

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO RIO GRANDE
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM COMPUTACIONAL

**MINERAÇÃO DE DADOS E NEUROCIÊNCIA: UM
ESTUDO DE SINAIS CEREBRAIS COM BASE NA
MODULAÇÃO DA AMPLITUDE**

por

Bianca Parulla Marques

Dissertação para obtenção do Título de
Mestre em Modelagem Computacional

Orientadora: Prof^a. Dr^a. Diana Francisca Adamatti
Co-orientador: Prof. Dr. Cleo Zanella Billa

Rio Grande, 2019



UNIVERSIDADE FEDERAL DO RIO GRANDE - FURG
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM COMPUTACIONAL
Av. Itália km 8 Campus Carreiros - CEP 96201-900 - Rio Grande/RS
E-mail: ppgmc@furg.br - Tel: (53) 3293 5055



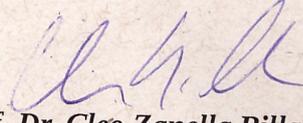
ATA DE DEFESA DE DISSERTAÇÃO

ATA Nº 17/2018

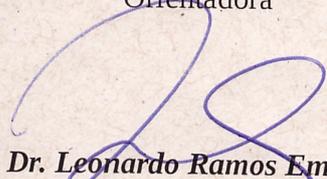
Ao vigésimo sexto dia do mês de outubro, no auditório do C3 na FURG, realizou-se a décima sétima sessão de defesa de dissertação do ano de dois mil e dezoito do Programa de Pós-Graduação em Modelagem Computacional. A mestranda **Bianca Parulla Marques** apresentou e defendeu a dissertação "**Mineração de dados e neurociências: um estudo dos sinais cerebrais com base na modulação de amplitude**" como requisito parcial para a obtenção do grau de Mestre em Modelagem Computacional. A sessão foi aberta às dez horas pela Prof^a. Dr^a. Diana Francisca Adamatti, orientadora da dissertação e presidente da Comissão de Avaliação. A referida Comissão esteve integrada, além da presidente e orientadora, pelo Prof. Dr. Cleo Zanella Billa – coorientador da FURG; pelo Prof. Dr. Leonardo Ramos Emmendorfer, da FURG; pelo Prof. Dr. Marcelo Rita Pias, da FURG e pelo Prof. Dr. Vinícius Rosa Cota, da UFSJ. Após cada integrante da Banca ter arguido a mestranda, e este ter tido a oportunidade de responder a cada um, a Comissão reuniu-se para deliberar sobre o conceito a ser atribuído ao trabalho. A Comissão de Avaliação considerou o candidato **APROVADO** neste requisito parcial para a obtenção do grau de Mestre em Modelagem Computacional. Por fim, o presidente publicou e encerrou a sessão da qual eu, Wiliam Correa Marques, na qualidade de Coordenador do Programa de Pós-Graduação em Modelagem Computacional, lavrei a presente ata que vai assinada por todos os membros da Comissão de Avaliação. Rio Grande, vinte e seis de outubro de 2018.


Prof^a. Dr^a. Diana Francisca Adamatti

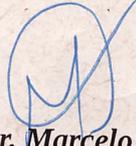
Orientadora


Prof. Dr. Cleo Zanella Billa

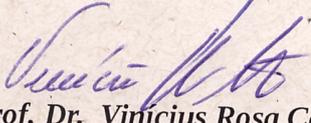
Coorientador


Prof. Dr. Leonardo Ramos Emmendorfer

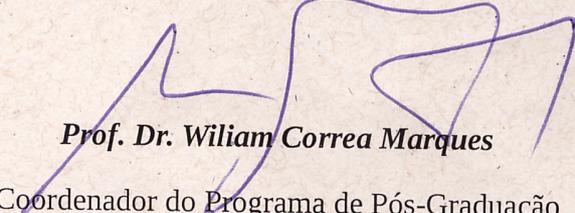
Membro Interno


Prof. Dr. Marcelo Rita Pias

Membro Interno


Prof. Dr. Vinícius Rosa Cota

Membro Externo


Prof. Dr. Wiliam Correa Marques
Coordenador do Programa de Pós-Graduação
em Modelagem Computacional

Ficha catalográfica

M357m Marques, Bianca Parulla.

Mineração de dados e neurociência : um estudo de sinais cerebrais com base na modulação da amplitude / Bianca Parulla Marques. – 2019.

71f.

Dissertação (mestrado) – Universidade Federal do Rio Grande – FURG, Programa de Pós-Graduação em Modelagem Computacional, Rio Grande/RS, 2019.

Orientadora: Dra. Diana Francisca Adamatti.

Coorientador: Dr. Cleo Zanella Billa.

1. Neurociências 2. Sinais cerebrais 3. Mineração de dados
4. Árvores de decisão I. Adamatti, Diana Francisca II. Billa, Cleo Zanella III. Título.

CDU 004.62:612.8

*À minha família,
pelo apoio sempre para
com as minhas decisões.
Serei eternamente grata.
Amo vocês!*

RESUMO

MARQUES, Bianca Parulla. **MINERAÇÃO DE DADOS E NEUROCIÊNCIA: UM ESTUDO DE SINAIS CEREBRAIS COM BASE NA MODULAÇÃO DA AMPLITUDE**. 2019. 71 f. Dissertação (Mestrado) – Programa de Pós-Graduação em Modelagem Computacional. Ministério da Educação, Rio Grande.

A neurociência é o estudo do sistema nervoso, sendo o encéfalo seu principal componente. Associado aos Sistemas BCI (Brain Computer Interface), é possível a comunicação do encéfalo com o computador. Uma das formas de coleta dos sinais cerebrais é a partir da eletroencefalografia (EEG). As ondas cerebrais podem ser classificadas usando sua frequência, amplitude, forma e posição no couro cabeludo. As aplicações EEG se concentram em uma faixa de 0,1 Hz a 100 Hz, onde os sinais de EEG são classificados com base em sua faixa de frequência e estão relacionados entre as ondas: Alfa, Beta, Teta, Gama e Delta. Este trabalho foca na modulação da amplitude da faixa de frequência da Banda Beta, que é a faixa de ondas cerebrais que situa-se entre 13 Hz e 30 Hz, associada ao estado de atenção. Os dados estudados referem-se ao reconhecimento de objetos espaciais por pessoas videntes e cegas, de forma a entender quais áreas são ativadas durante o reconhecimento. Para análise dos dados, são utilizadas técnicas de classificação, dando destaque aos algoritmos de Árvores de Decisão J48 e Random Tree, para a extração de outros tipos de padrões e informações. Os resultados obtidos sugerem que a modulação da amplitude para a Banda Beta não apresenta resultados melhores do que a banda completa, nem relacionado a confirmar ou a negar as hipóteses, nem relacionado a ter um percentual de acerto maior dos algoritmos de classificação.

Palavras-chave: Neurociências, Sinais Cerebrais, Mineração de Dados, Árvores de decisão.

ABSTRACT

MARQUES, Bianca Parulla. **DATA MINING AND NEUROSCIENCE: A STUDY OF CEREBRAL SIGNS BASED ON MODULATION OF AMPLITUDE..** 2019. 71 f. Dissertação (Mestrado) – Programa de Pós-Graduação em Modelagem Computacional. Ministério da Educação, Rio Grande.

Neuroscience is the study of the nervous system and the brain being is the main component. Associated with the BCI (Brain Computer Interface) Systems, it is possible to communicate with computers. One of the ways of collecting brain signals is from electroencephalography (EEG). Brain waves can be classified using their frequency, amplitude, shape and position on the scalp. EEG applications focus on a range of 0.1 Hz to 100 Hz, where EEG signals are classified based on their frequency range and are related between the waves: Alpha, Beta, Theta, Range and Delta. This work focuses on modulating the amplitude of the Beta Band frequency range, which is the range of brain waves that lies between 13 Hz and 30 Hz, associated with attention state. The data studied refer to the recognition of spatial objects by sighted and blind people, in order to understand which areas are activated during recognition. For data analysis, classification techniques are used, such as the algorithms of Decision Tree J48 and Random Tree, for the extraction of other types of patterns and information. The results suggest that the amplitude modulation for the Beta Band does not present better results than the complete band, neither related to confirm or to deny the hypothesis, nor related to having a higher accuracy percentage of the classification algorithms.

Keywords: Neuroscience, Brain Signals, Data Mining, Decision Trees.

LISTA DE FIGURAS

Figura 1	Lobos dos hemisférios cerebrais e suas funções específicas	17
Figura 2	Ondas Delta (δ)	18
Figura 3	Ondas Theta (θ)	18
Figura 4	Ondas Alpha (α)	18
Figura 5	Ondas Beta (β)	19
Figura 6	Processo de Descoberta de Conhecimento em Banco de Dados (KDD)	20
Figura 7	Modelo de uma Árvore de Decisão	25
Figura 8	Modelo de uma Matriz de Confusão	26
Figura 9	Metodologia Proposta.	31
Figura 10	Autômato para aplicação da modulação da banda Beta e conversão dos dados para CSV.	32
Figura 11	Módulo <i>Temporal Filter</i>	33
Figura 12	Módulo <i>Simple DSP</i>	33
Figura 13	Dados brutos.	34
Figura 14	Dados após a rotulação das classes.	34
Figura 15	Modelo de dados no formato ARFF.	35
Figura 16	Configuração da tarefa de classificação para o algoritmo J48.	37
Figura 17	Configuração da tarefa de classificação para o algoritmo Random Tree.	38
Figura 18	Hierarquia dos testes realizados com a ferramenta Weka (BASTOS, 2017).	38
Figura 19	Configuração do algoritmo J48.	39
Figura 20	Configuração do algoritmo Random Tree.	40
Figura 21	Árvore gerada a partir do algoritmo Random Tree com os dados da banda completa do Sujeito 1.	44
Figura 22	Árvore gerada a partir do algoritmo J48 com os dados da modulação banda Beta do Sujeito 1.	45
Figura 23	Árvore gerada a partir do algoritmo Random Tree com os dados da modulação banda Beta do Sujeito 1.	46
Figura 24	Árvore gerada a partir do algoritmo Random Tree com os dados da banda completa do Sujeito 2.	50
Figura 25	Árvore gerada a partir do algoritmo J48 com os dados da modulação banda Beta do Sujeito 2.	51
Figura 26	Árvore gerada a partir do algoritmo Random Tree com os dados da modulação banda Beta do Sujeito 2.	52
Figura 27	Árvore gerada a partir do algoritmo Random Tree com os dados da banda completa do Sujeito 3.	56

Figura 28	Árvore gerada a partir do algoritmo J48 com os dados da modulação banda Beta do Sujeito 3.	57
Figura 29	Árvore gerada a partir do algoritmo Random Tree com os dados da modulação banda Beta do Sujeito 3.	58
Figura 30	Árvore gerada a partir do algoritmo Random Tree com os dados da banda completa do Sujeito 4.	62
Figura 31	Árvore gerada a partir do algoritmo J48 com os dados da modulação banda Beta do Sujeito 4.	63
Figura 32	Árvore gerada a partir do algoritmo Random Tree com os dados da modulação banda Beta do Sujeito 4.	63

LISTA DE TABELAS

Tabela 1	Áreas cerebrais, eletrodos e suas funções (BASTOS, 2017).	17
Tabela 2	Trabalhos Relacionados	29
Tabela 3	Resultados gerados através dos testes realizados para os dados do Sujeito 1.	43
Tabela 4	Áreas ativadas do Sujeito 1.	47
Tabela 5	Resultados gerados através dos testes realizados para os dados do Sujeito 2.	48
Tabela 6	Áreas ativadas do Sujeito 2.	52
Tabela 7	Resultados gerados através dos testes realizados para os dados do Sujeito 3.	54
Tabela 8	Áreas ativadas do Sujeito 3.	59
Tabela 9	Resultados gerados através dos testes realizados para os dados do Sujeito 4.	60
Tabela 10	Áreas ativadas do Sujeito 4.	64
Tabela 11	Comparação entre algoritmos utilizando a faixa da banda completa.	64
Tabela 12	Comparação entre algoritmos utilizando a faixa da banda Beta.	65
Tabela 13	Comparação entre algoritmo e banda.	65
Tabela 14	Resultados das instâncias classificadas corretamente com o algoritmo Random Forest para os dados do Sujeito 1.	66
Tabela 15	Resultados das instâncias classificadas corretamente com o algoritmo Random Forest para os dados do Sujeito 2.	66
Tabela 16	Resultados das instâncias classificadas corretamente com o algoritmo Random Forest para os dados do Sujeito 3.	66
Tabela 17	Resultados das instâncias classificadas corretamente com o algoritmo Random Forest para os dados do Sujeito 4.	66

LISTA DE ABREVIATURAS E SIGLAS

EEG	Eletroencefalograma
BCI	Brain Computer Interface
DCBD	Descoberta de Conhecimento de Banco de Dados
KDD	Knowledge Discovery in Databases
MD	Mineração de Dados
ERD	Dessincronização Relacionada a Eventos
ERS	Sincronização Relacionada a Eventos
GDF	Graph Data Format
CSV	Comma-Separated Value

SUMÁRIO

1	INTRODUÇÃO	12
1.1	Objetivos	13
1.2	Estrutura do Texto	13
2	REFERENCIAL TEÓRICO	14
2.1	Sistema Nervoso e Neurociência	14
2.1.1	Sistema Nervoso: Aprendizado do encéfalo	14
2.1.2	Neurociências: Divisão e Funções do encéfalo	15
2.1.3	Padrões de Atividade Cerebral	18
2.2	Descoberta de Conhecimento e Mineração de Dados	19
2.2.1	Seleção dos Dados	20
2.2.2	Tipos de Dados	20
2.2.3	Pré-Processamento de Dados	21
2.2.4	Aprendizado Supervisionado e Não-Supervisionado	23
2.2.5	Árvores de Decisão	24
3	TRABALHOS RELACIONADOS	27
4	METODOLOGIA	30
4.1	Hipótese	31
4.2	Pré-processamento dos dados	31
4.2.1	Limpeza dos dados e aplicação dos rótulos	34
4.2.2	Transformação dos dados	35
4.3	Escolha das técnicas de MD	36
4.4	Aplicação e execução dos algoritmos	37
5	ANÁLISE E VALIDAÇÃO DOS RESULTADOS	41
5.1	Sujeito 1 - Vidente	41
5.2	Sujeito 2 - Vidente	47
5.3	Sujeito 3 - Cego	52
5.4	Sujeito 4 - Cego	59
5.5	Validação dos Resultados Obtidos	64
6	CONCLUSÕES E TRABALHOS FUTUROS	67
	REFERÊNCIAS	69

1 INTRODUÇÃO

A neurociência é o estudo científico do sistema nervoso, onde abrange seu funcionamento, desenvolvimento e estrutura (LENT, 2008). O encéfalo, sendo o principal componente do sistema nervoso, desenvolve os processos mentais como concentração, pensamento, capacidade de aprendizado e controle motor, praticados por meio dos neurônios, que atualmente podem ser explicados pela neurociência (COSENZA; GUERRA, 2011).

O eletroencefalograma (EEG) é baseado em registros de atividade elétrica cerebral medidos na superfície do escalpo (EYSENCK; KEANE, 2007). A eletroencefalografia tem sido o método mais utilizado para aquisição de sistemas BCI (Brain Computer Interface), baseado na captação e análise de sinais cerebrais, através de dados EEG (SILVEIRA, 2013). Como exemplo, para a captura de sinais EEG, é a inserção de uma touca na cabeça do indivíduo, e nela são acoplados os eletrodos, que facilmente, fazem o contato e enviam os estímulos cerebrais para a máquina, computador ou prótese.

As ondas cerebrais podem ser classificadas usando sua frequência, amplitude, forma e posição no couro cabeludo. As aplicações EEG se concentram em uma faixa relativamente estreita, a partir de 0,1 Hz a 100 Hz. Sinais de EEG são classificados com base em sua faixa de frequência e estão relacionados entre as ondas: Alfa, Beta, Teta, Gama e Delta (MUNIAN; SIVALINGAM; JAYARAMAN, 2014).

A Mineração de Dados é o processo de descoberta de informações em grandes depósitos de dados, onde se utilizam técnicas para extrair padrões úteis desses dados (TAN et al., 2009). Essas técnicas e ferramentas tornaram-se instrumentos de estudo da área de pesquisa denominada Descoberta de Conhecimento de Banco de Dados (DCBD), ou do inglês KDD (Knowledge Discovery in Databases). Esta área se tornou interdisciplinar, com interesse de diversos campos de pesquisa, como inteligência artificial, reconhecimento de padrões, estatística, aprendizado de máquina entre outros (FAYYAD, 1996).

Segundo FAYYAD (1996), KDD é o “processo, não trivial, de extração de informações implícitas, previamente desconhecidas e potencialmente úteis, a partir dos dados armazenados em um banco de dados”. Este trabalho foca na etapa de mineração de dados, que é um componente absoluto do KDD e, em geral, é a conversão de dados brutos

em informações úteis (TAN et al., 2009).

O foco desse trabalho é usar técnicas de mineração de dados na banda beta dos sinais cerebrais de EEG, que é associada ao estado de atenção, precisão visual e coordenação, para a descoberta de padrões. Foram utilizadas as técnicas de classificação: Árvores de Decisão J48 e Random Tree. A base de dados de sinais cerebrais analisada foi a mesma utilizada no trabalho de BASTOS (2017), onde foi aplicada a técnica de classificação de árvores de decisão J48 na faixa de frequência da banda completa.

1.1 Objetivos

O objetivo principal deste trabalho é identificar as principais áreas de ativação do encéfalo a partir da descoberta de padrões utilizando algoritmos de mineração de dados.

Para atingir o objetivo geral, os demais objetivos específicos devem ser alcançados:

- Encontrar um modelo que represente o processamento do encéfalo a partir da classificação dos sinais da banda Beta de EEG usando o algoritmo de árvores de decisão J48;
- Encontrar um modelo que represente o processamento do encéfalo a partir da classificação dos sinais da banda Beta de EEG usando o algoritmo Random Tree;
- Encontrar um modelo que represente o processamento do encéfalo a partir da classificação dos sinais da banda completa de EEG usando o algoritmo Random Tree;
- Comparar os modelos obtidos com o trabalho BASTOS (2017).

1.2 Estrutura do Texto

O texto está estruturado em 6 capítulos, disposto da seguinte maneira: capítulo 1 Introdução, capítulo 2 Referencial Teórico, no capítulo 3 são apresentados os Trabalhos Relacionados, a Metodologia é abordada no capítulo 4, Análise e Validação dos Resultados no capítulo 5 e no capítulo 6 as Conclusões e Trabalhos Futuros.

2 REFERENCIAL TEÓRICO

Neste capítulo é apresentado o referencial teórico necessário para o desenvolvimento do trabalho proposto.

2.1 Sistema Nervoso e Neurociência

2.1.1 Sistema Nervoso: Aprendizado do encéfalo

O encéfalo é o principal componente do sistema nervoso, é por meio do seu funcionamento que somos capazes de aprender e de modificar nosso comportamento a medida do tempo. Processos mentais como o pensamento, capacidade de julgamento e a atenção, são resultados da sua ação (COSENZA; GUERRA, 2011). Capacidades estas que só são possíveis por meio de circuitos neurais, os neurônios. Um neurônio dispara impulsos constantemente, e as informações só podem ser transmitidas para uma outra célula através do axônio, que é o prolongamento neural. As sinapses são os locais onde são feitas essas comunicações, através da liberação de uma substância química chamada neurotransmissor. As sinapses, na aprendizagem têm uma grande relevância, pois são os locais que regulam a passagem de informações no sistema nervoso (COSENZA; GUERRA, 2011).

De acordo com COSENZA; GUERRA (2011), pode-se afirmar que “todos temos vias motoras e sensoriais que seguem o mesmo padrão” e que estas vias “estão previstas nas informações genéticas de nossas células e são construídas enquanto nosso organismo se desenvolve dentro do útero materno”, mas no que entende-se sobre a nossa espécie, não existem dois encéfalos iguais.

O desenvolvimento do sistema nervoso é iniciado nas primeiras semanas de vida embrionária, no formato de um tubo constituído por uma parede de células-tronco, onde darão origem aos neurônios e as células auxiliares, denominadas gliais. Nesta primeira fase, a ocorrência mais importante é divisão das células-tronco constantemente, que permite a formação de novos neurônios que, irão aumentar consideravelmente seu pequeno número inicial, chegando a bilhões em um prazo de tempo curto. A medida que vão ficando mais espessas as paredes do tubo, na região que dará origem ao encéfalo, os novos neurônios deslocam-se ocupando os espaços predestinados geneticamente.

Completada a fase de diferenciação, a conexão com outros neurônios começa a se iniciar para que sejam criados circuitos relevantes para a execução de distintas funções. Portanto os axônios, vão abrindo caminho, desviando-se de obstáculos, até que tudo esteja no seu devido lugar. No procedimento de estruturação do encéfalo, um número bem maior que o fundamental de neurônios são formados, e muitos descartados por não atingirem as ligações corretas.

Por fim, os neurônios convenientemente localizados, emitem prolongamentos e tem seus axônios nos lugares definidos, ocorrendo então a sinaptogênese, que é a formação das sinapses que irão completar os circuitos nervosos. Essas primeiras fases do desenvolvimento do sistema nervoso são essenciais para que se estabeleçam depois as funções que as diversas estruturas vão desempenhar (COSENZA; GUERRA, 2011).

O sistema nervoso possui uma característica importante, que é a sua capacidade de fazer e desfazer ligações entre os neurônios durante toda a vida, chamada de plasticidade.

Em conformidade com COSENZA; GUERRA (2011), segundo a neurociência cognitiva, aprendizagem e memória estão fortemente relacionadas, sendo a memória a base orgânica para que ocorra a aprendizagem. Memorizar não é o mesmo que aprender, mas para que ocorra aprendizagem, é preciso que as informações fiquem armazenadas na nossa memória. Para IZQUIERDO (2004): “A memória é o processo pelo qual aquilo que é aprendido persiste ao longo do tempo”.

LENT (2010) diz que a capacidade de armazenar informações e utilizá-las posteriormente está relacionada a memória, e a aprendizagem é apenas o processo de aquisição das informações. CARVALHO (2007), fundamentada em LENT (2001), destaca que a aprendizagem é decisiva para o pensamento, o qual deriva da capacidade de lidar com as informações das áreas de associação motora, sensorial e mnemônica. Já o pensamento é o referencial para a orientação do comportamento, pois seu processamento envolve o recebimento, a percepção, a compreensão, o armazenamento, a manipulação, o monitoramento e o controle essenciais para lidar com o fluxo constante de dados objetivando planejar ações.

2.1.2 Neurociências: Divisão e Funções do encéfalo

A neurociência é o estudo científico do sistema nervoso, abrangendo seu funcionamento, desenvolvimento e estrutura (LENT, 2008). Por convenção é uma ciência da área da biologia, mas por contribuir para áreas como medicina, educação, computação, tornou-se interdisciplinar.

O encéfalo sendo o principal componente do sistema nervoso, desenvolve os processos mentais como concentração, pensamento, capacidade de aprendizado e controle motor, praticados por meio dos neurônios, que atualmente podem ser explicados pela neurociência.

O encéfalo humano é dividido em dois hemisférios, direito e esquerdo como os quais

apresentam muitas dobraduras denominadas circunvoluções, giros ou folhas, que são separados por sulcos ou fissuras. Cada hemisfério é responsável por um conjunto de funções, que acabam por trabalhar juntos (GAZZANIGA; HEATHERTON, 2005).

Estudos destacam especialidades nos hemisférios, que podem ser bem distintas, mas não são exclusividades funcionais. De forma geral, o hemisfério direito controla a parte esquerda do corpo e vice-versa. Aparentemente, os hemisférios são iguais, mas se diferem em alguns aspectos como as funcionalidades que de certa maneira são definidas geneticamente, aprimoradas na maturação com a interação com o meio, derivando assim as especializações de cada hemisfério (VALLADO et al., 2004).

Segundo LENT (2001), o hemisfério esquerdo é melhor na realização mental de cálculos matemáticos, no comando da escrita e na compreensão dela através da leitura, já o hemisfério direito é melhor na percepção de sons musicais e reconhecimento de faces. Apesar de o hemisfério esquerdo ser melhor na realização mental de cálculos matemáticos, o hemisfério direito é melhor no reconhecimento de relações espaciais quantitativas, que são úteis para nosso deslocamento. Já hemisfério esquerdo participa dessa função, mas com a habilidade no reconhecimento de relações espaciais qualitativas. Embora cada hemisfério tenha sua especialidade, eles muitas vezes se complementam, assim como o hemisfério direito é especialmente capaz de identificar categorias gerais de objetos e seres vivos, o esquerdo é que detecta as categorias específicas.

Além da divisão do encéfalo em dois hemisférios, anatomistas agrupam o encéfalo em grandes regiões denominadas Lobos, dos quais seus limites nem sempre são exatos, mas passam uma ideia de localização de cada região dos lobos. Os lobos são classificados em quatro visíveis externamente e um posicionado no interior dos grandes sulcos do encéfalo, o sulco lateral (LENT, 2010) totalizando em cinco.

Os quatro lobos visíveis são: frontal que é relacionado com a frente, parietal que se localiza sob o osso craniano do mesmo nome, temporal relacionado com a têmpora, occipital sob o osso craniano homônimo (GAZZANIGA; HEATHERTON, 2005) e o lobo da ínsula, que só pode ser visto no momento em que se abre o sulco lateral (LENT, 2008).

Cada lobo tem características específicas, como mostra a Figura 1. Com base nos estudos de GAZZANIGA; HEATHERTON (2005), o lobo temporal é a região mais ventral do córtex cerebral, que é importante para processar as informações auditivas e também para a memória; o lobo occipital é uma região do córtex cerebral na parte posterior do encéfalo que é destinada quase que exclusivamente para o sentido da visão; o lobo parietal fica em frente aos lobos occipitais e atrás dos lobos frontais, que é importante para o sentido do tato e para a noção espacial de um ambiente; e por fim, o lobo frontal, é a região anterior do córtex cerebral envolvida no planejamento e no movimento.

Durante a coleta de sinais EEG, no escalpo do couro cabeludo são dispostos eletrodos, que capturam dados de cada lobo. Segundo BASTOS (2017), com o auxílio da Prof^a. Dr^a. Fernanda Antoniolo Hammes de Carvalho, que é especialista nesta área, baseando-se na

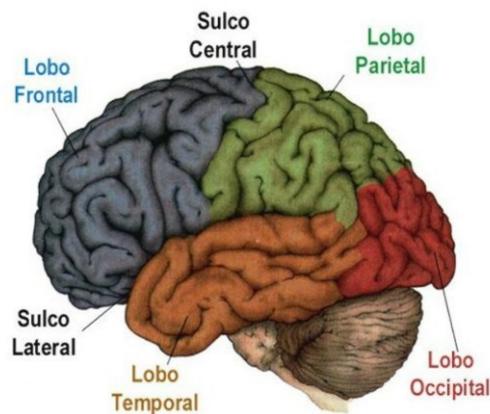


Figura 1: Lobos dos hemisférios cerebrais e suas funções específicas (GAZZANIGA; HEATHERTON, 2005).

literatura em neurociências formulou a Tabela 1, em que apresenta as regiões cerebrais, bem como os eletrodos espacialmente posicionados, referentes as suas funcionalidades.

Tabela 1: Áreas cerebrais, eletrodos e suas funções (BASTOS, 2017).

Regiões	Eletrodos	Funções
Lobo Frontal	FP1, FP2, FZ, F7, F3, F4, F8, FT9, FT10, FC5, FC1, FC2, FC6	Funções executivas (gerenciamento dos recursos cognitivos/emocionais diante de uma dada tarefa).
Lobo Temporal	T7, TP9, T8, TP10	Percepção de movimentos biológicos.
Lobo Parietal	P7, P3, PZ, P4, P8, C3, CZ, C4, CP1, CP2, CP6, CP5	Percepção somatossensorial, representações espaciais e percepções táteis.
Lobo Occipital	O1, OZ, O2	Visão (Inclusive durante um diálogo).

2.1.3 Padrões de Atividade Cerebral

As ondas cerebrais podem ser classificadas usando sua frequência, amplitude, forma e posição de eletrodos no couro cabeludo. As aplicações de EEG (Eletroencefalograma) se concentram em uma faixa relativamente estreita, a partir de 0,1 Hz a 100 Hz. Os sinais EEG são classificados com base em sua faixa de frequência (MUNIAN; SIVALINGAM; JAYARAMAN, 2014).

Ondas Delta (δ): estão na faixa de frequência de 0.5 a 3.5 Hz. São as ondas mais lentas em comparação com outras, geralmente ocorrem em sono profundo e, às vezes, quando acordadas e também em estado mental em coma.



Figura 2: Ondas Delta (δ)
(WEBSTER, 2009).

Ondas Theta (θ): estão na faixa de frequência de 3.5 a 7.5 Hz. Surgem durante o pensamento criativo, estressado e profundo estado de hipnose.

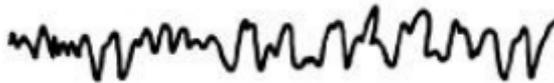


Figura 3: Ondas Theta (θ)
(WEBSTER, 2009).

Ondas Alpha (α): estão na faixa de frequência de 7.5 a 12 Hz. Dominam em estados mentais relaxados e calmos.



Figura 4: Ondas Alpha (α)
(WEBSTER, 2009).

Ondas Beta (β): estão na faixa de frequência de 13 a 30 Hz. Associado ao estado de atenção, precisão visual e coordenação.



Figura 5: Ondas Beta (β)
(WEBSTER, 2009).

Ondas Gamma (γ): estão na faixa de frequência acima de 30 Hz. Funções motoras, trabalho simultâneo e outras tarefas múltiplas ocorrem nesta faixa de frequência (MUNIAN; SIVALINGAM; JAYARAMAN, 2014).

As ondas MU (μ) está associada a atividades motoras, e também é encontrada na faixa de frequência da onda Alfa, mas onde a amplitude máxima é registrada sobre o córtex motor, então basicamente, desencadeia quando há um movimento real ou há uma intenção de se mover (MUNIAN; SIVALINGAM; JAYARAMAN, 2014).

2.2 Descoberta de Conhecimento e Mineração de Dados

Mineração de Dados é o processo de descoberta de informações em grandes depósitos de dados, onde se utilizam técnicas para extrair padrões úteis desses dados (TAN et al., 2009).

Com o avanço da tecnologia de coleta e armazenamento de dados, foi-se acumulando uma grande quantidade de dados, então descobriu-se necessário a combinação de métodos tradicionais de análise de dados com algoritmos aprimorados a fim de extrair conhecimentos importantes desses dados. Essas técnicas e ferramentas tornaram-se instrumentos de estudo e pesquisa denominada Descoberta de Conhecimento de Banco de Dados (DCBD) ou, do inglês, KDD (Knowledge Discovery in Databases) e tornou-se interdisciplinar, com interesse de diversos campos de pesquisa, como inteligência artificial, reconhecimento de padrões, estatística, aprendizado de máquina entre outros (FAYYAD, 1996).

FAYYAD (1996) diz que KDD é o “processo, não trivial, de extração de informações implícitas, previamente desconhecidas e potencialmente úteis, a partir dos dados armazenados em um banco de dados”. A mineração de dados é um componente absoluto do KDD, e em geral, é a conversão de dados brutos em informações úteis (TAN et al., 2009), conforme ilustra a Figura 7.

O processo de KDD pode ser interativo e iterativo. Interativo, pois o usuário pode intervir e controlar o curso das atividades. Iterativo, por ser uma sequência finita de operações, onde o resultado de cada uma é dependente dos resultados das que a precedem (PRASS, 2004).

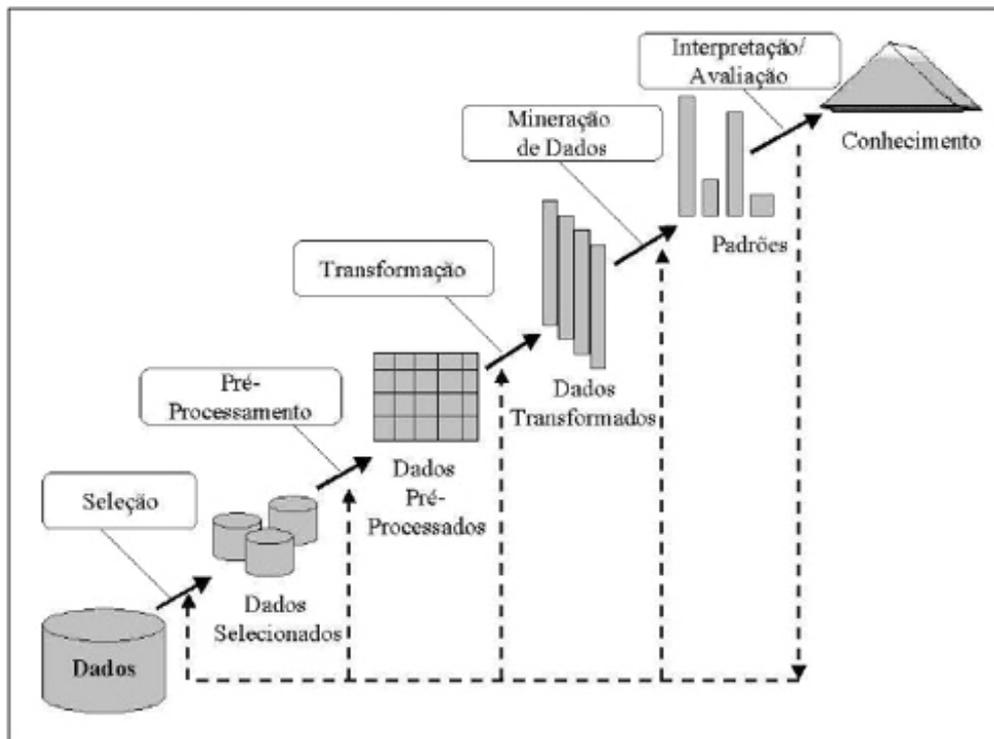


Figura 6: Processo de Descoberta de Conhecimento em Banco de Dados (KDD) (FAYYAD, 1996).

2.2.1 Seleção dos Dados

Os dados devem ser selecionados, previamente conhecidos, explorados e preparados para definir o método mais adequado de aplicação. Existem dois tipos de categorização dos dados, que podem ser quantitativos e qualitativos.

- Quantitativos: valores numéricos e podem ser discretos e contínuos.
- Qualitativos: valores nominais e ordinais e são categóricos.

A maior parte das técnicas de mineração de dados admitem falhas nos dados, mas questões de qualidade geralmente melhoram as análises resultantes, como a eliminação de ruídos, inconsistência ou duplicação dos dados. Essa técnica é descrita como pré-processamento dos dados, ou seja, os dados são analisados e preparados de modo que se adaptem melhor a uma ferramenta ou técnica específica de mineração de dados (TAN et al., 2009). As técnicas de pré-processamento serão melhor definidas na Seção 2.4.3.

2.2.2 Tipos de Dados

De acordo com TAN et al. (2009), um conjunto de dados pode ser uma coleção de objetos de dados. Objeto de dados podem ser chamados de *registros*, *ponteiros*, *vetores*, *padrões*, *eventos*, *casos*, *exemplos*, *observações* ou *entidades*, e por sua vez, são especificados por um número de atributos que contém suas características básicas. Os atributos

também podem ser conhecidos como *variáveis*, *características*, *campo*, *recurso* ou *dimensão*, e são descritos como uma propriedade ou característica de um objeto.

2.2.3 Pré-Processamento de Dados

O pré-processamento dos dados melhora a qualidade dos dados para que se tornem adequados para a aplicação de técnicas de mineração e, até mesmo, se adaptem melhor para com o uso de alguma ferramenta. Estas técnicas podem ser agrupadas nos seguintes grupos de tarefas:

- Eliminação manual de atributos;
- Integração de dados;
- Amostragem de dados;
- Balanceamento de dados;
- Limpeza de dados;
- Redução de dimensionalidade;
- Transformação de dados.

As técnicas de pré-processamento são frequentemente utilizadas para melhorar a qualidade dos dados por meio da eliminação ou minimização dos problemas citados (FACELI et al., 2011).

Eliminação Manual de Atributos: Atributos visivelmente considerados irrelevantes, como um atributo que não possui informação que diferencie os objetos, tenha o mesmo valor para todos os objetos.

Integração de Dados: Os dados devem ser integrados quando devem ser utilizados na mesma aplicação e estão distribuídos em conjuntos de dados distintos. Na integração, é essencial identificar os objetos presentes nos distintos conjuntos a serem combinados, problema esse conhecido como problema de identificação de entidade. Para a minimização do problema, é comum o uso de metadados na base de dados, que são dados sobre dados, que ao caracterizar as principais características dos dados, podem ser empregados para evitar erros no processo de integração. Por sua vez origina-se um grande repositório de dados, denominado *data warehouse*, que atua como uma base de dados centralizada.

Amostragem de Dados: Grande quantidade de objetos podem gerar problemas por razão de saturação de memória e aumento do tempo computacional, sendo preciso ajustar os parâmetros do modelo. Então, deve haver o equilíbrio entre a eficiência

computacional e acurácia do modelo (taxa de predições corretas). Quanto maior o número de dados utilizados, maior é a acurácia do modelo e menor sua eficiência computacional, pois o grande número de objetos pode tornar extenso o processamento. Para balancear eficiência e acurácia, seleciona-se então, uma amostra representativa ou um subconjunto dos dados.

Balanceamento de Dados: Os dados são considerados desbalanceados quando o número de objetos varia para diferentes classes, considerando assim, classes majoritárias contendo maior número de objetos, e classes minoritárias com menor número de objetos, problema típico da área de classificação de dados. Para ser permitida, a acurácia preditiva de um classificador para dados desbalanceados, deve ser maior que a acurácia obtida atribuindo todo novo objeto à classe majoritária. Algoritmos sustentados com dados desbalanceados tendem a favorecer a classificação de novos objetos na classe majoritária. Para lidar com o problema de desbalanceamento dos dados, são utilizadas as seguintes técnicas:

- Redefinir o tamanho do conjunto de dados, acrescentando objetos na classe minoritária ou eliminando objetos na classe majoritária;
- Empregar custos distintos de classificação para as diferentes classes;
- Induzir um modelo para uma classe.

Para o acréscimo de objetos novos, existe o risco destes demonstrarem situações que nunca ocorrerão, induzindo um modelo inadequado para os dados, além de que pode ocorrer um problema conhecido como *overfitting*, onde o modelo é superajustado aos dados de treinamento. Já para a eliminação de dados da classe majoritária, dados relevantes podem ser perdidos, podendo levar ao problema de *underfitting*, em que o modelo induzido não se ajusta aos dados de treinamento.

Limpeza de Dados: O conjunto de dados também pode apresentar dificuldades para com a qualidade dos dados, como ruídos, inconsistência, redundância ou quando estão incompletos:

- Dados ruidosos possuem erros ou valores diferentes do esperado, como a presença de *outliers*, que são valores que estão além dos limites aceitáveis, ou muito distintos dos demais valores observados para o mesmo atributo;
- Dados inconsistentes são os que possuem valores incompatíveis em seus atributos, como quando atributos tem os dados de entrada iguais e a saída é diferente.
- Dados redundantes, podem ser atribuídos aos objetos quando um objeto é muito semelhante a outro do mesmo conjunto de dados, ou seja, seus atributos possuem valores muito semelhantes ou iguais e atribuídos aos atributos

quando o valor do atributo para todos objetos pode ser deduzido a partir do valor de um ou mais atributos.

- Dados incompletos são ausências de valores para alguns atributos de alguns objetos, como quando:
 - Atributo não é considerado importante;
 - Valor desconhecido;
 - Irrelevância da informação;
 - Inexistência de um valor;
 - Problema com o equipamento de coleta, transmissão ou armazenamento.

Redução de Dimensionalidade: Para que dados com um grande número de atributos possam ser utilizados por aplicações de alguns algoritmos, a quantidade de atributos deve ser significativamente reduzida de forma a melhorar o desempenho do modelo induzido e reduzir do custo computacional da aplicação. Para solucionar o problema, existem diferentes técnicas como *Reconhecimento de Padrões, Estatística e Teoria da Informação*. Essas técnicas podem ser divididas em duas abordagens:

- Agregação, onde faz-se a combinação dos atributos originais por meio de funções lineares e eliminam-se as redundâncias;
- Seleção de atributos, onde os atributos mais importantes são identificados e faz-se então a eliminação de atributos irrelevantes e redução de ruídos, facilitando assim a visualização dos dados e reduzindo o tempo de processamento.

Transformação de Dados: Algumas técnicas de aprendizagem de máquina têm limitação quanto a manipulação de valores de determinados tipos, como aceitar somente tipo numérico ou somente tipo simbólico. Para solucionar este problema, existem três abordagens diferentes:

- Conversão Simbólico-Numérico;
- Conversão Numérico-Simbólico;
- Transformação de atributos numéricos.

2.2.4 Aprendizado Supervisionado e Não-Supervisionado

Os métodos de mineração de dados são divididos em aprendizado supervisionado (preditivo) e não-supervisionado (descritivo). O método supervisionado utiliza um atributo alvo para induzir o modelo, onde são considerados os principais métodos preditivos as regras de classificação e regressão. Já os métodos não-supervisionados não precisam de uma pré-categorização para os registros, ou seja, não é necessário um atributo alvo, sendo os principais métodos descritivos as regras de associação e agrupamento (clustering) (CAMILO; SILVA, 2009).

2.2.4.1 Métodos de Aprendizado Supervisionado

Classificação: Classificação é o processo de encontrar um conjunto de modelos (funções) que descrevem e distinguem classes ou conceitos, com o propósito de utilizar o modelo para prever a classe de objetos que ainda não foram classificados. O modelo construído baseia-se na análise prévia de um conjunto de dados de amostragem ou dados de treinamento, contendo objetos corretamente classificados (DE AMO, 2004). Árvores de decisão, algoritmos genéticos, redes bayesianas, vizinhos mais próximos, entre outros, são algoritmos muito utilizados para classificação.

Regressão ou Estimação: A regressão é semelhante à classificação, no entanto apropriada para registro identificado somente por valor numérico e não valor categórico. Estimando-se assim o valor de uma certa variável analisando os valores das demais (CAMILO; SILVA, 2009).

Predição: A tarefa de predição é similar às tarefas de classificação e regressão, porém ela visa descobrir o valor futuro de um determinado atributo (CAMILO; SILVA, 2009).

2.2.4.2 Métodos de Aprendizado Não-Supervisionado

Associação: As regras de associação representam combinações de itens mais frequentes em uma base de dados (GONÇALVES, 2005). É uma das tarefas mais conhecidas devido aos resultados satisfatórios (CAMILO; SILVA, 2009).

Agrupamento (Clustering): Um agrupamento (ou cluster) é uma coleção de registros similares entre si, mas distintos dos outros registros nos demais agrupamentos. Esta tarefa difere da classificação pois não necessita que os registros sejam previamente categorizados (aprendizado não-supervisionado) e não tem a intenção de classificar, estimar ou prever o valor de uma variável, ela apenas identifica os grupos de dados similares (CAMILO; SILVA, 2009) (PRASS, 2004).

2.2.5 Árvores de Decisão

As Árvores de Decisão estão dentro do método supervisionado de classificação, segundo MONARD; BARANAUSKAS (2003), as árvores de Decisão são uma estrutura de dados definida recursivamente como:

- um nó folha que corresponde a uma classe; ou
- um nó de decisão que contém um teste sobre algum atributo. Para cada resultado do teste existe uma aresta para uma subárvore. Cada subárvore tem a mesma estrutura que a árvore.

A Figura 8 demonstra um exemplo de árvore de decisão simbólica, onde as variáveis 'N' são os nós de decisão, para cada nó associa-se um atributo, as variáveis 'a' e 'b' representam os limites dos atributos que dividem a decisão nos caminhos da árvore, estes podem ser nominais ou numerais e as variáveis 'Classe' representam as folhas da árvore, que mostram a classificação do objeto.

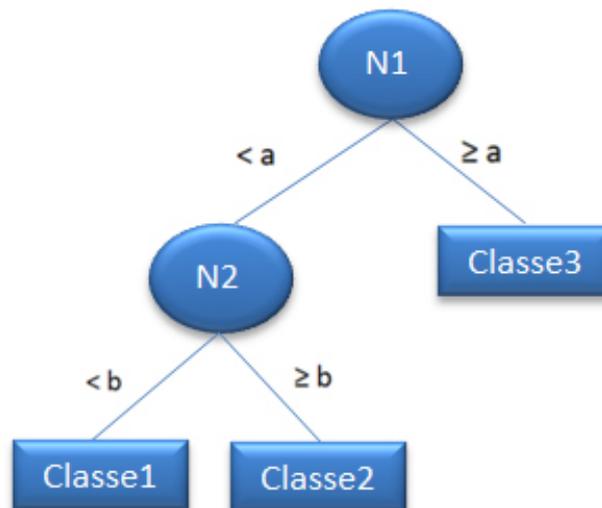


Figura 7: Modelo de uma Árvore de Decisão

- **Matriz de Confusão**

Segundo TAN et al. (2009), a classificação do desempenho de um modelo é fundamentado nas contagens de registros de testes previstos corretamente e incorretamente pelo modelo.

A matriz de confusão de uma hipótese oferece uma medida efetiva do modelo de classificação, ao mostrar o número de classificações corretas versus as classificações preditas para cada classe, sobre um conjunto de exemplos (MONARD; BARANAUSKAS, 2003). A Figura 9 mostra um exemplo de uma matriz de confusão, onde a diagonal principal com os valores em amarelo são as classes classificadas corretamente e os outros valores são as classes classificadas incorretamente.

Algoritmo J48: O algoritmo J48 tem a finalidade de gerar uma árvore de decisão baseada em um conjunto de dados de treinamento, para classificar as instâncias no conjunto de teste. Para a montagem da árvore, o algoritmo J48 utiliza a abordagem de dividir-para-conquistar, onde um problema complexo é dividido em subproblemas mais simples, aplicando recursivamente a mesma estratégia a cada subproblema, dividindo o espaço definido pelos atributos em subespaços, associando-se a eles uma classe (LIBRELOTTO; MOZZAQUATRO, 2014).

	Classe 0	Classe 1	Classe 2
Classe 0	15	3	0
Classe 1	2	13	3
Classe 2	5	1	12

Figura 8: Modelo de uma Matriz de Confusão

Algoritmo Random Tree: Random Tree é uma árvore induzida aleatoriamente a partir de um conjunto de árvores possíveis, usando m atributos aleatórios em cada nó, cada árvore tem uma chance igual de ser amostrada. Podem ser geradas eficientemente e a combinação de grandes conjuntos de *Random Trees* geralmente leva a modelos precisos (OSHIRO, 2013).

3 TRABALHOS RELACIONADOS

Este capítulo apresenta os trabalhos que mais se aproximam a pesquisa realizada neste trabalho, focando no uso de mineração de dados para sinais cerebrais e também no entendimento da modulação de amplitude.

Importante salientar que após a revisão da literatura, visando encontrar trabalhos com objetivos semelhantes ao tema deste estudo, na área da computação não foram encontrados trabalhos sobre a modulação da amplitude de bandas espectrais para classificação através da mineração de dados. Contudo, existem diversos trabalhos na área da medicina que abordam o estudo da modulação de amplitude de bandas espectrais e diversos na área de computação sobre técnicas de classificação. Os trabalhos considerados mais relevantes para essa pesquisa são apresentados nesse capítulo.

O trabalho de BASTOS (2017) apresenta um estudo sobre metodologia de análise EEG através da mineração de dados. É apresentado um estudo de caso utilizando sinais EEG de indivíduos cegos e videntes durante a execução de uma atividade que estimulou a habilidade espacial. A fim de verificar a hipótese de que videntes utilizam o sentido da visão, mesmo com os olhos vendados, e os cegos utilizam o sentido do tato para identificar objetos geométricos espaciais, foi feito um experimento e através dos dados dos sinais cerebrais gravados foram aplicadas técnicas de árvores de decisão para compreender as principais áreas envolvidas nas atividades cerebrais durante a execução da tarefa de reconhecimento.

Já o trabalho de PFURTSCHELLER et al. (2006) estuda a reatividade dos ritmos EEG (ritmo Mu) em associação com a imaginação da mão direita, mão esquerda, pé e movimento da língua em uma amostra de nove sujeitos saudáveis, utilizando a análise através de sessenta eletrodos. O ritmo Mu (μ) foi dessincronizado em todos sujeitos durante a imagem motora da mão, enquanto isso, um aumento do ritmo Mu (μ) da área da mão, foi observado durante a imagem motora do pé ou da língua na maioria dos sujeitos. A frequência dos componentes mais reativos foi de $11.7 \text{ Hz} \pm 0.4$. Enquanto os componentes dessincronizados eram de banda larga e centralizados em $10.9 \text{ Hz} \pm 0.9$, os componentes sincronizados eram estreitos e exibiam frequências mais altas a $12.0 \text{ Hz} \pm 1.0$. A discriminação entre as quatro tarefas de geração de imagens motoras, com base na classificação

de testes EEG únicos melhoraram quando, além da dessincronização relacionada a eventos (ERD - Event-Related Desynchronization), os padrões de sincronização relacionados a eventos (ERS - Event-Related Synchronization) foram induzidos em pelo menos uma ou duas tarefas. Isso implica que tais fenômenos de EEG podem ser utilizados em uma interface encéfalo-computador (BCI - Brain Computer Interface) de várias classes, operado simplesmente por imagens motoras.

O trabalho de GARCIA et al. (2007) apresenta um método de extração de características presentes em sinais adquiridos sobre o córtex motor, e classifica os estados mentais referentes a essas características, aplicando-os no controle de uma cadeira de rodas robótica. Dentre as muitas técnicas de processamento de sinais conhecidas, foi utilizado Mapas Auto-Organizáveis, uma vez que tem essa técnica tem se mostrado muito eficaz em aplicações referentes à mineração de dados e à extração de característica.

Em VAZ (2016) é apresentada uma avaliação dos principais extratores de características utilizados para classificar ensaios de imagética motora, onde os dados foram obtidos por meio de EEG (eletroencefalografia) e apresentam interferências de provenientes atividades oculares, monitorados pelo EOG (eletrooculografia). Os sinais coletados contêm informações sobre três canais EEG (C3, Cz e C4) e outros três de EOG, visando a classificação imagética motora da mão esquerda e da mão direita. Foi realizado o pré-processamento dos canais utilizando a técnica de Análise de Componentes Independentes (ICA), em conjunto com um limiar de correlação para a remoção de componentes associados a artefatos oculares, após foram avaliadas abordagens para a extração de características como: Árvores Diádica de bandas de Frequência (ADBF), Padrões Espaciais Comuns (CSP), Padrões Espectro-Espaciais Comuns (CSSP), Padrões Esparsos Espectro-Espaciais Comuns (CSSSP), Padrões Espaciais Comuns com Banco de Filtros (FBCSP), Padrões Esparsos Espectro-Espaciais Comuns com Banco de Filtros (FBCSSP), Padrões Esparsos Espectro-Espaciais Comuns com Banco de Filtros (FBCSSSP). Foi considerado a Técnica de Seleção de Características baseada em Informação Mútua (MIFS), para a escolha de atributos mais relevantes para o conjunto de dados escolhido na etapa de classificação e finalmente as Máquinas de Vetores de Suporte (SVM) foram escolhidas para a classificação das intenções dos usuários.

O trabalho de VASCONCELOS et al. (2011) aborda um estudo sobre as respostas intermodais em áreas sensoriais primárias, onde foram realizadas gravações neuronais no córtex somatossensorial primário (S1) e no córtex visual primário (V1) de ratos, à medida que realizavam tarefas com objetos no escuro. Durante a livre exploração de objetos, as respostas V1 e S1 continham quantidades comparáveis de informações sobre a identidade do objeto. Durante a execução de uma tarefa de discriminação tátil, foi mais lento e menos robusto em V1 do que em S1. No entanto, as respostas táteis V1 correlacionaram-se significativamente com o desempenho nas sessões. Em conjunto, os resultados apoiam a noção de que as áreas sensoriais primárias têm uma preferência por

uma determinada modalidade, mas podem se engajar em processamento modal cruzado significativo, dependendo da demanda da tarefa.

Em FOLADOR; ANDRADE (2018) é realizada uma revisão da literatura e identificação de técnicas de mineração de dados que têm sido utilizadas na extração de informações de séries temporais e que resultem do registro de sinais na área da engenharia biomédica e sinais motores da doença de Parkinson. Os resultados da revisão sugerem que os métodos mais utilizados na mineração de dados representados por séries temporais são: Árvore de Decisão, Redes Neurais, Redes Bayesianas e Modelos Ocultos de Markov. Também foi mostrado que existe uma necessidade da popularização do uso destas técnicas na avaliação de séries temporais representando os sinais motores da Doença de Parkinson, em especial o surgimento de aplicações relacionados ao monitoramento prolongado dos sinais motores da doença, e conseqüente o aumento no volume de dados a serem processados.

A Tabela 2 apresenta de forma sumarizada os trabalhos apresentados.

Tabela 2: Trabalhos Relacionados

Autor(es)	Objetivos
BASTOS (2017)	Análise de Sinais EEG através da Mineração de Dados utilizando a técnica de Classificação com Árvores de Decisão J48.
PFURTSCHELLER et al. (2006)	Estudo de tarefas de imagética motora.
GARCIA et al. (2007)	Mineração de dados baseado na extração de características de sinais EEG e classificação dos estados mentais.
VAZ (2016)	Avaliação dos principais extratores de características utilizados para classificar ensaios de imagética motora, por meio de dados EEG.
VASCONCELOS et al. (2011)	Um estudo sobre as respostas intermodais em áreas sensoriais primárias.
FOLADOR; ANDRADE (2018)	Uma revisão da literatura e identificação de técnicas de mineração de dados utilizadas na extração de informações de séries temporais.
Este trabalho	Análise comparativa baseada no trabalho de (BASTOS, 2017), utilizando a modulação de banda espectral e aplicação de técnicas de classificação com árvores de decisão J48 e Random Tree.

4 METODOLOGIA

O trabalho de (BASTOS, 2017) apresentou um estudo de caso utilizando sinais EEG de indivíduos cegos e videntes durante a execução de uma atividade que estimulou a habilidade espacial. A fim de verificar a hipótese de que videntes e cegos utilizam áreas diferentes do cérebro para visualizar objetos espaciais, em videntes, prioritariamente o lobo occipital é ativado e em cegos, prioritariamente, o lobo parietal é ativado. Foi feito um experimento e através dos dados dos sinais cerebrais gravados foram aplicadas técnicas de árvores de decisão com o algoritmo J48 para compreender as principais áreas cerebrais envolvidas durante a execução da tarefa de reconhecimento de objetos geométricos. A metodologia mostrou-se viável, gerando, ao final, modelos de árvores de decisão que permitiram confirmar a hipótese de que videntes e cegos utilizam áreas diferentes do cérebro para “visualizar” objetos espaciais. Em videntes, prioritariamente o lobo occipital é ativado e em cegos, prioritariamente, o lobo parietal é ativado.

BASTOS (2017) utilizou árvores de decisão para análise EEG, portanto foi aplicado somente um algoritmo de árvores de decisão J48, e a análise foi feita a partir da faixa de frequência entre 3.5 e 30 Hz, que permite adquirir as ondas cerebrais: Delta, Teta, Alfa e Beta. Contudo, este trabalho propõe fazer uma análise comparativa dos resultados, com base nas ondas cerebrais da faixa de frequência Beta que vai de 13 a 30 Hz, que está associada ao estado de atenção, bem como aplicar o algoritmo de árvores de decisão Random Tree para avaliar os resultados perante ao J48.

A metodologia utilizada neste trabalho está definida em 5 etapas e é baseada na metodologia proposta no trabalho de (BASTOS, 2017). O objetivo é aplicar técnicas de classificação com algoritmos de árvores de decisão, bem como utilizar a modulação da amplitude da faixa da banda Beta em uma base de dados cerebrais. Para tal, a Figura 10 ilustra o fluxograma da metodologia proposta. A última etapa da metodologia é explicada no capítulo 5.

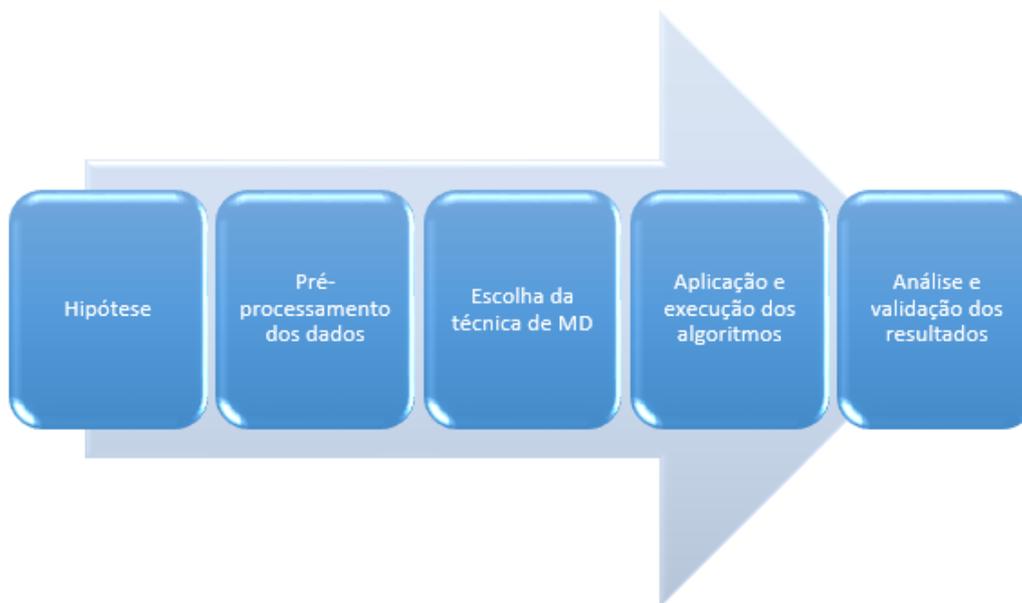


Figura 9: Metodologia Proposta.

4.1 Hipótese

São definidas três hipóteses para este trabalho, que será comparado aos resultados obtidos em (BASTOS, 2017):

- Os resultados de instâncias classificadas corretamente são melhores utilizando o algoritmo Random Tree;
- Os resultados de instâncias classificadas corretamente são melhores utilizando somente a faixa de frequência da banda Beta;
- Os resultados de instâncias classificadas corretamente são melhores utilizando o algoritmo Random Tree e utilizando a faixa de frequência da banda Beta.

4.2 Pré-processamento dos dados

A coleta de sinais cerebrais foi realizada através da eletroencefalografia, um método de coleta de dados que mede a atividade elétrica no encéfalo para ver como ela se relaciona às tarefas cognitivas e perceptivas. Estes dados foram coletados por meio da ferramenta ActiChamp², que é um sistema de amplificação modular, que integra componentes finais para análise eletrofisiológica como EEG, ERP e BCI, em conjunto com o acticap 32 canais, que é uma touca inserida no couro cabeludo do indivíduo, que apresenta 32 eletrodos, os quais são expostos no padrão internacional “10-20”. O acticap é conectado

²<https://brainvision.com/>

a um módulo do amplificador, permitindo a transmissão dos impulsos elétricos captados pelos eletrodos (BASTOS, 2017).

Os dados são armazenados no banco de dados do projeto, onde estão disponíveis para a utilização em pesquisas, tornando possível o acesso para a realização deste trabalho. O pré-processamento dos dados, como já mencionado na seção 2.2.3, é fundamental para deixar os dados preparados e adequados para a aplicação das técnicas, conforme suas necessidades e exigências (TAN et al., 2009).

Os sinais cerebrais foram retirados do banco de dados do projeto, e foram coletados sem a utilização de filtros, sendo gravados diretamente no formato GDF (*Graph Data Format*). Sendo assim, os filtros foram aplicados no momento em que foi feita a aplicação da modulação da faixa de frequência da banda Beta e para melhor leitura dos sinais, e por conseguinte foram convertidos para o formato CSV (*Comma-Separated Value*). Para a aplicação dos filtros, modulação da banda e transformação dos dados, foi utilizada plataforma Openvibe¹ que é uma plataforma de software que permite projetar, testar e utilizar sistemas BCI (PRADA, 2010). A Figura 11 mostra o autômato desenvolvido e utilizado no software para tais ocorrências.

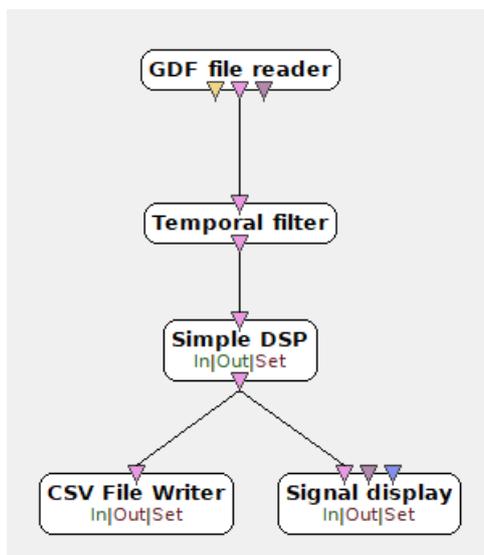


Figura 10: Autômato para aplicação da modulação da banda Beta e conversão dos dados para CSV.

Os autômatos são constituídos por módulos que designam funções. O módulo *GDF file reader* é utilizado para a leitura dos sinais; o módulo *Temporal filter* filtra os sinais EEG brutos, como representado na Figura 12. Inicialmente, os dados foram filtrados pelo método *Butterworth* de quarta ordem. Neste módulo também é definida a frequência que será utilizada para este trabalho, de 13 Hz a 30 Hz, para análise somente da banda Beta. A Figura 13 mostra a caixa *Simple DSP* que é utilizada para aplicar uma fórmula matemática

¹<http://openvibe.inria.fr/>

para cada amostra de entrada, e apresentar o sinal resultante na saída, portanto para retirar o sinal negativo dos dados foi aplicada a equação $x*x$.

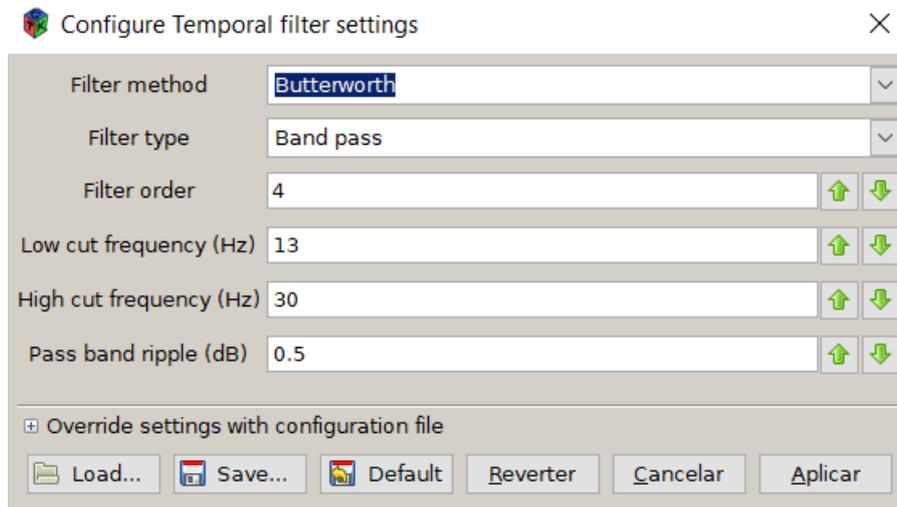


Figura 11: Módulo *Temporal Filter*.

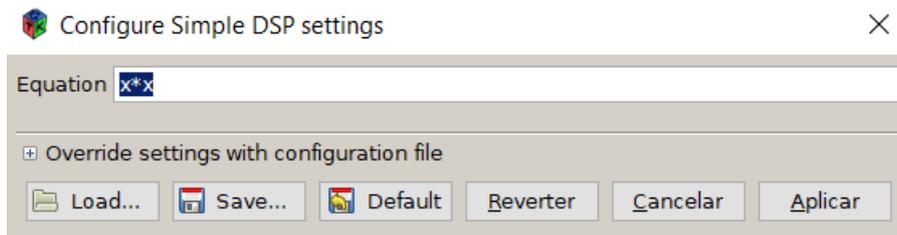


Figura 12: Módulo *Simple DSP*.

Logo após a transformação dos dados, as etapas de pré-processamento realizadas foram de limpeza e transformação dos dados CSV para o formato ARFF (*Attribute Relation File Format*), para a leitura no software de mineração de dados.

4.2.1 Limpeza dos dados e aplicação dos rótulos

A partir dos dados coletados de (BASTOS, 2017), foram aplicados os mesmos processos para a limpeza e rotulação dos dados. Inicialmente para o processo de limpeza dos dados foram retiradas as colunas consideradas desnecessárias para o estudo, a coluna “Time” foi retirada logo após a adição das classes (cubo, pirâmide e paralelogramo) que foram marcadas de acordo com o tempo, e os “channels” foram substituídos pelos nomes dos eletrodos (canais) referentes. A Figura 14 ilustra os dados brutos e a Figura 15, os dados após a rotulação das classes de objetos.

Time (s)	Channel 1	Channel 2	Channel 3	Channel 4	Channel 5
6,945	2,41E-11	9,31E-12	2,56E-10	1,03E-10	9,45E-11
6,947	4,38E+05	7,39E+05	1,24E+06	3,27E+06	1,21E+05
6,949	3,77E+07	6,30E+07	1,02E+08	1,96E+08	9,91E+06
6,951	8,03E+08	1,32E+09	2,05E+09	3,04E+09	2,19E+08
6,953	7,66E+09	1,24E+10	1,83E+10	2,24E+10	2,29E+09
6,955	4,21E+10	6,72E+10	9,51E+10	9,98E+10	1,42E+10
6,957	1,53E+11	2,41E+11	3,28E+11	3,01E+11	5,77E+10
6,959	3,94E+11	6,16E+11	8,11E+11	6,61E+11	1,65E+11
6,961	7,54E+11	1,18E+12	1,50E+12	1,10E+12	3,42E+11
6,963	1,08E+12	1,71E+12	2,12E+12	1,38E+12	5,23E+11

Figura 13: Dados brutos.

FP1	FZ	F3	F7	FT9	class
1,905	6,996	10,59	3,626	0,78	cubo
2,423	7,894	13,575	9,412	2,906	cubo
1,252	5,109	11,076	11,918	3,599	cubo
0,076	1,496	5,783	10,462	2,658	cubo
0,356	0,011	1,74	7,752	1,582	cubo
1,32	0,509	0,208	6,034	1,236	cubo
1,462	1,064	0	5,86	1,789	cubo
0,701	0,792	0,001	6,803	3,391	cubo
0,069	0,261	0,013	7,732	5,553	cubo
0,058	0,025	0,035	7,164	6,7	cubo

Figura 14: Dados após a rotulação das classes.

4.2.2 Transformação dos dados

Para o uso dos dados no software de mineração de dados WEKA², os mesmos devem estar no formato ARFF, que basicamente é um conjunto de instâncias que dividem um mesmo conjunto de atributos (PAYNTER et al., 2008). De modo geral, o formato ARFF denota de dois segmentos: cabeçalho e dados. No cabeçalho está descrito a *relation* que é a denominação do problema; *attribute* que são os atributos; e o tipo de dado. Nos dados (*data*) são os dados contidos nas instâncias dos atributos relacionados no cabeçalho, conforme Figura 16.

```
@relation Sujeito

@attribute FP1 numeric
@attribute FZ numeric
@attribute F3 numeric
@attribute F7 numeric
@attribute FT9 numeric
@attribute class {cubo,paralelo,piramide}

@data
1.905,6.996,10.59,3.626,0.78,cubo
2.423,7.894,13.575,9.412,2.906,cubo
1.252,5.109,11.076,11.918,3.599,cubo
0.076,1.496,5.783,10.462,2.658,cubo
0.356,0.011,1.74,7.752,1.582,cubo
1.32,0.509,0.208,6.034,1.236,cubo
1.462,1.064,0.0,5.86,1.789,cubo
0.701,0.792,0.001,6.803,3.391,cubo
0.069,0.261,0.013,7.732,5.553,cubo
0.058,0.025,0.035,7.164,6.7,cubo
```

Figura 15: Modelo de dados no formato ARFF.

²<https://www.cs.waikato.ac.nz/ml/weka/>

4.3 Escolha das técnicas de MD

A escolha da técnica de MD depende do problema, do tipo dos dados e suas características. Elas são aplicadas por objetivos específicos, dependendo do que se deseja buscar ou extrair dos dados (BASTOS, 2017).

Nesse trabalho, as técnicas escolhidas referem-se a classificação por Árvores de Decisão e foi embasada na comparação dos resultados com o trabalho de (BASTOS, 2017). Portanto, os algoritmos utilizados foram J48 e Random Tree, já que o trabalho visa a compreensão dos dados através da árvore de decisão e não somente da classificação correta das instâncias, tendo em vista que os dois algoritmos utilizados geram visualmente uma árvore de decisão.

A ferramenta operada para a execução do estudo, o software WEKA, contém diversos métodos implementados para mineração de dados como as tarefas de:

- Classificação com as técnicas de: Árvore de decisão induzida, Regras de aprendizagem, Naive Bayes, Tabelas de decisão, Regressão local de pesos, Perceptron, Perceptron multicamada e SVM;
- Predição Numérica com as técnicas de: Regressão linear, Geradores de árvores modelo, Regressão local de pesos, Aprendizado baseado em instâncias e Tabela de decisão;
- Agrupamento com as técnicas de: EM, Cobweb, SimpleKMeans, DBScan e CLOPE;
- Associação com as técnicas de: Apriori, FPGrowth, PredictiveApriori e Tertius.

Esta ferramenta foi escolhida por ser de fácil manuseio e ser ofertada gratuitamente. As Figuras 17 e 18 representam a interface da ferramenta para o campo da tarefa de classificação através dos algoritmos J48 e Random Tree, respectivamente.

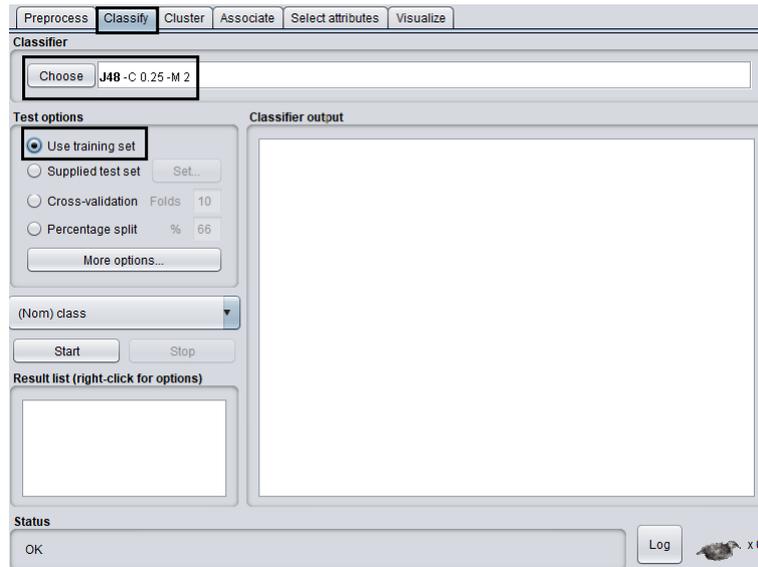


Figura 16: Configuração da tarefa de classificação para o algoritmo J48.

4.4 Aplicação e execução dos algoritmos

Para a realização do trabalho em que este estudo é fundamentado, foram realizados sete testes com os dados e as técnicas escolhidas, conforme a Figura 19. Estes testes basearam-se na variação dos seguintes parâmetros: número de canais, valores dos sinais e a ordem de execução das tarefas (BASTOS, 2017).

Os testes aplicados nos dados coletados de sinais cerebrais foram os seguintes:

- **Teste 1:** O primeiro teste foi feito com os 32 canais. Inicialmente através dos registros do tempo foi realizado a classificação dos objetos entre cubo, pirâmide e paralelogramo. Após a classificação dos objetos, foi feita a limpeza retirando a coluna do tempo e o cabeçalho com o nome dos canais. Por fim os dados CSV foram transformados para o formato ARFF para serem executados no Weka.
- **Teste 2:** O segundo teste foi realizado a partir dos dados do teste 1, porém foram testados somente os dados dos canais considerados prioritários para a execução da tarefa, ou seja, as áreas consideradas essenciais para a manipulação e identificação de objetos geométricos. Ao todo são 15 canais, que envolvem os lobos parietal, occipital e central. Os canais são: P7, P3, PZ, P4, P8, C3, CZ, C4, CP1, CP2, CP6, CP5, O1, O2 e OZ. Os canais não utilizados foram retirados diretamente na ferramenta Weka.
- **Teste 3:** Este teste foi feito a partir de um arquivo contendo somente as três primeiras amostras de cada sujeito, com os 32 canais. A seleção das amostras foram realizadas através da conferência do tempo em que cada indivíduo realizou cada atividade (amostra). Logo foram realizadas as etapas de pré-processamento como a limpeza dos dados, inserção das classes e transformação dos dados para ARFF.

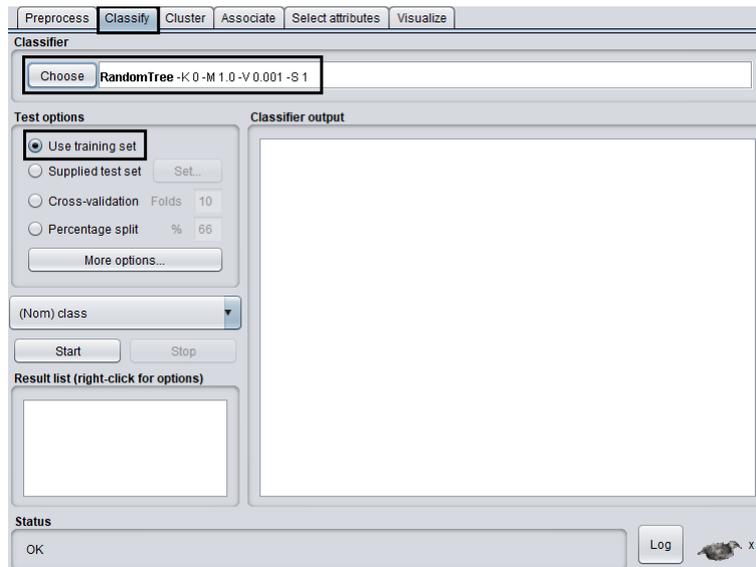


Figura 17: Configuração da tarefa de classificação para o algoritmo Random Tree.

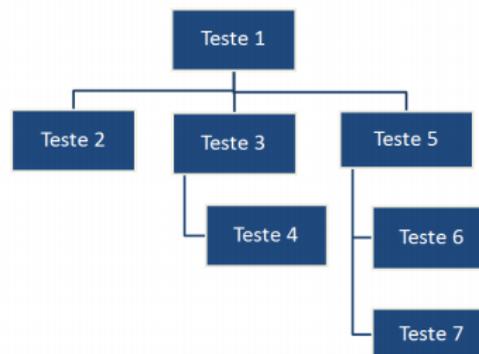


Figura 18: Hierarquia dos testes realizados com a ferramenta Weka (BASTOS, 2017).

- **Teste 4:** Foram utilizadas as últimas três amostras de cada indivíduo, com os 32 canais cada uma. Passando pelas etapas de limpeza, classificação dos dados e a transformação para o formato ARFF.
- **Teste 5:** O teste 5 foi realizado através do conjunto de 32 canais. Os dados foram compilados em um sistema que agrupou os maiores valores contidos a cada 10 linhas de cada coluna (canal), criando assim um novo arquivo CSV somente com os maiores valores de cada eletrodo, em um intervalo de 10. Os dados foram limpos e logo transformados para ARFF.
- **Teste 6:** Este teste foi realizado com base no teste 5, em que foram testados os maiores valores dos canais, no entanto foram utilizados somente os 15 canais considerados prioritários para a realização da tarefa. Como já citado, os canais considerados foram: P7, P3, PZ, P4, P8, C3, CZ, C4, CP1, CP2, CP6, CP5, O1, O2 e OZ.

- **Teste 7:** Os modelos gerados deste teste foram realizados a partir do agrupamento dos dados, com os maiores valores de cada grupo de cada um dos 32 canais, baseado no teste 5. Foram gerados 6 modelos para cada amostra, com as janelas de 25, 50 e 100 variando a porcentagem do número mínimo de objetos entre 10% e 1% da quantidade de instancias de cada amostra.

Para a aplicação dos testes, foi utilizada a mesma configuração do algoritmo J48 do trabalho de (BASTOS, 2017), nos algoritmos utilizados neste estudo. A opção utilizada para o teste foi “*Use training set*” que aplica o teste no mesmo conjunto de treinamento, que será usado para construir o classificador. Em *GenericObjectEditor* foi alterado o “*MinNumObj*” que determina o número mínimo de instâncias por folha, utilizou-se 1% e 10% do número de instâncias de cada amostra. Os outros parâmetros utilizados foram os definidos como padrão da ferramenta. As Figuras 20 e 21 a seguir mostram, respectivamente, as configurações do algoritmo J48 e do algoritmo Random Tree.

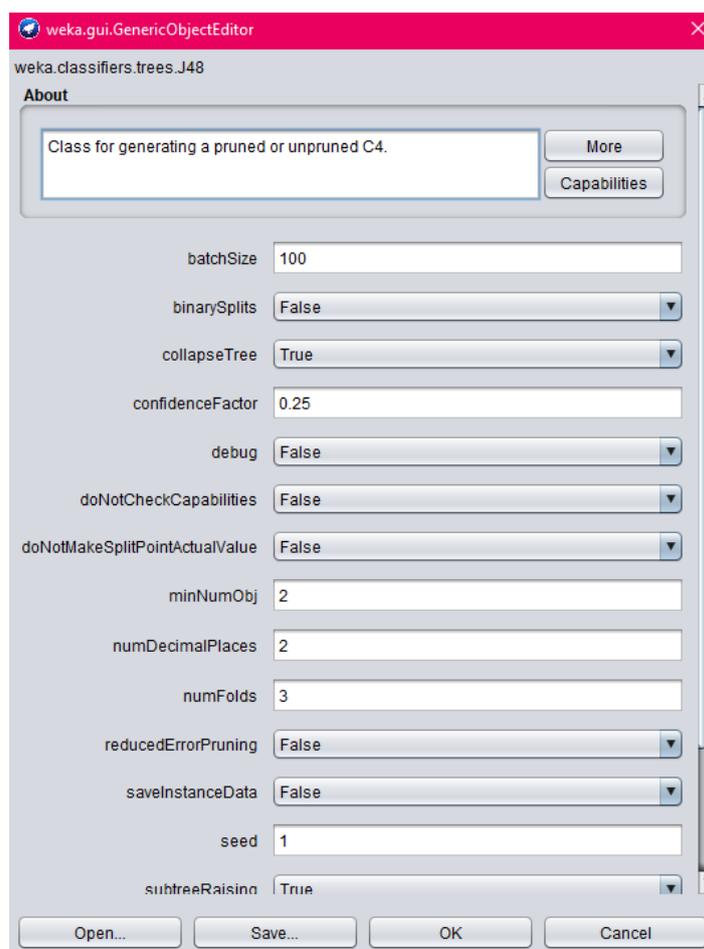


Figura 19: Configuração do algoritmo J48.

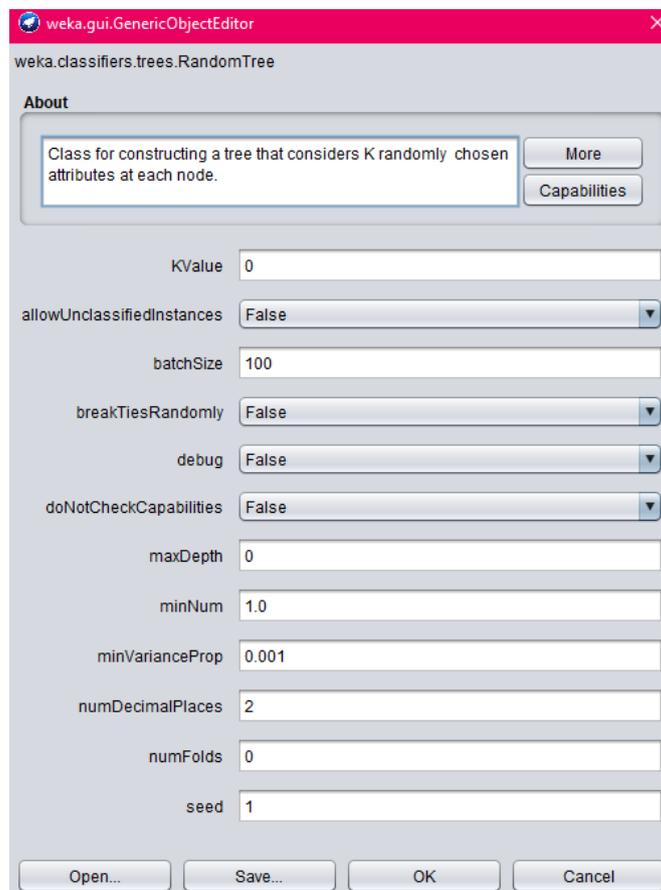


Figura 20: Configuração do algoritmo Random Tree.

5 ANÁLISE E VALIDAÇÃO DOS RESULTADOS

Este capítulo apresenta a compilação dos resultados obtidos sobre todos os 7 (sete) testes realizados. As tabelas 3, 5, 7 e 9 destacam em amarelo os melhores resultados obtidos após a aplicação do algoritmo J48 com a frequência da banda Beta, e a aplicação do algoritmo Random Tree na frequência dita "completa" e também na faixa de frequência da banda Beta. Os resultados do trabalho de BASTOS (2017) estão nas colunas destacadas em cinza. Como parâmetros de escolha dos melhores resultados, baseando-se nas hipóteses deste estudo, considerou-se o tamanho da árvore ser pequeno, para interpretar melhor através da visualização, combinado com a taxa de instâncias classificadas corretamente.

Os dados apresentados referem-se a 4 (quatro) sujeitos, sendo dois videntes e dois cegos congênitos, que foram coletados pelo trabalho de BASTOS (2017) e seguem a mesma hipótese:

Indivíduos videntes e cegos utilizam diferentes áreas do encéfalo para “visualizar” objetos espaciais. Videntes, utilizam o lobo occipital, que é responsável pela visualização, e cegos prioritariamente o lobo parietal, área responsável pela percepção tátil.

O trabalho de BASTOS (2017) utilizou o algoritmo J48 de árvores de decisão para classificação e visualização desses dados. E tendo como base nesta hipótese, a seguir cada sujeito é analisado individualmente.

5.1 Sujeito 1 - Vidente

Na Tabela 3 tem-se os dados do Sujeito 1, para a comparação de algoritmos com a frequência de banda completa pode-se observar através do teste 7, que a janela de maximização de 100 com 1% de instâncias por folha, obteve uma classificação de 73.74% corretamente, mas a árvore gerada de tamanho 219 é considerada de difícil visualização e interpretação. Portanto, no mesmo teste com 10% de instâncias por folha, obteve-se uma porcentagem menor de instâncias classificadas corretamente (46.19%), mas a árvore de tamanho 25 tem melhor entendimento do caminho que os canais percorrem para atingir os nodos folhas, que neste caso são os objetos Cubo, Paralelogramo e Pirâmide.

Para os testes aplicados com a modulação da banda Beta, os melhores resultados foram encontrados no teste 4, tendo em vista os tamanhos das árvores: para o algoritmo J48 a árvore é de tamanho 13 com 36.34% das instâncias classificadas corretamente; para o Random Tree uma árvore tamanho 23 com 37.07% de classificação. Salienta-se que com a aplicação do algoritmo J48 obteve-se uma classificação maior, de 42.17%, porém a árvore de tamanho 165, e para o algoritmo Random Tree uma classificação de 69.97% com a árvore gerada de tamanho 201.

Tabela 3: Resultados gerados através dos testes realizados para os dados do Sujeito 1.

Teste	Canais	Máx	Objetos por Folha	Banda Completa						Banda Beta					
				Instâncias	J48		Random		Tree	Instâncias	J48		Random		Tree
					Tamanho Árvore	Instâncias Corretas	Tamanho Árvore	Instâncias Corretas			Tamanho Árvore	Instâncias Corretas	Tamanho Árvore	Instâncias Corretas	
1	32	-	1%	87516	131	41.28%	231	41.23%	90584	167	39.13%	303	39.65%		
			10%		11	38.10%	23	37.95%		15	35.64%	39	35.68%		
2	15	-	1%	87516	115	40.60%	273	40.80%	90584	81	36.79%	285	38.93%		
			10%		13	37.38%	25	37.32%		9	35.28%	33	35.56%		
3	32	-	1%	47622	147	42.56%	271	41.68%	47592	157	41.51%	291	41.78%		
			10%		9	38.06%	21	36.47%		11	36.28%	51	36.82%		
4	32	-	1%	39894	111	43.83%	267	44.21%	42992	165	42.17%	289	42.91%		
			10%		11	38.81%	27	38.94%		13	36.34%	23	37.07%		
5	32	10	1%	8708	41	40.93%	293	48.40%	9059	9	35.69%	339	49.34%		
			10%		7	38.75%	33	40.75%		3	34.84%	47	38.45%		
6	15	10	1%	8708	21	39.08%	345	49.41%	9059	1	34.20%	283	47.40%		
			10%		9	38.31%	33	40.60%		1	34.20%	33	37.70%		
7	32	25	1%	3549	105	49.28%	249	54.75%	3624	1	34.40%	251	54.52%		
			10%		7	38.65%	25	42.78%		1	34.40%	63	42.08%		
			1%	1823	135	65.66%	229	61.96%	1812	1	34.49%	213	61.03%		
			10%		9	43.06%	23	44.60%		1	34.49%	53	44.70%		
		100	1%	1037	127	70.00%	219	73.74%	906	1	34.54%	201	69.97%		
			10%		15	46.19%	25	49.08%		1	34.54%	35	48.23%		

As Figuras 21, 22 e 23 apresentam respectivamente as árvores geradas pelo algoritmo Random Tree com a banda completa, algoritmo J48 com a modulação da banda Beta, e o algoritmo Random Tree com a modulação da banda Beta.

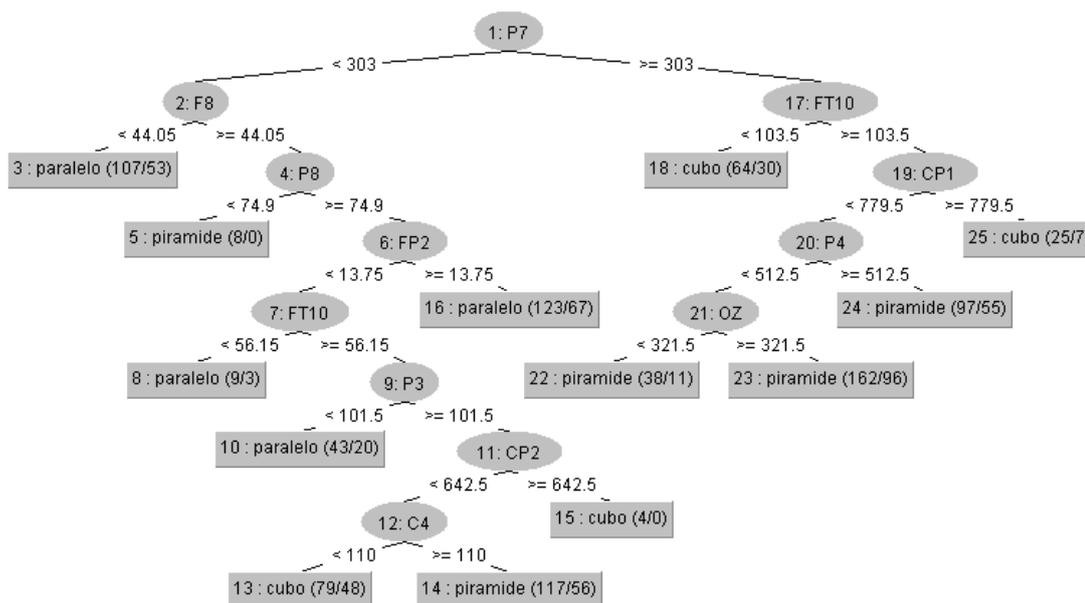


Figura 21: Árvore gerada a partir do algoritmo Random Tree com os dados da banda completa do Sujeito 1.

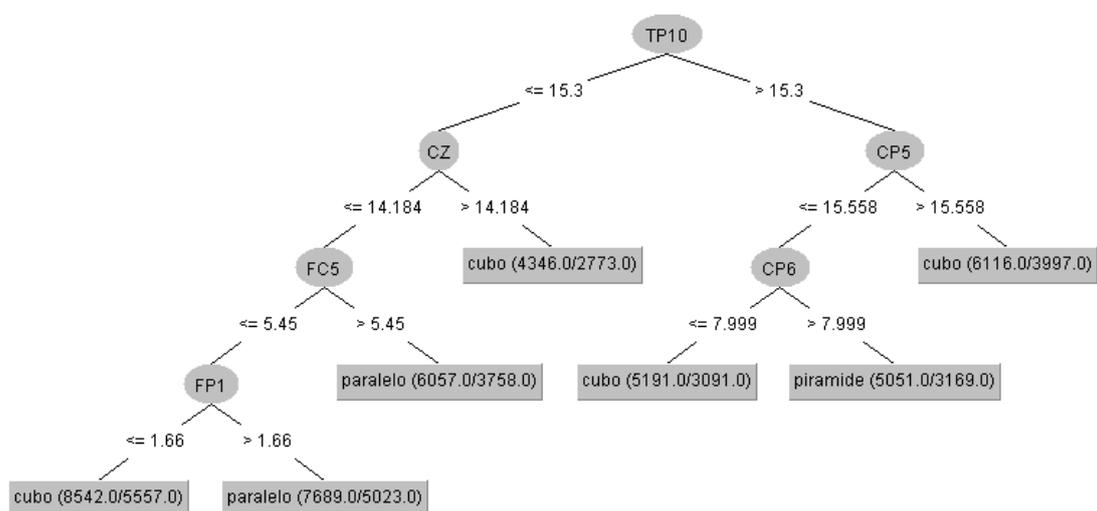


Figura 22: Árvore gerada a partir do algoritmo J48 com os dados da modulação banda Beta do Sujeito 1.

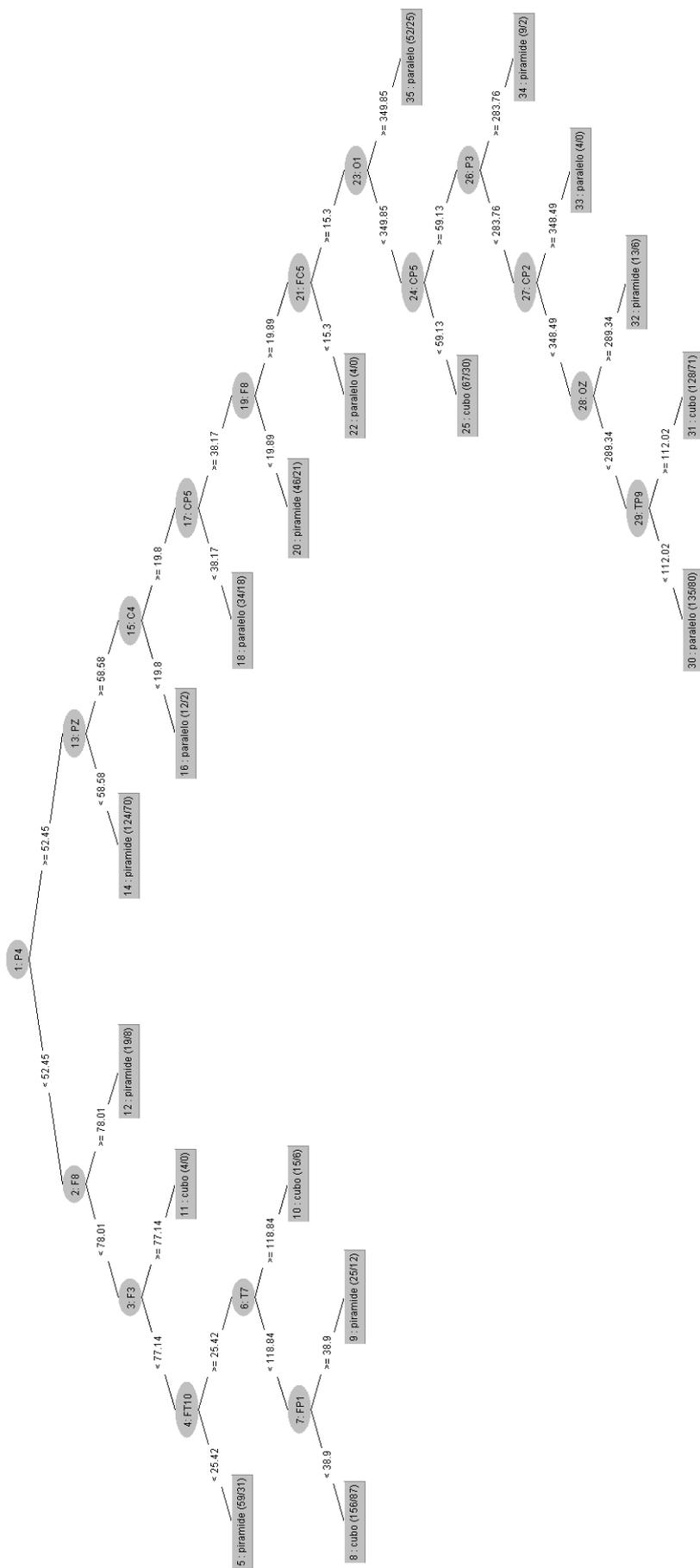


Figura 23: Árvore gerada a partir do algoritmo Random Tree com os dados da modulação banda Beta do Sujeito 1.

A Tabela 4 mostra as áreas que foram ativadas durante a atividade, aplicando o algoritmo Random Tree para os dados de banda completa e aplicando os algoritmos J48 e Random Tree nos dados com a modulação da banda Beta. As áreas ativadas retiradas do trabalho de (BASTOS, 2017) estão na célula destacada em cinza.

Tabela 4: Áreas ativadas do Sujeito 1.

Algoritmo	Banda Completa	Banda Beta
J48	T, P, F, O	T, P, F
Random Tree	P, F, O	P, F, T, O

Para o Sujeito 1, vidente, os eletrodos que mostraram relevância significativa, conforme a Tabela 4, e baseando-se na Tabela 1 foram: Temporal, o qual trata da percepção de movimentos biológicos; Parietal, que é responsável pelo sentido tátil; Frontal, responsável pelas funções executivas e gerenciamento dos recursos cognitivos; e Occipital onde ocorre o processamento das imagens. Como o sujeito é vidente, esperava-se ativação dos lobos Occipitais, baseando-se na hipótese de BASTOS (2017). Contudo, com a aplicação do algoritmo J48 com a modulação da banda Beta, não foi estimulado o Lobo Occipital no momento do teste.

5.2 Sujeito 2 - Vidente

Na Tabela 5, pode-se observar os dados do sujeito 2, através dos modelos gerados com o teste 7, na janela de maximização de 100 com 10% de instâncias por folha, foi o resultado escolhido, para todos os algoritmos e conjunto de dados utilizados, baseado principalmente nos tamanhos das árvores.

Na comparação do algoritmo para com a banda completa, o Random Tree classificou 49.47% das instâncias corretamente, gerando uma árvore de tamanho 29. Já para os testes com a modulação da banda Beta, o algoritmo J48 classificou 44.74% das instâncias e a árvore tamanho 13; o algoritmo Random Tree gerou uma árvore tamanho 23 ,com 42.70% das instâncias classificadas.

Tabela 5: Resultados gerados através dos testes realizados para os dados do Sujeito 2.

Teste	Canais	Máx	Objetos por Folha	Banda Completa						Banda Beta					
				Instâncias	J48		Random		Tree	Instâncias	J48		Random		Tree
					Tamanho Árvore	Instâncias Corretas	Tamanho Árvore	Instâncias Corretas			Tamanho Árvore	Instâncias Corretas	Tamanho Árvore	Instâncias Corretas	
1	32	-	1% 10%	123883	95 11	40.83% 38.55%	241 31	41.00% 38.56%	122622	151 11	39.34% 35.93%	305 35	39.95% 36.22%		
2	15	-	1% 10%	123883	79 9	40.10% 38.14%	263 19	40.51% 38.20%	122622	111 7	38.20% 35.38%	297 27	38.82% 35.73%		
3	32	-	1% 10%	65669	109 11	42.70% 39.60%	211 41	42.94% 39.85%	65613	145 11	41.33% 37.09%	253 23	41.65% 37.10%		
4	32	-	1% 10%	58214	137 13	43.09% 39.27%	269 31	43.71% 39.53%	57009	139 13	41.42% 37.37%	305 33	42.41% 37.59%		
5	32	10	1% 10%	12341	61 11	42.41% 39.36%	269 27	47.48% 40.57%	12263	73 13	41.67% 38.02%	269 25	46.97% 38.02%		
6	15	10	1% 10%	12341	23 7	38.77% 38.19%	351 53	47.38% 41.24%	12263	19 7	37.20% 36.52%	313 37	46.22% 37.98%		
7	32	25	1%	5108	107	49.72%	287	54.56%	4906	55	43.57%	259	52.71%		
			10%		11	40.40%	35	42.97%		5	36.46%	27	39.80%		
		50	1%	2554	71	50.66%	225	59.24%	2453	55	45.25%	231	58.86%		
		100	10%	1405	11	42.48%	31	45.03%		7	39.05%	31	41.41%		
			1%		119	63.13%	213	68.28%	1227	97	58.51%	215	67.72%		
			10%		11	44.19%	29	49.47%		13	44.74%	23	42.70%		

As Figuras 24, 25 e 26 apresentam, respectivamente, as árvores geradas pelo algoritmo Random Tree com a banda completa, algoritmo J48 com a modulação da banda Beta, e o algoritmo Random Tree com a modulação da banda Beta.



Figura 24: Árvore gerada a partir do algoritmo Random Tree com os dados da banda completa do Sujeito 2.

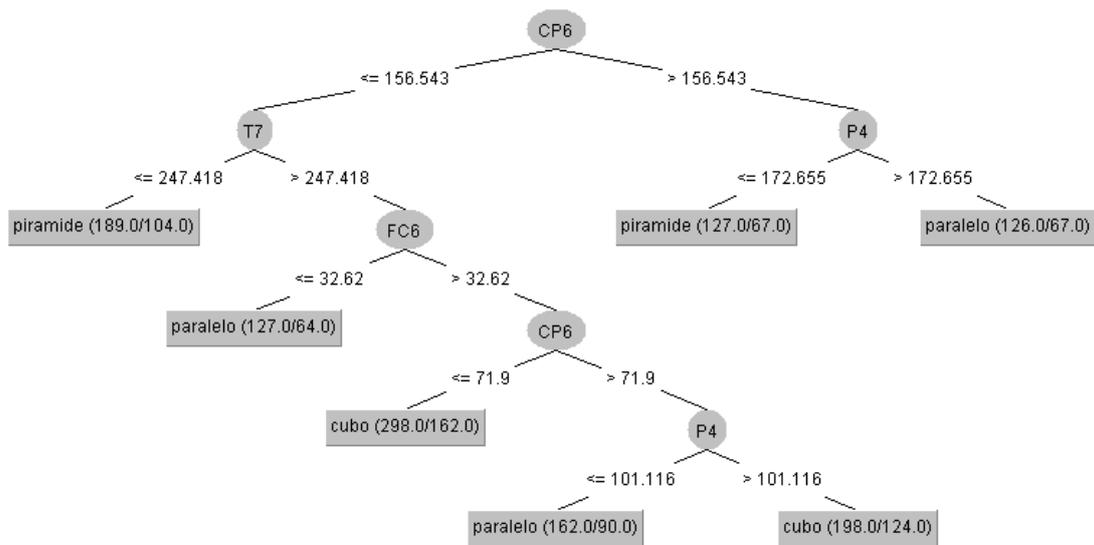


Figura 25: Árvore gerada a partir do algoritmo J48 com os dados da modulação banda Beta do Sujeito 2.

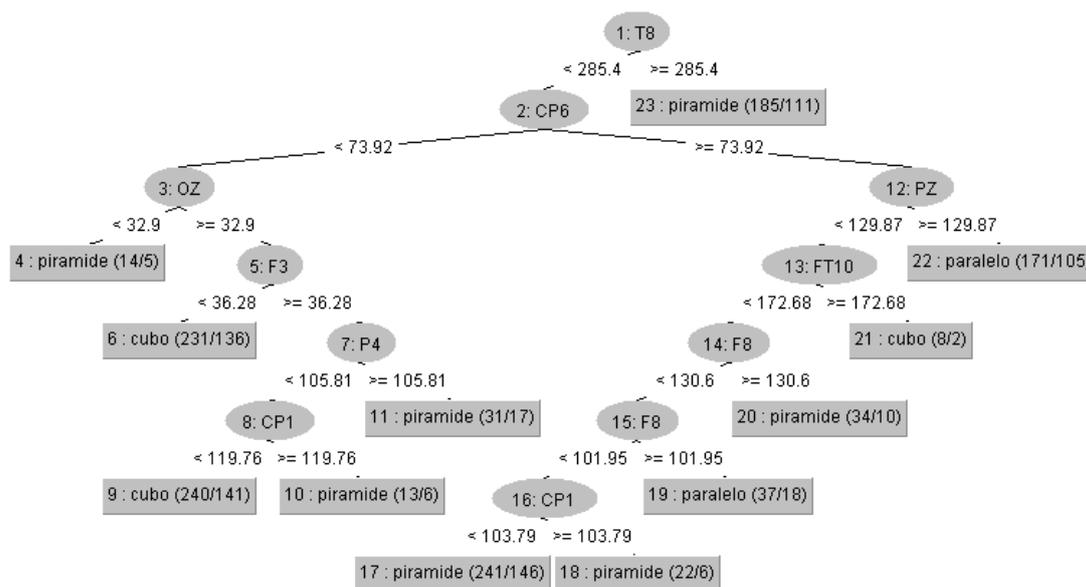


Figura 26: Árvore gerada a partir do algoritmo Random Tree com os dados da modulação banda Beta do Sujeito 2.

A Tabela 6 mostra as áreas que foram ativadas durante a atividade, aplicando o algoritmo Random Tree para os dados de banda completa e aplicando os algoritmos J48 e Random Tree nos dados com a modulação da banda Beta. As áreas ativadas retiradas do trabalho de (BASTOS, 2017) estão na célula destacada em cinza.

Tabela 6: Áreas ativadas do Sujeito 2.

Algoritmo	Banda Completa	Banda Beta
J48	F, T, O, P	P, T, F
Random Tree	P, T, F	F, T, O, P

Para o Sujeito 2, vidente, os eletrodos que mostraram relevância significativa, conforme a Tabela 6, e baseando-se na Tabela 1 foram: Frontal, responsável pelas funções executivas e gerenciamento dos recursos cognitivos; Temporal, o qual trata da percepção de movimentos biológicos; Occipital onde ocorre o processamento das imagens e Parietal, que é responsável pelo sentido tátil. Como o sujeito é vidente, esperava-se ativação dos lobos Occipitais, baseando-se na hipótese de BASTOS (2017). Contudo, com a aplicação do algoritmo Random Tree com a utilização da banda completa e o algoritmo J48 com a utilização da banda Beta, não ocorreu ativação do Lobo Occipital.

5.3 Sujeito 3 - Cego

A Tabela 7 referente ao sujeito 3, mostra que, apesar de apresentar árvores menores do que a escolhida, também foi feito um balanceamento entre a classificação e o tamanho

da árvore. Portanto, o melhor resultado para a banda completa gerado pelo algoritmo Random Tree foi considerado o teste 7 com a janela de maximização de 50 com 10% de instâncias por folha, com uma árvore tamanho 33 e a classificação de 41.85%.

Considerando os resultados a partir da modulação da banda Beta, o algoritmo J48 obteve poucas classificações acima de 40%, o que é considerado uma taxa baixa de acerto, mas foi escolhido para representação o teste 5 com 10% de instâncias por folha, que gerou uma árvore de tamanho 11 com 37.29% de classificação. Já para o algoritmo Random Tree considerou-se o teste 7 com a janela de 100 e a maximização de 10% o melhor resultado, pois a árvore gerada foi de tamanho 27, com 46.63% de classificação correta.

Tabela 7: Resultados gerados através dos testes realizados para os dados do Sujeito 3.

Teste	Canais	Máx	Objetos por Folha	Banda Completa						Banda Beta					
				Instâncias	J48		Random		Tree	Instâncias	J48		Random		Tree
					Tamanho Árvore	Instâncias Corretas	Tamanho Árvore	Instâncias Corretas			Tamanho Árvore	Instâncias Corretas	Tamanho Árvore	Instâncias Corretas	
1	32	-	1%	164348	131	37.89%	287	38.68%	100998	93	37.65%	265	39.29%		
			10%		9	35.67%	21	35.72%		13	35.66%	21	35.87%		
2	15	-	1%	164348	87	36.70%	337	38.00%	100998	91	37.54%	295	38.70%		
			10%		11	35.38%	33	35.28%		11	35.35%	23	35.85%		
3	32	-	1%	85622	137	39.38%	285	40.07%	53512	131	40.77%	313	41.71%		
			10%		9	35.70%	23	35.88%		13	36.56%	31	36.78%		
4	32	-	1%	78726	153	40.50%	269	40.72%	47488	135	42.36%	283	42.69%		
			10%		11	37.00%	29	37.27%		11	37.29%	39	38.40%		
5	32	10	1%	16384	31	36.50%	325	44.44%	10100	55	40.63%	265	47.10%		
			10%		11	36.73%	29	37.39%		13	37.68%	35	38.58%		
6	15	10	1%	16384	11	35.26%	355	44.98%	10100	11	36.50%	339	46.85%		
			10%		5	35.88%	35	37.47%		1	34.16%	81	38.68%		
7	32	25	1%	5108	53	40.78%	261	51.27%	4040	61	46.21%	255	54.52%		
			10%		3	36.42%	31	39.21%		3	37.40%	33	40.29%		
		50	1%	2554	15	54.24%	249	57.71%	2020	3	34.80%	221	59.95%		
			10%		11	38.86%	33	41.85%		1	33.96%	25	43.21%		
		100	1%	1405	71	52.07%	215	62.98%	1010	15	37.82%	205	70.59%		
			10%		15	43.00%	49	47.40%		1	33.76%	27	46.63%		

As Figuras 27, 28 e 29 apresentam, respectivamente, as árvores geradas pelo algoritmo Random Tree com a banda completa, algoritmo J48 com a modulação da banda Beta, e o algoritmo Random Tree com a modulação da banda Beta.

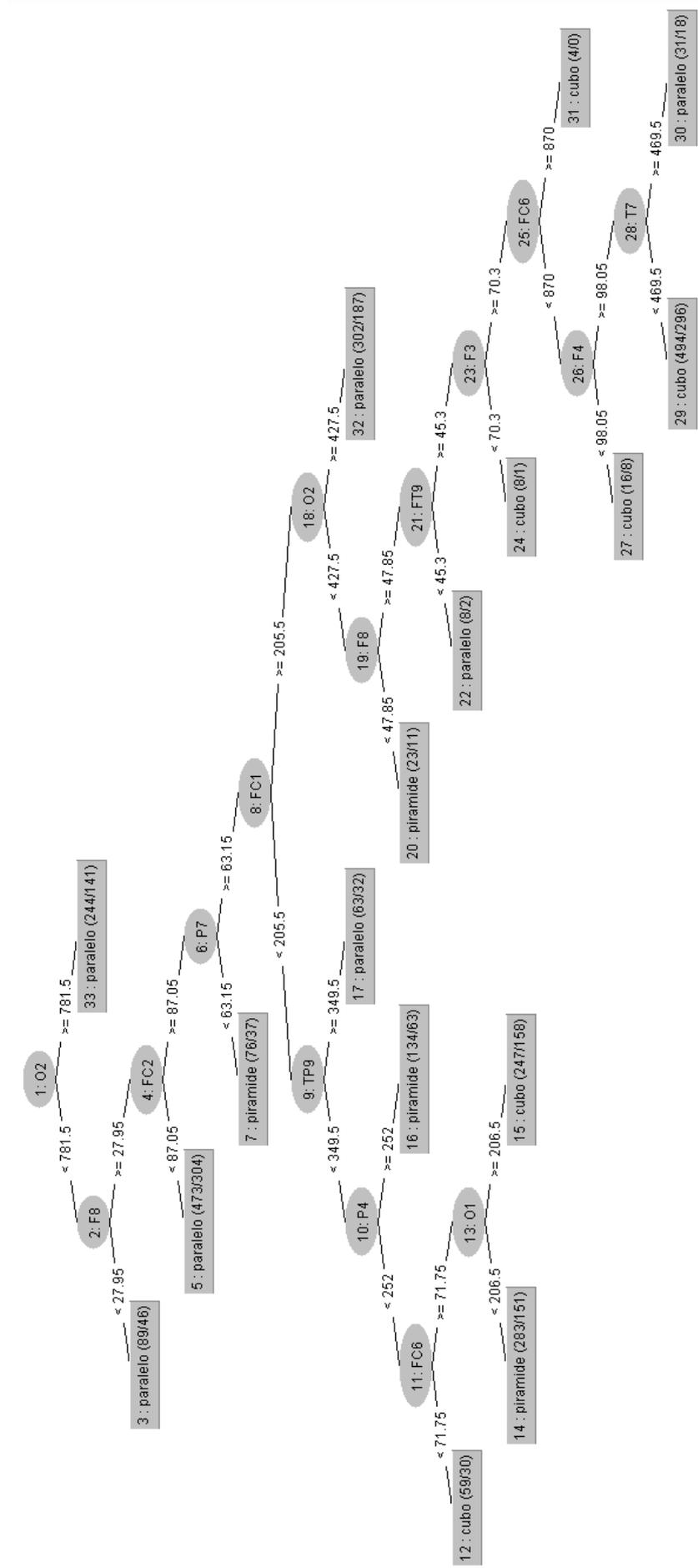


Figura 27: Árvore gerada a partir do algoritmo Random Tree com os dados da banda completa do Sujeito 3.

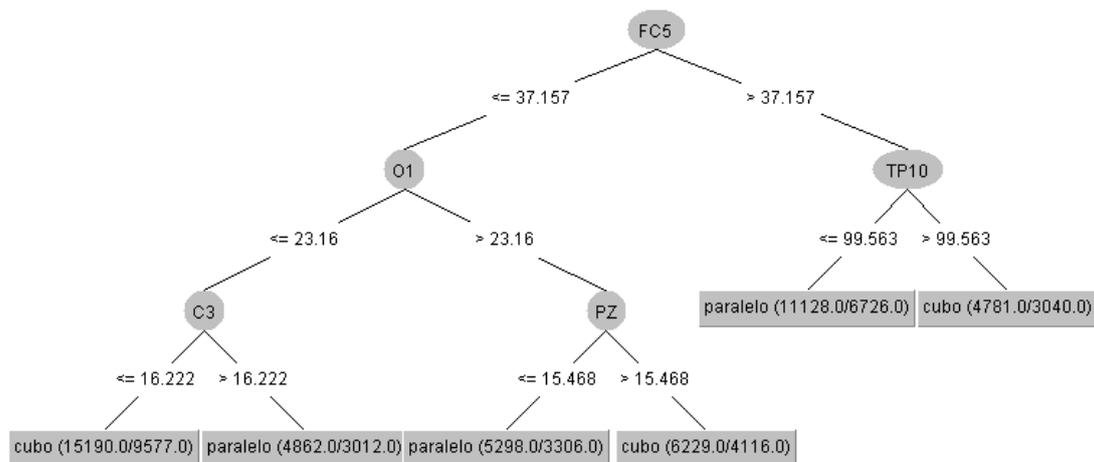


Figura 28: Árvore gerada a partir do algoritmo J48 com os dados da modulação banda Beta do Sujeito 3.

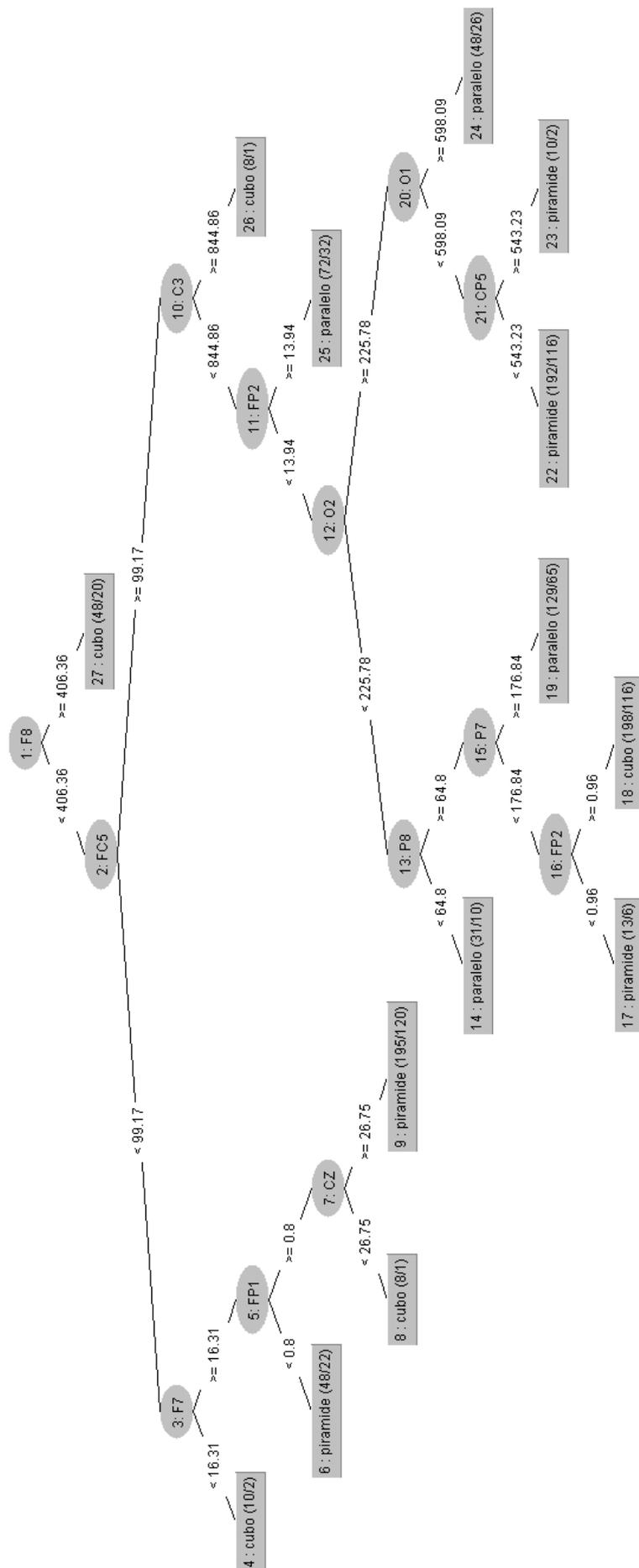


Figura 29: Árvore gerada a partir do algoritmo Random Tree com os dados da modulação banda Beta do Sujeito 3.

A Tabela 8 mostra as áreas que foram ativadas durante a atividade, aplicando o algoritmo Random Tree para os dados de banda completa e aplicando os algoritmos J48 e Random Tree nos dados com a modulação da banda Beta. As áreas ativadas retiradas do trabalho de (BASTOS, 2017) estão na célula destacada em cinza.

Tabela 8: Áreas ativadas do Sujeito 3.

Algoritmo	Banda Completa	Banda Beta
J48	T, P, F	F, O, P, T
Random Tree	O, F, T, P	F, O, P

Para o Sujeito 3, cego, os eletrodos que mostraram relevância significativa, conforme a Tabela 8, e baseando-se na Tabela 1 foram: Temporal, o qual trata da percepção de movimentos biológicos; Parietal, que é responsável pelo sentido tátil; Frontal, responsável pelas funções executivas e gerenciamento dos recursos cognitivos e Occipital onde ocorre o processamento das imagens. Como o sujeito é cego, esperava-se ativação dos Lobos Parietais baseando-se na hipótese de BASTOS (2017). Contudo, em todos testes houve ativação dos eletrodos do Lobo Parietal.

5.4 Sujeito 4 - Cego

Os dados referentes ao sujeito 4 estão dispostos na tabela 9. Para o algoritmo Random Tree com dados da banda completa, considerou-se o teste 7 com a janela de 50 e 10% de instâncias por folha o melhor resultado, pois gerou uma árvore tamanho 19 com 45.44% de classificação correta. Já para a banda a modulação da banda Beta, obteve-se para o algoritmo J48 uma árvore de tamanho 7 e 44.73% de classificação com o teste 7 maximização de 100 com 10% de instâncias por folha. Para o Random Tree, a árvore escolhida foi a de tamanho 19 com 44.61% de instâncias classificadas corretamente no teste 7 com a janela de 100 e 10% de instâncias por folha.

Tabela 9: Resultados gerados através dos testes realizados para os dados do Sujeito 4.

Teste	Canais	Máx	Objetos por Folha	Banda Completa						Banda Beta							
				Instâncias	J48		Random		Tree		Instâncias	J48		Random		Tree	
					Tamanho Árvore	Instâncias Corretas	Tamanho Árvore	Instâncias Corretas	Tamanho Árvore	Instâncias Corretas		Tamanho Árvore	Instâncias Corretas	Tamanho Árvore	Instâncias Corretas	Tamanho Árvore	Instâncias Corretas
1	32	-	1%	104728	95	46.38%	257	46.05%	103969	121	41.56%	283	41.93%	10%			
			10%		7	44.01%	43	44.27%		11	38.57%	35	39.05%				
2	15	-	1%	104728	51	45.32%	313	45.60%	103969	95	40.19%	235	70.90%	10%			
			10%		7	43.91%	39	44.48%		9	37.32%	53	38.55%				
3	32	-	1%	54450	133	43.90%	315	46.37%	53974	135	43.04%	293	43.56%	10%			
			10%		13	39.38%	17	41.24%		11	37.59%	21	38.06%				
4	32	-	1%	50278	107	45.77%	277	50.99%	49995	131	44.53%	245	44.56%	10%			
			10%		11	40.94%	29	48.53%		9	40.78%	37	41.17%				
5	32	10	1%	10423	93	45.17%	291	51.30%	10398	113	46.98%	255	48.73%	10%			
			10%		9	39.09%	27	44.81%		11	39.49%	31	41.69%				
6	15	10	1%	10423	15	37.00%	317	50.05%	10398	51	41.94%	291	48.49%	10%			
			10%		7	37.00%	53	45.05%		9	38.27%	47	40.30%				
7	32	25	1%	4347	43	48.19%	255	56.94%	4159	73	47.12%	239	54.72%	10%			
			10%		7	43.82%	27	45.99%		7	39.84%	25	41.11%				
		50	1%	2173	65	51.54%	241	63.48%	2080	37	46.92%	213	59.27%	10%			
			10%		5	45.54%	19	45.44%		11	42.45%	37	46.05%				
100	100	1%	1086	153	70.62%	221	73.09%	1040	33	44.42%	203	70.09%	10%				
		10%		3	44.56%	23	48.18%		7	41.73%	19	44.61%					

As Figuras 30, 31 e 32 apresentam, respectivamente, as árvores geradas pelo algoritmo Random Tree com a banda completa, algoritmo J48 com a modulação da banda Beta e o algoritmo Random Tree com a modulação da banda Beta.

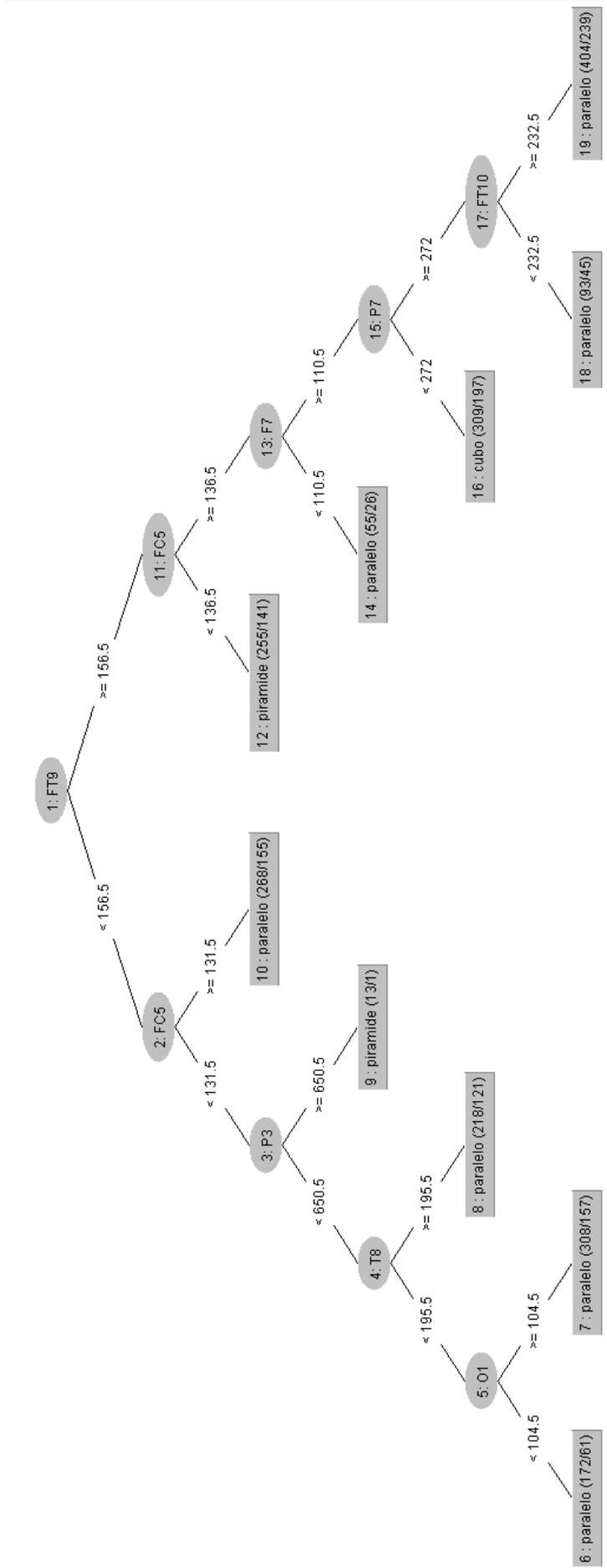


Figura 30: Árvore gerada a partir do algoritmo Random Tree com os dados da banda completa do Sujeito 4.

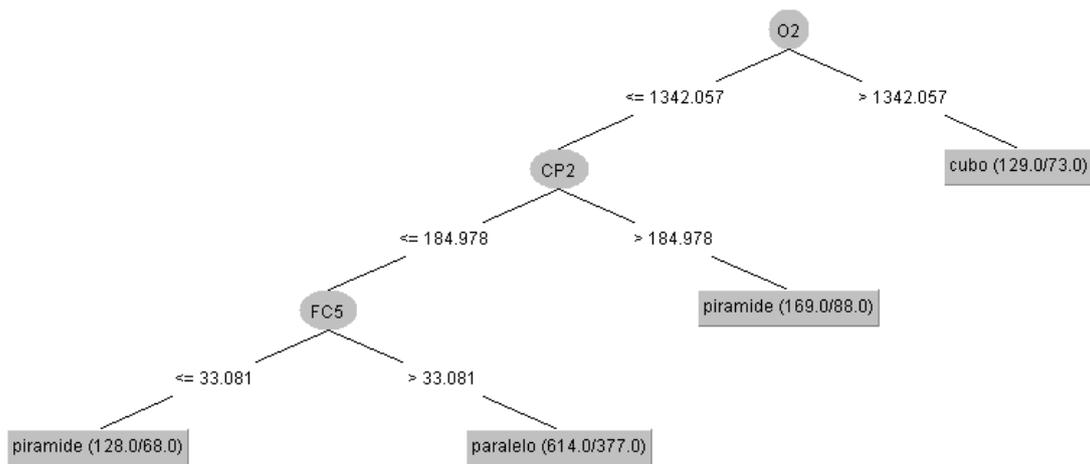


Figura 31: Árvore gerada a partir do algoritmo J48 com os dados da modulação banda Beta do Sujeito 4.

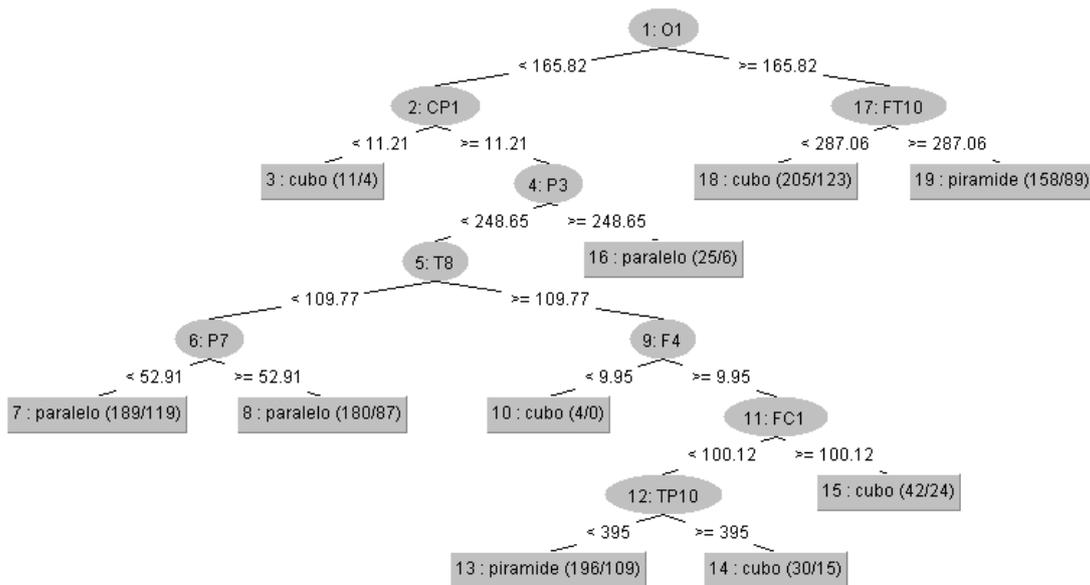


Figura 32: Árvore gerada a partir do algoritmo Random Tree com os dados da modulação banda Beta do Sujeito 4.

A Tabela 10 mostra as áreas que foram ativadas durante a atividade, aplicando o algoritmo Random Tree para os dados de banda completa e aplicando os algoritmos J48 e Random Tree nos dados com a modulação da banda Beta. As áreas ativadas retiradas do trabalho de (BASTOS, 2017) estão na célula destacada em cinza.

Tabela 10: Áreas ativadas do Sujeito 4.

Algoritmo	Banda Completa	Banda Beta
J48	O, P	O, P, F
Random Tree	P, F, O, T	O, P, T, F

Para o Sujeito 4, cego, os eletrodos que mostraram relevância significativa, conforme a Tabela 8, e baseando-se na tabela 1 foram: Occipital onde ocorre o processamento das imagens; Parietal, que é responsável pelo sentido tátil; Frontal, responsável pelas funções executivas e gerenciamento dos recursos cognitivos e Temporal, o qual trata da percepção de movimentos biológicos. Como o sujeito é cego, esperava-se ativação dos Lobos Parietais baseando-se na hipótese de BASTOS (2017). Contudo, pode-se observar que em todos testes houveram ativação dos eletrodos do Lobo Parietal.

5.5 Validação dos Resultados Obtidos

Seguindo as hipóteses propostas, segue a validação dos resultados de acordo com cada sujeito:

- Os resultados de instâncias classificadas corretamente são melhores utilizando o algoritmo Random Tree.

De acordo com a Tabela 11, onde os resultados são da aplicação dos algoritmos nos dados com utilização da banda completa, pode-se observar que as instâncias classificadas corretamente para os Sujeitos 1 e 2 foram maiores para o algoritmo Random Tree em relação ao J48. Contudo, para o Sujeito 3 diminuiu um pouco o percentual de acerto, e para o Sujeito 4 praticamente se manteve igual.

Tabela 11: Comparação entre algoritmos utilizando a faixa da banda completa.

Sujeito	J48	Random Tree
1	46.19%	49.08%
2	44.19%	49.47%
3	43.00%	41.85%
4	45.54%	45.44%

- Os resultados de instâncias classificadas corretamente são melhores utilizando somente a faixa de frequência da banda Beta.

De acordo com a Tabela 12, onde os resultados são dos algoritmos com a utilização somente da banda Beta, com exceção do Sujeito 2 com o algoritmo Random Tree, todos os outros obtiveram um resultado melhor na classificação das instâncias corretas.

Tabela 12: Comparação entre algoritmos utilizando a faixa da banda Beta.

Sujeito	J48	Random Tree
1	36.34%	48.23%
2	44.74%	42.70%
3	37.68%	46.63%
4	41.73%	44.61%

- Os resultados de instâncias classificadas corretamente são melhores utilizando o algoritmo Random Tree e utilizando a faixa de frequência da banda Beta.

De acordo com a Tabela 13, onde os resultados do algoritmo J48 são da utilização da banda completa e os resultados do algoritmo Random Tree da banda Beta, pode-se observar que para os Sujeitos 1 e 3 com o algoritmo Random Tree, obteve-se um resultado melhor na classificação. Já para os sujeitos 2 e 4, o melhor resultado foi com o algoritmo J48.

Tabela 13: Comparação entre algoritmo e banda.

Sujeito	J48	Random Tree
1	46.19%	48.23%
2	44.19%	42.70%
3	43.00%	46.63%
4	45.54%	44.61%

De forma a apenas analisar o percentual de acerto dos dados, com banda completa ou banda Beta, foi aplicado o algoritmo de Árvores de Decisão Random Forest, que consiste em uma técnica de agregação de classificadores do tipo árvore de decisão, construídos de forma que sua estrutura seja composta de maneira aleatória. Para determinar a classe de uma instância, o método combina o resultado de várias árvores de decisão, por meio de um mecanismo de votação. Ao final, cada árvore apresenta uma classificação, ou um voto para uma classe. A classificação final é dada pela classe que recebeu o maior número de votos entre todas as árvores da floresta (TELOKEN et al., 2016). Contudo, esse algoritmo não apresenta uma árvore, visto que é uma composição de diversas árvores, sendo impossível visualizar o caminho.

As Tabelas 14, 15, 16 e 17 apresentam os resultados das instâncias classificadas corretamente com a aplicação do algoritmo Random Forest. Os testes foram feitos a partir dos dados do teste 7, mencionado na seção 4.4. O Random Forest apresentou uma classificação correta das instâncias mais alta. Mas, como este algoritmo não gera a árvore, não é possível analisar os caminhos e ver a ativação dos eletrodos e as classes classificadas.

Tabela 14: Resultados das instâncias classificadas corretamente com o algoritmo Random Forest para os dados do Sujeito 1.

Teste	Canais	Max	Banda Completa		Banda Beta	
			Instâncias	Instâncias Corretas	Instâncias	Instâncias Corretas
7	32	25	3549	92.88%	3624	93.46%
		50	1823	93.71%	1812	93.32%
		100	1037	92.69%	906	92.93%

Tabela 15: Resultados das instâncias classificadas corretamente com o algoritmo Random Forest para os dados do Sujeito 2.

Teste	Canais	Max	Banda Completa		Banda Beta	
			Instâncias	Instâncias Corretas	Instâncias	Instâncias Corretas
7	32	25	5108	93.72%	4906	93.00%
		50	2554	93.62%	2453	93.27%
		100	1405	92.17%	1227	93.48%

Tabela 16: Resultados das instâncias classificadas corretamente com o algoritmo Random Forest para os dados do Sujeito 3.

Teste	Canais	Max	Banda Completa		Banda Beta	
			Instâncias	Instâncias Corretas	Instâncias	Instâncias Corretas
7	32	25	5108	93.08%	4040	93.86%
		50	2554	94.04%	2020	93.06%
		100	1405	92.88%	1010	93.26%

Tabela 17: Resultados das instâncias classificadas corretamente com o algoritmo Random Forest para os dados do Sujeito 4.

Teste	Canais	Max	Banda Completa		Banda Beta	
			Instâncias	Instâncias Corretas	Instâncias	Instâncias Corretas
7	32	25	4347	94.10%	4159	93.82%
		50	2173	93.65%	2080	93.26%
		100	1086	94.56%	1040	93.94%

6 CONCLUSÕES E TRABALHOS FUTUROS

Este trabalho teve como objetivo utilizar a mineração de dados, mais especificamente uma técnica de classificação, para gerar modelos do comportamento do encéfalo. Foi feita uma comparação entre os algoritmos de classificação J48 e Random Tree de Árvores de Decisão, mediante o uso de bandas espectrais diferentes.

A partir do trabalho de (BASTOS, 2017), surgiram as hipóteses deste estudo, onde foi utilizada a mesma base de dados, porém aplicada a modulação da amplitude da banda Beta, que é a banda espectral de estado de atenção, precisão visual e coordenação e também aplicado o algoritmo Random Tree de Árvores de Decisão nos dados, para fazer a comparação do desempenho dos algoritmos.

Através da análise dos modelos gerados, pode-se concluir que a modulação de amplitude para a tarefa de classificação e interpretação dos sinais cerebrais não apresentou resultados melhores do que a banda completa, nem relacionado a confirmar ou a negar as hipóteses do trabalho de (BASTOS, 2017), nem relacionado a ter um percentual de acerto maior. Os resultados apresentados pelo algoritmo Random Forest foram melhores, se analisado apenas o percentual de acerto, mas não permite uma análise de quais eletrodos foram utilizados para criação da árvore, ou seja, dos "caminhos" realizados pelo encéfalo para realização da atividade. Assim, os resultados sugerem que essa técnica é mais adequada para gerar modelos com maior acurácia que o J48, porém com menos informação para a análise do comportamento do encéfalo.

Como trabalhos futuros, vislumbra-se:

- Aplicar outras técnicas de mineração de dados, como SVM, Clusterização, Naive Bayes e Redes Neurais, de forma verificar os percentuais de acerto de cada uma das técnicas em relação as já testadas. Contudo, esse trabalho tem um aspecto mais computacional do que ligado a neurociências, pois não auxiliará no entendimento de quais lobos do encéfalo foram ativados.
- Aplicar técnicas de análise de dados temporais, visto que neste trabalho, no pré-processamento, os tempos foram excluídos dos dados da mineração, mas podem apresentar resultados promissores.

- Aplicar técnicas estatísticas de análise multivariada, para descobrir correlação entre variáveis, e assim, realizar um pré-processamento onde apenas variáveis "mais significativas" estejam presentes.

REFERÊNCIAS

- BASTOS, N. d. S. **Metodologia para Análise EEG Utilizando Árvores de Decisão: Um Estudo de Caso Para Reconhecimento de Objetos Espaciais**. 87 f. 2017. Dissertação (Mestrado em Ciência da Computação) — .
- CAMILO, C. O.; SILVA, J. C. d. Mineração de dados: Conceitos, tarefas, métodos e ferramentas. **Universidade Federal de Goiás (UFG)**, [S.l.], p.1–29, 2009.
- CARVALHO, F. A. H. d. **Reaprender a aprender: a pesquisa como alternativa metacognitiva**. 152 f. 2007. Tese (Doutorado em Ciência da Computação) — .
- COSENZA, R.; GUERRA, L. M e GUERRA, Leonor B. **Neurociência e Educação: como o cérebro aprende. Porto Alegre: Artmed**, [S.l.], 2011.
- DE AMO, S. Técnicas de mineração de dados. **Jornada de Atualização em Informatica**, [S.l.], 2004.
- EYSENCK, M.; KEANE, M. Limitações da atenção e do desempenho. **Manual de psicologia cognitiva**, [S.l.], v.5, 2007.
- FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. Inteligência Artificial: Uma abordagem de aprendizado de máquina. **Rio de Janeiro: LTC**, [S.l.], v.2, p.192, 2011.
- FAYYAD, a. o. From data mining to knowledge discovery in databases. **AI magazine**, [S.l.], v.17, n.3, p.37, 1996.
- FOLADOR, J.; ANDRADE, A. Mineração De Dados Aplicada à Séries Temporais Resultantes De Sinais Motores Da Doença De Parkinson. In: V CONGRESSO BRASILEIRO DE ELETROMIOGRAFIA E CINESIOLOGIA E X SIMPÓSIO DE ENGENHARIA BIOMÉDICA. CENTER CONVENTION UBERLÂNDIA., 2018, Uberlândia/MG. **Anais...** [S.l.: s.n.], 2018. p.660–663.
- GARCIA, F. D.; MÜLLER, S. M. T.; BASTOS FILHO, T. F.; SARCINELLI FILHO, M. Uso de Mapas Auto-Organizáveis em Classificação de Estados Mentais para Aplicação em uma Interface Cérebro-Computador. , [S.l.], 2007.

- GAZZANIGA, M. S.; HEATHERTON, T. F. **Ciência psicológica: mente, cérebro e comportamento**. [S.l.]: Artmed, 2005.
- GONÇALVES, E. C. Regras de associação e suas medidas de interesse objetivas e subjetivas. **UFF-Universidade Federal Fluminense, in INFOCOMP–Journal of Computer Science, Rio de Janeiro, RJ**, [S.l.], 2005.
- IZQUIERDO, I. **A arte de esquecer: cérebro, memória e esquecimento**. [S.l.]: Vieira & Lent, 2004.
- LENT, R. Visão das coisas: Estrutura e função do sistema visual. ____ **Cem**, [S.l.], 2001.
- LENT, R. A estrutura do sistema nervoso. **Neurociência da Mente e do Comportamento, Rio de Janeiro: Guanabara Koogan, 2008b**. ISBN, [S.l.], v.297246211, p.19–42, 2008.
- LENT, R. Cem bilhões de neurônios? 2. **São Paulo: Atheneu.[Links]**, [S.l.], 2010.
- LIBRELOTTO, S. R.; MOZZAQUATRO, P. M. Análise dos algoritmos de mineração J48 e Apriori aplicados na detecção de indicadores da qualidade de vida e saúde. **Revista Interdisciplinar de Ensino, Pesquisa e Extensão**, [S.l.], v.1, n.1, 2014.
- MONARD, M. C.; BARANAUSKAS, J. A. Indução de regras e árvores de decisão. **Sistemas Inteligentes. Rezende, SO Editora Manole Ltda**, [S.l.], p.115–140, 2003.
- MUNIAN, S.; SIVALINGAM, S.; JAYARAMAN, V. Analysis of Real Time EEG Signals. , [S.l.], p.1–53, 2014.
- OSHIRO, T. M. **Uma abordagem para a construção de uma única árvore a partir de uma Random Forest para classificação de bases de expressão gênica**. 2013. Tese (Doutorado em Ciência da Computação) — Universidade de São Paulo.
- PAYNTER, G.; TRIGG, L.; FRANK, E.; KIRKBY, R. Attribute-relation file format (ARFF). **Online] <http://www.cs.waikato.ac.nz/ml/weka/arff.html>**, [S.l.], 2008.
- PFURTSCHELLER, G.; BRUNNER, C.; SCHLÖGL, A.; DA SILVA, F. L. Mu rhythm (de) synchronization and EEG single-trial classification of different motor imagery tasks. **NeuroImage**, [S.l.], v.31, n.1, p.153–159, 2006.
- PRADA, B. M. L. **Interface cérebro-computador não invasiva baseada em OpenVibe**. 2010. Dissertação (Mestrado em Ciência da Computação) — Universidade de Aveiro.
- PRASS, F. S. **Estudo comparativo entre algoritmos de análise de agrupamentos em data mining**. 71 f. 2004. Dissertação (Mestrado em Ciência da Computação) — -Área de Concentração de Sistemas de Computação, Universidade Federal de Santa Catarina, Florianópolis.

SILVEIRA, J. d. Á. **Análise de sinais cerebrais utilizando árvores de decisão. 103 f.** 2013. Dissertação (Mestrado em Ciência da Computação) — .

TAN, P.-N. et al. **Introduction to data mining.** [S.l.]: Pearson Education India, 2009.

TELOKEN, A. et al. Estudo Comparativo entre os algoritmos de Mineração de Dados Random Forest e J48 na tomada de Decisão. **Simpósio de Pesquisa e Desenvolvimento em Computação**, [S.l.], v.2, n.1, 2016.

VALLADO, S. Y.; DELGADO, C.; SOUZA, D.; GUAGLIARDI, M.; SILVA, V.; LINS, F. R. W. Correlação entre hemisfericidade e o aprendizado psicomotor em tarefas de comunicação gestuais. **Fitness & Performance Journal**, [S.l.], v.3, n.4, p.207–210, 2004.

VASCONCELOS, N.; PANTOJA, J.; BELCHIOR, H.; CAIXETA, F. V.; FABER, J.; FREIRE, M. A. M.; COTA, V. R.; MACEDO, E. A. de; LAPLAGNE, D. A.; GOMES, H. M. et al. Cross-modal responses in the primary visual cortex encode complex objects and correlate with tactile discrimination. **Proceedings of the National Academy of Sciences**, [S.l.], p.201102780, 2011.

VAZ, Y. **Extração de características para a classificação de imagética motora em interfaces cérebro-computador. 111 f.** 2016. Tese (Doutorado em Ciência da Computação) — Universidade de São Paulo.

WEBSTER, J. G. **Medical instrumentation application and design.** [S.l.]: John Wiley & Sons, 2009.