

INSTITUTO POLITÉCNICO DE LISBOA

**ESCOLA SUPERIOR DE TECNOLOGIA DA SAÚDE DE
LISBOA**

**Classificação de tumores cerebrais com
algoritmos de *Machine Learning***

MONIQUE JOAQUIM WITT GARZILLO

Doutora Ana Luísa da Costa Ferreira Vieira, ESTeSL
Professor Doutor Luís Carvalho Freire, ESTeSL

Mestrado em Tecnologias de Física Médica

Lisboa, 2022

INSTITUTO POLITÉCNICO DE LISBOA
ESCOLA SUPERIOR DE TECNOLOGIA DA SAÚDE DE
LISBOA

Classificação de tumores cerebrais com
algoritmos de *Machine Learning*

MONIQUE JOAQUIM WITT GARZILLO

Orientadora: Doutora Ana Luísa da Costa Ferreira Vieira, ESTeSL

Orientador: Doutor Luís Carvalho Freire, ESTeSL

Júri

Presidente: Doutor Nuno Teixeira, ESTeSL

Arguente: Doutor Augusto Marques Ferreira da Silva, UA

Mestrado em Tecnologias de Física Médica

(esta versão incluiu as críticas e sugestões feitas pelo júri)

Lisboa, 2022

Classificação de tumores cerebrais com algoritmos de *Machine Learning*

Mestrado em Tecnologias de Física Médica – 1ª edição

A Escola Superior de Tecnologia da Saúde de Lisboa tem o direito de arquivar e publicar esta dissertação de mestrado, através de exemplares impressos ou de forma digital e de a divulgar através de repositórios científicos e admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e que tal não viole nenhuma restrição imposta por artigos publicados que nela tenham sido incluídos.

Agradecimentos

Poucos sentimentos humanos são tão transbordantes e plenificam tanto quanto a gratidão. Ela advém da alegria e alívio ímpares que sente uma pessoa que se encontra na escuridão de momentos difíceis e sobre ela é jogada a luz de alguém que lhe estende a mão. Clichés são clichés por algum motivo, como é o caso da frase sempre usada em agradecimentos formais e informais “palavras não são o suficiente para eu agradecer”. Embora de facto não o sejam, essas palavras de agradecimento que ficarão para sempre manifestas neste trabalho que tanto me custou são o que tenho de melhor neste momento para me expressar.

As primícias sempre para o Senhor, que me trouxe até esse instante rodeada de mais bênçãos na forma de pessoas e oportunidades gentis de aprendizado que eu mereceria, e que me deu forças mesmo nos dias mais difíceis. Ao meu melhor amigo, companheiro e marido Magnum, que nunca mediu esforços para me ajudar a trabalhar pelos meus sonhos, abraçando como seus vários deles e deixando tudo para trás para me seguir nessa empreitada de vir morar do outro lado do Atlântico. Meu amor, você esteve comigo dia após dia, enfrentando comigo situações que jamais pensaríamos passar, e prosperando e crescendo de uma maneira maravilhosa. Só você e eu sabemos o quanto. Aos meus pais, Luís e Mirian, que tantas vezes deixaram seus próprios projetos e estudos de lado para que nós o tivéssemos. Se hoje eu estou concluindo esse mestrado, é porque vocês não duvidaram de mim por um instante sequer, sempre me incentivaram a voar tão longe quanto eu pudesse e deram todo o apoio que eu precisei e muito mais. Espero que tenha valido tudo a pena. Ao Henrique e à Letícia, meus segundos pais, que nos adotaram como filhos desde o primeiro dia, os agradecimentos se estendem a vocês de maneira integral. Sou uma pessoa abençoada por ter vocês ao meu lado. Aos meus irmãos Fábio, Jaqueline, Orlando e Rafaela, vocês já estão todos crescidos, mas como irmã mais velha eu sempre espero ser um bom exemplo para todos vocês e poder ser o apoio quando vocês precisarem, mas a verdade é que cada um é exemplo para mim com sua personalidade única e apoio quando eu precisei. Sou grata a Deus pela honra de ser irmã de vocês. Gabi, muito obrigada também, por toda a torcida e por tirar dúvidas aleatórias sobre as partes do cérebro por mensagem. Aos meus avôs Armin e Orlando, que infelizmente não puderam me ver concluindo o mestrado, mas que estão sempre comigo por meio das orações que fizeram e exemplo que deram em vida. Às minhas avós Dóris e Hilda, pelo apoio e amor incondicional, por toda a ajuda, palavras, sabedoria, risadas, livros e panos de prato pintados à mão. Aos

meus tios, Josefa, Luís Antônio e Marcus, por serem mais uma fonte de amor incondicional, dedicação, conselhos e orações. Agradeço aos meus sogros por sempre me acolherem como filha e por me presentear com um homem como o Magnum, e aos meus cunhados por sempre terem me tratado como se fosse da família.

Embora as pessoas mais importantes da minha vida estejam a mais de 8000 quilômetros de distância, esse trabalho não teria sido concretizado sem algumas das pessoas que conheci em Portugal e que me apoiaram durante esse mestrado. Inês, muito obrigada por todas as conversas, taças de vinho, passeios e amizade incondicional. Minha vida aqui não teria sido a mesma sem ti. Dani, infelizmente você teve que deixar o mestrado a meio, mas a amizade que construímos vai perdurar. Muito obrigada por tudo.

Por último, e com certeza com um lugar de destaque, à professora Ana Luísa, que orientou esse trabalho, meu mais sincero agradecimento por sua competência, objetividade, eficiência, pela confiança depositada em mim e principalmente por ter proposto um tema que me chamou tanto a atenção e com o qual gostei tanto de trabalhar. Penso que não poderia ter escolhido orientadora melhor.

Para finalizar – poucos entenderão, mas é suposto – espero do fundo do coração ter orgulhado minha tribo.

A Inteligência Artificial (IA) e a *Data Science* são dois dos assuntos mais debatidos atualmente na Saúde. A IA tem sido considerada como uma parte essencial na resolução de problemas nesse campo (Patel *et al.*, 2008), e mesmo um quarto paradigma da Ciência como um todo, junto com as componentes Teórica, Experimental e Computacional (Cao, 2017).

Neste trabalho, são feitos testes com algoritmos de *Machine Learning* para a classificação multiclasse de três *datasets* de imagens de Ressonância Magnética de tumores cerebrais.

O objetivo principal dos testes com *datasets* diferentes é verificar a influência que os dados têm sobre a acurácia de cada um dos algoritmos; foram testados diferentes algoritmos e analisado qual deles (e com quais ajustes de parâmetros, quando há) que apresenta melhor desempenho; foram também testados diferentes *splits* dos dados para averiguar o quanto a divisão deles interfere no resultado da acurácia.

Foram utilizados três *datasets* (conjuntos de imagens) com 4 categorias de tumor cerebral: glioma, meningioma, tumor da glândula pituitária e sem tumor. O primeiro *dataset* é composto por 2870 imagens diferentes, o segundo por 7020 imagens, e o terceiro é um subconjunto do segundo, com 2870 imagens, com a mesma distribuição por categorias que o primeiro. Os algoritmos usados foram o *Decision Tree* (com os parâmetros definidos por defeito), o *Random Forest* (com variação do número de estimadores); o *Linear Discriminant Analysis* (que não tem parâmetros para ajustar); o *Support Vector Machine* com os 4 *kernels* diferentes disponíveis (Linear, Polinomial, RBF e Sigmoide) e variando o parâmetro de regularização; e o *k-Nearest Neighbors*, variando o número de vizinhos. Além disso, em cada caso também foram testados três *splits* diferentes – separação entre dados de treino e teste: 70%, 80% e 90% dos dados para treino.

Como medida de desempenho, foi usada a acurácia, e realizados testes de significância estatística entre resultados. O algoritmo com melhores resultados foi o *Random Forest*, que atingiu 96,425%, seguido pelo *kernel* RBF do *Support Vector Machine*, com 93,615%. Os resultados são comparáveis a outros trabalhos publicados que apresentam metodologias semelhantes.

Palavras-chave: *Machine Learning*, classificação, tumor cerebral

Artificial Intelligence (AI) and Data Science are two of the most debated subjects currently in Health. AI has been considered an essential part of problem solving in this field (Patel, *et al.*, 2008), and even a fourth paradigm of Science as a whole, along with the Theoretical, Experimental and Computational components (Cao, 2017).

In this work, tests are conducted with Machine Learning algorithms for the multiclass classification of three datasets of MRI images of brain tumors.

The main objective of tests with different datasets is to verify the influence that the data have on the accuracy of each of the algorithms; different algorithms were tested and analyzed which algorithm (and with which parameter settings, if any) had the best performance; different splits were also tested to see how much the division interferes with the accuracy result.

Three datasets (image sets) with 4 brain tumor categories were tested: glioma, meningioma, pituitary gland tumor, and no tumor. The first dataset is composed of 2870 different images, the second of 7020 images, and the third is a subset of the second, with 2870 images, with the same distribution by categories as the first. The algorithms used for testing were Decision Tree (with parameters defined by default), Random Forest (with variation in the number of estimators); Linear Discriminant Analysis (which has no parameters to adjust); Support Vector Machine with 4 different kernels available (Linear, Polynomial, RBF and Sigmoid) and varying the regularization parameter; and the k-Nearest Neighbors, varying the number of neighbors. Furthermore, in each case three different splits were also tested – separation between training and test data: 70%, 80% and 90% of the training data.

Accuracy was used as the performance measure and statistical significance tests were performed between results.

The algorithm with the best results was Random Forest, which reached an accuracy of 96.425%, followed by the Support Vector Machine's RBF kernel, with 93.615%.

The results are comparable to other published works that present similar methodologies.

Keywords: Machine Learning, classification, brain tumor

Índice Geral

Agradecimentos.....	i
Resumo	iii
Abstract	iv
Índice Geral	v
Índice de Tabelas	ix
Índice de Gráficos.....	xii
Índice de Figuras	xiv
1 Introdução.....	1
2 Tumores Cerebrais	5
2.1. Epidemiologia das Neoplasias Cerebrais	5
2.2. Estrutura Encefálica	6
2.3. Neoplasias	8
2.3.1 Etiologia das neoplasias no Sistema Nervoso Central	10
2.4 Classificação e Incidência de Tumores do Sistema Nervoso Central	12
2.5. Diagnóstico de Tumores Cerebrais	14
2.5.1. Biópsia.....	15
2.5.2. Imagem por Ressonância Magnética.....	18
2.5.3 Tomografia por Emissão de Positrões (PET).....	20
3. <i>Data Science</i> em Diagnóstico de Tumores Cerebrais	25
3.1. Inteligência Artificial e Ciência dos Dados	25
3.2. <i>Machine Learning</i>	28
3.2.1 Supervisionado.....	29
3.2.2 Não-Supervisionado	30
3.3. Algoritmos de <i>Machine Learning</i>	30
3.3.1 <i>Decision Tree</i>	31
3.3.2 <i>Random Forest</i>	32
3.3.3 <i>Linear Discriminant Analysis (LDA)</i>	33
3.3.4 <i>Support Vector Machine</i>	34
3.3.5 <i>k – Nearest Neighbors</i>	37
3.4. <i>Machine Learning</i> em Diagnóstico de Tumores Cerebrais	38
3.4.1 Trabalhos relacionados	39
4. Materiais e Métodos.....	47

4.1.	<i>Dataset</i>	47
4.2.	Algoritmos	49
4.2.1	Decision Tree	50
4.2.2	Random Forest.....	50
4.2.3	Linear Discriminant Analysis.....	51
4.2.4	Support Vector Machine	51
4.2.5	k-Nearest Neighbors (kNN).....	52
4.3.	Plataforma e Código.....	52
4.4.	Métricas	53
5.	Resultados.....	57
5.1	<i>Decision Tree</i>	57
5.2	<i>Random Forest</i>	59
5.3	<i>Linear Discriminant Analysis</i>	61
5.4	<i>Support Vector Machine</i>	62
5.4.1	Linear	63
5.4.2	Radial Basis Function (RBF).....	64
5.4.3	Polinomial.....	66
5.4.4	k – Nearest Neighbors (kNN).....	70
6.	Discussão de Resultados.....	73
6.1	Algoritmos Individuais	74
6.1.1	Decision Tree	74
6.1.2	Random Forest.....	76
6.1.3	LDA	81
6.1.4	Support Vector Machine	82
6.1.4.1	Linear	83
6.1.4.2	RBF	84
6.1.4.3	Polinomial.....	85
6.1.4.4	Sigmoide	86
6.1.5	kNN	87
6.2	Entre Algoritmos.....	90
6.2.1	Dataset.....	90
6.2.2	<i>Split</i>	92
6.3	Entre <i>kernels</i> do SVM	94
6.4	Comparação com outros trabalhos.....	95
7.	Conclusão.....	101

8. Referências.....	107
Anexo A – Tabela de Classificação de Tumores do SNC da OMS (Louis, <i>et al.</i> , 2016)	119
Apêndice A – Código utilizado	121

Índice de Tabelas

Tabela 2.1 – Lista de classificações dos agentes carcinogénicos (IARC, 2021).	11
Tabela 2.2 – Incidência das principais neoplasias no SNC.	13
Tabela 2.3 – Frequência dos tumores primários do SNC.	14
Tabela 2.4 – Classe, nome e mecanismo dos radiotraçadores usados em PET.	22
Tabela 2.5 – Processos patológicos diferentes passagem de tempo da irradiação do SNC.	23
Tabela 3.1 – Datasets utilizados em Amin J. , Sharif, Raza, Saba, & Anjum, 2019.	40
Tabela 3.2 – Comparação entre os resultados obtidos por Pinto, Pereira, Rasteiro, & Silva e Amin, Sharif, Raza, Saba, & Anjum (2019).	42
Tabela 3.3 – Classificação e quantidade de imagens do dataset utilizado em Polly, Shil, Hoassain, Ayman, & Jang (2018).	42
Tabela 3.4 – Comparação da acurácia dos autores com outros trabalhos – traduzida e adaptada de Polly, Shil, Hoassain, Ayman, & Jang (2018).	44
Tabela 3.5 – Acurácia (Acc) e Área sob a curva (Auc) para cada kernel em cada dataset.	45
Tabela 3.6 – Comparação de resultados entre Amin, Sharif, Yasmin, & Fernandes (2017) com trabalhos semelhantes.	46
Tabela 4.1 – Distribuição das categorias dos datasets.	48
Tabela 5.1 – Resultados e medidas de dispersão para acurácia de Decision Tree.	58
Tabela 5.2 – Resultados e medidas de dispersão para acurácia de Random Forest – split 70%.	59
Tabela 5.3 – Resultados e medidas de dispersão para acurácia de Random Forest – split 80%.	60
Tabela 5.4 – Resultados e medidas de dispersão para acurácia de Random Forest – split 90%.	61
Tabela 5.5 – Resultados e medidas de dispersão para acurácia de Linear Discriminant Analysis.	62
Tabela 5.6 – Resultados e medidas de dispersão para acurácia do kernel Linear – split 70%.	63
Tabela 5.7 – Resultados e medidas de dispersão para acurácia do kernel RBF – split 70%.	64
Tabela 5.8 – Resultados e medidas de dispersão para acurácia do kernel RBF – split 80%.	65

Tabela 5.9 – Resultados e medidas de dispersão para acurácia do kernel RBF – split 90%.....	66
Tabela 5.10 – Resultados e medidas de dispersão para acurácia do kernel Polinomial – split 70%.....	67
Tabela 5.11 – Resultados e medidas de dispersão para acurácia do kernel Polinomial – split 80%.....	68
Tabela 5.12 – Resultados e medidas de dispersão para acurácia do kernel Polinomial – split 90%.....	69
Tabela 5.13 – Resultados e medidas de dispersão para acurácia do kNN com 3 vizinhos, para os três splits.....	70
Tabela 5.14 - Resultados e medidas de dispersão para acurácia do kNN com 5 vizinhos, para os três splits.....	71
Tabela 5.15 – Resultados e medidas de dispersão para acurácia do kNN com 10 vizinhos, para os três splits.....	72
Tabela 6.1 – Teste de significância estatística para os resultados do algoritmo Decision Tree.....	74
Tabela 6.2 – Percentuais de acerto do algoritmo Decision Tree por categoria nos Datasets 1 e 2.....	76
Tabela 6.3 – Testes de significância estatística para os resultados do algoritmo Random Forest.....	77
Tabela 6.4 – Percentuais de acerto do algoritmo Random Forest por categoria nos Datasets 1, 2 e 3.....	80
Tabela 6.5 – Testes de significância estatística para os resultados do algoritmo LDA.....	82
Tabela 6.6 – Testes de significância estatística para os resultados do kernel linear do SVM.....	83
Tabela 6.7 – Testes de significância estatística para os resultados do kernel RBF do SVM.....	84
Tabela 6.8 – Testes de significância estatística para os resultados do kernel Polinomial do SVM.....	85
Tabela 6.9 – Testes de significância estatística para os resultados do kernel Sigmoide do SVM.....	86
Tabela 6.10 – Testes de significância estatística para os resultados do algoritmo kNN.....	87
Tabela 6.11 – Percentuais de acerto do algoritmo k-Nearest Neighbors por categoria nos Datasets 1, 2 e 3.....	89
Tabela 6.12 – Testes de significância estatística do melhor desempenho em cada algoritmo quanto ao dataset.....	91

Tabela 6.13 – Sumário dos resultados quanto ao dataset por algoritmo.....	92
Tabela 6.14 – Testes de significância estatística do melhor desempenho em cada algoritmo quanto ao split.....	93
Tabela 6.15 – Sumário dos resultados quanto ao split por algoritmo	94
Tabela 6.16 – Testes de significância estatística do melhor desempenho em cada kernel do SVM quanto ao parâmetro de regularização C.	95
Tabela 6.17 – Sumarização dos resultados de Amin, Sharif, Raza, Saba, & Anjum (2019).....	96
Tabela 6.18 – Sumarização de resultados de alguns algoritmos.	96
Tabela 6.19 – Comparação dos resultados com Amin, Sharif, Raza, Saba, & Anjum (2019).....	97
Tabela 6.20 – Sumarização dos resultados de Amin, Sharif, Yasmin, & Fernandes (2017)	98
Tabela 6.21 – Sumarização dos resultados para os kernels linear, polinomial e RBF do SVM com $C = 1$, para os três datasets.....	99
Tabela 6.22 – Comparação com os resultados de Amin, Sharif, Yasmin e Fernandes (2017).....	99

Índice de Gráficos

Gráfico 2.1 – Incidência de cancro no SNC por 100 mil habitantes (IARC, 2020)	6
Gráfico 5.1 – Acurácia do algoritmo <i>Decision Tree</i> para os três <i>datasets</i> e os três <i>splits</i>	58
Gráfico 5.2 – Acurácia do algoritmo <i>Random Forest</i> para os três <i>datasets</i> no <i>split</i> 70%	59
Gráfico 5.3 – Acurácia do algoritmo <i>Random Forest</i> para os três <i>datasets</i> no <i>split</i> 80%	60
Gráfico 5.4 – Acurácia do algoritmo <i>Random Forest</i> para os três <i>datasets</i> no <i>split</i> 90%	61
Gráfico 5.5 – Acurácia do algoritmo <i>Linear Discriminant Analysis</i> para os três <i>datasets</i> e os três <i>splits</i>	62
Gráfico 5.6 – Acurácia do <i>kernel</i> linear para os <i>datasets</i> D1 e D3 no <i>split</i> 70% e quatro valores de parâmetro de regularização	63
Gráfico 5.7 – Acurácia do <i>kernel</i> RBF para os três <i>datasets</i> no <i>split</i> 70% e quatro valores de parâmetro de regularização	64
Gráfico 5.8 – Acurácia do <i>kernel</i> RBF para os três <i>datasets</i> no <i>split</i> 80% e quatro valores de parâmetro de regularização	65
Gráfico 5.9 – Acurácia do <i>kernel</i> RBF para os três <i>datasets</i> no <i>split</i> 90% e quatro valores de parâmetro de regularização	66
Gráfico 5.10 – Acurácia do <i>kernel</i> Polinomial para os três <i>datasets</i> no <i>split</i> 70% e quatro valores de parâmetro de regularização	67
Gráfico 5.11 – Acurácia do <i>kernel</i> Polinomial para os três <i>datasets</i> no <i>split</i> 80% e quatro valores de parâmetro de regularização	68
Gráfico 5.12 – Acurácia do <i>kernel</i> Polinomial para os três <i>datasets</i> no <i>split</i> 90% e quatro valores de parâmetro de regularização	69
Gráfico 5.13 – Acurácia do kNN para os três <i>datasets</i> e os três <i>splits</i> , para 3 vizinhos	70
Gráfico 5.14 – Acurácia do kNN para os três <i>datasets</i> e os três <i>splits</i> , para 5 vizinhos	71
Gráfico 5.15 – Acurácia do kNN para os três <i>datasets</i> e os três <i>splits</i> , para 10 vizinhos	72

Índice de Figuras

Figura 2.1 – Regiões do Encéfalo – Secção sagital, metade direita (VanPutte, et al., 2016).	7
Figura 2.2 – Categorias de Cancro (SEER Training Modules).	9
Figura 2.3 – Aparato para a realização da biópsia estereotáxica. (Trammit Medical, s.d.).	16
Figura 2.4 – Realização da biópsia cerebral estereotáxica (Andrada, 2013).	17
Figura 2.5 - Detalhes do procedimento da biópsia estereotáxica (DeVita, Jr, Lawrence, & Rosenberg, 2015).	17
Figura 2.6 – Astrocitoma de baixo grau em RM nas modalidades de T1 com contraste (esquerda) e FLAIR (direita) (DeVita, Jr, Lawrence, & Rosenberg, 2015).	19
Figura 2.7 - imagem de ressonância magnética multimodal do cérebro na ponderação em T1 (a), T2 (b), T1 com contraste (c) e FLAIR (d) (Usman & Rajpoot, 2017).	20
Figura 3.1 – Tamanho Anual da Datasfera Global (Reinsel , Gantz , & Rydning, 2017).	27
Figura 3.2 – Categorias de algoritmos de ML conforme a natureza dos dados de treino (El Naqa, Ruijang, & Murphy, 2015).	29
Figura 3.3 – Ilustração de uma árvore de decisão (Didática Tech, 2020).	31
Figura 3.4 – Esquematização do algoritmo <i>Random Forest</i> (Pawar, 2020).	32
Figura 3.5 – Ilustração da classificação do algoritmo LDA (Sickit Learn, 2021).	34
Figura 3.6 – Representação do Hiperplano com duas dimensões (a) e três dimensões (b) (Gandhi, 2018).	35
Figura 3.7 – Representação gráfica da aplicação do <i>kernel</i> linear (Awasthi, 2020).	36
Figura 3.8 – Representação gráfica da aplicação do kernel polinomial (Awasthi, 2020).	36
Figura 3.9 – Representação gráfica da aplicação do kernel RBF gaussiano (Awasthi, 2020).	37
Figura 3.10 – Representação gráfica da aplicação do kernel sigmoide (Awasthi, 2020).	37
Figura 3.11 – Representação do algoritmo kNN com 3 vizinhos (Samavedam, 2020).	38
Figura 3.12 – Imagem sem pré-processamento (a) e com remoção do crânio (b) (Amin, Sharif, Yasmin, & Fernandes, 2017).	45
Figura 4.1 – Imagem em tons de cinzento e sua representação em escala de intensidade (Albuquerque & Albuquerque, 2012).	48
Figura 4.2 – Imagens do <i>Dataset 2</i>	49

Figura 4.3 – Implementação do algoritmo de classificação <i>Decision Tree</i> .	50
Figura 4.4 – Implementação do algoritmo de classificação <i>Random Forest</i> com 5000 árvores.	51
Figura 4.5 – Implementação do algoritmo LDA.	51
Figura 4.6 – Implementação do algoritmo SVM com <i>kernel</i> linear.	51
Figura 4.7 – Implementação do algoritmo SVM com <i>kernel</i> polinomial.	52
Figura 4.8 – Implementação do algoritmo SVM com <i>kernel</i> RBF gaussiano.	52
Figura 4.9 – Implementação do algoritmo SVM com <i>kernel</i> sigmoide.	52
Figura 4.10 – Implementação do algoritmo kNN com 10 vizinhos.	52
Figura 4.11 – Interface da plataforma <i>Google Colaboratory</i> .	53
Figura 4.12 – Interface dos <i>outputs</i> do algoritmo <i>Decision Tree</i> no Google Colaboratory	54
Figura 4.13 – Exemplo de Matriz Confusão multiclasse.	55
Figura 6.1 – Matriz Confusão para o <i>Dataset 1</i> com o algoritmo <i>Decision Tree</i> e <i>split</i> de 90%.	75
Figura 6.2 – Matriz Confusão para o <i>Dataset 2</i> com o algoritmo <i>Decision Tree</i> e <i>split</i> de 90%.	75
Figura 6.3 – Matriz confusão do <i>Dataset 1</i> com o algoritmo <i>Random Forest</i> com 100 árvores e <i>split</i> de 90%.	78
Figura 6.4 – Matriz confusão do <i>Dataset 2</i> com o algoritmo <i>Random Forest</i> com 100 árvores e <i>split</i> de 90%.	79
Figura 6.5 – Matriz confusão do <i>Dataset 3</i> com o algoritmo <i>Random Forest</i> com 100 árvores e <i>split</i> de 90%.	79
Figura 6.6 – Localização dos principais tumores malignos do cérebro (Miami Neuroscience Center , s.d.).	80
Figura 6.7 – Meningiomas pertencente ao <i>Dataset 1</i> .	81
Figura 6.8 – Matriz confusão do <i>Dataset 1</i> com o algoritmo <i>k-Nearest Neighbors</i> com 5 vizinhos e <i>split</i> de 90%.	88
Figura 6.9 – Matriz confusão do <i>Dataset 2</i> com o algoritmo <i>k-Nearest Neighbors</i> com 5 vizinhos e <i>split</i> de 90%.	88
Figura 6.10 – Matriz confusão do <i>Dataset 3</i> com o algoritmo <i>k-Nearest Neighbors</i> com 5 vizinhos e <i>split</i> de 90%	89
Figura 7.1 – Resultados do método proposto para a detecção de tumores de Amin J. , Sharif, Yasmin, & Fernandes (2017).	105
Figura 7.2 – Resultados do método proposto por Amin J. , Sharif, Raza, Saba, & Anjum (2019).	105
Figura 7.3 – Glioma antes e depois do redimensionamento.	106

1 Introdução

O cancro é uma enfermidade que aflige a humanidade “desde quando sequer havia palavras para descrevê-la, e feitiços e encantamentos eram as armas que se tinha para lutar contra ela” (Oliveto, 2015). A mais antiga evidência de um tumor maligno em homínídeos foi encontrada num osso de um neandertal de 35 mil anos na região de Stetten, na Alemanha (Oliveto, 2015). Já o primeiro registo escrito de que se tem notícia é de um papiro egípcio datado de 2600 a.C., em que constavam o relato de 48 doenças diferentes, dentre as quais a doença 45 foi descrita como “massas salientes no peito e que se espalharam pelo peito, frias, duras e densas como uma fruta” (Reis, 2020). A medicina oncológica pouco evoluiu nos séculos seguintes, até à descoberta dos raios X por Röntgen, em 1895. Recorrendo ao Raio-X fez-se a primeira tentativa de tratamento no ano seguinte. Ao longo do século XX e do começo do século XXI, várias modalidades de imagens e tratamentos foram sendo desenvolvidos e aprimorados, e atualmente dispõe-se de uma gama ampla de opções para diagnosticar, tratar e acompanhar a doença.

A Imagem por Ressonância Magnética e a Tomografia Computadorizada são duas modalidades de imagem que permitem detetar até 95% dos casos de cancro (Hospital Sírio-Libanês, 2019), para além de exames como a biópsia. Estas modalidades de imagem, atualmente, fornecem aos profissionais de saúde dezenas, centenas e até milhares de imagens em um mesmo exame, como nas atuais modalidades dinâmicas

4D (Lee & Fujita, 2020). A tarefa de examinar tantas imagens de vários pacientes todos os dias é repetitiva, cansativa e monótona, o que pode levar a erros no diagnóstico das lesões. Por essa razão, cada vez mais se procura a automatização e a integração sistemática desse tipo de processo, que fica interligado com vários sistemas e interfaces, organiza operações, gera resultados muito precisos, em cada vez menos tempo, permitindo a realocação de recursos – inclusive humanos – de modo que se otimizem as suas tarefas (Estúdio ABC, 2017). Dessa forma, um médico poderia dedicar o tempo destinado à análise de muitas imagens de um mesmo paciente a tarefas que requerem o elemento humano, como o contacto e atenção maiores com o paciente, ou mesmo a atender mais pacientes com mais atenção.

Se hoje em dia se repete exaustivamente que “dados são o novo petróleo”, também já se sabe que todo o “petróleo” é inútil sem que haja refinamento e processamento. No caso dos dados, ferramentas como a *Machine Learning* – aprendizagem de máquina – têm sido amplamente discutidas, utilizadas e desenvolvidas para utilização em todos os campos possíveis, desde *marketing*, passando pela indústria e até na saúde. De maneira simples, *Machine Learning* refere-se a um conjunto de processos em que o computador aprende uma determinada tarefa com dados que são separados para treino, e depois o seu desempenho é posto à prova com os dados de teste.

Trabalhos de investigação (Amien, Adb-elrehman, & Ibrahim (2013); Amin, Sharif, Yasmin & Fernandes (2017); Zacharaki, *et al.* (2009); dentre outros) têm mostrado resultados bastante promissores quanto à tarefa de processar e melhorar imagens médicas com o intuito de ser um apoio confiável ao diagnóstico de cancro cerebrais. O desenvolvimento e a investigação nessa área permitirá a concretização do cenário descrito no parágrafo anterior: tarefas que máquinas são capazes de realizar serão a elas delegadas, e as tarefas inerente e insubstituivelmente humanas – como um contacto mais próximo entre paciente e médico – ficarão para nós.

Estudantes, professores e investigadores são os principais responsáveis pela pesquisa e inovação nos mais diversos ramos da ciência, de modo que é salutar o incentivo à investigação em temas atuais e relevantes à sociedade como um todo.

À vista disso, esta dissertação de mestrado propõe-se a testar vários algoritmos de *Machine Learning* e respetivos parâmetros em imagens de tumores cerebrais, com o intuito de classificá-las quanto ao tipo de tumor presente.

A divisão do trabalho está organizada da seguinte maneira: o Capítulo 2 é dedicado à explanação sobre tumores cerebrais, começando com um resumo sobre epidemiologia

dos tumores, a anatomia do cérebro, depois as neoplasias e sua classificação e diagnóstico. No Capítulo 3 encontram-se explicações sobre inteligência artificial e *Data Science*, seu uso em Medicina e o funcionamento de cada um dos algoritmos usados. O Capítulo 4 é dedicado aos materiais e métodos utilizados para o desenvolvimento deste trabalho: a plataforma utilizada para os algoritmos, os parâmetros e *datasets*. O Capítulo 5 trata dos resultados obtidos, divididos por tipo de algoritmo. O Capítulo 6 apresenta a discussão dos resultados, testes de significância estatística e comparação com outros dois autores. Por fim, o Capítulo 7 é a conclusão do trabalho.

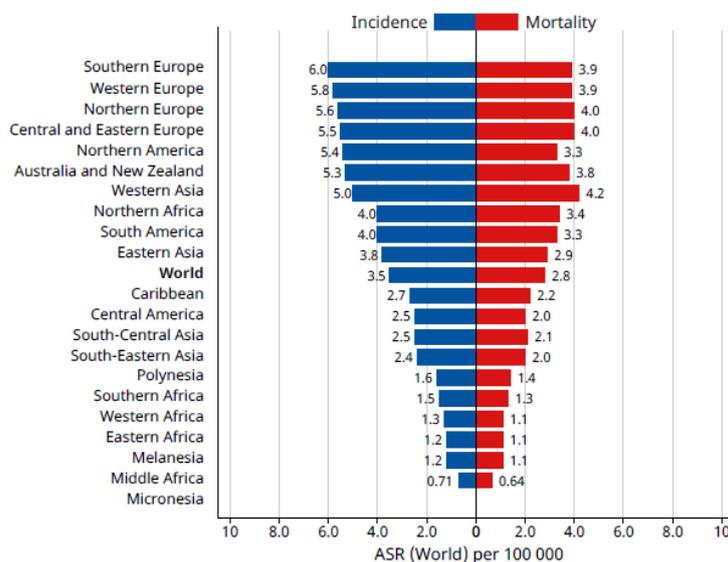
2 Tumores Cerebrais

2.1. Epidemiologia das Neoplasias Cerebrais

Segundo a Organização Mundial da Saúde, cerca de uma em cada seis mortes a nível global estão relacionadas ao cancro, e a incidência da doença deve aumentar em 63% até 2040, atingindo a marca de 29 milhões de casos anuais (IARC, 2018). Os cancros no Sistema Nervoso Central (SNC) correspondem a cerca de 1,8% desse total em adultos, e dentre esses, os cerebrais são cerca de 88% (INCA, 2018). A incidência em crianças é muito maior: 23% dos cancros diagnosticados são no SNC (Crocetti, et al., 2012). Um diagnóstico de cancro é sempre uma situação muito delicada e intimidante para o paciente e, embora a incidência no SNC seja relativamente baixa, quando estamos a lidar com o cérebro, o assunto torna-se ainda mais sério, dada a natureza e importância da região, e os índices de mortalidade. Em 2020, foram registados no mundo 308.102 casos de cancro no cérebro e SNC, com uma mortalidade de 251.329 (IARC, 2020), o que corresponde a assustadores 81,6%. A Europa, embora represente cerca de 9,5% da população mundial (Countymeters, 2021), apresenta 21,8% dos casos de cancro no cérebro e SNC, e 21,4% das mortes (IARC, 2020), o que deixa o continente com a maior taxa da patologia por habitante do mundo. Ao considerar os países da Europa Meridional, sub-região que compreende atualmente Portugal e mais 18 países – Albânia, Andorra, Bósnia e Herzegovina, Cidade do Vaticano, Chipre, Croácia, Gibraltar, Grécia, Eslovénia, Espanha, Itália, Kosovo, Macedónia, Malta, Montenegro, San Marino, Sérvia e Turquia (Southern European Countries 2021, s.d.) – tem-se a

maior taxa de incidência por 100 mil habitantes do mundo, conforme se vê no Gráfico 2.1.

Gráfico 2.1 –Incidência de cancro no SNC por 100 mil habitantes (IARC, 2020).



2.2. Estrutura Encefálica

O SNC é compreendido pela medula espinal e pelo encéfalo. Ambos são envolvidos pelas meninges, que são membranas protetoras, e pelo líquido cefalorraquidiano (Matthews, et al., 2020). O encéfalo é a parte protegida pelo crânio e formada pelo tronco encefálico, cerebelo, diencéfalo e cérebro, constituindo o “principal sítio para processamento de informação, iniciação das respostas e integração de processos mentais” (VanPutte, et al., 2016) e, por essa razão, é muitas vezes comparado a um computador.

O encéfalo humano pesa cerca de 1,5 kg em adultos, e é formado por milhares de milhões de células, sendo que o SNC tem a maior diversidade de células dos sistemas do organismo humano: ao passo que o coração e o fígado possuem, por exemplo, entre 3 a 5 tipos de células diferentes, o encéfalo possui mais de uma centena delas (Gomes, Tortelli, & Diniz, 2013), dentre as quais os gliócitos (células da glia) e os neurónios, sendo que estes últimos são mais conhecidos. As sinapses ocorrem entre os neurónios, e permitem que estímulos elétricos e químicos sejam transmitidos de um neurónio a outro, processo por meio do qual as atividades cognitivas e motoras são executadas (Britannica, The Editors of Encyclopaedia, 2020). Antigamente, pensava-se que a

função das células gliais era meramente estrutural, servindo como “cola” (de onde vem o nome glia) para os neurónios (Weaver II & Doyle, 2019). No entanto, a função da glia também está ativamente relacionada à sinalização cerebral. As células da glia são divididas em dois grandes grupos com características morfológicas e funcionais diferentes, conforme a origem embrionária: a microglia (origem mesodérmica) e a macroglia (ectodérmica). As micróglias têm terminações com que se unem aos neurónios, em busca de sinapses mortas ou mesmo neurónios que estejam a morrer e que precisam ser eliminados, ou de agentes externos que possam causar dano. Ou seja, elas exercem função imunológica no sistema nervoso (Fapesp, 2017). A macroglia é compreendida pela oligodendroglia – que compreende os oligodendrócitos e é responsável pela mielinização dos axónios; pela ependimoglia – compreendida pelos ependimócitos, células epiteliais pigmentares da retina e células do plexo coroide; e a astroglia, grupo que inclui os astrócitos, que são a maior parte das células da glia e que são a principal fonte de fatores de crescimento para neurónios, dentre outras funções mais específicas, como guiar a migração neuronal durante o desenvolvimento das camadas do Sistema Nervoso.

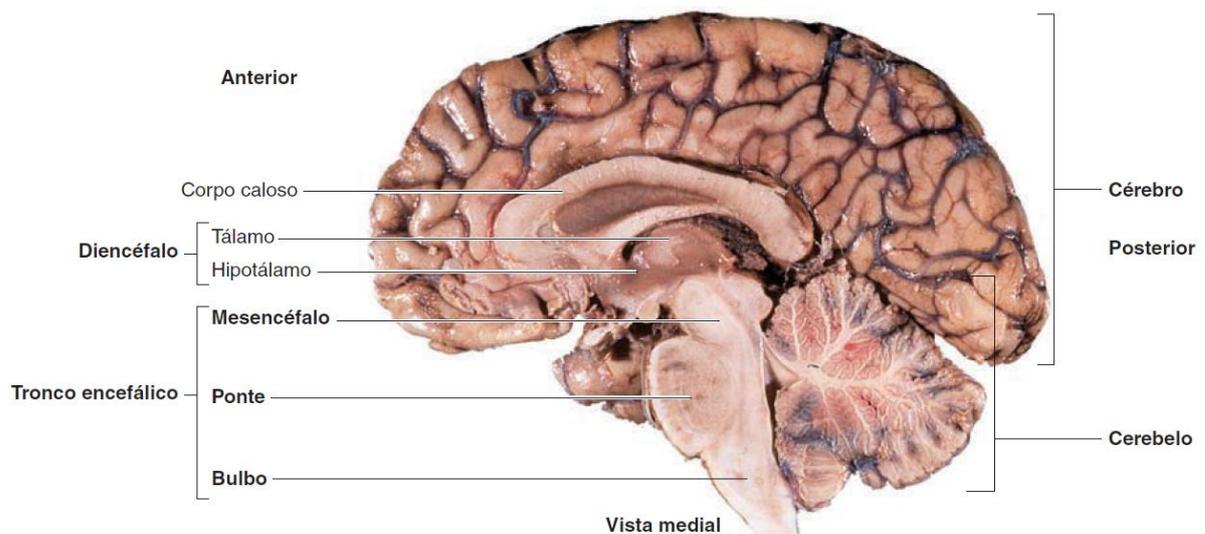


Figura 2.1 – Regiões do Encéfalo – Secção sagital, metade direita (VanPutte, et al., 2016).

Cada região ilustrada na Figura 2.1 é responsável por uma ou mais funções, que podem ser motoras, sensoriais, homeostáticas, mentais de transmissão ou comunicação, dentre várias outras.

2.3. Neoplasias

As células humanas são constituídas por três partes principais: a membrana celular (parte mais externa, que traça os limites do corpo da célula e faz trocas com o meio exterior), o citoplasma (corpo da célula, onde se encontram as organelas celulares) e o núcleo (parte que contém os cromossomos). Os cromossomos são compostos por genes, que são segmentos das moléculas de DNA, por meio dos quais as informações para o funcionamento geral das células são enviadas (INCA, 2021). Por vezes, ocorrem mutações nesses genes, que podem ser resultado de erros de cópia aquando da divisão celular, por exposição a agentes químicos mutagénicos ou a radiações ionizantes, dieta, tabaco, ou mesmo vírus (Oncoguia, 2015). A própria célula também pode incorrer em mutações deliberadamente, o que é chamado de hipermutação (Vassallo). As mutações mudam as informações que as células recebem, que passam a desempenhar determinada tarefa de maneira diferente dependendo do gene em que ocorrem, podendo produzir resultados benéficos, maléficos, ou mesmo um resultado não perceptível (Oncoguia, 2015). O cancro surge a partir de uma mutação nos proto-oncogenes, que em células saudáveis são inativos. Esses proto-oncogenes, quando ativados, transformam as células normais em cancerosas. As células cancerosas são aquelas com alterações nas suas instruções sobre seu crescimento, reprodução, funcionamento e morte (apoptose). Por isso, elas multiplicam-se de forma descontrolada e desordenada, bem mais rapidamente que as células normais. Ao acúmulo dessas células se dá o nome de tumor. Aos poucos, por meio dos vasos sanguíneos ou linfáticos, ocorre a invasão de todo o organismo por essas células. Por serem menos especializadas do que as células correspondentes saudáveis, as células cancerosas não desempenham corretamente sua função no órgão ou tecido em que estão e, com a sua gradual substituição, o órgão como um todo passa a perder sua função (INCA, 2020).

Uma massa tumoral (qualquer aumento de volume limitado em um determinado órgão ou área do organismo) é chamada de neoplasia (do grego *néos*, “novo”, e *plásis*, “formação – dicionário Infopédia da Língua Portuguesa da Porto Editora). Quando essa massa não é invasiva – algumas células perdem o controle do crescimento mas continuam reproduzindo as células originais e não se espalham para outros lugares do corpo – o tumor é dito *benigno*, e quando ele possui essa capacidade de invasão é *maligno* (Alberts, et al., 2008) – algumas células do órgão alteram totalmente suas características morfológicas e funcionais, crescem descontroladamente em tamanho e forma e infiltram tecidos vizinhos, inclusive vasos sanguíneos e linfáticos levando essas células doentes para outros órgãos e não mais respondem à apoptose (morte celular

programada). No primeiro caso, normalmente, a simples remoção ou ablação do tumor já é suficiente para curar o paciente, ao passo que no segundo caso a remissão costuma ser mais onerosa e demorada (A F. Hoffmann–La Roche AG, 2019).

Ao que genericamente chama-se “cancro”, na verdade é um termo que abrange mais de 100 doenças diferentes, que podem ou não dar origem a massas tumorais, que são agrupadas em seis categorias histológicas diferentes (*SEER Training Modules*), conforme pode-se ver resumido na Figura 2.2.

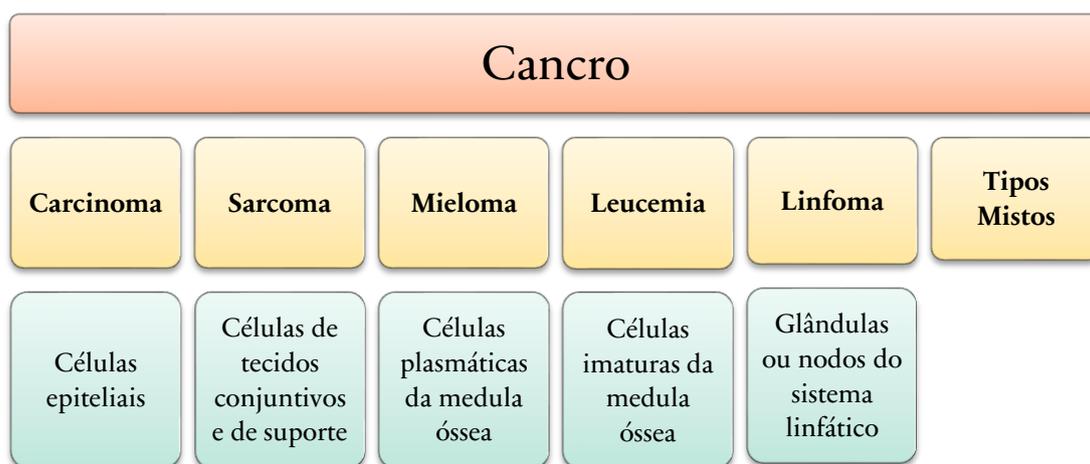


Figura 2.2 – Categorias de Cancro (*SEER Training Modules*).

Embora os profissionais da saúde tendam a nomear as neoplasias com base nos tipos histológicos citados acima, o público em geral encontra-se mais familiarizado com a nomenclatura baseada nos sítios primários, isto é, o local anatómico onde o cancro se originou (*SEER Training Modules*). Dentre os sítios mais comuns se encontram a pele, os pulmões, mamas, próstata, cólon e reto, cérvix e útero.

Essa nomenclatura baseada no sítio do tumor primário, embora mais intuitiva para os pacientes e público leigo, não é tão acurada quanto a baseada no tipo histológico. Os cancros de pele, por exemplo, são divididos em três grupos principais: os de células basais (carcinoma basocelular), de células escamosas (carcinoma espinocelular) e melanomas (*Sociedade Brasileira de Dermatologia, 2017*). Segundo o cirurgião oncologista Doutor Henrique Reis, especializado em cabeça e pescoço, “embora possam ser informalmente denominados da mesma forma (cancro de pele ou no cérebro, por exemplo), as abordagens podem ser completamente diferentes. O tipo de cirurgia, qual o tipo de tratamento radioterápico ou quimioterápico, o número de

sessões, outras abordagens como imunoterapia (ou alguns tratamentos concomitantes), tudo isso pode variar consoante o caso em específico, baseado em protocolos muito bem definidos internacionalmente.”

Por essa razão, é salutar que o diagnóstico seja feito da maneira mais precisa possível e, para isso, é importante que se disponha de equipamentos e materiais que não somente supram essas necessidades, mas que também facilitem e agilizem o trabalho dos especialistas.

2.3.1 Etiologia das neoplasias no Sistema Nervoso Central

Dentre os fatores etiológicos, para os tumores do SNC, não há uma identificação definitiva de um agente causador externo, sendo que fatores de risco são identificados em uma minoria dos casos. Os fatores mais comuns de malignidade em outras zonas anatómicas – nomeadamente dieta, álcool, sedentarismo, tabaco e vírus – não são considerados significativos quando se trata de tumores no SNC (DeVita, Jr, Lawrence, & Rosenberg, 2015).

De acordo com Ohgaki & Kleihues (2005), o único fator ambiental que é inequivocamente associado a um risco aumentado de tumor cerebral é a irradiação terapêutica com raios X e exposição a radiações ionizantes. Em particular, crianças irradiadas no tratamento de leucemia linfoblástica aguda “apresentam um risco significativamente elevado de desenvolver gliomas e tumores neuroectodermis primitivos, frequentemente dentro de 10 anos após a terapia” (Ohgaki & Kleihues, 2005). Há um aumento do risco de gliomas, meningiomas e neurinomas e tumores da hipófise em profissionais que lidam com radiações ionizantes, pacientes submetidos à radioterapia, inclusive para outros tumores; pacientes submetidos a radiografia, radioscopia e tomografia em grande quantidade, pessoas expostas a radiações causadas por acidentes nucleares ou uso de armas nucleares, pessoas com deficiências imunitárias - hereditárias, primárias, adquiridas (AIDS) ou relacionadas ao uso de drogas/medicamentos (imunossupressores, quimioterápicos ou outros) e por repercussão sistêmica da presença de outro cancro, infecções virais – alguns vírus aumentam o risco de desenvolver cancro do SNC, como o HIV, que aumenta o risco de Linfoma primário de SNC – outros vírus têm sido estudados como fatores de risco, mas até agora o resultado é inconclusivo (INCA, 2018).

Dentre outros fatores ambientais, estudos epidemiológicos indicam o aumento da incidência de tumores cerebrais para certos tipos de ocupação sem que, no entanto, houvesse a identificação de um agente ambiental que fosse incontestavelmente carcinogénico (Ohgaki & Kleihues, 2005). Algumas dessas profissões são médicos, bombeiros e agricultores (Carozza, et al., (2000) e De Ross, et al., (2003)).

Em relação à predisposição genética hereditária, os fatores de risco são: neurofibromatose tipos 1 e 2, esclerose tuberosa, retinoblastoma, síndrome de Li-Fraumeni, síndrome de Von-Hippel Lindau, síndrome de Turcot, síndrome de Gorlin e síndrome de Cowden (INCA, 2018).

O uso de telemóveis e outros aparelhos eletrónicos muitas vezes é apontado como tendo relação com tumores cerebrais, mas como o seu amplo uso ainda é recente, não é possível fazer qualquer análise de longo termo. A Comissão Internacional de Proteção contra Radiação Não-Ionizante classificou os campos eletromagnéticos do âmbito das radiofrequências (no qual se localizam energeticamente as radiações emitidas pelos aparelhos de telefonia móvel) na categoria 2B, isto é, “Possivelmente” carcinogénico para humanos (ICNIRP, 1998). As categorias são indicadas na Tabela 2.1, apresentada abaixo.

Tabela 2.1 – Lista de classificações dos agentes carcinogénicos (IARC, 2021).

Classificação	Definição	Agentes
1	Carcinogénico a humanos	121
2A	Provavelmente Carcinogénico a humanos	89
2B	Possivelmente Carcinogénico a humanos	319
3	Não classificável quanto à carcinogenicidade	500

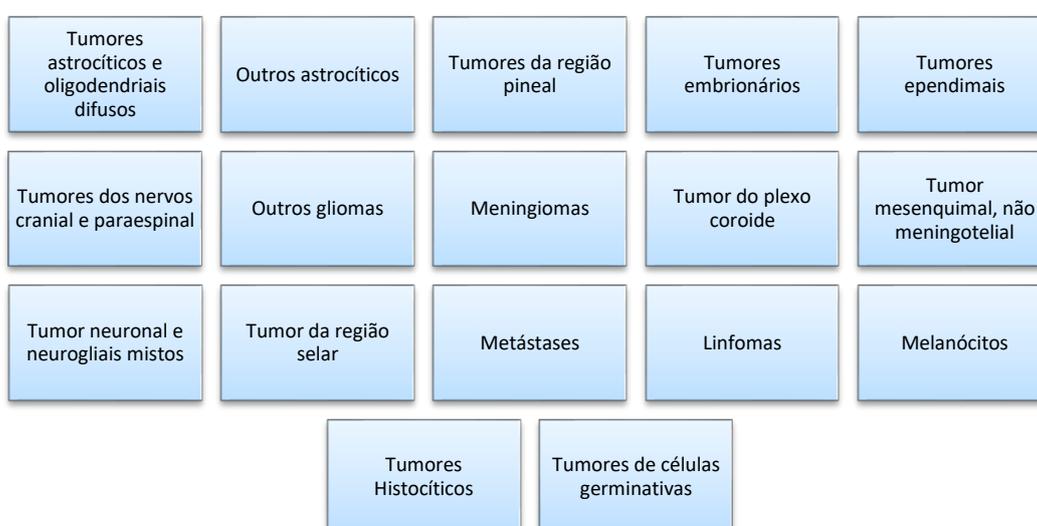
As radiofrequências encontram-se na categoria 2, que é descrita pelo documento como “inclui agentes com uma gama de evidências para carcinogénese em humanos e em animais experimentais”, e a subcategoria B são “agentes para o qual evidência em humanos não está disponível, mas para os quais é evidência suficiente de carcinogénese em animais experimentais” (tradução livre).

Essas categorias, segundo um documento explicativo sobre o relatório (IARC, 2019), indicam a força da evidência de que o agente é capaz de causar cancro mas, conforme é enfatizado, não mede a probabilidade de que de facto a doença se manifeste (tecnicamente, o risco) em razão de um certo nível de exposição ao agente.

2.4 Classificação e Incidência de Tumores do Sistema Nervoso Central

A última publicação da Organização Mundial da Saúde (OMS) a respeito da Classificação de Tumores do Sistema Nervoso Central aconteceu em 2016, estabelecendo renovações conceituais e práticas aos critérios de classificação anteriores – a publicação anterior havia sido em 2007. Nesse documento renovado (Louis, et al., 2016), a OMS usa parâmetros moleculares em acréscimo aos histológicos, “formulando um conceito sobre como os diagnósticos de tumores do SNC devem ser estruturados na era molecular”. Assim, o documento “*The 2016 World Health Organization Classification of tumors of Central Nervous System*” (Louis, et al., 2016) classifica os tumores do Sistema Nervoso Central nas categorias principais sumarizadas no Quadro 2.1.

Quadro 2.1 – Classificação de Tumores do SNC da OMS (Louis, et al., 2016).



Cada uma dessas categorias principais ainda é formada por subcategorias de tumores, relativos a grau e especificidade, que podem ir de I a IV, com base na observação das células em microscópio (*American Cancer Society*, 2020):

- Graus I e II: tendem a crescer lentamente e são menos propensos a infiltrarem-se em tecidos adjacentes;
- Graus III e IV: tendem a crescer rapidamente e têm maior probabilidade de se infiltrarem em tecidos adjacentes.

No Anexo A – Tabela de Classificação de Tumores do SNC da OMS – encontra-se a tabela de classificação elaborada pela OMS em 2016.

As Tabelas 2.2 e 2.3 indicam as distribuições de frequências das principais neoplasias do sistema nervoso central pelo tipo histopatológico, obtido por meio da base de dados Registro Central de Tumores Cerebrais dos Estados Unidos. A Tabela 2.2 traz a incidência de maneira geral (tumores primários e secundários) e a Tabela 2.3 em relação somente aos tumores primários. Ambas foram traduzidas e adaptadas de DeVita, Jr, Lawrence, & Rosenberg (2015).

Tabela 2.2 – Incidência das principais neoplasias no SNC.

Tipo	Incidência
Meningioma	30,1%
Glioblastoma	20,3%
Astrocitomas	9,8%
Glândula Pituitária	6,3%
Outros	33,5%

Tabela 2.3 – Frequência dos tumores primários do SNC.

Tipo	Incidência
Glioblastoma Multiforme	40,6%
Astrocitomas infiltrante	42,4%
Astrocitomas pilocísticos	1,5%
Oligodendroglioma	3,5%
Oligoastrocitomas mistos	1,7%
Ependimoma	3%
Meduloblastoma	3,6%
Meningioma	2%
Outros	1,9%

2.5. Diagnóstico de Tumores Cerebrais

O diagnóstico de cancro inclui intentar identificar qual o local de origem da malignidade e os tipos de células envolvidas (SEER Training Modules). Normalmente, os sintomas de um cancro na região do encéfalo se apresentam de maneira gradual, podendo, porém, se apresentar de maneira repentina (Hospital Sírio-Libanês, 2019), e podem ser generalizados ou específicos. Os generalizados são sintomas que estão relacionados principalmente com o fator expansivo e compressivo do tumor sobre as estruturas que, por estarem confinadas aos limites do crânio, não têm como se deslocar (Lusíadas, 2020). Outro fator importante é o inchaço no cérebro devido a bloqueios no fluxo do líquido cefalorraquidiano. Nesse grupo estão incluídos dores de cabeça, náuseas, vômitos, visão embaçada, problemas no equilíbrio, alteração de comportamento e, em casos avançados, letargia ou mesmo coma. Os específicos estão mais relacionados com o controlo das funções desempenhadas pelas áreas afetadas, e incluem ataques epiléticos, convulsões, fraqueza, sensações incomuns ou dormência em partes do corpo, perda de audição, problemas de equilíbrio, fraqueza de músculos faciais ou dificuldade para engolir, perda da visão, comprometimento da coordenação motora e dificuldade em executar tarefas cotidianas, dificuldades em falar ou compreender o que é falado, alterações na personalidade ou raciocínio, movimentos anormais e perda do controlo da bexiga ou intestino (Hospital Sírio-Libanês, 2019).

Em princípio, exames físicos e neurológicos são realizados com o intuito de descobrir a causa das anomalias fisiológicas e motoras. Caso a suspeita seja de um tumor no cérebro, a realização de exames imagiológicos torna-se necessária para a confirmação.

Os mais comuns são a Ressonância Magnética (RM) e a Tomografia Computadorizada (TC), que permitem detetar até 95% dos casos. Dentre exames imagiológicos complementares se encontram a espectroscopia por ressonância magnética, a tomografia por emissão de positrões (PET) e angiografia. (Hospital Sírio-Libanês, 2019). A seguir à imagem por Ressonância Magnética (RM) para a análise da localização e morfologia do tumor, é feita a biópsia para avaliação da malignidade da neoplasia (Zorzi, s.d.).

Em conversa com o Dr. Henrique Reis, foi dito que a punção lombar faz parte da rotina e muitas vezes também se faz necessária para a verificação da existência de células malignas no líquido cefalorraquidiano, pois isso evidencia a disseminação das células neoplásicas, ou mesmo a presença dessas células em casos de cancro sem massa tumoral, como no caso da neuroleucemia, meningite carcinomatosa ou falsos positivos. Em razão de a maioria dos tumores cerebrais em adultos ser secundária, é recomendada ainda a realização de exames de imagem na região do tórax para verificar a existência de tumores primários em outros órgãos.

2.5.1. Biópsia

Ao ser evidenciada uma massa cerebral pelos exames imagiológicos, o neurocirurgião fará o planeamento para a abordagem cirúrgica. É no momento da operação (intraoperatório) que se decidirá pela ressecção total ou parcial da massa, nesse último caso para se realizar o exame histopatológico (biópsia).

A biópsia é a análise laboratorial de uma pequena amostra de tecido ou células para a determinação de patologias, dentre elas o cancro (American Cancer Society, 2015). A técnica a ser utilizada para a retirada da amostra depende da localização do tumor e da natureza da lesão – neoplásica ou infecciosa, por exemplo. Para as neoplasias cerebrais, os principais tipos de biópsias realizadas são a estereotáxica e a cirúrgica (American Cancer Society, 2020).

A biópsia cirúrgica, também chamada de craniotomia, é um procedimento cirúrgico invasivo, que tem por objetivo não somente retirar amostras de tecido, mas simultaneamente fazer a ressecção de toda massa tumoral, ou pelo menos da maior parte, usando apenas o bisturi, dispositivos de sucção simples ou a vácuo, dependendo do caso (American Cancer Society, 2020). Por ser um procedimento invasivo, há riscos inerentes à técnica, os quais incluem até mesmo o óbito do paciente. Em casos de cirurgias cranianas eletivas, as complicações neurológicas mais comuns no pós-

operatório são diminuição do nível de consciência, vasoespasma cerebral, convulsões, hematoma intraparenquimatoso, hemiparesia e reoperação, com um índice global de mortalidade de 5,5%. Para o caso de cirurgias não eletivas – de emergência – os sintomas do pós-operatório podem incluir hipertensão intracraniana, déficits motores, hematoma subdural e hemorragia intraparenquimatosa, com uma taxa de mortalidade de cerca de 26% (Siqueira & Diccini, 2017).

A biópsia estereotáxica – ilustrada nas Figuras 2.3, 2.4 e 2.5 é uma técnica minimamente invasiva (embora ainda seja invasiva) que recorre à localização do alvo tumoral por meio de coordenadas tridimensionais (cartesianas ou polares), que é utilizada em casos em que os riscos de uma cirurgia são muito altos, para o implante de isótopos radioativos para tratamento por braquiterapia ou para a execução de radiocirurgia, e representam cerca de 20% das intervenções neurocirúrgicas (Pittella, 2008). O procedimento é realizado anestesiando-se localmente (para o caso de adultos) ou de forma geral (para pacientes pediátricos), e então uma estrutura rígida é usada para manter o posicionamento correto. É feita uma incisão no couro cabeludo, na qual se insere uma cânula ou sonda (que normalmente tem 1,4 a 2 mm de diâmetro) até que se atinja o alvo. A intervenção é guiada por ressonância magnética ou tomografia computadorizada para posicionar precisamente a agulha que recolherá a amostra de tecido (American Cancer Society, 2020). A depender do tipo de lesão, podem ser removidos vários fragmentos, extraídos a diferentes distâncias do centro do alvo tumoral (Pittella, 2008). Normalmente, o tempo de internamento é de 24 horas. Os riscos também dependem da localização do tumor, se a lesão é profunda ou situada próxima à linha média, quando podem se infiltrar ou já se infiltraram em áreas eloquentes ou estruturas vitais, como o córtex cerebral (American Cancer Society, 2020), (Pittella, 2008).



Figura 2.3 – Aparato para a realização da biópsia estereotáxica. (Trammit Medical, s.d.).



Figura 2.4 – Realização da biópsia cerebral estereotáxica (Andrada, 2013).

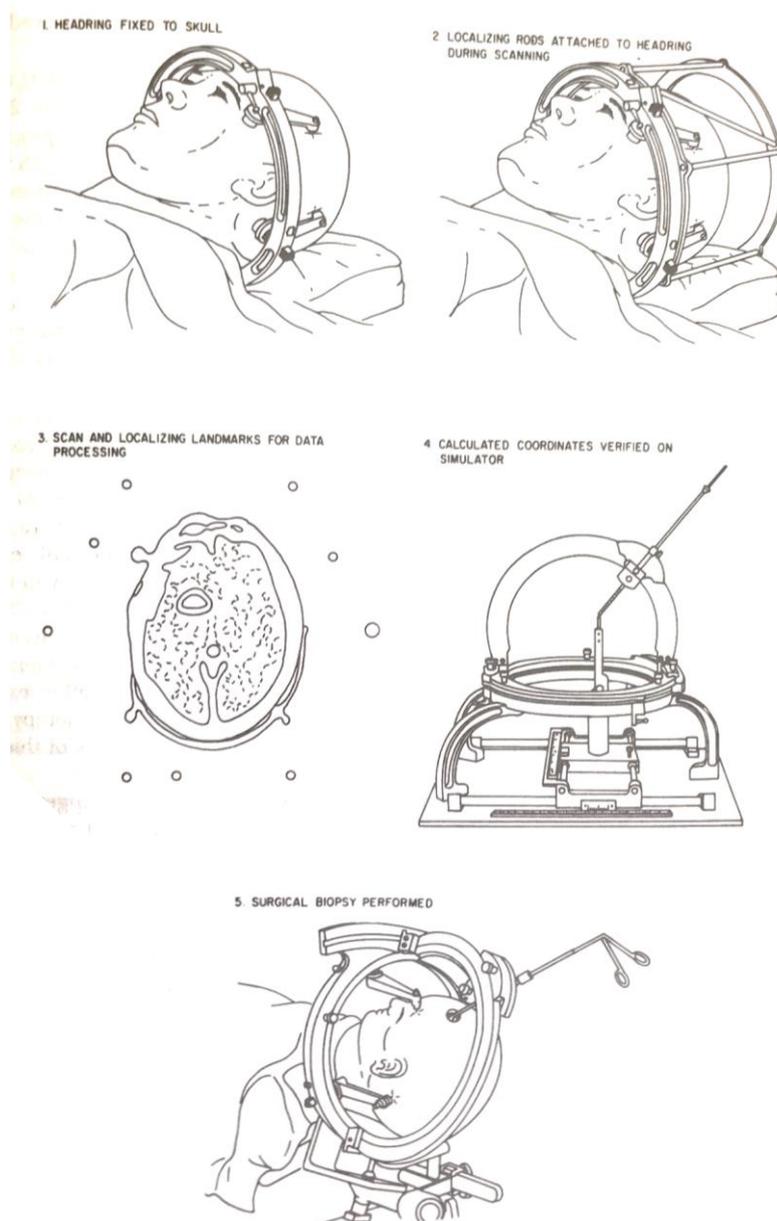


Figura 2.5 - Detalhes do procedimento da biópsia estereotáxica (DeVita, Jr, Lawrence, & Rosenberg, 2015).

Apesar de ser considerado um procedimento seguro (principalmente em comparação com a biópsia cirúrgica), há riscos inerentes ao processo. Os índices de complicações gerais variam muito conforme o estudo, encontrando-se entre 1,1% a 12,1% do número global de procedimentos. Dentre as principais complicações se encontram hemorragias subaracnoídea, intracerebral e intraventricular (1,7% a 9,2%) e traumas decorrentes da biópsia. Deteriorações neurológicas transitórias ocorrem em 1,3% a 10% dos pacientes e as prolongadas ou permanentes em 0,6% a 6,5%, e sangramentos diminutos ou assintomáticos em 1,5% a 9,9% dos pacientes. Um número maior de tentativas na obtenção das biópsias também são fatores de risco significativos para o surgimento de complicações como hemorragias ou edemas (Pittella, 2008). Outro fator de risco é o relacionado à habilidade e experiência do neurocirurgião.

Além desses riscos inerentes a processos invasivos, a biópsia e a punção lombar são procedimentos dolorosos para o paciente e requerem muito tempo e recursos (Amin, Sharif, Yasmin, & Fernandes, 2017). Outros riscos se referem a problemas de acurácia na caracterização do tumor, uma vez que usualmente o tecido neoplásico é heterogêneo, apresentando por vezes mais de um grau de malignidade, ou mesmo se sobrepondo a tecidos sadios, erro na amostragem ou variações na interpretação (Zacharaki, et al., 2009).

2.5.2. Imagem por Ressonância Magnética

A neuroimagem desempenha um papel essencial no diagnóstico e planejamento do tratamento para tumores cerebrais. Ela pode ser dividida em três amplas categorias, conforme a questão clínica em específico: 1) Diagnóstico; 2) Planejamento pré-operatório ou de terapia; 3) Avaliação pós-tratamento. A Imagem por Ressonância Magnética (RM) com e sem agente de contraste é normalmente o teste imagiológico de escolha (Cha, 2009). A Tomografia Computadorizada (TC) é mais utilizada para os casos em que o paciente não pode (por causa de implante de *pacemaker*, fragmentos de metal ou cliques cirúrgicos paramagnéticos) ou não quer (por causa de claustrofobia, por exemplo) fazer a RM.

A imagem por RM é uma das melhores técnicas para diagnóstico de patologias do cérebro, não somente por não fazer uso da radiação ionizante, mas também pela sua capacidade de fornecer informações detalhadas sobre as estruturas e função cerebral, podendo ser utilizada tanto na fase pré-tratamento quanto no acompanhamento e

seguimento (Mohsen, El-Dahshan, El-Horbaty, & Salem, 2017). Pela possibilidade de se fazerem imagens anatômicas com diversos tipos de contraste ou mesmo imagens funcionais, o que aumenta muito o detalhe com que se consegue visualizar as estruturas e função cerebral. Dentre essas sequências, encontram-se as imagens ponderadas em T1, em T2, em T1 com contraste de gadolínio e FLAIR (*fluid-attenuation inversion recovery*) (Usman & Rajpoot, 2017).

Normalmente, as imagens sagitais ponderadas em T1 com contraste são usadas para a classificação de tumores primários como astrocitomas, glioblastoma multiforme (GBM), meduloblastoma e meningioma. Já as axiais ponderadas em T2 são usadas para identificar tumores secundários, que são as metástases (Sachdeva, Kumar, Gupta, Khandelwal, & Ahuja, 2013). O contraste melhora a capacidade de discernimento dos tumores de outras entidades patológicas, um tipo de tumor de outro ou mesmo diferenciação de graus de tumores. Algumas outras patologias podem, no entanto, ser confundidas com tumores nos exames de RM anatômica, como abscessos bacteriais, doenças inflamatórias como sarcoidose, lesão tumefativa desmielinizante e isquemia aguda (DeVita, Jr, Lawrence, & Rosenberg, 2015).

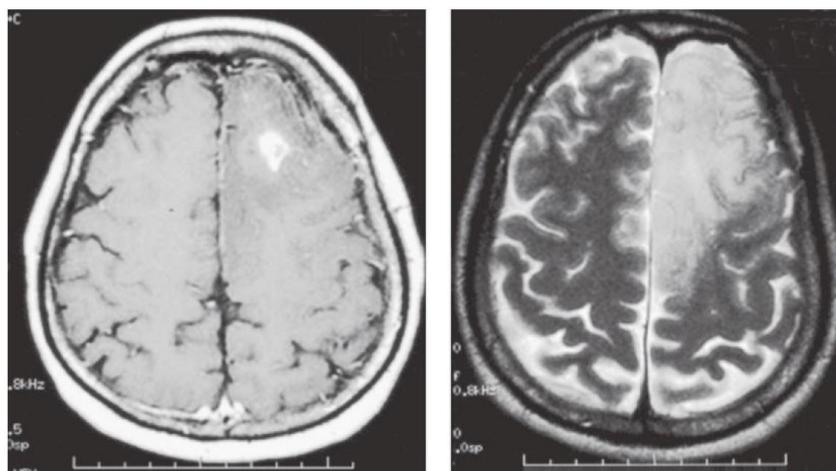


Figura 2.6 – Astrocitoma de baixo grau em RM nas modalidades de T1 com contraste (esquerda) e FLAIR (direita) (DeVita, Jr, Lawrence, & Rosenberg, 2015).

Na Figura 2.6 pode-se constatar a diferença na aplicação da sequência de RM correta: normalmente astrocitomas de baixo grau são difíceis de se evidenciar, e na ponderação em T1 com contraste a extensão infiltrativa da neoplasia é consideravelmente subestimada. Já a técnica FLAIR (*fluid-attenuated inversion recovery*) mostra-se bastante mais útil nesse tipo de neoplasia, conforme a Figura 2.6.

Algumas técnicas novas de RM incluem a Espectroscopia por Ressonância Magnética, RM com contraste dinâmico, RM com perfusão e difusão e RM funcional, que permitem não somente a distinção mais clara entre tumores e outras patologias como as citadas acima, como também a diferenciação e caracterização de tumores, em termos de tipo, grau, permeabilidade, vascularização e agressividade (Brandão, Shiroishi, & Law, 2013).

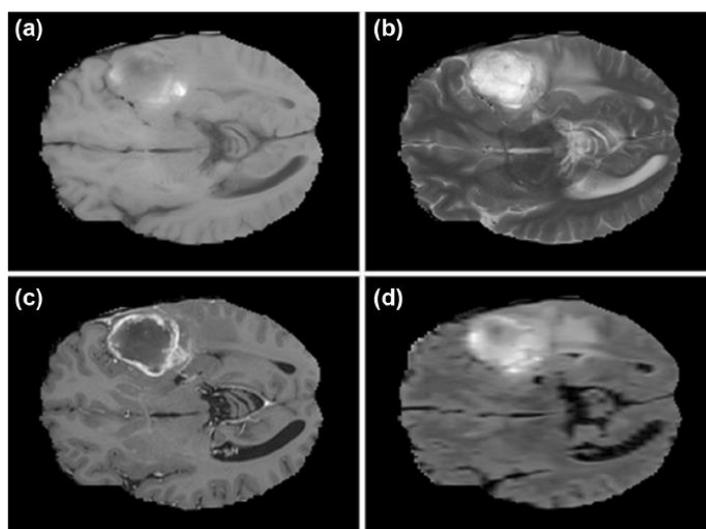


Figura 2.7 – imagem de ressonância magnética multimodal do cérebro na ponderação em T1 (a), T2 (b), T1 com contraste (c) e FLAIR (d) (Usman & Rajpoot, 2017).

2.5.3 Tomografia por Emissão de Positrões (PET)

O delineamento exato do volume tumoral, a distinção das características importantes do tumor na sua totalidade e a avaliação da resposta ao regime de tratamento em curso são essenciais para uma gestão bem-sucedida de todo o processo do paciente.

Entretanto, o uso dessas modalidades de imagens para seguimento após tratamentos cirúrgicos, quimioterapêuticos e radioterapêuticos pode ser limitante porque elas normalmente não conseguem diferenciar tumores recorrentes de radionecrose, por exemplo. Além disso, um entendimento ótimo da biologia tumoral é crucial para o correto procedimento em relação ao paciente, e para o desenvolvimento de terapias moleculares que se direcionem especificamente à neoplasia e reduzam a morbidade e a mortalidade.

Portanto, o interesse pela imagem de Tomografia por Emissão de Positrões (PET) advém da sua potencialidade de retratar vários processos bioquímicos que são críticos

para a compreensão da fisiopatologia e metabolismo das neoplasias em geral (incluindo as cerebrais), o que é particularmente valioso para a medição da proliferação celular atípica e investigação de tumores em estágios iniciais. A depender do radiotraçador, vários processos moleculares podem ser visualizados e quantificados. Portanto, a PET com seus diferentes radiotraçadores pode desempenhar um papel auxiliar mas significativo no seguimento das terapias (Roesler, Dini, & Isolan, 2021). Na Tabela 2.4 encontram-se indicados os radiotraçadores usados na imagem PET e o mecanismo envolvido.

A imagem PET com o marcador ^{18}F -FDG tem sido amplamente explorada (Fink, Muzi, Peck, & Krohn, 2015), sendo que a maioria dos estudos clínicos sobre imagem PET são baseados na administração desse radiofármaco, por ser um análogo da glicose, síntese relativamente simples, tempo de semivida longo (cerca de duas horas) e porque o seu mecanismo de captação é bem compreendido. O metabolismo da glicose pelas células pode ser medido por esse radiotraçador porque a sua quantidade no meio intracelular é proporcional à taxa de transporte de glicose e fosforilação intracelular, e as células malignas normalmente tem maiores taxas de metabolismo de glicose aeróbica, quando comparadas às células saudáveis. O aumento na taxa metabólica de glicose não está somente relacionada à aceleração do crescimento do tumor mas também ao seu processo de malignização (Basu & Alavi, 2009).

Tabela 2.4 – Classe, nome e mecanismo dos radiotraçadores usados em PET (Basu & Alavi, 2009).

Classe Principal	Nome do Traçador	Mecanismo Molecular
A	Fluorodeoxiglicose (FDG)	Metabolismo de glicose
B	[¹¹ C] Metionina (MET), fluoretil-L-tirosina (FET) e L-3, 4-dihidroxi-6-[¹⁸ F] fluorofenilalanina (FDOPA), L-1-[¹¹ C] tirosina (TYR), e L-3-[¹⁸ F] fluoro-a-metiltirosina (FMT)	Análogos de aminoácidos
C	[¹¹ C] PET colina (CHO)	Componentes da membrana celular radiomarcadas
D	[¹⁸ F] Fluorotimidina	Nucleosídeos marcados
E	[¹⁸ F] Fluoromisonidazol	Marcadores de imagem de hipóxia
F	[⁶⁸ Ga] DOTA-TOC PET	Marcadores de imagem do recetor de somatostatina

Dentre as aplicações da FDG na neuro-oncologia se encontram a deteção, delineação e determinação do grau do tumor, prognose da doença e planeamento da biópsia estereotáxica.

Entretanto, a alta captação de glicose – e, portanto, também do FDG – pelas células pode ocorrer também em casos de abscessos cerebrais, desmielinização inflamatória aguda e casos de lesão por radiação, ou seja, no geral em casos de alta concentração de células inflamatórias do sistema imunitário, como neutrófilos e macrófagos, o que pode levar a uma situação de diagnóstico falso-positivo (Basu & Alavi, 2009).

Essa questão é particularmente importante quando se trata de um paciente que já teve algum tumor cerebral e teve que ser tratado com radiação, uma vez que é imprescindível a correta distinção entre um tumor recorrente e radionecrose e outras lesões causada pelo tratamento radioterápico (Roesler, Dini, & Isolan, 2021). Na Tabela 2.5 constam as diferentes categorias de manifestações clínicas após a irradiação do SNC consoante o tempo decorrido.

Tabela 2.5 – Processos patológicos diferentes passagem de tempo da irradiação do SNC (Langleben & Segall, 2000).

	Aguda	Adiantada	Tardia	Crônica
Pós-irradiação	Horas - semanas	Semanas – meses	Meses - anos	-
Sintomas	Efeito de massa e consequente aumento de pressão intracranial (dor de cabeça, vômito, consciência alterada)	Misto entre sintomas agudos e tardios	Déficits neurológicos focais dependendo do tamanho e localização da lesão	Déficits neurológicos focais estáveis, bem como efeitos globais que causam mudança de personalidade e demência
MRI/CT	Edema; massa tumoral realçada	Formação de cisto de matéria branca	Atrofia cortical periventricular difusa; Edema Massa com realce expansivo com hiperintensidade focal em imagens T2 e densidade de proteínas	Calcificação vascular distrófica com alterações císticas; atrofia cortical; telangiectasia
PET	Hipometabolismo	Hipometabolismo	Hipometabolismo	Hipometabolismo
Patologia	Edema; placas desmielinizantes; fibrina exsudativa de junção cinza-branca; necrose de coagulação hemorrágica e proliferação vascular	Misto entre a aguda e tardia	Necrose fibrinoide do meio e pequenas artérias com proliferação vascular; giros inchados; necrose de coagulação hemorrágica da matéria cinza; substância branca afetada mais do que cinza; edema axonal, desmielinização multifocal, hialinização e gliose reativa	Gliose reativa; formação de cisto subcortical; vasculite, arterite, aterosclerose aumentada e calcificação vascular distrófica com mudanças císticas

A lesão por radiação pode ser aguda, adiantada ou tardia. A tardia ocorre em 5% a 37% dos pacientes tratados com radiação, e aumenta conforme o aumento da dose, e se manifesta entre 4 meses e 10 anos após a irradiação, sendo que dentro dos dois primeiros anos ocorrem 70% dos casos (Langleben & Segall, 2000). Os efeitos do tratamento com radiação também são mais pronunciados quando combinados com quimioterapia, no tratamento de gliomas de alto grau.

Alguns estudos com FDG mostraram que ele pode não ser o radiofármaco ideal na distinção entre radiolesões (Langleben & Segall, 2000), enquanto outros se mostram promissores (Fink, Muzi, Peck, & Krohn (2015) e Roesler, Dini, & Isolan (2021)). Embora o FDG seja o radiofármaco mais usado, estudos têm sido feitos com outras substâncias com o intuito de diminuir a incidência de falsos positivos nos exames imagiológicos de PET. Takenaka *et al.* (2014) estudou a acurácia de radiofármacos para diferenciação entre tumores recorrentes e radionecrose usando ^{11}C -metionina, ^{11}C -colina e ^{18}F -Fluorodeoxiglicose (FDG) em 50 pacientes tratados com radioterapia. Dos 50 pacientes, 17 tinham glioma grau 3 recorrente, 17 tinham glioblastoma (grau 4) recorrente e os 16 restantes tinham radionecrose, segundo o sistema de classificação da OMS já citado anteriormente. Os resultados da curva ROC (*Receiver Operating Characteristic* – características operacionais do recetor) foram de 92,5 para a Metionina, 81.4 para a colina e 77.4 para a FDG. De modo simples, a curva ROC é uma representação gráfica do desempenho de um classificador binário, e quanto mais perto de 100, melhor (Prates, 2020). A metionina mostrou também maior sensibilidade e especificidade do que os outros dois radiofármacos.

3. Data Science em Diagnóstico de Tumores Cerebrais

3.1. Inteligência Artificial e Ciência dos Dados

A Inteligência Artificial não é um tema exatamente novo; já nos anos 1950, o pioneiro britânico da computação, Alan Turing, propôs um teste – o Teste de Turing – para avaliar se as máquinas seriam capazes de demonstrar algum tipo de inteligência “artificial”, ponto em que o seu comportamento seria praticamente indistinguível daquele de um ser humano (Granatyr, 2016). Na mesma década, em 1956, o Professor John McCarthy elaborou o termo “Inteligência Artificial”, baseando-se em que “cada aspecto do aprendizado ou qualquer outra propriedade da inteligência pode em princípio ser tão precisamente descrita que a sua simulação pode ser feita por uma máquina (Panesar, 2019)”.

É inequívoca a receptividade da IA em setores diversos e, de maneira similar, na Medicina. Já em 2009, em uma publicação na revista *Artificial Intelligence in Medicine* (Patel, et al., 2008), Peter Szolovits afirma que:

“IA em Medicina é vista atualmente muito menos como um campo separado e mais como um componente essencial da informática biomédica e uma das metodologias que podem

ajudar a resolver problemas na saúde. (...) estamos reconhecendo que o sistema de assistência à saúde não tem uma base estática dos nossos esforços, mas deve aprender a partir de suas próprias experiências e se empenhar em implementar processos contínuos de melhorias que podem melhorar significativamente os resultados na saúde, enquanto de alguma forma mantêm sob controle o crescimento inexorável dos gastos em saúde”.

Assim como em todos os setores, busca-se e investe-se na adoção de novas tecnologias com o objetivo de diminuir custos e tempo gasto, ao passo em que se aumentam a eficiência e a qualidade. Uma grande parte desse aumento será devido ao aperfeiçoamento e personalização da Medicina possibilitados pela Inteligência Artificial, segundo Leite (2019).

Cerca de 79% dos erros médicos estão, segundo Lobo (2018), relacionados com problemas na relação médico-paciente: num exame clínico deficiente, na falha de avaliação dos dados do paciente, ou falta de exames comprovando a hipótese diagnóstica feita. Em saúde, os dados são obtidos a partir de fontes diversas: prontuários eletrônicos, dados administrativos, *wearable devices* e corporais, genomas e determinantes sociais. É com esse propósito que é utilizada a *Data Science*, a Ciência de Dados: por meio de IA, várias etapas ou competências são envolvidas, como o reconhecimento de padrões e imagens, entendimento de linguagem aberta escrita e falada, percepção de relações e nexos, seguir algoritmos de decisão propostos por especialistas para ter a capacidade de entender conceitos e não apenas de processar estatisticamente os dados, adquirir “raciocínios” pela capacidade de integrar novas experiências e, então, se autoaperfeiçoar (“*self learning*”) por meio da resolução de problemas ou da realização de tarefas.

A quantidade de informação disponível a respeito de todos os aspectos da existência humana e a rapidez e facilidade com que se pode aceder-lhe nunca foram tão grandes, e só tendem a crescer. Uma estimativa da IBM de 2012 em uma mensagem compartilhada na conta oficial da empresa na rede social Facebook assegurou que 90% dos dados globais haviam sido criados nos 2 anos anteriores (IBM, 2012). Se neste ano o volume de dados gerados foi mais de 5 zettabytes¹, em 2021 esse número deve atingir

¹ Um zetta byte equivale a 1×10^{21} bytes ou, por extenso, 1.000.000.000.000.000.000 de bytes.

os 60 zettabytes, uma vez que os tipos e fontes de dados estão se tornando cada vez mais numerosos e complexos, em razão do uso e aplicação crescentes de Inteligência Artificial (IA), *smartphones*, redes sociais e IoT (*Internet of Things*) (IBM, s.d.). Esse crescimento exponencial pode ser visualizado na Figura 3.1.

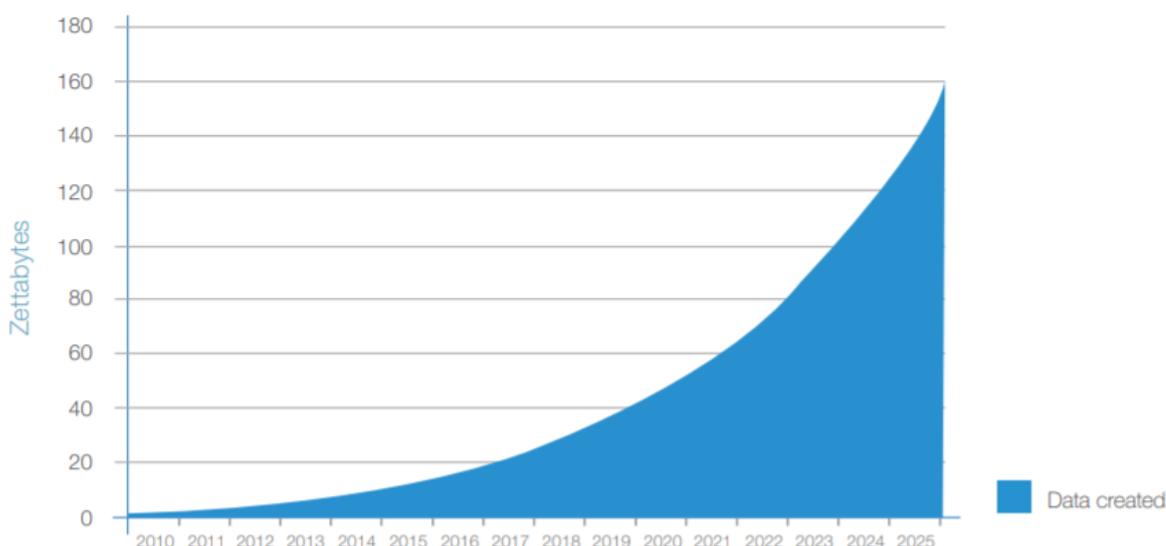


Figura 3.1 – Tamanho Anual da Datasfera Global (Reinsel , Gantz , & Rydning, 2017).

Em 2020, devido à crise multifacetada provocada pela pandemia de Covid-19, a Inteligência Artificial e suas ramificações (*Data Science, Big Data, Machine Learning*, dentre outras) passaram por uma grande validação e potencialização em todos os setores possíveis (Bilharvas, 2020), e em 2021 e 2022 o cenário é muito parecido. Projeções do IDC (IDC - International Data Corporation, 2020) indicam um crescimento anual de cerca de 17% nos investimentos no mercado de IA, ultrapassando os U\$ 300 mil milhões até 2025.

A chamada *Data Science* tem sido considerada (Cao, 2017) como um quarto paradigma da ciência contemporânea, que unifica os paradigmas predecessores: as Ciências Teóricas, as Empíricas (ou Experimentais) e as Computacionais. Mais ainda, já foi repetida de maneira exaustiva a frase “*data is the new oil*” (dados são o novo petróleo, em tradução livre). Uma simples busca na internet por essa exata frase em português conduz a uns impressionantes 17 400 000 resultados. Em inglês, chegam a ser

1 930 000 000 resultados². Para além dessa comparação, os dados “dirigem ou mesmo determinam o futuro da ciência, tecnologia, economia e possivelmente tudo em nosso mundo hoje e amanhã” (Cao, 2017).

3.2. *Machine Learning*

O desenvolvimento de ferramentas que recolhem dados, analisam-nos à procura de padrões que sejam relevantes para o problema e retornam uma resposta tem diversas designações: inteligência artificial, reconhecimento de padrões, mineração de dados, análise preditiva, *machine learning*, dentre outros. Cada área usa uma abordagem diferente, para se atingir um objetivo comum: fazer uma predição acurada (Kuhn & Johnson, 2013).

As análises estatísticas tradicionais são usadas com a intenção de se predizer quantitativamente qual a influência de uma variável em um resultado de interesse como, por exemplo, qual o peso de certo mineral na acidez da água, o que resulta na necessidade de um grande número de dados para cada variável incluída no modelo. *Machine Learning* – Aprendizagem Máquina – é uma aplicação da IA que permite que os computadores “aprendam” a partir da inserção de dados (Hurwitz & Kirsch, 2018), ao invés de programação convencional. A aprendizagem, nesse contexto, pode ser vista como estimar as dependências dos dados (IBM Cloud Education, 2020). Diferentemente dos processos estatísticos tradicionais citados acima, os algoritmos de *Machine Learning* (ML) aprendem de maneira iterativa a partir dos dados para que melhorem – ou atualizem – continuamente os resultados obtidos. Um processo iterativo pode ser entendido como aquele que é repetido de maneira contínua, de forma a gerar um resultado parcial a cada repetição, o qual será utilizado na repetição seguinte (Dicionário Informal, 2010). De modo simples, o algoritmo de ML é submetido a um processo de treino a cada iteração, o que melhora seguidamente (embora haja um limite de melhoria) o seu desempenho. Assim, mesmo que haja mudança nas variáveis (ou na relação entre elas), o algoritmo é capaz de realizar o acompanhamento da sua evolução, de processar esses novos dados (praticamente em tempo real caso seja um algoritmo *online*) e de se autoajustar (Figura 3.2).

² Acesso feito em 13/07/2021, a partir do mecanismo de pesquisa do Google, para “dados são o novo petróleo” e “*data is the new oil*”.

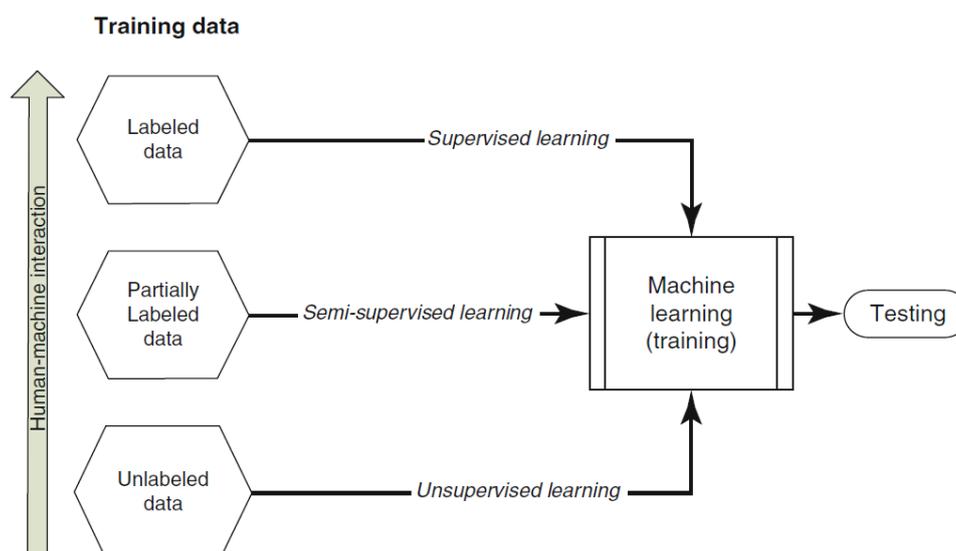


Figura 3.2 – Categorias de algoritmos de ML conforme a natureza dos dados de treino (El Naqa, Ruijiang, & Murphy, 2015).

A “aprendizagem” dos algoritmos de ML pode ser dividida em três categorias, de acordo com a natureza da rotulagem dos dados (*data labeling*): supervisionados (dados rotulados), não-supervisionados (não rotulados) e semissupervisionados (parcialmente rotulados). Abaixo, descreve-se com mais detalhe os dois tipos mais comuns.

3.2.1 Aprendizagem Supervisionada

De maneira simples, na aprendizagem supervisionada tenta-se prever uma variável dependente a partir de variáveis independentes. Diz-se “supervisionado” porque, no conjunto de dados de treino, a resposta (ou respostas) desejada está contida, ou seja, o resultado desejado do processo de aprendizagem é conhecido. Os dados então são rotulados com as previsões ou classes (Honda, Facure, & Yaohao, 2017) e os problemas são divididos em Regressão ou Classificação.

Na regressão, o objetivo é prever os resultados em um conjunto contínuo de saída, ou seja, fazer o mapeamento dos dados de entrada para alguma função contínua. Por exemplo, a previsão dos preços de casas em função de variáveis como tamanho, localização e o estado de conservação é um problema de regressão, já que o conjunto de saída “Preço” é composto por valores contínuos (Barros, 2016).

Os problemas de classificação envolvem a previsão de resultados discretos, ou seja, a categorização dos dados. A questão que está a ser apresentada nessa dissertação é um problema de classificação, pois envolve encaixar cada um dos dados de entrada (os exames que contém as diversas imagens de RM) em categorias definidas (sem tumor, meningioma, glioma ou tumor da glândula pituitária).

3.2.2 Aprendizagem Não-Supervisionada

Esta abordagem é usada em problemas em que há pouca ou nenhuma ideia da aparência que devem ter os resultados. O algoritmo analisa os dados sozinho e começa a estabelecer padrões, com base nas semelhanças (ou ausência dessas semelhanças) entre os dados. Como não se sabe quais deveriam ser os resultados, não há como fazer o cálculo da precisão.

Esta abordagem é mais utilizada quando não se tem todos os dados. Por exemplo, ao se tentar identificar os diferentes perfis de clientes que vão às compras em um mercado, não é possível saber previamente quantos são tais perfis ou quais suas características. Nesse caso, o algoritmo poderia fazer a análise dos registos de compras para identificar padrões e relações entre a compra de certos produtos (Honda, Facure, & Yaohao, 2017).

3.3. Algoritmos de *Machine Learning*

Um problema importante e geral em ML, que tem ampla aplicação, é aprender ou inferir uma relação funcional entre um conjunto de variáveis de atributo e a resposta associada ou variáveis-alvo para que possamos prever a resposta para qualquer conjunto de atributos (Rogers & Girolami, 2012). Para que essa relação funcional seja inferida com precisão e, provavelmente o mais importante, resolva o problema, a escolha do algoritmo deve ser ponderada consoante tal problema.

Para que se crie um modelo, a classificação requer um conjunto de dados (*dataset*) com vários exemplos de entrada e saída a partir do qual se possa “aprender”. O modelo usará o *dataset* de treino para calcular a melhor forma de mapear os exemplos dos dados em *labels* (categorias) de classe específicas. Dessa forma, o conjunto de dados

de treino deve ser suficientemente representativo do problema em questão, tendo vários exemplos de cada classe.

Para os problemas de classificação multiclasse, alguns dos algoritmos de ML mais populares são o k-NN, *Decision Trees*, *Random Forest*, *Naive Bayes* e *Gradient Boosting*, *SVM* e Regressão Logística. A classificação utilizando ML refere-se a um problema de modelo preditivo no qual uma classe, rótulo ou *label* é predita ou atribuída para um certo dado de *input* (Brownlee, 2020).

3.3.1 *Decision Tree*

Decision Tree (DT), ou Árvore de decisão, é um algoritmo de ML bastante utilizado para problemas de regressão e classificação. É um algoritmo não paramétrico de aprendizagem supervisionada, cujo objetivo é criar um modelo que preveja a classificação atribuída a uma variável alvo aprendendo regras simples de decisão inferidas a partir das características dos dados (SciKit Learn, s.d.). Os vários pontos de decisão criados pelo algoritmo são os “nós” da árvore, a partir dos quais o resultado seguirá um caminho ou outro, que são os “ramos” da árvore (Didática Tech, 2020). A árvore pode ser vista como uma aproximação constante por partes. Quanto mais profunda a árvore (o que se define pelo parâmetro `max_depth`), mais complexas as regras de decisão e mais ajustado é o modelo aos dados utilizados.

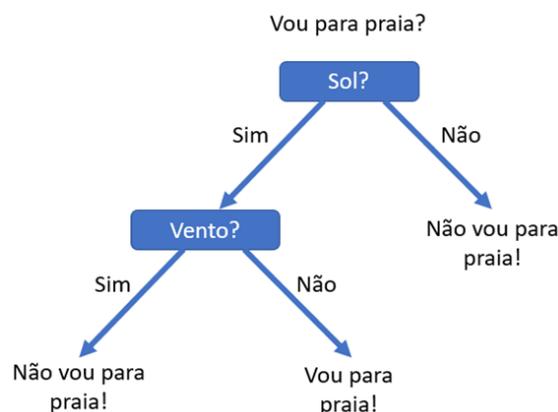


Figura 3.3 – Ilustração de uma árvore de decisão (Didática Tech, 2020).

Na Figura 3.3 é possível observar de uma maneira simplificada o funcionamento de uma árvore de decisão para uma pergunta, e os possíveis caminhos que se pode tomar a partir das respostas dadas em cada nó.

3.3.2 *Random Forest*

Random Forest (RF) significa literalmente “Floresta Aleatória”, por se tratar de um conjunto de árvores de decisão, e também é um algoritmo de aprendizagem supervisionada (Donges, 2021). O número de árvores ($n_{estimators}$) do conjunto é um dos parâmetros a definir, constituindo o mais importante, e o valor estabelecido por defeito são 100 (SciKit Learn, s.d.).

Na Figura 3.4 encontra-se um esquema do funcionamento do algoritmo, com três árvores. Cada árvore classifica individualmente os dados, e ao final a classe mais escolhida é tomada como a classificação final.

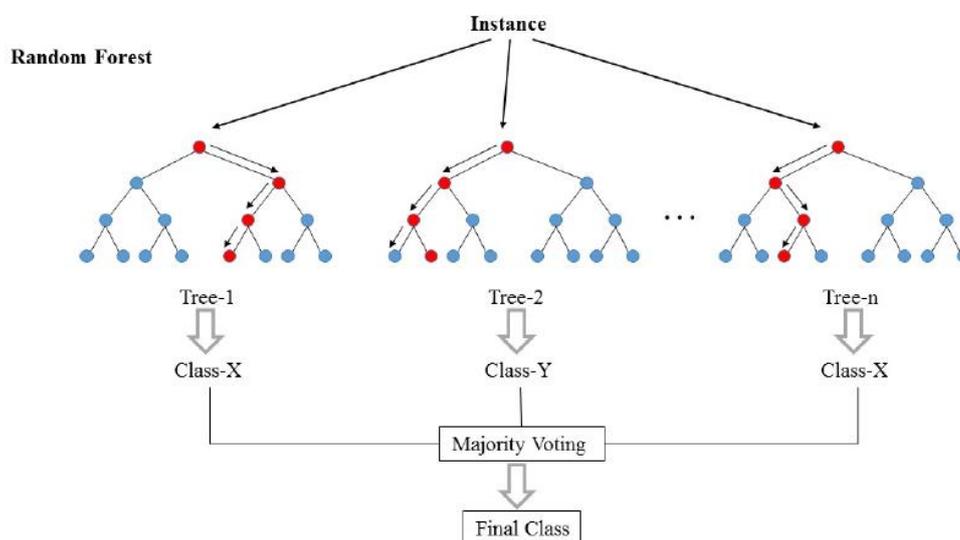


Figura 3.4 – Esquemática do algoritmo *Random Forest* (Pawar, 2020).

Alguns dos outros parâmetros são (SciKit Learn, s.d.):

- **Criterion:** é o critério com o qual dividir cada nó. Pode ser “Gini” (de defeito) ou “Entropia”;
- **max_depth** (profundidade máxima das árvores individuais – quanto mais profunda, maior a chance de *overfitting*). O valor de defeito é “none” (nenhum);
- **min_samples_split:** é o número mínimo de amostras requeridas para dividir um nó. O valor de defeito é 2;
- **min_samples_leaf:** o número mínimo de amostras necessárias para estar em um nó de folha. O número de defeito é 1.

As árvores constituem um conjunto diversificado de classificadores que é criado pela introdução de aleatoriedade na construção do classificador. A previsão do conjunto de árvores (floresta) é dada como a previsão média (no caso de regressão) ou do resultado que mais aparece (no caso de classificação) dos classificadores individuais. A vantagem de ter um conjunto de árvores ao invés de uma só é a diminuição da variância e do *overfitting* do modelo, de forma que, de maneira geral, quanto mais árvores, melhor (Didática Tech, 2020). A desvantagem em relação ao algoritmo de DT ocorre em relação ao tempo de computação, que aumenta com o número de árvores inseridas no modelo, e nem sempre um aumento significativo no número de árvores leva a um aumento significativo na acurácia.

3.3.3 *Linear Discriminant Analysis (LDA)*

“Análise Discriminante Linear” é um classificador de ML supervisionado com uma fronteira de decisão linear, gerada por ajustes de densidades condicionais de classes aos dados. É um classificador com “soluções de formato fechado que podem ser facilmente e rapidamente calculadas, são inerentemente multiclasse, têm funcionado bem na prática e não tem hiperparâmetros para ajustar” (Sickit Learn, 2021). Hiperparâmetros são os parâmetros do modelo que devem ser definidos antes do treino do modelo. A LDA é uma ferramenta usada para a classificação, redução de dimensão e visualização dos dados. A redução de dimensões visa remover características redundantes e dependentes, transformando um espaço dimensional maior num espaço dimensional menor (Silva & Macedo, 2020).

Há três passos básicos para o funcionamento desse algoritmo: primeiro, calcula-se a separabilidade entre as diferentes classes, o que é chamado de variância entre classe, a qual é definida como a distância entre a média das diferentes classes. Depois, calcula-se a variância intraclasse, que é a distância entre a média e a amostra de cada classe. Por último, constrói-se um espaço de dimensão menor para maximizar a primeira etapa (variação entre classes) e minimizar a segunda etapa (variação dentro da classe). Ou seja, o algoritmo encontra os discriminantes lineares que representam os eixos que maximizam a separação entre as diferentes classes (Mehta, 2021). A Figura 3.5 ilustra o processo de classificação do algoritmo LDA.

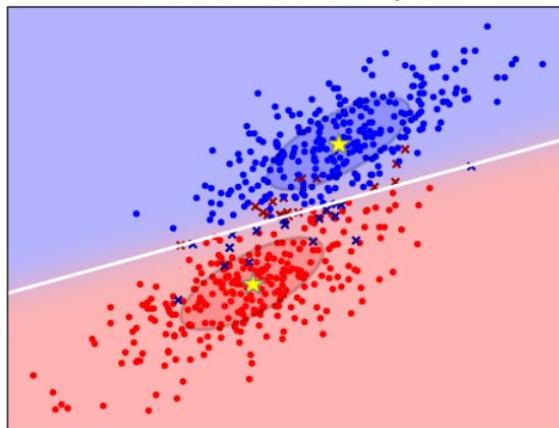


Figura 3.5 – Ilustração da classificação do algoritmo LDA (Sickit Learn, 2021).

A predição de classe é feita pela estimativa da probabilidade de que um certo conjunto de *inputs* pertença a cada classe. A classe de saída é aquela com maior probabilidade.

3.3.4 *Support Vector Machine*

As Máquinas Suporte Vetorial (SVM) constituem uma família de algoritmos supervisionados usados para problemas de regressão, detecção de *outliers* e classificação, sendo mais amplamente usado para a classificação. O objetivo do algoritmo é encontrar um hiperplano no espaço N -dimensional (sendo N o número de atributos) que classifica distintamente os dados. Para fazer a separação dos dados em classes, há muitos hiperplanos possíveis para serem escolhidos. O algoritmo procura encontrar um plano com a máxima margem possível, ou seja, a máxima distância entre os pontos dos dados das classes. A maximização da distância de margem faz que haja mais robustez, de modo que os próximos dados sejam classificados com mais confiança.

Os hiperplanos são fronteiras de decisão que ajudam a classificar os dados. Conforme o lado do hiperplano em que um determinado dado se encaixa, a ele é atribuída uma certa classe. O hiperplano é uma linha caso o número de atributos seja 2, e um plano caso seja 3. Hiperplanos com dimensões maiores não podem ser representados em uma imagem (Gandhi, 2018). Na Figura 3.6, é possível visualizar a separação linear em duas e três dimensões entre as classes de dados.

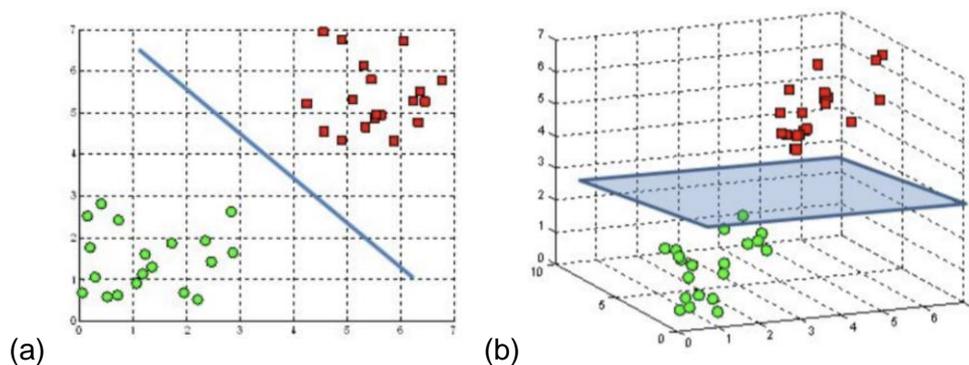


Figura 3.6 – Representação do Hiperplano com duas dimensões (a) e três dimensões (b) (Gandhi, 2018).

Entretanto, muitos grupos de dados não podem ser separados dessa forma. Por isso, existem outros tipos de *kernel*, que fazem separações de outros formatos, com formulações matemáticas diferentes (Zhang, 2018). Dentre os *kernels* disponíveis para implementação na biblioteca *SciKit Learn* em linguagem *Python* que foram utilizados, encontram-se o linear, o polinomial, o RBF gaussiano e o sigmoide. É possível também criar *kernels* personalizados (SciKit Learn, s.d.).

Dentre os parâmetros a serem ajustados, dois dos mais importantes são o Parâmetro de Regularização C , que define a otimização da margem do hiperplano (quanto maior o valor de C , menor a margem e mais ajustado aos dados) e o Parâmetro γ , que define o quão influente um dado de treino será – quanto maior o valor, mais o algoritmo considerará os pontos próximos ao hiperplano, e quanto menor, mais são considerados os pontos distantes (Zoltan, 2018). O parâmetro C varia entre 0,1 e 100, ao passo que o γ varia entre 0,0001 e 10.

Cada *kernel* funciona da seguinte forma:

- Linear: é o tipo mais básico de *kernel*, normalmente unidimensional. É o preferido em problemas de classificação de textos, já que esses casos podem ser facilmente separados de maneira linear (Awasthi, 2020). É definido matematicamente como o produto interno $k(x,y) = x \cdot y$ (no qual x e y são os dados a classificar) mais uma constante C (SciKit Learn, s.d.) – ver Figura 3.7.

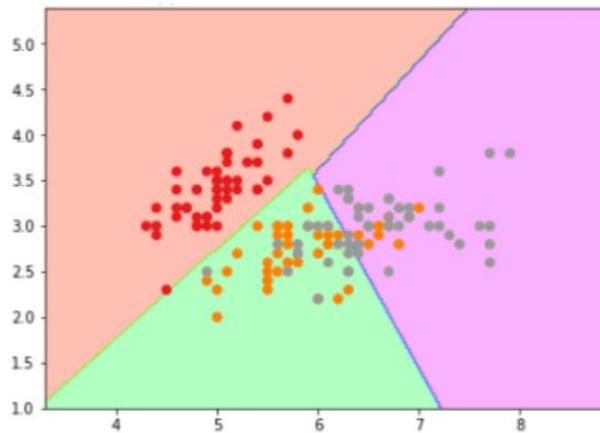


Figura 3.7 – Representação gráfica da aplicação do *kernel* linear (Awasthi, 2020).

- Polinomial: é uma versão mais generalizada do *kernel* linear, com polinómios com grau maior do que 1. É mais usado em problemas em que os dados de treino são normalizados (Joshi, 2012). Esse *kernel* não somente olha para as *features* dos dados para determinar a similaridade entre eles, mas também combinações entre elas. Sendo n o número de *features* e d o grau do polinómio, o *kernel* fornece n^d *features* expandidas (Zhang, 2018). A formulação matemática é $k(x, y) = (\gamma \langle x, y \rangle + r)^d$, na qual γ é o parâmetro *gamma*, e r é um coeficiente numérico (SciKit Learn, s.d.) – ver Figura 3.8.

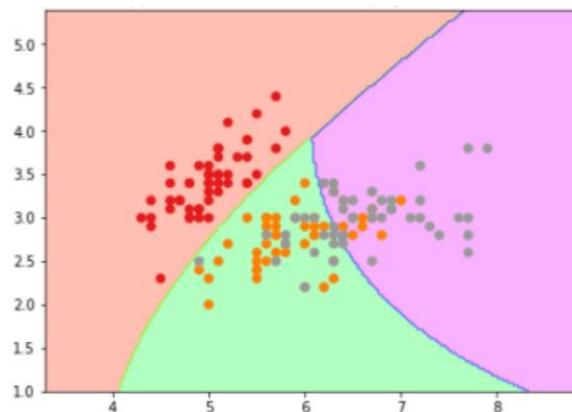


Figura 3.8 – Representação gráfica da aplicação do *kernel* polinomial (Awasthi, 2020).

- RBF: o *Gaussian Radial Basis Function* – função de base radial gaussiana – é um *kernel* usado para propósitos gerais, sendo definido como o *default* no algoritmo SVM. É a forma mais generalizada e, assim, um dos *kernels* mais amplamente usados, por ter uma distribuição gaussiana e bons resultados (Matsuzaki, 2020). Apresenta a formulação matemática de uma gaussiana: $k(x, y) = e^{-\gamma(x-y)^2}$ – ver Figura 3.9.

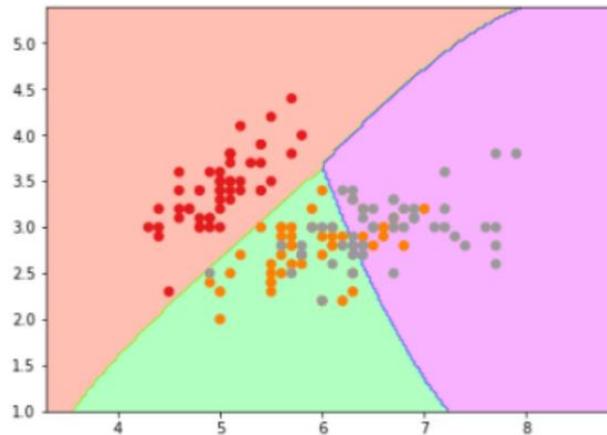


Figura 3.9 – Representação gráfica da aplicação do kernel RBF gaussiano (Awasthi, 2020).

- Sigmoide: é também conhecida como o *kernel* Tangente Hiperbólica (por causa da formulação matemática) ou *Multilayer Perceptron*, porque é usada frequentemente em algoritmos de Redes Neurais como função de ativação para neurónios artificiais, e é equivalente, em SVM, a uma rede neural de duas camadas. Matematicamente, é definido como: $k(x, y) = \tanh(\gamma \langle x, y \rangle + r)$ – ver Figura 3.10.

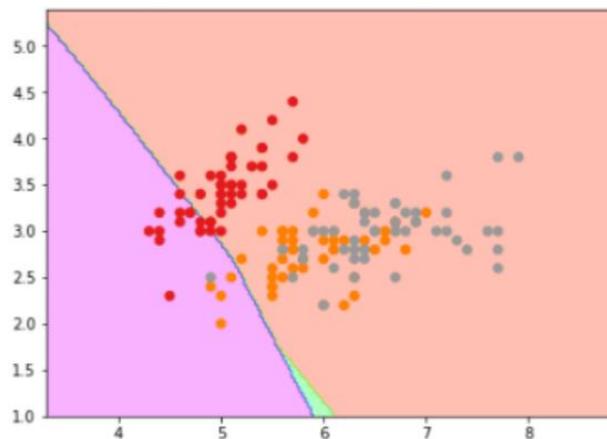


Figura 3.10 – Representação gráfica da aplicação do kernel sigmoide (Awasthi, 2020).

3.3.5 k – Nearest Neighbors

É um dos algoritmos mais simples e computacionalmente rápidos usados em ML para problemas de classificação e regressão. Esse algoritmo lê os dados de treino e classifica os novos dados baseando-se na semelhança entre eles (Singh, 2018). Significa, literalmente, os “ k vizinhos mais próximos”, pelo que este algoritmo assume que dados similares se encontram próximos uns aos outros. Normalmente, a distância euclidiana é a escolhida para calcular a proximidade entre os pontos (Harrison, 2018). O número

de vizinhos a serem analisados é um parâmetro a determinar. Por defeito, a biblioteca *SciKit Learn* define o número k de vizinhos como sendo 10. O valor ótimo de k é altamente dependente dos dados; no geral, um k maior diminui os efeitos de ruídos, mas acaba por fazer as fronteiras de classificação menos distintas (SciKit Learn, s.d.) – ver Figura 3.11.

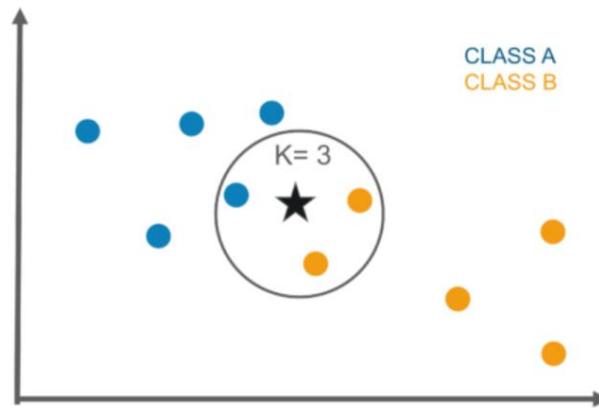


Figura 3.11 – Representação do algoritmo kNN com 3 vizinhos (Samavedam, 2020).

3.4. *Machine Learning* em Diagnóstico de Tumores Cerebrais

O tratamento de um paciente com cancro envolve uma série de etapas desde as partes iniciais de diagnóstico, a passar pela decisão e implementação do tratamento, verificação da resposta ao tratamento e acompanhamento na fase da remissão. Os graus de complexidade de cada etapa variam conforme o caso em específico e podem incluir muitas interações entre pessoas e máquina. A integração de algoritmos de *machine learning* pode auxiliar na otimização e automatização de várias partes desses processos, podendo inclusive melhorar tanto a segurança quanto os resultados do tratamento (El Naqa, Ruijang, & Murphy, 2015).

A Imagem Médica é um recurso extremamente importante em várias etapas no processo de tratamento de um paciente, como a deteção, caracterização, estadiamento, verificação da resposta ao tratamento e monitorização de recorrência, bem como auxílio essencial em intervenções cirúrgicas e radioterapia (Lee & Fujita, 2020). Essa importância é corroborada pelo facto de atualmente existirem diversas modalidades de imagens, de forma que se pode escolher qual a mais adequada para a situação, como o raio X, tomografia computadorizada, ultrassom e ressonância magnética.

Com o aumento da diversidade e complexidade dos exames de imagem, o número de imagens geradas para um dado paciente aumentou dramaticamente de algumas unidades no caso de imagens 2D para centenas com imagens em 3D a milhares com as atuais imagens dinâmicas 4D (Lee & Fujita, 2020). Dada a existência de um número tão grande de imagens, é esperado que os radiologistas e médicos possam deixar passar alguma lesão.

Uma vantagem importante da automação dos processos seria a substituição das pessoas em tarefas laboriosas e repetitivas, que causam cansaço físico e mental e por vezes levam a esses erros. Outra vantagem (mais significativa) é a de ser potencialmente mais capaz de aprender a identificar padrões mais sutis ou complicados que uma pessoa é capaz de observar.

Por isso mesmo, uma das principais propostas para aplicação de *Machine Learning* em Medicina é no desenvolvimento e melhoria dos *CADx System*, que são os Sistemas de Diagnóstico Auxiliado por Computador (do inglês *Computer Aided Diagnosis*) e *CADe System (Computer Aided Detection)* (El Naqa, Ruijang, & Murphy, 2015) que, dentre outras atribuições, fazem o processamento de imagens médicas digitais e ajudam na identificação e classificação de lesões e patologias em áreas diversas da Medicina (Lobo, 2017).

A ML desempenha um papel fundamental nos CADx e CADe, porque lesões e tumores nos órgãos e tecidos são muito complexos para serem descritos de maneira precisa por uma equação simples, de forma que esses modelos requerem um grande número de parâmetros que devem ser determinados por dados (El Naqa, Ruijang, & Murphy, 2015). A classificação de objetos candidatos a lesão em classes com base em características como área, contraste com as regiões circundantes e circularidade é um dos principais usos da ML em CAD.

3.4.1 Trabalhos relacionados

No primeiro estudo apresentado – *Brain tumor detection using statistical and machine learning method* – Amin J. , Sharif, Raza, Saba, & Anjum (2019) – foram utilizados dois *datasets* públicos (BRATS 2013 e 2015) contendo imagens de gliomas de alto grau (HGG – *high-grade glioma*) e baixo grau (LGG – *low-grade glioma*) e o terceiro foi obtido no hospital local, o Hospital Multan de Nishtar no Paquistão, e contém imagens

classificadas como positivo ou negativo para a existência de tumor, sem distinção de tipo de tumor no caso dos positivos – ver Tabela 3.1.

Tabela 3.1 – *Datasets* utilizados em (Amin J. , Sharif, Raza, Saba, & Anjum, 2019).

<i>Database</i>	<i>Modalidades</i>	<i>Slices</i>	<i>Gliomas</i>
BRATS 2013	FLAIR e T2	155 <i>slices</i> em cada modalidade	20 HGG e 10 LGG
BRATS 2015	FLAIR e T2	155 <i>slices</i> em cada modalidade	220 HGG e 54 LGG
Local	T2	86 imagens	49 tumor e 37 não-tumor

A proposta do trabalho de Amin, Sharif, Raza, Saba, & Anjum (2019) é a utilização de três fases com abordagens diferentes, cada uma seguida então pela classificação:

- a) Realce da lesão: emprega-se o filtro Wiener para a remoção de ruído e realce da região da lesão;
- b) Segmentação da lesão: aplica-se o método de agrupamento PF (*potential field*) para descobrir o subconjunto de pixels do tumor. Em seguida, usa-se o método do limiar global para a segmentação e então a morfologia matemática para a refinação dos resultados da segmentação;
- c) Extração das características: as *features* de textura GWT (*Gabor wavelet transform*) e LBP (*Local binary pattern*) são obtidos a partir de cada imagem segmentada e, então, são fundidas para melhorar os resultados de discriminação.

Os classificadores utilizados foram os seguintes:

- i) SVM com *kernel* de função quadrática para classificação binária;
- ii) kNN com 5 vizinhos para medir a acurácia da classificação;
- iii) *Decision Tree* é usado numa amostra dos dados para computar os atributos de interdependência baseado nos dados de *input*, por meio da criação de uma estrutura de árvore e imputa atributos e rótulos;
- iv) *Random Forest* é uma abordagem híbrida com, neste caso, 200 árvores (as *Decision Trees* do item anterior);

- v) *Naive Bayes* é um classificador que se beneficia de classes independentes de associação entre características que pertencem à mesma família probabilística.

As três etapas foram realizadas, após cada qual foi feita uma análise de performance:

- Na primeira, é aplicado um filtro do tipo Wiener nas imagens para a redução do ruído e realce da lesão, para que se faça uma melhor segmentação da região de interesse, seguida pelas análises de performance por PSNR (*peak signal-to-noise ratio* – razão sinal-ruído de pico), SSIM (*structural similarity index* – índice de similaridade estrutural) e MSE (*mean square error* – erro quadrado médio).
- Na segunda, o método de *clustering* (agrupamento) com PF (*potential field* – campo potencial) é usado para encontrar o subconjunto de pixels referentes ao tumor na imagem de RM. As intensidades de valores dos pixels criam agrupamentos de campo potencial, por exemplo, valores finitos de intensidade de pixels são necessários para calcular o agrupamento de PF em cada pixel. O limiar global é aplicado para a segmentação do tumor, além da erosão morfológica para remover pixels extra e dilatação morfológica para a recuperação de pixels perdidos.
- No terceiro passo, os atributos de LBP e GWT de cada fatia segmentada. 50 atributos são então combinados nas duas categorias, 20 para GWT e 30 para LBP, para criar um vetor fusão das duas categorias.

Os melhores resultados foram então comparados por Amin J. , Sharif, Raza, Saba, & Anjum com um outro estudo de referência (Pinto, Pereira, Rasteiro, & Silva, 2018), em termos de sensibilidade (SE), coeficiente de similaridade de dados (DSC) e valor preditivo positivo (PPV), segundo as equações abaixo, nas quais TP são os positivos verdadeiros (*true positive*), FP são os falsos positivos (*false positive*) e FN são os falsos negativos (*false negative*):

$$(a) SE = \frac{TP}{TP + FN} \quad (b) PPV = \frac{TP}{TP + FP} \quad (c) DSC = \frac{2TP}{FN + 2TP + FP}$$

Equação 3.1 – cálculo da sensibilidade (a), valor preditivo positivo (b) e coeficiente de similaridade de dados (c).

Nesses termos, os resultados obtidos foram os apresentados na Tabela 3.2.

Tabela 3.2 – Comparação entre os resultados obtidos por Pinto, Pereira, Rasteiro, & Silva e Amin, Sharif, Raza, Saba, & Anjum (2019).

REFERÊNCIA	ANO	SE	DSC	PPV
Pinto, Pereira, Rasteiro, & Silva	2018	0,84 ± 0,07	0,85 ± 0,02	0,88 ± 0,06
Amin, Sharif, Raza, Saba, & Anjum	2019	0,97	0,88	0,86

Os autores afirmam que, na avaliação mais detalhada da performance, a fusão de atributos junto ao classificador kNN (resultado apresentado na Tabela 3.2) superam os demais.

O segundo estudo, *Detection and Classification of HGG and LGG Brain Tumor Using Machine Learning* (Polly, Shil, Hoassain, Ayman, & Jang, 2018), apresenta um conjunto de 440 imagens de Ressonância Magnética ponderadas em T2 selecionadas dos *datasets* públicos BRATS17, BRATS13 e Midas: *healthy human brain database*, conforme a Tabela 3.3.

Tabela 3.3 – Classificação e quantidade de imagens do *dataset* utilizado em (Polly, Shil, Hoassain, Ayman, & Jang, 2018).

Classificação	Quantidade	Subclassificação
Normal	100	—
Tumor	340	HGG
		LGG

Segundo os autores, “o sistema proposto é projetado para a detecção precisa e classificação de imagens de RM cerebrais normais e anormais e em seguida, a classificação das ressonâncias magnéticas anormais em tumor de glioma HGG ou LGG” (livre tradução).

A metodologia proposta por Polly, Shil, Hoassain, Ayman, & Jang consiste, portanto, em duas fases, cada qual com um grupo diferente de imagens para classificações binárias: na primeira, as imagens são classificadas somente em “normal” (100 imagens) e “anormal” (180 imagens com tumor), sendo que dessas 280 imagens 130 (70 normais e 60 anormais) são usadas para treino e as outras 150 (30 normais e 120 anormais) para teste. Na segunda, 160 imagens “anormais” são classificadas em glioma de alto

grau (HGG), ou de baixo grau (LGG). Foram utilizadas para treino 30 imagens de HGG e 30 de LGG, e para teste 50 HGG e 50 LGG.

O processo consistiu nos seguintes passos:

1) A imagem é lida pelo sistema, o qual aplica uma binarização de *otsu* para a conversão para imagem binária (preto e branco). Os limiares da imagem são estabelecidos então a depender da variância mínima dentro das classes (*clusters*) e reconvertida para uma forma legível de imagem.

2) Depois, uma clusterização de *k-means* (agrupamento de k-médias) é aplicada para a segmentação, fazendo uso de distância euclidiana para medir a proximidade do ponto ao centroide. Uma nova ligação é feita entre o mais próximo novo centroide e os mesmos pontos de conjunto de dados. Um *loop* é então gerado. Como resultado deste *loop*, os locais com semelhantes valores de pixel são separados em dois ou mais clusters.

3) Então, a extração de atributos é feita via DWT (*Discrete Wavelet Transform*) e a redução de atributos via PCA (*Principal Component Analysis*).

4) Finalmente, um SVM (*Support Vector Machine*) é usado para a classificação. O *kernel* não foi especificado.

Os resultados foram classificados baseados nos parâmetros sensibilidade, especificidade e acurácia.

$$(a) SE = \frac{TP}{TP + FN} \quad (b) ES = \frac{TN}{TN + FP} \quad (c) Acc = \frac{TP + TN}{TN + FN + TP + FP}$$

Equação 3.2 – (a) sensibilidade (b) especificidade (c) e acurácia.

Para a classificação dos tumores em gliomas de alto ou baixo grau, foi alcançada a acurácia de 99%, sensibilidade de 100% e especificidade de 98,03%. Os autores também compararam a acurácia do seu trabalho com a de outros autores – ver Tabela 3.4 – não utilizando as outras medidas (sensibilidade e especificidade).

Tabela 3.4 – Comparação da acurácia dos autores com outros trabalhos – traduzida e adaptada de (Polly, Shil, Hoassain, Ayman, & Jang, 2018).

TRABALHO	ACURÁCIA
Batra & Kaushik, 2017	98.2 %
Amien, Adb-elrehman, & Ibrahim, 2013	98.6 %
Devasena & Hemalatha, 2013	98.8 %
Cheng, Chen, Yang, & Leng, 2012	91.49 %
Polly, Shil, Hoassain, Ayman, & Jang, 2018	99 %

Ao contrário dos anteriores, no terceiro estudo (Amin, Sharif, Yasmin, & Fernandes, 2017) a classificação dos tumores foi multiclasse, e não somente binária (tumor/não-tumor). Foram usados conjuntos de dados diferentes (Local, Harvard e RIDER) e três *kernel functions* diferentes (linear, cúbica e gaussiana) para o algoritmo de classificação *Support Vector Machine* (SVM), o que se aproxima mais da proposta deste trabalho.

O *dataset* local é composto por 85 imagens obtidas no Hospital Multan de Nashtar, sendo 46 com tumor e 39 sadios. O *dataset* de Harvard é composto por 100 imagens, sendo 65 com tumor e 35 saudáveis. Por fim, o *dataset* RIDER compreende exames de 126 pacientes, com vários graus de Astrocitomas, Glioblastomas Multiformes (grau 4) e Oligodendriomas, sendo 36 pacientes com grau 1, 41 com grau 2, 26 com grau 3 e 23 com grau 4. Os outros dois *datasets* não se encontram tão detalhados no trabalho, que foram avaliados por meio da detecção dos tumores (classificação binária), e não da classificação multiclasse. A técnica utilizada consiste em três passos principais, a saber, o pré-processamento das imagens, extração de atributos e análise de classificação.

Na primeira fase, são aplicados diferentes métodos para segmentar a região de interesse. São aplicados os métodos de BSE (*Brain Surface Extractor*) para identificar limiares e fazer operações morfológicas. Aqui são usados para remover o crânio, conforme é possível ver na Figura 3.12. A imagem também passa por um filtro Gaussiano para redução do ruído e suavização. A segunda etapa consiste em fazer a extração dos atributos de textura, intensidade e forma das regiões de interesse (candidatas à lesão). Os atributos extraídos são Energia, Variância, Entropia, Dissimilaridade, Homogeneidade, *Cluster Shade*, Proeminência do *Cluster*, Área, Perímetro e Circularidade.

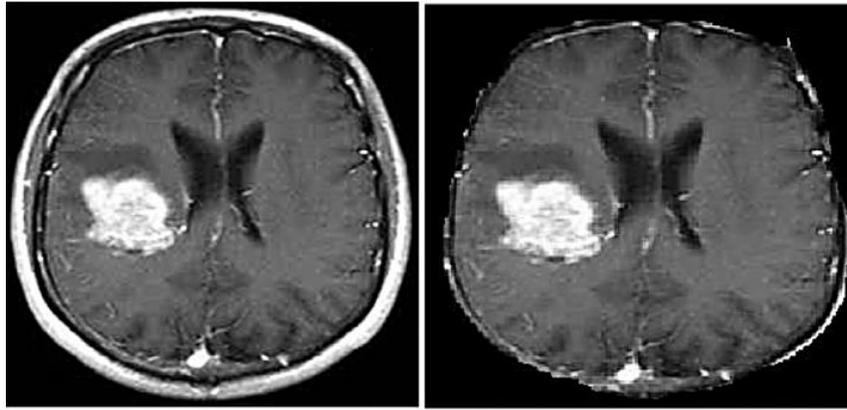


Figura 3.12 – Imagem sem pré-processamento (a) e com remoção do crânio (b) (Amin, Sharif, Yasmin, & Fernandes, 2017).

Na terceira e última etapa são testados três *kernels* (Linear, Gaussiano e Cúbico) da família de classificadores SVM. Os autores ainda utilizam a estratégia de validação cruzada dos resultados para uma melhor especulação dos classificadores, em 5, 10, 15, 20, 25 e 30 vezes. Os resultados globais encontram-se na Tabela 3.5, retirada diretamente do artigo, na qual Acc é a acurácia e Auc a área sob a curva.

Datasets	Cross validation	Linear		Cubic		Gaussian	
		ACC	AUC	ACC	AUC	ACC	AUC
Local	5 fold	98.8%	0.98	93.5%	0.94	96.1%	0.97
	10 fold	92.2%	1.00	79.9%	0.91	97.6%	0.99
	15 fold	96.0%	0.97	93.3%	0.94	85.9%	0.99
	20 fold	96.6%	0.98	94.6%	0.94	89.4%	1.00
	25 fold	96.0%	0.99	94.0%	0.94	87.1%	1.00
	30 fold	95.5%	0.96	77.3%	0.97	98.1%	0.98
RIDER	5 fold	81.3%	0.86	95.0%	1.00	93.3%	1.00
	10 fold	75.0%	0.73	96.7%	1.00	88.3%	1.00
	15 fold	85.9%	0.87	86.2%	0.77	84.0%	0.87
	20 fold	79.1%	0.86	80.0%	0.92	82.8%	1.00
	25 fold	90.0%	0.88	100%	1.00	89.3%	0.93
	30 fold	78.1%	0.76	85.0%	0.88	96.8%	0.97
Harvard	5 fold	98.1%	1.00	93.5%	0.96	95.5%	0.97
	10 fold	99.4%	1.00	87.7%	0.95	96.8%	0.97
	15 fold	98.7%	0.99	81.2%	0.90	97.4%	0.99
	20 fold	96.1%	0.94	75.3%	0.89	98.7%	1.00
	25 fold	95.5%	0.93	83.1%	0.91	86.4%	0.98
	30 fold	94.8%	1.00	78.6%	0.90	92.2%	1.00

Tabela 3.5 – Acurácia (Acc) e Área sob a curva (Auc) para cada *kernel* em cada *dataset* (Amin J. , Sharif, Yasmin, & Fernandes, 2017).

Para o mesmo *kernel*, percebe-se que há variações nos valores das métricas de desempenho consoante o *dataset*, o que mostra que o conjunto escolhido tem grande influência no resultado da acurácia dos classificadores, sendo que a menor foi

alcançada com o *kernel* linear, no conjunto RIDER, validação cruzada 10 vezes (75%), contra 99,4% para o conjunto de Harvard, sob as mesmas condições. Os autores ainda comparam seu trabalho com outros semelhantes, em termos da acurácia (ACC), especificidade (ESP) e sensibilidade (SEN).

Tabela 3.6 – Comparação de resultados entre Amin, Sharif, Yasmin, & Fernandes (2017) com trabalhos semelhantes. Traduzida e adaptada do artigo original.

TRABALHO	ACC	ESP	SEN
Sudharani, Sarma, & Prasad, 2016	89,2%	90%	88,9%
Nabizadeh, John, & Wright, 2014	91,5%	-	-
Subashini, Sahoo, Sunil, & Easwaran, 2016	91%	-	-
Eman, Mohammed, & Rashid, 2015	95,06%	-	-
Nabizadeh & Kubat, 2015	79,3 ± 0,3	81,3 ± 1,2	77,3 ± 1,4
Amin, Sharif, Yasmin, & Fernandes, 2017	97,1% (média)	98%	91,9%

Os autores, assim como nos dois trabalhos previamente apresentados, exibem resultados superiores ao comparar seu método – ver Tabela 3.6 – ao de outros semelhantes.

4. Materiais e Métodos

4.1. *Dataset*

Dataset significa literalmente “conjunto de dados”, e constitui a principal matéria-prima para as análises de dados. São representados por dados em formato de matriz, na qual as linhas são os registos dos acontecimentos e as colunas são as características (*features*) desses acontecimentos (Hoppen & Prates, 2018).

As imagens são “lidas” pelo computador como um conjunto de números binários, que são dispostos em linhas e colunas e então podem ser analisados como dados (Pedrosa, 2020). Uma imagem digital pode ser vista como uma matriz de níveis de cinzento, normalmente em escala de 8 bits, que fornece 2^8 tons de cinzento – 256 tons, do preto ao branco, portanto (TechLib, s.d.) – ver Figura 4.1.

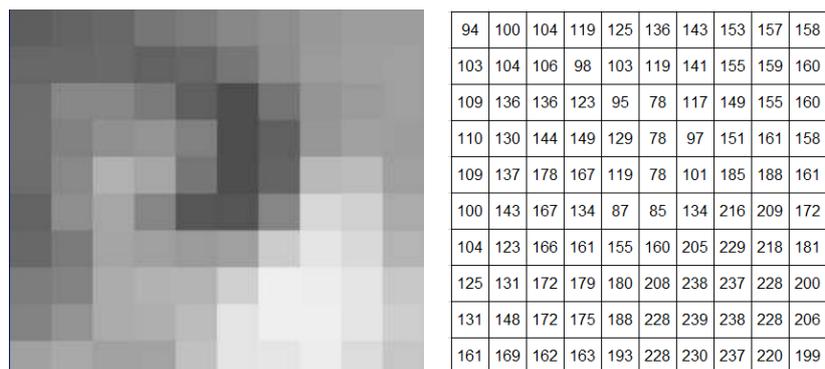


Figura 4.1 – Imagem em tons de cinzento e sua representação em escala de intensidade (Albuquerque & Albuquerque, 2012).

Para este trabalho, foram usados três grupos de imagens, sendo dois *datasets* públicos diferentes e o terceiro é um subconjunto do segundo *dataset*. Eles foram obtidos em duas plataformas: o Kaggle (Kaggle Inc, 2021) e o GitHub (GitHub Inc, 2021). Ambas são gratuitas, permitem o repositório de códigos-fonte, ficheiros e, no caso do Kaggle, também permite a competição entre pessoas que estão a aprender ou a trabalhar com ML (Wikipedia, 2021).

Os *datasets* são constituídos por imagens de RM de quatro categorias: três de tumores cerebrais diferentes e uma com cérebros sem lesão (normal). O primeiro *dataset* (Bhuvaji, 2020) é constituído por 2870 imagens, o segundo (Nickparvar, 2021) por 7020 imagens e o terceiro por uma seleção do segundo que fosse igual em número e distribuição nas categorias ao primeiro. Essa seleção foi feita aleatoriamente com o propósito de se observar as diferenças na classificação consoante o número de imagens (*datasets* 2 e 1 versus 3) e consoante as imagens em si (*datasets* 1 versus 3). As imagens de Ressonância Magnética contidas em cada categoria são de cortes nos diversos eixos (axial, sagital e coronal) e com ponderações diversas, conforme ilustrado na Figura 4.2. A distribuição dos *datasets* por categoria encontra-se na Tabela 4.1.

Tabela 4.1 – Distribuição das categorias dos *datasets*.

Dataset	Glioma	Meningioma	Glândula Pituitária	Sem Tumor	Total
1	826	822	827	395	2870
2	1628	1645	1757	2000	7020
3	826	822	827	395	2870

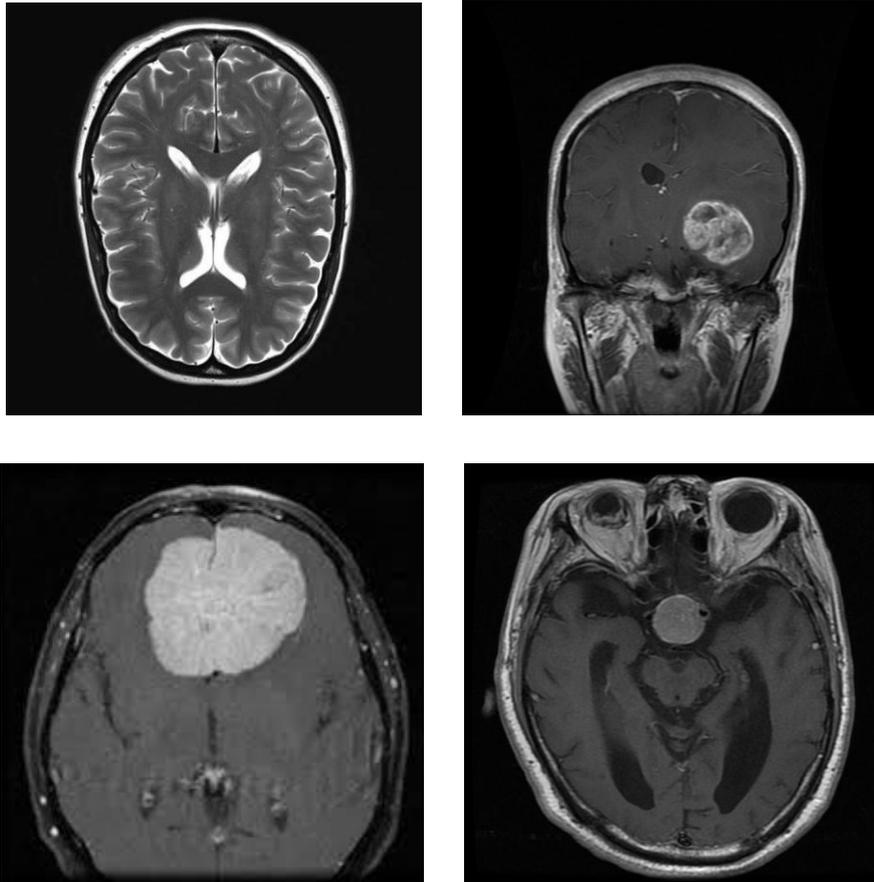


Figura 4.2 – Imagens do *Dataset 2*: cérebro normal (acima à esquerda); glioma (acima à direita); meningioma (abaixo à esquerda) e tumor da glândula pituitária (abaixo à direita).

Os algoritmos de ML requerem que uma percentagem dos dados seja usada como dados de treino e o restante para então se fazerem os testes. Com cada um dos três *datasets* foram testados os mesmos algoritmos com três *splits* (divisões) diferentes, para ver como o aumento dos dados de treino influencia nas métricas de classificação: 70%, 80% ou 90% dos dados para treino e os restantes 30%, 20% e 10% para testes.

4.2. Algoritmos

A proposta aqui apresentada é a classificação de imagens de tumores cerebrais em diversas classes, sendo que cada imagem corresponde a uma única classe. Ou seja, não se trata de uma classificação binária, na qual a proposta seria apenas que o algoritmo previsse se é ou não tumor, mas também a de colocar a imagem na categoria correta. Não há um guia ou teoria definitivos sobre como escolher o tipo de algoritmo mais adequado a cada situação; o que é geralmente recomendado é que se façam vários testes para descobrir qual o tipo de algoritmo (e parâmetros deste algoritmo)

resultam no melhor desempenho para uma dada tarefa (Brownlee, 2020). Dessa forma, foram postos em teste alguns dos algoritmos de preferência geral para o desempenho da tarefa de classificação multiclasse, sendo que a escolha foi baseada principalmente tendo em consideração a escolha feita em alguns dos artigos (Amien, Adb-elrehman, & Ibrahim; Amin, Sharif, Yasmin, & Fernandes; Zacharaki, *et al.*; Amin, Sharif, Raza, Saba, & Anjum) analisados aquando da fase de fundamentação teórica.

Os algoritmos usados para a classificação das imagens são o *Decision Tree* (DT) – Árvore de Decisão; *Random Forest* (RF) – Floresta Aleatória; *Linear Discriminant Analysis* (LDA) – Análise Discriminante Linear; *Support Vector Machine* (SVM –Máquina Vetor de Suporte) e *k-Nearest Neighbors* (*k*-vizinhos mais próximos). Vejamos de seguida cada um deles:

4.2.1 *Decision Tree*

Neste estudo, os parâmetros utilizados para este algoritmo foram os pré-definidos (SciKit Learn, s.d.), de modo que a implantação na plataforma *Google Colaboratory* foi a apresentada na Figura 4.3.

```
▶ classifierDT = tree.DecisionTreeClassifier()  
classifierDT.fit(X_train, Y_train)  
ClassificationDT = classifierDT.predict(X_test)  
print("Acurácia: %.3f" % (accuracy_score(ClassificationDT, Y_test)*100))
```

Figura 4.3 – Implementação do algoritmo de classificação *Decision Tree*.

4.2.2 *Random Forest*

O parâmetro variado foi o número de árvores, sendo que os testes foram feitos com 100, 1000, 2000 e 5000 árvores. Os demais parâmetros utilizados foram os que estão definidos por defeito. Dessa forma, a implementação do algoritmo foi a que se apresenta na Figura 4.4.

```

▶ classifierRF5000 = RandomForestClassifier(n_estimators=5000)
classifierRF5000.fit(X_train, Y_train)
ClassificationRF5000 = classifierRF5000.predict(X_test)
print("Acurácia: %.3f" % (accuracy_score(ClassificationRF5000, Y_test)*100))

```

Figura 4.4 – Implementação do algoritmo de classificação *Random Forest* com 5000 árvores.

4.2.3 Linear Discriminant Analysis

Conforme explicado na Seção 3.3.3, esse algoritmo não requer o ajuste de quaisquer parâmetros. Por esse motivo, a implementação deste algoritmo foi feita tal como se indica na Figura 4.5.

```

▶ clfLDA = LinearDiscriminantAnalysis()
clfLDA.fit(X_train, Y_train)
ClassificationLDA = clfLDA.predict(X_test)
print("Acurácia: %.3f" % (accuracy_score(ClassificationLDA, Y_test)*100))

```

Figura 4.5 – Implementação do algoritmo LDA.

4.2.4 Support Vector Machine

Para todos os *kernels*, os valores testados do parâmetro *C* foram 1 (valor de defeito), 10, 50 e 100, e para o parâmetro *gamma* foi utilizado o valor de defeito que é o *scale* ($1 / (\text{features} * X.\text{var}())$). Os demais parâmetros utilizados também foram os de defeito. Nas Figuras 4.6, 4.7, 4.8 e 4.9 são apresentadas, respectivamente, as implementações com *kernels* linear, polinomial, gaussiano e sigmoide:

```

▶ C = 1.0
clflinear = svm.SVC(kernel='linear', C=C)
clflinear.fit(X_train, Y_train)
ClassificationSVMlinear = clflinear.predict(X_test)
print("Acurácia: %.3f" % (accuracy_score(ClassificationSVMlinear, Y_test)*100))

```

Figura 4.6 – Implementação do algoritmo SVM com *kernel* linear.

```

▶ C = 1.0
clfpoli = svm.SVC(kernel='poly', C=C)
clfpoli.fit(X_train, Y_train)
ClassificationSVMpoli = clfpoli.predict(X_test)
print("Acurácia: %.3f" %(accuracy_score(ClassificationSVMpoli, Y_test)*100))

```

Figura 4.7 – Implementação do algoritmo SVM com *kernel* polinomial.

```

▶ C = 1.0
clfrbf = svm.SVC(kernel='rbf', C=C)
clfrbf.fit(X_train, Y_train)
ClassificationSVMrbf = clfrbf.predict(X_test)
print("Acurácia: %.3f" %(accuracy_score(ClassificationSVMrbf, Y_test)*100))

```

Figura 4.8 – Implementação do algoritmo SVM com *kernel* RBF gaussiano.

```

▶ C = 1.0
clfsig = svm.SVC(kernel='sigmoid', C=C)
clfsig.fit(X_train, Y_train)
ClassificationSVMsig = clfsig.predict(X_test)
print("Acurácia: %.3f" %(accuracy_score(ClassificationSVMsig, Y_test)*100))

```

Figura 4.9 – Implementação do algoritmo SVM com *kernel* sigmoide.

4.2.5 *k*-Nearest Neighbors (*k*NN)

O valor por defeito desse algoritmo são 10 *pixels* vizinhos. Foram testados os valores de 10, 5 e 3 vizinhos para cada um dos três *datasets*; na Figura 4.10, é exemplificada a implementação para 10 *pixels* vizinhos.

```

▶ clfKNN10 = KNeighborsClassifier(n_neighbors=10)
clfKNN10.fit(X_train, Y_train)
ClassificationKNN10 = clfKNN10.predict(X_test)
print("Acurácia: %.3f" %(accuracy_score(ClassificationKNN10, Y_test)*100))

```

Figura 4.10 – Implementação do algoritmo *k*NN com 10 vizinhos.

4.3. Plataforma e Código

Todos os algoritmos foram testados no *Google Colaboratory*, que é “um serviço de nuvem gratuito hospedado pelo próprio Google para incentivar a pesquisa de

Aprendizagem Máquina e Inteligência Artificial” (Santos, 2020). Esta plataforma centra-se na “criação e execução de códigos em *Python*, diretamente num navegador, sem a necessidade de nenhum tipo de instalação de *software* numa máquina” (Roveda, 2019) – ver Figura 4.11.



Figura 4.11 – Interface da plataforma *Google Colaboratory*.

Nesta plataforma, é possível indicar linhas de código ou de texto, as quais são executadas como se se estivesse a utilizar uma máquina local. Para utilizar a plataforma, é apenas necessário ter uma conta Google. É possível também permitir o acesso da plataforma ao Google Drive do usuário, de forma que não é necessário estar a fazer *upload* dos ficheiros a cada sessão.

O código utilizado foi adaptado de Kucukcirak (2021). O código original foi criado para a classificação de tumores cerebrais usando *Deep Learning*, que é um tipo de algoritmo diferente mas que pode ser usado com o mesmo fim. O código encontra-se no Apêndice A – Código utilizado.

4.4. Métricas

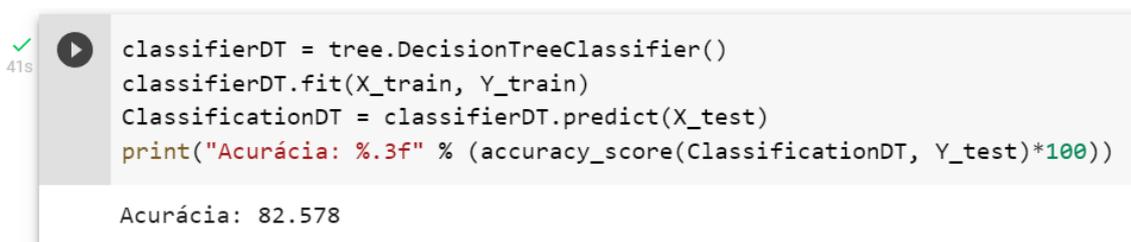
A principal métrica usada nos artigos analisados na Secção 3.4.1 é a acurácia, que é definida como a quantidade de verdadeiros positivos e negativos (TP e TN – *True Positives* e *True Negatives*) a dividir pela quantidade total de verdadeiros positivos e negativos mais os falsos positivos e negativos (FP e FN – *False Positives* e *False Negatives*), segundo a equação:

$$Acc = \frac{TP + TN}{TN + FN + TP + FP}$$

Equação 4.1 – Cálculo da Acurácia.

Um resultado “Verdadeiro Positivo” é contabilizado quando uma lesão é classificada como pertencendo à classe a que de facto pertence; um “Verdadeiro Negativo” ocorre quando a lesão é classificada como não pertencendo a uma classe que de facto não pertence; um “Falso Positivo” é quando a lesão é classificada como pertencendo a uma classe que não pertence; e o “Falso Negativo” é quando uma lesão é classificada como não pertencente a uma classe que na verdade pertence (Mariano, 2021). A acurácia representa a proximidade do resultado com o valor de referência. Assim, quanto maior a acurácia, mais próximo da realidade (ou do valor de referência) é o resultado encontrado. Níveis de acurácia entre 30% e 90% são considerados médios, o que representam níveis moderados de risco de que os resultados não sejam condizentes com os valores reais. Níveis acima de 90% representam alta precisão e baixo risco, sendo que o objetivo é atingir esse patamar de acurácia (González, 2019).

A acurácia é obtida diretamente por meio de um comando após a classificação feita por cada algoritmo. Na Figura 4.12 encontra-se ilustrada a obtenção da acurácia com o algoritmo *Decision Tree*.



```
✓ 41s ▶ classifierDT = tree.DecisionTreeClassifier()
classifierDT.fit(X_train, Y_train)
ClassificationDT = classifierDT.predict(X_test)
print("Acurácia: %.3f" % (accuracy_score(ClassificationDT, Y_test)*100))

Acurácia: 82.578
```

Figura 4.12 – Interface dos *outputs* do algoritmo *Decision Tree* no Google Colaboratory.

Para a complementação da acurácia, apresenta-se para alguns dos resultados a Matriz Confusão, que é uma tabela na qual constam as frequências de classificação para cada classe (Souza, 2019). No caso de um problema multiclasse, ao invés de apresentar os Verdadeiros Positivos e Negativos e Falsos Positivos e Negativos, vai cruzar as informações por classe. Assim, do total de elementos da classe 1, por exemplo, será mostrado quantos de facto foram classificados como pertencendo à classe 1, e quantos às classes 2, 3 e 4. Depois, é realizado o mesmo procedimento para as restantes classes, sempre em confronto com as demais. Assim, é possível perceber onde exatamente estão a ocorrer os maiores erros de classificação do algoritmo. Na Figura 4.13 está exemplificada uma Matriz Confusão como as que serão apresentadas posteriormente. Na primeira linha vê-se, respetivamente, os gliomas que foram classificados corretamente como gliomas, e os que foram classificados incorretamente

como meningiomas, sem tumor ou tumor da glândula pituitária. As demais linhas seguem esse padrão.

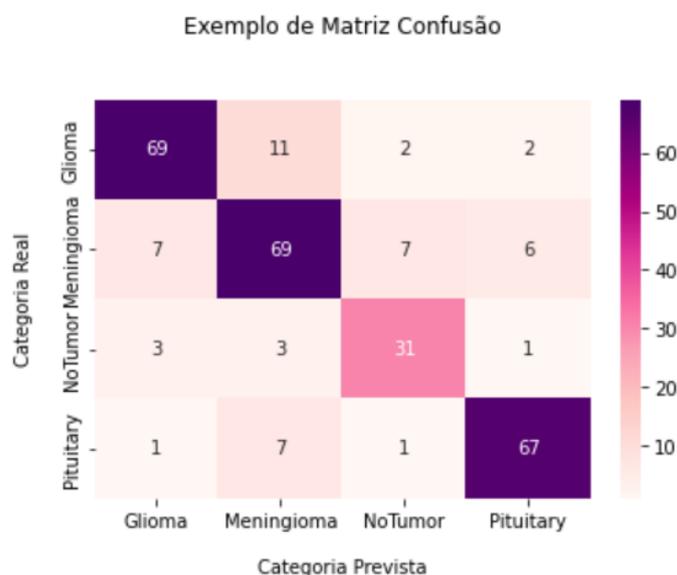


Figura 4.13 – Exemplo de Matriz Confusão multiclasse.

Além da acurácia e da matriz confusão, são apresentadas nas Tabelas 5.1 a 5.15 a medida de tendência central (Média) e duas de dispersão (Desvio Padrão e Amplitude dos dados) que foram calculadas por meio do Microsoft Excel. As medidas de dispersão são parâmetros estatísticos calculados para que se possa determinar o quão variáveis os dados de um conjunto são (Gouveia, s.d.). As definições das medidas apresentadas são:

- Média: é a soma de todos os valores do conjunto a dividir pelo número de elementos do conjunto.
- Amplitude: a diferença entre a maior e a menor medida de um conjunto de dados;
- Desvio padrão: é a raiz quadrada da variância, que por sua vez é a medida dos quadrados das diferenças das observações e a média do conjunto:

$$\sigma^2 = \frac{\sum(x_i - \bar{x})^2}{n}$$

Equação 4.2 – Desvio padrão de um conjunto de medidas.

Além dessas medidas, para que se pudesse de facto depreender a relevância entre as diferenças nas medidas, foram realizados testes de Significância Estatística entre grupos de medidas obtidas por algoritmos diferentes em um mesmo *split* e entre *datasets* diferentes.

O teste realizado foi o de *t*-Student, que é um teste de hipótese utilizado para avaliar se há diferença significativa entre as médias de dois conjuntos de dados (Franco, 2015). Os testes estatísticos também foram feitos no Microsoft Excel. As distribuições dos dados foram testadas em uma ferramenta *online* que realiza o teste de Shapiro-Wilk a respeito da normalidade da distribuição dos dados (Statistics Kingdom, s.d.). A depender da distribuição dos dados (se simétrica ou assimétrica), o teste realizado foi uni ou bicaudal, o que será especificado em cada caso.

A significância estatística foi testada para $\alpha = 0,01$, sendo que a hipótese nula (hipótese a ser testada é que não há diferença entre as médias dos dados). Dessa forma, sempre que o *p-value* for menor que α , rejeita-se a hipótese nula – ou seja, assume-se que há diferença estatisticamente significativa entre as médias – e quando for maior ou igual a α , aceita-se a hipótese nula – de que não há diferença significativa entre as médias.

5. Resultados

Nesta secção serão apresentados os resultados da classificação das imagens dos três *datasets* para cada um dos algoritmos e seus parâmetros em termos de acurácia, bem como as medidas de dispersão média, desvio padrão e amplitude.

5.1 *Decision Tree*

Para o algoritmo *Decision Tree*, foram testados os três *datasets*, cada um com os *splits* 70%, 80% e 90% dos dados usados para treino. Abaixo encontra-se o Gráfico 5.1 com as acurácias obtidas para cada um dos *splits*.

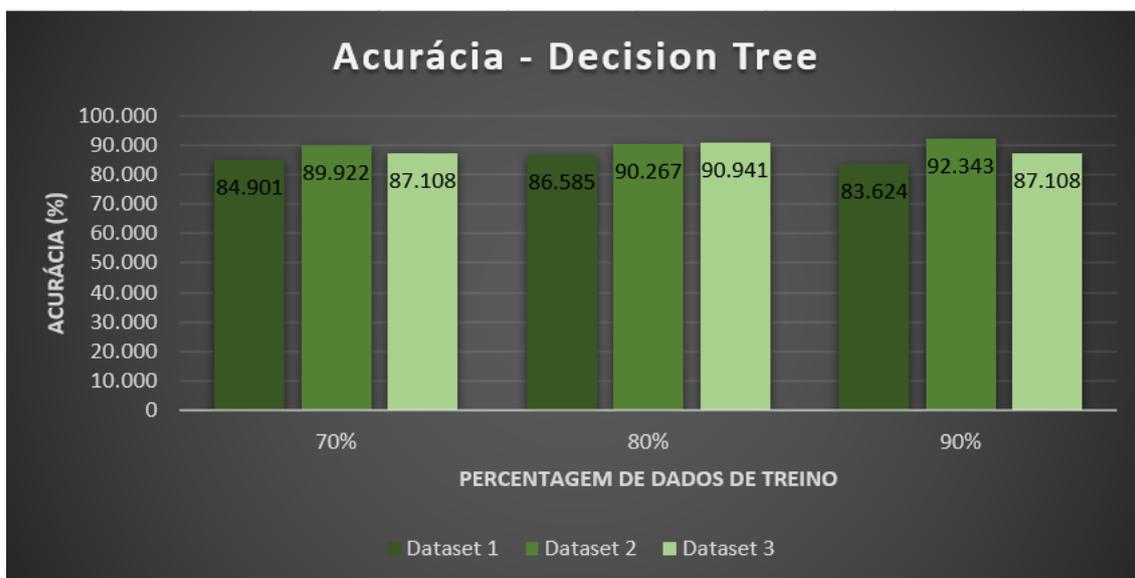


Gráfico 5.1 – Acurácia do algoritmo *Decision Tree* para os três *datasets* e os três *splits*.

Para dois dos três casos (70% e 90%), a melhor acurácia foi obtida com o *dataset 2*, que é constituído por mais imagens (7020), com o melhor resultado no *split* 90%, no qual se observa a acurácia de 92,343%. A menor diferença no *split* de 70% é de 2,814% (entre o D2 e o D3) e a maior é de 5,021% (entre o D1 e o D3). Para o *split* de 90%, a diferença mínima é de 5,235% (entre o D2 e o D3) e máxima de 8,719% (entre o D1 e o D3). E para o *split* de 80%, a menor diferença foi de menos de 0,7% (entre o D2 e o D3) e a maior foi de 4,356% (entre o D1 e o D3). O *dataset 1* atingiu a menor acurácia nos três *splits* testados, sendo a menor para o *split* 90%, em que se alcançou o valor de 83,624%.

Na Tabela 5.1 encontram-se os resultados obtidos para este algoritmo para os três *datasets* (D1, D2 e D3) mais as medidas de dispersão média, desvio padrão (DP) e amplitude dos resultados.

Tabela 5.1 – Resultados e medidas de dispersão para acurácia de *Decision Tree*.

<i>Split</i>	D1 (%)	D2(%)	D3(%)	Média(%)	DP(%)	Amplitude
70%	84.901	89.922	87.108	87.310	2.517	5.021%
80%	86.585	90.267	90.941	89.264	1.914	4.356%
90%	83.624	92.343	87.108	87.692	4.389	8.719%

5.2 Random Forest

Para o algoritmo *Random Forest*, foram testados os três *datasets*, com os três *splits* (70%, 80% e 90%) e quatro variações no parâmetro “estimadores” (ou número de árvores): 100, 1000, 2000 e 5000. Abaixo, encontra-se o Gráfico 5.2 com as acurácias para cada uma das variações no número de estimadores para o *split* 70%.



Gráfico 5.2 – Acurácia do algoritmo *Random Forest* para os três *datasets* no *split* 70%.

Para este *split*, em todos os casos o *dataset* com o qual se obteve a melhor acurácia foi o D3, o segundo melhor o D2 e a menor acurácia foi a do D1. Em todos os casos, foram alcançadas acurácias superiores a 90%, chegando muito próximo de 97% para o caso do D3 com todos os estimadores, e mínimos de cerca de 91% para o D1 com todos valores de estimadores.

Na Tabela 5.2, encontram-se os resultados obtidos para este *split* mais a média, desvio padrão (DP) e amplitude dos resultados consoante o número de estimadores.

Tabela 5.2 – Resultados e medidas de dispersão para acurácia de *Random Forest* – *split* 70%.

Estimadores	D1(%)	D2(%)	D3(%)	Média (%)	DP(%)	Amplitude
100	91.580	95.489	96.733	94.601	2.689	5.153
1000	91.754	95.548	96.951	94.751	2.689	5.197
2000	91.841	95.548	96.821	94.737	2.587	4.980
5000	91.812	95.537	96.864	94.738	2.619	5.052

Abaixo, encontra-se o Gráfico 5.3 com as acurácias para cada uma das variações no número de estimadores para o *split* 80%.



Gráfico 5.3 – Acurácia do algoritmo *Random Forest* para os três *datasets* no *split* 80%.

Para este *split*, assim como no anterior, em todos os casos a maior acurácia foi atingida com o D3, sendo o máximo global de 96,908% com 1000 estimadores, mas com superioridades muito pequenas em relação ao D2. O D1 também obteve a pior performance para todos os números de estimadores desse *split*, embora o mínimo global esteja em 91,855% para 2000 estimadores. Na Tabela 5.3, encontram-se os resultados obtidos para este *split* mais a média, desvio padrão (DP) e amplitude dos resultados consoante o número de estimadores.

Tabela 5.3 – Resultados e medidas de dispersão para acurácia de *Random Forest* – *split* 80%.

Estimadores	D1(%)	D2(%)	D3(%)	Média (%)	DP(%)	Amplitude
100	92.030	95.943	96.733	94.902	2.518	4.703
1000	92.117	96.050	96.908	95.025	2.555	4.791
2000	91.855	95.996	96.864	94.905	2.677	5.009
5000	91.943	96.068	96.864	94.958	2.642	4.921

No Gráfico 5.4 encontram-se as acurácias para o *split* 90%.



Gráfico 5.4 – Acurácia do algoritmo *Random Forest* para os três *datasets* no *split* 90%.

Para este *split*, obteve-se a maior acurácia com o D3 para 3 das 4 variações de parâmetros (100, 1000 e 2000), embora a superioridade em relação ao D2 não seja muito notável (sempre em torno de 0,5%). A máxima global foi alcançada com 2000 estimadores, sendo de 96,951% com o D3. No outro caso (5000 estimadores), a superioridade do D2 foi de quase 5% em relação ao D3. Novamente, como nos outros dois *splits*, o D1 foi o que atingiu menor acurácia em todos os casos, estando aproximadamente entre 81% e 82%. Os resultados, média e medidas de dispersão são apresentados na Tabela 5.4.

Tabela 5.4 – Resultados e medidas de dispersão para acurácia de *Random Forest* – *split* 90%.

Estimadores	D1(%)	D2(%)	D3(%)	Média (%)	DP(%)	Amplitude
100	81.882	96.296	96.821	91.666	8.478	14.939
1000	82.578	96.403	96.864	91.948	8.118	14.286
2000	81.882	96.474	96.951	91.769	8.566	15.069
5000	81.882	96.296	91.725	89.968	7.366	14.414

5.3 Linear Discriminant Analysis

Para o algoritmo LDA, foram testados os três *splits* (70%, 80% e 90%) para cada um dos três *datasets* (D1, D2 e D3). Os parâmetros utilizados foram os de defeito. No Gráfico 5.5, encontram-se os resultados obtidos para a acurácia, em percentagem.

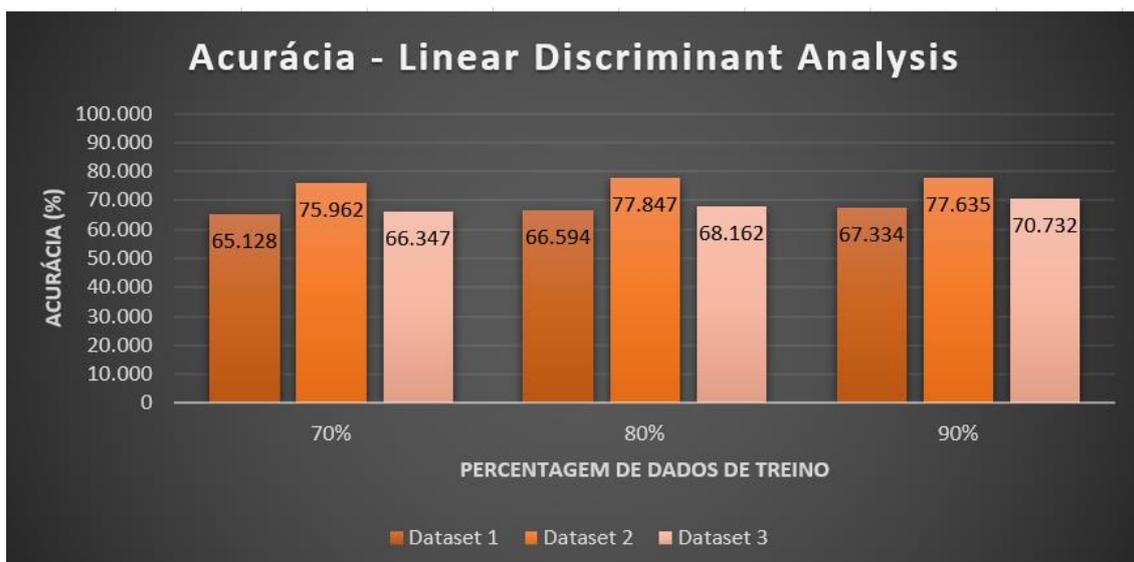


Gráfico 5.5 – Acurácia do algoritmo *Linear Discriminant Analysis* para os três *datasets* e os três *splits*.

Este algoritmo atingiu acurácias mais baixas que os anteriores, com uma máxima global de 77,635% para o D2 com um *split* de 90%. O D2 apresentou as maiores acurácias para os três *splits*, e D1 as menores, com uma mínima global de 65,128% nos *splits* 70% e 80%.

Tabela 5.5 – Resultados e medidas de dispersão para acurácia de *Linear Discriminant Analysis*.

<i>Split</i>	D1 (%)	D2(%)	D3(%)	Média(%)	DP(%)	Amplitude
70%	65.128	75.962	66.347	69.146	5.934	10.834
80%	66.594	77.847	68.162	70.868	6.095	11.253
90%	67.334	77.635	70.732	71.900	5.249	10.301

O algoritmo LDA atingiu valores relativamente baixos de acurácia, e grandes amplitudes nas medidas – ver Tabela 5.5 – ultrapassando os 10% em todos os *splits*, o que resulta em desvios-padrão altos também.

5.4 *Support Vector Machine*

O uso deste algoritmo implica a escolha de um *kernel*, pelo que foram testados quatro *kernels* diferentes, e com cada um quatro variações do Parâmetro de Regularização C (1, 10, 50 e 100), para cada um dos três *splits* (70%, 80% e 90%).

5.4.1 Linear

Com este *kernel*, foram testados os três *splits* com os *datasets* D1 e D3, para quatro valores do parâmetro de regularização C . Não há resultados a apresentar de D2 em razão de, em várias tentativas, demorar cerca dez horas de computação na plataforma, sem que ainda não houvesse sido calculado nenhum resultado. Este *dataset* foi, portanto, considerado inviável para este *kernel*. Para efeito de comparação, o tempo que os outros *kernels* demoraram com esse *dataset* foi pouco mais de 3 horas. Os valores de acurácia obtidos se encontram no Gráfico 5.6.

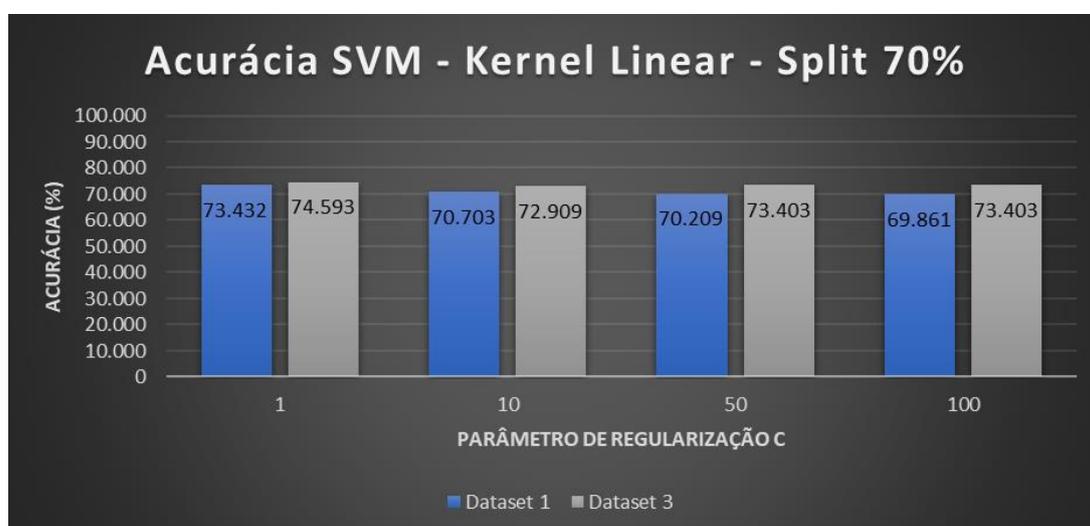


Gráfico 5.6 – Acurácia do *kernel* linear para os *datasets* D1 e D3 no *split* 70% e quatro valores de parâmetro de regularização.

Tabela 5.6 – Resultados e medidas de dispersão para acurácia do *kernel* Linear – *split* 70%.

C	D1(%)	D2(%)	D3(%)	Média (%)	DP(%)	Amplitude
1	73.432	-	74.593	74.013	0.821	1.161
10	70.703	-	72.909	71.806	1.560	2.206
50	70.209	-	73.403	71.806	2.258	3.194
100	69.861	-	73.403	71.632	2.505	3.542

O *kernel* linear apresentou medidas de acurácia relativamente baixas (comparadas às obtidas com os outros *kernels*, mas apresentou uma boa uniformidade, com valores

mais baixos de desvio padrão, como é possível ver na Tabela 5.6, e médias consideravelmente parecidas.

5.4.2 Radial Basis Function (RBF)

Com este *kernel*, foram testados os três *splits* (70%, 80% e 90%) para cada um dos três *datasets*, para quatro valores do parâmetro de regularização C . Os resultados estão no Gráfico 5.7.

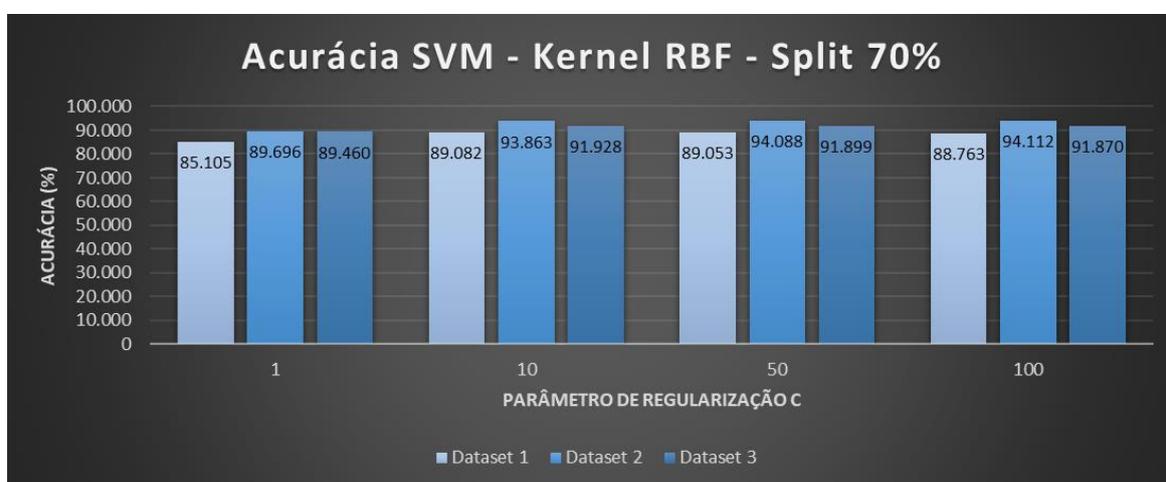


Gráfico 5.7 – Acurácia do *kernel* RBF para os três *datasets* no *split* 70% e quatro valores de parâmetro de regularização.

O D2 alcançou os maiores valores de acurácia para todos os valores de parâmetro de regularização, com uma máxima global neste *split* de 94,112% para $C = 100$, ao passo que o D1 atingiu os menores, com uma mínima global de 85,105 para $C = 1$.

Tabela 5.7 – Resultados e medidas de dispersão para acurácia do *kernel* RBF – *split* 70%.

C	D1(%)	D2(%)	D3(%)	Média (%)	DP(%)	Amplitude
1	85.105	89.696	89.460	88.087	2.585	4.591
10	89.082	93.863	91.928	91.624	2.405	4.781
50	89.053	94.088	91.899	91.680	2.525	5.035
100	88.763	94.112	91.870	91.582	2.686	5.349

A maior amplitude entre a maior e a menor acurácia também foi apresentada para $C = 100$, parâmetro com o qual se conseguiu a maior acurácia, e a menor amplitude foi com o parâmetro $C = 1$, com o qual se conseguiu o menor máximo dentre as variações de C nesse *split* – ver Tabela 5.7.

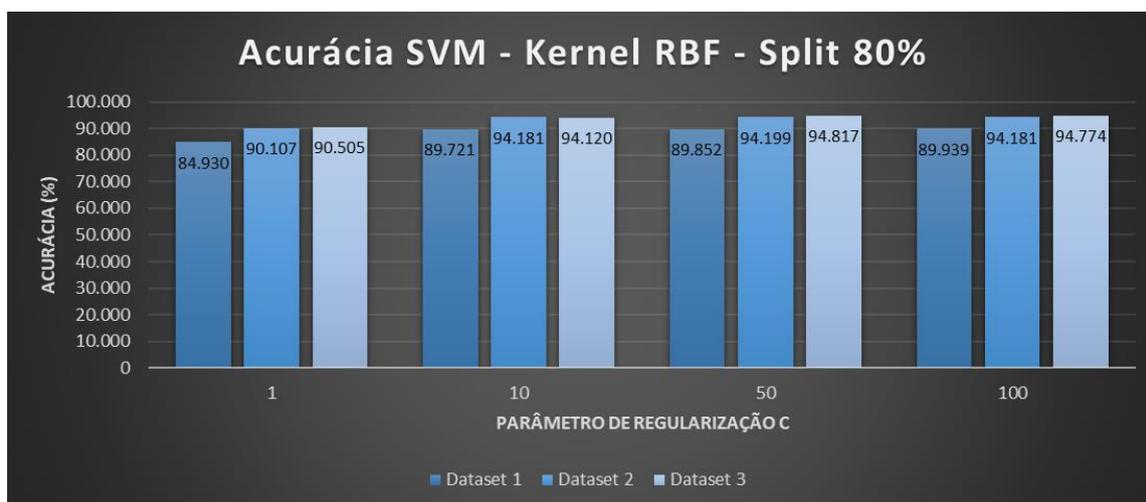


Gráfico 5.8 – Acurácia do *kernel* RBF para os três *datasets* no *split* 80% e quatro valores de parâmetro de regularização.

Neste *split*, foram obtidos valores máximos com o D2 e $C = 10$, e D3 para os outros três casos, com a máxima global de 94,817% com D3 e $C = 50$. Novamente, os valores mais baixos foram os obtidos com D1, com a mínima global de 84,930% para $C = 1$ – conferir Gráfico 5.8.

Tabela 5.8 – Resultados e medidas de dispersão para acurácia do *kernel* RBF – *split* 80%.

C	D1(%)	D2(%)	D3(%)	Média (%)	DP(%)	Amplitude
1	84.930	90.107	90.505	88.514	3.110	5.177
10	89.721	94.181	94.120	92.674	2.558	4.460
50	89.852	94.199	94.817	92.956	2.706	4.347
100	89.939	94.181	94.774	92.965	2.637	4.242

Nesse *split*, os valores de média são um pouco maiores do que nos anteriores, mas a amplitude dos dados e o desvio padrão também, como se apresenta na Tabela 5.8, ainda que tenham sido atingidos valores superiores de amplitude e desvio padrão com outros algoritmos que não conseguiram acurácias tão elevadas.

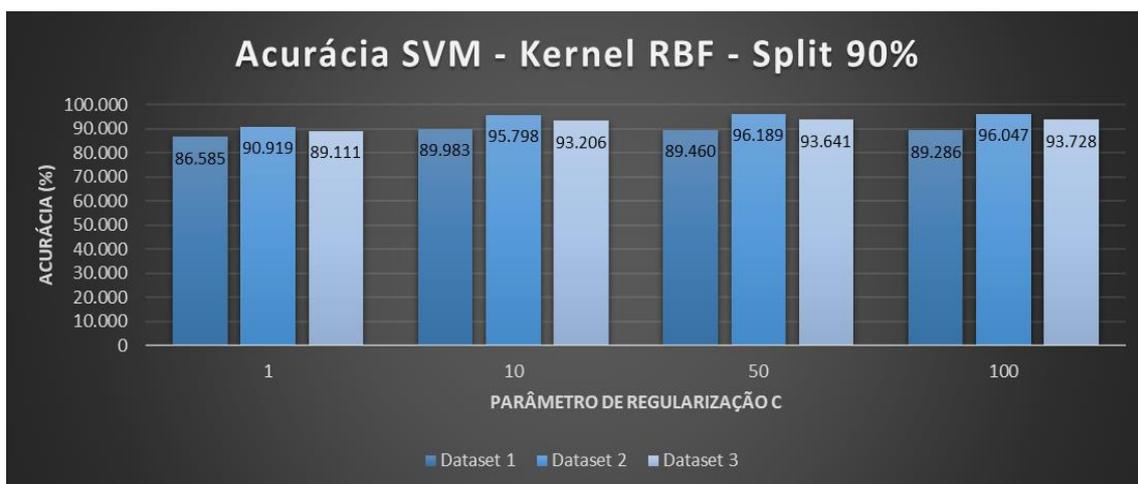


Gráfico 5.9 – Acurácia do *kernel* RBF para os três *datasets* no *split* 90% e quatro valores de parâmetro de regularização.

Neste *split*, os máximos foram obtidos todos com o D2, e os mínimos com o D1 – ver Gráfico 5.9 – sendo que o máximo global foi de 96,189% com $C = 50$ (que é um pouco maior que os outros dois *splits*), e o mínimo global foi de 86,585% (também um pouco maior que as mínimas globais anteriores), com $C = 1$.

Tabela 5.9 – Resultados e medidas de dispersão para acurácia do *kernel* RBF – *split* 90%.

C	D1(%)	D2(%)	D3(%)	Média (%)	DP(%)	Amplitude
1	86.585	90.919	89.111	88.872	2.177	4.334
10	89.983	95.798	93.206	92.996	2.913	5.815
50	89.460	96.189	93.641	93.097	3.397	6.729
100	89.286	96.047	93.728	93.020	3.436	6.761

Por outro lado, o desvio padrão e a amplitude dos dados também foram um pouco maiores para este *split* do que para os anteriores, chegando a quase 3,5% para o DP e mais de 6,5% de amplitude – conferir Tabela 5.9.

5.4.3 Polinomial

Com este *kernel*, foram testados os três *datasets*, e com cada um deles quatro valores diferentes do parâmetro de regularização C . Abaixo, apresenta-se um gráfico para cada um dos três *splits*, e em cada um deles as acurácias conforme os valores de C , com as

respetivas tabelas dos resultados e medidas de dispersão. O grau do polinómio usado foi o 3, que vem por defeito.

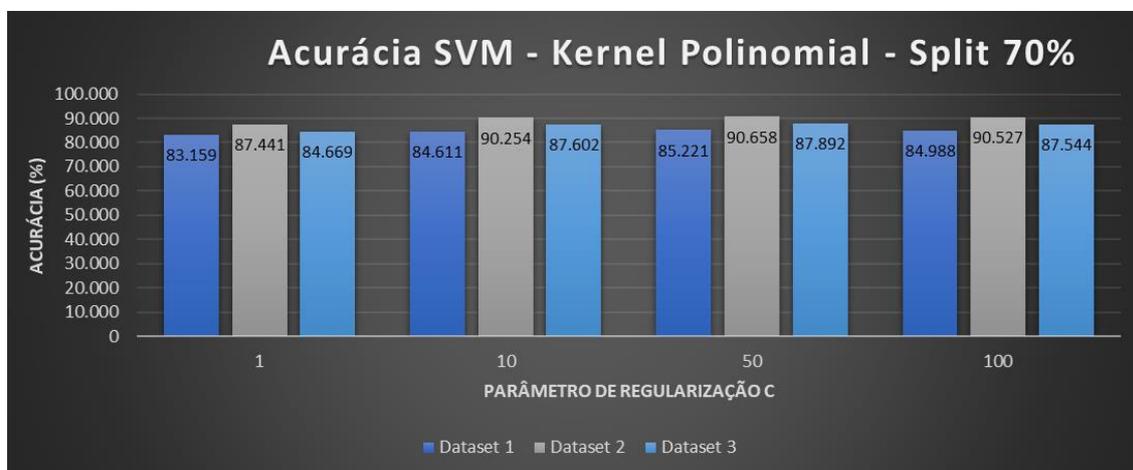


Gráfico 5.10 – Acurácia do *kernel* Polinomial para os três *datasets* no *split* 70% e quatro valores de parâmetro de regularização.

No Gráfico 5.10, vê-se que o D2 obteve a maior acurácia neste *split* com todos os valores de C , com uma máxima global de 90,658% para $C = 50$. O D1 atingiu as menores acurácias em todos os casos, com uma mínima global de 83,159% para $C = 1$, não ultrapassando muito os 85% (para $C = 50$).

Tabela 5.10 – Resultados e medidas de dispersão para acurácia do *kernel* Polinomial – *split* 70%.

C	D1(%)	D2(%)	D3(%)	Média (%)	DP(%)	Amplitude
1	83.159	87.441	84.669	85.090	2.172	4.282
10	84.611	90.254	87.602	87.489	2.823	5.643
50	85.221	90.658	87.892	87.924	2.719	5.437
100	84.988	90.527	87.544	87.686	2.772	5.539

Para este *split*, o desvio padrão e amplitude máximos foram obtidos para $C = 10$, ao passo que a maior média de acurácia foi alcançada com $C = 50$ – conferir Tabela 5.10.

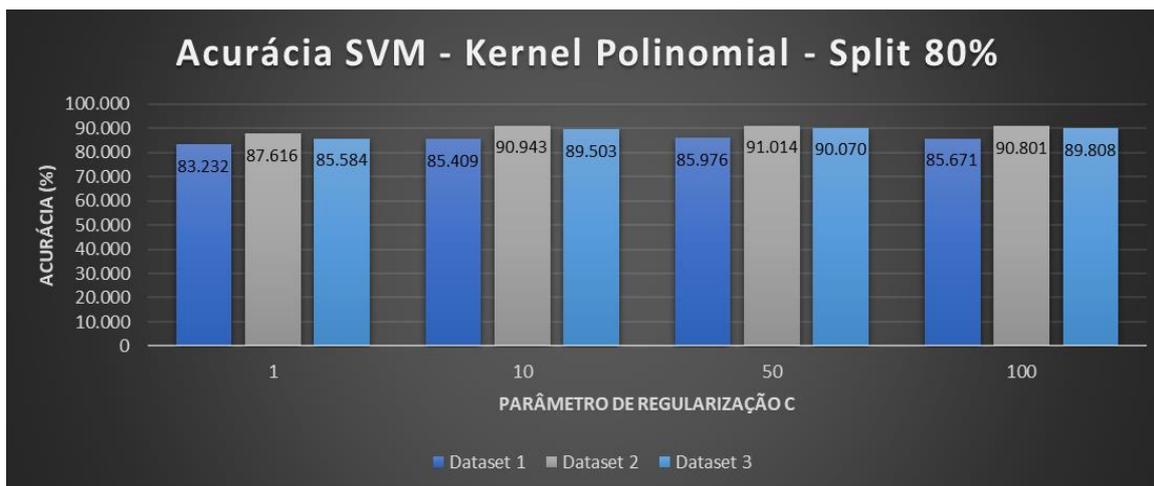


Gráfico 5.11 – Acurácia do *kernel* Polinomial para os três *datasets* no *split* 80% e quatro valores de parâmetro de regularização.

No *split* 80%, cujos resultados se mostram no Gráfico 5.11, os maiores valores para a acurácia foram obtidos com o D2, sendo que a máxima global de 91,014% também se alcançou com $C = 50$. Os valores mais baixos foram todos com o D1, sendo que a mínima global foi de 83,232%, com $C = 1$. O D3 alcançou as acurácias intermediárias em todos os casos e, como se pode observar nos casos anteriores, os valores sempre se aproximam mais das acurácias mais altas de D2 do que as mais baixas de D1.

Tabela 5.11 – Resultados e medidas de dispersão para acurácia do *kernel* Polinomial – *split* 80%.

C	D1(%)	D2(%)	D3(%)	Média (%)	DP(%)	Amplitude
1	83.232	87.616	85.584	85.477	2.194	4.384
10	85.409	90.943	89.503	88.618	2.871	5.534
50	85.976	91.014	90.070	89.020	2.678	5.038
100	85.671	90.801	89.808	88.760	2.721	5.130

Para este *split*, a maior média foi observada para $C = 50$, embora o maior desvio padrão (e amplitude) tenha sido com $C = 10$. As médias também foram maiores do que as do *split* anterior, ainda que os valores de desvio padrão e de amplitude tenham ficado próximos (Tabela 5.11).

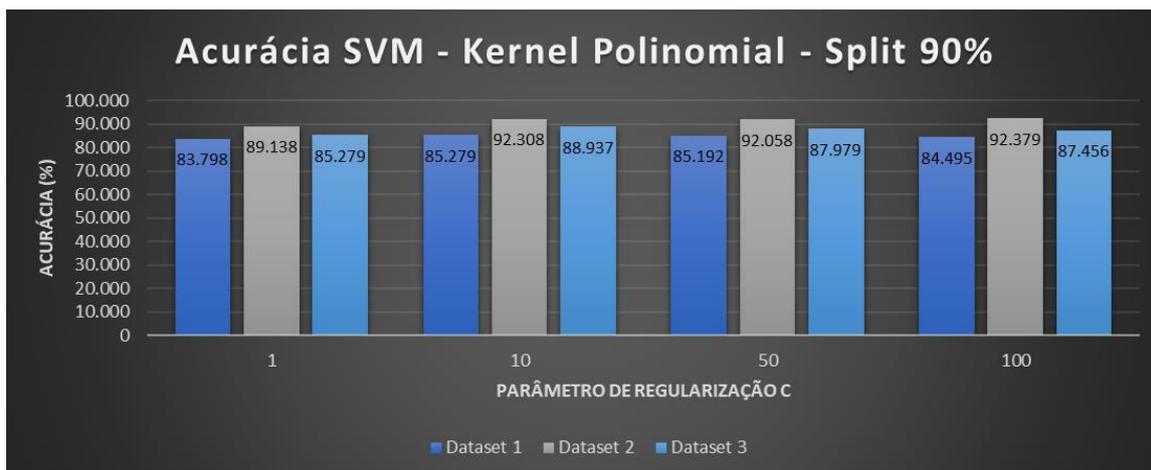


Gráfico 5.12 – Acurácia do *kernel* Polinomial para os três *datasets* no *split* 90% e quatro valores de parâmetro de regularização.

No último *split*, assim como nos dois anteriores, o D2 atingiu as maiores acurácias, seguido pelo D3, e o D1 apresentou as menores (ver Gráfico 5.12). A máxima global foi de 92,379%, com $C = 100$, praticamente igual ao valor de 92,308% conseguido com $C = 10$, diferindo dos *splits* anteriores em que se foram observados com $C = 50$. A mínima global foi de 83,798%, também com $C = 1$ como nos outros dois *splits*.

Tabela 5.12 – Resultados e medidas de dispersão para acurácia do *kernel* Polinomial – *split* 90%.

C	D1(%)	D2(%)	D3(%)	Média (%)	DP(%)	Amplitude
1	83.798	89.138	85.279	86.072	2.757	5.340
10	85.279	92.308	88.937	88.841	3.515	7.029
50	85.192	92.058	87.979	88.410	3.453	6.866
100	84.495	92.379	87.456	88.110	3.982	7.884

Como se observa na Tabela 5.12, as médias das acurácias para cada valor do parâmetro C são relativamente semelhantes ao do *split* anterior, mas com medidas de desvio padrão consideravelmente maiores, uma vez que se pode observar amplitudes maiores entre as máximas e mínimas, de quase 8% no caso de $C = 100$.

5.4.4 *k* – Nearest Neighbors (*k*NN)

Com este algoritmo, foram testados os três *datasets*, com os três *splits*, e três diferentes números de vizinhos: 3, 5 e 10. Os Gráfico 5.13, 5.14 e 5.15 foram feitos segundo o número de vizinhos.

Com 3 vizinhos (Gráfico 5.13), os valores máximos de acurácia foram alcançados com o D2 para os três *splits*, sendo que a máxima global foi de 93,162%, no *split* de 90%. Os valores mínimos também foram todos atingidos com D1, sendo que a mínima global foi de 83,634%, no *split* 90%.

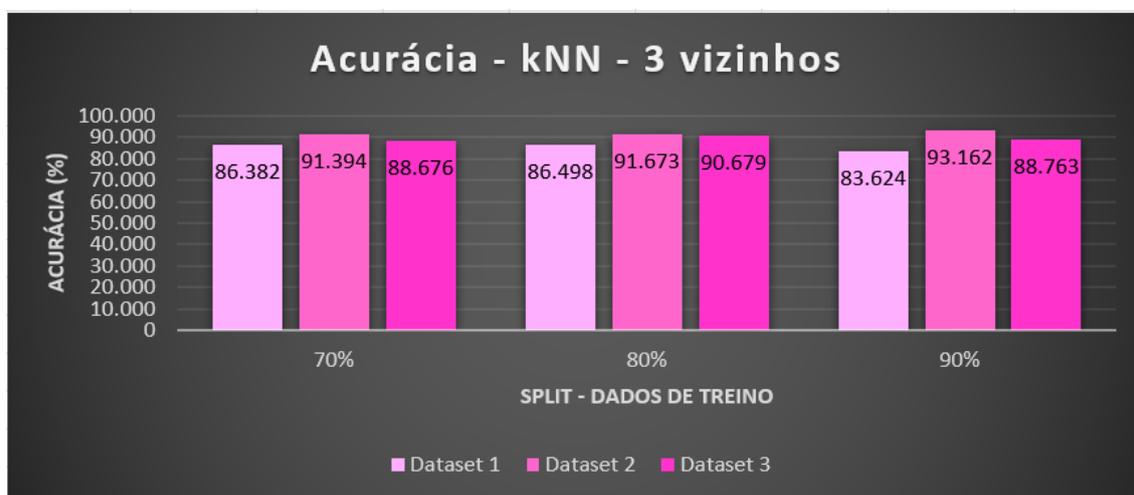


Gráfico 5.13 – Acurácia do kNN para os três *datasets* e os três *splits*, para 3 vizinhos.

Tabela 5.13 – Resultados e medidas de dispersão para acurácia do kNN com 3 vizinhos, para os três *splits*.

<i>Split</i>	D1 (%)	D2(%)	D3(%)	Média(%)	DP(%)	Amplitude
70%	86.382	91.394	88.676	88.817	2.509	5.012
80%	86.498	91.673	90.679	89.617	2.746	5.175
90%	83.624	93.162	88.763	88.516	4.774	9.538

As médias foram bastante semelhantes entre os três *splits* (Tabela 5.13), embora para o de 90% o desvio padrão e a amplitude das medidas sejam quase o dobro dos encontrados nos outros dois *splits*, dada a menor uniformidade dos dados.

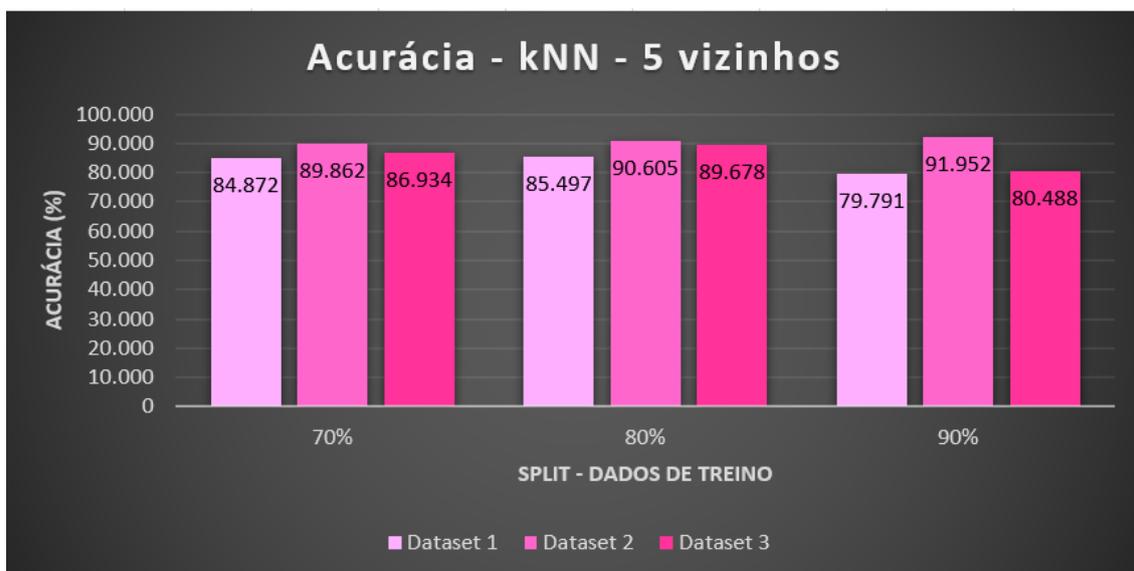


Gráfico 5.14 – Acurácia do kNN para os três *datasets* e os três *splits*, para 5 vizinhos.

Com um número um pouco maior de vizinhos – 5 – manteve-se o padrão de serem apresentadas as maiores acurácias com o D2 e as menores com o D1 (Gráfico 5.14), embora os valores sejam um pouco menores do que com 3 vizinhos. A máxima global ficou em 91,952% no *split* de 90%, e a mínima global ficou em 79,791%, no mesmo *split*.

Tabela 5.14 – Resultados e medidas de dispersão para acurácia do kNN com 5 vizinhos, para os três *splits*.

<i>Split</i>	D1 (%)	D2(%)	D3(%)	Média(%)	DP(%)	Amplitude
70%	84.872	89.862	86.934	87.223	2.507	5.012
80%	85.497	90.605	89.678	88.593	2.721	5.175
90%	79.791	91.952	86.672	86.138	6.098	9.538

As médias são um pouco menores com 5 vizinhos do que com 3, embora o desvio padrão e a amplitude tenham acabado por ter um comportamento parecido: mais baixas nos *splits* 70% e 80%, e a dobrar no *split* 90% (Tabela 5.14), sem que haja aumento da média, mas porque o valor mínimo de acurácia decresce consideravelmente nesse *split*.

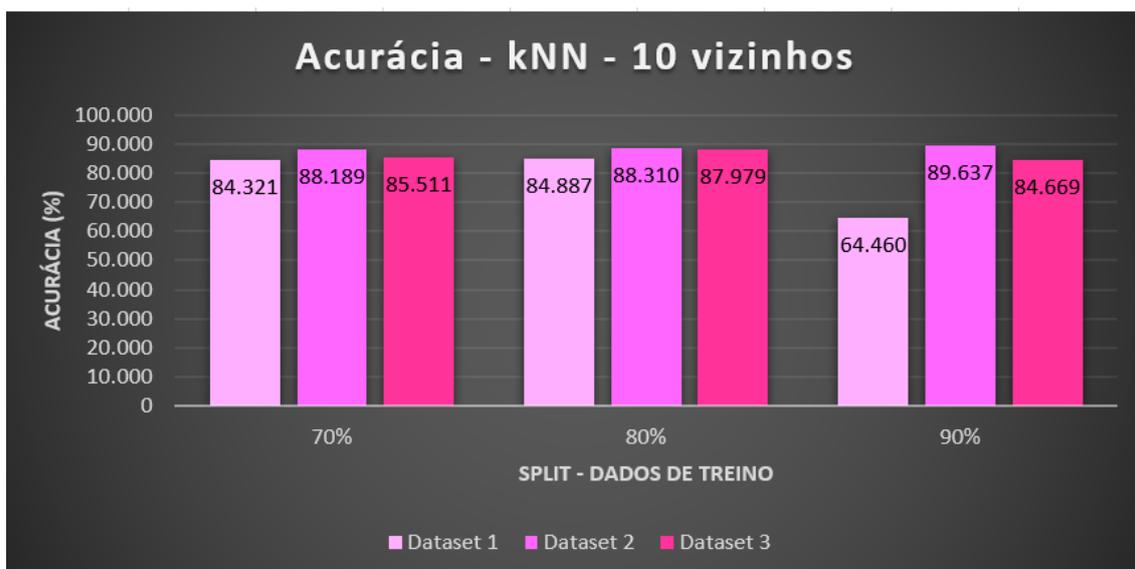


Gráfico 5.15 – Acurácia do kNN para os três *datasets* e os três *splits*, para 10 vizinhos.

Com o maior número de vizinhos – 10 – os valores da acurácia tendem a decrescer, o que foi observado. O D2 obteve as maiores acurácias em todos os *splits*, e D1 as menores, assim como nos casos anteriores. A máxima e a mínima globais também se encontraram no *split* 90%, sendo, respetivamente, 89,637% e 64,460% (Gráfico 5.15), muito abaixo do valor encontrado com 5 vizinhos.

Tabela 5.15 – Resultados e medidas de dispersão para acurácia do kNN com 10 vizinhos, para os três *splits*.

<i>Split</i>	D1 (%)	D2(%)	D3(%)	Média(%)	DP(%)	Amplitude
70%	84.321	88.189	85.511	86.007	1.981	3.868
80%	84.887	88.310	87.979	87.059	1.888	3.423
90%	64.460	89.637	84.669	79.589	13.335	25.177

Obteve-se valores de média, desvio padrão e amplitude bastante discrepantes no *split* 90%, em razão da acurácia do D1 ser muito abaixo que as dos outros *datasets*, e mesmo em relação aos valores obtidos por D1 nos outros dois *splits*, como se vê na Tabela 5.15. O teste foi repetido e confirmado.

6. Discussão de Resultados

Os testes de significância estatística foram feitos dentre os resultados de um mesmo algoritmo, entre resultados de algoritmos diferentes e entre os *kernels* do SVM. Para um mesmo algoritmo, foram comparados: *splits* diferentes e *datasets* diferentes, tomados dois a dois, e dentre parâmetros diferentes, quando existem para o algoritmo em questão. Entre algoritmos diferentes, foram comparados os melhores valores em termos de *dataset* e de *split*, e no caso do SVM, foram comparados os melhores resultados em relação ao parâmetro de regularização C . Em alguns casos, são apresentadas matrizes de confusão para que se perceba o motivo de haver diferença estatisticamente significativa entre resultados. Além desses testes, os resultados foram comparados com o de outros estudos já apresentados na Secção 3.4.1.

A hipótese testada é a de que não há diferença na média entre os *datasets*, *splits*, parâmetros de regularização ou entre os estudos, o que equivale a dizer que não há diferença entre usar um ou outro *dataset*, *split*, valor de parâmetro ou método de estudo. O valor de α utilizado foi de 0,01 (nível de confiança 99%). Quando o valor de *p-value* for menor que α , rejeita-se a hipótese, o que permite concluir que há diferença entre as médias dos conjuntos testados.

6.1 Algoritmos Individuais

6.1.1 *Decision Tree*

Foram feitos testes entre os três *datasets* (dois a dois), para verificar se há diferença na média simples da acurácia na utilização de cada um deles, e entre as médias simples dos três *splits* (também dois a dois). As médias dos *datasets* são as médias simples entre os resultados obtidos com os *splits* de 70%, 80% e 90% para o mesmo *dataset*, e as médias dos *splits* são as médias simples entre os resultados obtidos com D1, D2 e D3 com o mesmo *split*.

Como a única diferença ocorreu na comparação entre os *datasets* 1 e 2 (ver Tabela 6.1) são apresentadas as matrizes confusão dos dois *datasets*, com o mesmo *split* (90%), que foi onde houve maior diferença entre as acurácias.

Tabela 6.1 – Teste de significância estatística para os resultados do algoritmo *Decision Tree*.

Teste (A/B)	Média A	Média B	Tipo de Teste	<i>p-value</i>	H ₀	CONCLUSÃO
D1/D2	85,037	90,844	Bicaudal	0,00708	Rejeita-se	$\mu_2 > \mu_1$
D1/D3	85,037	88,386	Bicaudal	0,11774	Aceita-se	$\mu_1 = \mu_3$
D2/D3	90,844	88,386	Bicaudal	0,19633	Aceita-se	$\mu_2 = \mu_3$
70% / 80%	87,310	89,264	Bicaudal	0,38085	Aceita-se	$\mu_{70\%} = \mu_{80\%}$
70% / 90%	87,310	87,692	Bicaudal	0,90439	Aceita-se	$\mu_{70\%} = \mu_{90\%}$
80% / 90%	89,264	87,692	Bicaudal	0,62217	Aceita-se	$\mu_{80\%} = \mu_{90\%}$

Dessa forma, verifica-se que, apesar de as médias terem valores diferentes, não são estatisticamente diferentes em nenhum dos casos, a não ser na comparação entre os *datasets* D1 e o D2. Neste caso, se rejeita a hipótese de que as médias são iguais, de modo que a média de D1 pode ser considerada menor que a de D2.

Uma das razões para que a média da acurácia seja maior para D2 é porque se trata de um *dataset* diferente, com mais do que o dobro do número de imagens de D1, e com características (como intensidade, gradiente, média e variância dos níveis de cinzento) nas imagens que facilitam a tomada de decisão correta por parte do algoritmo.

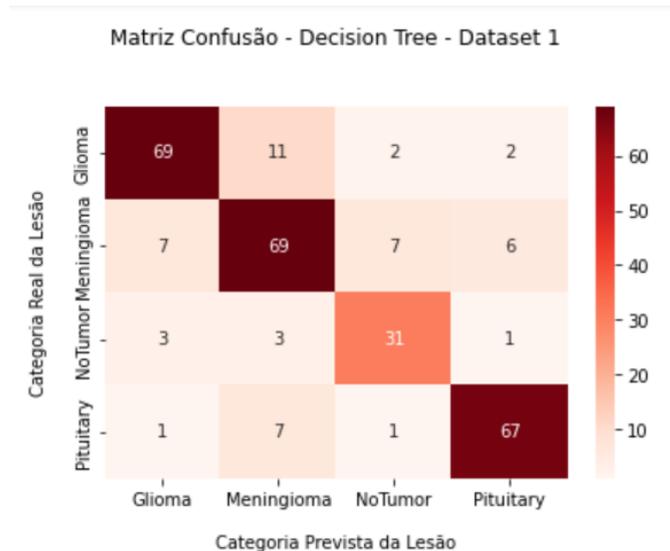


Figura 6.1 – Matriz Confusão para o *Dataset 1* com o algoritmo *Decision Tree* e *split* de 90%.

O *Dataset 1* é constituído por 2870 imagens, de modo que o número de imagens usadas para testar o algoritmo são 10%, ou seja, 287 imagens. Conforme se infere a partir da Figura 6.1 acima, dessas 287 imagens há 84 gliomas (primeira linha), 89 meningiomas (segunda linha), 38 sem tumor (terceira linha) e 76 tumores da glândula pituitária (quarta linha). A diagonal principal da matriz contém as classificações acertadas do algoritmo para cada tipo de lesão.

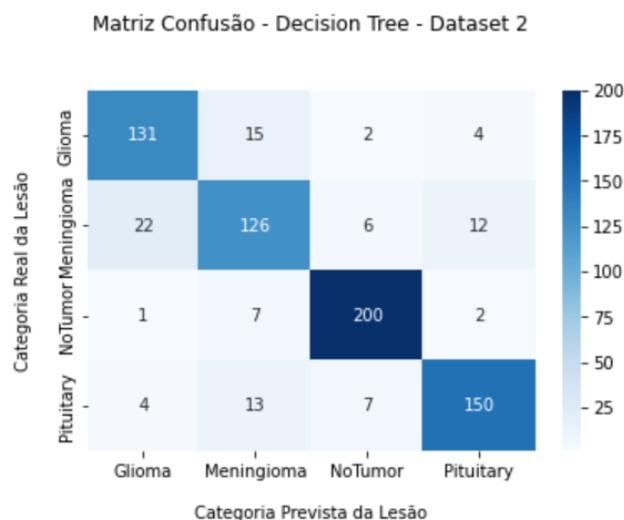


Figura 6.2 – Matriz Confusão para o *Dataset 2* com o algoritmo *Decision Tree* e *split* de 90%.

O *Dataset 2* é constituído por 7020 imagens, de modo que o número de imagens usadas para testar o algoritmo são 702 imagens. Dessas, há 152 gliomas, 166 meningiomas,

210 sem tumor e 174 tumores da glândula pituitária. A Figura 6.2 mostra a matriz confusão para o *Dataset 2*.

Na Tabela 6.2 abaixo, constam os percentuais de acerto de cada *Dataset* por categoria de lesão. O percentual de acerto foi calculado como o número de lesões classificadas corretamente dentro da categoria à qual pertence, a dividir pelo número total de lesões da categoria.

Tabela 6.2 – Percentuais de acerto do algoritmo *Decision Tree* por categoria nos *Datasets 1* e *2*.

Dataset	Glioma	Meningioma	Sem Tumor	Pituitária
D1	82,1%	77,5%	81,6%	88,2%
D2	86,2%	75,9%	95,2%	86,2%

Constata-se, dessa forma, que a maior disparidade nos acertos se dá nas imagens sem tumor, sendo quase 15% de diferença. No caso do meningioma e da pituitária, o algoritmo até atingiu taxas de acerto ligeiramente maiores com o *Dataset 1*, e ligeiramente menor para o glioma, mas a diferença para a categoria Sem Tumor foi o que mais contribuiu para que houvesse uma superioridade estatística para o uso do *Dataset 2* com esse algoritmo.

O erro se deu principalmente com a classificação equivocada de 3 imagens como gliomas (7,9%) e 3 como meningiomas. Para o *Dataset 2*, o principal erro foi a troca de classificação com Meningiomas, que representam apenas 3,5% a menos na taxa de acerto.

6.1.2 *Random Forest*

Foram realizados os testes de significância em pares com os três *datasets*, entre os *splits*, e entre os pares de números de estimadores (árvores), para verificar se um aumento no número de árvores leva a um aumento na acurácia. As médias apresentadas dos *datasets* são as médias simples entre acurácias obtidas com os diferentes *splits* e número de árvores, para um mesmo *dataset*. As médias dos *splits* são as médias simples entre as acurácias obtidas com os diferentes *datasets* e número de árvores, para um mesmo *split*. Da mesma forma, as médias relativamente ao número de árvores são as médias simples calculadas a partir das acurácias obtidas para os

diferentes *datasets* e *splits*, com o mesmo número de árvores. Os resultados encontram-se na Tabela 6.3.

Tabela 6.3 – Testes de significância estatística para os resultados do algoritmo Random Forest

Teste (A/B)	Média A	Média B	Tipo de Teste	<i>p-value</i>	H ₀	CONCLUSÃO
D1/D2	88,596	95,971	Bicaudal	0,000265	Rejeita-se	$\mu_1 < \mu_2$
D1/D3	88,596	96,425	Unicaudal	0,000065	Rejeita-se	$\mu_1 < \mu_3$
D2/D3	95,971	96,425	Bicaudal	0,322632	Aceita-se	$\mu_2 = \mu_3$
70% / 80%	94,706	94,947	Bicaudal	0,794278	Aceita-se	$\mu_{70\%} = \mu_{80\%}$
70% / 90%	94,706	91,338	Bicaudal	0,136454	Aceita-se	$\mu_{70\%} = \mu_{90\%}$
80% / 90%	94,947	91,338	Bicaudal	0,056092	Aceita-se	$\mu_{80\%} = \mu_{90\%}$
100/1000	93,723	93,908	Unicaudal	0,467825	Aceita-se	$\mu_{100} = \mu_{1000}$
100/2000	93,723	93,804	Unicaudal	0,486288	Aceita-se	$\mu_{100} = \mu_{2000}$
100/5000	93,723	93,221	Unicaudal	0,414260	Aceita-se	$\mu_{100} = \mu_{5000}$
1000/2000	93,908	93,804	Unicaudal	0,481888	Aceita-se	$\mu_{1000} = \mu_{2000}$
1000/5000	93,908	93,221	Unicaudal	0,381430	Aceita-se	$\mu_{1000} = \mu_{5000}$
2000/5000	93,804	93,221	Unicaudal	0,401184	Aceita-se	$\mu_{2000} = \mu_{5000}$

Para o algoritmo *Random Forest*, constata-se que o único fator de diferenciação estatística foi o D1, que alcançou uma acurácia menor na comparação com os outros dois *datasets*, D2 e D3. Quanto menor o valor de *p-value*, com mais certeza se pode rejeitar a hipótese nula.

O *dataset* D2 é o que possui a maior quantidade de imagens, o que poderia justificar o facto de alcançar uma acurácia maior que D1. Porém, D3 é um subconjunto de D2, com o mesmo número de imagens de D1, logo a quantidade só por si não explica suficientemente a diferença na acurácia, uma vez que estatisticamente não há diferença entre D2 e D3, que apresentam também números diferentes de imagens. Assim, conclui-se que os resultados são bastante dependentes dos dados empregados.

A utilização de um número bastante maior de estimadores (10, 20 ou 50 vezes maior) não resultou em qualquer acréscimo significativo na acurácia das classificações, ao passo que o tempo despendido em computação aumenta significativamente: com 100 estimadores, o tempo requerido foi de 43 segundos, e com 5000 foi 35 minutos e 47

segundos. Assim, pode-se considerar que o valor de defeito de estimadores (100) é bastante apropriado.

Como houve diferença estatística, dentre as comparações feitas, somente entre os *datasets*, serão analisadas as matrizes confusão dos três *datasets*, utilizando os parâmetros de *split* de 90% por ser onde se observou as maiores diferenças entre as acurácias obtidas e por não haver desigualdade estatística entre esse *split* e os demais, e com o parâmetro de número de árvores igual a 100, por ser mais rápido computacionalmente e por não ter havido diferença no resultado consoante o número de árvores.

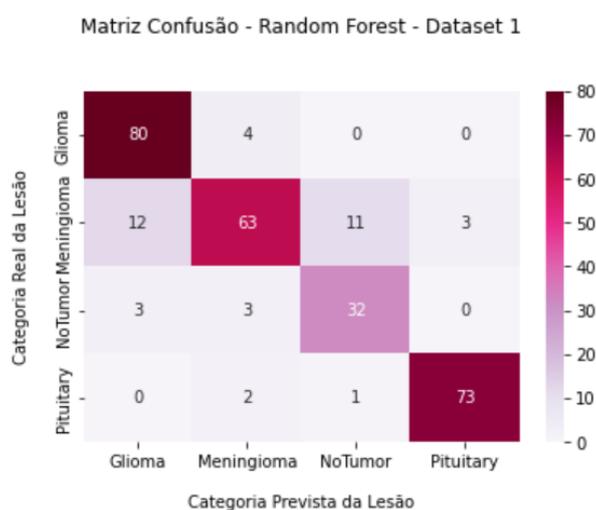


Figura 6.3 – Matriz confusão do *Dataset 1* com o algoritmo *Random Forest* com 100 árvores e *split* de 90%.

As distribuições das lesões por categoria para o *Dataset 1* são 84 gliomas, 89 meningiomas, 38 sem tumor e 76 tumores da glândula pituitária, como se pode verificar na matriz confusão (Figura 6.3).

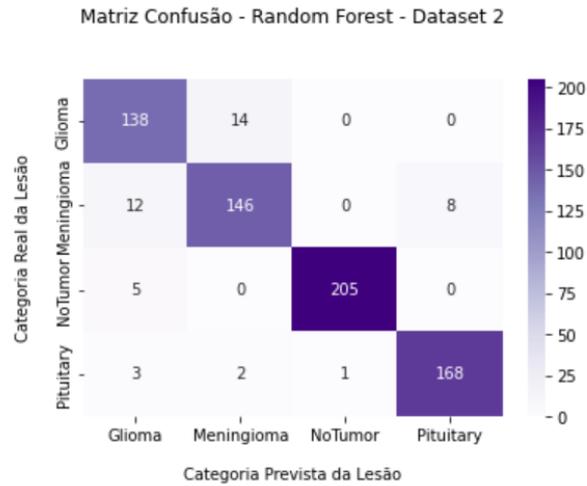


Figura 6.4 – Matriz confusão do *Dataset 2* com o algoritmo *Random Forest* com 100 árvores e *split* de 90%.

Já para o *Dataset 2*, a distribuição das lesões por categorias também se encontra igual à da secção anterior, sendo 152 gliomas, 166 meningiomas, 210 sem tumor e 174 pituitária (Figura 6.4). Por fim, a matriz confusão do *Dataset 3* encontra-se na Figura 6.5.

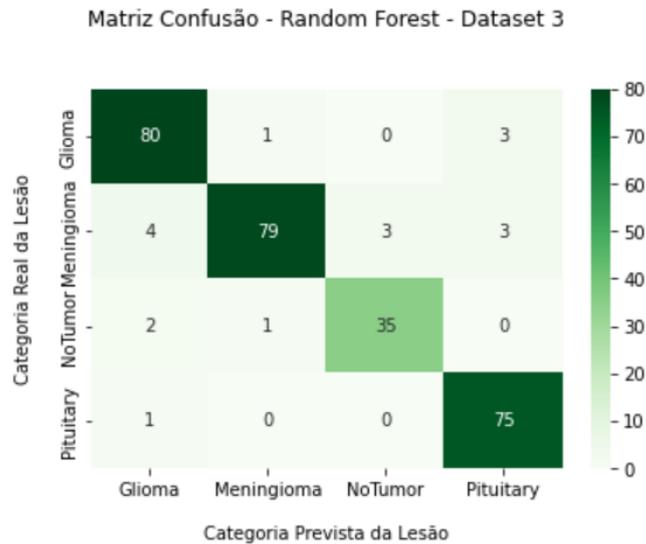


Figura 6.5 – Matriz confusão do *Dataset 3* com o algoritmo *Random Forest* com 100 árvores e *split* de 90%.

O *Dataset 3* possui o mesmo número de imagens que o *Dataset 1*, distribuídas da mesma forma pelas categorias. Abaixo encontram-se os percentuais de acerto do algoritmo para cada *Dataset* divididos por categorias de lesão; ver Tabela 6.4.

Tabela 6.4 – Percentuais de acerto do algoritmo *Random Forest* por categoria nos *Datasets* 1, 2 e 3.

Dataset	Glioma	Meningioma	Sem Tumor	Pituitária
D1	95,2%	70,8%	84,8%	96,1%
D2	90,1%	87,9%	97,6%	96,6%
D3	95,2%	88,8%	92,1%	98,7%

Nas duas primeiras linhas da Tabela 6.3 consta que com o *Dataset* 1 o *Random Forest* atingiu acurácias significativamente menores que com os outros dois, os quais entre si não mostram diferença. Na Tabela 6.4 acima observa-se que os resultados do D1 foram bastante inferiores aos outros dois na categoria Meningioma, em 17,1% e 18%. Nessa categoria, 13,5% dos tumores foram erroneamente classificados como glioma e 12,4% como sem tumor. Os restantes 3,3% correspondem à classificação de tumor da glândula pituitária.

É interessante observar que seria mais esperado haver maior equívoco entre gliomas e meningiomas, em razão da natureza e localização anatômica desses tumores. As classificações dos tumores da glândula pituitária obtiveram excelentes resultados, que podem ser bastante explicados pela localização bastante específica dessa estrutura, como se pode ver na Figura 6.6.

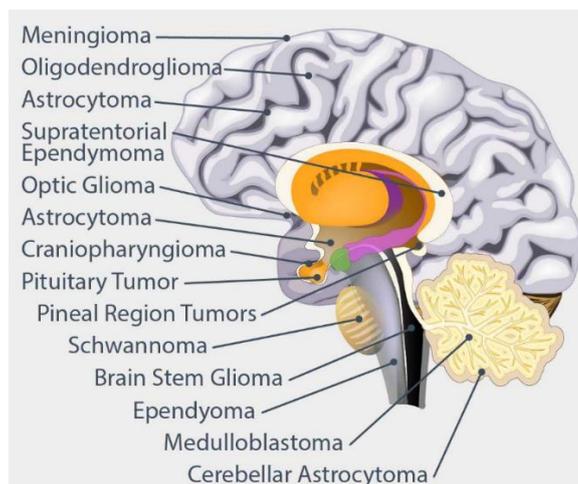


Figura 6.6 – Localização dos principais tumores malignos do cérebro (Miami Neuroscience Center , s.d.).

O que é discrepante é a grande quantidade de imagens de meningiomas que foram classificadas como sem tumor, em uma quantidade praticamente equivalente à classificação errônea como gliomas. Uma razão possível para que isso tenha ocorrido

pode o facto de alguns meningiomas estarem ainda em estágio bastante inicial, sendo de difícil diagnóstico. Outra possibilidade são tumores que, por estarem localizados em regiões que não são típicas de meningiomas, foram colocados nas categorias erradas.

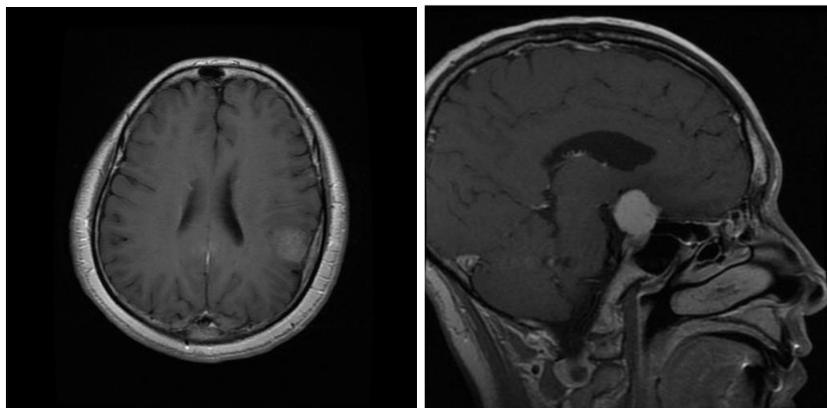


Figura 6.7 – Meningiomas pertencente ao *Dataset 1*.

Na Figura 6.7, a imagem à esquerda mostra uma região de contraste pertencente ao tumor que não é tão evidente quanto outras imagens do *Dataset 1*, o que pode ter causado ao algoritmo classificar meningiomas desse tipo como sendo sem tumor. Na imagem da direita o que fica evidenciado é um meningioma que, segundo o Médico Oncologista Doutor Henrique Reis, por estar localizado em uma região próxima à hipófise e outras estruturas, ao invés de na parte superior do crânio – onde se encontra a maior parte das meninges – pode ter sido classificado erroneamente.

Não pôde ser feita, neste trabalho, a procura por saber exatamente quais imagens acabaram por ser classificadas de maneira equivocada, pelo que as imagens na Figura 6.7 estão a ser utilizadas como ilustração das hipóteses.

6.1.3 LDA

Foram realizados testes de significâncias com resultados dos três datasets, e dos três *splits*. Esse algoritmo não tem parâmetros para ajustar. Na Tabela 6.5, as médias dos *datasets* são as médias simples entre os resultados obtidos com os *splits* de 70%, 80% e 90% para o mesmo *dataset*, e as médias dos *splits* são as médias simples entre os resultados obtidos com D1, D2 e D3 com o mesmo *split*.

Tabela 6.5 – Testes de significância estatística para os resultados do algoritmo LDA.

Teste (A/B)	Média A	Média B	Tipo de Teste	<i>p-value</i>	H ₀	CONCLUSÃO
D1/D2	66,352	77,148	Bicaudal	0,000254	Rejeita-se	$\mu_2 > \mu_1$
D1/D3	66,352	68,414	Bicaudal	0,244456	Aceita-se	$\mu_1 = \mu_3$
D2/D3	77,148	68,414	Bicaudal	0,008387	Rejeita-se	$\mu_2 > \mu_3$
70% / 80%	69,146	70,868	Bicaudal	0,743566	Aceita-se	$\mu_{70\%} = \mu_{80\%}$
70% / 90%	69,146	71,900	Bicaudal	0,579500	Aceita-se	$\mu_{70\%} = \mu_{90\%}$
80% / 90%	70,868	71,900	Bicaudal	0,834921	Aceita-se	$\mu_{80\%} = \mu_{90\%}$

Com este algoritmo, houve diferenças estatísticas que indicam a superioridade do *dataset* D2 em comparação aos dois outros. A diferença entre D2 e D3 é somente o número de imagens, uma vez que esta é um subconjunto daquela, enquanto em relação a D1, D2 é um conjunto diferente e mais numeroso. Assim, pode-se inferir que o número de imagens também é um fator decisivo no desempenho desse algoritmo, além da diferença entre os conjuntos de dados em si. Não se observam diferenças significativas em relação aos *splits* utilizados, o que é equivalente a dizer que, embora as acurácias obtidas pelos *splits* 70%, 80% e 90% variem entre 69,146 e 71,900, a diferença não é suficientemente expressiva para que se justifique o uso de um em detrimento de outro.

O tempo aqui também não é um fator relevante, uma vez que a distinção de *splits* não implica uma variação no tempo de computação, como pode acontecer com algoritmos que requerem o ajuste de parâmetros.

6.1.4 *Support Vector Machine*

Foram realizados testes dentre os parâmetros, *splits* e *datasets* variados internamente aos resultados de cada *kernel*. Para todos os *kernels*, as médias apresentadas dos *datasets* são as médias simples entre acurácias obtidas com os diferentes *splits* e parâmetros de regularização C , para um mesmo *dataset*. As médias dos *splits* são as médias simples entre as acurácias obtidas com os diferentes *datasets* e parâmetro C , para um mesmo *split*. Da mesma forma, as médias relativamente ao parâmetro C são as médias simples calculadas a partir das acurácias obtidas para os diferentes *datasets* e *splits*, com o mesmo valor de C .

6.1.4.1 Linear

Foram realizados os testes de significância (ver Tabela 6.6) para os resultados obtidos com os *datasets* D1 e D3, com os três *splits* e para cada um dos valores do parâmetro de regularização C . Os testes estatísticos foram realizados com cada um dos resultados dos pares de *datasets*, *splits* e parâmetro C .

Tabela 6.6 – Testes de significância estatística para os resultados do *kernel* linear do SVM.

Teste (A/B)	Média A	Média B	Tipo de Teste	<i>p-value</i>	H_0	CONCLUSÃO
D1/D3	70,504	73,869	Unicaudal	0,000007	Rejeita-se	$\mu_1 < \mu_3$
70% / 80%	72,314	72,501	Bicaudal	0,877171	Aceita-se	$\mu_{70\%} = \mu_{80\%}$
70% / 90%	72,314	71,744	Bicaudal	0,575289	Aceita-se	$\mu_{70\%} = \mu_{90\%}$
80% / 90%	72,501	71,744	Bicaudal	0,558832	Aceita-se	$\mu_{80\%} = \mu_{90\%}$
$C = 1/C = 10$	74,213	71,845	Bicaudal	0,039094	Aceita-se	$\mu_1 = \mu_{10}$
$C = 1/C = 50$	74,213	71,455	Bicaudal	0,032547	Aceita-se	$\mu_1 = \mu_{50}$
$C = 1/C = 100$	74,213	71,242	Bicaudal	0,029750	Aceita-se	$\mu_1 = \mu_{100}$
$C = 10/C = 50$	71,845	71,455	Bicaudal	0,736550	Aceita-se	$\mu_{10} = \mu_{50}$
$C = 10/C=100$	71,845	71,242	Bicaudal	0,628464	Aceita-se	$\mu_{10} = \mu_{100}$
$C = 50/C=100$	71,455	71,242	Bicaudal	0,877913	Aceita-se	$\mu_{50} = \mu_{100}$

O D1 obteve resultados significativamente menores que D3, com um *p-value* bastante baixo. Em outras palavras, é possível dizer com uma relevância bastante superior a 99% (valor de significância testado) que a acurácia alcançada com o primeiro *dataset* é inferior ao do terceiro. Ambos possuem o mesmo número de imagens, com a distribuição exatamente igual dentre as categorias.

O algoritmo não se mostrou sensível a todas as demais variações (valores diferentes de *split* e de parâmetro de regularização). O *p-value* encontrado nas comparações entre o parâmetro $C = 1$ e os demais é consideravelmente mais baixo do que os encontrados nas comparações dentre os demais valores (entre o 10, 50 e 100). Quanto mais alto o *p-value* (valor máximo = 1), mais forte a evidência para se apoiar a hipótese alternativa – nesse caso, hipótese de que há diferença nas médias – e os valores bastante elevados encontrados nas comparações entre $C = 10$, $C = 50$ e $C = 100$ suportam muito fortemente a hipótese alternativa. Embora os *p-value* encontrados para $C = 1$ suportem a hipótese

alternativa com pelo menos 99% de confiança, apresentam uma evidência bem menor do que os outros valores.

Desse modo, o valor de defeito do algoritmo ($C = 1$) pode ser considerado razoável, embora não se deva excluir testes com valores maiores.

6.1.4.2 RBF

O *kernel* RBF mostrou-se sensível quanto à variação dos dados, apresentando resultados inferiores na comparação entre o D1 e os outros dois *datasets*, com uma significância superior a 99%. Não se pode assumir que houve diferença entre os resultados de D2 e D3, o que pode indicar que o número de imagens (considerando que vêm de um mesmo conjunto) não é um fator relevante. Os *splits* usados também não se mostraram um fator relevante, exibindo nas comparações *p-values* relativamente altos.

Tabela 6.7 – Testes de significância estatística para os resultados do *kernel* RBF do SVM

Teste (A/B)	Média A	Média B	Tipo de Teste	<i>p-value</i>	H ₀	CONCLUSÃO
D1/D2	88,480	93,615	Unicaudal	0,000002	Rejeita-se	$\mu_1 < \mu_2$
D1/D3	88,480	92,422	Unicaudal	0,000023	Rejeita-se	$\mu_1 < \mu_3$
D2/D3	93,615	92,422	Bicaudal	0,175512	Aceita-se	$\mu_2 = \mu_3$
70% / 80%	90,743	91,777	Bicaudal	0,390754	Aceita-se	$\mu_{70\%} = \mu_{80\%}$
70% / 90%	90,743	91,996	Bicaudal	0,311396	Aceita-se	$\mu_{70\%} = \mu_{90\%}$
80% / 90%	91,777	91,996	Bicaudal	0,865664	Aceita-se	$\mu_{80\%} = \mu_{90\%}$
C = 1/C = 10	88,491	92,431	Bicaudal	0,002568	Rejeita-se	$\mu_1 < \mu_{10}$
C = 1/C = 50	88,491	92,578	Bicaudal	0,002863	Rejeita-se	$\mu_1 < \mu_{50}$
C = 1/C = 100	88,491	92,522	Bicaudal	0,003384	Rejeita-se	$\mu_1 < \mu_{100}$
C = 10/C = 50	92,431	92,578	Bicaudal	0,902236	Aceita-se	$\mu_{10} = \mu_{50}$
C = 10/C=100	92,431	92,522	Bicaudal	0,939675	Aceita-se	$\mu_{10} = \mu_{100}$
C = 50/C=100	92,578	92,522	Bicaudal	0,964857	Aceita-se	$\mu_{50} = \mu_{100}$

Tal como no *kernel* anterior, o parâmetro de regularização com valor 1 apresentou desempenho estatisticamente inferior na comparação com todos os outros três valores (Tabela 6.7), a ponto de nesse caso se poder rejeitar a hipótese nula, apesar de $C = 1$ ser o valor definido por defeito no algoritmo. Considerando os outros valores (10, 50 e

100), um aumento no parâmetro não significou aumento da acurácia, apesar de este parâmetro determinar o quanto o modelo se ajusta aos dados. Um valor alto como 50 ou 100 (valor máximo de ajuste) acaba por causar *overfitting* no modelo, razão pela qual não houve incremento na acurácia com o aumento do valor do parâmetro.

6.1.4.3 Polinomial

A seguir, é apresentada a Tabela 6.8, na qual estão os parâmetros e resultados dos testes de significância estatística de comparação entre desempenho de *datasets*, *splits* e parâmetro de regularização C .

Tabela 6.8 – Testes de significância estatística para os resultados do *kernel* Polinomial do SVM

Teste (A/B)	Média A	Média B	Tipo de Teste	<i>p-value</i>	H_0	CONCLUSÃO
D1/D2	84,753	90,428	Bicaudal	7,73E-09	Rejeita-se	$\mu_1 < \mu_2$
D1/D3	84,753	87,694	Bicaudal	8,94E-05	Rejeita-se	$\mu_1 < \mu_3$
D2/D3	90,428	87,694	Bicaudal	0,000715	Rejeita-se	$\mu_2 > \mu_3$
70% / 80%	87,047	87,969	Bicaudal	0,398829	Aceita-se	$\mu_{70\%} = \mu_{80\%}$
70% / 90%	87,047	87,858	Bicaudal	0,495222	Aceita-se	$\mu_{70\%} = \mu_{90\%}$
80% / 90%	87,969	87,858	Bicaudal	0,927173	Aceita-se	$\mu_{80\%} = \mu_{90\%}$
$C = 1/C = 10$	85,546	88,316	Bicaudal	0,029909	Aceita-se	$\mu_1 = \mu_{10}$
$C = 1/C = 50$	85,546	88,451	Bicaudal	0,020496	Aceita-se	$\mu_1 = \mu_{50}$
$C = 1/C = 100$	85,546	88,185	Bicaudal	0,040162	Aceita-se	$\mu_1 = \mu_{100}$
$C = 10/C = 50$	88,316	88,451	Bicaudal	0,916359	Aceita-se	$\mu_{10} = \mu_{50}$
$C = 10/C=100$	88,316	88,185	Bicaudal	0,921842	Aceita-se	$\mu_{10} = \mu_{100}$
$C = 50/C=100$	88,451	88,185	Bicaudal	0,838516	Aceita-se	$\mu_{50} = \mu_{100}$

Esse *kernel* mostrou-se bastante sensível aos dados, exibindo acurácias diferentes em todas as comparações entre os *datasets*. O D1 se mostrou inferior na comparação com os outros dois, sendo que D2 atingiu uma acurácia superior em comparação a D3. Nesse caso, pode-se afirmar que o número de imagens usadas é relevante para este *kernel*. As variações de *split* não apresentaram relevância para a acurácia. A variação dos valores do parâmetro de regularização também não teve efeito significativo sobre os resultados.

6.1.4.4 Sigmoide

Na Tabela 6.9 encontram-se registadas as comparações entre *datasets*, *splits* e parâmetros de regularização para o *kernel* sigmoide. São mostradas as médias, o tipo de teste estatístico realizado, o *p-value* obtido, se a hipótese nula foi aceite ou rejeitada, e a conclusão.

Tabela 6.9 – Testes de significância estatística para os resultados do *kernel* Sigmoide do SVM

Teste (A/B)	Média A	Média B	Tipo de Teste	<i>p-value</i>	H ₀	CONCLUSÃO
D1/D2	58,478	58,277	Unicaudal	0,051279	Aceita-se	$\mu_1 = \mu_2$
D1/D3	58,478	57,056	Unicaudal	6,82E-10	Rejeita-se	$\mu_1 > \mu_3$
D2/D3	58,277	57,056	Bicaudal	2,944E-08	Rejeita-se	$\mu_2 > \mu_3$
70% / 80%	58,108	57,924	Unicaudal	0,273723	Aceita-se	$\mu_{70\%} = \mu_{80\%}$
70% / 90%	58,108	57,779	Bicaudal	0,202692	Aceita-se	$\mu_{70\%} = \mu_{90\%}$
80% / 90%	57,924	57,779	Unicaudal	0,329046	Aceita-se	$\mu_{80\%} = \mu_{90\%}$
C = 1/C = 10	58,167	57,884	Bicaudal	0,412119	Aceita-se	$\mu_1 = \mu_{10}$
C = 1/C = 50	58,167	57,848	Bicaudal	0,350838	Aceita-se	$\mu_1 = \mu_{50}$
C = 1/ C = 100	58,167	57,980	Bicaudal	0,572656	Aceita-se	$\mu_1 = \mu_{100}$
C = 10/ C = 50	57,884	57,848	Bicaudal	0,920526	Aceita-se	$\mu_{10} = \mu_{50}$
C = 10/ C = 100	57,884	57,980	Bicaudal	0,786152	Aceita-se	$\mu_{10} = \mu_{100}$
C = 50/ C = 100	57,848	57,980	Bicaudal	0,706276	Aceita-se	$\mu_{50} = \mu_{100}$

Com o último *kernel*, houve diferença significativa na comparação entre D3 com os outros dois *datasets*, sendo este inferior em relação aos outros dois casos, o que foi um resultado diferente dos observados com os *kernels* anteriores em que houve diferença entre os *datasets*. Nesses casos, D1 foi aquele que sempre atingiu os resultados inferiores em comparação aos outros dois. As variações de *split* e parâmetro de regularização, por sua vez, não resultaram em qualquer aumento da acurácia. Ao contrário dos dois *kernels* anteriores, não houve uma diferença considerável no *p-value* encontrado para $C = 1$, tendo sido encontrados valores da mesma ordem de grandeza. Isso mostra que o *kernel* sigmoide não é tão sensível às variações do parâmetro de regularização quanto os anteriores, e que neste caso o valor de defeito de C é uma escolha apropriada.

6.1.5 kNN

Com este algoritmo, foram observadas diferenças estatísticas somente em relação ao *dataset* usado. O D2 obteve resultados superiores na comparação com os outros dois, com um nível de confiança superior a 99%. Entre os *datasets* D1 e D3, que tem o mesmo número de imagens, não houve diferença na acurácia alcançada com confiança de 99%, mas o *p-value* calculado permitiria afirmar que há diferença com um nível de confiança de 95%, que também é um resultado considerável.

Tabela 6.10 – Testes de significância estatística para os resultados do algoritmo kNN.

Teste (A/B)	Média A	Média B	Tipo de Teste	<i>p-value</i>	H ₀	CONCLUSÃO
D1/D2	82,259	90,532	Unicaudal	0,003563	Rejeita-se	$\mu_1 < \mu_2$
D1/D3	82,259	87,729	Unicaudal	0,024758	Aceita-se	$\mu_1 = \mu_3$
D2/D3	90,532	87,729	Bicaudal	0,004920	Rejeita-se	$\mu_2 > \mu_3$
70% / 80%	87,349	88,423	Bicaudal	0,355985	Aceita-se	$\mu_{70\%} = \mu_{80\%}$
70% / 90%	87,349	84,748	Bicaudal	0,408700	Aceita-se	$\mu_{70\%} = \mu_{90\%}$
80% / 90%	88,423	84,748	Bicaudal	0,252613	Aceita-se	$\mu_{80\%} = \mu_{90\%}$
3 / 5	88,983	87,318	Bicaudal	0,316599	Aceita-se	$\mu_3 = \mu_5$
3 / 10	88,983	84,218	Unicaudal	0,055429	Aceita-se	$\mu_3 = \mu_{10}$
5 / 10	87,318	84,218	Unicaudal	0,148003	Aceita-se	$\mu_5 = \mu_{10}$

Na comparação com os outros resultados, os *p-value* foram relevantemente altos (ver Tabela 6.10), o que reforça a aceitação da hipótese nula. Na comparação do desempenho entre os números de vizinhos 3 e 10, o *p-value* é consideravelmente mais baixo em relação aos outros (ainda que a hipótese de não haver diferença na média seja também aceite). Isso pode significar que, ainda que o valor de defeito para o número de vizinhos – 10 – possa ser considerado adequado, é possível que o teste com valores menores de vizinhos não deva ser excluído.

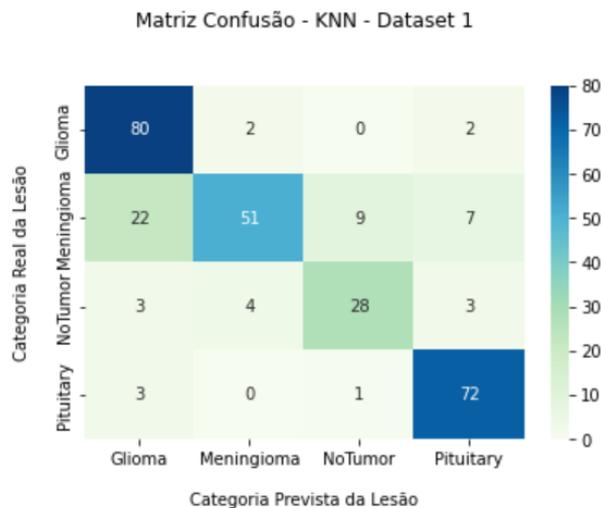


Figura 6.8 – Matriz confusão do *Dataset 1* com o algoritmo *k-Nearest Neighbors* com 5 vizinhos e *split* de 90%.

Como acontece com os algoritmos *Decision Tree* e *Random Forest*, no *Dataset 1* há 84 Gliomas, 89 Meningiomas, 38 Sem Tumor e 76 de Glândula Pituitária. A matriz confusão é apresentada na Figura 6.8.

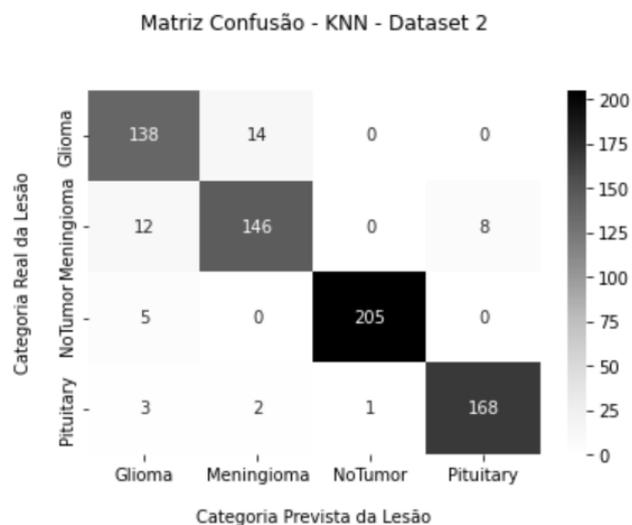


Figura 6.9 – Matriz confusão do *Dataset 2* com o algoritmo *k-Nearest Neighbors* com 5 vizinhos e *split* de 90%.

A distribuição do número de imagens entre as categorias também segue a mesma dos outros algoritmos no *Dataset 2* (Figura 6.9) e no *Dataset 3* (Figura 6.10).

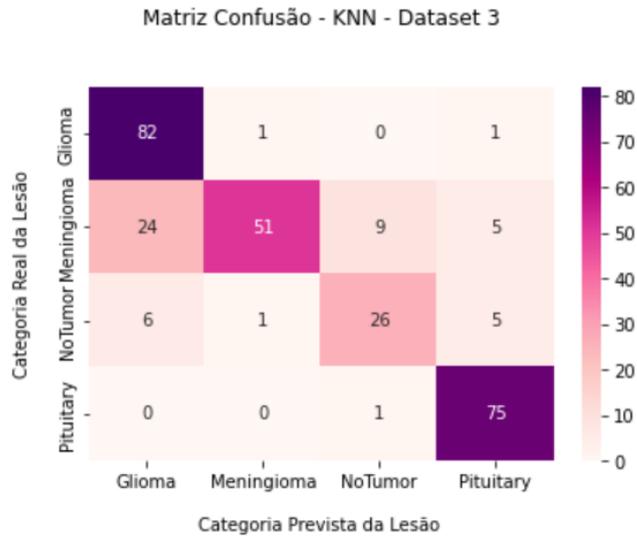


Figura 6.10 – Matriz confusão do *Dataset 3* com o algoritmo *k-Nearest Neighbors* com 5 vizinhos e *split* de 90%

Segue a Tabela 6.11 com as percentagens de acerto por algoritmo e por categoria de lesão.

Tabela 6.11 – Percentuais de acerto do algoritmo *k-Nearest Neighbors* por categoria nos *Datasets 1, 2 e 3*.

Dataset	Glioma	Meningioma	Sem Tumor	Pituitária
D1	95,2%	57,3%	73,7%	94,7%
D2	90,8%	87,9%	97,6%	96,6%
D3	97,6%	57,3%	68,4%	98,7%

O D2 foi superior aos outros dois *datasets* em três das quatro categorias em relação ao D1 (exceto Glioma) em dois em relação ao D3, e observam-se diferenças principalmente nas categorias de Meningiomas (30,6% a menos para os dois) e Sem Tumor (23,9% a menos para D1 e 29,2% para D3). Apesar de não ser superior em todas as categorias, naquela em que o *dataset 2* foi inferior a diferença era pequena (6,8% no máximo), ao passo que naquelas em que foi superior a diferença é bastante considerável, o que explica a superioridade estatística em relação aos outros dois *datasets*.

Em ambos os casos, na categoria Meningioma a confusão maior se deu em favor de Gliomas perfazendo 24,5% em D1 e 27% em D3. As outras duas categorias contribuíram com valores entre 6% e 10% do total. Como já referido anteriormente na

secções 6.1.1 e 6.1.2, é expectável que a maior confusão no caso de Meningiomas seja em favor de Gliomas, o que ocorre com este algoritmo. No D1, o erro na categoria Sem Tumor foi distribuído de maneira praticamente uniforme entre as outras 3 categorias, sendo 7,9% para Gliomas e Pituitária e 10,5% para Meningiomas. Já em D3 a confusão maior se deu com as categorias Glioma (15,8%) e Pituitária (13,2%).

6.2 Entre Algoritmos

Foram comparadas as medidas de acurácia entre algoritmos diferentes, considerando aos pares o melhor resultado de cada um quanto ao *split*, ao *dataset* e, no caso do algoritmo SVM, quanto aos diferentes *kernels*.

6.2.1 Dataset

Foram comparados dois a dois os melhores desempenhos quanto ao *dataset* de cada algoritmo, para que se possa verificar a superioridade na acurácia. Na Tabela 6.12 encontram-se os algoritmos comparados, as médias de cada um, o tipo de teste utilizado (uni ou bicaudal), o *p-value*, o aceite ou rejeição da hipótese nula e a conclusão. Para facilitar a visualização, sumarizaram-se os resultados na Tabela 6.13. No cruzamento encontra-se o algoritmo que se saiu melhor. Em caso de não haver diferença estatisticamente significativa, foi escrito “igual”.

Tabela 6.12 – Testes de significância estatística do melhor desempenho em cada algoritmo quanto ao *dataset*.

Teste (A/B)	Média A	Média B	Tipo de Teste	<i>p-value</i>	H ₀	CONCLUSÃO
DT/RF	90,844	96,425	Unicaudal	0,003822	Rejeita-se	$\mu_{DT} < \mu_{RF}$
DT/LDA	90,844	77,148	Bicaudal	0,000142	Rejeita-se	$\mu_{DT} > \mu_{LDA}$
DT/Linear	90,844	73,869	Unicaudal	0,000129	Rejeita-se	$\mu_{DT} > \mu_{Linear}$
DT/RBF	90,844	93,615	Unicaudal	0,018981	Aceita-se	$\mu_{DT} = \mu_{RBF}$
DT/Poli	90,844	90,428	Bicaudal	0,664936	Aceita-se	$\mu_{DT} = \mu_{Poli}$
DT/Sig	90,844	58,478	Unicaudal	0,000275	Rejeita-se	$\mu_{DT} > \mu_{Sig}$
DT/kNN	90,844	90,532	Bicaudal	0,756653	Aceita-se	$\mu_{DT} = \mu_{kNN}$
RF/LDA	96,425	77,148	Unicaudal	6,24E-06	Rejeita-se	$\mu_{RF} > \mu_{LDA}$
RF/Linear	96,425	73,869	Unicaudal	8,26E-22	Rejeita-se	$\mu_{RF} > \mu_{Linear}$
RF/RBF	96,425	93,615	Unicaudal	0,000837	Rejeita-se	$\mu_{RF} > \mu_{RBF}$
RF/Poli	96,425	90,428	Unicaudal	1,77E-09	Rejeita-se	$\mu_{RF} > \mu_{Poli}$
RF/Sig	96,425	58,478	Unicaudal	2,63E-17	Rejeita-se	$\mu_{RF} > \mu_{Sig}$
RF/kNN	96,425	90,532	Unicaudal	1,57E-07	Rejeita-se	$\mu_{RF} > \mu_{kNN}$
LDA/Linear	77,148	73,869	Unicaudal	0,004484	Rejeita-se	$\mu_{LDA} > \mu_{Linear}$
LDA/RBF	77,148	93,615	Unicaudal	1,47E-07	Rejeita-se	$\mu_{LDA} < \mu_{RBF}$
LDA/Poli	77,148	90,428	Bicaudal	1,13E-05	Rejeita-se	$\mu_{LDA} < \mu_{Poli}$
LDA/Sig	77,148	58,478	Unicaudal	0,000515	Rejeita-se	$\mu_{LDA} > \mu_{Sig}$
LDA/kNN	77,148	90,532	Bicaudal	3,33E-06	Rejeita-se	$\mu_{LDA} < \mu_{kNN}$
Linear/RBF	73,869	93,615	Unicaudal	9,56E-16	Rejeita-se	$\mu_{Linear} < \mu_{RBF}$
Linear/Poli	73,869	90,428	Unicaudal	6,75E-18	Rejeita-se	$\mu_{Linear} < \mu_{Poli}$
Linear/Sig	73,869	58,478	Unicaudal	7,13E-15	Rejeita-se	$\mu_{Linear} > \mu_{Sig}$
Linear/kNN	73,869	90,532	Unicaudal	2,25E-13	Rejeita-se	$\mu_{Linear} < \mu_{kNN}$
RBF/Poli	93,615	90,428	Unicaudal	0,000338	Rejeita-se	$\mu_{RBF} > \mu_{Poli}$
RBF/Sig	93,615	58,478	Unicaudal	4,62E-15	Rejeita-se	$\mu_{RBF} > \mu_{Sig}$
RBF/kNN	93,615	90,532	Unicaudal	0,000893	Rejeita-se	$\mu_{RBF} > \mu_{kNN}$
Poli/Sig	90,428	58,478	Unicaudal	4,95E-16	Rejeita-se	$\mu_{Poli} > \mu_{Sig}$
Poli/kNN	90,428	90,532	Bicaudal	0,889365	Aceita-se	$\mu_{Poli} = \mu_{kNN}$
Sig/kNN	58,478	90,532	Unicaudal	5,15E-12	Rejeita-se	$\mu_{Sig} < \mu_{kNN}$

Tabela 6.13 – Sumário dos resultados quanto ao *dataset* por algoritmo.

Algoritmo	DT	RF	LDA	SVM-Linear	SVM-RBF	SVM-Poli	SVM-Sig	kNN
DT	-	RF	DT	DT	Igual	Igual	DT	Igual
RF	RF	-	RF	RF	RF	RF	RF	RF
LDA	DT	RF	-	LDA	RBF	Poli	LDA	kNN
Linear	DT	RF	LDA	-	RBF	Poli	Linear	kNN
RBF	Igual	RF	RBF	RBF	-	RBF	RBF	RBF
Poli	Igual	RF	Poli	Poli	RBF	-	Poli	Igual
Sig	DT	RF	LDA	Linear	RBF	Poli	-	kNN
kNN	Igual	RF	kNN	kNN	RBF	Igual	kNN	-

Observa-se que o algoritmo *Random Forest* foi superior em comparação a todos os outros algoritmos. O *Support Vector Machine* com o *kernel* RBF também apresentou um desempenho superior em todas as comparações, exceto com o RF (em que foi inferior) e com o *Decision Tree*, em que os dois se igualaram.

O que apresentou o pior desempenho foi o SVM com o *kernel* Sigmoides, que foi inferior em todas as comparações com os demais algoritmos.

6.2.2 Split

Foram comparados os melhores desempenhos de cada algoritmo em relação ao *split*, para que se possa verificar se há superioridade na acurácia. Os resultados encontram-se na Tabela 6.14.

Tabela 6.14 – Testes de significância estatística do melhor desempenho em cada algoritmo quanto ao *split*.

Teste (A/B)	Média A	Média B	Tipo de Teste	<i>p-value</i>	H ₀	CONCLUSÃO
DT/RF	87,310	94,948	Unicaudal	0,008553	Rejeita-se	$\mu_{DT} < \mu_{RF}$
DT/LDA	87,310	71,900	Bicaudal	0,019481	Aceita-se	$\mu_{DT} = \mu_{LDA}$
DT/Linear	87,310	72,501	Bicaudal	0,001104	Rejeita-se	$\mu_{DT} > \mu_{Linear}$
DT/RBF	87,310	91,996	Bicaudal	0,052823	Aceita-se	$\mu_{DT} = \mu_{RBF}$
DT/Poli	87,310	87,969	Bicaudal	0,716372	Aceita-se	$\mu_{DT} = \mu_{Poli}$
DT/Sig	87,310	58,108	Bicaudal	0,002496	Rejeita-se	$\mu_{DT} > \mu_{Sig}$
DT/kNN	87,310	88,423	Bicaudal	0,551228	Aceita-se	$\mu_{DT} = \mu_{kNN}$
RF/LDA	94,948	71,900	Unicaudal	0,008793	Rejeita-se	$\mu_{RF} > \mu_{LDA}$
RF/Linear	94,948	72,501	Unicaudal	3,93E-11	Rejeita-se	$\mu_{RF} > \mu_{Linear}$
RF/RBF	94,948	91,996	Unicaudal	0,008027	Rejeita-se	$\mu_{RF} > \mu_{RBF}$
RF/Poli	94,948	87,969	Unicaudal	3,91E-07	Rejeita-se	$\mu_{RF} > \mu_{Poli}$
RF/Sig	94,948	58,108	Unicaudal	3,55E-16	Rejeita-se	$\mu_{RF} > \mu_{Sig}$
RF/kNN	94,948	88,423	Unicaudal	4,98E-06	Rejeita-se	$\mu_{RF} > \mu_{kNN}$
LDA/Linear	71,900	72,501	Bicaudal	0,868095	Aceita-se	$\mu_{LDA} = \mu_{Linear}$
LDA/RBF	71,900	91,996	Bicaudal	0,023957	Aceita-se	$\mu_{LDA} = \mu_{RBF}$
LDA/Poli	71,900	87,969	Bicaudal	0,035896	Aceita-se	$\mu_{LDA} = \mu_{Poli}$
LDA/Sig	71,900	58,108	Bicaudal	0,045157	Aceita-se	$\mu_{LDA} = \mu_{Sig}$
LDA/kNN	71,900	88,423	Bicaudal	0,034193	Aceita-se	$\mu_{LDA} = \mu_{kNN}$
Linear/RBF	72,501	91,996	Bicaudal	1,53E-10	Rejeita-se	$\mu_{Linear} < \mu_{RBF}$
Linear/Poli	72,501	87,969	Bicaudal	3,53E-09	Rejeita-se	$\mu_{Linear} < \mu_{Poli}$
Linear/Sig	72,501	58,108	Bicaudal	2,05E-06	Rejeita-se	$\mu_{Linear} > \mu_{Sig}$
Linear/kNN	72,501	88,423	Bicaudal	6,32E-09	Rejeita-se	$\mu_{Linear} < \mu_{kNN}$
RBF/Poli	91,996	87,969	Bicaudal	0,003147	Rejeita-se	$\mu_{RBF} > \mu_{Poli}$
RBF/Sig	91,996	58,108	Bicaudal	1,25E-13	Rejeita-se	$\mu_{RBF} > \mu_{Sig}$
RBF/kNN	91,996	88,423	Bicaudal	0,008877	Rejeita-se	$\mu_{RBF} > \mu_{kNN}$
Poli/Sig	87,969	58,108	Bicaudal	8,27E-14	Rejeita-se	$\mu_{Poli} > \mu_{Sig}$
Poli/kNN	87,969	88,423	Bicaudal	0,690778	Aceita-se	$\mu_{Poli} = \mu_{kNN}$
Sig/kNN	58,108	88,423	Bicaudal	3,97E-11	Rejeita-se	$\mu_{Sig} < \mu_{kNN}$

Para facilitar a visualização, os resultados foram sumarizados na Tabela 6.15.

Tabela 6.15 – Sumário dos resultados quanto ao *split* por algoritmo

Algoritmo	DT	RF	LDA	SVM-Linear	SVM-RBF	SVM-Poli	SVM-Sig	kNN
DT	-	RF	Igual	DT	Igual	Igual	DT	Igual
RF	RF	-	RF	RF	RF	RF	RF	RF
LDA	Igual	RF	-	Igual	Igual	Igual	Igual	Igual
Linear	DT	RF	Igual	-	RBF	Poli	Linear	kNN
RBF	Igual	RF	Igual	RBF	-	RBF	RBF	RBF
Poli	Igual	RF	Igual	Poli	RBF	-	Poli	Igual
Sig	DT	RF	Igual	Linear	RBF	Poli	-	kNN
kNN	Igual	RF	Igual	Sig	RBF	Igual	kNN	-

Tal como aconteceu com a comparação entre resultados em relação aos *datasets*, o algoritmo RF mostrou-se superior em relação a todos os outros algoritmos, seguido pelo *kernel* RBF do SVM, que foi inferior ao RF, e igualou com DT e LDA. É de salientar que o facto de se poder dizer que não há, estatisticamente, diferença de média entre o RBF e o LDA, mesmo tendo uma diferença numericamente grande entre os dois (maior inclusive que outros casos em que a hipótese da não-diferença foi rejeitada) deve-se a dois fatores: primeiro, a grande variância dos dados do *kernel* RBF, que atingiu uma ordem de grandeza a mais que os outros; e em segundo lugar o facto da hipótese ter sido testada para 99% - sendo que é bastante comum se aceitar um nível de confiança de 95% como sendo suficiente, situação em que o *p-value* seria baixo o suficiente para que se rejeitasse a hipótese nula. Portanto, embora para este estudo considere-se que o RBF e o LDA não tenham diferença estatística na média para o *split*, deve-se olhar para este resultado com uma certa cautela.

6.3 Entre *kernels* do SVM

Foram comparados os melhores resultados em termos de parâmetro de regularização entre os *kernels* do algoritmo *Support Vector Machine*. Os resultados encontram-se na Tabela 6.16.

Tabela 6.16 – Testes de significância estatística do melhor desempenho em cada *kernel* do SVM quanto ao parâmetro de regularização C.

Teste (A/B)	Média A	Média B	Tipo de Teste	<i>p-value</i>	H ₀	CONCLUSÃO
Linear/RBF	74,213	92,578	Bicaudal	3,23E-10	Rejeita-se	$\mu_{\text{Linear}} < \mu_{\text{RBF}}$
Linear/Poli	74,213	88,451	Bicaudal	7,77E-09	Rejeita-se	$\mu_{\text{Linear}} < \mu_{\text{Poli}}$
Linear/Sig	74,213	58,167	Bicaudal	4,32E-07	Rejeita-se	$\mu_{\text{Linear}} > \mu_{\text{Sig}}$
RBF/Poli	92,578	88,451	Bicaudal	0,004024	Rejeita-se	$\mu_{\text{RBF}} > \mu_{\text{Poli}}$
RBF/Sig	92,578	58,167	Bicaudal	2,7E-11	Rejeita-se	$\mu_{\text{RBF}} > \mu_{\text{Sig}}$
Poli/Sig	88,451	58,167	Bicaudal	8,91E-11	Rejeita-se	$\mu_{\text{Poli}} > \mu_{\text{Sig}}$

O *kernel* RBF mostrou-se superior na comparação com todos os outros *kernels*, o que era esperado tendo em consideração os resultados mostrados nas secções anteriores. Em seguida, o *kernel* Polinomial foi o que atingiu os melhores resultados, sendo inferior somente ao RBF. Também conforme visto anteriormente, o Sigmoide foi o que apresentou o pior desempenho, sendo inferior em todas as comparações.

Há que se mencionar que não houve “empates” estatísticos nesse teste, o que significa que o algoritmo SVM é bastante sensível em relação ao *kernel* que se utiliza, e que o RBF é o que apresenta melhores resultados.

6.4 Comparação com outros trabalhos

Nesta secção, os resultados obtidos serão comparados com os resultados de dois dos trabalhos de outros autores apresentados na Secção 3.4.1.

No trabalho de Amin, Sharif, Raza, Saba, & Anjum (2019), os autores consideram que o melhor resultado foi obtido com o algoritmo kNN com 5 vizinhos. A Tabela 6.17 abaixo sumariza as acurácias obtidas com cada um dos três *datasets* testados (chamados BRATS 2013, BRATS 2015 e Local), com os algoritmos em comum usados neste trabalho.

Tabela 6.17 – Sumarização dos resultados de Amin, Sharif, Raza, Saba, & Anjum (2019).

Algoritmo / Dataset	BRATS (2013)	BRATS (2015)	Local
SVM (Polinomial – 2º grau)	94%	97%	97%
DT	96%	97%	94%
kNN (5 vizinhos)	100%	97%	93%
RF (200 estimadores)	96%	97%	95%

Os resultados diferem quanto ao *dataset* utilizado, atingindo uma acurácia máxima de 100% em um deles, com o algoritmo kNN. Os parâmetros utilizados também foram diferentes: com o SVM foi usado somente o *kernel* Polinomial, e com grau 2, e o RF foi testado com 200 estimadores somente.

Os resultados apresentados na secção 6.1 com os mesmos algoritmos e com os parâmetros o mais parecidos possível aparecem na Tabela 6.18.

Tabela 6.18 – Sumarização de resultados de alguns algoritmos.

Algoritmo / Dataset	D1	D2	D3
SVM (Polinomial – 3º grau)	84.753%	90.428%	87.694%
DT	85.037%	90.844%	88.386%
kNN (5 vizinhos)	85.497%	91.952%	89.678%
RF (100 estimadores)	92.030%	96.296%	96.821%

Os resultados para um mesmo algoritmo com métodos diferentes foram comparados com um teste estatístico, nos mesmos moldes dos feitos nas secções anteriores, para que se verifique se há diferença significativa entre as propostas.

Na Tabela 6.19 encontram-se as comparações entre os mesmos algoritmos, sendo que a identificação 1 é para os resultados deste estudo e 2 para Amin, Sharif, Raza, Saba, & Anjum.

Tabela 6.19 – Comparação dos resultados com Amin, Sharif, Raza, Saba, & Anjum (2019).

Teste (A/B)	Média A	Média B	Tipo de Teste	<i>p-value</i>	H ₀	CONCLUSÃO
SVM1/SVM2	87,625%	96,000%	Unicaudal	0,011131	Aceita-se	$\mu_{SVM1} = \mu_{SVM2}$
DT1/DT2	88,089%	95,667%	Bicaudal	0,028224	Aceita-se	$\mu_{DT1} < \mu_{DT2}$
kNN1/kNN2	89,042%	96,667%	Unicaudal	0,025676	Aceita-se	$\mu_{kNN1} < \mu_{kNN2}$
RF1/RF2	95,049%	96,000%	Unicaudal	0,299563	Aceita-se	$\mu_{RF1} = \mu_{RF2}$

Em todos os casos, pode-se aceitar a hipótese de que não há diferença entre as médias, com um grau de confiança de 99%. Como aconteceu em secções anteriores, é válido assinalar os casos em que o *p-value* não é baixo o suficiente para que se rejeite a hipótese nula com 99% de confiança, mas seria caso o valor de confiança escolhido fosse o de 95%. Essa situação também ocorreu nas comparações entre os resultados do SVM, do DT e do kNN. O resultado do RF foi o que se pôde aceitar com mais certeza, dado o alto *p-value* encontrado.

O trabalho apresentado por Amin, Sharif, Yasmin, & Fernandes (2017) usa três *kernels* diferentes do SVM para fazer uma classificação multiclasse, com três *datasets* diferentes. Além disso, faz a validação cruzada (5, 10, 15, 20, 25 e 30 vezes) dos resultados – ver Tabela 6.20. Como não referem quais os valores dos parâmetros de regularização, assume-se que usaram o de defeito, $C = 1$. Os *splits* utilizados também não foram especificados.

Tabela 6.20 – Sumarização dos resultados de Amin, Sharif, Yasmin, & Fernandes (2017)

Kernel – Validações Cruzadas / Dataset		Local (%)	Harvard (%)	RIDER (%)
Linear	5 vezes	98,8	98,1	81,3
	10 vezes	92,2	99,4	75,0
	15 vezes	96,0	98,7	85,9
	20 vezes	96,6	96,1	79,1
	25 vezes	96,0	95,5	90,0
	30 vezes	95,5	94,8	78,1
Polinomial Cúbico	5 vezes	81,3	93,5	95,0
	10 vezes	75,0	87,7	96,7
	15 vezes	85,9	81,2	86,2
	20 vezes	79,1	75,3	80,0
	25 vezes	90,0	83,1	100
	30 vezes	78,1	78,6	85,0
RBF	5 vezes	96,1	95,5	93,3
	10 vezes	97,6	96,8	88,3
	15 vezes	85,9	97,4	84,0
	20 vezes	89,4	98,7	82,8
	25 vezes	87,1	86,4	89,3
	30 vezes	98,1	92,2	96,8

Os resultados apresentados na secção 6.1 para os *kernels* linear, polinomial e RBF com parâmetro de regularização $C = 1$ e todos os *splits* são sumarizados na Tabela 6.21.

Tabela 6.21 – Sumarização dos resultados para os *kernels* linear, polinomial e RBF do SVM com $C = 1$, para os três *datasets*.

Algoritmo - Split / Dataset		D1	D2	D3
Linear	70%	73.432	X	74.593
	80%	73.171	X	77.047
	90%	72.474	X	74.564
Polinomial Cúbico	70%	83.159	87.441	84.669
	80%	83.232	87.616	85.584
	90%	83.798	89.138	85.279
RBF	70%	85.105	89.696	89.460
	80%	84.930	90.107	90.505
	90%	86.585	90.919	89.111

A seguir, foram feitas comparações estatísticas para se verificar se há diferença significativa entre os resultados encontrados. Os parâmetros usados para o teste são os mesmos que nas secções anteriores.

Na Tabela 6.22, a identificação 1 refere-se aos resultados deste estudo, e a identificação 2 ao estudo de Amin, Sharif, Yasmin e Fernandes.

Tabela 6.22 – Comparação com os resultados de Amin, Sharif, Yasmin e Fernandes.

Teste (A/B)	Média A	Média B	Tipo de Teste	<i>p-value</i>	H_0	CONCLUSÃO
Linear 1/Linear 2	74,214	91,506	Unicaudal	1,66E-08	Rejeita-se	$\mu_1 < \mu_2$
Cúbico 1/Cúbico 2	85,546	85,094	Unicaudal	0,40671	Aceita-se	$\mu_1 = \mu_2$
RBF 1/RBF2	88,491	91,983	Unicaudal	0,01328	Aceita-se	$\mu_1 = \mu_2$

O *kernel* linear apresentou um resultado significativamente inferior ao do estudo de Amin, Sharif, Yasmin e Fernandes, com um índice de confiança muito superior a 99%. O polinomial foi o que apresentou menor diferença, com um *p-value* bastante alto, de modo que se pode considerar que os resultados são iguais. O RBF foi o que apresentou o maior resultado dentre os três *kernels* utilizados nos dois estudos, e também resultou em um *p-value* com valor suficiente para que se possa considerá-los como não tendo diferença com 99% de confiança, embora esse valor seja menor que o *p-value* do *kernel*

polinomial em uma ordem de grandeza. Além disso, a hipótese nula não poderia ser aceita caso o grau de confiança usado fosse de 95%.

7. Conclusão

A proposta principal deste trabalho consiste na testagem de vários algoritmos de *Machine Learning* com *datasets* constituídos por imagens de Ressonância Magnética de três tipos de tumores cerebrais – glioma, meningioma e tumor da glândula pituitária – e um quarto subconjunto com imagens de cérebro sem qualquer lesão. Os algoritmos foram incumbidos de realizar uma classificação multiclasse, isto é, avaliar se há lesão na imagem e, caso haja, categorizá-la corretamente em uma das três classes de tumor. A métrica de avaliação utilizada foi a acurácia em percentagem, que é a razão entre os resultados acertados – verdadeiro positivo e verdadeiro negativo – e o total de resultados – verdadeiros positivos e negativos mais os falsos positivos e negativos.

Dentre os objetivos específicos do trabalho estavam as avaliações da influência de cada um dos *datasets* na acurácia do algoritmo, de um parâmetro (em casos que havia parâmetros para ajustar) e dos diferentes *splits* – que é a divisão entre os dados de treino e de teste, e por fim a comparação com os resultados de trabalhos semelhantes.

O primeiro *dataset* (D1) é constituído por 2870 imagens, sendo 826 gliomas, 822 meningiomas, 827 tumores da glândula pituitária e 395 sem tumor; o segundo (D2) é constituído por 7020 imagens, sendo 1628 gliomas, 1645 meningiomas, 1757 tumores da glândula pituitária 2000 sem tumor; e o terceiro (D3) é um subconjunto de D2, construído para ter o mesmo número de imagens e distribuição dentro das categorias

que D1, para que ambos pudessem ser mais efetivamente comparados sem o enviesamento do número de imagens.

As divisões (*splits*) feitas nos *datasets* foram de 70% dados de treino e 30% de teste, 80% de treino e 20% de teste e 90% de treino e 10% de teste.

Os algoritmos testados foram:

- *Decision Tree* (DT): foi testado para os três *datasets*, cada um com os três *splits*, e usados os parâmetros padrão do algoritmo.
- *Random Forest* (RF): foi testado para os três *datasets*, cada um com os três *splits*, e com quatro números diferentes de estimadores (árvores) – 100, 1000, 2000 e 5000. Os demais parâmetros foram os de defeito.
- *Linear Discriminant Analysis* (LDA): foi testado para os três *datasets*, cada um com os três *splits*. Não há parâmetros para ajustar.
- *Support Vector Machine* (SVM): foi testado para os três *datasets* (exceto para o *kernel* linear, com o qual foram testados somente D1 e D3) cada um com os três *splits*, com os quatro *kernels* disponíveis (linear, polinomial cúbico, RBF e sigmoide), e cada um com quatro valores de parâmetro de regularização c diferentes (1, 10, 50 e 100). Para os demais parâmetros foram mantidos os valores padrão.
- *k-Nearest Neighbors* (kNN): foi testado para os três *datasets*, cada um com os três *splits*, e com e valores diferentes de vizinhos (3, 5 e 10). Para os demais parâmetros foram mantidos os valores padrão.

No algoritmo DT, o maior fator de diferença foi o D2, que alcançou acurácia superior ao D1 com um grau de confiança superior a 99%. Não houve diferença estatisticamente significativa entre D1 e D3 ou entre D2 e D3, nem entre os três *splits*.

No RF, D1 alcançou uma acurácia menor que os outros dois *datasets*, enquanto D2 e D3 obtiveram médias estatisticamente iguais. Entre os *splits*, também não houve diferença estatisticamente significativa, nem entre o número de estimadores, mesmo entre 100 e 5000, pelo que o valor padrão do algoritmo (100) pode ser considerado adequado.

Com o algoritmo LDA, o D2 foi superior aos outros dois algoritmos, ao passo que D1 e D3 foram estatisticamente iguais entre si. Entre os *splits* também não houve diferenças.

Os resultados do algoritmo SVM foram analisados separadamente por *kernel*, inicialmente. O *kernel* linear atingiu superioridade de acurácia com o D3 em detrimento ao D1. Em relação aos *splits* e parâmetros de regularização utilizados, não houve diferença significativa. O D1 foi inferior aos outros dois *datasets* com o *kernel* RBF, ao passo que D2 e D3 não atingiram diferenças relevantes entre si. Os *splits* também não mostraram quaisquer diferenças. O parâmetro de regularização $C = 1$ se mostrou inferior a todos os outros três, porém não houve diferença entre os valores de 10, 50 e 100, o que mostra que o valor padrão do algoritmo pode não ser a melhor escolha. O polinomial atingiu acurácias diferentes para todos os *datasets*, sendo o D1 inferior aos outros dois e D3 inferior ao D2. Entre *splits* não houve diferença, nem entre os valores de parâmetro de regularização. No sigmoide, o D3 foi inferior aos outros dois, e não houve diferença entre D1 e D2. Entre *splits* e valores de parâmetro de regularização também não houve, pelo que o valor padrão pode ser uma boa escolha no caso desses dois últimos *kernels*.

O D2 atingiu maior acurácia que os outros dois com o algoritmo kNN. Entre os outros dois *datasets* não houve diferenças, nem entre os *splits* nem entre o número de vizinhos. O número padrão de vizinhos (10) pode ser considerado então pertinente.

Dentre os algoritmos, o *Random Forest* foi o que obteve melhor resultado, sendo superior a todos os outros nos testes de média com todos os *datasets* e de média com os *splits*. O *kernel* RBF do SVM foi o que apresentou o segundo melhor desempenho nos dois testes. Foram realizados também testes entre os *kernels* do SVM com os resultados de parâmetro de regularização, em que o RBF também se mostrou superior aos outros três. A grande desvantagem do RBF em relação ao *Random Forest* é o tempo de processamento: enquanto o primeiro pode demorar cerca de três horas, o segundo requer alguns minutos.

Na comparação com o estudo de Amin, Sharif, Raza, Saba e Anjum (2019), obteve-se resultados estatisticamente iguais com o SVM e com o RF, e inferiores com o DT e com o kNN. Já com o estudo de Amin, Sharif, Yasmin e Fernandes (2017), em que foram testados *kernels* do SVM, obtiveram-se resultados sem diferença estatística com o polinomial e com o RBF, e bastante inferiores com o linear.

Em ambos os estudos, os autores utilizaram técnicas de pré-processamento da imagem para o aumento da acurácia dos classificadores, com a aplicação de filtros, por exemplo,

o que não foi realizado nesse trabalho, e que pode explicar as diferenças na média. Em relação ao estudo de Amin, Sharif, Raza, Saba e Anjum, outra questão que pode explicar a diferença é que o trabalho desses autores envolve a classificação binária dos tumores, ou seja, apenas se há ou não há tumor na imagem, que é uma tarefa de classificação que pode ser considerada “mais simples” do que a multiclasse, e por isso também é capaz de atingir valores mais altos de acurácia. Nota-se na Tabela 6.20 que os autores Amin, Sharif, Yasmin e Fernandes, que também fizeram classificação multiclasse, obtiveram valores em torno de 75% a 85%, como vários dos que foram obtidos nesta dissertação.

A principal limitação deste trabalho se deu justamente por conta desse fator: o pré-processamento. Todos os trabalhos analisados na secção 3.4.1 passam por essa etapa, que certamente faz parte das perspectivas e objetivos em relação a aprendizados futuros, numa possível continuação deste trabalho.

De seguida, algum tipo de pré-processamento como sugerido no parágrafo anterior, no grupo todo ou nas imagens erroneamente categorizadas, poderia ser realizado com o intuito de tentar melhorar a acurácia. Nas Figuras 7.1 e 7.2 a seguir é possível visualizar as etapas de pré-processamento das imagens do trabalho com os quais se fizeram as comparações.

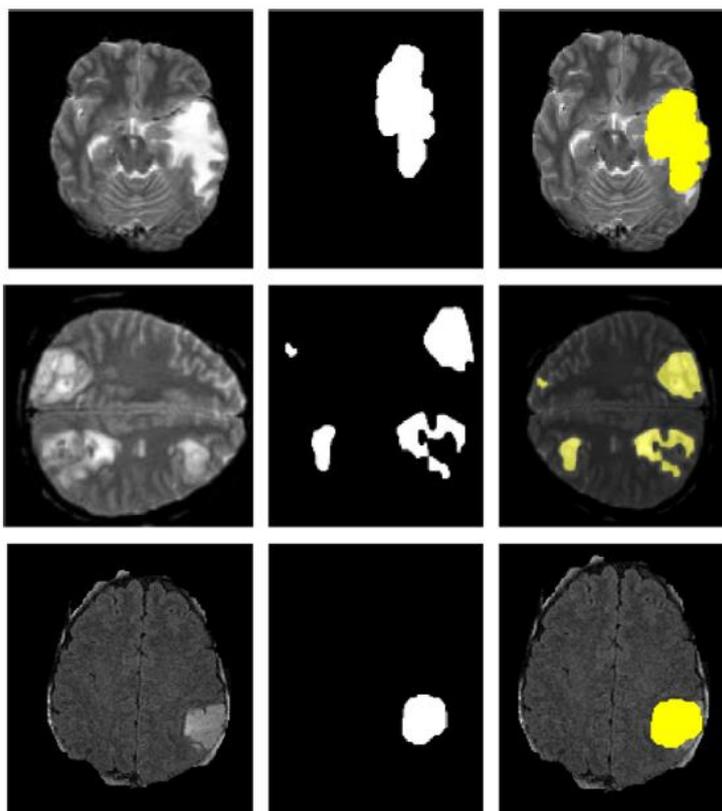


Figura 7.1 – Resultados do método proposto para a detecção de tumores de Amin J. , Sharif, Yasmin, & Fernandes.

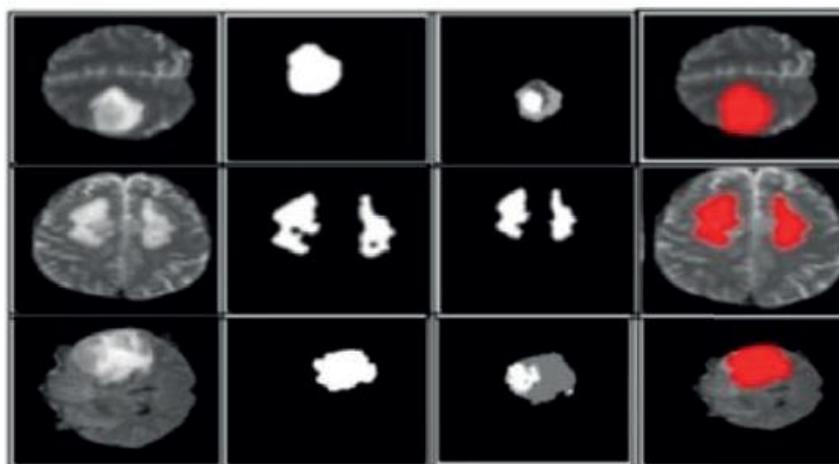


Figura 7.2 – Resultados do método proposto por Amin J. , Sharif, Raza, Saba, & Anjum.

O único processo pelo qual as imagens passam antes de serem classificadas é o redimensionamento, para que todas fiquem com o mesmo tamanho (150x150). Neste processo, não há perda significativa de detalhes das imagens, conforme se pode

observar na Figura 7.3 abaixo, que mostra uma imagem sem o redimensionamento, de tamanho 512x512 (esquerda) e a mesma imagem já redimensionada (direita), com 150x150.

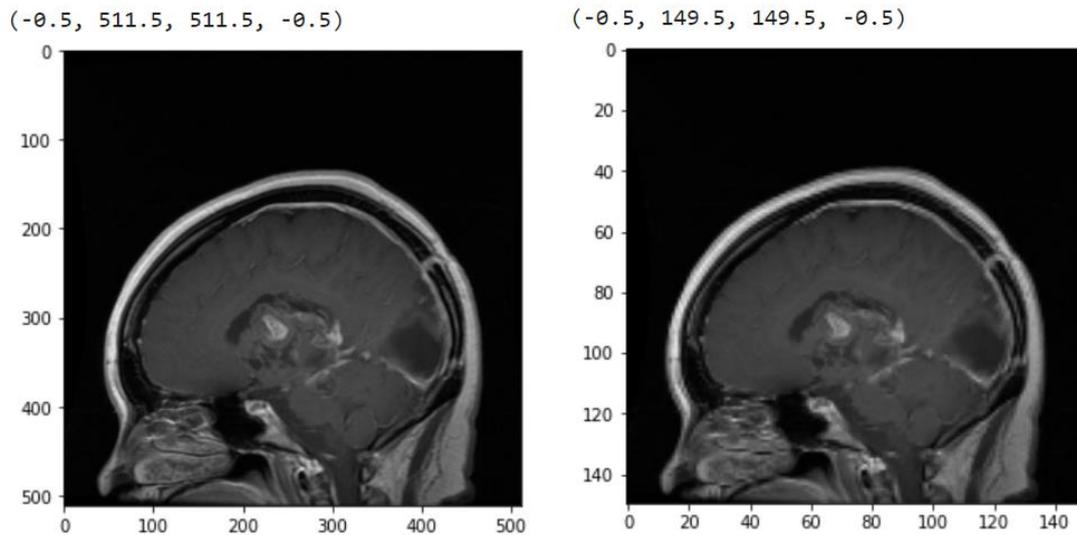


Figura 7.3 – Glioma antes e depois do redimensionamento.

É possível observar que a imagem redimensionada é um pouco menos nítida que a primeira, mas apenas se se olhar com bastante atenção se perceberá os contornos um pouco mais esbatidos e alguns “pixelizados”, mas a princípio não é essa a razão por que alguns resultados foram inferiores, mas sim os já mencionados.

8. Referências

- A F. Hoffmann–La Roche AG. (14 de maio de 2019). *Nem todo tumor é câncer*. Obtido em 22 de julho de 2021, de Roche Brasil: https://www.roche.com.br/pt/farmaceutica/areas_terapeuticas/oncologia/imunoterapia/nem-todo-tumor-e-cancer.html
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2008). *Molecular biology of the cell* (5ª ed.). Nova Iorque: Garland Science, Taylor Francis Group.
- Albuquerque, M. P., & Albuquerque, M. P. (16-27 de julho de 2012). Processamento de Imagens. *IX Escola do CBPF - Centro Brasileiro de Pesquisas Físicas*. Rio de Janeiro. Obtido em 9 de novembro de 2021, de CBPF - Centro Brasileiro de Pesquisas Físicas: https://mesonpi.cat.cbpf.br/e2012/arquivos/g06/Aula1_G06.pdf
- American Cancer Society. (30 de julho de 2015). *Testing Biopsy and Cytology Specimens for Cancer*. Obtido em 23 de julho de 2021, de American Cancer Society: <https://www.cancer.org/content/dam/CRC/PDF/Public/6105.00.pdf>
- American Cancer Society. (5 de maio de 2020). *Signs and Symptoms of Adult Brain and Spinal Cord Tumors*. Obtido em 12 de março de 2021, de American Cancer Society: <https://www.cancer.org/cancer/brain-spinal-cord-tumors-adults/detection-diagnosis-staging/signs-and-symptoms.html>
- American Cancer Society. (5 de maio de 2020). *Surgery for Adult Brain and Spinal Cord Tumors*. Obtido em 24 de agosto de 2021, de <https://www.cancer.org/cancer/brain-spinal-cord-tumors-adults/treating/surgery.html>
- American Cancer Society. (2020). *Types of Brain and Spinal Cords Tumors in Adults*. Obtido em 21 de agosto de 2021, de <https://www.cancer.org/cancer/brain-spinal-cord-tumors-adults/about/types-of-brain-tumors.html>

- Amien, M. B., Adb-elrehman, A., & Ibrahim, W. (Junho de 2013). An intelligent model for automatic brain-tumor diagnosis based on MRI images. *International Journal of Computer Applications*, 72(23), pp. 21-24.
- Amin, J., Sharif, M., Raza, M., Saba, T., & Anjum, M. A. (2019). Brain tumor detection using statistical and machine learning method. *Computer Methods and Programs in Biomedicine*, 177, pp. 69-79. doi:<https://doi.org/10.1016/j.cmpb.2019.05.015>
- Amin, J., Sharif, M., Yasmin, M., & Fernandes, S. L. (2017). A distinctive approach in brain tumor detection and classification using MRI. *Pattern Recognition Letters*, 000, pp. 1-10.
- Andrada, B. (1 de setembro de 2013). *Biópsia Cerebral Estereotáxica*. Obtido em 31 de agosto de 2021, de Blog do Neurocirurgião Jovem: <http://blogdoneurocirurgiao.blogspot.com/2013/09/biopsia-cerebral-estereotaxica.html>
- Awasthi, S. (17 de dezembro de 2020). Seven Most Popular SVM Kernels. *Data Aspirant*. Obtido em 6 de novembro de 2021, de <https://dataaspirant.com/svm-kernels/#t-1608054630732>
- Barros, P. (7 de abril de 2016). Aprendizagem de Máquina: Supervisionada ou Não Supervisionada? *Opensanca*. Obtido em 10 de setembro de 2021, de <https://medium.com/opensanca/aprendizagem-de-maquina-supervisionada-ou-n%C3%A3o-supervisionada-7d01f78cd80a>
- Basu, S., & Alavi, A. (2009). Molecular Imaging (PET) of Brain Tumors. *Neuroimaging Clinics of North America*, 19, pp. 625-646. doi:10.1016/j.nic.2009.08.012
- Batra, A., & Kaushik, D. (Junho de 2017). SECTUBIM: Automatic Segmentation and Classification of Tumeric Brain MRI Images using FHS (FCM, HWT and SVM). *International Journal of Engineering Science and Computing*, 7(6), pp. 13190-13194.
- BBC. (s.d.). *How do digital images work?* Obtido em 12 de março de 2021, de BBC: <https://www.bbc.co.uk/bitesize/topics/zf2f9j6/articles/z2tgr82>
- Bhuvaji, S. (24 de maio de 2020). *GitHub*. Obtido em 9 de novembro de 2021, de <https://github.com/sartajbhuvaji/brain-tumor-classification-dataset>
- Bilharvas, V. (28 de dezembro de 2020). Pandemia impulsiona uso de Big Data e Inteligência Artificial. (InforChannel, Entrevistador) Obtido em 21 de julho de 2021, de <https://inforchannel.com.br/2020/12/28/pandemia-impulsiona-uso-de-big-data-e-inteligencia-artificial/>
- Brandão, L. A., Shiroishi, M. S., & Law, M. (2013). Brain Tumors - A Multimodality Approach with Diffusion-Weighted Imaging, Diffusion Tensor Imaging, Magnetic Resonance Spectroscopy, Dynamic Susceptibility Contrast and Dynamic Contrast-Enhanced Magnetic Resonance Imaging. *Magnetic Resonance Imaging Clinics of North America*, 21, pp. 199-239. doi:<http://dx.doi.org/10.1016/j.mric.2013.02.003>
- Britannica, The Editors of Encyclopaedia. (21 de março de 2020). Brain. Encyclopedia Britannica. Obtido em 4 de agosto de 2021, de <https://www.britannica.com/science/brain>

- Brownlee, J. (8 de abril de 2020). 4 Types of Classification Tasks in Machine Learning. *Machine Learning Mastery*. Obtido em 31 de outubro de 2021, de <https://machinelearningmastery.com/types-of-classification-in-machine-learning/>
- Cao, L. (Junho de 2017). Data Science - A Comprehensive Overview. *ACM Computing Surveys*, 50(3). doi:<http://dx.doi.org/10.1145/3076253>
- Carozza, S. E., Wrench, M., Miike, R., Newman, B., Olshan, A. F., Savitz, D. A., . . . Lee, M. (2000). Occupation and Adult Gliomas. *American Journal of Epidemiology*, 152, pp. 838-846. Obtido de <https://doi.org/10.1093/aje/152.9.838>
- Cha, S. (julho de 2009). Neuroimaging in Neuro-Oncology. *Neurotherapeutics: The Journal of the American Society for Experimental NeuroTherapeutics*, 6, pp. 465-477. doi:<https://doi.org/10.1016/j.nurt.2009.05.002>
- Cheng, J., Chen, X., Yang, H., & Leng, M. (2012). An enhanced k-means algorithm using agglomerative hierarchical clustering strategy., (pp. 407-410).
- Countymeters. (4 de agosto de 2021). *População Mundial*. Obtido de Countymeters: <https://countymeters.info/pt>
- Crocetti, E., Trama, A., Stiller, C., Caldarella, A., Soffiatti, R., Jaal, J., . . . Brandes, A. (2012). Epidemiology of glial and non-glial brain tumours in Europe. *European Journal of Cancer*, 48, pp. 1532-1542. doi:10.1016/j.ejca.2011.12.013
- De Ross, A. J., Stewart, P. A., Linet, M. S., Heineman, E. F., Dosemeci, M., Wilcosky, T., . . . Inskip, P. D. (2003). Occupation and the risk of adult glioma in the United States. *Cancer Causes and Control*, 14, pp. 139-150. doi:10.1023/a:1023053916689
- Devasena, C. L., & Hemalatha, M. (2013). Efficient computer aided diagnosis of abnormal parts detection in magnetic resonance images using hybrid abnormality detection algorithm. *Central European Journal of Computer Science*, 3(3), pp. 117-128.
- DeVita, Jr, V. T., Lawrence, T. S., & Rosenberg, S. A. (2015). *Cancer: Principles and Practice of Oncology* (10^a ed.). Philadelphia, PA, EUA: Wolters Kluwer Health.
- Dicionário Informal. (22 de Fevereiro de 2010). *Iterativo*. Obtido em 21 de julho de 2021, de Dicionário Informal: <https://www.dicionarioinformal.com.br/iterativo/>
- Didática Tech. (2020). Como funciona o algoritmo Árvore de Decisão. Obtido em 31 de outubro de 2021, de <https://didatica.tech/como-funciona-o-algoritmo-arvore-de-decisao/>
- Didática Tech. (2020). O que é e como funciona o algoritmo Random Forest. Obtido em 01 de novembro de 2021, de <https://didatica.tech/o-que-e-e-como-funciona-o-algoritmo-randomforest/>
- Donges, N. (22 de julho de 2021). A Complete Guide to the Random Forest Algorithm. Built In. Obtido em 2 de novembro de 2021, de <https://builtin.com/data-science/random-forest-algorithm>
- El Naqa, I., Ruijiang, L., & Murphy, M. J. (2015). *Machine Learning in Radiation Oncology - Theory and Applications*. Switzerland: Springer Nature.

- Eman, A.-M., Mohammed, E., & Rashid, A.-A. (2015). Brain tumor segmentation based on a hybrid clustering technique. *Egyptian Informatics Journal*, 16(1), pp. 71-81.
- Estúdio ABC. (7 de julho de 2017). *Saúde entra na indústria 4.0*. Obtido em 28 de novembro de 2021, de Exame: <https://exame.com/tecnologia/saude-entra-na-era-da-industria-4-0/>
- Fapesp, P. M. (05 de 07 de 2017). Células de apoio do sistema nervoso ajudam a entender o Alzheimer. Obtido em 17 de agosto de 2021, de <https://jornal.usp.br/ciencias/ciencias-da-saude/celulas-de-apoio-do-sistema-nervoso-ajudam-a-entender-o-alzheimer/>
- Filho, K. d. (21 de setembro de 1998). *Imagens Digitais*. Obtido em 12 de março de 2021, de Departamento de Astronomia do Instituto de Física da UFRGS: <http://astro.if.ufrgs.br/med/imagens/node35.htm>
- Fink, J. R., Muzi, M., Peck, M., & Krohn, K. A. (2015). Multimodality Brain Tumor IMaging: MR Imaging, PET, and PET/MR Imaging. *The Journal of Nuclear Medicine*, 56, pp. 1554-1561. doi:10.2967/jnumed.113.131516
- Franco, G. (23 de março de 2015). Como utilizar o teste t de student. *SOS Estatística*. Obtido em 22 de novembro de 2021, de <https://sosestatistica.com.br/como-utilizar-o-teste-t-de-student/>
- Gandhi, R. (7 de junho de 2018). Support Vector Machine - Introduction to Machine Learning Algorithms. *Towards Data Science*. Obtido em 4 de novembro de 2021, de <https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>
- GitHub Inc. (2021). *GitHub*. Obtido de <https://github.com/>
- Gomes, F. C., Tortelli, V. P., & Diniz, L. (2013). Glia: dos velhos conceitos às novas funções de hoje e as que ainda virão. *Estudos Avançados*, 77, pp. 61-84.
- González, M. (23 de setembro de 2019). O que é acurácia? Entenda o conceito e sua importância. *idblog*. Obtido em 21 de novembro de 2021, de <https://blog.idwall.co/o-que-e-acuracia/>
- Gouveia, R. (s.d.). Medidas de Dispersão. *Toda Matéria*. Obtido em 22 de novembro de 2021, de <https://www.todamateria.com.br/medidas-de-dispersao/>
- Granatyr, J. (2016). *Expert Academy*. Obtido de <https://iaexpert.academy/2016/07/19/historico-da-ia-teste-de-turing/>
- Harrison, O. (10 de setembro de 2018). Machine Learning Basics with the K-Nearest Neighbors Algorithm. *Towards Data Science*. Obtido em 8 de novembro de 2021, de <https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761>
- Honda, H., Facure, M., & Yaohao, P. (27 de julho de 2017). Os Três Tipos de Aprendizado de Máquina. *LAMFO - Laboratório de Aprendizado de Máquina em Finanças e Organizações*. Obtido em 10 de setembro de 2021, de <https://lamfo-unb.github.io/2017/07/27/tres-tipos-am/>
- Hoppen, J., & Prates, W. (23 de abril de 2018). Datasets, o que são e como utilizá-los. *Aquarela*. Obtido em 9 de novembro de 2021, de <https://www.aquarela.com/datasets-o-que-sao-e-como-utiliza-los/>

- Hospital Sírio-Libanês. (2019). Câncer Cerebral. Obtido em 19 de agosto de 2021, de <https://www.hospitalsiriolibanes.org.br/hospital/especialidades/centro-oncologia/cerebral/Paginas/diagnosticos.aspx>
- Hurwitz, J., & Kirsch, D. (2018). *Machine Learning for Dummies*. Hoboken, New Jersey: John Wiley & Sons, Inc. .
- Hustinx, R., Pourdehnad, M., Kaschten, B., & Alavi, A. (2005). PET imaging for differentiating recurrent brain tumor from radiation necrosis. *Radiologic Clinics of North America*, 43(1), pp. 35-47.
- IARC. (2018). *GLOBOCAN - Global Cancer Incidence, Mortality and Prevalence*. International Agency for Research on Cancer - World Health Organization. Obtido em 21 de julho de 2021, de https://www.iarc.who.int/wp-content/uploads/2018/09/pr263_E.pdf
- IARC. (2019). Agents Classified by the IARC Monographs. 1-123. IARC Working Group. Obtido em 18 de agosto de 2021, de https://monographs.iarc.fr/wp-content/uploads/2019/02/List_of_Classifications.pdf
- IARC. (10 de dezembro de 2019). *IARC Monographs on the Identification of Carcinogenic Hazards to Humans*. Obtido em 20 de agosto de 2021, de https://monographs.iarc.who.int/wp-content/uploads/2018/07/QA_ENG.pdf
- IARC. (2019). *List of classifications by cancer site with sufficient or limited evidence in humans*. International Agency for Research on Cancer - World Health Organization. Obtido em 18 de agosto de 2021, de https://monographs.iarc.who.int/wp-content/uploads/2019/07/Classifications_by_cancer_site.pdf
- IARC. (2020). *GOBOCAN - Global Cancer Incidence, Mortality and Prevalence*. International Agency for Research on Cancer - World Health Organization. Obtido em 4 de agosto de 2021, de <https://gco.iarc.fr/today/data/factsheets/cancers/31-Brain-central-nervous-system-fact-sheet.pdf>
- IARC. (22 de julho de 2021). List of Classifications - Agents classified by the IARC Monographs. *IARC Monographs on the Identification of carcinogenic hazards to humans*, 1-129. Obtido em 18 de agosto de 2021, de <https://monographs.iarc.who.int/list-of-classifications>
- IBM. (19 de 03 de 2012). *90% of the data in the world today has been created in the last two years*. Obtido em 21 de 07 de 2021, de <https://www.facebook.com/IBM/posts/90-of-the-data-in-the-world-today-has-been-created-in-the-last-two-years/293229680748471/>
- IBM Cloud Education. (15 de julho de 2020). What is Machine Learning? Obtido em 18 de novembro de 2021, de <https://www.ibm.com/cloud/learn/machine-learning>
- IBM. (s.d.). *What is big data analytics?* Obtido em 21 de julho de 2021, de IBM: https://www.ibm.com/analytics/hadoop/big-data-analytics?mhsrc=ibmsearch_a&mhq=big%20data
- ICNIRP. (1998). Guidelines for limiting exposure to time varying electric, magnetic and electromagnetic fields (up to 300GHz). *Health Physics*, 74(4), pp. 494-522. Obtido de <http://www.icnirp.org/cms/upload/publications/ICNIRPemfgdlesp.pdf>

- IDC - International Data Corporation. (2020). *IDC Forecasts Strong 12.3% Growth for AI Market in 2020 Amidst Challenging Circumstances*. Framingham, Massachussets. Obtido em 21 de julho de 2021, de <https://www.idc.com/getdoc.jsp?containerId=prUS46757920>
- INCA. (2018). *Câncer do sistema nervoso central - Versão para Profissionais de Saúde*. Obtido em 2 de março de 2021, de Instituto Nacional do Câncer - Ministério da Saúde - Brasil: <https://www.inca.gov.br/tipos-de-cancer/cancer-do-sistema-nervoso-central/profissional-de-saude>
- INCA. (27 de janeiro de 2020). Como se comportam as células cancerosas? (I. N.-M. Brasil, Ed.) Brasil. Obtido em 10 de agosto de 2021, de <https://www.inca.gov.br/como-se-comportam-celulas-cancerosas>
- INCA. (4 de agosto de 2021). Como surge o câncer? Brasil: Instituto Nacional do Câncer - Ministério da Saúde do Brasil. Obtido em 10 de agosto de 2021, de <https://www.inca.gov.br/como-surge-o-cancer>
- Joshi, P. (1 de setembro de 2012). Kernel Functions for Machine Learning. *Perpetual Enigma*. Obtido em 6 de novembro de 2021, de <https://prateekvjoshi.com/2012/09/01/kernel-functions-for-machine-learning/>
- Jung, J.-h., & Ahn, B.-C. (2018). Current Radiopharmaceuticals for Positron Emission Tomography. *Brain Tumor Research and Treatment*, 6(2), pp. 47-53. doi:<https://doi.org/10.14791/btrt.2018.6.e13>
- Kaggle Inc. (2021). *Kaggle*. Obtido de <https://www.kaggle.com/>
- Kucukcirak, E. (março de 2021). *Kaggle*. Obtido em abril de 2021, de <https://www.kaggle.com/erkamk/brain-tumor-classification>
- Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. Nova Iorque: Springer Science + Business Media. doi:DOI 10.1007/978-1-4614-6849-3 1
- Langleben, D. D., & Segall, G. M. (2000). PET in Differentiation of Recurrent Brain Tumor from Radiation Injury. *The Journal of Nuclear Medicine*, 41, pp. 1861-1687.
- Lee, G., & Fujita, H. (2020). *Deep Learning in Medical Image Analysis - Challenges and Applications*. Switzerland: Springer Nature.
- Leite, C. d. (2019). Inteligência artificial, radiologia, medicina de precisão e medicina personalizada. *Colégio Brasileiro de Radiologia e Diagnóstico por Imagem*, 6(52), pp. VII-VIII.
- Lobo, L. C. (2017). Inteligência Artificial e Medicina. *Revista Brasileira de Educação Médica*, 41(2), pp. 185-193.
- Lobo, L. C. (2018). Inteligência artificial, o futuro da medicina e a educação médica. *Revista Brasileira de Educação Médica*, 42(3), pp. 3-8.
- Louis, D. N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W. K., . . . Ellison, D. W. (2016). The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathologica*, 131(6), pp. 803-820.
- Lusiadas. (2020). Tumor cerebral: sintomas, diagnóstico, tratamento e prognóstico. Obtido em 19 de agosto de 2021, de

<https://www.lusiadas.pt/blog/doencas/cancro/tumor-cerebral-sintomas-diagnostico-tratamento-prognostico>

- Mariano, D. (25 de abril de 2021). *Métricas de avaliação em machine learning*. Obtido de Diego Mariano: <https://diegomariano.com/metricas-de-avaliacao-em-machine-learning/>
- Matsuzaki, T. (1 de junho de 2020). Mathematical Introduction for SVM and Kernel Functions. *tsmatz*. Obtido em 6 de novembro de 2021, de <https://tsmatz.wordpress.com/2020/06/01/svm-and-kernel-functions-mathematics/>
- Matthews, P. B., Ratcliff, G., Nathan, P. W., Noback, C. R., Lentz, T. L., Haines, D. E., . . . Rudge, P. (9 de abril de 2020). Human Nervous System. Encyclopedia Britannica. Obtido em 4 de agosto de 2021, de <https://www.britannica.com/science/human-nervous-system>
- Mehta, A. (2021). Everything You Need to Know About Linear Discriminant Analysis. *Digital Vidya*. Obtido em 3 de novembro de 2021, de <https://www.digitalvidya.com/blog/linear-discriminant-analysis/>
- Miami Neuroscience Center . (s.d.). *Over 120 types of Brain Tumors*. Obtido de Miami Neuroscience Center: <https://miamineurosciencecenter.com/en/conditions/brain-tumors/types/>
- Mohsen, H., El-Dahshan, E.-S. A., El-Horbaty, E.-S. M., & Salem, A.-B. M. (2017). Classification using deep learning neural networks for brain tumors. *Future Computing and Informatics Journal*, 3, pp. 68-71.
- Nabizadeh, N., & Kubat, M. (2015). Brain tumor detection and segmentation in MR images: Gabor wavelet vs. statistical features. *Computers & Electrical Engineering*, 45, pp. 286-301.
- Nabizadeh, N., John, N., & Wright, C. (2014). Histogram-based gravitational optimization algorithm on single MR modality for automatic brain lesion detection and segmentation. *Expert Systems of Applications*, 41(17), pp. 7820-7836.
- Nickparvar, M. (2021). *Kaggle*. Obtido em 9 de novembro de 2021, de <https://www.kaggle.com/masoudnickparvar/brain-tumor-mri-dataset>
- Ohgaki, H., & Kleihues, P. (2005). Epidemiology and etiology of gliomas. *Acta Neuropathologica*, 109, pp. 93-108. doi:10.1007/s00401-005-0991-y
- Oliveto, P. (6 de setembro de 2015). *Primeiro caso de tumor maligno em humanos data de 35 mil anos*. Obtido em 28 de novembro de 2021, de Correio Braziliense: https://www.correiobraziliense.com.br/app/noticia/ciencia-e-saude/2015/09/06/interna_ciencia_saude,497488/primeiro-caso-de-tumor-maligno-em-humanos-data-de-35-mil-anos.shtml
- Oncoguia. (23 de setembro de 2015). Alterações nos Genes. (E. Oncoguia, Ed.) Obtido em 10 de agosto de 2021, de <http://www.oncoguia.org.br/conteudo/alteracoes-nos-genes/8160/73/>
- Oncoguia. (15 de dezembro de 2020). Tipos de tumores cerebrais/SNC. Obtido em 12 de agosto de 2021, de <http://www.oncoguia.org.br/conteudo/tipos-de-tumores-cerebrais-snc/894/293/>

- Organização Pan-Americana da Saúde. (Outubro de 2020). *Câncer*. Obtido em 21 de julho de 2021, de Organização Pan-Americana da Saúde: <https://www.paho.org/pt/topicos/cancer>
- Panesar, A. (2019). *Machine Learning and AI for Healthcare*. Nova Iorque: Apress Midia.
- Patel, V. L., Szolovitz, P., Stefanelli, M., Shortliffe, E. H., Berthold, M. R., Bellazzi, R., & Abu-Hanna, A. (2008). The coming of age of artificial intelligence in medicine. *Artificial Intelligence in Medicine*(46), pp. 5-17.
- Pawar, U. (4 de dezembro de 2020). Lets Open the Black Box of Random Forests. (A. Vidha, Ed.) Obtido em 2 de novembro de 2021, de <https://www.analyticsvidhya.com/blog/2020/12/lets-open-the-black-box-of-random-forests/>
- Pedrosa, L. (18 de maio de 2020). Imagens binárias. *MC102QR - Algoritmos e Programação de Computadores*. Obtido em 9 de novembro de 2021, de <https://www.ic.unicamp.br/~lehilton/mc102qr/tarefas/10-imagens-binarias.html>
- Pinto, A., Pereira, S., Rasteiro, D., & Silva, C. A. (5 de maio de 2018). Hierarchical Brain Tumor Segmentation using Extremely Randomized Trees. *Pattern Recognition*. doi:10.1016/j.patcog.2018.05.006
- Pittella, J. E. (2008). Biópsia estereotáxica no diagnóstico de tumores cerebrais e lesões não neoplásicas: indicações, acurácia e dificuldades diagnósticas. *Jornal Brasileiro de Patologia Médica Laboratorial*, 44(4), pp. 343-354.
- Polly, F. P., Shil, S. K., Hoassain, M. A., Ayman, A., & Jang, Y. M. (2018). Detection and Classification of HGG and LGG Brain Tumor Using Machine Learning. *32nd International Conference on Information Networking (ICOIN 2018)* (pp. 813-817). Chiang Mai, Tailândia: IEEE Computer Society. doi:Detection and Classification of HGG and LGG Brain
- Porto Editora. (s.d.). Neoplasia. *Dicionário Infopédia da Língua Portuguesa*. Obtido em 20 de agosto de 2021, de <https://www.infopedia.pt/dicionarios/lingua-portuguesa/neoplasia>
- Prates, W. R. (28 de novembro de 2020). Curva ROC e AUC em Machine Learning. *Ciência e Negócios*. Obtido em 11 de outubro de 2021, de <https://cienciaenegocios.com/curva-roc-e-auc-em-machine-learning/>
- Reinsel, D., Gantz, J., & Rydning, J. (2017). Data Age 2025: The Evolution of Data to Life-Critical. *An IDC White Paper*. Obtido de <https://www.import.io/wp-content/uploads/2017/04/Seagate-WP-DataAge2025-March-2017.pdf>
- Reis, B. (1 de setembro de 2020). *Câncer - a trajetória da doença*. Obtido em 28 de novembro de 2021, de Superinteressante: <https://super.abril.com.br/especiais/cancer-a-trajetoria-da-doenca/>
- Roesler, R., Dini, S., & Isolan, G. (24 de setembro de 2021). Brain tumours, brain metastases, and neuroinflammation: Insights from neuroimaging studies. *Authorea*. doi:DOI: 10.22541/au.163252299.90555284/v1
- Rogers, S., & Girolami, M. (2012). *A First Course in Machine Learning*. Boca Raton, Florida: CRC Press.

- Roveda, U. (novembro de 2019). Google Colab: o que é, como usar e quais são as vantagens? *Kenzie*. Obtido em 21 de novembro de 2021, de <https://kenzie.com.br/blog/google-colab/>
- Sachdeva, J., Kumar, V., Gupta, I., Khandelwal, N., & Ahuja, C. K. (4 de maio de 2013). Segmentation, Feature Extraction, and Multiclass Brain. *Journal of Digital Imaging*. doi:10.1007/s10278-013-9600-0
- Samavedam, A. (30 de dezembro de 2020). Data Classification Using k-Nearest Neighbors. *Laptrinhx*. Obtido em 8 de novembro de 2021, de <https://laptrinhx.com/data-classification-using-k-nearest-neighbors-3224663474/>
- Santos, T. G. (19 de abril de 2020). Google Colab: O que é, Tutorial de Como Usar e Criar Códigos. *Alura*. Obtido em 21 de novembro de 2021, de <https://www.alura.com.br/artigos/google-colab-o-que-e-e-como-usar>
- SciKit Learn. (s.d.). Decision Trees. *SciKit-Learn - Machine Learning in Python*. Obtido em 31 de outubro de 2021, de <https://scikit-learn.org/stable/modules/tree.html#id2>
- SciKit Learn. (s.d.). Nearest Neighbors. Obtido em 8 de novembro de 2021, de <https://scikit-learn.org/stable/modules/neighbors.html#nearest-neighbors-classification>
- SciKit Learn. (s.d.). sklearn.ensemble.RandomForestClassifier. Obtido em 1 de novembro de 2021, de <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>
- SciKit Learn. (s.d.). sklearn.tree.DecisionTreeClassifier. Obtido em 31 de outubro de 2021, de <https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html>
- SciKit Learn. (s.d.). Support Vector Machines. Obtido em 6 de novembro de 2021, de <https://scikit-learn.org/stable/modules/svm.html#kernel-functions>
- SEER Training Modules. (s.d.). Cancer Classification. US National Institutes of Health, National Cancer Institute. Obtido em 10 de agosto de 2021, de <https://training.seer.cancer.gov/disease/categories/classification.html>
- SEER Training Modules. (s.d.). Cancer Diagnosis. US National Institutes of Health, National Cancer Institute. Obtido em 12 de agosto de 2021, de <https://training.seer.cancer.gov/disease/diagnosis/>
- SEER Training Modules. (s.d.). Cancer Types by Site. US National Institutes of Health, National Cancer Institute. Obtido em 12 de agosto de 2021, de <https://training.seer.cancer.gov/disease/categories/site.html>
- Sickit Learn. (2021). Linear and Quadratic Discriminant Analysis. Obtido em 3 de novembro de 2021, de https://scikit-learn.org/stable/modules/lda_qda.html
- Silva, J. F., & Macedo, A. I. (10 de outubro de 2020). Análise Discriminante Linear. *LAMFO - Laboratório de Aprendizado de Máquina em Finanças e Organizações - Universidade de Brasília*. Obtido em 3 de novembro de 2021, de <https://lamfo-unb.github.io/2020/10/10/An%C3%A1lise-Discriminante-Linear/>

- Singh, V. (24 de janeiro de 2018). Machine Learning K-Nearest Neighbors (KNN) Algorithm in Python - An Introduction. *Quantinsti*. Obtido em 8 de novembro de 2021, de <https://blog.quantinsti.com/machine-learning-k-nearest-neighbors-knn-algorithm-python/>
- Siqueira, E. M., & Diccini, S. (2017). Complicações pós-operatórias em neurocirurgia eletiva e não eletiva. *Acta Paulista de Enfermagem*, 30(1), pp. 101-108.
- Sociedade Brasileira de Dermatologia. (2017). Câncer da pele. Obtido em 19 de agosto de 2021, de <https://www.sbd.org.br/dermatologia/pele/doencas-e-problemas/cancer-da-pele/64/>
- Southern European Countries 2021*. (s.d.). Obtido em 4 de agosto de 2021, de World Population Review: <https://worldpopulationreview.com/country-rankings/southern-european-countries>
- Souza, E. G. (27 de março de 2019). *Entendendo o que é Matriz de Confusão com Python*. Obtido de Medium: <https://medium.com/data-hackers/entendendo-o-que-%C3%A9-matriz-de-confus%C3%A3o-com-python-114e683ec509>
- Statistics Kingdom. (s.d.). Shapiro-Wilk Test Calculator. Obtido em 22 de novembro de 2021, de <https://www.statskingdom.com/shapiro-wilk-test-calculator.html>
- Subashini, M. M., Sahoo, S. K., Sunil, V., & Easwaran, S. (2016). A non-invasive methodology for the grade identification of astrocytoma using image processing and artificial intelligence techniques. *Expert Systems of Applications*, 43, pp. 186-196.
- Sudharani, K., Sarma, T. C., & Prasad, K. S. (2016). Advanced morphological technique for automatic brain tumor detection and evaluation of statistical parameters. *Procedia Technology*, 24, pp. 1374-1387.
- Takenaka, S., Asano, Y., Shinoda, J., Nomura, Y., Yonezawa, S., Miwa, K., . . . Iwama, T. (2014). Comparison of 11C-Methionine, 11C-Choline, and 18F-Fluorodeoxyglucose-Positron Emission Tomography for Distinguishing Glioma Recurrence from Radiation Necrosis. *Neurologia Medico-Chirurgica*, 54, pp. 280-289. doi:10.2176/nmc.0a2013-0117
- TechLib. (s.d.). Tons de cinza. Obtido em 9 de novembro de 2021, de <https://techlib.wiki/definition/grayscale.html>
- Techopedia. (20 de janeiro de 2013). *Volume Pixel (Volume Pixel or Voxel)*. Obtido em 12 de março de 2021, de Techopedia: <https://www.techopedia.com/definition/2055/volume-pixel-volume-pixel-or-voxel>
- Trammit Medical. (s.d.). Obtido em 31 de agosto de 2021, de <http://www.trammit.com.br/equipamentos-para-neurologia-/2841-locacao-de-sistema-estereotaxico-para-neurocirurgia-e-neuromodulacao.html>
- Usman, K., & Rajpoot, K. (2017). Brain tumor classification from multimodality MRI using wavelets and Machine Learning. *Pattern Analysis and Applications*, 20, pp. 871-881.
- VanPutte, C. L., Regan, J. L., Russo, A. F., Seeley, R., Stephens, T., & Tate, P. (2016). *Anatomia e Fisiologia de Seeley* (10ª ed.). Porto Alegre, RS, Brasil: AMGH Editora Ltda.

- Vassallo, J. (s.d.). Histofisiologia do folículo linfóide e da produção de células B. Obtido em 10 de agosto de 2021, de <http://anatpat.unicamp.br/talinfonodo.html>
- Weaver II, E. A., & Doyle, H. H. (2019). Cells fo the Brain, grades 9-12 (Portuguese). The Dana Alliance for Brain Initiatives. Obtido de The Dana Alliance for Brain Initiatives: https://dana.org/wp-content/uploads/2019/05/Cells-of-the-Brain_Grades-9-12_Portuguese.pdf
- Wikipedia. (15 de abril de 2021). GitHub. Obtido em 20 de novembro de 2021, de <https://pt.wikipedia.org/wiki/GitHub>
- Wikipedia. (6 de novembro de 2021). Kaggle. Obtido em 20 de novembro de 2021, de <https://en.wikipedia.org/wiki/Kaggle>
- Zacharaki, E. I., Wang, S., Chawla, S., Yoo, D. S., Wolf, R., Melhem, E. R., & Davatzikos, C. (2009). Classification of Brain Tumor Type and Grade Using MRI in a Machine Learning Scheme. *Magnetic Resonance in Medicine*, 62, pp. 1609-1618.
- Zhang, G. (11 de novembro de 2018). What is the kernel trick? Why is it important? *Medium*. Obtido em 5 de novembro de 2021, de <https://medium.com/@zxr.nju/what-is-the-kernel-trick-why-is-it-important-98a98db0961d>
- Zoltan, C. (13 de novembro de 2018). SVM and Kernel SVM. *Towards Data Science*. Obtido em 6 de novembro de 2021, de <https://towardsdatascience.com/svm-and-kernel-svm-fed02bef1200#tuning>
- Zorzi, R. (s.d.). *Tumores Cerebrais em Adultos*. Obtido em 2 de março de 2021, de Dra Raquel Zorzi: <https://www.draraquelzorzi.com.br/doencas/detalhes/57/Tumores%20cerebrais%20em%20adultos>

Anexo A – Tabela de Classificação de Tumores do SNC da OMS (Louis, *et al.*, 2016)

WHO classification of tumours of the central nervous system

Diffuse astrocytic and oligodendroglial tumours		Neuronal and mixed neuronal-glial tumours	
Diffuse astrocytoma, IDH-mutant	9400/3	Dysembryoplastic neuroepithelial tumour	9413/0
Gemistocytic astrocytoma, IDH-mutant	9411/3	Gangliocytoma	9492/0
<i>Diffuse astrocytoma, IDH-wildtype</i>	9400/3	Ganglioglioma	9505/1
Diffuse astrocytoma, NOS	9400/3	Anaplastic ganglioglioma	9505/3
		Dysplastic cerebellar gangliocytoma (Lhermitte-Duclos disease)	9493/0
Anaplastic astrocytoma, IDH-mutant	9401/3	Desmoplastic infantile astrocytoma and ganglioglioma	9412/1
<i>Anaplastic astrocytoma, IDH-wildtype</i>	9401/3	Papillary glioneuronal tumour	9509/1
Anaplastic astrocytoma, NOS	9401/3	Rosette-forming glioneuronal tumour	9509/1
		<i>Diffuse leptomeningeal glioneuronal tumour</i>	
Glioblastoma, IDH-wildtype	9440/3	Central neurocytoma	9506/1
Giant cell glioblastoma	9441/3	Extraventricular neurocytoma	9506/1
Gliosarcoma	9442/3	Cerebellar liponeurocytoma	9506/1
<i>Epithelioid glioblastoma</i>	9440/3	Paraganglioma	8693/1
Glioblastoma, IDH-mutant	9445/3*		
Glioblastoma, NOS	9440/3		
		Tumours of the pineal region	
Diffuse midline glioma, H3 K27M-mutant	9385/3*	Pineocytoma	9361/1
		Pineal parenchymal tumour of intermediate differentiation	9362/3
Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	9450/3	Pineoblastoma	9362/3
Oligodendroglioma, NOS	9450/3	Papillary tumour of the pineal region	9395/3
Anaplastic oligodendroglioma, IDH-mutant and 1p/19q-codeleted	9451/3	Embryonal tumours	
<i>Anaplastic oligodendroglioma, NOS</i>	9451/3	Medulloblastomas, genetically defined	
		Medulloblastoma, WNT-activated	9475/3*
<i>Oligoastrocytoma, NOS</i>	9382/3	Medulloblastoma, SHH-activated and TP53-mutant	9476/3*
<i>Anaplastic oligoastrocytoma, NOS</i>	9382/3	Medulloblastoma, SHH-activated and TP53-wildtype	9471/3
		Medulloblastoma, non-WNT/non-SHH	9477/3*
Other astrocytic tumours		<i>Medulloblastoma, group 3</i>	
Pilocytic astrocytoma	9421/1	<i>Medulloblastoma, group 4</i>	
Pilomyxoid astrocytoma	9425/3	Medulloblastomas, histologically defined	
Subependymal giant cell astrocytoma	9384/1	Medulloblastoma, classic	9470/3
Pleomorphic xanthoastrocytoma	9424/3	Medulloblastoma, desmoplastic/nodular	9471/3
Anaplastic pleomorphic xanthoastrocytoma	9424/3	Medulloblastoma with extensive nodularity	9471/3
		Medulloblastoma, large cell / anaplastic	9474/3
		Medulloblastoma, NOS	9470/3
Ependymal tumours			
Subependymoma	9383/1	Embryonal tumour with multilayered rosettes, C19MC-altered	9478/3*
Myxopapillary ependymoma	9394/1	<i>Embryonal tumour with multilayered rosettes, NOS</i>	9478/3
Ependymoma	9391/3	Medulloepithelioma	9501/3
Papillary ependymoma	9393/3	CNS neuroblastoma	9500/3
Clear cell ependymoma	9391/3	CNS ganglioneuroblastoma	9490/3
Tanycytic ependymoma	9391/3	CNS embryonal tumour, NOS	9473/3
Ependymoma, <i>RELA</i> fusion-positive	9396/3*	Atypical teratoid/rhabdoid tumour	9508/3
Anaplastic ependymoma	9392/3	<i>CNS embryonal tumour with rhabdoid features</i>	9508/3
Other gliomas		Tumours of the cranial and paraspinal nerves	
Chordoid glioma of the third ventricle	9444/1	Schwannoma	9560/0
Angiocentric glioma	9431/1	Cellular schwannoma	9560/0
Astroblastoma	9430/3	Plexiform schwannoma	9560/0
Choroid plexus tumours			
Choroid plexus papilloma	9390/0		
Atypical choroid plexus papilloma	9390/1		
Choroid plexus carcinoma	9390/3		

Melanotic schwannoma	9560/1	Osteochondroma	9210/0
Neurofibroma	9540/0	Osteosarcoma	9180/3
Atypical neurofibroma	9540/0		
Plexiform neurofibroma	9550/0	Melanocytic tumours	
Perineurioma	9571/0	Meningeal melanocytosis	8728/0
Hybrid nerve sheath tumours		Meningeal melanocytoma	8728/1
Malignant peripheral nerve sheath tumour	9540/3	Meningeal melanoma	8720/3
Epithelioid MPNST	9540/3	Meningeal melanomatosis	8728/3
MPNST with perineurial differentiation	9540/3		
		Lymphomas	
Meningiomas		Diffuse large B-cell lymphoma of the CNS	9680/3
Meningioma	9530/0	Immunodeficiency-associated CNS lymphomas	
Meningothelial meningioma	9531/0	AIDS-related diffuse large B-cell lymphoma	
Fibrous meningioma	9532/0	EBV-positive diffuse large B-cell lymphoma, NOS	
Transitional meningioma	9537/0	Lymphomatoid granulomatosis	9766/1
Psammomatous meningioma	9533/0	Intravascular large B-cell lymphoma	9712/3
Angiomatous meningioma	9534/0	Low-grade B-cell lymphomas of the CNS	
Microcystic meningioma	9530/0	T-cell and NK/T-cell lymphomas of the CNS	
Secretory meningioma	9530/0	Anaplastic large cell lymphoma, ALK-positive	9714/3
Lymphoplasmacyte-rich meningioma	9530/0	Anaplastic large cell lymphoma, ALK-negative	9702/3
Metaplastic meningioma	9530/0	MALT lymphoma of the dura	9699/3
Chordoid meningioma	9538/1		
Clear cell meningioma	9538/1	Histiocytic tumours	
Atypical meningioma	9539/1	Langerhans cell histiocytosis	9751/3
Papillary meningioma	9538/3	Erdheim-Chester disease	9750/1
Rhabdoid meningioma	9538/3	Rosai-Dorfman disease	
Anaplastic (malignant) meningioma	9530/3	Juvenile xanthogranuloma	
		Histiocytic sarcoma	9755/3
		Germ cell tumours	
Mesenchymal, non-meningothelial tumours		Germinoma	9064/3
Solitary fibrous tumour / haemangiopericytoma**		Embryonal carcinoma	9070/3
Grade 1	8815/0	Yolk sac tumour	9071/3
Grade 2	8815/1	Choriocarcinoma	9100/3
Grade 3	8815/3	Teratoma	9080/1
Haemangioblastoma	9161/1	Mature teratoma	9080/0
Haemangioma	9120/0	Immature teratoma	9080/3
Epithelioid haemangioendothelioma	9133/3	Teratoma with malignant transformation	9084/3
Angiosarcoma	9120/3	Mixed germ cell tumour	9085/3
Kaposi sarcoma	9140/3		
Ewing sarcoma / PNET	9364/3	Tumours of the sellar region	
Lipoma	8850/0	Craniopharyngioma	9350/1
Angiolipoma	8861/0	Adamantinomatous craniopharyngioma	9351/1
Hibernoma	8880/0	Papillary craniopharyngioma	9352/1
Liposarcoma	8850/3	Granular cell tumour of the sellar region	9582/0
Desmoid-type fibromatosis	8821/1	Pituicytoma	9432/1
Myofibroblastoma	8825/0	Spindle cell oncocyoma	8290/0
Inflammatory myofibroblastic tumour	8825/1		
Benign fibrous histiocytoma	8830/0	Metastatic tumours	
Fibrosarcoma	8810/3		
Undifferentiated pleomorphic sarcoma / malignant fibrous histiocytoma	8802/3		
Leiomyoma	8890/0		
Leiomyosarcoma	8890/3		
Rhabdomyoma	8900/0		
Rhabdomyosarcoma	8900/3		
Chondroma	9220/0		
Chondrosarcoma	9220/3		
Osteoma	9180/0		

The morphology codes are from the International Classification of Diseases for Oncology (ICD-O) (742A). Behaviour is coded /0 for benign tumours; /1 for unspecified, borderline, or uncertain behaviour; /2 for carcinoma in situ and grade III intrasphinctal neoplasia; and /3 for malignant tumours. The classification is modified from the previous WHO classification, taking into account changes in our understanding of these lesions.
*These new codes were approved by the IARC/WHO Committee for ICD-O.
Italic: Provisional tumour entities. **Grading according to the 2013 WHO Classification of Tumours of Soft Tissue and Bone.

Apêndice A – Código utilizado

Importação das bibliotecas

```
import pandas as pd
import seaborn as sns
from sklearn import svm
from sklearn.model_selection import train_test_split
import os
import numpy as np
import cv2
import matplotlib.pyplot as plt
from sklearn.ensemble import RandomForestClassifier
from sklearn import tree
from sklearn.metrics import accuracy_score
from sklearn.model_selection import cross_val_score
from keras.utils.np_utils import to_categorical
from google.colab.patches import cv2_imshow
from sklearn.model_selection import KFold
from sklearn.neighbors import KNeighborsClassifier
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
```

Fornecimento do caminho

```
DATADIR = r"/content/drive/MyDrive/Estudos/Mestrado_MTFM/TESE/Dados/Dataset2/Subgrupo"
CATEGORIES = ["Glioma", "Meningioma", "NoTumor", "Pituitary"]
for category in CATEGORIES:
    path = os.path.join(DATADIR, category)
    for img in os.listdir(path):
        img_array = cv2.imread(os.path.join(path, img))
        break
    break
```

Redimensionamento das imagens

```
IMG_SIZE = 150
new_array = cv2.resize(img_array, (IMG_SIZE, IMG_SIZE))
```

Criando os arrays

```
▶ training_data = []

def create_training_data():
    for category in CATEGORIES:
        path = os.path.join(DATADIR, category)
        class_num = CATEGORIES.index(category)
        for img in os.listdir(path):
            try:
                img_array = cv2.imread(os.path.join(path, img), cv2.IMREAD_GRAYSCALE)
                new_array = cv2.resize(img_array, (IMG_SIZE, IMG_SIZE))
                training_data.append([new_array, class_num])
            except Exception as e:
                pass
    create_training_data()
```

Dimensionando os arrays

```
▶ X = []
y = []
for features, label in training_data:
    X.append(features)
    y.append(label)
X = np.array(X).reshape(-1, IMG_SIZE, IMG_SIZE)
print(X.shape)
X = X/255.0
X = X.reshape(-1, 150, 150, 1)
```

Fornecendo as classificações, o número e o split (0.8)

```
✓ [1] tumor_classifications = os.listdir(r"/content/drive/MyDrive/Estudos/Mestrado MTFM/TESE/Dados/Dataset2/Subgrupo")
```

```
[ ] y = to_categorical(y, num_classes = 4)
```

```
[ ] X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20, random_state=0)
```

Redimensionando os arrays de treino e teste

```
▶ X_train = X_train.reshape(9184, 5625)
X_test = X_test.reshape(2296, 5625)
y_train = y_train.reshape(9184, 1)
y_test = y_test.reshape(2296, 1)
```

Algoritmo *Decision Tree*

```
▶ classifierDT = tree.DecisionTreeClassifier()
classifierDT.fit(X_train, y_train)
ClassificationDT = classifierDT.predict(X_test)
print("Acurácia: %.3f" % (accuracy_score(ClassificationDT, y_test)*100))
```

Acurácia: 90.941

Algoritmo *Random Forest* para 1000 árvores

```
▶ classifierRF = RandomForestClassifier(n_estimators=1000)
classifierRF.fit(X_train, y_train)
Classification = classifierRF.predict(X_test)
print("Acurácia: %.3f" % (accuracy_score(Classification, y_test)*100))
```

Acurácia: 96.908

Algoritmo *Linear Discriminant Analysis*

```
▶ clfLDA = LinearDiscriminantAnalysis()
clfLDA.fit(X_train, y_train)
ClassificationLDA = clfLDA.predict(X_test)
print("Acurácia: %.3f" % (accuracy_score(ClassificationLDA, y_test)*100))
```

Acurácia: 68.162

Algoritmo SVM com *kernel* Linear e C = 1

```
▶ C = 1.0
clf = svm.SVC(kernel='linear', C=C)
clf.fit(X_train, y_train)
ClassificationSVM = clf.predict(X_test)
print("Acurácia: %.3f" % (accuracy_score(ClassificationSVM, y_test)*100))
```

```
/usr/local/lib/python3.7/dist-packages/sklearn/utils/validation.py:760: D
y = column_or_1d(y, warn=True)
Acurácia: 74.608
```

Algoritmo SVM com *kernel* RBF e C = 10

```
▶ C = 10.0
clfrbf10 = svm.SVC(kernel='rbf', C=C)
clfrbf10.fit(X_train, y_train)
ClassificationSVMrbf10 = clfrbf10.predict(X_test)
print("Acurácia: %.3f" % (accuracy_score(ClassificationSVMrbf10, y_test)*100))
```

```
/usr/local/lib/python3.7/dist-packages/sklearn/utils/validation.py:760: DataCo
y = column_or_1d(y, warn=True)
Acurácia: 93.206
```

Algoritmo SVM com *kernel* Polinomial e C = 50

```
▶ C = 50.0
clfpoli50 = svm.SVC(kernel='poly', C=C)
clfpoli50.fit(X_train, y_train)
ClassificationSVMpoli50 = clfpoli50.predict(X_test)
print("Acurácia: %.3f" %(accuracy_score(ClassificationSVMpoli50, y_test)*100))

/usr/local/lib/python3.7/dist-packages/sklearn/utils/validation.py:760: DataCon
  y = column_or_1d(y, warn=True)
Acurácia: 87.979
```

Algoritmo com *kernel* Sigmoide e C = 100

```
▶ C = 100.0
clfsig100 = svm.SVC(kernel='sigmoid', C=C)
clfsig100.fit(X_train, y_train)
ClassificationSVMsig100 = clfsig100.predict(X_test)
print("Acurácia: %.3f" %(accuracy_score(ClassificationSVMsig100, y_test)*100))

/usr/local/lib/python3.7/dist-packages/sklearn/utils/validation.py:760: DataCon
  y = column_or_1d(y, warn=True)
Acurácia: 56.794
```

Algoritmo kNN com 10 vizinhos

```
▶ clfKNN = KNeighborsClassifier(n_neighbors=10)
clfKNN.fit(X_train,y_train)
ClassificationKNN = clfKNN.predict(X_test)
print("Acurácia: %.3f" %(accuracy_score(ClassificationKNN, y_test)*100))

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:2: DataConve
Acurácia: 84.669
```