



ISEL



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

INSTITUTO POLITÉCNICO DE LISBOA

Determinação das Características Radiómicas Informativas e o seu Impacto em Modelos de Apoio à Decisão Clínica aplicada à Ressonância Magnética da Próstata

ANDREIA PINTO COSTA

(Licenciada em Imagem Médica e Radioterapia)

Dissertação para obtenção do grau de Mestre em Engenharia Biomédica

Orientadores:

Doutor João António Alves Santinha

Doutora Lina da Conceição Capela de Oliveira Vieira

Júri:

Presidente: Doutor António Jorge Duarte de Castro Silvestre

Vogais:

Doutor Nuno Cruz Garcia

Doutor João António Alves Santinha

Setembro de 2025

Determinação das Características Radiómicas Informativas e o seu Impacto em Modelos de Apoio à Decisão Clínica aplicada à Ressonância Magnética da Próstata

ANDREIA PINTO COSTA
(Licenciada em Imagem Médica e Radioterapia)

Dissertação para obtenção do grau de Mestre em Engenharia Biomédica

Orientadores:

Doutor João António Alves Santinha (Champalimaud Clinical Centre, Champalimaud Foundation, Lisboa)

Doutora Lina da Conceição Capela de Oliveira Vieira (Escola Superior de Saúde de Lisboa/IPL e H&TRC, Lisboa)

Júri:

Presidente: Doutor António Jorge Duarte De Castro Silvestre (Instituto Superior de Engenharia de Lisboa)

Vogais:

Doutor Nuno Cruz Garcia (Faculdade de Ciências da Universidade de Lisboa)

Doutor João António Alves Santinha (Champalimaud Clinical Centre, Champalimaud Foundation, Lisboa)

Setembro de 2025

Agradecimentos

“Tenho a impressão de ter sido uma criança brincando à beira-mar, divertindo-me em descobrir uma pedrinha mais lisa ou uma concha mais bonita do que as outras, enquanto o imenso oceano de verdade continua misterioso diante dos meus olhos.” - Isaac Newton (Spence, 1820, p. 158).

De acordo com a sequência cronológica dos acontecimentos, agradeço à Coordenadora do Mestrado em Engenharia Biomédica, Professora Doutora Cecília Calado, que permitiu a transferência de mestrado.

À Professora Doutora Lina Vieira, pela confiança depositada, humanismo, carinho, sabedoria nas palavras e pela oportunidade que proporcionou de conhecer a notável investigação realizada no campo da radiômica pelo Doutor João Santinha.

Ao Doutor João Santinha, um agradecimento especial pela paciência e partilha de conhecimento, lapidou o meu caminho na radiômica e na programação. Aos dois orientadores, deixo a minha profunda admiração como investigadores, mentores e pela forma como estimularam as minhas capacidades, incutindo-me o desejo de explorar novos domínios, graças ao vosso constante incentivo. “Guardo comigo a frase que um dia me disseram: não precisas de encontrar a roda sozinha”. Um Bem-haja!

À minha mãe Elisabete Costa, pai Alcides Costa, irmã Rafaela Costa, cunhado Vilson Grilo e aos meus avós pela imensa paciência, amor e força. São a força motriz mais poderosa da minha vida, proporcionando tudo aquilo que tão orgulhosamente sou e conquistei. Às minhas fiéis companheiras Luna e Julinha pelo carinho e pelas esperas persistentes, companheiras silenciosas nas minhas jornadas.

Às minhas amigas, que sempre me apoiaram e me deram uma palavra de apoio em momentos de alegria e angústia Alexandra Santinho, Daniela Santos, Irina Neves, Rita Silva e Tamára Cocharra.

À dona Lurdes, que acreditou e cultivou a semente frágil e prematura como se dela fosse brotar a flor mais bonita da primavera. Muito obrigada!

Declaração de integridade

Declaro que esta(e) dissertação / trabalho de projeto / relatório de estágio é o resultado da minha investigação pessoal e independente. O seu conteúdo é original e todas as fontes listadas nas referências bibliográficas foram consultadas e estão devidamente mencionadas no texto. Mais declaro que todas as referências científicas e técnicas relevantes para o desenvolvimento do trabalho estão devidamente citadas e constam das referências bibliográficas.

O autor



Assinado por: Andreia Pinto
Costa
Identificação: B115364849
Data: 2025-09-19 às 15:53:37

Lisboa, 30 de Setembro de 2025

Determinação das Características Radiômicas Informativas e o seu Impacto em Modelos de Apoio à Decisão Clínica aplicada à Ressonância Magnética da Próstata

Resumo

Descritores de propriedades são parâmetros quantitativos extraídos de imagens médicas, os quais têm-se demonstrado promissores no desenvolvimento de modelos preditivos de *machine learning* (ML) aplicados ao apoio à decisão clínica. Contudo, a reprodutibilidade dos descritores, a elevada dimensionalidade dos dados, a seleção de recursos robustos e a generalização dos modelos constituem desafios metodológicos relevantes. Este estudo avaliou o impacto de diferentes metodologias de seleção de descritores radiômicos na construção de modelos de ML para a classificação de lesões prostáticas. Para tal, foi utilizado o *dataset PROSTATEx*, disponibilizado pela base de dados *The Cancer Imaging Archive*, composto por exames retrospectivos de *multiparametric magnetic resonance imaging* (mpMRI), em sequência T2, constituído por 204 sujeitos e 299 lesões segmentadas. Foram extraídos descritores de primeira-ordem, textura e forma, obtidos a partir da imagem original e de imagens derivadas de filtros. Consideraram-se dois conjuntos de dados: original e randomizado (obtido pela randomização da intensidade dos *píxeis*). As análises estatísticas incluíram: correlação *Spearman* e coeficiente de correlação intraclasse (ICC). Combinação dos resultados obtidos, através de operações de interseção e união, e análise univariada, para identificar descritores com maior significância com a variável a prever. Posteriormente, foram desenvolvidos quatro modelos de ML, cada um baseado numa das metodologias de seleção de descritores, comparados através de métricas de desempenho. Os resultados mostraram que a correlação *Spearman* identificou 1009 descritores não redundantes tendo sido eliminados 386, enquanto o ICC considerou 789 descritores informativos, com eliminação de 620. A união dos conjuntos resultou em 1154 descritores, enquanto a interseção obteve 650. Relativamente à análise univariada, nenhum descritor foi estatisticamente significativo. Este estudo demonstra que a aplicação de métodos de seleção de descritores informativos pode contribuir para melhorar o desempenho dos modelos de ML no contexto da classificação de lesões prostáticas.

Palavras-chave: Radiômica, Métodos Seleção de Variáveis, Modelos de *Machine Learning*

Identification of Informative Radiomic Features and their Impact on Clinical Decision Support Models Applied to Prostate Magnetic Resonance Imaging

Abstract

Feature descriptors are quantitative parameters extracted from medical images, which have shown promise in the development of machine learning (ML) predictive models to support clinical decision-making. However, the reproducibility of descriptors, the high dimensionality of the data, the selection of robust features, and the generalizability of models represent significant methodological challenges. This study evaluated the impact of different radiomic feature selection methodologies on the construction of ML models for the classification of prostate lesions. For this purpose, the PROSTATEx dataset, made available through The Cancer Imaging Archive, was used, consisting of retrospective *multiparametric magnetic resonance imaging* (mpMRI) examinations in T2-weighted sequences from 204 subjects and 299 segmented lesions. First-order, texture, and shape descriptors were extracted from both the original image and filter-derived images. Two datasets were considered: original and randomized (generated by pixel intensity randomization). Statistical analyses included Spearman correlation and the intraclass correlation coefficient (ICC). The results from these methods were combined through intersection and union operations, along with univariate analysis, to identify descriptors with the greatest significance in relation to the target variable. Subsequently, four ML models were developed, each based on one of the feature selection methodologies, and compared using performance metrics. The results showed that Spearman correlation identified 1009 non-redundant descriptors, with 386 eliminated, whereas ICC retained 789 informative descriptors, eliminating 620. The union of the sets resulted in 1154 descriptors, while the intersection yielded 650. With regard to univariate analysis, no descriptor reached statistical significance. This study demonstrates that the application of informative feature selection methods can contribute to improving the performance of ML models in the context of prostate lesion classification.

Keywords: Radiomics, Feature Selection Methods, Machine Learning Models

Publicações

Esta dissertação final de mestrado deu origem às seguintes publicações em conferências científicas, sob a forma de comunicação oral:

- Costa A, Santinha J, Vieira L Determinação das características radiómicas não-informativas e o seu impacto em modelos de apoio à decisão clínica Congresso Internacional de Radiologia de Coimbra (resumo e comunicação oral).
- Costa A, Santinha J, Vieira L Ressonância Magnética da próstata e radiómica: efeito da seleção de descritores no desempenho de modelos preditivos XXI Congresso Nacional ATARP (Associação Portuguesa dos Técnicos de Radiologia, Radioterapia e Medicina Nuclear) (resumo e comunicação oral).

Lista de Símbolos e de siglas

Alfabeto romano

S_{pixel} Intensidade do sinal em cada *píxel*

Alfabeto grego

μROI Média da intensidade da região de interesse

σROI Desvio-padrão correspondente

Siglas

AAPM *American Association of Physicists in Medicine*

AUC *Area Under the Curve*

CCC *Concordance Correlation Coefficient*

DCE *Dynamic Contrast-Enhanced*

EANM *European Association of Nuclear Medicine*

F1-Score *F1-Score*

GLCM *Gray-level co-occurrence Matrix*

GLDM *Gray-Level Dependence Matrix*

GLRLM *Gray-level run length Matrix*

GLSZM *Gray-level size zone Matrix*

IA *Artificial Intelligence*

IBSI *Image Biomarker Standardisation Initiative*

ICC *Intra-Class Correlation Coefficient*

LoG *Laplacian of Gaussian*

MCC *Matthews Correlation Coefficient*

ML *Machine Learning*

mpMRI *Multiparametric Magnetic Resonance Imaging*

MRI *Magnetic Resonance Imaging*

mRMR *Minimum Redundancy - Maximum Relevance*

NGTDM *Neighborhood Gray-Tone Difference Matrix*

NPV *Negative Predictive Value*

PCa *Prostate Cancer*

PCA *Principal Component Analysis*

PI-RADS *Prostate Imaging- Reporting and Data System*

PPV *Positive Predictive Value*

PSA	<i>Prostate Specific Antigen</i>
QIBA	<i>Quantitative Imaging Biomarker Alliance</i>
QIN	<i>Quantitative Imaging Network</i>
ROI	<i>Region of Interest</i>
RSNA	<i>Radiological Society of North America</i>
SL	<i>Supervised Learning</i>
SNR	<i>Signal-to-Noise Ratio</i>
T1	<i>T1-Weighted Imaging</i>
T2	<i>T2-Weighted imaging</i>
TCIA	<i>The Cancer Imaging Archive</i>
TE	<i>Echo Time</i>
TR	<i>Repetition Time</i>
TRUS	<i>Transrectal Ultrasound-Guided Biopsy</i>
US	<i>Ultrasonography</i>
WHO	<i>World Health Organization</i>

Índice

Capítulo I	1
1 Introdução	1
1.1 Objetivos	3
1.2 Contribuições.....	4
1.3 Estrutura da Dissertação	5
Capítulo II	7
2 Fundamentação Teórica	7
2.1 Anatomia da Próstata	7
2.2 Cancro da Próstata	9
2.3 Ressonância Magnética Multiparamétrica da Próstata.....	10
2.4 Fundamentos Radiômica	12
2.5 Extração de Descritores Radiômicos	13
2.5.1 <i>Descritores de Propriedade: primeira-ordem, forma e textura</i>	13
2.6 Modelos de <i>Machine Learning</i>	17
2.6.1 <i>Tipos de Aprendizagem dos Modelos</i>	20
2.6.1.1 <i>Classificação da Imagem por Machine Learning</i>	22
2.7 Fluxo de Trabalho da Radiômica	24
2.7.1 <i>Parâmetros de Aquisição de Imagem</i>	24
2.7.2 <i>Pós-Processamento de Imagem</i>	26
2.7.3 <i>Segmentação- Identificação de Regiões de Interesse</i>	28
2.7.4 <i>Extração de Descritores Radiômicos</i>	30
2.7.5 <i>Seleção de Descritores Radiômicos</i>	31
2.7.6 <i>Desenvolvimento do Modelo</i>	34
Capítulo III	39
3 Materiais e Métodos	39
3.1 Amostra	39
3.1.1 <i>Base de Dados The Cancer Imaging Archive (TCIA)</i>	39
3.1.1.1 <i>Protocolo de Aquisição de Imagem</i>	40

3.1.1.2	<i>Segmentação do Dataset PROSTATEx</i>	41
3.1.1.3	<i>Critérios de Seleção da Amostra em Estudo</i>	41
3.1.1.4	<i>Dados da Biópsia Histopatológica</i>	42
3.1.1.5	<i>Critérios de Seleção da Subamostra em Estudo</i>	42
3.1.1.6	<i>Caracterização da Subamostra em Estudo</i>	43
3.1.1.7	<i>Conjunto de Dados Definidos para Estudo</i>	44
3.2	Definição do Pós-Processamento de Imagem	45
3.3	Metodologias de Cálculo dos Descritores Radiômicos	46
3.3.1	<i>Extração de Descritores Radiômicos</i>	47
3.4	Data	51
3.5	Seleção de Descritores Radiômicos	52
3.5.1	<i>Análise Estatística Multivariada</i>	52
3.5.1.1	<i>Critério de Inclusão de Descritores Informativos</i>	53
3.5.1.2	<i>Operação entre Conjuntos de Dados e Determinação de Inputs</i>	53
3.5.2	<i>Análise Estatística Univariada</i>	54
3.5.2.1	<i>Critérios de Inclusão para o Input Descritores Informativos e Preditivos</i>	54
3.6	Construção dos Modelos de <i>Machine Learning</i>	55
3.6.1	<i>Conjunto de Modelos</i>	55
3.6.2	<i>Parâmetros de Construção dos Modelos</i>	55
3.6.3	<i>Avaliação do desempenho dos Modelos</i>	57
Capítulo IV	59
4	Resultados	59
4.1	Análise Estatística	59
4.1.1	<i>Correlação de Spearman entre Descritores de Primeira-ordem e Textura vs. Forma</i>	59
4.1.2	<i>Coefficiente de Correlação Intraclasse entre Descritores Radiômicos do Conjunto de Dados Original e Randomizado</i>	66
4.1.3	<i>Combinação dos Conjuntos: Operação de Interseção e União</i>	73
4.1.4	<i>Análise Univariada Radiômica</i>	73
4.2	Modelos de <i>Machine Learning</i>	75
4.2.1	<i>Modelo I: Conjunto de dados Originais</i>	75

4.2.2	<i>Modelo II: Input_Interseção</i>	76
4.2.3	<i>Modelo III: Input_União</i>	77
4.2.4	<i>Modelo IV: Input Descritores Informativos e Preditores</i>	78
Capítulo V	81
5	Discussão	81
Capítulo VI	95
6	Conclusões e Perspetivas Futuras	95
6.1	Limitações do presente trabalho	96
	Referências Bibliográficas	99
	Apêndices	104

Índice de figuras

Figura 2.1-Imagem de MRI da próstata em corte coronal. Ilustração das diferentes zonas anatómicas: zona central (em laranja), zona transicional (em azul) e zona periférica (em roxo).	8
Figura 2.2-Ilustração de MRI pélvica em corte axial, anatomia da glândula da próstata e estruturas adjacentes.	8
Figura 2.3- Exemplo ilustrativo do cálculo de descritores de textura a partir de matrizes GLCM (horizontal). ...	14
Figura 2.4- Exemplo ilustrativo do cálculo de descritores de textura a partir de matrizes GLRLM (horizontal)...	15
Figura 2.5- Exemplo ilustrativo do cálculo de descritores de textura a partir de matrizes GLSZM.	15
Figura 2.6- Exemplo ilustrativo do cálculo de descritores de textura a partir de matrizes NGTDM.	16
Figura 2.7-Representação visual dos descritores de propriedades de forma, primeira-ordem e textura	16
Figura 2.8- Etapas do desenvolvimento de uma modelo de <i>machine learning</i> . O desenvolvimento do modelo de <i>machine learning</i> tem como objetivo a resolução de diferentes tipos de problemas, sendo estes definidos por três elementos: (i) tarefa a ser aprendida, (ii) a métrica de desempenho a ser otimizada e (iii) o modo como o modelo adquire dados e interações para aprimorar aprendizagem.	17
Figura 2.9- Principais algoritmos de <i>machine learning</i> diferenciados entre estratégias de aprendizagem supervisionada e não supervisionada.....	21
Figura 2.10- Fluxo geral da radiômica: aquisição de imagem, pós-processamento de imagem, extração e seleção de descritores de propriedades e desenvolvimento do modelo.	24
Figura 3.1-Frequência absoluta dos grupos <i>Gleason</i> das lesões prostáticas.	42
Figura 3.2-Caracterização da população em estudo.	43
Figura 3.3-Classificação da biópsia histopatológica da amostra: 76 lesões classificadas em clinicamente significativa e 223 em não clinicamente significativa.	43
Figura 3.4- Imagem original e randomizada <i>PROSTATEx_0203 t2</i> após aplicação da operação de randomização das intensidades de pixels através das bibliotecas: <i>pandas, numpy, SimpleITK</i>	44
Figura 3.5- Aplicação dos filtros LoG (sigma 0,6 mm, 2mm, 3mm), exponencial, logaritmo, <i>Square, Square-Root, Wavelet</i> de decomposição em 2 níveis.....	48
Figura 4.1 -Frequência absoluta de descritores de propriedade de primeira-ordem e de textura, classificados como redundantes e não redundantes baseados na correlação de <i>Spearman</i>	60
Figura 4.2 -Frequência absoluta de descritores informativos e não-informativos resultante da análise do coeficiente de correlação intraclasse entre conjunto de dados original vs. randomizados.	66
Figura 4.3- Diagrama de <i>Venn</i> com ilustração dos conjuntos operados sobre união e interseção.	73
Figura 4.4- Gráfico da dispersão dos valores-p e a AUC no conjunto de dados. originais.....	74
Figura 4.5- O gráfico <i>Violin Plot</i> e área sob a curva ROC ilustrativo do descritor, <i>wavelet2-LL_firstorder_Minimum</i> demonstra uma distribuição semelhante entre as classes <i>TRUE\FALSE</i> , o que indica poder preditivo baixo para a classificação isoladamente. Adicionalmente, a área sob a curva ROC	

sugere um poder preditivo fraco ilustrado pelo traçado da curva sugestivo de uma classificação quase aleatória.....	74
Figura 4.6-Representação dos resultados do conjunto de teste entre os Modelos I, II, III, IV.	79
Figura 4.7-Gráfico Modelo II SHAP para o desfecho clínico significativo. Cada ponto representa uma amostra, em que a posição no eixo x indica o contributo da variável radiômica na probabilidade do desfecho clínico. A cor dos pontos reflete a intensidade da variável (azul: baixo) e (vermelho: alto). As variáveis estão ordenadas por importância média absoluta.....	80
Figura B.1-Frequência absoluta da idade da população em estudo (59).....	105
Figura D.1- <i>Correlation Plot</i> entre descritores de propriedades de primeira-ordem vs. forma $\rho > 0,8$	106
Figura D.2- <i>Correlation Plot</i> entre descritores de propriedades de GLDM vs. forma $\rho > 0,8$	107
Figura D.3- <i>Correlation Plot</i> entre descritores de propriedades de textura GLCM vs. forma $\rho > 0,8$	107
Figura D.4- <i>Correlation Plot</i> entre descritores de propriedades de textura GLSZM vs. forma $\rho > 0,8$	108
Figura D.5- <i>Correlation Plot</i> entre descritores de propriedades de textura GLRLM vs. forma $\rho > 0,8$	108
Figura D.6- <i>Correlation Plot</i> entre descritores de propriedades de textura NGTDM vs. forma $\rho > 0,8$	109
Figura E.1- <i>Correlation Plot</i> entre descritores de propriedades radiômicas ICC $\rho > 0,7$	110

Índice de Quadros

Quadro 2.1- Sequências de <i>mpMRI</i> abordadas quanto à sequência, princípio físico e principais contributos imagiológicos apresentados para cada fase de estudo <i>mpMRI</i> da próstata.	10
Quadro 2.2- Categorias de avaliação <i>PI-RADS</i> ™ versão 2 descrevem a probabilidade crescente de malignidade da próstata clinicamente significativo.	11
Quadro 2.3- Tipologias dos algoritmos supervisionados de <i>machine learning</i> utilizados neste estudo/dissertação: <i>Random Forest</i> , Regressão Logística e <i>Support Vector Machines</i>	22
Quadro 2. 4- Descritores radiômicos baseados na documentação <i>PyRadiomics</i>	30
Quadro 2.5- Matriz de confusão resume o desempenho do modelo, através dos seguintes elementos: real positivo e negativo vs. previsto positivo e negativo.	34
Quadro 3.1-Parâmetros de aquisição das imagens do estudo <i>PROSTATEx</i>	40
Quadro 3.2-Parâmetros de pós-processamento de extração estabelecidos no <i>PyRadiomics</i> versão 3.0.1.	46
Quadro 3.3- Parâmetros de extração definidos no <i>PyRadiomics</i> versão 3.0.1 dos descritores de propriedades primeira-ordem, forma e textura em tipos de imagens distintos.	47
Quadro 3.4- <i>FeaturesClass</i> de primeira-ordem definida por 18 descritores através do <i>software PyRadiomics</i> ..	49
Quadro 3.5- <i>FeaturesClass</i> de forma definidos por 14 descritores através do <i>software PyRadiomics</i>	50
Quadro 3.6- <i>FeaturesClass</i> Textura definidos pelas matrizes GLCM, GLRLM, GLSZM, NGTDM e GLDM através do <i>software PyRadiomics</i>	50
Quadro 4.1- Descritores de propriedade primeira-ordem redundantes determinados pela correlação de <i>Spearman</i>	60
Quadro 4.2- Descritores de propriedade de textura classe GLCM redundantes, determinados pela correlação de <i>Spearman</i>	61
Quadro 4.3-Descritores de propriedade de textura classe GLDM redundantes, determinados pela correlação de <i>Spearman</i>	62
Quadro 4.4- Descritores de propriedade de textura classe GLRLM redundantes, determinados pela correlação de <i>Spearman</i>	62
Quadro 4.5-Descritores de propriedades de textura classe GLSZM redundantes, determinados pela correlação de <i>Spearman</i>	63
Quadro 4.6-Descritores de propriedades de textura classe NGTDM redundante, determinados pela correlação de <i>Spearman</i>	64
Quadro 4.7- Lista de descritores radiômicos de primeira-ordem não redundantes resultantes da análise estatística correlação de <i>Spearman</i> para as classes descritoras tanto na imagem original como na imagem derivada de filtros.	64
Quadro 4.8-Lista de descritores radiômicos de textura não redundantes resultantes da análise estatística correlação de <i>Spearman</i> para as classes descritoras tanto na imagem original como na imagem derivada de filtros.	65

Quadro 4.9-Descritores de propriedade primeira-ordem não-informativos, determinados pelo ICC.	67
Quadro 4.10-Descritores de propriedade textura classe GLCM não-informativos, determinados pelo ICC.	68
Quadro 4.11-Descritores de propriedade textura classe GLDM não-informativos, determinados pelo ICC.	69
Quadro 4.12-Descritores de propriedade textura classe GLRLM não-informativos, determinados pelo ICC.	70
Quadro 4.13-Descritores de propriedade textura classe GLSZM não-informativos, determinados pelo ICC.	71
Quadro 4.14-Descritores de propriedade textura classe NGTDM não-informativos, determinados pelo ICC.	72
Quadro 4.15- Lista de descritores de propriedades informativos resultantes da análise estatística ICC para as várias classes descritoras tanto para a imagem original como para a imagem derivada de filtros.	72
Quadro 4.16-Descritores radiômicos selecionados através do valor superior de área sob a curva <i>ROC</i>	75
Quadro 4.17- Métricas de desempenho de treino e teste através do modelo de <i>machine learning random forest</i> em 5 <i>folds</i> desenvolvido através do conjunto de dados originais.	76
Quadro 4.18-Métricas de desempenho de treino e teste através do modelo de <i>machine learning random forest</i> em 5 <i>folds</i> desenvolvido através do conjunto de dados intersetados.	77
Quadro 4.19-Métricas de desempenho de treino e teste através do modelo de <i>machine learning</i> regressão logística em 5 <i>folds</i> desenvolvido através do conjunto de união.	78
Quadro 4.20 -Métricas de desempenho de treino e teste através do modelo de <i>machine learning Support Vector Machine</i> em 5 <i>folds</i> desenvolvido através do conjunto de dados originais.	79
Quadro A.1-Classificação de Grupos Gleason, com a respetiva descrição histopatológica da próstata associadas a cada categoria.	104
Quadro C.1-Imagens derivadas de filtros, descrição das transformações realizadas sobre as intensidades dos píxeis.	105

1 Introdução

O cancro da próstata (*PCa do inglês Prostate Cancer*) é considerado a segunda neoplasia maligna mais frequente em homens a nível mundial e a quinta causa de morte, apresentando tendências crescentes de incidência nas próximas décadas. O diagnóstico precoce constitui a base de um tratamento eficaz, estando associado a uma taxa de sobrevida estimada em 10 anos, próxima de 99% (1–3).

Do ponto de vista clínico, a PCa caracteriza-se por um espectro heterogéneo de comportamento clínico e morfológico, que pode variar desde neoplasias de baixo grau e evolução lenta, sem progressão para doença clinicamente significativa, até formas altamente agressivas, com elevada propensão para disseminação metastática. Esta heterogeneidade representa um desafio substancial para a prática clínica, particularmente no que concerne ao diagnóstico preciso e atribuição de uma estratégia terapêutica mais adequada (4).

O diagnóstico de PCa é determinado por um conjunto de exames clínicos, laboratoriais e imagiológicos. No entanto, as metodologias convencionais, como a dosagem do antigénio específico da próstata (*PSA do inglês Prostate Specific Antigen*), o toque retal, as biópsias aleatórias, apresentam limitações importantes em termos de sensibilidade e especificidade, conduzindo por vezes a resultados imprecisos de sobrediagnóstico ou sobretratamento. Neste contexto, as técnicas de imagem médica têm vindo a assumir um papel cada vez mais relevante como ferramentas de apoio ao diagnóstico (2,3,5,6).

Entre as diferentes modalidades de imagem que contribuem para o diagnóstico de PCa, a ressonância magnética multiparamétrica (*mpMRI do inglês Multiparametric Magnetic Resonance Imaging*), destaca-se como a técnica de referência, combinando informações detalhada para melhorar a deteção e caracterização das lesões prostáticas. A mpMRI tem demonstrado impacto significativo na redução de procedimentos invasivos e na previsão do

comportamento biológico da doença, consolidando-se como uma ferramenta robusta para a prática clínica contemporânea (2,5,7).

Paralelamente, os avanços na área da imagem médica, resultantes da evolução dos dispositivos, da melhoria da qualidade de imagem, do aumento da capacidade computacional, do desenvolvimento de métodos de análise de imagem e da expansão de ferramentas baseadas em inteligência artificial permitiram uma mudança de paradigma na análise imagiológica. O que outrora se baseava numa interpretação qualitativa, fundamentada em descritores semânticas, transformou-se numa ciência quantitativa, suportada na extração sistemática de descritores quantitativos a partir de imagens médicas (8–10). Este desenvolvimento fundamenta o surgimento da radiômica como área em expansão verificado pelo crescimento exponencial, com mais de 1500 publicações em 2020 incluindo o termo radiômica (11).

A radiômica é definida como o processo de conversão de imagens médicas digitais em dados quantitativos de elevada dimensionalidade, através da extração de vários conjuntos de descritores radiômicos (12–14). Com a premissa de que as imagens contêm informação subjacente que reflete a fisiopatologia e heterogeneidade tumoral, a qual pode ser relevada por meio de análises quantitativas avançadas. O fluxo de trabalho da radiômica compreende, de forma sistemática, várias etapas: aquisição de imagens, segmentação das regiões de interesse, extração de descritores radiômicos, seleção de descritores relevantes e desenvolvimento de modelos preditivos, frequentemente suportados por algoritmos de *machine learning* (10,15–17).

Entre essas etapas, a seleção de descritores assume um papel crítico, constituindo-se como foco do presente trabalho. A elevada dimensionalidade dos dados radiômicos, as redundâncias entre descritores e a ausência de métodos transparentes podem comprometer a robustez e a generalização dos modelos, conduzindo a problemas de *overfitting*, perda de interpretabilidade e resultados enviesados (12,18).

Adicionalmente, é importante destacar que os descritores de propriedades podem incorporar dependências subjacentes, o que pode prejudicar a precisão preditiva e introduzir instabilidades, uma vez que modelos de *machine learning* podem integrar múltiplos substitutos do mesmo descritor. Além disso, diferentes classes de descritores recorrem frequentemente a métricas semelhantes, o que tende a originar dependências adicionais entre classes distintas. Este conjunto de desafios reforça a importância do desenvolvimento e da avaliação de metodologias rigorosas de seleção de variáveis no âmbito da radiômica (18).

Apesar dos avanços alcançados, a aplicação clínica da radiômica enfrenta limitações substanciais. A padronização insuficiente dos protocolos de aquisição de imagem, a variabilidade interinstitucional e a reprodutibilidade limitada de determinados descritores dificultam a translação dos modelos para a prática clínica. Adicionalmente, a ausência de consenso em torno das metodologias de seleção de descritores constitui ainda uma barreira adicional (18).

1.1 Objetivos

Esta dissertação tem como **objetivo geral**:

Avaliar o impacto de diferentes métodos de seleção de descritores radiômicos informativos baseados na correlação de *Spearman* e no coeficiente de correlação intraclasse (ICC) por meio da construção de modelos de *machine learning* aplicados a um *dataset* de imagens de lesões na próstata em mpMRI.

Objetivos Específicos:

- I. Determinar descritores de propriedades de primeira-ordem e textura com baixa a moderada correlação com descritores de forma através da correlação de *Spearman*.
- II. Determinar descritores de propriedades com baixa a moderada correlação entre descritores de propriedades através do coeficiente de correlação intraclasse entre o conjunto original e randomizado.
- III. Avaliar diferentes formas de combinar os conjuntos de propriedades obtidos pelos métodos descritos em I e II e analisar a melhor metodologia a adotar através de modelos de *machine learning*.
- IV. Determinar descritores de propriedades significativamente associados com a variável a prever através da análise univariada e a sua utilização em modelos de *machine learning*.

1.2 Contribuições

Com este estudo, pretende-se contribuir para o aumento do conhecimento relativamente ao impacto das metodologias de seleção de descritores radiómicos. Para tal, desenvolveu-se quatro modelos através do conjunto de dados originais e após aplicação de metodologias de seleção de descritores provenientes de imagens de mpMRI de lesões de próstata disponíveis na base de dados pública *The Cancer Imaging Archive* (TCIA).

Apesar deste estudo se basear numa amostra limitada ($n = 204$ sujeitos), os resultados obtidos, alinhados com a literatura existente, (3,11,15,19–24) contribuem para a investigação de metodologias baseadas em boas práticas no desenvolvimento de modelos preditivos com alta reprodutibilidade, confiabilidade, precisão e capacidade de generalização. Esses fatores são essenciais para o avanço do campo da radiómica, especialmente na probabilidade de uma tradução clínica bem-sucedida.

1.3 Estrutura da Dissertação

A dissertação, encontra-se organizada em cinco capítulos:

O **Capítulo I** enquadra o problema e enuncia os objetivos que nortearam o processo de investigação e apresenta a estrutura da dissertação.

No **Capítulo II** é realizada uma revisão teórica dos aspetos que constituem a base teórica desta dissertação nomeadamente:

- Breve contexto sobre fundamentos teóricos referentes à anatomia da próstata, cancro da próstata, mpMRI, radiómica, extração de descritores radiómicos e modelos de *machine learning*.
- Principal fluxo de trabalho da radiómica: parâmetros de aquisição de imagem, pós-processamento de imagem, segmentação, extração de descritores radiómicos, seleção de descritores radiómicos e desenvolvimento do modelo.

No **Capítulo III** é descrito os materiais e métodos utilizados para a realização do presente estudo nomeadamente: amostra, base de dados, protocolos de aquisição de imagem, segmentação, critérios e caracterização da amostra e subamostra, pós-processamento de imagem, metodologias de cálculos, dados, análise estatística e construção dos modelos.

No **Capítulo IV** são apresentados os resultados obtidos individualmente para os diferentes métodos de seleção de descritores, seguidos da utilização destes como entradas para o desenvolvimento de modelos de *machine learning*. Estes modelos foram então treinados e validados com partição de dados selecionada, sendo posteriormente avaliados os impactos das metodologias adotadas nas métricas de desempenho.

No **Capítulo V** segue-se a discussão dos resultados obtidos comparativamente com a literatura científica disponível até ao momento. Em forma de conclusão é mencionado os pontos centrais do estudo.

No **Capítulo VI** Por fim, os objetivos gerais e específicos são respondidos e as limitações inerentes à realização do estudo são apresentadas, são sugeridas algumas perspetivas futuras relacionadas com o tema da dissertação, com base nas oportunidades científicas identificadas ao longo do estudo.

2 Fundamentação Teórica

2.1 Anatomia da Próstata

A próstata é uma glândula exócrina fibromuscular densa inserida no sistema gênito-urinário masculino localizada na zona pélvica, mais propriamente anterior ao reto e inferior à bexiga. Com formato cônico composta por base, contínua ao colo da bexiga e ápice, anexo à fáscia superior do diafragma urogenital. Apresenta-se circundada à uretra proximal na pelve menor e apresenta uma cápsula fibrosa que a envolve, juntamente com nervos, o plexo vascular e uma camada visceral de fáscia pélvica (25).

Uma vez que a composição histológica varia nas diferentes porções da glândula, a próstata classifica-se em três zonas como ilustrado pela Figura 2.1: a zona central, localizada na base da glândula e envolvendo os ductos ejaculatórios, a zona periférica, que constitui cerca de 70% do volume prostático, envolve parcialmente a zona central e circunda a parte distal da uretra prostática, e a zona de transição, uma pequena região glandular que envolve a porção da uretra compreendida entre a bexiga e o verumontanum (25).

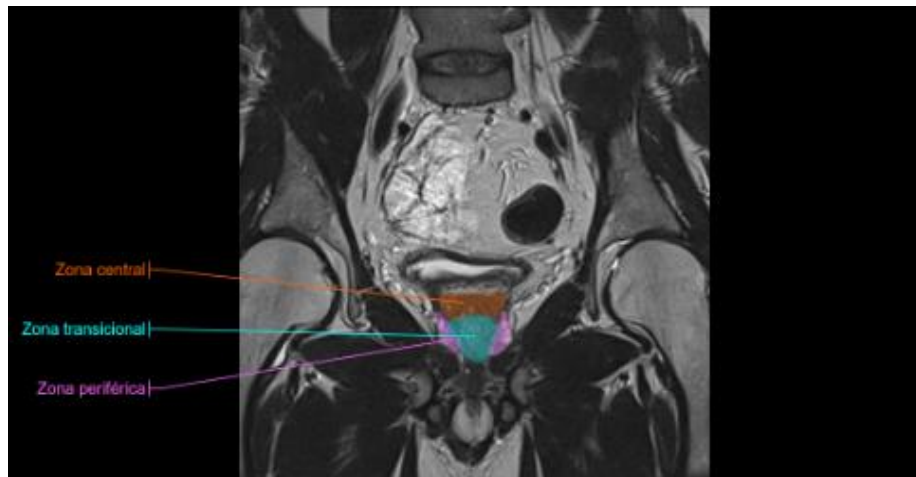


Figura 2.1-Imagem de MRI da próstata em corte coronal. Ilustração das diferentes zonas anatômicas: zona central (em laranja), zona transicional (em azul) e zona periférica (em roxo) (Micheau Antoine & Hoa Denis, 2025).

Na prática clínica, a nível anatómico, a próstata divide-se em dois lobos laterais (direito e esquerdo) e um lobo mediano (25). A sua principal função é a secreção e o armazenamento de fluido seminal, constituinte do esperma (25).

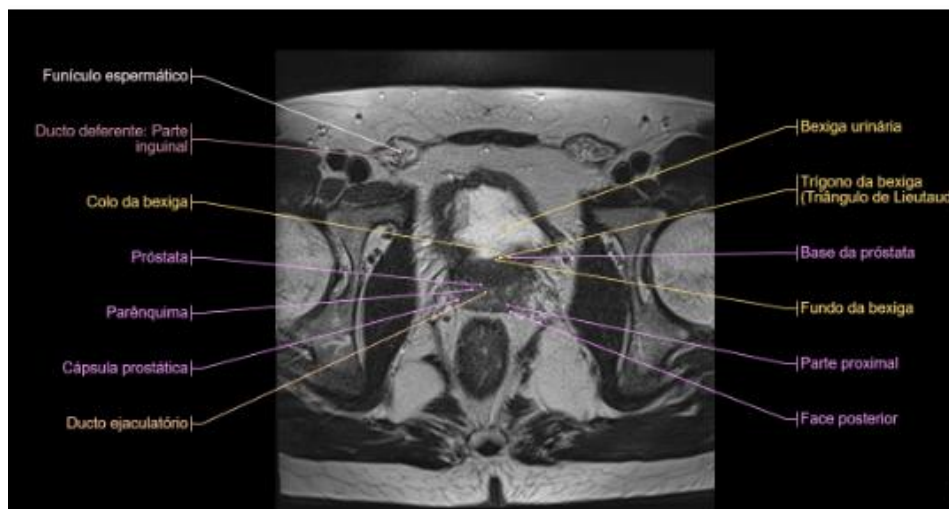


Figura 2.2-Ilustração de MRI pélvica em corte axial, anatomia da glândula da próstata e estruturas adjacentes (Micheau Antoine & Hoa Denis, 2025).

2.2 Cancro da Próstata

O PCa é considerado a segunda neoplasia maligna mais comum em homens, com uma taxa global em constante aumento, sendo a sua maior prevalência a partir dos 55 anos. A nível mundial, segundo dados referentes a 2020 estima-se que sejam diagnosticados com cancro da próstata 1,4 milhões de homens anualmente e a taxa de mortalidade estima-se em cerca de 375 304, sendo a sua maior prevalência em países desenvolvidos e raça negra (1,2).

Atualmente esta patologia é a terceira causa de morte no sexo masculino, com previsão de aumento de casos até 2040, a nível da Europa e Portugal, segundo a Organização Mundial de Saúde (*WHO do inglês World Health Organization*) (26).

O diagnóstico patológico baseia-se na combinação de diferentes exames e procedimentos, tais como: o toque retal, exames sanguíneos análise da *PSA*, a biópsia guiada por ultrassonografia transretal (*TRUS do inglês transrectal ultrasound-guided biopsy*) e técnicas de imagens como a mpMRI (2,5,6).

Segundo dados provenientes de autópsias e estudos epidemiológicos, estima-se que aproximadamente um em cada três homens com idades superiores a 50 anos apresente alterações histológicas compatíveis com cancro da próstata (4). Além disso, aproximadamente 94% dos doentes classificados de baixo risco, acabam por ser submetidos a terapêuticas radicais, envolvendo toda a glândula prostática. Abordagem terapêutica, possivelmente desproporcional ao perfil de risco, devido sobretudo às limitações atuais no diagnóstico, que dificulta não apenas a deteção precoce da doença, mas também a estratificação precisa do risco e a adequada avaliação prognóstica. Essa dificuldade é acentuada pela heterogeneidade clínica e biológica associada às lesões prostáticas (4).

As técnicas de imagem médica assumem papel central, uma vez que possibilitam a caracterização macroscópica de processos anatómicos e fisiopatológicos de forma não invasiva, acessível e amplamente disponível na prática clínica na atualidade (27–29). Cada modalidade de imagem resulta da aplicação de diferentes princípios físicos distintos, o que confere sensibilidade e especificidades diferenciadas para a deteção e caracterização de condições específicas (2,30).

Entre as modalidades atualmente disponíveis para o estudo do cancro da próstata, destaca-se a mpMRI, permitindo reduzir cerca de 25% a necessidade de biópsias adicionais (2).

2.3 Ressonância Magnética Multiparamétrica da Próstata

A mpMRI consolidou-se como uma das principais modalidades de imagem de ressonância magnética (MRI *do inglês Magnetic Resonance Imaging*) para a detecção primária ou exclusão do PCa clinicamente significativo. Resultados promissores têm sido demonstrados no diagnóstico, localização de lesões, estratificação do risco e estadiamento (2,7,31,32).

A elevada resolução espacial e relação sinal-ruído, associado à administração de contraste paramagnético, conferem à mpMRI uma caracterização tecidual detalhada (33). Contribuindo de forma, distinta das técnicas de biópsias, permitindo a integração de informação anatômica, funcional com sensibilidade e especificidade elevada (7,31).

A aplicação de diferentes gradientes magnéticos e pulsos de radiofrequência, permite a obtenção de diferentes imagens de MRI, através da parametrização da aquisição. Cada sequência fornece informações distintas, mas complementares ao diagnóstico médico. O estudo mpMRI da próstata consiste, na combinação de várias sequências, incluindo imagens: ponderadas em T1 (*do inglês T1-Weighted Imaging*), T2 (*do inglês T2-Weighted imaging*), DWI (*do inglês Diffusion-Weighted Imaging*), DCE (*do inglês Dynamic Contrast-Enhanced*) e em contextos específicos, imagens espectroscópicas (7,31).

Quadro 2.1- Sequências de mpMRI abordadas quanto à sequência, princípio físico e principais contributos imagiológicos apresentados para cada fase de estudo mpMRI da próstata. Adaptada das seguintes fontes bibliográficas: Demirel & Davis, (2018) e Ghai & Haider, (2015).

Sequências de mpMRI		
Sequências	Princípio Físico	Avaliação Imagiológica
T1	Tempos de relaxamento dos prótons baseados em tempo de eco e tempo de repetição curtos, que determinam a taxa de retorno dos prótons excitados ao equilíbrio.	Identificar hemorragia pós-biópsia na próstata e nas vesículas seminais, detecção de disseminação linfonodal e metástases quando associado à administração de contraste.
T2	Tempos de relaxamento dos prótons baseados em tempo de eco e tempo de repetição longos, que determinam a taxa de retorno dos prótons excitados ao equilíbrio ou perda de fase entre si.	Determina baixa intensidade de sinal no tecido tumoral. Alta resolução espacial na definição da cápsula prostática, permitindo avaliar a extensão extraprostática e o estadiamento. Insuficiência na avaliação da zona de transição e central.

DWI	Propriedade de difusão dos prótons dos átomos de água, que reflete o movimento aleatório das moléculas de água nos tecidos.	Estratificação de risco, permite diferenciar de outras patologias distintas, tais como: anomalias benignas, prostatite, fibrose, tecido cicatricial e hemorragia pós-biópsia. Avaliação de tumores na zona de transição.
DCE	Administração intravenosa de um bólus de contraste endovenoso, avaliação da acumulação.	Avaliação da angiogénese tumoral através de contraste à base de gadolínio, baseado no realce precoce de alta amplitude e vascularização antes e durante a injeção.
Espectroscopia	Absorção de energia na gama das radiofrequências por núcleos de uma molécula, submetidos a um forte campo magnético.	Avaliar o risco de malignidade de uma região de interesse através da relação colina, creatina e citrato como biomarcador metabólico.

T1: T1-weighted imaging, T2: T2-weighted imaging, DWI: Dynamic Contrast-Enhanced, DCE: Diffusion-Weighted Imaging

Com o desenvolvimento dos estudos em mpMRI, sistemas de interpretação estruturados foram implementados para melhorar a padronização, terminologia na avaliação das imagens e variabilidade entre observadores (4,6). O sistema PI-RADS (*do inglês Prostate Imaging Reporting and Data System*), representa a principal ferramenta nesse âmbito, ao estabelecer critérios objetivos de diagnóstico e atribuição de categorias compreendidas entre 1 a 5, em que as categorias crescentes correspondem a uma probabilidade progressiva das lesões serem clinicamente significativas, como é demonstrado pelo Quadro 2.2 (7,31,34).

Quadro 2.2- Categorias de avaliação PI-RADS TM versão 2 descrevem a probabilidade crescente de malignidade da próstata clinicamente significativo. Adaptada da seguinte fonte bibliográfica Demirel & Davis, (2018).

Categorias de avaliação PI-RADS TM

PI-RADS 1	Muito baixo, altamente improvável que haja lesões clinicamente significativo
PI-RADS 2	Baixo, improvável que haja lesões clinicamente significativas
PI-RADS 3	Intermédio, a presença de lesões clinicamente significativas é ambígua
PI-RADS 4	Alto, provável que haja lesões clinicamente significativas
PI-RADS 5	Muito alto, altamente provável que haja lesões clinicamente significativas

PI-RADS: Prostate Imaging Reporting and Data System

Apesar dos avanços alcançados, a avaliação da próstata continua a ser um processo controverso, subjetivo e inconsistente, o que evidencia a necessidade de métodos complementares para aumentar a reprodutibilidade e a precisão diagnóstica. Neste contexto, as imagens de mpMRI da próstata assumem-se como uma excelente fonte de dados, na medida em que integram informação anatômica e funcional. Tornando-se um recurso de grande relevância para a área da análise de dados como a radiômica, cuja finalidade é a extração de biomarcadores quantitativos e o desenvolvimento de novos conhecimentos. Tal abordagem resulta da confluência entre duas áreas fundamentais: a ciência da computação e estatística, proporcionando uma base sólida para apoiar processos (6,7,11,30,31).

2.4 Fundamentos Radiômica

O termo radiômica provém da junção da palavra “radio” com o sufixo “ômica”, em que o primeiro termo se refere à área da radiologia mais propriamente direcionado à ciência da aquisição de imagens através do uso de diferentes formas de energia para produção de imagens médicas. Já o sufixo “ômica” é usado como termo da área da genômica atribuído aos estudos e mapeamento do genoma humano (33).

A radiômica é uma abordagem quantitativa aplicada à imagem médica, cujo objetivo é a extração computacional de descritores de propriedades por meio de cálculo matemático avançado (10,15–17). O processo de análise de dados, por meio da radiômica, inicia-se com a extração de descritores radiômicos, seleção de um conjunto de variáveis robustas com relação a dados biológicos e desenvolvimento de modelos de *machine learning* (17,28,30)

Adicionalmente, o uso de técnicas de inteligência artificial (*IA do inglês artificial intelligence*) proporcionam os meios para detetar, correlacionar e analisar padrões em imagens médicas e associá-las a descritores moleculares e/ou genéticos, proporcionando o desenvolvimento de sistemas confiáveis de suporte à decisão clínica (9,16,28,33,35).

Através dos descritores radiômicos é possível obter informações relevantes sobre fenótipos de tecidos, microambientes, entre outras, com valor distinto e complementar em comparação com outros dados “ômicos” ou até testes clínicos (8,9).

A radiômica destaca-se por ser um campo de pesquisa promissor, visto que, oferece um número ilimitado de biomarcadores de imagem que permitem auxiliar na detecção de

patologias, avaliar o prognóstico das lesões, prever a resposta à terapêutica e monitorizar a progressão da doença (16,36).

Estudos em radiômica aplicados a imagens de mpMRI têm demonstrado a capacidade de distinguir entre lesões benignas e malignas (2,30), prever a agressividade tumoral (2,30), melhorar o desempenho do radiologista na avaliação do risco e aumentar a acurácia das pontuações PI-RADS (30).

2.5 Extração de Descritores Radiômicos

2.5.1 Descritores de Propriedade: primeira-ordem, forma e textura

Descritores de propriedades de imagens, também designados por recursos, analisam tanto tecidos patológicos como saudáveis a partir de imagens médicas de alta resolução espacial, quantificando parâmetros de: primeira-ordem, textura e forma como ilustrado pela Figura 2.7 (16,33). Embora a capacidade visual identifique alguma destas características para classificar imagens, é conhecido que o olho humano é limitado a alterações subtis que estão abaixo do limiar de deteção perceptível, tornando-as indetetáveis e consequentemente, limitando o potencial das imagens médicas (12). Um descritor de propriedade consiste num vetor, cujos elementos radiômicos caracterizam uma região de interesse de uma imagem, determinando diferentes propriedades, as quais serão analisadas em seguida (10,14,37,38).

Os descritores de primeira-ordem mais comumente designados como descritores estatísticos de histograma de nível de cinzas baseiam-se na análise de *píxeis* únicos e permitem calcular estatísticas fundamentais de intensidade dentro de uma região de interesse sem considerar relações espaciais, com base em métricas como média, mediana, valor máximo, valor mínimo, desvio-padrão, assimetria, uniformidade, entropia, curtose, entre outros descritores (11,16,18,38–40).

Os descritores de propriedades de textura fundamentam-se na análise das variações espaciais dos níveis de intensidade de *píxeis* ou *voxéis*. Essa abordagem permite quantificar a dispersão espacial entre os elementos da imagem refletindo as transições abruptas de diferentes intensidades. Contribuindo para avaliação detalhada de padrões de heterogeneidade das lesões e agressividade, aspeto de particular interesse na área da oncologia. (16,18)

As matrizes mais comumente utilizadas para descrever características de textura responsáveis por quantificar a heterogeneidade nos *píxeis* ou *voxéis* são os métodos: GLCM (do inglês *Gray-level co-occurrence matrix*), GLDM (do inglês *Gray-level dependence matrix*), GLRLM (do inglês *Gray-level run length matrix*), GLSZM (do inglês *Gray-level size zone matrix*), NGTDM (do inglês *Neighborhood Gray- Tone Difference matrix*) que iremos de seguida descrever: (39,40)

- A **GLCM** é uma matriz de coocorrência de nível de cinza que avalia a frequência com que pares de *píxeis* ou *voxéis* com determinados níveis de cinza, ocorrem numa determinada relação espacial de distância e direção na região de interesse (ROI do inglês *region of interest*) como é demonstrado pela Figura 2.3 (14,39). Esta classe permite calcular métricas como a entropia (mede a heterogeneidade), a energia ou segundo momento angular (mede a uniformidade), o contraste (mede a diferença entre os níveis de cinza entre *píxeis* ou *voxéis*), entre outros descritores. A partir desta matriz, derivam vinte e quatro descritores passíveis de cálculo (14).

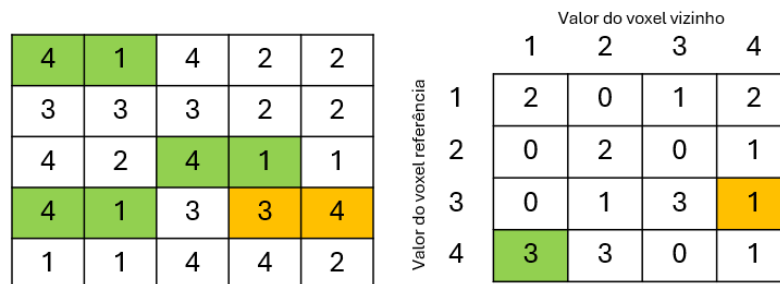


Figura 2.3- Exemplo ilustrativo do cálculo de descritores de textura a partir de matrizes GLCM (horizontal). Adaptada da seguinte fonte bibliográfica: Marius E. et. al. (2020).

- A **GLDM** é uma matriz de dependência de nível de cinza que quantifica a dependência de níveis de cinza, através de uma distância entre os *píxeis* ou *voxéis* conectados aos *píxeis* ou *voxéis* centrais. A partir desta matriz, derivam quatorze descritores passíveis de cálculo (14,39).
- A **GLRLM** é uma matriz de comprimento de sequência de nível de cinza que avalia os *gray-level runs* que consistem no número de *píxeis* ou *voxéis* consecutivos que apresentam o mesmo nível de cinza sobre uma determinada direção como é ilustrado na Figura 2.4 (14,39). Alguns dos cálculos a partir desta matriz avaliam fração de *píxeis* ou *voxéis* consecutivos com o mesmo nível de cinza, variação dos

níveis de cinza dos *gray-level runs*, entre outras classes. A partir desta matriz, derivam dezasseis descritores passíveis de cálculo (14,39).

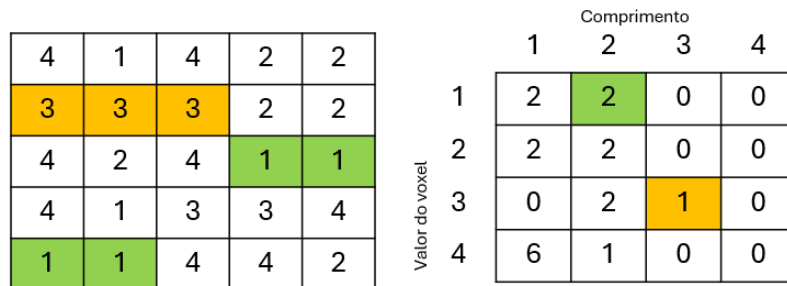


Figura 2.4- Exemplo ilustrativo do cálculo de descritores de textura a partir de matrizes GLRLM (horizontal). Adaptada da seguinte fonte bibliográfica: Marius E. (2020).

- A **GLSZM** é uma matriz de tamanho de zonas de cinza que avalia os pares de *píxeis* ou *voxéis* com o mesmo nível de cinza em zonas, conforme ilustrado pela Figura 2.5. Algumas das características medidas incluem a distribuição de pequenas e grandes zonas, bem como a distribuição dos níveis de cinza dessas zonas no volume de interesse. A partir desta matriz, derivam dezasseis descritores passíveis de cálculo (14).

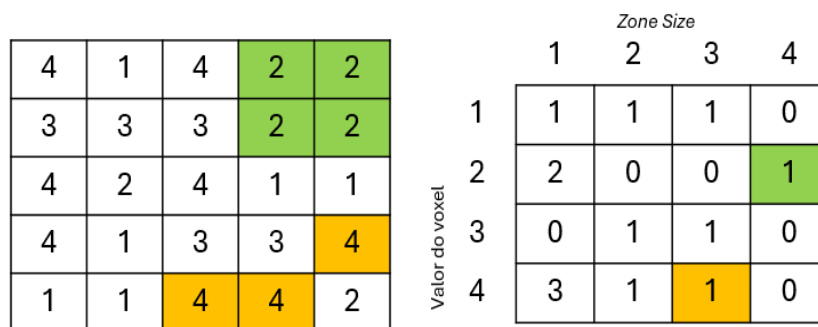


Figura 2.5- Exemplo ilustrativo do cálculo de descritores de textura a partir de matrizes GLSZM. Adaptada da seguinte fonte bibliográfica: Marius E. (2020).

- A **NGTDM** é uma matriz de diferença de nível de cinzentos de vizinhança que quantifica a soma das diferenças entre o nível de cinzento de um *píxel* ou *voxel* e o valor médio de cinzento dos vizinhos, dentro de uma determinada distância conforme ilustrado pela Figura 2.6. Entre as características avaliadas destacam-se a ocupação associada às transições abruptas de tons de cinzento, o contraste, a resistência e a aspereza, que avalia a uniformidade num determinado local. A partir desta matriz, derivam cinco descritores passíveis de cálculo (14).

4	1	4	2	2
3	3	3	2	2
4	2	4	1	1
4	1	3	3	4
1	1	4	4	2

Figura 2.6- Exemplo ilustrativo do cálculo de descritores de textura a partir de matrizes NGTDM. Adaptada da seguinte fonte bibliográfica: Marius E. (2020).

Adicionalmente, descritores de textura podem ser aplicados em domínios estatísticos de ordem superior, que consistem em operações mais complexas realizadas entre mais de três *píxeis* ou *voxéis*, podendo incluir a aplicação de métodos estatísticos, filtros ou transformações matemáticas nas imagens, tais como: decomposição *wavelet* em diferentes níveis, a transformada de *Laplace* e funcionais de *Minkowski*, entre outros métodos. Com o objetivo de identificar padrões de imagens, reduzir o ruído e aumentar o detalhe na imagem (10,16,18,40,41).

Por fim, os descritores de propriedades de forma, ou morfológicos, descrevem propriedades associadas à geometria da região de interesse, definida em duas ou três dimensões. Estes recursos incluem medidas como a compacidade, esfericidade, volume, área, entre outros cálculos (11,14,18,40).

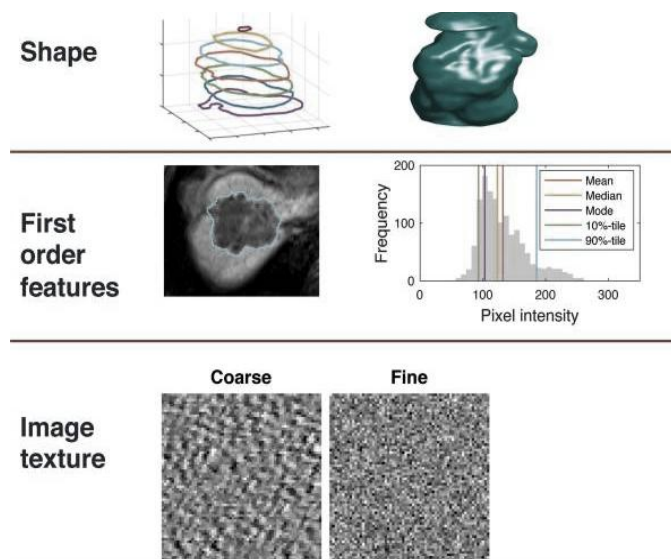


Figura 2.7-Representação visual dos descritores de propriedades de forma, primeira-ordem e textura (Shur et al., 2021).

2.6 Modelos de *Machine Learning*

O desenvolvimento de modelos preditivos em radiômica considera múltiplos fatores, entre os quais se destaca a definição de um propósito clínico relevante, a dimensionalidade e qualidade dos dados disponíveis, bem como a proporcionalidade entre ambos. Esses aspectos são fundamentais para avaliar a viabilidade do estudo e estimar a probabilidade de sucesso. Assim, o desenho experimental deve ser estruturado de forma, a assegurar resultados robustos, reproduzíveis e passíveis de interpretações confiáveis (12,18).

De forma geral, o processo de desenvolvimento de modelos radiômicos pode ser definido em seis etapas principais, como ilustrado na Figura 2.8: **(i)** recolha e preparação dos dados, **(ii)** seleção de descritores radiômicos, **(iii)** seleção do modelo e dos hiperparâmetros, **(iv)** treino do modelo de *machine learning*, **(v)** avaliação do desempenho e **(vi)** validação externa (18).



Figura 2.8- Etapas do desenvolvimento de um modelo de *machine learning*. O desenvolvimento do modelo de *machine learning* tem como objetivo a resolução de diferentes tipos de problemas, sendo estes definidos por três elementos: (i) tarefa a ser aprendida, (ii) a métrica de desempenho a ser otimizada e (iii) o modo como o modelo adquire dados e interações para aprimorar aprendizagem. Adaptada da seguinte fonte bibliográfica Alzubi et al. (2018)

- (i)** A etapa de recolha e preparação de dados corresponde ao processo de extração de descritores radiômicos, que serão utilizados como *input* do modelo, os quais devem ser sujeitos a controlo de qualidade e organização, de modo a constituírem uma base sólida para as etapas sucessivas (18,24).
- (ii)** Após a recolha dos dados, ocorre a seleção de descritores radiômicos, um processo crítico, uma vez que a extração de descritores frequentemente gera bases de dados de elevada dimensionalidade. Este efeito requer a aplicação de métodos estatísticos e computacionais destinados à redução da

dimensionalidade e à seleção de descritores reprodutíveis, relevantes e robustos, alinhados com o objetivo do estudo nomeadamente, que correlacionem os dados com processos biológicos subjacentes (7,13,18,24,30,31).

O efeito da elevada dimensionalidade, também designado como maldição da dimensionalidade, é particularmente relevante na análise de dados, sobretudo em cenários nos quais existe um grande número de variáveis preditoras (p) e um diminuto número de amostras independentes (n). Essa desproporção aumenta o risco de *overfitting*, processo que ocorre quando o modelo perde a capacidade de generalização ao memorizar padrões específicos dos dados de treino, resultando em sobreestimativa do desempenho (6,9,18,28,39,42,43). Adicionalmente, a elevada dimensionalidade implica maiores exigências em termos de tamanho amostral (n) para alcançar significância clínica, maior demanda de recursos computacionais e maior suscetibilidade a falsos positivos devido à multiplicidade de hipóteses (44).

Em contexto, de amostras limitadas, esta limitação pode ser parcialmente superada por meio da partilha de dados multicêntricos ou do uso de bases de dados públicas, estratégias que permitem aumentar a diversidade dos dados. Uma vez, que modelos treinados com maior conjunto de amostras demonstram ser mais estáveis e confiáveis, quando validados sobre dados externos, possivelmente porque variações e *outliers* afetam os modelos de forma distinta, tornando-os mais robustos (45).

- (iii) A seleção do modelo e dos hiperparâmetros constitui uma etapa em que são definidos os algoritmos mais adequados para responder ao problema em estudo, considerando as características dos dados e os objetivos clínicos estabelecidos. Nesta etapa, procede-se à otimização de hiperparâmetros, processo essencial para ajustar o desempenho dos algoritmos e maximizar a capacidade preditiva e de generalização (24). A escolha criteriosa dos modelos e da respetiva construção influencia diretamente a robustez, a reprodutibilidade e a aplicabilidade clínica dos resultados obtidos. Embora os estudos realizados por (18,44) verifiquem que as combinações de distintas metodologias obtêm resultados promissores em diferentes modalidades imagiológicas, trata-se ainda de uma área crítica de investigação. Esse facto sublinha a importância de relatórios de desempenho de diferentes abordagens, contribuindo para o avanço da investigação em radiómica (18,44).

- (iv) A etapa de treino do modelo de *machine learning* consiste em utilizar o conjunto de descritores radiômicos para treinar o modelo, com o objetivo de prever um diagnóstico, prognóstico ou desfecho clínico, por meio da otimização do algoritmo (44). Este processo baseia-se na implementação de capacidades de aprendizagem dos dados, dando suporte ao processo de tomada de decisão através de previsões futuras (24).

- (v) A avaliação do desempenho constitui uma etapa essencial no desenvolvimento de modelos, uma vez que permite estimar o erro de teste associado ao método de aprendizagem adotado. Este processo consiste em avaliar o desempenho do modelo por meio de métricas, antes da sua implementação (9,24). Para conjuntos de dados extensos, pode-se recorrer ao método *hold-out*, no qual o conjunto de dados é dividido em subconjuntos de treino e teste, no entanto, os resultados podem variar substancialmente em função do acaso da divisão. Já em situações de dimensão amostral menor, aplicam-se técnicas de validação cruzada, como *leave-one-out* ou *k-fold*, que proporcionam maior robustez estatística (18). Modelos que apresentam elevada precisão, sensibilidade e especificidade são considerados aptos para aplicação em novos conjuntos de dados independentes (18).

- (vi) A validação externa corresponde à aplicação do modelo em conjuntos de dados independentes, provenientes de diferentes instituições ou populações, permitindo avaliar a sua capacidade de generalização e robustez em contexto clínico distinto. Esta etapa possibilita a avaliação independente da reprodutibilidade e da aplicabilidade prática dos modelos radiômicos, assegurando, que o desempenho não se restringe ao conjunto de treino e teste. Modelos que mantêm métricas favoráveis quando submetidos à validação externa são considerados mais estáveis e clinicamente translacionais, reforçando a sua utilidade na medicina personalizada (18,45).

2.6.1 Tipos de Aprendizagem dos Modelos

Após a extração de descritores de propriedades, estes tornam-se variáveis preditoras utilizadas na construção de modelos através de ferramentas de IA. Esses modelos baseiam-se no ajuste de funções matemáticas e estatísticas que estabelecem a relação entre o desfecho e os preditores, visando não apenas a previsão de respostas em novas observações, mas também a compreensão das interações existentes entre desfechos clínicos e as variáveis explanatórias (16,24,46).

A IA, ramo da ciência de computação, dedica-se a tarefas originalmente realizadas pela inteligência humana (47,48). Entre as suas ferramentas, destacam-se as técnicas de *machine learning*, um conjunto de algoritmos computacionais com capacidade de aprendizagem que visam classificar achados imagiológicos e/ou prever desfechos clínicos relevantes. Estas técnicas são fundamentais no reconhecimento autónomo de padrões em conjuntos de dados, utilizando estratégias de aprendizagem supervisionada, não supervisionada e semi-supervisionada, com o mínimo de intervenção humana, conforme ilustrado na Figura 2.9 (16–18,39,45,46).

Os modelos supervisionados baseiam-se em amostras previamente classificadas Y_i para, a partir delas, calcular os parâmetros do classificador, permitindo depois atribuir uma classificação a novos dados ($X_{1i}, X_{2i}, \dots, X_{pi}, i = 1, 2, \dots, n$) não utilizados no treino (16). Vários métodos estatísticos integram este domínio de aprendizagem, tais como: *decision tree*, *random forest*, *support vector machine*, regressão logística, *deep learning*, *Naive Bayes*, redes neuronais artificiais, entre outros (11,16,24,46).

Por sua vez, os modelos não supervisionados recorrem a amostras sem classificação para criar grupos de informações oculta a partir de descritores de similaridade ou de padrões. Um exemplo típico de aprendizagem não supervisionada envolve métodos de *clustering*, redução da dimensionalidade, entre outros (11,16,24,46).

Por último, os modelos semi-supervisionados são utilizados em situações em que a classificação é difícil obter, combinando dados classificados com não classificados no processo de treino e recorrendo, assim, a toda a informação disponível (11,16,24,46).

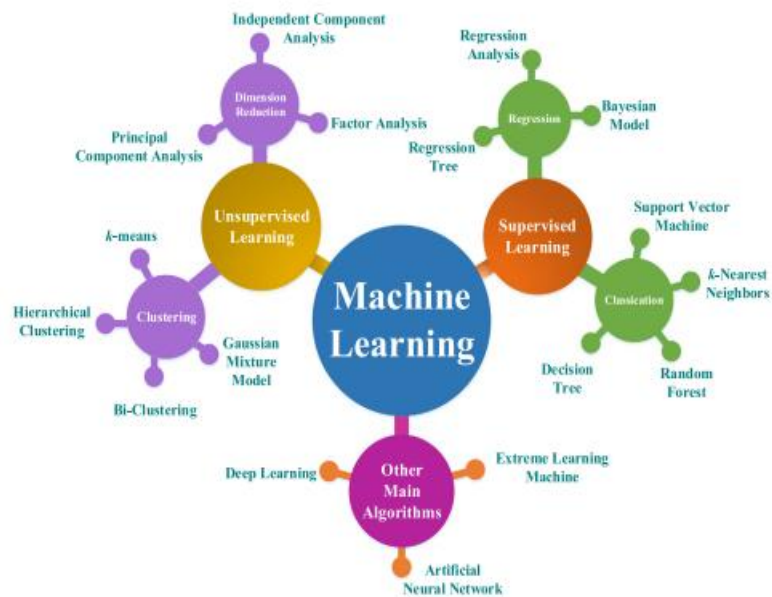


Figura 2.9- Principais algoritmos de *machine learning* diferenciados entre estratégias de aprendizagem supervisionada e não supervisionada J. A. Fajemisin et al. (2024)

2.6.1.1 Classificação da Imagem por *Machine Learning*

Tal como abordado no ponto 2.6.1, os modelos podem ser classificados de acordo com o tipo de aprendizagem adotado. Na sequência, apresentam-se no Quadro 2.3 de forma sintetizada, os conceitos fundamentais relativos aos classificadores de *machine learning* utilizados para classificação de lesões nas imagens mpMRI da próstata, no presente estudo.

Quadro 2.3- Tipologias dos algoritmos supervisionados de *machine learning* utilizados neste estudo/dissertação: *Random Forest*, Regressão Logística e *Support Vector Machines*.

Classificador	Tipo	Descrição
<i>Random Forest</i>	SL	<p>A aprendizagem baseia-se num conjunto composto por múltiplas árvores de decisão, treinadas a partir de subconjuntos de dados obtidos por <i>bootstrapping</i>. Cerca de 2/3 dos dados aleatórios da amostra são utilizados para treino e os restantes dados <i>out-of-bag</i> permitem o cálculo do erro da previsão e acurácia.</p> <p>O percurso ao longo da árvore de decisão inicia-se no nó, raiz e estende-se até às folhas, onde é realizada a decisão final. Em cada nó, são aplicados testes de decisões com base em subconjuntos dos dados de treino, permitindo a divisão progressiva do espaço de decisão. Durante o treino, as variáveis utilizadas em cada nó são selecionadas aleatoriamente a partir de uma subamostra do conjunto de treino, o que contribui para a identificação dos descritores mais relevantes e reduz a correlação entre as árvores do modelo. Os ramos ligam os nós da árvore e direcionam os dados com base nos resultados obtidos no nó anterior, conduzindo-os ao nó seguinte. As folhas representam os nós terminais da árvore e correspondem à atribuição de uma classe ao exemplo analisado. Na fase de teste, cada árvore do conjunto emite um voto, sendo atribuída ao exemplo a classe que obtiver a maioria dos votos (46).</p>

<p>Regressão Logística</p>	<p>SL</p>	<p>Técnica de modelação estatística para prever classes binárias, ou seja, que assumem valores de 0 ou 1. Baseia-se num modelo paramétrico, uma vez, que é definido por um número finito de parâmetros os coeficientes de regressão que normalmente podem estar associados a cada preditor mais uma constante. Este método não consegue detetar relações não lineares complexas entre variáveis dependentes e independentes “automaticamente” nem consegue identificar todas as possíveis interações entre as variáveis preditivas.</p> <p>Expressa através da probabilidade: $\gamma = (1/e^{-e^{-(a+bX)}})$</p> <p>Em que $p > 0,5$ classe 1, se $p < 0,5$ classe 2, X representa o vetor com as variáveis independentes (preditores) e Y a variável dependente (a ser predita) (46).</p>
<p>Support Vector Machines</p>	<p>SL</p>	<p>O <i>classificador Support Vector Machines</i> consiste num modelo linear de classificação, o qual pode ser descrito como aquele em que a saída é obtida a partir de uma combinação linear das variáveis de entrada.</p> <p>Expressa de acordo, com a seguinte expressão de probabilidade: $y = a_0 + a_1x_1$</p> <p>Em que x_1 corresponde aos preditores e a_0, a_1 a parâmetros do modelo. A regra de classificação é estabelecida com base no resultado da função: se $y > 0$, o elemento é atribuído à classe 1, se $y < 0$ é atribuído à classe 2. Esse tipo de abordagem pode ser generalizado para espaços de maiores dimensões nos quais o modelo mapeia pontos de dados e define fronteiras lineares de separação entre classes. O objetivo é encontrar um hiperplano que divida os dados com a maior margem possível em um espaço N-dimensional (49,50).</p>

SL: Supervised Learning

2.7 Fluxo de Trabalho da Radiômica

Os processos envolvidos no fluxo de trabalho da radiômica iniciam-se com a definição dos parâmetros de aquisição de imagem, seguidos do pós-processamento, segmentação de volumes de interesse, cálculo de descritores radiômicos, análise de dados, seleção de descritores radiômicos e, por fim, a construção de modelos preditivos (6,10,11,33,37,38,41,51).

A Figura 2.10 apresenta o fluxo geral de trabalho da radiômica, ilustrando as principais etapas que constituem um estudo (38). Em seguida, cada etapa será abordada de forma detalhada.

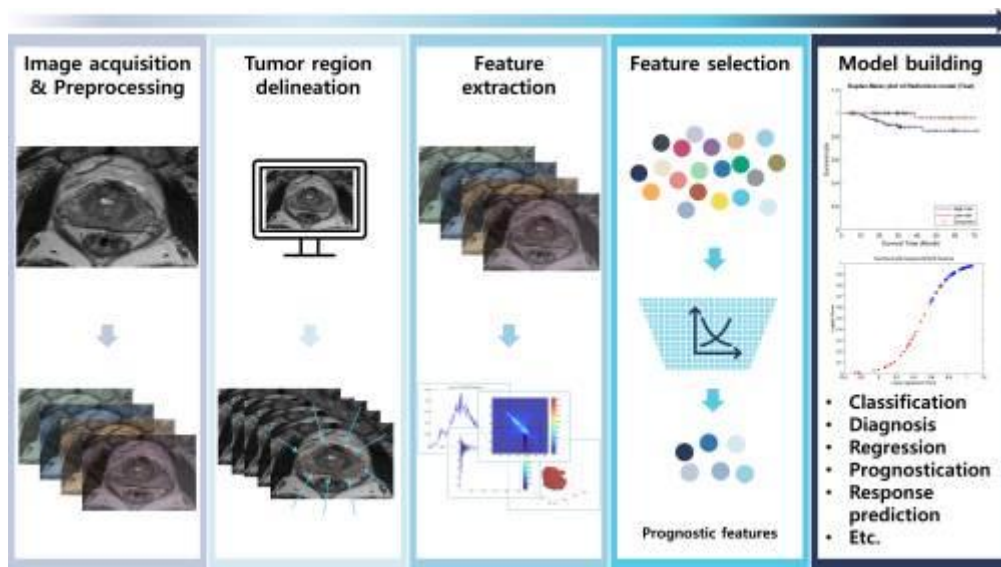


Figura 2.10- Fluxo geral da radiômica: aquisição de imagem, pós-processamento de imagem, extração e seleção de descritores de propriedades e desenvolvimento do modelo, S. Abrol et al. (2017).

2.7.1 Parâmetros de Aquisição de Imagem

No contexto da prática da imagiologia, as imagens médicas são adquiridas em múltiplas instituições, recorrendo a diferentes protocolos de aquisição, com distintos parâmetros de aquisição e reconstrução (36,38,39,52). Estes fatores exercem influência direta sobre as propriedades físicas e quantitativas das imagens, introduzindo uma fonte de variabilidade nos dados (10,11,13,16,38,39).

Exemplos relevantes de parâmetros com influência sobre as imagens de MRI, são apresentados de seguida:

- Resolução espacial e temporal, relação sinal-ruído (SNR do inglês *Signal-to-Noise Ratio*), tipo de sequência, tempo de eco (TE do inglês *Echo Time*) tempo de repetição (TR do inglês *Repetition Time*), intensidade de campo, entre outros (13,18,41).

Adicionalmente, estudos do tipo teste-reteste demonstraram que os descritores radiômicos apresentam sensibilidade a múltiplas fontes de variabilidades não controladas, incluindo o posicionamento do paciente, movimentos involuntários e variações fisiológicas. Estas influências refletem-se nos estudos através de variações no coeficiente de correlação de concordância (CCC do inglês *Concordance Correlation Coefficient*). Além disso, teste-reteste realizados em fantasmas apresentam melhores resultados, possivelmente devido à ausência de variabilidades não controladas (44).

Como resultado, fontes de heterogeneidades têm sido amplamente reconhecidas na literatura (18,20,36,38,39,52) como uma limitação no domínio da radiômica, uma vez que introduzem ruído sistemático e aleatório nos conjuntos de dados, comprometendo a robustez dos descritores extraídos e a sua generalização (18,36,38,39,52). Como, consequência, os descritores extraídos podem não refletir, de forma fidedigna, as propriedades biológicas dos tecidos, apresentado instabilidades significativas mesmo em medições repetidas sob condições aparentemente semelhantes (41).

Para mitigar os efeitos adversos da variabilidade institucional e entre equipamentos, tem-se defendido a uniformização de protocolos de aquisição e reconstrução (13,16,38,39,44,51), bem como a implementação de mecanismos de controlo de qualidade de dados (12). E o incentivo de estudos com recurso a fantasmas ou amostras populacionais, de modo a permitir avaliação sistemática da estabilidade dos descritores face a diferentes fontes de variações (20,39,41).

Neste sentido, diversas entidades internacionais, tais como a Rede de Imagens Quantitativas (QIN, *Quantitative Imaging Network*), Aliança de biomarcadores de Imagem Quantitativa (QIBA, *Quantitative Imaging Biomarker Alliance*), Associação Americana de Físicos em Medicina (AAPM, *American Association of Physicists in Medicine*), Associação Europeia de Medicina Nuclear (EANM, *European Association of Nuclear Medicine*), Sociedade Radiológica da América do Norte (RSNA, *Radiological Society of North America*), entre outras, têm promovido o desenvolvimento de diretrizes normativas com vista à padronização de procedimentos (12).

Paralelamente, investigações têm explorado a aplicação de técnicas de harmonização radiômica, incluindo calibração de algoritmos, normalizações estatísticas, como forma de compensar os efeitos da variabilidade, no entanto, apesar de promissoras, estas abordagens permanecem em investigação, sendo necessário a sua validação (44).

No que respeita à origem dos dados, além das imagens institucionais, observa-se uma crescente utilização de repositórios públicos, que agregam conjunto de dados provenientes de múltiplas instituições. Estas bases de dados, embora introduzam inevitavelmente variabilidade adicional, contribuem positivamente para o desenvolvimento de modelos preditivos com maior capacidade de generalização para diferentes populações clínicas (12,14).

A partilha de imagens médicas, metadados e dados clínicos, sob critérios de anonimização e conformidade ética, revela-se, portanto, indispensável à expansão da radiômica devido à necessidade de *big data* (42). Contudo, a ausência de padronização em dados retrospectivos, frequentemente utilizados, representa um desafio substancial à reprodutibilidade dos estudos (44).

2.7.2 Pós-Processamento de Imagem

Na sequência das variações introduzidas durante a aquisição e reconstrução de imagens médicas, o pós-processamento de imagem constitui uma etapa crucial no fluxo de trabalho da radiômica. Esta fase tem como principal objetivo homogeneizar os diferentes conjuntos de dados, de modo a assegurar a consistência e a fiabilidade dos descritores de propriedades posteriormente extraídos (10).

A aplicação de abordagens sistematizadas e padronizadas nesta etapa é essencial para garantir a reprodutibilidade e a robustez (12).

A imagem médica digital consiste numa função $f(x, y)$ traduzida matricialmente, em que cada interseção das coordenadas x com o y corresponde à unidade mais básica denominada como *píxel* da imagem. Cada *píxel* está associado a um valor de intensidade, que representa um nível da escala de cinza. No caso de volumes tridimensionais, a unidade básica torna-se o *voxel*, cuja dimensão inclui a espessura de corte como profundidade (17). As ferramentas de análise e pós-processamento operam diretamente sobre esta estrutura matricial de *píxeis* e *voxéis*, sendo, portanto, sensíveis a qualquer variação técnica ou biológica (39).

Durante o pós-processamento, aplicam-se várias operações com o intuito de garantir que variações visuais observadas, como o contraste e brilho, reflitam efetivamente alterações das propriedades dos tecidos e não artefactos ou interferências técnicas (39).

Entre os procedimentos comumente utilizados para normalização e padronização dos dados, destacam-se a **(i)** normalização, **(ii)** reamostragem, **(iii)** interpolação, **(iv)** registo de imagem e **(v)** discretização (38,39). Embora úteis, estas técnicas devem ser aplicadas com cautela, uma vez que, apesar de poderem melhorar a qualidade de imagem, também podem mascarar ou degradar assinaturas radiómicas (11).

- (i)** A normalização das intensidades consiste em transformar os valores de intensidades dos *píxeis* para uma determinada faixa de valores de 0 e 1, minimizando variações associadas às variabilidades de aquisição (12,38). Em imagens de MRI, a intensidade de sinal é arbitrária, em consequência, os valores podem variar substancialmente entre diferentes aquisições. Assim, a normalização da intensidade de sinal é recomendada, embora não exista consenso sobre o método ideal, uma das abordagens mais utilizadas é a normalização por *z-score*, expressa pela fórmula matemática:

$$z = \frac{SI_{pixel} - \mu_{ROI}}{\sigma_{ROI}} \quad (2.1)$$

Em que, SI_{pixel} representa a intensidade do sinal em cada *píxel*, μ_{ROI} média da intensidade da região de interesse e σ_{ROI} desvio-padrão correspondente (11).

- (ii)** A reamostragem é utilizada para corrigir diferenças na resolução das imagens, uniformizando o tamanho dos *voxéis* para uma resolução isotrópica. Este procedimento permite a padronização da resolução dos *voxéis* em todos os conjuntos de dados e, assim, homogeneizar os dados de entrada (10,33,38). Segundo o apurado por M. Shafiq-UI-hassan et al. (2017) esta técnica é eficaz na obtenção de descritores de propriedades mais reprodutíveis em conjuntos de dados com diferentes tamanhos de *voxéis* (45,51).
- (iii)** A abordagem de interpolar para obter um *voxel* isotrópico é maioritariamente necessária para as etapas seguintes, nomeadamente na extração de descritores de propriedades baseados em textura. Desta forma, estes tornam-

se rotacionais, permitindo aumentar a reprodutibilidade da extração de descritores e reduzir variações direcionais (10,12,33).

- (iv) O registo de imagem consiste na operação de alinhar imagens, no caso de serem adquiridas em diferentes modalidades de imagens médicas (38).
- (v) A discretização consiste em agrupar os valores de intensidades das imagens em intervalos definidos, designados por *bins* para normalizar os valores de intensidades (10,12,33). Este procedimento melhora a eficiência do cálculo dos descritores radiómicos uma vez, que permite reduzir o ruído nas imagens. A escolha do número de níveis de discretização é fundamental dado que, aumentar o número de níveis de discretização pode aumentar o detalhe da imagem, mas também pode adicionar ruído indesejável, por outro lado diminuir os níveis de discretização pode remover informação importante de textura (10,11,15,33).

A suscetibilidade dos descritores de propriedades radiómicos a alterações do tamanho do voxel, tipo de interpolação, número de *bins*, entre outros, representa um desafio adicional à reprodutibilidade dos resultados (39). Por este motivo, a literatura (39,40) recomenda a adoção de protocolos de pós-processamento rigorosamente documentados e reprodutíveis, incluindo detalhes de cada etapa realizada (39,40).

Adicionalmente, com vista à harmonização e padronização dos processos de pós-processamento de imagem, foi desenvolvida pela Iniciativa de Padronização de Biomarcadores de Imagem (*IBSI do inglês Image Biomarker Standardization Initiative*), um conjunto de normas e metodologias para pós-processamento e extração de descritores, com intuito de reduzir a variabilidade e aumentar a confiança na aplicação clínica entre instituições (40,53).

2.7.3 Segmentação- Identificação de Regiões de Interesse

A segmentação consiste numa técnica de pós-processamento de imagem que visa, delinear um volume ou região denominado como ROI, definindo o conjunto de *píxeis* a serem avaliados posteriormente na análise radiómica (10,16,17,28,38).

Esta etapa é considerada uma das mais críticas, uma vez que os processos subsequentes dependem fortemente da região definida. A segmentação pode ser realizada por diversas abordagens, nomeadamente: manual, automática e semi-automática (10,16,18,38).

A segmentação manual é a mais utilizada e baseia-se na capacidade visual do operador para delinear as regiões, o que compromete a eficiência e a reprodutibilidade devido às variações intra e interobservador. Além disso, exige um tempo considerável para execução, tornando-a inviável para grandes conjuntos de dados (10,16,18,38,39,41).

Como alternativa, as técnicas semi-automáticas e automáticas visam automatizar o processo de segmentação, oferecendo maior rapidez, precisão e reprodutibilidade na definição dos limites das regiões (41). A segmentação automática utiliza algoritmos, como métodos de: *thresholding*, crescimento de regiões, *clustering* e técnicas avançadas de *deep learning*. Embora diminua a variabilidade, a generalização destes algoritmos ainda enfrenta limitações, especialmente em regiões com contornos irregulares (10,39).

A segmentação semi-automática combina algoritmos automáticos com a intervenção do operador, baseando-se numa abordagem híbrida entre segmentação manual e automática (38).

Os métodos de segmentação à semelhança de outras etapas do fluxo de trabalho da radiômica apresentam um impacto direto na reprodutibilidade e na fiabilidade dos descritores radiômicos extraídos. Como verificamos, existem várias abordagens de segmentação que apresentam variações significativas quanto à precisão, consistência e suscetibilidade a erros (39).

Embora os métodos manuais ainda sejam amplamente usados na prática clínica e na investigação, são particularmente vulneráveis ao viés intra e interobservador, o que pode introduzir incertezas substanciais na definição dos volumes e comprometer a consistência dos descritores. Por outro lado, os métodos automáticos, apesar de eficazes em reduzir essa variabilidade, apresentam limitações em contextos clínicos mais complexos, como lesões com contornos heterogêneos ou mal definidos, onde a precisão e a fiabilidade da segmentação ainda não são suficientes (39).

Os métodos semi-automáticos surgem como uma solução intermédia, conciliando a eficiência dos algoritmos com a supervisão do operador. Esta abordagem híbrida equilibra a repetibilidade dos processos automáticos com o controlo manual, sendo frequentemente adotada em estudos de radiômica onde a precisão é fundamental (39).

2.7.4 Extração de Descritores Radiômicos

Após a etapa anterior e posterior ao pós-processamento de imagem, segue-se a extração dos descritores de propriedades de dados de elevadas dimensões (10). Os descritores radiômicos são extraídos dos volumes e subvolumes de interesse definidos previamente, tanto em imagens originais e/ou imagens derivadas, obtidas através da aplicação de filtros de imagem das quais resultam o cálculo das variáveis numéricas predictoras (47).

Estes descritores de propriedades são definidos matematicamente o que permite que sejam interpretados numericamente e posteriormente biologicamente (38,54)

No Quadro 2.4 encontram-se esquematizados os tipos de descritores de propriedades mais comumente utilizados nos estudos de radiômica, organizados em classes. No entanto, no ponto 2.5.1 apresentam-se, de forma detalhada, os descritores de propriedades adotados no estudo/dissertação.

Quadro 2.4- Descritores radiômicos baseados na documentação *PyRadiomics*. Adaptada das seguintes fontes bibliográficas: Cho et al., (2022) e PyRadiomics, (2016), consultado 5 de agosto 2025.

Classe	Medida de análise	Descritores de propriedade
Primeira- ordem	Análise ao histograma de níveis de cinza	<i>10Percentile, 90Percentile, Energy, Entropy, Interquartile Range, Kurtosis, Maximum, Mean Absolute Deviation, Mean, Median, Minimum, Range, Robust Mean Absolute Deviation, Root Mean Squared, Skewness, Total Energy, Uniformity, Variance.</i>
Textura	Análise de textura GLCM	<i>Autocorrelation, Cluster Prominence, Cluster Shade, Cluster Tendency, Contrast, Correlation, Difference Average, Difference Entropy, Difference Variance, Inverse Difference, Inverse Difference Moment, Inverse Difference Moment Normalized, Inverse Difference Normalized, Informational Measure Of Correlation1, Informational Measure of Correlation2, Inverse Variance, Joint Average, Joint Energy, Joint Entropy, Maximal Correlation Coefficient, Maximum Probability, Sum Average, Sum Entropy, Sum Squares.</i>
	Análise de textura GLRLM	<i>Short Run Emphasis, Long Run Emphasis, Gray Level Non-Uniformity, Gray Level Non-Uniformity Normalized, Run Length Non-Uniformity, Run Length Non-Uniformity Normalized, Run Percentage, Gray Level Variance, Run Variance, Run Entropy, Low Gray Level Run Emphasis, High Gray Level Run Emphasis, Short Run Low Gray Level Emphasis, Short Run High Gray Level Emphasis, Long Run Low Gray Level Emphasis, Long Run High Gray Level Emphasis..</i>

	Análise de textura NGTDM	<i>Coarseness, Contrast, Busyness, Complexity, Strength.</i>
	Análise de textura GLDM	<i>Small Dependence Emphasis, Large Dependence Emphasis, Gray Level Non-Uniformity, Dependence Non-Uniformity, Dependence Non-Uniformity Normalized, Gray Level Variance, Dependence Variance, Dependence Entropy, Low Gray Level Emphasis, High Gray Level Emphasis, Small Dependence Low Gray Level Emphasis, Small Dependence High Gray Level Emphasis, Large Dependence Low Gray Level Emphasis, Large Dependence High Gray Level Emphasis.</i>
	Análise de textura GLSZM	<i>Gray Level Non-Uniformity, Gray Level Non-Uniformity Normalized, Gray Level Variance, High Gray Level Zone Emphasis, Large Area Emphasis, Large Area High Gray Level Emphasis, Large Area Low Gray Level Emphasis, Low Gray Level Zone Emphasis, Size Zone Non-Uniformity, Size Zone Non-Uniformity Normalized, Small Area Emphasis, Small Area High Gray Level Emphasis, Small Area Low Gray Level Emphasis, Zone Entropy, Zone Percentage, Zone Variance.</i>
Forma	Descreve a geometria da região de interesse	<i>Elongation, Flatness, Least Axis Length, Least Axis Length, Maximum 2D Diameter Column, Maximum 2D Diameter Row, Maximum 2D Diameter Slice, Maximum 2D Diameter, Mesh Volume, Sphericity, Surface Area, Surface Volume Ratio, Voxel Volume.</i>

GLCM: *Gray-level co-occurrence matrix*, **GLRLM:** *Gray-level run length matrix*, **NGTDM:** *Neighborhood Gray- Tone Difference*
GLDM: *Gray-level dependence matrix*, **GLSZM:** *Gray-level size zone matrix*

2.7.5 Seleção de Descritores Radiômicos

Conforme abordado no tópico 2.6 ponto (ii), a etapa subsequente à extração de descritores radiômicos é essencial, uma vez que o número de descritores possíveis de extrair é virtualmente ilimitado, dependendo do ajuste dos parâmetros, das classes de descritores radiômicos e da aplicação de filtros na imagem (12,28). Embora valiosos, esses dados geram desafios estatísticos e computacionais significativos. Nesse sentido, torna-se indispensável a aplicação de métodos, com vista a identificação de subconjuntos de dados relevantes. Contribuindo para a melhoria do desempenho preditivo dos modelos, redução de redundâncias, redução de *overfitting*, otimização dos processos de treino, maximização da demanda computacional, aumento da interpretabilidade dos modelos e favorecimento da implementação para a prática clínica (40,45,54,55).

Neste contexto, são removidos os descritores não-informativos que apresentam elevada correlação entre si sem contribuição informativa relevante (11,54). Assegurando desta forma, que o subconjunto de descritores selecionados são o mais robustos possível, isto é, capazes de manter repetibilidade em condições semelhantes e reprodutibilidade sobre variações resultantes dos subprocessos envolvidos no fluxo de trabalho (11,45,56).

Para que um descritor de propriedade seja um potencial biomarcador de imagem, além da robustez, deve demonstrar um desempenho consistente, desenvolvendo resultados confiáveis e repetíveis (11,51).

Diversos métodos podem ser utilizados para a seleção de descritores radiômicos, podendo ser classificados em supervisionados e não supervisionados. A escolha da abordagem depende de fatores como a natureza dos *inputs* e/ou desfecho clínico, o tamanho amostral, a complexidade dos descritores, os objetivos do estudo e o tipo de modelo adotado (57). Essas abordagens diferem quanto à integração no algoritmo de aprendizagem, à complexidade computacional e à capacidade de detecção de redundâncias e correlações entre descritores (57).

Entre os métodos supervisionados, frequentemente considerados como preferenciais, destacam-se três técnicas: (i) filtragem, (ii) *wrapper* e (iii) *embedded* (11,29,38,45,54,55).

- (i) Os métodos de filtragem consistem em classificar variáveis de acordo com as suas propriedades estatísticas, selecionando as mais relevantes com base em associações diretas (38). Entre as ferramentas aplicadas encontram-se o Coeficiente de Correlação de *Pearson*, Coeficiente de Correlação Intraclasse (ICC do inglês *intraclass correlation coefficient*), CCC, método redundância Mínima - relevância máxima (*mRMR* do inglês *minimum Redundancy - Maximum Relevance*), bem como testes não paramétricos (29,54,55).
- (ii) Os métodos *wrapper* avaliam diretamente o desempenho preditivo do modelo através de *machine learning*, selecionando subconjuntos de descritores que maximizam esse desempenho (29,54) Embora mais exigentes em termos computacionais, são altamente eficazes. Podem recorrer a algoritmos como o método *Branch and Bound* ou mais simples, como os algoritmos sequenciais, que iniciam o processo sem ou com recursos e vão adicionando ou removendo consecutivamente até maximizar a função objetivo que será o maior desempenho predito (54).

- (iii) Os métodos *embedded* realizam a seleção de descritores durante o processo de treino do modelo, sendo o método Lasso um dos mais utilizados (54).

Por outro lado, os métodos não supervisionados são particularmente relevantes para a redução da dimensionalidade, uma vez que identificam padrões e relações entre dados sem depender do desfecho clínico. Entre as técnicas mais utilizadas encontram-se (i) técnicas de análise de *clusters* e (ii) redução de dimensionalidade (10,28).

- (i) A análise de *clusters* consiste numa técnica que permite agrupar descritores altamente correlacionados e selecionar o descritor mais representativo de cada grupo, recorrendo a métodos estatísticos, ou de *machine learning* ou de visualização de dados (44).
- (ii) A análise de componentes principais (PCA do inglês *Principal Component Analysis*) é uma técnica amplamente utilizada para reduzir a dimensionalidade de um conjunto de dados, preservando simultaneamente a variabilidade através da identificação de padrões e da sua representação, extraindo a informação mais relevante e expressando-a como um novo conjunto de variáveis, denominadas componentes principais (29,44,46).

Diversas abordagens podem ser utilizadas para a seleção de descritores, cada uma com potencial para identificar os subconjuntos mais informativos. A aplicação desses métodos é flexível e pode inclusive envolver estratégias híbridas que recorrem à aplicação sequencial de dois ou mais métodos ou até métodos que integram os resultados de múltiplas abordagens, de modo a maximizar a robustez e a estabilidade do subconjunto de descritores selecionados (9).

No que respeita a escolha do método mais adequado, a literatura evidencia um intenso debate e contínuo desenvolvimento de novas propostas (9). No entanto, é consensual que diferentes métodos de seleção conduzem, inevitavelmente, a subconjuntos distintos de descritores. Contudo, não existe um método universalmente superior, cada abordagem é capaz de identificar os descritores mais relevantes dentro do conjunto global disponível (9).

2.7.6 Desenvolvimento do Modelo

Em termos gerais, a aprendizagem supervisionada pode ser utilizada para resolver dois tipos de problemas: classificação e regressão. Sendo os de classificação os mais aplicados em radiômica no ramo da oncologia, através do cálculo de probabilidade de uma amostra pertencer a um resultado de classificação, como o diagnóstico patológico (55). Os modelos matemáticos apresentam distintos níveis de complexidade, capacidade computacional e poder preditivo, sempre associado a uma determinada taxa de erro, representada pela ocorrência de classificações incorretas (55).

Como descrito no ponto 2.6, este tipo de algoritmos de *machine learning* recebem como *input* os descritores radiômicos, que se comportam como variáveis independentes, sobre os quais a classificação já se encontra associada. Pretendendo-se assim que o algoritmo aprenda os padrões característicos de cada grupo, de modo a posteriormente classificar futuras observações, com base no conhecimento adquirido (13).

Para avaliar a precisão do processo de aprendizagem, utiliza-se o conjunto de teste, que idealmente deve corresponder, pelo menos, a um terço do tamanho da amostra (11). O desempenho dos modelos é frequentemente analisado por meio da matriz de confusão, a qual resume os resultados comparando as previsões realizadas com os valores reais, conforme ilustrado no Quadro 2.5.

Quadro 2.5- Matriz de confusão resume o desempenho do modelo, através dos seguintes elementos: real positivo e negativo vs. previsto positivo e negativo. Adaptada da seguinte fonte bibliográfica: *PyRadiomics*, (2016), consultado 5 de agosto 2025.

	Previsto Positivo	Previsto Negativo
Real Positivo	Verdadeiro Positivo (TP)	Falso Negativo (FN)
Real Negativo	Falso Positivo (FP)	Verdadeiro Negativo (TN)

A análise do desempenho permite não apenas aferir a qualidade e a aplicabilidade de um modelo, mas também comparar diferentes algoritmos de *machine learning* em termos de capacidade preditiva, interpretabilidade dos dados e viabilidade de implementação (45).

Entre as principais métricas utilizadas na avaliação de modelos em problemas binários ou multiclasse, destacam-se (58,59):

- **Sensibilidade**, representa a fração de classes positivas previstas corretamente, avalia a capacidade em identificar corretamente os casos positivos.

$$TPR = \frac{TP}{TP+FN} \quad (2.2)$$

Onde TP é o número de classes positivas previstas corretamente, e FN o número de exemplos positivos classificados como negativos pelo classificador.

- **Especificidade**, avalia a capacidade do classificador de prever a classe negativa.

$$TNR = \frac{TN}{TN+FP} \quad (2.3)$$

Onde TN é o número de classes negativas previstas corretamente, e FP o número de exemplos negativos classificados como positivos pelo classificador.

- **Acurácia**, determina a percentagem de previsões corretas identificadas nas duas previsões, é uma métrica de desempenho do modelo.

$$Acúrcia = \frac{TP+TN}{TP+FP+TN+FN} \quad (2.4)$$

- **F1-Score**, é a média harmônica entre precisão (PPV) e sensibilidade.

$$F1_score = \frac{2 \times \text{precisão} \times \text{sensibilidade}}{\text{precisão} + \text{sensibilidade}} \quad (2.5)$$

- **Valores Positivos Preditivos (PPV)**, determina a percentagem de previsões corretas entre todas as previsões positivas.

$$PPV = \frac{TP}{TP+FP} \quad (2.6)$$

- **Valor Preditivo Negativo (NPV)**, determina a percentagem de classe classificadas como negativas que realmente pertencem à classe negativa.

$$NPV = \frac{TN}{TN+FN} \quad (2.7)$$

- **Área sob a curva ROC (AUC)**, quantifica a área sob a curva ROC, que representa a relação entre a TPR e TNR de um classificador. Um valor de AUC próximo a 1 indica um excelente desempenho do modelo classificador.
- **Coefficiente de Correlação de Matthews**, determina e avalia a qualidade geral das previsões, tem por base as quatro categorias integrantes da matriz de confusão.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (2.8)$$

- **Kappa de Cohen**, avalia o grau de concordância entre as previsões do modelo e os valores reais, corrigida pela concordância do acaso.

$$\mathcal{K} = \frac{p_0 - p_e}{1 - p_e} \quad (2.9)$$

Onde, p_0 é a acurácia observada e p_e a acurácia esperada pelo acaso.

- **Acurácia balanceada**, consiste na média entre a sensibilidade e a especificidade.

$$Acurácia\ Balanceada = \frac{Sensibilidade + Especificidade}{2} \quad (2.10)$$

Avaliar classificadores em saúde exige mais do que métricas estatísticas, alguns dos quais foram brevemente mencionados ao longo da fundamentação teórica. Devem ser considerados na análise: a quantidade e qualidade de dados, a adequação do tamanho das amostras, a partição de conjuntos de treino e teste e análise de equilíbrio entre classes (24,57,60).

Adicionalmente, é essencial integrar o impacto clínico dos erros, adaptar os modelos às necessidades clínicas e envolver os profissionais na interpretação dos resultados. Métricas isoladas podem induzir a interpretações imprecisas. Por exemplo, num conjunto de dados com 16% de sujeitos com patologia e 84% sem, um classificador que atribuísse a todos os casos a classe negativa apresentaria acurácia de 84%, mas seria clinicamente irrelevante,

por não distinguir as classes. Tal como uma sensibilidade de 90% pode ser aceitável num contexto clínico e em outros não (60).

Outro aspeto fundamental refere-se ao controle de *overfitting* e *underfitting* se o erro diminuir no conjunto de treino, mas aumentar no conjunto de validação, frequentemente estamos perante *overfitting*. Por outro lado, quando o modelo apresenta baixo desempenho em ambos os conjuntos, possivelmente estamos perante *underfitting*, evidenciando que o modelo é excessivamente simplista para capturar relações subjacentes nos dados (57).

No contexto clínico, torna-se indispensável a análise das complexidades e a sua interpretabilidade. Um modelo que apresente resultados satisfatórios, mas cuja lógica interna seja opaca, pode comprometer a confiança, dificultando a identificação de erros e limitando a sua aplicação em ambientes clínicos e a integração em sistemas regulamentados (4,11,12,29,30).

3 Materiais e Métodos

3.1 Amostra

3.1.1 Base de Dados *The Cancer Imaging Archive* (TCIA)

Para o presente estudo/dissertação, foram extraídos dados da base pública TCIA (do inglês *The Cancer Imaging Archive*), um repositório internacional de imagens médicas que reúne estudos de carácter multicêntrico e longitudinal (61,62). Este arquivo constitui um recurso de referência para: a investigação e o desenvolvimento de iniciativas educacionais no domínio das patologias oncológicas (61,62).

No seu primeiro ano, o TCIA arquivou aproximadamente 3,3 milhões de imagens, organizadas segundo a patologia, a modalidade de imagem/tipo e o objetivo de investigação, garantindo assim uma estrutura sistemática de apoio à pesquisa científica (61,62). Adicionalmente, dados complementares relacionados às imagens, como desfechos clínicos, detalhes do tratamento, genómica e análise de imagem, são também disponibilizados quando acessíveis (61,62). Estes recursos encontram-se à disposição da comunidade científica, para auxiliar na investigação sobretudo de áreas direcionadas para a radiómica, *machine learning*, IA, correlações clínico-genómicas e ensaios clínicos (61,62).

No âmbito da presente dissertação, foi selecionado o *dataset PROSTATEx*, um conjunto de dados retrospectivo de mpMRI da próstata, destinado ao desenvolvimento e à avaliação de métodos quantitativos de análise de imagem para a classificação diagnóstica de cancro da próstata clinicamente significativo. Disponíveis em <https://www.cancerimagingarchive.net/collection/prostatex/>, DOI: 10.7937/K9TCIA.2017.MURS5CL (consultado em 23 de Outubro de 2023) (62). Por se

tratar de dados de domínio público, não foi necessária aprovação do comitê de ética. As aquisições de imagem foram realizadas no Centro Médico Universitário *Radboud* designado *Radboudumc*, localizado em *Nijmegen*, Holanda. O *dataset PROSTATEx* foi desenvolvido entre 21 de novembro de 2016 a 16 de fevereiro de 2017 (62).

3.1.1.1 Protocolo de Aquisição de Imagem

Para a aquisição das imagens, os estudos de mpMRI incluíram as seguintes sequências ponderadas: T2, DCE, DWI e densidade protônica. As imagens foram obtidas em dois equipamentos de MRI de 3 Tesla, ambos do fabricante *Siemens Medical Systems*, nomeadamente, os modelos *MAGNETOM Trio* e *Skyra*, sem bobina endorretal. Os protocolos de aquisição adotados basearam-se nos parâmetros técnicos apresentados no Quadro 3.1 (62).

Quadro 3.1-Parâmetros de aquisição das imagens do estudo *PROSTATEx*. Adaptada da seguinte fonte bibliográfica: (Litjens G. (2017))

Parâmetros de aquisição do estudo <i>PROSTATEx</i>		
Ponderação	Sequência	Parâmetros técnicos
T2	<i>Turbo Spin Echo</i>	<ul style="list-style-type: none"> Resolução: 0,5mm Espessura de Corte: 3,6mm
DCE	<i>Turbo Flash Gradient Echo</i>	<ul style="list-style-type: none"> Resolução: 1,5mm Espessura de Corte: 4mm Resolução Temporal: 3,5ms
Densidade Protônica	<i>Turbo Flash Gradient Echo</i>	<ul style="list-style-type: none"> Tempo de eco e repetição: variável <i>Flip angle</i>: variável
DWI	<i>Single-shot Echo Planar Imaging</i>	<ul style="list-style-type: none"> Resolução: 2mm Espessura de Corte: 3,6mm Gradientes de difusão: 3 direções Níveis de sensibilidade (b=50, 400 e 800 s/mm²) Mapas de Coeficiente de Difusão Aparente

T2: *T2-weighted imaging*, **DCE:** *Diffusion-Weighted Imaging*, **DWI:** *Dynamic Contrast-Enhanced*

3.1.1.2 Segmentação do *Dataset PROSTATEx*

A segmentação das imagens mpMRI da próstata foram realizados manualmente em imagens axiais ponderadas em T2 e em baixo valor b de DWI, por um radiologista especializado em 10 anos em urologia e um médico do internato, independentemente (23,62).

As distintas zonas segmentadas tiveram como objetivo reduzir a complexidade das etapas subsequentes de classificação (62). O *dataset PROSTATEx* disponibiliza quatro segmentações distintas, correspondendo a diferentes regiões anatómicas:

- Glândula prostática completa;
- Zona periférica;
- Zona anexa, incluindo: zona de transição, zona central e estroma anterior da próstata;
- Lesões prostáticas;

Após a segmentação manual, as máscaras foram binarizadas, com fundo definido como zero, tendo sido posteriormente submetidas a métodos automatizados de clusterização para correção e remoção de imprecisões através da FSL (*FMRIB Software library*) (versão 5.0) (62). Todas as segmentações estão disponíveis gratuitamente no repositório online: (https://rcuocolo.github.io/PROSTATEx_masks/)(63).

No que respeita às lesões prostáticas, estas foram segmentadas nas sequências T2 e DWI em alto valor b, totalizando 299 lesões, dado que alguns pacientes apresentam múltiplas regiões de interesse. Cada máscara da lesão possui um indentificador único, permitindo a correspondência precisa entre as imagens originais e as respetivas segmentações (23,62,63).

3.1.1.3 Critérios de Seleção da Amostra em Estudo

O conjunto de dados recolhidos inicialmente no Centro Médico Universitário englobou 348 exames de mpMRI de próstata, obtidos de 347 sujeitos, dos quais 183 não apresentavam evidência patológica e 165 tinham diagnóstico de cancro da próstata, mais concretamente 187 lesões (64).

A seleção da amostra contemplou exclusivamente 204 exames de mpMRI com lesões classificadas como PI-RADS ≥ 2 , tendo sido excluídos 144 exames que não cumpriam o critério. Lesões com pontuações inferiores a PI-RADS 2 foram desconsideradas, dado que,

no centro de recolha dos dados, a incidência de lesões clinicamente significativas nesta categoria é inferior a 5% (62).

3.1.1.4 Dados da Biópsia Histopatológica

O *dataset* dispõe adicionalmente de informação histopatológica relativa às lesões incluídas no grupo PI-RADS ≥ 3 . Estas lesões foram submetidas a biópsias e classificadas de acordo com os grupos de *Gleason*, disponíveis para consulta na Figura 3.1.

Um achado é considerado clinicamente significativo quando o *Gleason Score* atribuído é superior ou igual a 7, correspondendo aos grupos de *Gleason* 2,3, 4 e 5 respetivamente (62). Para uma melhor compreensão da categorização dos grupos *Gleason* é disponibilizado o Quadro A.1 de apoio no apêndice A.

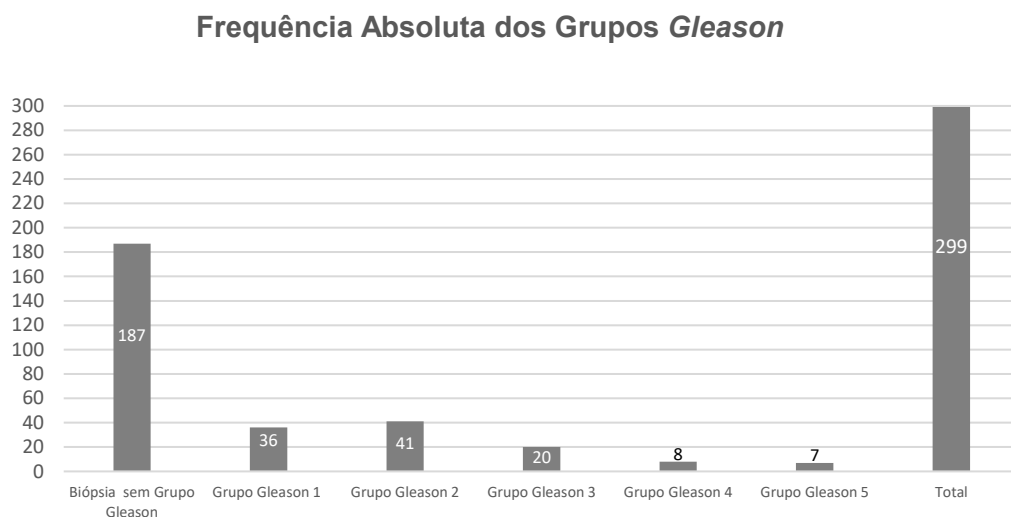


Figura 3.1-Frequência absoluta dos grupos *Gleason* das lesões prostáticas. Adaptada da seguinte fonte bibliográfica: Lijens G et al. (2017), consultado 3 de Janeiro 2025.

3.1.1.5 Critérios de Seleção da Subamostra em Estudo

Critérios de Inclusão da Coorte e Sequência de Estudo

A população em estudo nesta dissertação compreende 204 exames, correspondente à coorte de lesões prostáticas disponibilizadas em mpMRI, na sequência ponderada em T2. Todas as outras coortes presentes na base de dados *PROSTATEx* foram excluídos.

Para a população mencionada, foram utilizados exclusivamente as imagens mpMRI, as máscaras das lesões e o ficheiro classes com os dados da biópsia histopatológica.

3.1.1.6 Caracterização da Subamostra em Estudo

Após a aplicação dos critérios de inclusão relativos à coorte, à sequência e aos dados em estudo, a amostra foi composta por 299 lesões. Entre estas, 76 associadas a biópsias classificadas como clinicamente significativas, e 223 a lesões consideradas não clinicamente significativas. Na Figura 3.2 e 3.3 é apresentada a caracterização da população e das lesões em estudo, com base nos dados TCIA. Para uma caracterização mais detalhada do intervalo de idades, é disponibilizada a Figura B.1 no apêndice B.

Género	Espécie	Idade	Antigénio Específico da Próstata Total	Volume das Lesões
Masculino	Homo sapiens	63,47 ± 7,1 anos	13,62 ± 9,18	1,42 ± 0,41 ml

Figura 3.2-Characterização da população em estudo. Adaptada da seguinte fonte bibliográfica: Mylona et al. (2024) e Litijens et al. (2017), consultado 4 janeiro de 2025.

Classificação da Biópsia Histopatológica

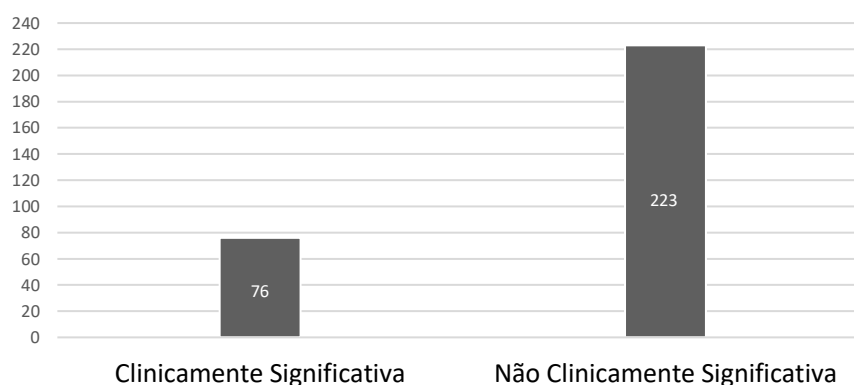


Figura 3.3-Classificação da biópsia histopatológica da amostra: 76 lesões classificados em clinicamente significante e 223 em não clinicamente significante. Adaptada da seguinte fonte bibliográfica: Litijens et al. (2017), consultado 4 janeiro 2025.

3.1.1.7 Conjunto de Dados Definidos para Estudo

Para o estudo foram considerados dois conjuntos de dados, designados originais e dados randomizados, diferenciando-se entre si pelo método de obtenção. O conjunto de dados originais foi obtido através da base de dados TCIA, previamente caracterizada no ponto 3.1.1.5. Já o conjunto de dados randomizados foram desenvolvidos a partir da duplicação dos dados originais, submetidos a uma operação de randomização da intensidade dos *píxeis* aplicada a todo o conjunto da amostra.

Essa transformação preserva as propriedades estatísticas de intensidade, mas elimina a informação estrutural e anatômica, resultando em imagens sem conteúdo morfologicamente interpretável. A Figura 3.4 ilustra este processo, apresentando o resultado da operação de randomização aplicada à imagem *PROSTATEx_0203 t2*, em comparação com a sua versão original.

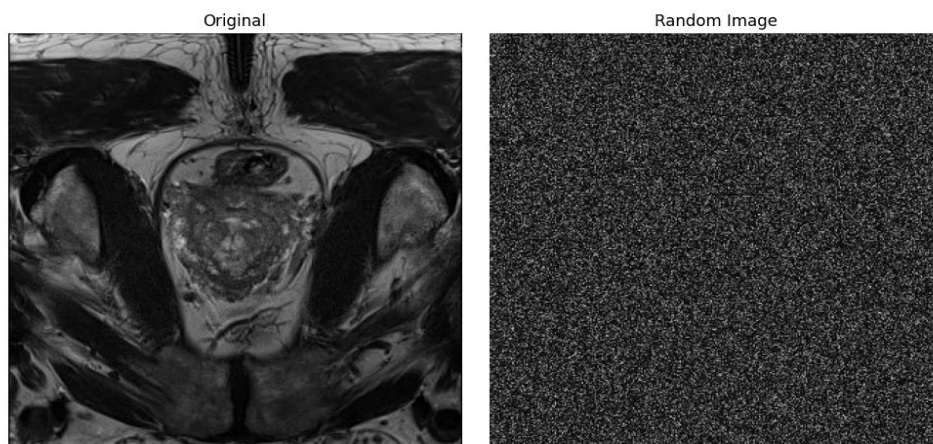


Figura 3.4 Imagem original e randomizada *PROSTATEx_0203 t2* após aplicação da operação de randomização das intensidades de píxeis através das bibliotecas: *pandas*, *numpy*, *SimpleITK*.

3.2 Definição do Pós-Processamento de Imagem

As imagens resultantes dos conjuntos anteriores foram submetidas a um processo estruturado de pós-processamento de imagem, apresentados no Quadro 3.2, que inclui etapas de normalização, otimização do desempenho computacional, padronização das dimensões do *voxel*, ajuste das máscaras e discretização. A seleção destes parâmetros baseou-se no compromisso de padronizar a qualidade de imagem, minimizar as variabilidades inerentes aos processos de aquisição e garantir a reprodutibilidade na extração dos descritores.

Inicialmente, procedeu-se à normalização da intensidade dos *píxeis*, recorrendo ao parâmetro *normalize*, que ajusta a média para 0 e o desvio-padrão para 1. Ainda com o intuito de garantir que a maioria dos valores presentes, nas intensidades dos *píxeis* normalizadas seja igual ou superior a 0 recorreremos ao *voxelArrayShift* para assegurar que 99,7% dos valores ocupem um valor acima de 0, este processo é crucial para o cálculo de alguns descritores de propriedades baseados em primeira-ordem.

De modo a otimizar o desempenho computacional, recorreu-se ao parâmetro *preCrop*, que permite reduzir, o número de dados processados e eliminar regiões irrelevantes da imagem.

Com o objetivo de padronizar as dimensões do *voxel*, aplicaram-se os parâmetros *sitkBSpline* e *resampledPixelSpacing*, definindo uma resolução de [0,6, 0,6, 0]. Como os *voxéis* são anisotrópicos, a extração foi realizada em duas dimensões, utilizando os parâmetros *force2D* e *force2Ddimension*, uma vez, que a escolha da resolução de imagem é particularmente importante. A reamostragem, melhora a reprodutibilidade dos descritores e evita que diferenças nos valores extraídos resultem de variações da resolução. Assim optou-se pela reamostragem no plano, com extração bidimensional, em detrimento da tridimensional, que poderia introduzir perda de informação (no caso da redução de resolução) ou a inclusão de informação não adquirida (no caso do aumento de resolução).

A redefinição e o ajuste das máscaras foram realizados com os parâmetros *resegmentMode* e *resegmentRange*, que removem intensidades fora do intervalo $média \pm 3 \times desvio\ padrão$.

Por último, foi aplicada a discretização de intensidades, assegurando que o número de *bins* utilizado na extração de descritores permanecesse entre 30 e 130, mediante avaliação prévia da gama de intensidades do volume de interesse. Esta etapa contribui para a redução de ruído e para a melhoria da reprodutibilidade dos descritores extraídos. De acordo com a literatura (15,30), recomenda-se a utilização de uma largura de *bins* que resulte num

número entre 16 e 128, o que tem demonstrado garantir boa consistência na reprodutibilidade dos descritores, sem variações estatisticamente significativas.

Quadro 3.2-Parâmetros de pós-processamento de extração estabelecidos no *PyRadiomics* versão 3.0.1.

Parâmetros de Extração de Descritores	
<i>Normalize</i>	<i>true</i>
<i>NormalizeScale</i>	100
<i>VoxelArrayShift</i>	300
<i>PreCrop</i>	<i>true</i>
<i>Interpolator</i>	' <i>sitkBSpline</i> '
<i>ResampledPixelspacin</i>	[0,6, 0,6, 0]
<i>Force2D</i>	<i>true</i>
<i>Force2Ddimension</i>	0
<i>GeometryTolerance</i>	1.e+3
<i>ResegmentMode</i>	' <i>sigma</i> '
<i>ResegmentRange</i>	[-3, 3]
<i>BinWidth</i>	0,2
<i>label</i>	1

3.3 Metodologias de Cálculo dos Descritores Radiômicos

Para atingir os objetivos propostos nesta dissertação, foi necessário projetar e implementar código utilizando o *Google Colab*, um ambiente baseado em *Jupyter Notebook*. O cálculo dos descritores radiômicos foram extraídos através da biblioteca de código aberto *PyRadiomics* versão 3.0.1 (<https://github.com/Radiomics/PyRadiomics/tree/3.0.1>) (65). Este *software* foi selecionado devido às várias vantagens, nomeadamente: facilidade de partilha intercomunidade, compatibilidade com vários formatos de imagem (*Dicom*, *Nifti*, entre outros), acesso livre, reprodutibilidade e padronização com as diretrizes IBSI (53,66).

3.3.1 Extração de Descritores Radiômicos

A extração de descritores radiômicos foi realizada em ambos os conjuntos de dados, de acordo com os parâmetros de extração apresentados no Quadro 3.3. A extração contemplou as imagens originais e derivadas de filtros, obtidas através da aplicação dos seguintes: *Laplaciano do Gaussiano (LoG do inglês Laplacian of Gaussian)* ($\sigma = 0,6, 2$ e 3), *wavelet* (decomposto em dois níveis), *square*, *square-root*, logaritmo e exponencial. A utilização desses filtros permitiu desenvolver novas representações das imagens, possibilitando uma melhor quantificação e detecção de padrões clinicamente relevantes.

Quadro 3.3- Parâmetros de extração definidos no *PyRadiomics* versão 3.0.1 dos descritores de propriedades primeira-ordem, forma e textura em tipos de imagens distintos.

Parâmetros de Extração de Descritores		Classe de Descritores
Tipo de Imagem		
Original	{}	Forma
<i>LoG</i>	sigma: [0,6, 2, 3]	Primeira-ordem
<i>Wavelet</i>	2 níveis	GLCM
<i>Square</i>	{}	GLRLM
<i>Square-Root</i>	{}	GLSZM
Logaritmo	{}	GLDM
Exponencial	{}	NGTDM

GLCM: *Gray-level co-occurrence matrix*, **GLRLM:** *Gray-level run length matrix*, **NGTDM:** *Neighborhood Gray- Tone Difference* **GLDM:** *Gray-level dependence matrix*, **GLSZM:** *Gray-level size zone matrix*, **LoG:** *Laplacian of Gaussian*

A Figura 3.5 exemplifica a aplicação dos referidos filtros à imagem original. Para uma compreensão mais aprofundada da aplicação dos filtros, são disponibilizadas informações complementares no Quadro C.1 no apêndice C.

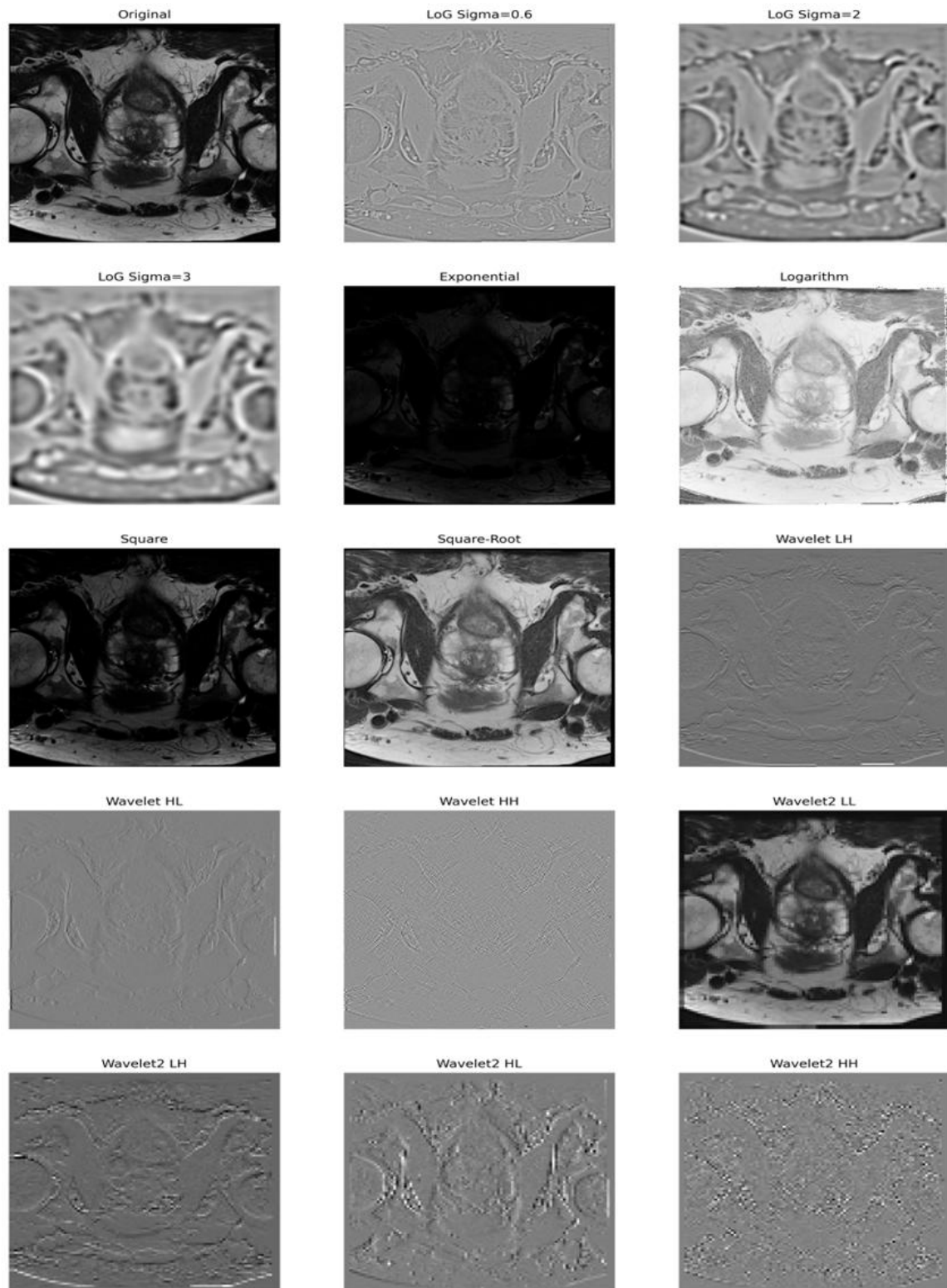


Figura 3.5- Aplicação dos filtros LoG (sigma 0,6 mm, 2mm, 3mm), exponencial, logaritmo, *Square*, *Square-Root*, *Wavelet* de decomposição em 2 níveis.

Foram ainda determinadas as classes e os descritores radiômicos a serem extraídos, definidas pelo parâmetro *FeaturesClasses* apresentadas no Quadro 3.4, 3.5 e 3.6. Essas classes compreendem: primeira-ordem, forma e textura, sendo esta última subdividida nas matrizes: GLCM, GLRLM, GLSZM, NGTDM e GLDM. Concretamente, extraídas sobre um conjunto de 18 descritores de primeira-ordem apresentadas no Quadro 3.4.

Quadro 3.4- *FeaturesClass* de primeira-ordem definida por 18 descritores através do *software PyRadiomics*. Adaptada das seguintes fontes bibliográficas: Zwanenburg (2020) e *PyRadiomics*, consultado 20 janeiro de 2025.

Primeira-Ordem
<i>10Percentile</i>
<i>90Percentile</i>
<i>Energy</i>
<i>Entropy</i>
<i>InterquartileRange</i>
<i>Kurtosis</i>
<i>Maximum</i>
<i>MeanAbsoluteDeviation</i>
<i>Mean</i>
<i>Median</i>
<i>Minimum</i>
<i>Range</i>
<i>RobustMeanAbsoluteDeviation</i>
<i>RootMeanSquared</i>
<i>Skewness</i>
<i>TotalEnergy</i>
<i>Uniformity</i>
<i>Variance</i>

Foram considerados 14 descritores de forma, apresentadas no Quadro 3.5, cuja extração foi realizada apenas a partir das imagens originais, não sendo aplicado às imagens filtradas. Tal deve-se ao facto de esta propriedade depender diretamente da geometria da lesão, preservando assim a fidelidade anatómica e a reprodutibilidade dos descritores.

Quadro 3.5-*FeaturesClass* de forma definidos por 14 descritores através do *software PyRadiomics*. Adaptada das seguintes fontes bibliográficas: Zwanenburg (2020) e *PyRadiomics*, consultado 20 janeiro de 2025.

Descritores de Propriedades de Forma

Elongation
Flatness
LeastAxisLength
MajorAxisLength
Maximum2DDiameterColumn
Maximum2DDiameterRow
Maximum2DDiameterSlice
Maximum3DDiameter
MeshVolume
MinorAxisLength
Sphericity
SurfaceArea
SurfaceVolumeRatio
VoxelVolume

Foram incluídos 75 descritores de textura, apresentadas no Quadro 3.6, organizadas de acordo com as seguintes matrizes: 24 descritores de GLCM, 16 descritores de GLRLM e GLSZM, 14 descritores de GLDM, 5 descritores de NGTDM.

Quadro 3.6-*FeaturesClass* Textura definidos pelas matrizes GLCM, GLRLM, GLSZM, NGTDM e GLDM através do *software PyRadiomics*. Adaptada da seguinte fonte bibliográfica: Zwanenburg (2020) e *PyRadiomics*, consultado 20 janeiro de 2025.

GLCM	GLRLM	GLSZM	NGTM	GLDM
<i>Autocorrelation</i>	<i>GrayLevelNonUniformity</i>	<i>GrayLevelNonUniformity</i>	<i>Busyness</i>	<i>DependenceEntropy</i>
<i>ClusterProminence</i>	<i>GrayLevelNonUniformityNormalized</i>	<i>GrayLevelNonUniformityNormalized</i>	<i>Coarseness</i>	<i>DependenceNonUniformity</i>
<i>ClusterShade</i>	<i>GrayLevelVariance</i>	<i>GrayLevelVariance</i>	<i>Complexity</i>	<i>DependenceNonUniformityNormalized</i>
<i>ClusterTendency</i>	<i>HighGrayLevelRunEmphasis</i>	<i>HighGrayLevelZoneEmphasis</i>	<i>Contrast</i>	<i>DependenceVariance</i>
<i>Contrast</i>	<i>LongRunEmphasis</i>	<i>LargeAreaEmphases</i>	<i>Strength</i>	<i>GrayLevelNonUniformity</i>
<i>Correlation</i>	<i>LongRunHighGrayLevelEmphasis</i>	<i>LargeAreaHighGrayLevelEmphasis</i>		<i>GrayLevelVariance</i>
<i>DifferenceAverage</i>	<i>LongRunLowGrayLevelEmphasis</i>	<i>LargeAreaLowGrayLevelEmphasis</i>		<i>HighGrayLevelEmphasis</i>
<i>DifferenceEntropy</i>	<i>LowGrayLevelRunEmphasis</i>	<i>LowGrayLevelZoneEmphasis</i>		<i>LargeDependenceEmphasis</i>
<i>DifferenceVariance</i>	<i>RunEntropy</i>	<i>SizeZoneNonUniformity</i>		<i>LargeDependenceHighGrayLevelEmphasis</i>
<i>Id</i>	<i>RunLengthNonUniformity</i>	<i>SizeZoneNonUniformityNormalized</i>		<i>LargeDependenceLowGrayLevelEmphasis</i>
<i>Idm</i>	<i>RunLengthNonUniformityNormalized</i>	<i>SmallAreaEmphases</i>		<i>LowGrayLevelEmphasis</i>
<i>Idmn</i>	<i>RunPercentage</i>	<i>SmallAreaHighGrayLevelEmphasis</i>		<i>SmallDependenceEmphasis</i>

<i>Idn</i>	<i>RunVariance</i>	<i>SmallAreaLowGrayLevelEmphasis</i>		<i>SmallDependenceHighGrayLevelEmphasis</i>
<i>Imc1</i>	<i>ShortRunEmphasis</i>	<i>ZoneEntropy</i>		<i>SmallDependenceLowGrayLevelEmphasis</i>
<i>Imc2</i>	<i>ShortRunHighGrayLevelEmphasis</i>	<i>ZonePercentage</i>		
<i>InverseVariance</i>	<i>ShortRunLowGrayLevelEmphasis</i>	<i>ZoneVariance</i>		
<i>JointAverage</i>				
<i>JointEnergy</i>				
<i>JointEntropy</i>				
<i>MaximumProbability</i>				
<i>MCC</i>				
<i>SumAverage</i>				
<i>SumEntropy</i>				
<i>SumSquares</i>				

3.4 Data

Com base nos processos anteriores descritos para a extração, foi obtido um total de 421291 descritores radiômicos. Estes descritores resultaram da extração para cada lesão de 93 descritores que incluem (primeira-ordem, GLCM, GLDM, GLRLM, GLSZM e NGTDM), aplicados a 15 subtipos de imagens (original e derivadas de filtros), incluindo também 14 descritores de forma extraídos exclusivamente das imagens originais.

Este conjunto representa uma ampla variedade de atributos e uma riqueza significativa de dados. No entanto, apresenta desafios metodológicos, nomeadamente o número limitado de amostras, o desequilíbrio entre classes e a elevada dimensionalidade das variáveis preditoras, aspetos que exigem cuidados específicos na construção de modelos de *machine learning*.

3.5 Seleção de Descritores Radiômicos

3.5.1 Análise Estatística Multivariada

A seleção de descritores radiômicos, neste estudo, foi realizada por meio de análise estatística multivariada, visando alcançar os objetivos definidos. Para esta análise, foram utilizados os testes estatísticos: correlação de *Spearman* e coeficiente de correlação intraclasse, implementados em código *python* com recurso às bibliotecas *Pandas*, *Scipy* e *Pingouin* (versão 0.5.5) (67–69). A aplicação da análise estatísticas recorreu a métodos sequenciais sobre critérios estabelecidos que determinaram:

- I. Teste de correlação *Spearman* com vista, na determinação de correlações baixas a moderadas entre descritores de propriedade de primeira-ordem e textura vs. forma aplicado ao conjunto de dados originais.

A. Critério de Seleção de Descritores Não-Redundantes:

Inclusão apenas das variáveis descritoras com correlação de *Spearman* menor a 0,8 ($\rho < 0,8$).

B. Caracterização do Critério:

A condição definida pretende garantir a independência e identificar multicolinearidades entre descritores de propriedades de forma e primeira-ordem/textura, através do limiar implementado. Considerou-se que valores nominais próximos de 0,2 identificam correlações fracas ou inexistentes, valores superiores representam correlações moderadas e valores acima de 0,8 caracterizam correlações elevadas (70).

- II. Determinação do coeficiente de correlação intraclasse de modo, a determinar correlações baixas a moderadas entre descritores radiômicos aplicados ao conjunto de dados original e randomizado.

A. Critérios de Seleção de Descritores Informativos:

Inclusão apenas das variáveis descritoras com coeficiente de correlação intraclasse menores ou iguais a 0,7 ($\rho < 0,7$).

B. Caracterização do Critério:

A condição estabelecida tem como objetivo avaliar a confiabilidade entre associações e identificar correlações baixas a moderadas entre descritores radiômicos no conjunto de dados, o que provavelmente é expectável uma vez, que existe uma deterioração morfológica provocada pela operação de randomização. Considerou-se que valores menores a 0,5 indicam fraca concordância, valores entre 0,7 e 0,9 correspondem a concordância moderada e valores superiores a 0,9 caracterizam elevada concordância (71).

3.5.1.1 Critério de Inclusão de Descritores Informativos

Os descritores de propriedades que apresentem simultaneamente valores elevados de correlação de *Spearman* entre propriedades de forma vs. primeira-ordem/textura e ICC entre os dois conjuntos original e randomizado, foram considerados redundantes e não-informativos, dado que, não acrescentam valor discriminativo aos modelos. Assim, apenas os descritores radiômicos que cumprem os critérios de seleção são selecionados como informativos e combinados através de operações entre conjuntos.

3.5.1.2 Operação entre Conjuntos de Dados e Determinação de *Inputs*

Para o tratamento dos conjuntos de dados anteriores, foram aplicadas operações de interseção e união que originaram posteriormente dois *inputs*:

- ***Input_Interseção***: Operação de interseção ($A \cap B$), consideradamente mais conservadora e rigorosa uma vez, que apenas inclui descritores de propriedades que sejam validados pelas duas análises estatísticas.
- ***Input_União***: Operação de união ($A \cup B$), menos conservador, incorpora um maior número de descritores radiômicos, abrangendo aqueles que estão presentes em pelo menos um dos testes estatísticos.

3.5.2 Análise Estatística Univariada

Para a análise estatística univariada, foram utilizadas as bibliotecas: *Scipy*, *Statsmodels*, *Scikit-learn*, *Pandas*, *Numpy* e os respectivos testes estatísticos sobre o conjunto de dados originais: (69,72–76)

- a) Teste de normalidade, avaliado por meio do teste estatístico de *Shapiro-Wilk* representado sobre as hipóteses:

H0: Distribuição normal

H1: Distribuição não normal

Critério de Seleção: Se $p \leq 0,05$, rejeitamos H0, distribuição não é normal, porém se $p > 0,05$, aceitamos H0, assumimos normalidade.

- b) Testes de hipóteses entre as variáveis preditoras e as variáveis a serem preditas, por base nos testes: Teste-t e *Mann-Whitney U*.

Critério de Seleção: Se respeitar uma distribuição normal utiliza-se o Teste-t, porém se a distribuição não for normal aplica-se o teste *Mann-Whitney U*.

- c) Análise do desempenho das variáveis preditoras como classificadores, por meio da avaliação: AUC, *Cutoff-Threshold*, Sensibilidade, Especificidade e Valor-p.

Como múltiplas hipóteses foram testadas, a probabilidade de ocorrência de erros do tipo I aumenta. Para controlar esse risco, aplicaram-se métodos de correção para testes múltiplos, utilizando-se o método de *Bonferroni*.

3.5.2.1 Critérios de Inclusão para o *Input* Descritores Informativos e Preditivos

A. Foi considerado como critério de inclusão, para o *input* de descritores informativos e preditivos, o seguinte critério:

Valor-p $< 0,05$, após aplicação do método de correção de comparações múltiplas *Bonferroni*.

Caso o critério anterior não fosse satisfeito, adotou-se uma metodologia alternativa, baseada na seleção de um descritor radiômico por cada 15 sujeitos. Esta seleção teve por base os descritores com valores superiores de área sob a curva ROC, formulando assim

uma hipótese exploratória de carácter intencional e contribuindo para a avaliação do desempenho de metodologias orientadas pelo critério discriminatório da AUC.

3.6 Construção dos Modelos de *Machine Learning*

3.6.1 Conjunto de Modelos

A construção dos modelos seguiu a mesma metodologia geral, variando apenas no algoritmo de *machine learning* aplicado aos diferentes conjuntos de dados (*inputs*). Para avaliar o impacto dos métodos de seleção de descritores radiômicos, foram implementados quatro modelos distintos:

- **Modelo I:** Conjunto de dados originais
- **Modelo II:** *Input*_Interseção
- **Modelo III:** *Input*_União
- **Modelo IV:** *Input* descritores informativos e preditivos

3.6.2 Parâmetros de Construção dos Modelos

A construção dos modelos envolveu o uso de diversas bibliotecas, entre as quais se destacam a *Scikit-learn*, *Numpy*, *Panda*, *Matplotlib*, *SHAP*, *gdown* (67,72,76).

As variáveis consideradas no modelo foram definidas da seguinte forma:

- χ = *variáveis predictoras, descritores radiômicos (inputs)*
- Y = *variável dependente a ser predita, desfecho clínico (significância clínica)*

O conjunto de dados como referido anteriormente, continha pacientes com múltiplas lesões, o que exigiu procedimentos adicionais para associar corretamente cada conjunto de descritores radiômicos ao respetivo desfecho clínico. Para tal, foi definido um identificador único *Finding ID*, que seguiu o formato: *ProstateX-000n°_Finding n°*, em que *n°* corresponde ao número de identificação do paciente e *Finding n°* ao número sequencial da lesão.

A divisão dos dados foi realizada através do método *Stratified Group K-Fold*, com quatro *folds*, garantindo simultaneamente estratificação e independência entre dados. Contudo, neste estudo foi considerada a partição aproximadamente $\frac{3}{4}$ dos sujeitos para treino e $\frac{1}{4}$ para teste.

Para a seleção do classificador, foram avaliadas diferentes abordagens, incluindo:

- *Support Vector Machine*
- Regressão logística
- *Random Forest*
- *Naive Bayes*
- *K-Nearest Neighbors*
- *Gradient Boosting*

O processo de avaliação foi desenvolvido por meio de validação cruzada aninhada: a validação interna, decomposta em dois *folds*, foi utilizada para a aplicação do *GridSearchCV*, responsável pela seleção de hiperparâmetros, já a validação externa, decomposta em cinco *folds*, permitiu estimar o desempenho médio dos modelos em dados independentes através do *cross_val_score*. Como métrica, foi adotado o *Matthews Correlation Coefficient*, reportado sob a forma de valor médio e desvio-padrão. Esta escolha justifica-se por se tratar de uma métrica robusta em cenários de amostras desbalanceadas, disponibilizando maior poder informativo relativamente a métricas mais tradicionais, como a acurácia balanceada, *Kappa de Cohen* e acurácia (77).

O classificador selecionado compreende o maior valor médio de MCC aliado à menor variabilidade entre os *folds*, equilibrando assim o desempenho e a robustez.

Após a seleção do classificador, a aprendizagem é otimizada com base no melhor hiperparâmetro identificado pelo *GridSearchCV*, implementado através da validação cruzada em cinco *folds* (treino em quatro *folds* e validação no *fold* restante). Hiperparâmetros mais baixos resultam em margens mais amplas e menor propensão ao *overfitting*, enquanto valores mais elevados reduzem a margem, aumentando a penalização sobre erros de classificação e o risco de *overfitting*. O objetivo foi, portanto, identificar o ponto de equilíbrio entre *underfitting* e *overfitting*, assegurando o melhor desempenho generalizável possível.

3.6.3 Avaliação do desempenho dos Modelos

A avaliação do desempenho dos modelos foi realizada com base em diferentes métricas de desempenho, incluindo:

- Área sob a curva ROC (AUC)
- *Coeficiente de Correlação Matthews* (MCC)
- *Kappa de Cohen*
- *F1-score*
- Acurácia
- Sensibilidade
- Especificidade
- Acurácia balanceada
- Valor preditivo positivo
- Valor preditivo negativo

4 Resultados

4.1 Análise Estatística

4.1.1 Correlação de *Spearman* entre Descritores de Primeira-ordem e Textura vs. Forma

A análise estatística realizada através da correlação de *Spearman* ao conjunto de dados originais resultaram em [14 descritores de forma x1395 descritores de primeira-ordem e textura] o que representa um total de 19530 coeficientes de correlações analisados para responder ao primeiro objetivo específico. Foram obtidos os respectivos resultados apresentados na Figura 4.1, considerando o critério de seleção de descritores, o que resultou na seleção de apenas 1009 descritores primeira-ordem e textura, tendo sido eliminados 386 redundantes.

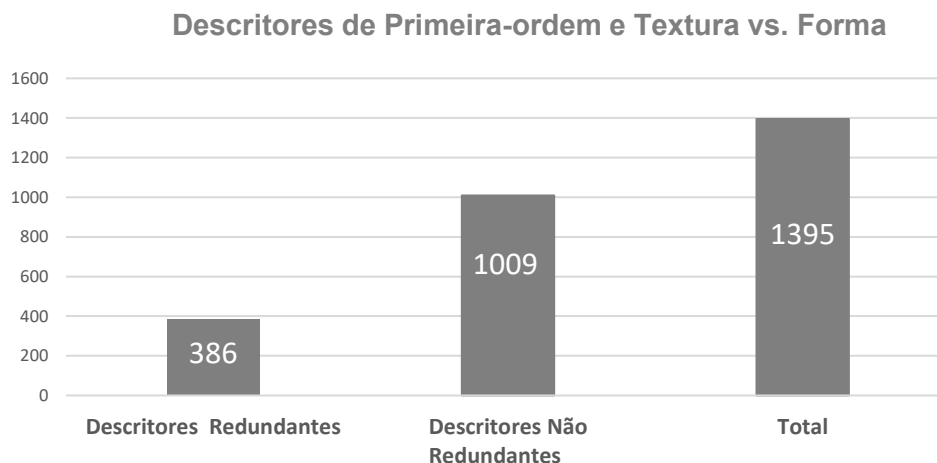


Figura 4.1-Frequência absoluta de descritores de propriedade de primeira-ordem e de textura, classificados como redundantes e não redundantes baseados na correlação de *Spearman*.

De seguida, são apresentados detalhadamente os descritores de propriedades de primeira-ordem e de textura com moderada a elevada correlação com a propriedade de forma, tanto na imagem original como nas imagens derivadas de filtros. De acordo com os dados obtidos, a classe primeira-ordem revelou-se redundante em quatro descritores: *TotalEnergy*, *Energy*, *Entropy* e *Uniformity*, resultando em 36 descritores redundantes, o que corresponde a uma proporção de 4 em 18 descritores da classe.

Relativamente aos subtipos de imagem, foram identificados os mesmos quatro descritores previamente observados na imagem original. Contudo, nas imagens derivadas de filtros essa correspondência, não se verificou, em particular nos filtros: LoG ($\sigma = 0,6, 2, 3$), *square* e exponencial como é observado pelo Quadro 4.1. Adicionalmente, disponibiliza-se na Figura D.1 em apêndice D um mapa de correlação, de modo a evidenciar as correlações entre os descritores de propriedade primeira-ordem e forma.

Quadro 4.1- Descritores de propriedade primeira-ordem redundantes determinados pela correlação de *Spearman*.

Filtros	Descritores de propriedade de primeira-ordem
LoG ($\sigma = 0,6, 2, 3$)	<i>Energy, Total Energy</i>
Square	
Exponencial	
Wavelet	<i>Energy, Entropy, TotalEnergy, Uniformity</i>
Square-root	
Original	

Referente à classe de textura, a matriz GLCM resultou num maior número de descritores redundantes nomeadamente, para os seguintes descritores: *DifferenceEntropy*, *Idmn*, *Idn*, *Imc1*, *Imc2*, *JointEnergy*, *JointEntropy*, *MCC*, *MaximumProbability* e *SumEntropy* totalizando 82 descritores redundantes, o que corresponde a uma proporção de 10 em 24 descritores da classe.

Para o subtipo de imagem original foram identificados seis descritores nomeadamente: *Imc1*, *JointEntropy*, *JointEnergy*, *MCC*, *MaximumProbability* e *SumEntropy*, no entanto, para as imagens derivadas de filtros, não se verificou o mesmo padrão, tal como ilustrado pelo Quadro 4.2. Adicionalmente, disponibiliza-se na Figura D.3 em apêndice D um mapa de correlação, de modo a evidenciar as correlações entre os descritores de propriedade GLCM e forma.

Quadro 4.2- Descritores de propriedade de textura classe GLCM redundantes, determinados pela correlação de *Spearman*.

Filtros	Descritores de propriedade de GLCM
LoG ($\sigma = 0,6$)	<i>Idmn</i> , <i>Imc1</i> , <i>JointEnergy</i> , <i>JointEntropy</i> , <i>MaximumProbability</i> , <i>SumEntropy</i>
LoG ($\sigma = 2, 3$)	<i>Idmn</i> , <i>Idn</i> , <i>Imc1</i> , <i>JointEnergy</i> , <i>JointEntropy</i> , <i>MaximumProbability</i> , <i>SumEntropy</i>
Wavelet	<i>Idmn</i> , <i>Idn</i> , <i>Imc1</i> , <i>Imc2</i> , <i>JointEnergy</i> , <i>JointEntropy</i> , <i>MCC</i> , <i>MaximumProbability</i> , <i>SumEntropy</i>
Square-root	<i>DifferenceEntropy</i> , <i>Imc1</i> , <i>JointEntropy</i> , <i>JointEnergy</i> , <i>MCC</i> , <i>MaximumProbability</i> , <i>SumEntropy</i>
Logaritmo	<i>JointEntropy</i> , <i>JointEnergy</i> , <i>MCC</i> , <i>MaximumProbability</i> , <i>SumEntropy</i>

Relativamente, à matriz GLDM, foram identificadas seis classes redundantes, nomeadamente: *DependenceEntropy*, *DependenceNonUniformity*, *GrayLevelNonUniformity*, *LargeDependenceLowGrayLevelEmphasis*, *LowGrayLevelEmphasis* e *SmallDependenceLowGrayLevelEmphasis* resultando em 79 descritores redundantes, numa proporção na classe de 6 em 14 descritores.

Para a imagem original foram determinadas as mesmas classes evidenciadas anteriormente, porém, nas imagens derivadas de filtros observaram-se resultados distintos apresentados, no Quadro 4.3. Adicionalmente, disponibiliza-se na Figura D.2 em apêndice D um mapa de correlação, de modo a evidenciar as correlações entre os descritores de propriedade GLDM e forma.

Quadro 4.3-Descritores de propriedade de textura classe GLDM redundantes, determinados pela correlação de *Spearman*.

Filtros	Descritores de propriedade de GLDM	
LoG ($\sigma = 0,6$)	<i>DependenceNonUniformity,</i> <i>LargeDependenceLowGrayLevelEmphasis,</i> <i>SmallDependenceLowGrayLevelEmphasis</i>	<i>GrayLevelNonUniformity,</i> <i>LowGrayLevelEmphasis,</i>
LoG ($\sigma = 2,3$) Wavelet Square-root Logaritmo Original	<i>DependenceNonUniformity,</i> <i>LargeDependenceLowGrayLevelEmphasis,</i> <i>SmallDependenceLowGrayLevelEmphasis,</i> <i>DependenceEntropy</i>	<i>GrayLevelNonUniformity,</i> <i>LowGrayLevelEmphasis,</i>
Exponencial	<i>DependenceNonUniformity, GrayLevelNonUniformity</i>	
Square	<i>DependenceNonUniformity</i>	

No âmbito da matriz GLRLM, resultaram sete descritores redundantes, nomeadamente: *GrayLevelNonUniformity,* *GrayLevelNonUniformityNormalized,* *LongRunLowGrayLevelEmphasis,* *LowGrayLevelRunEmphasis,* *RunEntropy,* *RunLengthNonUniformity* e *ShortRunLowGrayLevelEmphasis* totalizando 80 descritores, com uma proporção na classe de 7 em 16 descritores.

Para a imagem original foram observadas todas as classes previamente determinadas anteriormente, contudo, nas imagens derivadas de filtros, verificou-se observações distintas, conforme ilustrado no Quadro 4.4. Adicionalmente, disponibiliza-se na Figura D.5 em apêndice D um mapa de correlação, de modo a evidenciar as correlações entre os descritores de propriedade GLRLM e forma.

Quadro 4.4- Descritores de propriedade de textura classe GLRLM redundantes, determinados pela correlação de *Spearman*.

Filtros	Descritores de propriedade de GLRLM	
LoG ($\sigma = 0,6, 2, 3$)	<i>GrayLevelNonUniformity,</i> <i>LowGrayLevelRunEmphasis,</i> <i>ShortRunLowGrayLevelEmphasis</i>	<i>LongRunLowGrayLevelEmphasis,</i> <i>RunLengthNonUniformity,</i>
Wavelet Square-root Original	<i>RunLengthNonUniformity,</i> <i>ShortRunLowGrayLevelEmphasis,</i> <i>LongRunLowGrayLevelEmphasis,</i> <i>GrayLevelNonUniformityNormalized</i>	<i>GrayLevelNonUniformity,</i> <i>LowGrayLevelRunEmphasis,</i> <i>RunEntropy,</i>
Square Exponencial	<i>GrayLevelNonUniformity, RunLengthNonUniformity</i>	

Logaritmo	<i>RunLengthNonUniformity,</i> <i>ShortRunLowGrayLevelEmphasis,</i> <i>LongRunLowGrayLevelEmphasis, RunEntropy,</i>	<i>GrayLevelNonUniformity,</i> <i>LowGrayLevelRunEmphasis,</i>
------------------	---	---

Em relação à matriz GLSZM, verificou-se sete classes redundantes respectivamente: *GrayLevelNonUniformity,* *GrayLevelNonUniformityNormalized,* *LargeAreaLowGrayLevelEmphasis,* *LowGrayLevelZoneEmphasis,* *SizeZoneNonUniformity,* *SmallAreaLowGrayLevelEmphasis* e *ZoneEntropy,* resultando em 81 descritores redundantes com uma proporção na classe de 7 em 16 descritores.

Para a imagem original foram determinadas as mesmas classes anteriormente observadas, porém, para as imagens derivadas de filtros verificou-se padrões distintos apresentados de seguida, no Quadro 4.5. Adicionalmente, disponibiliza-se na Figura D.4 em apêndice D um mapa de correlação, de modo a evidenciar as correlações entre os descritores de propriedade GLSZM e forma.

Quadro 4.5-Descritores de propriedades de textura classe GLSZM redundantes, determinados pela correlação de *Spearman*.

Filtros	Descritores de propriedade de GLSZM	
LoG ($\sigma = 0,6$)	GrayLevelNonUniformity, LowGrayLevelZoneEmphasis, SmallAreaLowGrayLevelEmphasis.	LargeAreaLowGrayLevelEmphasis, SizeZoneNonUniformity,
LoG ($\sigma = 2, 3$) Logaritmo	GrayLevelNonUniformity, LowGrayLevelZoneEmphasis, SmallAreaLowGrayLevelEmphasis, ZoneEntropy.	LargeAreaLowGrayLevelEmphasis, SizeZoneNonUniformity,
Wavelet Square-root Original	GrayLevelNonUniformity, SmallAreaLowGrayLevelEmphasis, LargeAreaLowGrayLevelEmphasis, GrayLevelNonUniformityNormalized	SizeZoneNonUniformity, LowGrayLevelZoneEmphasis, ZoneEntropy,
Square	GrayLevelNonUniformity, SizeZoneNonUniformity	
Exponencial	GrayLevelNonUniformity	

Relativamente à matriz NGTDM, foram identificadas quatro classes redundantes, nomeadamente: *Busyness,* *Coarseness,* *Contrast* e *Strength* com 28 descritores, correspondendo a uma proporção na classe de 4 em 5 descritores.

Para a imagem original, apenas as classes *Coarseness* e *Contrast* se mostraram redundantes, contudo nas restantes imagens filtradas diferentes resultados são

observados, ilustrados pelo Quadro 4.6. Adicionalmente, disponibiliza-se na Figura D.6 em apêndice D um mapa de correlação, de modo a evidenciar as correlações entre os descritores de propriedade NGTDM e forma.

Quadro 4.6-Descritores de propriedades de textura classe NGTDM redundante, determinados pela correlação de *Spearman*.

Filtros	Descritores de propriedade NGTDM
LoG ($\sigma = 0,6, 2, 3$)	<i>Coarseness, Contrast</i>
Original	
Wavelet	<i>Coarseness, Contrast, Strength, Busyness</i>
Square	
Exponencial	<i>Coarseness</i>
Square-root	<i>Contrast</i>

Com base nos resultados, constatou-se que classes baseadas em primeira-ordem e GLCM preservaram menor redundância, com 14 descritores não redundantes. Seguem-se os descritores das matrizes GLRLM e GLSZM, com 9 descritores cada, e por fim, as matrizes GLDM e NGTDM, com 8 e 1, respetivamente. De seguida, são apresentados os descritores não redundantes organizados por classes no Quadro 4.7 e 4.8.

Quadro 4.7- Lista de descritores radiômicos de primeira-ordem não redundantes resultantes da análise estatística correlação de *Spearman* para as classes descritoras tanto na imagem original como na imagem derivada de filtros.

Primeira-Ordem
<i>10Percentile</i>
<i>90Percentile</i>
<i>InterquartileRange</i>
<i>Kurtosis</i>
<i>Maximum</i>
<i>MeanAbsoluteDeviation</i>
<i>Mean</i>
<i>Median</i>
<i>Minimum</i>
<i>Range</i>
<i>RobustMeanAbsoluteDeviation</i>
<i>RootMeanSquared</i>

Skewness

Variance

Quadro 4.8-Lista de descritores radiômicos de textura não redundantes resultantes da análise estatística correlação de *Spearman* para as classes descritoras tanto na imagem original como na imagem derivada de filtros.

GLCM	GLRM	GLSZM	NGTDM	GLDM
Autocorrelation	GrayLevelVariance	GrayLevelVariance	Complexity	DependenceNonUniformityNormalized
ClusterProminence	HighGrayLevelRunEmphasis	HighGrayLevelZoneEmphasis		DependenceVariance
ClusterShade	LongRunEmphasis	LargeAreaEmphasis		GrayLevelVariance
ClusterTendency	LongRunHighGrayLevelEmphasis	LargeAreaHighGrayLevelEmphasis		HighGrayLevelEmphases
Contrast	RunLengthNonUniformityNormalized	SizeZoneNonUniformityNormalized		LargeDependenceEmphasis
Correlation	RunPercentage	SmallAreaEmphasis		LargeDependenceHighGrayLevelEmphasis
DifferenceAverage	RunVariance	SmallAreaHighGrayLevelEmphasis		SmallDependenceHighGrayLevelEmphasis
DifferenceVariance	ShortRunEmphasis	ZonePercentage		SmallDependenceEmphasis
Id	ShortRunHighGrayLevelEmphasis	ZoneVariance		
Idm				
InverseVariance				
JointAverage				
SumAverage				
SumSquares				

4.1.2 Coeficiente de Correlação Intraclasse entre Descritores Radiômicos do Conjunto de Dados Original e Randomizado

A análise estatística, baseada no coeficiente de correlação intraclasse, foi realizada sobre o conjunto de dados originais (n=299) e sobre um conjunto de dados randomizados equivalente (n=299), ambos compostos por 1409 descritores radiômicos. Foram avaliadas 596 combinações entre os conjuntos de dados originais e randomizados para cada um dos 1409 descritores radiômicos, resultando num total de 839764 coeficientes de correlação analisados para responder ao segundo objetivo específico. Desta análise, identificam-se 789 descritores com propriedades informativos e 620 descritores considerados não-informativos, conforme ilustrado na Figura 4.2.

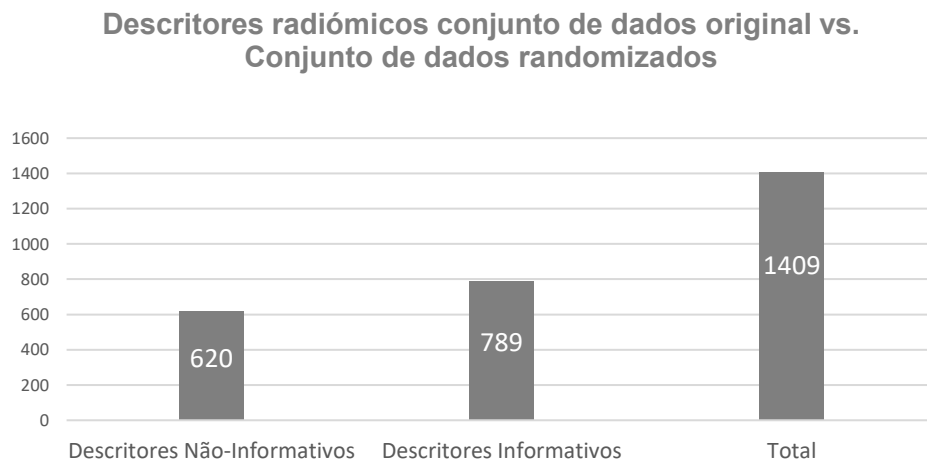


Figura 4.2-Frequência absoluta de descritores informativos e não-informativos resultante da análise do coeficiente de correlação intraclasse entre conjunto de dados original vs. randomizados.

De seguida, é apresentado detalhadamente os descritores radiômicos não-informativos entre os conjuntos, tanto na imagem original como na imagem derivada de filtros. De acordo com os dados obtidos a classe primeira-ordem revelou-se não-informativa nos seguintes descritores: *10Percentile*, *90Percentile*, *Energy*, *Entropy*, *InterquartileRange*, *Maximum*, *Mean*, *MeanAbsoluteDeviation*, *Minimum*, *Range*, *RobustMeanAbsoluteDeviation*, *RootMeanSquared*, *TotalEnergy*, *Uniformity* e *Variance* totalizando 122 descritores não-informativos, correspondendo a uma proporção na classe de 15 em 18 descritores.

No que se refere aos resultados obtidos para os subtipos de imagem, foram observados padrões particulares, demonstrados de seguida pelo Quadro 4.9. Adicionalmente, disponibiliza-se na Figura E.1 em apêndice E um gráfico de barras, de modo a evidenciar

as correlações entre os conjuntos de dados originais e randomizados nos descritores de propriedades de primeira-ordem.

Quadro 4.9-Descritores de propriedade primeira-ordem não-informativos, determinados pelo ICC.

Filtros	Descritores de propriedade de primeira-ordem
Original	<i>90Percentile, Energy, InterquartileRange, MeanAbsoluteDeviation, Range, RobustMeanAbsoluteDeviation, TotalEnergy, Uniformity, Variance</i>
LoG ($\sigma = 0,6$)	<i>10Percentile, 90Percentile, Energy, InterquartileRange, MeanAbsoluteDeviation, Range, RobustMeanAbsoluteDeviation, TotalEnergy, Uniformity, Variance</i>
LoG ($\sigma = 2$)	<i>Energy, Entropy, TotalEnergy, Uniformity</i>
LoG ($\sigma = 3$)	<i>Energy, TotalEnergy</i>
Wavelet	<i>10Percentile, 90Percentile, Energy, InterquartileRange, Maximum, MeanAbsoluteDeviation, Minimum, Range, RobustMeanAbsoluteDeviation, RootMeanSquared, TotalEnergy, Uniformity, Variance</i>
Square	<i>90Percentile, Energy, InterquartileRange, MeanAbsoluteDeviation, Mean, RobustMeanAbsoluteDeviation, RootMeanSquared, TotalEnergy, Variance</i>
Square-root	<i>Entropy, InterquartileRange, MeanAbsoluteDeviation,</i>
Logaritmo	<i>RobustMeanAbsoluteDeviation, Uniformity, Variance</i>
Exponencial	<i>Energy, TotalEnergy</i>

No que se refere aos descritores de forma, todos os atributos incluídos demonstraram elevada concordância, conforme previamente esperado conceptualmente.

Relativamente à matriz GLCM, foi possível identificar o conjunto de descritores não-informativos: *Autocorrelation, ClusterProminence, ClusterTendency, Contrast, Correlation, DifferenceAverage, DifferenceEntropy, DifferenceVariance, Id, Idm, Idmn, Idn, Imc1, Imc2, InverseVariance, JointAverage, JointEnergy, JointEntropy, MaximumProbability, SumAverage, SumEntropy e SumSquares* totalizando 173 descritores de propriedades não-informativos, representando uma proporção na classe de 22 em 24 descritores.

Para os diferentes subtipos de imagem, foram determinadas as seguintes observações, apresentadas no Quadro 4.10. Adicionalmente, disponibiliza-se na Figura E.1 em apêndice E um gráfico de barras, de modo a evidenciar as correlações entre os conjuntos de dados originais e randomizados nos descritores de propriedades GLCM.

Quadro 4.10-Descritores de propriedade textura classe GLCM não-informativos, determinados pelo ICC.

Filtro	Descritores de propriedade de GLCM
Original	Autocorrelation, ClusterProminence, ClusterTendency, Contrast, Correlation, DifferenceAverage, DifferenceVariance, Id, Idm, InverseVariance, JointAverage, JointEnergy, JointEntropy, MaximumProbability, SumAverage, SumEntropy, SumSquares
LoG ($\sigma = 0,6$)	ClusterProminence, ClusterTendency, Contrast, Correlation, DifferenceAverage, DifferenceVariance, Id, Idm, InverseVariance, JointEntropy, JointEntropy, MaximumProbability, SumSquares
LoG ($\sigma = 2$)	Idmn, Idn, Imc1, JointEnergy, JointEntropy, MaximumProbability, SumEntropy
LoG ($\sigma = 3$)	Id, Idmn, Idn, JointEnergy, JointEntropy, DifferenceEntropy
Wavelet	Autocorrelation, ClusterProminence, ClusterTendency, Contrast, Correlation, DifferenceAverage, DifferenceEntropy, DifferenceVariance, Id, Idm, InverseVariance, JointAverage, JointEnergy, JointEntropy, MaximumProbability, SumAverage, SumSquares, SumSquares
Square-root	ClusterProminence, ClusterTendency, Contrast, Correlation, DifferenceAverage, DifferenceVariance, Idmn, Idn, JointEnergy, JointEntropy, MaximumProbability, SumEntropy, SumSquares.
Logaritmo	Contrast, Correlation, DifferenceAverage, DifferenceVariance, Idmn, Idn, JointEnergy, JointEntropy, MaximumProbability, SumEntropy, SumSquares
Square	Contrast, Correlation, DifferenceAverage, DifferenceVariance, JointAverage, SumAverage, SumSquares
Exponencial	Correlation, DifferenceAverage, DifferenceEntropy, Id, InverseVariance, SumEntropy

Referente às matrizes GLDM, observou-se a presença de todas as classes como não-informativas, totalizando 90 descritores, correspondentes a uma proporção na classe de 14 em 14 descritores. Adicionalmente, as seguintes observações foram obtidas para os subtipos de imagem, conforme ilustrado no Quadro 4.11 Disponibiliza-se na Figura E.1 em apêndice E um gráfico de barras, de modo a evidenciar as correlações entre os conjuntos de dados originais e randomizados nos descritores de propriedades GLDM.

Quadro 4.11-Descritores de propriedade textura classe GLDM não-informativos, determinados pelo ICC.

Filtros	Descritores de propriedade de GLDM
Original	<i>DependenceNonUniformity, GrayLevelVariance, HighGrayLevelEmphasis, LargeDependenceHighGrayLevelEmphasis, SmallDependenceHighGrayLevelEmphasis, SmallDependenceLowGrayLevelEmphasis</i>
LoG ($\sigma = 0,6$)	DependenceNonUniformity, GrayLevelVariance, LargeDependenceLowGrayLevelEmphasis, LowGrayLevelEmphasis, SmallDependenceLowGrayLevelEmphasis
LoG ($\sigma = 2$)	DependenceEntropy, DependenceNonUniformity, GrayLevelNonUniformity, LowGrayLevelEmphasis, SmallDependenceLowGrayLevelEmphasis
LoG ($\sigma = 3$)	DependenceNonUniformity, LowGrayLevelEmphasis, SmallDependenceLowGrayLevelEmphasis
Wavelet	DependenceEntropy, DependenceNonUniformity, DependenceNonUniformityNormalized, DependenceVariance, GrayLevelVariance, HighGrayLevelEmphasis, LargeDependenceHighGrayLevelEmphasis, LowGrayLevelEmphasis, LargeDependenceLowGrayLevelEmphasis, SmallDependenceHighGrayLevelEmphasis, SmallDependenceLowGrayLevelEmphasis, SmallDependenceEmphasis, LargeDependenceEmphasis
Square	DependenceNonUniformity, GrayLevelVariance, HighGrayLevelEmphasis, LargeDependenceHighGrayLevelEmphasis, SmallDependenceHighGrayLevelEmphasis
Square-root	DependenceEntropy, DependenceNonUniformity, GrayLevelVariance, HighGrayLevelEmphasis, LargeDependenceHighGrayLevelEmphasis, SmallDependenceHighGrayLevelEmphasis, SmallDependenceLowGrayLevelEmphasis
Logaritmo	DependenceEntropy, DependenceNonUniformity, GrayLevelVariance, SmallDependenceLowGrayLevelEmphasis
Exponencial	SmallDependenceLowGrayLevelEmphasis

No âmbito da matriz GLRLM, verificou-se, de forma análoga às matrizes anteriores, a presença de todos os descritores como não-informativos, totalizando 98 descritores redundantes, equivalentes a uma proporção na classe de 16 em 16 descritores. Para cada subtipo de imagem, as observações determinadas estão apresentadas no Quadro 4.12. Adicionalmente, disponibiliza-se na Figura E.1 em apêndice E um gráfico de barras, de

modo a evidenciar as correlações entre os conjuntos de dados originais e randomizados nos descritores de propriedades GLRLM.

Quadro 4.12-Descritores de propriedade textura classe GLRLM não-informativos, determinados pelo ICC.

Filtros	Descritores de propriedade de GLRLM
Original	GrayLevelNonUniformityNormalized, GrayLevelVariance, HighGrayLevelRunEmphasis, LongRunHighGrayLevelEmphasis, RunLengthNonUniformity, ShortRunHighGrayLevelEmphasis
LoG ($\sigma = 0,6$)	GrayLevelNonUniformityNormalized, GrayLevelVariance, LongRunLowGrayLevelEmphasis, LowGrayLevelRunEmphasis, RunLengthNonUniformity, ShortRunLowGrayLevelEmphasis
LoG ($\sigma = 2$)	GrayLevelNonUniformity, GrayLevelNonUniformityNormalized, LongRunLowGrayLevelEmphasis, LowGrayLevelRunEmphasis, RunEntropy, RunLengthNonUniformity, ShortRunLowGrayLevelEmphasis
LoG ($\sigma = 3$)	LowGrayLevelRunEmphasis, RunLengthNonUniformity, ShortRunLowGrayLevelEmphasis
Wavelet	GrayLevelNonUniformityNormalized, GrayLevelVariance, HighGrayLevelRunEmphasis, LongRunEmphasis, LongRunHighGrayLevelEmphasis, LongRunLowGrayLevelEmphasis, RunEntropy, LowGrayLevelRunEmphasis, RunLengthNonUniformity, RunLengthNonUniformityNormalized, RunPercentage, RunVariance ShortRunHighGrayLevelEmphasis, ShortRunLowGrayLevelEmphasis, ShortRunEmphasis
Square	GrayLevelVariance, HighGrayLevelRunEmphasis, LongRunHighGrayLevelEmphasis, RunLengthNonUniformity, ShortRunHighGrayLevelEmphasis
Squere-root	GrayLevelNonUniformityNormalized, GrayLevelVariance, HighGrayLevelRunEmphasis, LongRunHighGrayLevelEmphasis, RunEntropy, RunLengthNonUniformity, ShortRunHighGrayLevelEmphasis
Logaritmo	GrayLevelNonUniformityNormalized, GrayLevelVariance, RunEntropy, RunLengthNonUniformity

Em relação à matriz GLSZM, de forma semelhante às matrizes GLDM e GLRLM, foram identificadas todas as classes como não-informativas, totalizando 106 descritores numa proporcionalidade na classe de 16 em 16 descritoras. Os subtipos de imagem deram origem às seguintes observações, organizadas no Quadro 4.13. Adicionalmente, disponibiliza-se na Figura E.1 em apêndice E um gráfico de barras, de modo a evidenciar as correlações

entre os conjuntos de dados originais e randomizados nos descritores de propriedades GLSZM.

Quadro 4.13-Descritores de propriedade textura classe GLSZM não-informativos, determinados pelo ICC.

Filtros	Descritores de propriedade de GLSZM	
Original	<i>GrayLevelNonUniformityNormalized,</i> <i>HighGrayLevelZoneEmphasis,</i> <i>LowGrayLevelZoneEmphasis,</i> <i>SmallAreaHighGrayLevelEmphasis, SmallAreaLowGrayLevelEmphasis</i>	<i>GrayLevelVariance,</i> <i>LargeAreaHighGrayLevelEmphasis,</i> <i>SizeZoneNonUniformity,</i>
LoG ($\sigma = 0,6$)	<i>GrayLevelNonUniformityNormalized,</i> <i>LargeAreaLowGrayLevelEmphasis,</i> <i>SizeZoneNonUniformity, SmallAreaLowGrayLevelEmphasis</i>	<i>GrayLevelVariance,</i> <i>LowGrayLevelZoneEmphasis,</i>
LoG ($\sigma = 2$)	<i>GrayLevelNonUniformity,</i> <i>LargeAreaLowGrayLevelEmphasis,</i> <i>SizeZoneNonUniformity, SmallAreaLowGrayLevelEmphasis, ZoneEntropy</i>	<i>GrayLevelNonUniformityNormalized,</i> <i>LowGrayLevelZoneEmphasis,</i>
LoG ($\sigma = 3$)	<i>LowGrayLevelZoneEmphasis,</i> <i>SmallAreaLowGrayLevelEmphasis</i>	<i>SizeZoneNonUniformity,</i>
Wavelet	<i>GrayLevelNonUniformityNormalized,</i> <i>HighGrayLevelZoneEmphasis,</i> <i>LargeAreaLowGrayLevelEmphasis,</i> <i>SizeZoneNonUniformity,</i> <i>SmallAreaLowGrayLevelEmphasis,</i> <i>SizeZoneNonUniformityNormalized, SmallAreaEmphasis, ZonePercentage, ZoneVariance, ZoneEntropy</i>	<i>GrayLevelVariance,</i> <i>LargeAreaHighGrayLevelEmphasis,</i> <i>LowGrayLevelZoneEmphasis,</i> <i>SmallAreaHighGrayLevelEmphasis,</i> <i>LargeAreaEmphasis,</i>
Square	<i>GrayLevelVariance,</i> <i>LargeAreaHighGrayLevelEmphasis,</i> <i>SmallAreaHighGrayLevelEmphasis</i>	<i>HighGrayLevelZoneEmphasis,</i> <i>SizeZoneNonUniformity,</i>
Square-root	<i>GrayLevelNonUniformityNormalized,</i> <i>HighGrayLevelZoneEmphasis,</i> <i>LowGrayLevelZoneEmphasis,</i> <i>SmallAreaHighGrayLevelEmphasis,</i> <i>ZoneEntropy</i>	<i>GrayLevelVariance,</i> <i>LargeAreaHighGrayLevelEmphasis,</i> <i>SizeZoneNonUniformity,</i> <i>SmallAreaLowGrayLevelEmphasis,</i>
Logaritmo	<i>SizeZoneNonUniformity, LowGrayLevelZoneEmphasis,</i> <i>GrayLevelNonUniformityNormalized,</i> <i>SmallAreaLowGrayLevelEmphasis</i>	<i>GrayLevelVariance,</i> <i>ZoneEntropy,</i>
Exponencial	<i>GrayLevelNonUniformity, SizeZoneNonUniformityNormalize</i>	

No que se refere às matrizes NGTDM, foram identificadas duas classes não- informativas, designadamente *Busyness* e *Strength*, totalizando 12 descritores, o que corresponde a uma proporcionalidade na classe de 2 em 5 descritores. As seguintes observações foram obtidas para os subtipos de imagem, conforme apresentado no Quadro 4.14. Adicionalmente, disponibiliza-se na Figura E.1 em apêndice E um gráfico de barras, de modo a evidenciar as correlações entre os conjuntos de dados originais e randomizados nos descritores de propriedades NGTDM.

Quadro 4.14-Descritores de propriedade textura classe NGTDM não-informativos, determinados pelo ICC.

Filtros	Descritores de propriedade de NGTDM
Original	
LoG ($\sigma = 2$)	<i>Busyness</i>
Logaritmo	
Square-root	
Wavelet	<i>Busyness, Strength</i>

Com base nos resultados, constatou-se que as classes baseadas em primeira-ordem e NGTDM preservaram maior poder informativo, com 3 descritores informativos e por fim a classe GLCM com 2. De seguida, são apresentados os descritores radiômicos informativos organizados por classes no Quadro 4.15.

Quadro 4.15- Lista de descritores de propriedades informativos resultantes da análise estatística ICC para as várias classes descritoras tanto para a imagem original como para a imagem derivada de filtros.

Primeira-ordem	GLCM	GLRLM	GLSZM	NGTDM	GLDM
<i>Kurtosis</i>	<i>ClusterShade</i>			<i>Coarseness</i>	
<i>Median</i>	<i>MCC</i>			<i>Complexity</i>	
<i>Skewners</i>				<i>Contrast</i>	

4.1.3 Combinação dos Conjuntos: Operação de Interseção e União

A avaliação realizada sobre os métodos de combinação de conjuntos baseados nos operadores de interseção e união, para responder ao terceiro objetivo específico, aplicados aos conjuntos resultantes das análises estatísticas multivariadas, revelou que a operação de união ($A \cup B$) totalizou 1154 descritores de propriedades informativos, enquanto, a operação de interseção ($A \cap B$) resultou em 650.

Os resultados obtidos evidenciam diferenças no número de descritores identificados. Nesse sentido, destaca-se a interseção como a metodologia de combinação mais adequada para a seleção de descritores, uma vez que possibilita a obtenção de conjuntos de dados mais consistentes e robustos, o que decorre do facto de serem incluídos apenas os descritores comuns entre os conjuntos, conforme ilustrado pelo diagrama de *Venn* apresentado na Figura 4.3.

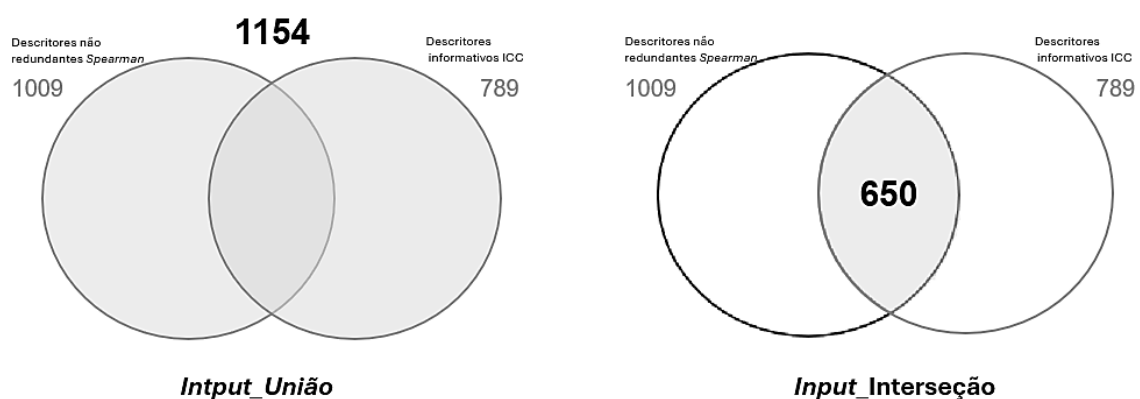


Figura 4.3- Diagrama de *Venn* com ilustração dos conjuntos operados sobre união e interseção.

4.1.4 Análise Univariada Radiômica

Com o objetivo de determinar descritores radiômicos significativamente associados com a variável a prever, procedeu-se à análise univariada. Como resultado desta abordagem, foram determinados 337 descritores de propriedades radiômicos com capacidade de classificação moderada a elevada com o desfecho clínico compreendendo valores de área sob a curva de 0,578 a 0,633 como é possível observar na Figura 4.4, onde evidencia uma relação entre o valor-p e a AUC, quanto maior a significância estatística (considerando limiar $\alpha = 0,05$) do valor-p maior tendência para valores de AUC superiores.

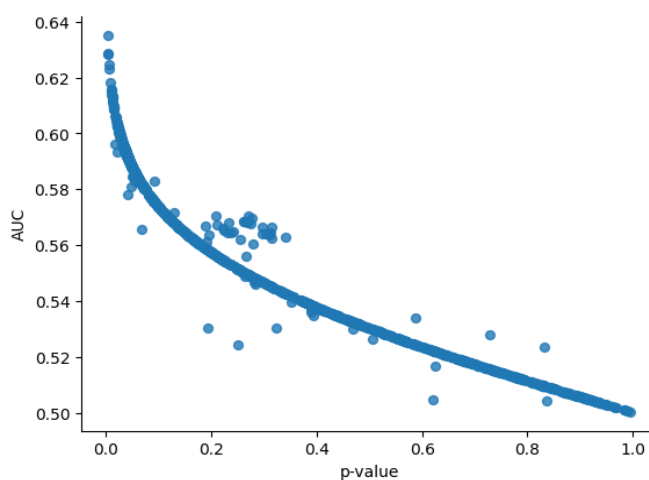


Figura 4.4- Gráfico de dispersão dos valores-p e a AUC no conjunto de dados originais.

Contudo, após a correção de comparações múltiplas foram determinados 0 descritores radiômicos com valor de significância estatística $< 0,05$, ou seja, isoladamente nenhum descritor possui um poder preditivo moderado ou elevado com o desfecho clínico. Ainda assim, independente de não ser significativamente estatístico foi observado o desempenho do descritor *wavelet2-LL_firstorder_Minimum* com o valor-p inicial de 0,002, AUC 0,635, sensibilidade 0,600, especificidade 0,600 e valor-p corrigido de 1, estabelecido como o descritor mais significativo entre o conjunto de descritores radiômicos dos dados originais. Como resultado o poder de classificação de desfecho clínico revelou-se fraco e redundante, contrariando a avaliação realizada inicial, demonstrada pela Figura 4.5.

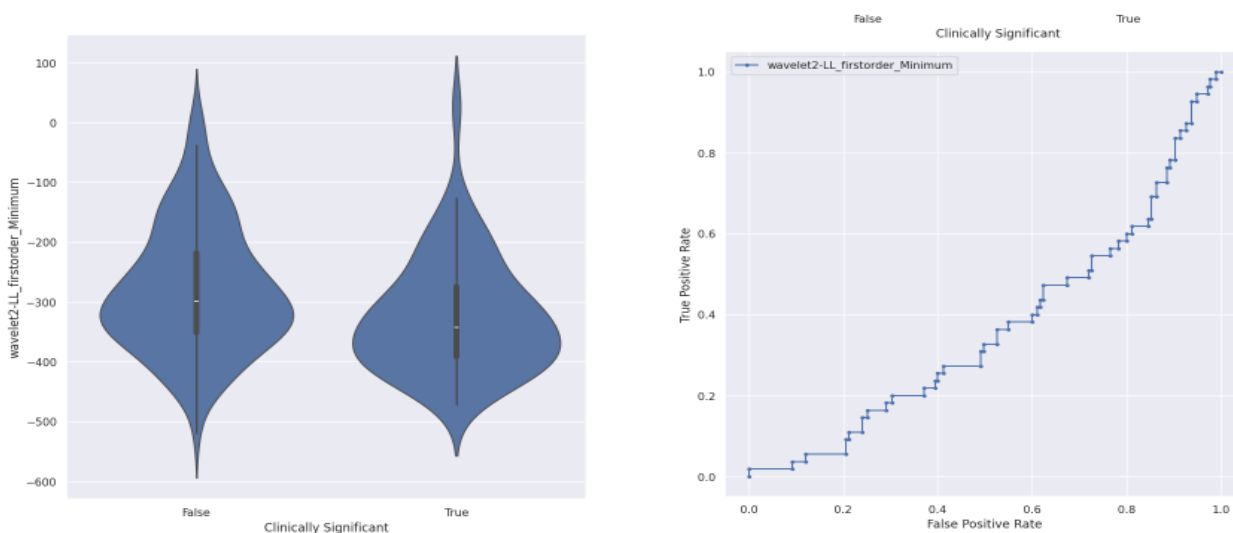


Figura 4.5- O gráfico *Violin Plot* e área sob a curva ROC ilustrativo do descritor, *wavelet2-LL_firstorder_Minimum* demonstra uma distribuição semelhante entre as classes, o que indica poder preditivo baixo para a classificação isoladamente. Adicionalmente, a área sob a curva ROC sugere um poder preditivo fraco ilustrado pelo traçado da curva sugestivo de uma classificação quase aleatória.

Em consequência, do critério de seleção de *input* de descritores informativos e preditivos não ter sido satisfeito, optou-se pela seleção de um descritor radiômico em cada 15 sujeitos, utilizando a área sob a curva com valores superiores, o que resultou nos seguintes descritores apresentados no Quadro 4.16.

Quadro 4.16-Descritores radiômicos selecionados através do valor superior de área sob a curva ROC.

Subamostra de Descritores de Propriedades	AUC
<i>Wavelet2-LL_firstorder_Minimum</i>	0,635
<i>Squareroot_glcm_ClusterShade</i>	0,628
<i>Logarithm_glcm_ClusterShade</i>	0,628
<i>Wavelet2-LH_glcm_MCC</i>	0,624
<i>Wavelet-HH_glcm_ClusterShade</i>	0,622
<i>Wavelet2-LH_glcm_Imc2</i>	0,618
<i>Exponential_gldm_GrayLevelNonUniformity</i>	0,615
<i>Log-sigma-3-mm-gldm_GrayLevelNonUniformity</i>	0,615
<i>Log-sigma-3-mm-glcm_Imc1</i>	0,615
<i>Log-sigma-3-mm-glrlm_GrayLevelNonUniformity</i>	0,613
<i>logarithm_firstorder_Skewness</i>	0,613
<i>Log-sigma-3-mm-glcm_Correlation</i>	0,613
<i>Squareroot_firstorder_Skewness</i>	0,613
<i>Exponential_glrlm_GrayLevelNonUniformity</i>	0,612
<i>Log-sigma-3-mm-glszm_GrayLevelNonUniformity</i>	0,611

4.2 Modelos de *Machine Learning*

Para avaliar o impacto de diferentes métodos de seleção, foram desenvolvidos e avaliados vários modelos de *machine learning*. De seguida, são apresentados de forma, sequencial os diferentes *inputs*:

4.2.1 Modelo I: Conjunto de dados Originais

Com base no conjunto de dados original, foram comparados distintos algoritmos de *machine learning*, na sequência da seleção do melhor modelo. O classificador *random forest* destacou-se, tendo evidenciado um melhor desempenho médio na validação cruzada aninhada com valor médio de 0,142 e desvio-padrão 0,203. Embora o desvio-padrão revele alguma instabilidade entre *folds*, o desempenho médio relativo superou o da regressão logística com valor médio 0,104 e *Support Vector Machine* 0,075.

Foi definida uma divisão em 5 *folds* baseada na validação cruzada, utilizando os seguintes hiperparâmetros otimizados: profundidade máxima 1 e número de estimadores 1024. A validação cruzada permitiu observar o desempenho médio do modelo ao longo dos diferentes *folds* e após o treino o modelo foi avaliado num conjunto independente. De seguida, as respetivas métricas de desempenho são apresentadas no Quadro 4.17.

Quadro 4.17- Métricas de desempenho de treino e teste através do modelo de *machine learning random forest* em 5 *folds* desenvolvido através do conjunto de dados originais.

Métricas de Desempenho do Modelo I		
	Treino	Teste
AUC (Área sob a curva ROC)	0,616 ± 0,048	0,644
MCC (Coef. Correlação <i> Matthews </i>)	0,182 ± 0,075	0,297
<i> Kappa de Cohen </i>	0,175 ± 0,098	0,297
<i> F1-score </i>	0,416 ± 0,056	0,500
Acurácia	0,761 ± 0,031	0,710
Sensibilidade	0,527 ± 0,036	0,476
Especificidade	0,966 ± 0,042	0,812
PPV (Valor Preditivo Positivo)	0,540 ± 0,408	0,526
NPV (Valor Preditivo Negativo)	0,820 ± 0,019	0,780
Acurácia Balanceada	0,601 ± 0,038	0,644

4.2.2 Modelo II: *Input_* Interseção

Com base no conjunto de dados da correlação de *Spearman* e ICC intersetados, foram comparados distintos algoritmos de *machine learning* , na sequência da seleção do melhor modelo. O classificador *random forest* destacou-se, tendo resultado em um melhor desempenho em relação aos restantes classificadores, com valor médio na validação cruzada aninhada de 0,073 e desvio-padrão 0,153. Embora, o desvio-padrão apresente maior instabilidade em relação aos restantes classificadores, o desempenho geral ainda assim é superior.

A nível de arquitetura foi definido uma divisão em 5 *folds* recorrendo à validação cruzada utilizando os seguintes hiperparâmetros otimizados: profundidade máxima 4 e número de estimadores 2. De seguida, as respetivas métricas de desempenho de treino e teste são apresentadas abaixo apresentadas no Quadro 4.18.

Quadro 4.18-Métricas de desempenho de treino e teste através do modelo de *machine learning random forest* em 5 *folds* desenvolvido através do conjunto de dados intersetados.

Métricas de Desempenho do Modelo II		
	Treino	Teste
AUC (Área sob a curva ROC)	0,613 ± 0,070	0,674
MCC (Coef. Correlação <i>Matthews</i>)	0,160 ± 0,125	0,331
<i>Kappa de Cohen</i>	0,155± 0,121	0,326
<i>F1-score</i>	0,403 ± 0,039	0,553
Acurácia	0,752 ± 0,040	0,696
Sensibilidade	0,709 ± 0,068	0,619
Especificidade	0,971 ± 0,036	0,729
PPV (Valor Preditivo Positivo)	0,640± 0,445	0,500
NPV (Valor Preditivo Negativo)	0,817± 0,043	0,814
Acurácia Balanceada	0,580 ± 0,063	0,674

4.2.3 Modelo III: *Input_União*

Com base no conjunto de dados da correlação de *Spearman* e ICC operados sobre união, foram comparados distintos algoritmos de *machine learning*, com intuito de determinar o melhor modelo classificador. A regressão logística destacou-se, tendo resultando em um melhor desempenho em relação aos restantes classificadores, com valor médio na validação cruzada aninhada de 0,077 e desvio-padrão 0,038. Embora, o modelo de classificação *random forest* também apresentasse um desempenho competitivo com valor médio 0,064 porém, apresentava maior variação entre *folds* de valor 0,139.

A nível dos parâmetros definidos no modelo foram divididos em 5 *folds* recorrendo à validação cruzada utilizando os seguintes hiperparâmetros otimizados: regularização L1 e inverso da força de regularização 10. De seguida, as respetivas métricas de desempenho do treino e teste são apresentadas no Quadro 4.19.

Quadro 4.19-Métricas de desempenho de treino e teste através do modelo de *machine learning* regressão logística em 5 *folds* desenvolvido através do conjunto de união.

Modelo III		
	Treino	Teste
AUC (Área sob a curva ROC)	0,576 ± 0,058	0,676
MCC (Coef. Correlação <i>Matthews</i>)	0,106 ± 0,113	0,326
<i>Kappa de Cohen</i>	0,092± 0,101	0,283
<i>F1-score</i>	0,386± 0,000	0,567
Acurácia	0,561 ± 0,052	0,623
Sensibilidade	1,000 ± 0,000	0,810
Especificidade	0,560 ± 0,062	0,542
PPV (Valor Preditivo Positivo)	0,286± 0,053	0,436
NPV (Valor Preditivo Negativo)	0,804± 0,044	0,867
Acurácia Balanceada	0,562 ± 0,066	0,676

4.2.4 Modelo IV: *Input* Descritores Informativos e Preditores

Com base no conjunto de dados originais com aplicação da análise estatística univariada radiômica, foram comparados diferentes algoritmos de construção de modelos de *machine learning*, com o objetivo de selecionar o modelo a desenvolver com os respetivos dados. O modelo *Support Vector Machine* destacou-se, tendo resultado num melhor desempenho em relação aos restantes classificadores. Com valor médio na validação cruzada aninhada de 0,250 e desvio-padrão 0,117. Embora, tenha apresentado maior valor de média, o desvio-padrão foi relativamente alto o que indica variabilidade entre *folds*. A nível de parâmetros utilizados, o presente classificador recorreu à regularização de 100.0, *Gamma* = 0,01 e *kernel RBF*. Para a construção do modelo, os dados foram divididos em 5 *folds* recorrendo à validação cruzada. De seguida, as respetivas métricas de desempenho do treino e teste são apresentadas abaixo no Quadro 4.20.

Quadro 4.20-Métricas de desempenho de treino e teste através do modelo de *machine learning Support Vector Machine* em 5 *folds* desenvolvido através do conjunto de dados originais.

MODELO IV		
	Treino	Teste
AUC (Área sob a curva ROC)	0,641 ± 0,118	0,533
MCC (Coef. Correlação <i>Matthews</i>)	0,283± 0,045	0,086
<i>Kappa de Cohen</i>	0,243± 0,043	0,077
<i>F1-score</i>	0,349± 0,056	0,258
Acurácia	0,778 ± 0,016	0,667
Sensibilidade	0,327± 0,093	0,190
Especificidade	1,000± 0,000	0,875
PPV (Valor Preditivo Positivo)	0,617± 0,100	0,400
NPV (Valor Preditivo Negativo)	0,801± 0,010	0,712
Acurácia Balanceada	0,599± 0,021	0,533

Para tornar a representação da informação mais acessível a Figura 4.6 reúne o desempenho das métricas entre os modelos I, II, III, IV.

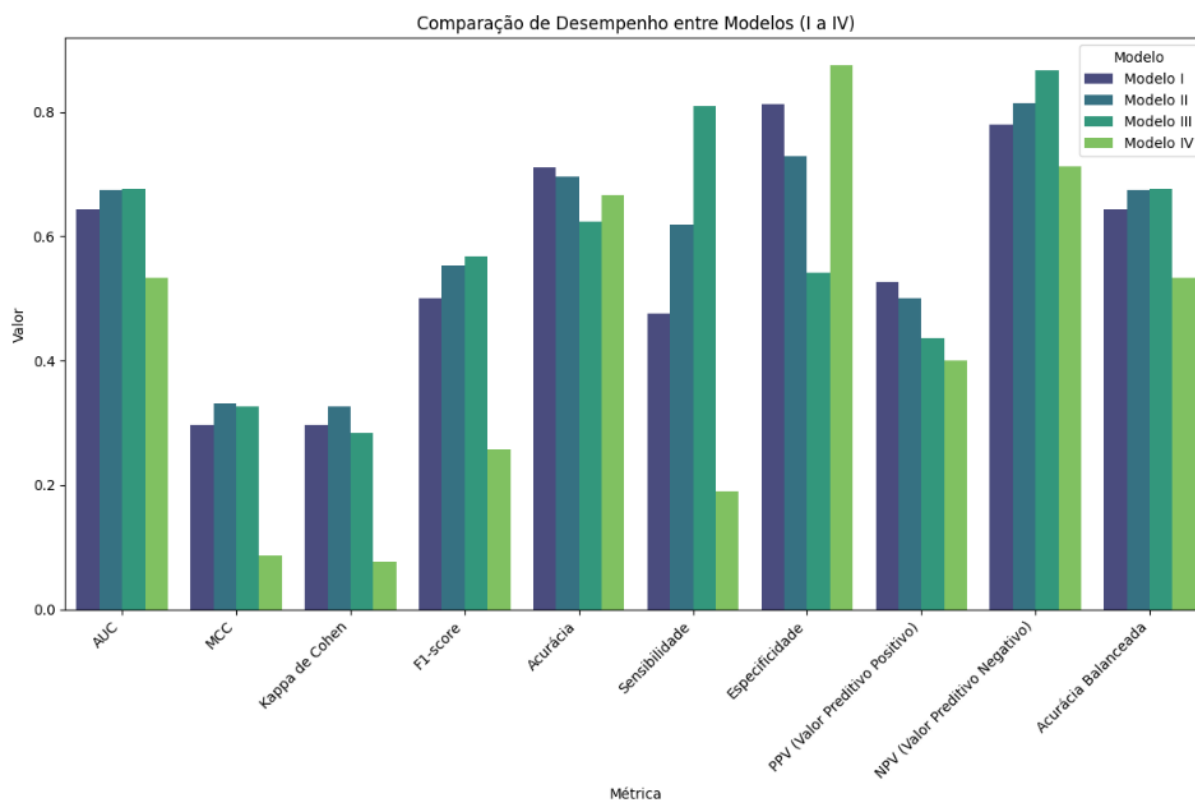


Figura 4.6-Representação dos resultados do conjunto de teste entre os Modelos I, II, III, IV.

Com base nas métricas de desempenho dos modelos, identificou-se o modelo II como o mais razoável em termos de equilíbrio entre métricas globais de desempenho com principal destaque para, *MCC*, *Kappa de Cohen*, acurácia, PPV e especificidade.

Desta forma, foram determinados neste modelo os descritores com maior contributo para o desfecho clínico. A Figura 4.7 apresenta os resultados, entre os quais se destacam os descritores radiómicos: *exponential_glrIm_ShortRunEmphasis*, *wavelet2-LL_firstorder_Skewness*, *logarithm_firstorder_RootMeanSquared*, *exponential_glcm_Imc1*, *logarithm_firstorder_Mean*, *exponential_ngtdm_complexity*, *log-sigma-2-mm-glcm-Imc2*, *wavelet2-HL-glszm-SmallAreaHighGrayLevelEmphasis* e *wavelet-HL-gldm-DependenceVariance*.

Observou-se que valores elevados de *exponential_glrIm_ShortRunEmphasis*, *wavelet2-LL_firstorder_Skewness*, *exponential_glcm_Imc1*, *log-sigma-2-mm-glcm-Imc2* e *wavelet-HL-gldm-DependenceVariance* está associado a uma predição com valor 1 (lesão clinicamente significativa). Por outro lado, valores menores *logarithm_firstorder_RootMeanSquared*, *logarithm_firstorder_Mean*, *exponential_ngtdm_complexity* e *wavelet2-HL-glszm-SmallAreaHighGrayLevelEmphasis* influenciam o modelo para a predição com valor 1 (lesão clinicamente significativa).

Os restantes 638 descritores resultaram num impacto reduzido no modelo, reforçando que a predição clínica está concentrada em um conjunto de descritores relevantes.

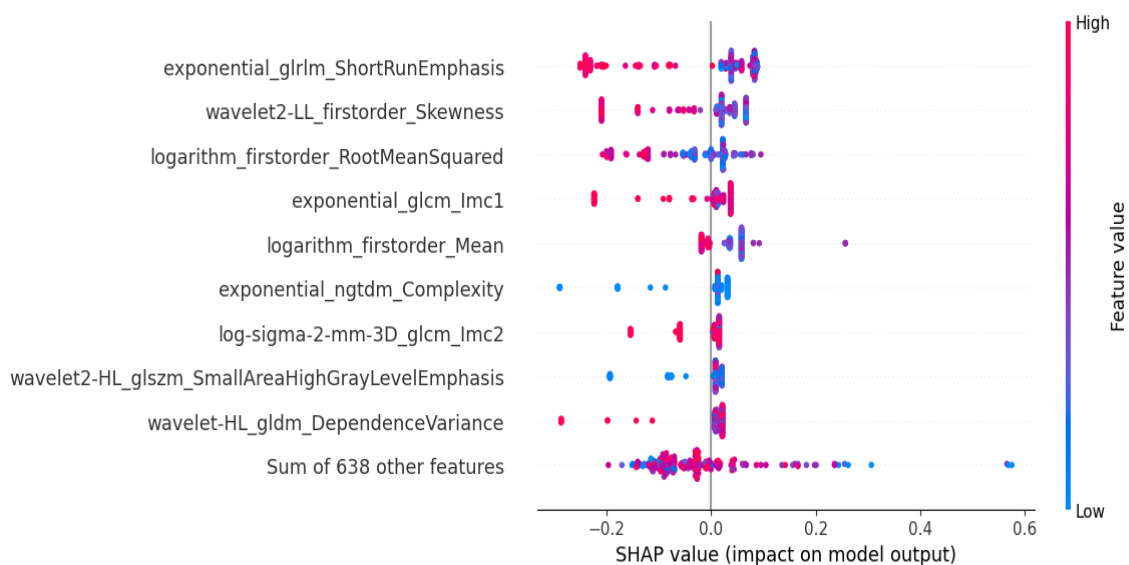


Figura 4.7-Gráfico Modelo II SHAP para o desfecho clínico significativo. Cada ponto representa uma amostra, em que a posição no eixo x indica o contributo da variável radiómica na probabilidade do desfecho clínico. A cor dos pontos reflete a intensidade da variável (azul: baixo) e (vermelho: alto). As variáveis estão ordenadas por importância média absoluta.

5 Discussão

Neste trabalho, verificou-se que os métodos baseados na análise estatística multivariada, como a correlação de *Spearman* e o ICC, são eficazes para a seleção de descritores radiômicos e redução de colinearidades. Esses resultados são corroborados pelo estudo de A. Perniciano et al. (29), que entre 45 estudos analisados, 8 deles basearam-se em testes estatísticos de *Spearman* ou ICC como metodologias de seleção.

A análise estatística pela correlação de *Spearman* eliminou 386 descritores de propriedades de primeira-ordem e textura, considerados redundantes por apresentarem moderada a elevada correlação com a propriedade de forma. O estudo desenvolvido por Rodrigues A. et al.(23) também incluiu esta metodologia de seleção, embora o tenha associado à validação cruzada através do *RapidMiner Studio*, baseado no coeficiente de correlação de *Pearson*. Por outro lado, o estudo de S. J. Doran et al. (78) também aplicou a eliminação de descritores radiômicos redundantes utilizando a correlação *Spearman*, por meio do algoritmo *findCorrelation* recorrendo à biblioteca *caret*, com limiar de 0,9. Com o objetivo de eliminar descritores redundantes entre descritores das classes e entre elas, priorizando descritores com interpretações mais simples e diretas pela seguinte ordem: forma, primeira-ordem e textura (78).

Em termos de aplicação, considera-se que o método implementado neste estudo é possivelmente mais simples e interpretável comparativamente aos métodos de (23,78) uma vez, que possui menor complexidade computacional e independência de ferramentas externas. Paralelamente, o estudo de Shur J. et al. (11) destacou a simplicidade desta análise estatística, mas ressaltou precauções em relação ao limiar aplicado, uma vez que este influencia diretamente o desempenho dos resultados.

Os resultados deste estudo permitiram identificar descritores redundantes por classes, salientando a correlação entre descritores de primeira-ordem e de textura com a propriedade de forma. Semelhantemente, foi determinado pelo estudo de Welch et al.(19), que provavelmente descritores de textura podem apenas refletir informações de volume, o que implica que adicionalmente outros descritores de forma possam englobar redundâncias (19). De forma paralela, o estudo de Bianchini L. et al.(20) verificou, em 930 descritores não relacionados à forma, em média 152, entre três observações realizadas, apresentaram elevada correlação com descritores de forma (20).

Por meio dos mapas de calor de correlação apresentados no Apêndice C, observa-se que descritores de forma, como *MeshVolume*, *MinorAxisLenght*, *SurfaceArea*, *SurfaceVolumeRatio* e *VoxelVolume*, apresentam elevada correlação com descritores de propriedades de primeira-ordem e de textura o que corrobora as observações realizadas por Welch et al. (19).

Na classe primeira-ordem identificaram-se os seguintes descritores redundantes *Energy*, *Entropy*, *TotalEnergy* e *Uniformity* determinando apenas quatorze descritores não redundantes. Adicionalmente, observou-se um padrão variável nas imagens derivadas de filtros, resultado em menor número de descritores redundantes para os filtros exponencial, *square* e *LoG* ($\sigma = 0,6, 2, 3$). O estudo de Welch et al. (19) apesar de utilizar uma metodologia distinta para análise de modelos de prognóstico, também verificou que os descritores *Energy* e *Gray Level Non-Uniformity*, em modelos de *Cox* não capturavam informações relevantes do fenótipo tumoral associado à propriedade de textura, mas sim ao volume tumoral, como fator prognóstico (19). Apesar da metodologia distinta, o descritor *Energy* observou-se entre os descritores redundantes mais frequentes na classe, o que reforça a sua possível redundância, seja devido ao próprio cálculo matemático ou pela interação com subprocessos do fluxo radiômico, como o pós-processamento de imagem.

Relativamente às matrizes de textura, a classe GLCM determinou o maior conjunto de descritores redundantes nomeadamente *DifferenceEntropy*, *Idmn*, *Idn*, *Imc1*, *Imc2*, *JointEnergy*, *JointEntropy*, *MCC*, *MaximumProbability* e *SumEntropy*, o que resultou em apenas quatorze descritores não redundantes. Complementarmente, evidenciou-se variações entre o número de descritores identificados redundantes para as imagens derivadas de filtros, entre eles destacam-se com menor número o filtro logaritmo e *LoG* ($\sigma = 0,6$). O estudo realizado por Çınarer G et al. (79), embora restrito à avaliação de correlações na classe GLCM, identificou correlações elevadas entre o descritor *Contrast* e *DifferenceEntropy*, *DifferenceAverage* e *DifferenceEntropy*, o que reforça a relevância em avaliar também correlações entre descritores da mesma classe (79).

Similarmente, o estudo Muftah A et al. (80) determina igualmente a correlação entre descritores da classe e entre classes nomeadamente, primeira-ordem e textura.

Referente à matriz GLDM verificou-se um número inferior de descritores redundantes, entre eles *DependenceEntropy*, *DependenceNonUniformity*, *GrayLevelNonUniformity*, *LargeDependenceLowGrayLevelEmphasis*, *LowGrayLevelEmphasis* e *SmallDependenceLowGrayLevelEmphasis*, resultando em apenas oito descritores não redundantes. De forma semelhante às matrizes anteriores, identificou-se inconsistência no número de descritores redundantes entre imagens derivadas de filtros, com destaque para os filtros exponencial e *square*, que preservaram maior número de descritores não redundantes.

Em relação à matriz GLRLM, constatou-se um conjunto ligeiramente superior de descritores redundantes em comparação com a classe GLDM e inferior à classe GLCM. Entre eles, destacam-se *GrayLevelNonUniformity*, *GrayLevelNonUniformityNormalized*, *LongRunLowGrayLevelEmphasis*, *LowGrayLevelRunEmphasis*, *RunEntropy*, *RunLengthNonUniformity* e *ShortRunLowGrayLevelEmphasis*, resultando em nove descritores não redundantes. À semelhança do observado nas classes anteriores, o número de descritores redundantes variou nas imagens derivadas de filtros, sendo mais conservador o filtro exponencial e *square*. Paralelamente ao observado, o estudo de Vallières M. et al.(22) identificou o descritor *GrayLevelNonUniformity*, extraído sobre a imagem original e derivada de filtro *wavelet*, entre os descritores fortemente correlacionados com o volume tumoral através do teste de *Pearson*, o que corroborou o mesmo achado deste estudo (22).

Em termos da matriz GLSZM identificaram-se os respetivos descritores redundantes *GrayLevelNonUniformity*, *GrayLevelNonUniformityNormalized*, *LargeAreaLowGrayLevelEmphasis*, *LowGrayLevelZoneEmphasis*, *SizeZoneNonUniformity*, *SmallAreaLowGrayLevelEmphasis* e *ZoneEntropy*, resultando em apenas nove descritores não redundantes. Também neste caso, verificou-se variações no padrão de redundância em imagens derivadas de filtros, com destaque para os filtros exponencial e *square*, que preservaram maior número de descritores não redundantes.

Para a matriz NGTDM verificou-se o seguinte conjunto de descritores redundantes *Busyness*, *Coarseness*, *Contrast* e *Strength*, resultando em apenas um descritor não redundante. Adicionalmente, nas imagens derivadas de filtros, os menores conjuntos de descritores redundantes foram identificados nos filtros exponencial, *square* e *square-root*.

De forma global, a análise dos resultados evidencia um impacto direto da aplicação de filtros na redundância de descritores radiômicos.

Em particular, as imagens derivadas de LoG ($\sigma = 2, 3$), *wavelet* e *square-root* apresentaram os maiores conjuntos de descritores redundantes, enquanto o filtro exponencial e *square* mostraram-se mais conservadores, preservando maior número de descritores não redundantes de primeira-ordem e textura.

Adicionalmente, resultante da análise estatística ICC foram determinados 620 descritores não-informativos demonstrando semelhantemente ao método anterior contribuir para a seleção de descritores informativos. De acordo, com os estudos de (11,19) a aplicação de metodologias baseadas em perturbação de dados e análise estatística é fundamental para identificar dependências subjacentes e assim, excluir descritores não-informativos. De forma semelhante, os resultados aqui apresentados reforçam a pertinência do método de seleção adotado.

Relativamente aos descritores de primeira-ordem identificaram-se os seguintes descritores não-informativos 10Percentile, 90Percentile, Energy, Entropy, InterquartileRange, Maximum, Mean, MeanAbsoluteDeviation, Minimum, Range, RobustMeanAbsoluteDeviation, RootMeanSquared, TotalEnergy, Uniformity, Variance, resultando em apenas três descritores informativos. Adicionalmente, foi observado um padrão variável entre as imagens derivadas de filtros, com menor número de descritores não-informativos para os filtros LoG ($\sigma = 3$), *wavelet2-LL* e exponencial.

Referente à classe textura a matriz GLCM, observou um maior número de descritores não-informativos entre os quais *Autocorrelation*, *ClusterProminence*, *ClusterTendency*, *Contrast*, *Correlation*, *DifferenceAverage*, *DifferenceEntropy*, *DifferenceVariance*, *Id*, *Idm*, *Idmn*, *Idn*, *Imc1*, *Imc2*, *InverseVariance*, *JointAverage*, *JointEnergy*, *JointEntropy*, *MaximumProbability*, *SumAverage*, *SumEntropy* e *SumSquares*, resultando em somente dois descritores informativos. De forma semelhante às observações anteriores, verificou-se variabilidade no padrão de descritores entre imagens derivadas de filtros, contudo, os filtros *square*, LoG ($\sigma = 2, 3$) e exponencial apresentaram o menor número de descritores não-informativos.

Em relação às matrizes GLDM, GLRLM e GLSZM, não foram identificados descritores informativos, contudo, observou-se uma menor evidencia de descritores não-informativos nas imagens derivadas de filtros. Especificamente, para GLDM e GLSZM, verificou-se um menor número designadamente no filtro LoG ($\sigma = 3$) e exponencial enquanto, a classe GLRLM determinou para os seguintes filtros LoG ($\sigma = 3$) e logaritmo.

Em termos da matriz NGTDM, verificou-se os seguintes descritores não-informativos *Busyness* e *Strength* resultando em três descritores informativos.

Além disso, não foram observadas diferenças significativas no número de descritores baseados na aplicação de filtros.

Perante os resultados, verificou-se, à semelhança dos obtidos com a correlação de *Spearman*, que a aplicação de filtros exponencial e LoG ($\sigma = 3$) resultam num menor número de classes identificadas como não-informativas. Por outro lado, a imagem original e filtros, LoG ($\sigma = 0,6, 2$), *walevet* e *square-root* apresentam padrões variáveis, sugerindo a necessidade de avaliação caso a caso.

Os resultados gerais da análise estatística multivariada reforçam a importância da seleção criteriosa de descritores, considerando as particularidades de cada tipo de imagem. Constatou-se que, para uma mesma classe, diferentes tipos de imagens podem produzir conjuntos de descritores redundantes e não-informativos, evidenciando um padrão variável. Os estudos de (11,21) apontam que a aplicação desta metodologia tende a resultar em conjuntos de descritores menos redundantes e mais interpretáveis. Paralelamente, Papanikolaou N. et al.(21) sugere a utilização de correlações visualizadas em mapas de calor como etapa integrada em processos mais complexos de seleção e redução de dimensionalidade. No presente estudo, são disponibilizados no Apêndice D os mapas de calor referentes à correlação de *Spearman*, no entanto, para o ICC, devido às limitações de representação gráfica de grandes volumes de dados, optou-se por uma abordagem alternativa, recorrendo a gráficos de barras disponíveis no Apêndice E.

Como principal resultado da análise anterior, determinou-se os seguintes descritores com baixa a moderada correlação com a propriedade de forma ou entre conjuntos originais e randomizados, nomeadamente:

- **Primeira-ordem:** *Median, Skewness, Kurtosis*
- **GLCM:** *Clustershade*
- **NGTDM** *Complexity*

Resultados compatíveis com os obtidos neste estudo já haviam sido relatados anteriormente na literatura, relativamente ao descritor de primeira-ordem *skewness* pelo estudo Adiraju R. et al. (81), documentado como um potencial biomarcador para caracterização tumoral, embora associado a limitações na capacidade de determinar interações não lineares mais complexas.

Relativamente ao método de avaliar diferentes formas de combinar os conjuntos de propriedades obtidos, observou-se um número distinto de descritores: 650 para o conjunto interseção e 1154 para o conjunto união.

Com base nos resultados, sugere-se que o conjunto de interseção constitui um método mais robusto e seletivo, em virtude da natureza metodológica da operação.

O estudo de Bianchini L et al.(20), baseado nos mesmos métodos estatísticos, com limiar de 0,9 para o ICC, aplicado sobre diferentes conjuntos de dados, provenientes de três equipamentos de MRI (1.5T *GE Healthcare* e 1.5T, 3T *Philips*) adquiridos em fantasmas, identificou 19 descritores radiômicos comuns correlacionados com a propriedade de forma, através da operação de interseção. Entre os descritores identificados destacaram-se: primeira-ordem (*Energy, TotalEnergy*), GLRLM (*RunLengthNonUniformity*), GLSZM (*GrayLevelNonUniformity*) e NGTDM (*Coarseness*) (20). Estes resultados corroboram os achados do presente estudo, sugerindo que as correlações observadas não decorrem apenas da origem do conjunto de dados, mas sim de uma propriedade intrínseca dos descritores radiômicos. O facto de o mesmo padrão surgir em análises retrospectivas e em avaliações de reprodutibilidade e repetibilidade em fantasmas, reforça que esses descritores mantêm um comportamento consistente independentemente da origem dos dados.

A análise estatística univariada realizada neste estudo identificou inicialmente 337 descritores radiômicos com valores de significância estatística compreendidos entre 0,002 e 0,049. Contudo, após a aplicação da correção baseada em métodos *Bonferroni*, nenhum descritor manteve significância estatística, resultando em valores-p corrigidos de 1,0 e áreas sob a curva ROC entre 0,578 e 0,633. Este resultado demonstra a necessidade de desenvolver abordagens metodológicas mais robustas para obtenção de dados confiáveis e clinicamente aplicáveis, como já apontado por Kumar V. et al (13). A metodologia de *Bonferroni* extremamente conservadora, embora minimize falsos positivos, pode possivelmente comprometer a seleção de predições moderados. Apesar de algumas variáveis terem apresentado um desempenho ligeiramente superior ao acaso em termos de AUC, o efeito estatístico foi demasiado pequeno para se manter após correções, o que reforça a ausência de evidencia estatística para a utilização como biomarcadores (13).

Foram igualmente abordados os métodos univariados por Wu W et al.(82), ao estudo realizado através de imagens de tomografia computadorizada de planeamento para avaliação de lesões pulmonares, resultando em 53 descritores radiômicos significativamente associados com a variável a prever, porém após aplicação de métodos de correção de comparações múltiplas, determinou-se uma predominância de descritores de propriedades derivados do filtro *wavelet*. A discrepância entre os resultados aqui observados e os obtidos no estudo de Wu W et al.(82) pode estar associada a diferentes fatores, incluindo a modalidade de imagem, a variabilidade de população, a anatomia em análise e crucialmente, o método de correção utilizado.

Enquanto o presente estudo recorre ao ajuste de *Bonferroni*, Wu W et al. (82) aplica procedimentos de *False Discovery Rate*, menos restritivos. Possivelmente a abordagem metodológica demonstra que a escolha de métodos de correção impacta diretamente a identificação de potenciais biomarcadores.

Adicionalmente, o mesmo autor salienta que metodologias baseadas exclusivamente na seleção de descritores significativamente estatísticos, pode incluir descritores radiômicos correlacionados, comprometendo desta forma, o desempenho e generalização dos modelos (82).

Outro ponto relevante por parte do autor Shur J et al. (11) é a averiguação de que dificilmente um único descritor radiômico apresenta desempenho isolado suficiente para ser considerado biomarcador clínico. Assim, análises univariadas assumem um papel referencial e comparativo, mas não substituem abordagens multivariadas, nas quais interações entre descritores podem integrar informações adicionais relevantes.

Apesar da ausência de significância estatística após correção, os resultados baseados nos valores superiores de AUC evidenciaram que descritores derivados de filtros *wavelet*, *wavelet2*, LoG ($\sigma=3$), exponencial, logaritmo e *square-root* concentraram os melhores desempenhos, ainda que moderados, obtendo valor máximo de 0,635 para o descritor *wavelet2-LL_firstorder_Minimum*. Destacou-se adicionalmente os descritores: primeira-ordem (*Minimum*, *Skewness*), *GLCM* (*ClusterShade*, *MCC*, *Imc2*, *Imc1*, *Correlation*), *GLDM* (*GrayLevelNonUniformity*), *GLRLM* (*GrayLevelNonUniformity*), *GLSZM* (*GrayLevelNonUniformity*). Estes achados reforçam a literatura, que demonstra que imagens derivadas de filtros de vários tipos, sigmas e níveis podem realçar padrões invisíveis na imagem original, como já descrito por A. Demircioğlu et al.(83), produzindo desempenho semelhante ou superior ao obtido com descritores extraídos exclusivamente da imagem original. Contudo, estas abordagens aumentam a dimensionalidade e a redundância dos dados, exigindo métodos de seleção rigorosos (83).

Relativamente, à escolha da área sob a curva ROC, como métrica de avaliação apesar de apropriada para estimar o desempenho preditivo, conforme apontado por Shur J. et al. (11), necessita da aplicação adicional de métodos de correção, como *Bonferroni* ou procedimentos de controlo da taxa de descobertas falsas (*Benjamini-Hochberg* e *Benjamini-Yekutieli*), para garantir robustez estatística.

Da análise comparativa, às métricas de avaliação de desempenho dos modelos, verificou-se que o modelo I apresentou desempenho razoável, com algumas métricas evidenciando capacidade limitada de generalização, embora com indicadores potencialmente interessantes em cenários específicos.

A AUC no conjunto de teste foi de 0,644, valor ligeiramente superior à média observada no treino ($0,616 \pm 0,048$), sugerindo uma capacidade razoável de discriminação entre desfechos clínicos, ainda que abaixo do valor ideal.

Adicionalmente, métricas como o MCC 0,297 e o *Kappa de Cohen* 0,297 reforçam a percepção de desempenho fraco do modelo, com previsões baixas e concordância razoáveis.

O *F1-score* 0,500 apresentou um valor equilibrado entre PPV e a sensibilidade, adicionalmente a acurácia também demonstrou um valor superior de 0,710, o que demonstrou que o modelo acertou em 71% das classificações, no entanto, o valor no conjunto de teste foi ligeiramente menor do que no treino ($0,761 \pm 0,031$).

A sensibilidade é inferior comparativamente tanto à especificidade como ao treino ($0,527 \pm 0,036$) e o teste (sensibilidade=0,476), o que sugere dificuldades na classificação da classe clinicamente significativo. Porém a especificidade, apesar de superior evidenciou um valor inferior relativamente ao treino ($0,966 \pm 0,042$) e teste (especificidade=0,812). Ainda assim, este valor continua superior à sensibilidade, indicando um potencial na classificação de lesões clinicamente não significativas. O que também é verificado, pela métrica, valor preditivo negativo 0,780. De modo, geral apesar do conjunto de dados ser ausente de métodos de seleção e apresentar alta dimensionalidade os resultados sugerem uma razoável capacidade de classificação de lesões não clinicamente significativa, apresentando limitações relevantes na classificação de lesões positivas.

Comparativamente aos modelos com métodos de seleção o modelo II apresentou desempenho geral superior ao modelo I, especialmente em métricas tais como: AUC (Modelo I = 0,644) para 0,674, o que indica uma melhor capacidade de classificação e equilíbrio entre sensibilidade e especificidade refletida pela acurácia balanceada 0,674 e *F1-score* 0,553.

De modo notável, a sensibilidade também aumentou (Modelo I = 0,476) para 0,619, indicando uma melhor capacidade de distinção de lesões clinicamente significativa. Porém a especificidade evidenciou uma diminuição (Modelo I = 0,812) para 0,729.

As métricas MCC e *Kappa de Cohen* também resultaram em valores superiores 0,331 e 0,326, reforçando a previsão e concordância moderada. Em contrapartida, a acurácia verificou uma diminuição de (Modelo I = 0,710) para 0,696, o que pode ser influenciado pelo aumento da sensibilidade e a diminuição da especificidade, o que demonstra um maior equilíbrio entre erros do tipo I e II. Os valores de PPV e NPV também permitem demonstrar uma melhoria nos valores NPV enquanto, os valores de PPV diminuem cerca de 0,026.

Em síntese, o recurso a métodos de seleção baseados na operação de interseção resultou num modelo, mais balanceado e robusto especialmente em métricas de concordância, mesmo sujeito a resultados inferiores a nível de especificidade e acurácia.

De seguida, comparativamente ao modelo I e II, o presente modelo III apresentou um comportamento distinto, com foco especialmente na sensibilidade. No conjunto de treino, a sensibilidade alcançou o valor de 1,000, o que indica que o modelo classificou todos os dados positivos corretamente. No entanto, o desempenho no conjunto de teste diminuiu para (sensibilidade = 0,810). Apesar da excelente sensibilidade, a especificidade teve uma diminuição substancial 0,542, o menor valor entre os modelos anteriores (Modelo I = 0,812) e (Modelo II = 0,729). O que indica uma dificuldade na classificação de lesões clinicamente não significativas. Consequentemente, o valor preditivo positivo foi o mais baixo entre os modelos 0,436, o que significa que apenas 43,6% das previsões positivas estavam corretas. Por outro lado, o valor preditivo negativo foi o mais alto 0,867 coerente com a elevada sensibilidade.

Em termos de *F1-score*, o modelo obteve melhores resultados entre os modelos o que evidencia uma otimização entre a PPV e a sensibilidade. A nível da acurácia balanceada também foi superior 0,676, à semelhança com o modelo II.

Métricas de concordância e previsão, como MCC e *Kappa de Cohen*, mantiveram-se dentro de valores similares aos observados no modelo II. Contudo, a nível da AUC 0,676 foi a mais alta entre os modelos, sugerindo que o modelo é superior a distinguir as classes.

Da avaliação dos resultados anteriores, diferentes potenciais foram evidenciados entre a ausência de métodos de seleção e seleção baseada na análise estatística multivariada e operações entre conjuntos. O modelo I apresentou a maior especificidade 0,812 entre os modelos, contudo, a nível da sensibilidade foi verificado um diminuto desempenho 0,476. Adicionalmente, a AUC foi a menor entre os modelos II e III 0,644 enquanto, os restantes modelos II e III respetivamente obtiveram 0,674 para a interseção e 0,676 para a união. Adicionalmente, não foram observadas discrepâncias fortemente significativas entre os valores médios e desvios-padrões entre os conjuntos de treino e teste. Em contrapartida, os modelos com aplicação de métodos de seleção apresentaram melhor equilíbrio geral entre sensibilidade, especificidade e métricas de concordância o que também foi averiguado pelo estudo de Muftah A et al. (80).

Contudo, é necessário considerar que o conjunto de dados evidencia prevalências distintas, com maior ocorrência para a classe não clinicamente significativa.

De seguida, iremos abordar os resultados do modelo IV construído sobre o conjunto de descritores de propriedades informativos e preditivos. Segundo, os resultados obtidos o desempenho do modelo demonstrou ser inferior entre os modelos. Nomeadamente nas seguintes métricas: AUC, MCC, *Kappa de Cohen*, *F1-score*, sensibilidade, PPV, NPV e acurácia balanceada. A área sob a curva ROC foi apenas, 0,533, a mais baixa entre os modelos, indicando que o modelo apresenta dificuldade na distinção de classes, com desempenho próximo ao limiar reconhecido como aleatório. A diminuição de desempenho do modelo é também observada a nível das métricas: MCC 0,086 e *Kappa de Cohen* 0,077 que refletem insuficiência entre a concordância e previsões. A sensibilidade, à semelhança com o observado até ao momento foi a menor entre os modelos 0,190 sugerindo limitações significativas na capacidade de identificar corretamente classes clinicamente significativas. Contudo, a especificidade resultou no valor superior entre os modelos 0,875 em excluir lesões clinicamente significativas.

Este desequilíbrio prejudica a métrica *F1-score* 0,258 o valor mais baixo entre modelos, demonstrando que o modelo não mantém equilíbrio entre a previsões positivas realmente positivas e a sensibilidade. A acurácia balanceada de 0,533 também resulta no menor valor entre os modelos, o que reforça a percepção de desempenho limitado e enviesado.

Com os resultados observados é possível concluir que o método baseado na análise estatística univariada para seleção de descritores de propriedades com maior poder informativo e preditivo resultou em métricas de desempenho com valores menores exceto para a especificidade o que possivelmente, corresponde ao número superior de classes negativas entre a classificação histopatológica da amostra. Isto sugere que os modelos construídos sobre a seleção rigorosa de descritores tais com áreas sob a curva superiores e seleção controlada de descritores por número de amostra, pode diluir o poder preditivo geral do modelo.

Como referência de desempenho para os modelos, os resultados foram comparados ao desafio internacional *SPIE-AAPM-NCI Prostate MR Classification Challenge* (84) derivado do conjunto de dados de *PROSTATEx*, baseados na submissão de 71 métodos desenvolvidos para prever lesões clinicamente significativas ou não, avaliados sobre a área sob a curva ROC. Foi determinado que a maioria dos métodos submetidos superou a AUC definida como classificador aleatório (AUC = 0,5), estes resultados encontram-se alinhados com os modelos desenvolvidos nos presentes estudos, tendo o modelo IV evidenciado o resultado mais baixo. Dentro das melhores submissões, identificam-se as AUC 0,87 e 0,84 com erro-padrão a variar entre 0,027 e 0,032 (84).

A discrepância de desempenho em relação ao presente estudo pode ser parcialmente atribuída à partição de dados, à metodologia de seleção do algoritmo de *machine learning*, à seleção de descritores radiômicos e à otimização de hiperparâmetros.

Um ponto relevante observado no desafio é a diversidade de abordagens metodológicas, aproximadamente metade dos participantes utilizou técnicas tradicionais de extração para entrada de classificadores, enquanto a outra recorreu a redes neurais convolucionais (84). Além disso, diferentes combinações de ponderações de imagens foram exploradas, o que demonstra que as escolhas das sequências exercem impacto no desempenho preditivo (84). Paralelamente, verificou-se que a fusão de modelos como estratégias *ensemble* apresentou vantagens no desempenho da classificação. Da combinação dos nove melhores modelos a AUC aumentou para 0,90 enquanto, a fusão de 25 melhores modelos obteve 0,92, superando significativamente qualquer modelo individual (84).

Comparativamente a estes resultados, observa-se que os modelos desenvolvidos neste estudo, ainda que com desempenhos mais baixos, se enquadram na faixa inferior dos valores reportados no desafio internacional. Isso sugere, que embora metodologias tradicionais de seleção e modelos *machine learning* mais simples apresentem potencial discriminativo, abordagens mais avançadas tais como estratégias *ensemble* e redes neurais oferecem maior desempenho, no entanto, são menos interpretáveis o que pode dificultar a implementação na prática clínica. Dessa forma, uma possível etapa promissora para futuros trabalhos será a integração dos métodos aqui explorados com abordagens mais complexas, tal como evidenciado por Mylona E et. al. (3).

Adicionalmente, Mylona E. et al. (3) no *setting* 1, com recurso a conjuntos de dados *ProstateNET* da base de dados *Prostate Imaging Archive* aplicou metodologias distintas de seleção de descritores radiômicos, em combinação com diferentes classificadores de *machine learning*. Para a sequência T2, foram reportados valores médios de AUC entre 0.62 e 0.68 com desvio-padrão 0,04, 0,05, bem como *F1-score* entre 0.65 e 0.74 com desvio-padrão 0,04, 0,05. Importa ainda salientar que Mylona E. et al. (3) também realizou avaliação externa com o conjunto de dados *PROSTATEx2*, o que resultou em desempenhos de AUC compreendidas entre 0,60 e 0,65 com desvio-padrão 0,02, 0,01, e para *F1-score* entre 0,40 e 0,44 com desvio-padrão 0,03, 0,01. Os modelos I, II, III do presente estudo, apresentaram valores de concordância aproximadamente com os respetivos resultados, permanecendo consistentes com a literatura no que respeita às AUCs reportadas, contudo em termos de *F1-score* apresentam discrepâncias. Porém, pode ser explicado pelas diferenças metodológicas significativas, permitindo apenas uma aproximação parcial entre os resultados.

Todavia, o modelo IV apresentou um desempenho baixo, mesmo em comparação com os cenários menos favoráveis descritos por Mylona E. et al. (3). Situações semelhantes foram descritas por Armato S. et al (84) no *SPIE-AAPM-NCI Prostate MR Classification Challenge*, em que alguns métodos também registaram desempenhos inferiores. No entanto, como o estudo não especificou em que condições esses resultados ocorreram, torna-se impossível estabelecer comparações diretas.

A análise dos descritores com maior contributo para o desfecho clínico no modelo II, avaliado através do gráfico SHAP, identificou oito descritores radiômicos, com destaque para: primeira-ordem (*Skewness*, *Mean*, *RootMeanSquared*) e GLCM (*Imc1*, *Imc2*).

Achados semelhantes foram parcialmente relatados por Muftah A et al. (80), que, utilizando o conjunto de dados *PROSTATEx* e dados internos de mpMRI de próstata em estágio inicial, demonstraram que o desempenho do algoritmo *random forest* dependia sobretudo de 14 descritores radiômicos em ambos os conjuntos. No caso do *PROSTATEx*, oito desses descritores correspondiam a propriedades de primeira ordem, incluindo *Median*, *Skewness* e *Kurtosis*.

Embora a *Kurtosis* não tenha sido considerada no presente estudo uma vez, que foi extraída a partir da sequência DWI, o descritor *Skewness* foi igualmente identificado com elevado contributo e adicionalmente, como não redundante e informativo. Importa ainda salientar que, tanto no conjunto *PROSTATEx* como nos dados internos, Muftah A et al. (80) observaram predominância de descritores de primeira-ordem, sugerindo que a relevância destes não depende exclusivamente do conjunto de dados utilizado.

Ainda que o método de seleção de descritores radiômicos exerça maior impacto no desempenho dos modelos do que propriamente a escolha do algoritmo de *machine learning* (80) este achado reforça a importância crítica da etapa de seleção de descritores em estudos radiômicos (3).

Na literatura, observa-se uma grande diversidade de estratégias metodológicas, Perniciano A. et al. (29) ao analisar 45 estudos radiômicos, verificaram que 23 aplicaram explicitamente um método de seleção ou compararam técnicas distintas para identificar a mais adequada, enquanto outros 23 recorreram a metodologias consecutivas de seleção. Este panorama evidencia que ainda não existe consenso quanto à abordagem ideal. Pelo contrário, o debate permanece ativo, refletindo tanto a complexidade intrínseca da determinação de descritores robustos como a variabilidade entre conjuntos de dados e objetivos clínicos (9).

É importante salientar adicionalmente que estudos retrospectivos enfrentam limitações adicionais, dado que os descritores radiômicos são sensíveis a múltiplas fontes de variabilidade. Além de que, armazenar, analisar, organizar, atribuir (metadados e dados clínicos) é uma tarefa desafiadora que necessita de controle de qualidade para garantir integridade dos dados fornecidos (84). O estudo de L. Bianchini et al. (20) paralelamente demonstrou de forma clara como fatores variacionais promovem variabilidades inerentes nos descritores. Essa falta de controle pode explicar, em parte, a ausência de desempenhos superiores em alguns modelos. Assim talvez seja adequado o estudo de dados prospectivos, concebidos de forma a controlar sistematicamente essas variabilidades, de modo a obter resultados mais consistentes e generalizáveis.

Com base nos resultados obtidos e na literatura disponível, observa-se que valores de AUC de 0,87 para o melhor classificador individual e de 0,92 através da fusão de modelos, reportados no desafio *SPIE-AAPM-NCI Prostate MR Classification Challenge*, evidenciam o potencial da radiômica como método complementar na predição do PCa, sobretudo na redução de biópsias desnecessárias (84). Esta abordagem mostra-se promissora para a distinção entre lesões clinicamente significativas e não significativas, embora apresente limitações na atribuição de informação adicional, como o Grau de *Gleason* (84).

Na última década, o PSA tem sido amplamente utilizado como biomarcador para o diagnóstico e prognóstico do PCa. Contudo, o estudo de Chen S. et al (85) demonstrou que menos de 30% dos homens com níveis de PSA entre 4-10 ng/ml apresentam confirmação patológica, o que releva uma elevada proporção de biópsias desnecessárias, sobrediagnóstico e desperdícios de recursos médicos. Apesar da utilidade destes métodos convencionais, persistem limitações quanto à sua precisão e latência diagnóstica. Por contraste, a radiômica, pela sua natureza não invasiva e custo relativamente reduzido, apresenta grande potencial de aplicação clínica. No entanto, a sua implementação ainda enfrenta desafios relevantes, incluindo baixa reprodutibilidade, significância estatística limitada e ausência de padronização rigorosa baseada em boas práticas em *machine learning* (11,43,44).

Estes aspetos reforçam a necessidade de investigações contínuas em métodos de seleção que consolidem a robustez e a aplicabilidade clínica desta abordagem emergente.

6 Conclusões e Perspetivas Futuras

O presente estudo teve como objetivo avaliar o impacto de métodos de seleção de descritores radiômicos, baseados na análise estatística multivariada e univariada, na construção de modelos de *machine learning* aplicados a imagens de mpMRI da próstata. A investigação permitiu abordar os principais desafios metodológicos da radiômica, particularmente associados à elevada redundância e à presença de descritores não-informativos, fatores que comprometem a robustez e a generalização dos modelos preditivos.

Relativamente ao primeiro objetivo específico, foi possível identificar 1009 descritores de primeira-ordem e de textura com correlação baixa a moderada com descritores de forma. Além disso, através da correlação de *Spearman*, foram identificados 386 descritores redundantes, evidenciando a utilidade deste método na redução de redundâncias entre descritores.

No segundo objetivo, a aplicação do ICC entre o conjunto original e o randomizado demonstrou capacidade adicional, resultando em 789 descritores com correlação baixa a moderada e na identificação de 620 descritores não-informativos, reforçando o papel complementar deste método na seleção de variáveis relevantes.

Quanto ao terceiro objetivo, a análise comparativa entre diferentes métodos de combinação de conjuntos demonstrou que o conjunto de interseção, com 650 descritores, produziu modelos mais equilibrados, com melhor compromisso entre sensibilidade e especificidade. Já o conjunto união, com 1154 descritores, apresentou maior sensibilidade e área sob a curva ROC, embora à custa de uma diminuição acentuada da especificidade. Estes

resultados sugerem que a interseção constitui a metodologia mais robusta e seletiva para reduzir a dimensionalidade e aumentar a interpretabilidade dos modelos.

No quarto objetivo, a análise univariada não identificou descritores estatisticamente significativos após correção de comparações múltiplas, o que evidencia as limitações do uso exclusivo desta abordagem para seleção de variáveis preditivas. Adicionalmente, a utilização de descritores selecionados unicamente com base em valores superiores de AUC resultou no pior desempenho entre os modelos, confirmando a inadequação dessa estratégia de forma isolada.

De forma integrada e geral, os resultados obtidos reforçam que metodologias de seleção de descritores radiômicos, assentes em análises estatísticas rigorosas, como *Spearman* e ICC, contribuem para a construção de modelos mais robustos, ainda que impliquem *trade-offs* em métricas como a acurácia e a especificidade. Estes achados representam um avanço metodológico na área e oferecem orientações úteis para futuras investigações na descoberta de biomarcadores radiômicos robustos, interpretáveis e clinicamente aplicáveis.

No cancro da próstata, as limitações dos métodos diagnóstico atuais destacam a radiômica como uma abordagem complementar, não invasiva e promissora, pela sua capacidade de distinguir lesões clinicamente significativas de não clinicamente significativas, no entanto, a consolidação clínica requer transparência metodológica que permita o desenvolvimento de modelos mais interpretáveis.

Os resultados apresentados contribuem para o avanço metodológico na área e oferecem orientações úteis para futuras investigações na descoberta de biomarcadores radiômicos robustos, interpretáveis e clinicamente aplicáveis. Contudo, é fundamental que esta metodologia seja validada em *datasets* independentes e aplicada a diferentes patologias, modalidades de imagem e regiões anatómicas, de forma a aferir a sua robustez e a capacidade de generalização dos resultados.

6.1 Limitações do presente trabalho

Entre as principais limitações do presente estudo, destaca-se o número reduzido de amostras, o que impossibilitou a calibração dos algoritmos e a realização de validação externa. Além disso, em função da abrangência da análise conduzida, não foi possível aplicar a metodologia a um conjunto de imagens anatómicas distintas, provenientes de outra base de dados pública, o que teria permitido avaliar de forma mais robusta e abrangente. Adicionalmente, utilizou-se apenas uma sequência de ponderação entre as várias

disponíveis, restringindo a possibilidade de investigar a influência de diferentes sequências de aquisição. Por fim, não foram reportadas análises de significância estatística, erro-padrão ou intervalos de confiança na avaliação de desempenho, o que limita a interpretação de diferenças subtis entre métricas.

Referências Bibliográficas

1. Wang L, Lu B, He M, Wang Y, Wang Z, Du L. Prostate Cancer Incidence and Mortality: Global Status and Temporal Trends in 89 Countries From 2000 to 2019. *Front Public Health*. 2022 Feb 16;10.
2. Zhu X, Shao L, Liu Z, Liu Z, He J, Liu J, et al. MRI-derived radiomics models for diagnosis, aggressiveness, and prognosis evaluation in prostate cancer. Vol. 24, *Journal of Zhejiang University: Science B*. Zhejiang University; 2023. p. 663–81.
3. Mylona E, Zaridis DI, Kalantzopoulos C, Tachos NS, Regge D, Papanikolaou N, et al. Optimizing radiomics for prostate cancer diagnosis: feature selection strategies, machine learning classifiers, and MRI sequences. *Insights Imaging*. 2024 Dec 1;15(1).
4. Kelloff GJ, Choyke P, Coffey DS. Challenges in clinical prostate cancer: Role of imaging. Vol. 192, *American Journal of Roentgenology*. 2009. p. 1455–70.
5. Demirel HC, Davis JW. Multiparametric magnetic resonance imaging: Overview of the technique, clinical applications in prostate biopsy and future directions. Vol. 44, *Turkish Journal of Urology*. AVES Ibrahim KARA; 2018. p. 93–102.
6. Kendrick J, Francis R, Hassan GM, Rowshanfarzad P, Jeraj R, Kasisi C, et al. Radiomics for Identification and Prediction in Metastatic Prostate Cancer: A Review of Studies. Vol. 11, *Frontiers in Oncology*. Frontiers Media S.A.; 2021.
7. Ghai S, Haider MA. Multiparametric-MRI in diagnosis of prostate cancer. Vol. 31, *Indian Journal of Urology*. Medknow Publications; 2015. p. 194–201.
8. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, Van Stiphout RGPM, Granton P, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012 Mar;48(4):441–6.
9. Rogers W, Seetha T, Refaee S, Lieverse TIY, Granzier RWY, Ibrahim R. Radiomics: from qualitative to quantitative imaging. Vol. 93, *Br J Radiol*. 2020.
10. van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B. Radiomics in medical imaging—“how-to” guide and critical reflection. Vol. 11, *Insights into Imaging*. Springer; 2020.
11. Shur JD, Doran SJ, Kumar S, Papanikolaou N. Radiomics in Oncology A Practical Guide. *RadioGraphics* [Internet]. 2021 Oct 27 [cited 2025 Aug 7];41:1717–32. Available from: <https://pubs.rsna.org/doi/full/10.1148/rg.2021210037>
12. Elmahdy M, Sebro R. Radiomics analysis in medical imaging research. Vol. 70, *Journal of Medical Radiation Sciences*. John Wiley and Sons Ltd; 2023. p. 3–7.
13. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: The process and the challenges. *Magn Reson Imaging*. 2012 Nov;30(9):1234–48.

14. Mayerhoefer ME, Materka A, Langs G, Häggström I, Szczypiński P, Gibbs P, et al. Introduction to radiomics. *Journal of Nuclear Medicine*. 2020 Apr 1;61(4):488–95.
15. Santinha J, Pinto dos Santos D, Laqua F, Visser JJ, Groot Lipman KBW, Dietzel M, et al. *ESR Essentials: radiomics—practice recommendations by the European Society of Medical Imaging Informatics*. European Radiology. Springer Science and Business Media Deutschland GmbH; 2024.
16. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than pictures, they are data. *Radiology*. 2016 Feb 1;278(2):563–77.
17. Koenigkam Santos M, Raniery Ferreira Júnior J, Tadao Wada D, Priscilla Magalhães Tenório A, Henrique Nogueira Barbosa M, Mazzoncini De Azevedo Marques P. Artificial intelligence, machine learning, computer-aided diagnosis, and radiomics: Advances in imaging towards to precision medicine. *Radiol Bras*. 2019 Nov 1;52(6):387–96.
18. Abrol S, Kotrotsou A, Salem A, Zinn PO, Colen RR. Radiomic Phenotyping in Brain Cancer to Unravel Hidden Information in Medical Images. *Topics in Magnetic Resonance Imaging*. 2017 Feb;26(1):43–53.
19. Welch ML, McIntosh C, Haibe-Kains B, Milosevic MF, Wee L, Dekker A, et al. Vulnerabilities of radiomic signature development: The need for safeguards. *Radiotherapy and Oncology*. 2019 Jan 1;130:2–9.
20. Bianchini L, Santinha J, Loução N, Figueiredo M, Botta F, Origi D, et al. A multicenter study on radiomic features from T2-weighted images of a customized MR pelvic phantom setting the basis for robust radiomic models in clinics. *Magn Reson Med*. 2021 Mar 1;85(3):1713–26.
21. Papanikolaou N, Matos C, Koh DM. How to develop a meaningful radiomic signature for clinical use in oncologic patients. *Cancer Imaging [Internet]*. 2020;20(1):33. Available from: <https://doi.org/10.1186/s40644-020-00311-4>
22. Vallières M, Kay-Rivest E, Perrin LJ, Liem X, Furstoss C, Aerts HJWL, et al. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. *Sci Rep [Internet]*. 2017;7(1):10117. Available from: <https://doi.org/10.1038/s41598-017-10371-5>
23. Rodrigues A, Santinha J, Galvão B, Matos C, Couto FM, Papanikolaou N. Prediction of Prostate Cancer Disease Aggressiveness Using Bi-Parametric Mri Radiomics. *Cancers (Basel) [Internet]*. 2021;13(23). Available from: <https://www.mdpi.com/2072-6694/13/23/6065>
24. Alzubi J, Nayyar A, Kumar A. *Machine Learning from Theory to Algorithms: An Overview*. In: *Journal of Physics: Conference Series*. Institute of Physics Publishing; 2018.
25. Singh O, Bolla SR. *Anatomy, Abdomen and Pelvis, Prostate*. StatPearls [Internet], editor. 2023.
26. WHO. *Data Platform Prostate Cancer*. 2025. World Health Organization.
27. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. 2020 May 1;295(2):328–38.
28. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, De Jong EEC, Van Timmeren J, et al. Radiomics: The bridge between medical imaging and personalized medicine. Vol. 14, *Nature Reviews Clinical Oncology*. Nature Publishing Group; 2017. p. 749–62.
29. Perniciano A, Loddo A, Di Ruberto C, Pes B. Insights into radiomics: impact of feature selection and classification. *Multimed Tools Appl*. 2024 Nov 15;
30. Mylona E, Zaridis DI, Kalantzopoulos C, Tachos NS, Regge D, Papanikolaou N, et al. Optimizing radiomics for prostate cancer diagnosis: feature selection strategies, machine learning classifiers, and MRI sequences. *Insights Imaging*. 2024 Dec 1;15(1).

31. Demirel HC, Davis JW. Multiparametric magnetic resonance imaging: Overview of the technique, clinical applications in prostate biopsy and future directions. Vol. 44, Turkish Journal of Urology. AVES Ibrahim KARA; 2018. p. 93–102.
32. Micheau Antoine, Hoa Denis. Anatomy, Abdomen and Pelvis, Prostate. 2025.
33. Avanzo M, Stancanello J, El Naqa I. Beyond imaging: The promise of radiomics. Vol. 38, Physica Medica. Associazione Italiana di Fisica Medica; 2017. p. 122–39.
34. Chatfield M, C. Weinreb J. PI-RADS™ Prostate Imaging-Reporting and Data System 2015 version 2 [Internet]. 2015 [cited 2024 Jul 23]. Available from: <https://www.acr.org/-/media/ACR/Files/RADS/Pi-RADS/PIRADS-V2.pdf>
35. Trivizakis E, Papadakis GZ, Souglakos I, Papanikolaou N, Koumakis L, Spandidos DA, et al. Artificial intelligence radiogenomics for advancing precision and effectiveness in oncologic care (Review). *Int J Oncol*. 2020 Jul 1;57(1):43–53.
36. Kim H, Park CM, Lee M, Park SJ, Song YS, Lee JH, et al. Impact of reconstruction algorithms on CT radiomic features of pulmonary tumors: Analysis of intra- and inter-reader variability and inter-reconstruction algorithm variability. *PLoS One*. 2016 Oct 1;11(10).
37. Kim H, Goo JM, Ohno Y, Kauczor HU, Hoffman EA, Gee JC, et al. Effect of Reconstruction Parameters on the Quantitative Analysis of Chest Computed Tomography. Vol. 34, *Journal of Thoracic Imaging*. Lippincott Williams and Wilkins; 2019. p. 92–102.
38. Cho HH, Kim CK, Park H. Overview of radiomics in prostate imaging and future directions. Vol. 95, *British Journal of Radiology*. British Institute of Radiology; 2022.
39. Pengpeng X, Yi X, U. Joseph S, Akos VS, Joseph G, Basel Y, et al. Radiomics The next Frontier of Cardiac Computed Tomography. *Circ Cardiovasc Imaging*. 2021 Mar;14(3).
40. Ge G, Zhang J. Feature selection methods and predictive models in CT lung cancer radiomics. Vol. 24, *Journal of Applied Clinical Medical Physics*. John Wiley and Sons Ltd; 2023.
41. Rizzo S, Botta F, Raimondi S, Origgi D, Fanciullo C, Morganti AG, et al. Radiomics: the facts and the challenges of image analysis. Vol. 2, *European Radiology Experimental*. Springer; 2018.
42. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: The process and the challenges. *Magn Reson Imaging*. 2012 Nov;30(9):1234–48.
43. Parmar C, Grossmann P, Rietveld D, Rietbergen MM, Lambin P, Aerts HJWL. Radiomic machine-learning classifiers for prognostic biomarkers of head and neck cancer. *Front Oncol*. 2015;5(DEC).
44. Park JE, Park SY, Kim HJ, Kim HS. Reproducibility and generalizability in radiomics modeling: Possible strategies in radiologic and statistical perspectives. *Korean J Radiol*. 2019 Jul 1;20(7):1124–37.
45. Fajemisin JA, Gonzalez G, Rosenberg SA, Ullah G, Redler G, Latifi K, et al. Magnetic Resonance-Guided Cancer Therapy Radiomics and Machine Learning Models for Response Prediction. *Tomography*. 2024 Sep 2;10(9):1439–54.
46. Gao K, Mei G, Piccialli F, Cuomo S, Tu J, Huo Z. Julia language in machine learning: Algorithms, applications, and open issues. Vol. 37, *Computer Science Review*. Elsevier Ireland Ltd; 2020.
47. Park JE, Park SY, Kim HJ, Kim HS. Reproducibility and generalizability in radiomics modeling: Possible strategies in radiologic and statistical perspectives. *Korean J Radiol*. 2019 Jul 1;20(7):1124–37.
48. Sollini M, Bartoli F, Marciano A, Zanca R, Start RHJA, Erba PA. Artificial intelligence and hybrid imaging: the best match for personalized medicine in oncology. Vol. 4, *European Journal of Hybrid Imaging*. Springer Science and Business Media B.V.; 2020.

49. Pisner DA, Schnyer DM. Chapter 6 - Support vector machine. In: Mechelli A, Vieira S, editors. Machine Learning [Internet]. Academic Press; 2020. p. 101–21. Available from: <https://www.sciencedirect.com/science/article/pii/B9780128157398000067>
50. Yang GR, Wang XJ. Artificial Neural Networks for Neuroscientists: A Primer. *Neuron*. 2020 Sep;107(6):1048–70.
51. Shafiq-UI-hassan M, Zhang GG, Latifi K, Ullah G, Hunt DC, Balagurunathan Y, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. *Med Phys*. 2017 Mar 1;44(3):1050–62.
52. Berenguer R, Del Rosario Pastor-Juan M, Canales-Vázquez J, Castro-García M, Villas MV, Legorburo FM, et al. Radiomics of CT features may be nonreproducible and redundant: Influence of CT acquisition parameters. *Radiology*. 2018 Aug 1;288(2):407–15.
53. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology*. 2020 May;295(2):328–38.
54. Chandrashekar G, Sahin F. A survey on feature selection methods. *Computers and Electrical Engineering*. 2014 Jan;40(1):16–28.
55. Ge G, Zhang J. Feature selection methods and predictive models in CT lung cancer radiomics. Vol. 24, *Journal of Applied Clinical Medical Physics*. John Wiley and Sons Ltd; 2023.
56. Meyer M, Ronald J, Vernuccio F, Nelson RC, Ramirez-Giraldo JC, Solomon J, et al. Reproducibility of CT radiomic features within the same patient: Influence of radiation dose and CT reconstruction settings. *Radiology*. 2019;293(3):583–91.
57. Pudjihartono N, Fadason T, Kempa-Liehr AW, O’Sullivan JM. A Review of Feature Selection Methods for Machine Learning-Based Disease Risk Prediction. Vol. 2, *Frontiers in Bioinformatics*. Frontiers Media SA; 2022.
58. Powers D. Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation. *Mach Learn Technol*. 2008 Sep;2.
59. Trevor Hastie, Robert Tibshirani, Jerome H. Friedman. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. 2nd ed. Springer; 2009.
60. Uribe CF, Mathotaarachchi S, Gaudet V, Smith KC, Rosa-Neto P, Bénard F, et al. Machine learning in nuclear medicine: Part 1—Introduction. *Journal of Nuclear Medicine*. 2019 Apr 1;60(4):451–8.
61. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, et al. The cancer imaging archive (TCIA): Maintaining and operating a public information repository. *J Digit Imaging*. 2013 Dec;26(6):1045–57.
62. Litjens G, Debats O, Barentsz J, Karssemeijer N, Huisman H. Dataset The Cancer Imaging Archive. 2017. SPIE-AAPM PROSTATEx Challenge Data.
63. Cuocolo R, Comelli A, Stefano A, Benfante V, Dahiya N, Stanzione A, et al. Deep Learning Whole-Gland and Zonal Prostate Segmentation on a Public Dataset. *Journal of Magnetic Resonance Imaging*. 2021 Aug 26;54(2):452–9.
64. Litjens G, Debats O, Barentsz J, Karssemeijer N, Huisman H. Computer-Aided Detection of Prostate Cancer in MRI. *IEEE Trans Med Imaging*. 2014 May;33(5):1083–92.
65. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* [Internet]. 2017 Nov 1;77(21):e104–7. Available from: <https://aacrjournals.org/cancerres/article/77/21/e104/662617/Computational-Radiomics-System-to-Decode-the>

66. Van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* 2017 Nov 1;77(21):e104–7.
67. McKinney W. *Data Structures for Statistical Computing in Python*. In 2010. p. 56–61.
68. Vallat R. Pingouin: statistics in Python. *J Open Source Softw.* 2018 Nov 19;3(31):1026.
69. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods.* 2020 Mar 2;17(3):261–72.
70. Akoglu H. User's guide to correlation coefficients. Vol. 18, *Turkish Journal of Emergency Medicine*. Emergency Medicine Association of Turkey; 2018. p. 91–3.
71. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med.* 2016 Jun 1;15(2):155–63.
72. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research.* 2012 Sep;12.
73. McKinney W. *Data Structures for Statistical Computing in Python*. In 2010. p. 56–61.
74. McKinney W. *Data Structures for Statistical Computing in Python*. In 2010. p. 56–61.
75. Seabold S, Perktold J. *Statsmodels: Econometric and Statistical Modeling with Python*. Proceedings of the 9th Python in Science Conference. 2010 Sep;2010.
76. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. *Nature.* 2020 Sep 17;585(7825):357–62.
77. Chicco D, Jurman G. The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification. *BioData Min.* 2023 Dec 1;16(1).
78. Doran SJ, Kumar S, Orton M, d'Arcy J, Kwaks F, O'Flynn E, et al. "Real-world" radiomics from multi-vendor MRI: an original retrospective study on the prediction of nodal status and disease survival in breast cancer, as an exemplar to promote discussion of the wider issues. *Cancer Imaging [Internet]*. 2021;21(1):37. Available from: <https://doi.org/10.1186/s40644-021-00406-6>
79. Çınarer G, Emiroglu B. Classification of brain tumours using radiomic features on MRI. *New Trends and Issues Proceedings on Advances in Pure and Applied Sciences.* 2020 Sep;80–90.
80. Muftah A, Schirmer SM, Langbein FC. Texture Feature Analysis for Classification of Early-Stage Prostate Cancer in mpMRI. 2024 Jun 21; Available from: <http://arxiv.org/abs/2406.15571>
81. Adiraju RV, Kalyani K, Suryanarayana G, Zakariah M, Almazyad AS. Radiomics: Assessing Significance and Correlation with Ground-Truth Data in Precision Medicine in Lung Adenocarcinoma. *Bioengineering [Internet]*. 2025;12(6). Available from: <https://www.mdpi.com/2306-5354/12/6/576>
82. Wu W, Parmar C, Grossmann P, Quackenbush J, Lambin P, Bussink J, et al. Exploratory study to identify radiomics classifiers for lung cancer histology. *Front Oncol.* 2016;6(MAR).
83. Demircioğlu A. The effect of preprocessing filters on predictive performance in radiomics. *Eur Radiol Exp.* 2022 Dec 1;6(1).
84. Armato SG, Huisman H, Drukker K, Hadjiiski L, Kirby JS, Petrick N, et al. PROSTATEx Challenges for computerized classification of prostate lesions from multiparametric magnetic resonance images. *Journal of Medical Imaging.* 2018 Nov 10;5(04):1.
85. Chen S, Jian T, Chi C, Liang Y, Liang X, Yu Y, et al. Machine Learning-Based Models Enhance the Prediction of Prostate Cancer. *Front Oncol.* 2022 Jul 6;12.

Apêndices

Apêndice A: Classificação de Grupo *Gleason*

Quadro A.1-Classificação de Grupos *Gleason*, com a respetiva descrição histopatológica da próstata associadas a cada categoria. Fonte:(7,20,26,27).

Score grupo 1	<i>Gleason</i> score ≤ 6	Apenas glândulas individuais, discretas e bem formadas.
Score grupo 2	Pontuação de <i>Gleason</i> 3+4=7	Glândulas predominantemente bem formadas, com um menor componente de glândulas malformadas, fundidas e cribriformes.
Score grupo 3	Pontuação de <i>Gleason</i> 4+3=7	Glândulas predominantemente malformadas, fundidas, cribriformes com um menor componente de glândulas bem formadas.
Score grupo 4	Pontuação de <i>Gleason</i> 4+4=8, 3+5=8, 5+3=8	Apenas glândula cribriformes, malformadas e fundidas ou glândulas predominantemente bem formadas e um componente menor sem glândulas ou glândulas predominantemente ausentes e um componente menor de glândulas bem formadas.
Score grupo 5	Pontuação de <i>Gleason</i> 9 e 10	Ausência de formação de glândulas ou presença de necrose, com ou sem glândulas malformadas, fundidas e cribriformes.

Apêndice B: Frequência absoluta da idade da população em estudo

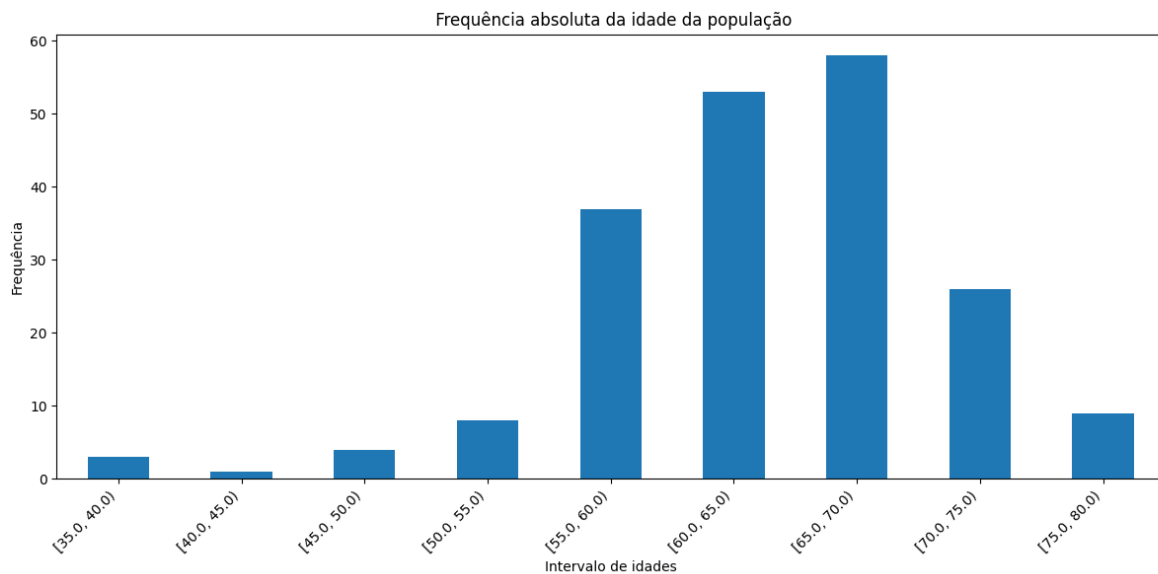


Figura B.1-Frequência absoluta da idade da população em estudo (59).

Apêndice C: Aplicação de Filtros sobre a Imagem Original

Quadro C.1-Imagens derivadas de filtros, descrição das transformações realizadas sobre as intensidades dos píxeis Fonte:(5,10,13,15).

Aplicação de filtros	
Filtros	Descrição
Wavelet	Em cada nível de classe, o filtro realiza decomposições, combinando filtros passa-alto e passa-baixo em cada uma das três dimensões na imagem.
LoG	Controlado pelo parâmetro sigma, que define se as texturas enfatizadas serão mais finas ou mais grossas. Esse filtro destaca regiões com variações no nível de cinza, tornando os limites mais evidentes.
Square	Eleva as intensidades da imagem ao quadrado e, em seguida, realiza um redimensionamento linear para os valores no intervalo original.

Square-root	Aplica a raiz quadrada às intensidades absolutas da imagem e redimensiona linearmente os valores para o intervalo original.
Logaritmo	Calcula o logaritmo da intensidade absoluta acrescida de 1 da imagem e redimensiona os valores resultantes para o intervalo original.
Exponencial	Aplica a função exponencial diretamente às intensidades da imagem original.

Apêndice D: Mapas de Calor análise estatística correlação de Spearman

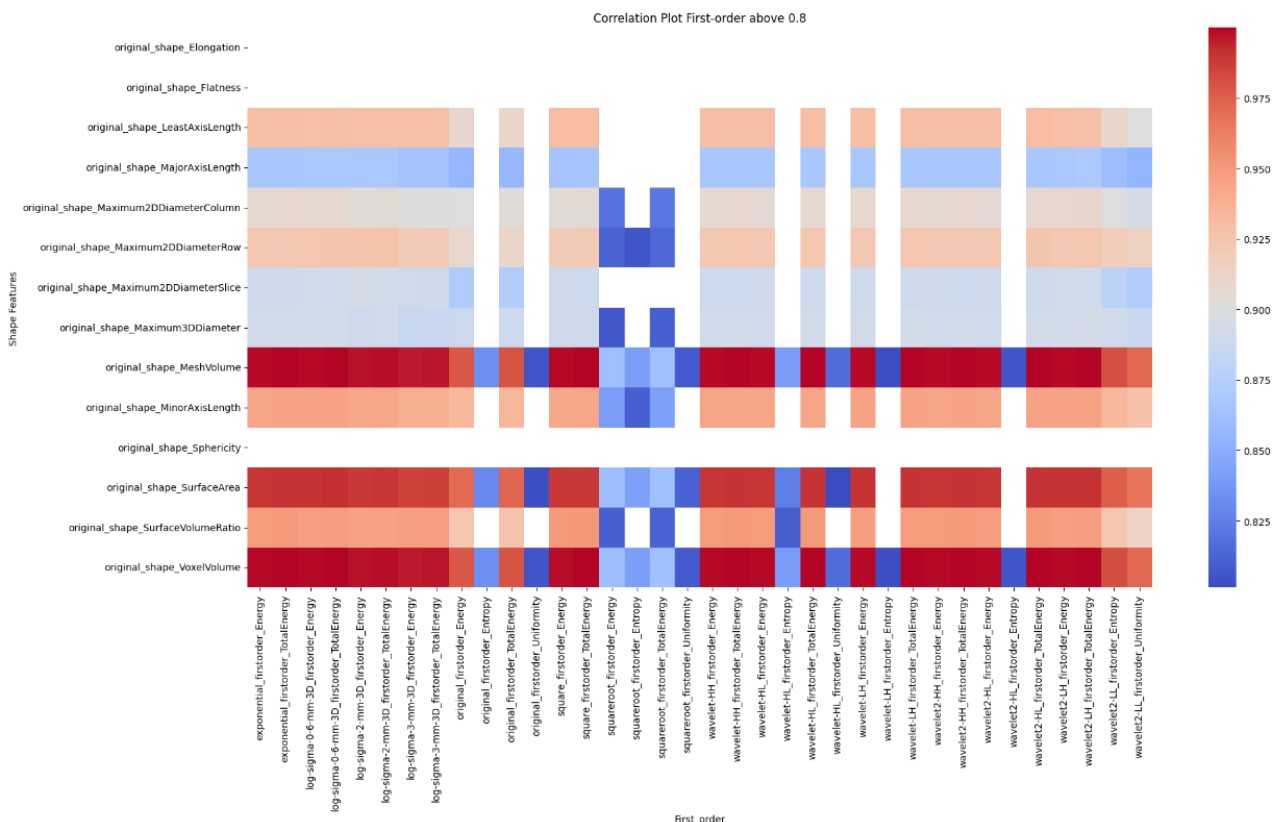


Figura D.1- *Correlation Plot* entre descritores de propriedades de primeira-ordem vs. forma $\rho > 0,8$.

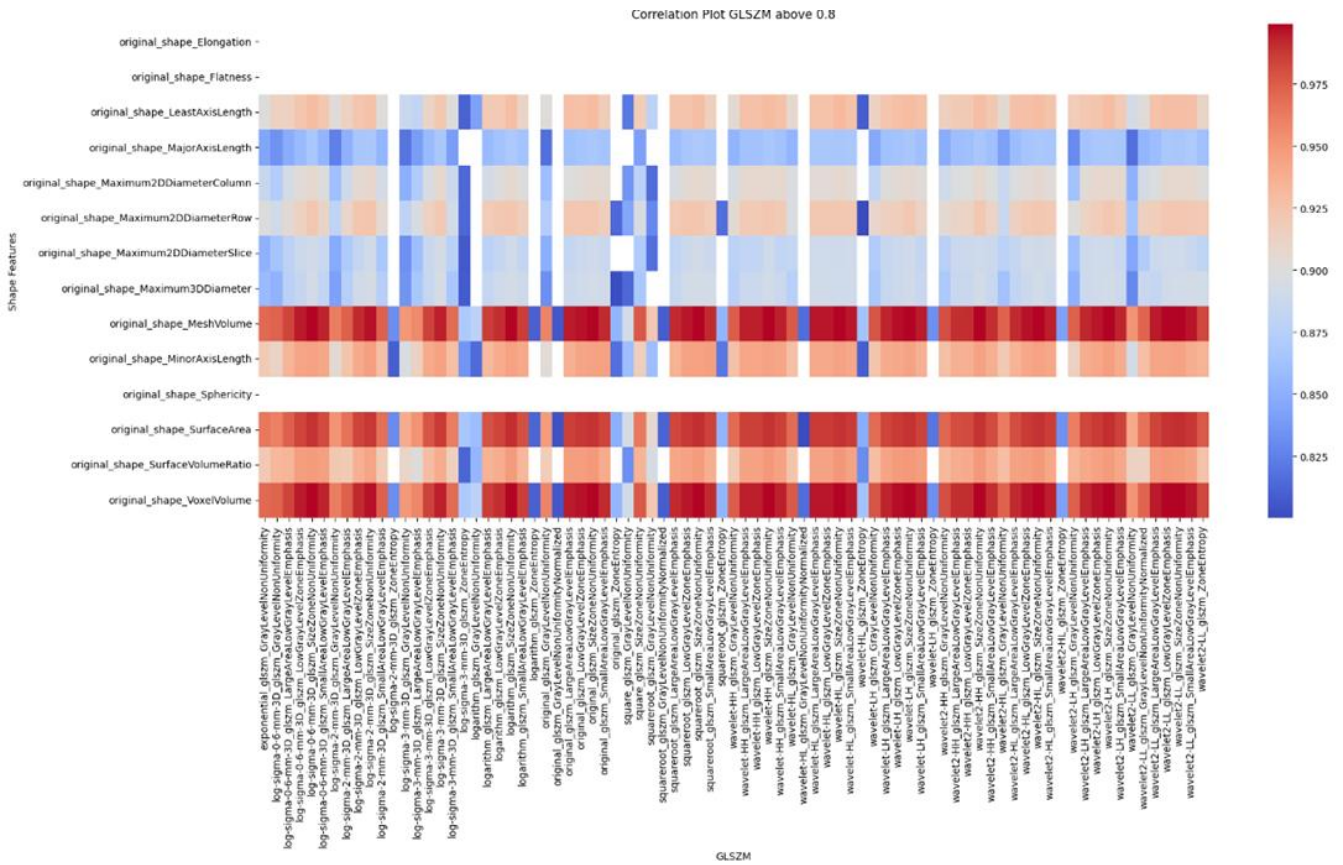


Figura D.4-*Correlation Plot* entre descritores de propiedades de textura GLSZM vs. forma $\rho > 0,8$.

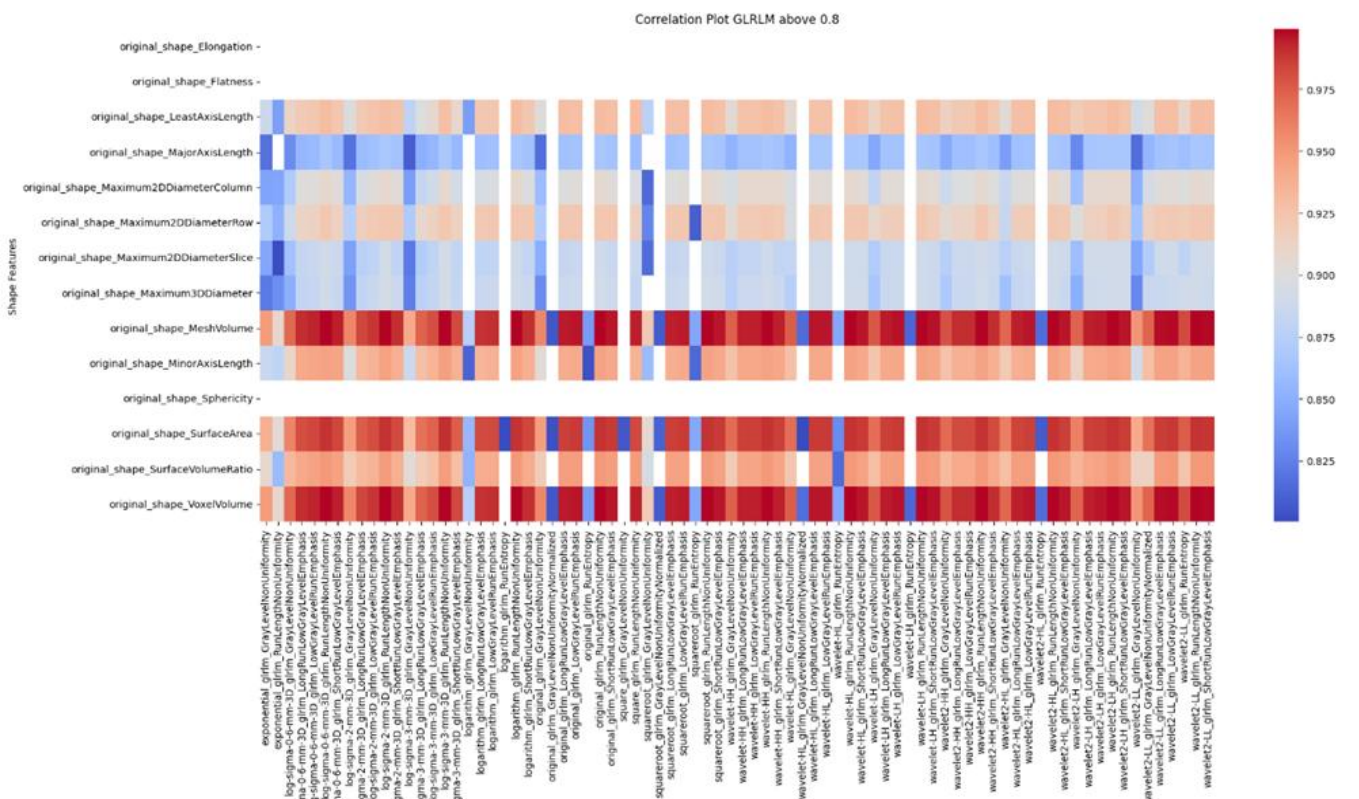


Figura D.5-*Correlation Plot* entre descritores de propiedades de textura GLRLM vs. forma $\rho > 0,8$.

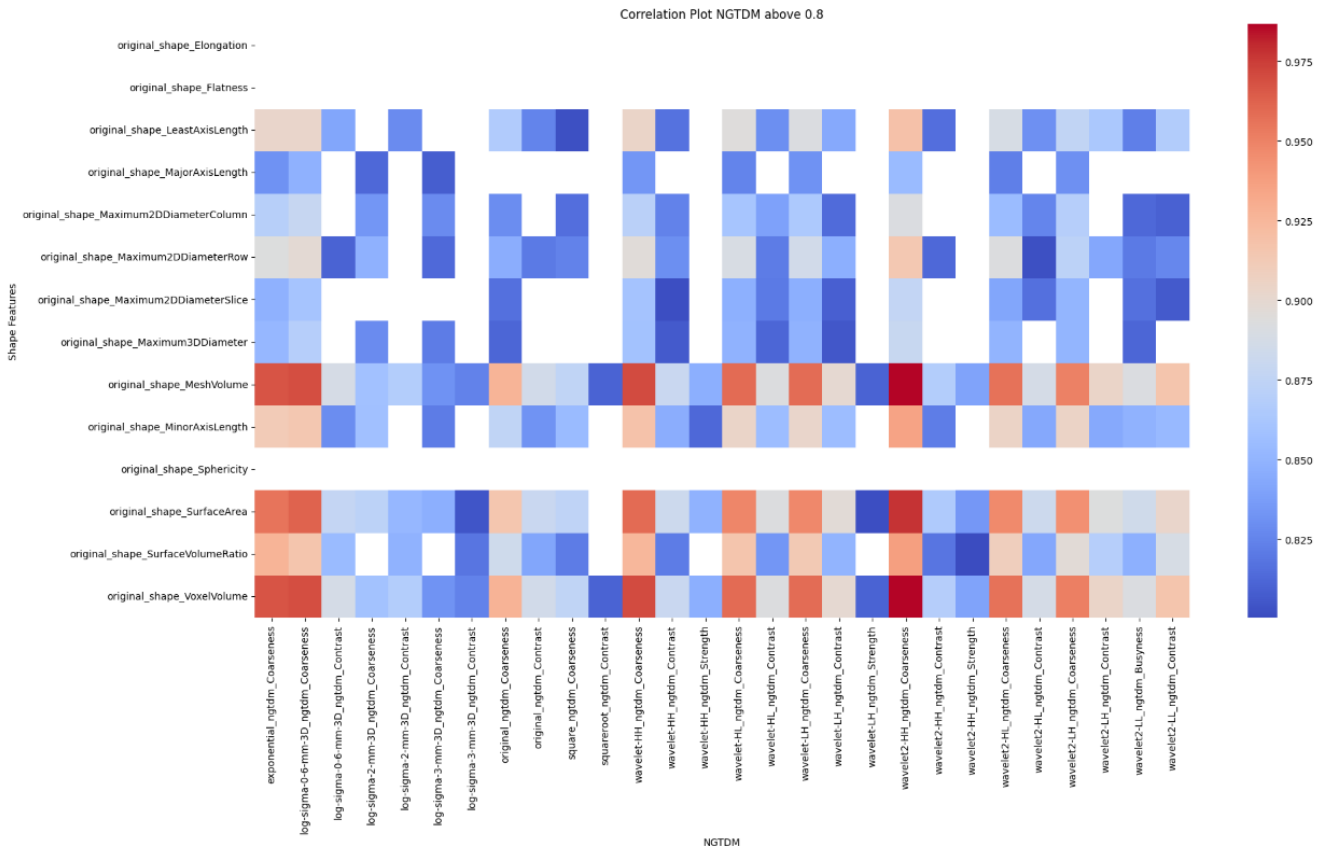


Figura D.6-Correlation Plot entre descriptores de propiedades de textura NGTDM vs. forma $\rho > 0,8$.

Apêndice E: Gráfico de barras análise estatística ICC

Number of Features ICC above 0.7

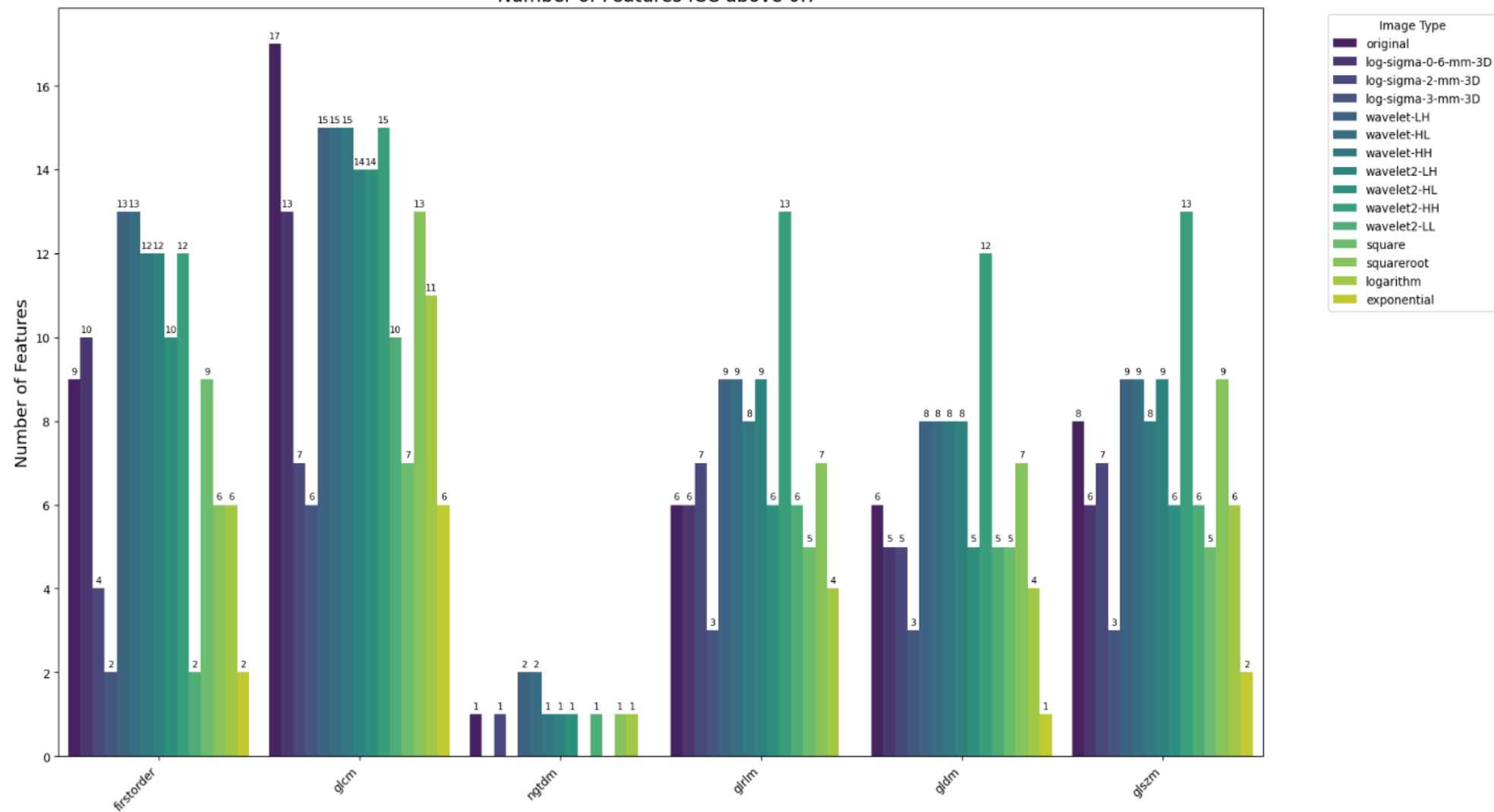


Figura E.1-Correlation Plot entre descritores de propriedades radiômicas ICC $\rho > 0,7$.

