes OFF e CTRL. Nesse caso, a maior acurácia obtida foi de 56%. Assim, devido a esse resultado próximo do acaso, a terceira abordagem foi proposta, em que o módulo da FFT não foi elevado ao quadrado e a maior acurácia obtida foi de 58%.

Portanto, foi sugerida a quarta proposta para o repouso, em que os sinais foram subtraídos, em vez dos coeficientes da FFT. Nesse caso, para o módulo da FFT não elevado ao quadrado, a acurácia foi de 60%, e com o módulo ao quadrado, foi de 70%, com o classificador baseado em modelos de regressão logística. Nessa análise preliminar, apesar das abordagens definidas parecerem não ter conseguido discriminar bem pacientes de indivíduos controle para todos os casos, alguns resultados indicaram que o estudo da lateralização da doença de Parkinson na busca por biomarcadores durante o repouso se mostrou promissor, mesmo utilizando características

simples, como média e mediana.

A Figura 8 apresenta a matriz de confusão para o melhor caso do repouso. Os zeros representam a classe CTRL e os uns a classe OFF. A cor rosa indica as previsões incorretas do modelo e a cor verde, as previsões corretas. Dessa forma, o classificador identificou corretamente 19 dos 25 indivíduos de controle e 16 dos pacientes com Parkinson.

Figura 8 – Matriz de confusão para a frequência beta dos canais C3-C4 no repouso.



Fonte: A Autora

5.1.2 Oddball Paradigm

As Tabelas 6 e 7 mostram as acurácias obtidas com as partes dos sinais relativas à tarefa do Oddball considerando uma parte de aproximadamente 100 segundos do sinal referente a vários tipos de estímulos. É possível observar que os resultados foram melhores do que aqueles referentes ao repouso, indicando uma maior diferenciação dos canais opostos do cérebro dos indivíduos de controle e dos indivíduos com Parkinson durante a realização dessa tarefa. Na Tabela 6, LR se refere à regressão logística (*logistic regression*). É possível observar que, para CTRL x ON, o melhor resultado foi para a frequência beta de F3-F4, com o kNN alcançando 70% de acurácia. Já para CTRL x OFF, o melhor resultado, encontrado para alfa de Fp1-Fp2, foi obtido

com o *k*NN com 74%, podendo ser um indicativo da diferença de ativação do cérebro em pacientes com e sem uso de levodopa nas últimas 15 horas.

Tabela 6 – Aprendizado de máquina para CTRL x ON e para a análise espectral do Oddball Paradigm.

Canais	Frequência	Acurácia (%)	Classificador
Fp1-Fp2	Alfa	56	LR
	Beta	64	DA
	Gama	58	DA
F3-F4	Alfa	54	SVM
	Beta	70	<i>k</i> NN
	Gama	60	SVM
C3-C4	Alfa	62	<i>k</i> NN
	Beta	66	<i>k</i> NN
	Gama	54	<i>k</i> NN
P3-P4	Alfa	68	SVM
	Beta	58	<i>k</i> NN
	Gama	56	<i>k</i> NN

Fonte: A Autora.

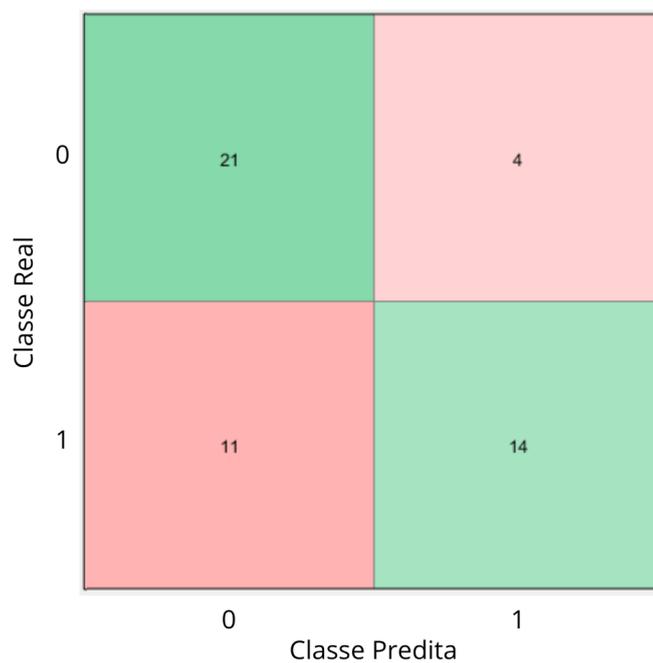
Tabela 7 – Aprendizado de máquina para CTRL x OFF e para a análise espectral do Oddball Paradigm.

Canais	Frequência	Acurácia (%)	Classificador
Fp1-Fp2	Alfa	74	kNN
	Beta	58	SVM
	Gama	68	DT
F3-F4	Alfa	60	kNN
	Beta	58	DA
	Gama	60	DA
C3-C4	Alfa	54	SVM
	Beta	62	kNN
	Gama	52	SVM
P3-P4	Alfa	54	SVM
	Beta	58	SVM
	Gama	62	SVM

Fonte: A Autora.

A Figura 9 mostra a matriz de confusão para a frequência beta dos canais F3-F4 e CTRL x ON. Nesse caso, o modelo acertou 21 dos 25 indivíduos controles e 14 dos 25 pacientes com Parkinson sob efeito da medicação. Já na Figura 10, é possível observar a matriz de confusão para a frequência alfa dos canais Fp1-Fp2, para o caso CTRL x OFF. Nota-se que o modelo indicou corretamente 21 dos indivíduos de controle, mas dessa vez acertou 18 dos pacientes com Parkinson. Esse resultado corrobora com o trabalho (57), que também obtém uma maior acurácia quando o grupo de controle é comparado com os pacientes que não tomaram a medicação, utilizando a mesma base de dados estudada no presente trabalho. Isso indica que o uso da levodopa reduz a capacidade de diferenciação dos indivíduos com e sem Parkinson pelo classificador.

Figura 9 – Matriz de confusão para a frequência beta dos canais F3-F4 na tarefa do Oddball para CTRL x ON.



Fonte: A Autora

Figura 10 – Matriz de confusão para a frequência alfa dos canais Fp1-Fp2 na tarefa do Oddball para CTRL x OFF.



Fonte: A Autora

Diante de resultados indicativos de que a lateralização cortical auxilia na diferenciação de pacientes com e sem Parkinson, uma abordagem comparativa foi proposta e os resultados são mostrados a seguir.

5.2 TODOS OS CANAIS

A Tabela 8 mostra os resultados utilizando todos os canais e toda a extensão do sinal relativa à tarefa do Oddball. Exibe também os resultados referentes à subtração dos canais opostos e ao hemisfério do cérebro mais afetado do indivíduo. Estão na tabela as acurácias (ACC) e o melhor classificador para cada caso, obtidos por meio do *classification learner*, pois como essa ferramenta permite o uso de vários classificadores ao mesmo tempo, com ela foi possível comparar o desempenho desses classificadores. Nas Tabelas 8 e 9, EN vem de *ensemble bagged trees*, que é um conjunto de árvores de decisão ensacadas para classificação, de forma a evitar o *overfitting* e melhorar a generalização (39). O *overfitting* é um problema do aprendizado de máquina supervisionado em que o modelo se adequa perfeitamente aos dados de treinamento, mas não funciona muito bem para dados de teste (64).

Tabela 8 – Acurácias (%) e melhores classificadores para todos os canais, subtração de canais opostos e hemisfério mais afetado do cérebro.

Frequência	Classes	Todos		Subtração		Hemisf. + afetado	
		ACC	Classif.	ACC	Classif.	ACC	Classif.
Alfa	CTRL x ON	76	DT	78	SVM	56	EN
	CTRL x OFF	54	kNN	76	LD	62	kNN
Beta	CTRL x ON	70	DT	70	SVM	58	SVM
	CTRL x OFF	66	kNN	70	SVM	62	LD
Gama	CTRL x ON	66	kNN	70	SVM	74	EN
	CTRL x OFF	66	SVM	70	SVM	64	SVM

Fonte: A Autora.

A Tabela 9 mostra os resultados utilizando o hemisfério menos afetado do cérebro do indivíduo, o hemisfério direito e o hemisfério esquerdo.

Tabela 9 – Acurácias (%) e melhores classificadores para o hemisfério menos afetado do cérebro, hemisfério direito e hemisfério esquerdo.

Frequência	Classes	Hemisf. - afetado		Hemisf. direito		Hemisf. esquerdo	
		ACC	Classif.	ACC	Classif.	ACC	Classif.
Alfa	CTRL x ON	62	LR	68	DT	64	SVM
	CTRL x OFF	50	kNN	68	SVM	58	LD
Beta	CTRL x ON	66	EN	62	EN	58	SVM
	CTRL x OFF	66	EN	68	SVM	64	EN
Gama	CRTL x ON	64	EN	60	LR	64	kNN
	CTRL x OFF	56	kNN	60	SVM	60	kNN

Fonte: A Autora.

As acurácias apresentadas nas Tabelas 8 e 9 foram as maiores obtidas no estudo. Dessas duas tabelas, é possível perceber que o melhor método de lateralização é da subtração entre os hemisférios opostos, pois todas as acurácias se mostraram superiores em relação às obtidas usando todos os canais. Além disso, os melhores classificadores para esses dados parecem ser kNN, SVM, DT e EN, sendo que esses dois últimos são tipos de árvores aleatórias.

Assim, em busca de resultados padronizados com o mesmo classificador, foi utilizada a técnica do *bootstrapping*. A Tabela 10 mostra as acurácias para todos os canais, subtração entre canais opostos e hemisfério do cérebro mais afetado para os três classificadores (kNN, SVM e RF). Já a Tabela 11 mostra as acurácias para o hemisfério do cérebro menos afetado, hemisfério direito e hemisfério esquerdo para os três classificadores.

Tabela 10 – Acurácias (%) do bootstrapping para a análise do Oddball Paradigm.

Frequência	Classes	Todos			Subtração			Hemisf. + afetado		
		kNN	SVM	RF	kNN	SVM	RF	kNN	SVM	RF
Alfa	CTRL x ON	59	53	61	62	65	68	48	47	54
	CTRL x OFF	52	54	60	58	57	65	45	50	47
Beta	CTRL x ON	52	57	55	53	60	64	45	53	47
	CTRL x OFF	50	61	52	53	63	66	45	52	46
Gama	CTRL x ON	44	48	50	58	59	71	51	58	52
	CTRL x OFF	50	50	46	50	54	58	48	61	49

Fonte: A Autora.

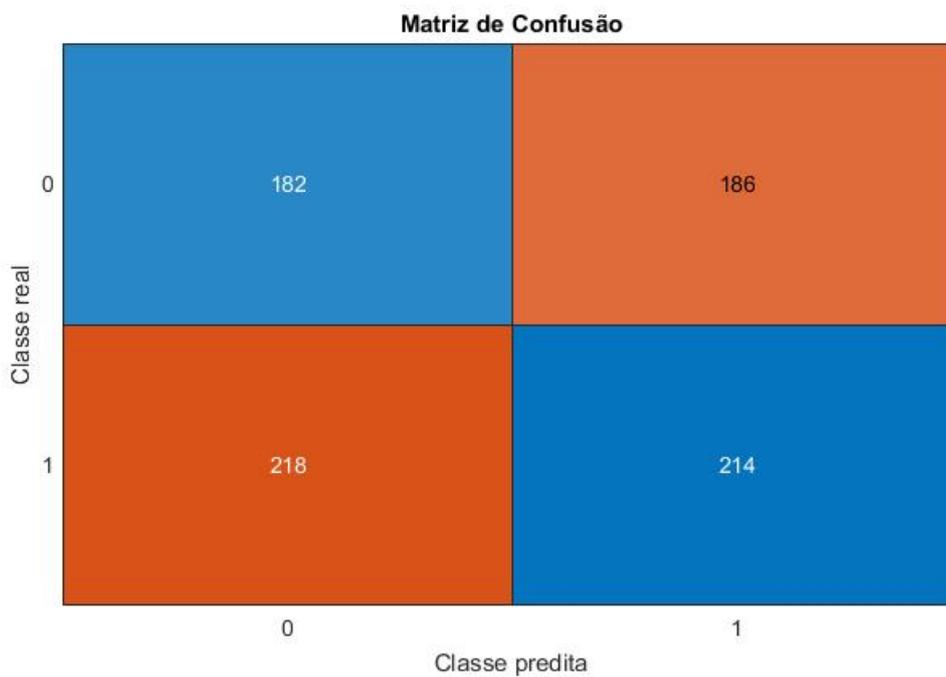
Tabela 11 – bootstrapping para a análise espectral do Oddball Paradigm.

Frequência	Classes	Hemisf. - afetado			Hemisf. direito			Hemisf. esquerdo		
		kNN	SVM	RF	kNN	SVM	RF	kNN	SVM	RF
Alfa	CTRL x ON	51	53	56	53	54	63	51	48	53
	CTRL x OFF	49	44	48	53	57	55	52	47	56
Beta	CTRL x ON	58	62	59	52	59	52	48	50	50
	CTRL x OFF	54	60	53	54	64	53	50	51	52
Gama	CTRL x ON	50	50	44	49	50	50	46	49	48
	CTRL x OFF	47	44	46	53	41	48	51	48	51

Fonte: A Autora.

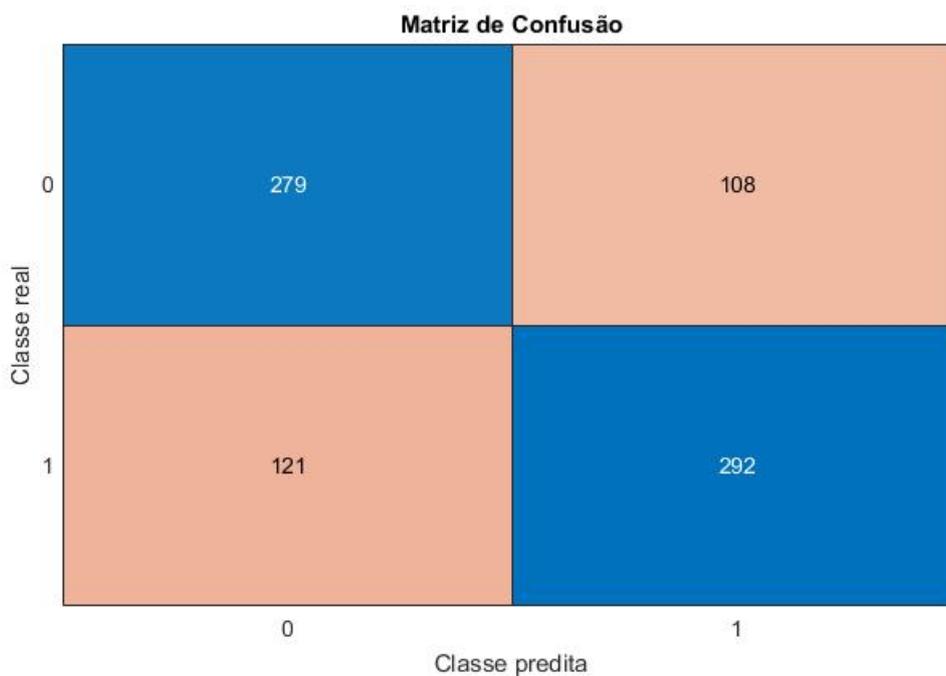
Dessa forma, a técnica de lateralização de subtrair os hemisférios opostos se sobressaiu mais uma vez, para os três classificadores, com destaque para o RF, que obteve as melhores acurácias. As Figuras 11 e 12 apresentam as matrizes de confusão para a frequência gama, CTRL x ON, com o RF como classificador, que foi o melhor caso, para todos os canais e a subtração dos canais opostos respectivamente. Nesse caso, o 0 corresponde à classe ON e o 1 refere-se ao grupo controle. As cores que tendem para o vermelho representam as predições feitas incorretamente, mudando sua intensidade de acordo com a quantidade de erros. Já as cores azuis indicam as predições corretas feitas pelo modelo.

Figura 11 – Matriz de confusão para a frequência gama CTRL x ON e RF considerando todos os canais



Fonte: A Autora

Figura 12 – Matriz de confusão para a frequência gama CTRL x ON e RF considerando a subtração dos canais



Fonte: A Autora

A cada iteração do *bootstrapping*, foram utilizados 17 dos 25 indivíduos controle

e 17 dos 25 com Parkinson para treinamento e os 8 restantes de cada grupo foram usados para teste. Por isso, as matrizes de confusão apresentam 400 instâncias em cada classe: são os 8 indivíduos de teste vezes as 50 iterações. Então, nota-se que o modelo acertou 214 das 400 instâncias referentes a casos controle utilizando todos os canais, enquanto acertou 292 fazendo uso da subtração de canais opostos, bem como 162 das 400 instâncias dos casos DP ON para todos os canais e 279 para a subtração. Foi realizado o teste t entre as acurácias para todos os canais e para a subtração em cada frequência (alfa, beta e gama) e para CTRL x ON e CTRL x OFF para avaliar se havia diferença estatística entre os resultados desses dois casos. Os resultados do p-valor estão listados na Tabela 12.

Tabela 12 – P-valor para todos os canais x subtração entre canais opostos.

Frequências	Classes	Classificador		
		kNN	SVM	RF
Alfa	CTRL x ON	0,1514	1,2966E-8	0,0036
	CTRL x OFF	0,0096	0,0474	0,0250
Beta	CTRL x ON	0,6795	0,0970	9,2890E-5
	CTRL x OFF	0,0049	0,3136	1,1508E-8
Gama	CTRL x ON	5,7217E-8	1,9886E-8	6,9707E-17
	CTRL x OFF	0,8283	0,0525	5,9759E-7

Fonte: A Autora.

O único classificador que apresentou diferença estatística entre as acurácias para todos os casos foi o RF, que se mostrou o melhor classificador para esses dados.

Como as características utilizadas até então eram medidas estatísticas simples, optou-se por testar algo mais robusto. Dessa forma, foram aplicadas as frequências da bicoerência aos classificadores. Inicialmente, foi empregado o *classification learner* para verificar quais os melhores classificadores. A Tabela 13 mostra os resultados obtidos.

Tabela 13 – Acurácias (%) e melhores classificadores para os coeficientes da bicoerência como características.

Frequências	Classes	Todos		Subtração	
		ACC	Classif.	ACC	Classif.
Alfa	CTRL x ON	70	kNN	82	SVM
	CTRL x OFF	72	EN	70	kNN
Beta	CTRL x ON	68	kNN	74	SVM
	CTRL x OFF	66	DT	72	kNN
Gama	CTRL x ON	72	LD	70	kNN
	CTRL x OFF	64	LD	78	EN

Fonte: A Autora.

Nesse caso, foi obtida uma acurácia de 82% para CTRL x ON na frequência alfa utilizando o SVM e a maioria dos resultados da subtração ficaram maiores em relação a todos os canais. Da mesma forma que para as medidas estatísticas como características, os classificadores que mais se repetiram foram kNN, SVM e os relacionados ao RF. Por isso, foi aplicada a técnica do *bootstrapping* para esses três classificadores. A Tabela 14 mostra esses resultados.

Tabela 14 – Acurácias (%) e melhores classificadores para os coeficientes da bicoerência como características.

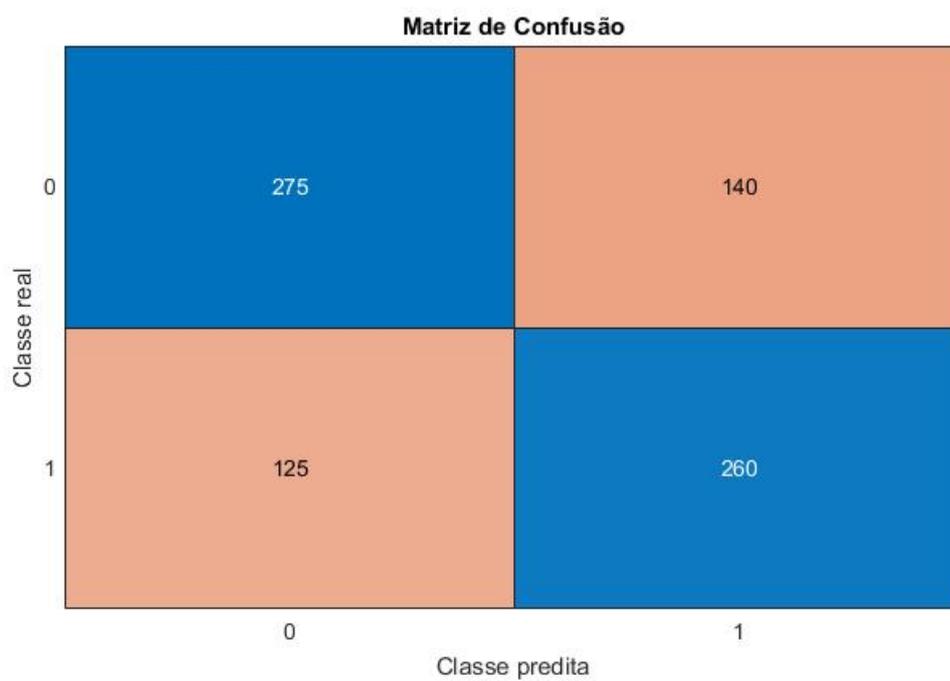
Frequências	Classes	Todos			Subtração		
		kNN	SVM	RF	kNN	SVM	RF
Alfa	CTRL x ON	62	64	66	68	70	69
	CTRL x OFF	60	59	63	61	59	64
Beta	CTRL x ON	58	64	63	64	69	65
	CTRL x OFF	51	63	56	62	64	63
Gama	CTRL x ON	54	62	55	63	65	59
	CTRL x OFF	46	57	49	54	57	60

Fonte: A Autora.

Os resultados dos três classificadores continuam se apresentando melhores para a subtração, com exceção apenas de CTRL x OFF nas frequências alfa e gama, que se

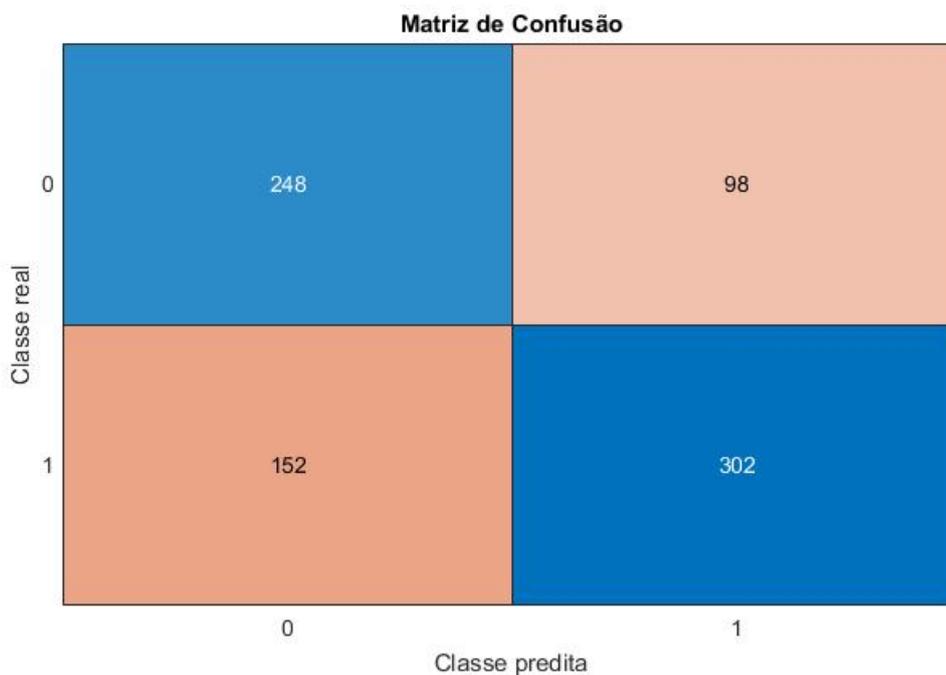
mantiveram iguais para o SVM. As Figuras 13 e 14 mostram as matrizes de confusão para a frequência alfa, CTRL x ON, usando o SVM como classificador, para todos os canais e subtração entre canais opostos respectivamente.

Figura 13 – Matriz de confusão para a frequência alfa CTRL x ON e SVM considerando todos os canais



Fonte: A Autora

Figura 14 – Matriz de confusão para a frequência alfa CTRL x ON e SVM considerando a subtração dos canais



Fonte: A Autora

Observa-se que o modelo acertou 250 das 400 instâncias referentes a casos controle utilizando todos os canais, enquanto acertou 302 fazendo uso da subtração de canais opostos, bem como, 275 das 400 instâncias dos casos DP ON para todos os canais e 260 para a subtração. Para verificar se há diferença estatística entre as acurácias para todos os canais e para a subtração, foi realizado o teste t, como mostra a Tabela 15.

Tabela 15 – P-valor para todos os canais x subtração entre canais opostos utilizando o HOS.

Frequências	Classes	Classificador		
		kNN	SVM	RF
Alfa	CTRL x ON	0,0474	0,0328	0,0050
	CTRL x OFF	0,6314	0,8596	0,4993
Beta	CTRL x ON	0,0051	0,0499	0,3639
	CTRL x OFF	2,8688E-7	0,4261	0,0016
Gama	CTRL x ON	2,6171E-6	0,1931	0,0186
	CTRL x OFF	0,0023	1,0000	2,0647E-6

Fonte: A Autora.

Nesse caso, o classificador que apresentou melhor diferenciação estatística entre os dois casos foi o kNN, em que o p-valor só apresentou um resultado acima de 5% para CTRL x OFF na frequência alfa.

Realizando-se uma comparação entre os dois métodos de extração de características, a Tabela 16 mostra os resultados do teste t entre as acurácias de todos os canais para o método das medidas estatísticas e o método do HOS e a Tabela 17 mostra a comparação entre as acurácias da subtração de canais opostos para os dois casos.

Tabela 16 – P-valor para as acurácias de todos os canais utilizando medidas estatísticas x todos os canais utilizando HOS.

Frequências	Classes	Classificador		
		kNN	SVM	RF
Alfa	CTRL x ON	0,0230	2,5628E-11	0,3323
	CTRL x OFF	6,1876E-4	7,8275E-4	0,1897
Beta	CTRL x ON	0,0073	1,9404E-4	2,5038E-4
	CTRL x OFF	0,4611	0,2625	0,0469
Gama	CTRL x ON	1,1365E-6	1,9027E-9	1,1054E-4
	CTRL x OFF	0,0262	2,2677E-4	0,1527

Fonte: A Autora.

Tabela 17 – P-valor para as acurácias da subtração de canais opostos utilizando medidas estatísticas x subtração de canais opostos utilizando HOS.

Frequências	Classes	Classificador		
		kNN	SVM	RF
Alfa	CTRL x ON	0,0106	0,0257	0,4454
	CTRL x OFF	0,1225	0,4673	0,3791
Beta	CTRL x ON	5,6384E-7	4,1673E-4	0,9038
	CTRL x OFF	0,0017	0,4573	0,0525
Gama	CTRL x ON	0,0077	3,7670E-4	7,6890E-7
	CTRL x OFF	0,1381	0,1474	0,4207

Fonte: A Autora.

Considerando o caso em que todos os canais são utilizados, as duas técnicas apresentam diferença estatística para todas as situações no kNN e SVM, com exceção de CTRL x OFF na frequência beta. Já no RF, não houve diferença estatística para a frequência alfa e CTRL x OFF na frequência gama. Já quando se considera a subtração entre canais opostos, não há diferenciação estatística em CTRL x OFF de alfa e gama em nenhum dos classificadores, além de beta CTRL x OFF no SVM e RF e alfa e beta CTRL x ON no RF. É possível observar que, a despeito de uma maior robustez e complexidade, as frequências do HOS não se mostraram mais eficientes enquanto características dos sinais para identificar se a lateralização cortical melhora a classificação de pacientes com e sem doença de Parkinson, já que as acurácias utilizando o *bootstrapping* se mostraram inferiores ou iguais aquelas obtidas usando medidas estatísticas simples. Com o HOS, é possível perceber também que a assimetria do cérebro melhora a identificação da DP, mas as acurácias não se mostraram superiores àquelas obtidas usando medidas mais simples como características. Além disso, o HOS possui um custo computacional bem mais elevado, levando em média seis vezes mais tempo para ser computado do que as medidas estatísticas.

Em (57), a resposta do sinal de EEG a eventos foi analisada como possível biomarcador da doença de Parkinson. Os autores utilizaram os sinais de EEG referentes à tarefa do Oddball e empregaram o classificador SVM linear, tendo como entradas os 50 primeiros termos da transformada rápida de Fourier do trecho de sinal imediata-

mente após um estímulo, podendo esse estímulo ser o sinal alvo, o sinal padrão ou o sinal divergente do *3-Stimulus Auditory Oddball Task*. A melhor acurácia alcançada utilizando como classes os grupos CTRL e OFF foi 82%. Esse valor foi obtido quando as entradas do classificador correspondiam apenas aos parâmetros extraídos de estímulos divergentes. Para os outros estímulos, a acurácia foi de aproximadamente 70%. Já para classificar os grupos CTRL e ON, a melhor acurácia foi 75%, também alcançada quando foram avaliados apenas os trechos de sinal correspondentes aos estímulos divergentes. Nota-se que, no trabalho citado, os autores usaram momentos específicos do sinal de forma a separar os tipos de estímulo. Esse tipo de abordagem não foi realizada no presente trabalho e a parte do sinal que foi analisada contém todos os tipos de estímulo mais de uma vez. Mesmo assim, acurácias de 78% e até 82% foram alcançadas com o *classification learner* e 71% com o *bootstrapping*, indicando que o estudo da lateralização do cérebro é promissor no processo da busca por biomarcadores da doença de Parkinson.

Alguns trabalhos que buscaram diagnóstico da DP utilizando também sinais de EEG obtiveram acurácias maiores que o presente estudo ao comparar pacientes com Parkinson a um grupo de controle, como (65) e (66), com acurácias de 99.2% e 98%, respectivamente. Entretanto, esses trabalhos utilizaram técnicas bem mais complexas de extração de características e de classificação, demandando mais tempo e capacidade computacional, bem como fizeram uso de bases de dados distintas. Em (65), os autores fizeram uso do HOS e de redes neurais convolucionais (CNN, do inglês *convolutional neural networks*). Já em (66), os autores aplicaram dados segmentados dos sinais de EEG a três redes neurais artificiais diferentes, fundindo os resultados ao final do processo. Isso enfatiza a importância do estudo da lateralização cerebral, pois mesmo com características simples, como média e mediana, e classificadores mais básicos, este trabalho pôde alcançar acurácias acima de 70%.

6 CONCLUSÃO E TRABALHOS FUTUROS

Este trabalho buscou investigar o uso da lateralização cortical no desenvolvimento de um sistema para auxiliar no diagnóstico da doença de Parkinson utilizando sinais de EEG. Para isso, foi utilizada uma base de dados pública contendo os sinais de EEG de um grupo de controle e de pacientes com Parkinson sob efeito da medicação dopaminérgica, bem como sem o seu uso recente, submetidos ao *3-Stimulus Auditory Oddball Task* e em estado de repouso. O estudo foi dividido em duas partes.

Na primeira parte, os sinais foram pré-processados e quatro pares de canais opostos (Fp1-Fp2, F3-F4, C3-C4 e P3-P4) foram subtraídos para que a lateralização da DP fosse investigada. Medidas estatísticas foram extraídas do sinal no domínio da frequência para o repouso e foram empregados classificadores de aprendizado de máquina, implementando as medidas estatísticas como características para diferenciar as classes CTRL e ON e CTRL e OFF nas frequências alfa, beta e gama. Empregando um classificador baseado em modelos de regressão logística para o momento do sinal referente ao repouso, foi obtida uma acurácia de 70% para a frequência beta dos canais C3-C4 em CTRL x OFF. Já utilizando a parte do sinal referente ao momento da realização da tarefa do Oddball, o *kNN* obteve uma acurácia de 70% para a frequência beta dos canais F3-F4 considerando CTRL x ON e 74% para a frequência alfa de Fp1-Fp2 usando CTRL e OFF como classes. Com isso, notou-se que é possível utilizar a lateralização para buscar biomarcadores para a doença de Parkinson. Foi observado também que a administração da medicação dopaminérgica reduz a acurácia da classificação, indicando que a substância interfere na resposta dos pacientes ao repouso e à tarefa do Oddball.

Na segunda parte, após pré-processamento, foram usados os 60 canais disponíveis coletados durante a tarefa do Oddball. Inicialmente foram utilizadas como características medidas estatísticas simples dos sinais e o *classification learner* foi empregado para identificação dos melhores classificadores para esses dados. A fim de buscar a forma mais eficiente de lateralização dos sinais, as classificações de CTRL x ON e CTRL x OFF nas frequências alfa, beta e gama foram realizadas para as seguintes situações: todos os canais, subtração entre canais de hemisférios opostos, hemisfério mais afetado do cérebro, hemisfério menos afetado, hemisfério direito e hemisfério esquerdo do cérebro. Com isso, foi possível identificar os melhores clas-

sificadores, sendo eles: *k*NN, SVM e RF. Em seguida, essa mesma abordagem foi aplicada para esses três classificadores por meio da técnica *bootstrapping*. Assim, foi observado que a forma de lateralização mais eficiente foi a subtração de canais opostos, em que todas as acurácias foram maiores que aquelas referentes a todos os canais para o três classificadores. Pela observação dos resultados e por meio de um teste t entre as acurácias para todos os canais e acurácias para subtração, o melhor classificador foi o RF.

Na tentativa de melhorar o desempenho do modelo, foram empregadas como características as frequências do HOS. Utilizando o *classification learner*, os melhores classificadores continuaram sendo *k*NN, SVM e RF, que também foram usados com a técnica do *bootstrapping*. Os resultados para a subtração de canais opostos também se mostraram superiores aos resultados para todos os canais. Entretanto, o desempenho do modelo se manteve parecido com aquele obtido utilizando apenas medidas estatísticas como características, mostrando que o HOS não se mostrou uma boa técnica para esses dados, já que seu custo computacional é bem mais elevado. Dessa forma, é possível concluir que a assimetria cortical melhora sim o desempenho do modelo para identificação de indivíduos com e sem doença de Parkinson.

Por fim, sugere-se que, em trabalhos futuros, as partes dos sinais referentes a cada tipo de estímulo do Oddball sejam separadas para serem analisadas individualmente. Além disso, outras ferramentas, como transformada Wavelet e Higher Order Crossings (HOC), podem ser utilizadas para extração de características a fim de melhorar a acurácia da classificação. Ademais, outros tipos de classificadores, como redes neurais de diferentes tipos, podem ser investigados. Outra sugestão é a aquisição de sinais de EEG de indivíduos com Parkinson aqui na região do Vale do São Francisco, usando a mesma abordagem, para investigar questões regionais nesses sinais.

REFERÊNCIAS

- 1 SILVA, G. et al. Parkinson disease early detection using EEG channels cross-correlation. *International Journal of Applied Engineering Research*, v. 15, n. 3, p. 197–203, 2020.
- 2 POEWE, W. et al. Parkinson disease. *Nature reviews Disease primers*, Nature Publishing Group, v. 3, n. 1, p. 1–21, 2017.
- 3 KHOSHNEVIS, S. A.; SANKAR, R. Classification of the stages of Parkinson's disease using novel higher-order statistical features of EEG signals. *Neural Computing and Applications*, Springer, p. 1–13, 2020.
- 4 OH, S. L. et al. A deep learning approach for Parkinson's disease diagnosis from EEG signals. *Neural Computing and Applications*, Springer, p. 1–7, 2018.
- 5 MEI, J.; DESROSIERS, C.; FRASNELLI, J. Machine learning for the diagnosis of Parkinson's disease: a review of literature. *Frontiers in aging neuroscience*, Frontiers Media SA, v. 13, p. 633752, 2021.
- 6 SAKAR, B. E. et al. Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. *IEEE Journal of Biomedical and Health Informatics*, IEEE, v. 17, n. 4, p. 828–834, 2013.
- 7 SHETTY, S.; RAO, Y. Svm based machine learning approach to identify Parkinson's disease using gait analysis. In: IEEE. *2016 International Conference on Inventive Computation Technologies (ICICT)*. [S.l.], 2016. v. 2, p. 1–5.
- 8 ANJUM, M. F. et al. Linear predictive coding distinguishes spectral EEG features of Parkinson's disease. *Parkinsonism & Related Disorders*, Elsevier, v. 79, p. 79–85, 2020.
- 9 PICARD, R. W.; VYZAS, E.; HEALEY, J. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 23, n. 10, p. 1175–1191, 2001.
- 10 OLIVEIRA, A. P. S. de et al. Early diagnosis of Parkinson's disease using EEG, machine learning and partial directed coherence. *Research on Biomedical Engineering*, Springer, v. 36, n. 3, p. 311–331, 2020.
- 11 MALIK, A. S.; AMIN, H. U.; ULLAHAMIN, H. Visual and cognitive fatigue during learning. *Designing EEG Experiments for Studying the Brain*, Academic Press, n. 8, p. 123–135, 2017.
- 12 RIEDERER, P. et al. Lateralisation in Parkinson disease. *Cell and tissue research*, Springer, v. 373, n. 1, p. 297–312, 2018.
- 13 WEI, W. et al. Emotion recognition based on weighted fusion strategy of multichannel physiological signals. *Computational intelligence and neuroscience*, Hindawi, v. 2018, 2018.

- 14 JENKE, R.; PEER, A.; BUSS, M. Feature extraction and selection for emotion recognition from EEG. *IEEE Transactions on Affective computing*, IEEE, v. 5, n. 3, p. 327–339, 2014.
- 15 HUANG, Y. et al. Combining facial expressions and electroencephalography to enhance emotion recognition. *Future Internet*, MDPI, v. 11, n. 5, p. 105, 2019.
- 16 DRISSI, T. B. et al. Diagnosis of Parkinson's disease based on wavelet transform and mel frequency cepstral coefficients. *International Journal of Advanced Computer Science and Applications*, v. 10, n. 3, 2019.
- 17 YEAN, C. W. et al. An emotion assessment of stroke patients by using bispectrum features of EEG signals. *Brain sciences*, Multidisciplinary Digital Publishing Institute, v. 10, n. 10, p. 672, 2020.
- 18 CHATURVEDI, M. et al. Quantitative EEG (qEEG) measures differentiate Parkinson's disease (PD) patients from healthy controls (HC). *Frontiers in aging neuroscience*, Frontiers, v. 9, p. 3, 2017.
- 19 VANNESTE, S.; SONG, J.-J.; RIDDER, D. D. Thalamocortical dysrhythmia detected by machine learning. *Nature communications*, Nature Publishing Group, v. 9, n. 1, p. 1–13, 2018.
- 20 CAVANAGH, J. F. et al. The Patient Repository for EEG Data+ Computational Tools (pred+ ct). *Frontiers in neuroinformatics*, Frontiers, v. 11, p. 67, 2017.
- 21 YUVARAJ, R.; ACHARYA, U. R.; HAGIWARA, Y. A novel Parkinson's disease diagnosis index using higher-order spectra features in EEG signals. *Neural Computing and Applications*, Springer, v. 30, n. 4, p. 1225–1235, 2018.
- 22 CHALLA, K. N. R. et al. An improved approach for prediction of Parkinson's disease using machine learning techniques. In: IEEE. *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*. [S.l.], 2016. p. 1446–1451.
- 23 PENG, B. et al. A multilevel-roi-features-based machine learning method for detection of morphometric biomarkers in Parkinson's disease. *Neuroscience letters*, Elsevier, v. 651, p. 88–94, 2017.
- 24 ARMSTRONG, M. J.; OKUN, M. S. Diagnosis and treatment of Parkinson disease: a review. *Jama*, American Medical Association, v. 323, n. 6, p. 548–560, 2020.
- 25 HORNYKIEWICZ, O. A brief history of levodopa. *Journal of neurology*, Springer, v. 257, n. 2, p. 249–252, 2010.
- 26 TRAN, T. N. et al. Levodopa-induced dyskinesia: clinical features, incidence, and risk factors. *Journal of Neural Transmission*, Springer, v. 125, p. 1109–1117, 2018.
- 27 ZANGEROLAME, F. M. *Monitoramento analógico e digital de sinais elétricos cerebrais*. 64 p. Dissertação (Mestrado) — Centro Federal de Educação Tecnológica Celso Suckow da Fonseca (CEFET), Rio de Janeiro, 2009.

- 28 SCHWARTZ, A. B. et al. Brain-controlled interfaces: movement restoration with neural prosthetics. *Neuron*, Elsevier, v. 52, n. 1, p. 205–220, 2006.
- 29 CRAIK, A.; HE, Y.; CONTRERAS-VIDAL, J. L. Deep learning for electroencephalogram (EEG) classification tasks: a review. *Journal of neural engineering*, IOP Publishing, v. 16, n. 3, p. 031001, 2019.
- 30 ALARCAO, S. M.; FONSECA, M. J. Emotions recognition using EEG signals: A survey. *IEEE Transactions on Affective Computing*, IEEE, v. 10, n. 3, p. 374–393, 2017.
- 31 BANDEIRA, V. Redução de sensores de EEG e detecção de movimento motor através de redes neurais artificiais. 2017.
- 32 TAVARES, M. C.; ENG, M.; BIOMÉDICA, E. EEG e potenciais evocados—uma introdução. *Contronic Sistemas Automáticos Ltda*, p. 1–13, 2011.
- 33 LIMA, I. N. O. et al. Desenvolvimento de uma plataforma para o processamento estatístico de sinal eletroencefalográfico (EEG). Universidade Federal de Uberlândia, 2018.
- 34 JUNG, T.-P. et al. Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, Wiley Online Library, v. 37, n. 2, p. 163–178, 2000.
- 35 PROCHAZKA, A.; KUKAL, J.; VYSATA, O. Wavelet transform use for feature extraction and EEG signal segments classification. In: IEEE. *2008 3rd International symposium on communications, control and signal processing*. [S.l.], 2008. p. 719–722.
- 36 MCCLELLAN, J.; PARKS, T.; RABINER, L. A computer program for designing optimum fir linear phase digital filters. *IEEE Transactions on Audio and Electroacoustics*, IEEE, v. 21, n. 6, p. 506–526, 1973.
- 37 BRUNNER, C.; DELORME, A.; MAKEIG, S. EEGlab—an open source Matlab toolbox for electrophysiological research. *Biomedical Engineering/Biomedizinische Technik*, De Gruyter, v. 58, n. SI-1-Track-G, p. 000010151520134182, 2013.
- 38 FARIA, R. R. A. Medidas de tendência central. 2017.
- 39 MATHWORKS Help Center. Disponível em: <<https://www.mathworks.com/help/index.html>>. Acesso em: 06 de setembro 2022.
- 40 LATHI, B. P. *Sinais e sistemas lineares-2*. [S.l.]: Bookman, 2006.
- 41 MELO, C. H. S. *Combinando sinais de EEG com respostas a questionários afetivos e expressões faciais para avaliação de diferentes estados emocionais*. 90 p. Dissertação (Mestrado) — Universidade Federal do Vale do São Francisco, Petrolina-PE, 2019.
- 42 NIKIAS, C. L.; MENDEL, J. M. Signal processing with higher-order spectra. *IEEE Signal processing magazine*, IEEE, v. 10, n. 3, p. 10–37, 1993.

- 43 HOSSEINI, S. A. et al. Higher order spectra analysis of EEG signals in emotional stress states. In: IEEE. *2010 Second international conference on information technology and computer science*. [S.l.], 2010. p. 60–63.
- 44 SWAMI ANANTHRAM, J. M. M.; NIKIAS, C. L. Higher-order spectral analysis toolbox. *The Mathworks Inc 3*, p. 22–26, 1998.
- 45 NAQA, I. E.; MURPHY, M. J. What is machine learning? In: *machine learning in radiation oncology*. [S.l.]: Springer, 2015. p. 3–11.
- 46 ALPAYDIN, E. *Introduction to Machine learning*. [S.l.]: MIT press, 2020.
- 47 PATEL, H. H.; PRAJAPATI, P. Study and analysis of decision tree based classification algorithms. *International Journal of Computer Sciences and Engineering, ISROSET: International Scientific Research Organization for Science . . .*, v. 6, n. 10, p. 74–78, 2018.
- 48 SATO, L. et al. Análise comparativa de algoritmos de árvore de decisão do sistema WEKA para classificação do uso e cobertura da terra. In: . [S.l.: s.n.], 2013.
- 49 BREIMAN, L. Random forests. *Machine learning*, Springer, v. 45, p. 5–32, 2001.
- 50 GONZALEZ, L. d. A. Regressão logística e suas aplicações. Universidade Federal do Maranhão, 2018.
- 51 IBM Documentation. Disponível em: <<https://www.ibm.com/docs/en>>. Acesso em: 06 de setembro 2022.
- 52 CASTRO, C. L. d.; BRAGA, A. P. Supervised learning with imbalanced data sets: an overview. *Sba: Controle & Automação Sociedade Brasileira de Automatica, SciELO Brasil*, v. 22, p. 441–466, 2011.
- 53 MOSTILE, G. et al. Electroencephalographic lateralization, clinical correlates and pharmacological response in untreated Parkinson's disease. *Parkinsonism & Related Disorders*, Elsevier, v. 21, n. 8, p. 948–953, 2015.
- 54 WELCH, P. The use of Fast Fourier Transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics*, IEEE, v. 15, n. 2, p. 70–73, 1967.
- 55 YUVARAJ, R. et al. On the analysis of EEG power, frequency and asymmetry in Parkinson's disease during emotion processing. *Behavioral and brain functions*, Springer, v. 10, p. 1–19, 2014.
- 56 OPENNEURO. Disponível em: <<https://openneuro.org/>>. Acesso em: 06 de setembro 2022.
- 57 CAVANAGH, J. F. et al. Diminished EEG habituation to novel events effectively classifies Parkinson's patients. *Clinical Neurophysiology*, Elsevier, v. 129, n. 2, p. 409–418, 2018.

- 58 TOMLINSON, C. L. et al. Systematic review of levodopa dose equivalency reporting in Parkinson's disease. *Movement disorders*, Wiley Online Library, v. 25, n. 15, p. 2649–2653, 2010.
- 59 TALK:CHAPTER 01: Rejecting artifacts. Disponível em: <https://sccn.ucsd.edu/wiki/Talk:Chapter_01:_Rejecting_Artifacts>.
- 60 NOLAN, H.; WHELAN, R.; REILLY, R. B. Faster: fully automated statistical thresholding for EEG artifact rejection. *Journal of neuroscience methods*, Elsevier, v. 192, n. 1, p. 152–162, 2010.
- 61 COMON, P. Independent component analysis, a new concept? *Signal processing*, Elsevier, v. 36, n. 3, p. 287–314, 1994.
- 62 YI, G.-S. et al. Complexity of resting-state EEG activity in the patients with early-stage Parkinson's disease. *Cognitive Neurodynamics*, Springer, v. 11, n. 2, p. 147–160, 2017.
- 63 PANE, E. S.; WIBAWA, A. D.; PURNOMO, M. H. Improving the accuracy of EEG emotion recognition by combining valence lateralization and ensemble learning with tuning parameters. *Cognitive processing*, Springer, v. 20, n. 4, p. 405–417, 2019.
- 64 YING, X. An overview of overfitting and its solutions. In: IOP PUBLISHING. *Journal of physics: Conference series*. [S.l.], 2019. v. 1168, p. 022022.
- 65 LEE, S. et al. A convolutional-recurrent neural network approach to resting-state EEG classification in Parkinson's disease. *Journal of neuroscience methods*, Elsevier, v. 361, p. 109282, 2021.
- 66 SHABAN, M. Automated screening of Parkinson's disease using deep learning based electroencephalography. In: IEEE. *2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)*. [S.l.], 2021. p. 158–161.