



Centro de Ciências Exatas e Tecnologias - CCET  
Programa de Pós-graduação em Engenharia Elétrica

# Classificação de Janelas de Eletroencefalograma com e sem Crises Epilépticas

Davi Costa Nascimento

São Luis - MA, 2021



Davi Costa Nascimento

# Classificação de Janelas de Eletroencefalograma com e sem Crises Epilépticas

Dissertação apresentada ao Programa de Pós-graduação em Engenharia Elétrica do Centro de Ciências Exatas e Tecnologias, da Universidade Federal do Maranhão, como requisito para a obtenção do grau de Mestre em Engenharia Elétrica.

Orientador: Allan Kardec Duailibe Barros Filho

São Luis, Fevereiro de 2021

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).  
Núcleo Integrado de Bibliotecas/UFMA

Costa Nascimento, Davi.

Classificação de Janelas de Eletroencefalograma com e sem Crises Epilépticas / Davi Costa Nascimento. - 2021.  
57 f.

Orientador(a): Allan Kardec Duailibe BarrosFilho.  
Dissertação (Mestrado) - Programa de Pós-graduação em Engenharia Elétrica/ccet, Universidade Federal do Maranhão, Sistema de Conferência Google Meet, 2021.

1. Aprendizado de máquina. 2. Eletroencefalograma. 3. Epilepsia. 4. Estatística. I. Kardec Duailibe BarrosFilho, Allan. II. Título.

CCET - Centro de Ciências Exatas e Tecnologias  
Universidade Federal do Maranhão

Dissertação apresentada ao Programa de Pós-graduação em Engenharia Elétrica intitulada ***Classificação de Janelas de Eletroencefalograma com e sem Crises Epilépticas*** de autoria de Davi Costa Nascimento, aprovada pela banca examinadora constituída pelos seguintes professores:

---

Prof. Dr. Allan Kardec Duailibe Barros Filho -  
Orientador  
Programa de Pós-graduação em Engenharia Elétrica  
- UFMA

---

Prof. Dr. Jonathan Araújo Queiroz -  
Membro da banca  
Programa de Pós-graduação em Engenharia Elétrica  
- UFMA

---

Prof. Dr. Ginalber Luiz de Oliveira Serra-  
Membro da banca  
Programa de Pós-graduação em Engenharia Elétrica  
- UFMA

---

Prof. Dr. Carlos Alberto Bezerra Tomaz -  
Membro da banca  
Coordenação de Pesquisa em Neurociências - Ceuma

São Luis, 23 de Fevereiro de 2021



*"When you change yourself, you change the world"*  
***Silvera, Gojira***



## **DEDICATÓRIA**

Dedico este trabalho a todas as pessoas que sofrem com a epilepsia. E que de alguma forma ele possa tornar suas vidas mais tranquilas e confortáveis.

## AGRADECIMENTOS

Primeiramente a Deus, por seus infinitos livramentos e sua misericórdia sobre mim e aqueles que amo. Ele pavimentou os caminhos para chegar até aqui.

Aos meus pais, Davi e Angela, por todos os sacrifícios feitos por mim e minha irmã, para que esse e outros objetivos pudessem se realizar. Amo vocês.

A Thamires. Minha parceira de todos os momentos, com quem compartilho frustrações diárias, pequenas conquistas e todos os sonhos.

Aos amigos e companheiros do *squad* Maestro da *Pulse* (Grupo Mateus). Em especial a Luciana Lindoso, que todos os dias me desafia a aplicar o conhecimento acadêmico no "mundo real", e por todos os seus incentivos e *feedbacks* necessários.

Agradeço ainda aos membros do Laboratório de Processamento da Informação Biológica (PIB - UFMA). Principalmente a Gean Sousa e Jonathan Queiroz, fortes incentivadores e auxiliares desse trabalho. Agradecimentos especiais ao professor Allan Kardec Barros, pela oportunidade de fazer parte de sua equipe e por sua confiança em meu potencial.

## RESUMO

A epilepsia é um distúrbio neurológico crônico, o qual potencializa a ocorrência de convulsões em suas vítimas. Tal quadro clínico acarreta transtornos no cotidiano de pacientes, que em alguns casos, a ocorrência de crises pode promover lesões, traumatismos ou mesmo morte súbita. Um diagnóstico prévio, bem como um tratamento adequado, possibilita aos pacientes uma vida mais confortável e até mesmo o desaparecimento de crises. Portanto, há a necessidade do desenvolvimento de tecnologias e metodologias que possam agilizar e simplificar o diagnóstico da epilepsia, como também novas opções para o tratamento e acompanhamento de pacientes. Esse trabalho tem o objetivo de desenvolver um modelo generalizado de classificação de segmentos de eletroencefalograma com presença ou ausência de crises epiléticas. A metodologia adotada na extração de características consiste no cálculo dos momentos estatísticos de segunda, terceira e quarta ordem. Características essas que foram extraídas de janelas de 1 segundo de arquivos da base de dados CHB-MIT (Children's Hospital Boston). A classificação dos vetores de características obtidos se deu em duas etapas, sendo a primeira a partir dos momentos estatísticos, a segunda com a rotação dos mesmo vetores, utilizando a análise de componentes principais (PCA). O modelo de classificação proposto obteve para os vetores de características e para as componentes oriundas do PCA, acurácia de 86.4% e 94.6%, respectivamente, ambas utilizando o algoritmo k-NN. Acreditasse que com devidas adaptações melhorias o modelo pode ser embarcado em um dispositivo, para classificação de janelas em tempo real.

**Palavras-chave:** Epilepsia, Eletroencefalograma, Aprendizado de máquina, Estatística.

## ABSTRACT

Epilepsy is a chronic neurological disorder, which enhances the occurrence of seizures in its victims. This clinical picture causes daily disorders of patients, that in some cases, the occurrence of seizures can promote injuries, trauma or even sudden death. A previous diagnosis, as well as an adequate treatment, allows the patients a more comfortable life and even the disappearance of seizures. Therefore, there is a need to develop technologies and methodologies that can streamline and simplify the diagnosis of epilepsy, as well as new options for the treatment and monitoring of patients. This work aims to develop a generalized model for the classification of electroencephalogram segments with the presence or absence of epileptic seizures. The methodology adopted in the extraction of features consists in the calculation of the statistical moments of second, third and fourth order. These features were extracted from 1-second windows of files from the CHB-MIT database (Children's Hospital Boston). The classification of the feature vectors obtained consisted of two stages, the first from the statistical moments, the second with the rotation of the same vectors, using the principal component analysis (PCA). The proposed classification model obtained for the feature vectors and for the components coming from the PCA, accuracy of 86.4% and 94.6%, respectively, both using the k-NN algorithm. It was believed that with appropriate adaptations and improvements, the model can be embedded in a device, for classification of windows in real time.

**Key-words:** Epilepsy, Electroencephalogram, Machine Learning, Statistics.

## LISTA DE FIGURAS

2.1	Classificação Internacional das Crises Epilépticas. . . . .	21
2.2	Nomenclatura de regiões cerebrais. . . . .	23
2.3	Montagem de EEG no sistema 10-20. . . . .	24
2.4	Montagem de EEG no sistema 10-10. . . . .	24
2.5	Ilustração gráfica do eletroencefalograma. . . . .	25
3.1	Classificação de função de distribuição a partir da assimetria. . . . .	27
3.2	Classificação de função de distribuição a partir da curtose. . . . .	28
4.1	Ilustração de classificação com o $k$ -NN. . . . .	29
4.2	Limites de classificação não otimizados. . . . .	30
4.3	Ilustração de hiperplano de classificação do SVM. . . . .	31
4.4	Transformação realizada em conjunto de dados não linear. . . . .	32
4.5	Arquitetura de rede neural <i>Perceptron</i> . . . . .	35
4.6	Exemplo de Rede Neural MLP . . . . .	36
6.1	Montagem de EEG adotada . . . . .	40
6.2	Janelamento aplicado, exemplificado na matriz $\mathbf{A}$ . . . . .	41
6.3	Fluxograma para ilustração da extração de características e classificação de janelas de EEG. . . . .	45
7.1	Gráfico tridimensional de vetores de características. . . . .	46
7.2	Gráfico tridimensional de componentes principais . . . . .	47
7.3	Gráfico de Curvas ROC de classificadores dos vetores de características. . . . .	49
7.4	Gráfico de Curvas ROC de classificadores das componentes do PCA . . . . .	51

## LISTA DE TABELAS

2.1	Mudanças de nomenclaturas de eletrodos do sistema 10-20 para o sistema 10-10 . . . . .	24
4.1	Tabela com funções kernel mas comuns . . . . .	33
6.1	Informações quantitativas sobre base de dados CHB-MIT. . . . .	39
7.1	Métricas de classificação dos vetores de característica. . . . .	48
7.2	Métricas de classificação das componentes do PCA. . . . .	50
8.1	Comparativo de performance de metodologias anteriores com o a performance obtida neste trabalho. . . . .	53

## LISTA DE ABREVIATURAS

EEG	-	Eletroencefalograma
EKG	-	Eletrocardiograma
K-NN	-	K-Nearest Neighbors
SVM	-	Support Vector Machine
LDA	-	Linear Discriminant Analysis
OMS	-	Organização Mundial da Saúde
ILAE	-	International League against Epilepsy
QDA	-	Quadratic Discriminant Analysis
MLP	-	Multi-layer Perceptron
RBF	-	Radial Basis Function
PCA	-	Principal Component Analysis

## Conteúdo

<b>1</b>	<b>INTRODUÇÃO</b>	<b>18</b>
1.1	Trabalhos relacionados . . . . .	19
1.2	Objetivo Geral . . . . .	20
1.3	Objetivos Específicos . . . . .	20
<b>2</b>	<b>EPILEPSIA E ELETROENCEFALOGRAFIA</b>	<b>21</b>
2.1	Crises Epilépticas . . . . .	21
2.2	Eletroencefalograma - EEG . . . . .	23
<b>3</b>	<b>MOMENTOS ESTATÍSTICOS</b>	<b>26</b>
3.1	O Segundo Momento Estatístico e a Variância . . . . .	26
3.2	O Terceiro Momento Estatístico e a Assimetria . . . . .	26
3.3	O Quarto Momento Estatístico e a Curtose . . . . .	27
<b>4</b>	<b>CLASSIFICADORES</b>	<b>29</b>
4.1	k-Vizinhos Próximos - k-NN . . . . .	29
4.2	Máquina de Vetores de Suporte - SVM . . . . .	30
4.3	Naive Bayes . . . . .	33
4.4	Análise Quadrática Discriminante - QDA . . . . .	34
4.5	Rede Neural Artificial Perceptron de Múltiplas Camadas - RNA - MLP . . . . .	34
<b>5</b>	<b>ANÁLISE DE COMPONENTES PRINCIPAIS- PCA</b>	<b>37</b>
<b>6</b>	<b>MATERIAIS E MÉTODOS</b>	<b>38</b>
6.1	Base de Dados . . . . .	38
6.2	Pré-processamento . . . . .	39
6.3	Extração de Características . . . . .	40
6.4	Classificação . . . . .	43
6.5	Métricas de Avaliação . . . . .	43
<b>7</b>	<b>RESULTADOS</b>	<b>46</b>

7.1	Avaliação da Extração de Características . . . . .	46
7.2	Classificação . . . . .	48
<b>8</b>	<b>DISCUSSÃO</b>	<b>52</b>
8.1	Trabalhos Futuros . . . . .	53
<b>9</b>	<b>CONCLUSÃO</b>	<b>54</b>
	<b>REFERÊNCIAS</b>	<b>54</b>

# 1 INTRODUÇÃO

Doenças e anomalias cerebrais acometem boa parcela da população mundial. Suas origens são diversas, podendo seu surgimento ser provocado desde traumas ou lesões até mesmo fatores genéticos e hereditários. Dentre as principais doenças cerebrais podemos destacar o Mal de Alzheimer, que trata-se de um distúrbio cognitivo neurodegenerativo progressivo, o qual gradualmente decrementa as funções cognitivas de pacientes e resultando em sua morte [5]. A doença de Alzheimer é responsável por 60% a 70% dos casos de comprometimento cognitivo progressivo em pacientes idosos. A prevalência total da doença de Alzheimer nos Estados Unidos é estimada em 2.3 milhões [4].

Outra anomalia cerebral é a esclerose múltipla, que é uma doença incapacitante e não traumática, que pode acometer tanto adultos quanto os mais jovens [22]. Trata-se uma doença complexa que pode ser causada por vários fatores ambientais, dentre eles a falta de vitamina D ou luz ultravioleta B (UVB), infecção pelo vírus *Epstein-Barr* (Herpesvírus humano 4), obesidade e tabagismo [3]. No momento que o sistema imunológico passa a desconhecer lipídeos e proteínas do próprio indivíduo, o mesmo passa a atacá-los. Isso vem a provocar a degeneração de axônios, impedindo comunicação entre os neurônios, devido a impossibilidade de passagem de impulsos elétricos, comprometendo funções motoras e cognitivas [12].

A epilepsia também figura como uma conhecida doença cerebral. Ela conceituada como uma predisposição persistente do cérebro de promover crises recorrentes [36]. Tais crises tratam-se de ocorrências transitórias de sinais, devido a uma intensa atividade elétrica neural e síncrona do cérebro. Essa anormalidade na atividade neural acarreta em indivíduos alterações na consciência e até eventos involuntários de natureza muscular, psíquica e sensitiva.

Dados da Organização Mundial da Saúde estimam que em torno de 50 milhões de pessoas possuem algum grau de epilepsia em todo mundo. E que 70% dessas pessoas poderiam não mais sofrer a ocorrência de crises epiléticas se tivessem um diagnóstico prévio ou devido tratamento [1].

Devido ao caráter imprevisível da ocorrência crises, e também ao seu difícil controle, o cotidiano de pacientes acometidos de epilepsia é impactado de forma significativa, havendo alterações em âmbito social e até mesmo financeiro. Em casos mais extremos a ocorrência de crises pode aumentar o risco de um indivíduo sofrer queimaduras, lacerações, fraturas cranianas e até mesmo morte súbita e inesperada [32].

Uma das ferramentas mais comuns para avaliar e monitorar crises epiléticas é a análise do eletroencefalograma do couro cabeludo (EEG), que é um registro não invasivo de múltiplos canais da atividade elétrica do cérebro. Por ele podem ser identificados

padrões no comportamento neural para um diagnóstico detalhado e estabelecer um tratamento adequado. Diante desse quadro o desenvolvimento de tecnologias que facilitem o diagnóstico da epilepsia e o seu tratamento se tornam válidos e necessários. A seguir listaremos alguns trabalhos relacionados aos desenvolvimento dessa pesquisa.

## 1.1 Trabalhos relacionados

Uma das principais bases de dados para apoio de pesquisas relacionadas a epilepsia é a CHB-MIT [31]. Tal base fundamentou trabalhos que se tornaram as principais referências sobre a identificação e classificação de segmentos de crises epiléticas na última década. Dentre os primeiros trabalhos estão os de Shoeb, que inicialmente apresentou uma bordagem baseada em aprendizado de máquina, afim de estabelecer classificadores específicos para detectar o início de crises epiléticas de determinado paciente [32]. Obteve então uma taxa de acerto de 96% em 173 convulsões, com latência média de 3 segundos. Em uma segunda proposta foi abordada uma problemática inversa, um modelo baseado em aprendizado de máquina para detecção do término de crises epiléticas [34]. O qual detectou o final de 132 crises em 133, com latência em  $10.3 \pm 5.5$  segundos. Ao combinar os modelos de detecção de início e término de crises foi possível estimar a duração de 85% delas, com latência média de 15 segundos.

Em demais contribuições podemos destacar Khan [21], o qual se baseou na extração de características como coeficiente de variação, coeficiente de variação normalizado e coeficiente de variação relativo. Utilizando o classificador k-NN foi obtido um percentual de acurácia de 98% na detecção automática de crises.

Fergus [10] propôs um modelo baseado em aprendizado de máquina supervisionado para a classificação de registros com e sem crises epiléticas. Seu modelo obteve melhores resultados utilizando também o classificador k-NN, alcançando métricas de classificação de 93% de sensibilidade, 94% de especificidade e acurácia de 98%.

Também em um modelo baseado em aprendizado de máquina, Nandy apresentou uma metodologia para detecção automática de crises epiléticas [26]. Utilizando uma mescla de características extraídas dos domínios do tempo e da frequência, juntamente com o classificador SVM com kernel RBF, foi obtido o percentual de 97.05% de acurácia.

Como supracitado, a comunidade acadêmica desprende uma variedade de esforços para identificação crises baseadas e análise de padrões eletroencefalográficos. Ocasionalmente em possíveis opções para acompanhamento no tratamento de pacientes que sofrem de crises epiléticas, bem como o auxílio ao diagnóstico de pessoas com suspeita de epilepsia.

Ressalta-se que a maioria das metodologias propostas para classificação de segmentos eletroencefalográficos limita-se a uma classificação de caráter individual. Em outras palavras, classificadores treinados com dados de um paciente específico, que apresentam

pouca ou nenhuma generalização.

## **1.2 Objetivo Geral**

Desenvolver um modelo, baseado em aprendizado de máquina, para a classificação generalizada de segmentos de eletroencefalogramas com ou sem crises epiléticas.

## **1.3 Objetivos Específicos**

- Avaliar resultados de classificação de vetores de características;
- Realizar comparativo entre resultados deste trabalho e resultados anteriormente propostos.

## 2 EPILEPSIA E ELETROENCEFALOGRAFIA

Crises epilépticas se conceituam como ocorrências transitórias de sinais, dada uma intensa atividade elétrica neuronal no cérebro [36]. Tal atividade é capaz de provocar em alterações na consciência até eventos involuntários de natureza musculares, psíquica, sensitiva e autonômica.

De acordo com a International League against Epilepsy (ILAE), epilepsia defini-se como a predisposição persistente do cérebro de gerar crises epilépticas recorrentes [37, 38]. É definido ainda um estado de epilepsia pela ocorrência de até duas crises não provocadas em um intervalo de mais de 24 horas, desde que o cérebro demonstre alguma condição que predisponha crises, tal como lesões cerebrais ou anormalidades eletroencefalográficas. Havendo identificação desses fatores o profissional médico é autorizado a prosseguir com o tratamento de um paciente [16].

### 2.1 Crises Epilépticas

Há grande variedade de manifestações de crises epilépticas, devido a seu caráter imprevisível e transitório. Podem ocorrer subitamente e com curta duração, porém há casos em o indivíduo pode ser submetido a crises por mais de 5 minutos. Sendo assim, a Commission on Classification and Terminology of the International League Against Epilepsy, definiu uma classificação em três grupos [2], Figura 2.1.

**Figura 2.1:** Classificação Internacional das Crises Epilépticas.



**Fonte:** Yacubian, 2014.

Crises parciais (focais ou locais) são aquelas que tem por distinção fundamental a

preservação de consciência em seu princípio. São subdivididas em crises parciais simples (caracterizadas por sinais motores, alucinações e sintomas autonômicos), crises parciais complexas (iniciam com características de crises simples seguindo até leve alteração na consciência) e crises secundariamente generalizadas (que evoluem gradativamente do tipo simples, ao tipo complexo e se estabelecendo em uma crise generalizadas).

Crises generalizadas são caracterizadas pelo envolvimento dos dois hemisférios cerebrais. Se originam em alguma região específica do cérebro e rapidamente se propagam para outras. Assim como as crises parciais, as crises generalizadas também têm sua subdivisão [38], descrita resumidamente a seguir.

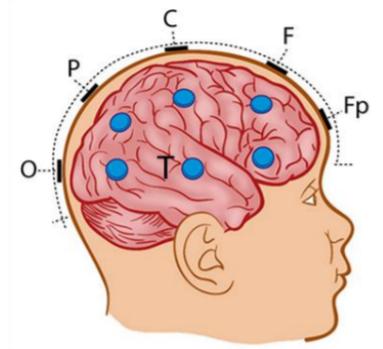
- *Crises de Ausência:* Constituem em breves lapsos de consciência com início e término abruptos. Acompanhadas por leves manifestações motoras, piscamento, aumento e diminuição de tônus musculares. Geralmente manifestadas durante a infância;
- *Crises Mioclônicas:* Caracterizadas por breves contrações musculares, com duração menor que 200 ms. As contrações musculares não são rítmicas, atingindo diferentes grupos musculares [38] ;
- *Crises Atônicas:* Têm a característica de perda do tônus da musculatura postural de forma súbita com duração de um a dois segundos. As manifestações musculares podem ir desde discreta queda da cabeça até uma súbita queda ao solo. Por tais características o seu reconhecimento clínico é dificultado se o paciente estiver sentado ou deitado [38];
- *Crises Tônicas:* Caracterizadas por contração muscular axial ou global, com duração média de cinco a dez segundos. A perda de consciência ocorre frequentemente e a sua recuperação coincide com o final da descarga eletroencefalográfica [38];
- *Crises Tônico-Clônicas:* Constituem em duas fases bem definidas. A fase tônica em que o paciente estende os membros superiores e inferiores, mantendo os braços e cruzados em frente ao corpo. A fase tônica que dura de cinco a dez segundos e é seguida por um tremor sutil produzido pelas flexões pequenas e rápidas ao nível do cotovelo [38].

## 2.2 Eletroencefalograma - EEG

O eletroencefalograma (EEG) trata-se de um método utilizado para registrar a atividade elétrica do cérebro [35]. Através de eletrodos que são distribuídos no couro cabeludo do paciente ocorre a aquisição dos valores de tensão do cérebro. O posicionamento dos eletrodos no couro cabeludo do paciente é chamado de montagem, que são classificadas pelos sistemas internacionais 10-20 e 10-10 [24].

A montagem 10-20 é composta por 21 eletrodos, na qual cada eletrodo possui denominação padrão composta por uma letra, referente a região cerebral coberta pelo eletrodo, e um número indicando a lateralização. Então, temos Fp para região fronto-polar, F para frontal, C para central, P para região parietal, T para temporal e O para occipital. A Figura 2.2 ilustra as respectivas siglas referentes as regiões do cérebro.

**Figura 2.2:** Nomenclatura de regiões cerebrais.



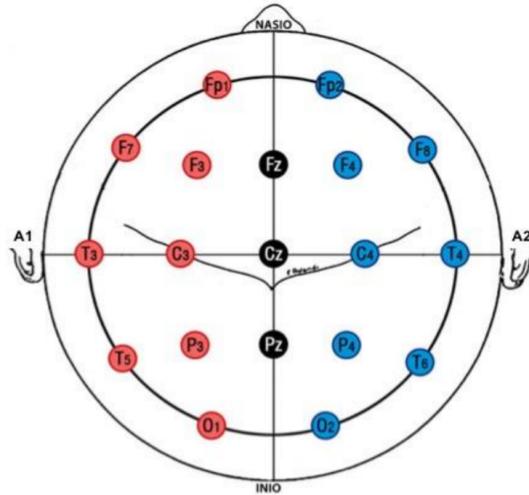
**Fonte:** Marinho, 2017.

Resta ainda diferenciar as áreas do hemisfério cerebrais, agora utilizando números pares para o hemisfério direito (Fp2, F4, F8, C4, P4, T4, T6 e O2) e ímpares no hemisfério esquerdo (Fp1, F3, F7, C3, P3, T3, T5 e O1). Eletrodos da região média da cabeça têm a denominação Z (z de zero), sendo eles: Fz, Cz e Pz. E os eletrodos auriculares chamados de A1 e A2, respectivamente, esquerda e direita. A Figura 2.3 ilustra por completo a montagem de EEG no padrão 10-20.

O sistema 10-10 tem a finalidade de uso em registros prolongados de atividade neural ou para estudos de regiões específicas do cérebro. A colocação dos eletrodos é baseada na mesma referência do sistema 10-20, porém a adição de eletrodos intermediários. Ao todo o sistema usa 75 eletrodos (21 do sistema 10-20 mais 54 intermediários). Com a finalidade de adaptação aos eletrodos intermediários, quatro eletrodos do sistema 10-20 tiveram sua nomenclatura altera, tal como a Tabela 2.1.

A distribuição de eletrodos no couro cabeludo definida pelo sistema 10-10 para montagem de EEG é ilustrada pela Figura 2.4.

**Figura 2.3:** Montagem de EEG no sistema 10-20.

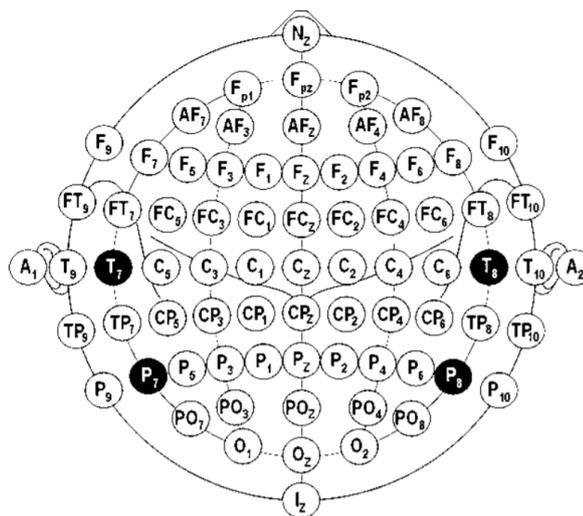


**Fonte:**Marinho, 2017.

**Tabela 2.1:** Mudanças de nomenclaturas de eletrodos do sistema 10-20 para o sistema 10-10

Sistema 10-20	Sistema 10-10
T3	T7
T4	T8
T5	P7
T6	P8

**Figura 2.4:** Montagem de EEG no sistema 10-10.

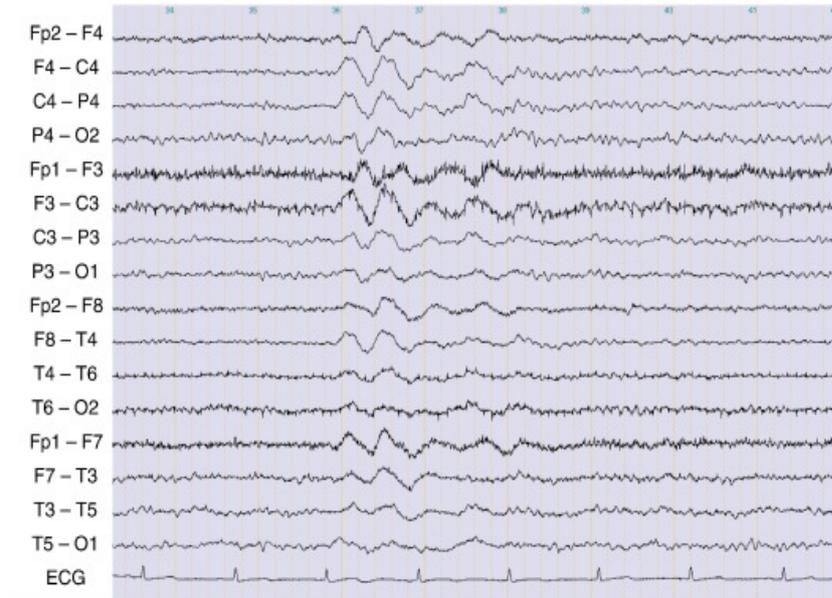


**Fonte:** Marinho, 2017.

A gravação de EEG no couro cabeludo exibe a diferença de potenciais elétricos entre

duas regiões diferentes no córtex cerebral sobrejacente da cabeça que é o mais próximo do eletrodo de registro [35]. Onde cada canal do EEG corresponde a diferença de potencial elétrico dos eletrodos que cobrem diferentes regiões do cérebro em função do tempo, tal como a Figura 2.5.

**Figura 2.5:** Ilustração gráfica do eletroencefalograma.



**Fonte:** Hussain, 2017.

A partir dos dados coletados pelo EEG é possível realizar análises clínicas visuais no comportamento neural de um paciente, bem como, realizar estudos mais detalhados através dos valores numéricos de tensão obtidos.

### 3 MOMENTOS ESTATÍSTICOS

Um momento estatístico de ordem  $n$  de uma variável aleatória  $X$ , é definido pela seguinte integral [28]:

$$m_n = E\{\mathbf{x}^n\} = \int_{+\infty}^{-\infty} x^n f(x) dx \quad (3.1)$$

Sendo também  $f(x)$  a função de densidades de  $X$  e  $E$  o operador esperança. Partindo disso é definido também o  $k$ -ésimo momento central ( $k > 1$ ) 3.2

$$\mu_k = E\{(\mathbf{x} - m_1)^k\} \quad (3.2)$$

#### 3.1 O Segundo Momento Estatístico e a Variância

A variância de uma variável aleatória  $X$  é definida como o momento central de segunda ordem, de modo que:

$$\sigma^2 = \mu_2. \quad (3.3)$$

Partindo disso é possível estabelecer a distância de cada valor do conjunto em relação ao valor médio [9].

#### 3.2 O Terceiro Momento Estatístico e a Assimetria

A assimetria, também denominada como coeficiente de assimetria, define o grau de assimetria de uma distribuição de probabilidades. Em outras palavras trata-se do afastamento de uma distribuição da unidade de simetria [11]. É definida por:

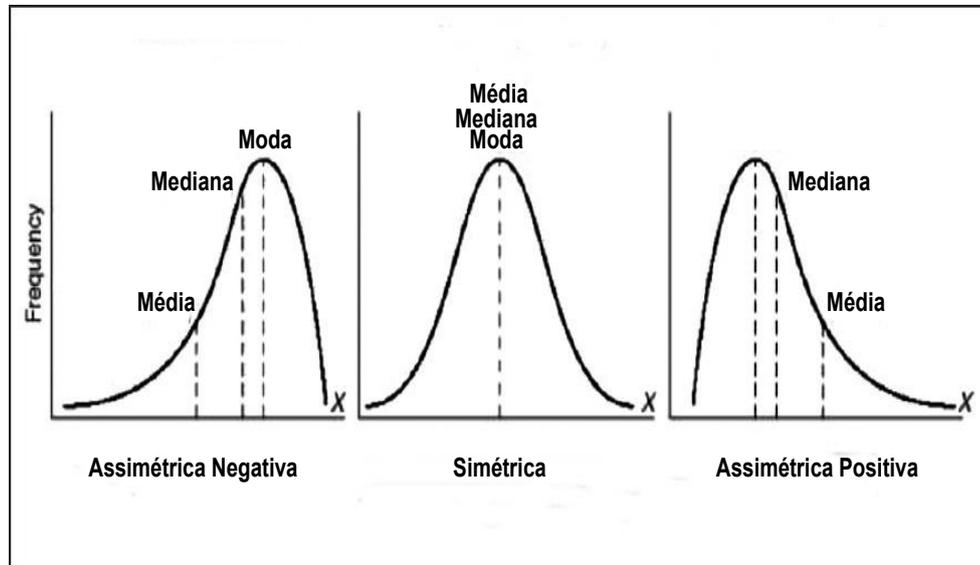
$$\gamma = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} \quad (3.4)$$

O coeficiente de assimetria pode ainda ser classificado levando em consideração os valores de média ( $\mu$ ), mediana( $\tilde{x}$ ) e moda( $m_o$ ):

- *Simétrica*: para  $\mu = \tilde{x} = m_o$ .
- *Assimétrica positiva*: para  $m_o < \tilde{x} < \mu$ .
- *Assimétrica negativa*: para  $\mu < \tilde{x} < m_o$ .

A Figura 3.1 ilustra a classificação de distribuições com base em assimetria.

**Figura 3.1:** Classificação de função de distribuição a partir da assimetria.



Fonte: Adaptada pelo autor

Observações importantes devem ser feitas a cerca da assimetria:

- Uma distribuição simétrica não necessariamente será uma distribuição gaussiana ou assumirá o formato de sino;
- Dados reais muitas vezes apresentam valores extremos em uma cauda;
- Dados reais podem ter distribuições bimodais (e multimodais).

### 3.3 O Quarto Momento Estatístico e a Curtose

A curtose, também chamada de coeficiente de curtose, trata-se da medida de intensidade dos picos de uma distribuição de probabilidades. A mesma é definida por:

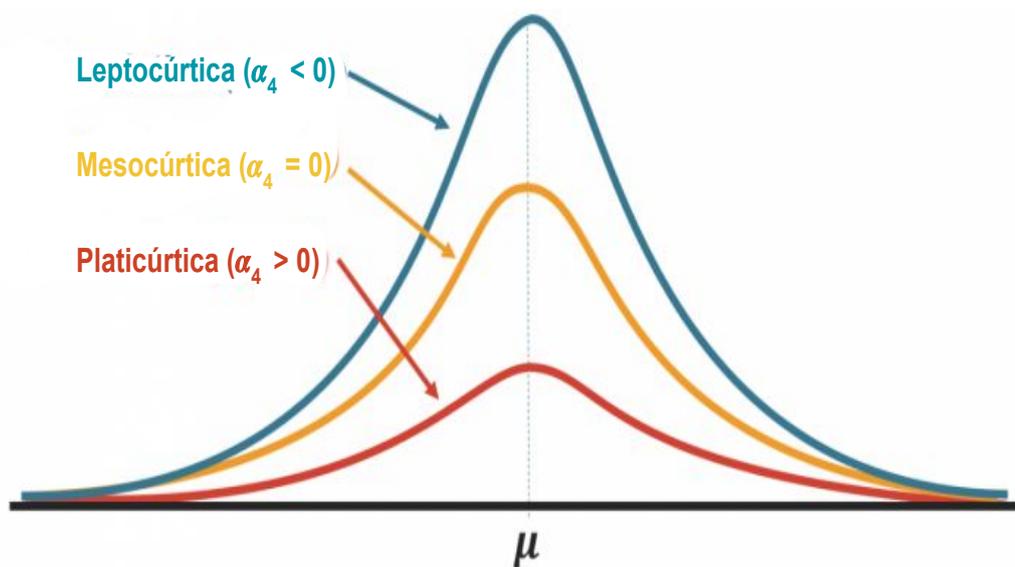
$$\kappa = \frac{\mu_4}{\mu_2^2} - 3 \quad (3.5)$$

Em outras palavras a partir do coeficiente de curtose é possível indicar a concentração dos valores da distribuição em torno do centro desta. Tendo obtido esse valor a função de de distribuição de probabilidades pode ser classificador em:

- *Mesocúrtica*: para  $\alpha_4 = 0$ , a função de distribuição tem o mesmo achatamento da distribuição normal.
- *Leptocúrtica*: para  $\alpha_4 > 0$ , a função de distribuição possui a curva da função de distribuição mais afunilada com um pico mais alto do que a distribuição normal.
- *Platicúrtica*: para  $\alpha_4 < 0$ , a função de distribuição é mais achatada do que a distribuição normal.

A Figura 3.2 ilustra a classificação da função de distribuição com base nos valores de curtose.

**Figura 3.2:** Classificação de função de distribuição a partir da curtose.



Fonte: Portal Action: curtose

## 4 CLASSIFICADORES

### 4.1 k-Vizinhos Próximos - k-NN

O k-vizinho mais próximo (k-NN, do inglês k-Nearest Neighbor ) está entre os mais populares e mais simples algoritmos de aprendizado de máquina. Em síntese, ocorre a memorização do conjunto de treinamento e ao ser inserido um novo dado, a este será atribuído o rótulo de seus vizinhos mais próximos [14]. Inicialmente a proximidade de um novo dado em relação aos seus vizinhos é definida pela distância euclidiana de seus vetores de atributos [8]

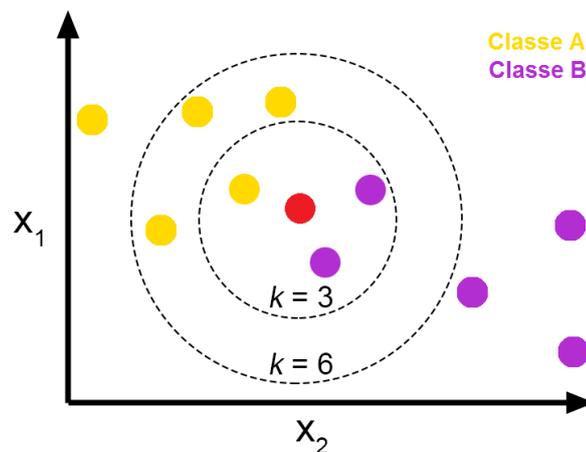
$$d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_{l=1}^d (x_i^l - x_j^l)^2} \quad (4.1)$$

onde  $\mathbf{x}_i$ ,  $\mathbf{x}_j$  são objetos representados por vetores em um determinado espaço, e  $x_i^l$  e  $x_j^l$  são elementos dos mesmo vetores correspondentes a valores da coordenada  $l$  (atributos).

É estabelecido um valor para  $k$ , o qual deve ser um número ímpar e maior que 1. Com base nas distâncias euclidianas já calculadas é atribuído ao novo dado o rótulo de maior ocorrência entre os  $k$  vetores de atributos com menor distância em relação ao novo dado.

Um breve exemplo de classificação com o algoritmo  $k$ -NN [20] é ilustrado pela figura 4.1.

**Figura 4.1:** Ilustração de classificação com o  $k$ -NN.



Fonte: José, 2018.

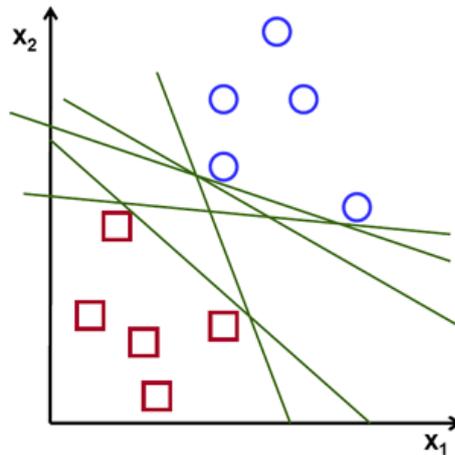
Tendo em vista o cálculo da distância euclidiana, ao novo dado será atribuído o rótulo da Classe B se  $k = 3$ , porém tendo  $k = 6$ , o novo dado pertencerá a Classe A.

Como dito anteriormente, o k-NN trata-se de um algoritmo de fácil implementação, podendo ser aplicável em problemas complexos de classificação e naturalmente incremental, visto que basta apenas inserir novos exemplos de treinamento na memória [8]. Um aspecto negativo desse classificador é que seu método de aprendizado é preguiçoso (*lazy*), não é aplicada uma discriminação de dados mas há apenas a memorização dos dados de treino.

## 4.2 Máquina de Vetores de Suporte - SVM

O objetivo de classificadores binários e lineares é estabelecer um limite satisfatório entre duas classes. Porém tais limites podem não ser ótimos, tais como a Figura 4.2 <sup>1</sup>

**Figura 4.2:** Limites de classificação não otimizados.



**Fonte:** OpenCV: Introduction to Support Vector Machines.

A máquina de vetores de suporte (Support Vector Machine - SVM) visa estabelecer o hiperplano de separação ideal o qual maximiza a margem da base de treinamento [27].

Temos  $\mathbf{X}$  como um conjunto de dados com  $n$  objetos  $\mathbf{x}_i$ , seus rótulos  $i_i$  e  $X$  como espaço de entrada e  $Y = -1, +1$ . Partindo do pressuposto que  $\mathbf{X}$  é linearmente separável sendo possível separar as classes  $-1$  e  $+1$  por um hiperplano.

A equação do hiperplano é definida por

$$h(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b, \quad (4.2)$$

sendo  $\mathbf{w}$  o vetor normal ao hiperplano e  $b$  a distância do hiperplano em relação à origem [8].

Podemos então dividir o espaço  $X$  em regiões  $\mathbf{w} \cdot \mathbf{x} + b > 0$  e  $\mathbf{w} \cdot \mathbf{x} + b < 0$ . Para

<sup>1</sup>Disponível em: <https://docs.opencv.org>

isso é definida uma função sinal  $g(\mathbf{x})$  que será a saída da função *sigmoid*, tendo como parâmetro  $h(\mathbf{x})$ , tal como a equação a seguir

$$g(\mathbf{x}) = \text{sgn}(h(\mathbf{x})) \begin{cases} +1, & \text{se } \mathbf{w} \cdot \mathbf{x} + b > 0 \\ -1, & \text{se } \mathbf{w} \cdot \mathbf{x} + b < 0 \end{cases} \quad (4.3)$$

Define-se então, o hiperplano canônico em relação ao conjunto  $X$ , de forma que exemplos próximos ao hiperplano  $\mathbf{w} \cdot \mathbf{x} + b = 0$  satisfaçam a equação

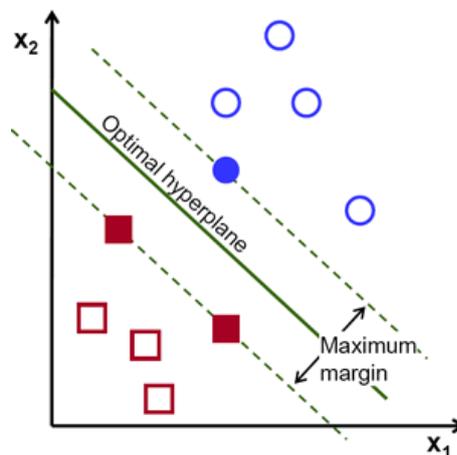
$$|\mathbf{x} \cdot \mathbf{x}_i + b| = 1. \quad (4.4)$$

As condições são resumidas pela seguinte expressão:

$$y_i(\mathbf{x} \cdot \mathbf{x}_i + b) - 1 \geq 0, \forall (\mathbf{x}_i, y_i) \in X \quad (4.5)$$

De forma que é possível a separação de classes por um hiperplano otimizado, estabelecendo a maior distância entre eles, tal como a Figura 4.3 <sup>2</sup>.

**Figura 4.3:** Ilustração de hiperplano de classificação do SVM.

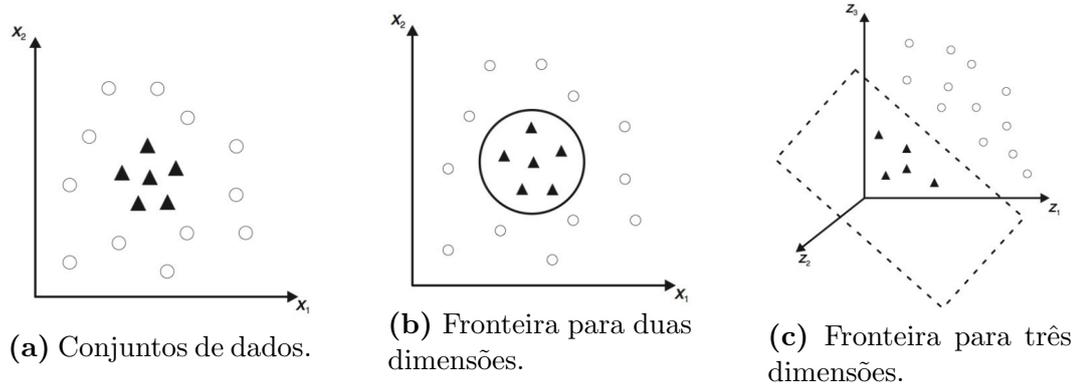


**Fonte:** OpenCV: Introduction to Support Vector Machines.

Há casos nos quais não é possível dividir os dados satisfatoriamente por um hiperplano de treino. Essa situação é exemplificada pela Figura 4.4, em que o uso de uma curva é mais adequado para fazer a separação de classes.

<sup>2</sup>Disponível em: <https://docs.opencv.org>

**Figura 4.4:** Transformação realizada em conjunto de dados não linear.



**Fonte:** aqui vai a fonte

É necessário então realizar o mapeamento do conjunto de treinamento em seu espaço original, referenciado como de entradas, para um novo espaço de maior dimensão, este denominado de espaço da características (*feature space*) [18]. Esse procedimento é chamado de teorema de Cover [23], definido por:

$$\Phi : X \rightarrow \mathfrak{S} \quad (4.6)$$

No qual  $X$  corresponde ao espaço de entradas e  $\mathfrak{S}$  ao espaço de características. por uma escolha adequado de  $\Phi$ , o teorema afirma que o espaço  $X$  pode ser transformado em um espaço de características  $\mathfrak{S}$  no qual há alta probabilidade dos objetos serem linearmente separáveis. De modo que, precisão serem satisfeitas duas duas condições:

- a transformação precisa ser não linear;
- a dimensão do espaço de características seja suficientemente alta.

Sabendo que  $\mathfrak{S}$  pode ter alta dimensão, ou mesmo infinita, o cálculo de  $\Phi$  pode ser extremamente custoso. Para facilitar esse processo são utilizadas as funções kernel  $K$ , as quais recebem dois ponto  $\mathbf{x}_i$  e  $\mathbf{x}_j$  no espaço de entradas e obtém-se o produto escalar desse objetos no espaço de características [19], tendo então:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j). \quad (4.7)$$

A Tabela 4.1 lista as funções mais populares utilizadas em SVM's não lineares.

**Tabela 4.1:** Tabela com funções kernel mas comuns

Tipo de kernel	Função $K(\mathbf{x}_i \cdot \mathbf{x}_j)$	Parâmetros
Polinomial	$(\delta(\mathbf{x}_i \cdot \mathbf{x}_j) + \kappa)^d$	$\delta, \kappa$ e $d$
RBF	$exp(-\delta \ \mathbf{x}_i - \mathbf{x}_j\ ^2)$	$\delta$
Sigmoidal	$tanh(\delta(\mathbf{x}_i \cdot \mathbf{x}_j) + \kappa)$	$\delta$ e $\kappa$

### 4.3 Naive Bayes

Trata-se de um popular algoritmo de classificação também conhecido como “*Idiot’s Bayes*” [15]. Baseado na teoria bayesiana que parte da suposição de que cada característica de uma classe em particular é independente de qualquer uma das outras características. Seu processo de classificação ocorre realizado conforme a seguir:

Havendo um conjunto de treinamento para  $n$  classes com vetor de características  $Y$  e rótulos de classe associados. O atributo  $Y$  pertence a classe com maior probabilidade a *posteriori* e é dada como

$$P(C_i|Y) > P(C_j|Y) \text{ para } 1 \leq j \leq n, j \neq i \quad (4.8)$$

onde

$$P(C_i|Y) = \frac{P(Y|C_i)P(C_i)}{P(Y)} = P(y_1|C_i) \times \dots \times P(y_n|C_i). \quad (4.9)$$

Tomando por base a Teoria de Bayes temos que  $P(C_i)$  são as probabilidades a *priori* da classe,  $P(Y)$  é a probabilidade a *priori* de  $Y$ ,  $P(C_i|Y)$  é a probabilidade posterior e  $P(Y|C_i)$  é a probabilidade a *posteriori* de  $Y$  condicionada em  $C_i$ . Sendo  $P(Y)$  uma constante para todas as classes, apenas o numerador de  $P(C_i|Y)$  precisa ser maximizado. Afim de reduzir o cálculo de uma estimativa de  $P(Y|C_i)$ , o classificador assume que os atributos são condicionalmente independentes uns dos outros. De modo que

$$P(Y|C_i) = \prod_{k=1}^n P(y_k|C_i). \quad (4.10)$$

As probabilidades  $P(y_1|C_i), \dots, P(y_n|C_i)$  são determinadas a partir do conjunto de treinamento e  $y_k$  representam o valor de um atributo para o conjunto de dados  $Y$ . Para estimar o rótulo de classe de  $Y$ ,  $P(Y|C_i)P(C_i)$  é avaliado para cada classe  $C_i$ . O classificador identifica o rótulo da classe de o atributo  $Y$  é  $C_i$  com base na condição fornecida abaixo:

$$P(Y|C_i)P(C_i) > P(Y|C_j)P(C_j) \text{ para } 1 \leq j \leq n, j \neq i \quad (4.11)$$

#### 4.4 Análise Quadrática Discriminante - QDA

A Análise Quadrática Discriminante é um algoritmo de classificação binário, que visa definir uma transformação das características de entrada [39]. QDA parte do pressuposto os dados obedecem uma distribuição normal multivariada, cujo a função de densidade de probabilidade é definida por

$$p(t) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(t-u)^T \Sigma^{-1}(t-u)\right\} \quad (4.12)$$

onde  $t$  é um vetor  $p$ -dimensional,  $u$  o vetor médio e  $\Sigma$  é uma matriz de covariância de dimensão  $p \times p$ . A probabilidade a posteriori é obtida pelo critério Bayesiano

$$p(\omega|t) = \frac{p(t|\omega_i)p(\omega_i)}{p(t)}, i = 1, \dots, M \quad (4.13)$$

em que  $p(\omega|t)$  é a função de densidade de probabilidade da amostra,  $p(\omega)$  a probabilidade a priori da classe  $\omega_i$  e  $p(t)$  é uma função de densidade mista independente de classe. Aplica-se então o logaritmo da equação  $p(t|\omega)p(\omega_i)$ . Se as probabilidades anteriores de todas as classes são iguais, todo  $p(\omega_i)$  é igual para todo  $i$ . Remove-se o item não relacionado a  $i$  e é obtida a função discriminante quadrática

$$d(t) = p(\omega_i|t) = (t - u_i)^T \sum_i^{-1} (t - u_i) + \ln |\sum_i| \quad (4.14)$$

#### 4.5 Rede Neural Artificial Perceptron de Múltiplas Camadas - RNA - MLP

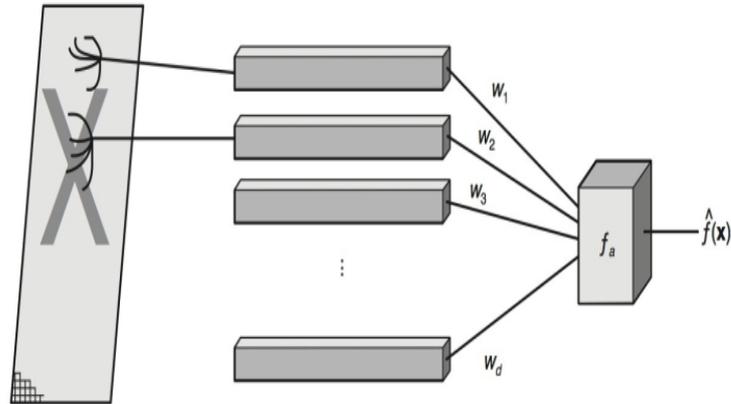
O *Perceptron* trata-se da arquitetura mais simplista de redes neurais. Tendo sua simplicidade observada pela sua constituição contendo apenas uma camada neural, havendo somente um único neurônio artificial na camada [30]. A Figura 4.6 ilustra a arquitetura de uma rede neural *Perceptron*.

Para o seu treino usa-se um algoritmo supervisionado de correção de erro e uma função de ativação. Na qual para um objeto  $\mathbf{x}_i$  os pesos são ajustados de acordo com:

$$w_j(t+1) = w_j(t) + \eta x_i^j (y_i - \hat{f}(\mathbf{x}_i)) \quad (4.15)$$

Onde  $w_j(t)$  trata-se do peso da  $j$ -ésima conexão de entrada no instante de tempo

**Figura 4.5:** Arquitetura de rede neural *Perceptron*.



**Fonte:** OpenCV: Introduction to Support Vector Machines.

$t$ ,  $\eta$  corresponde a taxa de aprendizagem,  $x_i^j$  é o valor do  $j$ -ésimo atributo do vetor de entrada  $\mathbf{x}_i$ ,  $\mathbf{x}_i$  refere-se a saída da rede neural no instante  $t$  e por fim,  $y_i$  é a saída desejada da rede (o rótulo de  $\mathbf{x}_i$ ) [8].

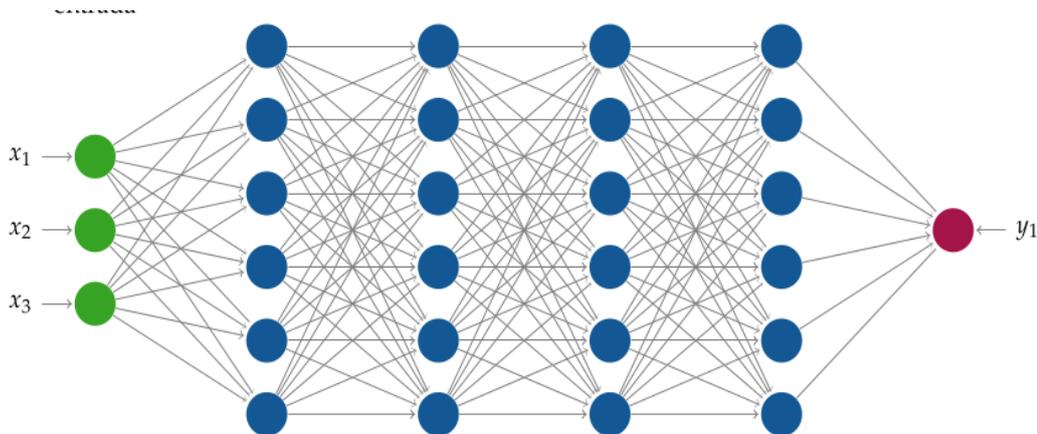
Com relação a função de ativação podemos destacar as de uso mais comum:

- Identidade:  $f(z) = z$ ;
- Logística:  $f(z) = \frac{1}{1+e^{-z}}$ ;
- Tangente Hiperbólica:  $f(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$ ;
- ReLu:  $f(z) = \max(0, z)$ ;
- Leaky ReLu:  $f(n) = \begin{cases} 0.01 & \text{se } z < 0 \\ z & \text{se } z \geq 0 \end{cases}$

Entretanto uma simples rede *Perceptron* é uma alternativa aplicável a classificação de dados linearmente separáveis. Para dados de maior complexidade e de caráter não linear existe a alternativa de adicionar uma ou mais camadas intermediárias, as quais podem permitir a aproximação de qualquer função. São as chamadas redes neurais *Perceptron* multicamandas (MLP, do inglês *multilayer perceptron*) [6], tendo sua arquitetura ilustrada pela Figura.

A partir de 4.15 e sabendo apenas os pesos para neurônios da camada de saída, é necessária a estimação de pesos de camadas intermediárias. Para isso usa-se o algoritmo *back-propagation*, que estima o erro dos neurônios de determinada camada observando o erro dos neurônios de uma camada posterior [17]. De forma que o erro da camada em que se encontra o neurônio é calculado a seguir:

**Figura 4.6:** Exemplo de Rede Neural MLP



**Fonte:** OpenCV: Introduction to Support Vector Machines.

$$\delta_l = \begin{cases} f'_a e_l, & \text{se } n_l \in C_{sai} \\ f'_a \sum w_{lk} \delta_k, & \text{se } n_l \in C_{int} \end{cases} \quad (4.16)$$

De forma que  $n_l$  é o  $l$ -ésimo neurônio,  $C_{sai}$  é a camada de saída,  $C_{int}$  representando a camada de intermediária,  $f'_a$  sendo a derivada parcial da função de ativação do respectivo neurônio e por fim  $e_l$  sendo o erro quadrático cometido pelo neurônio de saída. Sendo o erro quadrático definido pela seguinte forma:

$$e_l = \frac{1}{2} \sum_{q=1}^k (y_q - \hat{f}_q)^2 \quad (4.17)$$

A derivada aplicada é capaz de medir a contribuição de cada peso no erro da rede para a classificação de uma objeto  $\mathbf{x}$ . Se o resultado for um peso positivo, o peso prova um aumento na diferença entre a saída da rede e a saída desejada. E sendo a derivada negativa, o peso contribui para que a saída da rede seja mais próxima da desejada.

## 5 ANÁLISE DE COMPONENTES PRINCIPAIS- PCA

A análise de componentes principais (PCA) se refere ao processo pelo qual os componentes principais são calculados e o uso subsequente desses componentes na compressão dos dados [7]. O PCA é uma abordagem não supervisionada, pois envolve apenas um conjunto de recursos  $x_1, x_2, \dots, x_p$ , e nenhuma resposta associada  $y$ .

Em uma abordagem simplista, o processo de aquisição de componentes principais se inicia com o cálculo da covariância entre no mínimo duas dimensões ( $X$  e  $Y$ ) [25]:

$$\text{cov}(X, Y) = \frac{\sum_{i=1}^n [(X_i - \bar{X})(Y_i - \bar{Y})]}{n} \quad (5.1)$$

onde  $X$  e  $Y$  correspondem a listas de dados, em que  $X$  é a primeira e  $Y$  a segunda dimensão,  $\bar{X}$  e  $\bar{Y}$  suas médias, e por fim  $n$  o número de dados.

Havendo mais de duas dimensões é necessária a aplicação da covariância em cada par de dimensões, daí surgindo a matriz de covariância. Onde, por exemplo, tendo três dimensões ( $x$ ,  $y$  e  $z$ ), a matriz de covariância é definida por:

$$\begin{pmatrix} \text{cov}(x, x) & \text{cov}(x, y) & \text{cov}(x, z) \\ \text{cov}(y, x) & \text{cov}(y, y) & \text{cov}(y, z) \\ \text{cov}(z, x) & \text{cov}(z, y) & \text{cov}(z, z) \end{pmatrix} \quad (5.2)$$

Em seguida, para  $M$  amostras de um determinado conjunto de dados, o vetor médio é definido por:

$$m_x = \frac{1}{M} \sum_{i=1}^M x_i. \quad (5.3)$$

E ainda tendo a matriz de covariância de um certo conjunto de dados com  $M$  amostras, temos que o vetor médio de 5.3 é dado por:

$$C_x = \frac{1}{M} \sum_{i=1}^M x_i x_i^T - m_x m_x^T \quad (5.4)$$

$$y = \mathbf{A}(x - m_x) \quad (5.5)$$

Além de produzir variáveis derivadas para uso em problemas de aprendizagem supervisionada, o PCA também serve como ferramenta para visualização de dados (visualização das observações ou visualização das variáveis).

## 6 MATERIAIS E MÉTODOS

Neste capítulo está descrita toda a metodologia adotada no presente trabalho. Desde as etapas de pré-processamento, extração de características e classificação.

### 6.1 Base de Dados

Neste trabalho tratamos da base de dados do Children's Hospital Boston - CHB-MIT [31], a qual contém registros de EEG de pacientes pediátricos com convulsões intratáveis. Afim de avaliar a candidatura dos pacientes para intervenção cirúrgica e também caracterizar suas convulsões, os pacientes por dias foram monitorados após a retirada da sua medicação anticonvulsiva.

A base de dados é constituída por gravações de EEG's de 22 casos coletados de 22 pacientes. Sendo 5 homens, com idades entre 3 e 22 anos; e mulheres com idades entre 1,5 e 19 anos. Para cada caso foram coletados entre 9 e 42 arquivos de *.edf*, com gravações entre uma e quatro horas de duração. Todos os sinais possuem frequência amostral de 256 Hz com resolução de 16 *bits*. Ao todo a base de dados conta 664 registros, sendo 129 rotulados com presença de pelo menos uma crise epilética. Assim foram registrados o total de 198 episódios de crises epiléticas nos arquivos da base. A Tabela 6.1 resume informações sobre cada um dos casos, como gênero, idade e número de crises.

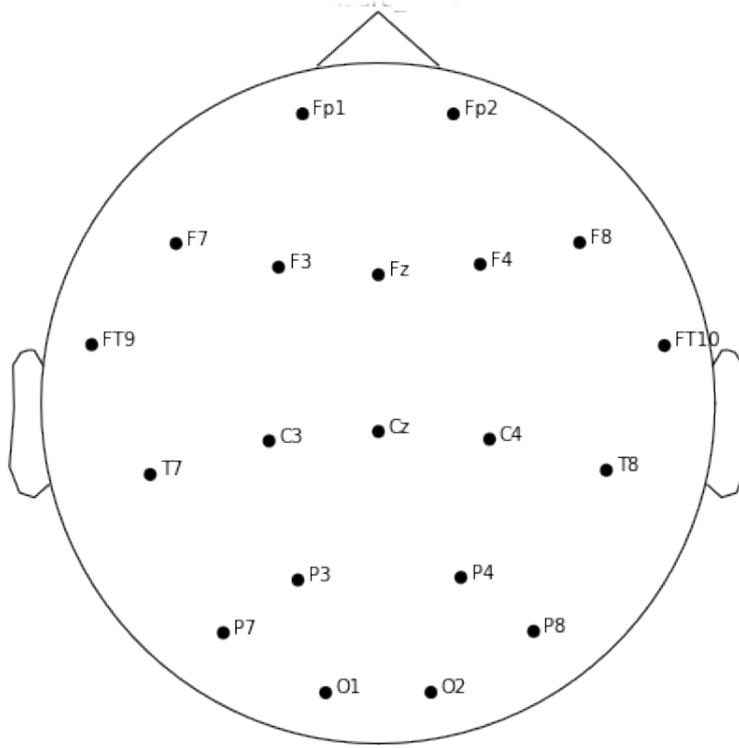
**Tabela 6.1:** Informações quantitativas sobre base de dados CHB-MIT.

Caso	Gênero	Idade	Número de Crises
chb01	F	11	7
chb02	M	11	3
chb03	F	14	7
chb04	M	22	4
chb05	F	7	5
chb06	F	1.5	10
chb07	F	14.5	3
chb08	M	3.5	5
chb09	F	10	4
chb10	M	3	7
chb11	F	12	3
chb12	F	2	27
chb13	F	3	10
chb14	F	9	8
chb15	M	16	20
chb16	F	7	8
chb17	F	12	3
chb18	F	18	6
chb19	F	19	3
chb20	F	6	8
chb21	F	13	4
chb22	F	9	3
chb23	F	6	7
chb24	Desconhecido	Desconhecida	16

## 6.2 Pré-processamento

O processo metodológico deste trabalho se inicia com uma análise exploratória da base de dados. Imediatamente foram identificadas características particulares em alguns EEG's, tais como canais repetidos, canais com sinais gerados artificialmente e também outros tipos de sinais biológicos, tais como eletrocardiogramas e estímulo do nervo vago. Canais semelhantes a descrição anterior foram removidos. Também foram identificadas diferentes configurações de montagens de EEG dentro da base de dados. Foi feita a identificação da montagem com maior ocorrência entre os arquivos (FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8, P8-O2, FZ-CZ, CZ-PZ, P7-T7, T7-FT9, FT9-FT10 e FT10-T8). Nas etapas seguintes, EEG's como montagens diferentes da ilustrada pela Figura 6.1 não foram utilizados.

**Figura 6.1:** Montagem de EEG adotada



**Fonte:** Elaborada pelo autor.

### 6.3 Extração de Características

Tendo posse dos EEG's rotulados como positivos para crises epiléticas, foram retirados seus respectivos períodos com presença de crises, tal como descrito na documentação da base de dados. Os seguimentos retirados foram concatenados, afim de forma uma dada matriz  $\mathbf{A}$ .

$$\mathbf{A}_{m,n} = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{bmatrix} \quad (6.1)$$

Onde  $m$  corresponde ao número de canais contidos no EEG e  $n$  ao número de amostras.

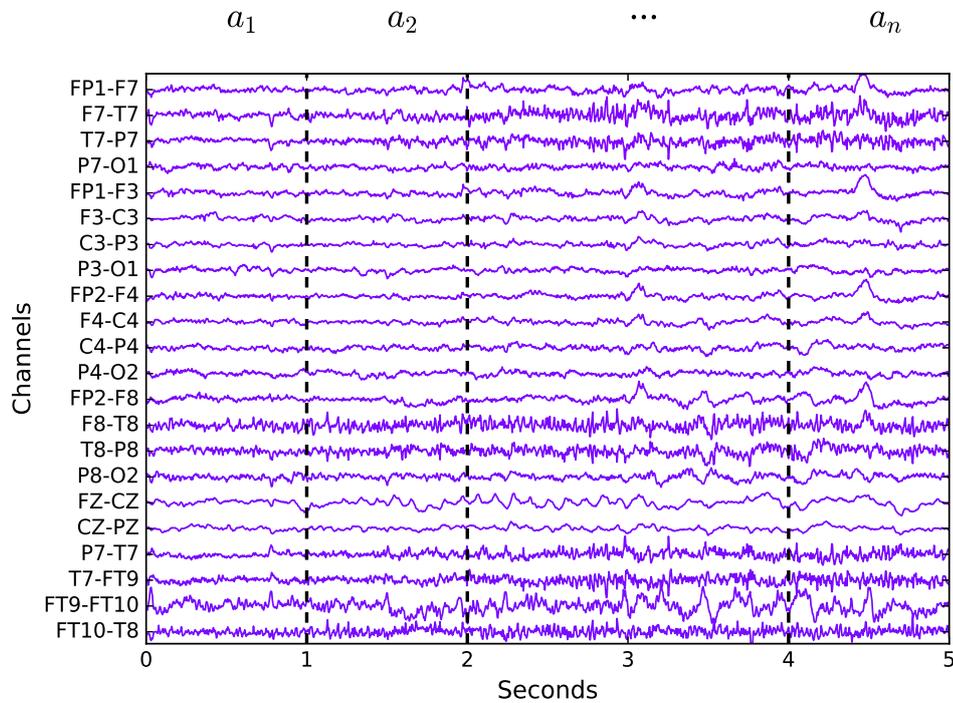
Semelhantemente, dos demais EEG's rotulados sem crises epiléticas foram retirados segmentos de períodos com ausência de crises, gerando a matriz  $\mathbf{B}$ . De forma que a matriz  $\mathbf{B}$  obedece a mesma ordem da matriz  $\mathbf{A}$ .

$$\mathbf{B}_{m,n} = \begin{bmatrix} b_{1,1} & \cdots & b_{1,n} \\ \vdots & \ddots & \vdots \\ b_{m,1} & \cdots & b_{m,n} \end{bmatrix} \quad (6.2)$$

A partir disso foi realizado o cálculo de momentos estatísticos sobre cada uma das janelas de EEG resultante pelo processo de concatenação descrito.

Em cada uma linha de 6.1 e 6.2 foi aplicado um janelamento, que consiste no cálculo produto do período de duração de 1 segundo pela sua frequência amostral de 256 Hz. Tal janelamento é descrito pela Figura 6.2.

**Figura 6.2:** Janelamento aplicado, exemplificado na matriz  $\mathbf{A}$



**Fonte:** Elaborada pelo autor.

E assim foram calculadas as medidas de variância, assimetria e curtose, nas matrizes 6.1 e 6.2. De forma que foram geradas novas matrizes  $\mathbf{A}_{\sigma^2}$ ,  $\mathbf{A}_{\gamma}$ ,  $\mathbf{A}_{\kappa}$  a partir de métricas extraídas de 6.1.

$$\mathbf{A}_{\sigma^2} = \begin{bmatrix} \sigma_{1,1}^2 & \cdots & \sigma_{1,k}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{m,1}^2 & \cdots & \sigma_{m,k}^2 \end{bmatrix} \quad (6.3)$$

$$\mathbf{A}_\gamma = \begin{bmatrix} \gamma_{1,1} & \cdots & \gamma_{1,k} \\ \vdots & \ddots & \vdots \\ \gamma_{m,1} & \cdots & \gamma_{m,k} \end{bmatrix} \quad (6.4)$$

$$\mathbf{A}_\kappa = \begin{bmatrix} \kappa_{1,1} & \cdots & \kappa_{1,k} \\ \vdots & \ddots & \vdots \\ \kappa_{m,1} & \cdots & \kappa_{m,k} \end{bmatrix} \quad (6.5)$$

E semelhantemente,  $\mathbf{B}_{\sigma^2}$ ,  $\mathbf{B}_\gamma$ ,  $\mathbf{B}_\kappa$  geradas pelas características estatísticas de 6.2

$$\mathbf{B}_{\sigma^2} = \begin{bmatrix} \sigma_{1,1}^2 & \cdots & \sigma_{1,k}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{m,1}^2 & \cdots & \sigma_{m,k}^2 \end{bmatrix} \quad (6.6)$$

$$\mathbf{B}_\gamma = \begin{bmatrix} \gamma_{1,1} & \cdots & \gamma_{1,k} \\ \vdots & \ddots & \vdots \\ \gamma_{m,1} & \cdots & \gamma_{m,k} \end{bmatrix} \quad (6.7)$$

$$\mathbf{B}_\kappa = \begin{bmatrix} \kappa_{1,1} & \cdots & \kappa_{1,k} \\ \vdots & \ddots & \vdots \\ \kappa_{m,1} & \cdots & \kappa_{m,k} \end{bmatrix} \quad (6.8)$$

Por fim os vetores de característica são obtidos pelo cálculo da média das colunas de cada uma das matrizes 6.3, 6.4 e 6.5. Como descrito pelas equações a seguir:

$$\mathbf{A}v_{\sigma^2} = \left( \frac{\sum_{i=1}^m \mathbf{A}_{\sigma^2 i,1}}{m}, \dots, \frac{\sum_{i=1}^m \mathbf{A}_{\sigma^2 k,n}}{m} \right), \quad (6.9)$$

$$\mathbf{A}v_\gamma = \left( \frac{\sum_{i=1}^m \mathbf{A}_{\gamma i,1}}{m}, \dots, \frac{\sum_{i=1}^m \mathbf{A}_{\gamma k,n}}{m} \right), \quad (6.10)$$

$$\mathbf{A}v_\kappa = \left( \frac{\sum_{i=1}^m \mathbf{A}_{\kappa i,1}}{m}, \dots, \frac{\sum_{i=1}^m \mathbf{A}_{\kappa k,n}}{m} \right). \quad (6.11)$$

De modo análogo o mesmo processo foi realizado nas matriz 6.6, 6.7 e 6.8

$$\mathbf{B}v_{\sigma^2} = \left( \frac{\sum_{i=1}^m \mathbf{B}_{\sigma^2 i,1}}{m}, \dots, \frac{\sum_{i=1}^m \mathbf{B}_{\sigma^2 k,n}}{m} \right), \quad (6.12)$$

$$\mathbf{B}v_\gamma = \left( \frac{\sum_{i=1}^m \mathbf{B}_{\gamma i,1}}{m}, \dots, \frac{\sum_{i=1}^m \mathbf{B}_{\gamma k,n}}{m} \right), \quad (6.13)$$

$$\mathbf{B}v_\kappa = \left( \frac{\sum_{i=1}^m \mathbf{B}_{\kappa i,1}}{m}, \dots, \frac{\sum_{i=1}^m \mathbf{B}_{\kappa k,n}}{m} \right). \quad (6.14)$$

Assim foram obtidos os vetores de característica oriundos de 6.1 e 6.2, que serão parâmetros de entrada dos algoritmos de classificação apresentados posteriormente.

## 6.4 Classificação

A etapa de classificação consiste em duas partes. A primeira utilizando os vetores de característica dos dois grupos de janelas imediatamente após a sua concepção. E na segunda etapa ocorreu a classificação das componentes do PCA, oriundas a partir dos vetores de característica das janelas de EEG.

Em ambas as etapas foram utilizados os classificadores já mencionados: k-Vizinhos Próximos (k-NN), Máquina de Vetores de Suporte (SVM), Análise Quadrática Discriminante (QDA), Rede Neural Perceptron de Múltiplas Camadas (MLP) e o Naive Bayes. Os hiper-parâmetros definidos em cada um dos classificadores foram definidos da seguinte forma:

- SVM: kernel *RBF*,  $C = 10.0$ ,  $\gamma = \frac{1}{n \times \sigma^2(X)}$  ;
- k-NN:  $k = 5$ ;
- MLP: duas camadas ocultas com 5 e 2 neurônios respectivamente, função de ativação linear retificada e otimização de pesos pela função Adam;
- Naive Bayes: Método Gaussiano;
- QDA: forma padrão;

Nos dois cenários foram definidos 80% dos vetores de característica para o treino dos classificadores e 20% restantes para a etapa de testes. E para validação cruzada foi utilizado o método k-fold [15], com  $k = 10$ .

## 6.5 Métricas de Avaliação

Foram estabelecidas as métricas de desempenho dos classificadores aplicados. Tais métricas são respectivamente a acurácia, sensibilidade e especificidade.

$$ac = \frac{VP + VN}{VP + VN + FN + FP} \times 100. \quad (6.15)$$

$$sens = \frac{VP}{VP + FN} \times 100. \quad (6.16)$$

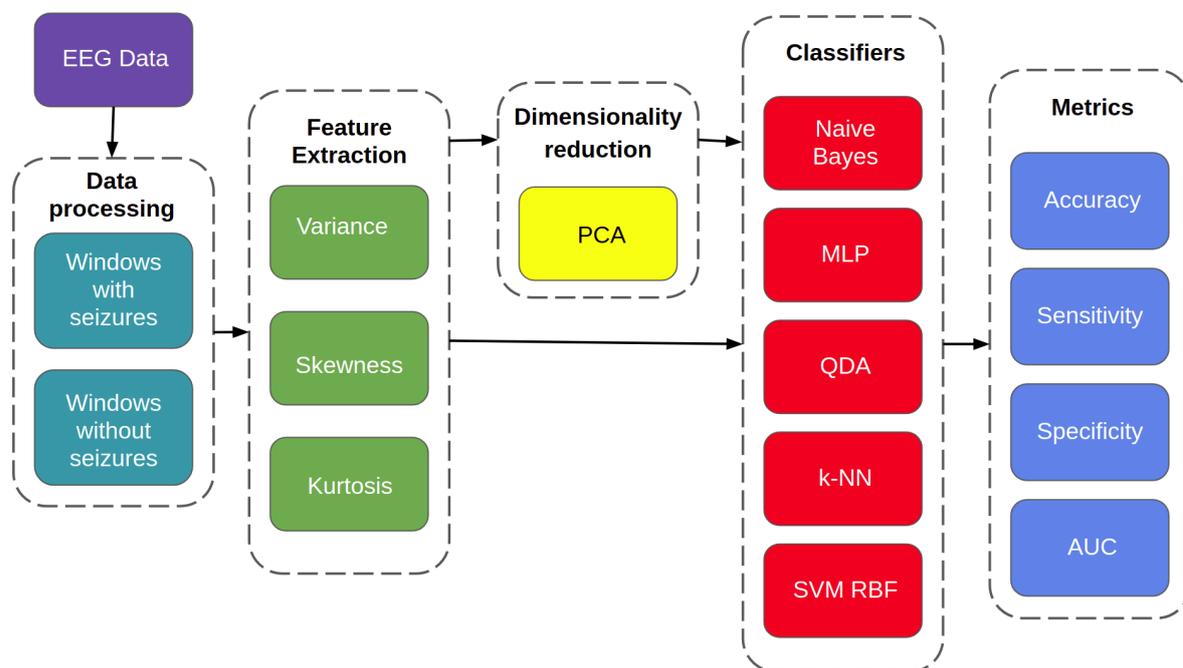
$$esp = \frac{VN}{VN + FP} \times 100. \quad (6.17)$$

De modo que  $VP$  corresponde a quantidade de janelas rotuladas com verdadeiros positivos para presença de crises epilépticas,  $VN$  ao número de verdadeiros negativos para ausência e crises,  $FP$  para registros com crise classificados pelos algoritmos com ausência de crises e  $FN$  para classificações positivas para registros sem episódios de epilepsia.

Afim de estabelecer um melhor comparativo entre o desempenho de cada classificador, adotou-se ainda as curvas ROC (*Receiving Operating Characteristics*) [8]. Trata-se de um espaço bidimensional com eixos representados pela taxa de falsos positivos ( $TFP$ ) e a taxa de verdadeiros positivos ( $TVP$ ). De modo que o desempenho de cada classificador toma por base a área abaixo da curva ROC: AUC (Area Under ROC Curve). Tal medida produz valores entre 0 e 1, de modo que valores próximos de 1 são considerados melhores.

Por fim, a sequência de passos da metodologia proposta nesse trabalho, desde a extração de características até as etapas de classificação podem ser resumidas pelo fluxograma ilustrado pela Figura 6.3.

**Figura 6.3:** Fluxograma para ilustração da extração de características e classificação de janelas de EEG.



**Fonte:** Elaborada pelo autor.

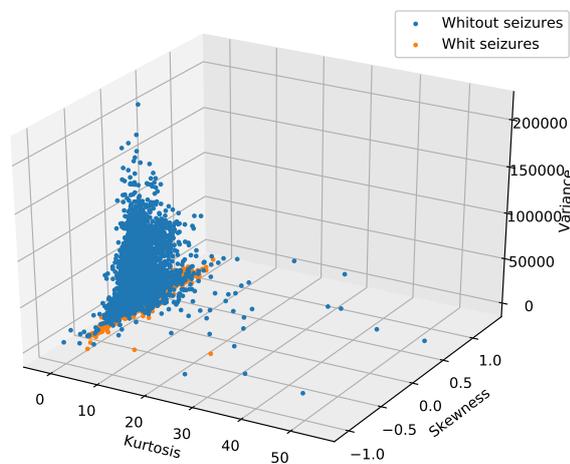
## 7 RESULTADOS

### 7.1 Avaliação da Extração de Características

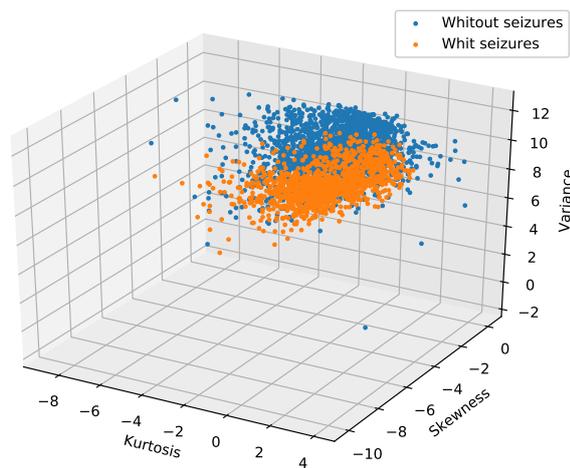
Realizadas todas as etapas descritas na metodologia, inicialmente foi feita uma análise visual sobre os dois grupos de janelas. A Figura 7.1a ilustra um gráfico de dispersão com eixos compostas pelas características extraídas das janelas de EEG (variância, assimetria e curtose).

**Figura 7.1:** Gráfico tridimensional de vetores de características.

(a) Características estatísticas.



(b) Log das características estatísticas.



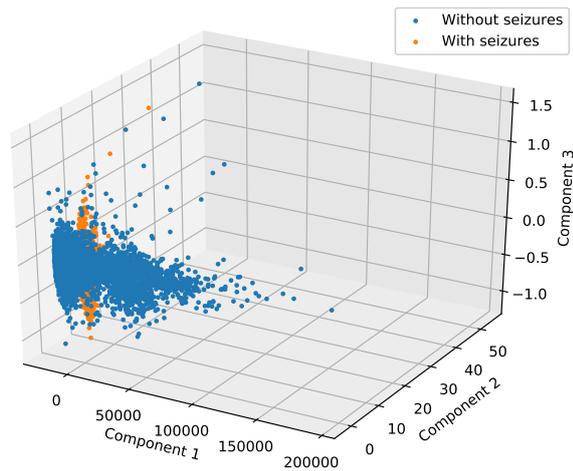
**Fonte:** Elaborada pelo autor.

É possível notar a aglomeração dos grupos. Após a aplicação de uma transformação logarítmica percebe-se certa separação entre os dois grupos de janelas, tal como na Figura 7.1b.

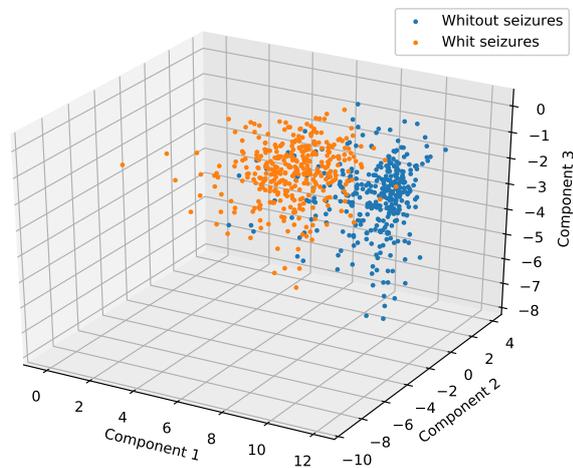
Logo após foi feita a aplicação do PCA com o redução em 3 componentes principais. Isso foi feito afim de alterar a distribuição dos vetores de características. As Figuras 7.2a e 7.2b ilustram gráficos de dispersão em 3 dimensões dos vetores de característica e das componentes obtidas pelo PCA, respectivamente.

**Figura 7.2:** Gráfico tridimensional de componentes principais

(a) Componentes do PCA.



(b) Log das componentes do PCA.



**Fonte:** Elaborada pelo autor.

É possível observar um reposicionamento de janelas com crises na Figura 7.2a. E após a transformação logarítmica nas componentes do PCA a separação entre os grupos torna-se mais expressiva, tal como a Figura 7.2a.

## 7.2 Classificação

Como mencionado anteriormente, foram feitas duas etapas de classificação: classificação dos vetores de característica das janelas de EEG, e classificação das componentes oriundas das do PCA a partir dos vetores de característica.

Na Tabela 7.1 estão contidos os percentuais das métricas de classificadores (acurácia, sensibilidade, especificidade e  $AUC$ ) na etapa de classificação dos vetores de característica.

**Tabela 7.1:** Métricas de classificação dos vetores de característica.

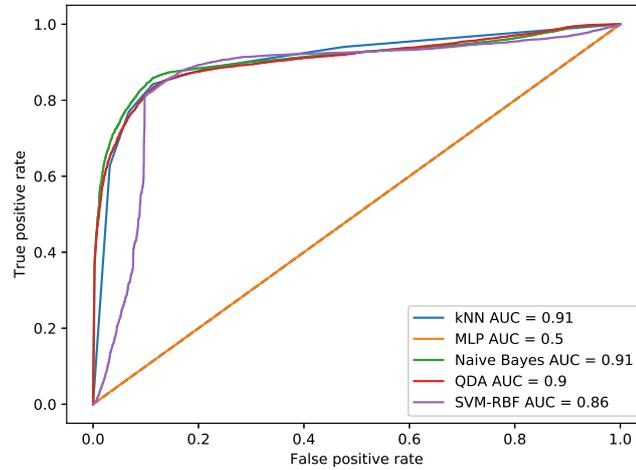
Métricas	SVM-RBF	k-NN	QDA	MLP	Naive Bayes
Acurácia	85.5	<b>86.4</b>	82.6	50.0	83.2
Sensibilidade	84.1	88.7	95.3	30.0	96.5
Especificidade	86.9	84.2	69.88	70.0	70.0
$AUC$	0.86	0.91	0.9	0.5	0.91

Afim de estabelecer uma comparativo de desempenho entre os classificadores a Figura 7.3a ilustra a curva  $ROC$  de cada um deles.

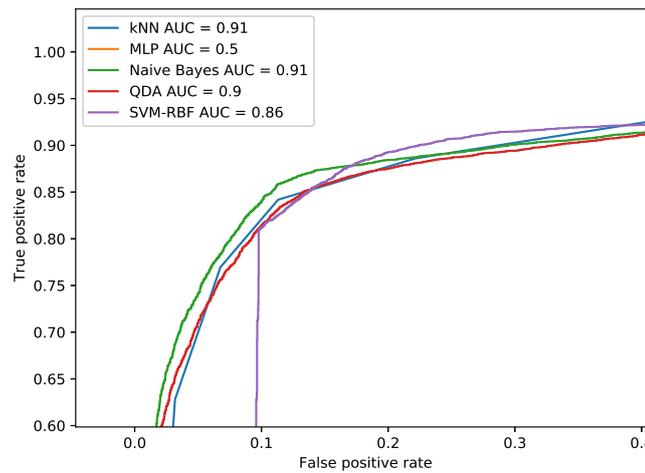
Na Figura 7.3b é possível observar mais detalhadamente as diferenças entre as curvas de cada classificador.

**Figura 7.3:** Gráfico de Curvas ROC de classificadores dos vetores de características.

**(a)** Curva ROC dos classificadores dos vetores de característica



**(b)** Zoom na curva ROC dos classificadores dos vetores de característica



**Fonte:** Elaborada pelo autor.

De forma semelhante foram calculadas as métricas dos classificadores que tiveram como parâmetros as componentes resultantes do PCA. A Tabela 7.2 mostra seus respectivos valores.

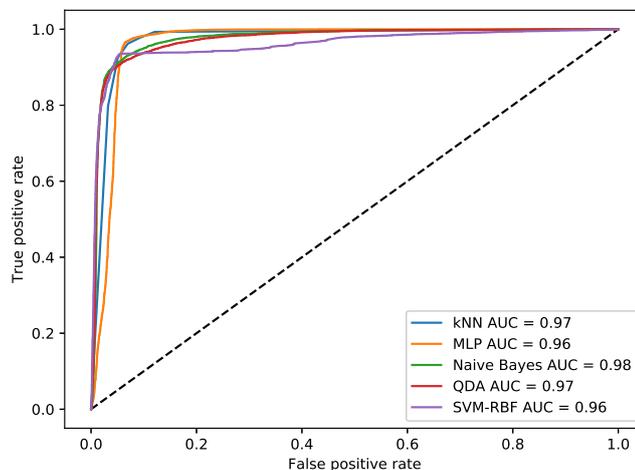
**Tabela 7.2:** Métricas de classificação das componentes do PCA.

Métricas	SVM-RBF	k-NN	QDA	MLP	Naive Bayes
Acurácia	92.4	<b>94.6</b>	91.1	94.3	91.8
Sensibilidade	96.3	93.0	87.2	90.5	87.3
Especificidade	88.5	96.2	69.88	98.1	96.3
<i>AUC</i>	0.96	0.97	0.97	0.96	0.98

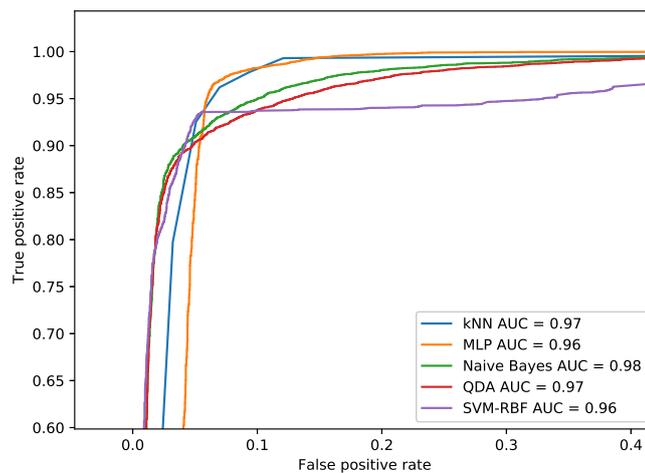
A Figura 7.4a ilustra o comparativo entre os desempenhos de cada classificador das componentes resultantes do PCA, através da curva *ROC*. Tal como a Figura 7.4b detalha as diferenças entre as curvas dos classificadores.

**Figura 7.4:** Gráfico de Curvas ROC de classificadores das componentes do PCA

**(a)** Curva ROC de classificação das componentes do PCA



**(b)** Zoom na curva ROC dos classificadores dos vetores de característica



**Fonte:** Elaborada pelo autor.

## 8 DISCUSSÃO

De forma visual, é possível notar que após a extração de características ocorre um maior agrupamento das janelas de EEG. De tal forma que é possível notar a aglomeração de janelas com crise e uma alta dispersão das janelas sem crise no eixo da variância na Figura 7.1a. A separação entre os dois grupos se evidencia após aplicação da transformação logarítmica, Figura 7.1b.

Foi realizado um *teste T* com os vetores de característica das janelas. No teste foi adotada a hipótese nula de que as amostras são de uma distribuição normal [29]. Essa hipótese foi rejeitada. Foi realizado um novo teste estatístico paramétrico Mann-Whitney U-test [13]. Foi adotada a hipótese nula de que os grupos são de distribuições contínuas e com medianas iguais. Tal hipótese foi rejeitada para as três características extraídas das janelas de EEG. Os  $p$ -valores obtidos foram 0,  $2.6^{-33}$  e  $2.4^{-11}$ , para variância, assimetria e curtose, respectivamente.

Após a aplicação do PCA houve uma mudança na disposição dos dados, tal como na Figura 7.2a. A separação dos grupos fica ainda mais evidente após aplicação da transformação logarítmica, como mostra a Figura 7.2b. Na classificação das componentes do PCA, os percentuais dos classificadores aumentaram, com destaque para os índices da MLP.

Na classificação dos vetores de característica todos os classificadores obtiveram valores de acurácia semelhantes, com exceção da MLP (acurácia de 50%). Apesar dos altos índices obtidos pelos dois classificadores, o seu custo e tempo de convergência os impediu de serem aplicados em futuras aplicações para classificação de janelas em tempo real.

O Naive Bayes e QDA tiveram desempenho semelhante. Na classificação dos vetores de características o Naive Bayes obteve  $AUC = 0.91$  e acurácia de 83.2%; e na classificação das componentes do PCA obteve  $AUC = 0.98$  e acurácia de 91.8%. Já o QDA classificando vetores de características alcançou  $AUC = 0.9$ , acurácia de 82.6%; e classificando componentes do PCA obteve  $AUC = 0.97$  e acurácia de 91.1%. Com resultados tão semelhantes e assumindo que as matrizes de covariância do QDA são diagonais, temos que as entradas de cada classe são condicionalmente independentes. Logo o QDA será um classificador equivalente ao Naive Bayes Gaussiano. Porém ressalta-se que nos testes estatísticos anteriores a hipótese nula, de que os vetores de característica são de distribuições normais, foi rejeitada. Sabendo que o Naive Bayes Gaussiano admite que seus dados de entrada sejam de distribuições normais, podemos considerar que as métricas obtidas foram um acaso específico para os dados utilizados. Por fim podemos desconsiderar o uso dos dois classificadores.

Tanto nas classificações dos vetores de característica quanto nas dos componentes do PCA, o k-NN os melhores percentuais de acurácia, respectivamente 86.4% e 94%. Tendo em vista o uso da distância euclidiana para a classificação e seus índices de acurácia, podemos concluir que ambos os grupos foram bem separados. O que denota eficiência da metodologia de extração de características adotada.

Em comparação a metodologias anteriores podemos notar semelhanças. Assim como Fergus et.al[10] a melhor acurácia obtida foi através do k-NN. Porém, foram utilizadas apenas características no domínio do tempo, as quais podem ser computacionalmente melhor obtidas. Nandy et.al [26] obteve acurácia mais elevada, porém ainda utilizando a frequência. E apesar disso usou o SVM-RBF, geralmente utilizado em classificações de dados não linearmente separáveis, o que denota que a sua extração de características não foi tão satisfatória. Khan et.al [21] utilizou características igualmente fáceis de serem calculadas, porém a sua acurácia foi menor se comparada a do presente trabalho.

A Tabela 8.1 ilustra um comparativo entre os melhores resultados obtidos neste trabalho em relação as soluções citadas, tomando por base os percentuais de acurácia de cada um.

**Tabela 8.1:** Comparativo de performance de metodologias anteriores com o a performance obtida neste trabalho.

Autor	Características	Classificador	Acurácia
Fergus, 2014	Frequência média e picos	k-NN	98%
Khan,2012	COV, RCOV e NCOV	LDA	91%
Nandy, 2019	Tempo e frequência	SVM	97.05%
Este Trabalho	variância, assimetria e curtose	k-NN	94.6%

## 8.1 Trabalhos Futuros

A partir da metodologia de extração de característica proposta, bem como, o uso de algoritmos de classificação, seria possível elaborar um modelo de identificação de início e fim de crises epiléticas, semelhante aos propostos por Shoeb [32, 33]. Tal metodologia poderia ser portada para um dispositivo embarcado. Para assim efetuar classificação de janelas em tempo real. Dessa forma seria possível auxiliar a triagem de pacientes com suspeita de epilepsia.

## 9 CONCLUSÃO

Neste trabalho foi proposto um modelo para classificação de janelas de eletroencefalogramas com e sem epilepsia. A metodologia proposta se destaca por sua generalização, pois são utilizados segmentos de sinais de pacientes com diversas particularidades e em períodos de tempo distintos.

A extração de características se diferencia pois são usadas exclusivamente características obtidas no domínio do tempo, sem necessidade de técnicas mais complexas afim de obter informações no domínio da frequência. Aplicada a etapa de extração foram obtidos vetores estatisticamente distintos.

No que diz respeito a classificação dos vetores de características a aplicação do PCA proporcionou maior separação entre os grupos. De forma que foram obtidos altos índices de acurácia, com destaque para o classificador k-NN com 94.6%.

A partir da metodologia proposta também torna-se possível implementação de um modelo de identificação de início e término de crises epilépticas. Com uma menor complexidade matemática e uma implementação mais amigável em dispositivo embarcado.

A implementação da metodologia mostra-se relevante como alternativa para auxílio a triagem de pacientes com suspeita de epilepsia. Pois são utilizadas características de fácil aquisição e foram necessários segmentos de EEG de 1 segundo, o que reduziria o tempo de em que pacientes são submetidos a coleta de sinais.

## REFERÊNCIAS

- [1] World Health Organization. <https://www.who.int/news-room/fact-sheets/detail/epilepsy>. Acesso em: 18 de dez de 2020.
- [2] D. Angeles et al. Proposal for revised clinical and electroencephalographic classification of epileptic seizures. *Epilepsia*, 22(4):489–501, 1981.
- [3] A. Ascherio. Environmental factors in multiple sclerosis. *Expert review of neurotherapeutics*, 13(sup2):3–9, 2013.
- [4] R. Brookmeyer, S. Gray, and C. Kawas. Projections of alzheimer’s disease in the united states and the public health impact of delaying disease onset. *American journal of public health*, 88(9):1337–1342, 1998.
- [5] J. L. Cummings and G. Cole. Alzheimer disease. *Jama*, 287(18):2335–2338, 2002.
- [6] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989.
- [7] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern classification and scene analysis*, volume 3. Wiley New York, 1973.
- [8] K. Faceli, A. C. Lorena, J. Gama, and A. C. P. d. L. F. Carvalho. *Inteligência Artificial: Uma Abordagem de Aprendizagem de Máquina*. LTC, 2011.
- [9] L. Farber. *Estatística aplicada*, 4a edição, 2010.
- [10] P. Fergus, D. Hignett, and A. Hussain. Automatic epileptic seizure detection using scalp eeg and advanced artificial intelligence techniques. *BioMed Research International*, 2015.
- [11] J. S. Fonseca and G. A. Martins. *Curso de Estatística*. Atlas, 2010.
- [12] J. O. F. Freitas and C. R. R. A. d. Aguiar. Avaliação das funções cognitivas de atenção, memória e percepção em pacientes com esclerose múltipla. *Psicologia: Reflexão e Crítica*, 25(3):457–466, 2012.
- [13] J. D. Gibbons and S. Chakraborti. *Nonparametric statistical inference*. Springer, 2011.
- [14] T. Hastie. *Understanding Machine Learning: From Theory to Algorithms*. Springer, 2009.
- [15] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2009.

- [16] W. Hauser, J. Annegers, and L. T. Kurland. Prevalence of epilepsy in rochester, minnesota: 1940-1980. *Epilepsia*, 32:429–445, 1991.
- [17] S. Haykin. *Redes neurais: princípios e prática*. Bookman Editora, 2007.
- [18] M. Hearst, B. Scholkopf, S. Dumais, E. Osuna, and J. Platt. Trends and controversies- support vector machines. *IEEE Intelligent Systems*, 13(4):18–28, 1998.
- [19] R. Herbrich. *Learning kernel classifiers: theory and algorithms*. MIT press, 2001.
- [20] I. José. Knn (k-nearest neighbors). <https://towardsdatascience.com/knn-k-nearest-neighbors-1-a4707b24bd1d>. Acesso em: 25 de jun de 2019.
- [21] Y. Khan, O. Farooq, and P. Sharma. Automatic detection of seizure onset in pediatric eeg. *International Journal of Embedded Systems and Applications*, 2:81–89, 2012.
- [22] G. Kobelt, A. Thompson, J. Berg, M. Gannedahl, J. Eriksson, M. S. Group, and E. M. S. Platform. New insights into the burden and costs of multiple sclerosis in europe. *Multiple Sclerosis Journal*, 23(8):1123–1136, 2017.
- [23] M. Kubat. Neural networks: a comprehensive foundation by simon haykin, macmillan, 1994, isbn 0-02-352781-7. *The Knowledge Engineering Review*, 13(4):409–412, 1999.
- [24] T. F. Marinho. Recomendações da sbnc para localização de eletrodos e montagens de eeg. *Epilepsia*, 2017.
- [25] G. Milone. *Estatística: geral e aplicada*. Pioneira Thomson Learning, 2004.
- [26] A. Nandy and M. A. Alahe. Feature extraction and classification of eeg signals for seizure detection. In *International Conference on Robotics, Electrical and Signal Processing Techniques*, 2019.
- [27] OpenCV. Introduction to Support Vector Machines. [https://docs.opencv.org/3.4/d1/d73/tutorial\\_introduction\\_to\\_svm.html](https://docs.opencv.org/3.4/d1/d73/tutorial_introduction_to_svm.html). Acesso em: 20 de jun de 2019.
- [28] A. Papoulis and S. U. Pillai. *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education, 2002.
- [29] E. S. Pearson, R. B. D “AGOSTINO, and K. O. Bowman. Tests for departure from normality: Comparison of powers. *Biometrika*, 64(2):231–246, 1977.
- [30] F. Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.

- [31] A. Shoeb. CHB-MIT Scalp EEG Database. <https://physionet.org/pn6/chbmit/>. Acesso em: 21 de jun de 2019.
- [32] A. Shoeb and J. Guttag. Application of machine learning to epileptic seizure detection. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, pages 975–982, 2010.
- [33] A. Shoeb and J. Guttag. Application of machine learning to epileptic seizure detection. 2010.
- [34] A. Shoeb, A. Kharbouch, and J. Soegaard. A machine-learning algorithm for detecting seizure termination in scalp eeg. *Epilepsy and Behavior*, 22:36–43, 2011.
- [35] W. O. Tatum, A. M. Husain, S. R. Banbadis, and P. W. Kaplan. *Handbook of EEG interpretation*. Demos, 2008.
- [36] D. Thurman and E. Beghi. Epileptic seizures and epilepsy: definitions proposed by the international league against epilepsy (ilae) and the international bureau for epilepsy (ibe). *Epilepsia*, 46:470–472, 2005.
- [37] D. Thurman and E. Beghi. Standards for epidemiologic studies and surveillance of epilepsy. *Epilepsia*, 52:2–26, 2011.
- [38] E. Yacubian and S. Kochen. *Crises epilépticas*. Leitura Médica Ltda, 2014.
- [39] Y. Zhou, B. Zhang, G. Li, T. Tong, and X. Wan. Gd-rda: a new regularized discriminant analysis for high-dimensional data. *Journal of Computational Biology*, 24(11):1099–1111, 2017.