

**UNIVERSIDADE DE SÃO PAULO  
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO**

**RAFAEL VASCONCELOS BARROS**

**RADIÔMICA NA DIFERENCIAÇÃO ENTRE LESÕES DE BAIXO E ALTO GRAU  
NA RESSONÂNCIA MAGNÉTICA DE PRÓSTATA**

**RIBEIRÃO PRETO**

**2023**

**RAFAEL VASCONCELOS BARROS**

**RADIÔMICA NA DIFERENCIAÇÃO ENTRE LESÕES DE BAIXO E ALTO GRAU  
NA RESSONÂNCIA MAGNÉTICA DE PRÓSTATA**

Dissertação apresentada à Faculdade de Medicina  
de Ribeirão Preto da Universidade de São Paulo  
para e Titulação de Mestrado Profissional  
Área de concentração: Radiologia  
Orientador: Prof. Dr. Valdair Francisco Muglia

RIBEIRÃO PRETO

2023

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

BARROS, RV.

Radiômica na diferenciação entre lesões de baixo e alto grau na ressonância magnética de próstata

47 p.; 3 il; 30 cm

Dissertação de mestrado profissional apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo. Área de concentração: Radiologia

Orientador: Prof. Dr. Valdair Francisco Muglia

Palavras-chave: Ressonância Magnética; inteligência artificial; PyRadiomics; Sklearn; Random Forest; SVM; Logistic Regression; neoplasia de próstata

**FOLHA DE APROVAÇÃO**

BARROS, RV. Radiômica na diferenciação entre lesões de baixo e alto grau na ressonância magnética de próstata. Dissertação apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo para titulação de mestrado profissional. Área de concentração: Radiologia

Aprovado em:

\_\_\_\_ / \_\_\_\_ / \_\_\_\_\_

**BANCA EXAMINADORA**

Dr.

Instituição

\_\_\_\_\_

\_\_\_\_\_

Julgamento

Assinatura

\_\_\_\_\_

\_\_\_\_\_

Dr.

Instituição

\_\_\_\_\_

\_\_\_\_\_

Julgamento

Assinatura

\_\_\_\_\_

\_\_\_\_\_

Dr.

Instituição

\_\_\_\_\_

\_\_\_\_\_

Julgamento

Assinatura

\_\_\_\_\_

\_\_\_\_\_

## AGRADECIMENTOS

A Deus, pela dádiva da vida.

Aos meus pais, pelo amor incondicional todos os dias.

A Luísa, minha companheira de vida, pelo apoio em todas as jornadas.

Ao meu orientador, Prof. Valdair Muglia, pela amizade e dedicação ao ensino.

Ao meu colega Júlio, por ajudar a concretizar ideias no computador.

Aos colegas Marcelo Raposo, Carlos Mikio e André Benez, que se dedicaram a fazer esse projeto sair do papel.

Por fim, a todos os amigos que fiz em Ribeirão Preto, que contribuíram de alguma forma nesse projeto.

“O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001”.

## **NÚCLEO DE PESQUISA/SETOR**

Centro de Ciências da Imagem e Física Médica da Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo – USP

### **LOCAL DE EXECUÇÃO**

Unidade Campus do Hospital das Clínicas de Ribeirão Preto – Universidade de São Paulo – USP

## **PESQUISADORES RESPONSÁVEIS**

Dr. Rafael Vasconcelos Barros

Pós-graduando em Radiologia e Diagnóstico por Imagem da Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo – USP

Dr. Júlio César Nather Júnior

Doutorando em Radiologia e Diagnóstico por Imagem da Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo – USP

Dr. Carlos Mikio Tanno Filho

Médico Adido do Setor de Radiologia e Diagnóstico por Imagem da Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo – USP

Dr. Marcelo de Almeida Braga Raposo do Amaral

Médico residente de Radiologia e Diagnóstico por Imagem da Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo - USP

Prof. Dr. Valdair Francisco Muglia

Docente do departamento de Imagens Médicas, Oncologia e Hematologia da Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo- USP.

## RESUMO

Barros RV. **Radiômica na diferenciação entre lesões de baixo e alto grau na ressonância magnética de próstata.** Faculdade de Medicina de Ribeirão Preto, Universidade da São Paulo, 2023.

**Objetivo:** Avaliar o desempenho da radiômica, com base em ferramentas de processamento de inteligência artificial com *machine learning*, para a diferenciação entre lesões de baixo/intermediário grau (ISUP 1 e 2) e lesões de alto grau (ISUP 3-5) em pacientes com confirmação histopatológica de câncer de próstata (CaP).

**Materiais e métodos:** Segmentamos, manualmente, as lesões de próstata visíveis em exames consecutivos de ressonância nuclear magnética multiparamétrica (RM) realizados em nossa instituição, de 2019 a 2022, usando o software 3D Slicer. Posteriormente, foram extraídas características de imagem utilizando a biblioteca PyRadiomics. Pacientes que apresentaram artefatos nas imagens de RM, lesões com escore de Gleason <6, lesões localmente avançadas, intervalo de tempo entre a RM e a biópsia ou cirurgia superior a 4 meses, lesões sem confirmação histopatológica no serviço ou ausência de lesões visíveis na RM foram excluídos do estudo. Após as exclusões, 146 pacientes antes de prostatectomia radical (26%) ou biópsia direcionada (74%) foram incluídos no estudo, dos quais 83 pacientes com lesões ISUP 1 (n=33) ou 2 (n=50) foram atribuídos ao grupo 1, e outros 63 pacientes com lesões ISUP 3 (n=18), 4 (n=18) ou 5 (n=27) foram atribuídos ao grupo 2. A proporção de pacientes nos conjuntos de treinamento e validação foi de 70% dos dados para treinamento e 30% para validação. Os modelos de *Support Vector Machine* (SVM), *Random Forest* (RF) e *Logistic Regression* (LR) foram usados para classificar os casos. A acurácia, matriz de confusão, precisão, recall, F1-score e área sob a curva característica de operação do receptor (AUC-ROC) foram avaliados para cada modelo. Um procedimento de *cross validation* foi realizado para avaliar a consistência e estabilidade dos modelos. **Resultados:** O *cross validation* para o modelo *support vector machine* mostrou acurácia média de 0.91(0,87-0,94), precisão de 0.92(0,87-0,98) e a curva AUC-ROC de 0.98(0,96- 0,99). O modelo *random forest* teve acurácia média de 0.91(0,88- 0,95), precisão de 0.98(0,95- 1,01) e a curva AUC-ROC de 0.97(0,95-0,99). O modelo *logistic regression* mostrou uma acurácia média de 0.93

(0,89-0,98), precisão de 0.96(0,92-1,00) e curva AUC-ROC de 0,96 (0,93-0,99).

**Conclusão:** A utilização de radiômica pode diferenciar com precisão lesões de baixo/intermediário grau (ISUP 1 e 2) de lesões de alto grau (ISUP 3-5) e pode ser uma ferramenta auxiliar importante para a confirmação da elegibilidade de pacientes que consideram uma abordagem conservadora como manejo primário do câncer de próstata. Estudos prospectivos e multi-institucionais adicionais são necessários para a validação de nossos resultados encorajadores.

**Palavras chave:** Ressonância Magnética; inteligência artificial; *PyRadiomics*; *Sklearn*; *Random Forest*; *SVM*; *Logistic Regression*



## ABSTRACT

BARROS RV. Radiomics in the differentiation between low and high-grade lesions on prostate magnetic resonance imaging

**Objectives:** To evaluate the performance of radiomics, based on artificial intelligence processing tools with machine learning, for the differentiation between low/intermediate (ISUP 1 and 2) and high-grade lesions (ISUP 3-5), in patient with histopathological confirmation of prostate cancer. **Materials and Methods:** We manually segmented MRI-visible prostate lesions from consecutive mpMRI exams performed at our institution, from 2019 to 2022 using the 3D Slicer software. Subsequently, image features were extracted using the PyRadiomics library. Patients presenting artifacts on MRI images, lesions with Gleason score <6, locally advanced lesions, time interval between MRI and biopsy or surgery longer than 4 months, lesions without histopathological confirmation in the service, or absence of visible lesions on MRI were excluded from the study. After exclusions, 146 patients before radical prostatectomy (26%) or targeted biopsy (74%) were included in the study, of which 83 patients with ISUP 1 (n=33) or 2(n=50) lesions were assigned to group 1, and another 63 patients with ISUP 3(n=18), 4(n=18) or 5(n=27) lesions, assigned to group 2. The proportion of patients in the training and validation sets used was 70% of the data for training and 30% for validation. The Support Vector Machine (SVM), Random Forest (RF), and Logistic Regression (LR) models were used to classify the cases. Accuracy, confusion matrix, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC) were evaluated for each model. A cross-validation procedure was performed to assess the consistency and stability of the models. **Results:** The cross-validation for the support Vector Machine model showed an average accuracy of 0.91 (0.87-0.94), precision of 0.92 (0.87-0.98), and AUC-ROC curve of 0.98 (0.96-0.99). The random forest model had an average accuracy of 0.91 (0.88-0.95), precision of 0.98 (0.95-1.01), and AUC-ROC curve of 0.97 (0.95-0.99). The logistic regression model showed an average accuracy of 0.93 (0.89-0.98), precision of 0.96 (0.92-1.00), and AUC-ROC curve of 0.96 (0.93-0.99). **Conclusion:** The use of radiomics can accurately differentiate low/intermediate-grade lesions (ISUP

1 and 2) from high-grade lesions (ISUP 3-5) and can be an important auxiliary tool for confirming the eligibility of patients considering a conservative approach as the primary management for prostate cancer. Further prospective and multi-institutional studies are required for the validation of our encouraging results.

**Keywords:** Magnetic Resonance, Artificial Intelligence, PyRadiomics, Sklearn; Random Forest; SVM; Logistic Regression; Prostate cancer

## LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS

<b>ADC</b>	<i>Apparent diffusion coefficient</i> – Coeficiente de difusão aparente
<b>CaP</b>	Câncer de próstata
<b>DCE</b>	<i>Dynamic contrast-enhanced</i> – Sequências dinâmicas com contraste
<b>DICOM</b>	<i>Digital imaging and communications in medicine</i> – Imagens digitais e comunicações em Medicina
<b>DP</b>	Desvio-padrão
<b>DWI</b>	<i>Diffusion-weighted Imaging</i> – Sequência com restrição à difusão
<b>ESUR</b>	<i>European Society of Urogenital Radiology</i> – Sociedade Europeia de Radiologia Urogenital
<b>FOV</b>	<i>Field of view</i> – Campo de visão
<b>GLCM</b>	<i>Grey level co-occurrence matrix</i> – Matriz de coocorrência em nível de cinza
<b>GLSZM</b>	<i>Grey level size zone matrix</i> – Matriz de zona de tamanho em nível de cinza
<b>GLDM</b>	<i>Grey level dependence matrix</i> – Matriz de dependência de nível de cinza
<b>GLRLM</b>	<i>Grey level run length matrix</i> – Matriz de comprimento de execução em nível de cinza
<b>NGTDM</b>	<i>Neighboring gray tone difference matrix</i> – Matriz de diferença entre tons de cinza vizinhos
<b>HC – FMRP – USP</b>	Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo
<b>IA</b>	Inteligência artificial
<b>IC</b>	Intervalo de Confiança
<b>ISUP</b>	<i>International Society of Urologic Pathologists</i> – Sociedade Internacional de Patologistas Urológicos
<b>LR</b>	<i>Logistic Regression</i>

<b>mpRNM</b>	Ressonância nuclear magnética multiparamétrica
<b>PI-RADS</b>	<i>Prostate Imaging Reporting &amp; Data System</i> – Sistema de dados e relatórios de imagem da próstata
<b>PSA</b>	<i>Prostate-specific antigen</i> – Antígeno prostático específico
<b>dPSA</b>	<i>PSA density</i> - Densidade de PSA
<b>PTX</b>	Prostatectomia radical
<b>RIS</b>	<i>Radiological information system</i> – Sistema de informações radiológicas
<b>RF</b>	<i>Random Forest</i>
<b>RM</b>	Ressonância magnética
<b>VOI</b>	<i>Volume of interest</i> – Volume de interesse
<b>SIH</b>	Sistema de informações hospitalares
<b>SVM</b>	<i>Support Vector Machine</i>
<b>PCA</b>	<i>Principal component analysis</i> – Análise do componente principal

## LISTA DE FIGURAS

<b>Figura 1</b> -Fluxograma mostrando número de exames iniciais, exclusões e o estudo de coorte final. ....	24
<b>Figura 2</b> -Segmentação da lesão prostática visualizada na RM no 3D <i>Slicer</i> com reconstrução volumétrica .....	28
<b>Figura 3</b> -Segmentação da lesão prostática no terço médio à esquerda visualizada na RM no 3D <i>Slicer</i> .....	28
<b>Figura 4</b> -Mapa ilustrativo do funcionamento do <i>PyRadiomics</i> .....	29
<b>Figura 5</b> -Histograma mostrando a contribuição das características mais importantes na assinatura radiômica. ....	35
<b>Figura 6</b> - Comparação entre os modelos no <i>cross-validation</i> .....	39

**LISTA DE TABELAS**

<b>Tabela 1</b> -Características dos pacientes do grupo 1 .....	34
<b>Tabela 2</b> -Características dos pacientes do grupo 2 .....	35
<b>Tabela 3</b> -Tabela de resultados no <i>validation test</i> – <i>Support vector machine</i> .....	36
<b>Tabela 4</b> -Tabela de resultados no <i>cross-validation</i> – <i>Support vector machine</i> .....	36
<b>Tabela 5</b> -Tabela de resultados no <i>validation test</i> – <i>Random forest</i> .....	36
<b>Tabela 6</b> -Tabela de resultados no <i>cross-validation</i> – <i>Random forest</i> .....	37
<b>Tabela 7</b> -Tabela de resultados no <i>validation test</i> – <i>Logistic regression</i> .....	37
<b>Tabela 8</b> -Tabela de resultados no <i>cross-validation</i> – <i>Logistic regression</i> .....	38
<b>Tabela 9</b> -Comparação entre resultados dos três modelos avaliados no <i>validation test</i> .....	38
<b>Tabela 10</b> -Comparação entre resultados dos três modelos avaliados no <i>cross- validation</i> .....	38

## SUMÁRIO

<b>FOLHA DE APROVAÇÃO .....</b>	<b>3</b>
<b>SUMÁRIO .....</b>	<b>14</b>
<b>1. INTRODUÇÃO.....</b>	<b>16</b>
<b>2. OBJETIVOS .....</b>	<b>22</b>
<b>2.1. Objetivo principal .....</b>	<b>22</b>
<b>2.2. Objetivo secundário.....</b>	<b>22</b>
<b>3. MATERIAIS E MÉTODOS.....</b>	<b>23</b>
<b>3.1. Tipo de estudo .....</b>	<b>23</b>
<b>3.2. Local do estudo .....</b>	<b>23</b>
<b>3.3. Coleta de dados.....</b>	<b>23</b>
<b>3.4. Comitê de Ética.....</b>	<b>23</b>
<b>3.5. Critérios de inclusão .....</b>	<b>23</b>
<b>3.6. Critérios de exclusão.....</b>	<b>23</b>
<b>3.7. Protocolo de Ressonância Magnética (RM) .....</b>	<b>25</b>
<b>3.8. Segmentação de ressonância magnética.....</b>	<b>25</b>
<b>3.9. Randomização dos pacientes .....</b>	<b>26</b>
<b>3.10. Extração de recursos .....</b>	<b>26</b>
<b>3.11. Análise e processamento dos dados .....</b>	<b>29</b>
<b>3.11.1. Análise exploratória dos dados .....</b>	<b>29</b>
<b>3.11.2. Redução de dimensionalidade.....</b>	<b>30</b>
<b>3.11.3. Visualização de dendrogramas e hierarquias de agrupamento .....</b>	<b>30</b>
<b>3.11.4. Preparação dos dados para treinamento e validação .....</b>	<b>30</b>
<b>3.11.5. Normalização e padronização dos dados.....</b>	<b>30</b>
<b>3.11.6. Treinamento e avaliação dos modelos de classificação .....</b>	<b>31</b>

<b>3.11.7. Cross-validation</b> .....	<b>31</b>
<b>3.11.8. Comparação e seleção do melhor modelo</b> .....	<b>31</b>
<b>3.11.9. Implementação e monitoramento do modelo</b> .....	<b>32</b>
<b>3.11.10. Padrão de referência</b> .....	<b>32</b>
<b>4. RESULTADOS</b> .....	<b>33</b>
<b>4.1. Support Vector Machine (SVM)</b> .....	<b>36</b>
<b>4.2. Random Forest (RF)</b> .....	<b>36</b>
<b>4.3. Logistic Regression (LR)</b> .....	<b>37</b>
<b>4.4. Comparação entre os modelos</b> .....	<b>38</b>
<b>5. DISCUSSÃO</b> .....	<b>40</b>
<b>6. CONCLUSÃO</b> .....	<b>42</b>
<b>7. REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	<b>43</b>



## 1. INTRODUÇÃO

O câncer de próstata (CaP) tem relevância significativa para os sistemas de saúde, visto que oscila entre a quarta e quinta causa de morte entre a população masculina mundial (1). No Brasil, representa a segunda causa de morte por câncer entre os homens e é a terceira neoplasia maligna mais incidente na população (2020 - INCA). Há ainda a estimativa de 72 mil novos casos por ano durante o triênio 2023-2025 (2023 - INCA). Segundo a OMS, 1 em cada 9 homens vai ser diagnosticado com CaP ao longo da vida. Desses, apenas 1 a cada 41 homens com câncer de próstata vai morrer por isso.

Considerando-se o grande número de neoplasias indolentes, houve a necessidade de se buscar uma definição de quais CaPs teriam relevância clínica, surgindo o conceito importante de câncer clinicamente significativo: Gleason>6 ou volume tumoral >0,5 cm<sup>3</sup> ou densidade de PSA (dPSA) > 0,15 (33).

Diante desse cenário, o diagnóstico e o tratamento precoces, além de técnicas de predição de quais pacientes necessitarão de tratamento mais agressivo, podem impactar positivamente o desfecho desses casos. Nesse contexto, emergem novas tecnologias de imagem, dentre elas métodos de inteligência artificial (IA), como radiômica, capaz de fazer extração e análise de dados de imagens radiológicas, que podem ser convertidas em modelos descritivos e preditivos combinando características de imagem e fenótipos da doença (1).

Atualmente, o rastreamento do câncer de próstata deve ser feito em homens a partir de 50 anos ou de 45 anos, caso sejam de raça negra ou possuam histórico da doença em parentes de primeiro grau, conforme recomendação da Sociedade Brasileira de Urologia (2). A investigação costuma ser feita por meio do exame de toque retal, da dosagem sérica de *prostate-specific antigen* (PSA) e, caso seja indicada, a biópsia de próstata pode dar o diagnóstico de câncer (3).

Dentre os casos de neoplasia maligna, a doença pode ser estratificada como de alto grau, de grau intermediário e de baixo grau de acordo com a classificação da *International Society of Urologic Pathologists* (ISUP), do PSA e do estadiamento clínico (1). Nesse cenário, o uso da ressonância magnética multiparamétrica (mpRNM) de próstata tem ganhado destaque, e os urologistas, de modo geral, defendem a realização desse exame antes da biópsia, visto que a combinação dos

dois é capaz de diagnosticar o câncer em até 96% dos casos de doença localizada, enquanto a biópsia guiada por outros métodos de imagem resulta em falsos negativos em cerca de 38-46% dos casos (1).

Considerando o uso desse método para avaliação da próstata, é preciso conhecer a anatomia do órgão e entender como será visualizado no exame de imagem. Anatomicamente, em plano sagital, a próstata é dividida em base, terço médio e ápice. Em plano axial, pode-se dividi-la em 4 zonas: estroma fibromuscular anterior, que não contém tecido glandular; zona de transição, que cerca a uretra e possui 5% de tecido glandular; zona central, que possui 20% do tecido glandular; e zona periférica, onde se encontram 70-80% do tecido glandular. Quase a totalidade dos cânceres de próstata são adenocarcinomas e se desenvolvem no tecido glandular, nas proximidades do ducto prostático. Por isso, 70% se originam da zona periférica, 20-25% acometem a zona periférica e a de transição e somente 5 a 10% são exclusivos da zona de transição (7).

Histologicamente, a próstata é avaliada por meio do material enviado para estudo após a biópsia, que deve coletar 8-12 fragmentos, preferencialmente 12, das 6 regiões: apical, mediana e basal, bilateralmente. Os fragmentos devem ser enviados para estudo em frascos separados e identificados. Vale salientar que a Ressonância Magnética (RM) é superior ao US para guiar essa coleta. Pelo sistema de Gleason, de acordo com a diferenciação glandular e com o padrão de crescimento em relação ao estroma, gradua-se a lesão prostática. Apesar de ser a gradação mais utilizada mundialmente por muito tempo, Gleason vem perdendo espaço, desde 2005, quando a ISUP definiu nova classificação aperfeiçoando o sistema prévio (5).

O escore de Gleason e o ISUP são sistemas de classificação utilizados para avaliar o grau de agressividade do câncer de próstata, objetivando auxiliar na determinação do melhor tratamento para cada caso.

O escore de Gleason, por meio de classificação histológica, avalia a aparência das células cancerígenas na biópsia da próstata, em uma escala que varia de 1, menos agressivo, a 5, mais agressivo. Por meio da soma das pontuações dos dois tipos histológicos mais predominantes na amostra estudada, obtém-se o escore de Gleason.

A Sociedade Internacional de Patologia Geniturinária propôs um sistema de classificação que também avalia o grau de agressividade do câncer de próstata a partir

da análise histológica do tecido tumoral, também visando estratificação de risco. Partindo do escore de Gleason, agrupou-se em cinco categorias, variando da melhor diferenciação para a pior diferenciação histológica: ISUP 1 (Gleason 6), ISUP 2 (Gleason 3 + 4 = 7), ISUP 3 (Gleason 4 + 3 = 7), ISUP 4 (Gleason 8- 4+4, 3+5, 5+3) e ISUP 5 (Gleason 9 e 10- 4+5, 5+4, 5+5). Lesões ISUP 1 são consideradas câncer de baixo risco, enquanto ISUP 5 é um câncer de alto risco (15).

Em resumo, a classificação ISUP é uma versão atualizada do que já era proposto pelo escore de Gleason: a avaliação da agressividade do câncer de próstata para auxiliar o médico assistente na decisão quanto à melhor conduta.

Em relação ao risco de progressão de doença para pacientes com câncer de próstata, uma das primeiras e ainda mais frequentemente usadas classificações foi introduzida em 1998 por D'Amico et al (13) que conta com três categorias. As diretrizes atuais da *National Comprehensive Cancer Network* (NCCN) para o câncer de próstata (versão 4.2019) subdividem ainda mais os pacientes com câncer de próstata de risco por meio de cinco categorias (34):

- Muito baixo risco: todas as características de baixo risco, mais no máximo 2 fragmentos positivos, com, no máximo 50% de acometimento.
- Baixo risco: caracterizado por PSA abaixo de 10 ng/ml, escore de Gleason menor que 7 e estadiamento clínico T1c ou T2a
- Risco intermediário: caracterizado por PSA entre 10 e 20 ng/ml, escore de Gleason igual a 7 ou estadiamento clínico T2b;
- Alto risco: caracterizado por PSA acima de 20 ng/ml, escore de Gleason maior que 7 ou estadiamento clínico T2c ou T3.
- Muito alto risco: duas ou três das características que definem alto risco

Diante do exposto, a vigilância ativa do câncer de próstata é uma estratégia de gerenciamento dos casos de baixo risco de progressão ou de disseminação de doença, em que não há necessidade de tratamento imediato, mas há indicação de monitoramento cuidadoso do tumor. Durante a vigilância ativa, o paciente é monitorado regularmente, por meio de exames laboratoriais, de exames de imagem e de biópsias periódicas. Havendo sinais de progressão da doença, o tratamento pode ser iniciado imediatamente.

Essa conduta conservadora tem se tornado uma opção de tratamento cada vez mais aceita para o câncer de próstata de baixo risco, pois evita os efeitos colaterais

associados ao tratamento mais agressivo, como incontinência urinária, impotência sexual, dentre outros.

Em pacientes classificados como ISUP 2, ou seja, com escore de Gleason 7 (3 + 4), a vigilância ativa ainda pode ser uma opção, apesar de já se tratar de doença de risco intermediário (14). Pacientes que apresentem PSA <10ng/mL, estágio clínico  $\leq$  T2a e um baixo número de fragmentos positivos ( $\leq$  3 fragmentos ou  $\leq$  50% por fragmento) são casos de ISUP 2 considerados favoráveis para tal conduta.

Para seguimento desses pacientes, a avaliação por meio de exame de imagem é de extrema importância, sendo a RM o método mais indicado. Esse exame tem ganhado espaço por fornecer informações metabólicas, sugerir celularidade alterada, além de poder auxiliar, de forma não invasiva, na caracterização de vascularização tecidual e tumoral. As diretrizes mais atuais de câncer de próstata incentivam o uso do método.

Como a interpretação da RM da próstata pode ser subjetiva e inconsistente, foi desenvolvido escore de suspeita para o câncer de próstata, *Prostate Imaging Reporting & Data System*, mais conhecido pela sua sigla: PI-RADS. Essa é uma escala de 1 a 5 pontos, baseada em critérios fixos, para melhorar a padronização da interpretação e relatório da RM. Em 2012, foi lançada sua primeira versão, que incluía diretrizes clínicas para a realização de RM, bem como com uma escala de Likert de 5 pontos para a interpretação da imagem (6). Em 2015, houve refinamento da escala, atualizada em 2019 para versão 2.1, fornecendo parâmetros de aquisição de imagens revisados e um sistema de pontuação que mantém a ideia original da versão 2 (7- 8).

O ACR-PIRADS versão 2.1 recomenda que a aquisição de imagem de ressonância magnética da próstata inclua sequências ponderadas em T1 (T1w), ponderadas em T2 (T2w), imagens ponderadas em difusão (DWI), ADC e dinâmicas com contraste (DCE). É importante seguir um protocolo padronizado para obter imagens de alta qualidade. A imagem multiparamétrica é recomendada para aumentar a sensibilidade e especificidade na detecção de lesões suspeitas da próstata (7-8).

Diante desse cenário e visando identificar, de maneira mais precoce, lesões de próstata suspeitas e qual o sítio mais preciso para realização da biópsia, além do PI-RADS v2.1, a IA ganha cada vez mais evidência na avaliação prostática por meio da RM (9). Por definição, a IA representa a capacidade das máquinas de mimetizar a função cognitiva dos seres humanos para identificar e resolver problemas (10). A necessidade de contar com o auxílio das máquinas na avaliação desses exames

surgiu do aumento da demanda por produtividade. A quantidade de informações obtidas por meio de imagens radiológicas vem crescendo de maneira desproporcional quando comparada ao número de médicos radiologistas treinados para interpretar tais exames (9).

A interpretação de imagens pode se beneficiar do conceito de *machine learning*: computadores são ensinados por radiologistas especialistas amplamente treinados a identificar imagens suspeitas e clinicamente significativas, de modo que reprodutibilidade e eficiência no diagnóstico. *Deep learning* é um subconjunto de *machine learning* e é a base da maioria das ferramentas de IA para interpretação de imagem. O sistema vai mudando a sua capacidade de interpretação à medida que mais substrato é adicionado como treinamento (10).

Considerando que uma mesma afecção pode se apresentar como imagens discretamente divergentes e ainda assim ter a mesma interpretação, a máquina deve ser ensinada que há padrões diferentes a serem armazenados e identificados para uma mesma doença, e esse é o conceito de *data augmentation*. Quanto maior a quantidade de imagens utilizadas como banco de dados para a máquina, maior a acurácia do método (10).

Diante disso, novas ferramentas começaram a ser utilizadas. Radiômica é a extração de características de imagens diagnósticas que gera parâmetros quantitativos e mensuráveis de imagem (10). O processo começa com a aquisição de um vasto banco de dados de imagens de pacientes sobre os quais a evolução é conhecida. Quanto maior o banco de dados, mais sensível e específica se torna a capacidade de acerto da máquina. O ideal é que o computador seja treinado com exames realizados em um mesmo aparelho, sob os mesmos protocolos, objetivando reduzir erros de interpretação em consequência de artefatos de imagem ou de técnica não padronizada (11).

Retomando a necessidade de interpretação de exames de imagem da próstata mais robustas e menos variáveis e, assim, tornar o diagnóstico mais precoce, surgiu a ideia de que o computador poderia ser capaz de identificar lesões suspeitas mais rapidamente. Além disso, a máquina também poderia classificar lesões como sugestivas de neoplasia de alto ou de baixo grau por meio de padrões de imagens, de modo que o radiologista poderia identificar o melhor local de punção para o diagnóstico histopatológico (11).

Surgiu, então, o *Pyradiomics*, um pacote *Python* (uma linguagem de programação de compreensão acessível e de código aberto), para extração de dados radiômicos de imagens médicas. O carregamento e o pré-processamento da imagem são feitos usando o *SimpleITK*. Os dados são convertidos em matrizes *NumPy*, que possuem métodos para a aplicação de funções aritméticas, para cálculos adicionais usando vários recursos, transformando a avaliação subjetiva de imagens em padrões de números, que podem ser objetivamente classificados. Esse método já é aceito para a pesquisa de câncer (12).

Deste modo, o avanço tecnológico e, conseqüentemente, a maior acurácia dos métodos diagnósticos de imagem possibilitam a identificação de casos de câncer que se beneficiariam de tratamentos mais precoces e/ou mais agressivos. De acordo com o PI-RADS v2.1 (8), a adaptação do conceito de CaP clinicamente significativo para a imagem ocorre quando há lesão coincidente nas 3 sequências (T2, DWI e DCE) e/ou  $ADC < 0,750 \times 10^{-3} \text{ mm}^2/\text{s}$ , além de volume  $> 0,5 \text{ cm}^3$  e/ou extensão extraprostática (EPE) (36, 37). Essa denominação é utilizada para diferenciar o CaP com alta morbidade e/ou morte dos subtipos indolentes, nos quais o tratamento pode resultar em mais danos do que a própria doença (32).

Dessa forma, a diferenciação entre CaP clinicamente significativo e clinicamente insignificante, principalmente através de métodos não invasivos, justifica a realização de estudos como este.

## **2. OBJETIVOS**

### **2.1. Objetivo principal**

Avaliar o desempenho da radiômica, com base em ferramentas de processamento de inteligência artificial com *machine learning*, para a diferenciação entre lesões de baixo/intermediário grau (ISUP 1 e 2) e lesões de alto grau (ISUP 3-5) em pacientes com confirmação histopatológica de câncer de próstata (CaP).

### **2.2. Objetivo secundário**

Identificar qual modelo de *machine learning* tem maior precisão diagnóstica na análise proposta.

### **3. MATERIAIS E MÉTODOS**

#### **3.1. Tipo de estudo**

Trata-se de um estudo retrospectivo, observacional, realizado em um único centro.

#### **3.2. Local do estudo**

O estudo foi realizado no Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo (HCFMRP-USP).

#### **3.3. Coleta de dados**

O Sistema de Informações Radiológicas (RIS) e o Sistema de Informações Hospitalares (HIS) de nossa instituição foram as fontes de pesquisa para o presente estudo. De forma retrospectiva, coletamos dados sobre pacientes submetidos à RM de próstata, no período de julho de 2019 a julho de 2022, e que possuem resultados histopatológicos por meio de biópsia direcionada ou de prostatectomia radical (PTX).

#### **3.4. Comitê de Ética**

O estudo foi aprovado pelo Comitê de Ética e Pesquisa (CEP) do HCFMRP-USP com o título: “Radiômica na diferenciação entre lesões de baixo e alto grau na ressonância magnética de próstata”, com dispensa do Termo de Consentimento Livre e Esclarecido (TCLE), devido à natureza retrospectiva. (CAAE: 69597023.0.0000.5440).

#### **3.5. Critérios de inclusão**

Foram incluídos no estudo pacientes que atendessem aos seguintes critérios:

- a) lesão maligna comprovada histologicamente;
- b) realização da RM antes de qualquer medida terapêutica.

#### **3.6. Critérios de exclusão**

Foram excluídos do estudo pacientes nas seguintes situações:

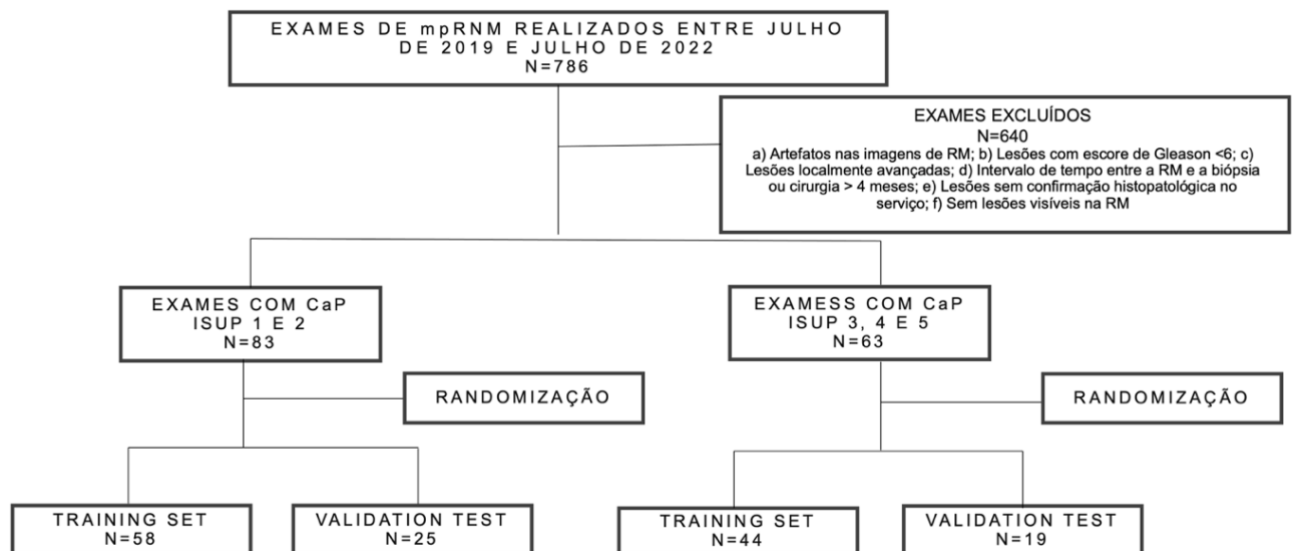
- a) artefatos nas imagens da RM;
- b) lesões com escore de Gleason <6;
- c) lesões localmente avançadas;



- d) tempo compreendido entre a realização da RM e comprovação histológica maior que 4 meses;
- e) lesões sem confirmação histopatológica no serviço;
- f) sem lesões visíveis na RM

Dos 786 exames iniciais, 25 foram excluídos por artefatos nas imagens da RM, 109 por lesões com escore de Gleason <6, 212 por lesões localmente avançadas, 83 pelo tempo compreendido entre a realização da RM e da biópsia ou cirurgia maior que 4 meses, 199 por lesões sem confirmação histopatológica no serviço e 12 por não haver lesões visíveis na RM. Dessa forma, 146 pacientes foram incluídos no estudo, dos quais 83 pacientes com lesões ISUP 1 ou 2, denominados de grupo 1; e outros 63 pacientes com lesões ISUP 3, 4 ou 5, denominados de grupo 2, conforme fluxograma (figura 1).

**Figura 1**-Fluxograma mostrando número de exames iniciais, exclusões e o estudo de coorte final.



Fonte: Arquivo próprio

### **3.7. Protocolo de Ressonância Magnética (RM)**

Todos os exames foram realizados com um aparelho de ressonância magnética de alta definição, 3,0 Tesla, e 16 canais (Philips Achieva, Best - Holanda), utilizando uma bobina corporal de 16 canais colocada ao redor da pelve. A escolha deste equipamento garante maior resolução e qualidade das imagens adquiridas, contribuindo para um diagnóstico mais preciso. O protocolo de ressonância magnética foi definido de acordo com o protocolo PI-RADS (Prostate Imaging - Reporting and Data System) versão 2.1 (8), que é um sistema padronizado de classificação das imagens de ressonância magnética da próstata e tem como objetivo melhorar a detecção, localização, caracterização e risco estratificado do câncer de próstata.

O protocolo PI-RADS inclui sequências ponderadas em T1 (T1w), ponderadas em T2 (T2w), imagens ponderadas em difusão (DWI), coeficiente de difusão aparente (ADC) e dinâmicas com meio de contraste (DCE). Essas sequências são selecionadas para proporcionar uma melhor visualização e diferenciação das estruturas anatômicas e patológicas da próstata, auxiliando assim na identificação e caracterização do câncer de próstata.

Todos os exames de ressonância magnética foram realizados e lidos por dois radiologistas, sendo um fellow e um radiologista com mais de 20 anos de experiência, garantindo a precisão e confiabilidade dos resultados. Cada lesão prostática foi categorizada pela aplicação do PI-RADS v2.1 (8), que permite a classificação das lesões em uma escala de 1 a 5, com base no risco de malignidade. Essa categorização ajuda a determinar a necessidade de intervenções adicionais, como biópsias, e a direcionar o tratamento apropriado para cada paciente.

Para a análise radiômica consecutiva, o mapa ADC (derivado de DWI com valor b de 0/1000), foi extraído no formato “*Digital Imaging and Communications in Medicine*” (DICOM). Os mapas ADC são uma ferramenta útil para avaliar a difusão das moléculas de água nos tecidos e, portanto, ajudam a identificar áreas de câncer, que geralmente apresentam difusão restrita.

### **3.8. Segmentação de ressonância magnética**

A segmentação das imagens de ressonância magnética é uma etapa crucial na análise radiômica, pois permite a identificação e quantificação de características específicas das lesões prostáticas. Em correlação direta com uma estação de trabalho institucional e os respectivos relatórios clínicos, as séries extraídas foram revisadas

por um investigador usando a plataforma de software 3D *Slicer* de código aberto (<http://slicer.org>, versão 4.9.0) (17-18). O 3D *Slicer* é um *software* avançado e amplamente utilizado na pesquisa médica para visualização e análise de imagens médicas, oferecendo diversas ferramentas para processamento e análise de dados.

Com posicionamento do *Volume of Interest* (VOI) consecutivo usando imagens ADC, a segmentação manual das lesões foi executada aplicando a ferramenta de pintura do editor de segmentos conforme mostra a Figura 2. A segmentação manual é um processo minucioso, que envolve a demarcação cuidadosa das áreas de interesse em cada fatia das imagens ADC. Essa etapa é crucial para garantir a precisão na identificação das características radiômicas das lesões prostáticas.

### **3.9. Randomização dos pacientes**

A randomização dos pacientes é um aspecto fundamental no processo de validação do modelo desenvolvido, pois permite a avaliação do desempenho do modelo em um conjunto independente de dados. Isso ajuda a garantir que os resultados obtidos sejam generalizáveis e não sejam apenas aplicáveis ao conjunto de dados utilizado no treinamento do modelo.

Para evitar viés de seleção e garantir a imparcialidade na divisão dos dados, os exames de pacientes foram randomizados em dois grupos: conjunto de treinamento e conjunto de validação. A divisão foi realizada utilizando uma técnica de amostragem aleatória simples, garantindo que cada paciente tivesse a mesma probabilidade de ser incluído em um dos dois grupos. A proporção de pacientes nos conjuntos de treinamento e validação pode variar, mas uma divisão comum é 70% dos dados para treinamento e 30% para validação.

### **3.10. Extração de recursos**

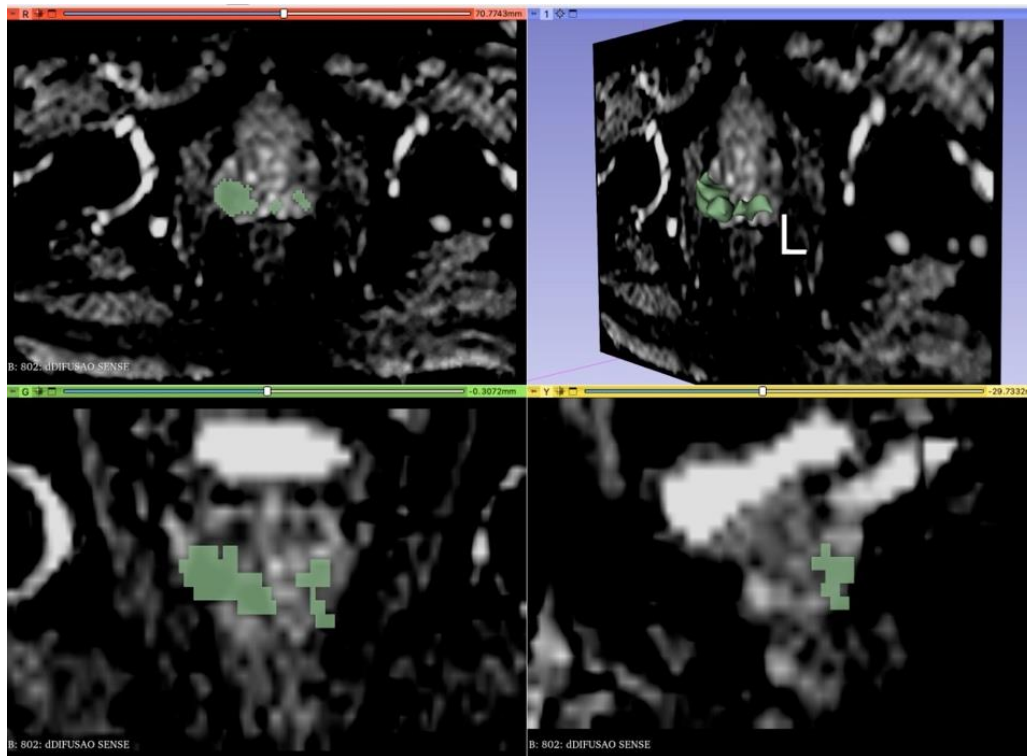
Utilizamos o pacote de código aberto *PyRadiomics* (12) que ganha cada vez mais estabelecimento como padrão de referência para análise radiômica (19). De sete classes de recursos, todos os recursos padrão foram extraídos: estatísticas de primeira ordem, baseadas em forma, matriz de coocorrência de nível de cinza (GLCM), matriz de comprimento de execução de nível de cinza (GRLLM), matriz de zona de tamanho de nível de cinza (GLSZM), matriz de dependência de nível de cinza (GLDM) e matriz de diferença de tons de cinza vizinhos (NGTDM) levando a 105 características/VOIs em ADC (<http://pyradiomics.readthedocs.io>) (12). Extraímos os

recursos com as configurações padrão no *PyRadiomics*, ou seja, sem reamostragem ou filtragem, sem recursos baseados em *wavelets*, largura de *bin* 25 e GLCM simétrico forçado (<http://pyradiomics.readthedocs.io>). A Figura 3 mostra um mapa ilustrativo do funcionamento do *PyRadiomics*.

A função de processamento de múltiplas pastas foi desenvolvida para iterar sobre todas as pastas contendo imagens médicas e máscaras, chamando a função de extração de características radiômicas para cada par de imagem e máscara. Os resultados das características extraídas foram armazenados em um *DataFrame* do *pandas*, com as linhas representando cada exame e as colunas correspondendo às características radiômicas extraídas. Além disso, um identificador único foi atribuído a cada caso, e uma coluna adicional foi incluída no *DataFrame* para indicar a categoria do tumor (alto ou baixo grau). Após o processamento de todas as pastas e da extração das características radiômicas, o *script* criou uma pasta para armazenar os resultados em formato CSV. O *DataFrame* contendo as características e os identificadores foi então exportado como um arquivo CSV, facilitando o uso e a análise dos dados em outras ferramentas e ambientes.

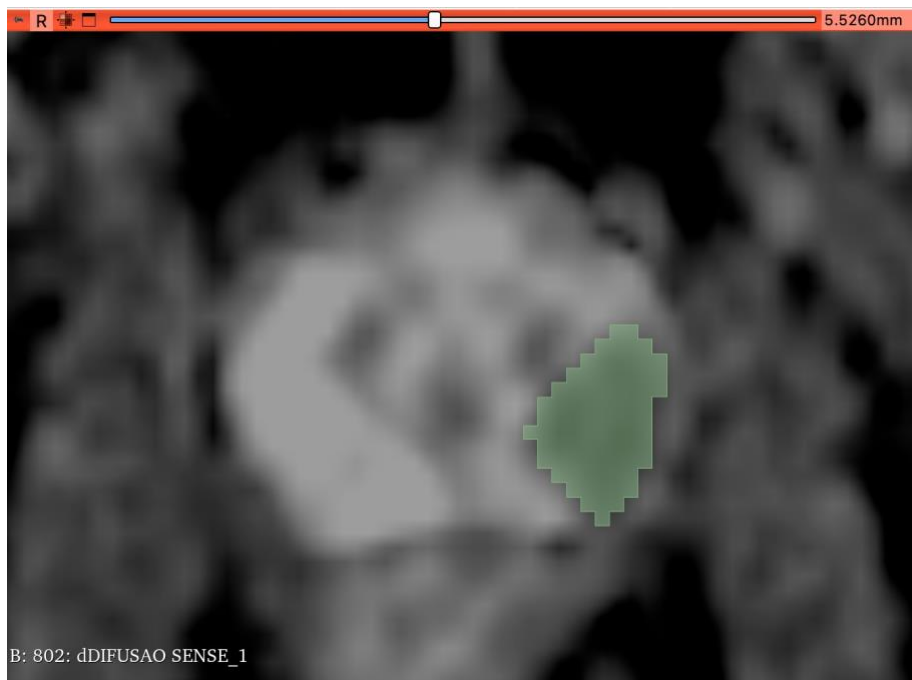
Durante o desenvolvimento e a execução do *script*, foram incluídos comentários e verificações de erros para garantir a correta extração e o correto armazenamento das características radiômicas. Isso incluiu a verificação da existência das pastas e arquivos de imagem e máscara, a validação dos parâmetros de extração e a inspeção visual dos resultados obtidos.

**Figura 2**-Segmentação da lesão prostática visualizada na RM no 3D *Slicer* com reconstrução volumétrica



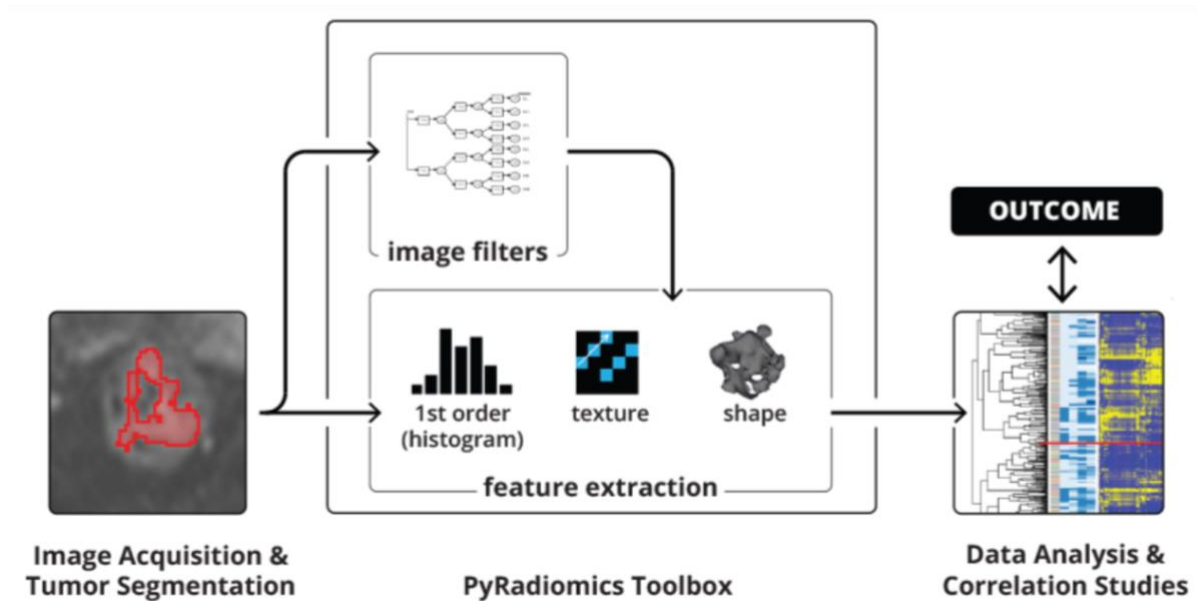
Fonte: Arquivo próprio

**Figura 3**-Segmentação da lesão prostática no terço médio à esquerda visualizada na RM no 3D *Slicer*



Fonte: Arquivo próprio

**Figura 4-**Mapa ilustrativo do funcionamento do *PyRadiomics*



Fonte: *PyRadiomics*

### 3.11. Análise e processamento dos dados

A análise e o processamento dos dados são etapas cruciais no desenvolvimento e avaliação de modelos de *machine learning* para diagnóstico e prognóstico em imagens médicas. Essas etapas envolvem a manipulação, limpeza, visualização e transformação dos dados para garantir a qualidade, a representatividade e a compreensibilidade das informações obtidas. A seguir, detalhamos as etapas e as decisões tomadas durante a análise e o processamento dos dados neste estudo.

#### 3.11.1. Análise exploratória dos dados

A análise exploratória dos dados permitiu uma compreensão inicial dos padrões e das relações presentes no conjunto de dados. Isso incluiu a geração de estatísticas descritivas, como média, mediana, desvio padrão e distribuição das características, e a visualização dos dados por meio de gráficos, como histogramas, *boxplots* e *scatter plots*. Essas análises auxiliaram na identificação de características relevantes, *outliers* e possíveis correlações entre as variáveis.

### **3.11.2. Redução de dimensionalidade**

A técnica *Principal Component Analysis* (PCA) foi aplicada para reduzir a dimensionalidade dos dados e identificar as direções de maior variância nos dados. Isso permitiu uma melhor visualização e interpretação dos dados, além de reduzir o tempo de treinamento e melhorar a eficiência dos modelos de classificação. A quantidade de componentes principais a serem mantidos foi selecionada com base na porcentagem acumulada de variância explicada, garantindo a preservação da maior parte das informações presentes nos dados originais.

### **3.11.3. Visualização de dendrogramas e hierarquias de agrupamento**

A visualização de dendrogramas e hierarquias de agrupamento auxiliou na compreensão dos padrões de agrupamento presentes nos dados e na identificação de possíveis subgrupos ou estruturas de dados. Essas visualizações podem ajudar a detectar relações não lineares entre as características e a identificar grupos de pacientes com perfis semelhantes, o que pode ser útil para a personalização do tratamento e a estratificação de risco.

### **3.11.4. Preparação dos dados para treinamento e validação**

O conjunto de dados foi dividido em dois subconjuntos randomicamente: treinamento e validação. Isso é importante para garantir que os modelos de classificação sejam treinados e avaliados em dados representativos e balanceados.

### **3.11.5. Normalização e padronização dos dados**

Os dados foram normalizados e padronizados para garantir que todas as características tivessem a mesma escala e distribuição. Isso é fundamental

para garantir a comparabilidade das características e a convergência dos algoritmos de *machine learning*.

Depois de processadas as imagens, os dados de resultado (*Features*) foram gravados em planilhas de Excel. Os dados nas planilhas (*Features*) foram processados e consumidos pelos modelos de *machine learning*. Modelos que foram treinados: *logistic regression*, *random forest* e *support vector machine*.

### **3.11.6. Treinamento e avaliação dos modelos de classificação**

Os modelos de classificação, incluindo *support vector machine*, *random forest* e *logistic regression*, foram treinados utilizando o conjunto de dados de treinamento e avaliados com base no conjunto de dados de validação. Métricas de desempenho, como acurácia, matriz de confusão, precisão, *recall*, F1-score e AUC-ROC, foram calculadas para avaliar o desempenho dos modelos e identificar os mais adequados para a tarefa de classificação.

### **3.11.7. Cross-validation**

Foi realizada uma validação cruzada com 10 *folds* para estimar a robustez e a generalização dos modelos de classificação. A validação cruzada envolve a divisão do conjunto de dados em k subconjuntos e a realização do treinamento e validação em cada *fold*, garantindo uma avaliação mais confiável do desempenho dos modelos. As métricas de desempenho foram calculadas para cada *fold* e as médias e intervalos de confiança foram reportados.

### **3.11.8. Comparação e seleção do melhor modelo**

Após treinar e avaliar os modelos de classificação e realizar a validação cruzada, os resultados foram comparados para identificar o modelo com o melhor desempenho geral. A seleção do melhor modelo considerou não apenas as métricas de desempenho, mas também a complexidade, a interpretabilidade e a capacidade de generalização dos modelos. Com base nos resultados, o



modelo selecionado pode ser usado para classificar novos casos e auxiliar na tomada de decisões clínicas.

### **3.11.9. Implementação e monitoramento do modelo**

Após a validação externa e a seleção do modelo final, o modelo pode ser implementado em sistemas clínicos para auxiliar na classificação de tumores de próstata e na tomada de decisões clínicas. É importante monitorar o desempenho do modelo continuamente e atualizá-lo conforme necessário, incorporando novos dados e refinando as características e hiperparâmetros para garantir a sua eficácia e relevância no ambiente clínico em constante evolução.

Para avaliação de qualidade, utilizamos o Checklist para Inteligência Artificial em Imagens Médicas (CLAIM) (35) , uma lista de verificação com 42 itens que compreende elementos para avaliar a reprodução, rigidez, qualidade e generalização ideais. Ele é considerado uma diretriz de "melhores práticas" para relatar algoritmos de IA na literatura.

### **3.11.10. Padrão de referência**

Utilizamos como padrão de referência o resultado histopatológico obtido através de biópsia por fusão cognitiva mais biópsia sistemática por sextantes, no mesmo procedimento, ou prostatectomia radical. Com base na topografia da lesão índice (dominante) na amostra, um radiologista com dois anos de experiência fez a correlação da imagem na RM durante a segmentação.

#### 4. RESULTADOS

Em cada grupo, os pacientes foram caracterizados por meio dos seguintes critérios: idade, método de coleta de material para estudo histopatológico (biópsia transretal ou prostatectomia radical), grau histológico da lesão prostática, volume prostático médio, PSA e dPSA.

No grupo 1, a idade dos pacientes analisados variou de 46 anos a 79 anos, com média de 64,9 e desvio padrão de 7,6 anos. Em relação ao método de obtenção de amostra para estudo histopatológico, a prostatectomia radical prevaleceu, sendo utilizada em 19 casos, o que equivale a 23% da amostra. Nos outros 64 casos (77%) desse grupo, a biópsia transretal com fusão cognitiva de imagens foi o método utilizado. 33 pacientes foram classificados como ISUP 1, e 50 pacientes foram classificados como ISUP 2. O volume prostático médio foi de 38,7 cm<sup>3</sup>, com variação de 16 cm<sup>3</sup> a 125 cm<sup>3</sup>, e dPSA médio foi de 0,22 ng/mL/cm<sup>3</sup>, variando entre 0,03 ng/mL/cm<sup>3</sup> e 0,87 ng/mL/cm<sup>3</sup>. Esses dados podem ser visualizados na tabela 1.

No grupo 2, a idade dos pacientes analisados variou de 41 anos a 86 anos, com média de 68,9 e desvio padrão de 8,6 anos. Dessa vez, o método mais frequente de obtenção de material para estudo histopatológico foi a biópsia transretal, presente em 43 casos, 68% da amostra. A prostatectomia radical foi o método utilizado nos outros 20 casos, representando 32% do Grupo, conforme mostrado na tabela 2. Quanto ao grau histológico, 18 pacientes foram classificados como ISUP 3, 18 pacientes foram classificados como ISUP 4 e 27 pacientes foram classificados como ISUP 5. O volume prostático médio foi de 42,0cm<sup>3</sup>, variando entre 16 cm<sup>3</sup> e 122 cm<sup>3</sup>. O PSA médio variou entre 1,1 ng/mL e 997 ng/mL, com média de 60,1 ng/mL. A densidade do PSA teve média 1,67 ng/mL/cm<sup>3</sup>, com variação entre 0,03 ng/mL/cm<sup>3</sup> e 29,3 ng/mL/cm<sup>3</sup>, conforme dados apresentados na tabela 2.

Os resultados da análise de classificação de tumores de próstata de alto e baixo grau utilizando diferentes modelos de *machine learning* são apresentados a seguir. Foram utilizados os modelos *support vector machine*, *random forest* e *logistic regression* para a classificação dos casos.

Foram avaliados a acurácia, a matriz de confusão, a precisão, o *recall*, o F1-score e a área sob a curva ROC (AUC-ROC) para cada modelo. Além disso, foi realizado o *cross-validation* com 10 *folds* para avaliar a consistência e estabilidade dos modelos.

Os resultados para os modelos *support vector machine*, *random forest* e *logistic regression* são apresentados abaixo. No teste de validação, o *support vector machine* obteve uma precisão de 0,90. A precisão do modelo foi de 0,89, o recall foi de 0,89 e o F1-score foi de 0,89. A AUC-ROC foi de 0,90. *Random forest* obteve uma precisão de 0,93. A precisão do modelo foi de 1,0, o recall foi de 0,84 e o F1-score foi de 0,91. A AUC-ROC foi de 0,92. *Logistic regression* obteve uma precisão de 0,93. A precisão do modelo foi de 0,94, o recall foi de 0,89 e o F1-score foi de 0,91. No *cross-validation* para os modelos *support vector machine*, *random forest* e *logistic regression* apresentou acurácia média de 0.91(0.87- 0.94), 0.91 (0.88- 0.95), e 0.93 (0.89- 0,98), respectivamente. A AUC-ROC no *cross-validation* foi de 0.98 (0,96- 0,99) para *support vector machine*, 0.97 (0,95- 0,99) para *random forest* e 0.96 (0,93- 0,99) para *logistic regression*.

Os resultados indicam que os três modelos utilizados apresentaram resultados semelhantes visto que não houve significância estatísticas nas diferenças encontradas. O *cross-validation* com 10 *folds* demonstrou consistência e estabilidade nos resultados, com intervalos de confiança estreitos para os três modelos avaliados. O desvio padrão foi relativamente baixo, o que sugere que os modelos são robustos e apresentam pouca variação nos resultados quando avaliados em diferentes conjuntos de dados de treinamento e teste.

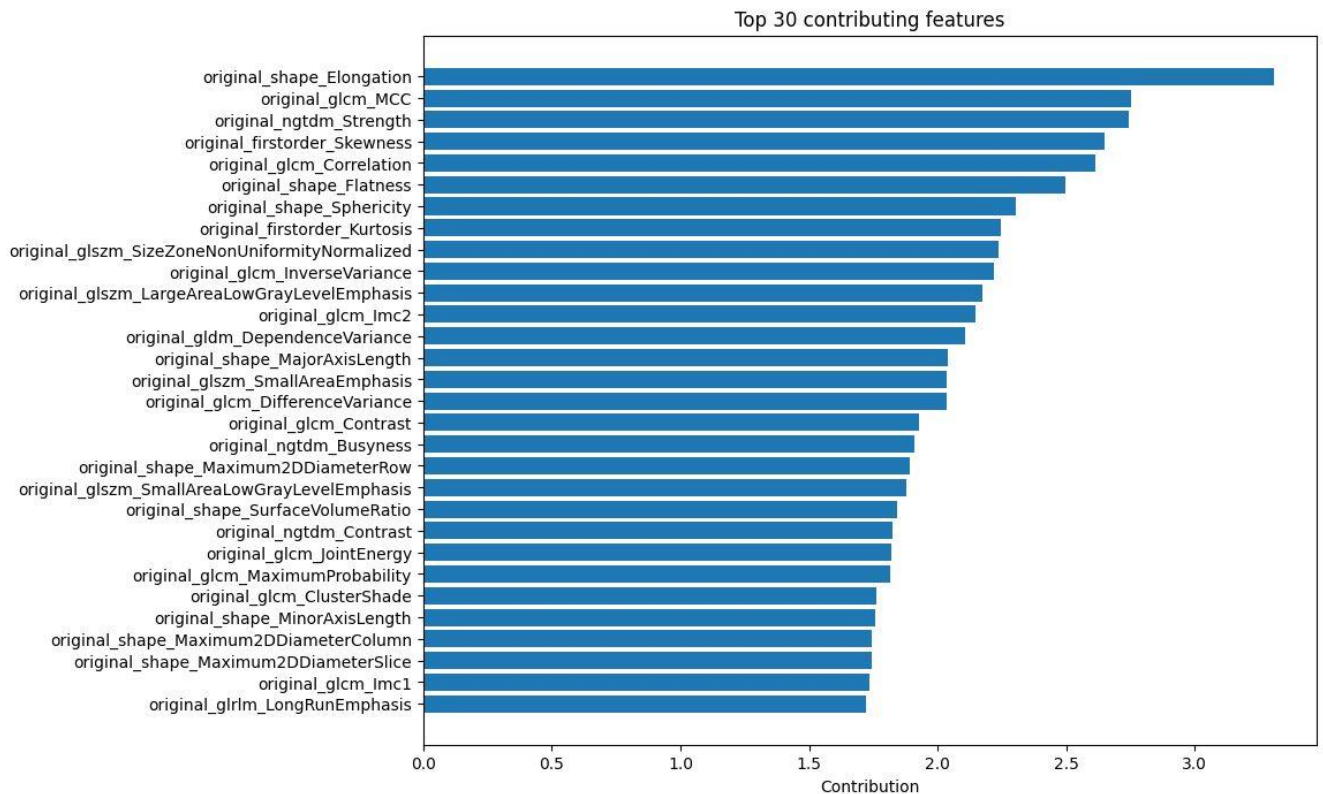
**Tabela 1**-Características dos pacientes do grupo 1

<b>Idade média</b>	64,9 (+/- 7,60) anos
<b>Método de coleta de material para estudo histopatológico</b>	Prostatectomia radical: 19 (23%) Biópsia transretal: 64 (77%)
<b>Grau histológico</b>	ISUP 1: 33 ISUP 2: 50
<b>Volume prostático médio</b>	38,7 (16-125) cm <sup>3</sup>
<b>PSA</b>	7,5 (1,5-33) ng/mL
<b>dPSA</b>	0,22 (0,03-0,87) ng/mL/cm <sup>3</sup>

**Tabela 2-**Características dos pacientes do grupo 2

<b>Idade média</b>	68,9 (+/- 8,62) anos
<b>Método de coleta de material para estudo histopatológico</b>	Prostatectomia radical: 20 (32%) Biópsia transretal: 43 (68 %)
<b>Grau histológico</b>	ISUP 3: 18 ISUP 4: 18 ISUP 5: 27
<b>Volume prostático médio</b>	42,0 (16-122) cm <sup>3</sup>
<b>PSA</b>	60,1 (1,1-997) ng/mL
<b>dPSA</b>	1,67 (0,03-29,3) ng/mL/cm <sup>3</sup>

Fonte: Arquivo próprio

**Figura 5-**Histograma mostrando a contribuição das características mais importantes na assinatura radiômica.

Fonte: Arquivo próprio

#### 4.1. **Support Vector Machine (SVM)**

No *validation test*, o *support vector machine* obteve uma precisão de 0,90. A precisão do modelo foi de 0,89, o recall foi de 0,89 e o F1-score foi de 0,89. A curva AUC-ROC foi de 0,90. No *cross-validation*, o esse modelo obteve acurácia média de 91 (desvio-padrão: 0.05, IC de 95%: [0.87- 0.94]), precisão de 0.92 (IC de 95%: [0,87- 0,98]) e a AUC-ROC de 0.98 (IC de 95%: [0,96- 0,99]). Os resultados desse modelo estão descritos nas tabelas 3 e 4.

**Tabela 3-**Tabela de resultados no *validation test* – *Support vector machine*

Parâmetro	Valor
Acurácia	0.90
Precisão	0.89
Recall	0.89
F1-score	0.89
AUC-ROC	0.90

**Tabela 4-**Tabela de resultados no *cross-validation* – *Support vector machine*

	Acurácia ( $\pm$ DP)	Precisão ( $\pm$ DP)	AUC-ROC ( $\pm$ DP)
SVM	0.91 ( $\pm$ 0.05)	0.92 ( $\pm$ 0.09)	0.98 ( $\pm$ 0.02)

SVM: *Support Vector Machine*; DP: desvio-padrão; AUC: *area under the curve*. ROC: *Receiver operator characteristic*

#### 4.2. **Random Forest (RF)**

No *validation test*, o *random forest* obteve uma precisão de 0,93. A precisão do modelo foi de 0,93, o recall foi de 0,84 e o F1-score foi de 0,91. A curva AUC-ROC foi de 0,92. No *cross-validation*, esse modelo obteve acurácia média de 91 (IC de 95%: [0.88- 0.95]), precisão de 0.98 (IC de 95%: [0,95- 1,01]) e a AUC-ROC de 0.97 (IC de 95%: [0,95- 0,99]). Os resultados desse modelo estão descritos nas tabelas 5 e 6.

**Tabela 5-**Tabela de resultados no *validation test* – *Random forest*

Parâmetro	Valor
Acurácia	0.93
Precisão	1.00
Recall	0.84
F1-score	0.91
AUC-ROC	0.92

**Tabela 6**-Tabela de resultados no *cross-validation* – *Random forest*

	Acurácia ( $\pm$ DP)	Precisão ( $\pm$ DP)	AUC-ROC ( $\pm$ DP)
<i>Random Forest</i>	0.91 ( $\pm$ 0.05)	0.98 ( $\pm$ 0.05)	0.97 ( $\pm$ 0.03)

DP: desvio-padrão; AUC: *area under the curve*. ROC: *Receiver operator characteristic*

### 4.3. *Logistic Regression* (LR)

No *validation test*, o *logistic regression* obteve uma precisão de 0,93. A precisão do modelo foi de 0,94, o recall foi de 0,89 e o F1-score foi de 0,91. A curva AUC-ROC foi de 0,92. No *cross-validation*, esse modelo obteve acurácia média de 0.93 (IC de 95%: [0.89- 0.98]), precisão de 0.96 (IC de 95%: [0,93- 0,99]) e a AUC-ROC de 0.96 (IC de 95%: [0,93- 0,99]). Os resultados desse modelo estão descritos nas tabelas 7 e 8.

**Tabela 7**-Tabela de resultados no *validation test* – *Logistic regression*

Parâmetro	Valor
Acurácia	0.93
Precisão	0.94
Recall	0.89
F1-score	0.91
AUC-ROC	0.92

**Tabela 8-**Tabela de resultados no *cross-validation* – *Logistic regression*

	Acurácia ( $\pm$ DP)	Precisão ( $\pm$ DP)	AUC-ROC ( $\pm$ DP)
<i>Logistic Regression</i>	0.93 ( $\pm$ 0.07)	0.96 ( $\pm$ 0.06)	0.96 ( $\pm$ 0.04)

DP: desvio-padrão; AUC: *area under the curve*. ROC: *Receiver operator characteristic*

#### 4.4. Comparação entre os modelos

Nas tabelas 9 e 10 e na figura 9, podemos observar a comparação entre os três modelos avaliados.

**Tabela 9-**Comparação entre resultados dos três modelos avaliados no *validation test*

	SVM	RF	LR
<b>Acurácia</b>	0.90	0.93	0.93
<b>Precisão</b>	0.89	1.00	0.94
<b>Recall</b>	0.89	0.84	0.89
<b>F1-score</b>	0.89	0.91	0.91
<b>AUC-ROC</b>	0.90	0.92	0.92

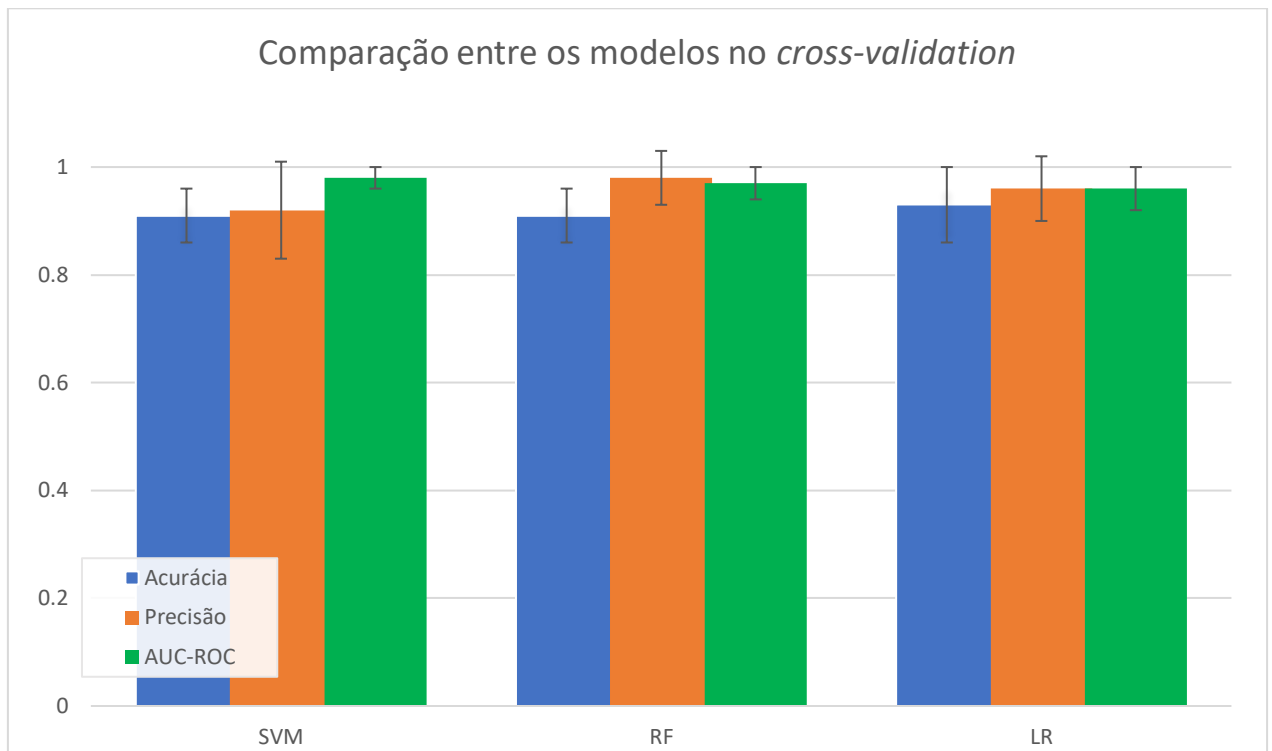
SVM: *Support Vector Machine*; RF: *Random Forest*; LR: *Logistic Regression*; AUC: *area under the curve*. ROC: *Receiver operator characteristic*

**Tabela 10-**Comparação entre resultados dos três modelos avaliados no *cross-validation*

	SVM( $\pm$ DP)	RF( $\pm$ DP)	LR( $\pm$ DP)
<b>Acurácia</b>	0.91( $\pm$ 0.05)	0,91( $\pm$ 0.05)	0,93( $\pm$ 0.07)
<b>Precisão</b>	0.92( $\pm$ 0.09)	0.98( $\pm$ 0.05)	0.96( $\pm$ 0.06)
<b>AUC-ROC</b>	0.98( $\pm$ 0.02)	0.97( $\pm$ 0.03)	0.96( $\pm$ 0.04)

SVM: *Support Vector Machine*; RF: *Random Forest*; LR: *Logistic Regression*; DP: desvio-padrão; AUC: *area under the curve*. ROC: *Receiver operator characteristic*

**Figura 6-** Comparação entre os modelos no *cross-validation*



SVM: *Support Vector Machine*; RF: *Random Forest*; LR: *Logistic Regression*; AUC: *area under the curve*. ROC: *Receiver operator characteristic*

Fonte: Arquivo próprio



## 5. DISCUSSÃO

Os modelos de *machine learning* para avaliação de características radiômicas na imagem por ressonância magnética são marcadores promissores da agressividade do câncer de próstata(20, 22, 27, 32). Os resultados aqui apresentados indicam que é possível utilizar características radiômicas extraídas de imagens médicas para a diferenciação de tumores de próstata de alto e de grau baixo/intermediário com alta acurácia e com desempenho consistente. Além disso, eles sugerem que a técnica de seleção de características é importante para aprimorar a eficiência dos modelos de classificação.

No nosso estudo, foram desenvolvidos, validados e comparados modelos de *machine learning* para prever o grau histológico do câncer de próstata na RM. Todos os modelos testados apresentaram alta acurácia, com resultados consistentes.

Os resultados encontrados estão de acordo com alguns estudos da literatura, como o de Bernatz, S. et al (20). Nesse estudo, com o objetivo de diferenciar o CaP clinicamente significativo do insignificante através da radiômica, foram selecionados 73 pacientes que foram submetidos à RM antes da prostatectomia ou da biópsia. Então, lesões de próstata na sequência ADC da RM foram segmentadas com o software *3D Slicer* e em seguida, foram extraídas características de imagem utilizando a biblioteca *PyRadiomics*. Como modelos de *machine learning*, utilizaram support vector machine (SVM), neural networks (NN), e random forest (RF), obtendo resultados relevantes na diferenciação de câncer clinicamente significativo, sendo o random forest o modelo que obteve maior precisão.

Jing et al. (32) desenvolveram e validaram um nomograma radiômico baseado em ressonância magnética para prever o câncer de próstata clinicamente significativo. 181 pacientes submetidos à RM pré-prostatectomia radical ou biópsia em três centros diferentes foram avaliados retrospectivamente, sendo 141 na coorte de treinamento e 60 na validação, ou seja, 78% e 22% respectivamente. Foram utilizados dois métodos diferentes de segmentação manual, incluindo a segmentação da próstata inteira em T2 e segmentação da lesão em DWI. As características radiômicas obtidas foram selecionadas para construir uma assinatura radiômica. Um nomograma final foi utilizado para avaliar o câncer de próstata clinicamente significativo, combinando a assinatura radiômica e o PI-RADS. Como resultado, esse nomograma superou a

avaliação subjetiva isolada de acordo com a análise ROC em todos os conjuntos de dados (todos com  $p < 0,05$ ).

Bonekamp et al.(27), em um estudo com 316 pacientes, sendo 183 na coorte de treinamento e 133 na coorte de teste, conseguiram mostrar que modelos de *machine learning* obtidos apenas no ADC melhoraram a diferenciação entre lesões prostáticas benignas e malignas. Nesse trabalho, foram segmentadas 462 lesões de próstata nas três sequências (T2, ADC e DWI) da RM. Então, com os dados extraídos através do *PyRadiomics*, modelos de *machine learning* foram desenvolvidos a partir tanto das três sequências como apenas do ADC. Esses foram comparados entre si e com a classificação PIRADS. O desempenho dos modelos foi significativamente melhor do que o PIRADS e foi igualmente bom entre o ADC e outras sequências.

No nosso trabalho, seguimos uma padronização de obtenção dos dados de imagem, selecionamos pacientes com idade, método de obtenção de fragmentos para estudo histopatológico e volume prostático semelhantes e utilizamos ferramentas de domínio público e já extensamente conhecidas na área para treinamento de máquina. Diante do exposto, a interpretação de exames de imagem de forma mais acurada e objetiva contribui para identificar, por meio de método não invasivo, CaP clinicamente significativo, possibilitando que procedimentos e terapêuticas mais agressivos sejam reservados para casos selecionados.

Ainda assim, a generalização dos nossos resultados ainda é incerta, e estudos de validação externa teriam que ser desenvolvidos, com equipamentos de RM distintos daqueles da nossa instituição e em diferentes populações. Enfatizamos a necessidade de cautela antes de aplicar diagnósticos assistidos por computador, pois o desempenho preditivo depende muito da seleção de características e algoritmos de *machine learning*, podendo até reduzir o desempenho da avaliação clínica. Modelos de previsão não invasivos podem ter o potencial de fazer parte de ferramentas de suporte à decisão para auxiliar os médicos na seleção de uma terapia adequada, mas precisamos ser cautelosos antes de sua implementação na rotina clínica. É importante realizar estudos futuros, com conjunto de dados maior, prospectivos e multicêntricos para validar e aprimorar os resultados apresentados neste estudo.

## 6. CONCLUSÃO

A utilização de radiômica pode diferenciar com precisão lesões de baixo/intermediário grau (ISUP 1 e 2) de lesões de alto grau (ISUP 3-5) e pode ser uma ferramenta auxiliar importante para a confirmação da elegibilidade de pacientes que consideram uma abordagem conservadora como manejo primário do câncer de próstata. Apesar de uma acurácia discretamente maior para *logistic regression*, todos métodos de *machine learning* foram equivalentes

## 7. REFERÊNCIAS BIBLIOGRÁFICAS

1. FERRO, M. et al. Radiomics in prostate cancer: an up-to-date review. **Therapeutic Advances in Urology**, v. 14, p. 1-37, jan. 2022.
2. SOCIEDADE BRASILEIRA DE UROLOGIA. Câncer de próstata. São Paulo(SP) SBU
3. Sociedade Brasileira de Urologia, Sociedade Brasileira de Medicina de Família e Comunidade. Projeto Diretrizes. Câncer de próstata: Diagnóstico. Brasília: Associação Médica Brasileira/Conselho Federal de Medicina; 2006.
4. HAMDY, F. C. et al. Fifteen-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Prostate Cancer. **New England Journal of Medicine**, 11 mar. 2023.
5. EPSTEIN, J. I. et al. The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. **The American Journal of Surgical Pathology**, v. 40, n. 2, p. 1, out. 2015.
6. Sociedade Brasileira de Patologia. Próstata – Carcinoma da próstata – (5ª edição – 2019) Disponível em: <https://www.sbp.org.br/manual-de-laudos-histopatologicos/prostata-carcinoma-prostata/>.
7. TURKBEY, B. et al. Prostate Imaging Reporting and Data System Version 2.1: 2019 Update of Prostate Imaging Reporting and Data System Version 2. **European Urology**, v. 76, n. 3, p. 340–351, 1 set. 2019.
8. **PI-RADS**. Disponível em: <<https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/PI-RADS>>.
9. SPOHN, S. K. B. et al. Radiomics in prostate cancer imaging for a personalized treatment approach - current aspects of methodology and a

- systematic review on validated studies. **Theranostics**, v. 11, n. 16, p. 8027–8042, 2021.
10. GILLIES, R. J.; KINAHAN, P. E.; HRICAK, H. Radiomics: Images Are More than Pictures, They Are Data. **Radiology**, v. 278, n. 2, p. 563–577, fev. 2016.
  11. EUROPEAN SOCIETY OF RADIOLOGY (ESR). What the radiologist should know about artificial intelligence – an ESR white paper. **Insights into Imaging**, v. 10, n. 1, 4 abr. 2019.
  12. VAN GRIETHUYSEN, J. J. et al. Computational Radiomics System to Decode the Radiographic Phenotype. **Cancer research**, v. 77, n. 21, p. e104–e107, 1 nov. 2017.
  13. D'AMICO, A. V. Biochemical Outcome After Radical Prostatectomy, External Beam Radiation Therapy, or Interstitial Radiation Therapy for Clinically Localized Prostate Cancer. **JAMA**, v. 280, n. 11, p. 969, 16 set. 1998.
  14. WILLEMSE, P.-P. M. et al. Systematic Review of Active Surveillance for Clinically Localised Prostate Cancer to Develop Recommendations Regarding Inclusion of Intermediate-risk Disease, Biopsy Characteristics at Inclusion and Monitoring, and Surveillance Repeat Biopsy Strategy. **European Urology**, v. 81, n. 4, p. 337–346, abr. 2022.
  15. VAN LEENDERS, G. J. L. H. et al. The 2019 International Society of Urological Pathology (ISUP) Consensus Conference on Grading of Prostatic Carcinoma. *American Journal of Surgical Pathology*, v. 44, n. 8, p. e87–e99, 26 maio 2020.
  16. ENIKEEV, D. et al. Active Surveillance for Intermediate-Risk Prostate Cancer: Systematic Review and Meta-analysis of Current Protocols and Outcomes. *Clinical Genitourinary Cancer*, v. 18, n. 6, p. e739–e753, 1 dez. 2020.

17. KUMAR, V. et al. Radiomics: the process and the challenges. **Magnetic Resonance Imaging**, v. 30, n. 9, p. 1234–1248, nov. 2012.
18. FEDOROV, A. et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. **Magnetic Resonance Imaging**, v. 30, n. 9, p. 1323–1341, nov. 2012.
19. BLEKER, J. et al. Multiparametric MRI and auto-fixed volume of interest-based radiomics signature for clinically significant peripheral zone prostate cancer. **European Radiology**, v. 30, n. 3, p. 1313–1324, 27 nov. 2019.
20. BERNATZ, S. et al. Comparison of machine learning algorithms to predict clinically significant prostate cancer of the peripheral zone with multiparametric MRI using clinical assessment categories and radiomic features. **European Radiology**, v. 30, n. 12, p. 6757–6769, 16 jul. 2020.
21. YAO, S.; JIANG, H.; SONG, B. Radiomics in prostate cancer: basic concepts and current state-of-the-art. **Chinese Journal of Academic Radiology**, v. 2, n. 3-4, p. 47–55, 12 nov. 2019.
22. YU, S. et al. Development and head-to-head comparison of machine-learning models to identify patients requiring prostate biopsy. **BMC Urology**, v. 21, n. 1, 16 maio 2021.
23. URRARO, F. et al. MRI Radiomics in Prostate Cancer: A Reliability Study. **Frontiers in Oncology**, v. 11, 21 dez. 2021.
24. JI, X. et al. Bi-parametric magnetic resonance imaging based radiomics for the identification of benign and malignant prostate lesions: cross-vendor validation. v. 44, n. 3, p. 745–754, 1 jun. 2021.
25. ZWANENBURG, A. et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. **Radiology**, v. 295, n. 2, p. 328–338, maio 2020.

26. YIP, S. S. F.; AERTS, H. J. W. L. Applications and limitations of radiomics. **Physics in Medicine and Biology**, v. 61, n. 13, p. R150–R166, 8 jun. 2016.
27. BONEKAMP, D. et al. Radiomic Machine Learning for Characterization of Prostate Lesions with MRI: Comparison to ADC Values. **Radiology**, v. 289, n. 1, p. 128–137, 31 jul. 2018.
28. KOCAK, B. et al. Radiomics with artificial intelligence: a practical guide for beginners. **Diagnostic and Interventional Radiology**, v. 25, n. 6, p. 485–495, 5 nov. 2019.
29. CUTAIA, G. et al. Radiomics and Prostate MRI: Current Role and Future Applications. **Journal of Imaging**, v. 7, n. 2, p. 34, 11 fev. 2021.
30. CHEN, T. et al. Prostate Cancer Differentiation and Aggressiveness: Assessment with a Radiomic-Based Model vs. PI-RADS v2. v. 49, n. 3, p. 875–884, 1 mar. 2019.
31. SMITH, C. P. et al. Radiomics and radiogenomics of prostate cancer. **Abdominal Radiology**, v. 44, n. 6, p. 2021–2029, 20 jun. 2018.
32. JING, G. et al. Prediction of clinically significant prostate cancer with a multimodal MRI-based radiomics nomogram. v. 12, 15 jul. 2022.
33. MATOSO, A.; EPSTEIN, J. I. Defining clinically significant prostate cancer on the basis of pathological findings. **Histopathology**, v. 74, n. 1, p. 135–145, 18 dez. 2018.
34. MOHLER, J. L. et al. Prostate Cancer, Version 2.2019, NCCN Clinical Practice Guidelines in Oncology. **Journal of the National Comprehensive Cancer Network**, v. 17, n. 5, p. 479–505, maio 2019.
35. MONGAN, J.; MOY, L.; KAHN, C. E. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers. **Radiology: Artificial Intelligence**, v. 2, n. 2, p. e200029, 1 mar. 2020.

36. FLAVELL, R. R. et al. Abnormal findings on multiparametric prostate magnetic resonance imaging predict subsequent biopsy upgrade in patients with low risk prostate cancer managed with active surveillance. **Abdominal Imaging**, v. 39, n. 5, p. 1027–1035, 17 abr. 2014.
37. SOMFORD, D. M. et al. Evaluation of Diffusion-Weighted MR Imaging at Inclusion in an Active Surveillance Protocol for Low-Risk Prostate Cancer. v. 48, n. 3, p. 152–157, 1 mar. 2013.
38. PARMAR, C. et al. Machine Learning methods for Quantitative Radiomic Biomarkers. **Scientific Reports**, v. 5, n. 1, 17 ago. 2015.