

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO

NATÁLIA SANTANA CHIARI CORREIA

Classificação computadorizada de imagens de ressonância magnética da coluna vertebral para diferenciação de fraturas vertebrais por compressão benignas e malignas utilizando uma abordagem 3D

Ribeirão Preto
2022

NATÁLIA SANTANA CHIARI CORREIA

Classificação computadorizada de imagens de ressonância magnética da coluna vertebral para diferenciação de fraturas vertebrais por compressão benignas e malignas utilizando uma abordagem 3D

Versão Corrigida

Tese apresentada ao Programa de Pós-Graduação em Medicina (Clínica Médica) da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo para obtenção do título de Doutora em Ciências.

Orientador: Prof. Dr. Paulo Mazzoncini de Azevedo Marques

Coorientador: Prof. Dr. Marcello Henrique Nogueira Barbosa

Ribeirão Preto
2022

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Chiari-Correia, Natália Santana

Classificação computadorizada de imagens de ressonância magnética da coluna vertebral para diferenciação de fraturas vertebrais por compressão benignas e malignas utilizando uma abordagem 3D. Ribeirão Preto, 2010.

128 p. : il. ; 30 cm

Tese de Doutorado apresentada à Faculdade de Medicina de Ribeirão Preto/USP. Área de concentração: Clínica Médica.

Orientador: Azevedo-Marques, Paulo Mazzoncini.

1. Coluna Vertebral 2. Redes Neurais Artificiais. 3. Fraturas Vertebrais por Compressão. 4. Atributos Radiômicos. 5. Aprendizado de Máquina. 6. Ressonância Magnética

NATÁLIA SANTANA CHIARI CORREIA

Classificação computadorizada de imagens de ressonância magnética da coluna vertebral para diferenciação de fraturas vertebrais por compressão benignas e malignas utilizando uma abordagem 3D

Tese apresentada ao Programa de Pós-Graduação em Medicina (Clínica Médica) da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo para obtenção do título de Doutora em Ciências.

Aprovado em: ____ / ____ / _____

BANCA EXAMINADORA

Prof.(a) Dr.(a) _____
Instituição _____
Julgamento _____

Prof.(a) Dr.(a) _____
Instituição _____
Julgamento _____

Prof.(a) Dr.(a) _____
Instituição _____
Julgamento _____

Prof.(a) Dr.(a) _____
Instituição _____
Julgamento _____

À minha família, pois tudo o que eu faço é por ela e para ela.

AGRADECIMENTOS

Aos meus pais, Romualdo e Ivaneide, pois mesmo com as dificuldades, sempre priorizaram a educação na vida de suas filhas. Devo tudo a vocês.

Ao meu esposo Rodolfo, amor da minha vida, meu maior incentivador, a pessoa que mais acredita na minha capacidade e que me apoiou em todos os momentos durante essa jornada. Eu amo você!

Às minhas lindas irmãs Tata e Angélica, minhas inspirações, que estão sempre vibrando com minhas conquistas e dispostas a me ajudar no que for preciso.

Ao meu sobrinho e afilhado Enzo, que eu amo mais que o infinito, por me fazer sentir uma super-heroína.

Aos meus sogros, José Paulo e Cidinha, e aos meus cunhados Edgar, Humberto, Henrique e Priscila, por todas as palavras de incentivo.

Ao meu coorientador Marcello, por toda sua paciência e disposição para compartilhar seus conhecimentos.

Ao meu orientador Paulo, por toda confiança em mim e no meu trabalho. O senhor é o melhor professor que eu já tive na vida!

Este estudo foi apoiado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, bolsa nº 88887.371177/2019-00 e Código 001) e Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, bolsa nº 2016/17078-0)

A Ciência salva. O negacionismo mata.

RESUMO

CHIARI-CORREIA, N. S. **Classificação computadorizada de imagens de ressonância magnética da coluna vertebral para diferenciação de fraturas vertebrais por compressão benignas e malignas utilizando uma abordagem 3D.** 2022. Tese (Doutorado) – Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2022.

Introdução: O estudo para diferenciação de fraturas vertebrais por compressão (FVCs) benignas (secundárias à osteoporose) e malignas (secundárias à infiltração neoplásica) tem sido objeto de pesquisa por mais de 20 anos e os critérios utilizados para realizar o diagnóstico diferencial em exames de ressonância magnética (RM), que é a modalidade mais relevante para o diagnóstico de doenças malignas da coluna vertebral, ainda são considerados inespecíficos. A Radiômica, que é a prática de extrair características quantitativas de imagens médicas para gerar dados que possam ser analisados por meio de técnicas de mineração, tem obtido ótimos resultados na criação de modelos para predição diagnóstica, prognóstica e de resposta à terapia. Assim, nosso objetivo nesse trabalho foi utilizar a abordagem radiômica para extrair informações quantitativas tridimensionais de exames de RM da coluna vertebral e investigar a possibilidade de encontrar uma assinatura radiômica que, associada a um modelo computacional, fosse capaz de diferenciar FVCs malignas e benignas.

Material e Métodos: Este estudo retrospectivo analisou exames de RM da coluna lombar na sequência sagital ponderada em T1 de 91 pacientes (36 homens e 55 mulheres, com idade média de $64,24 \pm 11,75$ anos), que foram diagnosticados com uma ou mais FVCs no período de 2010 a 2019 e que estavam sendo acompanhados no Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto (HCFMRP-USP). 146 corpos vertebrais foram segmentados e tiveram suas características quantitativas extraídas tridimensionalmente. Diferentes combinações metodológicas para cada etapa do reconhecimento de padrão em imagens foram exploradas e a configuração que obteve os melhores resultados foi a segmentação semiautomática realizada com o software 3D Slicer, a extração de atributos quantitativos tridimensionais utilizando a biblioteca *PyRadiomics* e a classificação com rede neural artificial *Multilayer Perceptron*. Para validar o modelo, foram utilizadas as técnicas de validação cruzada de 10 *folds* (validação interna) e a classificação de um conjunto de teste independente (validação externa). A avaliação foi realizada por meio das métricas de acurácia, área sob a curva ROC, sensibilidade e especificidade, com um intervalo de confiança de 95%.

Resultados: O conjunto de treinamento foi formado por 100 corpos vertebrais fraturados (54 fraturas benignas e 46 malignas) provenientes de 61 pacientes com idade média de $63,2 \pm 12,5$ anos e o conjunto de teste compreendeu 30 corpos vertebrais fraturados (15 fraturas benignas e 15 malignas) provenientes de 30 pacientes com idade média de $66,4 \pm 9,9$ anos. No teste de validação interna, realizado entre 30 modelos gerados por meio da randomização do conjunto de dados, as médias para acurácia, AUC ROC, sensibilidade e especificidade foram, respectivamente, 90,97% (IC de 95%: 90,22%, 91,72%), 0,97 (IC de

95%: 0,966, 0,974), 89,36% (IC de 95%: 88,18%, 90,54%) e 92,33% (IC de 95%: 91,34%, 93,32%). Na validação interna do melhor modelo, essas mesmas métricas alcançaram os seguintes valores: acurácia = 95%, AUC ROC = 0,98, sensibilidade = 93,5% e especificidade = 96,3%. No teste de validação externa, por sua vez, os valores para acurácia, AUC ROC, sensibilidade e especificidade foram, respectivamente, 93,3%, 0,97, 93,3% e 93,3%.

Conclusão: O modelo baseado em rede neural proposto neste estudo usando características radiômicas tridimensionais foi capaz de diferenciar fraturas vertebrais por compressão benignas e malignas com excelente desempenho, superando as limitações de estudos anteriores. Sua provável maior contribuição seria aumentar a eficiência diagnóstica dos radiologistas para essas duas condições clínicas, possibilitando uma eventual redução do número de biópsias realizadas e do tempo gasto para liberação do diagnóstico final.

Palavras-chave: coluna vertebral, redes neurais artificiais, fraturas vertebrais por compressão, atributos radiômicos, aprendizado de máquina, ressonância magnética.

ABSTRACT

CHIARI-CORREIA, N. S. **Computerized classification of spinal MRI images to differentiate benign from malignant vertebral compression fractures using a 3D approach.** 2022. Tese (Doutorado) – Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2022.

Introduction: The study to differentiate between benign (secondary to osteoporosis) and malignant (secondary to neoplastic infiltration) vertebral compression fractures (VCFs) has been the subject of research for more than 20 years and the criteria used to perform the differential diagnosis through magnetic resonance imaging (MRI), which is the most relevant modality for the diagnosis of malignant diseases of the spine, are still considered non-specific. Radiomics, which is the practice of extracting quantitative characteristics from medical images to generate data that can be analyzed using data mining, has obtained excellent results creating models for diagnostic, prognostic, and therapy responses predictions. Thus, our objective in this work was to use the radiomics approach to extract three-dimensional quantitative information from spinal MRIs and investigate the possibility of finding a radiomics signature that, associated with a computational model, would be able to differentiate malignant from benign FVCs.

Material and Methods: This retrospective study analyzed sagittal T1-weighted lumbar spine MRIs from 91 patients (36 men and 55 women, with an average age of 64.24 ± 11.75 years) diagnosed with benign or malignant VCFs from 2010 to 2019 who underwent treatment at the University Hospital. The 146 lumbar fractures were three-dimensionally segmented and had their 3D radiomic features extracted. Different methodological approaches for each stage of pattern recognition in images were explored and the configuration that obtained the best results was the semi-automatic segmentation performed by using 3D Slicer software, the extraction of 3D quantitative features by using PyRadiomics package, and the classification with the artificial neural network Multilayer Perceptron. To validate the model, the 10-fold cross-validation technique and an independent test set were used. The performance of the model was evaluated using the average with a 95% confidence interval for accuracy, the Roc AUC, sensitivity, and specificity.

Results: The training set consisted of 100 fractured vertebral bodies from 61 patients (average age of 63.2 ± 12.5 years) and the test set was comprised of 30 fractured vertebral bodies from 30 patients (average age of 66.4 ± 9.9 years). In the internal validation test, performed among 30 models generated through the randomized dataset, the averages for accuracy, ROC AUC, sensitivity and specificity were, respectively, 90,97% (95% CI: 90,22%, 91,72%), 0,97 (95% CI: 0,966, 0,974), 89,36% (95% CI: 88,18%, 90,54%) e 92,33% (95% CI: 91,34%, 93,32%). The internal validation test for the best model reached an accuracy of 95%, an AUC of 0.98, sensitivity of 93.5%, and specificity of 96.3%. In the external validation test, the model achieved an accuracy of 93.3%, an AUC of 0.97, a sensitivity of 93.3%, and a specificity of 93.3%.

Conclusion: The neural network-based model proposed in this study using 3D radiomic features could differentiate benign from malignant vertebral compression fractures with excellent performance, overcoming the limitations of previous studies. The likely greatest contribution of this work would be to increase the diagnostic efficiency of the radiologist for these two clinical conditions, allowing an eventual reduction of both the number of biopsies and the time taken to deliver the final diagnosis.

Keywords: spine, artificial neural network, vertebral compression fractures, radiomic features, machine learning, MRI.

LISTA DE FIGURAS

Figura 1 - Exemplo de uma região da coluna vertebral sem e com FVC.....	23
Figura 2 – Características associadas às fraturas malignas.	26
Figura 3 – Características associadas às fraturas benignas.	26
Figura 4 - Representação de uma imagem digital na forma de matriz de intensidades de cinza.	29
Figura 5 - Diferença de intensidade de sinal para líquido e gordura em imagens ponderadas em T1 e T2.	31
Figura 6 - Representação da classificação de uma nova instância utilizando o kNN	36
Figura 7 - Exemplo de árvore de decisão para seis atributos e sete classes.....	37
Figura 8 - Perceptron Multicamadas.....	40
Figura 9 - Criação de hiperplano do classificador SVM para classificação	41
Figura 10 - Esquema de treinamento e predição para os algoritmos <i>Bagging</i> e <i>Boosting</i>	43
Figura 11 - Esquema ilustrando uma validação cruzada de 10 <i>folds</i>	44
Figura 12 - Matriz de confusão e métricas de avaliação	45
Figura 13 - Exemplos de curva ROC com os respectivos valores de AUC	45
Figura 14 - Interface do RIS para busca de laudo por patologia	46
Figura 15 - Fluxograma da seleção de pacientes com critérios de exclusão	49
Figura 16 - Representação de uma imagem antes e depois do processo de equalização.	51
Figura 17 - Exemplos de <i>slices</i> com e sem aparição dos pedículos	52
Figura 18 - Segmentação tridimensional realizada no 3D <i>Slicer</i>	52
Figura 19 - Esquema sumarizado dos experimentos realizados.....	57
Figura 20 - Processo da análise de hiperparâmetros do experimento 5	62
Figura 21 - Esquemas 1 e 2 utilizados na seleção de características do experimento 5... ..	64
Figura 22 - Esquema 3 utilizado na seleção de características do experimento 5.....	64
Figura 23 - Gráfico da AUC ROC dos atributos que obtiveram valor > 0.700	71
Figura 24 - Acurácia alcançada em cada um dos subexperimentos do experimento 2 usando validação cruzada de 10 <i>folds</i>	73
Figura 25 - AUC ROC alcançada em cada um dos subexperimentos do experimento 2 usando validação cruzada de 10 <i>folds</i>	74

Figura 26 - Sensibilidade alcançada em cada um dos subexperimentos do experimento 2 usando validação cruzada de 10 <i>folds</i>	75
Figura 27 - Especificidade alcançada em cada um dos subexperimentos do experimento 2 usando validação cruzada de 10 <i>folds</i>	76
Figura 28 - Acurácia alcançada em cada um dos subexperimentos do experimento 3 utilizando a validação cruzada de 10 <i>folds</i>	78
Figura 29 - AUC ROC alcançada em cada um dos subexperimentos do experimento 3 utilizando a validação cruzada de 10 <i>folds</i>	79
Figura 30 - Sensibilidade alcançada em cada um dos subexperimentos do experimento 3 utilizando a validação cruzada de 10 <i>folds</i>	80
Figura 31 - Especificidade alcançada em cada um dos subexperimentos do experimento 3 utilizando a validação cruzada de 10 <i>folds</i>	81
Figura 32 - Gráfico comparativo dos resultados da classificação do experimento 4	86
Figura 33 - Gráfico comparativo da acurácia para cada combinação de parâmetros	87
Figura 34 - Gráfico comparativo da AUC ROC para cada combinação de parâmetros.....	87
Figura 35 - Quantidade de características selecionadas após cada execução do <i>wrapper</i>	89
Figura 36 - Gráficos de dispersão dos dados para todos os atributos selecionados exibidos em pares	93
Figura 37 - Resultados do teste de validação interna considerando 30 modelos	94
Figura 38 - Falsos negativos resultantes do melhor modelo no teste de validação interna	95
Figura 39 - Resultado do teste de validação externa	95
Figura 40 - Arquitetura do modelo com melhor performance	96

LISTA DE TABELAS

Tabela 1 - Características demográficas e clínicas dos pacientes incluídos no estudo	66
Tabela 2 - Resultado dos testes de hipótese para diferenciação de fraturas benignas e malignas	68
Tabela 3 - Diferença entre as proporções de cada família de atributos comparando o grupo total de características com o grupo selecionado ($p < 0,001$)	70
Tabela 4 - Proporção de características selecionadas por família de atributos	85
Tabela 5 - Distribuição das famílias de atributos entre o grupo original e o grupo das selecionadas	85
Tabela 6 - Resultados da classificação de cada algoritmo com seus respectivos ver de características selecionadas pela técnica <i>wrapper</i>	86
Tabela 7 - Comparação da performance da MLP entre a configuração alternativa e <i>default</i>	88
Tabela 8 - Pesos do modelo final	96
Tabela 9 - Teste de Normalidade Anderson-Darling	120
Tabela 10 - AUC ROC com IC de 95% para as 93 características extraídas das imagens.	123

LISTA DE QUADROS

Quadro 1 - Exemplos de conclusão de laudos recuperados a partir da consulta no RIS ..	47
Quadro 2 - Algoritmos de classificação utilizados para cada família e seus respectivos parâmetros <i>default</i>	59
Quadro 3 - Parâmetros do método <i>wrapper</i> utilizados no experimento 4.....	61
Quadro 4 - Hiperparâmetros da rede neural estudados no experimento 5.....	61
Quadro 5 - Método <i>wrapper</i> do experimento 5 (esquema 1)	63
Quadro 6 - Método <i>wrapper</i> do experimento 5 (esquema 2)	63
Quadro 7 - Método <i>wrapper</i> do experimento 5 (esquema 3)	63
Quadro 8 - Resultado da seleção de hiperparâmetros dos algoritmos de classificação....	77
Quadro 9 - Resultado da seleção de características utilizando o método wrapper	83
Quadro 10 - Atributos selecionados pelo <i>wrapper</i> após o Esquema 1.....	89
Quadro 11 - Atributos selecionados pelo <i>wrapper</i> após o Esquema 2.....	90
Quadro 12 - Atributos das Seleção 2 ranqueados após a execução do filtro ReliefF	90
Quadro 13 - Vetor de características final resultante do esquema 3 do wrapper	91
Quadro 14 -Ranking decrescente das características após aplicação dos filtros de seleção	125

LISTA DE ABREVIATURAS E SIGLAS

AUC	Area Under the Curve
BAG	Bagging
BST	Boosting
DA	Discriminant Analysis
DICOM	Digital Imaging and Communications in Medicine
DT	Decision Tree
FVC	Fratura Vertebral por Compressão
GLCM	Grey Level Cooccurrence Matrix
GLDM	Grey Level Dependence Matrix
GLRLM	Grey Level Run Length Matrix
GLSZM	Grey Level Size Zone Matrix
IC	Intervalo de Confiança
KNN	K-Nearest Neighbors
LDA	Linear Discriminant Analysis
LR	Logistic Regression
MLP	Multilayer Perceptron
NB	Naïve Bayes
NGTDM	Neighboring Gray Tone Difference Matrix
PACS	Picture Archiving and Communication System
QDA	Quadratic Discriminant Analysis
RF	Random Forest
RIS	Radiology Information System
RM	Ressonância Magnética
ROC	Receiver Operating Characteristic
ROI	Region of Interest
SVM	Support Vector Machine
TC	Tomografia Computadorizada

SUMÁRIO

1	INTRODUÇÃO	19
1.1	CONTEXTUALIZAÇÃO	19
1.2	JUSTIFICATIVA.....	20
1.3	HIPÓTESE E OBJETIVO	21
1.4	ORGANIZAÇÃO DA TESE.....	22
2	FUNDAMENTAÇÃO TEÓRICA	23
2.1	FRATURAS VERTEBRAIS POR COMPRESSÃO.....	23
2.1.1	Diagnóstico	25
2.1.2	Tratamento	27
2.2	RECONHECIMENTO DE PADRÕES EM IMAGENS MÉDICAS	28
2.2.1	Imagem Digital	28
2.2.2	Etapas do reconhecimento de padrões em imagens	29
2.2.2.1	<i>Aquisição da imagem</i>	29
2.2.2.2	<i>Segmentação</i>	31
2.2.2.3	<i>Extração de características quantitativas</i>	32
2.2.2.4	<i>Seleção de características</i>	34
2.2.2.5	<i>Aprendizado de Máquina</i>	35
3	MATERIAL E MÉTODOS	46
3.1	ASPECTOS ÉTICOS.....	46
3.2	BASE DE DADOS	46
3.3	PIPELINE DO RECONHECIMENTO DE PADRÕES EM IMAGENS MÉDICAS.....	50
3.3.1	Pré-processamento	50
3.3.2	Segmentação	51
3.3.3	Extração de características quantitativas	53
3.3.4	Seleção de características	54
3.3.5	Análise	55
3.4	EXPERIMENTOS	56
3.4.1	Experimento 1 - Análise Univariada	58
3.4.2	Experimento 2 - PyRadiomics + Filtros	58
3.4.3	Experimento 3 - U-Net + Filtros	60
3.4.4	Experimento 4 - PyRadiomics + Wrapper	60
3.4.5	Experimento 5 - Rede Neural + Wrapper	61
4	RESULTADOS	66

4.1	ANÁLISE CLÍNICA E DEMOGRÁFICA	66
4.2	BASE DE DADOS PARA PESQUISA REPRODUTÍVEL	67
4.3	EXPERIMENTOS	68
4.3.1	Experimento 1 - Análise Univariada.....	68
4.3.2	Experimento 2 - PyRadiomics + Filtros.....	72
4.3.3	Experimento 3 - U-Net + Filtros.....	77
4.3.4	Experimento 4 - PyRadiomics + <i>Wrapper</i>	83
4.3.5	Experimento 5 - Rede Neural + <i>Wrapper</i>.....	87
5	DISCUSSÃO.....	99
5.1	ANÁLISE METODOLÓGICA.....	99
5.1.1	Pré-processamento das imagens.....	99
5.1.2	Técnicas de segmentação e extração de atributos.....	99
5.1.3	Seleção de características	100
5.1.4	Algoritmos de classificação	101
5.2	ASSINATURA RADIÔMICA E MODELO PROPOSTO.....	102
5.3	TRABALHOS RELACIONADOS	104
5.4	LIMITAÇÕES DO ESTUDO.....	106
6	CONCLUSÃO	108
7	CONSIDERAÇÕES FINAIS	110
	REFERÊNCIAS.....	112
	APÊNDICE A	120
	APÊNDICE B	123
	APÊNDICE C.....	125
	ANEXO A.....	128

1 INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO

As fraturas vertebrais por compressão (FVC) podem ter diferentes etiologias, podendo ser de origem traumática, secundárias à osteoporose ou à infiltração neoplásica. Quando há história recente de trauma, o radiologista não tem dificuldades em determinar a etiologia e pode focar sua avaliação na presença de possíveis fraturas concomitantes, lesões ligamentares e instabilidade. No entanto, nos casos de fraturas não-traumáticas, o diagnóstico pode tornar-se um desafio, principalmente na população idosa (CUÉNOD et al., 1996).

O estudo para diferenciação de FVCs benignas (secundárias à osteoporose) e malignas (secundárias à infiltração neoplásica) tem sido objeto de pesquisa por mais de 20 anos, tornando possível ampliar a compreensão dos padrões de imagem que poderiam ajudar a discriminar entre essas duas condições (CUÉNOD et al., 1996; LECOUVET et al., 1997; LEE et al., 2001; SUH et al., 2018). A ressonância magnética (RM) tem se estabelecido como a técnica de imagem médica mais relevante para o diagnóstico de doenças malignas da coluna vertebral, principalmente por causa de sua alta sensibilidade às anormalidades da medula óssea, permitindo identificar seu acometimento de forma precoce, antes de haver destruição do tecido ósseo trabecular do osso esponjoso (LECOUVET, 2016; PORTER; SHIELDS; OLSON, 1986).

Os critérios usados na rotina clínica para distinguir fraturas benignas de malignas estão relacionados à morfologia do corpo vertebral e às anormalidades de intensidade de sinal da medula óssea, principalmente observadas nos exames ponderados em T1 e T2 (MAUCH et al., 2018; THAWAIT et al., 2012). No entanto, há uma sobreposição substancial de padrões de imagem na ressonância magnética apresentados por esses dois grupos distintos de doenças que afetam a coluna vertebral, e nenhum dos critérios qualitativos disponíveis são patognomônicos quando usados isoladamente (LECOUVET et al., 1997; TEHRANZADEH; TAO, 2004).

Na última década, os exames por imagem deixaram de ser considerados simples ferramentas diagnósticas na clínica oncológica e passaram a assumir um papel mais relevante no âmbito da medicina personalizada (AERTS et al., 2014; LAMBIN et al., 2012). Nesse contexto, o uso de características quantitativas extraídas das imagens, técnica

denominada Radiômica (do Inglês, *Radiomics*), tem um grande potencial no apoio à tomada de decisão clínica para orientação do caso, pois pode capturar a heterogeneidade intratumoral, fornecer uma visão abrangente do tumor e monitorar continuamente a progressão da doença e sua resposta à terapia (AERTS et al., 2014).

Nesse ecossistema das FVCs, pretendemos investigar neste trabalho se é possível definir um modelo computacional com variáveis radiômicas tridimensionais que sejam capazes de diferenciar fraturas vertebrais por compressão benignas e malignas em imagens de ressonância magnética da coluna vertebral.

1.2 JUSTIFICATIVA

A diferenciação de fraturas vertebrais por compressão benignas e malinas é um problema comum na radiologia clínica, e estabelecer o diagnóstico correto é de grande importância para se determinar o tratamento e o prognóstico (BAUR et al., 2002).

Embora o exame de ressonância magnética seja o método estabelecido para se realizar o diagnóstico, as características observadas para cada uma das condições muitas vezes se sobrepõem. No caso do mieloma múltiplo, por exemplo, lesões mielomatosas podem estar presentes nos corpos vertebrais com sinal medular normal. Além disso, quando o mieloma múltiplo se distribui difusamente na medula óssea, pode dar o aspecto de osteoporose, potencialmente pela ativação difusa dos osteoclastos (MAUCH et al., 2018). De acordo com (LECOUVET et al., 1997), FVCs secundárias ao mieloma múltiplo podem aparentar serem benignas em 38% dos casos.

Em um estudo recente para avaliar a concordância e acurácia diagnóstica na diferenciação de FVCs benignas e malignas usando exames de RM, (ARANA et al., 2020) concluíram que a acurácia diagnóstica e a concordância interobservador é, no máximo, moderada, independentemente da especialidade médica ou cirúrgica, anos de experiência clínica ou tipo de hospital. Eles ainda afirmam que esse resultado põe em dúvida a confiabilidade do uso de achados de RM juntamente com a história clínica como base para distinguir esses dois tipos de fraturas na rotina clínica.

O material proveniente de biópsias também é uma importante referência para o diagnóstico diferencial entre lesões malignas e benignas, no entanto, por se tratar de uma técnica invasiva, não deve ser utilizada como método padrão na rotina clínica. Além disso, já existe o entendimento de que, devido à heterogeneidade espacial e temporal dos

tumores, outros tipos de descritores precisam ser utilizados para permitir uma caracterização mais completa da lesão (AERTS et al., 2014; NAPEL et al., 2018).

A radiômica é uma abordagem recente que tem trazido bons resultados no auxílio ao diagnóstico e na avaliação prognóstica de várias doenças, principalmente na área da oncologia (AERTS et al., 2014; LAMBIN et al., 2012; PARMAR et al., 2015). Ela pode ser definida como a prática de extrair características quantitativas de imagens médicas para gerar dados que possam ser analisados por meio de técnicas de mineração (do Inglês, *data mining*). A radiômica tem como objetivo encontrar padrões e correlações que possam associar as imagens a outros indicadores clínicos, auxiliando, assim, os especialistas nas tomadas de decisão (KUMAR et al., 2012). Essa prática contrasta com a prática tradicional de tratar as imagens médicas como figuras destinadas exclusivamente à interpretação visual e tem sido usada, principalmente, para melhorar a tomada de decisão clínica em pacientes com câncer (GILLIES; KINAHAN; HRICAK, 2016).

Alguns poucos trabalhos prévios (ADELA ARPITHA; RANGARAJAN, 2020; AZEVEDO-MARQUES et al., 2015; CASTI et al., 2017; FRIGHETTO-PEREIRA et al., 2015, 2016) que utilizaram características quantitativas extraídas das imagens de RM para diferenciar FVCs benignas de malignas mostraram resultados promissores. No entanto, todos eles apresentam uma característica exploratória, não se aprofundando nas análises ao ponto de propor uma assinatura radiômica capaz de realizar essa diferenciação com uma confiabilidade adequada. Além disso, até onde sabemos, esse é o primeiro trabalho que utilizou uma abordagem tridimensional associada ao estudo radiômico para propor um modelo capaz de diferenciar FVC benignas e malignas em imagens de ressonância magnética.

1.3 HIPÓTESE E OBJETIVO

A pergunta que pretendemos responder com esse trabalho é: É possível definir um modelo computacional com variáveis radiômicas extraídas de imagens tridimensionais de ressonância magnética que seja capaz de diferenciar fraturas vertebrais por compressão benignas de malignas?

A nossa hipótese é a de que é possível encontrar uma assinatura radiômica tridimensional que, inserida em um modelo computacional, seja capaz de diferenciar fraturas vertebrais por compressão benignas e malignas.

Nesse sentido, o objetivo principal desse trabalho foi desenvolver um modelo computacional para diferenciação de fraturas vertebrais por compressão benignas e malignas a partir de atributos radiômicos tridimensionais extraídos de exames de ressonância magnética da coluna vertebral.

Os objetivos específicos desse trabalho foram:

- Organizar uma base com dados retrospectivos de pacientes diagnosticados com fratura vertebral por compressão seguindo o modelo recomendado para reprodutibilidade em estudos radiômicos;
- Realizar uma análise exploratória para investigar os métodos mais viáveis de reconhecimento de padrão em imagens médicas para o nosso problema;
- Desenvolver o modelo a partir das abordagens escolhidas;
- Avaliar a performance do modelo em conjunto com as variáveis radiômicas e compará-lo com resultados anteriormente publicados.

1.4 ORGANIZAÇÃO DA TESE

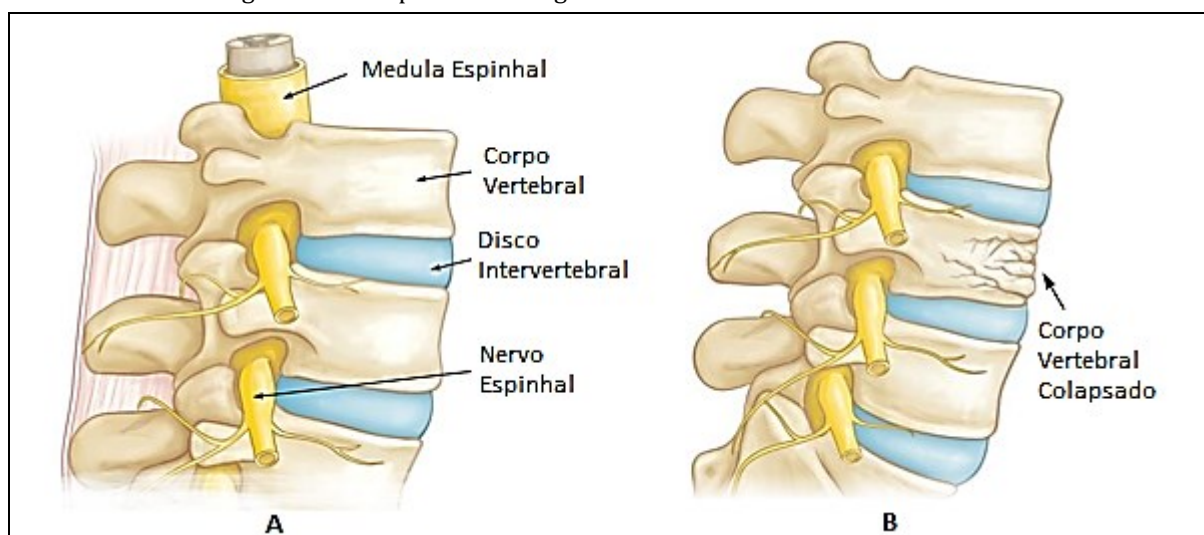
A tese está organizada da seguinte maneira: O Capítulo 2 apresenta a fundamentação teórica necessária para a compreensão do trabalho, contemplando duas seções principais. A primeira traz informações sobre as fraturas vertebrais por compressão, apresentando suas causas, diagnóstico e tratamento; a segunda aborda o reconhecimento de padrão em imagens médicas e detalha cada uma de suas etapas. O Capítulo 3 apresenta o material e os métodos utilizados no desenvolvimento do trabalho, incluindo a fonte de dados, o suporte ferramental, a abordagem metodológica para cada uma das etapas do reconhecimento de padrões e o *design* experimental. O Capítulo 4 apresenta os resultados produzidos no presente trabalho. O Capítulo 5 traz uma discussão sobre diferentes aspectos do estudo, desde a escolha das abordagens metodológicas até a interpretação dos resultados e sua comparação com trabalhos da literatura. O Capítulo 6 apresenta a conclusão do trabalho destacando as principais contribuições e, por fim, o Capítulo 7 contempla as considerações finais, incluindo produções geradas direta e indiretamente durante o doutorado, orientações realizadas e demais atividades exercidas que contribuíram para a evolução da autora enquanto pesquisadora.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 FRATURAS VERTEBRAIS POR COMPRESSÃO

As fraturas vertebrais por compressão ocorrem quando o corpo vertebral colapsa, o que pode causar dor intensa (levando à incapacidade de realizar as atividades da vida diária), deformidade e perda de altura (OLD; CALVERT, 2004). A Figura 1A ilustra a anatomia de um corpo vertebral normal e a Figura 1B de um corpo vertebral com fratura por compressão.

Figura 1 - Exemplo de uma região da coluna vertebral sem e com FVC



Nota. – A) Parte de uma coluna vertebral intacta, mostrando a medula espinhal, o corpo vertebral sem fratura, o disco intervertebral e o nervo espinhal. B) Exemplo de uma fratura vertebral por compressão (seta), que apresenta a forma de uma cunha, com a parte frontal rachada e encurtada enquanto a parte de trás permanece intacta. *Fonte:* Adaptado de (PARK; JENIS, 2021)

A Fratura Vertebral por Compressão é uma condição comum, principalmente em idosos. Um estudo de (CLARK et al., 2009) que avaliou a prevalência de fraturas vertebrais na população feminina em países da América do Sul mostrou que, no Brasil, a prevalência dessas fraturas em mulheres acima de 50 anos é de 14,2%. No caso de mulheres acima de 80 anos, essa prevalência aumenta para 25%.

Além das limitações físicas, as fraturas por compressão vertebral podem produzir uma carga psicossocial e emocional no idoso que já enfrenta perdas de função independente. Isso pode fazer com que a pessoa se preocupe demasiadamente, uma vez que há um risco substancial do paciente sofrer novas fraturas subsequentes de todos os tipos e morbidade adicional (MELTON III et al., 1999).

As fraturas vertebrais por compressão estão intrinsecamente associadas à osteoporose e muitos dos fatores de risco são os mesmos, podendo ser categorizados em fatores não modificáveis e potencialmente modificáveis. Os fatores de risco não modificáveis incluem idade avançada, sexo feminino, raça branca, presença de demência, suscetibilidade a quedas, história de fraturas na idade adulta e história de fraturas em parente de primeiro grau. Fatores de risco potencialmente modificáveis incluem estar em situação de abuso físico, uso de álcool e/ou tabaco, presença de osteoporose e/ou deficiência de estrogênio, menopausa precoce ou ovariectomia bilateral, amenorreia na pré-menopausa há mais de um ano, fragilidade, deficiência visual, atividade física insuficiente, baixo peso corporal e deficiência de cálcio e/ou vitamina D na dieta (OLD; CALVERT, 2004).

Embora a etiologia mais comum seja a osteoporose, outros fatores como trauma e neoplasias (primárias e secundárias) também podem levar às fraturas vertebrais por compressão (ALEXANDRU, 2012).

Com relação à neoplasia secundária, ou metástase, ela pode ser entendida como espalhamento do câncer a partir do lugar de onde ele começou (chamado de sítio primário) para outra parte do corpo (MACCAURO et al., 2011). Aproximadamente 70% dos pacientes com câncer desenvolvem metástases (KLIMO, 2004). O osso é o terceiro órgão mais afetado pelo câncer metastático (COLEMAN, 2001; MACEDO et al., 2017; SELVAGGI; SCAGLIOTTI, 2005) sendo a coluna vertebral o sítio mais comum das metástases ósseas (JACOBS; PERRIN, 2001).

Um tipo de câncer que é considerado de origem primária por alguns autores por se tratar de um câncer da medula óssea é o mieloma múltiplo. Trata-se de uma neoplasia maligna das células plasmáticas presentes na medula óssea, representando cerca de 1% das neoplasias totais e 10% das malignidades hematológicas (MOREAU et al., 2017; RAJKUMAR; KUMAR, 2016). No entanto, neste tipo de câncer, a destruição óssea é a principal característica clínica em mais de 80% dos pacientes (SELVAGGI; SCAGLIOTTI, 2005).

Embora a observação mais comum seja de pacientes que evoluíram para metástases ósseas após terem recebido o diagnóstico do câncer do sítio primário, em 25-30% dos casos a primeira manifestação da malignidade pode ser a lesão no osso, que frequentemente vem acompanhada de fratura (KATAGIRI et al., 1999; PICCIOLI et al., 2015). É nesse ponto que o diagnóstico diferencial das fraturas benignas e malignas feito

pelo radiologista torna-se essencial, uma vez que cada uma dessas doenças (osteoporose e neoplasias) apresentam prognósticos e tratamentos completamente diferentes.

2.1.1 Diagnóstico

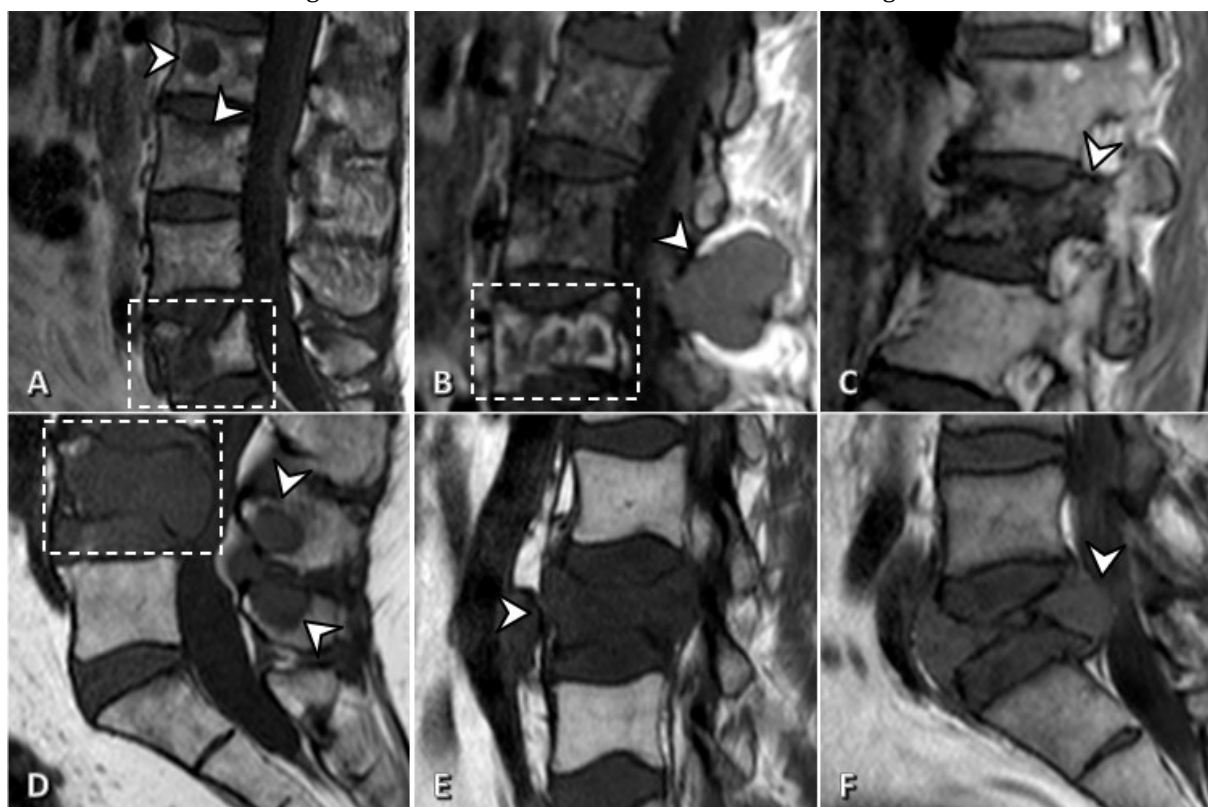
Como mencionado anteriormente, o exame de ressonância magnética, principalmente nas ponderações T1 e T2, tem tido papel fundamental na rotina clínica dos radiologistas para a realização do diagnóstico diferencial das fraturas vertebrais por compressão (THAWAIT et al., 2012).

Algumas características comumente associadas às fraturas malignas são: i) metástases em outras vértebras; ii) presença de uma massa de tecido mole epidural ou paravertebral; iii) sinal anormal da medula envolvendo o pedículo ou outros elementos posteriores; iv) substituição total do sinal de gordura da medula óssea por tecido de baixo sinal na sequência ponderada em T1 e v) uma protuberância convexa da parede posterior do corpo vertebral (MAUCH et al., 2018; THAWAIT et al., 2012). Essas características podem ser observadas na Figura 2.

No caso das fraturas benignas, algumas características frequentemente associadas a elas são: i) coexistência de outras fraturas benignas; ii) pedículo ou elemento vertebral posterior com sinal normal; iii) convexidade posterior focal (retropulsão) e iv) uma banda horizontal de baixo sinal nas ponderações T1 e T2 (MAUCH et al., 2018; THAWAIT et al., 2012). Essas características são mostradas na Figura 3.

Esses critérios usados atualmente por radiologistas são baseados em análises subjetivas e qualitativas, sujeitos à variabilidade interpessoal e dependentes da experiência e do estado de atenção. Quando o diagnóstico não é fechado por meio das análises clínicas e de imagem, eventualmente é solicitado que o paciente realize uma biópsia para que o diagnóstico diferencial seja definido.

Figura 2 – Características associadas às fraturas malignas.



Nota.— A) presença de outras metástases; B) presença de massa paravertebral; C) envolvimento do pedículo; D) envolvimento de outros elementos posteriores; E) hipossinal difuso e F) convexidade posterior difusa. *Fonte:* CCIFM/HCFMRP-USP.

Figura 3 – Características associadas às fraturas benignas.



Nota.— A) não-envolvimento do pedículo; B) Retropulsão pósterio-superior com aspecto angulado e C) hipossinal em banda horizontal *Fonte:* CCIFM/HCFMRP-USP

2.1.2 Tratamento

O tratamento bem-sucedido de fraturas vertebrais por compressão tem o potencial de melhorar a qualidade de vida, aumentar a expectativa de vida independente e/ou produtiva e prevenir complicações médicas sobrepostas (SHAH et al., 2018).

Como passo inicial, no caso dos pacientes diagnosticados com osteoporose, é fundamental que seja feito o tratamento para essa doença a fim de prevenir novas fraturas bem como estabilizar as já existentes. Um exemplo seria a administração de medicamentos para melhorar a densidade óssea, como é o caso dos bifosfonatos (OLD; CALVERT, 2004; SHAH et al., 2018).

As opções de tratamento para das FVCs podem ser divididas em não cirúrgicas (tratamento conservador) e cirúrgicas. No caso do tratamento conservador, repouso por um curto período pode ser recomendado, no entanto a inatividade prolongada deve ser evitada, especialmente em pacientes idosos. (BABB; CARLSON, 2006). A administração de analgésicos para controle da dor e a realização de fisioterapia também são opções do tratamento conservador (ALEXANDRU, 2012; SHAH et al., 2018). Além disso, um aspecto importante que muitas vezes é subestimado é o uso de órteses para diminuir a dor e auxiliar na cicatrização de fraturas por compressão vertebral. A órtese pode ser utilizada imediatamente, mas só deve ser continuada por dois a três meses (BABB; CARLSON, 2006). Já no caso de pacientes com fraturas malignas, sessões de radioterapia podem ser indicadas se o tumor for radiosensível. Estudos mostram que a radioterapia proporcionou alívio da dor em aproximadamente 50% dos pacientes com FVC secundário ao mieloma, câncer de próstata ou mama (OLD; CALVERT, 2004).

A abordagem cirúrgica pode ser indicada quando o paciente não responde ao tratamento conservador, quando há evidência de déficit neurológico iminente ou existente, ou quando a deformidade da coluna é extremamente grave (ALEXANDRU, 2012; OLD; CALVERT, 2004; SHAH et al., 2018). De maneira geral, o que se faz é injetar cimento ortopédico chamado polimetilmetacrilato (PMMA) na vertebra colapsada para estabilizar e fortalecer o corpo vertebral. As duas abordagens mais utilizadas são a vertebroplastia e a cifoplastia. No caso da vertebroplastia, o procedimento não restaura a forma ou a altura da vértebra comprimida, uma vez que o produto é injetado diretamente na vértebra acometida, preenchendo as lacunas dentro osso. Já na cifoplastia, o cimento é injetado em uma cavidade criada por um balão ou um dilatador mecânico, refazendo,

eventualmente, o formato da vértebra antes da doença (SHAH et al., 2018). Em ambos os casos, após a secagem, esse cimento fornecerá suporte e estabilidade para o osso, reduzindo consideravelmente a dor devido à estabilidade proporcionada. Além disso, ambos os procedimentos podem ser indicados como etapa prévia nos tratamentos onde é necessária a fixação de hastes e parafusos (GU et al., 2015).

2.2 RECONHECIMENTO DE PADRÕES EM IMAGENS MÉDICAS

A técnica de Reconhecimento de Padrões em Imagens é a base computacional para solução de diversos problemas nas mais diversas áreas do conhecimento, como, por exemplo, na geografia, astronomia, arqueologia, física e medicina (GONZALEZ; WOODS, 1992). Particularmente na área da radiologia, ela vem sendo amplamente utilizada para construção de modelos computacionais para auxílio ao diagnóstico, análise prognóstica e previsão de resposta à terapia (KOENIGKAM-SANTOS et al., 2019).

Nesta seção, apresentaremos os principais elementos presentes nessa área. No entanto, por se tratar de um assunto abrangente e complexo, focaremos nossa explicação nas técnicas utilizadas especificamente para esse trabalho. Iniciaremos as explicações com o conceito de imagem digital e, em seguida, apresentaremos os processos e etapas que constituem a abordagem tradicional do reconhecimento de padrões em imagens usando *machine learning*.

2.2.1 Imagem Digital

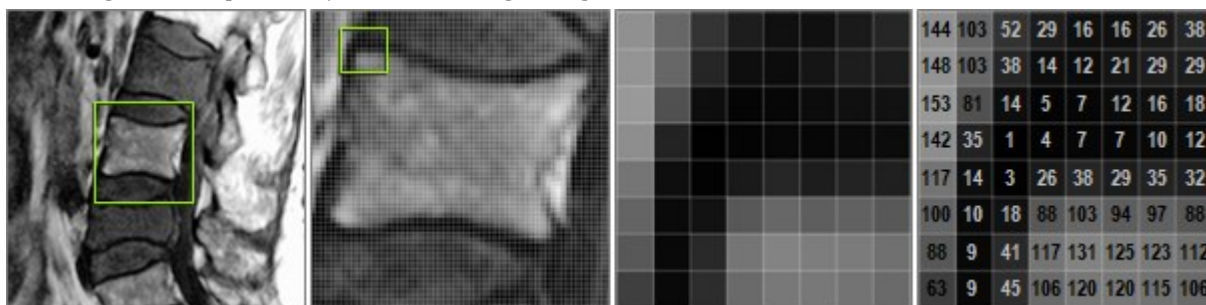
Uma imagem digital bidimensional pode ser representada como uma função $f(x, y)$, onde x e y denotam as coordenadas espaciais e o valor de f no ponto (x, y) corresponde ao brilho (nível de cinza) da imagem neste ponto (GONZALEZ; WOODS, 1992).

Outra forma de representar uma imagem digital é por meio de uma matriz, onde os índices das linhas e colunas identificam os pontos dessa imagem e os valores dos elementos descrevem o nível de cinza da imagem nesses pontos. A esses pontos damos o nome de *pixel*, que pode ser entendido ainda como a menor unidade pictórica de uma imagem digital.

A quantidade de intensidades de cinza de uma imagem irá depender do número de bits necessários para armazenar cada pixel da imagem, e é expressa pela relação 2^n , onde

n é o número de bits. Uma imagem de 8 bits por pixel, por exemplo, apresentará um máximo de 256 níveis de cinza, onde 0 corresponde à cor preta e 255 à cor branca. A Figura 4 representa uma imagem digital na forma de matriz.

Figura 4 - Representação de uma imagem digital na forma de matriz de intensidades de cinza.



Fonte: da Autora.

2.2.2 Etapas do reconhecimento de padrões em imagens

O processo tradicional de reconhecimento de padrões em imagens consiste das seguintes etapas: aquisição da imagem, pré-processamento, segmentação, extração de características quantitativas, seleção de características e análise (aprendizado de máquina) (BISHOP, 2006; GONZALEZ; WOODS, 1992).

2.2.2.1 Aquisição da imagem

A aquisição da imagem é a primeira etapa do processo de reconhecimento de padrões em imagens. Nos casos das imagens médicas, essa aquisição se dá de maneiras diferentes, a depender da técnica utilizada. No entanto, todas elas requerem alguma forma de energia para produzir a imagem, energia esta que deve ser capaz de penetrar os tecidos do corpo humano.

Existem várias modalidades de imageamento médico, cada qual com sua própria aplicabilidade na medicina. A primeira delas foi a radiografia, que foi criada pelo físico Wilhelm Roentgen, quando este descobriu os raios-x em 8 de Novembro de 1895 (BUSHBERG; BOONE, 2011).

Atualmente, as modalidades mais comuns são: Radiografia, Mamografia, Tomografia Computadorizada (do Inglês, *Computed Tomography - CT*), Tomografia Computadorizada por Emissão de Fóton Único (do Inglês, *Single Photon Emission Computed Tomography -*

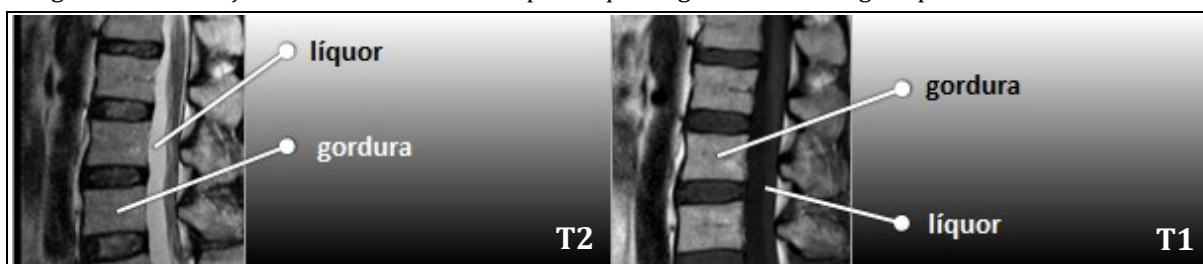
SPECT), Tomografia por Emissão de Póstron (do Inglês, *Positron Emission Tomography - PET*), Ultrassonografia e Ressonância Magnética (do Inglês, *Magnetic Resonance - MR*) (BUSHBERG; BOONE, 2011).

No caso da ressonância magnética, que é a modalidade que estamos utilizando neste trabalho, o processo físico de aquisição das imagens é complexo e abrangente, envolvendo conceitos de eletromagnetismo, supercondutividade e processamento de sinais (BUSHBERG; BOONE, 2011; MAZZOLA, 2009).

Em linhas gerais, a imagem padrão da RM é resultado da interação dos prótons dos átomos de hidrogênio presentes nos tecidos do corpo humano com um forte campo magnético gerado pela máquina de ressonância e com pulsos de radiofrequência que são emitidos localmente de acordo com a região a ser analisada. Cada tecido do corpo absorve essa onda de radiofrequência e a retransmite com tempos e intensidades diferentes. Um receptor acoplado à máquina faz, então, a leitura do sinal retransmitido pelo paciente e um computador, também acoplado à máquina, faz o mapeamento do sinal recebido para os níveis de cinza da imagem (BUSHBERG; BOONE, 2011).

Variações no pulso de radiofrequência e na interpretação do sinal recebido pelo receptor permitem a criação de imagens com diferentes tipos de contrastes para um mesmo tecido (por exemplo, um tecido pode ter baixo brilho em uma determinada sequência de RM e alto brilho em outra). Isso faz da ressonância magnética um dos exames mais sensíveis e detalhados, capaz de diferenciar uma grande quantidade de tecidos do corpo humano. Neste trabalho estamos utilizando a sequência de RM ponderada em T1, onde T1 se refere ao tempo de relaxação da componente magnética dos prótons de hidrogênio presentes nos tecidos. Nesta sequência, enquanto a gordura, que é um dos principais tecidos que compõem o osso esponjoso, produz um hipersinal, o líquido apresenta hipossinal. Na ponderação T2, em contrapartida, o líquido apresenta um hipersinal ao passo que a gordura apresenta uma intensidade de sinal cinza-médio, como mostrado na Figura 5. Essa diferença de sinal para um mesmo tecido utilizando sequências diferentes permite que o radiologista possa diferenciar estruturas e lesões com maior precisão.

Figura 5 - Diferença de intensidade de sinal para líquido e gordura em imagens ponderadas em T1 e T2.



Fonte: da Autora.

2.2.2.2 Segmentação

A segmentação pode ser definida como a divisão de uma imagem em partes ou objetos constituintes. O nível dessa subdivisão irá depender do problema a ser resolvido. O que se espera com a segmentação é isolar a região de interesse (do Inglês, *region of interest - ROI*) do restante da imagem (GONZALEZ; WOODS, 1992).

Algoritmos de segmentação para imagens monocromática geralmente são baseados em duas propriedades básicas dos valores de níveis de cinza: descontinuidade e similaridade. Na primeira categoria, o algoritmo particiona a imagem quando ele identifica uma alteração abrupta no nível de cinza. Algumas aplicações que precisam segmentar pontos isolados ou detectar bordas costumam utilizar esta abordagem.

Na categoria de similaridade, destacam-se as abordagens de limiarização e crescimento de região (GONZALEZ; WOODS, 1992). A limiarização é uma das abordagens mais importantes e mais simples de segmentação. Geralmente ela é utilizada quando se pretende separar objetos brilhantes de um fundo escuro, por exemplo. A função de limiarização para segmentar uma imagem de entrada f tendo como saída uma imagem g , pode ser definida de acordo com a Equação (1):

$$g(x, y) = \begin{cases} f(x, y) & \text{se } f(x, y) > T \\ 0 & \text{se } f(x, y) \leq T \end{cases} \quad (1)$$

onde T é o valor de limiar e $f(x, y)$ é o nível de cinza no ponto (x, y) .

A abordagem de crescimento de região, por sua vez, agrupa pixels ou sub-regiões em regiões maiores. Uma das técnicas mais simples dessa abordagem é conhecida como agregação de pixel, que começa com a marcação de “pontos sementes” e, a partir deles, agrega a cada semente novos pixels vizinhos que possuam propriedades similares (de

nível de cinza ou textura). Um dos problemas dessa abordagem é que a segmentação irá depender da marcação adequada das sementes iniciais, que devem ser bem representativas da região de interesse (GONZALEZ; WOODS, 1992).

Um outro método que vem ganhando espaço em aplicações na radiologia é a segmentação baseada em *deep learning*. A rede neural convolucional U-Net proposta por (RONNEBERGER; FISCHER; BROX, 2015) é considerada, atualmente, o estado da arte em segmentação de imagens biomédicas. Essa rede, assim como outras redes neurais convolucionais que realizam segmentação semântica, classifica cada pixel da imagem com um rótulo específico (que poderia ser, por exemplo, ROI ou *background*). No entanto, uma vantagem da U-Net é que ela produz segmentações precisas mesmo com poucas imagens no conjunto de treinamento. Além disso, ela responde bem às técnicas de *data augmentation*, bem como à aplicação de transformadas de rotação, translação e espelhamento das imagens (RONNEBERGER; FISCHER; BROX, 2015).

2.2.2.3 Extração de características quantitativas

Depois de a imagem ter sido segmentada, a ROI deve ser representada e descrita quantitativamente para, em seguida, ser processada computacionalmente. Esses descritores contêm informações sobre três aspectos da imagem: forma, intensidade de cinza e textura.

Descritores de forma trazem informações sobre as características externas (bordas) do objeto representado, como por exemplo, tamanho, orientação da linha que une os pontos extremos da ROI ou o número de concavidades do contorno. Já os descritores de intensidade de cinza e textura representam características internas do objeto e, portanto, não devem ser sensíveis a variações de tamanho, translação e rotação (GONZALEZ; WOODS, 1992).

Atributos de intensidade de tons de cinza (também chamados de atributos de primeira ordem) descrevem propriedades relacionadas ao histograma da imagem, que pode ser definido como um gráfico de distribuição das frequências de níveis de cinza da imagem. A partir desse gráfico, é possível calcular diversas medidas estatísticas como energia, entropia, média, variância, curtose, assimetria, entre outras (GONZALEZ; WOODS, 1992; VELHO; FRERY; GOMES, 2009).

Com relação à textura, apesar desta não possuir uma definição formal, seus descritores medem propriedades como suavidade, aspereza ou rugosidade da imagem ou região. Uma das abordagens mais utilizadas para extração de características de textura utiliza medidas estatísticas que são calculadas a partir de matrizes geradas com informações sobre o relacionamento entre os pixels da imagem (GONZALEZ; WOODS, 1992; NIOCHE et al., 2018). Algumas que podemos destacar são:

- **Matriz de coocorrência de níveis de cinza** (do inglês, *Grey Level Cooccurrence Matrix - GLCM*): descrita inicialmente em (HARALICK, 1979), leva em consideração os arranjos de pares de pixels para calcular os índices de textura. Numa matriz de coocorrência, um elemento (i, j) dessa matriz representa o número de vezes que a combinação dos níveis de cinza i e j ocorre em dois pixels na imagem, separados por uma distância de δ pixels ao longo do ângulo θ . Alguns atributos que podemos calcular a partir dessa matriz são, por exemplo, homogeneidade, entropia, contraste e correlação;
- **Matriz de dependência de níveis de cinza** (do inglês, *Grey Level Dependence Matrix - GLDM*): Uma matriz de dependência de nível de cinza quantifica dependências de nível de cinza em uma imagem. Uma dependência no nível de cinza é definida como o número de pixels conectados à distância δ que são dependentes do pixel central. Um pixel vizinho com nível de cinza j é considerado dependente do pixel central com nível de cinza i se $|i - j| \leq \alpha$. Em uma matriz de dependência de nível de cinza $P(i, j)$, o elemento (i, j) é o número de vezes que um pixel com nível de cinza i com j pixels dependentes aparece na imagem.
- **Matriz de comprimento de cadeia de níveis de cinza** (do inglês, *Grey Level Run Length Matrix - GLRLM*): fornece o tamanho de cadeias homogêneas para cada nível de cinza. O tamanho de uma cadeia homogênea é definido como o comprimento (em número de pixels) de pixels consecutivos que possuem o mesmo valor de nível de cinza. Nesse tipo de matriz, o elemento (i, j) descreve o número de cadeias com nível de cinza i e comprimento j que ocorrem na imagem ao longo de um ângulo θ .

- **Matriz de tamanho de zona de níveis de cinza** (do inglês, *Grey Level Size Zone Matrix - GLSZM*): fornece informações sobre o tamanho das zonas homogêneas para cada nível de cinza. O tamanho de uma zona homogênea é definido como o número de pixels conectados que compartilham a mesma intensidade de nível de cinza. Um elemento (i, j) dessa matriz corresponde ao número de zonas com nível de cinza i e tamanho j que ocorrem na imagem.
- **Matriz de diferença de tons de cinza vizinhos** (do inglês, *Neighboring Gray Tone Difference Matrix - NGTDM*): quantifica a diferença entre um valor cinza e a média do valor de cinza de seus vizinhos a uma distância δ . A soma das diferenças absolutas para o nível de cinza i é armazenada na matriz.

2.2.2.4 Seleção de características

A seleção de características é uma etapa fundamental do reconhecimento de padrões em imagens, uma vez que em algumas aplicações são extraídos centenas de atributos da imagem. Portanto, técnicas de redução de dimensionalidade precisam ser utilizadas para filtrar esses atributos de modo que apenas permaneçam no vetor de características os atributos que são relevantes para o problema a ser resolvido (KUMAR, 2014; TANG; ALELYANI; LIU, 2014).

Os métodos de seleção de características podem retornar dois tipos de saída: no primeiro, os atributos são ranqueados de acordo com algum critério de avaliação e, após essa etapa, os atributos que não obtiveram uma boa colocação são eliminados do conjunto; no segundo tipo, o método já retorna o subconjunto de atributos que obteve a melhor avaliação.

Os métodos de seleção de características podem ser categorizados em filtros, *wrapper* e embarcados. Os métodos baseados em filtros são independentes de qualquer algoritmo de aprendizado. Eles utilizam as características dos dados para avaliar a importância do atributo. Esses métodos geralmente são mais eficientes computacionalmente do que os métodos *wrapper*. No entanto, devido à falta de um algoritmo de aprendizado específico orientando a fase de seleção, os atributos selecionados podem não ser ideais para os algoritmos de aprendizado de destino (LI et al., 2018). Existem dezenas de filtros que podem ser utilizados para selecionar

características, porém a escolha vai depender de aspectos como o tipo de aprendizado (supervisionado binário, supervisionado multiclases ou não supervisionado), o tipo de saída (*ranking* ou subconjunto), o tipo de atributo (contínuo, discreto ou categórico) e o tipo de dados (estruturado, não estruturado, semiestruturado, dados linkados, entre outros).

A abordagem conhecida como *wrapper*, por sua vez, consiste em gerar subconjuntos de atributos a partir do vetor de características original e avaliar o desempenho de cada subconjunto tendo como base um classificador específico. Ao final, o subconjunto com menor erro na classificação será o escolhido (TANG; ALELYANI; LIU, 2014). Para criação dos subconjuntos de atributos é necessário utilizar algum método de busca. Um exemplo são os algoritmos da família *Sequential Forward Selection – SFS*. Nesse método, o subconjunto de atributos (que inicialmente está vazio) vai incorporando os atributos um a um até que seja satisfeito algum critério de parada pré-estabelecido. A cada inclusão, o modelo de classificação é executado e, então, avaliado. Ao final, o subconjunto que contribuiu para o melhor desempenho do classificador é selecionado.

Finalmente, os métodos embarcados executam uma seleção de características dentro do próprio classificador, avaliando o subconjunto selecionado com base nas métricas da própria classificação. Um exemplo de classificador que utiliza esse método é o *Random Forest*, que será detalhado nas próximas seções.

2.2.2.5 *Aprendizado de Máquina*

O Aprendizado de Máquina (do inglês, *Machine Learning*) é uma área da Inteligência Artificial na qual o computador é capaz de tomar decisões baseadas em exemplos, sem a necessidade de ser explicitamente programado para resolver o problema.

Existem duas formas principais de aprendizado: o supervisionado e o não supervisionado. Em um modelo de aprendizado supervisionado, o algoritmo aprende em um conjunto de dados rotulado, fornecendo uma chave de resposta que o algoritmo pode usar para avaliar sua precisão nos dados de treinamento. Um modelo não supervisionado, por outro lado, fornece dados não rotulados que o algoritmo tenta entender extraindo recursos e padrões por conta própria (BISHOP, 2006; JAMES et al., 2013).

Problemas que visam separar o conjunto de dados em diferentes grupos (ou classes) previamente rotulados são chamados de problemas de classificação. Tais problemas

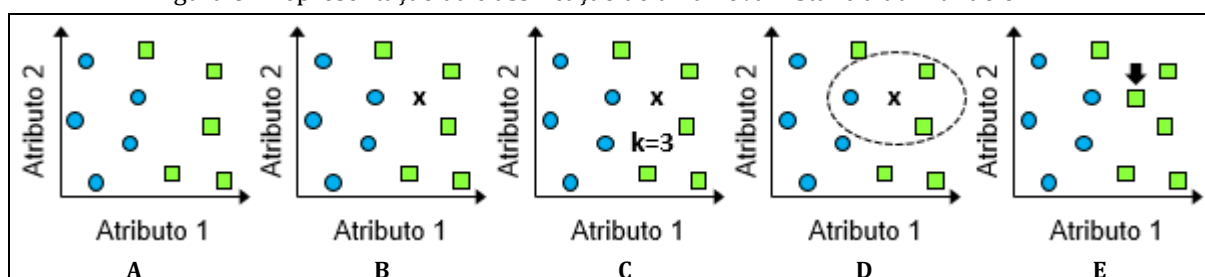
podem ser solucionados utilizando diferentes métodos (LOOG, 2018). Neste trabalho, as famílias de algoritmos utilizadas foram: regra dos vizinhos mais próximos, árvores de decisão, análise discriminante, classificadores bayesianos, rede neural artificial, regressão logística, máquinas de vetores de suporte, *random forest*, *bagging* e *boosting*. Abaixo serão descritos cada um deles.

a) Regra dos vizinhos mais próximos

A regra do vizinho mais próximo (do inglês, *k nearest neighbors* – *kNN*) (COVER; HART, 1967) é possivelmente o classificador com mais apelo intuitivo.

Segundo a regra do kNN, para classificar um objeto novo basta determinar as distâncias entre seu vetor de características descritivo e os vetores de características no conjunto de treinamento e atribuir o objeto à mesma classe dos elementos que possuem as características mais próximas. Na maioria das vezes, a distância euclidiana é usada para determinar o vizinho mais próximo no conjunto de dados de treinamento, mas, em princípio, qualquer outra medida de distância pode ser empregada (COVER; HART, 1967). A Figura 6 exemplifica a classificação de um novo elemento utilizando a abordagem kNN.

Figura 6 - Representação da classificação de uma nova instância utilizando o kNN



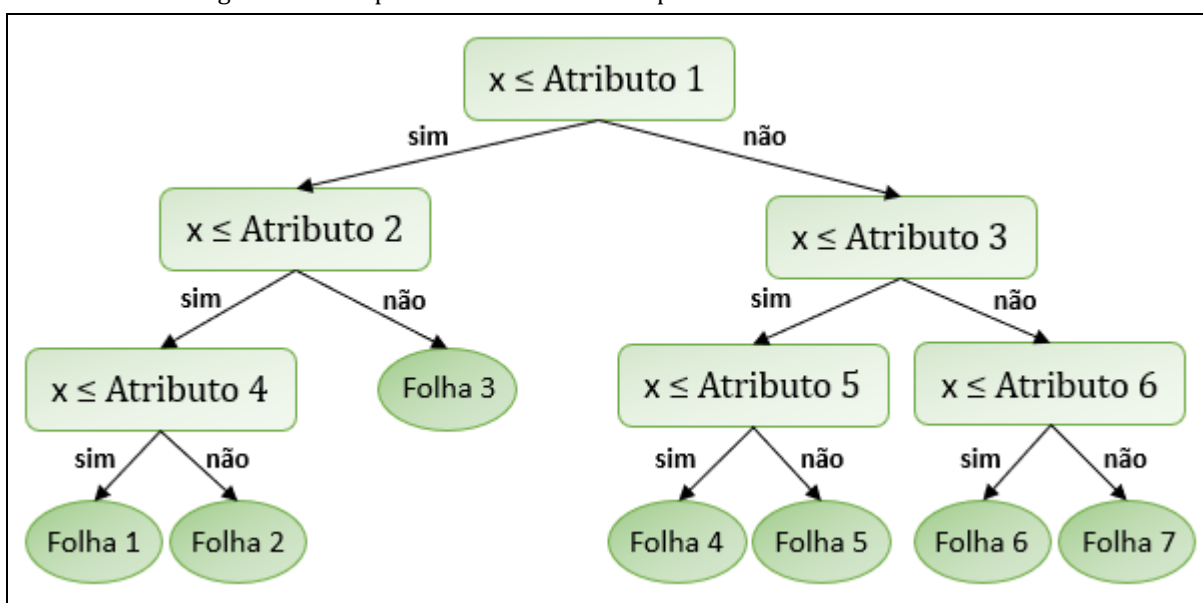
Fonte: da Autora.

A Figura 6A representa o conjunto de treinamento e suas classes (círculo e quadrado) para um par arbitrário de atributos correspondendo aos eixos x e y. Quando um novo elemento é inserido (Figura 6B), de acordo com a quantidade de vizinhos especificada (Figura 6C) ele irá selecionar os mais próximos (Figura 6D). A classificação final será a mesma da maioria dos vizinhos selecionados (Figura 6E).

b) Árvores de Decisão

No algoritmo de árvore de decisão (do Inglês, *Decision Tree*) existem dois elementos principais: a folhas, que indicam as classes do problema, e os nós de decisão, que especificam algum teste que deve ser aplicado para um único valor de atributo, de modo que para cada possível resultado do teste, uma ramificação é criada (QUINLAN, 2014). A classificação de um novo caso se inicia pelo nó raiz (que também é um nó de decisão) e vai percorrendo a estrutura da árvore de acordo com os resultados de cada teste até encontrar um nó folha, cuja classe é atribuída ao caso em questão. Uma representação genérica de uma árvore de decisão pode ser visualizada na Figura 7

Figura 7 - Exemplo de árvore de decisão para seis atributos e sete classes



Fonte: da Autora

A árvore de decisão pode ser utilizada tanto para atributos categóricos quanto para atributos numéricos e uma vantagem particular desse tipo de classificador é que ele retorna uma representação gráfica que pode ser muito útil para interpretação dos resultados.

c) Regressão Logística

Na Regressão Logística, ao invés de estimar a classe diretamente, o método estima a probabilidade de ser uma determinada classe, ou seja, é um classificador probabilístico. Ela prevê a probabilidade da ocorrência de um evento por meio do ajuste dos dados a uma função logística e é utilizada para problemas cujas saídas são variáveis categóricas binárias (CESSIE; HOUWELINGEN, 1992). Utilizando uma explicação um pouco mais formal, dado que um conjunto Y seja composto pelas classes binárias 0 e 1, a probabilidade para uma nova instância pertencer à classe 1, por exemplo, pode ser denotada por $\mathbb{P}(Y = 1|X = x)$, onde X é o conjunto de atributos e x são os valores dos atributos para a respectiva instância. Utilizando a função logística, essa probabilidade pode ser expressa por:

$$\mathbb{P}(Y = 1|X = x) = \frac{e^{(\beta_0 + \sum_{i=1}^d \beta_i x_i)}}{1 + e^{(\beta_0 + \sum_{i=1}^d \beta_i x_i)}} \quad (2)$$

onde d é o número de atributos no conjunto X e os valores de β são os parâmetros de ajuste da função. Durante o processo de treinamento, o que o algoritmo faz é tentar estimar os valores de β para encontrar um modelo capaz de classificar novas instâncias.

d) Classificadores Bayesianos

Outro classificador simples, mas poderoso, é o classificador Bayesiano (ou *Naïve Bayes*). Um classificador Bayesiano é um classificador probabilístico baseado na aplicação do teorema de Bayes para probabilidades condicionais e pode ser expresso como:

$$\mathbb{P}(Y = c|X = x) = \frac{f(x|Y = c)\mathbb{P}(Y = c)}{\sum_{k \in K} f(x|Y = k)\mathbb{P}(Y = k)} \quad (3)$$

onde K é o número total de classes, k são os possíveis rótulos, $\mathbb{P}(Y = k)$ é a probabilidade dos dados pertencerem à classe k e $f(x|Y = k)$ é a função densidade de probabilidade da instância pertencer à classe k . Durante o treinamento, o que o algoritmo faz é estimar os parâmetros da expressão (3). O parâmetro $\mathbb{P}(Y = k)$, para qualquer que seja o k , é fácil estimar, de modo que basta observar a distribuição de cada classe na amostra. No entanto, o desafio está em estimar o parâmetro $f(x|Y = k)$ e é aí que entra a suposição dos classificadores Bayesianos, que considera que todos os atributos são

independentes e não relacionados entre si. Dessa forma, $f(x|Y = k)$ pode ser calculada como o produto das distribuições marginais dado $Y = k$, o que pode ser expresso como mostra a Equação (4).

$$f(x|Y = k) = f(x_1, x_2, \dots, x_d|Y = k) = \prod_{i=1}^d f(x_i|Y = k) \quad (4)$$

Os classificadores Bayesianos podem ser treinados de forma muito eficiente, uma vez que o modelo escolhido é aplicado a cada atributo independentemente. Outro benefício desse classificador é que, uma vez que o modelo é aprendido, nenhum dado de treinamento precisa ser armazenado. Apenas os parâmetros do modelo são necessários. O classificador é construído multiplicando as probabilidades condicionais individuais de cada característica para obter a probabilidade total de uma classe. Em seguida, a classe com maior probabilidade é selecionada (SOLEM, 2012).

e) Análise Discriminante Linear

O método de Análise Discriminante Linear (do Inglês, *linear discriminant analysis* - LDA) também é um método probabilístico semelhante ao *Naïve Bayes*, no entanto, a suposição inicial sobre os atributos é diferente.

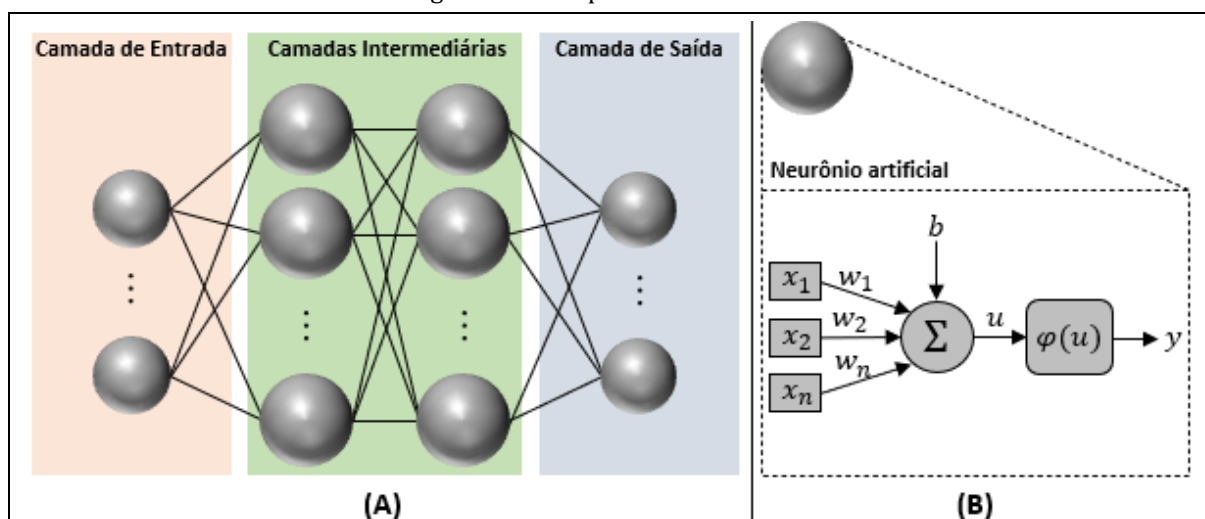
Enquanto o *Naïve Bayes* supõe que os atributos são independentes e não relacionados entre si, no LDA a suposição é a de que a distribuição conjunta dos atributos segue uma distribuição normal multivariada e, dessa forma, a dependência entre eles pode ser capturada e utilizada no cálculo do parâmetro $f(x|Y = k)$ do teorema de Bayes para estimar a probabilidade $\mathbb{P}(Y = c|X = x)$ (JAMES et al., 2013).

O fato dos classificadores Bayesianos assumirem a independência entre os atributos os tornam classificadores menos complexos do que o LDA, e por isso, o bias desses classificadores (que é o erro entre a predição e o valor real) tende a ser maior ao passo que a variância (erro entre duas predições com conjuntos diferentes) tende a ser menor. No LDA ocorre uma situação oposta, onde o bias tende a ser maior enquanto a variância tende a ser maior. O classificador ideal é aquele que consegue encontrar um equilíbrio entre bias e variância.

f) Redes Neurais Artificiais

A rede neural artificial é um dos algoritmos de *machine learning* mais populares e é inspirada no funcionamento do sistema nervoso central. Uma ideia mais intuitiva de como funcionam essas redes pode ser construída tomando como exemplo a rede neural artificial conhecida como perceptron multicamadas (do Inglês, *multilayer perceptron* - MLP), que é mostrada na Figura 8A.

Figura 8 - Perceptron Multicamadas



Fonte: da Autora.

Considerando um problema hipotético de classificação de imagens médicas, os neurônios da camada de entrada da MLP correspondem aos atributos extraídos das imagens, os da camada de saída correspondem aos desfechos (por exemplo, normal *versus* patológico) e os das camadas intermediárias são responsáveis pelo ajuste do modelo. A Figura 8B apresenta os principais elementos de um neurônio artificial, onde podemos destacar: i) as informações de entrada, representados pelo vetor $[x_1, x_2, \dots, x_n]$ (essas informações variam de acordo com a camada a qual o neurônio pertence, podendo ser os atributos das imagens, no caso dos neurônios da camada de entrada, ou as saídas das camadas anteriores, para os neurônios das demais camadas); ii) os pesos sinápticos, representados pelo vetor $[w_1, w_2, \dots, w_n]$; iii) o combinador linear Σ , que é responsável por gerar o potencial de ativação u a partir da soma dos valores de entrada ($x_1w_1 + x_2w_2 + \dots + x_nw_n + b$), onde b é o limiar de ativação e iv) a função de ativação, que é a

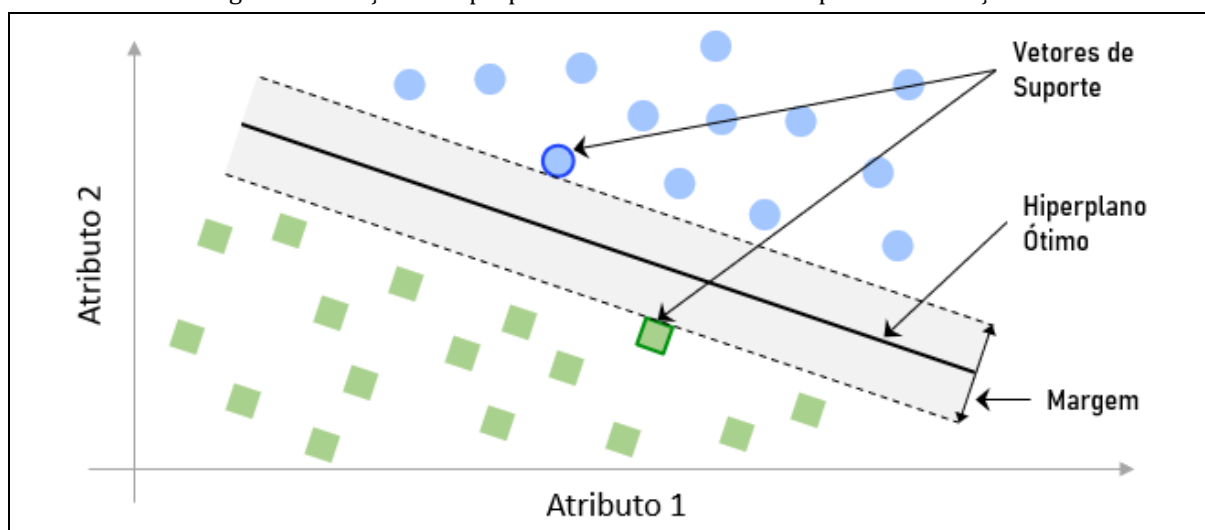
responsável por fazer com que o sinal se propague ou seja interrompido para aquele respectivo neurônio.

A força das redes neurais artificiais reside na integração de múltiplos neurônios nas múltiplas camadas ocultas. As saídas de uma camada servem como entradas da próxima camada. Na última camada de neurônios é implementada uma função de perda, que estima o erro da rede ao prever os rótulos de dados especificados, um processo chamado de *forward-propagation*. Com base no erro, pequenas mudanças são realizadas nos pesos da rede em um processo chamado de *back-propagation*. Repetidas iterações de *forward-propagation* e *back-propagation* em todo o conjunto de dados produzem, como resultado, uma rede otimizada.

g) Máquina de Vetores de Suporte

As máquinas de vetores de suporte (do Inglês, *support vector machine* - SVM) são um tipo poderoso de classificador que geralmente fornece resultados bastante precisos para muitos problemas de classificação. Trata-se de um classificador binário não probabilístico que, em sua forma mais simples, encontra um hiperplano de separação linear (um plano em espaços de dimensões superiores) com a melhor separação possível entre duas classes (SOLEM, 2012). Para isso, cada dado é plotado em um plano n-dimensional e o algoritmo tenta encontrar a maior distância entre os dados plotados e o hiperplano, ou seja, maximizar a margem entre eles, como mostrado na Figura 9.

Figura 9 - Criação de hiperplano do classificador SVM para classificação



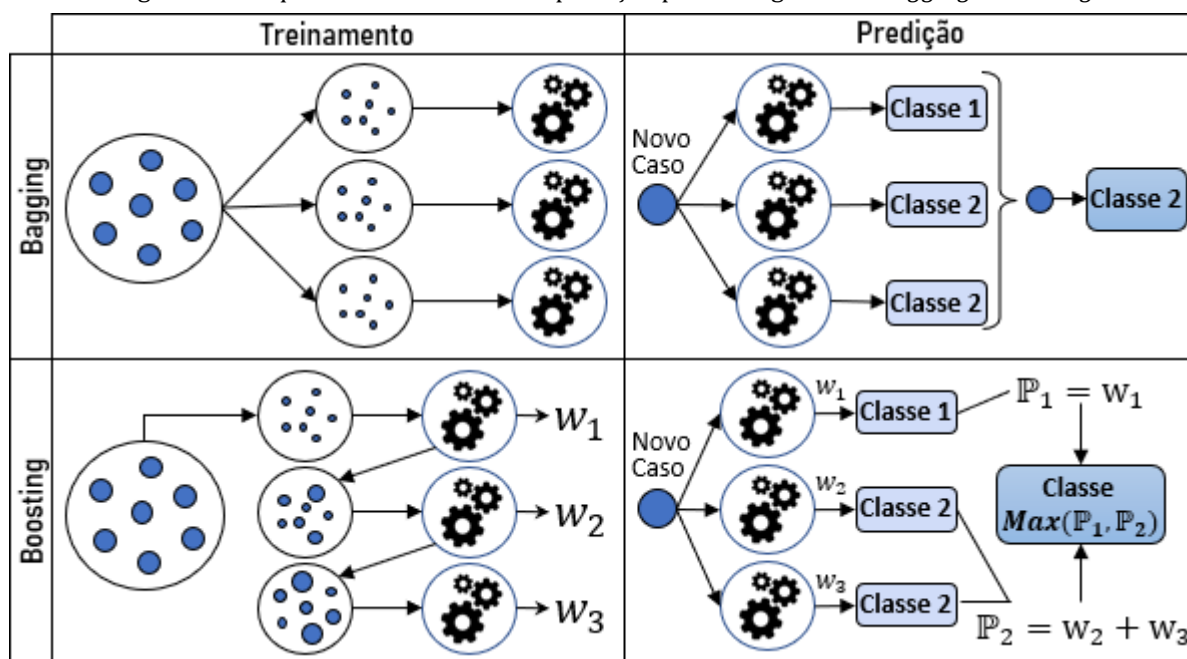
Fonte: da Autora.

h) Métodos *Ensemble*

Ensemble é um conceito de *machine learning* onde a ideia é treinar vários modelos usando o mesmo algoritmo de aprendizado (BISHOP, 2006; SOLEM, 2012). Para serem utilizados, esses métodos precisam de algum classificador base como, por exemplo, uma árvore de decisão. Três classificadores que utilizam essa abordagem e foram incluídos nesse trabalho são *Bagging*, *Boosting* e *Random Forest*.

No método *Bagging*, novos conjuntos de dados são criados utilizando a técnica *bootstrap*, que consiste em realizar uma reamostragem dos dados com reposição. Em seguida, são gerados n modelos, um para cada novo conjunto, e cada um é treinado paralelamente. Para classificar uma nova instância, ela é inserida em todos os modelos resultando em uma classe para cada modelo. A classe gerada pela maioria dos modelos será, então, atribuída à nova instância (BREIMAN, 1996).

O método *Boosting* também realiza uma reamostragem dos dados. No entanto, diferente do método *Bagging* no qual todas as instâncias têm a mesma probabilidade de aparecer no novo conjunto, no *Boosting* é atribuído um peso para as instâncias, de modo que as que foram classificadas erroneamente na primeira execução possuam um peso maior (maior representatividade no conjunto) do que as que foram classificadas corretamente. Esse processo implica que, no método *Boosting*, durante o treinamento os diferentes modelos sejam executados sequencialmente. Ao final do processo, um peso também é atribuído para cada modelo. Assim, para classificar uma nova instância, o número de votos para cada classe é expresso como a soma dos pesos dos modelos que tiveram como saída a respectiva classe. (FREUND; SCHAPIRE, 1996). A Figura 10 ilustra um esquema comparativo entre os métodos *Bagging* e *Boosting*.

Figura 10 - Esquema de treinamento e predição para os algoritmos *Bagging* e *Boosting*

Nota.— w_1, w_2 e w_3 são os pesos atribuídos para cada modelo. Fonte: da Autora.

O algoritmo *Random Forest*, por sua vez, também é considerado um método *Bagging*, no entanto, diferente do *Bagging* que utiliza todos os atributos para gerar os modelos, o *Random Forest* seleciona aleatoriamente os atributos que serão utilizados em cada árvore criada. Devido a essa seleção, as árvores são mais independentes umas das outras em comparação com o *Bagging*, o que geralmente resulta em melhor desempenho preditivo (JAMES et al., 2013).

Ainda com relação à criação de modelos de *machine learning*, nosso principal objetivo é desenvolver modelos que possuam a capacidade de generalização. Em outras palavras, o que desejamos é que o modelo aprenda uma tarefa a partir dos dados de entrada e possa, então, realizar previsões para novos dados com um erro aceitável. Para isso, existem algumas estratégias de divisão de dados que devem ser utilizadas para que seja evitado o *overfitting*, que ocorre quando o modelo está sobreajustado para os dados de treinamento e, portanto, não consegue generalizar para outros dados (BISHOP, 2006).

Uma das estratégias mais comuns consiste em dividir o conjunto de dados em três conjuntos (treinamento, validação e teste) de modo que cada um seja utilizado em um estágio diferente da criação do modelo. De maneira geral, o conjunto de treinamento é o conjunto usado para o ajuste dos parâmetros de um modelo como, por exemplo, os pesos

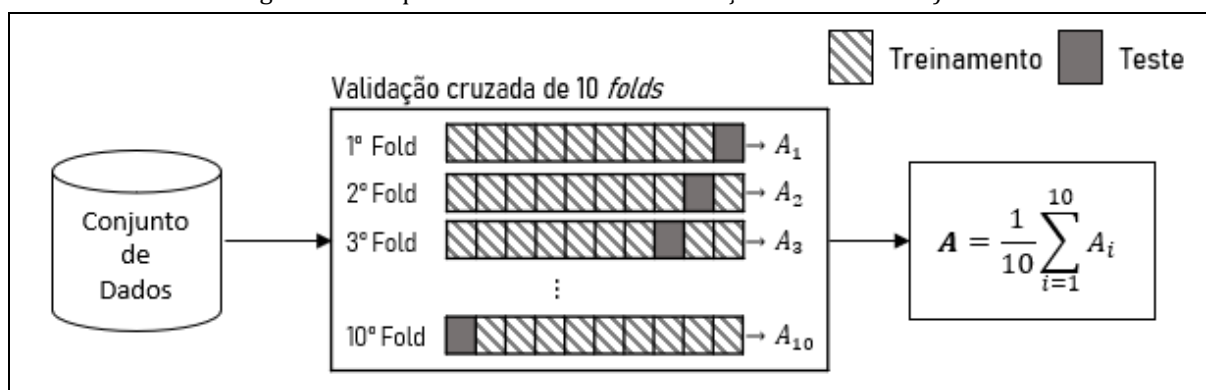
das interconexões entre os neurônios artificiais, ou os pesos de um modelo de regressão logística (JAMES et al., 2013).

O conjunto de validação pode ser utilizado para definir os hiperparâmetros do modelo, ou seja, aqueles que estão relacionados à arquitetura e que precisam ser configurados previamente (diferentemente dos parâmetros, que são fornecidos como resultado do modelo treinado). Ainda, o conjunto de validação pode ser utilizado para estimar a performance de um modelo na ausência de um conjunto de teste, como ocorre, por exemplo, com a validação cruzada, que será explicada adiante (JAMES et al., 2013).

O conjunto de teste, por sua vez, é utilizado para avaliar o desempenho e capacidade de generalização do modelo. Para isso, os dados devem pertencer a um conjunto totalmente independente. Isso significa que nenhuma de suas instâncias deve ser utilizada previamente durante as etapas de treinamento e/ou validação (JAMES et al., 2013).

Embora ter um conjunto de teste seja o cenário ideal para avaliar a performance de um modelo, em algumas situações, como é o caso da radiologia, muitas vezes é um desafio produzir um conjunto de dados com muitos casos com a devida qualidade e, por isso, é comum utilizar como estratégia de avaliação do modelo a técnica de validação cruzada (REFAEILZADEH; TANG; LIU, 2016). Essa técnica consiste em dividir o conjunto de dados em k partições (denominadas *folds*) de maneira que $k - 1$ partições são utilizadas para treinar o modelo e *fold* restante é utilizado para teste. O processo é repetido até que todos os *folds* tenham sido utilizados uma vez como teste. Esse processo é ilustrado na Figura 11.

Figura 11 - Esquema ilustrando uma validação cruzada de 10 *folds*



Nota.— A = avaliação. *Fonte:* da Autora

Com relação à avaliação do desempenho dos modelos, as métricas mais comumente utilizadas são as provenientes da matriz de confusão, como mostra a Figura 12.

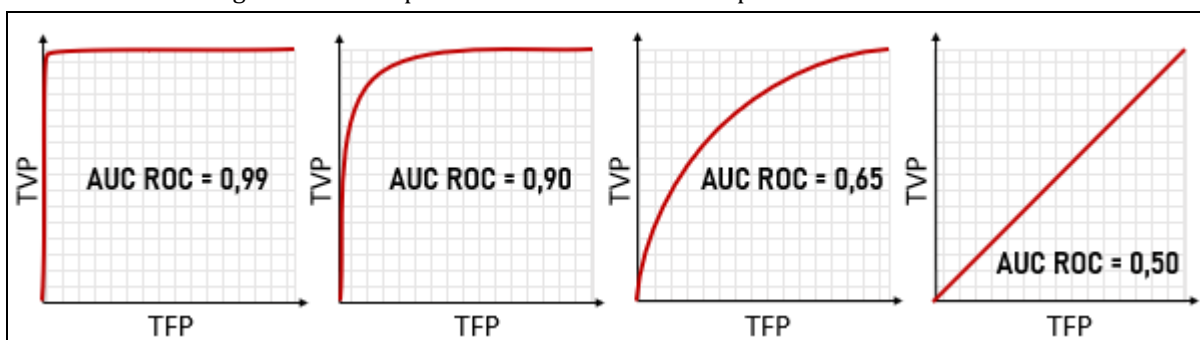
Figura 12 - Matriz de confusão e métricas de avaliação

		Valor Verdadeiro (padrão de referência)		
		positivo	negativo	
Valor Predito (predito pelo classificador)	positivo	VP Verdadeiro Positivo	FP Falso Positivo	$Acurácia = \frac{VP + VN}{n}$
	negativo	FN Falso Negativo	VN Verdadeiro Negativo	$Sensibilidade = \frac{VP}{VP + FN}$
				$Especificidade = \frac{VN}{VN + FP}$

Fonte: da Autora.

Outra métrica bastante comum para avaliar os modelos de *machine learning* é a área sob a curva ROC (ou AUC ROC). Ela é uma métrica interessante pois permite uma avaliação mais generalizada do modelo, não ficando restrita a um ponto de corte ótimo. O gráfico da curva ROC apresenta no eixo x a taxa de falsos positivos e no eixo y a taxa de verdadeiros positivos e plota os dados para diferentes limiares de classificação. O que a AUC ROC faz é resumir a informação da curva ROC resultando em um único valor, que é a área abaixo da curva. Quanto mais próxima a AUC ROC for de 1, melhor é a performance do modelo para discriminar os dados entre as duas classes e quanto mais próxima de 0,5, pior é a performance. Figura 13 mostra alguns exemplos de curva ROC com seus respectivos valores de AUC.

Figura 13 - Exemplos de curva ROC com os respectivos valores de AUC



Nota.— TVP = taxa de verdadeiros positivos; TFP = taxa de falsos positivos. Fonte: da Autora.

3 MATERIAL E MÉTODOS

3.1 ASPECTOS ÉTICOS

Este trabalho foi devidamente avaliado e aprovado pelo Comitê de Ética em Pesquisa (CEP) do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo – HCFMRP-USP, de acordo com o parecer de número 2.550.679 e o CAAE de número 84415318.8.0000.5440. A confirmação pode ser realizada pelo site da Plataforma Brasil fornecendo um desses dois números de identificação.

3.2 BASE DE DADOS

Para a realização deste trabalho foram utilizados exames de ressonância magnética da coluna vertebral ponderados em T1, no período de 2010 a 2019, de pacientes consecutivos que estavam em tratamento no Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo (HCFMRP-USP) e que foram diagnosticados com fratura vertebral por compressão benigna ou maligna na região lombar.

Para a seleção desses pacientes, foi realizada uma consulta no Sistema de Informação de Radiologia (do Inglês, *Radiology Information System* - RIS) por meio do módulo “Busca de Laudo por Patologia”. A Figura 14 apresenta os filtros que foram utilizados para a seleção dos pacientes.

Figura 14 - Interface do RIS para busca de laudo por patologia

The screenshot shows a software interface titled "Pesquisa por Patologia". It contains a search form with the following fields and values:

- Data Inicial:** 2010-01-01
- Exame:** 6
- Doença:** FRATURA
- Data Final:** 2019-12-31
- Região:** 13
- Médico Emissor:** COLUNA LOMBO SACRA

There are four buttons on the right side of the form: "Pesquisar" (with a magnifying glass icon), "Limpar" (with an eraser icon), "Excel" (with a spreadsheet icon), and "Sair" (with a door icon).

Fonte: da Autora

Após a exportação da lista dos pacientes resultantes da consulta em uma planilha do Excel, foi realizada uma varredura nos dados para eliminar tanto os registros duplicados quanto os casos em que o paciente não apresentava fratura nas vértebras lombares.

É importante ressaltar que como a consulta foi realizada utilizando a palavra “fratura”, foram recuperados casos em que o paciente poderia ter a presença de fratura em outras regiões anatômicas que não fossem os corpos vertebrais lombares. Além disso, os laudos que continham expressões como, por exemplo, “sem evidência de fratura”, também foram recuperados. Outra observação importante é que, embora na consulta no RIS tenham sido filtrados apenas exames de ressonância magnética da região lombo sacra, muitos dos resultados mostravam laudos de fraturas em regiões torácicas e cervicais. Isso ocorre pois, na rotina clínica, é comum que o médico solicite a realização de um exame da coluna total do paciente. Esse exame pode ser gerado por meio da união de exames adquiridos, isoladamente, das regiões lombar, torácica e cervical. Nesses casos, quando o médico realiza o laudo, ele registra o que observa na coluna como um todo, de modo que esse laudo fica anexado a cada um dos exames que foi adquirido separadamente. Portanto, é comum encontrarmos no PACS (do Inglês, *Picture Archiving and Communication System*) um exame da região lombar cujo laudo traga informações tanto da região lombar quanto das demais regiões. O Quadro 1 mostra um exemplo dos laudos recuperados a partir da consulta realizada no RIS.

Quadro 1 - Exemplos de conclusão de laudos recuperados a partir da consulta no RIS

ESTADIAMENTO DE NEOPLASIA DE MAMA COM ACOMETIMENTO DIFUSO DA MEDULA ÓSSEA COMPATÍVEL COM METÁSTASES DISSEMINADAS E FRATURAS PATOLÓGICAS DOS CORPOS VERTEBRAIS DE T10 E L4.
SINAIS DE FRATURA RECENTE DO CORPO VERTEBRAL DE T12 , SEM DESLOCAMENTO DO MURO POSTERIOR. ESPONDILOARTROSE LOMBAR SEM EVIDÊNCIAS DE ESTENOSE SIGNIFICATIVA DO CANAL VERTEBRAL OU DOS FORAMES NEURAIS.
EXAME SEM EVIDÊNCIAS DE LESÕES SUSPEITAS PARA NEOPLASIA. ACHADOS SUGESTIVOS DE FRATURAS POR INSUFICIÊNCIA NOS CORPOS VERTEBRAIS DE T7, T8 E T10 . ALTERAÇÕES DEGENERATIVAS COM ESTREITAMENTO FORAMINAL PRINCIPALMENTE EM L5-S1 À ESQUERDA.
FRATURAS DE CORPOS VERTEBRAIS D7, L3 E L5. O ASPECTO SUGERE FRATURAS DE INSUFICIÊNCIA (SECUNDÁRIO A INFILTRAÇÃO DA MEDULA ÓSSEA?). ALTERAÇÕES DEGENERATIVAS DA COLUNA VERTEBRAL, COM ESTENOSE DOS FORAMES NEURAIS L5-S1.
ESPONDILOLISTESE L5-S1 GRAU 1 ASSOCIADA A FRATURA DO ISTMO DIREITO DE L5. ALTERAÇÕES DEGENERATIVAS COM ESTENOSE DO CANAL VERTEBRAL LOMBAR.
LISTESE ANTERIOR L3-L4 GRAU I COM FRATURA DOS ÍSTMOS DE L3 . ALTERAÇÕES NOS NÍVEIS L3-L4 E L4-L5 COMPATÍVEIS COM SEQUELA DE ESPONDILODISCITE. FORMAÇÃO EPIDURAL COM EXTENSÃO INTRA-ÓSSEA EM S1 (COLEÇÃO? CONTEÚDO DISCAL MIGRADO?) HÉRNIA EXTRUSA L4-L5.
CONTROLE DE ARTRODESE VIA POSTERIOR LOMBOSSACRAL, COM SINAIS DE ESTENOSE DO CANAL VERTEBRAL AO NÍVEL DE S1-S2. ALTERAÇÕES DEGENERATIVAS DA COLUNA LOMBAR E DISCOPATIA. FRATURA DOS CORPOS VERTEBRAIS DE T9, T12 E L3.
FRATURA CRÔNICA ENVOLVENDO OS DOIS TERÇOS ANTERIORES DO CORPO VERTEBRAL DE T12 , INDETERMINADA. ALTERAÇÕES DEGENERATIVAS DA COLUNA VERTEBRAL LOMBOSSACRA SEM EVIDÊNCIAS DE COMPRESSÃO RADICULAR OU DE ESTENOSE DE CANAL.

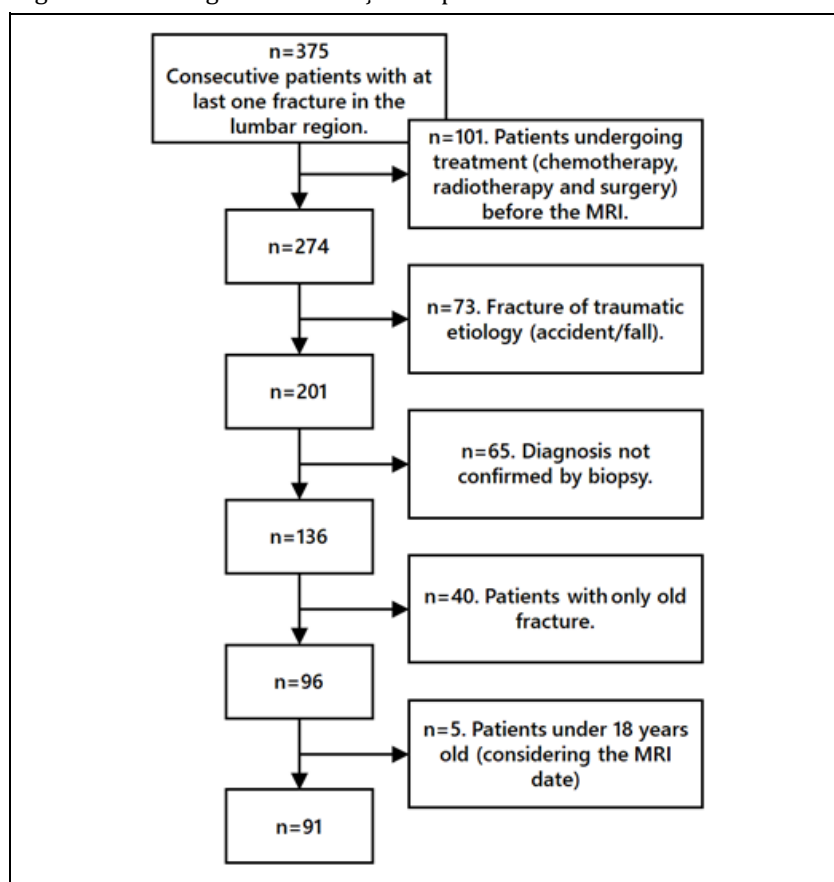
Nota.— As células em cinza apresentam exemplos de laudos que não contêm menção à fratura nos corpos vertebrais lombares e que foram excluídos. Os textos foram transcritos exatamente como está no laudo oficial.

Após a seleção dos casos que correspondiam ao critério de inclusão (pacientes com fraturas em corpos vertebrais lombares), foi feita uma última filtragem para eliminação dos casos de acordo com os seguintes critérios de exclusão: a) realização de tratamentos como quimioterapia, radioterapia e cirurgia; b) fraturas de etiologia traumática; c) ausência de confirmação histopatológica (para os casos de fraturas malignas); d) fratura antiga e e) pacientes menores de 18 anos (considerando a data de realização do exame).

Para checagem dos critérios de exclusão, foi realizada a revisão dos prontuários dos pacientes pela autora sob supervisão do Prof. Dr. Marcello Henrique Nogueira Barbosa, radiologista com 20 anos de experiência em radiologia do sistema musculoesquelético. Após a definição do conjunto final dos pacientes, o diagnóstico de fratura benigna ou maligna foi confirmado para cada paciente pelo Prof. Dr. Marcello Henrique Nogueira Barbosa, que realizou uma última revisão dos prontuários, incluindo história clínica, exames de biópsia, exames laboratoriais e exames de imagem complementares. O período mínimo de acompanhamento clínico e de imagem foi de dois anos.

O conjunto final foi composto por 91 pacientes (36 homens e 55 mulheres, com idade média de 64.24 ± 11.75 anos). A Figura 15 apresenta o fluxograma da seleção de pacientes.

Figura 15 - Fluxograma da seleção de pacientes com critérios de exclusão



Fonte: da Autora

Após a definição da lista final de pacientes que fariam parte do estudo, foi realizada a recuperação desses exames a partir do PACS da Instituição. Para isso, foi desenvolvido um *script* na linguagem *python* para realização do download automático dos exames a partir das informações contidas no cabeçalho das imagens referentes ao registro do paciente, à modalidade do exame e à data de realização do exame. Ainda nesse *script*, foi inserida uma função de anonimização de algumas informações presentes no cabeçalho DICOM¹, a saber: *PatientName*, *InstitutionAddress*, *InstitutionName*, *OperatorsName*, *ProtocolName*, *RequestingPhysician*, *RequestedPROCEDUREDescription* e *BodyPartExamined*. Isso garantiu que todos os exames usados nesse trabalho fossem devidamente anonimizados.

¹ *Digital Imaging and Communications in Medicine - DICOM* é o padrão internacional para transmitir, armazenar, recuperar, imprimir, processar e exibir informações de imagens médicas. <https://www.dicomstandard.org/>

3.3 PIPELINE DO RECONHECIMENTO DE PADRÕES EM IMAGENS MÉDICAS

Como explicado na Seção 2.2, uma vez que as imagens tenham sido recuperadas, as etapas para realizar o reconhecimento de padrões são: pré-processamento, segmentação, extração de características quantitativas, seleção de características e análise por meio do aprendizado de máquina.

Nesta seção, serão descritos os métodos e o suporte ferramental utilizados neste trabalho para cada uma dessas etapas. Vale destacar que em algumas situações, mais de um método foi utilizado para cada etapa, com a finalidade de explorá-los para, em seguida, definir o que melhor se ajustaria aos nossos dados e objetivos.

3.3.1 Pré-processamento

Após realizado o processo de recuperação dos exames de RM a partir do PACS, a etapa de pré-processamento das imagens foi implementada.

Inicialmente foi realizada a conversão da imagem de 16 bits/pixel para 8 bits/pixel, para que a quantização de intensidade de níveis de cinza abrangesse os valores de 0 a 255. Essa conversão foi realizada utilizando o filtro *RescaleIntensityImageFilter*, o qual aplica uma transformação linear para os níveis de intensidade da imagem de entrada em termos dos valores de mínimo e máximo que a imagem de saída deveria ter. Este filtro está inserido no software 3D Slicer² versão 4.8.1, que é um software para processamento de imagens médicas que se comporta de forma semelhante a um workstation típico da radiologia, sendo uma ferramenta de uso livre e não dependente de nenhum hardware específico. A Equação (5) expressa o mapeamento dos valores de intensidade utilizado pelo filtro *RescaleIntensityImageFilter*.

$$outputPixel = (inputPixel - inputMin) \cdot \frac{(outputMax - outputMin)}{(inputMax - inputMin)} + outputMin \quad (5)$$

Após a requantização das imagens, foi realizada a equalização dos histogramas, que foi utilizada para realçar detalhes por meio da redistribuição dos valores de ocorrência no histograma da imagem. Um histograma de imagem, representado graficamente por

² www.slicer.org

colunas verticais, mostra a distribuição das frequências de intensidades de uma imagem. O que a equalização faz é expandir esse histograma de modo que a imagem passe a conter, de modo mais equilibrado, sombras, meio tons e luminosidade, melhorando seu contraste e realçando detalhes. A Figura 16 apresenta imagens da coluna vertebral com seus respectivos histogramas. Na Figura 16A, a imagem não passou por nenhum tipo de processamento. Ao observar seu histograma, pode-se notar que a distribuição não ocupa toda a escala de níveis de cinza. Já na Figura 16B, é apresentada a imagem após o processo de equalização. Ao observar seu histograma, nota-se que houve uma abertura da distribuição e que a imagem passou a ter tons de cinza ao longo de toda a escala.

Figura 16 - Representação de uma imagem antes e depois do processo de equalização.



Fonte: da Autora

A equalização foi realizada por meio da ferramenta *AdaptiveHistogramEqualizationImageFilter*, também presente no software 3D Slicer versão 4.8.1, e foi configurada com os seguintes parâmetros: $\alpha = 0,3$, $\beta = 0,3$ e raio = 5.

3.3.2 Segmentação

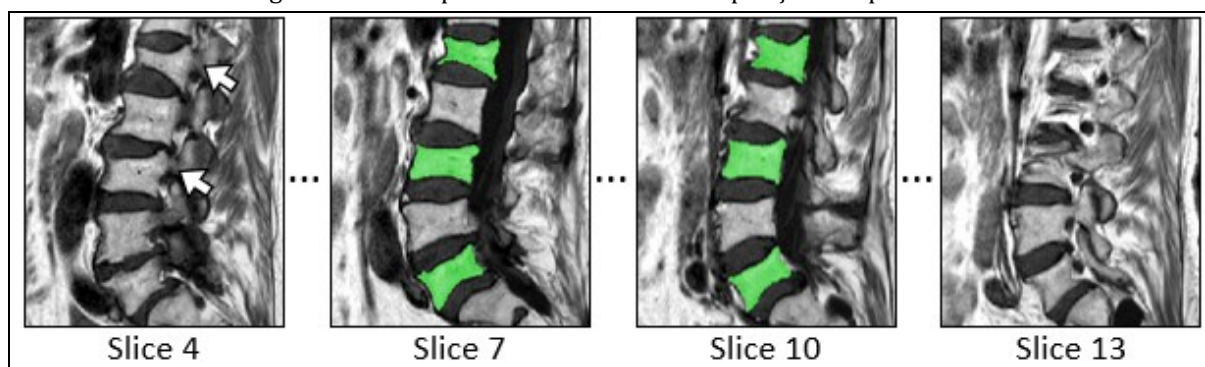
A segmentação consiste em delimitar nos exames de imagem as regiões de interesse (do Inglês, *regions of interest* - ROIs) que terão suas características posteriormente extraídas. Em nosso estudo, consideramos como região de interesse os corpos vertebrais fraturados da região lombar da coluna vertebral.

A segmentação foi realizada utilizando o software 3D Slicer por meio de algoritmos de limiarização, crescimento de região e erosão. Entretanto, por se tratar de processos semiautomáticos, eventuais marcações e ajustes manuais foram necessários. Após a

conclusão das segmentações, os resultados foram validados pelo Prof. Dr. Marcello Henrique Nogueira Barbosa.

Para a realização da segmentação, foram considerados apenas os *slices* nos quais os pedículos não apareciam, como mostra a Figura 17.

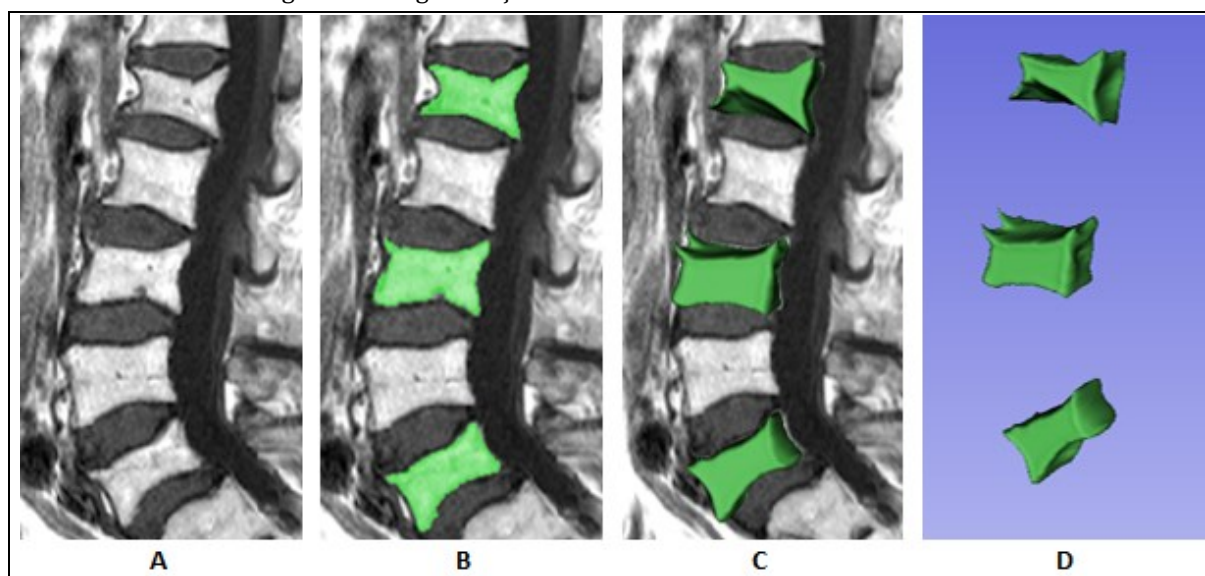
Figura 17 - Exemplos de *slices* com e sem aparição dos pedículos



Nota.— As setas brancas indicam a presença de alguns pedículos na imagem. Os *slices* 4 e 13 são exemplos de *slices* que não foram segmentados. Os *slices* 7 e 10 são exemplos de *slices* que foram segmentados. Fonte: da Autora.

A Figura 18A apresenta um recorte de ressonância magnética ponderado em T1 da coluna vertebral. Figura 18B mostra a segmentação de um único corte do exame. A Figura 18C apresenta a segmentação tridimensional sobreposta ao exame original e a Figura 18D mostra apenas a segmentação 3D.

Figura 18 - Segmentação tridimensional realizada no 3D Slicer



Fonte: da Autora.

3.3.3 Extração de características quantitativas

Para a extração das características quantitativas das imagens segmentadas dos corpos vertebrais foi utilizado o pacote *PyRadiomics*³, que é uma biblioteca livre, escrita em *python*, que permite o processamento e extração de atributos radiômicos de forma, textura e intensidade de nível de cinza a partir de imagens médicas (VAN GRIETHUYSEN et al., 2017).

Embora possua uma interface *front-end* embutida no *3D Slicer*, nós optamos por utilizar a interface *back-end*. Para isso, foi desenvolvido um *script* em que foi inserido o pacote *PyRadiomics*. Esse *script* possibilitou tanto realizar o processamento em lote quanto melhorar a customização do processo.

Na versão em que foi utilizada (v 0.9.7), a biblioteca extraía um total de 109 atributos, distribuídos da seguinte maneira: 19 de primeira ordem (atributos de intensidade de cinza), 16 de forma e 74 de textura. Os atributos de textura são calculados a partir das seguintes matrizes: coocorrência de níveis de cinza (23 atributos), comprimento de cadeia de níveis de cinza (16 atributos), tamanho de zona de níveis de cinza (16 atributos), diferença de tons de cinza vizinhos (5 atributos) e dependência de níveis de cinza (14 atributos).

Os atributos de forma não foram considerados neste trabalho, pois o número de *slices* utilizados para caracterização tridimensional de cada corpo vertebral foi definido com base na presença ou ausência dos pedículos. Essa seleção de cortes poderia potencialmente causar mudanças no formato original das imagens dos corpos vertebrais, inserindo um possível fator de confusão na análise tridimensional da forma. Logo, foram extraídas um total de 93 características para o presente trabalho.

Após a execução do *script* para a extração das características, o algoritmo retornou uma planilha no formato *.csv* contendo o valor de cada atributo para cada um dos corpos vertebrais. Esses dados foram então normalizados de acordo com a Equação (6)

$$\text{ValorNormalizado} = (\text{ValorOriginal} - \text{min}) / (\text{max} - \text{min}) \quad (6)$$

³ <https://pyradiomics.readthedocs.io/en/latest/>

Após a normalização, foi inserida uma última coluna na planilha contendo a classe à qual cada instância pertencia. Por fim, o arquivo foi salvo no formato `.csv` para ser inserido como vetor de características nas etapas seguintes.

3.3.4 Seleção de características

Para a etapa de seleção de características foram usados tanto o método de filtros quanto o método *wrapper*.

Com relação aos filtros, utilizamos quatro algoritmos, sendo que três (*Fisher_Score*, *ReliefF* e *Trace_Ratio*) eram baseados em similaridade e um (*t_Score*) era baseado em estatística. Os *scripts* de seleção de características foram implementados utilizando o pacote *scikit-feature*⁴, que é um repositório de algoritmos de seleção de características *open-source* em *python* desenvolvido pela Universidade do Estado do Arizona. Os *scripts* recebem como entrada o vetor com as características para serem selecionadas e retornam um arquivo `.csv` com as características e seus respectivos scores.

No método *wrapper*, os subconjuntos de características foram avaliados dentro do próprio classificador. Para isso, utilizamos o software WEKA⁵ (*Waikato Environment for Knowledge Analysis*) na versão 3.8.4, que disponibiliza uma coleção de algoritmos para aprendizado de máquina, contendo ferramentas de preparação dos dados, seleção de atributos, classificação, regressão, clusterização e visualização. Para cada um dos classificadores utilizados nesse trabalho, o método *wrapper* foi executado. Os parâmetros disponíveis para configuração do método *wrapper* no software WEKA são: o estado inicial do espaço de busca, o método de busca, a direção da busca, o critério de parada e a métrica de avaliação.

- **Estado inicial:** No caso do estado inicial, devemos informar como o nosso vetor de características será iniciado. Ele pode estar vazio, completo ou com apenas alguns atributos.
- **Método de busca:** Como método de busca temos as opções *GreedyStepwise* ou *BestFirst*. Ambos os algoritmos realizam uma busca gananciosa no espaço de atributos, no entanto, no *GreedyStepwise* a busca é interrompida quando um atributo, ao ser adicionado (ou excluído), reduzir a performance. No caso do

⁴ <https://jundongli.github.io/scikit-feature/>

⁵ <https://www.cs.waikato.ac.nz/~ml/weka/index.html>

BestFirst, uma função de *backtracking* é adicionada ao algoritmo permitindo que a avaliação da performance de cada adição/exclusão do atributo seja comparada com um número definido de performances passadas.

- **Direção da busca:** Temos as opções *forward*, *backward* e *bi-directional*. *Forward* é usada quando o estado inicial é vazio, dessa forma, os atributos vão sendo adicionados ao subconjunto de atributos para serem avaliados. A opção *Backward* é usada quando o estado inicial é completo, fazendo com que as características sejam eliminadas uma a uma. Com a opção *bi-directional*, são feitas duas buscas simultâneas, uma progressiva e uma regressiva, com o objetivo de que as duas se encontrem no meio do percurso.
- **Critério de parada:** Nesse parâmetro especificamos quantos atributos devem ser analisados consecutivamente sem que haja melhora na performance do algoritmo. Quando esse número é atingido, a busca é interrompida.
- **Métrica de avaliação:** Trata-se da medida que será utilizada para comparação das performances do algoritmo com cada um dos subconjuntos de atributos.

Os valores para cada um desses parâmetros foram definidos de maneira individual para cada experimento realizado neste trabalho e, por isso, serão detalhados nas próximas seções.

3.3.5 Análise

No presente trabalho foram realizadas tanto a análise univariada quanto a análise multivariada. Para a realização da análise univariada, foram utilizadas ferramentas estatísticas tradicionais, que serão detalhadas na seção experimentos.

Para a análise multivariada, utilizamos o aprendizado de máquina supervisionado. Parte das análises foi realizada no software WEKA e uma outra parte foi executada via *script* na linguagem *python*, por meio do pacote *scikit-learn*, que é uma biblioteca de aprendizado de máquina de código aberto. Tanto no WEKA quanto com o pacote *scikit-learn*, foram utilizados algoritmos de diferentes famílias, a saber: *Nearest Neighbors*, *Decision Tree*, *Discriminant Analysis*, *Naïve Bayes*, *Neural Network*, *Logistic Regression*, *Support Vector Machine*, *Random Forest*, *Bagging* e *Boosting*.

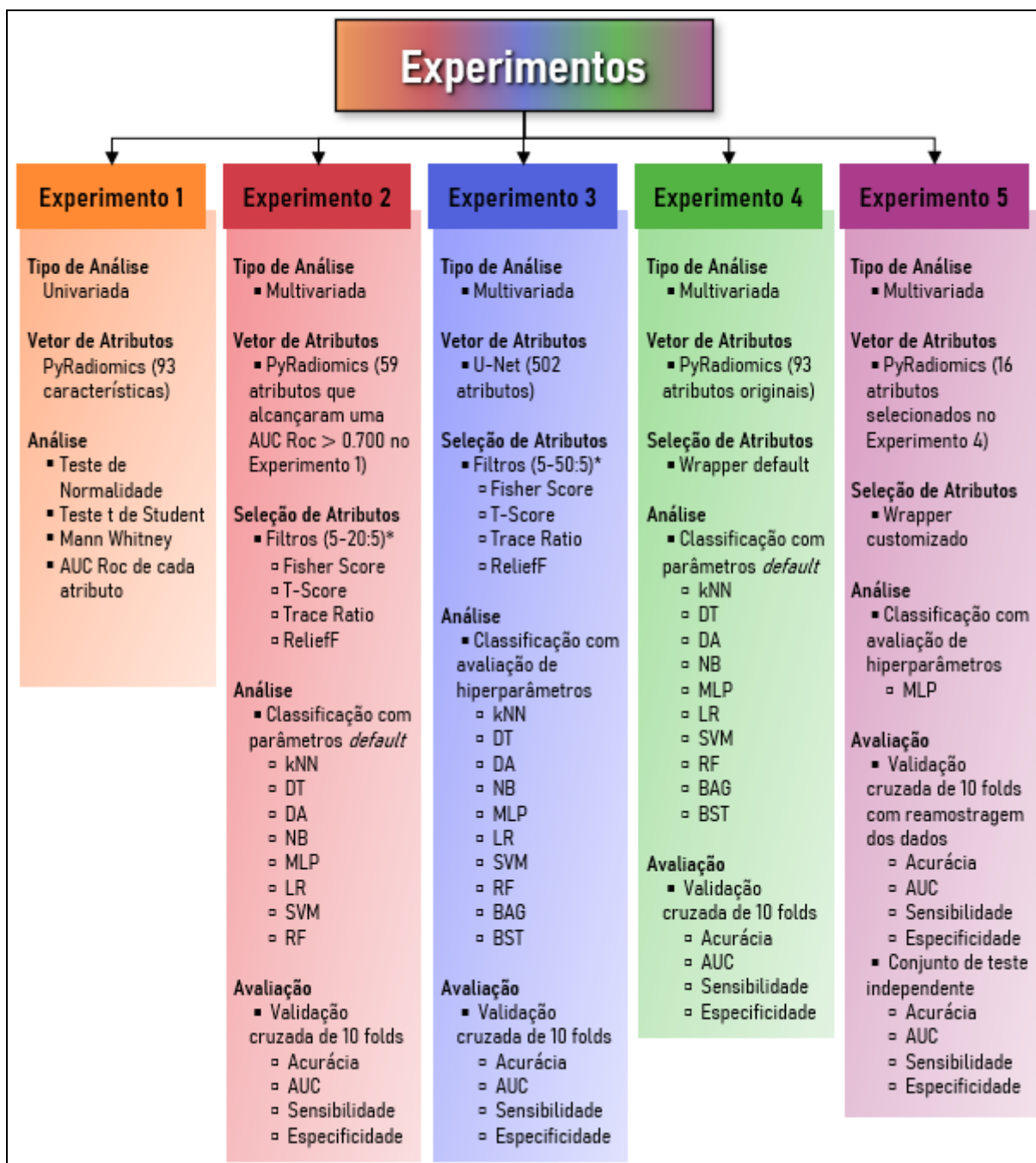
Detalhes sobre as configurações utilizadas em cada algoritmo e os métodos de treinamento, validação e teste serão informados nas próximas seções, as quais

descreverão os experimentos realizados neste trabalho. Além disso, as métricas utilizadas para avaliar os modelos bem como as demais avaliações estatísticas utilizadas serão detalhadas para cada um desses experimentos.

3.4 EXPERIMENTOS

No presente trabalho foram realizados cinco experimentos. No primeiro deles foi realizada uma análise univariada das 93 características extraídas por meio do pacote PyRadiomics. Do segundo ao quinto, foram realizadas análises multivariadas, de modo que para cada um foram definidos métodos específicos dentro de cada etapa do *pipeline* de reconhecimento de padrão em imagens. A Figura 19 ilustra o esquema sumarizado dos experimentos realizados neste trabalho e as subseções seguintes detalham cada um deles.

Figura 19 - Esquema sumarizado dos experimentos realizados



Nota.— kNN = k nearest neighbor; DT = decision tree; DA = discriminant analysis; NB = naïve bayes; MLP = multilayer perceptron; LR = logistic regression; SVM = support vector machine; RF = random forest; BAG = bagging; BST = boosting; AUC = area under curve.

* No formato (5-20:5), o primeiro número indica a quantidade de atributos inicial, o segundo indica a quantidade de atributos final e o terceiro o incremento de atributos. Ex: No Experimento 2, usamos o filtro Fisher Score para selecionar as características com a configuração (5-20:5). Isso significa que, para esse filtro, foram realizadas análises usando como vetor de características 4 subconjuntos de atributos, sendo o primeiro deles composto pelos 5 atributos mais bem ranqueados, o segundo pelos 10, o terceiro pelos 15 e o quarto pelos 20 atributos mais bem ranqueados.

Fonte: da Autora

3.4.1 Experimento 1 - Análise Univariada

Neste primeiro experimento, realizamos uma análise univariada dos atributos que foram extraídos das imagens utilizando a biblioteca *PyRadiomics* para avaliar a capacidade de cada um em diferenciar, isoladamente, as fraturas malignas e benignas. Como mencionado anteriormente, foram extraídos 93 atributos, dos quais 19 eram de primeira ordem e 74 eram de textura.

Para cada atributo foram realizadas duas análises: o teste de hipóteses para duas médias e o cálculo da área sob a curva ROC. Com relação ao teste de hipóteses, antes de realizar a análise, a normalidade dos dados foi testada para cada grupo (fraturas benignas e malignas) por meio do teste de Anderson-Darling com nível de significância de 5%. Para as variáveis cujos dados de ambos os grupos seguiram a distribuição normal, aplicamos o teste t de Student e para as demais, aplicamos o teste de Mann Whitney. Para a realização desses cálculos, usamos o módulo *stats* da biblioteca *scipy*, que fornece uma grande quantidade de ferramentas para cálculos estatísticos.

Para calcular a AUC ROC foi utilizado o estimador Wilcoxon-Mann-Whitney a partir do método DeLong-DeLong-Pearson. A ferramenta estatística utilizada foi a *Analyse-it*⁶ na versão *Ultimate Edition*, que conta com análises de performance diagnóstica como gráfico de curva ROC, gráfico de curva ROC múltiplo, sensibilidade/especificidade, razões de verossimilhança, valores preditivos, AUC, comparação de AUC, determinação de *threshold* entre outras.

Após as análises individuais de cada variável, foi realizada uma análise para cada família de atributos e suas proporções, que foram comparadas utilizando o teste de hipótese para duas proporções com nível de confiança de 95%, disponível no software *Minitab*⁷.

3.4.2 Experimento 2 - PyRadiomics + Filtros

Neste experimento foi realizada uma análise multivariada exploratória usando quatro filtros para seleção de características e oito algoritmos de classificação para investigar quais pares de métodos teriam maior eficiência na tarefa de diferenciar fraturas benignas

⁶ <https://analyse-it.com/products/ultimate>

⁷ <https://www.minitab.com/pt-br/>

e malignas. Para realizar este experimento, foram utilizadas apenas as características que obtiveram, isoladamente, uma AUC ROC ≥ 0.700 no Experimento 1.

Com relação à seleção de características, foram utilizados os quatro filtros apresentados na seção 3.3.4, a saber: *Fisher_Score*, *ReliefF*, *Trace_Ratio* e *t_Score*. Para cada um, foi gerado um ranking de características ordenadas em ordem decrescente em relação ao score obtido. Ao final, foram selecionadas as 5, 10, 15 e 20 características mais bem ranqueadas em cada um dos filtros, totalizando 16 subconjuntos de atributos (vetores de características) que foram testados com os algoritmos de classificação.

Para o aprendizado de máquina, foram utilizadas as seguintes famílias de algoritmos: *Nearest Neighbor*, *Decision Tree*, *Discriminant Analysis*, *Naïve Bayes*, *Artificial Neural Network*, *Logistic Regression*, *Support Vector Machine* e *Random Forest*. Para este experimento utilizamos os algoritmos disponíveis no WEKA com os parâmetros *default*. O Quadro 2 apresenta os algoritmos escolhidos para cada família e os parâmetros configurados como *default*.

Quadro 2 - Algoritmos de classificação utilizados para cada família e seus respectivos parâmetros *default*

Família	Algoritmo (parâmetros)
Nearest Neighbor	IBk (k=3; batchSize = 100; searchAlgorithm = LinearNNSearch)
Decision Tree	J48 (batchSize = 100; confidenceFactor = 0,25; minNumObj = 2)
Discriminant Analysis	LDA (batchSize = 100; ridge = 1,0 e ⁻⁶)
Naïve Bayes	Naive Bayes (batchSize = 100)
Neural Network	MLP (batchSize = 100; learningRate = 0,3; momentum = 0,2; hiddenLayers = (a+c)/2; epochs = 500)
Logistic Regression	Logistic (batchSize = 100; ridge = 1,0 e ⁻⁸)
Support Vector Machine	SMO (batchSize = 100; complexityParameter = 1,0; calibrator = Logistic; épsilon = 1,0 e ⁻¹² ; kernel = PolyKernel; toleranceParameter = 0,001)
Random Forest	RandomForest (bagSizePercent = 100; batchSize = 100; numIterations = 100; maxDepth = none;)

Neste experimento foi utilizado o conjunto de treinamento contendo 100 corpos vertebrais (54 com fraturas benignas e 46 com fraturas malignas) e para avaliar o classificador foi utilizada a validação cruzada de 10 *folds*. As medidas de acurácia, área sob a curva ROC, sensibilidade e especificidade foram, então, calculadas e inseridas em um *script* para criação de mapas de calor, permitindo uma melhor análise dos resultados do experimento.

3.4.3 Experimento 3 - U-Net + Filtros

Este experimento trata-se também de uma análise exploratória multivariada, e foi desenvolvido como parte do projeto do Doutorado Sanduíche realizado sob supervisão do Prof. Dr. Hiroyuki Yoshida no Massachusetts General Hospital da Harvard Medical School, no período de dezembro/2019 a maio/2020.

As características quantitativas foram fornecidas por seu grupo de pesquisa, que utilizou as imagens e segmentações semiautomáticas dos 100 corpos vertebrais fraturados do conjunto de treinamento como *input* da rede U-Net e retornou 502 características quantitativas em um arquivo *.csv*. A partir desses dados, foram realizados os processos de seleção de características e classificação.

Com relação à seleção de características, utilizamos os mesmos filtros do experimento 2 (*Fisher_Score*, *ReliefF*, *Trace_Ratio* e *t_Score*) para gerar 40 subconjuntos de atributos (vetores de características), que foram criados a partir da seleção das 5, 10, 15, 20, 25, 30, 35, 40, 45 e 50 características mais bem ranqueadas em cada um dos filtros.

A etapa de classificação foi realizada em *python*, por meio do pacote *scikit-learn*. O conjunto de dados utilizado foi o formado pelos 100 corpos vertebrais fraturados (46 fraturas malignas e 54 benignas). Antes da classificação, foi realizada a seleção dos hiperparâmetros dos algoritmos por meio da função *GridSearchCV*. Foram avaliados 10 algoritmos de classificação, a saber: *Nearest Neighbor*, *Decision Tree*, *Discriminant Analysis*, *Naïve Bayes*, *Artificial Neural Network*, *Logistic Regression*, *Support Vector Machine*, *Random Forest*, *Bagging* e *Boosting*. No total, foram realizados 400 subexperimentos, que foram avaliados por meio da validação cruzada de 10 *folds* com as métricas de acurácia, área sob a curva ROC, sensibilidade e especificidade.

3.4.4 Experimento 4 - PyRadiomics + Wrapper

Neste experimento foi realizada a análise exploratória dos algoritmos de classificação com o método de seleção de características *wrapper*. Para isso, foi utilizado o software WEKA com suas respectivas ferramentas e algoritmos. O vetor de características de entrada foi o vetor original contendo 93 atributos extraídos a partir dos 100 corpos vertebrais fraturados (54 com fraturas benignas e 46 com fraturas malignas).

Os algoritmos analisados foram: *Nearest Neighbor*, com $k = 5, 7$ e 9 , *Decision Tree*, *Discriminant Analysis*, *Naïve Bayes*, *Multilayer Perceptron*, *Logistic Regression*, *Support*

Vector Machine, *Random Forest*, *Bagging* e *Boosting*. Para cada um deles, antes de executar a classificação, foi realizada a seleção de características por meio do método *wrapper* com sua configuração *default*, como mostra o Quadro 3. A partir das características selecionadas, foi realizado o treinamento dos algoritmos e a avaliação por meio da validação cruzada de 10 *folds* com as medidas de acurácia, AUC ROC, sensibilidade e especificidade.

Quadro 3 - Parâmetros do método *wrapper* utilizados no experimento 4

Parâmetro	Valor
Estado Inicial	Vetor de características vazio
Método de busca	<i>BestFirst</i>
Direção	<i>Forward</i>
Critério de parada	5 expansões consecutivas sem melhora de performance
Métrica de avaliação	AUC ROC

3.4.5 Experimento 5 - Rede Neural + *Wrapper*

Neste experimento, foi realizada uma análise mais aprofundada do algoritmo de classificação *Multilayer Perceptron* e do método de seleção de características *wrapper*. Todos os testes foram feitos no software WEKA.

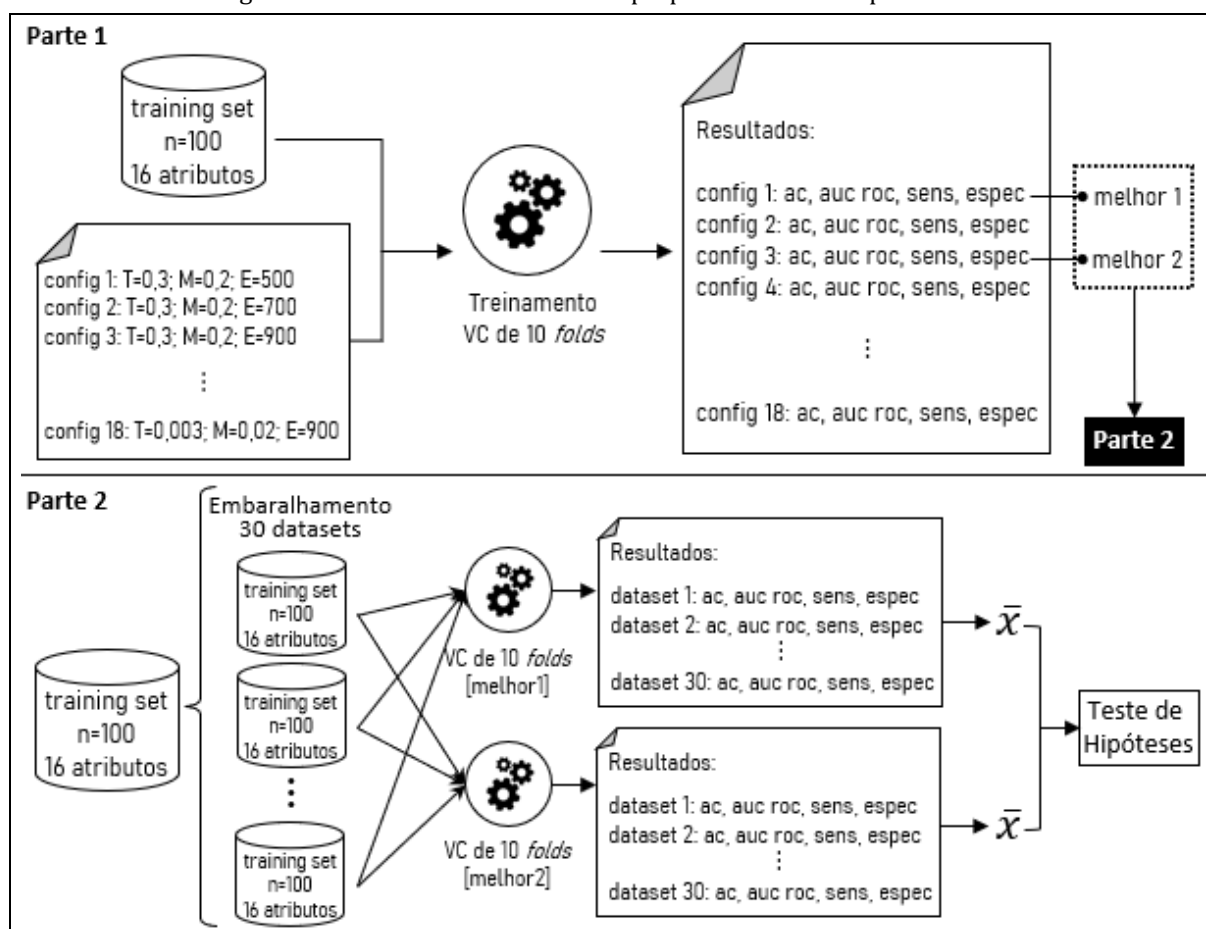
Inicialmente foi realizado um estudo dos hiperparâmetros do algoritmo, ou seja, variáveis que controlam o comportamento do algoritmo. Foram analisados os hiperparâmetros taxa de aprendizado, *momentum* e número de épocas, como mostra o Quadro 4.

Quadro 4 - Hiperparâmetros da rede neural estudados no experimento 5.

Hiperparâmetro	Valores Testados	Descrição
Taxa de aprendizado	0,3/0,03/0,003	Fator que está relacionado com a velocidade de aprendizado da rede. Valores muito pequenos tornam o aprendizado muito lento, enquanto valores muito altos impedem a convergência da rede.
Momentum	0,2/0,02	Fator usado para evitar oscilações na rede, reduzindo a instabilidade.
Épocas	500/700/900	É o número de vezes que todos os dados de treinamento são mostrados à rede durante o treinamento.

Para o estudo dos hiperparâmetros, o algoritmo MLP foi executado com o vetor de características resultante do experimento 4, contendo 16 atributos extraídos dos 100 corpos vertebrais fraturados, sendo 54 fraturas benignas e 46 fraturas benignas. Para avaliação, foi utilizada a técnica de validação cruzada de 10 *fold*s com as medidas de acurácia, AUC ROC, sensibilidade e especificidade. Após esse processo, as duas configurações de hiperparâmetros que obtiveram os melhores resultados foram comparadas por meio do teste de hipóteses. Para isso, o conjunto de treinamento foi embaralhando com a função *Resample*, presente no WEKA, e foram gerados 30 novos conjuntos de dados, os quais foram utilizados para testar a performance do MLP com as duas configurações de hiperparâmetros escolhidas. Finalmente, as médias das medidas de acurácia, AUC ROC, sensibilidade e especificidade foram comparadas por meio do teste de hipóteses. A Figura 20 ilustra o processo de estudo dos hiperparâmetros.

Figura 20 - Processo da análise de hiperparâmetros do experimento 5



Nota.— T = taxa de aprendizado; M = momentum; E = épocas; VC = validação cruzada; ac = acurácia; sens = sensibilidade; espec = especificidade. *Fonte:* da Autora.

Após a análise dos hiperparâmetros, foi realizado o estudo dos atributos. Para isso, o método *wrapper* foi executado 3 vezes com diferentes configurações. O Quadro 5 o Quadro 6 apresentam as configurações utilizadas nas duas primeiras execuções, que chamaremos de esquemas 1 e 2.

Quadro 5 - Método *wrapper* do experimento 5 (esquema 1)

Parâmetro	Valor
Estado Inicial	Vetor de características vazio (93 atributos)
Método de busca	<i>BestFirst</i>
Direção	<i>Forward</i>
Critério de parada	20 expansões consecutivas sem melhora de performance
Métrica de avaliação	AUC ROC

Quadro 6 - Método *wrapper* do experimento 5 (esquema 2)

Parâmetro	Valor
Estado Inicial	Vetor de características completo (Resultante do esquema 1)
Método de busca	<i>BestFirst</i>
Direção	<i>Backward</i>
Critério de parada	10 expansões consecutivas sem melhora de performance
Métrica de avaliação	AUC ROC

Durante a execução desses dois esquemas, foi utilizada a técnica de validação cruzada de 10 *folds* para avaliar o modelo. Após esse processo, os atributos resultantes foram ranqueados por meio do filtro ReliefF e avaliados com a validação cruzada de 10 *folds*. Os atributos que obtiveram um score médio $> 0,03$ foram utilizados como Estado Inicial na última execução do *wrapper*.

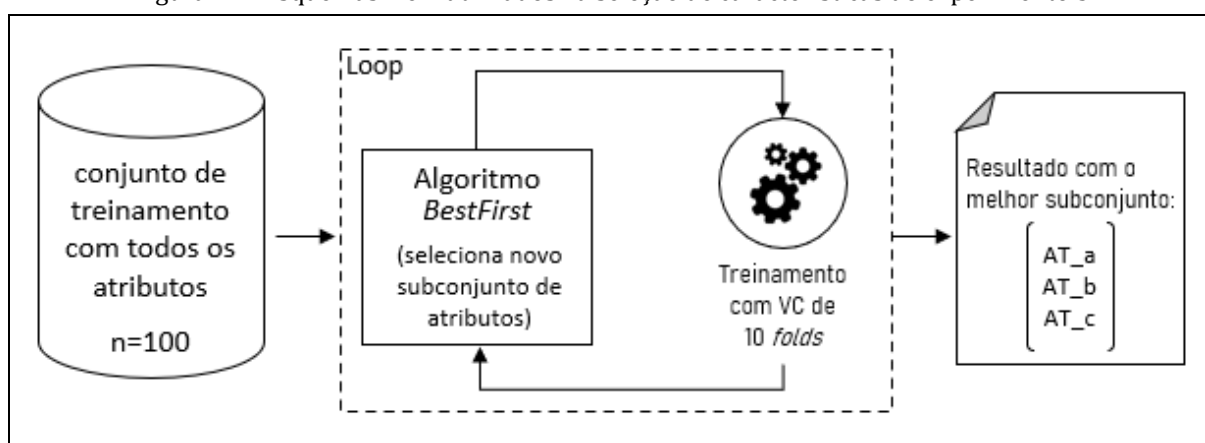
O esquema 3, por sua vez, foi configurado conforme mostra o Quadro 7.

Quadro 7 - Método *wrapper* do experimento 5 (esquema 3)

Parâmetro	Valor
Estado Inicial	Características que obtiveram score $> 0,03$ após filtro reliefF
Método de busca	<i>BestFirst</i>
Direção	<i>Forward</i>
Critério de parada	10 expansões consecutivas sem melhora de performance
Métrica de avaliação	AUC ROC

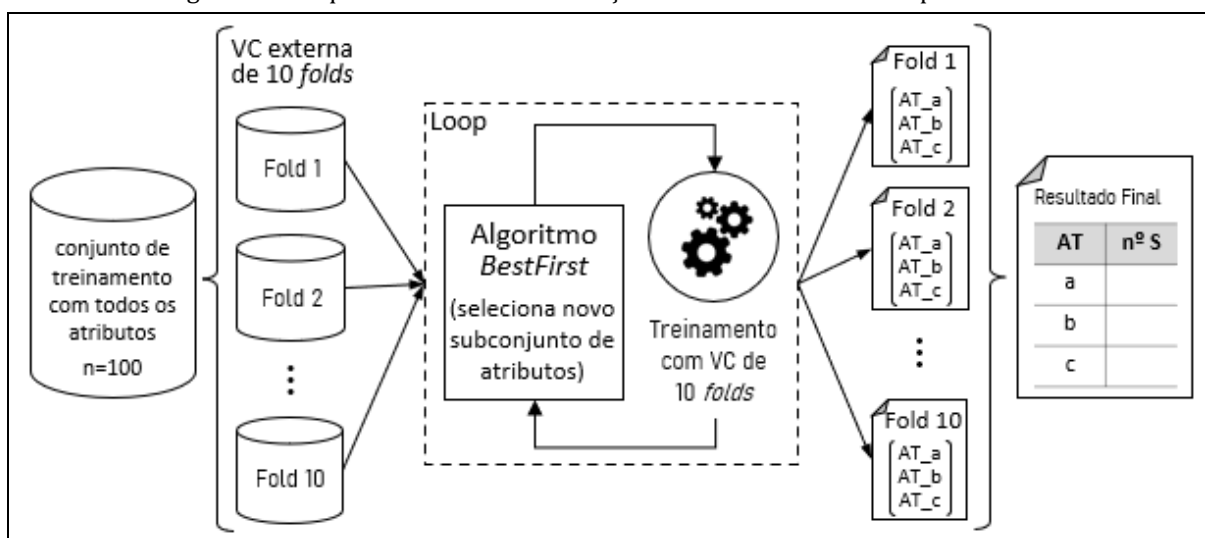
O design experimental dessa última configuração foi diferente das duas primeiras, pois foi adicionada uma etapa de validação cruzada externa. Isso significa que a validação cruzada de 10 *folds* não foi utilizada apenas para avaliar o modelo, mas também para particionar o conjunto inicial, de modo que o *wrapper* fosse executado para cada uma das 10 partições independentemente. Após a análise, foi gerada uma tabela contendo todas as características analisadas e o número de vezes que cada uma delas foi selecionada no total de partições. A Figura 21 ilustra o processo utilizado nos esquemas 1 e 2, e a Figura 22 o processo utilizado no esquema 3.

Figura 21 - Esquemas 1 e 2 utilizados na seleção de características do experimento 5



Nota.— VC = validação cruzada; AT = atributo. *Fonte:* da Autora.

Figura 22 - Esquema 3 utilizado na seleção de características do experimento 5



Nota.— VC = validação cruzada; AT = atributo; nº S = número de vezes em que o atributo em questão foi selecionado. *Fonte:* da Autora.

Após a finalização da seleção de atributos, foi realizado um teste de validação interna e um teste de validação externa. Para a validação interna, os dados originais pré-configurados com os atributos selecionados foram randomizados 30 vezes, de modo que cada novo conjunto de dados foi utilizado como entrada para a execução do modelo MLP. Foram, portanto, treinados 30 modelos diferentes, que foram testados com a validação cruzada de 10 *folds*. A performance geral foi avaliada utilizando as médias das medidas de acurácia, AUC ROC, sensibilidade e especificidade

O teste validação externa, por sua vez, teve como objetivo avaliar a performance do melhor modelo resultante do teste de validação interna. Para isso, foi utilizado um conjunto de teste independente, o qual não havia sido utilizado em nenhum momento previamente. Os dados desse conjunto foram então classificados em uma única execução e as medidas de acurácia, AUC ROC, sensibilidade e especificidade foram computadas.

4 RESULTADOS

4.1 ANÁLISE CLÍNICA E DEMOGRÁFICA

As informações demográficas e clínicas dos pacientes incluídos no estudo são mostradas da Tabela 1.

Tabela 1 - Características demográficas e clínicas dos pacientes incluídos no estudo

Parâmetro	Grupo de Treinamento	Grupo de Teste	Valor p
Nº de pacientes	61	30	
Idade	63.2±12.5	66.4±9.9	0,19
Sexo*			
Masculino	25/61 (41)	11/30 (37)	0,69
Feminino	36/61 (59)	19/30 (63)	0,83
Nº de pacientes por diagnóstico*			
Osteoporose (Fraturas Benignas)	32/61 (52)	15/30 (50)	0,83
Câncer (Fraturas Malignas)	29/61 (48)	15/30 (50)	0,83
Mieloma Múltiplo	12/29 (41)	3/15 (20)	...
Mama	8/29 (28)	3/15 (20)	...
Pulmão	2/29 (7)	1/15 (7)	...
CEC (orofaríngeo/epiglote/laringe)	2/29 (7)	2/15 (0)	...
Próstata	1/29 (3)	1/15 (7)	...
Leucemia Miofibroblástica Inflamatória	1/29 (3)	0/15 (0)	...
Colangiocarcinoma	1/29 (3)	0/15 (0)	...
Leucemia Linfocítica Crônica	1/29 (3)	0/15 (0)	...
Paraganglioma Metastático	1/29 (3)	0/15 (0)	...
Tireoide	0/29 (0)	1/15 (7)	...
Testículo	0/29 (0)	1/15 (7)	...
Bexiga	0/29 (0)	1/15 (7)	...
Rim	0/29 (0)	1/15 (7)	...
Linfoma Não-Hodgkin	0/29 (0)	1/15 (7)	...
Nº de pacientes por nº de FVCs lombares*			
1 Vértebra fraturada	39/61 (64)	22/30 (73)	0,35
2 Vértebras fraturadas	12/61 (20)	4/30 (13)	0,43
3 Vértebras fraturadas	6/61 (10)	1/30 (3)	0,20
4 Vértebras fraturadas	1/61 (2)	2/30 (7)	0,30
5 Vértebras fraturadas	3/61 (5)	1/30 (3)	0,71
Nº de vértebras fraturadas	100	46	
Nº de vértebras fraturadas usadas no ML	100	30	
Nº de vértebras fraturadas por classe*			
Fraturas Benignas	54/100 (54)	15/30 (50)	0,70
Fraturas Malignas	46/100 (46)	15/30 (50)	0,70
Nº de vértebras fraturadas por tipo*			
L1	29/100 (29)	6/30 (20)	0,30
L2	19/100 (19)	6/30 (20)	0,90
L3	19/100 (19)	5/30 (17)	0,77
L4	19/100 (19)	9/30 (30)	0,23
L5	14/100 (14)	4/30 (13)	0,93

Note.— CEC = carcinoma espinocelular; ML = *machine learning*; L = lombar; FVC = fratura vertebral por compressão. * Dados em parênteses são porcentagens.

Dos 91 pacientes incluídos neste estudo, 47 (52%) foram diagnosticados com fraturas secundárias à osteoporose, conhecidas como FVCs benignas e 44 (48%) foram diagnosticadas com fraturas secundárias ao câncer, chamadas de FVCs malignas. O tipo de câncer mais prevalente no conjunto de dados foi o mieloma múltiplo, com 15 casos (16,5%), seguido pelo câncer de mama com 11 casos (12%).

O grupo de treinamento continha 100 FVCs provenientes de 61 pacientes, das quais 54 eram fraturas benignas e 46 eram fraturas malignas. O grupo de teste, por sua vez, continha 30 FVCs provenientes de 30 pacientes, das quais 15 eram fraturas benignas e 15 eram fraturas malignas. Vale ressaltar que o número de fraturas identificadas no conjunto de teste totalizava 46, no entanto, optamos pelo uso de apenas uma fratura por paciente com a finalidade de evitar viés do resultado. Para os pacientes que possuíam mais de um corpo vertebral fraturado, a função *python random.choice* foi utilizada para escolher a vértebra que seria inserida no conjunto de teste.

Não houve diferença estatisticamente significativa entre os grupos de treinamento e teste para nenhum dos parâmetros informados na Tabela 1.

4.2 BASE DE DADOS PARA PESQUISA REPRODUTÍVEL

O armazenamento e compartilhamento dos dados do estudo é uma das ideias preconizadas pela radiômica. No presente trabalho, foi realizada uma organização de toda a base de dados, contendo não só os exames de RM originais, mas também as imagens processadas, os arquivos com as segmentações, o documento com as informações de aquisição das imagens, a planilha com as informações clínicas dos pacientes e as características extraídas das imagens. É importante destacar que todas as informações seguiram os padrões de anonimização conforme as diretrizes da Lei Geral de Proteção de Dados Pessoais - LGPD.

Até o momento, todo esse material está disponível no servidor de estudos radiômicos do Centro de Ciências das Imagens e Física Médica do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto (CCIFM-HCFMRP/USP). No entanto, após a conclusão da escrita do artigo que está sendo desenvolvido sobre a base, pretendemos disponibilizá-la publicamente. Essa disponibilização foi previamente solicitada ao Comitê de Ética juntamente com o projeto e, portanto, já está autorizada.

4.3 EXPERIMENTOS

4.3.1 Experimento 1 - Análise Univariada

Neste experimento foi realizada uma análise univariada das características quantitativas para investigar se alguma delas seria capaz de diferenciar as fraturas vertebrais por compressão benignas das malignas.

De acordo com os testes de hipóteses (teste-t para as variáveis com distribuição normal e Mann Whitney para as que não possuíam distribuição normal), das 93 variáveis quantitativas, 64 (68,8%) obtiveram diferença estatisticamente significativa, considerando $\alpha = 0,1\%$. A Tabela 2 apresenta os resultados dos testes indicando os valores p. O resultado do teste de normalidade está disponível no APÊNDICE A.

Tabela 2 - Resultado dos testes de hipótese para diferenciação de fraturas benignas e malignas

Característica Quantitativa	Média ± Desvio Padrão		valor p
	Benigna	Maligna	
firstorder_10Percentile	0.40 ± 0.21	0.24 ± 0.11	< 0,001
firstorder_90Percentile	0.49 ± 0.20	0.24 ± 0.14	< 0,001
firstorder_Energy	0.39 ± 0.23	0.39 ± 0.21	0.988
firstorder_Entropy	0.57 ± 0.19	0.38 ± 0.24	< 0,001
firstorder_InterquartileRange	0.42 ± 0.19	0.30 ± 0.20	0.002
firstorder_Kurtosis	0.39 ± 0.21	0.29 ± 0.12	0.011
firstorder_Maximum	0.56 ± 0.18	0.36 ± 0.18	< 0,001
firstorder_Mean	0.46 ± 0.21	0.23 ± 0.10	< 0,001
firstorder_MeanAbsoluteDeviation	0.46 ± 0.20	0.30 ± 0.22	< 0,001
firstorder_Median	0.47 ± 0.23	0.22 ± 0.11	< 0,001
firstorder_Minimum	0.58 ± 0.20	0.63 ± 0.20	0.196
firstorder_Range	0.62 ± 0.19	0.42 ± 0.21	< 0,001
firstorder_RobustMeanAbsoluteDeviation	0.43 ± 0.19	0.29 ± 0.20	0.002
firstorder_RootMeanSquared	0.47 ± 0.21	0.22 ± 0.11	< 0,001
firstorder_Skewness	0.54 ± 0.19	0.66 ± 0.13	< 0,001
firstorder_StandardDeviation	0.49 ± 0.21	0.31 ± 0.22	< 0,001
firstorder_TotalEnergy	0.39 ± 0.23	0.39 ± 0.21	0.988
firstorder_Uniformity	0.40 ± 0.18	0.58 ± 0.25	< 0,001
firstorder_Variance	0.42 ± 0.22	0.26 ± 0.21	< 0,001
glcm_Autocorrelation	0.50 ± 0.23	0.26 ± 0.18	< 0,001
glcm_ClusterProminence	0.30 ± 0.21	0.17 ± 0.17	< 0,001
glcm_ClusterShade	0.63 ± 0.15	0.71 ± 0.07	0.002
glcm_ClusterTendency	0.38 ± 0.19	0.24 ± 0.20	< 0,001
glcm_Contrast	0.41 ± 0.19	0.28 ± 0.16	< 0,001
glcm_Correlation	0.55 ± 0.17	0.45 ± 0.27	0.040
glcm_DifferenceAverage	0.49 ± 0.18	0.35 ± 0.17	< 0,001
glcm_DifferenceEntropy	0.52 ± 0.19	0.37 ± 0.18	< 0,001
glcm_DifferenceVariance	0.43 ± 0.20	0.27 ± 0.16	< 0,001
glcm_Id	0.45 ± 0.17	0.58 ± 0.18	< 0,001
glcm_Idm	0.47 ± 0.18	0.60 ± 0.18	< 0,001
glcm_Idmn	0.69 ± 0.12	0.60 ± 0.19	0.010
glcm_Idn	0.55 ± 0.15	0.50 ± 0.20	0.126

glcm_Imc1	0.63 ± 0.15	0.66 ± 0.26	0.207
glcm_Imc2	0.58 ± 0.16	0.48 ± 0.27	0.021
glcm_InverseVariance	0.78 ± 0.13	0.71 ± 0.21	0.094
glcm_JointAverage	0.57 ± 0.22	0.32 ± 0.20	< 0,001
glcm_JointEnergy	0.35 ± 0.18	0.52 ± 0.24	< 0,001
glcm_JointEntropy	0.55 ± 0.20	0.36 ± 0.23	< 0,001
glcm_MaximumProbability	0.39 ± 0.19	0.53 ± 0.23	0.002
glcm_SumAverage	0.57 ± 0.22	0.32 ± 0.20	< 0,001
glcm_SumEntropy	0.56 ± 0.18	0.38 ± 0.24	< 0,001
glcm_SumSquares	0.40 ± 0.21	0.25 ± 0.20	< 0,001
gldm_DependenceEntropy	0.51 ± 0.16	0.40 ± 0.23	0.008
gldm_DependenceNonUniformity	0.31 ± 0.19	0.48 ± 0.22	< 0,001
gldm_DependenceNonUniformityNormalized	0.39 ± 0.16	0.31 ± 0.16	0.010
gldm_DependenceVariance	0.43 ± 0.17	0.52 ± 0.20	0.020
gldm_GrayLevelNonUniformity	0.29 ± 0.16	0.49 ± 0.20	< 0,001
gldm_GrayLevelVariance	0.42 ± 0.21	0.26 ± 0.21	< 0,001
gldm_HighGrayLevelEmphasis	0.51 ± 0.22	0.27 ± 0.18	< 0,001
gldm_LargeDependenceEmphasis	0.42 ± 0.15	0.54 ± 0.17	< 0,001
gldm_LargeDependenceHighGrayLevelEmphasis	0.43 ± 0.20	0.27 ± 0.17	< 0,001
gldm_LargeDependenceLowGrayLevelEmphasis	0.19 ± 0.14	0.35 ± 0.18	< 0,001
gldm_LowGrayLevelEmphasis	0.24 ± 0.21	0.46 ± 0.27	< 0,001
gldm_SmallDependenceEmphasis	0.26 ± 0.15	0.17 ± 0.08	< 0,001
gldm_SmallDependenceHighGrayLevelEmphasis	0.27 ± 0.17	0.13 ± 0.09	< 0,001
gldm_SmallDependenceLowGrayLevelEmphasis	0.22 ± 0.14	0.34 ± 0.27	0.061
grlm_GrayLevelNonUniformity	0.30 ± 0.16	0.49 ± 0.20	< 0,001
grlm_GrayLevelNonUniformityNormalized	0.38 ± 0.19	0.57 ± 0.24	< 0,001
grlm_GrayLevelVariance	0.47 ± 0.23	0.28 ± 0.21	< 0,001
grlm_HighGrayLevelRunEmphasis	0.52 ± 0.22	0.29 ± 0.20	< 0,001
grlm_LongRunEmphasis	0.33 ± 0.14	0.46 ± 0.18	< 0,001
grlm_LongRunHighGrayLevelEmphasis	0.44 ± 0.20	0.28 ± 0.20	< 0,001
grlm_LongRunLowGrayLevelEmphasis	0.17 ± 0.14	0.34 ± 0.19	< 0,001
grlm_LowGrayLevelRunEmphasis	0.24 ± 0.21	0.45 ± 0.28	< 0,001
grlm_RunEntropy	0.48 ± 0.17	0.41 ± 0.27	0.136
grlm_RunLengthNonUniformity	0.35 ± 0.20	0.47 ± 0.24	0.008
grlm_RunLengthNonUniformityNormalized	0.45 ± 0.15	0.33 ± 0.13	< 0,001
grlm_RunPercentage	0.50 ± 0.15	0.39 ± 0.15	< 0,001
grlm_RunVariance	0.32 ± 0.14	0.44 ± 0.19	< 0,001
grlm_ShortRunEmphasis	0.51 ± 0.15	0.39 ± 0.15	< 0,001
grlm_ShortRunHighGrayLevelEmphasis	0.48 ± 0.22	0.26 ± 0.18	< 0,001
grlm_ShortRunLowGrayLevelEmphasis	0.24 ± 0.21	0.44 ± 0.30	< 0,001
glszm_GrayLevelNonUniformity	0.36 ± 0.20	0.52 ± 0.25	< 0,001
glszm_GrayLevelNonUniformityNormalized	0.28 ± 0.17	0.40 ± 0.23	0.011
glszm_GrayLevelVariance	0.46 ± 0.19	0.38 ± 0.23	0.018
glszm_HighGrayLevelZoneEmphasis	0.51 ± 0.17	0.35 ± 0.21	< 0,001
glszm_LargeAreaEmphasis	0.29 ± 0.18	0.49 ± 0.22	< 0,001
glszm_LargeAreaHighGrayLevelEmphasis	0.33 ± 0.20	0.40 ± 0.22	0.114
glszm_LargeAreaLowGrayLevelEmphasis	0.17 ± 0.14	0.39 ± 0.23	< 0,001
glszm_LowGrayLevelZoneEmphasis	0.16 ± 0.13	0.34 ± 0.27	< 0,001
glszm_SizeZoneNonUniformity	0.42 ± 0.25	0.40 ± 0.19	0.939
glszm_SizeZoneNonUniformityNormalized	0.42 ± 0.24	0.27 ± 0.16	0.001
glszm_SmallAreaEmphasis	0.49 ± 0.22	0.32 ± 0.15	< 0,001
glszm_SmallAreaHighGrayLevelEmphasis	0.51 ± 0.23	0.30 ± 0.19	< 0,001
glszm_SmallAreaLowGrayLevelEmphasis	0.17 ± 0.12	0.31 ± 0.24	0.004
glszm_ZoneEntropy	0.58 ± 0.20	0.63 ± 0.20	0.195
glszm_ZonePercentage	0.17 ± 0.15	0.11 ± 0.06	0.004
glszm_ZoneVariance	0.29 ± 0.18	0.48 ± 0.22	< 0,001
ngtdm_Busyness	0.16 ± 0.10	0.41 ± 0.22	< 0,001
ngtdm_Coarseness	0.34 ± 0.19	0.19 ± 0.15	< 0,001
ngtdm_Complexity	0.44 ± 0.19	0.26 ± 0.16	< 0,001

ngtdm_Contrast	0.48 ± 0.22	0.46 ± 0.16	0.748
ngtdm_Strength	0.34 ± 0.20	0.14 ± 0.11	< 0,001

A Tabela 3 apresenta a análise da diferença das proporções de cada família de atributos entre o grupo original e o grupo composto pelas características que obtiveram valor $p < 0,001$ no teste de hipóteses.

Tabela 3 - Diferença entre as proporções de cada família de atributos comparando o grupo total de características com o grupo selecionado ($p < 0,001$)

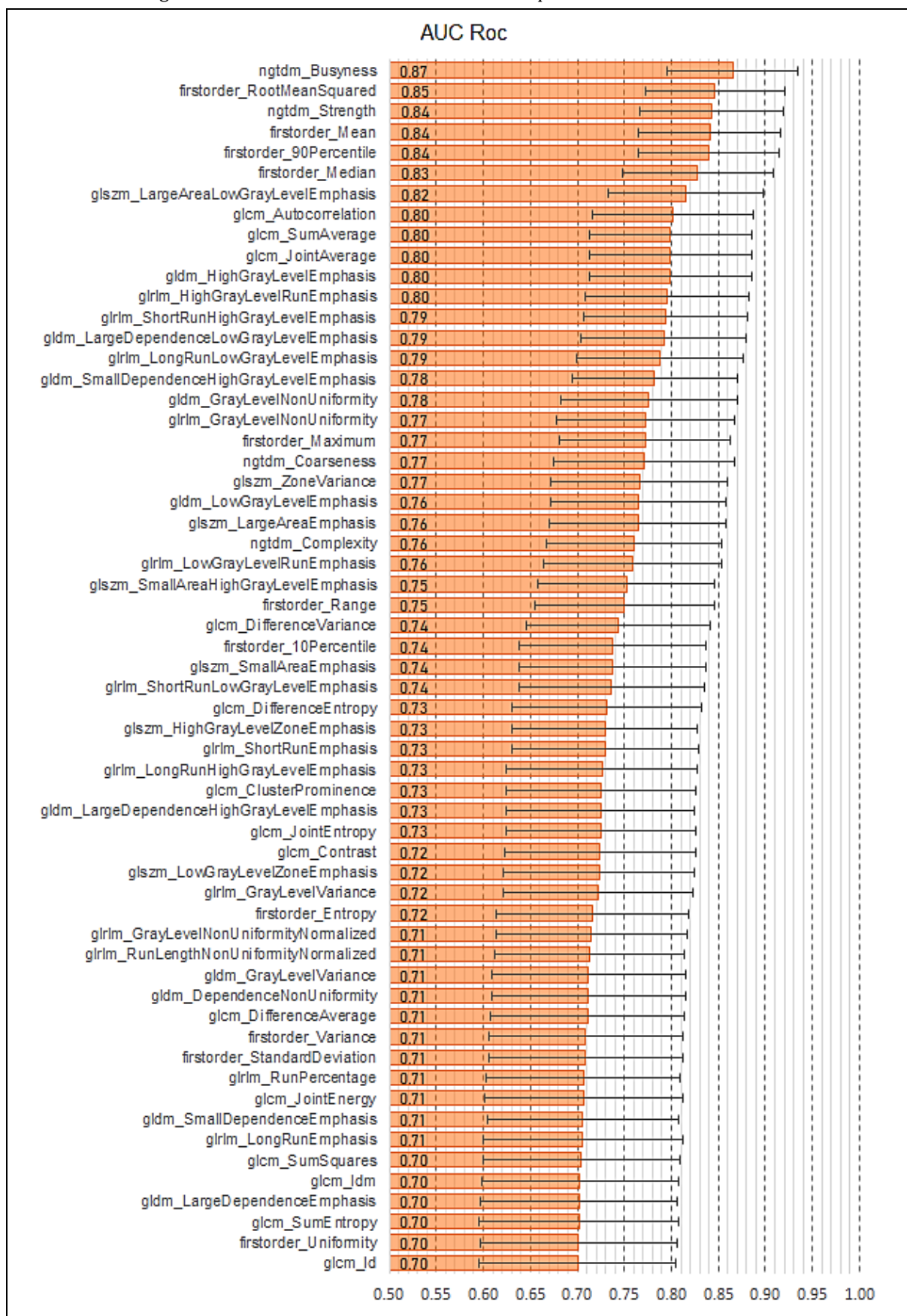
Família de atributos	Original (n = 93)	Teste de Hipótese valor $p < 0,001$ (n = 64)	Diferença entre as proporções (valor p)
Primeira ordem	19/93 (20.4%)	13/64 (20,3%)	0,98
Textura	74/93 (79.6%)	51/64 (79,7%)	0,98
GLCM	23/93 (24.7%)	15/64 (23,4%)	0,85
GLDM	14/93 (15.1%)	10/64 (15,6%)	0,92
GLRLM	16/93 (17.2%)	14/64 (21,9%)	0,47
GLSZM	16/93 (17.2%)	8/64 (12,5%)	0,41
NGTDM	5/93 (5.4%)	4/64 (6,3%)	0,82

Nota.— GLCM = grey level cooccurrence matrix; GLDM = grey level dependence matrix; GLRLM = grey level run length matrix; GLSZM = grey level size zone matrix; NGTDM = neighboring gray tone difference matrix

De acordo com a Tabela 3, não houve diferença estatisticamente significativa entre o grupo original e o com as características selecionadas com relação às famílias de atributos, ou seja, de acordo com o teste de hipóteses, todas as famílias tiveram atributos considerados relevantes para diferenciar fraturas benignas de malignas quando analisados isoladamente.

Outra análise realizada neste experimento foi o cálculo da área sob a curva ROC para cada uma das características quantitativas. A Figura 23 mostra o gráfico de barras com os valores das áreas sob a curva ROC e seus respectivos intervalos de confiança de 95% para os atributos que obtiveram uma AUC ROC $> 0,700$. Esse valor foi escolhido apenas por uma questão de exibição. Os resultados completos dessa análise estão disponíveis no APÊNDICE B. O cálculo da AUC ROC foi particularmente interessante pois, diferente do teste de hipóteses, é possível realizar uma ordenação dessas características e identificar quais obtiveram a melhor performance.

Figura 23 - Gráfico da AUC ROC dos atributos que obtiveram valor > 0.700



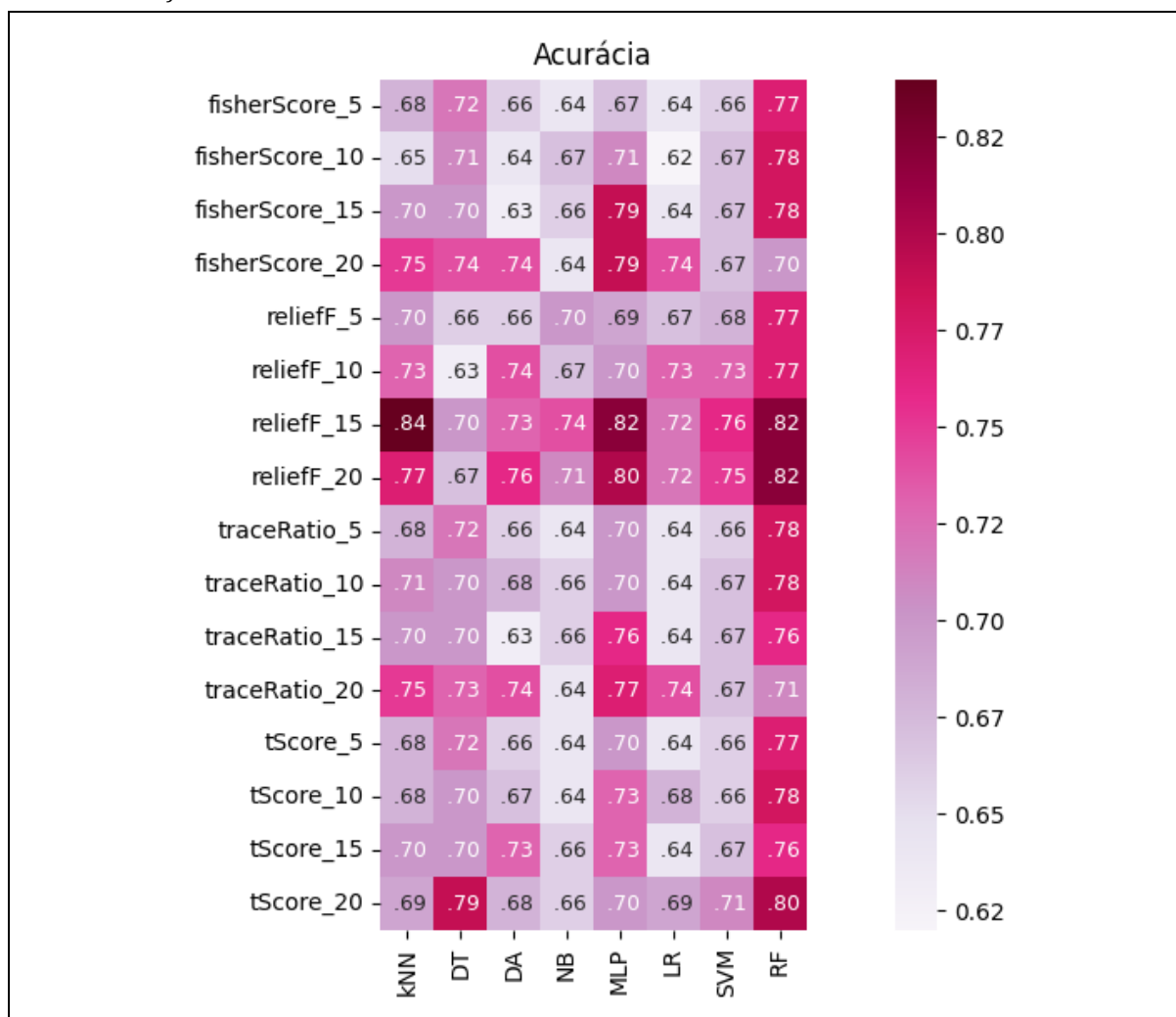
Fonte: da Autora.

4.3.2 Experimento 2 - PyRadiomics + Filtros

Neste experimento realizamos uma análise exploratória de quatro métodos de seleção de características baseados em filtros e de oito algoritmos de classificação pertencentes a diferentes famílias. Com relação aos métodos de seleção de atributos, utilizamos o *Fisher_Score*, *ReliefF*, *Trace_Ratio* e *t_Score*, selecionando as 5, 10, 15 e 20 características mais bem ranqueadas para cada um. O resultado das características ranqueadas utilizando os quatro filtros está disponível no APÊNDICE C.

Foram realizados, no total, 128 subexperimentos, cujos resultados são apresentados a seguir. A Figura 24, Figura 25, Figura 26 e Figura 27 mostram os resultados, respectivamente, para acurácia, área sob a curva ROC, sensibilidade (taxa de acerto para as fraturas malignas) e especificidade (taxa de acerto para as fraturas benignas) para cada um dos 128 subexperimentos.

Figura 24 - Acurácia alcançada em cada um dos subexperimentos do experimento 2 usando validação cruzada de 10 *folds*

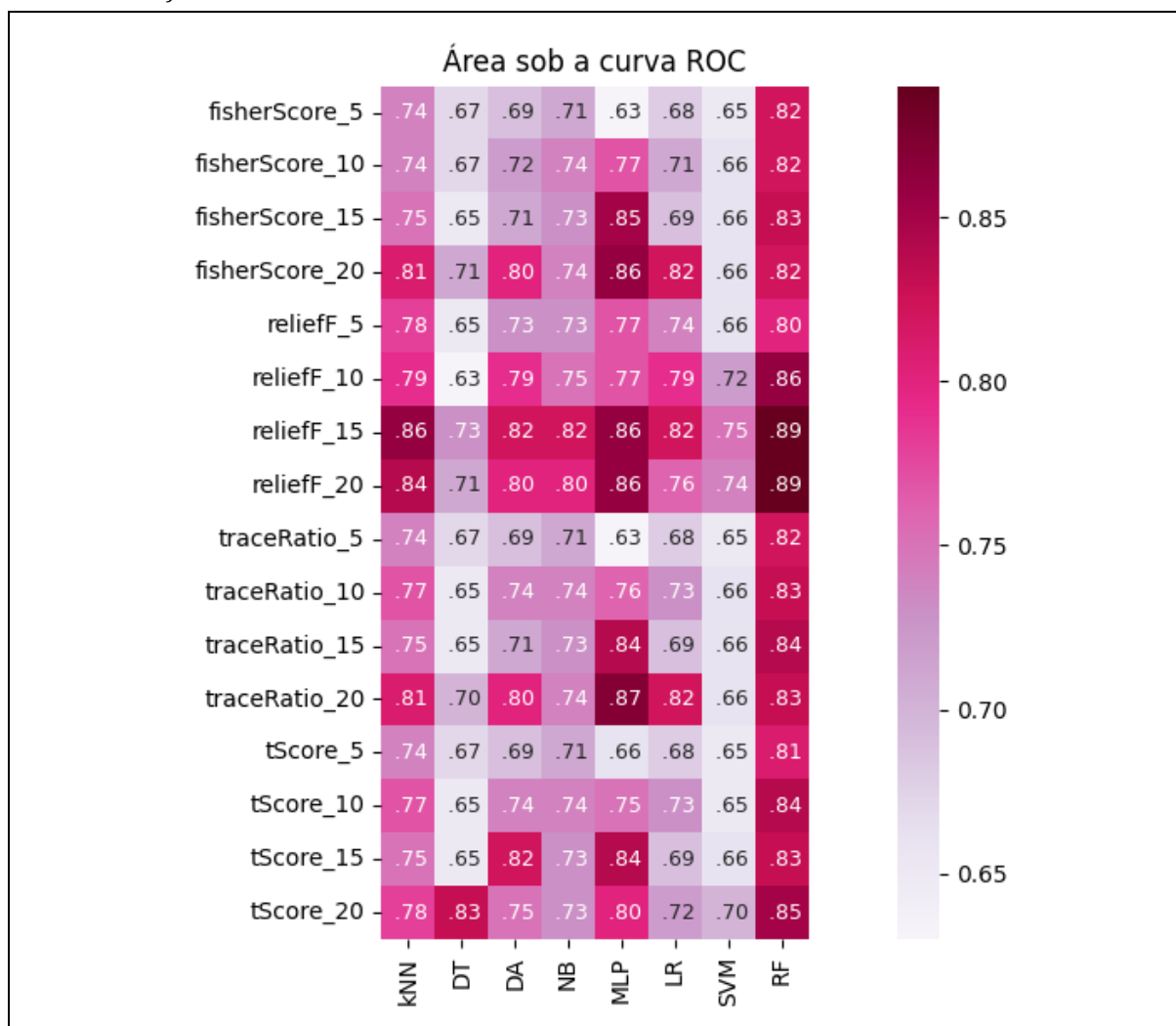


Nota.— KNN = k nearest neighbor; DT = decision tree; DA = discriminant analysis; NB = naïve bayes; MLP = multilayer perceptron; LR = logistic regression; SVM = support vector machine; RF = random forest.

Fonte: da Autora.

Com relação à acurácia, o algoritmo de classificação que apresentou melhor performance geral foi o *Random Forest*, alcançando uma acurácia de 82% com os vetores de características *reliefF_15* e *reliefF_20* (ou seja, os 15 e os 20 atributos mais bem ranqueados com o filtro *reliefF*). No entanto, a acurácia máxima foi obtida com o classificador kNN com $k=3$ e o *reliefF* com 15 atributos. A segunda melhor performance geral foi alcançada com a rede neural *Multilayer Perceptron*, atingindo os maiores valores também com os vetores de atributos *reliefF_15* e *reliefF_20*. Com relação aos métodos de seleção, o filtro *reliefF* foi o que apresentou os melhores resultados, principalmente com os vetores com 15 e com 20 atributos.

Figura 25 - AUC ROC alcançada em cada um dos subexperimentos do experimento 2 usando validação cruzada de 10 *folde*s

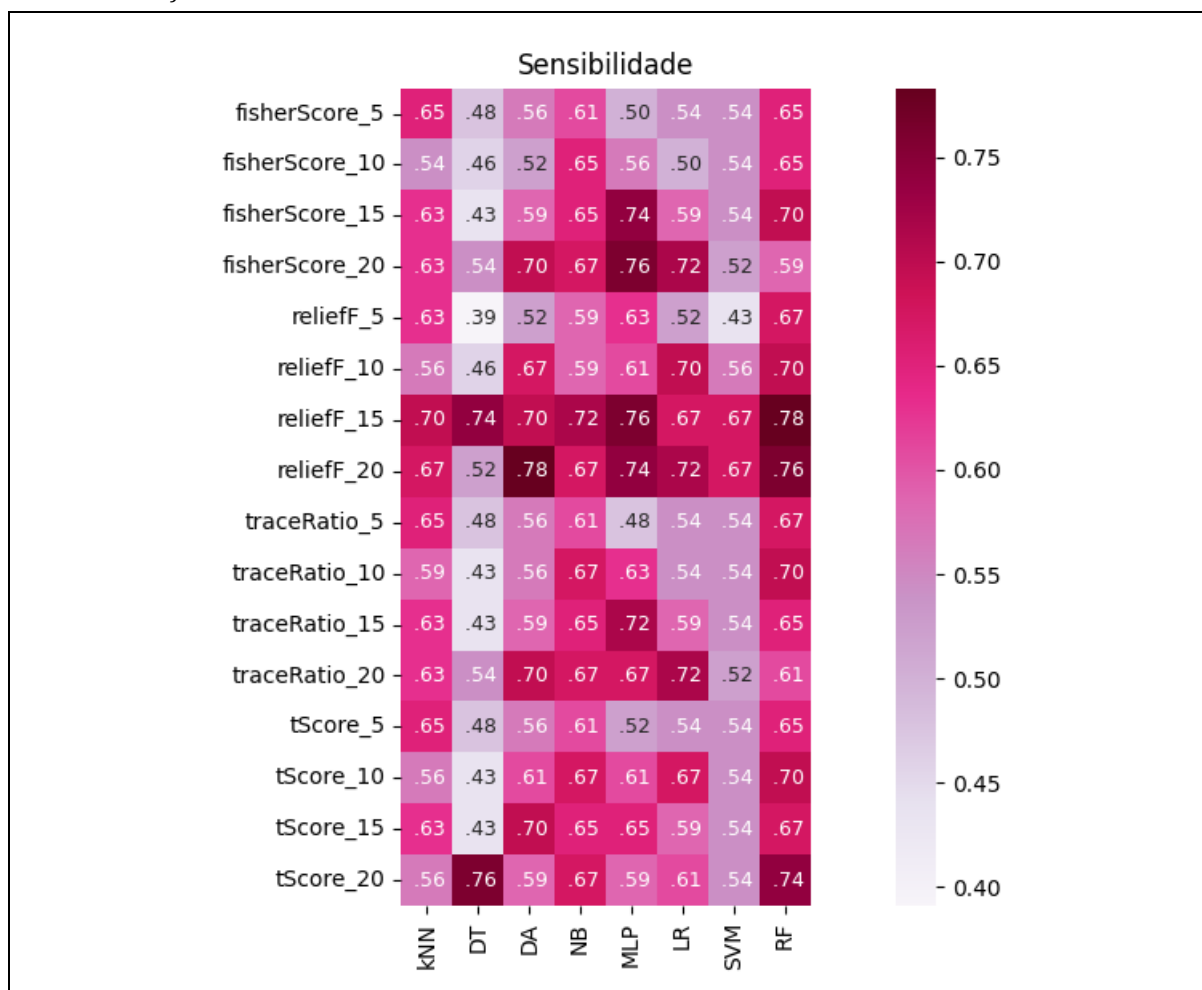


Nota.— KNN = k nearest neighbor; DT = decision tree; DA = discriminant analysis; NB = naïve bayes; MLP = multilayer perceptron; LR = logistic regression; SVM = support vector machine; RF = random forest.

Fonte: da Autora.

Os resultados considerando a área sob a curva ROC seguem o mesmo padrão dos apresentados para a acurácia. O algoritmo *Random Forest* apresentou os melhores resultados de maneira geral, independentemente do método de seleção, sendo o maior valor 0,89, usando os vetores reliefF_15 e reliefF_20. Em seguida, o melhor desempenho foi novamente da MLP, que inclusive superou o RF para alguns vetores de atributos em comparação direta, como, por exemplo, fisherScore_15, fisherScore_20 e traceRatio_20, este último alcançando uma AUC ROC de 0,87, enquanto RF alcançou 0,83. Com relação aos métodos de seleção de atributos, assim como para acurácia, os melhores desempenhos gerais foram dos vetores reliefF_15 e reliefF_20. SVM e DT foram os algoritmos que obtiveram o pior desempenho geral.

Figura 26 - Sensibilidade alcançada em cada um dos subexperimentos do experimento 2 usando validação cruzada de 10 *folds*

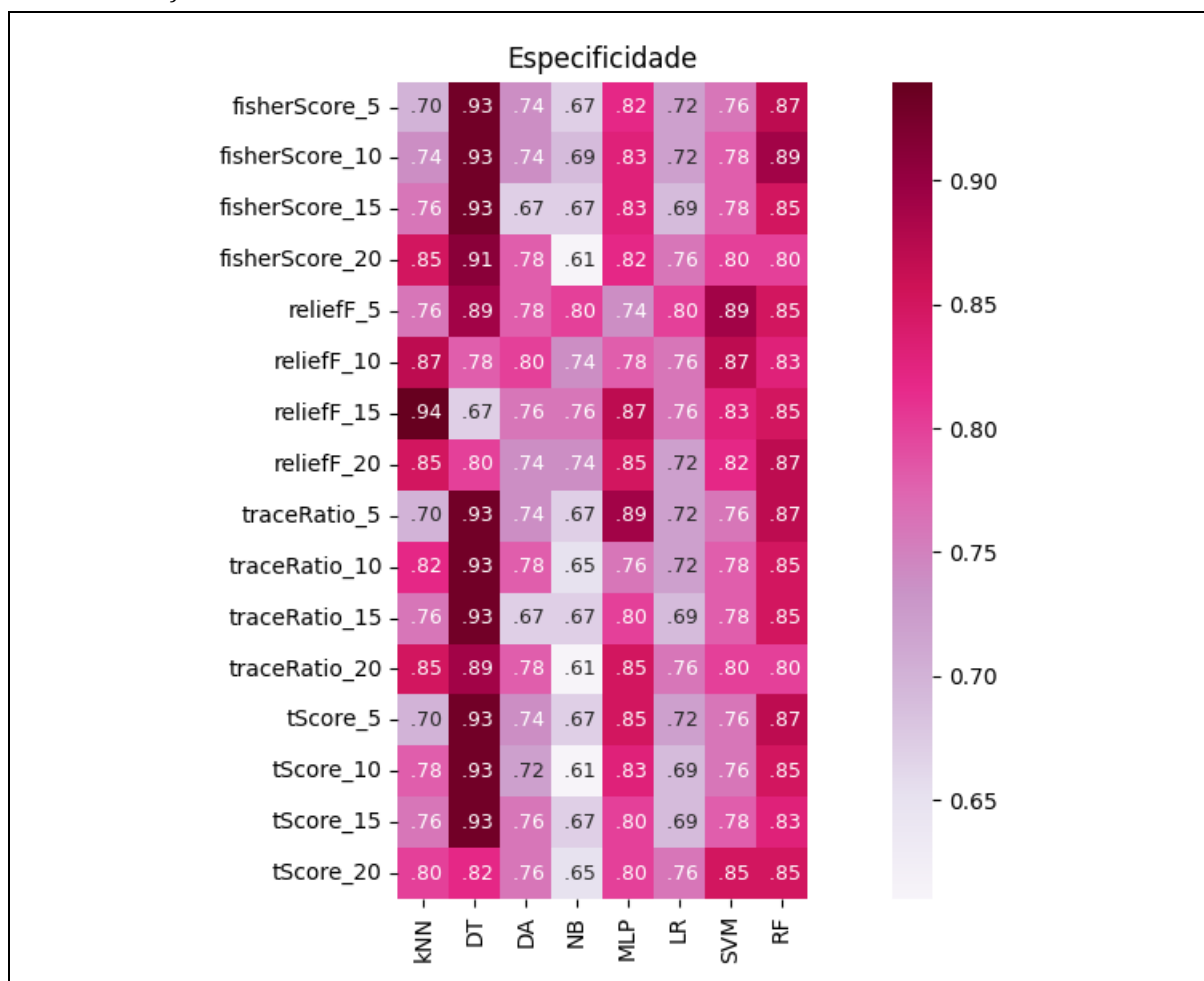


Nota.— KNN = k nearest neighbor; DT = decision tree; DA = discriminant analysis; NB = naïve bayes; MLP = multilayer perceptron; LR = logistic regression; SVM = support vector machine; RF = random forest.

Fonte: da Autora.

Com relação à medida de sensibilidade, ou a taxa de acerto para as fraturas malignas, o resultado geral foi inferior aos valores de acurácia e AUC ROC, sendo o valor máximo alcançado de 78%. Algumas informações que se destacam são o desempenho da Árvore de Decisão, que foi bem baixo no geral. No entanto, para dois vetores de características específicos (reliefF_15 e tScore_20), esse mesmo algoritmo atingiu o segundo melhor resultado, com uma sensibilidade de 76%. O algoritmo *Random Forest* permaneceu obtendo o melhor desempenho geral, com o valor máximo de sensibilidade de 78% com o vetor de características reliefF_15. O algoritmo MLP, embora não tenha tido o melhor desempenho geral, para alguns vetores de características (fisherScore_15, fisherScore_20 e traceRatio_15) foi o algoritmo que alcançou os melhores resultados. Vale destacar ainda que, juntamente com o algoritmo RF, DA atingiu a maior sensibilidade com um valor de 78% usando o vetor de características reliefF_20.

Figura 27 - Especificidade alcançada em cada um dos subexperimentos do experimento 2 usando validação cruzada de 10 *folde*s



Nota.— KNN = k nearest neighbor; DT = decision tree; DA = discriminant analysis; NB = naïve bayes; MLP = multilayer perceptron; LR = logistic regression; SVM = support vector machine; RF = random forest.

Fonte: da Autora.

Com relação à especificidade, ou taxa de acerto para as fraturas benignas, podemos fazer algumas observações. A melhor performance geral foi do algoritmo DT, com valores muito altos para quase todos os vetores de características. No entanto, o maior valor para especificidade foi alcançado pelo kNN e o vetor de atributos reliefF_15, atingindo um valor de 94%. A segunda melhor performance geral foi do algoritmo *Random Forest*, que obteve resultados consistentes para todos os vetores de características, sendo seguido pelo algoritmo MLP e depois pelo SVM. O algoritmo *Naïve Bayes* foi o que apresentou a pior performance geral. Com relação aos vetores de atributos, podemos dizer que os da família reliefF tiveram um desempenho geral um pouco melhor do que os demais. Um fator interessante é que o algoritmo DT, que obteve o melhor desempenho geral, alcançou os piores resultados justamente utilizando os vetores de características reliefF.

4.3.3 Experimento 3 - U-Net + Filtros

Neste experimento foram utilizadas características extraídas a partir da Rede Neural Convolutacional *U-Net*. Como essas características não possuem interpretação matemática, não faria sentido apresentá-las na forma de tabela.

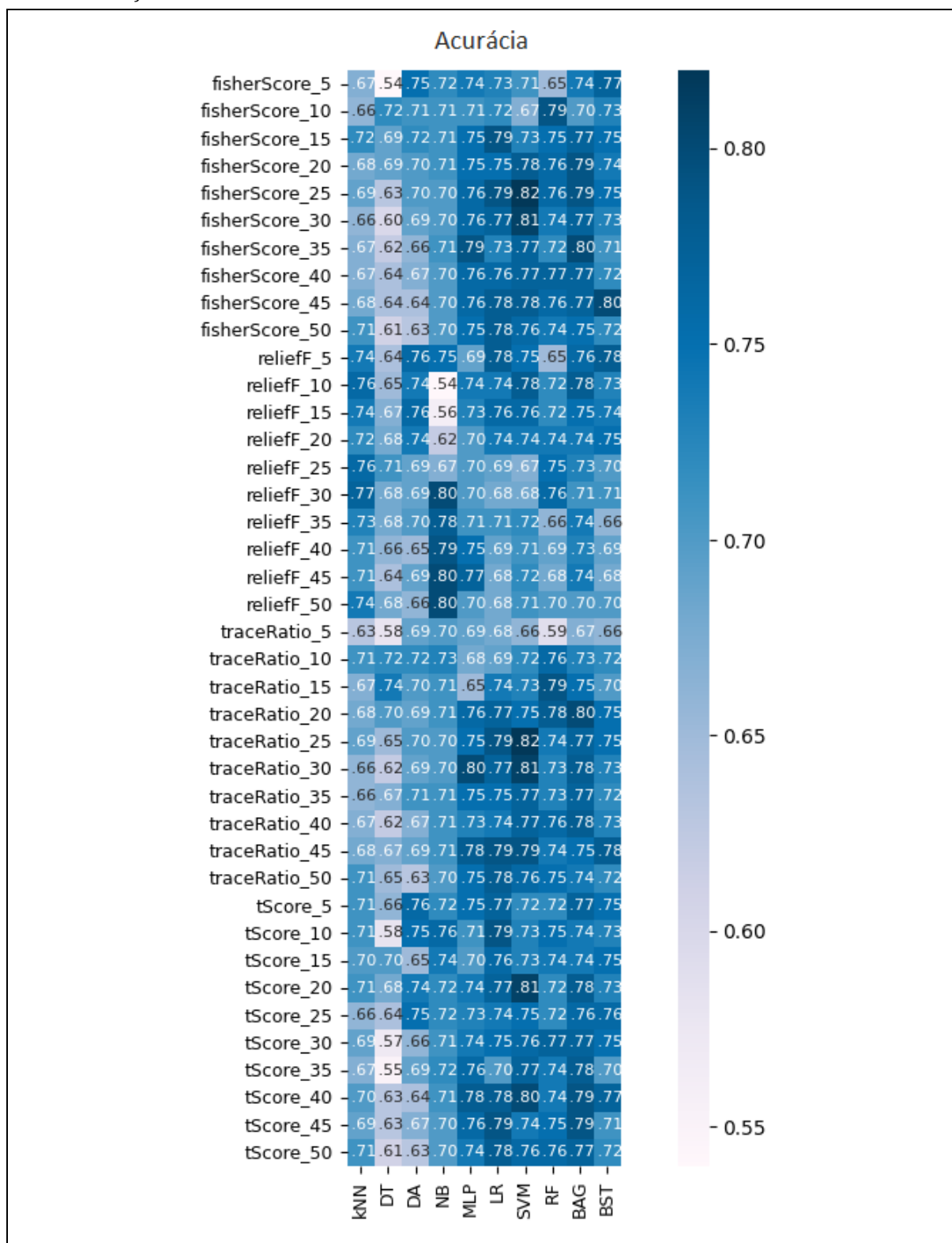
Conforme mencionado na seção de métodos, para este experimento utilizamos a função *GridSearchCV* para selecionar os melhores parâmetros em cada um dos algoritmos. Como vetor de características, foi utilizado o vetor original resultante do modelo *U-Net*, sem seleção de características prévia. O resultado da análise de hiperparâmetros, bem como o algoritmo utilizado para cada uma das famílias estão apresentados no Quadro 8.

Quadro 8 - Resultado da seleção de hiperparâmetros dos algoritmos de classificação

Classificador	Parâmetros selecionados pela função GridSearchCV
Nearest Neighbor	<code>KNeighborsClassifier(n_neighbors=9, algorithm='brute', p=2)</code>
Decision Tree	<code>DecisionTreeClassifier(criterion='gini', max_depth=None, min_sample_leaf=1)</code>
Discriminant Analysis	<code>LinearDiscriminantAnalysis(solver='svd')</code>
Naïve Bayes	<code>GaussianNB(priors=None, var_smoothing=1e-09)</code>
Neural Network	<code>MLPClassifier(activation='relu', batch_size=25, hidden_layer_sizes=(75,), learning_rate_init=0.1, max_iter=700, momentum=0.9, solver='sgd')</code>
Logistic Regression	<code>LogisticRegression(C=1000, solver='lbfgs')</code>
Support Vector Machine	<code>SVC(C=100, gamma=0.1, kernel='rbf')</code>
Random Forest	<code>RandomForestClassifier(bootstrap=True, max_features=None, n_estimators=100, warm_start=False)</code>
Bagging	<code>BaggingClassifier(base_estimator='SVC', n_estimators=50)</code>
Boosting	<code>AdaBoostClassifier(algorithm='SAMME', base_estimator='LogisticRegression', learning_rate=1.0, n_estimators=50)</code>

A Figura 28, Figura 29, Figura 30 e Figura 31 apresentam, respectivamente, os resultados para acurácia, área sob a curva ROC, sensibilidade e especificidade para os 400 subexperimentos realizados utilizando 10 algoritmos de classificação e 40 vetores de características provenientes de quatro filtros de seleção de atributos.

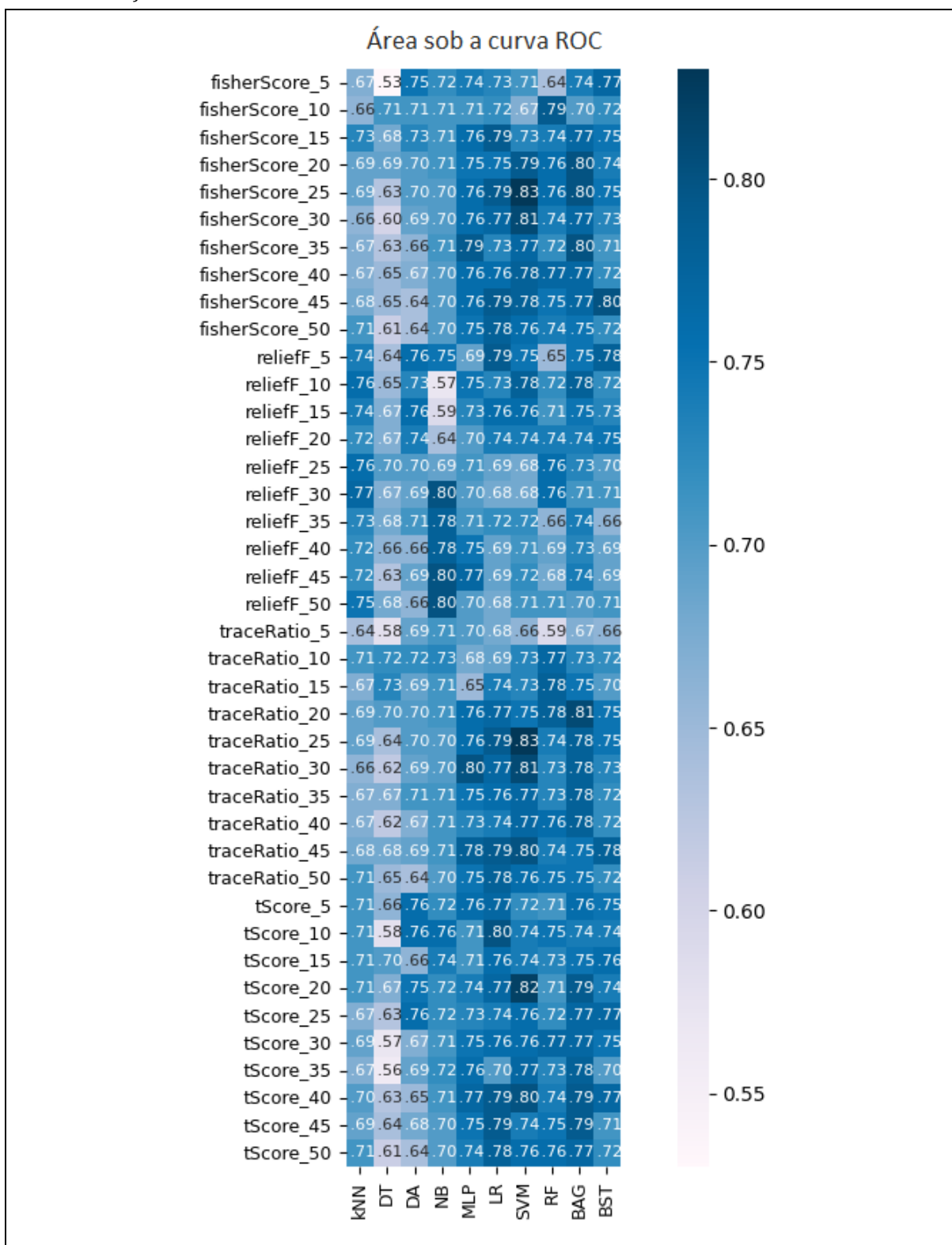
Figura 28 - Acurácia alcançada em cada um dos subexperimentos do experimento 3 utilizando a validação cruzada de 10 *folde*s



Nota.— KNN = k nearest neighbor; DT = decision tree; DA = discriminant analysis; NB = naïve bayes; MLP = multilayer perceptron; LR = logistic regression; SVM = support vector machine; RF = random forest; BAG = bagging; BST = boosting.

Fonte: da Autora.

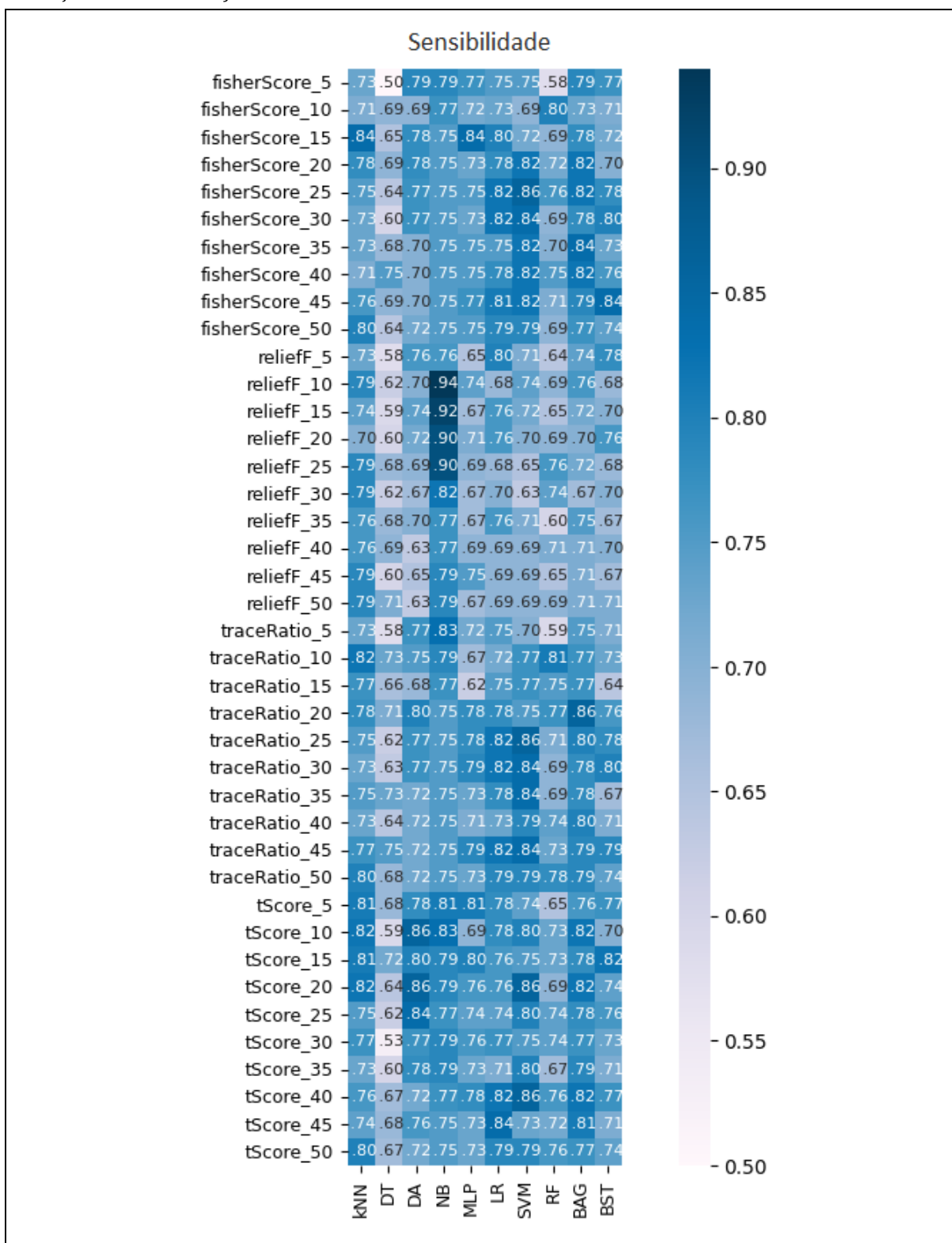
Figura 29 - AUC ROC alcançada em cada um dos subexperimentos do experimento 3 utilizando a validação cruzada de 10 *folds*



Nota.— KNN = k nearest neighbor; DT = decision tree; DA = discriminant analysis; NB = naïve bayes; MLP = multilayer perceptron; LR = logistic regression; SVM = support vector machine; RF = random forest; BAG = bagging; BST = boosting.

Fonte: da Autora.

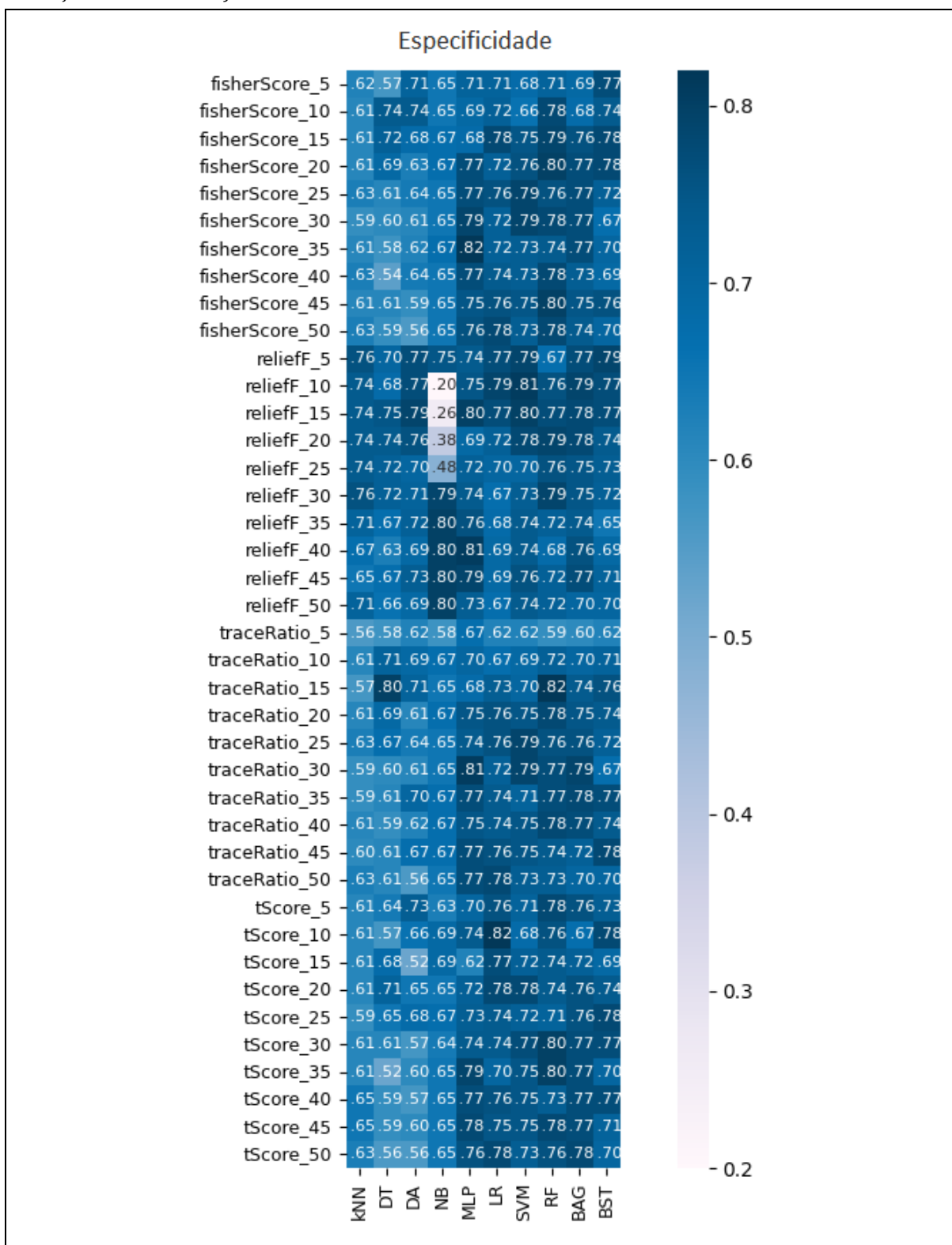
Figura 30 - Sensibilidade alcançada em cada um dos subexperimentos do experimento 3 utilizando a validação cruzada de 10 folds



Nota.— KNN = k nearest neighbor; DT = decision tree; DA = discriminant analysis; NB = naïve bayes; MLP = multilayer perceptron; LR = logistic regression; SVM = support vector machine; RF = random forest; BAG = bagging; BST = boosting.

Fonte: da Autora.

Figura 31 - Especificidade alcançada em cada um dos subexperimentos do experimento 3 utilizando a validação cruzada de 10 folds



Nota.— KNN = k nearest neighbor; DT = decision tree; DA = discriminant analysis; NB = naïve bayes; MLP = multilayer perceptron; LR = logistic regression; SVM = support vector machine; RF = random forest; BAG = bagging; BST = boosting.

Fonte: da Autora.

Com relação à acurácia, o melhor resultado foi alcançado pelo algoritmo SVM com os vetores `fisherScore_25` e `traceRatio_25`, atingindo um valor de 82%. Os algoritmos MLP, LR, SVM, RF, BAG e BST obtiveram os melhores desempenhos de maneira geral, no entanto, para a família de vetores de características `reliefF`, na faixa de 25 a 50 atributos, esses algoritmos tiveram um desempenho reduzido. Isso contrasta com o algoritmo NB, que obteve os melhores resultados dentre todos os algoritmos de classificação para esse mesmo grupo de vetores de características, alcançando uma acurácia de 80%.

Os resultados para a AUC ROC seguem o mesmo padrão da acurácia. O algoritmo SVM com os vetores `fisherScore_25` e `traceRatio_25` obteve o melhor resultado, alcançando uma AUC ROC de 0,83. Os vetores de características da família `reliefF`, na faixa de 25 a 50 atributos, também obtiveram os piores resultados de maneira geral, com exceção para o caso de sua utilização com o algoritmo NB, que novamente obteve os maiores valores justamente para essa faixa de atributos, alcançando uma AUC ROC de 0,80. Dentre os algoritmos com os piores desempenhos temos o kNN, DA, DT, sendo este último o que mostrou os piores resultados.

No caso da medida de sensibilidade, o algoritmo NB foi o que apresentou os melhores resultados com os vetores `reliefF` na faixa de 10 a 25, de modo que o maior valor foi de 94% para o vetor de atributos `reliefF_10`. Embora o algoritmo NB tenha apresentado os maiores valores de sensibilidade para um grupo pequeno de vetores de características, em termos de desempenho geral, o algoritmo SVM obteve os maiores resultados para vários vetores de características, atingindo uma sensibilidade de 86%. Ainda seguindo o padrão das outras medidas, o algoritmo DT foi o que apresentou os menores valores de sensibilidade.

Por fim, com relação à medida de especificidade, o maior destaque é do algoritmo NB, que apresentou os piores valores para as mesmas combinações em que apresentou os melhores resultados na análise de sensibilidade, ou seja, com os vetores `reliefF` na faixa de 10 a 25. Isso já era esperado, uma vez que os valores de sensibilidade estavam bem elevados e os da AUC ROC estavam baixos. A pior combinação foi do algoritmo NB com o vetor de atributos `reliefF_10`, que obteve uma especificidade de apenas 20%. Embora esse limite inferior tenha comprometido um pouco a interpretação visual do mapa de calor, ainda é possível observar que o vetor `traceRatio_5` foi o que apresentou os menores valores de especificidade geral. O maior valor de especificidade foi de 82%, e foi obtido por 3 combinações de algoritmos: *Multilayer Perceptron* com `fisherScore_35`, *Random*

gldm_LargeDependenceLowGrayLevelEmphasis	x	x							x			
gldm_SmallDependenceEmphasis						x	x					
gldm_SmallDependenceHighGrayLevelEmphasis			x		x	x						
gldm_SmallDependenceLowGrayLevelEmphasis									x			
glrlm_GrayLevelVariance											x	
glrlm_HighGrayLevelRunEmphasis				x								
glrlm_LongRunHighGrayLevelEmphasis												x
glrlm_LongRunLowGrayLevelEmphasis			x					x	x			
glrlm_RunEntropy		x								x		
glrlm_RunLengthNonUniformity										x		
glrlm_RunPercentage					x			x			x	
glrlm_RunVariance					x	x	x				x	
glrlm_ShortRunEmphasis				x					x			
glrlm_ShortRunLowGrayLevelEmphasis									x			
glszm_GrayLevelNonUniformity	x	x		x				x		x		
glszm_GrayLevelNonUniformityNormalized								x	x			
glszm_HighGrayLevelZoneEmphasis				x				x	x			
glszm_LargeAreaEmphasis									x			
glszm_LargeAreaLowGrayLevelEmphasis		x		x								
glszm_LowGrayLevelZoneEmphasis		x							x			
glszm_SizeZoneNonUniformity										x	x	
glszm_SmallAreaHighGrayLevelEmphasis								x				
glszm_SmallAreaLowGrayLevelEmphasis				x				x				
glszm_ZoneEntropy										x		
glszm_ZonePercentage						x	x					
ngtdm_Busyness			x					x	x	x		
ngtdm_Complexity							x					
ngtdm_Contrast	x	x								x		x
ngtdm_Strength		x	x				x		x			x
TOTAL DE ATRIBUTOS SELECIONADOS	6	10	7	7	8	10	9	14	16	13	15	7

Das 93 características extraídas das imagens, 58 (62,4%) foram selecionadas pelo menos por um algoritmo de classificação. Os atributos mais selecionados foram firstorder_10Percentile (7 seleções), glszm_GrayLevelNonUniformity (5 seleções), ngtdm_Strength (5 seleções), firstorder_Maximum (4 seleções), glcm_InverseVariance (4 seleções), glrlm_RunVariance (4 seleções), ngtdm_Busyness (4 seleções) e ngtdm_Contrast (4 seleções). A Tabela 4 apresenta a proporção de características selecionadas por família de atributos.

Tabela 4 - Proporção de características selecionadas por família de atributos

Família de atributos	Total	Selecionados	Proporção
Primeira Ordem	19	9	9/19 (47.4%)
Textura	74	49	49/75 (66,2%)
GLCM	23	16	16/23 (69.6%)
GLDM	14	8	8/14 (57.1%)
GLRLM	16	10	10/16 (62.5%)
GLSZM	16	11	11/16 (68.8%)
NGTDM	5	4	4/5 (80.0%)
TOTAL	93	58	58/93 (62,4%)

Nota.— GLCM = grey level cooccurrence matrix; GLDM = grey level dependence matrix; GLRLM = grey level run length matrix; GLSZM = grey level size zone matrix; NGTDM = neighboring gray tone difference matrix

Os atributos de textura foram mais selecionados proporcionalmente com relação aos atributos de primeira ordem. Dentre as subfamílias dos atributos de textura, os atributos provenientes da matriz de diferença de tons de cinza vizinhos (ngtdm) foram os mais selecionados proporcionalmente (80%), o que mostra que os algoritmos de classificação consideraram esses atributos relevantes para diferenciar fraturas benignas de malignas. Os atributos provenientes da matriz de dependência de níveis de cinza (gldm) foram os menos selecionados dentre as subfamílias de textura.

A distribuição das famílias de atributos entre o grupo original e o grupo com características selecionadas é mostrada na Tabela 5.

Tabela 5 - Distribuição das famílias de atributos entre o grupo original e o grupo das selecionadas

Família de atributos	Total (n=93)	Selecionados (n=58)	valor p
Primeira Ordem	19/93 (20,4%)	9/58 (15,5%)	0,439
Textura	74/93 (79,6%)	49/58 (84,5%)	0,438
GLCM	23/93 (24,7%)	16/58 (27,6%)	0,699
GLDM	14/93 (15,1%)	8/58 (13,8%)	0,829
GLRLM	16/93 (17,2%)	10/58 (17,2%)	0,995
GLSZM	16/93 (17,2%)	11/58 (18,9%)	0,785
NGTDM	5/93 (5,4%)	4/58 (6,9%)	0,709

Nota.— GLCM = grey level cooccurrence matrix; GLDM = grey level dependence matrix; GLRLM = grey level run length matrix; GLSZM = grey level size zone matrix; NGTDM = neighboring gray tone difference matrix

Não houve diferença estatisticamente significativa entre os grupos da Tabela 5 para nenhuma das famílias de atributos, ou seja, tanto o vetor de atributos original quanto o vetor com as características selecionadas contêm a mesma distribuição de atributos.

A Tabela 6 apresenta os valores para AUC ROC, acurácia, sensibilidade e especificidade de todos os classificadores utilizando seus respectivos vetores de

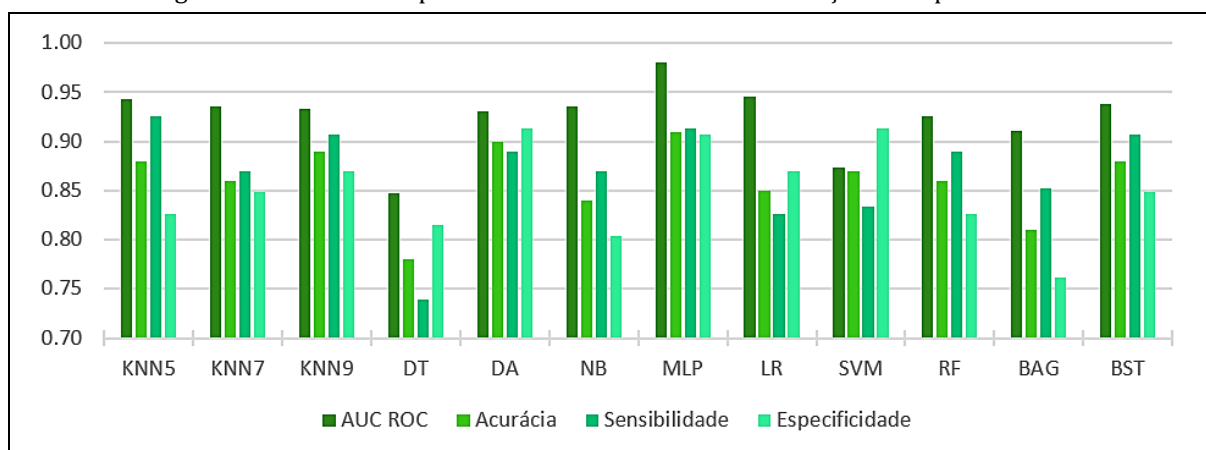
características com os atributos selecionados e validação cruzada de 10 *folde*s. A Figura 32 ilustra um gráfico comparativo entre todos os algoritmos de classificação e seus respectivos valores para AUC ROC, acurácia, sensibilidade e especificidade.

Tabela 6 - Resultados da classificação de cada algoritmo com seus respectivos vetores de características selecionados pela técnica *wrapper*

	KNN5	KNN7	KNN9	DT	DA	NB	MLP	LR	SVM	RF	BAG	BST
AUC ROC	0.94	0.94	0.93	0.85	0.93	0.94	0.98	0.95	0.87	0.93	0.91	0.94
Acurácia	88.0%	86.0%	89.0%	78.0%	90.0%	84.0%	91.0%	85.0%	87.0%	86.0%	81.0%	88.0%
Sensibilidade	92.6%	87.0%	90.7%	73.9%	88.9%	87.0%	91.3%	82.6%	83.3%	88.9%	85.2%	90.7%
Especificidade	82.6%	84.8%	87.0%	81.5%	91.3%	80.4%	90.7%	87.0%	91.3%	82.6%	76.1%	84.8%

Nota.— KNN = k nearest neighbor; DT = decision tree; DA = discriminant analysis; NB = naïve bayes; MLP = multilayer perceptron; LR = logistic regression; SVM = support vector machine; RF = random forest; BAG = bagging; BST = boosting.

Figura 32 - Gráfico comparativo dos resultados da classificação do experimento 4



Nota.— KNN = k nearest neighbor; DT = decision tree; DA = discriminant analysis; NB = naïve bayes; MLP = multilayer perceptron; LR = logistic regression; SVM = support vector machine; RF = random forest; BAG = bagging; BST = boosting. *Fonte:* da Autora.

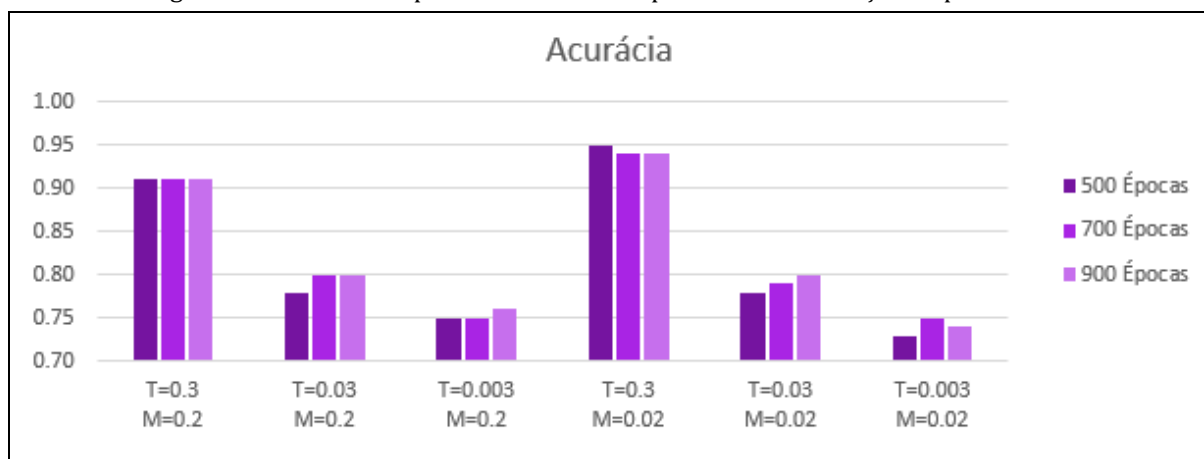
De maneira geral, o algoritmo MLP foi o que apresentou os melhores resultados, alcançando uma AUC ROC de 0,98 e acurácia de 91%. O algoritmo DA, embora não tenha apresentado o maior resultado para ROC AUC e acurácia, apresentou valores altos e consistentes, obtendo a maior especificidade com o valor de 91,3%. Esse mesmo valor de especificidade foi alcançado pelo algoritmo SVM, no entanto as demais medidas foram consideravelmente inferiores. O algoritmo DT foi o que obteve pior desempenho geral, com uma AUC ROC de 0,85 e acurácia de 78%.

A partir desses resultados, o algoritmo MLP foi o escolhido para prosseguir com as análises, aprofundando tanto o estudo dos hiperparâmetros do algoritmo quanto dos atributos do vetor de características.

4.3.5 Experimento 5 - Rede Neural + Wrapper

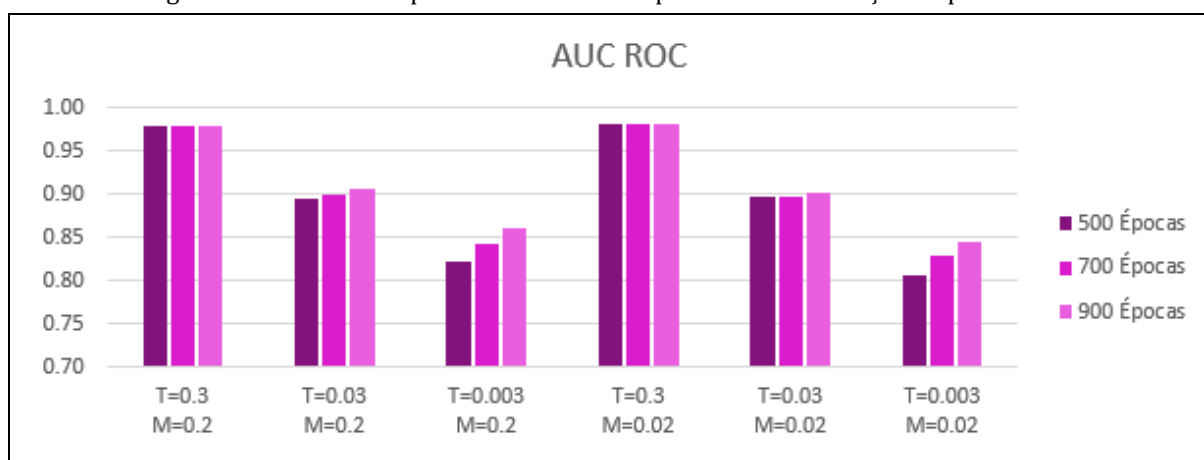
Iniciamos este experimento com o estudo dos hiperparâmetros que seriam utilizados para ajuste do aprendizado do algoritmo. Para isso, usamos como vetor de características os 16 atributos selecionados no experimento anterior por meio do método *wrapper* e como método de validação a validação cruzada de 10 *folds*. Os parâmetros testados foram a taxa de aprendizado, o *momentum* e o número de épocas. A Figura 33 e a Figura 34 apresentam, respectivamente, os valores para acurácia e AUC ROC para cada uma das combinações entre os parâmetros testados.

Figura 33 - Gráfico comparativo da acurácia para cada combinação de parâmetros



Nota.— T= taxa de aprendizado; M = *momentum*. Fonte: da Autora.

Figura 34 - Gráfico comparativo da AUC ROC para cada combinação de parâmetros



Nota.— T= taxa de aprendizado; M = *momentum*. Fonte: da Autora.

De acordo com as figuras acima, com relação às taxas de aprendizado, as melhores performances foram com o valor de 0,3, independentemente dos valores para o número de épocas e *momentum*. Desse modo, as próximas análises foram realizadas considerando esse valor de parâmetro como o escolhido.

Com relação ao número de épocas, para os casos onde a taxa de aprendizado era de 0,3, os resultados tanto para acurácia quanto para AUC ROC foram praticamente os mesmos, com exceção da combinação com $M=0,02$, onde o valor para acurácia foi discretamente superior para o número de épocas = 500.

Após a análise desse gráficos, optamos por realizar uma análise mais aprofundada para duas configurações de parâmetros: ($T=0,3$, $M=0,02$, $E=500$), que chamaremos de configuração alternativa e ($T=0,3$, $M=0,2$, $E=500$), que é a configuração *default*. Para isso, realizamos o embaralhamento do conjunto de treinamento criando 30 novos conjuntos, os quais foram treinados com cada uma das configurações e validados com o método de validação cruzada de 10 *folds*. As médias dos resultados para acurácia, AUC ROC, sensibilidade e especificidade foram, então, comparados por meio de um teste de hipóteses. Os resultados dessa análise são mostrados na Tabela 7.

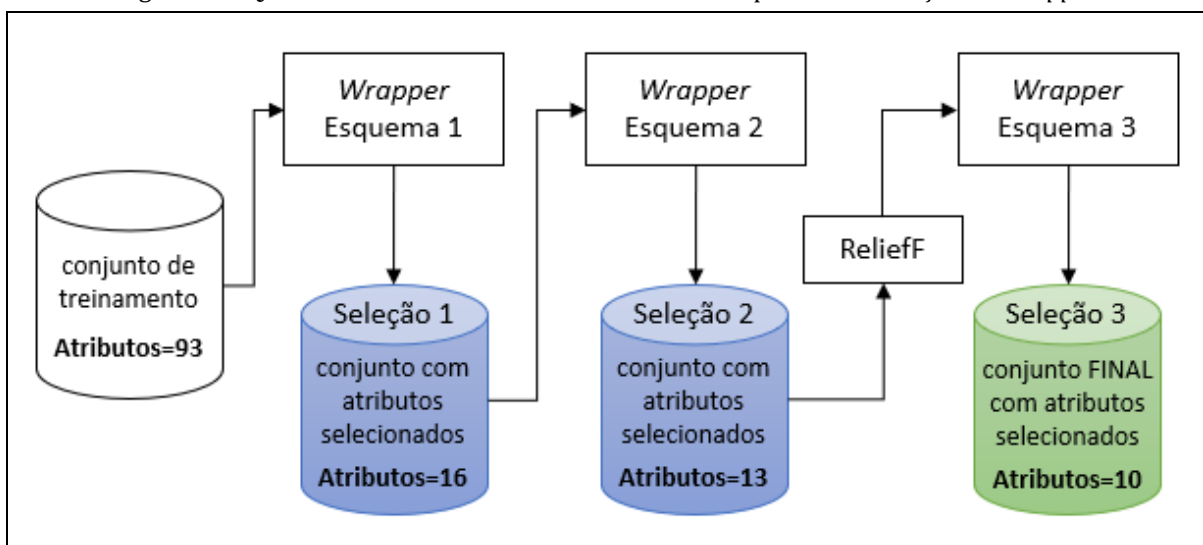
Tabela 7 - Comparação da performance da MLP entre a configuração alternativa e *default*.

Métrica	n	Média	Desvio Padrão	Valor p
Acurácia_A	30	0.89	0.027	0.527
Acurácia_D	30	0.88	0.017	
AUC ROC_A	30	0.95	0.015	0.598
AUC ROC_D	30	0.95	0.015	
Sensibilidade_A	30	0.86	0.034	0.851
Sensibilidade_D	30	0.87	0.026	
Especificidade_A	30	0.91	0.031	0.812
Especificidade_D	30	0.90	0.027	

Nota.— A= alternativa; D=*default*.

De acordo com a Tabela 7, não houve diferença estatisticamente significativa entre os grupos para nenhuma das medidas comparadas. Nesse sentido, optamos por seguir as análises com a configuração de hiperparâmetros *default*.

Com relação ao estudo do vetor de características, executamos o método wrapper consecutivamente com três diferentes esquemas. O número de atributos selecionados após cada uma das execuções é apresentado na Figura 35.

Figura 35 - Quantidade de características selecionadas após cada execução do *wrapper*.

Fonte: da Autora.

As características selecionadas após a execução do esquema 1 do *wrapper* são mostradas no Quadro 10.

Quadro 10 - Atributos selecionados pelo *wrapper* após o Esquema 1

firstorder_10Percentile
glcm_Id
glcm_Idm
glcm_InverseVariance
gldm_DependenceNonUniformityNormalized
gldm_LowGrayLevelEmphasis
gldm_SmallDependenceLowGrayLevelEmphasis
glrlm_LongRunLowGrayLevelEmphasis
glrlm_ShortRunEmphasis
glrlm_ShortRunLowGrayLevelEmphasis
glszm_GrayLevelNonUniformityNormalized
glszm_HighGrayLevelZoneEmphasis
glszm_LowGrayLevelZoneEmphasis
glszm_ZoneVariance
ngtdm_Busyness
ngtdm_Strength

Das 16 características selecionadas, uma era de primeira ordem, 3 eram provenientes da matriz de coocorrência de níveis de cinza (GLCM), 3 da matriz de dependência de níveis de cinza (GLDM), 3 da matriz de comprimento de cadeia de níveis de cinza (GLRLM), 4 da matriz de tamanho de zona de níveis de cinza (GLSZM) e 2 da matriz de diferença de tons de cinza vizinhos (NGTDM).

Esse vetor de características contendo 16 atributos foi utilizado como input para a execução do segundo esquema do *wrapper*, cujo resultado é apresentado no Quadro 11.

Quadro 11 - Atributos selecionados pelo *wrapper* após o Esquema 2

firstorder_10Percentile
glcm_Id
glcm_Idm
glcm_InverseVariance
gldm_DependenceNonUniformityNormalized
gldm_LowGrayLevelEmphasis
gldm_SmallDependenceLowGrayLevelEmphasis
glrlm_LongRunLowGrayLevelEmphasis
glrlm_ShortRunLowGrayLevelEmphasis
glszm_GrayLevelNonUniformityNormalized
glszm_HighGrayLevelZoneEmphasis
ngtdm_Busyness
ngtdm_Strength

O esquema 2 foi executado com o objetivo de eliminar os atributos que continuavam não sendo relevantes (embora já tivessem sido selecionados no esquema 1). Após sua execução, três características da Seleção 1 foram eliminadas, a saber: *glrlm_ShortRunEmphasis*, *glszm_LowGrayLevelZoneEmphasis* e *glszm_ZoneVariance*.

As 13 características resultantes do esquema 2 foram, então, ranqueadas utilizando o filtro ReliefF e validação cruzada de 10 *folds*. O resultado do ranking com os respectivos scores médios é mostrado no Quadro 12.

Quadro 12 - Atributos das Seleção 2 ranqueados após a execução do filtro ReliefF

Score médio	Atributo
0.067 ± 0.007	ngtdm_Busyness
0.061 ± 0.008	firstorder_10Percentile
0.046 ± 0.004	ngtdm_Strength
0.031 ± 0.006	gldm_LowGrayLevelEmphasis
0.023 ± 0.003	glrlm_LongRunLowGrayLevelEmphasis
0.023 ± 0.007	glrlm_ShortRunLowGrayLevelEmphasis
0.02 ± 0.005	glszm_GrayLevelNonUniformityNormalized
0.014 ± 0.002	glcm_Idm
0.015 ± 0.005	gldm_SmallDependenceLowGrayLevelEmphasis
0.014 ± 0.002	glcm_Id
0.015 ± 0.004	glcm_InverseVariance
0.014 ± 0.004	glszm_HighGrayLevelZoneEmphasis
0.007 ± 0.003	gldm_DependenceNonUniformityNormalized

As características que obtiveram um score > 0,03 após a execução do filtro ReliefF (que foram `ngtdm_Busyness`, `firstorder_10Percentile`, `ngtdm_Strength` e `gldm_LowGrayLevelEmphasis`) inicializaram o vetor de características na execução do terceiro e último wrapper. O vetor de características final resultante dessa última execução é apresentado no Quadro 13.

Quadro 13 - Vetor de características final resultante do esquema 3 do wrapper

Família de Atributos	Atributos selecionados e suas descrições
Primeira Ordem	10th Percentile: Valor que representa o décimo percentil da amostra.
GLCM	Inverse Difference: É uma medida da homogeneidade local de uma imagem. Quanto maior o valor desse atributo, mais homogênea é a imagem. Inverse Difference Moment: Semelhante à <i>Inverse Difference</i> , mas com pesos maiores para elementos próximos à diagonal principal. Inverse Variance: Mede a heterogeneidade da imagem.
GLDM	Low Gray Level Emphasis: Mede a distribuição dos valores de níveis de cinza mais baixos. Um valor alto para esse atributo significa que a imagem possui uma maior concentração de níveis de cinza de valores mais baixos. Small Dependence Low Gray Level Emphasis: Mede a distribuição conjunta de pixels com baixo valores de níveis de cinza que também possuem uma baixa dependência.
GLSZM	Gray Level Nonuniformity Normalized: Mede a variabilidade dos valores de nível de cinza na imagem. Um valor baixo para esse atributo indica maior similaridade nos valores de intensidade. High Gray Level Zone Emphasis: Mede a distribuição dos valores de nível de cinza mais altos. Um valor alto para esse atributo indica que as zonas homogêneas da imagem apresentam intensidade de cinza de valores mais elevados.
NGTDM	Busyness: É uma medida da mudança de um pixel para seu vizinho. Um valor alto para este atributo indica uma imagem com aspecto de areia fina, com rápidas mudanças de intensidade entre um pixel e sua vizinhança. Strength: Mede a facilidade com que as primitivas de uma imagem podem ser visualizadas. Seu valor é alto quando as primitivas são facilmente definidas e visíveis, ou seja, a imagem apresenta uma mudança lenta na intensidade, porém diferenças grosseiras maiores são facilmente identificadas.

Essas 10 características selecionadas a partir do experimento 5 são as variáveis que objetivávamos encontrar e que chamaremos de assinatura radiômica. A Figura 36 mostra o gráfico de dispersão dos dados de treinamento para pares desses atributos.

Ainda com o objetivo de testar o quão robusta e eficiente essa assinatura era para diferenciar FVCs benignas de malignas, realizamos um teste de validação interna e externa. No teste de validação interna, treinamos e testamos 30 diferentes modelos do tipo MLP e computamos as medidas de acurácia, AUC ROC, sensibilidade e especificidade com um intervalo de confiança de 95%. Esses resultados são mostrados na Figura 37.

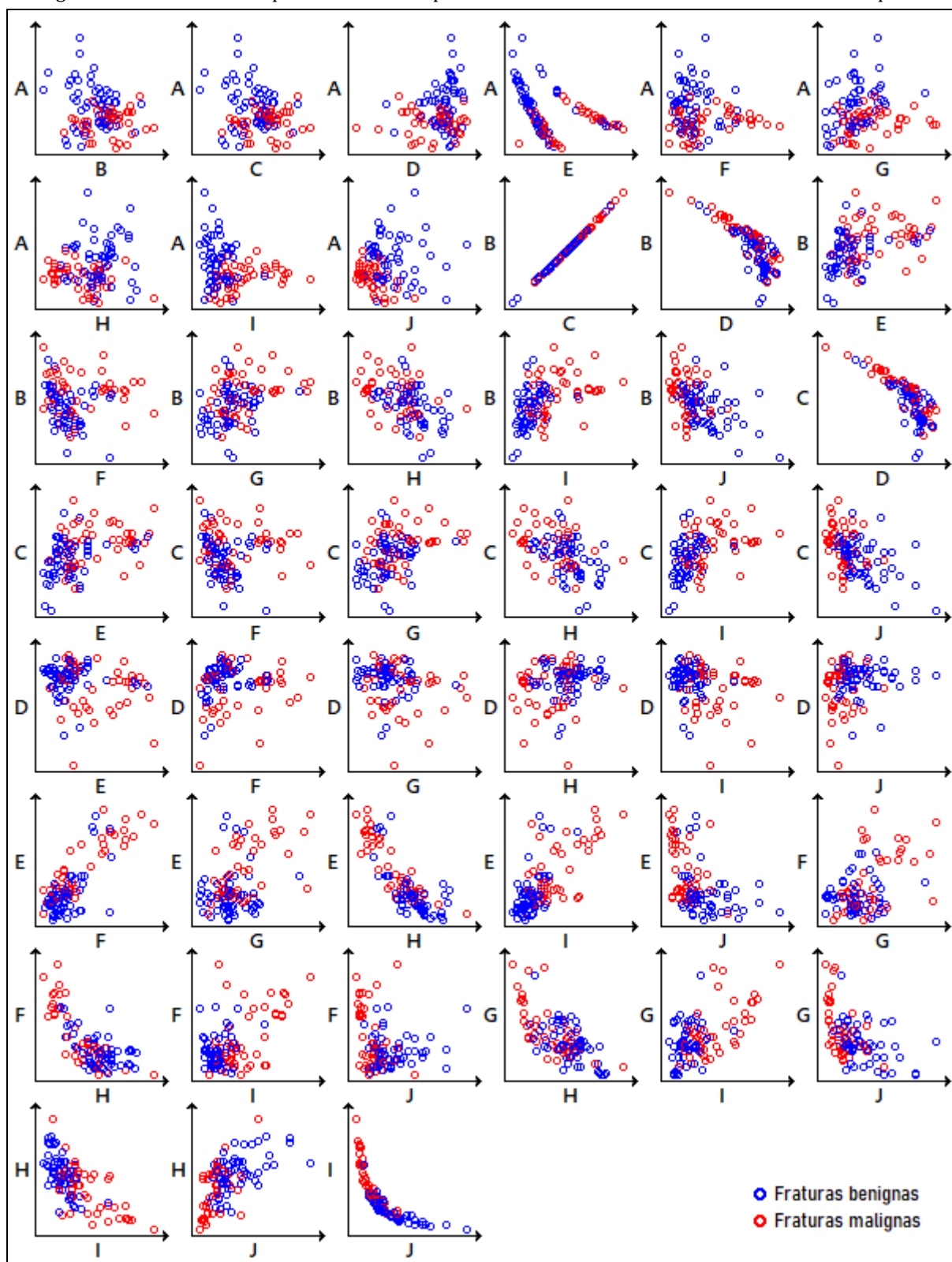
Os valores médios entre os 30 modelos para acurácia, AUC ROC, sensibilidade e especificidade foram, respectivamente, 90,97% (IC de 95%: 90,22%, 91,72%), 0,97 (IC de

95%: 0,966, 0,974), 89,36% (IC de 95%: 88,18%, 90,54%) e 92,33% (IC de 95%: 91,34%, 93,32%). Esses valores reforçam a ideia de que a assinatura radiômica encontrada no presente trabalho se mostrou robusta e capaz de diferenciar as FVCs malignas e benignas.

O melhor modelo dentre os 30 testados obteve resultados excelentes com a validação cruzada de 10 *folds*, alcançando os seguintes valores: acurácia = 95%, AUC ROC = 0,98, sensibilidade = 93,5% e especificidade = 96,3%. Esse modelo apresentou apenas 3 falsos negativos, que podem ser visualizados na Figura 38.

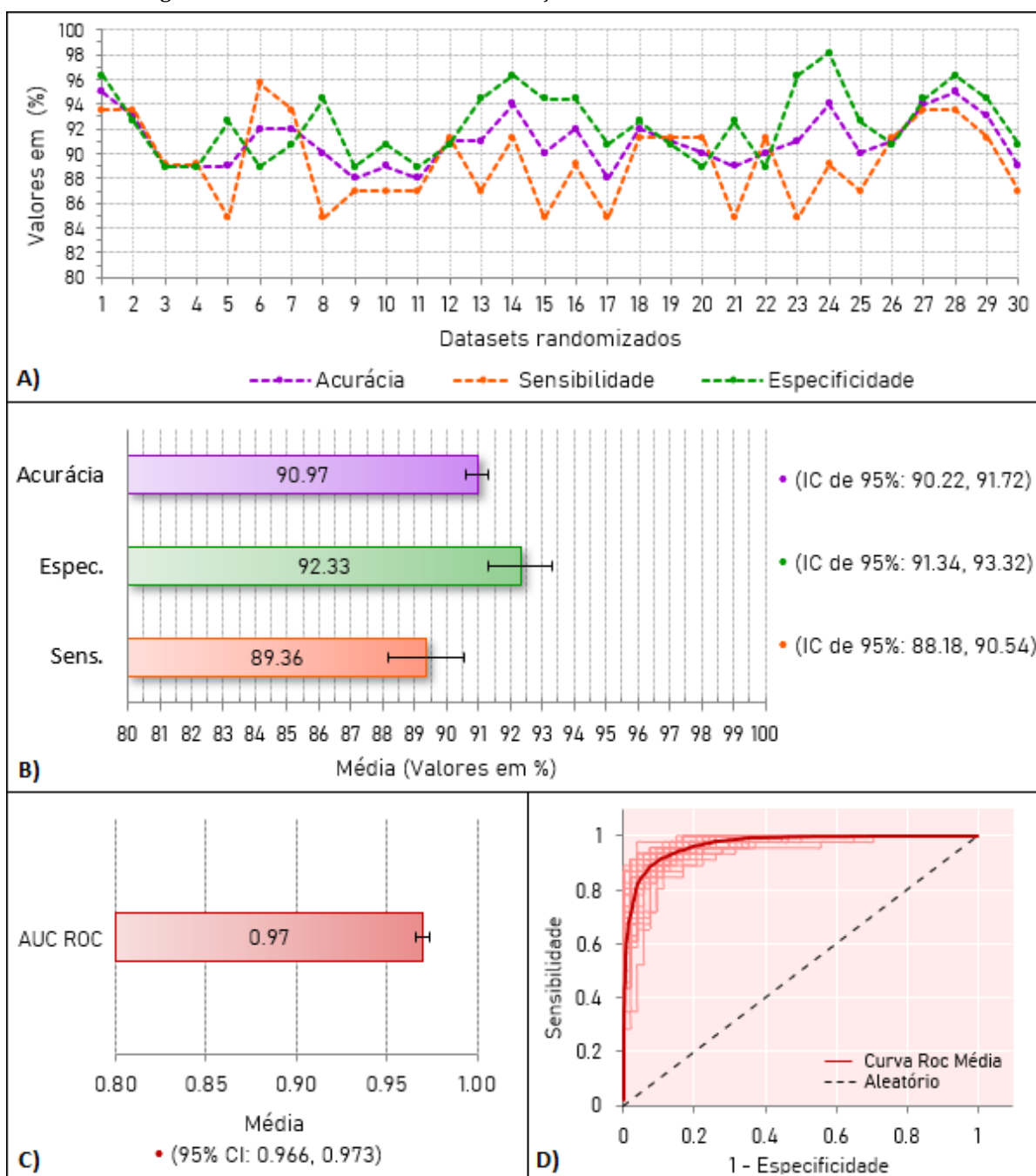
Após o teste de validação interna, realizamos um teste de validação externa, o qual utilizou o modelo que apresentou os melhores resultados na validação interna para testar um conjunto de dados independente, que ainda não havia sido utilizado em nenhum momento (nem no treinamento e nem na validação dos algoritmos). Esse conjunto era composto por 30 corpos vertebrais fraturados, dos quais 15 possuíam fratura maligna e 15 possuíam fratura benigna. A acurácia, AUC ROC, sensibilidade e especificidade da classificação do conjunto de teste foram, respectivamente, 93,3%, 0,97, 93,3% e 93,3%. Os resultados do teste de validação externa utilizando o melhor modelo podem ser visualizados na Figura 39 e a arquitetura da rede deste modelo é apresentada na Figura 40.

Figura 36 - Gráficos de dispersão dos dados para todos os atributos selecionados exibidos em pares



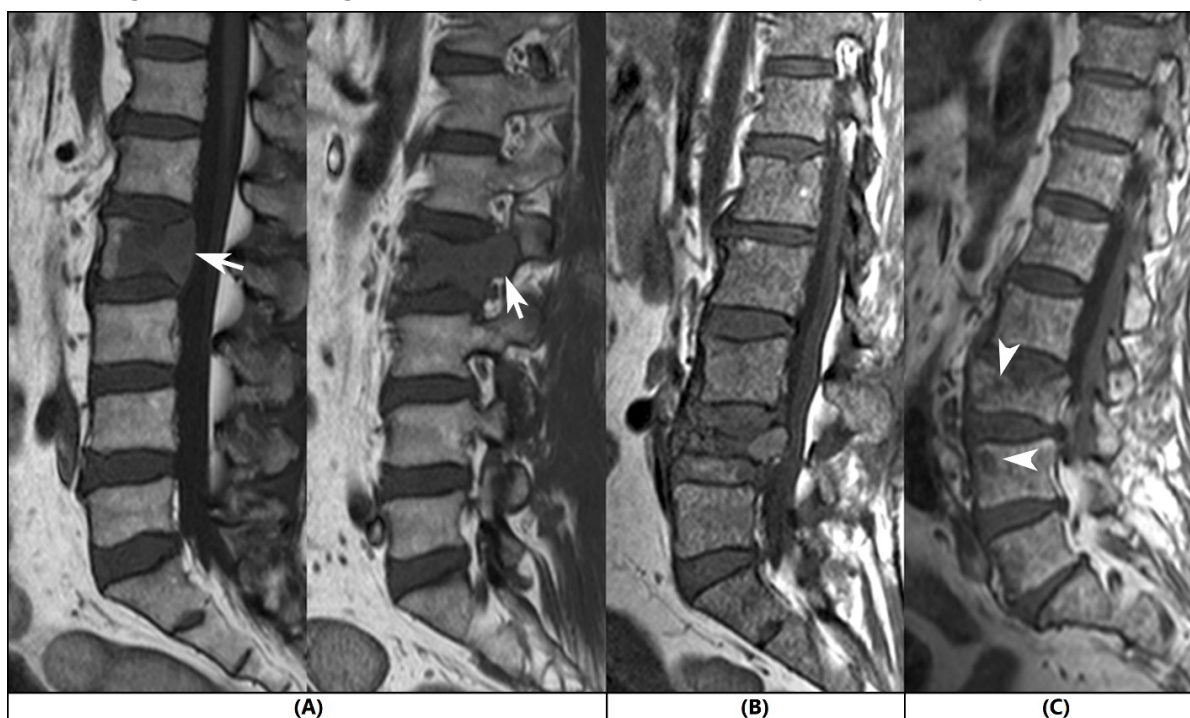
Nota.— A = 10th percentile; B = inverse difference; C = inverse difference moment; D = inverse variance; E = low gray level emphasis; F = small dependence low gray level emphasis; G = gray level nonuniformity normalized; H = high gray level zone emphasis; I = busyness; J = strength. *Fonte:* da Autora.

Figura 37 - Resultados do teste de validação interna considerando 30 modelos



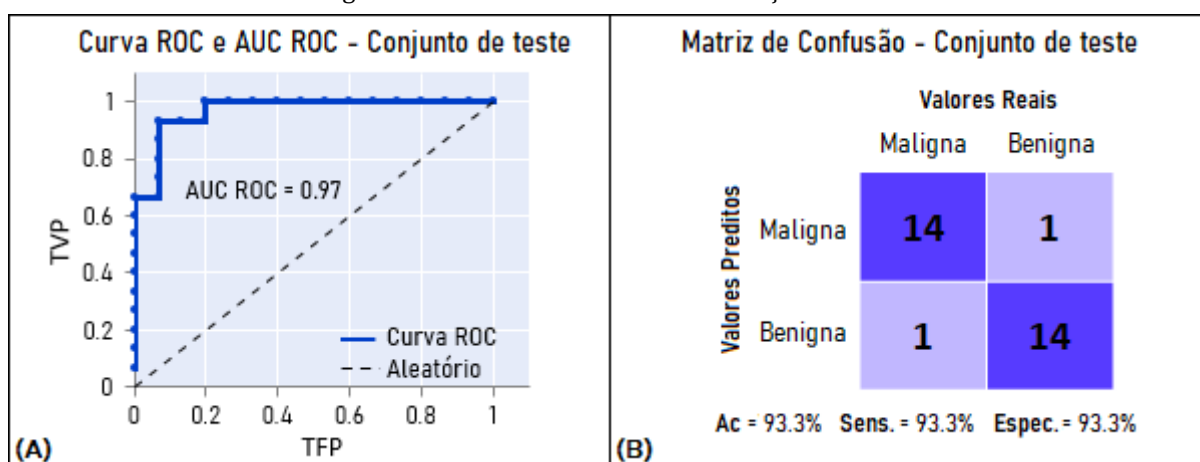
Nota.— Espec. = especificidade; Sens. = sensibilidade; IC = intervalo de confiança. *Fonte:* da Autora.

Figura 38 - Falsos negativos resultantes do melhor modelo no teste de validação interna



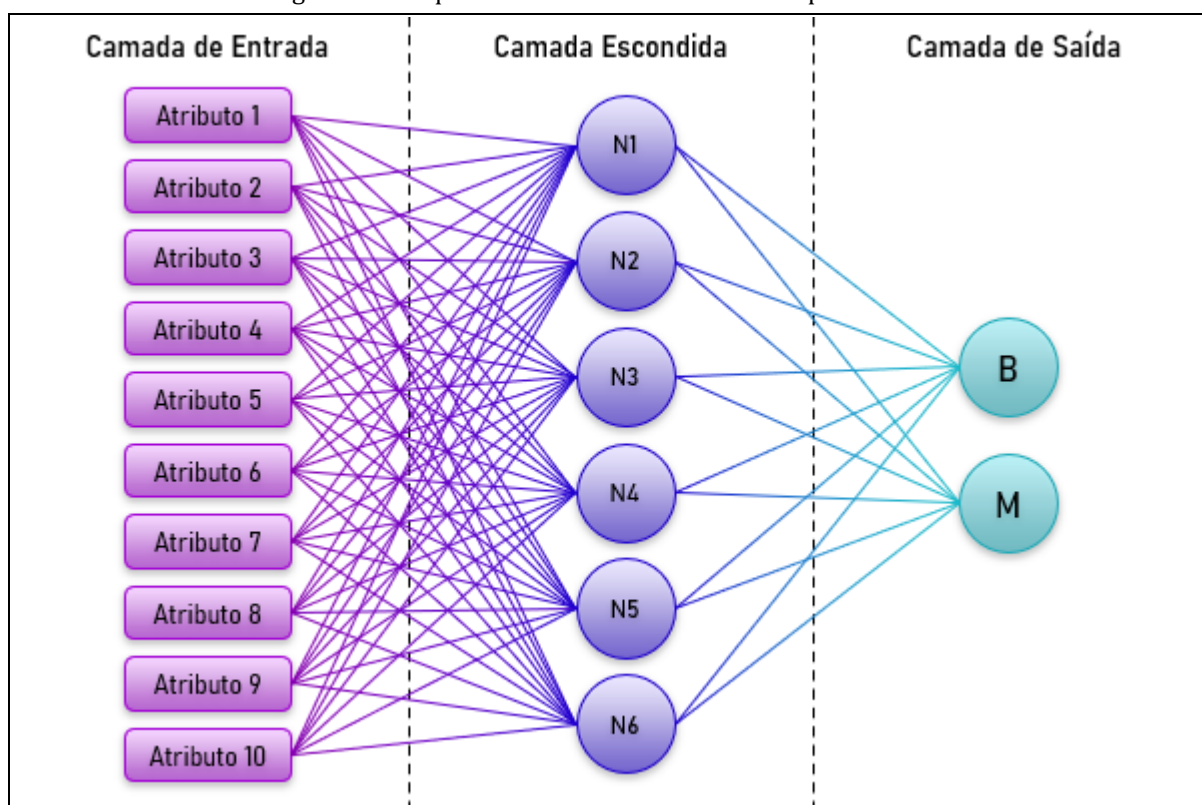
Nota.— Em (A) (homem de 59 anos), a fratura vertebral compressiva está associada a lesão nodular focal, abaulamento convexo da parede posterior (seta branca esquerda) e envolvimento do pedículo (seta branca direita). Em (B) (mulher de 52 anos) e (C) (mulher de 51 anos), a medula óssea exibe um padrão de infiltração do tipo de sal e pimenta. Os micronódulos são mais bem visualizados em (C) (pontas de seta brancas). A fratura do corpo vertebral de L4 em (B) mostra retropulsão da parede posterior. *Fonte:* da Autora.

Figura 39 - Resultado do teste de validação externa



Nota.— TVP = taxa de verdadeiros positivos; TFP = taxa de falsos positivos; Ac = acurácia; Sens. = sensibilidade; Espec. = especificidade. *Fonte:* da Autora.

Figura 40 - Arquitetura do modelo com melhor performance



Nota.— Atributo 1 = 10th percentile; Atributo 2 = inverse difference; Atributo 3 = inverse difference moment; Atributo 4 = inverse variance; Atributo 5 = low gray level emphasis; Atributo 6 = small dependence low gray level emphasis; Atributo 7 = gray level nonuniformity normalized; Atributo 8 = high gray level zone emphasis; Atributo 9 = busyness; Atributo 10 = strength; N = neurônio; B = fratura benigna; M = fratura maligna. *Fonte:* da Autora.

Na Figura 40, a camada de entrada corresponde aos 10 atributos selecionados, a camada oculta é composta por seis neurônios artificiais e a camada de saída é composta pelas classes fratura benigna e fratura maligna. Os valores dos pesos ajustados para cada interconexão (sinapse) da rede podem ser visualizados na Tabela 8.

Tabela 8 - Pesos do modelo final

Neurônio N1	Pesos	Neurônio N2	Pesos
INTERCONEXÃO		INTERCONEXÃO	
Atributo_1→N1	-3.2154	Atributo_1→N2	7.8978
Atributo_2→N1	-0.3999	Atributo_2→N2	0.0073
Atributo_3→N1	-0.1239	Atributo_3→N2	0.0173
Atributo_4→N1	7.5655	Atributo_4→N2	-0.6875
Atributo_5→N1	-6.7146	Atributo_5→N2	-1.6295
Atributo_6→N1	9.3134	Atributo_6→N2	-2.0289
Atributo_7→N1	7.0455	Atributo_7→N2	3.9376
Atributo_8→N1	2.8134	Atributo_8→N2	-3.1596
Atributo_9→N1	1.4576	Atributo_9→N2	-5.7082
Atributo_10→N1	-2.8006	Atributo_10→N2	4.3194
THRESHOLD	-1.9577	THRESHOLD	0.1548
Neurônio N3	Pesos	Neurônio N4	Pesos

INTERCONEXÃO		INTERCONEXÃO	
Atributo_1→N3	-1.6532	Atributo_1→N4	0.4986
Atributo_2→N3	-2.0551	Atributo_2→N4	-1.6393
Atributo_3→N3	-2.1176	Atributo_3→N4	-1.8124
Atributo_4→N3	-0.8456	Atributo_4→N4	-1.5268
Atributo_5→N3	3.3034	Atributo_5→N4	3.0756
Atributo_6→N3	-0.2236	Atributo_6→N4	-4.3502
Atributo_7→N3	-2.2581	Atributo_7→N4	-1.8618
Atributo_8→N3	-0.6972	Atributo_8→N4	-3.7455
Atributo_9→N3	1.2645	Atributo_9→N4	-1.0157
Atributo_10→N3	1.7179	Atributo_10→N4	2.8438
THRESHOLD	-0.5994	THRESHOLD	-0.2563
Neurônio N5	Pesos	Neurônio N6	Pesos
INTERCONEXÃO		INTERCONEXÃO	
Atributo_1→N5	5.5179	Atributo_1→N6	-0.0239
Atributo_2→N5	-1.4568	Atributo_2→N6	4.3483
Atributo_3→N5	-1.6194	Atributo_3→N6	4.1669
Atributo_4→N5	-1.6264	Atributo_4→N6	-6.1192
Atributo_5→N5	-0.6148	Atributo_5→N6	-0.1888
Atributo_6→N5	-2.9598	Atributo_6→N6	0.4997
Atributo_7→N5	2.1000	Atributo_7→N6	-4.1438
Atributo_8→N5	-3.2373	Atributo_8→N6	4.1383
Atributo_9→N5	-4.8556	Atributo_9→N6	0.5510
Atributo_10→N5	4.2871	Atributo_10→N6	0.4958
THRESHOLD	0.6262	THRESHOLD	-0.6599
Neurônio B	Pesos	Neurônio M	Pesos
INTERCONEXÃO		INTERCONEXÃO	
N1→B	-9.6385	N1→M	9.6315
N2→B	7.2378	N2→M	-7.2706
N3→B	4.0091	N3→M	-3.9888
N4→B	4.6684	N4→M	-4.6934
N5→B	5.7552	N5→M	-5.7131
N6→B	-8.1459	N6→M	8.1485
THRESHOLD	-0.1910	THRESHOLD	0.1961

Nota.— Atributo_1 = 10th percentile; Atributo_2 = inverse difference; Atributo_3 = inverse difference moment; Atributo_4 = inverse variance; Atributo_5 = low gray level emphasis; Atributo_6 = small dependence low gray level emphasis; Atributo_7 = gray level nonuniformity normalized; Atributo_8 = high gray level zone emphasis; Atributo_9 = busyness; Atributo_10 = strength;

De acordo com a definição matemática de um neurônio artificial (como o ilustrado na Figura 8B), temos que a saída y de um neurônio artificial arbitrário é expressa por:

$$y = \varphi(w_0 + x_1w_1 + x_2w_2 + \dots + x_nw_n) \quad (7)$$

onde $[x_1, x_2, \dots, x_n]$ são os valores de entrada (atributos), $[w_1, w_2, \dots, w_n]$ são os pesos das interconexões, w_0 é o bias (ou *threshold*) e a função φ é a função de ativação do neurônio. Para este modelo, a função de ativação utilizada foi a sigmoide, que é expressa como mostra a Equação (8).

$$\varphi(u) = \frac{1}{1 + e^{-u}} \quad (8)$$

Fazendo um paralelo entre os valores da Tabela 8 com a Equação (7), temos que a saída y do Neurônio N1, por exemplo, seria expressa como:

$$y = \varphi\{(-1,9577) + [\text{Atributo}_1 * (-3,2154) + \text{Atributo}_2 * (-0,3999) + \text{Atributo}_3 * (-0,1239) + \\ + \text{Atributo}_4 * (7,5655) + \text{Atributo}_5 * (-6,7146) + \text{Atributo}_6 * (9,3134) + \\ + \text{Atributo}_7 * (7,0455) + \text{Atributo}_8 * (2,8134) + \text{Atributo}_9 * (1,4576) + \\ + \text{Atributo}_{10} * (-2.8006)]\}$$

onde as variáveis ($\text{Atributo}_1, \text{Atributo}_2, \text{Atributo}_3, \dots$) correspondem às entradas $[x_1, x_2, \dots, x_n]$, os pesos $(-3,2154, -0,3999, -0,1239, \dots)$ correspondem aos pesos $[w_1, w_2, \dots, w_n]$ e o *threshold* = -1,9577 corresponde ao bias w_0 .

5 DISCUSSÃO

Neste estudo, investigamos a possibilidade de realizar o diagnóstico diferencial de fraturas vertebrais por compressão benignas e malignas por meio do uso de atributos radiômicos tridimensionais extraídos de imagens de ressonância magnética da coluna vertebral associados a um modelo computacional. Para isso, realizamos análises exploratórias com o objetivo de definir a melhor abordagem metodológica para depois aprofundarmos as análises. Nesta seção, a primeira parte será destinada à comparação direta dos métodos testados. Em seguida, faremos uma comparação dos nossos achados com os resultados encontrados na literatura e, finalmente, finalizaremos com as limitações encontradas durante o trabalho.

5.1 ANÁLISE METODOLÓGICA

5.1.1 Pré-processamento das imagens

O processo de equalização de histograma foi particularmente importante neste trabalho por estarmos utilizando imagens de ressonância magnética. As intensidades de sinal nas imagens de RM surgem de uma complexa interação de propriedades inerentes ao tecido, como tempos de relaxamento e parâmetros de aquisição. Portanto, é difícil obter informações sobre as propriedades físicas do tecido apenas a partir das intensidades do sinal de imagem de RM. Isso contrasta com as imagens de TC, nas quais a intensidade do sinal pode ser correlacionada com a densidade do tecido (KUMAR et al., 2012). Nesse sentido, notamos que mais importante do que a utilizar a intensidade do pixel em si, seria trabalharmos com a interação entre os pixels e o modo como eles se contrastam e, assim, concluímos que realizar uma operação de melhoramento de contraste poderia ser interessante em nosso estudo.

5.1.2 Técnicas de segmentação e extração de atributos

De acordo com os resultados dos experimentos 2 e 3, foi possível observar que, de maneira geral, a segmentação realizada pelo 3D Slicer com posterior extração de atributos de primeira ordem e de textura obteve resultados superiores quando comparada com os

resultados do experimento utilizando a rede U-Net para segmentar e extrair características. Como não realizamos uma análise individual para cada uma das etapas, não é possível saber se a diferença de performance é atribuída apenas ao método de segmentação, apenas ao método de extração de atributos ou aos dois métodos em conjunto. No entanto, o que podemos afirmar é que na área médica, onde é fundamental que todas as informações sejam disponibilizadas de forma transparente e que sejam passíveis de interpretação, trabalhar com variáveis que tenham significados matemáticos que possam ser correlacionados com achados biológicos é considerado o mais adequado.

No que se refere à segmentação, a literatura mostra que a rede U-Net tem obtido ótimos resultados (BYRA et al., 2020b, 2020a; FABIJANŠKA, 2018; HUANG et al., 2018; NORMAN; PEDOIA; MAJUMDAR, 2018; PARK et al., 2020; PUNN; AGARWAL, 2021). Um trabalho inicial que estamos fazendo propõe unir os métodos de segmentação utilizando a U-Net com a extração de características utilizando o PyRadiomics. Nossa hipótese é a de que essa união não traria tantos prejuízos para o desempenho do modelo e contribuiria para um possível estudo prospectivo em fase operacional, testando o modelo num cenário real da rotina clínica.

5.1.3 Seleção de características

O processo de seleção de características é fundamental em estudos de classificação ou regressão que utilizam *machine learning*, pois é a etapa que vai identificar as variáveis que, de fato, são relevantes para solucionar o problema em questão. Muitas vezes, a falta de atenção dada a esta fase faz com que vários trabalhos transmitam superficialidade e não sejam reproduzíveis. Performances altas alcançadas em trabalhos que não realizaram um bom estudo das características podem apresentar vieses de interpretação, uma que os bons resultados podem ser atribuídos a outros fatores como, por exemplo, *overfitting* do modelo.

De acordo com os resultados obtidos em nossos experimentos, a técnica *wrapper* se mostrou superior às técnicas baseadas em filtros, ainda que os atributos selecionados não tenham sido os mais relevantes (do ponto de vista estatístico) quando analisados individualmente no experimento 1. Isso pode ser explicado pela seguinte razão: classificadores induzidos por dados não possuem acesso à distribuição desses dados e, dessa forma, a relevância individual de um atributo não implica na otimização de um

subconjunto de características, assim como a irrelevância de um atributo do subconjunto não implica que este não possa ser o subconjunto ideal.

A técnica *wrapper* permite que um subconjunto ideal seja encontrado ainda que ele contenha características com relevância individual fraca, como demonstrado com nossos resultados. Dos 10 atributos selecionados no experimento final (que foram testados de diferentes formas e permaneceram mostrando estabilidade), três deles (*inverse variance*, *small dependence low gray level emphasis* e *gray level non uniformity normalized*) não apresentaram diferença estatisticamente significativa entre os grupos de fraturas benignas e malignas quando analisados isoladamente.

Outro ponto importante a se destacar é com relação ao modo como o método *wrapper* foi configurado. No experimento 5, por exemplo, utilizamos tanto a técnica de inclusão de atributos no vetor de características como a técnica de ablação. O *wrapper* não testa todos os subconjuntos possíveis por meio de uma busca exaustiva, pois isso seria extremamente custoso, uma vez que o número total de subconjuntos testados nesse tipo de busca é 2^n , onde n é o número de atributos total. O que o *wrapper* faz é utilizar uma função heurística para guiá-lo. Dessa forma, para se aumentar as chances de encontrar o menor e mais eficiente subconjunto de atributos, uma proposta foi utilizar o *wrapper* em sequência, com diferentes configurações e formas de validação, o que se mostrou uma ótima opção, resultando em uma assinatura radiômica estável contendo 10 atributos.

5.1.4 Algoritmos de classificação

A escolha do algoritmo de classificação vai depender do problema a ser modelado, do tipo de saída desejada e a natureza dos atributos. Em nosso trabalho, a performance de alguns algoritmos variou de acordo com a origem dos atributos (PyRadiomics ou U-Net) e o método de seleção de características utilizado. Por exemplo, o classificador *Support Vector Machine* obteve um dos piores desempenhos no experimento 2 (PyRadiomics + Filtros) e um dos melhores no experimento 3 (U-Net + filtros). O algoritmo *Discriminant Analysis*, por sua vez, saltou de um desempenho mediano nos experimentos 2 e 3 para o segundo melhor desempenho no experimento 4.

No entanto, para outros algoritmos, essa variabilidade não foi observada. Os classificadores *Multilayer Perceptron* e *Random Forest*, por exemplo, conquistaram sempre boas performances, independentemente das técnicas utilizadas nas etapas

anteriores. Esses algoritmos são robustos e capazes de lidar com problemas não linearmente separáveis. A superioridade do algoritmo *Random Forest* no experimento 2, por exemplo, pode ser explicada pelo fato dele já possuir uma seleção de características dentro do próprio algoritmo, como explicado na seção 2.2.2.4 ao descrever o método embarcado para seleção de atributos. No entanto, no experimento 4, quando utilizado o método *wrapper* para selecionar as características, a performance do *Multilayer Perceptron* se sobressaiu com relação aos demais algoritmos de classificação e, por isso, foi o classificador escolhido para dar continuidade nos estudos por meio de análises mais detalhadas.

5.2 ASSINATURA RADIÔMICA E MODELO PROPOSTO

O modelo final proposto (experimento 5) foi desenvolvido com base nas análises dos experimentos prévios, que serviram para definir a abordagem que seria utilizada. Optamos pelo vetor de características com os atributos da biblioteca PyRadiomics, que foram extraídos tridimensionalmente a partir dos corpos vertebrais segmentados no 3D Slicer. Para a etapa de *machine learning*, realizamos a seleção de atributos por meio do método *wrapper* e a classificação utilizando o algoritmo *Multilayer Perceptron*.

Nós conseguimos identificar um conjunto de atributos radiômicos contendo 10 características que, inseridas em uma rede neural, alcançaram excelente performance diagnóstica. Nossos resultados mostraram valores médios altos com intervalos de confiança estreitos para o conjunto de treinamento, indicando que os modelos gerados eram robustos e os atributos selecionados eram estáveis. Tais resultados são particularmente confiáveis uma vez que o embaralhamento aleatório do conjunto de treinamento foi realizado 30 vezes antes da validação cruzada de 10 *folds*. Isso é bastante relevante pois, em alguns classificadores, como é o caso das redes neurais, pequenas mudanças no conjunto de treinamento podem resultar em grandes mudanças nos resultados da classificação. Portanto, é importante utilizar técnicas de avaliação capazes de garantir a confiabilidade dos resultados obtidos. Embora a técnica de validação cruzada *k-fold* seja amplamente utilizada para avaliar a generalização de um modelo, quando realizada apenas uma vez no conjunto de treinamento, pode causar um viés relacionado à divisão do conjunto de dados. Isso significa que para uma única execução, o conjunto de dados poderia ser dividido de forma a gerar a melhor classificação,

apresentando, assim, um desempenho resultante acima das médias reais. Um embaralhamento aleatório anterior deve minimizar esse viés.

O melhor modelo proposto, apresentado na Tabela 8, apresentou excelente desempenho tanto para o conjunto de treinamento, usando validação cruzada de 10 *folds*, quanto para o conjunto de teste independente. Em relação ao desempenho baseado na validação cruzada, o modelo errou apenas cinco classificações (dois falso-positivos e três falso-negativos). Os casos falso-negativos eram provenientes de pacientes cujas fraturas eram secundárias ao mieloma múltiplo. Em dois desses casos, a medula óssea exibiu um padrão de infiltração do tipo sal-e-pimenta e, dessa forma, presumimos que a rede neural precisaria de mais casos no conjunto de dados para alcançar o treinamento adequado para reconhecer esse padrão específico. Além disso, o classificador não recebeu características extraídas das vértebras restantes, dos pedículos ou dos discos intervertebrais, que poderiam ser informações relevantes para diferenciar esse tipo específico de fratura maligna. Outro aspecto que vale ressaltar é que não extraímos ou analisamos características de forma e, portanto, os resultados da rede neural não puderam refletir as anormalidades do contorno da parede posterior presentes em dois dos três casos falso-negativos. No entanto, a literatura anterior sugeriu que as características de textura são mais importantes para diferenciar fraturas vertebrais benignas de malignas no diagnóstico auxiliado por computador (FRIGHETTO-PEREIRA et al., 2016).

Com relação ao conjunto de atributos radiômicos selecionados, das dez características resultantes, nove são baseadas em textura e uma é baseada em intensidade. Isso mostra que os padrões encontrados pela rede neural estão mais associados à relação espacial entre os voxels do que aos seus valores individuais de intensidade. Tal fenômeno era esperado, pois, ao contrário das imagens de TC, que permitem uma correlação entre a intensidade do sinal e o coeficiente de atenuação dos raios X do tecido, nas imagens de RM isso não é possível devido à natureza de sua aquisição. Nesse caso, as intensidades de sinal surgem da interação entre as propriedades do tecido, como o tempo de relaxamento e os parâmetros de aquisição.

Duas características da assinatura radiômica que se destacaram neste estudo foram *busyness* e *strength*. Como mencionado previamente no Quadro 13, um valor alto para *busyness* ocorre quando há mudanças rápidas de intensidade entre um pixel/voxel e seus vizinhos, o que significa que a frequência espacial das mudanças de intensidade é bastante

alta. *Strength* refere-se à facilidade com que uma primitiva pode ser claramente definível e visível. Em outras palavras, quanto maior o valor do atributo *strength*, mais fácil é definir e visualizar primitivas (AMADASUN; KING, 1989). Os resultados mostraram que FVCs malignas tendem a ter valor alto para *busyness* e baixo para *strength*, enquanto FVCs benignas tendem a apresentar características opostas.

5.3 TRABALHOS RELACIONADOS

Até onde sabemos, nosso trabalho produziu resultados superiores comparados aos estudos anteriores que usaram *machine learning* para diferenciar FVCs malignas e benignas.

Frighetto-Pereira et al. (FRIGHETTO-PEREIRA et al., 2015) avaliaram 103 FVCs usando o classificador kNN e validação cruzada de 10 *folds*. Para isso, eles usaram o corte central de exames de RM ponderados em T1 da coluna vertebral de 61 pacientes. Eles realizaram segmentação manual utilizando o software Adobe Photoshop⁸ e extraíram 14 características das imagens (*coefficient of variation, skewness, kurtosis* e 14 atributos sugeridos por Haralick et al. (HARALICK, 1979)), alcançando uma AUC ROC de 0,91 (para k=3 e 4 características) e uma acurácia de 86,41% (para k=1 e 13 características).

Azevedo-Marques et al. (AZEVEDO-MARQUES et al., 2015) realizaram segmentação manual de 73 FVCs (54 benignas e 19 malignas) provenientes de 47 pacientes e extraíram 19 atributos a partir das transformadas de Fourier e Wavelet. Usando o algoritmo kNN e validação cruzada de 10 *folds*, eles alcançaram uma acurácia média de 82,9% e uma AUC ROC média de 0,81. Os valores máximos alcançados foram acurácia = 94,7 e AUC ROC = 0,95 para k=3.

Frighetto et al. (FRIGHETTO-PEREIRA et al., 2016) usaram os classificadores kNN, *Naïve Bayes* e *RBF Network* para classificar 102 FVCs como malignas ou benignas a partir do corte central das RMs sagitais ponderadas em T1. Eles realizaram segmentação manual usando Adobe Photoshop e extraíram 27 características, sendo três estatísticas, 14 baseadas em textura e 10 baseadas em forma. Para avaliação, eles usaram validação cruzada de 10 *folds*, alcançando uma precisão de 85,3% e uma AUC ROC de 0,92 com o classificador Naïve Bayes.

⁸ <https://www.adobe.com/br/products/photoshop.html>

Casti et al. (CASTI et al., 2017) propuseram um CAD cooperativo (CCAD) para classificar 103 FVCs como benignas ou malignas. Eles executaram nove técnicas de segmentação diferentes em paralelo, extraíram 15 atributos baseados em forma a partir de cada resultado de segmentação e os inseriram em quatro classificadores: análise discriminante linear (do Inglês, *linear discriminant analysis* - LDA), análise discriminante quadrática (do Inglês, *quadratic discriminant analysis* - QDA), máquina de vetores de suporte (do Inglês, *support vector machine* - SVM), e k-vizinhos mais próximo (do Inglês, *k-nearest neighbor* - kNN), com k=5. Eles avaliaram os modelos usando a AUC ROC média em 100 subamostras aleatórias, onde 80% do conjunto de dados foi usado para treinamento e 20% para teste. A regressão logística com o método *stepwise* foi usada para seleção de características em cada subamostra de cada técnica de segmentação. Eles alcançaram uma precisão de 92% (QDA, SVM e k-NN) e uma AUC ROC de 0,95 (com QDA).

Finalmente, Arpitha et al. (ADELA ARPITHA; RANGARAJAN, 2020) propuseram um método de segmentação e classificação baseado em forma e textura para diferenciar FVCs benignas e malignas de 62 pacientes. Eles usaram o mesmo conjunto de atributos usados em (FRIGHETTO-PEREIRA et al., 2016) e (FRIGHETTO-PEREIRA et al., 2015). A técnica SMOTE foi utilizada para aumentar os dados, resultando em 108 fraturas benignas e 96 malignas. O algoritmo SVM foi escolhido como classificador e a validação cruzada de 6 *folds* foi usada para avaliar o modelo. Os autores não utilizaram acurácia e AUC ROC para avaliação, mas foi possível estimar a acurácia a partir do valor de revocação e do número de corpos vertebrais utilizados, obtendo-se um valor de aproximadamente 93,1%.

Todos os estudos supracitados avaliaram apenas a fatia central dos exames de RM sagitais ponderados em T1. No presente estudo, utilizamos uma abordagem tridimensional, que permitiu extrair mais informações das imagens (ORTIZ-RAMON et al., 2017). Além disso, nenhum desses estudos utilizou um conjunto de teste independente para validar os modelos propostos, restringindo a técnica de avaliação apenas à validação cruzada. Um conjunto de teste independente é necessário porque suas instâncias não são usadas anteriormente pelo classificador e, assim, seus resultados tornam-se mais confiáveis. No presente trabalho, nós não só realizamos a avaliação com um conjunto de teste independente como ainda incluímos nesse conjunto apenas um corpo vertebral fraturado por paciente, o que contribuiu ainda mais para a redução de viés na avaliação do modelo.

Foram encontrados ainda na literatura dois trabalhos recentes que utilizaram *machine learning* e *deep learning* para realizar o diagnóstico diferencial de FVCs benignas e malignas em imagens, porém utilizando imagens de tomografia computadorizada. (CHEE et al., 2021) avaliaram a predição de malignidade em fraturas vertebrais usando um modelo clínico-radiômico combinado em exames de TC. Eles utilizaram a regressão LASSO para selecionar as características e treinar o modelo por meio da validação cruzada de 3 *folds*. Ainda nesse estudo, foi utilizado um conjunto de teste independente para validar o modelo. Com um total de nove características selecionadas (oito de primeira ordem e 1 de textura), eles alcançaram uma AUC ROC de 0,941, acurácia de 94,5%, sensibilidade de 91,7% e especificidade de 96,8% para o conjunto de treinamento e uma AUC ROC de 0,852, acurácia de 87,3%, sensibilidade de 84,0% e especificidade de 90,0% para o conjunto de teste.

(LI et al., 2021) utilizaram *deep learning* com a arquitetura ResNet50 para diferenciar FVCs benignas e malignas. Eles utilizaram anotação manual para marcar a ROI (corpo vertebral fraturado) no corte do exame onde a anormalidade estava mais presente e, a partir dessa anotação, um software criou automaticamente a menor *bounding box* ao redor da ROI. O número de canais de entrada da rede foram 3, sendo a ROI do corte escolhido para anotação e os dois cortes vizinhos adjacentes. Utilizando validação cruzada de 10 *folds*, eles alcançaram uma acurácia de 85%, sensibilidade de 90% e especificidade de 79%.

Embora ambos os estudos tenham alcançado ótimos resultados, os nossos foram superiores tanto utilizando validação cruzada quanto o conjunto de teste independente (este último para o caso do trabalho de (CHEE et al., 2021), que foi o que realizou este tipo de análise). Como os próprios autores mencionam, a ressonância magnética é a modalidade de escolha para o diagnóstico diferencial das fraturas por compressão vertebral por ser mais sensível para detectar infiltração neoplásica da medula óssea. Além disso, uma vantagem substancial da RM é que ela não usa radiação ionizante.

5.4 LIMITAÇÕES DO ESTUDO

Nosso estudo possui algumas limitações. A primeira delas é que usamos apenas sequências do plano sagital ponderadas em T1. Na prática clínica, embora essa seja a configuração mais comum para avaliar FVCs, os radiologistas utilizam diferentes

sequências e planos anatômicos para obter uma caracterização mais completa da lesão. Além disso, em nosso trabalho avaliamos apenas as fraturas de vértebras lombares.

Outra limitação de nosso estudo foi a segmentação ter sido realizada por meio de um método semiautomático. Para uma possível implantação do modelo na prática clínica, seria necessário ter uma técnica totalmente automática para segmentar os corpos vertebrais fraturados ou, pelo menos, um método com a mínima interferência inicial do radiologista. Nossos esforços, nesse sentido, estão direcionados para o uso das redes U-Net para segmentação. Embora nosso trabalho tenha apresentado análises com informações provenientes dessa rede, não tivemos acesso à etapa da segmentação propriamente dita. Por isso, estamos trabalhando para concluir esse estudo e realizar uma análise comparativa do modelo aqui proposto utilizando tanto a segmentação pelo software 3D Slicer quanto pela rede neural U-Net.

Por fim, embora a avaliação com um conjunto de teste independente tenha sido realizada em nosso trabalho, esse conjunto incluiu apenas pacientes da mesma instituição do conjunto de treinamento. Em um cenário ideal, uma avaliação com pacientes de outra instituição, com um n considerável e com exames provenientes de outros aparelhos de ressonância magnética seria fundamental antes de se pensar em qualquer tipo de implantação do modelo na rotina clínica.

6 CONCLUSÃO

A radiologia vem passando por importantes mudanças com uma incorporação cada vez maior da tecnologia em sua rotina. A radiômica, embora ainda concentrada no âmbito da pesquisa, tem apresentado excelentes resultados e se mostrado como uma das ferramentas mais promissoras para conduzir a radiologia na direção da medicina de precisão. Este trabalho investigou se características radiômicas extraídas tridimensionalmente de exames de ressonância magnética e inseridas em um modelo computacional poderiam diferenciar fraturas vertebrais por compressão benignas e malignas.

Um modelo de rede neural artificial com base em 10 atributos radiômicos tridimensionais extraídos de exames de ressonância magnética da coluna vertebral foi desenvolvido, validado e testado, se mostrando capaz de classificar, com excelente performance, fraturas vertebrais por compressão benignas e malignas, superando os resultados encontrados na literatura. Este modelo tem potencial de servir como base para o desenvolvimento de uma ferramenta voltada para uso em ambiente clínico. Sua provável maior contribuição seria aumentar a eficiência diagnóstica dos radiologistas para essas duas condições clínicas, possibilitando uma eventual redução do número de biópsias realizadas e do tempo gasto para liberação do diagnóstico final. Isso traria enormes benefícios para os pacientes, uma vez que quanto mais precoce é o diagnóstico, maiores são as chances de cura. Para isso, no entanto, seria necessário cumprir algumas etapas como, por exemplo, incluir um método de segmentação automática, ampliar os testes por meio de estudos prospectivos simulando a rotina clínica, solicitar aprovação da Anvisa e garantir protocolos de monitoramento e avaliação contínuos.

Este trabalho apresentou ainda resultados de análises exploratórias comparando diferentes métodos para cada uma das etapas do reconhecimento de padrão em imagens médicas. Essas análises podem auxiliar e direcionar a escolha metodológica de trabalhos futuros, uma vez que os pesquisadores podem se basear nesses resultados para eliminar técnicas (ou conjunto de técnicas) que não obtiveram um bom desempenho.

Uma última contribuição de nosso trabalho foi a criação de uma base de dados seguindo as recomendações propostas por (KUMAR et al., 2012), que visa propiciar a reprodutibilidade de estudos radiômicos. A organização dessa base contemplou desde a adequação dos padrões e formatos propostos além do cumprimento de protocolos para a

devida anonimização dos exames, identificação única, inclusão de dados clínicos, segmentações, características quantitativas e demais documentos que possam caracterizar a base.

Como trabalhos futuros, algumas possibilidades seriam incluir fraturas de corpos vertebrais das regiões cervical e torácica e reavaliar o modelo realizando ajustes se necessário; investigar se a inclusão de outras sequências de RM e de outros cortes melhoraria a performance do modelo e investigar técnicas de detecção de objetos e segmentação automática com o objetivo de automatizar o *pipeline* completo.

7 CONSIDERAÇÕES FINAIS

Nesta seção, apresentaremos a produção bibliográfica gerada direta e indiretamente pela pesquisa bem como algumas atividades que foram desenvolvidas durante o doutorado e que contribuíram para a evolução da autora enquanto pesquisadora.

Produção Direta

CHIARI-CORREIA, N. S.; NOGUEIRA-BARBOSA, M. H.; AZEVEDO-MARQUES, P. M. A 3D radiomics-based artificial neural network model for vertebral compression fracture classification in MRI. Submetido para *European Radiology*, 2021.

Produção Indireta

DEL LAMA, R.; CANDIDO, R. M.; CHIARI-CORREIA, N. S.; NOGUEIRA-BARBOSA, M. H.; AZEVEDO-MARQUES, P. M.; , TINÓS, R. Computer-aided Diagnosis of Vertebral Compression Fractures using Convolutional Neural Networks and Radiomics. Aceito para publicação no *Journal of Digital Imaging*. 2022.

FERREIRA-JUNIOR, J. R.; KOENIGKAM-SANTOS, M.; VITA-GRAVES, C.; CHIARI-CORREIA, N. S.; CIPRIANO, F. E. G.; FABRO, A. T.; AZEVEDO-MARQUES, P. M. Quantifying intratumor heterogeneity of lung neoplasms with radiomics. *Clinical Imaging*, v. 74, p. 27-30, 2021.

CANDIDO, R. M.; CHIARI-CORREIA, N. S.; NOGUEIRA-BARBOSA, M. H.; AZEVEDO-MARQUES, P. M.; TINÓS, R. Impact of Feature Selection on Clustering Images of Vertebral Compression Fractures. *Anais do XVII Encontro Nacional de Inteligência Artificial e Computacional*. p. 752-763, 2020.

FERREIRA JUNIOR, J. R.; KOENIGKAM-SANTOS, M.; MACHADO, C. V. B.; FALEIROS, M. C.; CHIARI-CORREIA, N. S.; CIPRIANO, F. E. G.; FABRO, A. T.; AZEVEDO-MARQUES, P. M. Análise radiômica do câncer de pulmão para avaliação prognóstica do paciente e da heterogeneidade intratumoral. *Radiologia Brasileira*, v. 1, p. 1, 2020.

FERREIRA JUNIOR, J. R.; CHIARI-CORREIA, N. S.; MARQUES, P. M. A. Aprendizado de Máquina na Atenção à Saúde Humana. *Computação Brasil*, Porto Alegre, p. 37 - 40, 2019.

Estágio no Exterior

Doutorado Sanduíche realizado sob supervisão do Prof. Dr. Hiroyuki Yoshida no Massachusetts General Hospital da Harvard Medical School no período de 12/2019 a 05/2020 (ANEXO A).

Palestras

Segmentação em Imagens Médicas. *XVIII Semana Brasileira de Informática Biomédica*.
Ribeirão Preto, 2020.

Classificação computadorizada de câncer em imagens de RM. *XVI Semana Brasileira de Informática Biomédica*. Ribeirão Preto, 2018.

Inteligência Artificial Aplicada à Radiologia. *FATESA - Faculdade de Tecnologia em Saúde*.
Ribeirão Preto, 2018.

REFERÊNCIAS⁹

ADELA ARPITHA; RANGARAJAN, Lalitha. Computational techniques to segment and classify lumbar compression fractures. **La radiologia medica**, [S. l.], v. 125, n. 6, p. 551–560, 2020. DOI: 10.1007/s11547-020-01145-7. Disponível em: <http://link.springer.com/10.1007/s11547-020-01145-7>.

AERTS, Hugo J. W. L. et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. **Nature Communications**, [S. l.], v. 5, n. 1, p. 4006, 2014. DOI: 10.1038/ncomms5006. Disponível em: <http://www.nature.com/doi/10.1038/ncomms5006>.

ALEXANDRU, Daniela. Evaluation and Management of Vertebral Compression Fractures. **The Permanente Journal**, [S. l.], p. 46–51, 2012. DOI: 10.7812/TPP/12-037. Disponível em: <http://www.thepermanentejournal.org/issues/2012/fall/4906-management-of-vertebral-compression-fractures.html>.

AMADASUN, M.; KING, R. Textural features corresponding to textural properties. **IEEE Transactions on Systems, Man, and Cybernetics**, [S. l.], v. 19, n. 5, p. 1264–1274, 1989. DOI: 10.1109/21.44046. Disponível em: <http://ieeexplore.ieee.org/document/44046/>.

ARANA, Estanislao et al. Metastatic Versus Osteoporotic Vertebral Fractures on MRI: A Blinded, Multicenter, and Multispecialty Observer Agreement Evaluation. **Journal of the National Comprehensive Cancer Network**, [S. l.], v. 18, n. 3, p. 267–273, 2020. DOI: 10.6004/jnccn.2019.7367. Disponível em: <https://jnccn.org/view/journals/jnccn/18/3/article-p267.xml>.

AZEVEDO-MARQUES, P. M.; SPAGNOLI, H. F.; FRIGHETTO-PEREIRA, L.; MENEZES-REIS, R.; METZNER, G. A.; RANGAYAN, R. M.; NOGUEIRA-BARBOSA, M. H. Classification of vertebral compression fractures in magnetic resonance images using spectral and fractal analysis. In: 2015 37TH ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY (EMBC) 2015, **Anais [...]**. : IEEE, 2015. p. 723–726. DOI: 10.1109/EMBC.2015.7318464. Disponível em: <https://ieeexplore.ieee.org/document/7318464/>.

BABB, Aaron; CARLSON, Walter O. Vertebral compression fractures: treatment and evaluation. **South Dakota medicine : the journal of the South Dakota State Medical Association**, [S. l.], v. 59, n. 8, p. 343–5, 347, 2006. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/16941851>.

BAUR, Andrea; STÄBLER, Axel; ARBOGAST, Susanne; DUERR, Hans Roland; BARTL, Reiner; REISER, Maximilian. Acute Osteoporotic and Neoplastic Vertebral Compression Fractures: Fluid Sign at MR Imaging. **Radiology**, [S. l.], v. 225, n. 3, p. 730–735, 2002. DOI: 10.1148/radiol.2253011413. Disponível em: <http://pubs.rsna.org/doi/10.1148/radiol.2253011413>.

BISHOP, Christopher M. **Pattern recognition and machine learning**. New York:

⁹ De acordo com a Associação Brasileira de Normas Técnicas (ABNT NBR 6023).

Springer, 2006.

BREIMAN, Leo. Bagging predictors. **Machine Learning**, [S. l.], v. 24, n. 2, p. 123–140, 1996. DOI: 10.1007/BF00058655. Disponível em: <http://link.springer.com/10.1007/BF00058655>.

BUSHBERG, Jerrold T.; BOONE, John M. **The essential physics of medical imaging**. 2nd. ed. [s.l.] : Lippincott Williams & Wilkins, 2011.

BYRA, Michal; JAROSIK, Piotr; SZUBERT, Aleksandra; GALPERIN, Michael; OJEDA-FOURNIER, Haydee; OLSON, Linda; O'BOYLE, Mary; COMSTOCK, Christopher; ANDRE, Michael. Breast mass segmentation in ultrasound with selective kernel U-Net convolutional neural network. **Biomedical Signal Processing and Control**, [S. l.], v. 61, p. 102027, 2020. a. DOI: 10.1016/j.bspc.2020.102027. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S174680942030183X>.

BYRA, Michal; WU, Mei; ZHANG, Xiaodong; JANG, Hyungseok; MA, Ya-Jun; CHANG, Eric Y.; SHAH, Sameer; DU, Jiang. Knee menisci segmentation and relaxometry of 3D ultrashort echo time cones MR imaging using attention U-Net with transfer learning. **Magnetic Resonance in Medicine**, [S. l.], v. 83, n. 3, p. 1109–1122, 2020. b. DOI: 10.1002/mrm.27969. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/mrm.27969>.

CASTI, Paola; MENCATTINI, Arianna; NOGUEIRA-BARBOSA, Marcello H.; FRIGHETTO-PEREIRA, Lucas; AZEVEDO-MARQUES, Paulo Mazzoncini; MARTINELLI, Eugenio; DI NATALE, Corrado. Cooperative strategy for a dynamic ensemble of classification models in clinical applications: the case of MRI vertebral compression fractures. **International Journal of Computer Assisted Radiology and Surgery**, [S. l.], v. 12, n. 11, p. 1971–1983, 2017. DOI: 10.1007/s11548-017-1625-2. Disponível em: <http://link.springer.com/10.1007/s11548-017-1625-2>.

CESSIE, S. Le; HOUWELINGEN, J. C. Van. Ridge Estimators in Logistic Regression. **Applied Statistics**, [S. l.], v. 41, n. 1, p. 191, 1992. DOI: 10.2307/2347628. Disponível em: <https://www.jstor.org/stable/10.2307/2347628?origin=crossref>.

CHEE, Choong Guen; YOON, Min A.; KIM, Kyung Won; KO, Yusun; HAM, Su Jung; CHO, Young Chul; PARK, Bumwoo; CHUNG, Hye Won. Combined radiomics-clinical model to predict malignancy of vertebral compression fractures on CT. **European Radiology**, [S. l.], v. 31, n. 9, p. 6825–6834, 2021. DOI: 10.1007/s00330-021-07832-x. Disponível em: <https://link.springer.com/10.1007/s00330-021-07832-x>.

CLARK, P. et al. The prevalence of radiographic vertebral fractures in Latin American countries: the Latin American Vertebral Osteoporosis Study (LAVOS). **Osteoporosis International**, [S. l.], v. 20, n. 2, p. 275–282, 2009. DOI: 10.1007/s00198-008-0657-4. Disponível em: <http://link.springer.com/10.1007/s00198-008-0657-4>.

COLEMAN, R. E. Metastatic bone disease: clinical features, pathophysiology and treatment strategies. **Cancer Treatment Reviews**, [S. l.], v. 27, n. 3, p. 165–176, 2001. DOI: 10.1053/ctrv.2000.0210. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S030573720090210X>.

COVER, T.; HART, P. Nearest neighbor pattern classification. **IEEE Transactions on Information Theory**, [S. l.], v. 13, n. 1, p. 21–27, 1967. DOI: 10.1109/TIT.1967.1053964. Disponível em: <http://ieeexplore.ieee.org/document/1053964/>.

CUÉNOD, C. A.; LAREDO, J. D.; CHEVRET, S.; HAMZE, B.; NAOURI, J. F.; CHAPAUX, X.; BONDEVILLE, J. M.; TUBIANA, J. M. Acute vertebral collapse due to osteoporosis or malignancy: appearance on unenhanced and gadolinium-enhanced MR images. **Radiology**, [S. l.], v. 199, n. 2, p. 541–549, 1996. DOI: 10.1148/radiology.199.2.8668809. Disponível em: <http://pubs.rsna.org/doi/10.1148/radiology.199.2.8668809>.

FABIJAŃSKA, Anna. Segmentation of corneal endothelium images using a U-Net-based convolutional neural network. **Artificial Intelligence in Medicine**, [S. l.], v. 88, p. 1–13, 2018. DOI: 10.1016/j.artmed.2018.04.004. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0933365718300575>.

FREUND, Yoav; SCHAPIRE, Robert E. Experiments with a New Boosting Algorithm. **Proceedings of the 13th International Conference on Machine Learning**, [S. l.], p. 148–156, 1996. DOI: 10.1.1.133.1040. Disponível em: <http://www.public.asu.edu/~jye02/CLASSES/Fall-2005/PAPERS/boosting-icml.pdf>.

FRIGHETTO-PEREIRA, Lucas; MENEZES-REIS, Rafael; METZNER, Guilherme Augusto; RANGAYYAN, Rangaraj Mandayam; AZEVEDO-MARQUES, Paulo Mazzoncini; NOGUEIRA-BARBOSA, Marcello Henrique. Semiautomatic Classification of Benign Versus Malignant Vertebral Compression Fractures Using Texture and Gray-Level Features in Magnetic Resonance Images. In: 2015 IEEE 28TH INTERNATIONAL SYMPOSIUM ON COMPUTER-BASED MEDICAL SYSTEMS 2015, **Anais [...]**. : IEEE, 2015. p. 88–92. DOI: 10.1109/CBMS.2015.37. Disponível em: <http://ieeexplore.ieee.org/document/7167463/>.

FRIGHETTO-PEREIRA, Lucas; RANGAYYAN, Rangaraj Mandayam; METZNER, Guilherme Augusto; DE AZEVEDO-MARQUES, Paulo Mazzoncini; NOGUEIRA-BARBOSA, Marcello Henrique. Shape, texture and statistical features for classification of benign and malignant vertebral compression fractures in magnetic resonance images. **Computers in Biology and Medicine**, [S. l.], v. 73, p. 147–156, 2016. DOI: 10.1016/j.compbimed.2016.04.006. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0010482516300932>.

GILLIES, Robert J.; KINAHAN, Paul E.; HRICAK, Hedvig. Radiomics: Images Are More than Pictures, They Are Data. **Radiology**, [S. l.], v. 278, n. 2, p. 563–577, 2016. DOI: 10.1148/radiol.2015151169. Disponível em: <http://pubs.rsna.org/doi/10.1148/radiol.2015151169>.

GONZALEZ, Rafael C.; WOODS, Richard E. **Digital Image Processing**. [s.l.] : Addison-Wesley, 1992.

GU, Yu-Tong; ZHU, Dong-Hui; LIU, Hai-Fei; ZHANG, Feng; MCGUIRE, Robert. Minimally invasive pedicle screw fixation combined with percutaneous vertebroplasty for preventing secondary fracture after vertebroplasty. **Journal of Orthopaedic Surgery and Research**, [S. l.], v. 10, n. 1, p. 31, 2015. DOI: 10.1186/s13018-015-0172-1. Disponível em: <http://josr-online.biomedcentral.com/articles/10.1186/s13018-015->

0172-1.

HARALICK, R. M. Statistical and Structural Approaches to Texture. *[S. l.]*, v. 67, n. 5, p. 786–804, 1979.

HUANG, Qing; SUN, Jinfeng; DING, Hui; WANG, Xiaodong; WANG, Guangzhi. Robust liver vessel extraction using 3D U-Net with variant dice loss function. **Computers in Biology and Medicine**, *[S. l.]*, v. 101, p. 153–162, 2018. DOI: 10.1016/j.compbimed.2018.08.018. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0010482518302385>.

JACOBS, W. Bradley; PERRIN, Richard G. Evaluation and treatment of spinal metastases: an overview. **Neurosurgical Focus**, *[S. l.]*, v. 11, n. 6, p. 1–11, 2001. DOI: 10.3171/foc.2001.11.6.11. Disponível em: <https://thejns.org/view/journals/neurosurg-focus/11/6/foc.2001.11.6.11.xml>.

JAMES, Gareth; WITTEN, Daniela; HASTIE, Trevor; TIBSHIRANI, Robert. **An introduction to statistical learning**. New York: Springer, 2013.

KATAGIRI, Hirohisa; TAKAHASHI, Mitsuru; INAGAKI, Jiro; SUGIURA, Hideshi; ITO, Shigeki; IWATA, Hisashi. Determining the site of the primary cancer in patients with skeletal metastasis of unknown origin: A retrospective study. **Cancer**, *[S. l.]*, v. 86, n. 3, p. 533–537, 1999. DOI: 10.1002/(SICI)1097-0142(19990801)86:3<533::AID-CNCR24>3.0.CO;2-1.

KLIMO, P. Surgical Management of Spinal Metastases. **The Oncologist**, *[S. l.]*, v. 9, n. 2, p. 188–196, 2004. DOI: 10.1634/theoncologist.9-2-188. Disponível em: <http://theoncologist.alphamedpress.org/cgi/doi/10.1634/theoncologist.9-2-188>.

KOENIGKAM-SANTOS, Macel; FERREIRA-JUNIOR, José Raniery; WADA, Danilo Tadao; TENÓRIO, Ariane Priscilla Magalhães; NOGUEIRA-BARBOSA, Marcello Henrique; AZEVEDO-MARQUES, Paulo Mazzoncini. Artificial intelligence , machine learning , computer-aided diagnosis , and radiomics : advances in imaging towards to precision medicine. *[S. l.]*, p. 1–10, 2019. DOI: <http://dx.doi.org/10.1590/0100-3984.2019.0049>.

KUMAR, Vipin. Feature Selection: A literature Review. **The Smart Computing Review**, *[S. l.]*, v. 4, n. 3, 2014. DOI: 10.6029/smartcr.2014.03.007. Disponível em: <https://www.cc.gatech.edu/~hic/CS7616/Papers/Kumar-Minz-2014.pdf>.

KUMAR, Virendra et al. Radiomics: the process and the challenges. **Magnetic Resonance Imaging**, *[S. l.]*, v. 30, n. 9, p. 1234–1248, 2012. DOI: 10.1016/j.mri.2012.06.010. Disponível em: <http://dx.doi.org/10.1016/j.mri.2012.06.010>.

LAMBIN, Philippe et al. Radiomics: Extracting more information from medical images using advanced feature analysis. **European Journal of Cancer**, *[S. l.]*, v. 48, n. 4, p. 441–446, 2012. DOI: 10.1016/j.ejca.2011.11.036. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0959804911009993>.

LECOUVET, F. E.; VANDE BERG, B. C.; MALDAGUE, B. E.; MICHAUX, L.; LATERRE, E.; MICHAUX, J. L.; FERRANT, A.; MALGHEM, J. Vertebral compression fractures in multiple myeloma. Part I. Distribution and appearance at MR imaging. **Radiology**, *[S. l.]*, v. 204, n.

1, p. 195–199, 1997. DOI: 10.1148/radiology.204.1.9205246. Disponível em: <http://pubs.rsna.org/doi/10.1148/radiology.204.1.9205246>.

LECOUVET, Frédéric E. Whole-Body MR Imaging: Musculoskeletal Applications. **Radiology**, [S. l.], v. 279, n. 2, p. 345–365, 2016. DOI: 10.1148/radiol.2016142084. Disponível em: <http://pubs.rsna.org/doi/10.1148/radiol.2016142084>.

LEE, Young Joon; JEE, Won Hee; HA, Kee Yong; LEE, Bae Young; KIM, Yeon Shil; KIM, Bum Soo; SUH, Kyung Jin; CHOI, Kyu Ho. MR Distinction between Multiple Myeloma and Metastasis Involving the Spine. **Journal of the Korean Radiological Society**, [S. l.], v. 44, n. 2, p. 229, 2001. DOI: 10.3348/jkrs.2001.44.2.229. Disponível em: <http://synapse.koreamed.org/DOIx.php?id=10.3348/jkrs.2001.44.2.229>.

LI, Jundong; CHENG, Kewei; WANG, Suhang; MORSTATTER, Fred; TREVINO, Robert P.; TANG, Jiliang; LIU, Huan. Feature Selection. **ACM Computing Surveys**, [S. l.], v. 50, n. 6, p. 1–45, 2018. DOI: 10.1145/3136625. Disponível em: <http://dl.acm.org/citation.cfm?doid=3161158.3136625>.

LI, Yuan et al. Differential diagnosis of benign and malignant vertebral fracture on CT using deep learning. **European radiology**, [S. l.], v. 31, n. 12, p. 9612–9619, 2021. DOI: 10.1007/s00330-021-08014-5. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/33993335>.

LOOG, Marco. Supervised Classification: Quite a Brief Overview. In: **Machine Learning Techniques for Space Weather**. [s.l.] : Elsevier, 2018. p. 113–145. DOI: 10.1016/B978-0-12-811788-0.00005-6. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/B9780128117880000056>.

MACCAURO, Giulio; SPINELLI, Maria Silvia; MAURO, Sigismondo; PERISANO, Carlo; GRACI, Calogero; ROSA, Michele Attilio. Physiopathology of Spine Metastasis. **International Journal of Surgical Oncology**, [S. l.], v. 2011, p. 1–8, 2011. DOI: 10.1155/2011/107969. Disponível em: <http://www.hindawi.com/journals/ijso/2011/107969/>.

MACEDO, Filipa; LADEIRA, Katia; PINHO, Filipa; SARAIVA, Nadine; BONITO, Nuno; PINTO, Luísa; GONÇALVES, Francisco. Bone metastases: an overview. **Oncology Reviews**, [S. l.], v. 11, n. 1, 2017. DOI: 10.4081/oncol.2017.321. Disponível em: <http://www.oncologyreviews.org/index.php/or/article/view/321>.

MAUCH, J. T.; CARR, C. M.; CLOFT, H.; DIEHN, F. E. Review of the Imaging Features of Benign Osteoporotic and Malignant Vertebral Compression Fractures. **American Journal of Neuroradiology**, [S. l.], 2018. DOI: 10.3174/ajnr.A5528. Disponível em: <http://www.ajnr.org/lookup/doi/10.3174/ajnr.A5528>.

MAZZOLA, Alessandro A. Ressonância magnética: princípios de formação da imagem e aplicações em imagem funcional. **Revista Brasileira de Física Médica**, [S. l.], v. 3, n. 1, p. 117–129, 2009.

MELTON III, L. J.; ATKINSON, E. J.; COOPER, C.; O'FALLON, W. M.; RIGGS, B. L. Vertebral Fractures Predict Subsequent Fractures. **Osteoporosis International**, [S. l.], v. 10, n. 3, p. 214–221, 1999. DOI: 10.1007/s001980050218. Disponível em:

<http://link.springer.com/10.1007/s001980050218>.

MOREAU, P. et al. Multiple myeloma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. **Annals of Oncology**, [S. l.], v. 28, n. suppl_4, p. iv52–iv61, 2017. DOI: 10.1093/annonc/mdx096. Disponível em: https://academic.oup.com/annonc/article/28/suppl_4/iv52/3768071.

NAPÉL, Sandy; MU, Wei; JARDIM-PERASSI, Bruna V.; AERTS, Hugo J. W. L.; GILLIES, Robert J. Quantitative imaging of cancer in the postgenomic era: Radio(geno)mics, deep learning, and habitats. **Cancer**, [S. l.], v. 124, n. 24, p. 4633–4649, 2018. DOI: 10.1002/cncr.31630. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/cncr.31630>.

NIOCHE, Christophe et al. LIFEx: A Freeware for Radiomic Feature Calculation in Multimodality Imaging to Accelerate Advances in the Characterization of Tumor Heterogeneity. **Cancer Research**, [S. l.], v. 78, n. 16, p. 4786–4789, 2018. DOI: 10.1158/0008-5472.CAN-18-0125. Disponível em: <http://cancerres.aacrjournals.org/lookup/doi/10.1158/0008-5472.CAN-18-0125>.

NORMAN, Berk; PEDOIA, Valentina; MAJUMDAR, Sharmila. Use of 2D U-Net Convolutional Neural Networks for Automated Cartilage and Meniscus Segmentation of Knee MR Imaging Data to Determine Relaxometry and Morphometry. **Radiology**, [S. l.], v. 288, n. 1, p. 177–185, 2018. DOI: 10.1148/radiol.2018172322. Disponível em: <http://pubs.rsna.org/doi/10.1148/radiol.2018172322>.

OLD, JERRY L.; CALVERT, MICHELLE. Vertebral Compression Fractures in the Elderly. **American family physician**, [S. l.], v. 69, n. 1, p. 111–116, 2004.

ORTIZ-RAMON, Rafael; LARROZA, Andres; ARANA, Estanislao; MORATAL, David. A radiomics evaluation of 2D and 3D MRI texture features to classify brain metastases from lung cancer and melanoma. In: 2017 39TH ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY (EMBC) 2017, **Anais [...]**. : IEEE, 2017. p. 493–496. DOI: 10.1109/EMBC.2017.8036869. Disponível em: <https://ieeexplore.ieee.org/document/8036869/>.

PARK, Daniel K.; JENIS, Lowis G. **Osteoporosis and Spinal Fractures**. 2021. Disponível em: <https://orthoinfo.aaos.org/en/diseases--conditions/osteoporosis-and-spinal-fractures>. Acesso em: 10 jan. 2022.

PARK, Jongha; YUN, Jihye; KIM, Namkug; PARK, Beomhee; CHO, Yongwon; PARK, Hee Jun; SONG, Mijeong; LEE, Minho; SEO, Joon Beom. Fully Automated Lung Lobe Segmentation in Volumetric Chest CT with 3D U-Net: Validation with Intra- and Extra-Datasets. **Journal of Digital Imaging**, [S. l.], v. 33, n. 1, p. 221–230, 2020. DOI: 10.1007/s10278-019-00223-1. Disponível em: <http://link.springer.com/10.1007/s10278-019-00223-1>.

PARMAR, Chintan; GROSSMANN, Patrick; BUSSINK, Johan; LAMBIN, Philippe; AERTS, Hugo J. W. L. Machine Learning methods for Quantitative Radiomic Biomarkers. **Scientific Reports**, [S. l.], v. 5, p. 13087, 2015. DOI: 10.1038/srep13087. Disponível em: <http://dx.doi.org/10.1038/srep13087>.

PICCIOLI, Andrea; MACCAURO, Giulio; SPINELLI, Maria Silvia; BIAGINI, Roberto; ROSSI, Barbara. Bone metastases of unknown origin: epidemiology and principles of management. **Journal of Orthopaedics and Traumatology**, [S. l.], v. 16, n. 2, p. 81–86, 2015. DOI: 10.1007/s10195-015-0344-0. Disponível em: <http://link.springer.com/10.1007/s10195-015-0344-0>.

PORTER, B. A.; SHIELDS, A. F.; OLSON, D. O. Magnetic resonance imaging of bone marrow disorders. **Radiologic clinics of North America**, [S. l.], v. 24, n. 2, p. 269–89, 1986. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/3715001>.

PUNN, Narinder Singh; AGARWAL, Sonali. Modality specific U-Net variants for biomedical image segmentation: A survey. [S. l.], 2021. Disponível em: <http://arxiv.org/abs/2107.04537>.

QUINLAN, J. Ross. **C4. 5: programs for machine learning**. [s.l.] : Elsevier, 2014.

RAJKUMAR, S. Vincent; KUMAR, Shaji. Multiple Myeloma: Diagnosis and Treatment. **Mayo Clinic Proceedings**, [S. l.], v. 91, n. 1, p. 101–119, 2016. DOI: 10.1016/j.mayocp.2015.11.007. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0025619615008952>.

REFAEILZADEH, Payam; TANG, Lei; LIU, Huan. Cross-Validation. In: **Encyclopedia of Database Systems**. New York, NY: Springer New York, 2016. p. 1–7. DOI: 10.1007/978-1-4899-7993-3_565-2. Disponível em: http://link.springer.com/10.1007/978-1-4899-7993-3_565-2.

RONNEBERGER, Olaf; FISCHER, Philipp; BROX, Thomas. U-Net: Convolutional Networks for Biomedical Image Segmentation. [S. l.], 2015. Disponível em: <http://arxiv.org/abs/1505.04597>.

SELVAGGI, Giovanni; SCAGLIOTTI, Giorgio V. Management of bone metastases in cancer: A review. **Critical Reviews in Oncology/Hematology**, [S. l.], v. 56, n. 3, p. 365–378, 2005. DOI: 10.1016/j.critrevonc.2005.03.011. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1040842805000661>.

SHAH, Lubdha M. et al. ACR Appropriateness Criteria® Management of Vertebral Compression Fractures. **Journal of the American College of Radiology**, [S. l.], v. 15, n. 11, p. S347–S364, 2018. DOI: 10.1016/j.jacr.2018.09.019. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S154614401831161X>.

SOLEM, Jan Erik. **Programming Computer Vision with Python: Tools and algorithms for analyzing images**. [s.l.] : O'Reilly Media, 2012.

SUH, Chong Hyun; YUN, Seong Jong; JIN, Wook; LEE, Sun Hwa; PARK, So Young; RYU, Chang-Woo. ADC as a useful diagnostic tool for differentiating benign and malignant vertebral bone marrow lesions and compression fractures: a systematic review and meta-analysis. **European Radiology**, [S. l.], v. 28, n. 7, p. 2890–2902, 2018. DOI: 10.1007/s00330-018-5330-5. Disponível em: <http://link.springer.com/10.1007/s00330-018-5330-5>.

TANG, Jiliang; ALELYANI, Salem; LIU, Huan. Feature selection for classification: A review.

Data classification: Algorithms and applications, [S. l.], p. 37, 2014.

TEHRANZADEH, Jamshid; TAO, Cliff. Advances in MR imaging of vertebral collapse.

Seminars in Ultrasound, CT and MRI, [S. l.], v. 25, n. 6, p. 440–460, 2004. DOI:

10.1053/j.sult.2004.09.002. Disponível em:

<https://linkinghub.elsevier.com/retrieve/pii/S0887217104000642>.

THAWAIT, Shrey K.; MARCUS, Matthew A.; MORRISON, William B.; KLUFAS, Roman A.; ENG, John; CARRINO, John A. Research synthesis: what is the diagnostic performance of magnetic resonance imaging to discriminate benign from malignant vertebral compression fractures? Systematic review and meta-analysis. **Spine**, [S. l.], v. 37, n. 12, p. E736-44, 2012. DOI: 10.1097/BRS.0b013e3182458cac. Disponível em:

<http://asianspinejournal.org/journal/view.php?doi=10.4184/asj.2017.11.3.478>.

VAN GRIETHUYSEN, Joost J. M. et al. Computational Radiomics System to Decode the Radiographic Phenotype. **Cancer Research**, [S. l.], v. 77, n. 21, p. e104–e107, 2017. DOI: 10.1158/0008-5472.CAN-17-0339. Disponível em:

<http://cancerres.aacrjournals.org/lookup/doi/10.1158/0008-5472.CAN-17-0339>.

VELHO, L.; FRERY, A. ..; GOMES, J. **Image Processing for Computer Graphics and Vision**. Second Edi ed. [s.l.] : Springer Science & Business Media, 2009.

APÊNDICE A

Tabela 9 - Teste de Normalidade Anderson-Darling

Característica	Grupo Fraturas Benignas			Grupo Fraturas Malignas			Teste**
	Estat.	Valor crítico	DN*	Estat.	Valor crítico	DN*	
firstorder_10Percentile	0.284	0.739	V	0.425	0.732	V	Teste-t
firstorder_90Percentile	0.589	0.739	V	0.433	0.732	V	Teste-t
firstorder_Energy	0.542	0.739	V	0.448	0.732	V	Teste-t
firstorder_Entropy	0.749	0.739	F	0.587	0.732	V	Mann Whitney
firstorder_InterquartileRange	1.511	0.739	F	0.633	0.732	V	Mann Whitney
firstorder_Kurtosis	0.761	0.739	F	0.280	0.732	V	Mann Whitney
firstorder_Maximum	0.740	0.739	F	0.313	0.732	V	Mann Whitney
firstorder_Mean	0.641	0.739	V	0.104	0.732	V	Teste-t
firstorder_MeanAbsoluteDeviation	1.450	0.739	F	0.654	0.732	V	Mann Whitney
firstorder_Median	0.510	0.739	V	0.210	0.732	V	Teste-t
firstorder_Minimum	0.295	0.739	V	0.409	0.732	V	Teste-t
firstorder_Range	0.711	0.739	V	0.306	0.732	V	Teste-t
firstorder_RobustMeanAbsoluteDeviation	1.524	0.739	F	0.708	0.732	V	Mann Whitney
firstorder_RootMeanSquared	0.635	0.739	V	0.178	0.732	V	Teste-t
firstorder_StandardDeviation	1.221	0.739	F	0.638	0.732	V	Mann Whitney
firstorder_Skewness	0.360	0.739	V	0.410	0.732	V	Teste-t
firstorder_TotalEnergy	0.542	0.739	V	0.448	0.732	V	Teste-t
firstorder_Uniformity	0.618	0.739	V	0.641	0.732	V	Teste-t
firstorder_Variance	1.732	0.739	F	0.845	0.732	F	Mann Whitney
glcm_Autocorrelation	0.370	0.739	V	0.830	0.732	F	Mann Whitney
glcm_ClusterProminence	2.289	0.739	F	1.754	0.732	F	Mann Whitney
glcm_ClusterShade	1.232	0.739	F	2.444	0.732	F	Mann Whitney
glcm_ClusterTendency	1.992	0.739	F	0.790	0.732	F	Mann Whitney
glcm_Contrast	1.112	0.739	F	0.700	0.732	V	Mann Whitney
glcm_Correlation	0.228	0.739	V	0.317	0.732	V	Teste-t
glcm_DifferenceAverage	0.734	0.739	V	0.404	0.732	V	Teste-t
glcm_DifferenceEntropy	0.818	0.739	F	0.495	0.732	V	Mann Whitney
glcm_DifferenceVariance	1.006	0.739	F	0.932	0.732	F	Mann Whitney
glcm_Id	0.508	0.739	V	0.273	0.732	V	Teste-t
glcm_Idm	0.556	0.739	V	0.301	0.732	V	Teste-t
glcm_Idmn	0.463	0.739	V	0.471	0.732	V	Teste-t
glcm_Idn	0.526	0.739	V	0.245	0.732	V	Teste-t
glcm_Imc1	0.232	0.739	V	0.787	0.732	F	Mann Whitney
glcm_Imc2	0.293	0.739	V	0.351	0.732	V	Teste-t
glcm_InverseVariance	1.954	0.739	F	1.211	0.732	F	Mann Whitney
glcm_JointAverage	0.315	0.739	V	0.900	0.732	F	Mann Whitney
glcm_JointEnergy	0.341	0.739	V	0.564	0.732	V	Teste-t
glcm_JointEntropy	0.515	0.739	V	0.681	0.732	V	Teste-t

(continua)

Característica	Grupo Fraturas Benignas			Grupo Fraturas Malignas			Teste**
	Estat.	Valor crítico	DN*	Estat.	Valor crítico	DN*	
glcm_MaximumProbability	0.430	0.739	V	0.320	0.732	V	Teste-t
glcm_SumAverage	0.315	0.739	V	0.900	0.732	F	Mann Whitney
glcm_SumEntropy	0.843	0.739	F	0.485	0.732	V	Mann Whitney
glcm_SumSquares	1.848	0.739	F	0.887	0.732	F	Mann Whitney
gldm_DependenceEntropy	0.526	0.739	V	0.308	0.732	V	Teste-t
gldm_DependenceNonUniformity	0.776	0.739	F	0.153	0.732	V	Mann Whitney
gldm_DependenceNonUniformityNormalized	0.995	0.739	F	0.341	0.732	V	Mann Whitney
gldm_DependenceVariance	0.502	0.739	V	0.319	0.732	V	Teste-t
gldm_GrayLevelNonUniformity	0.420	0.739	V	0.243	0.732	V	Teste-t
gldm_GrayLevelVariance	1.739	0.739	F	0.834	0.732	F	Mann Whitney
gldm_HighGrayLevelEmphasis	0.297	0.739	V	0.816	0.732	F	Mann Whitney
gldm_LargeDependenceEmphasis	0.396	0.739	V	0.359	0.732	V	Teste-t
gldm_LargeDependenceHighGrayLevelEmphasis	0.367	0.739	V	0.512	0.732	V	Teste-t
gldm_LargeDependenceLowGrayLevelEmphasis	2.313	0.739	F	1.053	0.732	F	Mann Whitney
gldm_LowGrayLevelEmphasis	3.324	0.739	F	1.955	0.732	F	Mann Whitney
gldm_SmallDependenceEmphasis	2.074	0.739	F	0.318	0.732	V	Mann Whitney
gldm_SmallDependenceHighGrayLevelEmphasis	1.264	0.739	F	0.998	0.732	F	Mann Whitney
gldm_SmallDependenceLowGrayLevelEmphasis	1.728	0.739	F	1.730	0.732	F	Mann Whitney
glrlm_GrayLevelNonUniformity	0.354	0.739	V	0.447	0.732	V	Teste-t
glrlm_GrayLevelNonUniformityNormalized	0.697	0.739	V	0.613	0.732	V	Teste-t
glrlm_GrayLevelVariance	1.437	0.739	F	0.767	0.732	F	Mann Whitney
glrlm_HighGrayLevelRunEmphasis	0.273	0.739	V	0.869	0.732	F	Mann Whitney
glrlm_LongRunEmphasis	0.602	0.739	V	0.648	0.732	V	Teste-t
glrlm_LongRunHighGrayLevelEmphasis	0.136	0.739	V	0.750	0.732	F	Mann Whitney
glrlm_LongRunLowGrayLevelEmphasis	3.006	0.739	F	1.137	0.732	F	Mann Whitney
glrlm_LowGrayLevelRunEmphasis	3.780	0.739	F	2.114	0.732	F	Mann Whitney
glrlm_RunEntropy	0.499	0.739	V	0.558	0.732	V	Teste-t
glrlm_RunLengthNonUniformity	0.828	0.739	F	0.271	0.732	V	Mann Whitney
glrlm_RunLengthNonUniformityNormalized	0.443	0.739	V	0.239	0.732	V	Teste-t
glrlm_RunPercentage	0.378	0.739	V	0.290	0.732	V	Teste-t
glrlm_RunVariance	0.732	0.739	V	0.690	0.732	V	Teste-t
glrlm_ShortRunEmphasis	0.293	0.739	V	0.386	0.732	V	Teste-t
glrlm_ShortRunHighGrayLevelEmphasis	0.291	0.739	V	0.702	0.732	V	Teste-t
glrlm_ShortRunLowGrayLevelEmphasis	4.161	0.739	F	2.240	0.732	F	Mann Whitney
glszm_GrayLevelNonUniformity	0.407	0.739	V	0.174	0.732	V	Teste-t
glszm_GrayLevelNonUniformityNormalized	0.713	0.739	V	1.250	0.732	F	Mann Whitney
glszm_GrayLevelVariance	0.496	0.739	V	0.961	0.732	F	Mann Whitney
glszm_HighGrayLevelZoneEmphasis	0.296	0.739	V	0.857	0.732	F	Mann Whitney
glszm_LargeAreaEmphasis	1.041	0.739	F	0.284	0.732	V	Mann Whitney
glszm_LargeAreaHighGrayLevelEmphasis	0.945	0.739	F	1.328	0.732	F	Mann Whitney
glszm_LargeAreaLowGrayLevelEmphasis	2.341	0.739	F	0.884	0.732	F	Mann Whitney
glszm_LowGrayLevelZoneEmphasis	4.147	0.739	F	2.519	0.732	F	Mann Whitney

(continua)

Característica	Grupo Fraturas Benignas			Grupo Fraturas Malignas			Teste**
	Estat.	Valor crítico	DN*	Estat.	Valor crítico	DN*	
glszm_SizeZoneNonUniformity	0.875	0.739	F	0.271	0.732	V	Mann Whitney
glszm_SizeZoneNonUniformityNormalized	0.160	0.739	V	0.843	0.732	F	Mann Whitney
glszm_SmallAreaEmphasis	0.174	0.739	V	0.508	0.732	V	Teste-t
glszm_SmallAreaHighGrayLevelEmphasis	0.456	0.739	V	0.411	0.732	V	Teste-t
glszm_SmallAreaLowGrayLevelEmphasis	2.622	0.739	F	1.424	0.732	F	Mann Whitney
glszm_ZoneEntropy	0.203	0.739	V	0.645	0.732	V	Teste-t
glszm_ZonePercentage	4.257	0.739	F	0.442	0.732	V	Mann Whitney
glszm_ZoneVariance	1.060	0.739	F	0.273	0.732	V	Mann Whitney
ngtdm_Busyness	0.828	0.739	F	1.238	0.732	F	Mann Whitney
ngtdm_Coarseness	0.994	0.739	F	2.779	0.732	F	Mann Whitney
ngtdm_Complexity	0.653	0.739	V	0.525	0.732	V	Teste-t
ngtdm_Contrast	0.391	0.739	V	0.784	0.732	F	Mann Whitney
ngtdm_Strength	1.806	0.739	F	1.080	0.732	F	Mann Whitney

Nota.— *A hipótese nula H_0 é de que os dados seguem uma distribuição normal. Uma estatística de teste maior que o valor crítico significa que a hipótese H_0 deve ser rejeitada, ou seja, os dados não seguem uma distribuição normal. **Foi utilizado o teste-t apenas para os casos em que a variável seguia uma distribuição normal para ambos os grupos. Estat = estatística de teste; DN = distribuição normal; V = verdadeiro; F = Falso.

APÊNDICE B

Tabela 10 - AUC ROC com IC de 95% para as 93 características extraídas das imagens.

Atributos	AUC	IC de 95%	
		Inferior	Superior
ngtdm_Busyness	0.865	0.80	0.94
firstorder_RootMeanSquared	0.846	0.77	0.92
ngtdm_Strength	0.843	0.77	0.92
firstorder_Mean	0.841	0.76	0.92
firstorder_90Percentile	0.840	0.76	0.92
firstorder_Median	0.828	0.75	0.91
glszm_LargeAreaLowGrayLevelEmphasis	0.815	0.73	0.90
glcm_Autocorrelation	0.802	0.72	0.89
glcm_JointAverage	0.799	0.71	0.89
glcm_SumAverage	0.799	0.71	0.89
gldm_HighGrayLevelEmphasis	0.799	0.71	0.89
glrlm_HighGrayLevelRunEmphasis	0.795	0.71	0.88
glrlm_ShortRunHighGrayLevelEmphasis	0.794	0.71	0.88
gldm_LargeDependenceLowGrayLevelEmphasis	0.792	0.70	0.88
glrlm_LongRunLowGrayLevelEmphasis	0.788	0.70	0.88
gldm_SmallDependenceHighGrayLevelEmphasis	0.782	0.69	0.87
gldm_GrayLevelNonUniformity	0.776	0.68	0.87
firstorder_Maximum	0.772	0.68	0.86
glrlm_GrayLevelNonUniformity	0.772	0.68	0.87
ngtdm_Coarseness	0.771	0.67	0.87
glszm_ZoneVariance	0.766	0.67	0.86
gldm_LowGrayLevelEmphasis	0.765	0.67	0.86
glszm_LargeAreaEmphasis	0.764	0.67	0.86
ngtdm_Complexity	0.761	0.67	0.85
glrlm_LowGrayLevelRunEmphasis	0.758	0.66	0.85
glszm_SmallAreaHighGrayLevelEmphasis	0.752	0.66	0.85
firstorder_Range	0.750	0.65	0.85
glcm_DifferenceVariance	0.743	0.65	0.84
firstorder_10Percentile	0.738	0.64	0.84
glszm_SmallAreaEmphasis	0.737	0.64	0.84
glrlm_ShortRunLowGrayLevelEmphasis	0.736	0.64	0.83
glcm_DifferenceEntropy	0.731	0.63	0.83
glszm_HighGrayLevelZoneEmphasis	0.729	0.63	0.83
glrlm_ShortRunEmphasis	0.729	0.63	0.83
glrlm_LongRunHighGrayLevelEmphasis	0.726	0.62	0.83
glcm_ClusterProminence	0.725	0.62	0.83
glcm_JointEntropy	0.725	0.62	0.83
gldm_LargeDependenceHighGrayLevelEmphasis	0.725	0.62	0.83
glcm_Contrast	0.724	0.62	0.83
glszm_LowGrayLevelZoneEmphasis	0.723	0.62	0.82
glrlm_GrayLevelVariance	0.722	0.62	0.82
firstorder_Entropy	0.717	0.61	0.82
glrlm_GrayLevelNonUniformityNormalized	0.715	0.61	0.82
glrlm_RunLengthNonUniformityNormalized	0.713	0.61	0.81
gldm_GrayLevelVariance	0.712	0.61	0.82
gldm_DependenceNonUniformity	0.712	0.61	0.81
glcm_DifferenceAverage	0.711	0.61	0.81
firstorder_StandardDeviation	0.709	0.61	0.81
firstorder_Variance	0.709	0.61	0.81
glcm_JointEnergy	0.707	0.60	0.81

(continua)

Atributos	AUC	IC de 95%	
		Inferior	Superior
glrlm_RunPercentage	0.707	0.60	0.81
gldm_SmallDependenceEmphasis	0.706	0.60	0.81
glrlm_LongRunEmphasis	0.706	0.60	0.81
glcm_SumSquares	0.704	0.60	0.81
glcm_Idm	0.703	0.60	0.81
glcm_SumEntropy	0.702	0.60	0.81
gldm_LargeDependenceEmphasis	0.702	0.60	0.81
firstorder_Uniformity	0.701	0.60	0.81
glcm_Id	0.700	0.60	0.80
firstorder_MeanAbsoluteDeviation	0.698	0.59	0.80
firstorder_Skewness	0.695	0.59	0.80
glcm_ClusterTendency	0.694	0.59	0.80
glszm_GrayLevelNonUniformity	0.693	0.59	0.80
glszm_SizeZoneNonUniformityNormalized	0.691	0.59	0.80
glrlm_RunVariance	0.690	0.58	0.80
firstorder_RobustMeanAbsoluteDeviation	0.681	0.57	0.79
glcm_ClusterShade	0.680	0.57	0.79
firstorder_InterquartileRange	0.678	0.57	0.78
glcm_MaximumProbability	0.673	0.56	0.78
glszm_ZonePercentage	0.667	0.56	0.77
glszm_SmallAreaLowGrayLevelEmphasis	0.667	0.56	0.78
glrlm_RunLengthNonUniformity	0.655	0.55	0.76
gldm_DependenceNonUniformityNormalized	0.651	0.54	0.76
gldm_DependenceEntropy	0.650	0.54	0.76
firstorder_Kurtosis	0.649	0.54	0.76
glszm_GrayLevelNonUniformityNormalized	0.649	0.54	0.76
glszm_GrayLevelVariance	0.638	0.53	0.75
gldm_DependenceVariance	0.631	0.52	0.74
glcm_Idmn	0.624	0.51	0.74
glcm_Imc2	0.616	0.50	0.73
glcm_Correlation	0.612	0.50	0.73
gldm_SmallDependenceLowGrayLevelEmphasis	0.609	0.49	0.72
glrlm_RunEntropy	0.602	0.48	0.72
glcm_InverseVariance	0.598	0.48	0.71
glszm_LargeAreaHighGrayLevelEmphasis	0.592	0.48	0.70
glcm_Idn	0.583	0.47	0.70
glszm_ZoneEntropy	0.578	0.46	0.69
firstorder_Minimum	0.576	0.46	0.69
glcm_Imc1	0.574	0.45	0.69
ngtdm_Contrast	0.519	0.40	0.63
firstorder_Energy	0.512	0.40	0.63
firstorder_TotalEnergy	0.512	0.40	0.63
glszm_SizeZoneNonUniformity	0.505	0.39	0.62

APÊNDICE C

Quadro 14 - Ranking decrescente das características após aplicação dos filtros de seleção

Fisher Score	Trace Ratio	T Score	ReliefF
glcm_Id	glcm_Id	firstorder_Uniformity	glcm_SumSquares
firstorder_Uniformity	firstorder_Uniformity	glcm_SumEntropy	glcm_SumEntropy
glcm_SumEntropy	glcm_SumSquares	glcm_SumSquares	firstorder_Uniformity
glcm_Idm	glcm_Idm	glcm_Idm	glrlm_ShortRunLowGrayLevelEmphasis
glcm_SumSquares	glcm_SumEntropy	glcm_Id	gldm_LargeDependenceLowGrayLevelEmphasis
gldm_LargeDependenceEmphasis	glcm_JointEnergy	gldm_LargeDependenceEmphasis	glcm_Id
glrlm_LongRunEmphasis	glrlm_RunPercentage	gldm_SmallDependenceEmphasis	glcm_Idm
glcm_JointEnergy	gldm_SmallDependenceEmphasis	glrlm_RunPercentage	glrlm_GrayLevelVariance
glrlm_RunPercentage	glrlm_LongRunEmphasis	glcm_JointEnergy	gldm_GrayLevelNonUniformity
gldm_SmallDependenceEmphasis	firstorder_Variance	firstorder_Variance	glrlm_GrayLevelNonUniformityNormalized
firstorder_Entropy	gldm_LargeDependenceEmphasis	firstorder_Entropy	firstorder_Maximum
firstorder_Variance	firstorder_StandardDeviation	glrlm_LongRunEmphasis	firstorder_RootMeanSquared
firstorder_StandardDeviation	firstorder_Entropy	glcm_DifferenceAverage	firstorder_Mean
glcm_DifferenceAverage	glcm_DifferenceAverage	firstorder_StandardDeviation	gldm_LargeDependenceEmphasis
glrlm_GrayLevelNonUniformityNormalized	glrlm_GrayLevelNonUniformityNormalized	glrlm_GrayLevelNonUniformityNormalized	glrlm_HighGrayLevelRunEmphasis
glrlm_GrayLevelVariance	glrlm_GrayLevelVariance	glcm_ClusterProminence	glrlm_GrayLevelNonUniformity
glcm_ClusterProminence	glcm_ClusterProminence	glrlm_RunLengthNonUniformityNormalized	glcm_Contrast
gldm_DependenceNonUniformity	gldm_DependenceNonUniformity	glrlm_GrayLevelVariance	gldm_SmallDependenceEmphasis
gldm_GrayLevelVariance	gldm_GrayLevelVariance	glrlm_ShortRunEmphases	glcm_JointEntropy
glrlm_ShortRunEmphases	glrlm_ShortRunEmphases	glrlm_ShortRunLowGrayLevelEmphasis	glcm_JointEnergy
glszm_HighGrayLevelZoneEmphasis	glszm_HighGrayLevelZoneEmphasis	gldm_DependenceNonUniformity	glrlm_RunPercentage
glrlm_RunLengthNonUniformityNormalized	glcm_Contrast	gldm_GrayLevelVariance	glszm_LowGrayLevelZoneEmphasis
glcm_Contrast	glcm_JointEntropy	glszm_HighGrayLevelZoneEmphasis	glrlm_LongRunLowGrayLevelEmphasis
glcm_JointEntropy	glrlm_RunLengthNonUniformityNormalized	glcm_Contrast	gldm_HighGrayLevelEmphasis
glrlm_ShortRunLowGrayLevelEmphasis	glrlm_ShortRunLowGrayLevelEmphasis	glcm_JointEntropy	glszm_LargeAreaEmphasis
gldm_GrayLevelNonUniformity	glszm_LowGrayLevelZoneEmphasis	glszm_LowGrayLevelZoneEmphasis	firstorder_Entropy

(continua)

Fisher Score	Trace Ratio	T Score	Relieff
glszm_LargeAreaEmphas is	glszm_SmallAreaHighGra yLevelEmphasis	glszm_LargeAreaEmphas is	glrlm_LongRunEmphasis
gldm_LargeDependence HighGrayLevelEmphasis	gldm_GrayLevelNonUnifo rarity	gldm_GrayLevelNonUnifo rarity	glcm_SumAverage
firstorder_Maximum	glszm_LargeAreaEmphas is	gldm_LargeDependence HighGrayLevelEmphasis	firstorder_Variance
glrlm_LongRunHighGray LevelEmphasis	glcm_Autocorrelation	glszm_SmallAreaEmphasi s	firstorder_Median
glszm_LowGrayLevelZon eEmphasis	gldm_SmallDependenceH ighGrayLevelEmphasis	firstorder_Maximum	gldm_LowGrayLevelEmp hasis
ngtdm_Coarseness	firstorder_Maximum	glrlm_LongRunHighGray LevelEmphasis	firstorder_StandardDeviat ion
glszm_SmallAreaEmphasi s	glszm_LargeAreaLowGra yLevelEmphasis	gldm_LowGrayLevelEmp hasis	glrlm_LowGrayLevelRunE mphasis
gldm_LowGrayLevelEmp hasis	gldm_LargeDependence HighGrayLevelEmphasis	glcm_DifferenceEntropy	glcm_DifferenceAverage
glszm_SmallAreaHighGra yLevelEmphasis	ngtdm_Coarseness	glszm_SmallAreaHighGra yLevelEmphasis	glszm_HighGrayLevelZon eEmphasis
glcm_SumAverage	gldm_LowGrayLevelEmp hasis	glrlm_GrayLevelNonUnifo rarity	glcm_ClusterProminence
firstorder_RootMeanSqua red	glcm_DifferenceEntropy	ngtdm_Coarseness	ngtdm_Complexity
glcm_DifferenceEntropy	glszm_SmallAreaEmphasi s	glcm_DifferenceVariance	gldm_LargeDependence HighGrayLevelEmphasis
firstorder_Range	glcm_JointAverage	glrlm_HighGrayLevelRun Emphasis	ngtdm_Busyness
glrlm_GrayLevelNonUnifo rarity	glcm_DifferenceVariance	firstorder_Range	firstorder_90Percentile
firstorder_Mean	firstorder_Range	glcm_SumAverage	gldm_DependenceNonUn iformity
glrlm_HighGrayLevelRun Emphasis	glrlm_LongRunHighGray LevelEmphasis	firstorder_RootMeanSqua red	glrlm_LongRunHighGray LevelEmphasis
glcm_DifferenceVariance	ngtdm_Strength	firstorder_Mean	glrlm_RunLengthNonUnif ormityNormalized
glszm_LargeAreaLowGra yLevelEmphasis	glcm_SumAverage	gldm_SmallDependenceH ighGrayLevelEmphasis	glrlm_ShortRunHighGray LevelEmphasis
firstorder_10Percentile	firstorder_Mean	glszm_LargeAreaLowGra yLevelEmphasis	glszm_SmallAreaEmphasi s
glrlm_LowGrayLevelRunE mphasis	firstorder_RootMeanSqua red	glrlm_LowGrayLevelRunE mphasis	glcm_DifferenceVariance
gldm_SmallDependenceH ighGrayLevelEmphasis	glszm_ZoneVariance	firstorder_10Percentile	firstorder_10Percentile
gldm_LargeDependenceL owGrayLevelEmphasis	glrlm_GrayLevelNonUnifo rarity	glrlm_LongRunLowGrayL evelEmphasis	gldm_GrayLevelVariance
glrlm_LongRunLowGrayL evelEmphasis	glrlm_HighGrayLevelRun Emphasis	glszm_ZoneVariance	glcm_DifferenceEntropy
gldm_HighGrayLevelEmp hasis	glrlm_ShortRunHighGray LevelEmphasis	gldm_HighGrayLevelEmp hasis	glszm_ZoneVariance
glszm_ZoneVariance	glrlm_LowGrayLevelRunE mphasis	gldm_LargeDependenceL owGrayLevelEmphasis	glszm_LargeAreaLowGra yLevelEmphasis
glcm_JointAverage	firstorder_90Percentile	glcm_Autocorrelation	glcm_Autocorrelation
glrlm_ShortRunHighGray LevelEmphasis	ngtdm_Busyness	glcm_JointAverage	ngtdm_Strength

(continua)

Fisher Score	Trace Ratio	T Score	Relieff
ngtdm_Strength	gldm_LargeDependenceLowGrayLevelEmphasis	glrlm_ShortRunHighGrayLevelEmphasis	glcm_JointAverage
firstorder_Median	glrlm_LongRunLowGrayLevelEmphasis	firstorder_Median	ngtdm_Coarseness
glcm_Autocorrelation	gldm_HighGrayLevelEmphasis	ngtdm_Strength	glszm_SmallAreaHighGrayLevelEmphasis
firstorder_90Percentile	ngtdm_Complexity	ngtdm_Complexity	firstorder_Range
ngtdm_Busyness	firstorder_Median	firstorder_90Percentile	gldm_SmallDependenceHighGrayLevelEmphasis
ngtdm_Complexity	firstorder_10Percentile	ngtdm_Busyness	glrlm_ShortRunEmphasis

ANEXO A



MASSACHUSETTS
GENERAL HOSPITAL



HARVARD
MEDICAL SCHOOL

Department of Radiology
25 New Chardon Street, Suite 400C
Boston, MA 02114
Phone: 617-643-2326, Fax: 617-643-2743
Email: yoshida.hiro@mgh.harvard.edu

Hiroyuki Yoshida, Ph.D.
Director, 3D Imaging Research
Massachusetts General Hospital
Associate Professor of Radiology
Harvard Medical School

October 6, 2020

To: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)

Re: Processo nº 88887.371177/2019-00, EDITAL Nº 01/2019 - PrInt USP

Dear Sir/Madam,

The following is the feedback to your committee about the activities performed by Mrs. Natália Santana Chiari Correia while she attended the 3D Imaging Research in the Department of Radiology at the Massachusetts General Hospital (MGH) and Harvard Medical School (HMS) as a Ph.D. student during December 2019 and May 2020. She had a great number of accomplishments during such a short period.

- Sharing/Teaching of our personnel about techniques/software of image processing used by her group in Brazil, such as segmentation and feature extraction methods using 3D Slicer and the PyRadiomics package in MR exams.
- Development of an image database that consists of vertebral compression fractures MRI cases from her institution in Brazil to be used for experiments that were expected to be conducted according to the proposal.
- Pre-processing of the images using 3D Slicer to be used by the classification models.
- Development in python of six scripts for feature selection and 10 scripts for classification of the data.
- Beginning of writing a scientific journal paper to report the results obtained in the analyzes performed in the laboratory.

She demonstrated a high level of professionalism while working and studying at the MGH/HMS. Even in adverse situations such as those generated by a pandemic, she managed to get things worked out, showing her high level of commitment and motivation to research. She worked collaboratively with our group and acquired valuable knowledge that will certainly improve her research project.

Should you have any questions, please do not hesitate to contact me at yoshida.hiro@mgh.harvard.edu.

Sincerely,

Hiroyuki Yoshida, PhD
Director, 3D Imaging Research, MGH
Associate Professor of Radiology, HMS