

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO CEARÁ
PRÓ-REITORIA DE PESQUISA, PÓS-GRADUAÇÃO E INOVAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Williana Luzia Sousa Leite

**Extração de características com mistura de gaussianas para
classificação de AVC**

Fortaleza – CE

14 de abril de 2023

Williana Luzia Sousa Leite

Extração de características com mistura de gaussianas para classificação de AVC

Dissertação submetida ao Programa de Pós-Graduação em Ciência da Computação do Instituto Federal de Educação, Ciência e Tecnologia do Ceará, como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Linha de Pesquisa: Inteligência Artificial

Orientador: Prof. Dr. Carlos Maurício Jaborandy de Mattos Dourado
Júnior

Coorientador: Prof. Dr. Róger Moura Sarmento

Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)

Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação (PRPI)

Programa de Pós-Graduação em Ciência da Computação (PPGCC)

Fortaleza – CE

14 de abril de 2023

Dados Internacionais de Catalogação na Publicação

Instituto Federal do Ceará - IFCE

Sistema de Bibliotecas - SIBI

Ficha catalográfica elaborada pelo SIBI/IFCE, com os dados fornecidos pelo(a) autor(a)

L533e Leite, Williana.

Extração de características com mistura de gaussianas para classificação de AVC /
Williana Leite. - 2023.

86 f. : il. color.

Dissertação (Mestrado) - Instituto Federal do Ceará, Mestrado em Ciência da
Computação, Campus Fortaleza, 2023.

Orientação: Prof. Dr. Carlos Maurício Jaborandy de Mattos Dourado Júnior.

Coorientação: Prof. Dr. Róger Moura Sarmento.

1. Acidente vascular cerebral. 2. Tomografia computadorizada. 3. Extrator de
características. 4. Densidade radiológica. 5. Classificação de AVC. I. Título.

CDD 004



INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO CEARÁ
Av. 13 de maio, 2081 , - Bairro Benfica - CEP 60040-531 - Fortaleza - CE - www.ifce.edu.br

AVALIAÇÃO

Processo: 23256.001875/2023-27

Interessado: Williana Luzia Sousa Leite

WILLIANA LUZIA SOUSA LEITE

Extração de características com mistura de gaussianas para classificação de AVC

Dissertação submetida à Coordenação do Curso de Pós-graduação em Ciência da Computação do Instituto Federal de Educação, Ciência e Tecnologia do Ceará, como requisito parcial para a obtenção do título de mestre em Ciência da Computação, área de concentração Sistemas de Computação.

Aprovada em **16/03/2023**.

BANCA EXAMINADORA

Prof. Dr. Carlos Maurício Jaborandy de Mattos Dourado Júnior (Orientador)
Instituto Federal de Educação, Ciência e Tecnologia do Ceará - PPGCC/IFCE

Prof. Dr. Roger Moura Sarmento (Coorientador)
Instituto Federal de Educação, Ciência e Tecnologia do Ceará - IFCE

Prof. Dr. Francisco Nivando Bezerra
Instituto Federal de Educação, Ciência e Tecnologia do Ceará - PPGCC/IFCE

Profa Dra. Elizângela de Souza Rebouças
Instituto Federal de Educação, Ciência e Tecnologia do Ceará - IFCE Canindé

Prof. Dr. Antônio Carlos da Silva Barros

Universidade da Integração Internacional da Lusofonia Afro-Brasileira - UNILAB



Documento assinado eletronicamente por **Roger Moura Sarmento, Professor do Ensino Básico, Técnico e Tecnológico**, em 20/03/2023, às 15:48, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Francisco Nivando Bezerra, Professor do Ensino Básico, Técnico e Tecnológico**, em 21/03/2023, às 08:31, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Elizangela de Souza Reboucas, Professora do Ensino Básico, Técnico e Tecnológico**, em 21/03/2023, às 10:25, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Carlos Mauricio Jaborandy de Mattos Dourado Junior, Professor do Ensino Básico, Técnico e Tecnológico**, em 18/04/2023, às 13:44, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Antonio Carlos da Silva Barros, Usuário Externo**, em 18/04/2023, às 20:40, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade do documento pode ser conferida no site https://sei.ifce.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0 informando o código verificador **4673408** e o código CRC **D7E5D202**.

*Este trabalho é dedicado a todas as pessoas que me ajudaram,
direta ou indiretamente, nesses últimos dois anos.
Sem vocês isso não seria possível.*

...

*Em especial aos meus pais Fabiana e William,
que semearam a cultura do estudo na minha criação.
Esse foi o maior presente que vocês poderiam ter me dado.*

Agradecimentos

Cursar um mestrado acadêmico durante a pandemia, e ainda conciliar com o trabalho foi no mínimo mais desafiador do que eu imaginava. Mas eu tive a sorte de contar com muitas pessoas, cada uma a sua maneira, me ajudou a enfrentar os desafios e merecem toda a minha gratidão.

Ao meu orientador Prof. Dr. Carlos Maurício, sempre solícito e tranquilo, embora quase sempre tivéssemos que lidar com prazos apertados. Ao meu co-orientador Prof. Dr. Róger Sarmiento, que sempre trazia ideias e sugestões nas nossas reuniões. Aos dois agradeço pela disponibilidade, ensinamentos, paciência e confiança. Vocês foram simplesmente excepcionais.

Aos professores das disciplinas que cursei, os quais tiveram todo o cuidado em manter a qualidade de ensino mesmo nas situações adversas da pandemia. Agradeço por todo ensinamento.

Aos meus pais e a minha irmã Ana Giselle agradeço pelas palavras de incentivo e encorajamento nos bons e maus momentos. O apoio contínuo de vocês é o meu combustível nos dias difíceis.

Ao meu companheiro Hudson Maia, que esteve presente diariamente e apoiou cada um dos desafios que enfrentei. Foi um ouvinte espetacular, e o melhor parceiro que eu poderia ter ao meu lado. Agradeço por toda ajuda, paciência e resiliência.

As minhas amigas de vida Dr. Mariana Dehon e Prof. Dr. Tatiane Fernandes, por terem me ajudado inúmeras vezes ao longo desses anos. Vocês foram decisivas no meu desenvolvimento pessoal e profissional.

E por fim, e mais importante, agradeço a Deus por escutar as minhas orações e sempre me guiar no caminho correto, mesmo quando não faz parte dos meus planos. Por me conceder saúde e força de vontade pra conseguir superar as adversidades da vida, pois "tudo posso naquele que me fortalece"(Filipenses 4:13).

Muito obrigada!

*“Bem-aventurado o homem que acha sabedoria, e o homem que adquire conhecimento.
Pois a sabedoria é mais proveitosa do que a prata e rende mais do que o ouro.”
(Provérbios 3:13,14)*

Resumo

O Acidente Vascular Cerebral (AVC) é uma emergência médica com elevados índices de morbidade e mortalidade. Todos os anos 15 milhões de pessoas sofrem um AVC, das quais 5 milhões morrem e 5 milhões ficam permanentemente incapacitadas. No Brasil, o AVC voltou a ser a principal causa de morte, e no primeiro semestre de 2022, vitimou cerca de 50 mil pessoas. Para o diagnóstico de AVC, o exame da Tomografia Computadorizada (TC) é o mais indicado, pois é possível determinar o tipo, a extensão e a severidade do AVC. O tratamento depende de qual tipo de AVC o paciente sofreu, para isso é imprescindível que o diagnóstico seja feito o mais rápido possível após os primeiros sintomas. Cada minuto a partir dos primeiros sintomas é crítico, em torno de 1.9 milhões de células cerebrais são afetadas por minuto durante um AVC. Devido ao senso de urgência e acuracidade necessária no diagnóstico, é relevante um sistema computacional que auxilie os médicos na obtenção de informações imperceptíveis ao olho humano, tornando o diagnóstico mais preciso. Sendo assim, este trabalho propõe um novo método para extração de características que utiliza mistura de gaussianas e padrões de densidades radiológicas cerebrais, denominado de Análise de Mistura Gaussianas da Densidade do Tecido Cerebral ou *Mixture Gaussian Analysis of Brain Tissue Density* (MGABTD). O método busca extrair informações relevantes de imagens de TC cerebrais para identificação e classificação do tipo de AVC. Para isso foram desenvolvidas três variações diferentes do uso de mistura de gaussianas para esse propósito, chamadas de: $MGABTD_{pixel}$, $MGABTD_{percent}$ e $MGABTD_{neighborhood}$. O desempenho dos três extratores foram comparados entre si, visando determinar qual variação é mais indicada em cada cenário, e foram comparadas empiricamente com resultados obtidos nos trabalhos relacionados. Cada método de extração foi avaliado utilizando seis classificadores: MLP, SVM, KNN, OPF, RF e *Extremely Randomized Trees* (Extra Tree), considerando três classes de AVC: normal, isquêmico ou hemorrágico. O método $MGABTD_{pixel}$ obteve maior performance, com 99.9% de acurácia nos classificadores: *Extra Tree*, KNN, RF e SVM. Baseado na relevância dos resultados obtidos, pode-se afirmar que o MGABTD é uma técnica útil para extração de características de imagens de TC do cérebro, com potencial de integrar sistemas que auxiliem o diagnóstico médico para detecção e classificações de AVC.

Palavras-chave: Acidente vascular cerebral, Tomografia computadorizada, Extrator de características, Densidade radiológica, Classificação de AVC.

Abstract

A stroke is a medical emergency with high rates of morbidity and mortality. Every year 15 million people suffer a stroke, of which 5 million die and 5 million are permanently disabled. In Brazil, stroke was once again the leading cause of death, and in the first half of 2022, it killed around 50,000 people. Computed Tomography (CT) is the most indicated for the diagnosis of stroke, as it is possible to determine the stroke's type, extent, and severity. Treatment depends on which type of stroke the patient has suffered, so it is essential that the diagnosis be made as soon as possible after the first symptoms. Every minute from the first symptoms is critical, around 1.9 million brain cells are affected per minute during a stroke. Due to the sense of urgency and accuracy required in the diagnosis, a computational system is relevant to help physicians obtain information that is invisible to the human eye, making the diagnosis more accurate. Therefore, this work proposes a new method for extracting features that use a mixture of Gaussians and patterns of brain radiological densities, called Mixture Gaussian Analysis of Brain Tissue Density (MGABTD). The method seeks to extract relevant information from brain CT images for identifying and classifying the type of stroke. For this, three different variations of the use of Gaussian mixtures were developed for this purpose, called: $MGABTD_{pixel}$, $MGABTD_{percent}$ and $MGABTD_{neighborhood}$. The performance of the three extractors was compared with each other, aiming to determine which variation is most indicated in each scenario, and were empirically compared with results obtained in related works. Each extraction method was evaluated using six classifiers: MLP, SVM, KNN, OPF, RF, and *Extremely Randomized Trees* (Extra Tree), considering three-stroke classes: normal, ischemic, or hemorrhagic. The $MGABTD_{pixel}$ method achieved the best performance, with 99.9% accuracy in the classifiers: *Extra Tree*, KNN, RF, and SVM. Based on the relevance of the results obtained, it can be stated that the MGABTD is a useful technique for extracting features from CT images of the brain, with the potential to integrate systems that help medical diagnosis for stroke detection and classification.

Keywords: Stroke, CT scan, Features Extractor, Radiological density, Stroke classification.

Lista de ilustrações

Figura 1 – Diferentes tipos de AVC. Fonte: SBAVC.	30
Figura 2 – Ilustração dos subtipos de AVCi. Fonte: Adaptado de Clinic (2021). . .	31
Figura 3 – Ilustração dos subtipos de AVCh. Fonte: Adaptado de Chong (2020). .	33
Figura 4 – Ilustração de um tomógrafo. Fonte: Freitas (2013).	34
Figura 5 – Exemplo de imagens obtidos nos planos axial, coronal e sagital. Fonte: Adaptado de Bankman (2008).	35
Figura 6 – Ilustração da escala de <i>Hounsfield</i> . Fonte: Freitas (2013).	35
Figura 7 – Exemplo considerando a base de dados <i>Old Faithful</i> . Fonte: Bishop e Nasrabadi (2006).	42
Figura 8 – Exemplo de mistura com 3 gaussianas. Fonte: Bishop e Nasrabadi (2006).	43
Figura 9 – Ilustração do processo de classificação do KNN. Fonte: Elaborado pelo autor.	48
Figura 10 – Ilustração da rede PS. Fonte: Elaborado pelo autor.	49
Figura 11 – Ilustração da rede MLP para classificação. Fonte: (BISHOP; NASRA- BADI, 2006).	50
Figura 12 – Ilustração dos 2 tipos de erros que devem ser calculados na MLP. Fonte: (BISHOP; NASRABADI, 2006).	51
Figura 13 – Problema linearmente separável. Fonte: Elaborado pelo autor.	52
Figura 14 – Problema não linearmente separável, aplicando <i>kernel</i> . Fonte: Elaborado pelo autor.	53
Figura 15 – Identificação dos protótipos. Fonte: Elaborado pelo autor.	54
Figura 16 – Floresta de caminhos ótimos. Fonte: Elaborado pelo autor.	55
Figura 17 – Processo para determinar a classe de uma amostra desconhecida. Fonte: Elaborado pelo autor.	56
Figura 18 – Separação do espaço de decisão. Fonte: Bishop e Nasrabadi (2006). . .	56
Figura 19 – Estrutura de uma <i>Random Forest</i> . Fonte: Adaptado de Williams et al. (2020).	58
Figura 20 – Para cada imagem da base de dados, a região interna do crânio é extraída, redimensionada, dividida em dois hemisférios e para cada hemisfério o extrator MGABTD é aplicado. Por fim as características dos dois hemisférios são unidas gerando um vetor de características que é enviado para um classificador de padrões, determinando a classe da imagem de entrada. Fonte: Elaborado pelo autor.	61

Figura 21 – Amostra de TC do cérebro de cada uma das classes. Na Figura 21b o AVCh é representado pela mancha branca, enquanto na Figura 21c o AVCi é representado pela mancha mais escura próxima ao crânio. Fonte: Base de dados.	62
Figura 22 – Etapas da segmentação da região de interesse. Fonte: Sarmento et al. (2020).	63
Figura 23 – Metodologia após a imagem de entrada passar pela segmentação ROI e redimensionamento, utilizando o extrator $MGABTD_{pixel}$. Fonte: Elaborado pelo autor.	65
Figura 24 – Metodologia após a imagem de entrada passar pela segmentação ROI e redimensionamento, utilizando o extrator $MGABTD_{percent}$. Fonte: Elaborado pelo autor.	66
Figura 25 – Redefinição da classe de cada pixel.	66
Figura 26 – Metodologia após a imagem de entrada passar pela segmentação ROI e redimensionamento, utilizando o extrator $MGABTD_{neighborhood}$. Fonte: Elaborado pelo autor.	67
Figura 27 – Resultado extrator $MGABTD_{pixel}$. Fonte: Elaborado pelo autor.	74
Figura 28 – Resultado extrator $MGABTD_{percent}$. Fonte: Elaborado pelo autor.	75
Figura 29 – Resultado extrator $MGABTD_{neighborhood}$. Fonte: Elaborado pelo autor.	75
Figura 30 – Tempo médio de Treinamento. Fonte: Elaborado pelo autor.	77
Figura 31 – Tempo médio de projeção/teste. Fonte: Elaborado pelo autor.	77
Figura 32 – Importância dos pixels considerando o $MGABTD_{pixel}$ com Extra Tree. Fonte: Elaborado pelo autor.	79
Figura 33 – Importância dos atributos considerando impureza Gini com $MGABTD_{percent}$ com Extra Tree. Fonte: Elaborado pelo autor.	80
Figura 34 – Importância dos atributos considerando impureza Gini com $MGABTD_{neighborhood}$ com Extra Tree. Fonte: Elaborado pelo autor.	80
Figura 35 – Importância dos atributos considerando SHAP com $MGABTD_{percent}$ com Extra Tree. Fonte: Elaborado pelo autor.	81
Figura 36 – Importância dos atributos considerando SHAP com $MGABTD_{neighborhood}$ com Extra Tree. Fonte: Elaborado pelo autor.	81

Lista de tabelas

Tabela 1 – Configuração padrão dos classificadores.	72
Tabela 2 – Melhor configuração de parâmetros de cada extrator.	73
Tabela 3 – Resultados obtidos com as configurações ótimas de cada extrator-classificador.	76
Tabela 4 – Classificador recomendado considerando acuracidade e tempo de teste para cada variação de extrator proposto.	78
Tabela 5 – Comparação dos resultados obtidos com a literatura.	82

Lista de abreviaturas e siglas

1D	Unidimensional
2D	Bidimensional
3D	Tridimensional
ABTD	<i>Analysis of Brain Tissue Densities</i>
ACC	Acurácia
ANN	<i>Artificial Neural Network</i>
AVC	Acidente Vascular Cerebral
AVCi	Acidente Vascular Cerebral Isquêmico
AVCh	Acidente Vascular Cerebral Hemorrágico
CAD	<i>Computer-Aided Diagnosis</i>
CNN	<i>Convolutional Neural Network</i>
DICOM	<i>Digital Imaging and COmmunications in Medicine</i>
EM	<i>Expetetaion Maximization</i>
Extra Tree	<i>Extremely Randomized Trees</i>
FOF	<i>First Order Features</i>
GLCM	<i>Gray level co-occurrence matrix features</i>
GLRLM	<i>Gray level run length matrix features</i>
GMM	<i>Gaussian Mixture Model</i>
HO	<i>Hold Out</i>
HU	Unidade Hounsfield
HUM	<i>HU Moments</i>
IA	Inteligência Artificial
IRM	Imagem por Ressonância Magnética

KNN	<i>K-Nearest Neighbors</i>
LNP	<i>Local Neighbourhood Pattern</i>
LOO	<i>Leave One Out</i>
ML	<i>Machine Learning</i>
MLP	<i>Multilayer Perceptron</i>
MGABTD	<i>Mixture Gaussian Analysis of Brain Tissue Densit</i>
NPV	<i>Negative Predictive Value</i>
OPF	<i>Optimum Path Forest</i>
PABTD	<i>Parzen Analysis of Brain Tissue Densities</i>
PDI	Processamento Digital de Imagens
RBF	<i>Radial Basis Function</i>
RBM	<i>Restricted Boltzmann Machines</i>
RF	<i>Random Forest</i>
RNA	Rede Neural Artificial
ROI	<i>Region of Interest</i>
SVM	<i>Support Vector Machine</i>
SBAVC	Sociedade Brasileira de AVC
SHAP	<i>SHapley Additive exPlanations</i>
TC	Tomografia Computadorizada
TIA	<i>Transient Ischemic Attack</i>
TreeBagger	<i>Ensemble Bagged Trees</i>
WHO	<i>World Health Organization</i>
WSO	<i>World Stroke Organization</i>

Sumário

1	INTRODUÇÃO	23
1.1	Motivação	24
1.2	Descrição do Problema	26
1.3	Objetivos	27
1.3.1	Objetivos Gerais	27
1.3.2	Objetivos Específicos	27
1.4	Desafios	28
1.5	Produção Científica	28
1.6	Organização da Dissertação	28
2	ACIDENTE VASCULAR CEREBRAL	29
2.1	Tipos de AVC	30
2.1.1	AVC Isquêmico	30
2.1.2	AVC Hemorrágico	32
2.2	Diagnóstico e Tratamento	32
2.3	Exames de Neuroimagens	34
2.4	Estado da arte	36
3	RECONHECIMENTO DE PADRÕES	41
3.1	Mistura de Gaussianas	42
3.1.1	Algoritmo <i>Expectation Maximization</i>	43
3.1.2	Aplicações GMM	45
3.1.2.1	Classificador Bayesiano	45
3.2	K Vizinhos Mais Próximos	46
3.3	Rede Perceptron de Multicamadas	48
3.4	Máquina de Vetores de Suporte	52
3.5	Floresta de Caminhos Ótimos	53
3.6	Métodos baseados em Árvores de Decisão	56
3.6.1	Floresta Aleatória	58
3.6.2	Árvores Extremamente Aleatórias	58
4	METODOLOGIA E EXPERIMENTOS	61
4.1	Base de dados	62
4.2	Pré-processamento	63
4.3	Análise de Mistura Gaussianas da Densidade do Tecido Cerebral	63
4.3.1	Analysis of Brain Tissue Densities (ABTD)	64

4.3.2	$MGABTD_{pixel}$	64
4.3.3	$MGABTD_{percent}$	65
4.3.4	$MGABTD_{neighborhood}$	65
4.4	Possíveis configurações MGABTD	67
4.5	Significância Estatística	67
4.5.1	<i>5x2 cv paired t-test</i>	68
4.5.1.1	Friedman	69
5	RESULTADOS E DISCUSSÕES	71
5.1	Configuração ótima do MGABTD	71
5.2	Configuração ótima dos classificadores	74
5.2.1	Análise de Tempo de processamento	76
5.2.2	Seleção do melhor classificador para cada extrator	77
5.2.3	Importância dos Atributos	78
5.3	Comparação dos resultados com trabalhos relacionados	80
5.3.1	Comparação com método base ABTD	82
5.4	Análise geral dos resultados	83
6	CONCLUSÕES, CONTRIBUIÇÕES E TRABALHOS FUTUROS . .	85
6.1	Trabalhos Futuros	86
	REFERÊNCIAS	87

1 Introdução

A pandemia do COVID-19, iniciada no 1º semestre de 2020, foi um catalisador para a inovação tecnológica em diversos setores no mundo todo, principalmente no setor da saúde. A emergência do serviço público mundial e as frequentes interrupções das atividades econômicas colocaram a inovação tecnológica no centro do debate (LIU; SHI; YANG, 2022). Fazendo com que processos antes padronizadas, precisassem se adaptar a nova realidade, como foi o caso da Telemedicina que ganhou grande destaque (BAHL et al., 2020). No Brasil foram desenvolvidos aplicativos móveis, sites, atendimentos online, visando ajudar os profissionais da saúde e os pacientes (CELUPPI et al., 2021).

Sistemas computacionais que fornecem suporte ao diagnóstico médico também ganharam ainda mais notoriedade durante a pandemia. Diversas pesquisas foram desenvolvidas a fim de auxiliar os médicos no diagnóstico da COVID-19 (BRINDHA; KAVITHA; BHUSHAN, 2022). No entanto sistemas com o propósito de auxiliar o diagnóstico médico são objeto de estudo dos pesquisadores há muito tempo antes do COVID-19 (MARQUES, 2001). Para diversas patologias, pesquisadores tem se aprofundado nas áreas de Inteligência Artificial (IA) e Processamento Digital de Imagens (PDI), buscando desenvolver novos sistemas ou superar limitações de sistemas desenvolvidos anteriormente (DOI, 2007).

Dentre as diversas modalidades de exames que suportam o diagnóstico, a imagiologia desempenha um papel central na medicina atual. Segundo a World Health Organization (WHO), o exame de imagem é essencial para confirmar, avaliar e documentar o andamento de muitas doenças e respostas a tratamentos. Dentre os exames dessa modalidade, destacam-se a Tomografia Computadorizada (TC) e a Imagem por Ressonância Magnética (IRM). Ambas utilizam sistemas computacionais para gerar imagens internas das regiões de interesse do corpo (BUZUG, 2011).

A TC possui mais rapidez no exame, levando em torno de 15-20 minutos, com ambiente menos confinante do que a IRM, e imagens de alta qualidade. Como desvantagem o paciente é exposto à radiação, e se for uma tomografia com contraste e o paciente tiver problema renal, o material de contraste pode piorar a função renal (KWONG; YUCEL, 2003). A IRM possui um tempo de duração do exame muito maior, em torno de 45-90 minutos, possui um espaço mais confinante, alto ruído, e o paciente não pode ter objetos metálicos (SILVA; OLIVEIRA, 2017). Dessa forma, considerando velocidade, custo e disponibilidade a TC é a técnica mais indicada para diagnóstico de AVC (LUO et al., 2009).

A partir das imagens obtidas nos exames de TC e/ou IRM, sistemas computacionais podem ser desenvolvidos, visando segmentar, reconhecer e identificar detalhes de uma

região de interesse. Para isso geralmente os pesquisadores utilizam técnicas de Visão Computacional (VC), PDI e Reconhecimento de Padrões (RP) para extrair informações dos exames (BANKMAN, 2008). Com essas informações obtidas do sistema computacional o médico radiologista pode validar o diagnóstico final. Esse tipo de sistema computacional que auxilia o diagnóstico é chamado de Diagnóstico Auxiliado por Computador ou *Computer Aided Diagnosis* (CAD) (MARQUES, 2001).

O processo tradicional de análise dos exames de imagens, correspondem ao médico visualmente examinar cuidadosamente as imagens do exame em busca de alterações patológicas. No entanto, em alguns casos essas alterações são tão sutis que acabam se tornando imperceptíveis ao olho humano (WU et al., 2019). As imagens TC possuem uma quantidade de informação muito maior que o olho humano é capaz de compreender, sendo aproximadamente 4000 tons de cinza enquanto a capacidade humana é de aproximadamente 64 tons de cinza (BUZUG, 2011). Dessa forma, analisar computacionalmente as imagens pode fornecer características/padrões que o olho humano não consegue enxergar.

Diversas pesquisas estão sendo desenvolvidas visando criar/melhorar técnicas de extração e análise de informações dos exames, tendo em conta os benefícios apresentados dos sistemas CAD para o diagnóstico (YANASE; TRIANTAPHYLLOU, 2019). Sendo assim, essa pesquisa busca desenvolver um método para extração de características em conjunto com a classificação de padrões em imagens de TC do cérebro para identificação de Acidente Vascular Cerebral (AVC), onde um novo extrator de atributos é proposto.

1.1 Motivação

O Acidente Vascular Cerebral (AVC) é uma doença cardiovascular que ocorre pela alteração do abastecimento de sangue no cérebro, seja pela falta ou excesso de sangue, causando danos as células cerebrais (ZHANG et al., 2015). O AVC também é uma das principais causas de morbidade, incapacidade adquirida, hospitalizações e mortalidade no mundo (MACKAY; MENSAH; GREENLUND, 2004).

De acordo com a WHO, todos anos em todo o mundo aproximadamente 15 milhões de pessoas sofrem um AVC, desses 5 milhões morrem e outros 5 milhões ficam permanentemente incapacitados. Apesar da alta taxa de mortalidade, se diagnosticado rapidamente o tratamento pode evitar sequelas graves e até mesmo a morte (HANKEY, 2014).

Antes da pandemia do COVID-19 a maioria dos países Europeus já apresentavam estudos que indicavam um aumento na carga de AVC nos próximos 30 anos (WAFA et al., 2020). Isso, combinado com o efeito adverso da pandemia, provavelmente pode fazer com que o AVC tenha um impacto ainda maior na saúde pública no século XXI (KRALJ; ČUKELJ, 2022). No Brasil, segundo o Portal da Transparência dos Cartórios de Registro Civil, o AVC voltou a ser a principal causa de morte. No primeiro semestre de 2022 vitimou

cerca de 50 mil pessoas, tornando-se mais letal que o COVID-19 (PIAUÍ, 2022).

A expressão *Time is Brain* creditada ao neurologista Camilo Gomez, M.D. por volta de 1990, é uma expressão conhecida que representa o quanto é crucial obter ajuda imediatamente no caso de apresentar sintomas de AVC. Os principais sintomas são: rosto caído, dificuldade na fala e fraqueza/dormência em um dos braços. Cada minuto importa após os primeiros sintomas, o AVC ataca 1.9 milhões de células cerebrais por minuto (SAVER, 2006). Quando comparado com a taxa regular de perda de células, o cérebro envelhece 3,6 anos por hora sem tratamento imediato.

Os pacientes que chegam à emergência em até 3 horas após os primeiros sintomas de um AVC geralmente têm melhores resultados em termos de incapacidade a longo prazo do que aqueles que recebem cuidados atrasados (CDC; PREVENTION, 2008). Nos Estados Unidos, os custos relacionados ao AVC chegaram a quase 53 bilhões de dólares entre 2017 e 2018 (TSAO et al., 2022).

De todas as pessoas que sofrem um AVC em média 30% se recuperam, 40% ficam com sequelas que limitam atividades diárias e 30% morrem ou ficam incapacitados de realizar atividades simples. O risco de ter um primeiro derrame é quase duas vezes maior para os negros do que para os brancos, e os negros têm a maior taxa de morte por AVC (TSAO et al., 2022). A hipertensão arterial, fatores alimentares e o aumento da idade, são os principais fatores de risco do AVC (OLIVEIRA; ANDRADE, 2001).

O AVC é a principal causa de incapacidade funcional no mundo ocidental. A mobilidade dos sobreviventes com 65 anos ou mais cai em mais da metade (TSAO et al., 2022). Cerca de 40 a 50% das pessoas morrem nos 6 meses após o AVC. A maioria dos sobreviventes desenvolve deficiências neurológicas e incapacidades residuais significativas (PERLINI; FARO, 2005). Os principais fatores que impactam a sobrevivência do AVC é a gravidade do AVC, histórico de AVC passados, idade, insuficiência cardíaca e alta concentração de glicose no sangue e contagem de glóbulos brancos (HANKEY, 2003).

O tratamento depende de qual tipo de AVC o paciente possui, logo o diagnóstico deve ser o mais rápido possível devido a necessidade rápida de tratamento. A velocidade e o custo se tornam um dos principais fatores para o uso disseminado da TC no diagnóstico (WARDLAW; FARRALL, 2004). Com a TC é possível avaliar o fluxo e o volume sanguíneo, e com isso diagnosticar o AVC. Com a IRM o diagnóstico também é possível, no entanto o custo para implantação é maior e com isso a disponibilidade é mais provável em grandes centros hospitalares (KAK; SLANEY, 2001).

Considerando a possível indisponibilidade momentânea do radiologista, um sistema CAD que auxilie no diagnóstico do AVC tem potencial de ser produto muito útil aos profissionais da saúde, principalmente aos profissionais com pouca experiência. Sendo assim, o propósito desse trabalho é auxiliar o diagnóstico de AVC, por meio de imagens

TC do cérebro.

1.2 Descrição do Problema

O diagnóstico de AVC feito por meio de imagens de TC, além de conterem detalhes que podem ser imperceptíveis ao olho humano, são geradas muitas imagens, tornando o processo cansativo e suscetível ao erro por falha humana, principalmente quando aliado a alta carga de trabalho (BRADY, 2017). Pesquisadores do Laboratório de Atenção Visual de Harvard, fizeram um experimento para avaliar a cegueira desatenta em 24 radiologistas experientes. O experimento consistia em cada radiologista receber 5 TCs para interpretar, cada uma com 100 a 500 imagens, contendo uma média de 10 nódulos pulmonares. No último caso a imagem de um gorila escuro 48 vezes maior que os nódulos foi esmaecido e adicionada perto de um nódulo. Ao final, 20 dos 24 radiologistas não relataram ter visto o gorila, mesmo que 12 deles o olharam diretamente (DREW; VÕ; WOLFE, 2013).

Em geral a análise de imagens TC não é simples, e as dificuldades de interpretação do exame podem resultar em erros de diagnóstico. Essas divergências de interpretação ocorrem em maior grau entre médicos emergencistas e radiologistas nos exames de imagem (SIEGEL et al., 1998). Logo, o diagnóstico de AVC também é acometido por esse problema quando feito por exames de imagem. Além disso, a presença de sintomas "não tradicionais", corrobora o erro de diagnóstico do AVC (ARCH et al., 2016).

Os problemas mencionados até então, impulsionam a possibilidade de erros de diagnósticos de AVC. Os autores Newman-Toker et al. (2014) analisaram os erros de diagnósticos de AVC obtidos no departamento de emergência em nove estados dos EUA, por meio de investigação do histórico hospitalar. Os pacientes que tiveram o diagnóstico final de AVC e dias antes apresentaram sintomas e foram diagnosticados com outra patologia, foram considerados como erros. Nesse estudo houveram 13,9% de erros de diagnóstico.

Uma outra emergência médica que é constantemente associada a erros de diagnóstico é a Hemorragia Subaracnoidea (HSA). A HSA quando ocorre espontaneamente é considerada um AVC, e possui alto risco de morbidade e mortalidade (PATEL; PARIKH; OKORIE, 2021). A taxa de diagnósticos perdidos ou atrasados varia de 12% a 51% (OH et al., 2018; VERMEULEN; SCHULL, 2007).

Além do diagnóstico por meio da análise visual da tomografia, existe a possibilidade de analisar os padrões radiológicos obtidos na TC e compará-los a literatura para identificação e classificação de doenças cerebrais. Por exemplo, na literatura, a região da massa branca é delimitada pelos valores de 23HU a 34HU, onde a unidade HU representa quantitativamente a radiodensidade (YOUSEM, 2011). Dessa forma, bastaria buscar na tomografia valores HU que representasse a região de interesse para fazer o diagnóstico

(MÜHL-BENNINGHAUS et al., 2021), como por exemplo fazer o diagnóstico do AVC buscando o intervalo da região na TC. No entanto, essa abordagem de reconhecimento dos padrões radiológicos apresenta alguns problemas:

- A definição dessas regiões do cérebro, por meio dos intervalos não são padronizados na literatura, deixando o diagnóstico subjetivo, e causando confusão entre os especialistas;
- Algumas regiões possuem definições de intervalos muito próximas ou até mesmo sobrepostas. Como por exemplo as regiões massa cinzenta (32HU a 41 HU) e massa branca (23HU a 34HU), que de acordo o autor Yousem (2011) acabam ficando sobrepostas. E com isso os valores de 32HU a 34HU pertencem a mais de uma região.

Esses problemas acabam se tornando uma área intrigante e propícia para aplicação de análises computadorizadas e estudos mais aprofundados. Sendo assim, com o propósito de amenizar os problemas citados, torna-se necessário desenvolver um sistema capaz de auxiliar os médicos na avaliação de exames de TC, possibilitando uma identificação correta e precisa da região onde o AVC ocorreu.

1.3 Objetivos

1.3.1 Objetivos Gerais

Neste trabalho é proposto variações do uso da Mistura de Gaussianas como extrator de características para imagens de TC do cérebro. Os extratores de características resultantes de cada variação são chamados essencialmente de Análise de Mistura Gaussianas da Densidade do Tecido Cerebral ou *Mixture Gaussian Analysis of Brain Tissue Density* (MGABTD).

1.3.2 Objetivos Específicos

Os objetivos específicos dessa pesquisa são:

- Desenvolver variações de extratores de características que utilizam mistura de gaussianas;
- Aplicar as variações de extratores em bases de dados formadas por diferentes exames de TC do cérebro;
- Realizar um estudo comparativo entre as variações de extratores determinando qual variação é mais indicada em cada cenário;
- Classificar imagens TC do cérebro em: normal, AVC Hemorrágico ou AVC Isquêmico.

1.4 Desafios

Baseado nos problemas apresentados na Seção 1.2, os principais desafios encontrados são:

- Desenvolver um método que seja rápido o suficiente sem comprometer a precisão da classificação, devido a necessidade do diagnóstico rápido e preciso do AVC;
- Identificar a forma mais eficiente de utilizar a mistura de gaussiana na extração de características, afim de obter o melhor conjunto de características possível;
- Desenvolver um extrator de características automático baseados nos padrões radiológicos cerebrais que facilite o processo de reconhecimento e classificação de um AVC em imagens de TC do cérebro.

1.5 Produção Científica

Essa pesquisa resultou no seguinte artigo que foi aceito, publicado e apresentado presencialmente em conferência:

- **LEITE, W. L.; SARMENTO, R. M.; JUNIOR, C. M. D. Feature extraction with mixture gaussian for stroke classification.** In: IEEE. 2022 35th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). Natal, Brazil, 2022. v. 1, p. 91–96. *Qualis A3* em Ciência da Computação.

1.6 Organização da Dissertação

Esta dissertação possui o capítulo atual introdutório (Capítulo 1) e outros 5 capítulos.

Capítulo 2: visa apresentar os principais aspectos relativos ao acidente vascular cerebral (AVC). Além das definições, apresenta os tipos, diagnóstico e o estado da arte em extração de características em imagens de TC do cérebro.

Capítulo 3: busca apresentar os fundamentos teóricos da principal técnica utilizada na extração de características e das técnicas utilizadas para classificação de padrões.

Capítulo 4: explica a metodologia, e como funciona as variações de extração de características.

Capítulo 5: apresenta a comparação de performance entre as variações de extratores propostos e entre os trabalhos relacionados, assim como os resultados obtidos por meio de testes.

O Capítulo 6: apresenta a conclusão e expectativas para os trabalhos futuros.

2 Acidente Vascular Cerebral

O funcionamento normal do cérebro depende da perfusão sanguínea que fornece oxigênio e nutrientes (GROTTA et al., 2021). O Acidente Vascular Cerebral (AVC), é uma doença que ocorre devido ao entupimento ou rompimento dos vasos sanguíneos do sistema nervoso central. Com isso as células cerebrais param de receber oxigênio e podem ser danificadas ou morrerem. Segundo a *World Stroke Organization* (WSO), o AVC é considerado uma epidemia global, com alto índice de morbidade e mortalidade.

Ainda de acordo com a WSO, o AVC pode ocorrer em qualquer idade, alguns sintomas são:

- distúrbio da fala;
- fraqueza no braço;
- queda facial;
- início de convulsões;
- perda de coordenação;
- dor de cabeça severa;
- distúrbios visuais;
- tontura;
- perda de consciência;
- febre alta.

Sendo os 3 primeiros da lista, os sintomas mais comuns. O AVC pode durar alguns minutos, horas, ou até mesmo mais de 24h (WARLOW, 1998). Dependendo da região do cérebro atingida e da rapidez do tratamento, as consequências do AVC podem ser de curto ou longo prazo. Segundo a Sociedade Brasileira de AVC (SBAVC), quanto mais rápido iniciar o tratamento, maiores são as chances de diminuir/acabar com os sintomas do AVC. O AVC tem prevenção e consiste em realizar atividades físicas, manter uma dieta saudável, limitar o consumo de álcool e evitar o uso de cigarro. De acordo com O'Donnell et al. (2016) mais de 80% dos AVC podem ser prevenidos.

Pessoas que possuem pressão alta, diabetes, doença cardíaca ou colesterol alto, possuem maior risco de ter um AVC. Apesar do AVC ocorrer em qualquer idade, a incidência aumenta significativamente com a velhice. Pessoas negras e asiáticos, possuem maior incidência de AVC, especificamente no Brasil, os homens negros possuem 40% a mais de risco de óbito no AVC comparado a homens brancos e mulheres negras possuem 76% a mais de risco de óbito no AVC comparado a mulheres brancas (LOTUFO; BENSENOR, 2013). Segundo a SBAVC algumas causas menos populares de AVC são: malformações arteriovenosas, aneurismas, distúrbios da coagulação e uso de anticoagulantes ou trombolí-

ticos, transformações hemorrágicas de infartos e sangramento de tumores, ou por abuso de drogas (OLIVEIRA; ANDRADE, 2001).

2.1 Tipos de AVC

Segundo Gonçalves et al. (2004) e Manzi et al. (2003) existem dois tipos de AVC: Isquêmico (AVCi) e Hemorrágico (AVCh). Ainda existe o Ataque Isquêmico Transitório ou *Transient Ischemic Attack* (TIA), também conhecido como AVCi temporário. Na Figura 1 é ilustrado os dois tipos de AVC.

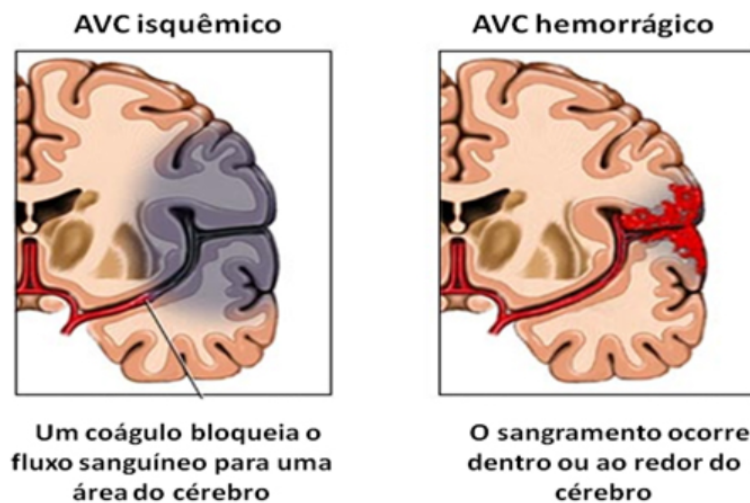


Figura 1 – Diferentes tipos de AVC. Fonte: SBAVC.

Os AVCs ainda possuem subtipos, sendo dois subtipos para o AVCi: Trombótico e Embólico, e também dois subtipos para o AVCh: Intracerebral e Subaracnóide (LEPPALA et al., 1999).

2.1.1 AVC Isquêmico

O AVCi ocorre quando um vaso sanguíneo é obstruído por um coágulo e o tecido cerebral é privado de sangue. Segundo Tsao et al. (2022) cerca de 87% de todos os AVCs são Isquêmicos. Quando o fluxo sanguíneo é bloqueado apenas por um período curto de tempo, geralmente não mais que 5 minutos, corresponde a um TIA. Segundo a *American Stroke Association* mais de um terço das pessoas que tiveram um TIA e não foram tratadas sofrem um AVC grave em 1 ano. Cerca de 10% a 15% das pessoas terão um AVC grave dentro de 3 meses após o TIA. O TIA, também é uma emergência médica e não há como diferenciar no início apenas pelos sintomas se o paciente sofre de AVCi ou TIA. O coágulo da TIA geralmente se dissolve sozinho ou é desalojado.

Segundo a WSO, quando o coágulo que obstruiu o vaso é formado na mesma região que o vaso sanguíneo é chamado de **Trombose Cerebral**. Existem dois tipos de trombose

cerebral: a venosa e a arterial. As trombozes arteriais são as mais frequentes, geralmente antecedem o TIA (cerca de 20% dentro do mesmo território arterial). A principal causa das trombozes cerebrais são os depósitos de gordura nos vasos sanguíneos e nas artérias chamados de aterosclerose.

Ainda de acordo com a WSO, o coágulo também pode ter sido formado em outra região do corpo, ter soltado uma parte e ter sido transportado pela corrente sanguínea até chegar em um vaso pequeno, esse cenário é chamado de **Embolia Cerebral**. Geralmente os coágulos da embolia cerebral são formados no coração ou nas artérias da parte superior do tórax e pescoço. A principal causa de embolia é o batimento cardíaco irregular chamado fibrilação atrial. Na Figura 2 é ilustrado os dois subtipos de AVCi.

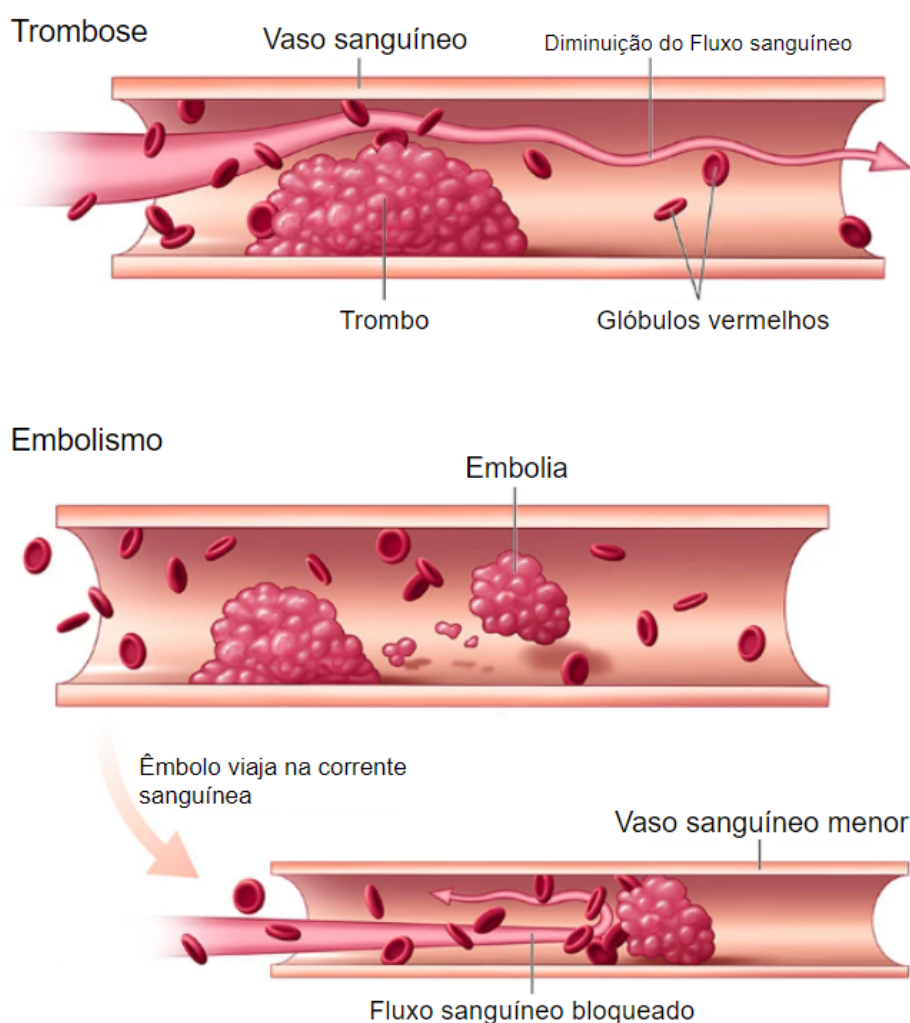


Figura 2 – Ilustração dos subtipos de AVCi. Fonte: Adaptado de Clinic (2021).

O fato de uma região está sofrendo um AVCi, não implica necessariamente que todas as células da região vão morrer de imediato. A sobrevivência/recuperação das células dependem diretamente do tempo e intensidade que o AVCi teve (MOZAFFARIAN et al., 2015). Existe uma região ao redor do tecido que está em processo de morte que, se o sangue voltar rapidamente, pode ser recuperada totalmente/parcialmente, essa região se

chama zona de penumbra. Identificar a zona de penumbra é importante para limitar a área da morte, e com isso as sequelas. Com essa informação é possível fazer uma avaliação qualitativa do tecido que ainda pode ser recuperado, e com isso conseguir prever com precisão o grau e o volume da região afetada (MENON et al., 2011; YOO et al., 2012). Entretanto, ainda não existe uma forma padronizada de mensurar e localizar a zona de penumbra (CHIN et al., 2002).

2.1.2 AVC Hemorrágico

O AVCh, ou hemorragia cerebral, ocorre quando um vaso se rompe e o sangue se acumula e cria pressão no tecido cerebral circundante. Esse rompimento pode ocorrer devido a vários fatores, sendo os principais: aneurisma, malformação vascular, hipertensão arterial, problemas na coagulação do sangue ou traumatismos. Apesar de menos comum, o AVCh é responsável por quase metade das mortes do AVC (SARBAZI et al., 2018).

O aneurisma é uma protuberância em uma região enfraquecida de um vaso sanguíneo. Se não for tratado, o aneurisma continua a enfraquecer até se romper e sangrar no cérebro. A ruptura do vaso sanguíneo também pode acontecer em decorrência do arteriosclerose, visto que as paredes perdem a elasticidade, ficam rígidas e finas (LEPPALA et al., 1999).

O sangramento pode ocorrer dentro do cérebro, sendo chamado de **Intracerebral**, ou pode ocorrer na superfície do cérebro, chamado de **Subaracnóide**. Na Figura 3 é ilustrado os dois subtipos de AVCh. O primeiro sinal mais comum de um AVCh do tipo subaracnóide, é uma forte dor de cabeça com início em uma fração de segundo e sem causa conhecida, geralmente chamada de *Thunderclap headache* (DUCROS; BOUSSER, 2013). A hemorragia intracerebral está desproporcionalmente associada a maior mortalidade e incapacidade de longo prazo em comparação com o AVCi (O'CARROLL; BROWN; FREEMAN, 2021).

2.2 Diagnóstico e Tratamento

Segundo a *American Stroke Association* o diagnóstico e tratamento do AVC geralmente é feito pelo médico neurologista, que será baseado nas informações de histórico médico do paciente, e variados tipos de testes. Existem vários testes que podem ser utilizados para o diagnóstico, os principais são: testes elétricos, de imagem e de fluxo sanguíneo. Os testes elétricos registram os impulsos elétricos do cérebro, os testes de imagem fornecem a imagem interna do cérebro, e os sanguíneos mostram qualquer problema que possa causar alterações no fluxo sanguíneo para o cérebro.

Nos exames de neuroimagens, existem duas técnicas mais comumente utilizadas: Imagem por Ressonância Magnética (IRM) ou Tomografia Computadorizada (TC). Descobrir qual o tipo de AVC que o paciente possui é importante visto que o tratamento é

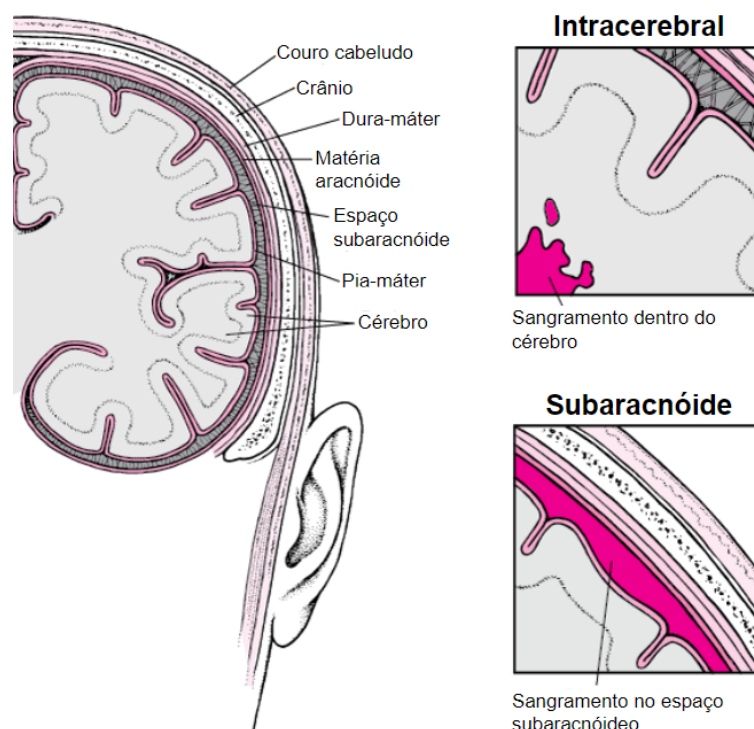


Figura 3 – Ilustração dos subtipos de AVCh. Fonte: Adaptado de Chong (2020).

diferente para cada tipo (RAFFIN et al., 2006). Em ambas as técnicas é possível descobrir qual o tipo de AVC, a IRM fornece uma imagem com melhor qualidade, no entanto a TC é mais rápida, e portanto geralmente é preferível dado a gravidade na demora do diagnóstico (WARDLAW; FARRALL, 2004).

Segundo a *American Stroke Association* os efeitos do AVC dependem de qual região foi afetada. Se o AVC ocorreu no hemisfério esquerdo, as possíveis consequências mais comuns são: paralisia do lado direito do corpo, problemas de fala, comportamento lento e cauteloso. Se o AVC ocorreu no hemisfério direito, os possíveis problemas mais comuns são: paralisia do lado esquerdo do corpo, problemas de visão, comportamento rápido e inquisitivo. Problemas em ambos os hemisférios podem causar perda de memória. Quando o AVC ocorre no tronco cerebral, dependendo da gravidade da lesão, pode afetar os dois lados do corpo.

De acordo com a SBAVC, o AVC é a principal causa de perda de independência em todo mundo, existem diversas outras sequelas que o AVC pode causar. Considerando o tratamento durante o AVC, novas drogas trombolíticos e anticoagulantes, como o ativador do plasminogênio tecidual (*recombinant tissue plasminogen activator*, rt-PA), demonstraram resultados promissores, podendo diminuir a extensão dos danos (TEIXEIRA; SILVA; FERREIRA, 2004; MASSENA, 2012). O tratamento pós-AVC depende das sequelas que o paciente teve, dessa forma, podem necessitar de vários especialistas como por exemplo: neuropsicólogos, fonoaudiólogo, fisioterapia, dentre outras especialidades.

2.3 Exames de Neuroimagens

A Tomografia Computadorizada (TC), proposta por Godfrey N. Hounsfield e Allan M. Comark em 1972, combina uma série de imagens de raios-X tiradas de diferentes ângulos ao redor do corpo e usa processamento do computador para criar imagens transversais do corpo (HOUNSFIELD, 1973; YANG, 2017).

Em geral a TC é mais detalhada que a radiografia. Na radiografia o raio-X é passado pelo corpo, uma parte é absorvida ou espalhada pela estrutura interna e o padrão de raio-X restante é enviado a um detector para que a imagem seja registrada para avaliação posterior (TUM; MIDDLETON, 2006). Na TC, o feixe de raio-X se move em um círculo ao redor do corpo, permitindo várias visualizações diferentes da mesma região de interesse e com mais detalhes (YANG, 2017). Na Figura 4 é ilustrado um tomógrafo, equipamento usado para fazer a TC.

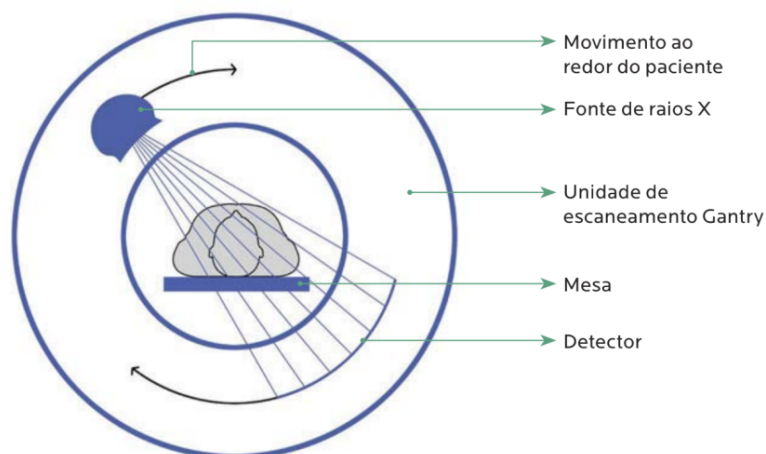


Figura 4 – Ilustração de um tomógrafo. Fonte: Freitas (2013).

Os planos anatômicos são superfícies 2D imaginárias utilizadas para dividir o corpo facilitando as descrições de localização e movimento. Existem 3 planos anatômicos comumente utilizados: axial, coronal e sagital. Em uma imagem de TC os cortes originais são nos planos axial e coronal, utilizando a reconstrução multiplanar (RMP), pode-se reconstruir os outros planos como: sagital, parassagital, tridimensional (FREITAS, 2013). Na Figura 5 é apresentado algumas imagens 2D de tomografias obtidas nos 3 planos anatômicos mencionados anteriormente.

As imagens da TC possuem escala diferente da que normalmente é utilizada para representar imagens digitais (RGB, com valores de 0 á 255), isso ocorre porque as imagens TC obedecem a uma escala diferente, a escala de *Hounsfield*. As imagens TC possuem grande variedade de tons de cinza representados pela escala de *Hounsfield*. A escala de *Hounsfield*, possui valores de 1000 a -1000, chamados de Unidade *Hounsfield* (HU), no

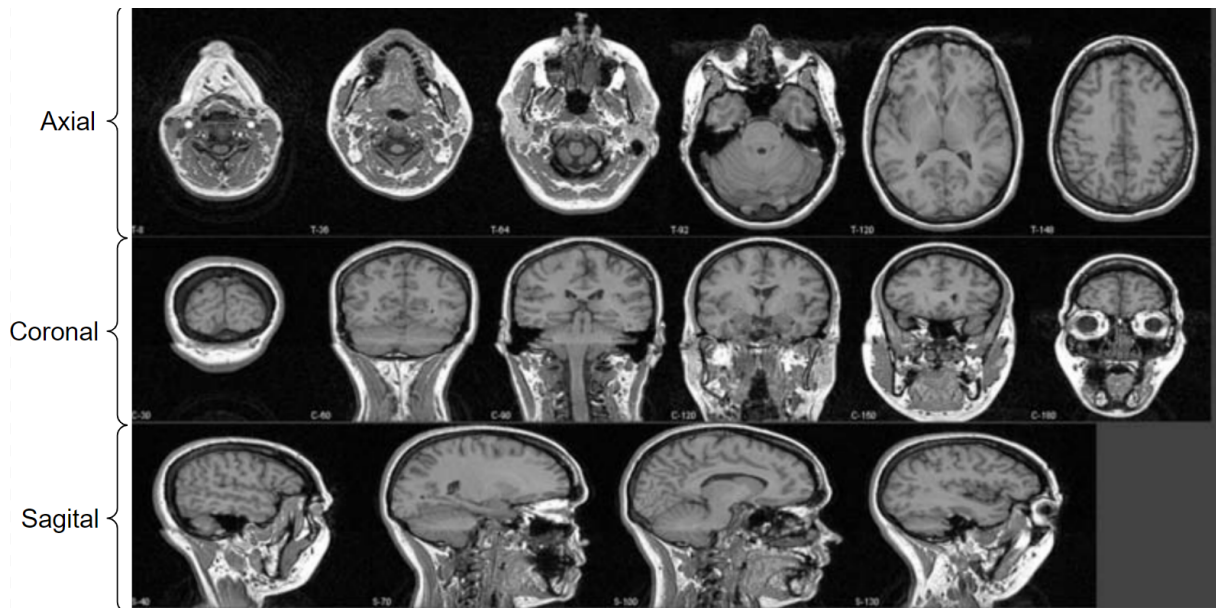


Figura 5 – Exemplo de imagens obtidos nos planos axial, coronal e sagital. Fonte: Adaptado de Bankman (2008).

qual 0 representa a água e -1000 o ar (FREITAS, 2013). Na Figura 6 é ilustrado a escala de *Hounsfield*.

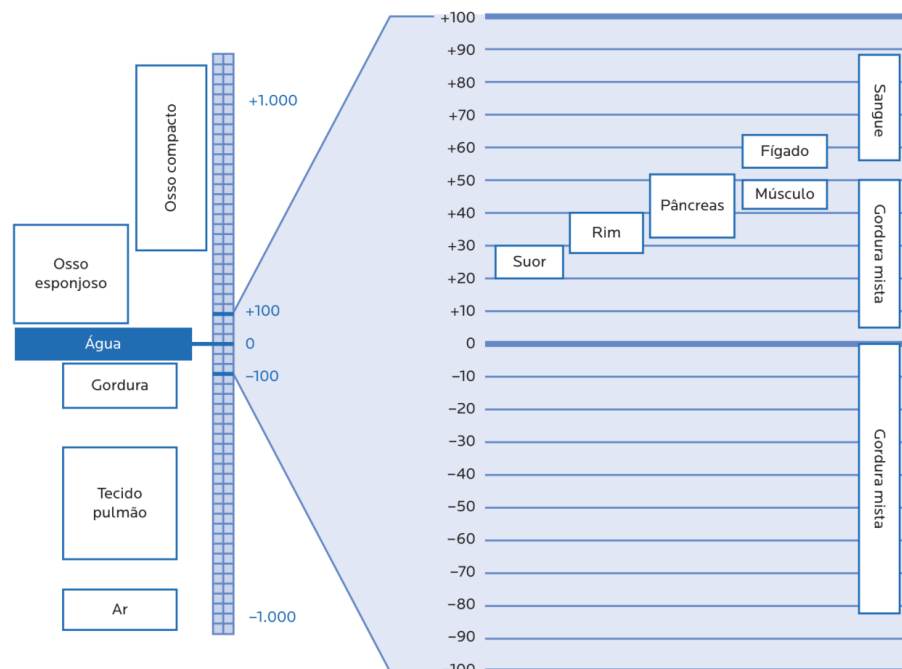


Figura 6 – Ilustração da escala de *Hounsfield*. Fonte: Freitas (2013).

Considerando a quantidade de informações que uma imagem TC possui, em conjunto com o uso de PDI, podem ser criadas representações com maior qualidade, seja realçando regiões de interesse, ou eliminando regiões que não são relevantes, e com isso é possível

reduzir o tempo de análise dos exames e melhorar a precisão do diagnóstico, auxiliando o médico especialista em suas interpretações (LUIZ et al., 2014).

2.4 Estado da arte

O Diagnóstico Auxiliado por Computador ou *Computer-Aided Diagnosis* (CAD) busca auxiliar os médicos especialistas no processo de análise de imagens médicas aumentando a sensibilidade e especificidade do diagnóstico/prognóstico. Os sistemas CAD tem sido foco de diversas pesquisas, visto que além de auxiliar especialistas os sistemas computacionais possuem a capacidade de lidar com grande quantidade de informações, em menor tempo (YANASE; TRIANTAPHYLLOU, 2019). Dentre outras modalidades de sistemas CAD, o diagnóstico baseado em imagem também tem apresentado resultados satisfatórios para diversas patologias (ZAHOOOR et al., 2020; LIAQAT et al., 2020; INAMDAR et al., 2021).

Misawa et al. (2017) testaram a eficiência de um sistema CAD na prática. Os autores desenvolveram um sistema CAD para verificar lesões colorretais a partir de imagens de endoscopia. O sistema desenvolvido possui como saída o diagnóstico de neoplásico ou não neoplásico. Para testar a eficiência do sistema CAD, foi selecionado uma base de dados com imagens, que foi avaliada pelo sistema CAD, quatro Endoscopistas especialistas e três estagiários. Com isso a precisão dos 3 diagnósticos foram comparadas estatisticamente por meio do teste exato de *Fisher* e o teste t de *Student*. Como resultado o sistema CAD obteve performance equivalente aos especialistas e superior a dos estagiários. Dessa forma os autores concluem que o sistema CAD proposto é uma poderosa ferramenta de tomada de decisão para endoscopistas menos experientes.

Sistemas CAD baseados em imagens, possuem uma etapa essencial, que é a escolha de quais atributos serão extraídos dos exames, de forma que ajudem a identificar a região de interesse (SARAIVA; BORCHARTT, 2014). A escolha/geração desses atributos podem ser feitas pelo radiologista ou de forma automática. A geração automática desses atributos é importante, visto que melhora a reprodutibilidade e a velocidade do processo em geral. O processo de identificação da região de interesse, pode ser interpretado como um problema de classificação de imagens, detecção de um objeto, segmentação, dentre outras abordagens, e essa definição depende de qual é a resposta esperada do sistema CAD.

No contexto do AVC, existem pesquisas nas frentes relacionadas ao prognóstico, diagnóstico e tratamento (SIRSAT; FERMÉ; CÂMARA, 2020). Sistemas computacionais conseguem capturar informações imperceptíveis ao olho humano, como por exemplo em Wu et al. (2019) que a partir da TC sem contraste, propôs um método para detectar precocemente o AVCi, e como resultado foram detectados 108 AVC invisíveis ao radiologista.

A classificação de AVC no contexto dessa pesquisa, refere-se ao problema de receber

como entrada uma imagem de TC, e um modelo computacional determinar se a imagem possui: AVCi, AVCh ou que não possui AVC (normal). A definição desse problema é caracterizado como um problema de classificação de imagens. Existem diversas abordagens utilizadas para resolver o problema da classificação de AVC, desde o uso de técnicas mais recentes como aprendizagem profunda ou *Deep Learning*, a métodos clássicos de extração de características e classificação. A seguir, serão apresentadas as diferentes abordagens utilizadas nos últimos anos para resolver o problema da classificação de AVC.

Kanchana e Menaka (2020) propôs o uso de **extratores clássicos**, visando a segmentação e classificação de AVCi. Onde utiliza uma nova abordagem para segmentar as regiões da imagem em normal e anormal, após a segmentação a região é enviada para os extratores: *First order features* (FOF), *Gray level run length matrix features* (GLRLM), *Gray level co-occurrence matrix features* (GLCM) e *HU Moments* (HUM), em seguida para selecionar apenas as características mais importantes foi utilizada a técnica ranqueamento de características, com o método estatístico ANOVA. Logo após quatro classificadores foram testados: *Logistic Regression* (LR), *Support Vector Machine* (SVM), *Random Forest* (RF) e uma rede neural, a melhor acurácia obtida entre as combinações foi de 99.79%. Badriyah et al. (2020) também propôs o uso GLCM após o processamento das imagens e testou com 8 classificadores: KNN, *Naive Bayes*, LR, *Decision Tree*, RF, MLP, abordagem com aprendizado profundo e SVM, utilizando a metodologia validação cruzada com 10 dobras e validação cruzada com LOO a melhor configuração foi com RF, chegando a 95.97% de acurácia.

Sarmiento et al. (2020) propuseram um novo método para extração de características do cérebro em imagens TC com o objetivo de melhorar a classificação de AVC. O novo extrator utiliza a janela de *parzen* para estimar a probabilidade de cada pixel pertencer a faixas pré-definidas que delimitam cada uma das classes usando a escala HU. Nessa abordagem a extração é feita por hemisfério e no fim as características são unidas em um único vetor e enviadas ao classificador. Os classificadores testados foram: Bayes, KNN, *Optimum Path Forest* (OPF), MLP e SVM. A configuração que obteve melhor performance foi o SVM com *kernel* RBF, alcançando acurácia de 98.41%. Em Vasconcelos et al. (2020), também foi proposto um novo extrator chamado ABTD-Adaptative, que considera a vizinhança dos pixels e as informações de intervalos das classes para fazer a extração. Foram testados: Bayes, KNN, OPF, MLP e SVM, a melhor acurácia obtida foi de 98.13% com o SVM com *kernel* RBF.

Considerando modelos de **Deep Learning**, um dos métodos mais populares entre os pesquisados são as Redes Neurais Convolucionais ou *Convolutional Neural Network* (CNN). As CNNs são Redes Neurais Artificiais que utilizam o processo de convolução matemática. Existem diversas arquiteturas conhecidas dessas redes na literatura, e normalmente essas redes são usadas como extrator de características das imagem, em conjunto com algum

classificador de padrões, que utiliza as características extraídas em vez de usar diretamente a imagem como entrada.

Os autores Dourado et al. (2019) e Dourado et al. (2020) propuseram 2 abordagens para classificação de AVC utilizando CNNs em imagens de TC. Na primeira abordagem, os autores propuseram o uso do extrator de características *NASNet Large* e do classificador de padrões Bayesiano para detecção de AVC, atingindo acurácia de 100% (DOURADO et al., 2019). Na segunda, foram testadas 12 arquiteturas conhecidas na literatura: *VGG16*, *VGG19*, *MobileNet*, *ResNet50*, *InceptionV3*, *Xception*, *InceptionResNetV2*, *DenseNet121*, *DenseNet201*, *DenseNet169*, *NASNetLarge* and *NASNetMobile* para a extração de características. Para a classificação de padrões foram testados 6: Bayesiano, MLP, *K-Nearest Neighbors* (KNN), RF e SVM. A configuração que obteve melhor resultado foi o extrator da *VGG16* e o KNN como classificador de padrões, atingindo 100% de acurácia (DOURADO et al., 2020). Seguindo essa mesma abordagem, Chen et al. (2022) propôs o uso de TC sem contraste, a combinação que obteve o melhor resultado foi a extração com a *Resnet50*, chegando a 98.7% de acurácia, mas o autor alerta para a demora no processo de extração de características.

A Tomografia por Impedância Elétrica (TIE) foi utilizada por Agnelli et al. (2020) para fazer a classificação de AVC em imagens TC. Os autores realizaram um comparativo entre utilizar diretamente a matriz *Dirichlet-to-Neumann* (DN) e a aplicação de funções *Virtual Hybrid Edge Detection* (VHED) antes de enviá-las para a classificação. Foram utilizadas as redes CNN e a CNN totalmente conectada. O formato que obteve melhor desempenho foi a aplicação das funções VHED em conjunto com a CNN totalmente conectada, alcançando uma acurácia de 99.7% no conjunto de teste.

O aprendizado multimodal, com técnicas de fusão de informação em arquiteturas de aprendizado profundo, também foi explorado na literatura pelos pesquisadores Roder et al. (2021), onde propuseram aplicar as transformadas de Fourier como entrada para a rede além das imagens, para que a rede possa aprender simultaneamente. O classificador utilizado foi o *Restricted Boltzmann Machines* (RBM), reconhecido por ser mais simples, obteve a melhor acurácia de 99.72%. Lo, Hung e Lin (2021) propôs o uso de CNNs profundas em TC sem contraste para a classificação entre paciente saudável e com AVCi. Com 1.254 imagens de 96 pacientes, foram testadas as redes utilizando o *transfer learning* da *Inception-V3*, *AlexNet* e *Resnet-101*, obtendo acurácia máxima de 97.12% com a *AlexNet*.

Os autores Gautam e Raman (2021a) também propuseram uma nova abordagem para extração de características de imagens TC: *Local Neighbourhood Pattern* (LNP) para extrair informações espaciais e quânticas das imagens. A extração baseia-se na comparação de pixels diagonais com a média de toda a intensidade da imagem e, os pixels restantes são comparados com a pixels diagonais em seu lado esquerdo. Para classificação de AVC foram testados os classificadores: SVM, KNN, e variações de árvores *ensemble*. A melhor acurácia

obtida foi de 82.65% com o *Ensemble Bagged Trees* (EBT), usando validação cruzada com 10 dobras. Ainda os mesmos autores propuseram outro formato para a classificação de AVC utilizando fusão de imagens e CNN, com resultados ainda melhores, onde desenvolveram uma CNN com nova arquitetura de 13 camadas, com tamanho bem menor que as redes mais famosas da literatura como *Alexnet*, *Resnet* dentre outras, tornou a classificação muito mais rápida. A melhor acurácia obtida em um dos conjuntos de teste foi de 98.77% (GAUTAM; RAMAN, 2021b).

A revisão bibliográfica mostra que algumas abordagens atingiram ótimas performances, como por exemplo os trabalhos baseados em *Deep Learning* que alcançaram 100% de acurácia (DOURADO et al., 2019; DOURADO et al., 2020). No entanto, um fator quem nem sempre é considerado nas pesquisas é o tempo de extração. A solução ideal para o problema de classificação de AVC, é um modelo que seja o mais acurado no menor tempo possível. Ponderar esses fatores de performance e tempo é uma tarefa difícil, visto que em geral para obter performance ótima, técnicas mais complexas são necessárias, no entanto normalmente possuem um custo computacional maior. Sendo assim, a suposição encontrada é que métodos menos complexos que os utilizados em *Deep Learning*, que possuam um tempo de processamento menor, apesar de provavelmente não serem tão precisos quanto essas redes, em alguns casos, podem ser preferíveis devido a necessidade rápida do diagnóstico que o AVC possui.

Tendo em vista também que o principal diferencial entre os trabalhos apresentados na revisão bibliográfica é a etapa de extração de características, nota-se a relevância e o quanto essa etapa pode impactar na qualidade dos resultados e no tempo de execução. Afinal, a região de interesse só será encontrada corretamente pelo classificador se os atributos fornecidos forem relevantes para identificar a região. Sendo assim, este trabalho propõe um novo método de extração de características em imagens de TC baseado nas faixas de densidade radiológica utilizando mistura de gaussianas, com o objetivo de identificar e classificar exames que apresentam incidência de AVC.

3 Reconhecimento de Padrões

A Inteligência Artificial (IA), é o campo de estudo que busca habilitar máquinas a executar tarefas comumente associadas a seres inteligentes. A IA é relevante em qualquer atividade intelectual, e hoje possui atuação em diversas áreas, como por exemplo: aprendizado, percepção, jogos e diagnósticos de doenças (RUSSELL, 2010). O reconhecimento de padrões busca descobrir de forma automática padrões nos dados por meio de algoritmos de computador, e com essas informações realizar ações, como por exemplo classificação dos dados, uma tarefa de IA contida na subárea Aprendizagem de Máquina (BISHOP; NASRABADI, 2006).

O aprendizado de máquina se propõe a resolver diversos tipos de problemas, os mais conhecidos são: supervisionados, não supervisionados e com reforço. O aprendizado não supervisionado consiste em não apresentar uma saída esperada durante o treinamento, dessa forma o algoritmo buscará encontrar padrões/agrupar os dados. O aprendizado por reforço ocorre por meio da tentativa e erro, buscando melhorar as estratégias para maximizar as recompensas (CHINNAMGARI, 2019). O aprendizado supervisionado, foco dessa pesquisa, são problemas que fornecem exemplos de amostras com suas saídas esperadas, e com essas informações os algoritmos aprendem os padrões do problema (BISHOP; NASRABADI, 2006). Quando a saída esperada é um número finito de categorias discretas, são chamados de problemas de classificação, e quando a saída é uma variável contínua, são chamados de problemas de regressão.

Em problemas de classificação, existem 2 fases: treinamento e teste. O treinamento é o momento em que o algoritmo vai aprender os padrões de um determinado conjunto de dados. Esse conjunto de dados, possui N linhas, onde cada linha é uma amostra/padrão que possui um conjunto de variáveis/atributos e uma saída esperada, popularmente chamada de classe. A fase de teste, ocorre após o treinamento quando se tem uma amostra e o objetivo é descobrir sua classe. A fase de teste também é utilizada para determinar a performance de um algoritmo, utilizando um conjunto de dados para validação. Para realizar esse processo de validação existem diversas técnicas que serão exploradas em maiores detalhes no Capítulo 5.

Considerando o contexto de classificação de imagens, o processo de treinamento consiste em: receber um conjunto imagens para treinamento, extrair as características dessas imagens, gerando um conjunto de dados que será enviado para o treinamento de um classificador de padrões. Enquanto que o processo de teste corresponde a: receber uma imagem de entrada, extrair as características e enviar as características extraídas para o classificador que fará o teste e retornará a classe da imagem de entrada.

Determinar qual classificador utilizar para resolver um problema, muitas vezes ocorre de maneira arbitrária, baseado no conhecimento teórico que o indivíduo tem do classificador, ou na habilidade técnica para implementar e/ou utilizar o classificador. No entanto por meios de experimentos pode-se chegar em resultados mais adequados ao problema. Com os experimentos pode-se analisar qual classificador obtém melhor performance, o tempo de treinamento e teste, que em geral são informações relevantes para decidir a melhor abordagem para um problema específico.

Sendo assim, nessa pesquisa utiliza essa abordagem dos experimentos para definir o classificador ideal. Foram selecionados classificadores provenientes de áreas distintas, visando diversificar ao máximo as possibilidades e contribuir para uma percepção geral da performance dos extratores mesmo em abordagens com bases teóricas diversas. Os classificadores escolhidos para realizar os experimentos da classificação de AVC e utilizados para construção do extrator, são descritos em maiores detalhes nas sessões seguintes.

3.1 Mistura de Gaussianas

A Mistura de Gaussianas é uma combinação linear de gaussianas, em que ao se determinar a quantidade correta de gaussianas, suas médias e covariâncias, quase qualquer densidade contínua pode ser aproximada com uma precisão arbitrária (BISHOP; NASRABADI, 2006). Na Figura 7, são apresentados os dados do conjunto *Old Faithful*, no qual contém dados sobre a erupção de um gêiser no estado de Wyoming-EUA, onde o eixo X corresponde a duração e o eixo y o tempo em minutos até a próxima erupção. Na Figura 7a nota-se que uma única gaussiana é incapaz de se ajustar com precisão aos dados, devido os dois aglomerados mais dominantes, enquanto que na Figura 7b, a combinação das duas gaussianas que representa melhor os dados.

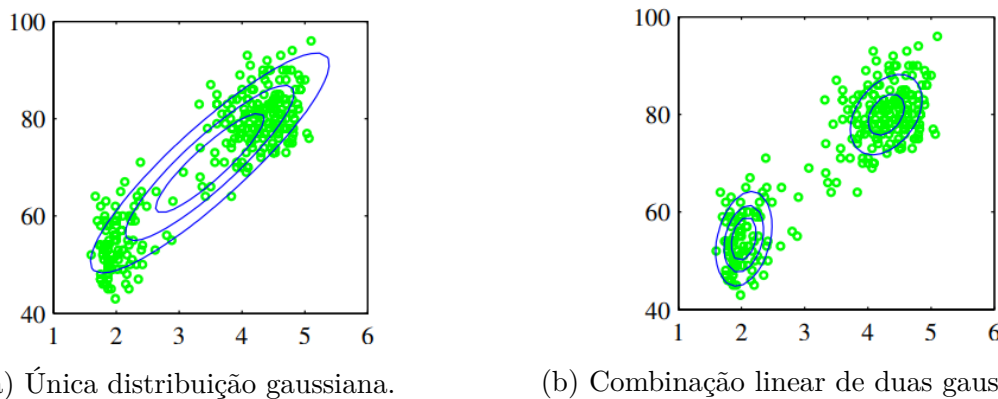


Figura 7 – Exemplo considerando a base de dados *Old Faithful*. Fonte: Bishop e Nasrabadi (2006).

Um Modelo de Mistura de Gaussianas ou *Gaussian Mixture Models* (GMM), é uma função de densidade probabilidade paramétrica, que representa uma soma ponderada

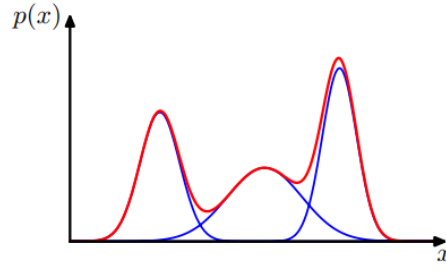


Figura 8 – Exemplo de mistura com 3 gaussianas. Fonte: Bishop e Nasrabadi (2006).

de densidades das componentes gaussianas (REYNOLDS, 2009; BISHOP; NASRABADI, 2006). Na Figura 8 é apresentado um exemplo com 3 gaussianas, linha azul, e a GMM na linha vermelha com a soma ponderada. Uma GMM é definida por:

$$p(x) = \sum_{i=1}^K w_i g(x|\mu_i, \Sigma_i) \quad (3.1)$$

Onde x é uma amostra da base de dados, um vetor D-dimensional. O w_i é o coeficiente da mistura, ou seja, o peso que cada componente gaussiana terá, e o K representa a quantidade de componentes. O somatório dos pesos $\sum_{i=1}^K w_i$ deve ser igual a 1, para satisfazer as propriedades de uma função de densidade probabilidade. Uma componente é definida por uma gaussiana multivariada:

$$g(x|\mu_i, \Sigma_i) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i) \right\} \quad (3.2)$$

Em que μ_i significa o vetor de média e Σ_i a matriz de covariância da i-ésima componente. O objetivo de uma GMM é representar a distribuição dos dados o mais preciso possível, para isso, é necessário estimar os parâmetros da GMM visando maximizar a sua verossimilhança de acordo com os parâmetros utilizados (MCLACHLAN; BASFORD, 1988). Existem dois algoritmos populares para maximizar os parâmetros: Expectation-Maximization (EM) e Maximum A Posteriori (MAP), na literatura o algoritmo EM ainda é mais utilizado que o MAP (DEMPSTER; LAIRD; RUBIN, 1977; REYNOLDS, 2009).

3.1.1 Algoritmo Expectation Maximization

O algoritmo EM, é um algoritmo iterativo que considera os dados de treinamento para otimização dos parâmetros: médias, matrizes de covariância e coeficientes de mistura de cada uma das componentes de cada classe. O algoritmo EM é dividido em 2 etapas: *E-Step* e o *M-Step*:

1. Inicialize os parâmetros: μ_i, Σ_i, w_i ;

2. *E-Step*: Calcule as responsabilidades para cada amostra x os valores μ_i, Σ_i, w_i , da iteração atual;
3. *M-Step*: Atualize os parâmetros μ_i, Σ_i, w_i de acordo com as responsabilidades obtidas no *E-Step*.
4. Repita os passos 2 e 3 por N iterações ou até que não exista mudança significativa de uma iteração e outra.

Existem várias abordagens para inicialização dos parâmetros, visto que quanto melhor as gaussianas estiverem bem posicionadas, menos iterações serão necessárias para ajustá-las para o lugar correto. Alguns exemplos de abordagem são: aleatória e utilizando o algoritmo k-means. A abordagem aleatória, para a média por exemplo pode tanto ser um valor totalmente aleatório, quanto ser definido de acordo com uma amostra aleatória. A abordagem utilizando o k-means, consiste em executar o algoritmo nos dados e utilizar os centroides como média, nessa pesquisa essa foi a abordagem de inicialização utilizada.

No E-step o cálculo da responsabilidade, que é a probabilidade de uma amostra pertencer a uma componente, é definida como:

$$\gamma_{nk} = \frac{w_i g(x|\mu_i, \Sigma_i)}{\sum_{i=1}^K w_i g(x|\mu_i, \Sigma_i)} \quad (3.3)$$

Para cada amostra n a responsabilidade γ_{nk} é calculada para cada uma das k componentes. Nessa etapa, cada amostra terá sua componente correspondente.

No M-step ocorre a atualização dos parâmetros de cada componente de acordo com os amostras da sua componente. As atualizações dos parâmetros: μ, Σ, w , seguem respectivamente, as seguintes fórmulas:

$$\hat{\mu}_k = \frac{1}{N_k} \sum_{n=1}^N y_{nk} x_n \quad (3.4)$$

$$\hat{\Sigma}_k = \frac{1}{N_k} \sum_{n=1}^N y_{nk} (x_n - \hat{\mu}_k)(x_n - \hat{\mu}_k)^T \quad (3.5)$$

$$\hat{w}_k = \frac{N_k}{N} \quad (3.6)$$

E assim ao longo de t iterações o algoritmo EM posiciona as componentes para cada uma das classes. Nota-se que o número de iterações do EM, é um parâmetro que impacta diretamente a qualidade da GMM, visto que ele determina o quão bem as componentes vão representar os dados. Poucas iterações podem ocasionar em componentes que não se ajustaram corretamente.

3.1.2 Aplicações GMM

Em essência uma GMM é um algoritmo de agrupamento baseado em modelos de mistura, logo sua aplicação primária se dá em problemas de clusterização, proveniente do aprendizado não supervisionado (BISHOP; NASRABADI, 2006). No aprendizado de máquina as principais aplicações da GMM são: detecção de anomalias (ZONG et al., 2018; LAXHAMMAR; FALKMAN; SVIESTINS, 2009; DING et al., 2019; LI et al., 2016), segmentação de imagens (GREENSPAN; RUF; GOLDBERGER, 2006; FARNOUSH; ZAR, 2008; BALAFAR, 2014; BAN; LIU; CAO, 2018; JI et al., 2012) e séries temporais (EIROLA; LENDASSE, 2013; DING et al., 2019). No entanto a GMM também pode ser utilizada em conjunto com outras técnicas, como por exemplo o classificador Bayesiano, para resolver problemas supervisionados de classificação (VERVERIDIS; KOTROPOULOS, 2005; BOLUKI et al., 2017; KAYABOL; KUTLUK, 2016).

3.1.2.1 Classificador Bayesiano

O classificador Bayesiano é um modelo probabilístico baseado no teorema de Bayes, utilizado para determinar a classe de uma amostras de acordo com a probabilidade a *posteriori* da amostra pertencer a uma classe. Nessa classificação, a tomada de decisão é baseada no princípio de escolher a opção mais provável ou de menor risco (custo esperado). A probabilidade a *posteriori* pode ser calculada pela regra de Bayes:

$$P(w_i | x) = \frac{p(x | w_i) P(w_i)}{p(x)}, \quad (3.7)$$

onde $P(w_i)$ é a probabilidade a priori da classe, representa o conhecimento a priori sobre a aparição de exemplos das classes do problema. $p(x | w_i)$ é a função de verossimilhança ou probabilidade condicional da classe, e $p(x)$ é a evidência que age como uma constante de normalização, definida por:

$$p(x) = \sum_{i=1}^c p(x | w_i) P(w_i), \quad (3.8)$$

em que c , é o número de classes (HART; STORK; DUDA, 2000).

O classificador Bayesiano precisa que a probabilidade a priori e sua função densidade de probabilidade ($p(x | w_i)$) sejam estimadas, para que seja considerado ótimo, em relação a minimização da probabilidade média do erro. Determinar a função densidade probabilidade, pode ser um desafio quando o vetor de característica é N dimensional, principalmente quando o conjunto de dados tem poucas amostras (HART; STORK; DUDA, 2000).

Em geral a distribuição mais utilizada é a Gaussiana, e quanto mais próxima da realidade for a função densidade probabilidade escolhida, mais o classificador Bayesiano se aproxima da perda média mínima de classificação (GONZALEZ; WOODS, 2009). A

função densidade probabilidade da GMM também pode ser utilizada em conjunto com o classificador Bayesiano. Para isso é criada uma GMM para cada classe, e a PDF da GMM de cada classe, é utilizada como estimador densidade de probabilidade no classificador de Bayes:

$$P(c_y|x) = \frac{p(x|c_y) * P(c_y)}{\sum_{y=1}^C p(x|c_y) * p(c_y)}, \quad (3.9)$$

onde $p(c_y|x)$ representa a probabilidade da amostra x pertencer a classe c_y , onde $y \in \{1, \dots, p\}$, em que p é o total de classes. $P(c_y)$ representa a probabilidade a priori de x pertencer a c_y e $p(x|c_y)$ representa a função densidade de probabilidade obtida na GMM da classe y . O denominador representa um fator de normalização, para garantir que as probabilidades de cada classe quando somadas totalizem 1. Para determinar a classe da amostra calcula-se a probabilidade da amostra pertencer a cada uma das classes $\max\{p(c_1|x), \dots, p(c_p|x)\}$ e seleciona a classe que tiver maior probabilidade.

Sendo assim, os principais parâmetros de um classificador Bayesiano com GMM, ou GMM para classificação, são: número de componentes de cada classes e quantidade de iterações do algoritmo EM. Como cada classe terá uma GMM é necessário informar previamente a quantidade de componentes que cada GMM/Classe terá. Determinar a quantidade de iterações não é uma tarefa simples, visto que depende da abordagem utilizada para inicialização das GMMs. Se a inicialização for muito boa, ou seja, conseguiu posicionar de primeira as componentes muito próximo dos lugares corretos, a quantidade de iterações para convergir para o melhor lugar das componentes será mínima, no entanto se a inicialização resultar nas componentes muito afastadas do lugar ótimo serão necessárias várias iterações no EM para posioná-las corretamente.

3.2 K Vizinhos Mais Próximos

O algoritmo K Vizinhos Mais Próximos, ou K Nearest Neighbors (KNN), é um algoritmo supervisionado não paramétrico, ou seja, não faz suposição sobre a distribuição dos dados, em que pode ser utilizado tanto para classificação quanto para regressão (FIX; HODGES, 1951; HART; STORK; DUDA, 2000). O KNN é uma variação do método Nearest Neighbors (NN), no qual identifica o vizinho mais próximo de uma amostra. A proximidade entre uma amostra e outra é determinada pela distância entre seus vetores n-dimensionais de característica. O KNN analisa os K vizinhos mais próximos, em vez de analisar apenas o vizinho mais próximo como o NN faz.

Cada amostra é analisada como um ponto no espaço, que possui um vetor n dimensional de características, a proximidade entre as amostras é definido pela distância entre dois pontos e para isso existem algumas métricas popularmente conhecidas:

- Euclidean: $d(x, y) = \sqrt{\sum_{i=1}^k (x_i - y_i)^2}$;
- Minkowsky: $d(x, y) = \left(\sqrt{\sum_{i=1}^k (|x_i - y_i|)^q} \right)^{\frac{1}{q}}$;
- Mahalanobis: $d(x, y) = \sum_{i=1}^k |x_i - y_i|$;
- Hamming: $d(x, y) = \sum_{i=1}^k |x_i - y_i| \begin{cases} x = y \rightarrow D_H = 0, \\ x \neq y \rightarrow D_H = 1. \end{cases}$

Dentre as métricas, a mais utilizada é a distância euclideana. No geral a distância entre dois pontos será bastante sensível a escala dos dados e/ou unidades físicas, por exemplo, um vetor que possua duas características: quantidade de quartos de uma casa e o valor do imóvel, por serem atributos com proporções muito diferentes a informação da distância pode ficar distorcida, e nesses casos é indicado padronizar a escala dos atributos. Existem diversas técnicas para padronizar a escala dos dados, a mais utilizada é a normalização popularmente conhecida como Min Max, no qual a nova escala do atributo será de 0 a 1. A normalização do atributo de uma amostra $x'_i \in [0, 1]$, é definida por:

$$x'_i = \frac{x_i - \min_{x_i}}{\max_{x_i} - \min_{x_i}} \quad (3.10)$$

no qual $\min_{x_i}$ e $\max_{x_i}$ são valores mínimos e máximos que o atributo possui.

Após a escolha da métrica que será utilizada para o cálculo da distância, e a normalização dos dados caso necessário, o processo para classificação pode ser apresentado. O processo de treinamento do KNN para classificação, é simplesmente armazenar as amostras de treinamento e suas respectivas classes na memória. O processo de previsão da classe para uma amostra não conhecida y é:

1. Calcular a distância da amostra y para todas as amostras de treinamento;
2. Selecionar as K amostras mais próximas da amostra y ;
3. Determinar qual a classe mais frequente dos vizinhos e atribuir essa classe a amostra y (COVER; HART, 1967).

Na Figura 9 é apresentado uma ilustração do processo descrito anteriormente. O ponto vermelho é a amostra desconhecida, e se o $K = 3$ a classe determinada para o ponto vermelho será a classe B, se $k = 5$ a classe determinada será a classe A. Para facilitar o processo de determinar a classe da amostra, é indicado valores impares para o K para evitar empates. Sendo assim, os principais parâmetros de um KNN para classificação é o K e a métrica utilizada para calcular a distância (ALFEILAT et al., 2019).

O valor ótimo para o parâmetro K varia de problema para problema, e encontrá-lo é uma tarefa computacionalmente cara. O método mais simples é variar o K e aplicar alguma técnica de validação, e com isso determinar o K que obtém a melhor métrica no

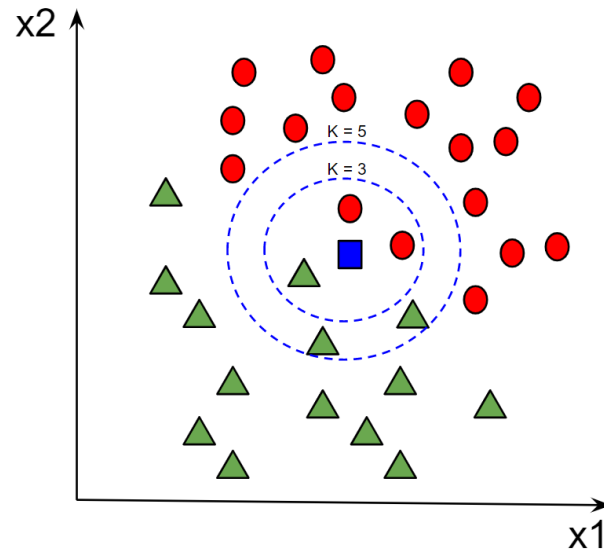


Figura 9 – Ilustração do processo de classificação do KNN. Fonte: Elaborado pelo autor.

conjunto de validação, ou o K que obtém a menor taxa de erro. A principal desvantagem do KNN é o custo computacional para manter a base de treinamento na memória, e por precisa computar a distância da nova amostra para todas as amostras de treinamento, mas em contra partida possui a vantagem de ser uma técnica simples, logo a interpretabilidade dos resultados é mais clara.

3.3 Rede Perceptron de Multicamadas

Uma Rede Perceptron de Multicamadas ou Multilayer Perceptron (MLP) é uma topologia específica de Rede Neural Artificial ou Artificial Neural Network (ANN). Mehrotra, Mohan e Ranka (1997) descreve as ANNs como um sistema computacional análogo as redes neurais biológicas. McCulloch e Pitts (1943) desenvolveram o primeiro modelo matemático de um neurônio artificial, que foi amplamente aplicado e modificado em trabalhos posteriores. Uma ANN é um grafo direcionado, que possui um conjunto de nós (vértices) conectados (arestas). A força de conexão entre os nós, os pesos, indica a extensão que um sinal ou atributo é amplificado ou diminuído por uma conexão.

Rosenblatt (1958) desenvolveu o Perceptron Simples (PS) a primeira ANN, contendo uma camada neural com um neurônio, que é capaz de resolver apenas problemas linearmente separáveis, visto que ANNs de 1 camada não são capazes de encontrar superfícies de decisão para problemas não linearmente separáveis (HAYKIN, 2001). Na Figura 10 é apresentada a rede PS, onde recebe como entrada os atributos $x_1, \dots, x_n$ de uma amostra, junto com o *bias* x_0 , que serão associadas aos pesos $w_0, \dots, w_n$, por meio da função aditiva ($\sum$), em seguida o resultado u , passará por uma função de ativação $\varphi(\cdot)$ e por fim será gerado a saída y .

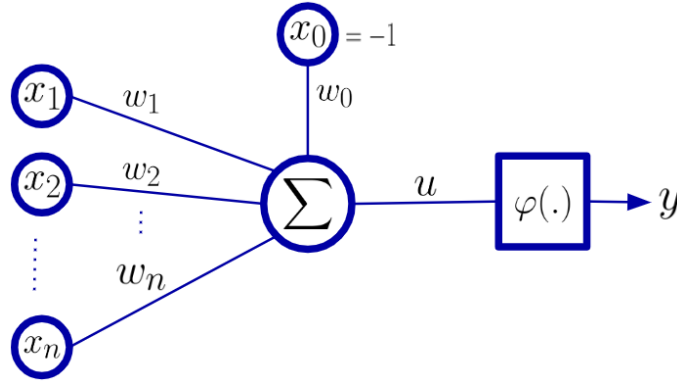


Figura 10 – Ilustração da rede PS. Fonte: Elaborado pelo autor.

Para determinar a classe y de uma amostra com x_n atributos, a rede PS aplica o seguinte cálculo:

$$y = \varphi\left(\sum_{i=0}^N w_i x_i\right) \quad (3.11)$$

onde a função de ativação $\varphi(\cdot)$, é uma função degrau. E com isso dado um conjunto de atributos de uma amostra a rede PS determina se é da classe 0 ou 1.

Como evolução do PS, a MLP é capaz de resolver problemas não linearmente separáveis, e consiste em: uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída, onde todos os neurônios de cada camada estão totalmente conectados (HAYKIN, 2001). Na Figura 11 é apresentado uma ilustração de uma MLP com apenas 1 camada oculta.

Ainda considerando a Figura 11, a camada de entrada é representada pelos neurônios $x_0, \dots, x_D$, ou seja, essa rede aceita um vetor de atributos de tamanho D como entrada. A camada de saída corresponde aos neurônios $y_1, \dots, y_K$, ou seja, essa rede consegue resolver problemas com múltiplas classes, cada neurônio representa uma classe específica. E por fim, a camada escondida, que recebe essa nomenclatura por não ser a camada de entrada e não ser a camada de saída é representada por M neurônios ($z_0, \dots, z_M$). Os neurônios x_0 e z_0 , correspondem as *bias* da camada escondida e de saída, respectivamente.

A MLP possui a arquitetura *feedforward*, ou seja, as conexões entre os nós não formam círculos. O treinamento de uma MLP possui 2 etapas: *forward*, onde determina a classe da entrada e a etapa do *backward* para atualizar os pesos de acordo com o erro obtido, onde o cálculo do erro só é viável quando é informado a classe desejada da amostra.

Na fase *forward*, os dados da camada de entrada são propagados para as camadas seguintes, até a camada de saída. Utilizando a sigmoide como um exemplo de função de ativação o valor obtido em y_k , dado um conjunto de entrada X , e um conjunto de pesos W , pode ser definido por:

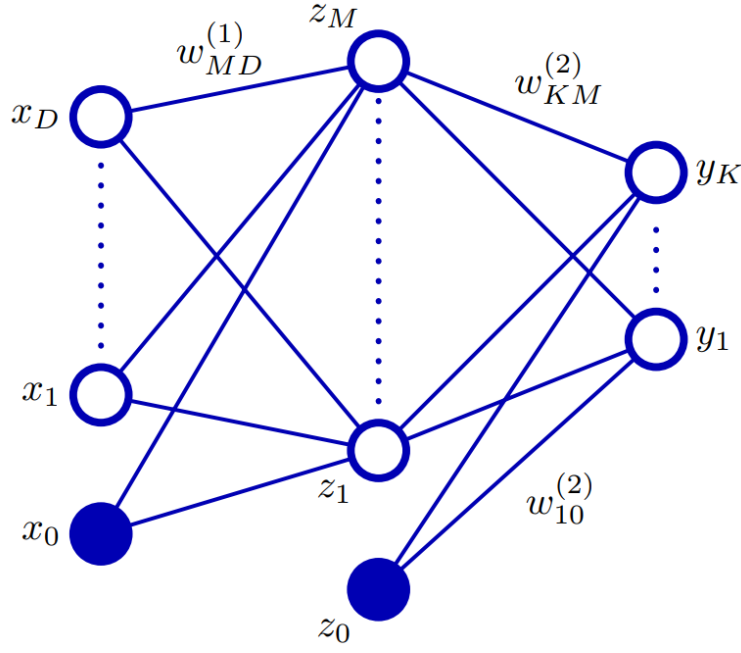


Figura 11 – Ilustração da rede MLP para classificação. Fonte: (BISHOP; NASRABADI, 2006).

$$y_k(X, W) = \sigma \left(\sum_{j=0}^M w_{kj}^{(2)} h \left(\sum_{i=0}^D w_{ji}^{(1)} x_i \right) \right) \quad (3.12)$$

dessa forma, na fase *forward* dado um conjunto de dados de entrada e de pesos uma saída é obtida (BISHOP; NASRABADI, 2006).

Na fase *backward*, para descobrir o erro dos pesos existem dois cenários: quando os pesos são relacionados a camada de saída, e quando os pesos são referentes as camadas escondidas. No caso dos pesos da camada de saída, basta calcular a diferença entre valor desejado e obtido, no caso dos neurônios das camadas escondidas, o erro precisa ser propagado para trás, visto que nas camadas escondidas não existe um valor esperado. Na Figura 12, é apresentado um exemplo dos dois cenários mencionados anteriormente, a seta azul representa o sentido que a fase *forward* aplica e a seta vermelha representa o sentido em que a fase *backward* utiliza, onde δ representa o erro do neurônio.

Ainda na Figura 12 o error dos neurônios $\delta_1, \dots, \delta_k$ são obtidos da seguinte maneira:

$$\delta_k = y_k - t_k, \quad (3.13)$$

onde, y_k é o valor obtido na fase *forward* e t_k é o valor esperado/desejado. Para determinar o erro de um neurônio escondido δ_j , o erro obtido em δ_k é propagado para a camada escondida seguindo a seguinte definição:

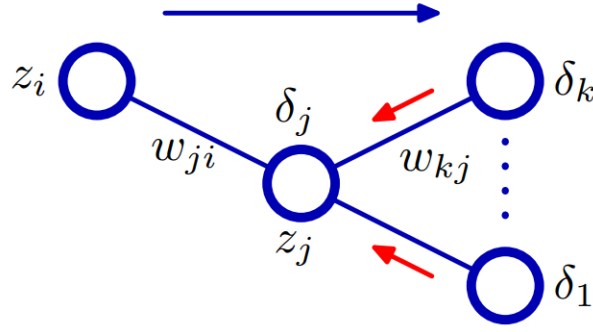


Figura 12 – Ilustração dos 2 tipos de erros que devem ser calculados na MLP. Fonte: (BISHOP; NASRABADI, 2006).

$$\delta_j = h'(a_j) \sum_k w_{kj} \delta_k, \quad (3.14)$$

onde a_j é referente a aplicação da função de ativação. Após a descoberta do erro da camada escondida e de saída, o algoritmo aplica a regra da cadeia para determinar a influência do erro (E) no peso: $\frac{\nabla E}{\nabla w_{ji}}$, e após aplicação da regra da cadeia chega-se a seguinte definição:

$$\frac{\nabla E}{\nabla w_{ji}} = \delta_j z_i \quad (3.15)$$

Uma abordagem simples de utilizar a informação do gradiente para atualização dos pesos, é definida abaixo:

$$w^{(t+1)} = w^{(t)} - \eta \nabla E(w^{(t)}), \quad (3.16)$$

onde, $w^{(t)}$ é o peso atual, $\nabla E(w^{(t)})$ é a derivada do peso em relação ao erro, e η é a taxa de aprendizagem (BISHOP; NASRABADI, 2006). A taxa de aprendizagem ou *learning rate*, determina a velocidade da convergência, quanto mais baixo mais lenta será a convergência. Sendo assim os principais parâmetros de uma MLP são: quantidade de camadas escondidas, quantidade de neurônios em cada uma das camadas escondidas, taxa de aprendizado e quantidade de iterações no treinamento.

Uma MLP com uma camada oculta, com os parâmetros corretos, é capaz de resolver problemas que estejam dispostos em regiões convexas, e uma MLP com duas camadas ocultas é capaz de resolver problemas em qualquer tipo de regiões geométricas (CYBENKO, 1989). No entanto a MLP apresenta alguns desafios, como determinar uma quantidade alta de parâmetros, tempo de convergência, e custo computacional.

3.4 Máquina de Vetores de Suporte

Uma Máquina de Vetores de Suporte ou *Support Vector Machines* (SVM), é um algoritmo de aprendizagem de máquina baseado na Teoria de Aprendizado Estatístico (TAE) (VAPNIK, 1999), que pode resolver problemas de classificação e regressão (HAYKIN, 2001). Proposta por Cortes e Vapnik (1995), uma SVM para classificação, busca encontrar o melhor hiperplano em um espaço n -dimensional, onde n é a quantidade de atributos, que separe as amostras em duas classes.

Na Figura 13a nota-se que várias linhas, hiperplanos de 2 dimensões, poderiam ser escolhidas e ainda assim os padrões seriam classificados corretamente. Para determinar o melhor hiperplano o SVM busca maximizar a margem. A margem é obtida como a distância dos padrões mais próximos até a linha de decisão. Os padrões que formam a margem, são chamados de vetores de suporte. Na Figura 13b é apresentado o hiperplano que maximiza a margem, onde a margem é representada pela distância entre as linhas tracejadas. O processo de maximizar a margem é necessário, visto que quanto maior a margem, menor o erro de generalização (HASTIE et al., 2009). Dessa forma, amostras acima do hiperplano serão classificadas como da classe positiva ($w^T x + b = 1$) e amostras abaixo do hiperplano serão classificadas como da classe negativa ($w^T x + b = -1$), na Figura 13b são representadas pela cor azul e vermelha respectivamente.

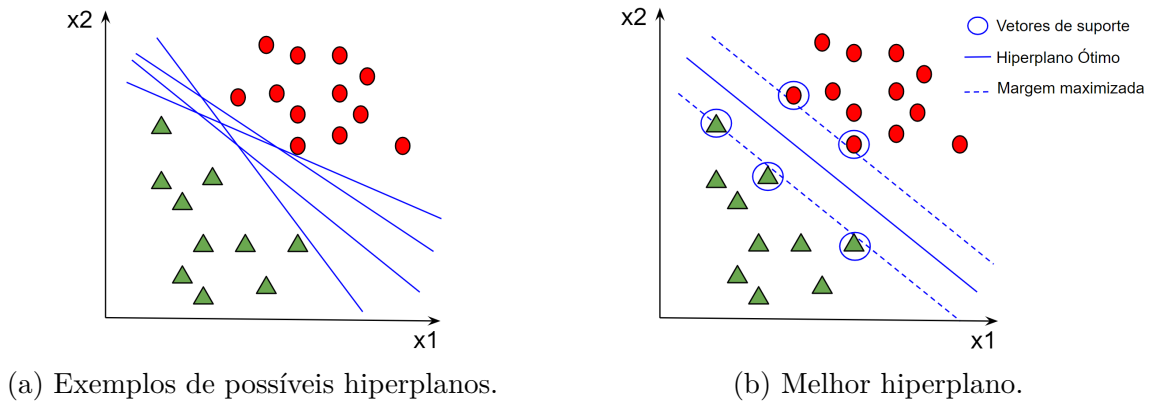


Figura 13 – Problema linearmente separável. Fonte: Elaborado pelo autor.

Para problemas não linearmente separáveis o SVM utiliza duas bordagens: aumenta a dimensionalidade dos dados, e considera tolerar alguns erros de classificação. Para aumentar a dimensionalidade o SVM aplica uma função *kernel*, nos dados. Na Figura 14 é apresentado um problema não linearmente separável, onde é aplicado a função *kernel*: $z = x^2 + y^2$, tornando possível encontrar um hiperplano que separe corretamente todos os padrões. Existem várias funções *kernel*, algumas mais conhecidas são:

- Polinomial: $(\langle x_i \cdot x_j \rangle + 1)^p$;
- Função Base Radial (Radial basis function, RBF): $\exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$;
- Sigmoidal: $\tanh(\beta_0 \langle x_i \cdot x_j \rangle + \beta_1)$

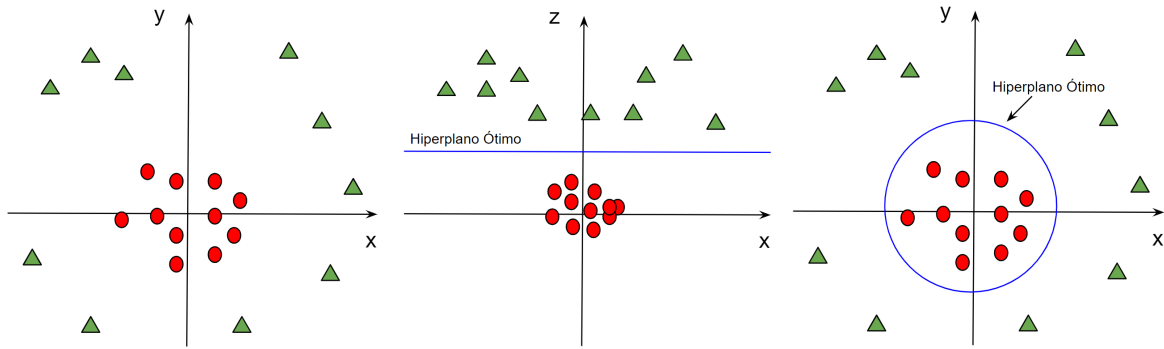


Figura 14 – Problema não linearmente separável, aplicando *kernel*. Fonte: Elaborado pelo autor.

O grau de tolerância, parâmetro chamado de C , com amostras que são classificadas erradas é um parâmetro importante do SVM. Quanto maior o C maior será a penalidade por classificar errado, logo mais amostras de treinamento serão classificadas corretamente, e com isso maior o risco de sobreajuste (*overfitting*) no conjunto de treinamento, dessa forma é indicado valores baixos de C .

Outro parâmetro também muito importante, é o γ , que define a influência que um padrão terá no treinamento. Se γ assumir um valor alto, a definição do hiperplano dependerá apenas dos pontos que estão próximos da linha, ignorando os padrões que estão distantes do limiar de decisão. Em contrapartida um valor baixo de γ resulta em uma curva mais linear, visto que os padrões mais distantes ganham um peso considerável. Para simplificar a escolha do γ , existem algumas abordagens conhecidas, como por exemplo, a *scale* que é definida por $\frac{1}{N \cdot \text{Var}(X)}$, onde N é a quantidade de atributos, e $\text{Var}(X) = E(x^2) - E(x)^2$.

O SVM foi originalmente desenvolvido para classificação binária, no entanto existem algumas técnicas para lidar com problemas de múltiplas classes. A primeira abordagem *one-versus-all* proposta por Scholkopf et al. (1997), cria N classificadores SVM, onde N é o número de classes, em que cada classificador utiliza uma classe como positiva e o restante como negativa. A segunda abordagem, proposta por Knerr, Personnaz e Dreyfus (1990) é chamada de *one-versus-one* e consiste em construir $\frac{N(N-1)}{2}$ classificadores, realizando todas as combinações de pares das N classes.

Existem diversas aplicações para as SVM's, principalmente no contexto de AVC (FARID et al., 2013; LAU; TONG; ZHU, 2009; CHEN; CHEN, 2011; FORKERT et al., 2015). Os resultados obtidos com as SVM's são consistentes e muitas vezes superiores aos obtidos por outros algoritmos de aprendizado, como as ANN's (STEINWART, 2002).

3.5 Floresta de Caminhos Ótimos

O classificador Floresta de Caminhos Ótimos ou *Optimum Path Forest* (OPF), é um algoritmo de aprendizado de máquina, que consegue resolver problemas de classificação e

clusterização. O OPF é baseado na Teoria dos Grafos, e para resolver a tarefa de classificação o rótulo é determinado analisando o caminho de menor custo. No grafo formado pelo OPF, cada amostra é representada por um nó/vértice, e cada aresta é ponderada por uma função de distância entre os vértices da aresta, como por exemplo a distância euclidiana (PAPA; FALCAO; SUZUKI, 2009).

A primeira etapa na fase de treinamento é determinar qual será a relação de adjacência entre as amostras, existem duas abordagens para isso: baseada nos k-vizinhos mais próximos ou considerar que todas as amostras são conectadas entre si. A segunda abordagem, remete a criação de um grafo de conexão completa, e essa foi a abordagem utilizada originalmente para esse classificador. Após determinar a relação de adjacência as conexões entre os nós devem ser criadas e ponderadas de acordo com a distância entre eles (PAPA et al., 2012). Na Figura 15a é apresentado um exemplo utilizando a relação de adjacência completa e ponderada pela distância euclidiana.

Agora com o grafo completo de todas as amostras de treinamento ponderadas pela distância, a segunda fase inicia-se, no qual os protótipos são definidos. Protótipo nesse contexto refere-se a uma amostra que seja exemplo da classe, cada classe deve ter um ou mais protótipos (PAPA et al., 2012). A heurística utilizada para determinar os protótipos foi escolher amostras que estão próximas do limiar de decisão, similar a lógica dos vetores de suporte apresentados na Sessão 3.4 do SVM. Para encontrar as amostras mais próximas de classes distintas, o algoritmo Árvore de Custo Mínimo ou *Minimum Spanning Tree* (MST) é aplicado. O MST retorna as arestas com menor custo possível que conecta todos os vértices do grafo original, e como o custo nesse cenário é a distância, isso significa que os vértices dessas arestas com menor custo estão próximos. Após aplicar o MST, as arestas que conectam vértices de classes diferentes, são escolhidas e seus vértices serão considerados os protótipos. Nas Figuras 15b é apresentado a aplicação do MST no grafo exemplo da Figura 15a, os vértices destacados com a linha azul, representam as amostras mais próximas de classes de diferentes. Na Figura 15c os protótipos escolhidos são destacados com a linha tracejada.

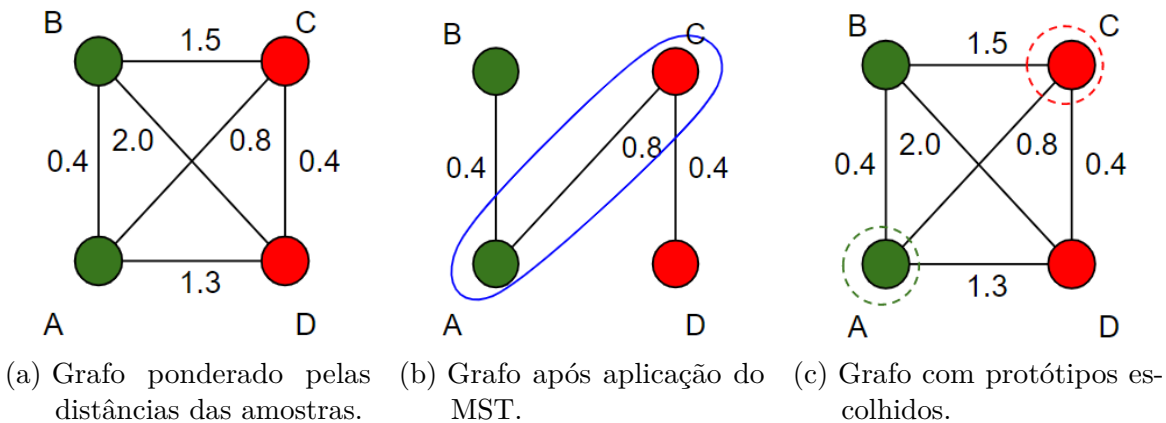


Figura 15 – Identificação dos protótipos. Fonte: Elaborado pelo autor.

Agora com os protótipos definidos a terceira fase do processo de treinamento é iniciada, onde o objetivo é conquistar amostrar. Quando uma amostra A conquista uma amostra B , isso significa que a classe da amostra B será igual a classe da amostra A . Uma amostra A conquistará uma amostra B quando a amostra A oferecer um caminho de menor custo. A função custo de um caminho, deve ser uma função monotonicamente crescente, no caso do classificador original do OPF a função *max* foi utilizada, ou seja, o custo de um caminho será a aresta de maior valor. O processo de conquistar ocorre por meio de uma fila de prioridades, onde a prioridade é o menor custo de um caminho que o vértice tem no momento. Na iteração inicial, os protótipos possuem custo 0 e os outros vértices possuem custo infinito, a cada iteração um vértice de maior prioridade, ou seja, o vértice de menor custo é extraído da fila. O vértice retirado da fila tentará conquistar todos os outros vértices, mas só conseguirá conquistar aqueles em que ele apresentar um custo menor, quando um vértice for conquistado o seu valor de custo deve ser atualizado. O processo acaba quando a fila fica vazia, e como resultado existem N árvores, onde N é a quantidade de protótipos, no qual cada árvore possui como raiz um protótipo. A coleção dessas árvores de caminhos ótimos forma a floresta de caminhos ótimos (PAPA; FALCAO; SUZUKI, 2009). Na Figura 16 é ilustrado a floresta de caminhos ótimos obtida a partir do grafo da Figura 15a.

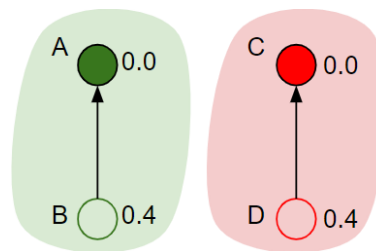


Figura 16 – Floresta de caminhos ótimos. Fonte: Elaborado pelo autor.

O processo de teste de uma amostra, ou seja, descobrir qual classe a amostra pertence, consiste em adicionar a amostra na floresta de caminhos ótimos, conecta-lá a todos os vértices ponderando pela distância, e verificar qual amostra possui menor custo, com isso a classe dessa amostra será a classe retornada. Na Figura 17 é apresentado um exemplo de como ocorre o processo para determinar a classe de uma amostra desconhecida. Na Figura 17a a nova amostra, representada pela cor azul é adicionada a floresta de caminhos ótimos. Na Figura 17b a amostra é determinada para a classe verde, visto que possui o caminho com menor custo. Ao final desse processo a amostra desconhecida é removida da floresta de caminhos ótimos.

Existem diversas aplicações do algoritmo OPF, principalmente no contexto de análise de imagens (ROCHA; CAPPABIANCO; FALCÃO, 2009; CAPPABIANCO et al., 2010; STEHLING; NASCIMENTO; FALCÃO, 2002; CAPPABIANCO et al., 2012; PAPA et al., 2009; PAPA; ROCHA, 2011). O OPF possui a vantagem de pouca parametrização,



(a) Amostra desconhecida é anexada a floresta. (b) Amostra escolhe com a classe determinada.

Figura 17 – Processo para determinar a classe de uma amostra desconhecida. Fonte: Elaborado pelo autor.



Figura 18 – Separação do espaço de decisão. Fonte: Bishop e Nasrabadi (2006).

basicamente precisa definir qual o método que será utilizado para calcular a distância. Quando comparado com outros classificadores como SVM, MLP e KNN, o OPF em algumas bases de dados obteve melhor performance (QUINTA et al., 2012).

3.6 Métodos baseados em Árvores de Decisão

A árvore de decisão ou *Decision Tree*, é uma árvore binária que representa um processo de decisão, similar ao percurso de uma árvore, no qual cada nó possui uma condição (BISHOP; NASRABADI, 2006). As árvores de decisão são capazes de resolver problemas de classificação e regressão, existem vários algoritmos conhecidos na literatura relacionados a essas árvores, os principais são: *classification and regression trees* (CART) (BREIMAN et al., 1984), ID3 (QUINLAN, 1986) e C4.5 (QUINLAN, 2014).

As árvores de decisão para classificação particionam o espaço em cubos para separar as amostras entre as classes, montando uma espécie de fluxograma onde cada nó possui uma condição referente aos atributos de entrada. Na Figura 18 é apresentado um exemplo contendo dois atributos: x_1 e x_2 .

Em geral os algoritmos constroem a árvore de cima para baixo, e em cada nível utilizam uma abordagem para definir qual atributo e ponto de corte deve ser utilizado. Existem várias medidas/conceitos utilizados para determinar a atributo que será escolhida, as principais são: índice Gini e Ganho de informação (RAILEANU; STOFFEL, 2004). A

intuição por trás de definir o atributo, é priorizar os atributos que mais ajudam a separar as amostras entre as classes.

O índice Gini ou impureza Gini, é utilizado no algoritmo CART, e representa a probabilidade de um atributo específico ser classificado incorretamente quando selecionado aleatoriamente. O índice de Gini é definido da seguinte maneira:

$$Gini\ Index = 1 - \sum_i^N (P_i)^2, \quad (3.17)$$

onde N é a quantidade de classe e P_i é a fração de itens rotulados com classe i no conjunto de dados. O índice Gini varia de 0 a 1, no qual 0 significa que todos os elementos pertencem apenas a 1 classe específica, ou só existe 1 classe; 1 indica a distribuição aleatória de elementos em várias classes; e 0,5 demonstra uma distribuição igualitária dos elementos em algumas classes. Para a construir a árvore quanto mais próximo de 0 melhor.

O ganho de informação, utilizado pelos algoritmos ID3 e C4.5, determina o quanto um atributo pode trazer de informação, quantificando o tamanho da incerteza/desordem do atributo por meio da entropia. A entropia é definido da seguinte maneira:

$$Entropy = - \sum_i^N P_i \log_2 P_i, \quad (3.18)$$

onde o ganho de informação é calculado como a diferença entre a entropia do nó pai e o somatório ponderado das entropias dos nós filhos. Para a construção da árvore, os atributos com maior ganho informação são escolhidos a cada iteração.

A árvore de decisão é uma abordagem simples para resolver o problema de classificação, garante interpretabilidade dos resultados, e possui baixo custo computacional quando comparado a abordagens mais complexas. No entanto, em geral a performance é inferior aos classificadores mais complexos, apresentam variabilidade alta e são muito sensíveis ao ruídos das amostras, o que torna suscetível ao *overfitting* (ALI et al., 2012).

Como proposta para contornar esse problema da tendência que as árvores de decisão tem ao *overfitting*, os métodos *Ensemble* ganharam popularidade. O método *Ensemble* utiliza vários modelos base visando obter melhor resultado que o modelo base teria sozinho (DIETTERICH, 2000). Existem vários classificadores baseados em árvores de decisão que utilizam aprendizagem *Ensemble*, como por exemplo: Floresta Aleatória e Árvores Extremamente Aleatórias. Para esses algoritmos a quantidade de estimadores é o principal parâmetro.

3.6.1 Floresta Aleatória

Uma Floresta Aleatória ou *Random Forest* (RF), é um algoritmo de aprendizado de máquina proposto por Breiman (2001), que utiliza o método *Ensemble* e é capaz de resolver problemas de classificação e regressão. O processo de treinamento de uma RF consiste em treinar seus N estimadores. Para treinar cada estimador uma base de dados é construída. Cada base de dados é construída utilizando a técnica *bagging* ou *bootstrap aggregating*, no qual seleciona k amostras com j atributos da base de dados original, de forma aleatória e com reposição. Com as N subamostras os N estimadores podem ser treinados, dessa forma os estimadores são independentes. Para determinar a classe de uma amostra o RF combina o resultado de todos os N estimadores e atribui à classe de maior frequência obtida na saída dos estimadores. Na Figura 19 é ilustrado um exemplo de como a RF é estruturada. Assim como na árvore de decisão, a escolha de qual atributo será utilizado em cada nó, e a o ponto de corte, é definido de maneira ótima.

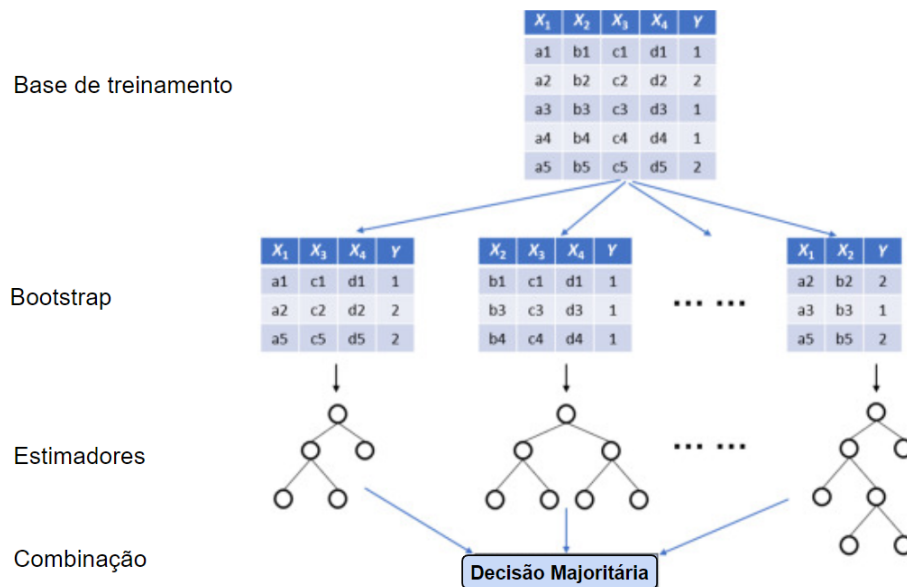


Figura 19 – Estrutura de uma *Random Forest*. Fonte: Adaptado de Williams et al. (2020).

Existem diversas aplicações do algoritmo RF na literatura, especificamente no contexto de AVC (FERNANDEZ-LOZANO et al., 2021; SUBUDHI; DASH; SABUT, 2020; MITRA et al., 2014; MCKINLEY et al., 2017). A RF possui a vantagem de ter poucos parâmetros, basicamente precisa definir a quantidade de estimadores. E tem a possibilidade de paralelizar o processamento, visto que cada estimador é independente o treinamento de cada um pode ser feito em paralelo.

3.6.2 Árvores Extremamente Aleatórias

As Árvores Extremamente Aleatórias ou *Extremely Randomized Trees* (Extra Tree), é um algoritmo de aprendizado de máquina proposto por Geurts, Ernst e Wehenkel (2006),

que utiliza o método *Ensemble* e é capaz de resolver problemas de classificação e regressão. O Extra Tree é muito similar as RF, a diferença entre eles é a formação das subamostras e a forma de construir os estimadores. O Extra Tree utiliza todas as amostras da base de dados para treinamento dos estimadores, no entanto cada estimador receberá uma base com j atributos aleatórios. No processo de construção da árvore, o ponto de corte em cada nó é escolhido aleatoriamente, enquanto na RF é definido de maneira ótima, essa diferença faz com que o Extra Tree seja mais rápido que o RF.

Existem diversas aplicações do algoritmo Extra Tree na literatura, especificamente no contexto de AVC (MAIER et al., 2015; FANG; XU; LIU, 2020; HASAN et al., 2021). O Extra Tree possui as mesmas vantagens da RF: poucos parâmetros e possibilidade de paralelização do treinamento, como também tem o treinamento mais rápido que a RF.

4 Metodologia e Experimentos

Neste Capítulo será apresentado informações sobre a base de dados utilizada nessa pesquisa, uma descrição detalhada sobre o novo método proposto para extração de características cerebrais em imagens TC, e possíveis configurações desse novo método. O novo extrator será aplicado para extrair as características de cada uma das imagens da base de dados, gerando uma nova base de dados com apenas as características extraídas, onde serão utilizadas para classificação.

O processo de classificação do AVC possui as seguintes macro-etapas: seleção da região de interesse, extração de características com o método MGABTD e, por fim, a classificação em: normal, AVCh ou AVCi. Na Figura 20 é apresentado a metodologia dessa pesquisa. Após a etapa de pré-processamento que seleção da região de interesse, a TC é redimensionada e em seguida separada em 2 partes: hemisfério esquerdo e hemisfério direito. O extrator MGABTD é aplicado em cada um dos hemisférios, em seguida a informação é unida e enviada para classificação.

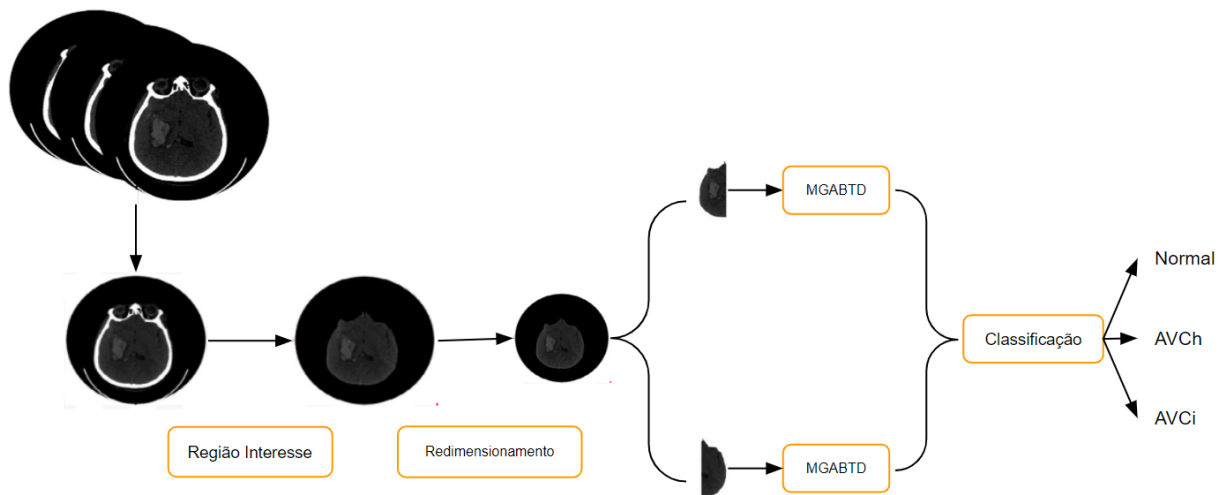


Figura 20 – Para cada imagem da base de dados, a região interna do crânio é extraída, redimensionada, dividida em dois hemisférios e para cada hemisfério o extrator MGABTD é aplicado. Por fim as características dos dois hemisférios são unidas gerando um vetor de características que é enviado para um classificador de padrões, determinando a classe da imagem de entrada. Fonte: Elaborado pelo autor.

A separação entre hemisférios fará com que a região do AVC se destaque mais considerando o escopo do hemisfério corrente, do que considerando o cérebro inteiro, visto que o AVC geralmente ocorre em um dos hemisférios (VASCONCELOS et al., 2020). Nas seções seguintes são apresentadas em detalhes as macro-etapas.

4.1 Base de dados

As imagens de TC foram obtidas com o apoio da Clínica Trajano Almeida - Diagnóstico por Imagem e do Hospital do Coração por meio de um tomógrafo e utilizada em estudos anteriores (DOURADO et al., 2019; DOURADO et al., 2020; SARMENTO et al., 2020; VASCONCELOS et al., 2020). O tomógrafo é calibrado em até 3 meses antes da realização dos exames, onde a densidade do ar é ajustada para -1000 HU e a água para 0 HU. O tomógrafo também é submetido a um corte tomográfico com o *phantom* de água (FILHO, 2013).

O modelo tomográfico utilizado para adquirir as imagens foi o *HiSpeed CT/e Dual*¹ do fabricante General Electric. No processo de aquisição de imagens, os cortes tomográficos foram realizados na base do plano axial, sob as seguintes condições: espessura do corte 0,7 mm, campo de visão 230 mm, tensão elétrica no tubo 120 Kv, corrente elétrica no tubo 80 mA, dimensões 512x512 *pixels* e janelamento do voxel 0,585x0, 585x1, 5 mm com uma quantificação de 16 bits, nível de janelamento de 40 HU e amplitude da janela igual a 80 HU (SARMENTO et al., 2020).

A base de dados é composta por imagens obtidas em 25 exames de TC do cérebro de pacientes sem AVC, com AVCi e com AVCh. Por questões éticas, a base de dados não fornece nenhuma identificação dos pacientes. A base de dados consiste em 473 imagens 2D, divididas entre 174 imagens de pacientes saudáveis (normais), 157 imagens com AVCi e 142 imagens com AVCh. As imagens não possuem padrões ou iluminação regular e as características estruturais são diferentes. As imagens são armazenadas utilizando o formato DICOM (Digital Imaging and COmmunications in Medicine). Na Figura 21 é apresentado um exemplo de imagem de cada uma das classes.

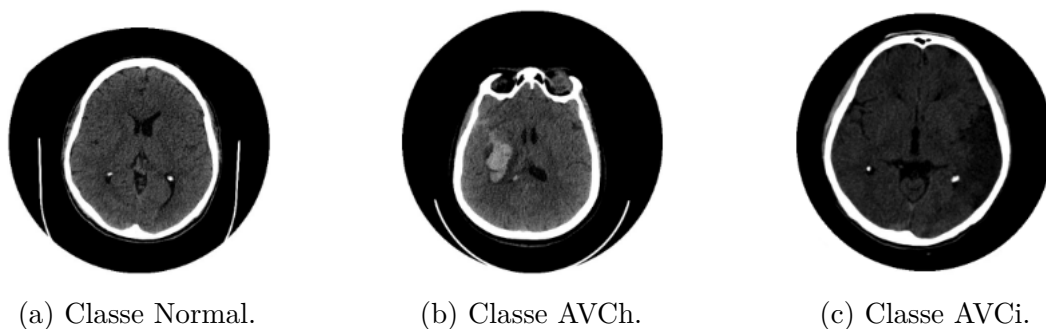


Figura 21 – Amostra de TC do cérebro de cada uma das classes. Na Figura 21b o AVCh é representado pela mancha branca, enquanto na Figura 21c o AVCi é representado pela mancha mais escura próxima ao crânio. Fonte: Base de dados.

DICOM é um padrão internacional criado pelo ACR (American College of Radiology) e pelo NEMA (National Electrical Manufacturers Association) que representa um conjunto de normas para tratamento, armazenamento e transmissão de imagens médicas no formato eletrônico para padronizar o formato das imagens de tomografias, ressonâncias

magnéticas, dentre outros. O padrão permite que equipamentos de imagens, computadores de marcas diferentes consigam lê as informações da mesma forma (PIANYKH, 2009). A biblioteca *pydicom*, compatível com a linguagem python, foi utilizada para a leitura das imagens DICOM. Após a leitura, cada pixel das imagens foram convertidos de 16 bits, para 8 bits. Isso ocorre para facilitar e acelerar a manipulação dessas imagens, visto que, um pixel com 16 bits possui valores variando de -32.768 a 32.768, e com 8 bits os valores variam de 0 a 255 (LAROBINA; MURINO, 2014). Esses valores se correlacionam com a escala de *Hounsfield* (HU), que mede quantitativamente a radiodensidade (DENOTTER; SCHUBERT, 2019).

4.2 Pré-processamento

A parte interna do crânio é a região mais importante da TC para o diagnóstico de AVC. Para segmentar essa região de interesse ou ROI (*Region of Interest*) foi utilizado o método apresentado em Sarmento et al. (2020). Na Figura 22 é apresentado visualmente as etapas da segmentação do ROI. Após a seleção da região de interesse, a imagem tem o seu tamanho reduzido para um determinado percentual previamente escolhido, essa redução é necessária para reduzir tempo de processamento.

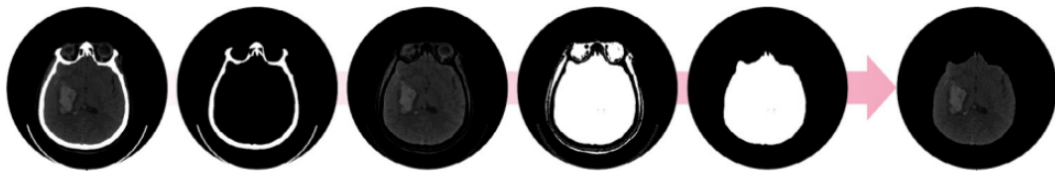


Figura 22 – Etapas da segmentação da região de interesse. Fonte: Sarmento et al. (2020).

A segmentação possui basicamente 4 etapas: carregar a imagem; aplicar uma máscara que evidência apenas os pixels com valores HU entre 0 e 120, referente a região interna do crânio (YOUSEM, 2011); com pixels do crânio evidenciados, uma subtração é feita da imagem original, assim removendo o crânio; em seguida a maior componente conexa é selecionada; e por fim, com apenas o crânio evidenciado a interseção com a imagem original é feita. Ao final do processo a imagem resultante não possui o crânio ou qualquer ruído que possa interferir na análise posterior do método proposto.

4.3 Análise de Mistura Gaussianas da Densidade do Tecido Cerebral

O método Análise de Mistura Gaussianas da Densidade do Tecido Cerebral ou *Mixture Gaussian Analysis of Brain Tissue Density* (MGABTD) consiste em criar uma GMM para classificação considerando as seguintes classes baseadas nas definições de (YOUSEM, 2011): Classe 1: Fluido Cerebrospinal (-99 a 5 HU); Classe 2: AVCi (6 a 22

HU); Classe 3: Massa Branca (23 a 34 HU); Classe 4: Massa Cinzenta (32 a 41 HU); Classe 5: AVCh (50 a 81 HU), e Classe 6: Calcificação (130 a 250 HU). Desconsiderando a classe Crânio, visto que foi removida na etapa da Seleção da região de interesse e adicionando a Classe 0 que representa o plano de fundo, ao todo são 7 classes.

Com a GMM treinada a extração de característica é aplicada em cada hemisfério, sendo assim a saída dessa etapa é uma matriz 2D, com as classes preditas de cada pixel do hemisfério esquerdo e outra matriz com as classes do hemisfério direito. A partir dessas matrizes com as classes foram propostas 3 variações de como transforma-las em atributos, são elas: $MGABTD_{pixel}$, $MGABTD_{percent}$ e $MGABTD_{neighborhood}$.

4.3.1 Analysis of Brain Tissue Densities (ABTD)

O ABTD é um método de extração de características proposto por Filho et al. (2017) no qual o vetor de característica é representado pela porcentagem de pixels em cada classe, considerando os intervalos HU apresentados na literatura (FILHO; CORTEZ; HOLANDA, 2011). Como mencionado na Seção 1.2, utilizar os intervalos propostos na literatura diretamente pode ser problemático, pelos casos de sobreposição e limiares diferentes de uma mesma região, sendo a principal limitação dessa técnica. O MGABTD é uma proposta de melhoria ao método ABTD, visto que em vez de utilizar os intervalos da literatura, utiliza a GMM como estimador de probabilidades.

4.3.2 $MGABTD_{pixel}$

Nessa abordagem, as matrizes com as classes do hemisfério esquerdo e direito são redimensionadas para 1D, tornando-se o vetor de classes do hemisfério esquerdo e o vetor de classes do hemisfério direito. Para combinar a informação dos dois vetores e construir um vetor unificado com os atributos de uma imagem, foi idealizado duas abordagens: concatenar os dois vetores ou somar os elementos dos dois vetores.

A 1ª alternativa, implica que o vetor final de atributos, seria do tamanho do vetor do hemisfério esquerdo mais o tamanho do vetor do hemisfério direito, mesmo considerando que a imagem foi reduzida, o vetor ainda ficaria muito grande, por exemplo, se for considerado uma redução de 30% do tamanho da imagem original, cada vetor de cada hemisfério teria 11.858 posições (pixels), logo o vetor final de atributos de uma imagem teria 23.716 posições. A 2ª alternativa, implica que o vetor final de atributos teria o mesmo tamanho dos vetores de cada hemisfério, ou seja, seguindo o exemplo anterior o vetor final de atributos teria 11.858 posições. Sendo assim, foram realizados alguns testes empíricos, e a 2ª alternativa se mostrou tão eficiente quanto a 1ª, logo a 2ª alternativa foi escolhida visto que possui menor tamanho. Na Figura 23 é apresentado visualmente como a 2ª alternativa será utilizada.

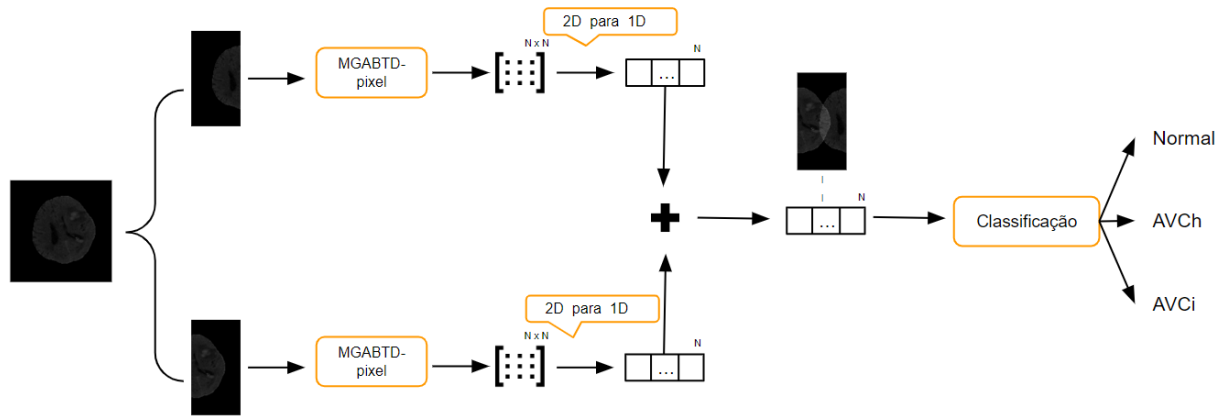


Figura 23 – Metodologia após a imagem de entrada passar pela segmentação ROI e redimensionamento, utilizando o extrator $MGABTD_{pixel}$. Fonte: Elaborado pelo autor.

Nessa abordagem, mesmo com a redução do tamanho da imagem de entrada, o ponto de atenção continua sendo o tamanho do vetor de atributos, no exemplo anterior, mesmo com a redução de 30%, o número de atributos ainda fica consideravelmente grande, e isso impacta diretamente no tempo de processamento dos classificadores.

4.3.3 $MGABTD_{percent}$

Nessa abordagem, similar a metodologia dos extratores ABTD e PABTD, é calculado o percentual de pixels de cada classe em cada matriz. No ABTD as classes dos pixels são determinadas pelas próprios limiares HU da literatura, e no método PABTD utiliza a Janela de Parzen para determinar a classe, enquanto o $MGABTD_{percent}$ utiliza uma GMM para obter a classe do pixel.

Após o cálculo do percentual de pixels de cada classe, desconsiderando a classe de plano de fundo, cada matriz resultará em um vetor com de 6 posições, onde cada posição i possui o percentual de pixels da classe i , em um determinado hemisfério. Por fim, os vetor são concatenados, gerando um vetor de atributos final contendo 12 posições fixas, onde as 6 primeiras são percentuais de cada classe referente ao hemisfério esquerdo e as 6 últimas referente ao hemisfério direito. Na Figura 24 é apresentado visualmente como ocorre o processo de extração com o $MGABTD_{percent}$.

4.3.4 $MGABTD_{neighborhood}$

Baseado no método de Vasconcelos et al. (2020), que analisa a vizinhança dos pixels, no $MGABTD_{neighborhood}$ após a obtenção das matrizes com as classes predita pela GMM, uma análise da vizinhança é realizada. Essa análise determina a nova classe de um pixel analisando a classe mais frequente da sua vizinhança-8. Na Figura 25 é apresentado

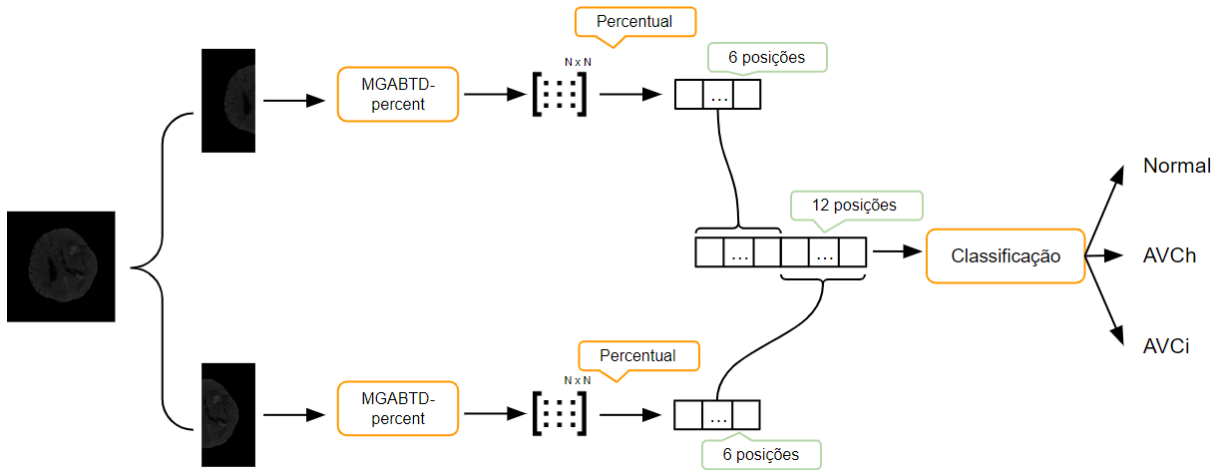
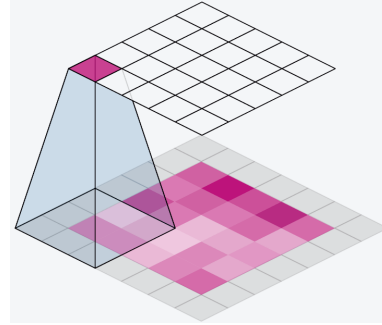
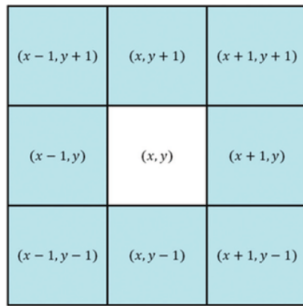


Figura 24 – Metodologia após a imagem de entrada passar pela segmentação ROI e redimensionamento, utilizando o extrator *MGABTD_{percent}*. Fonte: Elaborado pelo autor.

a vizinhança-8 de um pixel na posição (x, y) .



(a) Vizinhança-8 de um pixel (x, y) . Fonte: Yahya et al. (2018). (b) Exemplo janela deslizante. Fonte: Holbrook (2022).

Figura 25 – Redefinição da classe de cada pixel.

Com as novas classes de cada pixel, o mesmo processo que é feito no *MGABTD_{percent}* é aplicado aqui, ou seja, o percentual de pixels de cada classe é calculado para cada um dos hemisférios, gerando um vetor de 6 posições em cada hemisfério e por fim a concatenação dos vetores de cada hemisfério gera o vetor final de atributos com 12 posições. Na Figura 26 é apresentado visualmente como ocorre o processo de extração com o *MGABTD_{neighborhood}*. O principal ponto de atenção desse extrator é o processo de redefinição das classes analisando a vizinhança, essa janela deslizante tem um alto custo computacional, que pode impactar negativamente o tempo de extração.

Na Figura 26, nota-se que o processo a partir de cada hemisfério ocorre paralelamente, sem nenhuma dependência, dessa forma a implementação dessa etapa foi desenvolvida utilizando *Thread*, no qual uma *thread* foi utilizada para extração das características do hemisfério esquerdo e outra *thread* para o hemisfério direito, e só então

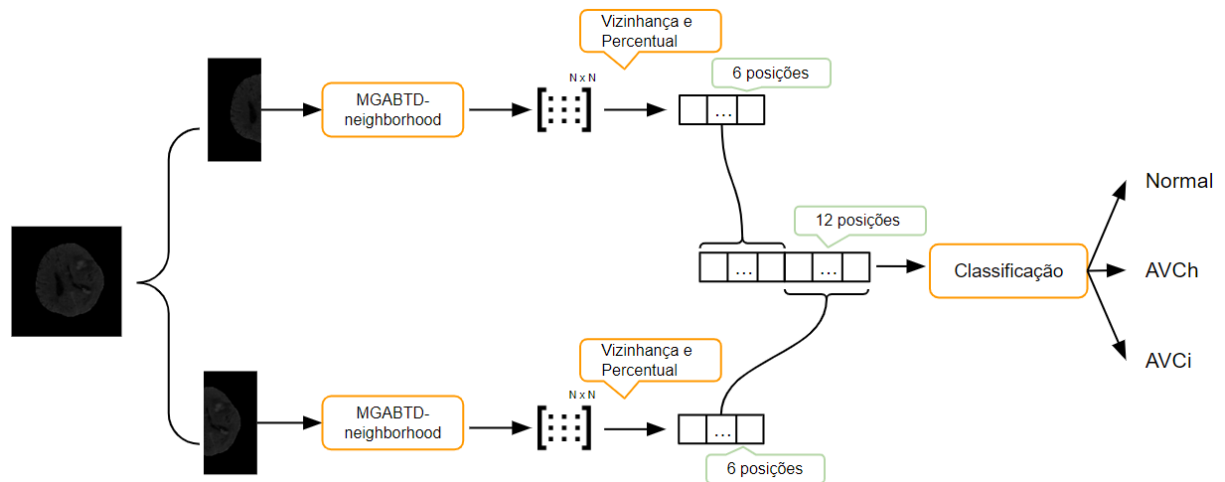


Figura 26 – Metodologia após a imagem de entrada passar pela segmentação ROI e redimensionamento, utilizando o extrator $MGABTD_{neighborhood}$. Fonte: Elaborado pelo autor.

quando ambas as *threads* concluem, a soma dos vetores ocorre. A *Thread* é uma forma de diminuir o tempo o tempo de processamento, visto que a etapa de análise da vizinhança é muito custosa.

4.4 Possíveis configurações MGABTD

Todas as 3 variações de extração do MGABTD apresentados, contém 3 principais parâmetros que precisam ser determinados previamente antes da extração: número de comportes por classe, número de iteração do EM-step e o percentual de redução do tamanho da imagem. Os dois primeiros parâmetros são necessários pois são relevantes para a definição da GMM, nesses parâmetros são aceitos números naturais. Para o parâmetro de números de componentes por classe, o número de componentes não deve exceder o número de amostra da classe. O parâmetro referente ao percentual de redução do tamanho da imagem, foi idealizado visando diminuir o tempo de extração dos extratores mantendo uma performance aceitável.

4.5 Significância Estatística

Para verificar a significância estatísticas dos resultados obtidos serão utilizados dois testes estatísticos: *5x2 cv paired t-test* e *Friedman*. O teste *5x2 cv paired t-test* será utilizado para verificar equivalência de performance entre os classificadores de um mesmo extrator. O teste de *Friedman* será utilizado para verificar equivalência de performance entre os extratores $MGABTD_{\{pixel, percent, neighborhood\}}$ e o método base ABTD.

4.5.1 5x2 cv paired t-test

Em Dietterich (1998) foram analisados 5 testes estatísticos com o objetivo de avaliar se um algoritmo A supera um algoritmo B em termos de performance, como resultado dos experimentos o *t-test* com validação cruzada se mostrou o mais poderoso, sendo a variação *5x2 cv paired t-test* mais poderosa ainda que o teste McNemar's. Sendo assim, para verificar se as diferenças de acurácia apresentada entre os classificadores são estatisticamente significativas, e não fruto do acaso estatístico, o teste de hipótese: *5x2 cross-validation paired t-test* foi aplicado. Saber quais classificadores possuem performance equivalente, é importante, visto que se são equivalentes o critério de escolha pode ser a velocidade seja no treinamento ou no teste, ou escolher o classificador que tem uma explicabilidade mais simples por exemplo.

O teste de hipótese *5x2 cv paired t-test*, considera a hipótese nula de que dois modelos possuem performance equivalente. A performance p é qualquer métrica que representa a qualidade do modelo, podendo ser por exemplo a acurácia. O teste consiste em analisar a performance de dois modelos (A e B) utilizando o método de validação cruzada com k -folds (DIETTERICH, 1998). O teste possui 5 iterações, em cada iteração, os dados de treino e teste são divididos usando $k = 2$, ou seja, 50% treino e 50% teste e a performance é calculada (p^1) para os dois modelos, em seguida os dados que eram de teste passam a ser de treino e os dados de treino passam a ser de teste (p^2):

$$p^{(1)} = p_A^{(1)} - p_B^{(1)} \quad (4.1)$$

$$p^{(2)} = p_A^{(2)} - p_B^{(2)} \quad (4.2)$$

Com isso a performance média ($\bar{p}$) da iteração e a variância (s^2) da diferença entre as performances dos modelos são calculadas:

$$\bar{p} = \frac{p^{(1)} + p^{(2)}}{2} \quad (4.3)$$

$$s^2 = \left(p^{(1)} - \bar{p}\right)^2 + \left(p^{(2)} - \bar{p}\right)^2 \quad (4.4)$$

Dessa forma, a estatística t é calculada sob a hipótese nula de que os modelos A e B têm desempenho igual:

$$t = \frac{p_1^{(1)}}{\sqrt{(1/5) \sum_{i=1}^5 s_i^2}} \quad (4.5)$$

Com a estatística t , o p -value pode ser calculado sob um nível de significância previamente escolhido, se o p - *value* for menor que α então a hipótese nula é rejeitada,

ou seja, os modelos não possuem performance equivalente, logo existe uma diferença significativa de performance entre os modelos A e B (DIETTERICH, 1998).

4.5.1.1 Friedman

Segundo Demšar (2006), um dos testes mais recomendados para comparação de múltiplas bases de dados, visto que cada extrator é uma base, seria o teste de *Friedman*. O teste de *Friedman* ranqueia os algoritmos do melhor para o pior para cada conjunto de dados considerando seu desempenho, e em seguida compara a média dos algoritmos ranqueados, sob a hipótese nula de que os algoritmos possuem performance equivalente.

Considerando um determinado α a ser definido previamente, se o $p - value$ for menor que α , então significa que existe uma diferença significativa de performance entre os extratores. No cenário deste trabalho a performance de um extrator é medida por meio de alguma métrica de desempenho obtida a partir da aplicação do extrator em conjunto com um classificador.

5 Resultados e Discussões

Este capítulo contém os resultados obtidos nos 3 métodos de extração propostos. Para determinar a eficiência e robustez dos extratores serão comparados aos resultados obtidos nos trabalhos relacionados mais recentes. Os testes foram realizados no ambiente Google Colaboratory, que permite escrever e executar códigos em python pelo navegador. A máquina disponibilizada pelo Colaboratory possui um processador Intel(R) Xeon(R) CPU @2.00GHz, 12GB de RAM, e uma *Graphics Processing Unit* (GPU) Tesla P100-PCIE-16GB. Os extratores foram implementados usando Python, versão 3.6.9.

A eficiência dos extratores foi mensurada por meio da performance dos classificadores em suas configurações ótimas, ou seja, se as características do extrator 1 obtém acurácia superior a acurácia obtida usando as características do extrator 2, logo o extrator 1 gera características mais relevantes para a classificação, e portanto é melhor que o extrator 2. Os classificadores utilizados foram: KNN, OPF, SVM, e MLP. Também foi utilizado classificadores baseados em árvore: Extra Tree e RF.

5.1 Configuração ótima do MGABTD

Os extratores MGABTD possuem parâmetros que devem ser definidos previamente, então para definir a configuração ótima desses parâmetros, também foi utilizado a performance da classificação para determinar se uma configuração de parâmetros é melhor ou não que outra.

Sendo assim uma busca em grade para cada extrator foi realizada. Os parâmetros da GMM testados foram: número de componentes por classe e número máximo de iterações no processo EM. Seguindo as lições aprendidas sobre a qualidade da classificação mesmo após o redimensionamento das imagens, discutidas em (SARMENTO et al., 2020), foram testadas as porcentagens de redução no tamanho das imagens. Os classificadores foram treinados com suas configurações padrão apresentadas na Tabela 1.

A performance de cada extrator-classificador da grade foi analisada considerando 6 métricas: acurácia (*acc*), precisão, *recall* e F1-Score. Essas métricas são calculadas baseadas nos conceitos de: Verdadeiros Positivos (VP), Falsos Positivos (FP), Verdadeiros Negativos (VN) e Falsos Negativos (FN). Se o problema for de classificação binária, o conceito verdadeiro/falso é referente a condição: *predito* = *real*, e o conceito de positivo/negativo são referentes as classes 1 e 0 respectivamente. No caso de problemas multiclasse, como o dessa pesquisa (Normal, AVCh, AVCi), os VP, FP, VN e FN são calculados para cada uma das classes, onde positivo se refere a classe analisada, e negativo qualquer outra classe.

Tabela 1 – Configuração padrão dos classificadores.

Classificador	Configuração Padrão
Extra Tree	Estimadores=100
KNN	K=5
OPF	Distância=log_squared_euclidean
RF	Estimadores=100
SVM	C=1
	Kernel=rbf
	Gamma=scale
MLP	Iterações=200
	Neurônios=100
	Ativação=relu

Seguindo esse raciocínio, os valores das métricas são contabilizados por classe, por exemplo, a precisão é calculada para a classe Normal, AVCh e AVCi, para obter um valor único da métrica, algumas estratégias podem ser usadas, a estratégia utilizada nessa pesquisa foi a média ponderada pelo tamanho da classe.

A acurácia representa o percentual de amostras que o modelo acertou de cada classe, sobre o total de amostras, essa métrica é influenciada pelo desbalanceamento das classes. A métrica precisão, é o percentual de amostras que o modelo projetou que seria de uma classe e realmente é da classe projetada. O *recall* é o total de amostras que o modelo projetou para uma classe sobre o total de amostras que eram verdadeiramente daquela classe. O F1-score é uma média harmônica da precisão e do *recall*. Cada configuração possível de cada extrator foi avaliada usando 3 técnicas diferentes: *Hold Out* (HO), Validação cruzada com k-folds (CV) e *Leave One Out* (LOO).

A técnica *Hold Out*, consiste em separar a base em duas, uma para treinamento e outra para teste seguindo uma proporção (DEVROYE; WAGNER, 1979). Nessa técnica, quanto maior o conjunto de teste, maior o viés do modelo, e se for um conjunto de dados pequeno, diminuir o tamanho do conjunto de treinamento pode distorcer o modelo, como em contrapartida também pode melhorar a variância do modelo. Existe também a possibilidade do conjunto de teste possuir amostras mais fáceis que o conjunto de treinamento e vice-versa, o que pode gerar uma interpretação equivocada dos resultados. Para amenizar essas questões, uma variação do método HO foi utilizada: *repeated hold-out*, em que o método HO é repetido n-vezes, essas repetições também podem ser chamadas de realizações. O método *repeated hold-out* foi utilizado com uma separação 70-30 ao longo de 200 realizações.

A técnica Validação Cruzada K-folds, consiste em dividir o conjunto de dados em k partes iguais, onde cada parte é chamada de dobra (*fold*), e ao longo de k iterações: cada iteração uma dobra é usada como conjunto de teste e o restante é usado como conjunto de treinamento (BURMAN, 1989). Dessa forma ao final das k iterações todos os dados

foram usados em algum momento para treino ou teste. Essa técnica também apresenta algumas questões relacionadas ao tamanho do k e ao viés que essa definição pode causar, uma das alternativas quanto á essas questões, é o uso do método *repeted k-folds*, em que a validação cruzada com k -folds é repetida N vezes. O método *repeted k-folds* foi utilizado com $k = 10$ ao longo de 100 iterações.

A técnica *Leave One Out*, é a validação cruzada k -folds onde $k = n$, em que n é o número de amostras, ou seja, ocorrerá k iterações, e para cada iteração todas as $n - 1$ amostras serão usadas para treino e 1 amostra será usada como teste. O principal ponto de atenção dessa técnica é custo computacional, visto que o modelo será treinado n vezes (KIM, 2009).

A melhor configuração obtida para cada extrator considerando as 3 técnicas de validação é apresentada na Tabela 2. O conjunto de valores testados como quantidade de componentes em cada classe varia de um a quantidade de amostras que a classe corrente possui. Os valores testados de iterações do algoritmo EM varia de: $\{1, 2, \dots, 20\}$. Como nessa pesquisa foi utilizada uma técnica de inicialização mais robusta que a aleatoriedade, números altos de iterações não produzem impacto significativo na performance. O conjunto de valores testados para o percentual de redução da imagem de entrada foram: $\{30, 40, 50, 60, 70, 80, 90\}$, iniciando por 30% devido ao resultados apresentados no trabalho de Sarmiento et al. (2020).

Tabela 2 – Melhor configuração de parâmetros de cada extrator.

Melhor Configuração	MGABTD pixel	MGABTD percent	MGABTD neighborhood
Nº de Componentes por Classe	Classe 0: 2	Classe 0: 2	Classe 0: 2
	Classe 1: 2	Classe 1: 2	Classe 1: 2
	Classe 2: 4	Classe 2: 2	Classe 2: 2
	Classe 3: 2	Classe 3: 2	Classe 3: 2
	Classe 4: 2	Classe 4: 2	Classe 4: 2
	Classe 5: 4	Classe 5: 2	Classe 5: 2
	Classe 6: 4	Classe 6: 2	Classe 6: 2
Iterações EM-step	5	5	5
Redução Tamanho	30%	30%	40%
Tempo médio extração	0,44s	0,37s	18,9s

Na Tabela 2 também é apresentado o tempo médio de extração de características em uma imagem que cada variação possui. Apesar do $MGABTD_{pixel}$ ter maior dimensionalidade, o tempo de extração é próximo ao do $MGABTD_{percent}$ que só possui 12 atributos. Nota-se que o extrator $MGABTD_{neighborhood}$ tem um tempo muito maior do que os outros extratores testados, apesar de também ter 12 atributos, esse extrator tem um custo computacional muito maior que os outros, pois além dos processamentos similares aos outros extratores, ele analisa os 8-vizinhos de cada pixel, e mesmo com a redução do

tamanho da imagem e paralelismo do processamento por hemisfério, ainda teve o tempo de extração bem maior que os outros extratores.

5.2 Configuração ótima dos classificadores

Após obter a melhor configuração de cada extrator, foi realizado a busca em grade dos classificadores. Os parâmetros escolhidos correspondem aos mais relevantes descritos no Capítulo ?? e são apresentados na Tabela 3. Determinar a melhor configuração de cada classificador de acordo com o extrator é necessário, pois o mesmo classificador pode necessitar de configurações diferentes para cada extrator, como por exemplo: a configuração de parâmetros ótima de uma MLP usando as características do extrator $MGABTD_{pixel}$, pode ser diferente da melhor configuração utilizando o extrator $MGABTD_{percent}$, logo o mesmo classificador teria configurações ótimas diferentes de acordo com as características de entrada. Na Tabela 3, na coluna Melhor Configuração, é apresentado os parâmetros ótimos obtidos da busca em grade.

A partir da melhor configuração de parâmetros de cada extrator-classificador, pode-se comparar a performance geral de cada abordagem proposta. Também foram utilizadas as 3 técnicas de validação: HO, CV e LOO. Para a técnica HO foi utilizado duas formas de separação da base: 80-20 e 70-30, e para a técnica CV foi utilizado 3 tipos de separação: 5, 10 e 15 dobras. Na Figuras 27, 28 e 29, são apresentadas as acurácias médias obtidas por cada extrator considerando cada técnica de validação. Essa visão é importante para observar que em geral existe pouca variação de performance entre as técnicas de validação. Analisando empiricamente apenas a acurácia, o extrator $MGABTD_{pixel}$ obteve melhor resultado, seguido pelo $MGABTD_{percent}$ e por fim o $MGABTD_{neighborhood}$.

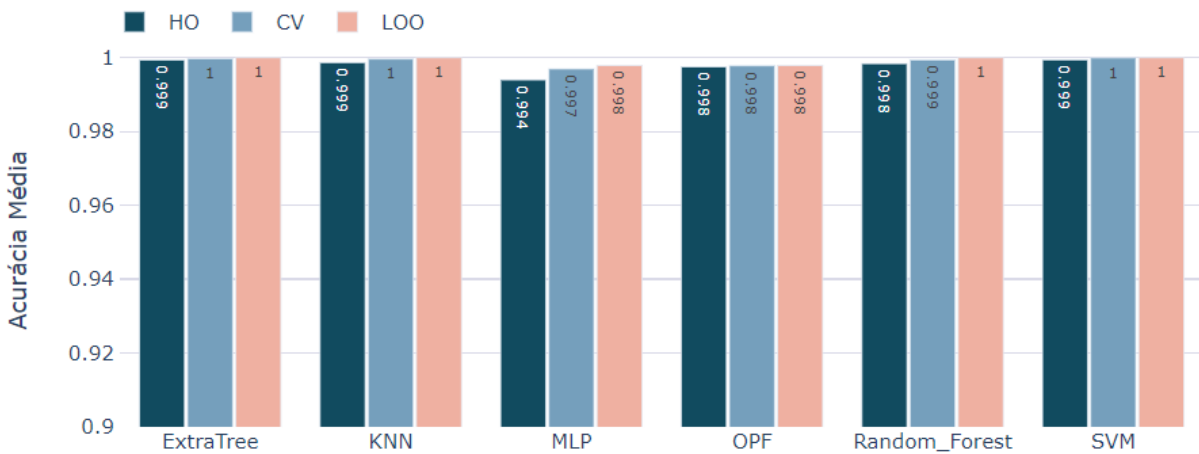


Figura 27 – Resultado extrator $MGABTD_{pixel}$. Fonte: Elaborado pelo autor.

Para tornar o comparativo entre os extratores mais robusto, em vez de escolher os resultados de apenas uma das técnicas de validação, foi calculada uma média aritmética



Figura 28 – Resultado extrator $MGABTD_{percent}$. Fonte: Elaborado pelo autor.

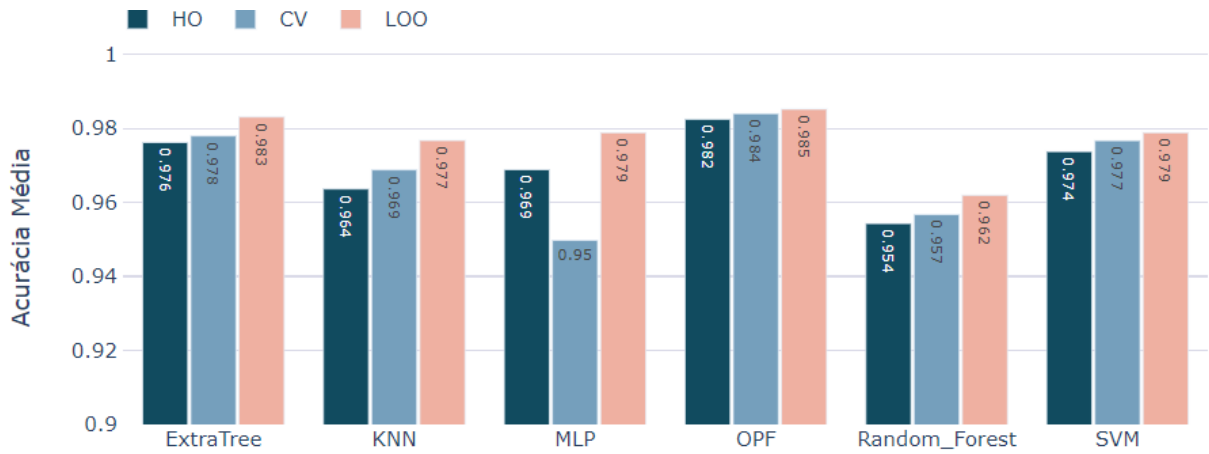


Figura 29 – Resultado extrator $MGABTD_{neighborhood}$. Fonte: Elaborado pelo autor.

das métricas considerando todas as 3 técnicas de validação: HO, CV e LOO. Por exemplo a acurácia final, representada pela coluna ACC na Tabela 3, é calculada como a média aritmética da acurácia média obtida no HO, no CV e no LOO, assim todas as 3 técnicas impactam nos resultados finais, em vez de escolher apenas uma delas. Além da acurácia, outras métricas também foram calculadas: precisão, *recall*, *f1-score*, *Negative Predictive Value* (NPV) e Similaridade Jaccard, todas seguindo a mesma lógica da média entre as técnicas de validação. A métrica NPV, é o percentual de amostras projetadas para a classe negativa que realmente eram da classe negativa. A similaridade Jaccard encontra a similaridade entre dois conjuntos, por meio da divisão do tamanho da interseção entre os conjuntos sobre o tamanho da união dos conjuntos, onde os conjuntos nesse contexto são os valores reais e valores preditos pelo modelo. Na Tabela 3 é apresentado o resultados de cada métrica considerando as configurações ótimas dos extratores e classificadores. Mesmo considerando a média das 3 técnicas de validação o extrator $MGABTD_{pixel}$, em geral obtém o melhor desempenho, seguido pelo $MGABTD_{percent}$ e o $MGABTD_{neighborhood}$.

Tabela 3 – Resultados obtidos com as configurações ótimas de cada extrator-classificador.

Extrator	Classificador	Melhor Configuração	ACC	Precisão	recall	F1-Score	NPV	Jaccard
MGABTD pixel	Extra Tree	Estimadores=100	0.999	0.999	0.999	0.999	0.999	0.999
	KNN	k=5	0.999	0.999	0.999	0.999	0.999	0.998
	OPF	Distância=euclidiana	0.997	0.997	0.997	0.997	0.998	0.996
	RF	Estimadores=300	0.999	0.999	0.999	0.999	0.999	0.998
	SVM	C=10 kernel=rbf gamma=0.0001	0.999	0.999	0.999	0.999	0.999	0.999
	MLP	Iterações=10000 Neurônios=50 Ativação=tanh	0.996	0.996	0.996	0.996	0.997	0.993
MGABTD percent	Extra Tree	Estimadores=100	0.984	0.985	0.984	0.984	0.987	0.973
	KNN	k=5	0.968	0.970	0.968	0.968	0.976	0.947
	OPF	Distância=euclidiana	0.992	0.992	0.992	0.992	0.994	0.987
	RF	Estimadores=100	0.960	0.960	0.960	0.960	0.972	0.936
	SVM	C=1000 kernel=rbf gamma=scale	0.985	0.985	0.985	0.985	0.989	0.975
	MLP	Iterações=10000 Neurônios=100 Ativação=relu	0.967	0.971	0.967	0.966	0.973	0.945
MGABTD neighborhood	Extra Tree	Estimadores=100	0.979	0.980	0.979	0.979	0.984	0.964
	KNN	k = 5	0.969	0.971	0.969	0.969	0.976	0.949
	OPF	Distância=euclidiana	0.983	0.984	0.983	0.983	0.988	0.973
	RF	Estimadores=150	0.957	0.959	0.957	0.957	0.968	0.931
	SVM	C=1000 kernel=rbf gamma=1.0	0.976	0.977	0.976	0.976	0.983	0.961
	MLP	Iterações=10000 Neurônios=150 Ativação=relu	0.965	0.974	0.965	0.965	0.969	0.941

5.2.1 Análise de Tempo de processamento

Um dos principais fatores que impactam o tempo de treinamento e previsão é a dimensionalidade dos dados, e os extratores $MGABTD_{pixel}$, $MGABTD_{percent}$ e $MGABTD_{neighborhood}$, possuem respectivamente, 11.858, 12 e 12 atributos. Nas Figuras 30 e 31 é apresentado, respectivamente, o tempo médio de treinamento e teste dos classificadores. O tempo de teste/projeção apresentado na Figura 31 é o tempo médio que o classificador leva para prevê a classe de uma amostra.

O extrator $MGABTD_{pixel}$ por ter maior dimensionalidade, apresenta os maiores tempos de treinamento e teste em comparação com os outros extratores. O $MGABTD_{percent}$ e $MGABTD_{neighborhood}$ possuem tempos de treinamento e teste similares, visto que ambos possuem o mesmo número de atributos. Nota-se que a MLP foi disparado o classificador mais demorado no treinamento, independente do extrator utilizado, e em contrapartida os classificadores: SVM, RF, Extra Tree e KNN, tiveram o treinamento mais rápido independente do extrator. Como o tempo de teste de todos os classificadores são inferiores a 1s, o tempo gasto em segundos para extrair as características de uma imagem e projetar qual a classe da imagem, é comandado pelo tempo de extração, visto que o tempo de teste tende a ser imperceptível.

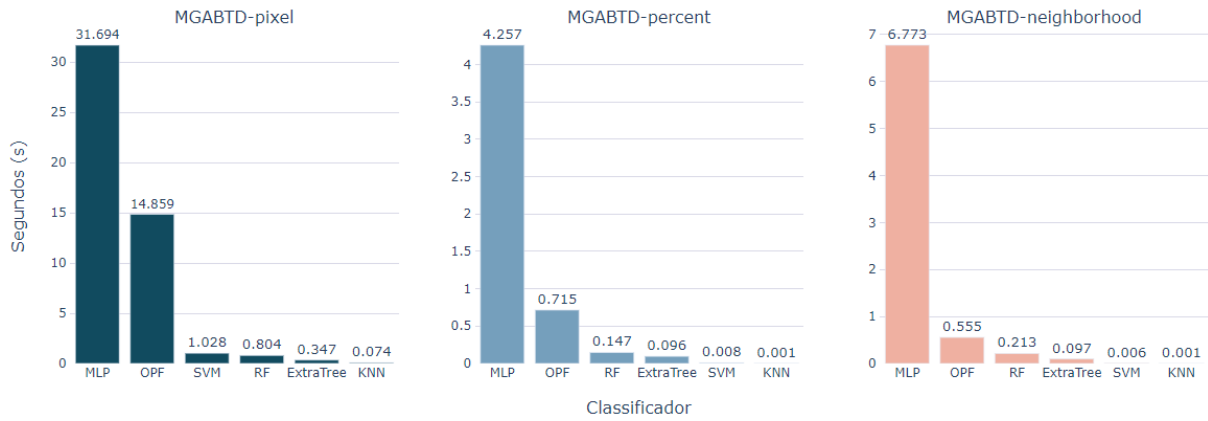


Figura 30 – Tempo médio de Treinamento. Fonte: Elaborado pelo autor.

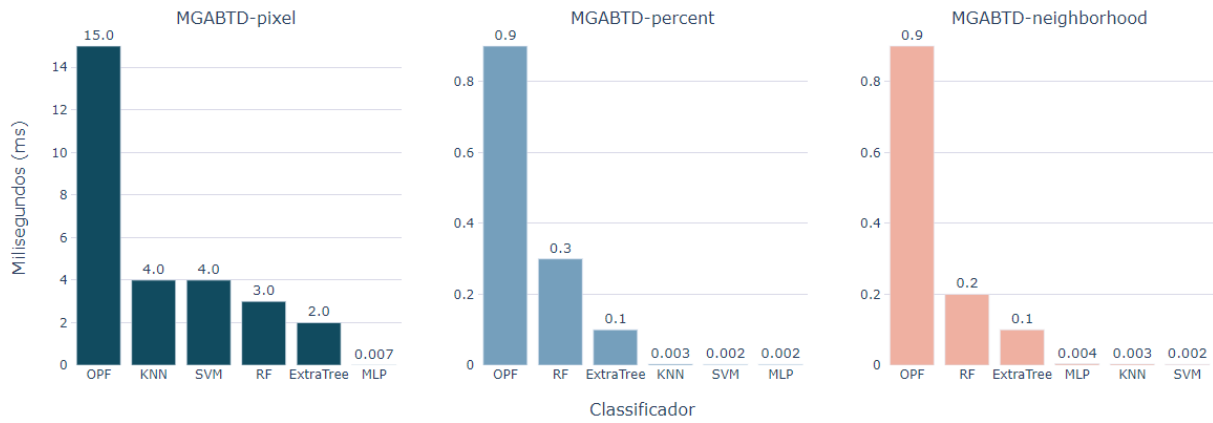


Figura 31 – Tempo médio de projeção/teste. Fonte: Elaborado pelo autor.

Considerando apenas o tempo de treinamento e teste, o classificador Extra Tree, foi o melhor classificador para o $MGABTD_{pixel}$. Para o $MGABTD_{percent}$ e $MGABTD_{neighborhood}$, o SVM foi o melhor classificador, seguido do KNN e Extra Tree.

5.2.2 Seleção do melhor classificador para cada extrator

Analisando as performances dos classificadores na Tabela 3, nota-se que para alguns classificadores as acurácias são bastante próximas. O teste $5x2$ *cv paired t-test* foi aplicado entre os classificadores de cada extrator considerando um nível de significância de $\alpha = 0.05$, e considerando a métrica acurácia como performance. Para o extrator $MGABTD_{pixel}$, o teste resultou que não há diferenças significativas entre o desempenho dos classificadores, ou seja, o ($p - value < \alpha$) logo possuem performance equivalente, e as diferenças de performance apresentadas na Tabela 3, são acaso estatístico. Para os classificadores dos extratores $MGABTD_{percent}$ e $MGABTD_{neighborhood}$, as conclusões foram similares, apenas os classificadores MLP e SVM apresentaram diferenças significativas ($p - value > \alpha$), e os classificadores: KNN, Extra Tree, OPF e RF mostraram-se equivalentes ($p - value < \alpha$).

Na Tabela 4 é apresentado o classificador recomendado para cada extrator.

Tabela 4 – Classificador recomendado considerando acuracidade e tempo de teste para cada variação de extrator proposto.

Extrator de Características	Classificador Recomendado
$MGABTD_{pixel}$	Extra Tree
$MGABTD_{percent}$	KNN e Extra Tree
$MGABTD_{neighborhood}$	KNN e Extra Tree

A escolha do classificador recomendado apresentado na Tabela 4, ocorreu considerando os fatores de equivalência de performance resultantes dos testes de hipótese e os tempos de treinamento e teste apresentados na seção anterior.

5.2.3 Importância dos Atributos

Agora que a melhor configuração de extração e classificação já foi determinada, pode-se analisar qual a importância que cada classificador atribuiu aos atributos criados pelo extrator. Com essa análise pode-se determinar o quão eficientes são os atributos para a classificação e posteriormente usar essa informação para retirar atributos menos importantes por exemplo, ou servir-se dessa informação para pensar em novos atributos, melhorias na modelagem ou acionáveis no mundo real.

Existem diversas técnicas para extrair a importância dos atributos de um classificador, sendo as 3 principais abordagens: baseadas em coeficiente, baseadas em árvore de decisão e baseadas em permutação. O objetivo de cada abordagem é extrair de alguma forma o quanto determinado atributo foi importante para determinar a classe das amostras em geral. A abordagem baseada em coeficiente, utiliza o próprio coeficiente como medida de importância, essa técnica é focada para modelos de regressão linear e logística. Na abordagem baseada em árvore o método mais popular é utilizar a impureza Gini como medida de importância. A abordagem baseada em permutação pode ser utilizada em qualquer classificador, a medida de importância do atributo será determinada sobre o quanto o modelo terá sua performance prejudicada se a variável for removida, onde a remoção da variável corresponde a simplesmente permutá-la aleatoriamente. Existe uma outra técnica que é alternativa a abordagem de permutação, que é o *SHapley Additive exPlanations (SHAP)* onde a importância dos atributos é baseada na magnitude das contribuições que o atributo trás, e isso é medido utilizando a teoria dos jogos cooperativos (MOLNAR, 2022).

Como o classificador Extra Tree foi determinado como o melhor ou um dos melhores para os 3 extratores, ele será utilizado para analisar as importâncias dos atributos. Por ser um classificador baseado em árvore o método utilizado para analisar a importância dos atributos será a impureza Gini. A importância dos atributos considerando o extrator

$MGABTD_{pixel}$, remete a importância dos pixels visto que os atributos representam a classe dos próprios pixels, na Figura 32 é apresentado os pixels que se mostraram mais importantes para a classificação. Pode-se interpretar por exemplo que a região dos pontos brancos é a região mais importante pra determinar se uma tomografia tem AVCi, AVCh ou não tem AVC. Como o tempo é um critério pouco favorável desse extrator, em trabalhos futuros pode-se testar a eficiência da classificação somente com essa região que se mostrou importante para a classificação, podendo diminuir significativamente o número de atributos, e com isso deixando extração, treinamento e teste mais rápidos.

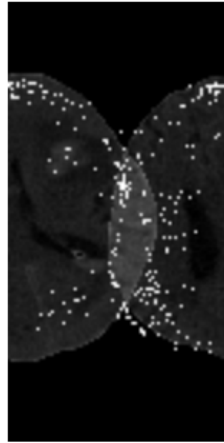


Figura 32 – Importância dos pixels considerando o $MGABTD_{pixel}$ com Extra Tree. Fonte: Elaborado pelo autor.

Em ambos os extratores $MGABTD_{percent}$ e $MGABTD_{neighborhood}$ os atributos são separados de acordo com hemisfério, nas Figuras 33 e 34 são apresentadas a importância dos atributos considerando a impureza Gini, os atributos nomeados no eixo x que possuem o sufixo *dir* são referentes ao hemisfério direito e *esq* são referentes ao hemisfério esquerdo. Em ambos os extratores nota-se que o top 3 atributos mais importantes são: massa cinzenta, massa branca e AVCh, e em contrapartida os atributos com menos importância são os atributos de fluido cerebrospinal e calcificação. Seguindo a mesma linha do $MGABTD_{pixel}$, em trabalhos futuros, pode-se testar a eficiência dos classificadores sem esses dois atributos de menor importância, se a performance não sofrer um impacto significativo, pode-se remover esses atributos dos extratores, gerando um tempo de extração, treinamento e teste mais rápido, visto que a dimensionalidade do dado foi diminuída.

Essa importância baseada na impureza Gini, não trás clareza sobre qual classe o atributo é mais relevante, para isso a abordagem do SHAP foi utilizada. Nas Figuras 35 e 36 são apresentadas a importância dos atributos considerando a abordagem SHAP. Essa visão, apesar de apresentar a magnitude da importância do atributo para cada classe, não é apresentado como essa importância se relaciona com a classe, por exemplo, na Figura 35 o atributo $AVCh_{esq}$ é mais importante para classificar uma amostra para a classe AVCi do que para a classe AVCh, empiricamente esse comportamento seria explicado

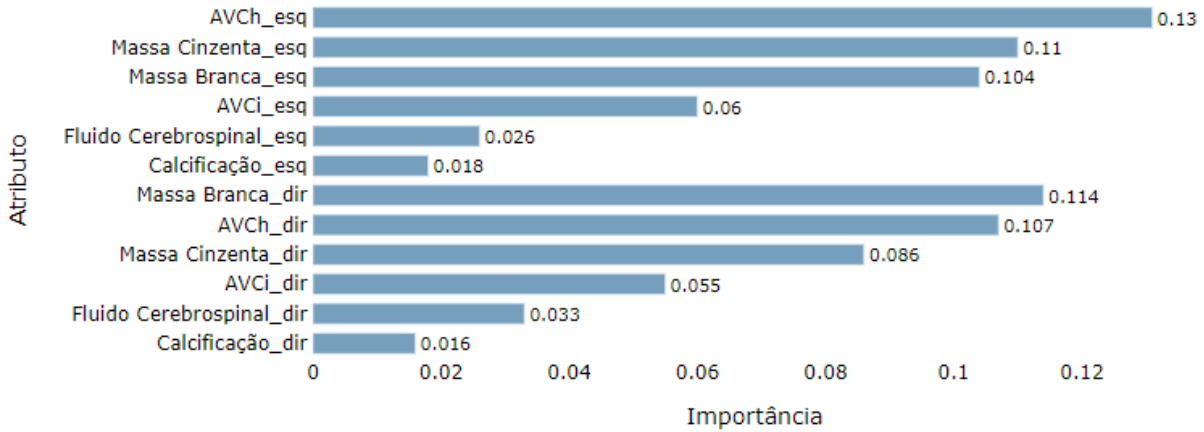


Figura 33 – Importância dos atributos considerando impureza Gini com $MGABTD_{percent}$ com Extra Tree. Fonte: Elaborado pelo autor.

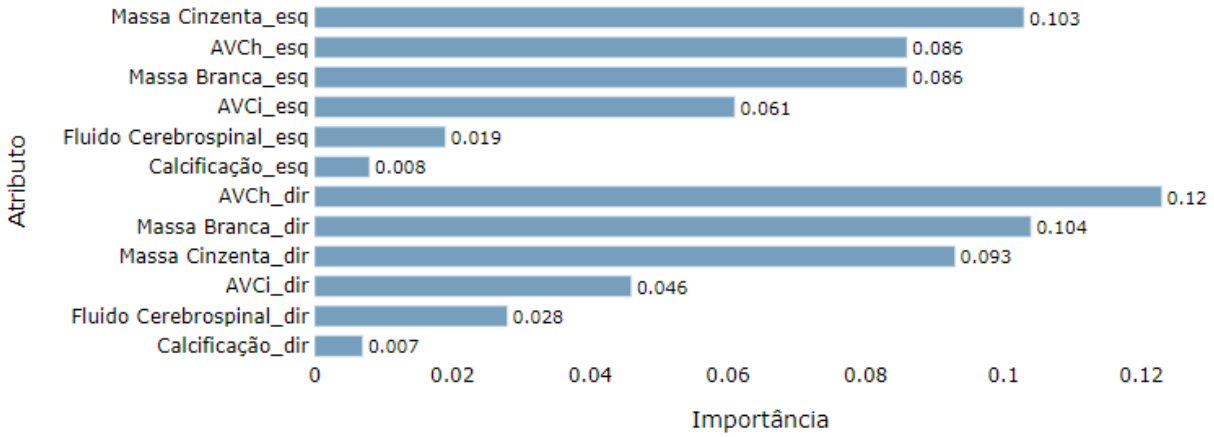


Figura 34 – Importância dos atributos considerando impureza Gini com $MGABTD_{neighborhood}$ com Extra Tree. Fonte: Elaborado pelo autor.

se a relevância do atributo estivesse ligada a diminuição desse atributo, ou seja, quanto mais baixo fosse o valor do $AVCh_{esq}$, mais essa informação seria importante para saber se tem AVCi, no entanto nessa visão, não existe uma medida/métrica que comprove essa afirmação, ou seja, não tem como saber a relação da importância do atributo com a classe. Mesmo utilizando uma técnica diferente para determinar importância, os atributos fluido cerebrospinal e calcificação continuam com a importância baixa, reforçando a hipótese de removê-los futuramente.

5.3 Comparação dos resultados com trabalhos relacionados

Nas seções anteriores, os extratores $MGABTD_{\{pixel, percent, neighborhood\}}$ tiveram suas performances analisadas entre si. Nessa seção os extratores MGABTD serão comparados aos trabalhos relacionados apresentados na Seção 2.4.

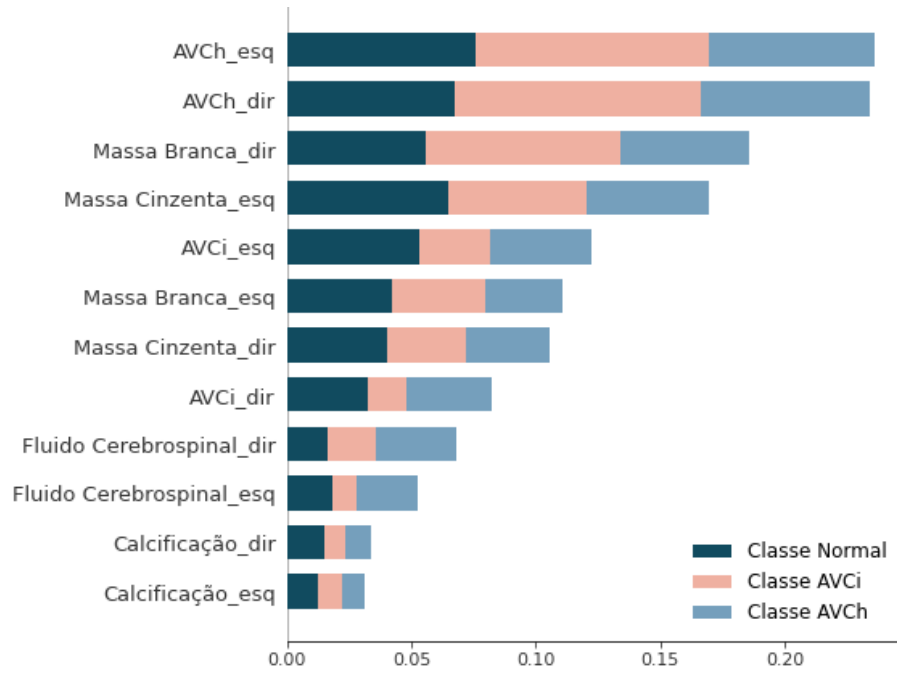


Figura 35 – Importância dos atributos considerando SHAP com $MGABTD_{percent}$ com Extra Tree. Fonte: Elaborado pelo autor.

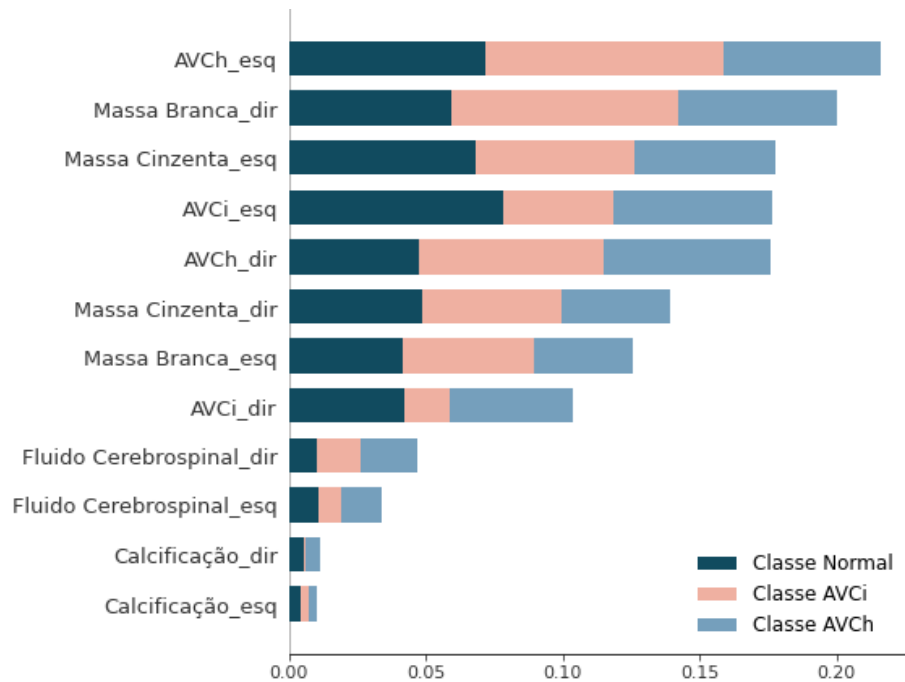


Figura 36 – Importância dos atributos considerando SHAP com $MGABTD_{neighborhood}$ com Extra Tree. Fonte: Elaborado pelo autor.

Na Tabela 5 é apresentado as acurácias obtidas em outras pesquisas para o problema de classificação de AVC. Desses trabalhos apresentados, os trabalhos (DOURADO et al., 2019; DOURADO et al., 2020; SARMENTO et al., 2020; VASCONCELOS et al., 2020) utilizaram a mesma base de dados que a base dessa pesquisa. Sendo assim, a análise

comparativa será focada nesses trabalhos que possuem a mesma base.

Tabela 5 – Comparação dos resultados obtidos com a literatura.

Fonte	Extrator	Classificador	ACC
(DOURADO et al., 2020)	VGG16	KNN	100%
(DOURADO et al., 2019)	NASNet Large	Bayesian	100%
-	MGABTD pixel	Extra Tree	99.9%
(KANCHANA; MENAKA, 2020)	FOF, GLRLM, GLCM, HUM	RF	99.79%
(RODER et al., 2021)	RBM		99.7%
-	MGABTD percent	OPF	99.2%
(GAUTAM; RAMAN, 2021b)	CNN		98.7%
(CHEN et al., 2022)	Resnet50		98.7%
(SARMENTO et al., 2020)	PABTD	SVM	98.4%
-	MGABTD neighborhood	OPF	98.3%
(VASCONCELOS et al., 2020)	ABTD Adaptative	SVM	98.1%
(LO; HUNG; LIN, 2021)	Deep CNN		97.12%
(BADRIYAH et al., 2020)	GLCM	RF	95.97%
(GAUTAM; RAMAN, 2021a)	LNP	TreeBagger	82.6%

Entre as pesquisas que possuem a mesma base de dados desse trabalho, somente as soluções de Dourado et. al, que são abordagens com *Deep Learning*, alcançam acurácia superior ao método $MGABTD_{pixel}$. Nesse caso um possível diferencial dos extratores $MGABTD_{pixel}$ e $MGABTD_{percent}$ seria o tempo de extração e/ou teste. Considerando os experimentos apresentados em Dourado et al. (2019), Dourado et al. (2020), 11 arquiteturas de *Deep Learning* atingiram 100%, dessas, 5 arquiteturas: DenseNet 121, DenseNet 169, Densenet 201, InceptionResNetV3 e NasNetLarge, tiveram tempo de extração/teste superior ao $MGABTD_{pixel}$ e $MGABTD_{percent}$, então apesar dos extratores não atingirem 100% acurácia, eles possuem uma previsão mais rápida nesses cenários.

5.3.1 Comparação com método base ABTD

Os extratores MGABTD são uma proposta de melhoria do método ABTD, então para verificar se de fato houve melhoria da performance do ABTD-7 em relação a performance dos extratores MGABTD, o teste de *Friedman* foi utilizado. O teste foi aplicado com um nível de significância de $\alpha = 0.05$ e utilizou o classificador Extra Tree resultando em um $p - value < \alpha$. Sendo assim os extratores ABTD-7 e as variações do MGABTD, não possuem performance equivalente, e como a acurácia média foi superior, de fato os extratores $MGABTD_{\{pixel, percent, neighborhood\}}$ são uma evolução do ABTD.

5.4 Análise geral dos resultados

Em síntese, dos 3 extratores propostos, o $MGABTD_{pixel}$ foi o que apresentou a melhor performance, seguido pelo $MGABTD_{percent}$, com um tempo de extração de 0.44s e 0.37s respectivamente. Qualquer um dos classificadores testados (KNN, OPF, MLP, Extra Tree, RF) com o extrator $MGABTD_{pixel}$ atingiram 99.9% de acurácia, sendo o classificador Extra Tree o que apresentou melhor custo-benefício referente ao tempo de treinamento (0.34s) e teste (2ms). O classificador Extra Tree também se mostrou a melhor combinação para o extrator $MGABTD_{percent}$, apesar de não ter tido a melhor performance quando comparado ao $MGABTD_{pixel}$, o $MGABTD_{percent}$ possui o tempo de treinamento (0.096s) e teste (0.1ms) muito mais rápido, devido a dimensionalidade dos atributos que é muito menor que o $MGABTD_{pixel}$. O $MGABTD_{neighborhood}$ teve um tempo de extração muito alto (18.9s), quando comparado a outros extratores e teve uma performance mediana, sendo assim o uso desse extrator não é indicado, visto que o $MGABTD_{percent}$ é uma alternativa com tempo de extração menor e performance melhor.

6 Conclusões, Contribuições e Trabalhos Futuros

Esta dissertação tem como principal contribuição o desenvolvimento de 3 extratores de características utilizando mistura de gaussianas e análise da densidade do tecido cerebral. Os extratores: $MGABTD_{pixel}$, $MGABTD_{percent}$ e $MGABTD_{neighborhood}$ representam variações do uso de GMM como uma evolução ao método ABTD.

Conforme apresentado, a principal contribuição desta abordagem é o uso de um estimador de probabilidade para determinar a classe do pixel. Os limiares apresentados na literatura geralmente vêm de observações médicas, que podem variar de pesquisa para pesquisa. Com o estimador de probabilidade, a definição da classe do pixel tende a se tornar mais robusta e confiável.

Para determinar a configuração ótima dos 3 extratores foram utilizados 6 classificadores consolidados no âmbito de ML visando a identificação e classificação de AVC em imagens de TC do cérebro. As imagens TC do cérebro foram compostas de exames de voluntários saudáveis e pacientes com AVCi e AVCh. Além disso, os extratores tiveram suas performances comparadas entre si, buscando determinar a variação mais indicada, e também foram comparados empiricamente com os trabalhos relacionados.

O extrator $MGABTD_{neighborhood}$ não apresentou vantagens significativas quando comparado as demais variações $MGABTD_{pixel}$ e $MGABTD_{percent}$, devido o tempo alto de extração de características (18s), independente da performance na classificação não se torna competitivo com outras soluções apresentadas na literatura. O extrator $MGABTD_{pixel}$ leva em média 0,44s para extrair as características e os classificadores Extra Tree, KNN, RF e SVM atingiram 99,9% de acurácia, precisão e F1-score. O extrator $MGABTD_{percent}$ faz a extração de características em média 0,37s e com o classificador OPF obtém 99,2% de acurácia, precisão e F1-score.

Na literatura apenas as abordagens com *Deep Learning* reportaram acurácia superior ao método $MGABTD_{pixel}$. Quase metade das arquiteturas apresentadas, que atingiram 100% de acurácia, tiveram um tempo de extração superior ao $MGABTD_{pixel}$. Sendo assim, considerando a necessidade rápida de diagnóstico do AVC, o método $MGABTD_{pixel}$ pode ser preferível a algumas arquiteturas de *Deep Learning*.

Dessa forma, pode-se afirmar que ambos os extratores: $MGABTD_{pixel}$, $MGABTD_{percent}$ são métodos eficientes para extração de características usando imagens TC do cérebro, considerando o desempenho e a velocidade da extração. Dessa maneira um dos extratores poderia ser utilizado em um sistema CAD, para auxiliar o diagnóstico de AVC. E com

isso diminuiria o tempo de análise do médico, em todas as imagens geradas no exame de TC, que é uma atividade cansativa e suscetível a erro humano.

6.1 Trabalhos Futuros

Considerando os resultados obtidos nas variações dos extratores propostos, a expectativa é que os trabalhos futuros sejam focados em validação e melhoria na performance.

As iniciativas relacionadas a validação, são focadas em obter mais segurança em relação a estabilidade da performance do método. As atividades futuras seriam:

- Aquisição de mais imagens para realização de mais uma etapa de validação. O objetivo é verificar a capacidade de generalização do método em outras bases de dados;
- Implantação do método proposto em um sistema CAD aplicado na prática, visando investigar a performance do método e avaliar se o método de fato auxiliou o diagnóstico.

As melhorias relacionadas a performance, buscam diminuir ainda mais o tempo de extração, mantendo ou melhorando a acurácia do método. As atividades futuras seriam:

- Segmentar ainda mais a região de interesse, analisando apenas regiões que mostraram ser relevantes para o diagnóstico, como apresentado na Seção 5.2.3. Por conseguinte, o tamanho da imagem será menor levando a um tempo de extração menor;
- Incluir técnicas de paralelização na etapa de projeção da classe dos pixels, e dependendo do quanto o tempo irá diminuir, é possível que a etapa de redução da imagem de entrada se torne desnecessária;
- Testar o uso de classificadores que tenham menor tempo de execução e verificar a performance obtida, como por exemplo: *Naive Bayes* e aplicação da técnica *ensemble* utilizando classificadores mais rápidos.

Espera-se que com estas melhorias, os resultados futuros sejam ainda mais precisos e rápidos, comparados com os obtidos neste trabalho.

Referências

- AGNELLI, J. P. et al. Classification of stroke using neural networks in electrical impedance tomography. *Inverse Problems*, IOP Publishing, v. 36, n. 11, p. 115008, 2020. Citado na página 38.
- ALFEILAT, H. A. A. et al. Effects of distance measure choice on k-nearest neighbor classifier performance: a review. *Big data*, Mary Ann Liebert, Inc., publishers 140 Huguenot Street, 3rd Floor New ... , v. 7, n. 4, p. 221–248, 2019. Citado na página 47.
- ALI, J. et al. Random forests and decision trees. *International Journal of Computer Science Issues (IJCSI)*, International Journal of Computer Science Issues (IJCSI), v. 9, n. 5, p. 272, 2012. Citado na página 57.
- ARCH, A. E. et al. Missed ischemic stroke diagnosis in the emergency department by emergency medicine and neurology services. *Stroke*, Am Heart Assoc, v. 47, n. 3, p. 668–673, 2016. Citado na página 26.
- BADRIYAH, T. et al. Machine learning algorithm for stroke disease classification. In: IEEE. *2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*. Istanbul, Turkey, 2020. p. 1–5. Citado 2 vezes nas páginas 37 e 82.
- BAHL, S. et al. Telemedicine technologies for confronting covid-19 pandemic: a review. *Journal of Industrial Integration and Management*, World Scientific, v. 5, n. 04, p. 547–561, 2020. Citado na página 23.
- BALAFAR, M. Gaussian mixture model based segmentation methods for brain mri images. *Artificial Intelligence Review*, Springer, v. 41, n. 3, p. 429–439, 2014. Citado na página 45.
- BAN, Z.; LIU, J.; CAO, L. Superpixel segmentation using gaussian mixture model. *IEEE Transactions on Image Processing*, IEEE, v. 27, n. 8, p. 4105–4117, 2018. Citado na página 45.
- BANKMAN, I. *Handbook of medical image processing and analysis*. Baltimore, MD, USA: Elsevier, 2008. Citado 3 vezes nas páginas 15, 24 e 35.
- BISHOP, C. M.; NASRABADI, N. M. *Pattern recognition and machine learning*. New York, USA: Springer, 2006. v. 4. Citado 8 vezes nas páginas 15, 41, 42, 43, 45, 50, 51 e 56.
- BOLUKI, S. et al. Constructing pathway-based priors within a gaussian mixture model for bayesian regression and classification. *IEEE/ACM transactions on computational biology and bioinformatics*, IEEE, v. 16, n. 2, p. 524–537, 2017. Citado na página 45.
- BRADY, A. P. Error and discrepancy in radiology: inevitable or avoidable? *Insights into imaging*, Springer, v. 8, p. 171–182, 2017. Citado na página 26.
- BREIMAN, L. Random forests. *Machine learning*, Springer, v. 45, p. 5–32, 2001. Citado na página 58.

BREIMAN, L. et al. Cart. *Classification and regression trees*, Wadsworth, 1984. Citado na página 56.

BRINDHA, R.; KAVITHA, A.; BHUSHAN, B. A systematic review literature on computer-aided detection methods for covid-19 detection in x-ray and ct image modalities. In: SAINI, H. S. et al. (Ed.). *Innovations in Electronics and Communication Engineering*. Singapore: Springer Singapore, 2022. p. 227–233. ISBN 978-981-16-8512-5. Citado na página 23.

BURMAN, P. A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. *Biometrika*, Oxford University Press, v. 76, n. 3, p. 503–514, 1989. Citado na página 72.

BUZUG, T. M. Computed tomography. In: _____. *Springer Handbook of Medical Technology*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. p. 311–342. ISBN 978-3-540-74658-4. Disponível em: <https://doi.org/10.1007/978-3-540-74658-4_16>. Citado 2 vezes nas páginas 23 e 24.

CAPPABIANCO, F. A. et al. Mr-image segmentation of brain tissues based on bias correction and optimum-path forest clustering. In: *Tech. rep.* Campinas, São Paulo: Institute of Computing University of Campinas, 2010. Citado na página 55.

CAPPABIANCO, F. A. et al. Brain tissue mr-image segmentation via optimum-path forest clustering. *Computer Vision and Image Understanding*, Elsevier, v. 116, n. 10, p. 1047–1059, 2012. Citado na página 55.

CDC, C. f. D. C.; PREVENTION. Awareness of stroke warning symptoms–13 states and the district of columbia, 2005. *MMWR. Morbidity and mortality weekly report*, v. 57, n. 18, p. 481–485, 2008. Citado na página 25.

CELUPPI, I. C. et al. Uma análise sobre o desenvolvimento de tecnologias digitais em saúde para o enfrentamento da covid-19 no brasil e no mundo. *Cadernos de Saúde Pública*, SciELO Brasil, v. 37, 2021. Citado na página 23.

CHEN, K.-C.; CHEN, C. Y.-C. Stroke prevention by traditional chinese medicine? a genetic algorithm, support vector machine and molecular dynamics approach. *Soft Matter*, Royal Society of Chemistry, v. 7, n. 8, p. 4001–4008, 2011. Citado na página 53.

CHEN, Y.-T. et al. Deep learning–based brain computed tomography image classification with hyperparameter optimization through transfer learning for stroke. *Diagnostics*, MDPI, v. 12, n. 4, p. 807, 2022. Citado 2 vezes nas páginas 38 e 82.

CHIN, A. Y. et al. Multiphasic perfusion ct in acute middle cerebral artery ischemic stroke: prediction of final infarct volume and correlation with clinical outcome. *Korean journal of radiology*, The Korean Radiological Society, v. 3, n. 3, p. 163–170, 2002. Citado na página 32.

CHINNAMGARI, S. K. *R Machine Learning Projects: Implement supervised, unsupervised, and reinforcement learning techniques using R 3.5*. Birmingham, UK: Packt Publishing Ltd, 2019. Citado na página 41.

CHONG, J. Y. *Intracerebral Hemorrhage*. 2020. Disponível em: <<https://www.msmanuals.com/home/brain,-spinal-cord,-and-nerve-disorders/stroke-cva/intracerebral-hemorrhage>>. Acesso em: 29 jan 2023. Citado 2 vezes nas páginas 15 e 33.

- CLINIC, C. *Thrombosis: Types, causes, symptoms treatment*. 2021. Disponível em: <<https://my.clevelandclinic.org/health/diseases/22242-thrombosis>>. Acesso em: 30 jan 2023. Citado 2 vezes nas páginas 15 e 31.
- CORTES, C.; VAPNIK, V. Support-vector networks. *Machine learning*, Springer, v. 20, n. 3, p. 273–297, 1995. Citado na página 52.
- COVER, T.; HART, P. Nearest neighbor pattern classification. *IEEE transactions on information theory*, IEEE, v. 13, n. 1, p. 21–27, 1967. Citado na página 47.
- CYBENKO, G. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, Springer, v. 2, n. 4, p. 303–314, 1989. Citado na página 51.
- DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, Wiley Online Library, v. 39, n. 1, p. 1–38, 1977. Citado na página 43.
- DEMŠAR, J. Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine learning research*, JMLR. org, v. 7, p. 1–30, 2006. Citado na página 69.
- DENOTTER, T. D.; SCHUBERT, J. Hounsfield unit. 2019. Citado na página 63.
- DEVROYE, L.; WAGNER, T. Distribution-free performance bounds for potential function rules. *IEEE Transactions on Information Theory*, IEEE, v. 25, n. 5, p. 601–604, 1979. Citado na página 72.
- DIETTERICH, T. G. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural computation*, MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info, v. 10, n. 7, p. 1895–1923, 1998. Citado 2 vezes nas páginas 68 e 69.
- DIETTERICH, T. G. Ensemble methods in machine learning. In: SPRINGER. *Multiple Classifier Systems: First International Workshop, MCS 2000 Cagliari, Italy, June 21–23, 2000 Proceedings 1*. Berlin, Heidelberg, 2000. p. 1–15. Citado na página 57.
- DING, N. et al. Real-time anomaly detection based on long short-term memory and gaussian mixture model. *Computers & Electrical Engineering*, Elsevier, v. 79, p. 106458, 2019. Citado na página 45.
- DOI, K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized medical imaging and graphics*, Elsevier, v. 31, n. 4-5, p. 198–211, 2007. Citado na página 23.
- DOURADO, C. M. et al. Deep learning iot system for online stroke detection in skull computed tomography images. *Computer Networks*, Elsevier, v. 152, p. 25–39, 2019. Citado 5 vezes nas páginas 38, 39, 62, 81 e 82.
- DOURADO, C. M. et al. An open ioht-based deep learning framework for online medical image recognition. *IEEE Journal on Selected Areas in Communications*, IEEE, v. 39, n. 2, p. 541–548, 2020. Citado 5 vezes nas páginas 38, 39, 62, 81 e 82.
- DREW, T.; VÕ, M. L.-H.; WOLFE, J. M. The invisible gorilla strikes again: Sustained inattention blindness in expert observers. *Psychological science*, Sage Publications Sage CA: Los Angeles, CA, v. 24, n. 9, p. 1848–1853, 2013. Citado na página 26.

- DUCROS, A.; BOUSSER, M.-G. Thunderclap headache. *Bmj*, British Medical Journal Publishing Group, v. 346, 2013. Citado na página 32.
- EIROLA, E.; LENDASSE, A. Gaussian mixture models for time series modelling, forecasting, and interpolation. In: SPRINGER. *International Symposium on Intelligent Data Analysis*. Berlin, Heidelberg, 2013. p. 162–173. Citado na página 45.
- FANG, G.; XU, P.; LIU, W. Automated ischemic stroke subtyping based on machine learning approach. *IEEE Access*, IEEE, v. 8, p. 118426–118432, 2020. Citado na página 59.
- FARID, N. et al. A comparative analysis for support vector machines for stroke patients. *Recent Adv. Inf. Sci*, v. 41, p. 71–76, 2013. Citado na página 53.
- FARNOUSH, R.; ZAR, P. B. Image segmentation using gaussian mixture model. *INTERNATIONAL JOURNAL OF ENGINEERING SCIENCE (ENGLISH)*, 2008. Citado na página 45.
- FERNANDEZ-LOZANO, C. et al. Random forest-based prediction of stroke outcome. *Scientific reports*, Springer, v. 11, n. 1, p. 1–12, 2021. Citado na página 58.
- FILHO, P. P. R. Métodos de contornos ativos crisp adaptativo 2d e 3d aplicados na segmentação dos pulmões em imagens de tomografia computadorizada do tórax. 2013. Citado na página 62.
- FILHO, P. P. R.; CORTEZ, P. C.; HOLANDA, M. A. Active contour modes crisp: new technique for segmentation of the lungs in ct images. *Revista Brasileira de Engenharia Biomédica*, v. 27, n. 4, p. 259–272, 2011. Citado na página 64.
- FILHO, P. P. R. et al. New approach to detect and classify stroke in skull ct images via analysis of brain tissue densities. *Computer methods and programs in biomedicine*, Elsevier, v. 148, p. 27–43, 2017. Citado na página 64.
- FIX, E.; HODGES, J. *Discriminatory Analysis, Nonparametric Discrimination: Consistency Properties* USAF School of Aviation Medicine, Randolph Field. Bath, United Kingdom, 1951. Citado na página 46.
- FORKERT, N. D. et al. Multiclass support vector machine-based lesion mapping predicts functional outcome in ischemic stroke patients. *PloS one*, Public Library of Science San Francisco, CA USA, v. 10, n. 6, p. e0129569, 2015. Citado na página 53.
- FREITAS, C. F. de. *Imaginologia: série abeno: Odontologia essencial-Parte clínica*. São Paulo, Brasil: Artes Medicas Editora, 2013. Citado 3 vezes nas páginas 15, 34 e 35.
- GAUTAM, A.; RAMAN, B. Brain strokes classification by extracting quantum information from ct scans. *Multimedia Tools and Applications*, Springer, p. 1–17, 2021. Citado 2 vezes nas páginas 38 e 82.
- GAUTAM, A.; RAMAN, B. Towards effective classification of brain hemorrhagic and ischemic stroke using cnn. *Biomedical Signal Processing and Control*, Elsevier, v. 63, p. 102178, 2021. Citado 2 vezes nas páginas 39 e 82.
- GEURTS, P.; ERNST, D.; WEHENKEL, L. Extremely randomized trees. *Machine learning*, Springer, v. 63, n. 1, p. 3–42, 2006. Citado na página 58.

- GONÇALVES, A. et al. Identificação de pacientes com risco de acidente vascular cerebral pela radiografia panorâmica. *JBC j. bras. clin. odontol. integr*, p. 162–164, 2004. Citado na página 30.
- GONZALEZ, R. C.; WOODS, R. C. *Processamento digital de imagens*. São Paulo: Pearson Educación, 2009. Citado na página 45.
- GREENSPAN, H.; RUF, A.; GOLDBERGER, J. Constrained gaussian mixture model framework for automatic segmentation of mr brain images. *IEEE transactions on medical imaging*, IEEE, v. 25, n. 9, p. 1233–1245, 2006. Citado na página 45.
- GROTTA, J. C. et al. *Stroke e-book: Pathophysiology, diagnosis, and management*. USA: Elsevier Health Sciences, 2021. Citado na página 29.
- HANKEY, G. J. Long-term outcome after ischaemic stroke/transient ischaemic attack. *Cerebrovascular diseases*, Karger Publishers, v. 16, n. Suppl. 1, p. 14–19, 2003. Citado na página 25.
- HANKEY, G. J. Secondary stroke prevention. *The Lancet Neurology*, Elsevier, v. 13, n. 2, p. 178–194, 2014. Citado na página 24.
- HART, P. E.; STORK, D. G.; DUDA, R. O. *Pattern classification*. New York, USA: Wiley Hoboken, 2000. Citado 2 vezes nas páginas 45 e 46.
- HASAN, T. et al. Exploring the performances of stacking classifier in predicting patients having stroke. In: IEEE. *2021 8th NAFOSTED Conference on Information and Computer Science (NICS)*. Hanoi, Vietnam, 2021. p. 242–247. Citado na página 59.
- HASTIE, T. et al. *The elements of statistical learning: data mining, inference, and prediction*. New York, USA: Springer, 2009. v. 2. Citado na página 52.
- HAYKIN, S. *Redes neurais: princípios e prática*. São Paulo, Brasil: Bookman Editora, 2001. Citado 3 vezes nas páginas 48, 49 e 52.
- HOLBROOK, R. *The Sliding Window*. 2022. Disponível em: <<https://www.kaggle.com/code/ryanholbrook/the-sliding-window>>. Acesso em: 29 jan 2023. Citado na página 66.
- HOUNSFIELD, G. N. Computerized transverse axial scanning (tomography): Part 1. description of system. *The British journal of radiology*, The British Institute of Radiology, v. 46, n. 552, p. 1016–1022, 1973. Citado na página 34.
- INAMDAR, M. A. et al. A review on computer aided diagnosis of acute brain stroke. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 21, n. 24, p. 8507, 2021. Citado na página 36.
- JI, Z. et al. Fuzzy local gaussian mixture model for brain mr image segmentation. *IEEE transactions on Information Technology in Biomedicine*, IEEE, v. 16, n. 3, p. 339–347, 2012. Citado na página 45.
- KAK, A. C.; SLANEY, M. *Principles of computerized tomographic imaging*. New York, USA: SIAM, 2001. Citado na página 25.
- KANCHANA, R.; MENAKA, R. Ischemic stroke lesion detection, characterization and classification in ct images with optimal features selection. *Biomedical Engineering Letters*, Springer, v. 10, n. 3, p. 333–344, 2020. Citado 2 vezes nas páginas 37 e 82.

- KAYABOL, K.; KUTLUK, S. Bayesian classification of hyperspectral images using spatially-varying gaussian mixture model. *Digital Signal Processing*, Elsevier, v. 59, p. 106–114, 2016. Citado na página 45.
- KIM, J.-H. Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. *Computational statistics & data analysis*, Elsevier, v. 53, n. 11, p. 3735–3745, 2009. Citado na página 73.
- KNERR, S.; PERSONNAZ, L.; DREYFUS, G. Single-layer learning revisited: a stepwise procedure for building and training a neural network. In: *Neurocomputing*. Paris, France: Springer, 1990. p. 41–50. Citado na página 53.
- KRALJ, V.; ČUKELJ, P. Stroke–will it be the next pandemic? *Medicus*, PLIVA HRVATSKA doo, v. 31, n. 1 Moždani udar, p. 7–14, 2022. Citado na página 24.
- KWONG, R. Y.; YUCEL, E. K. Computed tomography scan and magnetic resonance imaging. *Circulation*, Am Heart Assoc, v. 108, n. 15, p. e104–e106, 2003. Citado na página 23.
- LAROBINA, M.; MURINO, L. Medical image file formats. *Journal of digital imaging*, Springer, v. 27, n. 2, p. 200–206, 2014. Citado na página 63.
- LAU, H.-y.; TONG, K.-y.; ZHU, H. Support vector machine for classification of walking conditions of persons after stroke with dropped foot. *Human movement science*, Elsevier, v. 28, n. 4, p. 504–514, 2009. Citado na página 53.
- LAXHAMMAR, R.; FALKMAN, G.; SVIESTINS, E. Anomaly detection in sea traffic—a comparison of the gaussian mixture model and the kernel density estimator. In: *IEEE. 2009 12th International Conference on Information Fusion*. Seattle, WA, USA, 2009. p. 756–763. Citado na página 45.
- LEPPALA, J. M. et al. Different risk factors for different stroke subtypes: association of blood pressure, cholesterol, and antioxidants. *Stroke*, Am Heart Assoc, v. 30, n. 12, p. 2535–2540, 1999. Citado 2 vezes nas páginas 30 e 32.
- LI, L. et al. Anomaly detection via a gaussian mixture model for flight operation and safety monitoring. *Transportation Research Part C: Emerging Technologies*, Elsevier, v. 64, p. 45–57, 2016. Citado na página 45.
- LIAQAT, A. et al. Gastric tract infections detection and classification from wireless capsule endoscopy using computer vision techniques: A review. *Current Medical Imaging*, Bentham Science Publishers, v. 16, n. 10, p. 1229–1242, 2020. Citado na página 36.
- LIU, Z.; SHI, Y.; YANG, B. Open innovation in times of crisis: An overview of the healthcare sector in response to the covid-19 pandemic. *Journal of Open Innovation: Technology, Market, and Complexity*, MDPI, v. 8, n. 1, p. 21, 2022. Citado na página 23.
- LO, C.-M.; HUNG, P.-H.; LIN, D.-T. Rapid assessment of acute ischemic stroke by computed tomography using deep convolutional neural networks. *Journal of Digital Imaging*, Springer, v. 34, n. 3, p. 637–646, 2021. Citado 2 vezes nas páginas 38 e 82.
- LOTUFO, P. A.; BENSENOR, I. J. M. Race and stroke mortality in brazil. *Rev. Saude Publica*, v. 47, n. 6, p. 1201–1204, dec 2013. Citado na página 29.

- LUIZ, J. et al. Auxílio ao diagnóstico do glaucoma baseado em métodos de processamento digital de imagens. *XXIV Congresso Brasileiro de Engenharia Biomédica – CBEB*, UNESP - Sorocaba, 2014. Citado na página 36.
- LUO, S. et al. Automatic liver parenchyma segmentation from abdominal ct images using support vector machines. In: IEEE. *2009 ICME International Conference on Complex Medical Engineering*. Tempe, AZ, USA, 2009. p. 1–5. Citado na página 23.
- MACKAY, J.; MENSAH, G. A.; GREENLUND, K. *The atlas of heart disease and stroke*. USA: World Health Organization, 2004. Citado na página 24.
- MAIER, O. et al. Extra tree forests for sub-acute ischemic stroke lesion segmentation in mr sequences. *Journal of neuroscience methods*, Elsevier, v. 240, p. 89–100, 2015. Citado na página 59.
- MANZI, F. R. et al. Panoramic radiography as an auxiliary in detecting patients at risk for cerebrovascular accident (cva): a case report. *Journal of oral science*, Nihon University School of Dentistry, v. 45, n. 3, p. 177–180, 2003. Citado na página 30.
- MARQUES, P. M. d. A. Diagnóstico auxiliado por computador na radiologia. *Radiologia Brasileira*, SciELO Brasil, v. 34, p. 285–293, 2001. Citado 2 vezes nas páginas 23 e 24.
- MASSENA, V. A razão do uso de rtpa na trombólise do ave isquêmico. *J. bras. med*, p. 19–29, 2012. Citado na página 33.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, Springer, v. 5, n. 4, p. 115–133, 1943. Citado na página 48.
- MCKINLEY, R. et al. Fully automated stroke tissue estimation using random forest classifiers (faster). *Journal of Cerebral Blood Flow & Metabolism*, SAGE Publications Sage UK: London, England, v. 37, n. 8, p. 2728–2741, 2017. Citado na página 58.
- MCLACHLAN, G.; BASFORD, K. *Mixture models marcel dekker*. New York, 1988. Citado na página 43.
- MEHROTRA, K.; MOHAN, C. K.; RANKA, S. *Elements of artificial neural networks*. USA: MIT press, 1997. Citado na página 48.
- MENON, B. K. et al. Initial experience with the penumbra stroke system for recanalization of large vessel occlusions in acute ischemic stroke. *Neuroradiology*, Springer, v. 53, p. 261–266, 2011. Citado na página 32.
- MISAWA, M. et al. Accuracy of computer-aided diagnosis based on narrow-band imaging endocytoscopy for diagnosing colorectal lesions: comparison with experts. *International journal of computer assisted radiology and surgery*, Springer, v. 12, p. 757–766, 2017. Citado na página 36.
- MITRA, J. et al. Lesion segmentation from multimodal mri using random forest following ischemic stroke. *NeuroImage*, Elsevier, v. 98, p. 324–335, 2014. Citado na página 58.
- MOLNAR, C. *Interpretable Machine Learning: A guide for making black box models explainable*. 2. ed. Independently published, 2022. Disponível em: <<https://christophm.github.io/interpretable-ml-book>>. Citado na página 78.

- MOZAFFARIAN, D. et al. Heart disease and stroke statistics—2015 update: a report from the american heart association. *circulation*, Am Heart Assoc, v. 131, n. 4, p. e29–e322, 2015. Citado na página 31.
- MÜHL-BENNINGHAUS, R. et al. Utility of hounsfield unit in the diagnosis of tandem occlusion in acute ischemic stroke. *Neurological Sciences*, Springer, v. 42, p. 2391–2396, 2021. Citado na página 27.
- NEWMAN-TOKER, D. E. et al. Missed diagnosis of stroke in the emergency department: a cross-sectional analysis of a large population-based sample. *Diagnosis*, De Gruyter, v. 1, n. 2, p. 155–166, 2014. Citado na página 26.
- O'DONNELL, M. J. et al. Global and regional effects of potentially modifiable risk factors associated with acute stroke in 32 countries (INTERSTROKE): a case-control study. *Lancet*, Elsevier BV, v. 388, n. 10046, p. 761–775, aug 2016. Citado na página 29.
- OH, S.-y. et al. Initial misdiagnosis of aneurysmal subarachnoid hemorrhage: associating factors and its prognosis. *Acta Neurochirurgica*, Springer, v. 160, p. 1105–1113, 2018. Citado na página 26.
- OLIVEIRA, R. d. M. C. de; ANDRADE, L. A. F. de. Acidente vascular cerebral. *Rev Bras Hipertens*, v. 8, p. 280–290, 2001. Citado 2 vezes nas páginas 25 e 30.
- O'CARROLL, C. B.; BROWN, B. L.; FREEMAN, W. D. Intracerebral hemorrhage: a common yet disproportionately deadly stroke subtype. In: ELSEVIER. *Mayo Clinic Proceedings*. USA, 2021. v. 96, n. 6, p. 1639–1654. Citado na página 32.
- PAPA, J. P. et al. Efficient supervised optimum-path forest classification for large datasets. *Pattern Recognition*, Elsevier, v. 45, n. 1, p. 512–520, 2012. Citado na página 54.
- PAPA, J. P. et al. Fast and accurate holistic face recognition using optimum-path forest. In: IEEE. *2009 16th International Conference on Digital Signal Processing*. Santorini, Greece, 2009. p. 1–6. Citado na página 55.
- PAPA, J. P.; FALCAO, A. X.; SUZUKI, C. T. Supervised pattern classification based on optimum-path forest. *International Journal of Imaging Systems and Technology*, Wiley Online Library, v. 19, n. 2, p. 120–131, 2009. Citado 2 vezes nas páginas 54 e 55.
- PAPA, J. P.; ROCHA, A. Image categorization through optimum path forest and visual words. In: IEEE. *2011 18th IEEE International Conference on Image Processing*. Brussels, Belgium, 2011. p. 3525–3528. Citado na página 55.
- PATEL, S.; PARIKH, A.; OKORIE, O. N. Subarachnoid hemorrhage in the emergency department. *International Journal of Emergency Medicine*, BioMed Central, v. 14, n. 1, p. 1–8, 2021. Citado na página 26.
- PERLINI, N. M. O. G.; FARO, A. C. M. Cuidar de pessoa incapacitada por acidente vascular cerebral no domicílio: o fazer do cuidador familiar. *Revista da Escola de Enfermagem da USP*, SciELO Brasil, v. 39, p. 154–163, 2005. Citado na página 25.
- PIANYKH, O. *Digital Imaging and Communications in Medicine (DICOM): A Practical Introduction and Survival Guide*. Berlin, Germany: Soc Nuclear Med, 2009. Citado na página 63.

- PIAUÍ, G. *AVC volta a ser a principal causa de morte no Brasil e campanha alerta para a prevenção*. 2022. Disponível em: <<https://www.pi.gov.br/noticias/avc-volta-a-ser-a-principal-cao-de-morte-no-brasil-e-campanha-alerta-para-a-prevencao/>>. Acesso em: 29 jan 2023. Citado na página 25.
- QUINLAN, J. R. Induction of decision trees. *Machine learning*, Springer, Boston, USA, v. 1, p. 81–106, 1986. Citado na página 56.
- QUINLAN, J. R. *C4. 5: programs for machine learning*. San Mateo, California: Elsevier, 2014. Citado na página 56.
- QUINTA, L. N. B. et al. Floresta de caminhos ótimos na classificação de pólen. In: *WVC 2012-Workshop de Visão Computacional*. Goiânia, GO: Universidade Federal de Goiás, 2012. Citado na página 56.
- RAFFIN, C. N. et al. Clinical and interventional revascularization in the acute ischemic stroke: national opinion. *Arquivos de Neuro-psiquiatria*, v. 64, n. 2A, p. 342–348, 2006. Citado na página 33.
- RAILEANU, L. E.; STOFFEL, K. Theoretical comparison between the gini index and information gain criteria. *Annals of Mathematics and Artificial Intelligence*, Springer, v. 41, p. 77–93, 2004. Citado na página 56.
- REYNOLDS, D. A. Gaussian mixture models. *Encyclopedia of biometrics*, Berlin, Springer, v. 741, n. 659-663, 2009. Citado na página 43.
- ROCHA, L. M.; CAPPABIANCO, F. A.; FALCÃO, A. X. Data clustering as an optimum-path forest problem with applications in image analysis. *International Journal of Imaging Systems and Technology*, Wiley Online Library, Campinas, São Paulo, v. 19, n. 2, p. 50–68, 2009. Citado na página 55.
- RODER, M. et al. Enhancing shallow neural networks through fourier-based information fusion for stroke classification. In: *IEEE. 2021 34th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. Gramado, Rio Grande do Sul, Brazil, 2021. p. 378–385. Citado 2 vezes nas páginas 38 e 82.
- ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, American Psychological Association, v. 65, n. 6, p. 386, 1958. Citado na página 48.
- RUSSELL, S. J. *Artificial intelligence a modern approach*. USA: Pearson Education, Inc., 2010. Citado na página 41.
- SARAIVA, A. S.; BORCHARTT, T. B. Uma api para extração de características de imagens médicas. *Jornada de Informática do Maranhão - JIM*, 11 2014. Citado na página 36.
- SARBAZI, E. et al. Factors related to 6-month mortality after the first-ever stroke. *Journal of Education and Health Promotion*, Wolters Kluwer–Medknow Publications, v. 7, 2018. Citado na página 32.
- SARMENTO, R. M. et al. An iot platform for the analysis of brain ct images based on parzen analysis. *Future Generation Computer Systems*, Elsevier, v. 105, p. 135–147, 2020. Citado 8 vezes nas páginas 16, 37, 62, 63, 71, 73, 81 e 82.

- SAVER, J. L. Time is brain—quantified. *Stroke*, Am Heart Assoc, v. 37, n. 1, p. 263–266, 2006. Citado na página 25.
- SCHOLKOPF, B. et al. Comparing support vector machines with gaussian kernels to radial basis function classifiers. *IEEE transactions on Signal Processing*, IEEE, v. 45, n. 11, p. 2758–2765, 1997. Citado na página 53.
- SIEGEL, E. et al. Computerized follow-up of discrepancies in image interpretation between emergency and radiology departments. *Journal of Digital Imaging*, Springer, v. 11, p. 18–20, 1998. Citado na página 26.
- SILVA, F. M. S. da; OLIVEIRA, E. M. F. de. Comparação dos métodos de imagem (tomografia computadorizada e ressonância magnética) para o diagnóstico de acidente vascular encefálico. *Revista Enfermagem Contemporânea*, v. 6, n. 1, p. 81–89, 2017. Citado na página 23.
- SIRSAT, M. S.; FERMÉ, E.; CÂMARA, J. Machine learning for brain stroke: a review. *Journal of Stroke and Cerebrovascular Diseases*, Elsevier, v. 29, n. 10, p. 105162, 2020. Citado na página 36.
- STEHLING, R. O.; NASCIMENTO, M. A.; FALCÃO, A. X. A compact and efficient image retrieval approach based on border/interior pixel classification. In: *Proceedings of the eleventh international conference on Information and knowledge management*. New York, USA: Association for Computing Machinery, 2002. p. 102–109. Citado na página 55.
- STEINWART, I. Support vector machines are universally consistent. *Journal of Complexity*, Elsevier, v. 18, n. 3, p. 768–791, 2002. Citado na página 53.
- SUBUDHI, A.; DASH, M.; SABUT, S. Automated segmentation and classification of brain stroke using expectation-maximization and random forest classifier. *Biocybernetics and Biomedical Engineering*, Elsevier, v. 40, n. 1, p. 277–289, 2020. Citado na página 58.
- TEIXEIRA, R. A.; SILVA, L. de D.; FERREIRA, V. Tratamento trombolítico no acidente vascular cerebral isquêmico. *Revista Neurociências*, v. 12, n. 1, p. 5–17, 2004. Citado na página 33.
- TSAO, C. W. et al. Heart disease and stroke statistics—2022 update: a report from the american heart association. *Circulation*, Am Heart Assoc, v. 145, n. 8, p. e153–e639, 2022. Citado 2 vezes nas páginas 25 e 30.
- TUM, J.; MIDDLETON, A. *Radiography of cultural material*. Oxford, UK: Routledge, 2006. Citado na página 34.
- VAPNIK, V. N. An overview of statistical learning theory. *IEEE transactions on neural networks*, IEEE, v. 10, n. 5, p. 988–999, 1999. Citado na página 52.
- VASCONCELOS, F. F. et al. Artificial intelligence techniques empowered edge-cloud architecture for brain ct image analysis. *Engineering Applications of Artificial Intelligence*, Elsevier, v. 91, p. 103585, 2020. Citado 6 vezes nas páginas 37, 61, 62, 65, 81 e 82.
- VERMEULEN, M. J.; SCHULL, M. J. Missed diagnosis of subarachnoid hemorrhage in the emergency department. *Stroke*, Am Heart Assoc, v. 38, n. 4, p. 1216–1221, 2007. Citado na página 26.

- VERVERIDIS, D.; KOTROPOULOS, C. Emotional speech classification using gaussian mixture models and the sequential floating forward selection algorithm. In: IEEE. *2005 iee international conference on multimedia and expo*. Amsterdam, Netherlands, 2005. p. 1500–1503. Citado na página 45.
- WAFI, H. A. et al. Burden of stroke in europe: thirty-year projections of incidence, prevalence, deaths, and disability-adjusted life years. *Stroke*, Am Heart Assoc, v. 51, n. 8, p. 2418–2427, 2020. Citado na página 24.
- WARDLAW, J.; FARRALL, A. *Diagnosis of stroke on neuroimaging*. London, UK: British Medical Journal Publishing Group, 2004. 655–656 p. Citado 2 vezes nas páginas 25 e 33.
- WARLOW, C. Epidemiology of stroke. *The Lancet*, Elsevier, v. 352, p. S1–S4, 1998. Citado na página 29.
- WILLIAMS, B. et al. Data-driven model development for cardiomyocyte production experimental failure prediction. In: *Computer Aided Chemical Engineering*. Milano, Italy: Elsevier, 2020. v. 48, p. 1639–1644. Citado 2 vezes nas páginas 15 e 58.
- WU, G. et al. Early identification of ischemic stroke in noncontrast computed tomography. *Biomedical Signal Processing and Control*, Elsevier, v. 52, p. 41–52, 2019. Citado 2 vezes nas páginas 24 e 36.
- YAHYA, R. I. et al. Membrane computing for real medical image segmentation. *Aro-the Scientific Journal of Koya University*, v. 6, n. 2, p. 27–38, 2018. Citado na página 66.
- YANASE, J.; TRIANTAPHYLLOU, E. A systematic survey of computer-aided diagnosis in medicine: Past and present developments. *Expert Systems with Applications*, Elsevier, v. 138, p. 112821, 2019. Citado 2 vezes nas páginas 24 e 36.
- YANG, K.-H. *Basic finite element method as applied to injury biomechanics*. USA: Academic Press, 2017. Citado na página 34.
- YOO, A. J. et al. The penumbra stroke system: a technical review. *Journal of neurointerventional surgery*, British Medical Journal Publishing Group, v. 4, n. 3, p. 199–205, 2012. Citado na página 32.
- YOUSEM, D. M. *Neuroradiologia*. Rio de Janeiro: Elsevier Brasil, 2011. Citado 3 vezes nas páginas 26, 27 e 63.
- ZAHOOOR, S. et al. Breast cancer detection and classification using traditional computer vision techniques: a comprehensive review. *Current Medical Imaging*, Bentham Science Publishers, v. 16, n. 10, p. 1187–1200, 2020. Citado na página 36.
- ZHANG, C. R. et al. Determinants of white matter hyperintensity burden differ at the extremes of ages of ischemic stroke onset. *Journal of Stroke and Cerebrovascular Diseases*, Elsevier, v. 24, n. 3, p. 649–654, 2015. Citado na página 24.
- ZONG, B. et al. Deep autoencoding gaussian mixture model for unsupervised anomaly detection. In: *International conference on learning representations*. Vancouver, Canada: International Conference on Learning Representations, 2018. Citado na página 45.