



UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS I
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS E TECNOLOGIA EM SAÚDE
MESTRADO PROFISSIONAL EM CIÊNCIAS E TECNOLOGIA EM SAÚDE

ALBERTO LADEIA DE QUEIROZ FILHO

COMPUTER-AIDED DIAGNOSIS (CAD) NA AVALIAÇÃO DE NÓDULOS
PROSTÁTICOS ATRAVÉS DA RESSONÂNCIA MAGNÉTICA
MULTIPARAMÉTRICA

CAMPINA GRANDE – PB

2021

ALBERTO LADEIA DE QUEIROZ FILHO

***COMPUTER-AIDED DIAGNOSIS (CAD) NA AVALIAÇÃO DE NÓDULOS
PROSTÁTICOS ATRAVÉS DA RESSONÂNCIA MAGNÉTICA
MULTIPARAMÉTRICA***

Dissertação apresentada ao Programa de Pós-Graduação em Ciência e Tecnologia em Saúde da Universidade Estadual da Paraíba, em cumprimento às exigências para obtenção do título de Mestre em Ciência e Tecnologia em Saúde.

Orientador: Prof. Dr. Robson Pequeno de Sousa

CAMPINA GRANDE – PB

2021

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

Q3c Queiroz Filho, Alberto Ladeia de.

Computer-aided diagnosis (CAD) na avaliação de nódulos prostáticos através da ressonância magnética multiparamétrica [manuscrito] / Alberto Ladeia de Queiroz Filho. -2021.

128 p. : il. colorido.

Digitado.

Dissertação (Mestrado em Profissional em Ciência e Tecnologia em Saúde) - Universidade Estadual da Paraíba, Pró-Reitoria de Pós-Graduação e Pesquisa, 2021.

"Orientação : Prof. Dr. Robson Pequeno de Sousa, Departamento de Computação - CCT."

1. Ressonância magnética. 2. Diagnóstico por imagem. 3. Câncer de próstata. 4. Inteligência artificial. I. Título

21. ed. CDD 616.075 48

ALBERTO LADEIA DE QUEIROZ FILHO

**COMPUTER-AIDED DIAGNOSIS (CAD) NA AVALIAÇÃO DE NÓDULOS
PROSTÁTICOS ATRAVÉS DA RESSONÂNCIA MAGNÉTICA
MULTIPARAMÉTRICA**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência e Tecnologia em Saúde da Universidade Estadual da Paraíba como requisito para obtenção do título de Mestre em Ciência e Tecnologia em Saúde.

Dissertação aprovada em: 14/07/2021

BANCA EXAMINADORA:



Prof. Dr. Robson Pequeno de Sousa
Universidade Estadual da Paraíba (UEPB)



Prof. Dra. Kátia Elizabete Galdino
Universidade Estadual da Paraíba (UEPB)



Prof. Dra. Evelinda Marramom Trindade
Universidade de São Paulo (USP)

RESUMO

Contexto: a Ressonância Magnética da Próstata é método reconhecido de diagnóstico do câncer prostático, sobretudo quando se utiliza protocolo multiparamétrico. Entretanto, a grande quantidade de imagens obtidas a serem avaliadas e a complexidade desta requer médicos radiologistas treinados e longo tempo de análise, onerando o estudo. Nesta conjuntura, as técnicas de auxílio computacional ao diagnóstico médico (*computer aided diagnostic* - CAD) surgem não somente como estratégia de apoio laboral, mas também qualificando características de imagem não apreciadas ao olho humano e que eventualmente podem ser determinantes na elucidação clínica. *Objetivo:* a pesquisa visa identificar o conhecimento publicado da aplicação do CAD na Ressonância Magnética Multiparamétrica da próstata, comparando-o com os dados de metanálise de validação do protocolo recomendado pelo PIRADS, visando detecção do câncer prostático. *Fonte de dados:* emprego de expressões previamente definidas nos mecanismos de busca nas plataformas MEDLINE/PUBMED, ScienceDirect, IEEE, Cochrane Library e Portal de Periódicos da CAPES, nas línguas inglesa e portuguesa. *Seleção dos estudos:* catalogação preliminar das publicações, aplicação de filtro de seleção das publicações relevantes (critérios de inclusão e exclusão) e filtro de qualidade, e resgate de dados relevantes, outrora excluídos pelos filtros precedentes. *Síntese dos dados e resultados:* 36 artigos foram selecionados no grupo CAD, categorizado quanto ao tipo do classificador (SVM, *random forest*, redes neurais e outros), e 10 no grupo PIRADS. As variáveis estatísticas disponíveis foram área sob a curva ROC, sensibilidade, especificidade, com médias no grupo CAD de, respectivamente, 0,849; 0,799 e 0,796. No grupo PIRADS observou-se médias semelhantes, sendo 0,838; 0,817 e 0,71. Aplicado teste de Mann-Whitney de amostras independentes, não sendo observada diferença estatística significativa entre os grupos. *Conclusão:* a detecção do câncer de próstata através do CAD aplicado na Ressonância Magnética Multiparamétrica é sustentada, não havendo discrepância apreciável quando comparado com o processo já estabelecido (PIRADS), porém com possíveis ganhos quanto ao tempo de análise e esforço médico.

Palavras-chave: Ressonância Magnética Multiparamétrica da Próstata. CAD. Câncer de Próstata. Inteligência artificial.

ABSTRACT

Context: Magnetic Resonance Imaging (MRI) of the Prostate is a recognized method for prostate cancer diagnosis, particularly when using a multiparametric protocol. However, to evaluate the large number of complex images requires trained radiologists and elevated amount of time for analysis, difficulting its application. In this regard, computational aid techniques for medical diagnosis emerged not only as a work support strategy, but also qualifying image characteristics that are not appraisable by the human eye. Eventually, such characteristics details may be decisive in clinical ascertainment. *Objective:* to identify the computer aided diagnostic (CAD) applied in Multiparametric MRI published knowledge for detection of prostate cancer, and, to compare it with the published meta-analysis data of PIRADS for its ascertainment. *Data source:* we used established expressions defined in MEDLINE/PUBMED, ScienceDirect, IEEE, Cochrane Library and CAPES Journal Portal to search bibliographic databases in English and Portuguese languages. *Steps for studies selection:* an evidence matrix was built with the relevant publications regarding the inclusion and exclusion criteria, as well as the quality filter. Duplicated publications or opinion articles were excluded by titles and abstracts review. Eligible studies were retrieved for full text lecture and evaluation. The evidence matrix titles were then indexed with the important study variables, and specific data was retrieved by interviewing each full text selected study. *Data synthesis and results:* 36 articles were selected in the CAD group, categorized according to the type of classifier (SVM, random forest, neural networks, amid others), and 10 studies in the PIRADS group. The available statistical variables were "area under the receiver operating" (ROC) curve, sensitivity, specificity. The respective CAD group means were 0.849; 0.799 and 0.796. In the PIRADS group, similar means were observed, with 0.838; 0.817 and 0.71 respective estimates. Mann-Whitney test of independent samples was applied, with no statistically significant difference being observed between groups. *Conclusion:* the detection of prostate cancer through CAD applied in Multiparametric Magnetic Resonance is of value, with no discrepancy when compared with the established scoring process (PIRADS), adding possible gains in terms of analyses time length and radiologists' effort.

Keywords: Multiparametric Prostate Magnetic Resonance Imaging. CAD. Prostate Cancer. Artificial Intelligence.

LISTA DE ILUSTRAÇÕES

Figura 1 – Classificação Gleason.....	24
Figura 2 – Exemplos de imagens ponderadas em T2, DWI e ADC.....	30
Figura 3 – Etapas radiômicas do CAD.....	35
Figura 4 – Curva ROC.....	54
Figura 5 – Evolução do processo de seleção de artigos para análise sistemática.....	60
Figura 6 – Ponderações de RM utilizadas nos estudos, por percentual de publicações encontradas.....	64
Figura 7 – Comparação entre os tipos de técnicas utilizadas pelos CAD nos estudos, em percentual.....	65
Figura 8 – Regiões de interesse da aplicação CAD na próstata, em percentual de artigo.....;	67
Figura 9 – Regiões de interesse da aplicação CAD na próstata, em percentual de amostras.....	67
Figura 10 – Tipos de classificadores CAD, em percentual.....	68
Figura 11 – Número de artigos publicados por ano de publicação.	69
Figura 12 – Número de artigos publicados por ano de publicação e por tipo de classificador.....	70
Figura 13 – Histograma da variável área sob a curva ROC para o CAD (ROC_CAD)	73
Figura 14 – Boxplot e Q-Q plot da variável área sob a curva ROC para o CAD (ROC_CAD).....	73
Figura 15 – Histograma das variáveis sensibilidade (SENS_CAD) e especificidade (ESPEC_CAD).....	74
Figura 16 – Boxplot e Q-Q plot das variáveis sensibilidade (SENS_CAD) e especificidade (ESPEC_CAD).....	75
Figura 17 – Boxplot das variáveis área sob a curva ROC (ROC_CAD), sensibilidade (SENS_CAD) e especificidade (ESPEC_CAD), categorizadas pelo tipo do classificador CAD.....	76
Figura 18 – Q-Q plot da variável área sob a curva ROC (ROC_CAD), categorizado pelo tipo do classificador CAD.....	78
Figura 19 – Q-Q plot da variável sensibilidade (SENS_CAD), categorizado pelo tipo do classificador CAD.....	79

Figura 20 – Q-Q plot da variável especificidade (ESPEC_CAD), categorizado pelo tipo do classificador CAD.....	80
Figura 21 – Histograma das variáveis sensibilidade (SENS_PIRADS) e especificidade (ESPEC_PIRADS) referente aos estudos de validação do PIRADS.....	81
Figura 22 – Boxplot e Q-Q plot da variável área sob a curva ROC dos estudos de validação do PIRADS (ROC_PIRADS).....	82
Figura 23 – Boxplot e Q-Q plot da variável sensibilidade dos estudos de validação do PIRADS (SENS_ PIRADS).....	83
Figura 24 – Boxplot e Q-Q plot da variável especificidade dos estudos de validação do PIRADS (ESPEC_ PIRADS).....	83
Figura 25 – Resultado do teste de hipótese da variável área sob a curva ROC.....	86
Figura 26 – Resultado do teste de hipótese da variável sensibilidade.....	87
Figura 27 – Resultado do teste de hipótese da variável especificidade.....	88
Figura 28 – Box-plot do teste de medianas de amostras independentes das variáveis área sob a curva ROC, sensibilidade e especificidade, por grupos CAD e PIRADS.....	89
Figura 29 – Box-plot do teste de medianas de amostras independentes da variável área sob a curva ROC, por grupos CAD e PIRADS, segundo o tipo do classificador CAD.....	91
Figura 30 – Box-plot do teste de medianas de amostras independentes da variável sensibilidade, por grupos CAD e PIRADS, segundo o tipo do classificador CAD.....	92
Figura 31 – Box-plot do teste de medianas de amostras independentes da variável especificidade, por grupos CAD e PIRADS, segundo o tipo do classificador CAD.....	93
Figura 32 – Box-plot do teste de hipótese de Kruskal-Wallis de amostras independentes, para a variável sensibilidade, dos grupos CAD, PIRADS e EXPERIMENTO COM RADIOLOGISTAS.....	94
Figura 33 – Exemplo de gráfico curva de característica de operação.....	96
Figura 34 – Ferramenta de extração de dados.....	121
Figura 35 – Ficha de avaliação radiológica das imagens de MR-MP da próstata utilizada no EXPERIMENTO COM RADIOLOGISTAS.....	122
Figura 36 – Exemplo de anatomia patológica de espécime de próstata oriundo de prostatectomia total na coloração hematoxilina-eosina, identificando-se lesões compatíveis com CaP.....	124

LISTA DE QUADROS

Quadro 1 – Hipóteses do estudo	19
Quadro 2 – Padrões da classificação de Gleason	24
Quadro 3 – Escore de Gleason <i>versus</i> Prognóstico	26
Quadro 4 – Categorização PIRADS (versão 2)	34

LISTA DE TABELAS

Tabela 1 – Tabela de contingência.....	49
Tabela 2 – Técnicas multiparamétricas utilizadas nos estudos.....	65
Tabela 3 – Origem do produto para validação do diagnóstico imaginológico.....	66
Tabela 4 – Origem do produto para validação do diagnóstico imaginológico, por número de amostras.	66
Tabela 5 – Tipo de classificador CAD por quantidade de estudos publicados.....	68
Tabela 6 – Tipo de classificador por número de amostra e por número de publicações	70
Tabela 7 – Métricas de desempenho empregadas, por número de amostra e por número de publicações.....	71
Tabela 8 – Número de publicações, por método estatístico empregado e tipo do classificador CAD.....	71
Tabela 9 – Resultados gerais das métricas de desempenho empregadas.....	72
Tabela 10 – Teste de normalidade das variáveis área sob a curva ROC, sensibilidade e especificidade.....	72
Tabela 11 – Resultados dos testes de normalidade, por tipo do classificador CAD.....	77
Tabela 12 – Resultados gerais das métricas de desempenho dos estudos de validação do PIRADS.	81
Tabela 13 – Teste de normalidade das variáveis área sob a curva ROC, sensibilidade e especificidade, dos estudos de validação do PIRADS.	84
Tabela 14 – Experimento com radiologistas: detecção do CaP em imagens de RM-MP com validação pelo padrão-ouro (anatomia patológica).....	85
Tabela 15 – Resultado do teste de hipótese das variáveis área sob a curva ROC, sensibilidade e especificidade entre os grupos CAD e PIRADS, segundo tipo de classificador.....	90
Tabela 16 – Resultado do teste de hipótese de Kruskal-Wallis da variável sensibilidade entre os grupos CAD, PIRADS e EXPERIMENTO COM RADIOLOGISTAS.	94
Tabela 17 – Características dos artigos selecionados.....	108
Tabela 18 – Ponderações de RM utilizadas nos estudos.....	111
Tabela 19 – Região de análise do CAD por quantidade de estudos publicados.....	111
Tabela 20 – Região de análise do CAD por quantidade de estudos publicados.....	112

Tabela 21 – Número de artigos (estudos) publicados e número de amostras por ano de publicação do estudo, e seus correspondentes percentuais.....	113
Tabela 22 – Número de artigos publicados por ano de publicação e por tipo de classificador.....	114
Tabela 23 – Resultados dos classificadores na detecção do CaP.....	117
Tabela 24 – Resultados das métricas de desempenho empregadas, por tipo do classificador CAD.....	119
Tabela 25 – Resultados de estudos e metanálises para validação do PIRADS.....	120

LISTA DE ABREVIATURAS E SIGLAS

3D	Imagens tridimensionais
ACR	Colégio Americano de Radiologia – American College of Radiology
ADC	Coeficiente de difusão aparente
ANVISA	Agência Nacional de Vigilância Sanitária
ATS	Avaliações de Tecnologias em Saúde
BIRADS	<i>Breast Imaging – Reporting and Data System</i>
CAD	<i>Computer-Aided Diagnosis</i> – Diagnóstico assistido por computador
CaP	Câncer de próstata
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior do Ministério da Educação
CE	Critérios de exclusão
CI	Critérios de inclusão
DCE	<i>Dynamic contrast-enhanced</i>
DICOM	<i>Digital Imaging and Communications in Medicine</i> (Comunicação de Imagens Digitais em Medicina)
DWI	<i>Diffusion weighted images</i>
ESUR	European Society of Urogenital Radiology – Sociedade Européia de Radiologia Urogenital
FOV	<i>Field of view</i>
HPB	Hiperplasia Prostática Benigna
MRS	<i>MR spectroscopy</i> – espectroscopia de prótons por Ressonância Magnética
OR	<i>odds ratio</i>
PIRADS	Ou ainda PI-RADS – Prostate Imaging Reporting and Data System

PRF	Pulsos de radiofrequência
PSA	Antígeno prostático específico
RM	Ressonância Magnética
RM-MP	Ressonância Magnética Multiparamétrica
ROC	Área sob a curva
ROI	<i>Region of interest</i> – região de interesse
RR	Risco relativo
SVM	<i>Support vector machines</i>
T1	Tempo de magnetização longitudinal
T2	Perda da intensidade de magnetização transversal
T2W	Imagens ponderadas em T2
TE	Tempos de eco
TR	Tempo de repetição
US	Ultrassonografia
USTR	Ultrassonografia transretal

LISTA DE SÍMBOLOS

T Tesla

K^{trans} Constante de transferência de volume

SUMÁRIO

1	INTRODUÇÃO	14
1.1.	Escopo	14
1.2.	Epidemiologia	14
1.3.	Justificativa	16
1.4.	Objetivos geral e específicos	17
2	FUNDAMENTAÇÃO TEÓRICA	19
2.1.	Câncer de próstata: aspectos clínicos e diagnóstico	19
2.2.	Método de diagnóstico por imagem por ressonância magnética	26
2.3.	Computer-aided diagnosis	33
3	METODOLOGIA	42
3.1.	Aspectos gerais	42
3.2.	Avaliação do risco de viés	46
3.3.	Experimento com radiologistas	47
3.4.	Ferramentas de análise estatística	49
4	RESULTADOS E DISCUSSÃO	58
4.1.	Dados coletados	58
4.2.	Descrição das variáveis de interesse	61
4.3.	CAD	64
4.4.	PIRADS	80
4.5.	Resultado do experimento com radiologistas	84
4.6.	Análise comparativa CAD, PIRADS e radiologistas	86
5	CONCLUSÕES E POSSIBILIDADES DE DESENVOLVIMENTO	97
	REFERÊNCIAS	100
	APÊNDICES	107
	ANEXO	123

1. INTRODUÇÃO

1.1. Escopo

A Ressonância Magnética (RM) é um método de diagnóstico por imagem, não invasivo, que utiliza ondas de radiofrequência em meio a um campo magnético para interagir com prótons de hidrogênio do corpo humano, portanto sem radiação ionizante. A técnica multiparamétrica (RM multiparamétrica ou RM-MP) emprega protocolos bem estabelecidos para detecção precoce do câncer de próstata, que é a segunda causa de morte por neoplasia na população masculina [1], e, em se tratando de rastreamento, já se tem conhecimento há mais de 10 anos que supera em acurácia a dosagem dos níveis séricos de antígeno prostático específico (PSA) e o toque digital retal [2].

Dependendo da correlação clínica, a lesão suspeita para neoplasia detectada em um exame de RM-MP requer confirmação histológica, comumente pela aquisição de amostras do tecido prostático em biópsias guiadas, por exemplo, através ultrassonografia transretal [3].

A despeito da relativamente rápida realização do exame, a análise dos dados obtidos em uma RM-MP é trabalhosa e demorada, requerendo treinamento específico dos médicos radiologistas, com longa curva de aprendizagem [4]. É neste contexto que os sistemas de diagnóstico médico assistidos por computador (CAD) poderiam apoiar o radiologista no processo de identificação e caracterização das neoplasias de próstata através do estudo das imagens de RM-MP, reduzindo falsos negativos e indicações desnecessárias de biópsias [5].

1.2. Epidemiologia

Sendo o câncer de próstata (CaP) a terceira causa de morte masculina nos Estados Unidos (com 26.000 óbitos ao ano), também é a neoplasia mais frequente, com mais de 160.000 novos casos [1]. Globalmente, corresponde ao terceiro câncer mais frequentemente diagnosticado e a quinta principal causa de morte por neoplasia em homens [6], com incidência de 1,4 milhão de novos casos e 375.000 óbitos no ano de 2020. A incidência varia mais de 25 vezes em todo o mundo, sendo as maiores taxas registradas principalmente nos países desenvolvidos, em parte devido à ampla utilização de métodos diagnósticos que detectam tumores clinicamente importantes, bem como lesões de crescimento lento, mas que poderiam, de outra forma, escapar do diagnóstico [6].

Os dados estatísticos que suportam a incidência do CaP na população brasileira advêm dos serviços públicos de referência no tratamento, nem sempre contemplando os indivíduos usuários do sistema de saúde complementar. Ainda assim, segundo o *site* [7] do Instituto Nacional do Câncer – INCA, no ano de 2020 estima-se 65.840 novos casos, sendo que no ano anterior (2019) ocorreram 15.983 óbitos.

Na Bahia, segundo os dados do Ministério da Saúde [8], para o período de 2013 a 2019 (ano mais recente com dados disponíveis), o CaP foi a principal causa de óbito por câncer, sendo responsável por 19,5% dos casos (seguido do câncer pulmonar e do estômago, com 9,9% e 6,5%, respectivamente). Tal informação contrasta com a média nacional, na qual o principal sítio de morte por câncer é pulmão (14,0%) e a segunda é a próstata (13,3%). Analisando pelas regiões, no Nordeste o CaP aparece como primeiro (17,5%) seguido do pulmão (11,8%), enquanto que na região Sudeste o sítio pulmonar é o principal (13,6%), seguido pelo sítio prostático (11,9%), o que pode ser atribuído a diferentes acessos ao diagnóstico precoce, mudando a sobrevida e o desfecho da doença, mas também ao conhecido consumo tabágico masculino maior nas regiões Sul e Sudeste, relacionado ao desenvolvimento do câncer de pulmão [9].

Pelo menos desde o final do século passado e o início do atual, o rastreamento de indivíduos sadios (sem suspeita clínica de CaP) tem favorecido a descoberta de tumores ocultos nos países desenvolvidos que adotam essa prática na população em geral, o que favorece ocorrência de taxas de incidências mais altas nestes [10]. Os mais frequentes tipos de câncer (próstata, cólon-retal, mama feminina e pulmão) são até 5 vezes maiores nos países desenvolvidos em comparação com os países em desenvolvimento, em parte por diferentes perfis populacionais, tais como a evidência de maior casos avançados de CaP em obesos [11]. Embora possa ser atribuído a fatores de riscos desiguais, desiguais práticas diagnósticas devem também ser consideradas [12].

Outro fator, já bem conhecido há vários anos e que tem elevado a incidência de CaP nos países desenvolvidos, é a maior utilização de terapêutica cirúrgica no tratamento de condições benignas, como por exemplo, ressecção transuretral em estados de hiperplasia prostática naturais do envelhecimento, em que a aquisição de espécimes anatomopatológicos inesperadamente revela tumor [13]. De forma semelhante, o mesmo estudo implica ao refinamento das técnicas de biópsia também como fator envolvido no aumento da incidência de CaP nestes países. Pesquisas mais recentes [14] já mostram o benefício do surgimento de

métodos de imagem mais sensíveis que possibilitam precisão diagnóstica e otimização das estratégias de vigilância e terapêutica personalizada direcionada às alterações genômicas.

Globalmente, em revisão de 2019, mais de 60% dos CaP são diagnosticados em indivíduos idosos (maiores de 65 anos) [15], sendo este o seu principal fator de risco. História familiar de CaP está implicada no maior risco da doença, sendo estimado que de 10 a 15% dos pacientes tenham algum parente igualmente afetado. Indivíduos sadios parentes de primeiro grau de pacientes com CaP têm de 2 a 3 vezes mais chances de desenvolver a doença, que aumenta com a quantidade de casos familiares, porém reduz com a idade em que foi feito o diagnóstico do caso familiar afetado.

Há, ainda, risco relativo à etnia, sendo maior em afrodescendentes, e a relação com influência dietética, o que favoreceria a menor ocorrência de casos na Ásia, onde o consumo de fito-estrógenos é maior [15].

O diagnóstico do CaP antes dos 50 anos é muito raro (menos de 0,1% de todos os casos) [15]. A idade média dos pacientes é estimada entre 72 e 74 anos, sendo que 85% têm o diagnóstico após os 65 anos. O CaP latente (tumor desconhecido e sem apresentação clínica) é condição comum e há muito tempo conhecida: em estudos de autópsia, já era encontrado em mais de 75% dos homens com idade superior a 85 anos [16].

1.3. Justificativa

Um método de auxílio diagnóstico, economicamente viável e universalmente disponível, que favoreça o rastreamento do CaP e, portanto, seu diagnóstico precoce, no âmbito da saúde pública assistiria na redução das desigualdades regionais e sociais de acesso ao tratamento, aumentando a sobrevida da população. Embora a RM-MP não detenha todos estes requisitos, sobretudo devido aos elevados custos de implantação, manutenção e operação, qualquer tecnologia que simplifique o processo, promova facilidade de uso e agregue maior sensibilidade e especificidade ao exame será de importância nesta mitigação. As técnicas CAD despontam, portanto, como ferramenta neste sentido, com ampla aplicabilidade nas diversas áreas da saúde coletiva. Especificamente no CaP, o CAD favoreceria o acesso a diagnósticos mais precisos e rápidos, talvez menos onerosos, quando inserido dentro de protocolos estabelecidos e políticas sociais.

1.4. Objetivos geral e específicos

Estudos têm mostrado resultados valiosos do CAD como ferramenta auxiliar no diagnóstico e estadiamento do câncer, inclusive de próstata, mas a aplicabilidade na prática clínica ainda requer otimização dos processos de fluxos de trabalho, validação em casos reais e regulamentação [17] [18]. Como não há protocolo uniforme, seja no processamento da imagem ou no algoritmo computacional utilizado, faz-se necessário estimar quais técnicas de CAD, que estão sendo aplicadas nos diversos centros de pesquisas, são adequadas no diagnóstico de neoplasia prostática através da RM-MP. Neste caso, uma Revisão Sistemática dos dados de literatura poderia sintetizar os conhecimentos outrora publicados quanto a adequação do CAD na RM-MP como auxiliar na detecção do câncer de próstata e compará-lo com a acurácia de métodos bem estabelecidos (por exemplo, o estudo anátomo-patológico do espécime prostático) e de técnicas e protocolos não-CAD já testados e validados pela comunidade científica (como o PI-RADS, acrônimo para *Prostate Imaging Reporting and Data System*, que são padrões clínicos de alta qualidade na aquisição de imagens de RM-MP e de elaboração de relatórios estruturados com categorização de achados baseadas em escalas de Likert).

Especificamente na questão “análise” das imagens e informações adquiridas, são esperados os seguintes benefícios do CAD aplicado na RM-MP da próstata:

1. menor esforço na avaliação do grande volume de dados do estudo;
2. menor tempo de análise, detecção e caracterização das lesões nos exames;
3. menor variabilidade no diagnóstico, diminuindo as taxas de falso-negativo;
4. e melhorar significativamente o desempenho de radiologistas menos experientes no diagnóstico de câncer de próstata.

Quando se detém ao fato de que o câncer de próstata é a segunda neoplasia masculina mais diagnosticada no mundo, tais benefícios no processamento do exame (especificamente a fase de análise dos dados / interpretação de imagens) promovem impactos no contexto da saúde do paciente e na saúde coletiva, sobretudo:

1. menor tempo no resultado do exame;
2. maior segurança no diagnóstico;

3. redução do custo do estudo (menor tempo dispendido pelo médico radiologista), isto se a despesa do uso da ferramenta CAD não sobrepujar o eventual desconto oriundo da sua aplicação.

Logo, diante da variabilidade dos protocolos CAD aplicados em RM-MP na detecção do CaP e da expectativa dos benefícios do seu resultado, será adequado utilizá-lo neste contexto? Para responder tal pergunta, definiu-se como objetivo do estudo catalogar os artigos completos publicados (científicos em periódicos), entre janeiro de 2011 a abril de 2021, sobre a aplicação do CAD no diagnóstico de câncer de próstata através da RM-MP, com o intuito de realizar metanálise para sintetizar os conhecimentos publicados sobre a técnica. Portanto, como principal hipótese a ser testada, temos (quadro 1):

Quadro 1

Hipóteses do estudo	
H0	Não é factível o diagnóstico de câncer de próstata através do CAD em RM-MP.
H1	É factível o diagnóstico de câncer de próstata através do CAD em RM-MP.

* para o nível de 5% de significância estatística.

Fonte: o autor, 2021

Para fundamentá-la, é necessário verificar também se, em sendo factível, qual o impacto do método quanto a sensibilidade e especificidade, quais as técnicas de processamento de imagem com melhor acurácia e quais os algoritmos computacionais com melhor performance na detecção do câncer de próstata, sendo estes elencados como objetivos secundários do estudo.

Três cenários foram desenhados para serem utilizados como forma de comparação na análise dos quesitos acima:

CENÁRIO 1: CAD x anatomia patológica

CENÁRIO 2: CAD x PIRADS

CENÁRIO 3: PIRADS x anatomia patológica.

2. FUNDAMENTAÇÃO TEÓRICA

2.1. Câncer de próstata: Aspectos clínicos e diagnóstico

O diagnóstico do CaP ainda é um dilema, alvo de discussões e protocolos assistenciais em constantes mudanças e atualizações, quando do surgimento de novos conhecimentos.

Clinicamente, a apresentação do CaP é variada, com idade média de diagnóstico estimada em 68 anos, inúmeras vezes assintomática e sendo somente detectada em exames de rastreamento [19]. Quando seu crescimento é suficiente para exercer efeito de massa, pode ser observada obstrução ao fluxo urinário miccional. Hematúria e hematoespermia estão implicadas no comprometimento uretral / vesical e de vesículas seminais / ductos ejaculatórios, sugerindo extensão da lesão prostática além dos limites da glândula.

O rastreamento do CaP é realizado através da dosagem dos níveis séricos de antígeno prostático específico (PSA) e do toque digital retal, ambos com acurácia limitada. O diagnóstico definitivo advém de biópsias de próstata guiadas pela ultrassonografia transretal (USTR), cuja avaliação histopatológica fornece informações sobre o prognóstico da doença, através do escore de Gleason (classificação com forte correlação clínico-histológica). Entretanto, biópsias negativas não são infrequentes, já que menos de 1% da próstata é amostrada [12] [3].

O PSA é uma glicoproteína produzida pelas células do epitélio prostático. Por estar alterada em diversas condições (como na prostatite, no trauma e na hiperplasia prostática benigna), não pode ser utilizado como um marcador específico do CaP. É encontrada no soro na forma livre (PSA livre) e ligada a outras proteínas. Embora os níveis séricos do PSA aumentem com a presença do CaP, não é possível definir um valor exato para distinguir a lesão maligna das patologias benignas e, inclusive, há incidência do CaP em indivíduos com baixos níveis. Além disso, seus valores basais não se mantêm constante ao longo da vida, sendo maior com o aumento da idade. A Sociedade Brasileira de Urologia indica que:

- indivíduos com níveis séricos de PSA acima de 10 ng/mL e que não tenham clinicamente justificativa para seu aumento (por exemplo, prostatite aguda), devem ser investigados para CaP e a biópsia é recomendada;
- nos indivíduos com PSA menor que 2,5 ng/mL, a biópsia estaria indicada em situações especiais (como, por exemplo, toque retal alterado ou nódulo palpável);
- porém, nos indivíduos cujo nível sérico de PSA esteja entre 2,5 ng/mL e 10 ng/mL, outros parâmetros clínicos devem ser considerados antes de se indicar uma biópsia

desnecessária, como a velocidade de elevação deste PSA entre coletas sequenciais, a densidade do PSA (sua relação com o volume prostático estimado pela ultrassonografia transretal), e o PSA ajustado pela idade.

A relação PSA livre e o PSA total potencializa o efeito diagnóstico de CaP quando se compara com a utilização isolada do PSA, e é inversamente proporcional a agressividade do tumor e das dimensões prostática e tumoral [20].

O toque retal é considerado indispensável na avaliação do risco individual do CaP [21]. Requer profissionais experientes na técnica, com sensibilidade e especificidade de até 59% e 92%, respectivamente. Há concordância diagnóstica entre o toque retal e resultados positivos para CaP em prostatectomia, em 50% dos casos. Entretanto, há baixa adesão do toque retal durante a prática clínica (somente 22,3% dos pacientes se submetem ao toque retal durante as consultas urológicas) [20].

Dos métodos de diagnóstico por imagem disponíveis para a avaliação prostática, constitui-se a Ultrassonografia (US) e a Ressonância Magnética (RM) os mais utilizados.

A via suprapúbica da ultrassonografia da próstata é limitada na avaliação do parênquima, meramente fornece informação sobre sua dimensão, contudo é útil na determinação do resíduo urinário vesical pós-miccional, sobretudo nos paciente com hiperplasia prostática benigna e sintomas obstrutivos do trato urinário inferior, situação na qual o aumento das dimensões da glândula (geralmente por hiperplasia e não por neoplasia) promove dificuldade para o adequado esvaziamento da bexiga durante a micção.

Embora a ultrassonografia pela via retal (US transretal ou USTR) permita melhor caracterização da textura parenquimatosa prostática, não é observado significativo aumento da performance do método na detecção do CaP: enquanto a US por via suprapúbica apresenta questionável sensibilidade de 72,7% e especificidade de 72,2%, fornecendo $RV+ = 2,57$ (IC 95% de 1,84 a 3,60), a USTR oferece sensibilidade de 48% e especificidade de até 81% dos casos, com $RV+ = 2,53$ (IC 95% de 1,61 a 3,97) [20]. A USTR é valiosa no acompanhamento seriado do aumento das dimensões prostática, que é uma variável implicada na suspeição do CaP, mas também é útil como ferramenta no auxílio de biópsia, orientando o local adequado para o procedimento.

O custo elevado do método e sua disponibilidade limitada, quando comparado com a US, desfavorecem a aplicabilidade da RM como ferramenta de rastreamento. Utilizando-se protocolos específicos, como a técnica Difusão, sua sensibilidade e a especificidade são

estimadas em 72% e 81%, respectivamente, com $RV+ = 3,79$ (IC 95% de 2,48 a 5,78), portanto superior aos demais métodos. A utilização rotineira de técnica Difusão (*diffusion-weighted imaging* – DWI, ou Imagem por Ressonância Magnética ponderada em Difusão) aliado as imagens ponderadas em T2 e perfusionais, permite predizer a chance de uma lesão prostática se tratar de CaP, aplicando-se categorização de risco através do sistema de padronização PIRADS (*Prostate Imaging – Reporting and Data System*).

Como o comportamento clínico do câncer de próstata é variável, desde tumores indolentes até lesões altamente neoplásicas (portanto com agressividade local – extensão extraprostática – e com possibilidade disseminação a distância – metástases ganglionares e ósseas, por exemplo), é necessário estabelecer quais pacientes são eletivos para investigação e terapêutica invasiva. Ao passo que oferecer tratamento agressivo (como a cirurgia e radioterapia) em neoplasias de baixo escore Gleason pode resultar em danos aos pacientes (perda do controle miccional, disfunção erétil etc.), deixar de tratar energeticamente lesões de comportamento agressivo também compromete a sobrevida do paciente, sendo preciso adequado diagnóstico precoce e manejo. Alia-se a detecção cada vez mais frequente de CaP em paciente mais jovens e em estágio inicial, embora destes, 5% a 15% já tenham diagnóstico de doença localmente avançada [20].

Anatomicamente, a próstata foi dividida por McNeal [22] em três regiões principais, a zona periférica (região mais próxima do reto e que compõe cerca de 70% do parênquima prostático), a zona central (relacionada ao ducto ejaculatório e compondo 25% da massa glandular) e a zona de transição (situada na porção ventral glândula, em contato o estroma fibromuscular e com este representando os 5% restantes da glândula), todas elas com diferenças biológicas e histológicas. Nas técnicas de imagem, a identificação da zona periférica é clara tanto na US quanto na RM. Entretanto, os limites entre as zonas de transição e central são difíceis de se estabelecer, sendo comumente ambas tratadas como uma única região, denominada glândula interna. A zona periférica é a sede principal dos carcinomas da próstata, sendo encontrado de 70% a 80% dos casos, seguido de 20% a 30% da zona de transição [23]. A zona central é relativamente resistente ao CaP, sendo raramente encontrado nesta região. A zona transicional é comumente relacionada a Hiperplasia Prostática Benigna (HPB), condição na qual o aumento glandular leva aos sintomas urinários inferiores pelo efeito compressivo uretral. Também há diferença quanto a agressividade e prognóstico do CaP e sua localização, sendo pior na zona periférica [24]. No exame clínico, a localização frequente do CaP na zona

periférica favorece sua detecção pelo toque retal, enquanto dificulta quando situado na zona de transição.

A biópsia da próstata é o meio pelo qual se espera estabelecer, em definitivo, o diagnóstico do CaP. Usualmente é realizada por via transretal, guiada pela ultrassonografia, embora também possa ser guiada pela RM, com superioridade desta última na detecção das lesões na zona de transição (ZT). O momento adequado para sua indicação ainda é controverso entre os diversos estudos, porém, no Brasil, a Sociedade Brasileira de Urologia recomenda quando:

- o toque retal for alterado;
- o PSA for maior que 10 ng/mL;
- a relação entre o PSA livre e total < 10% nos indivíduos com PSA entre 4 e 10 ng/mL;
- o PSA estiver entre 2,5 e 4 ng/mL em paciente de risco ou jovens (considerando a velocidade de elevação) [20].

É mandatório a utilização de profilaxia antibacteriana antes do procedimento em todos os indivíduos, haja visto a agulha atravessar a parede anterior do reto antes de adentrar o parênquima prostático, podendo, portanto, carrear bactérias da flora intestinal. Recomenda-se ciprofloxacina ou levofloxacina, via oral, na dosagem de 500 mg, administrada uma hora antes do evento. Deve-se manter a profilaxia por dois dias (com uma dose a cada 24 horas) nos pacientes imunodeprimidos, diabéticos, em uso recente de imunossuppressores (corticóides) e naqueles com disfunção obstrutiva urinária grave ou com próstata com peso maior que 75 g. Não é um procedimento livre de complicações, sendo as imediatas mais frequentes a hematúria (62%) e o sangramento retal (2,1%), embora tardiamente possa ocorrer hematoespermia (9,8%), disúria persistente (7,2%), febre (2,9%), infecção (2,5%), prostatite aguda (1,8%) e sepses de origem urológica (0,1%) [20]. Esta última ainda é maior nos casos de ré-biopsia, quando há necessidade de repetir o estudo devido a resultado negativo [3].

Vários modelos foram preconizados quanto a técnica adequada para a biópsia de próstata, sendo a sextante inicialmente proposta, na qual três fragmentos são obtidos dos terços apical, médio e basal de cada lado da glândula. Outros modelos foram desenvolvidos, com amostragens estendidas para 12, 18 ou até mesmo 24 fragmentos, sendo a sextante com 12 amostras amplamente aceita. Embora a maior amostragem favoreça o diagnóstico de CaP, deve se considerar também o aumento das complicações advindas do procedimento e o número

elevado de resultados negativos, já que menos de 1% da próstata é amostrada, implicando em necessidade de rebiópsia. Direcionar o procedimento para lesões suspeitas elencadas por métodos de imagem tem mostrado resultados promissores na detecção de CaP mais agressivos, sobretudo quando se utiliza a fusão de imagens de RM com a ultrassonografia em tempo real no momento do procedimento, reduzindo no futuro a quantidade de amostras necessárias e evitando as complicações (sepse) de uma rebiópsia. Resultados de estudos randomizados multicêntricos de biópsia da glândula prostática guiada por fusão de imagens são esperados para que essa técnica seja aceita como substituta ao padrão sextante [3].

Para facilitar o entendimento e padronizar a descrição anatomopatológica do CaP, em 1966 Donald Gleason e colaboradores desenvolveram um sistema de pontuação [25], que posteriormente foi homenageado com seu nome. É baseada no aspecto arquitetural microscópico do tumor na amostra, que é obtida por biópsia ou produto de prostatectomia (parcial ou total), e apresenta constante evolução, sendo as últimas revisões em 2005, 2014 e 2017, voltadas sobretudo para “promover uniformidade na notificação do grau do câncer de próstata e no seu gerenciamento contemporâneo” [26]. A classificação é obtida através da pontuação em graus das seguintes características (quadro 2):

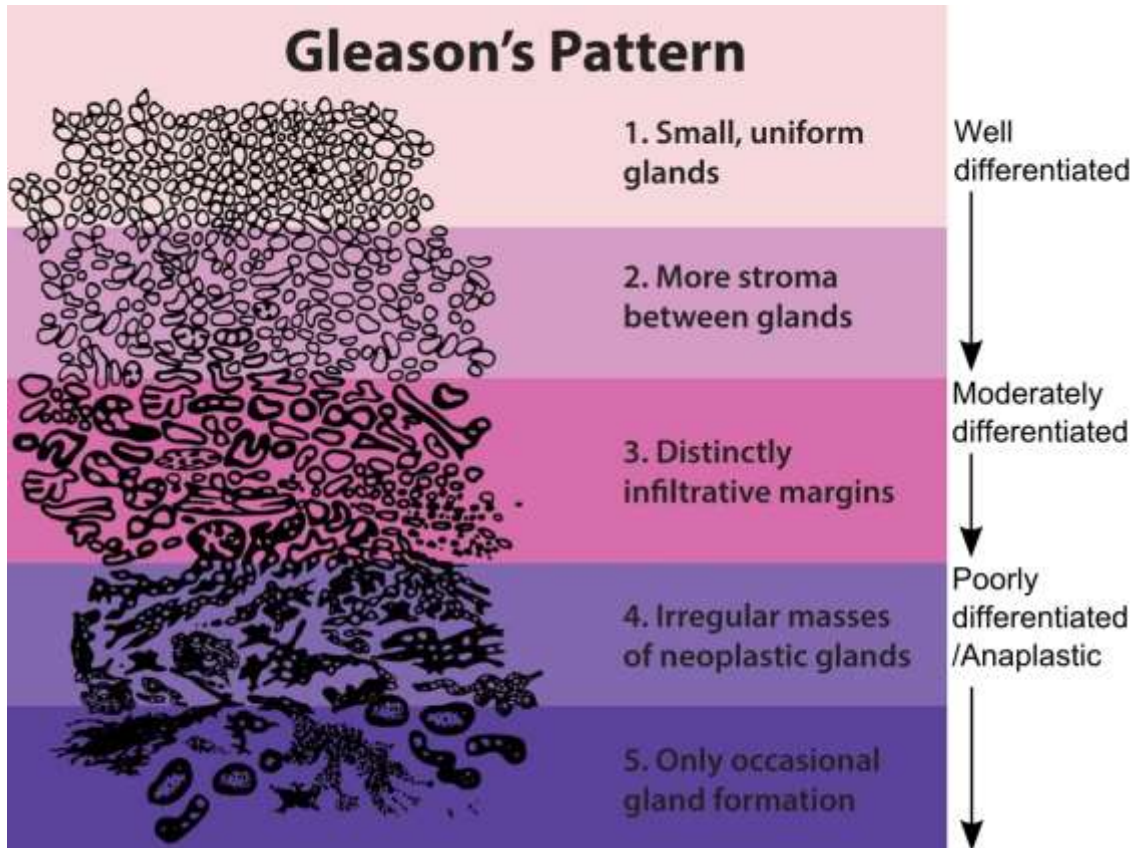
Quadro 2

Padrões da classificação de Gleason	
Aspecto microscópico do CaP segundo a classificação Gleason	Padrões
A amostra se parece muito com o tecido prostático normal, sendo as glândulas prostáticas pequenas, bem formadas e muito próximas uma das outras.	1
Embora a amostra ainda se assemelhe ao tecido normal, as células glandulares são maiores e há maior estroma (tecido conectivo) entre elas.	2
A amostra ainda apresenta células prostáticas reconhecíveis, mas já são mais escuras que o habitual e há sinais de extensão extracelular com invasão do estroma circunjacente.	3
A amostra possui poucas glândulas reconhecíveis como normais e há muitas células neoplásicas invadindo o tecido estromal.	4
A amostra não exhibe células normais reconhecíveis, somente tecido neoplásico.	5

Fonte: Epstein, 2017 [26].

Como se pode observar na figura 1, que exemplifica a padronização Gleason, quanto maior o grau de desdiferenciação, maior o aspecto anaplásico do tumor, portanto maior a sua agressividade.

Figura 1 – Classificação Gleason



Fonte: National Institutes of Health, 2019 [27].

Os padrões 1 e 2 são raramente observados, sendo o padrão 3 o mais comum. Na análise, o patologista utiliza dois graus distintos que, somados, definem o escore Gleason. O primeiro grau define o padrão dominante do tumor e o segundo grau o próximo padrão mais frequente na lesão. Há, ainda, um grau terciário que pode ser usado quando há um grupo menor de células cancerosas, geralmente mais agressivas, denominado no consenso de 2014 de padrão menor de alto grau, reservado aos produtos de prostatectomia total (não sendo, portanto, aplicado em biópsia). Por exemplo, se em uma amostra o padrão dominante é o 3 e o segundo mais observado é o 2, o escore Gleason é 5.

Logo, o escore Gleason vai geralmente variar de 2 (1+1) até 10 (5+5), sendo o primeiro representado pelos tumores mais bem diferenciados e o último pelos tumores menos diferenciados (quanto maior a desdiferenciação, maior a agressividade). A partir deste escore,

é possível classificar as lesões quanto ao comportamento biológico em tumores de baixo grau (bem diferenciados), grau intermediário e alto grau (moderado a pobremente diferenciado) [28].

Com o conhecimento do grau de desdiferenciação, foi possível agrupar os CaP de acordo com o escore Gleason e que melhor reflita o prognóstico baseado no comportamento biológico [29], conforme o quadro 3:

Quadro 3

Fonte: Pierorazio, 2013 [29]

Escore de Gleason <i>versus</i> Prognóstico	
Escore de Gleason	Grupo prognóstico
≤ 6	grau I
$3 + 4 = 7$	grau II
$4 + 3 = 7$	grau III
$4 + 4 = 8$	grau IV
9-10	grau V

Observa-se que um mesmo Gleason pode conter tumores diferentes, com prognósticos distintos. Por exemplo, um Gleason 3 ($3 + 4$) indica que a maior parte da amostra é o padrão 3, porém um mesmo Gleason 3 pode ter o padrão 4 como dominante ($4 + 3$), logo esse último tende a ser mais agressivo pela dominância do padrão 4. Portanto, a estratificação do grau de prognóstico permite separar adequadamente tumores com mesmo escore Gleason, favorecendo a abordagem terapêutica individualizada.

Um CaP com escore de Gleason menor que 6 geralmente vai apresentar bom prognóstico. Escores de 2 a 4 são achados incidentais de produtos de prostatectomia parcial (exérese da glândula interna periuretral – zona de transição) no contexto do tratamento cirúrgico da Hiperplasia Prostática Benigna (HPB). A maior parte dos diagnósticos de CaP são Gleason entre 5 e 7, oriundos de lesões palpáveis na zona periférica e passíveis de tratamento. CaP com Gleason entre 8 e 10 são lesões mais agressivas, portanto com menor chance de cura.

Há uma preocupação crescente quanto ao impacto econômico nos sistemas de saúde público e privado, oriundo do diagnóstico e tratamento do CaP. Evidente que os melhores resultados clínicos e financeiros são obtidos nos pacientes com detecção precoce de lesão restrita ao órgão. Entretanto, é interessante estabelecer qual seria o método de diagnóstico que ofereceria melhor custo-benefício na seleção destes pacientes, até mesmo quando aplicado ao rastreamento populacional. Um estudo recente [30], de 2018, realizado nos Estados Unidos da América, avaliou o custo do programa *US Medicare* (sistema de seguro de saúde gerido pelo governo daquele país voltado às pessoas maiores de 65 anos ou que atendam determinados critérios de participação) com o diagnóstico e tratamento do CaP local entre os homens com mais de 70 anos. Foi observada uma despesa média de 14.453 dólares americanos ao longo de três anos de tratamento por paciente. Entretanto, indivíduos com escore de Gleason até 6 submetidos a tratamento conservador por três anos, custaram apenas 1.194 dólares no mesmo período. O custo total do programa nos três anos analisados pela pesquisa com a detecção do CaP em idosos maiores de 70 anos foi de 1,2 bilhão de dólares, mas que poderia ser reduzido em 320 milhões de casos fossem implementadas medidas de vigilância ativa nos pacientes com escore de Gleason até 6. A maior parte deste valor é decorrente do tratamento, inclusive quando se utiliza terapias agressivas em pacientes com CaP menos agressivos. A redução do *sobrediagnóstico* do CaP implicando em *supertratamento* em indivíduos idosos que provavelmente morreriam de outra causa do que o câncer de próstata resultaria em economia substancial aos serviços de saúde.

2.2. Método de diagnóstico por imagem por ressonância magnética

Com os avanços tecnológicos, tanto em *software* quanto em *hardware*, a Ressonância Magnética (RM) se tornou ferramenta auxiliar no rastreamento e diagnóstico do CaP, ao mesclar imagens anatômicas com funcionais e metabólicas, associadas ao progressivo conhecimento na sua interpretação por radiologistas mais experientes, permitindo o diagnóstico precoce do câncer clinicamente significativo (e reduzindo sua mortalidade) e assegurando a diferenciação entre lesões benignas e tumorais com baixa agressividade biológica (e que comumente não trazem maiores risco a vida).

Sumariamente, a RM é realizada através de aquisição de imagens através do fenômeno da ressonância que ocorre quando se estimula prótons alinhados dentro de um campo magnético com pulsos de radiofrequência (PRF). Na vigência de mesma frequência entre a precessão dos

prótons e o PRF, o estado energético do próton é modificado, ocorrendo realinhamento dos prótons tanto perpendicular ao vetor do campo – magnetização longitudinal – como na precessão (congruência de fase) – magnetização transversal. Ao término do pulso, o próton tende a retornar ao estado inicial, quando então libera energia na forma de eletromagnética (chamado de sinal) que é captado pelo aparelho, filtrado, analisado e transformado em imagem. Este sinal é formado tanto pela intensidade do tempo de retorno ao estado energético inicial (alinhamento como campo magnético) como pela intensidade da perda da congruência de fase.

Na obtenção do campo magnético, o equipamento pode ser montado com ímãs permanentes ou temporários. Os permanentes são mais pesados e de menor força (até 0,3 Tesla, disponíveis comercialmente). Os temporários, embora mais caros, são de maior intensidade de campo (encontrados nos diversos fabricantes entre 1,5 e 3 T, apesar de haver campos maiores que 7 T para fins de pesquisa). Estes campos temporários são bobinas metálicas que quando ligadas a um circuito elétrico geram um campo magnético, com base no conhecimento teórico de que todo elétron em movimento promove um campo magnético e vice-versa. Como o fluxo de elétrons dentro de um fio metálico promove “atrito” e, portanto, calor, tais bobinas são imersas em hélio líquido, reduzindo a temperatura do condutor para próxima do zero absoluto, temperatura na qual este adquire propriedades supercondutoras, com mínimo atrito e pouca produção de calor. Quanto maior a força do campo magnético, mais prótons são recrutados para o fenômeno da ressonância magnética, portanto mais sinal é obtido e melhor será a qualidade da imagem.

A disponibilidade do próton no meio de estudo determinará a frequência dos PRF que o aparelho vai empregar. No corpo humano, o próton de hidrogênio é o mais abundante, logo mais disponível para serem recrutados e participarem do fenômeno da ressonância magnética. Mas também é possível realizar com fósforo, carbono, sódio e flúor, quando se adquire informações relevantes de estruturas ricas destes elementos e de seus metabolismos.

A manipulação do tempo entre o término do PRF e o momento no qual o aparelho irá captar o sinal (denominado tempo de eco) e no intervalo entre a repetição dos PRF (chamado tempo de repetição) determina aquisição de momentos diferentes do mesmo sinal (em termos de tempo de realinhamento do próton com o campo magnético e no tempo e intensidade da perda da congruência de fase), que traduzem diferentes aspectos do tecido biológico em estudo. Tais características do próton de hidrogênio em relação ao seu tempo de magnetização longitudinal e de perda da intensidade de magnetização transversal são denominadas T1 e T2, respectivamente. Por exemplo, uma mesma estrutura poderá ter aparência distinta entre as

imagens com ponderação em T1 e T2: a água é representada com *pixel* escuro em T1 e *pixel* com alto brilho em T2, enquanto a gordura aparecerá com alto brilho nestas ponderações. A partir das ponderações T1 e T2, e manipulando os tempos de eco e de repetição, é possível criar diversas sequências de pulso, que aliadas a inúmeros artifícios (tais como supressão de água ou gordura), geram uma variedade de tipos de ponderações.

Na atualidade dos protocolos de exames focados no estudo da próstata, a RM com ponderação T2, técnica Difusão (*diffusion weighted images – DWI*) e seus mapas derivados, técnica de avaliação dinâmica do realce pelo contraste (*dynamic contrast-enhanced – DCE*) e espectroscopia de prótons de hidrogênio provou ser o método eficaz na localização dos CaP de alto risco, ao combinar num mesmo exame as características anatômicas e funcionais das lesões suspeitas, favorecendo detecção e estadiamento, assim como definindo o sítio ideal de biópsia [31] [32].

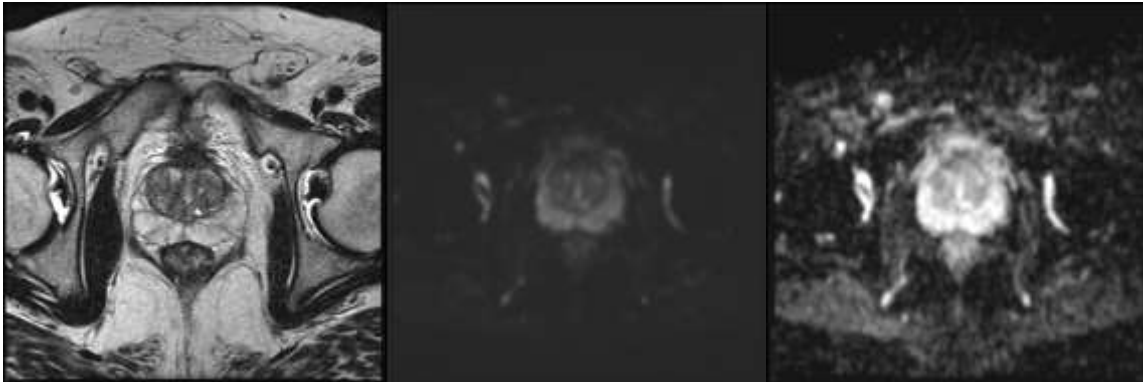
As imagens ponderadas em T2 fornecem as características anatômicas da próstata, notadamente na distinção zonal, sendo possível diferenciar a zona periférica da glândula interna ou central (dificilmente os métodos de imagem permitem distinguir as zonas transicional, central e periuretral entre si). A glândula interna é sede frequente dos nódulos de hiperplasia responsáveis pelo aumento do volume prostático na HPB, com aspecto heterogêneo de sua intensidade de sinal em T2, por vez semelhante ao nódulo de CaP. Para diferenciá-los, as lesões neoplásicas tendem a ser lentiformes, homoganeamente hipotensas em T2 (denotando maior densidade celular) e suas margens mal definidas e às vezes espiculadas. Invasão capsular, do estroma fibromuscular anterior ou da uretra favorecem a identificação do CaP. Por outro lado, os nódulos de hiperplasia costumam ser heterogeneamente hipointensos em T2, com halo de menor hipointensidade e contornos nítidos [33]. Porém, somente pelo padrão do baixo sinal em T2 não é possível assegurar especificidade ao método, pois outras condições podem apresentar aparência semelhante, tais como a hemorragia, atrofia glandular e o *status* pós-radioterapia.

As imagens ponderadas com a técnica Difusão (DWI) modificaram substancialmente a acurácia da RM na detecção do CaP, não só ao aumentar a especificidade, mas também ao atribuir sensibilidade ao método [34]. O fundamento físico da ponderação DWI é a observação do movimento browniano (aleatório) das moléculas de água, baseado na atenuação do sinal T2 Gradiente Echo na região em análise. Quanto mais moléculas de água possam se movimentar (ou seja, quanto mais livremente a água possa se difundir), mais perda do sinal original em T2 Gradiente Echo ocorrerá. Porém, se em uma determinada região houver algo que restrinja o movimento das moléculas de água, mais o sinal T2 Gradiente Echo inicial permanecerá.

Tomando por exemplo o encéfalo, a água que compõem o líquido cefalorraquidiano – líquor – pode se difundir com facilidade, restando pouco do sinal original do início da sequência, logo o sistema ventricular (cheio de líquor) será representado por *pixels* pretos (sem sinal). Em uma situação oposta, como em uma lesão cerebral isquêmica, a água no seu interior não poderá se movimentar com tanta facilidade devido a falência energética nas bombas de membrana celular no contexto da anóxia, logo continuará com o mesmo sinal inicial em T2 Gradiente Echo e será representada por *pixels* branco (alto sinal), clinicamente comentada como uma lesão com “difusibilidade restrita”. Por ser baseada em T2, as imagens DWI poderá ser “contaminada” pelo sinal original em T2 quando este for muito elevado, falseando o aspecto de restrição a mobilidade das moléculas de água. Para contornar tal situação, utiliza-se mais de uma sequência DWI, com ponderações diferentes de T2, e matematicamente se calcula o peso atribuído ao efeito T2 no sinal DWI, obtendo-se uma imagem cujas regiões com real restrição a difusibilidade da água são simuladas em *pixels* pretos e aquelas com difusibilidade facilitada (“contaminação T2”) são exibidas em *pixels* brancos. A este artifício se denomina Mapa de Coeficiente de Difusão Aparente (*apparent diffusion coefficient* – ADC).

O grau de difusibilidade das moléculas de água está associado ao arranjo das membranas do tecido em estudo, ao conteúdo do seu citoplasma e ao arranjo arquitetural do seu citoesqueleto. Em um tecido tumoral, no qual as células se reproduzem e crescem desordenadamente, é de se esperar que a densidade celular seja maior, e por consequência maior será a restrição à mobilidade das moléculas de água. Esse conhecimento permite utilizar as imagens por Difusão não só na detecção de lesões neoplásicas em diversos órgãos e tecidos corporais, mas também em categorizá-las em relação ao grau de restrição, sendo maior quanto mais intensa a proliferação celular e pior a agressividade biológica. Na avaliação do CaP, isso é traduzido em mapas ADC com baixo sinal nas lesões suspeitas para neoplasia em comparação ao sinal intermediário / alto do tecido sadio circunvizinho, assim como quanto mais baixo o sinal no mapa ADC, maior a correlação com o grau de agressividade tumoral e com o escore de Gleason [35].

Figura 2: exemplos de imagens ponderadas em T2, DWI e ADC.



Fonte: o autor, 2020

Na figura 2 é possível apreciar exemplo de imagens de RM-MP da próstata. Na imagem à esquerda, corte axial na ponderação T2; na central, corte axial na técnica Difusão (notar o aspecto ruidoso e com baixo sinal típico); e na imagem à direita, o respectivo mapa ADC daquela Difusão.

A utilização intravascular do meio de contraste paramagnético, em especial os derivados do ácido gadopentético (comercialmente denominado gadolínio), para caracterização do comportamento hemodinâmico de lesões, já é bem definido na prática radiológica. Sua fundamentação advém da neoangiogênese dos tecidos tumorais que, pela velocidade em que se dá, formam estruturas vasculares defeituosas em relação a integridade do endotélio, favorecendo a permeabilidade vascular e, portanto, o extravasamento da substância contrastante para o interstício celular. Na RM, o contraste paramagnético reduz o tempo de relaxamento na ponderação em T1 dos tecidos por ele embebidos, logo aumentando o sinal da região. Especificamente na próstata, utiliza-se protocolo dinâmico (*dynamic contrast enhanced* – DCE) cuja técnica é a injeção intravenosa rápida da substância contrastante seguida da aquisição seriada de blocos de imagens ao longo de um tempo definido. Na análise, se obtém curvas da intensidade da captação do contraste (realce) pela lesão ao longo do tempo de aquisição. Três tipos de curva são descritos [36]. A tipo 3, nos tumores mais neoangiogênicos, nos quais é observado precoce e intenso realce (*wash in*), com perda da contrastação nos segundos seguintes devido a depuração pelo sangue mais diluído (e, portanto, menos contrastado) que adentra a lesão nas aquisições subsequentes (efeito de clareamento ou “lavagem” – do inglês *wash out*). Já nos tumores menos angiogênicos é observado realce que permanece ao longo do tempo (curva tipo 2) e nas lesões benignas e no tecido circunjacente normal o realce é menor, lento e progressivo (curva tipo 1). Esta propriedade farmacodinâmica do contraste nos tumores

favorece não somente a localização destes como também sua caracterização em termos de agressividade biológica, já que o CaP tende a ser hemodinamicamente mais ativo que as lesões benignas [37]. Três modelos de abordagem são utilizados para análise do realce de uma lesão. O qualitativo é baseado na avaliação do padrão de curva, não sendo levado em consideração valores absolutos. Estatisticamente há grande variabilidade interobservador na diferenciação entre as curvas tipo 1 e 2, porém é excelente na caracterização da curva tipo 3. O modelo semi-quantitativo foi proposto para objetivar a análise qualitativa através de métodos que calculam os parâmetros da curva, tais como o tempo para o pico de realce, a taxa de realce, taxa de “lavagem”, etc, mas sua aplicabilidade é limitada devido falta de padronização, seja no protocolo de aquisição de imagem, seja no processo matemático implicado. Já o modelo quantitativo surgiu como ferramenta para sanar as limitações dos demais, utilizando a modelos farmacocinéticos conhecidos para comparação com o realce da lesão tumoral suspeita (por exemplo, a confrontação entre o realce de um compartimento conhecido e o da lesão). A vantagem desta abordagem é que os valores quantitativos não são dependentes dos protocolos de exame e dos operados do aparelho. Essas limitações de análise impediram que os métodos semiquantitativo e quantitativo fossem aceitos nas diretrizes atuais de diagnóstico e rastreamento do CaP [36]. Apesar disto, vários estudos reforçam que o contraste paramagnético agrega valor ao melhorar a detecção e distinção da agressividade do CaP, por avaliar características diferentes daquelas observadas nas ponderações T2 e DWI, além de já ter sido provado seu benefício em situações específicas, como no seguimento por imagem da terapêutica instituída (detecção de recorrências, estadiamento loco-regional etc.) [38].

A Espectroscopia de prótons por Ressonância Magnética (*MR spectroscopy* – MRS) obtém dados do metabolismo de um tecido através da estimativa de concentração de determinados metabólicos endógenos. Através de frequências ressonantes semelhantes ao do próton de hidrogênio, é possível estimar a concentração de diversas substâncias; em particular, na avaliação do CaP, a colina, a creatina, o citrato e a espermina (uma poliamina). A técnica aplica um *voxel* na região de interesse para o estudo espectroscópico, de onde se obtém um gráfico de amplitude x frequência, onde a frequência é específica para cada metabólito, descrita em ppm (parte por milhão), e a amplitude se relaciona com a sua concentração [34]. O citrato é produzido pela glândula prostática normal, sobretudo na zona periférica, e se encontra reduzido no CaP. Seu pico de frequência é 2,6 ppm. A colina é observada nos tecidos com alta proliferação de celular, como nas neoplasias, pelo maior metabolismo fosfolipídico da membrana. Seu pico é 3,22 ppm. A creatina reflete a reserva energética do tecido e, quando

aumentada, denota possibilidade de CaP. Como seu pico de 3,03 ppm é muito próximo do pico de colina, por vezes inseparáveis no gráfico espectral, na prática utiliza-se a relação (colina + creatina) / citrato, sendo valores acima de 0,5 suspeitos para neoplasia. A espermina é um metabólito de difícil caracterização devido a sua concentração (3,1 ppm) estar entre aos picos de colina e creatina. Em lesões suspeitas para o CaP costuma estar reduzida. O longo tempo para aquisição dos dados espectroscópicos, aliado a baixa resolução espacial e a susceptibilidade a artefatos, inviabilizou a espectroscopia nos protocolos e *guidelines*, apesar do seu valor diagnóstico [35] [39].

Os estudos recentes têm denominado a utilização combinada das ponderações T2, DWI e DCE como padrão no estudo prostático, sendo a técnica denominada RM Multiparamétrica (RM-MP). A associação da análise anatômica na ponderação T2 com a análise funcional na ponderação DWI e com a técnica DCE favorecem a diferenciação entre o tecido prostático benigno e canceroso, reduzindo o número de biópsias e evitando terapêuticas desnecessárias.

Em virtude da ausência de padronização nos protocolos de exame e nos relatórios de seus resultados, no ano de 2013, a Sociedade Européia de Radiologia Urogenital (*European Society of Urogenital Radiology – ESUR*) criou um protocolo único de classificação de lesões prostáticas quanto ao grau de suspeição de malignidade, denominado PI-RADS (*Prostate Imaging – Reporting and Data System*), semelhante a iniciativa BIRADS (*Breast Imaging – Reporting and Data System*), que desde 2003 é aplicada na avaliação do câncer de mama. Sua última versão, de 2015, já em conjunto com o Colégio Americano de Radiologia (*American College of Radiology – ACR*), definiu o uso das ponderações T2, DWI e DCE como recomendadas e classifica os achados em 5 categorias (quadro 4) [40]. Trata-se de um escore do tipo Likert, no qual a análise de características de imagem soma pontos para definir se o achado é concordante ou não com lesão suspeita para neoplasia. Em um estudo recente, que relaciona a categoria PIRADS reportada e a positividade de tumor clinicamente significativo, encontrou valores de 15,7% para categoria 2, 33% para 3, 70,5% para 4 e 90,7% para 5, mostrando boa correlação entre os achados de imagem e o estudo anatomopatológico [41].

Quadro 4

Categorização PIRADS (versão 2)	
Categoria	Probabilidade de neoplasia
1	Muito baixa probabilidade de neoplasia prostática clinicamente significante

2	Baixa probabilidade de neoplasia prostática clinicamente significativa
3	Achados indeterminados para neoplasia prostática clinicamente significativa
4	Moderada probabilidade de neoplasia prostática clinicamente significativa
5	Alta probabilidade de neoplasia prostática clinicamente significativa

Fonte: American College Of Radiology, 2015

Em síntese, os recentes avanços tecnológicos da RM no rastreamento e acompanhamento do CaP, através da adição de sequências funcionais que possibilitam maior acurácia na detecção e caracterização das lesões, tornam o método ferramenta central no manejo desta condição [42].

2.3. Computer-aided diagnosis

O *Computer-Aided Diagnosis* (CAD, ou diagnóstico auxiliado por computador) é um sistema computacional de análise de dados médicos que assiste os profissionais na interpretação de grandes volumes de informação em um curto intervalo de tempo. Embora alguns autores distingam os sistemas de detecção auxiliada por computador dos sistemas de diagnóstico auxiliado por computador, sendo que os primeiros se limitam a identificar estruturas e regiões conspícuas, e o segundo a análise propriamente dita, o CAD visa fornecer oferecer informações para apoiar uma decisão tomada pelo profissional.

Não há prejuízo da independência da prática médica ao se utilizar sistemas CAD, haja visto que a responsabilidade final é do profissional. O CAD participa como coadjuvante, detectando lesões cujas características iniciais de anormalidade não são passíveis de visualização humana e selecionando propriedades de um grande volume de informações, que manualmente demandaria longo tempo. Por exemplo, nos estudos mamográficos, o CAD pode identificar regiões de assimetria mamária suspeitas para câncer de mama em glândulas com densidade elevada e que a capacidade visual humana não distinguiria. Nos estudos tomográficos do tórax, nos quais o volume obtido num único exame ultrapassa mais de 300 imagens, o CAD pode rapidamente identificar todas as lesões nodulares milimétricas em poucos segundos, enquanto um profissional gastaria diversos minutos nesse rastreamento.

Na radiologia, as diversas modalidades digitais podem ser utilizadas pelos sistemas CAD e, inclusive, as modalidades não digitais (como a radiologia analógica) quando da sua digitalização. O sistema DICOM (*Digital Imaging and Communications in Medicine*) é o padrão para a comunicação e gerenciamento de informações de imagens médicas e propicia o formato adequado para análise das imagens pelo sistema CAD.

Está em voga nos últimos anos o termo *radiomics* (em português, equivalente ao neologismo radiômica), aplicado a extração de grande volume de informações de imagens médicas radiológicas através de algoritmos de caracterização de dados. Tais propriedades da imagem, também chamadas de características radiômicas, podem abrigar informações valiosas sobre doenças e que não são visíveis ao olho humano, abrindo um novo campo de conhecimento que poderá modificar tratamentos e prognóstico de diversas condições. Alguns autores citam que os dados radiômicos estão em uma forma “minerável”, e que a análise de fenótipos, imagens ou assinaturas de proteína / genes, por exemplo, pode criar modelos descritivos e preditivos de doenças [43]. Inicialmente aplicada na oncologia, e com estudo mais numerosos e avançados neste campo, a radiômica é útil em qualquer área médica na qual se possa extrair características de uma patologia se quer concebida.

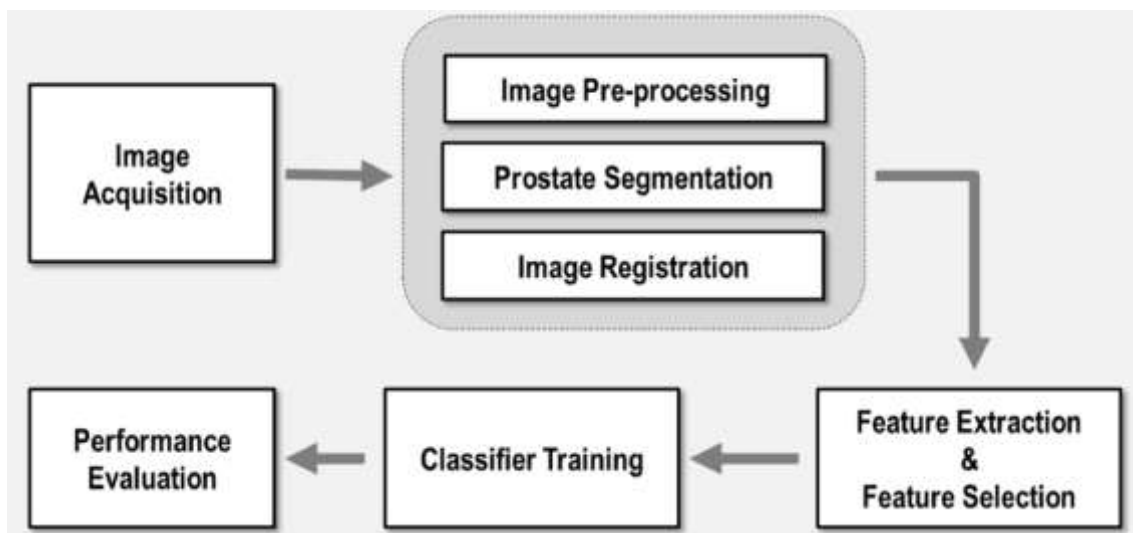
Ainda no campo oncológico, a radiômica apresenta vantagens em comparação com outros métodos convencionais de detecção e caracterização de tumores [44]. Por ser aplicada em imagens médicas já disponíveis, como ressonância magnética e tomografia computadorizada, a radiômica é não invasiva e não expõe o paciente a outro procedimento além daqueles elencados no diagnóstico de sua condição e no acompanhamento da patologia. Ao utilizar tanto imagens 2D quanto 3D do tecido amostrado, é possível quantificar a heterogeneidade intratumoral, aspecto interessante no caso do CaP, onde a variabilidade tumoral está relacionada adequada seleção do tratamento e ao prognóstico. Como as imagens médicas fazem parte de grandes acervos de armazenamento, pesquisas retrospectivas podem ser realizadas sobre essas imagens para melhorar a performance dos algoritmos de procura e detecção, assim como na extração de características específicas outrora não reconhecidas. Além do custo econômico reduzido frente aos benefícios gerados, pois além de não requerer novas aquisições de imagens além daquelas que naturalmente já seriam realizadas no contexto da avaliação oncológica, a radiômica basicamente utiliza da análise computacional de dados, demandando apenas gastos com *hardware* e *software* [45] [46] [47].

Especificamente no CaP, os esforços acadêmicos sobre a radiômica têm se voltado para a detecção tumoral, como ferramenta de análise radiológica, embora também possa ser aplicada

na decisão clínica entre terapêuticas diferentes baseada na análise das características de imagem, laboratoriais e genéticas.

O processo radiômico inclui uma sequência de instruções a serem aplicadas, compreendidas como etapas de pré-processamento, análise e pós-processamento. Inicialmente, após a aquisição da imagem digital, é realizada a “**normalização**” dela, que visa reduzir o ruído, remover a distorção e padronizar a intensidade de sinal para evitar variabilidade entre pacientes. A seguir é necessária a **segmentação** do objeto em estudo, na qual se delimita a região de interesse para análise CAD. A fase seguinte, denominada **registro**, é necessária para o correto alinhamento entre as diversas imagens adquiridas. A partir deste ponto, há a **extração** das propriedades da imagem e **seleção** das características pelo algoritmo computacional. As últimas etapas são o **treinamento** do classificador e a **avaliação** de sua performance (figura 3).

Figura 3 – Etapas radiômicas do CAD



Fonte: Sun et al, 2019 [44].

Pré-processamento: normalização

Considerando o CaP e as aquisições de protocolo multiparamétrico de um RM da próstata (T2W, DWI, DCE), inicialmente são aplicados filtros visando a “normalização” da imagem, pois a qualidade destas pode sofrer interferência da temperatura do ambiente e do tecido, da variação da intensidade do campo magnético, do ganho aplicado na bobina receptora, da posição da bobina em relação ao corpo humano, além de variabilidade entre as sequências

de pulso e os modelos de diferentes fabricantes. De um modo geral, é de necessária eliminação de ruído e homogeneização da intensidade de sinal.

Pré-processamento: segmentação

É indispensável reduzir a quantidade de informações que serão aplicadas ao classificador nas etapas finais da análise, evitando, assim sobrecarga de processamento de dados. Para tal, seleciona-se somente o tecido que sofrerá a apreciação CAD (no caso, a glândula prostática) e elimina-se o tecido circunjacente. As imagens de maior resolução e contraste são preferidas nesta etapa: das disponíveis no protocolo multiparamétrico (T2, DWI e DCE), a ponderação T2 é a escolhida e a segmentação é extrapolada para as demais sequências. Tal processo pode ser manual ou automatizado. Na técnica manual, o operador delinea o contorno prostático em cada imagem, sendo processo mais demorado e trabalhoso, embora adequado na aplicação de estudos científicos de pequena coorte [44].

Em grandes volumes de imagens, pode se utilizar métodos automatizados ou semiautomatizados de segmentação do tecido prostático, e em estudo de 2016 já demonstrava resultados promissores [18]. Há diversos modelos para esta abordagem, destacando-se os baseado em forma (na qual em um modelo pré-estabelecido são aplicados vetores de força para determinar o objeto), no contorno (no qual detecta-se as bordas e arestas do objeto para determiná-lo), em atlas (uma imagem de segmentação manual que já determinou o objeto é aplicada como referência em novos pacientes) e, mais recentemente, na abordagem em gráfico (em que um algoritmo de corte é aplicado em uma esfera de probabilidade do tecido tratar-se ou não do objeto em estudo).

Qualquer que seja a metodologia aplicada na segmentação, alguns complicadores podem ser observados. Na aquisição da imagem T2 deve haver adequado contraste entre o tecido prostático e a gordura circunjacente para adequado delineamento da glândula, o que é prejudicado nos pacientes muito magros ou até mesmo caquéticos em virtudes da doença cancerosa avançada. É importante que a mesma orientação, espessura de corte, gap e campo de visão (FOV, *field of view*) devem ser utilizados nas diversas aquisições de imagem para que seja apropriada a extrapolação dos limites prostáticos para as sequencias de baixa resolução (DWI e DCE). Mudança da localização prostática entre uma sequência e a seguinte também pode interferir na segmentação, seja por movimento voluntário do paciente (inquietaude) ou até mesmo por aumento da repleção vesical ou modificação do padrão respiratório do indivíduo, embora em menor intensidade. O peristaltismo de alças intestinais, que pode gerar artefatos e

degradar a qualidade das imagens adquiridas, pode ser evitado ou reduzido através da administração venosa de escopolamina ou substâncias análogas. A variabilidade da morfologia e dimensões da próstata entre os pacientes também deve ser considerada ao se pensar na aplicação de métodos automatizados que utilizem modelos pré-estabelecidos. Por fim, condições específicas de um aparelho de RM podem afetar o processo de segmentação, inviabilizando a aplicabilidade de um esquema de segmentação prostática validado para outros equipamentos.

Pré-processamento: registro ou cadastro

Fase do processo em que as diversas sequências serão alinhadas, fundamental para que ocorra a extração de características na etapa seguinte. É necessária a correspondência espacial entre a segmentação prostática obtida ao longo das diversas ponderações (T2, DWI e DCE), que é obtido pelo através das coordenadas das imagens. O padrão DICOM, no cabeçalho de cada imagem, detém informações relevantes sobre a aquisição daquela imagem, inclusive relacionadas a localização daquele corte em relação as coordenadas x , y e z do aparelho [18]. Deformação da próstata entre uma sequência e outra pode ocorrer e é dependente dos protocolos utilizados e da aplicação ou não de bobina endo-retal: por exemplo, protocolos mais longos podem gerar diferenças entre as primeiras aquisições e as mais tardias [44]. Em tais situações, algoritmos de registros que considerem a estas variabilidades podem ser aplicados.

Quando da validação de um sistema CAD, é interessante a comparação dos resultados da obtidos pela RM com o padrão-ouro (o estudo anátomo-patológico). Para tal, faz-se necessário o mesmo alinhamento seccional entre as imagens obtidas pela RM com o fatiamento da peça cirúrgica. Vários estudos com abordagens diferentes descrevem formas de registrar de forma adequada os cortes de RM e histológicos [18].

Extração de características

É o momento no qual os atributos que definem um achado patológico serão obtidos das imagens. No caso DICOM, características como intensidade de sinal, morfologia, textura e comportamento hemodinâmico (farmacocinética) podem ser extraídas. Em geral, as características podem ser classificadas em agnósticas e semânticas [46], as primeiras relacionadas a descritores quantitativos que captam a heterogeneidade de uma lesão e as segundas a descritores já utilizados no léxico radiológico, como intensidade de sinal, morfologia, limites, contornos etc., que já detém valor prognóstico estabelecido na literatura

(por exemplo, lesão com intensidade de sinal baixa nas imagens ponderadas em T2 que sugere CaP).

Os descritores quantitativos (atributos agnósticos) são obtidos através da análise estatística da imagem e habitualmente não fazem parte da prática radiológica, sendo classificados em saídas de primeira, segunda ou alta ordem.

As saídas de primeira ordem descrevem a distribuição dos valores obtidos em *voxels* sem considerar as relações espaciais. Um exemplo comum deste método é o histograma, no qual da intensidade de sinal dos *voxels* de uma região de interesse (do inglês *region of interest*, ROI) são extraídos valores de média, mediana, máximo, mínimo, uniformidade ou aleatoriedade (entropia), obliquidade (assimetria) e *kurtosis* (nivelamento) da curva de distribuição estatística.

As saídas de segunda ordem geralmente descrevem a característica de “textura” e está relacionada ao arranjo espacial dos *voxels* em uma lesão, portanto analisa estatisticamente a semelhança entre eles através da referência de valores de contraste parecidos ou diferentes. Em última análise, a “textura” matematicamente fornece a medida de heterogeneidade de uma lesão, muito útil na CaP, onde a variabilidade tumoral está relacionada a diferentes prognósticos.

Por fim, as saídas de alta ordem (ou ordem superior) que utiliza cálculos matemáticos mais avançados através da aplicação de elementos filtrantes às imagens com intuito de se extrair padrões de repetição. Por exemplo, *wavelets* (filtro que multiplica uma imagem por uma matriz de "ondas" lineares ou radiais complexas), filtros passa-banda gaussianos (para extrair áreas com padrões de textura mais grossos de uma imagem), fractais (onde são selecionados *voxels* de um valor específico quando é superposto à imagem um padrão definido) e a função de Minkowski (que elege padrões de *voxels* cuja intensidade de sinal ultrapassa determinado valor).

No tratamento das imagens, para a ponderação T2 é avaliada além da intensidade de sinal da lesão e da sua representação na escala de cinza, suas características anatômicas como, por exemplo, a distância para contorno prostático. Para a ponderação DWI, mas sensível na identificação do CaP, é utilizada a “característica de bolha” (*blobness*), que traduz o aspecto morfológico da lesão. Já para a ponderação DCE, é aplicada curva bi-exponencial de intensidade de sinal *versus* tempo para captar a característica farmacocinética da lesão e classificá-la segundo as curvas de padrão de benignidade e malignidade.

Seleção das características

É necessário otimizar a quantidade de informações que o sistema CAD irá processar.

Ao aumentar o volume de características, mais espaço entre os dados será demandado, tornando-os mais escassos e reduzindo a capacidade discriminativa entre eles (efeito conhecido como “maldição da dimensionalidade”) [44]. Além disso, aumenta-se a demanda por tempo e poder de processamento dos dados. A abordagem através de sistematização das prioridades evita tal problema, na qual classifica-se as características agnósticas e semânticas, determinando a contribuição de cada uma ao conjunto de informações obtidas e atribuindo peso a probabilidade estatística de cada.

Ao reduzir o número de características, a capacidade de generalização do modelo aumenta já que uma das dificuldades da abordagem agnóstica é que o número de elementos para análise pode se tornar tão complexo e grandioso que prejudique o julgamento do CAD (overfitting) [44].

O excesso de informação iguais (redundância) e a coexistência de características semelhantes que se correlacionem também pode provocar falha na análise (efeito de colinearidade / multicolinearidade). Modelos de programação CAD que utilizam subconjuntos menores de dados têm menor tempo de treinamento e performance aprimorada na interpretabilidade [44].

No processo de seleção das características aplica-se recursos estatísticos para escolher o subconjunto de atributos, tais como o teste t para duas amostras independentes, o valor de p para significância estatística etc. Outra abordagem para adequar o tamanho do subconjunto é decompor o atributo e reconstruí-lo de forma menor a partir da combinação de suas características originais [44].

Classificação

Nesta fase, enfim, ocorre a avaliação da característica extraída e o confronto com a resposta esperada. Envolve o treinamento, no qual é apresentado ao classificador imagens rotuladas para o seu aprendizado, e a testagem do conhecimento adquirido, quando é oferecido conjunto imagens não rotuladas.

Independente do sistema CAD utilizado, o banco de imagens disponíveis deve ser dividido em subconjunto de teste e subconjunto de validação, este reservado a parte e somente apresentado ao sistema quando do término da etapa de treinamento. Uma abordagem é

simplesmente particionar o banco de dados e utilizar cada subconjunto para um propósito específico. Outra possibilidade é a validação cruzada (*cross validation*), uma abordagem iterativa baseada em particionamento do banco de dados em subconjuntos, do qual um deles será reservado para aplicação do teste e os demais utilizados no treinamento, sendo o processo repetido até que todos os subconjuntos tenham sido empregados como teste [44].

No sistema CAD, dependendo da sua complexidade e objetivo, pode aplicar técnicas de inteligência artificial, aprendizado de máquina ou abordagens estatísticas, incluindo redes neurais, vetor de suporte e redes Bayesianas [46]. Qualquer que seja o método empregado, o objetivo final (*endpoint*) é a resposta ao problema proposto, comumente expresso em uma variável dicotômica (*sim* ou *não*) ou ainda numa probabilidade (por exemplo, percentual de chance de CaP em uma lesão). Das diferentes técnicas, máquinas de vetor de suporte (*support vector machines*, SVM) e florestas aleatórias (*random forests*) têm bom desempenho quando aplicadas a amostras de treinamento dicotômicas [18].

A técnica de regressão logística utiliza uma função sigmoideana para discriminar uma característica em um intervalo entre 0 e 1. Por exemplo, no CaP pode determinar a probabilidade de um ROI sobre uma lesão ser ou não tumor. Vários modelos já foram testados no CaP, com sensibilidade e especificidade de 87,3% e 78,4%, respectivamente [48]. Utilizando-se modelos de regressão logística baseados em imagens T2 e DWI (mapas ADC, mais precisamente), foi observada adequada caracterização do CaP independente da zona prostática acometida [49]. Nesta técnica, a importância da característica determinará o peso dela no cálculo do resultado, porém é mais difícil lidar com padrões mais complexos que envolvam vários atributos.

A máquina de vetor de suporte (SVM) é um método de classificação binária, linear e não probabilística, sendo muito utilizado no CAD. De um conjunto de dados, a SVM determina se uma característica encontrada na amostra faz parte de qual subconjunto de dados (sempre separando-as em duas classes) [44]. Em um gráfico, a SVM define a linha com os valores que separam duas classes distintas (também chamado de hiperplano), logo, quanto mais afastados dos valores que compõem esse plano, mais distintas são as classes; por outro lado, quanto mais complexo o padrão das características, mais difícil é definir o hiperplano, com menor o afastamento dos valores e menor a distinção das classes.

Baseada na técnica “árvore de decisão” (*decision tree*, na qual um fluxograma representa graficamente as possibilidades de cada entrada), o objetivo da “**floresta aleatória**” (*random forest*) é de descobrir a classificação de uma nova entrada [44]. Fornece saída tanto

binária quanto probabilística. Neste classificador, as características são separadas em subconjuntos de atributos, ou seja, ele cria diversas *decision trees*, cada uma delas com suas particularidades, e ao final combina o produto da classificação de todas, tornando um algoritmo mais forte que uma *decision tree* individual. Na comparação com a SVM na detecção do CaP através da RM com ponderações T2 e DWI, a “floresta aleatória” foi superior, tanto na identificação da lesão quanto na menor taxa de erro [50] [51].

Avaliação da performance (validação)

Por último, a validação do algoritmo é passo fundamental no desenvolvimento de um sistema CAD. O padrão-ouro utilizado, o estudo anátomo-patológico, serve de referência da performance do procedimento empregado, em geral obtidos de prostatectomia ou amostras de biópsias. Ressalta-se que há necessidade para adequada correlação da imagem com o produto *ex vivo*, como foi anteriormente exposto, do correto registro (alinhamento) das imagens de RM com os cortes seccionais histológicos (corados em hematoxilina-eosina e fixados em parafina, quando se tratar de produtos de prostatectomia). Pode-se, ainda, aplicar o conhecimento dos escores Gleason. É factível calcular a precisão do método através do cálculo do número de voxels ou ROIs que o sistema CAD apropriadamente classificou. Nos algoritmos que produzem resultados contínuos, a curva ROC (especificamente a área sob a curva) permite aferir a performance.

3. METODOLOGIA

3.1. Aspectos gerais

A população de estudo é a coleção de artigos científicos publicados em periódicos relevantes obtida pela aplicação da metodologia de seleção. Não foram considerados artigos incompletos, de conferência ou resumo de congresso.

O objetivo foi garantir acesso ao maior número de artigos que preenchessem critérios adequados para inclusão no estudo, através de pesquisa em diversas bibliotecas digitais por meio de mecanismos de busca específicos. Visando conceder acesso equiparável de qualquer artigo à seleção, as bibliotecas virtuais tinham que oferecer mecanismo de busca que permitissem utilizar expressões lógicas, ou artifício semelhante. Ainda foi necessário que, em suas bases de indexação, tivessem publicações na área de estudo e que permitissem busca no texto completo da publicação. Todas as publicações teriam que pertencer a alguma das editoras disponíveis no Portal de Periódicos da CAPES.

A pesquisa de artigos científicos disponíveis na literatura médica digital foi através de mecanismos de busca nas plataformas MEDLINE/PUBMED, ScienceDirect, IEEE, Cochrane Library e Portal de Periódicos da CAPES, utilizando-se expressões previamente definidas. Não houve busca manual, pois fontes não digitais foram excluídas.

A seguinte expressão de busca foi utilizada para pesquisa de artigos na língua inglesa:

("computer-aided diagnosis" OR "computer-aided detection" OR CAD) AND (MRI OR "magnetic resonance imaging") AND multiparametric AND "prostate cancer"

E, para língua portuguesa, a expressão de busca abaixo:

("diagnóstico assistido por computador" OR CAD) AND (RM OR "ressonância magnética") AND multiparamétrica AND "câncer de próstata"

Somente foram analisados os artigos cuja publicação atenda aos critérios definidos, e que tenham sido divulgados entre janeiro de 2011 a abril de 2021. O idioma inglês foi selecionado para realização desta pesquisa, haja visto uma inúmera quantidade de material

disponível nesta língua e por ser o idioma principal das editoras disponíveis no Portal de Periódicos da CAPES. O idioma português foi excluído como língua na pesquisa pois os resultados preliminares não encontraram qualquer publicação.

Após o resultado da expressão de busca nas plataformas citadas, todos os artigos retornados foram catalogados em matriz de análise criada para este fim e armazenada em diretório específico na nuvem, com cópia (backup) off-line. O pesquisador, então, analisou cada artigo retornado, selecionando aqueles relevantes através da aplicação de critérios de inclusão e exclusão. A fim de diminuir a chance de um artigo ser excluído prematuramente, em casos de dúvida durante a aplicação dos critérios de inclusão e exclusão, o mesmo foi incluído.

O processo de seleção de artigos foi composto de quatro etapas: (I) seleção e catalogação preliminar das publicações, conforme explicitado no parágrafo anterior; (II) aplicação do primeiro filtro de seleção das publicações relevantes, que é composto de critérios de inclusão e exclusão; (III) aplicação do segundo filtro de seleção, utilizando-se padrões de qualidade; e (IV) aplicação do terceiro filtro, que visa resgatar dados relevantes, outrora excluídos pelos filtros precedentes.

O primeiro filtro (de seleção das publicações relevantes) utilizou critérios de inclusão (CI) e de exclusão (CE), conforme abaixo:

Crítérios de Inclusão (CI): incluídos artigos que:

CI1: incluíram estimativa da adequação do CAD como ferramenta auxiliar na detecção e a caracterização de neoplasia prostática.

CI2: descreveram técnicas de segmentação, registro, extração de características e classificação de imagens de próstata.

CI3: abordaram a avaliação de desempenho e validação usando casos reais (por biópsia ou estudo anatomopatológico de peça cirúrgica – prostatectomia) para mensurar a performance e robustez do algoritmo utilizado no CAD.

CI4: abordaram a avaliação de desempenho e validação do CAD em relação a análise humana com protocolo PIRADS

CI5: abordaram a avaliação de desempenho e validação do CAD em relação a análise CAD com protocolo PIRADS

CI6: apresentaram recomendações para melhoria no processamento de imagens para CAD.

Crítérios de Exclusão (CE): excluídos artigos que:

CE1: não abordaram técnicas de processamento de imagens para CAD.

CE2: não abordaram segmentação, registro, extração de características ou classificação de imagens de próstata.

CE3: não utilizaram sequência T2, técnicas de Difusão e Perfusão no diagnóstico de câncer de próstata.

CE4: não aceitaram a padronização PIRADS.

Para avaliação da qualidade científica do artigo (segundo filtro) foram considerados os periódicos com classificação QUALIS CAPES A1, A2, B1 e B2 (conforme busca na plataforma Sucupira, disponível em <https://sucupira.capes.gov.br/>); periódicos com revisão por pares; originalidade do tema e estrutura do trabalho levando-se em consideração a aplicação do método científico.

O terceiro filtro (seleção dos dados relevantes – SDR) visava assegurar que os artigos selecionados são necessários à pesquisa pois eventualmente os filtros de inclusão e exclusão poderia escolher material inapropriado ou descartar artigo relevante. Ao fim da leitura dos artigos selecionados no primeiro e segundo filtros, aplicou-se os seguintes critérios:

SDR1: não deveriam ser selecionados artigos que descrevessem técnicas de diagnóstico de câncer de próstata por CAD, mas que não tivessem confirmação histológica da malignidade da lesão.

SDR2: poderiam ser selecionados artigos que, embora não utilizassem técnicas preconizadas pela padronização PIRADS ou não fizessem uso da sequência T2, das técnicas de Difusão e Perfusão, apresentassem inovações tecnológicas no diagnóstico do câncer de próstata.

Na matriz de síntese de dados utilizada, cada artigo, com suas referências completas, que tenha sido aprovado em determinado filtro, foi marcado na coluna correspondente a este como APROVADO; em contrário, utilizou-se o termo REPROVADO. Nesta matriz de análise, constavam:

i) Título do artigo.

- ii) Autores.
- iii) Ano da publicação.
- iv) Local da publicação.
- v) Nome da revista em que foi publicada.
- vi) Categoria plataforma Sucupira / CAPES.
- vii) Coluna de categorização em APROVADO ou REPROVADO.

O procedimento para extração dos dados, a partir da população de artigos oriundos dos mecanismos de busca e previamente filtrados / selecionados, obedeceu a seguinte sequência: **extração de dados, alocação em matriz de análise específica e sumarização**. Cada artigo foi individualmente “entrevistado” através de formulário específico (ferramenta de extração dos dados), conforme a figura 4 que se encontra no apêndice B, e com apoio da matriz de análise e síntese de dados, coletou-se os seguintes itens:

a) Dados gerais:

- i) Título do artigo.
- ii) Autores.
- iii) Ano da publicação.

b) Modalidades utilizadas pelo CAD:

- i) T2.
- ii) DWI.
- iii) ADC.
- iv) DCE
- v) Espectroscopia.

c) Método de validação:

- i) Biópsia.
- ii) Prostatectomia
- iii) Secções histológicas.
- iv) Seguimento

d) Região da próstata estudada:

- i) Próstata total.
- ii) Zona periférica.
- iii) Glândula central.
- iv) Combinações (assinalando-se todas aquelas opções correspondentes na ferramenta).

- e) **Tipo do classificador:**
 - i) SVM.
 - ii) Redes neurais.
 - iii) LDA / QDA.
 - iv) *Random forest*.
 - v) Outros, incluindo não declarados e comerciais.
- f) **Tamanho da amostra**
- g) **Resultados do estudo:**
 - i) Área sob a curva ROC.
 - ii) Sensibilidade.
 - iii) Especificidade.

Os resultados obtidos foram, por fim, resumidos e sumarizados em forma de tabelas, somente com a seleção final, visando detectar intercessão de conhecimento entre os artigos selecionados e agrupá-los conforme afinidade. O orientador do estudo realizou revisão dos procedimentos realizados.

No procedimento para análise, os dados obtidos sofreram avaliação quanto aos aspectos quantitativo e qualitativo. Na primeira, a extração dos dados da matriz de análise permitiu a avaliação quanto:

- a) Quantidade de artigos selecionados a participarem do estudo.
- b) Número de métodos utilizados para detecção de lesão neoplásica através da aplicação de CAD em RM-MP.
- c) Número de artigos com validação do CAD como ferramenta para detecção de neoplasia de próstata através da RM-MP.

Já a análise qualitativa verificou a performance do CAD em RM-MP na detecção de neoplasia prostática e produziu inferências estatísticas quanto a sua *performance* diagnóstica clínica.

3.2. Avaliação do risco de viés

Ao apreciar a população de artigos selecionados, notou-se que havia heterogeneidade nas publicações e, portanto, risco de viés quanto a inclusão de produto que não atendessem critérios de qualidade. A fim de mitigar tal risco, além dos critérios descritos no item 3.1

(considerados os periódicos com elevada classificação QUALIS CAPES, periódicos com revisão por pares, originalidade do tema e estrutura do trabalho), cada artigo foi avaliado individualmente, em sua leitura completa, quanto a:

- a) Seleção de sua amostra, observando-se o perfil populacional.
- b) Qualidade da imagem adquirida, levando-se em consideração espessura do corte, tempos de eco (TE) e de repetição (TR), e valores de B (para as imagens ponderadas em Difusão), fabricante, campo magnético, tipo de bobina e pós-processamento da imagem.
- c) Processamento de imagem CAD: etapas de segmentação, registro, extração de características e classificação.
- d) Qualidade do método de validação: estudo anatomopatológico, prostatectomia *versus* biópsia.

Tais informações, quando relevantes, eram anotadas no verso da ferramenta de extração ainda no processo de entrevista, visando posterior abordagem para mitigar potencial falha. Por exemplo, estudos que utilizaram de pacientes com escore de Gleason prévio de grau elevado, outros que usaram a biópsia guiada por US-TR como instrumento de seleção prévia e posterior estudo anátomo-patológico confirmatório, classificadores não declarados e comerciais etc. *Clusterização* da amostra é um artifício interessante e válido, porém se considerou que a escassez de artigos selecionados sobre o tema no contexto de uma subcategorização dos dados implicaria fortemente na perda do poder estatístico do estudo.

3.3. Experimento com radiologistas

Durante a realização do estudo, embora não fosse o objetivo, surgiu a oportunidade de avaliar a performance de alguns radiologistas quanto a detecção do CaP utilizando os parâmetros prescritos no protocolo PIRADS. Foi aproveitado banco de dados com imagens de RM multiparamétrica da próstata com as respectivas análises da anatomia patológica, disponível gratuitamente para fins acadêmicos em <http://i2cvb.github.io/>, oriundo de iniciativa de colaboração no desenvolvimento de CADs, inclusive já empregado em uma publicação [52] que fazia parte da amostra selecionada neste estudo. Ao término do experimento, foi proposta comparação com os resultados dos grupos CAD e PIRADS.

O banco de dados deste experimento foi composto por 26 pacientes, todos eles sabidamente portadores de câncer de próstata, confirmado por estudo de anatomia patológica realizado em produto de prostatectomia. Todas os pacientes foram submetidos, antes da cirurgia, a RM-MP em equipamento 3T e com bobina endorretal. As imagens no formato DICOM analisadas foram aquelas obtidas na ponderação T2, técnica Difusão (com o respectivo mapa ADC) e dinâmicas ponderadas em T1 após a administração do meio de contraste paramagnético.

Dois radiologistas, então, avaliaram cada exame em estação de trabalho Apple Macintosh com software OsiriX DICOM Viewer (Pixmeo SARL, Berna, Suíça), versão 12.0.0, licença ANVISA 8109273001. Da análise das imagens, a ficha demonstrada na figura 5, vide apêndice B, foi preenchida individualmente e de forma independente por cada radiologista, indicando-se no gráfico, por cores diferentes, a(s) lesão(ões) observada(s). Cada ficha era identificada pelo radiologista que a preencheu (A ou B) e pelo exame anonimizado através da codificação de MIP-PROSTATA-01-0001 a MIP-PROSTATA-01-0026. As respostas dos radiologistas foram comparadas com as respectivas imagens de anatomia patológica, em coloração hematoxilina-eosina, como exemplificado na figura 6, vide anexo. Devido ao caráter multicêntrico do CaP, com lesões simultâneas sincrônicas em um mesmo paciente, para fim de validação, considerou-se somente as duas principais lesões selecionadas pelos radiologistas que obtiveram classificação PIRADS igual ou maior que 3 (de forma semelhante ao proposto por alguns artigos que compõem a amostra de publicações PIRADS selecionada).

Do resultado obtido e matriciados foram construídas tabelas de contingência, conforme orientado por SOARES E SIQUEIRA [53]. Nestas tabelas, os achados do experimento foram confrontados com a anatomia patológica e categorizados quanto ao acerto ou erro do radiologista, para posterior análise estatística quanto ao grau de confiabilidade do julgamento de cada um deles. Foram consideradas, para efeito de análise estatística, a detecção do CaP ao nível do indivíduo e não apenas pela quantidade de lesões observadas. Abaixo, a forma de preenchimento das tabelas de contingência (tabela 1):

Tabela 1 – Tabela de contingência

Doença	Teste		Total
	Positivo	Negativo	
Presente	<i>a</i>	<i>b</i>	<i>a + b</i>
Ausente	<i>c</i>	<i>d</i>	<i>c + d</i>
Total	<i>a + c</i>	<i>b + d</i>	<i>n</i>

Onde:

- a* corresponde a teste positivo em indivíduo portador de doença
- b* corresponde a teste negativo em indivíduo portador de doença
- c* corresponde a teste positivo em indivíduo que não tem a doença
- d* corresponde a teste negativo em indivíduo que não tem a doença
- n* total da amostra

3.4. Ferramentas de análise estatística

Como o objetivo do estudo foi catalogar artigos completos publicados (científicos em periódicos) entre janeiro de 2020 e abril de 2021 sobre a aplicação do CAD no diagnóstico de câncer de próstata através da RM-MP, era interessante realizar metanálise sobre a síntese do conhecimento adquirido. E, para fundamentá-la, foi necessário verificar qual o impacto do método quanto a sensibilidade e especificidade, quais as técnicas de processamento de imagem com melhor acurácia e quais os algoritmos computacionais de melhor performance na detecção do câncer de próstata.

A Revisão Sistemática é um tipo de estudo secundário que visa reunir e sintetizar o conhecimento anteriormente publicado para responder uma pergunta específica, através da busca metódica em várias fontes [54]. É necessária metodologia específica, exata e estrita, tanto

na busca quanto na síntese dos dados, a fim de evitar vieses e assegurar confiáveis resultados que garantam o preciso acerto em uma tomada de decisão [55]. Por se tratar de uma Revisão Sistemática, a metanálise foi escolhida, entre os métodos estatísticos, como a mais adequada para sumarizar resultados de estudos independentes, através da combinação de dados relevantes de todos eles, medindo os efeitos de cada técnica CAD mais precisamente do que cada estudo em separado. Após essa avaliação preliminar, foi necessário agregar as mensurações obtidas com ponderação dos resultados individuais, admitindo-se que houvesse alguma heterogeneidade ou relevância científica entre eles. Assim, foi possível calcular as medidas de impacto e benefício da técnica CAD proposta.

O modelo de análise foi aquele de efeito fixo, onde se assumia que os estudos selecionados representavam uma amostra e que eventuais variações entre eles são resultantes do acaso (erro randômico).

Na apreciação dos dados obtidos, as técnicas estatísticas de mensuração da tendência central foram utilizadas na análise descritiva: a média, mediana, variância e normalidade calculadas e, considerando a distribuição não normal e a natureza não pareada dos dados encontrados, aplicados testes não paramétricos de amostras independentes. Todas as análises foram realizadas em *software* IBM SPSS Statistics para Mac OS, versão 23, sob licença temporária de avaliação.

As diversas variáveis disponíveis dos artigos publicados foram submetidas a análise descritiva. Para as variáveis SENSIBILIDADE, ESPECIFICIDADE e CURVA ROC, o seguinte roteiro foi aplicado visando as análises descritiva e inferencial:

- a. Síntese de média, mediana, desvio-padrão e coeficiente de variação.
- b. Produção de gráficos de normalidade: *box-plot* e Q-Q plot.
- c. Aplicação dos testes de normalidade de Kolmogorov-Smirnov e Shapiro-Wilk.
- d. Aplicação de testes de hipóteses:
 - a. Teste de Mediana de amostras independentes (teste U de Mann-Whitney de amostras independentes).
 - b. Teste de Mediana de amostras independentes (teste H de Kruskal-Wallis de amostras independentes).

A avaliação de dispersão da amostra foi necessária na interpretação da tendência central, sendo comumente utilizado o desvio-padrão e o coeficiente de variação [54]. O entendimento da dispersão através do seu **coeficiente de variação** (a razão do desvio-padrão pela média) permite estimar o quanto há de variabilidade relativa em torno da média e, indiretamente, o seu

grau de representatividade. Dados homogêneos têm menor valor do coeficiente de variação, portanto menor será a dispersão em torno da média:

- Se o valor do coeficiente de variação for menor ou igual a 15%, há baixa dispersão (os dados são homogêneos e existe pouca variabilidade entre eles).

- Se o valor do coeficiente de variação for entre 15% e 30%, há média dispersão.

- Porém, se o valor do coeficiente de variação for maior que 30%, há alta dispersão (os dados são heterogêneos e existe maior variabilidade entre eles).

De forma semelhante, o **desvio-padrão** também sintetiza medida de dispersão dos dados em torno da média amostral. Um pequeno desvio-padrão indica que os dados tendem a estar próximos da média, enquanto um valor grande de desvio-padrão aponta que os dados estão distribuídos por uma ampla faixa de valores.

Os gráficos de normalidade facilitam o entendimento do padrão da amostra. O **diagrama de caixa** (ou *box-plot*) é uma das maneiras gráficas de se representar, de forma resumida e objetiva, a distribuição da amostra em quartis. Em um eixo vertical, uma linha se estende do menor ao maior valor encontrado, interposta por uma caixa que contempla 50% dos valores amostrais e que, portanto, se inicia no primeiro quartil e termina no terceiro, sendo o segundo quartil indicado por uma linha transversal no seu interior (que corresponde a própria mediana) [55]. Valores discrepantes na amostra (também chamados de *outliers*) são plotados individualmente fora da linha. Visualmente, a dimensão da caixa e seus componentes contemplam o grau de dispersão e a assimetria dos dados.

Já os gráficos quantil-quantil (ou *Q-Q plot*) plotam a relação entre os quantis da amostra e os quantis esperados da normalidade, assim, também úteis para avaliar a suposição desta. Quando a distribuição gráfica dos pontos (que representam os valores da amostra) se aproximam de uma reta, supõe-se que há normalidade, porém, pontos que se desviam da linearidade implicam na suposição de variabilidade amostral [56].

Nesta análise, há etapa que visa identificar o tipo de distribuição dos dados e sua natureza quanto a parametricidade. Estatisticamente, tantos métodos gráficos, como o Q-Q plot, quanto ensaios numéricos podem ser aplicados nessa avaliação, sendo os testes de normalidade de **Kolmogorov-Smirnov** e **Shapiro-Wilk** os amplamente utilizados [57]. Se aconselha o teste de Kolmogorov-Smirnov para amostras iguais ou superiores a 50, enquanto para amostras de menor dimensão, o teste de Shapiro-Wilk é o mais indicado. A hipótese nula (H_0) considera

que a amostra testada advém de um mecanismo de distribuição aleatória, gaussiana, logo se tratando de dados paramétricos. Ao rejeitarmos a H_0 temos, portanto, uma distribuição não-gaussiana e dados não paramétricos. Tais testes fornecem um valor de probabilidade (p) da H_0 ser verdadeira:

- se o valor de (p) for $\leq 0,05$: rejeita-se a H_0 e a distribuição é não gaussiana (dados não paramétricos).
- se o valor de (p) for $> 0,05$: confirma-se a H_0 e a distribuição é gaussiana (dados paramétricos).

Os gráficos de normalidade facilitam o entendimento do padrão da amostra. O **diagrama de caixa** (ou *box-plot*) é uma das maneiras gráficas de se representar, de forma resumida e objetiva, a distribuição da amostra em quartis. Em um eixo vertical, uma linha se estende do menor ao maior valor encontrado, interposta por uma caixa que contempla 50% dos valores amostrais e que, portanto, se inicia no primeiro quartil e termina no terceiro, sendo o segundo quartil indicado por uma linha transversal no seu interior [55]. Valores discrepantes na amostra (também chamados de *outliers*) são plotados individualmente fora da linha. Visualmente, a dimensão da caixa e seus componentes contemplam o grau de dispersão e a assimetria dos dados.

Já os gráficos quantil-quantil (ou *Q-Q plots*) plotam a relação entre os quantis da amostra e os quantis esperados da normalidade, assim, também úteis para avaliar a suposição desta. Quando a distribuição gráfica dos pontos (que representam os valores da amostra) se aproximam de uma reta, supõe-se que há normalidade, porém, pontos que se desviam da linearidade implicam na suposição de anormalidade amostral [56].

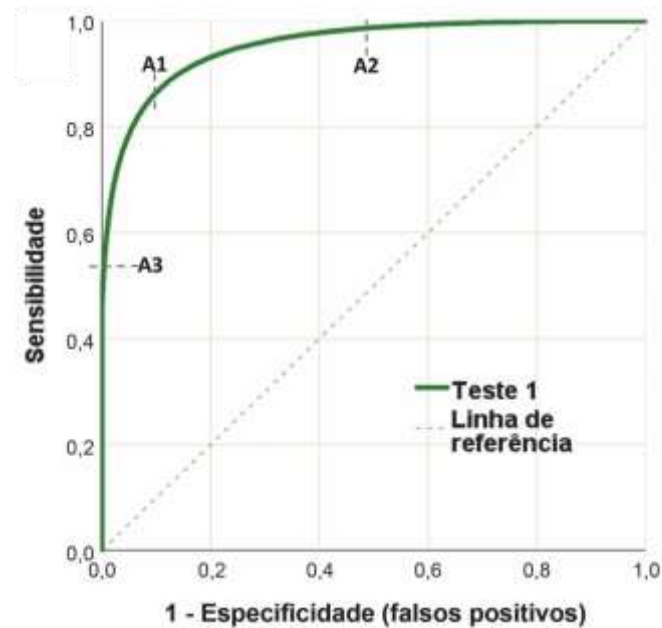
A aplicação do conceito de sensibilidade e especificidade de um exame é rotina na prática médica, como apoio na tomada de decisão frente as incertezas dos fatores que influenciam o processo saúde-doença de um indivíduo, sendo necessário a acurácia do exame a que será submetido. Já que não há método diagnóstico perfeito, sempre encontraremos indivíduos com teste positivo, mas que são saudáveis (**falso positivo**); e indivíduos com teste negativo, mas que são verdadeiros doentes (**falso negativo**), muitas vezes levando em consideração um exame sabidamente com acurácia e precisão superiores, comumente denominado padrão-ouro. A sensibilidade é calculada pela razão entre os indivíduos com teste positivo e doentes (**verdadeiro positivo**) pelo total de doentes; e a especificidade através da razão entre os indivíduos com teste negativo (**verdadeiro negativo**) e o total de não-doentes.

- Sensibilidade = verdadeiros positivos / doentes

- Especificidade = verdadeiros negativos / não-doentes

Logo, é suposto que na prática clínica, testes com elevada especificidade, se positivo, aumentam a chance do indivíduo ser doente; ao passo que testes com elevada sensibilidade, se negativo, diminuem a chance do indivíduo ser doente. Por exemplo, um teste com sensibilidade de 0,8 significa que é capaz de detectar 80% dos indivíduos doentes (se o teste for negativo, menor a possibilidade de estar doente); um teste com especificidade de 0,9 traduz que é capaz de detectar 90% dos sadios entre os não-doentes (se o teste for positivo, maior a possibilidade de ele estar doente). Idealmente, quanto mais próximo de 1 a sensibilidade e a especificidade de um teste, mais segurança clínica o teste fornece na decisão, por exemplo, entre tratar e não tratar um indivíduo com sintomas de uma determinada patologia.

Em se tratando de aprendizado de máquinas e classificação de um achado na imagem, quando queremos avaliar o desempenho de um determinado classificador, a área sob a curva ROC é uma das métricas estatísticas mais importantes. Diante de uma curva de probabilidade (por exemplo, se a lesão detectada é cancerosa ou não), a área sob a curva ROC representa o quantum de separação entre essas categorias, portanto é capaz de mensurar a capacidade do modelo em análise em distinguir as diferentes classes. Também é conhecida como AUC - ROC (*area under the curve - receiver operating characteristic*) ou ainda como Área Sob as Características de Operação do Receptor (AUROC). A área sob a curva ROC é um gráfico da taxa de verdadeiro positivos (portanto, a sensibilidade) em função da taxa de falso positivos (portanto, 1 - especificidade), vide a figura 7, abaixo.

Figura 4 – Curva ROC.

Fonte: Polo TCF, Miot HA, 2020 [58]

De acordo com POLO E MIOT [58], “quanto mais a curva ROC se aproxima do canto superior esquerdo, melhor é a qualidade do teste quanto à capacidade para discriminar os grupos”, sendo que “a linha de referência diagonal da curva ROC representa uma região de completa aleatoriedade do teste, incapaz de classificar tanto doentes como saudáveis (sensibilidade = especificidade)”. Considerando um intervalo de confiança de 95%, quando a área sob a curva ROC é igual a 0,5, teremos a nulidade da hipótese que aquele teste representa. Por outro lado, quanto mais próximo de 1, mais aquele teste terá alta sensibilidade e especificidade. No auxílio da interpretação, os dados encontrados em uma curva ROC podem ser categorizados em “péssimo” (0,5 - 0,6), “ruim” (0,6 - 0,7), “pobre” (0,7 - 0,8), “bom” (0,8 - 0,9) e “excelente” (> 0,9) [63]. A área sob a curva ROC não necessariamente depende de amostras com distribuição normal, porém precisa de um teste padrão-ouro (como o estudo anatomopatológico) para adequada classificação da amostra em casos e controles.

Portanto, serão utilizadas as variáveis SENSIBILIDADE, ESPECIFICIDADE e ÁREA SOB A CURVA ROC para efeito comparativo entre os grupos CAD, PIRADS e EXPERIMENTO COM RADIOLOGISTA, visando responder as hipóteses objetos deste estudo.

Como subsídio para análise inferencial, os testes de hipóteses permitem concluir sobre uma população a partir de dados uma amostra dela. Considerando o tamanho amostral pequeno dos dados das variáveis sensibilidade, especificidade e área sob a curva ROC, e a distribuição assimétrica destes ao redor da média, os **testes não paramétricos**, baseados nas posições das observações, mas não em suas grandezas numéricas, são mais indicados. Por exemplo, o teste de Mann-Whitney compara a mediana ao invés da média (como ocorre no teste t de Student), independente da distribuição populacional.

Na elaboração do **teste de Mann-Whitney**, é definida as hipóteses nulas (H_0) e alternativa (H_1):

- $H_0: n_1 = n_2$ [a mediana de uma população (n_1) é igual à mediana de outra (n_2)]
- $H_1: n_1 \neq n_2$ [a mediana de uma população (n_1) não é igual à mediana de outra (n_2)]
- $H_1: n_1 > n_2$ [a mediana de uma população (n_1) é maior que a mediana de outra (n_2)]
- $H_1: n_1 < n_2$ [a mediana de uma população (n_1) é menor que a mediana de outra (n_2)]

Caso a hipótese nula seja rejeitada enquanto verdadeira, se incorre no erro do tipo I, ao passo que aceitar a hipótese nula sendo ela falsa, incorre-se no erro do tipo II. Para mitigar o risco de cometer este tipo de erro, faz-se necessário estabelecer nos testes nível de significância tal que a ocorrência de tais erros sejam inferiores a 5%. Logo, quando o resultado do teste for menor que o nível de significância estabelecido (por exemplo, 5%), a H_0 será rejeitada e se supõe que existe diferença significativa entre as duas amostras; por outro lado, se assume que ao acaso há um risco de até 5% de se encontrar um erro do tipo I [59].

O teste de **teste de Mann-Whitney (U)** é realizado através da ordenação ascendente de cada dado da amostra, sendo-lhe atribuído uma nota (denominada “posto”), ou seja, aos “postos” menores são aqueles de menor valor numérico. O resultado estatístico é obtido pelo número de vezes que “postos” de uma segunda amostra precede os “postos” da primeira amostra ordenados crescentemente [60]. Abaixo, visualiza-se a equação do teste:

Onde:

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

R₁ =

$$U = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$$

soma dos postos da amostra 1

R₂ = soma dos postos da amostra 2

n₁ = número de casos da amostra 1

n₂ = número de casos da amostra 2

U = valor estatístico do teste

Havendo desigualdade entre os grupos, é necessário confrontar o valor encontrado da mediana com o valor crítico do teste, que é dado por tabelas específicas para cada nível de significância que se deseja:

- H₁: mediana n₁ ≠ mediana n₂ Rejeita H₀ se a mediana < valor crítico
- H₁: mediana n₁ > mediana n₂ Rejeita H₀ se a mediana n₂ < valor crítico
- H₁: mediana n₁ < mediana n₂ Rejeita H₀ se a mediana n₁ < valor crítico

O teste **H de Kruskal-Wallis** [61] pode ser considerado como uma extensão do teste de Mann-Whitney, usado quando há mais de dois grupos para serem analisados. Ao invés de testar igualdade de médias ou de medianas, o teste de Kruskal-Wallis avalia a hipótese de distribuição igual entre três ou mais amostras. O teste supõe que, em cada amostra, os dados são diferentes e independentes. Para tal, as mensurações obtidas em cada amostra são ordenadas (semelhante ao teste de Mann-Whitney) e, para cada posição observada (também denominada “posto”), um valor é atribuído. Mas complexo que seu correspondente para duas amostras, o teste de Kruskal-Wallis pode ser calculado de alguns modos, dependendo da quantidade de empates nos “postos”. No geral, aplica-se a seguinte equação, quando empates não são observados:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1)$$

Donde:

k = número de amostras comparadas

N = número total de medições observadas

R_i = soma dos “postos” de cada amostra

n_i = número de medições observadas em cada amostra

H = valor estatístico do teste

Neste teste, as hipóteses nulas (H₀) e alternativa (H₁) são:

- H₀: n₁ = n₂ = n₃ = n_x [as amostras n₁, n₂, n₃ ... n_x são iguais]
- H₁: há pelo menos uma amostra diferente

A análise deverá comparar o valor estatístico obtido ao valor crítico do teste, obtido na tabela do Qui-Quadrado (x²) para tantos graus de liberdade quantos sejam a quantidade de amostras menos 1. Se o valor H for superior ao valor crítico, deve-se então rejeitar a hipótese nula, logo considerar que uma das amostras é diferente. Para descobrir qual (ou quais) das amostras dentro os grupos é (ou são) a(s) diferente(s) é necessário aplicar comparações múltiplas entre cada grupo de amostra.

Os testes não paramétricos são a opção para aquelas circunstâncias na qual não há as condições necessárias para se utilizar um teste paramétrico [53], como por exemplo, ao depararmos com amostras gaussianas e não gaussianas, ou mesmo amostras muito pequenas para assegurarmos sua normalidade. Embora os testes não paramétricos não dependam do modelo de distribuição da amostra (sobretudo quando se tem uma amostra de tamanho reduzido), o poder estatístico é menor que seus correspondentes paramétricos, que são propostos para aplicação em distribuições normais.

4. RESULTADOS E DISCUSSÕES

4.1. Dados coletados

As palavras chaves escolhidas abrangiam quatro termos e suas variações: CAD, RM, PIRADS e “câncer de próstata”. Na primeira aplicação da expressão de busca, foi observado número restrito de artigo retornados (63) na ferramenta PUBMED. Como nem todo artigo retornado abordava a padronização PIRADS, a possibilidade de bons artigos sobre CAD e câncer de próstata terem sido excluídos obrigou a retirada do termo PIRADS do mecanismo de busca, o que então ampliou para 73 o número de artigos obtidos.

Nesse momento, foram identificados inicialmente quatro artigos que contemplavam o escopo desta pesquisa e altamente pertinentes, selecionados, portanto, como artigos potenciais candidatos ao Grupo Controle. Destes, procurou-se palavras-chaves importantes que compusessem a expressão de busca, que foram avaliadas em rodadas de testes.

Ainda nesta fase, observou-se a necessidade de adequadamente definir a sigla CAD. Para isso foi utilizado o catálogo de descritores MESH (<https://www.ncbi.nlm.nih.gov/mesh/>) do *The National Center for Biotechnology Information*. Não foi encontrado o descritor “*Computer-Aided Diagnosis*”, mas o mecanismo de busca retornou o descritor “*Diagnosis, Computer Assisted*” como semelhante, e os seguintes sinônimos:

Computer-Assisted Diagnosis

Computer Assisted Diagnosis

Computer-Assisted Diagnoses

Entretanto, diversos artigos utilizam o termo “*aided*” ao invés de “*assisted*”, inclusive em seus títulos, o que obrigou a manutenção de ambos nos testes de calibração.

Como teste das variáveis do mecanismo de busca, em 19/01/2019, inicialmente utilizamos a frase

(“computer-aided diagnosis” OR “computer-assisted diagnosis” OR CAD) AND MRI AND
"prostate cancer"

no mecanismo de busca ScienceDirect (www.sciencedirect.com), o que retornou 544 resultados. Como em algumas referências de artigos no Grupo Controle utilizavam a palavra-chave “*computer-aided detection*” como referência às técnicas CAD, esta foi acrescentado a busca, resultando na frase

(“computer-aided diagnosis” OR “computer-assisted diagnosis” OR “computer-aided detection” OR CAD) AND MRI AND "PROSTATE CANCER"

que retornou 558 resultados. Da mesma forma, acrescentamos o termo “*magnetic resonance imaging*” à expressão de busca, ampliando um pouco mais o universo disponível para 727 artigos.

A frase resultante desta primeira rodada foi:

(“computer-aided diagnosis” OR “computer-assisted diagnosis” OR “computer-aided detection” OR CAD) AND (MRI OR magnetic resonance imaging) AND "PROSTATE CANCER"

Ajustando a expressão de busca, ainda no mecanismo ScienceDirect, acrescentou-se aos termos de pesquisa “*multiparametric*”, visando estreitar artigos que utilizassem modelos PIRADS na aquisição e interpretação de imagens prostáticas. Foram retornados 84 resultados. A frase utilizada nesta rodada foi:

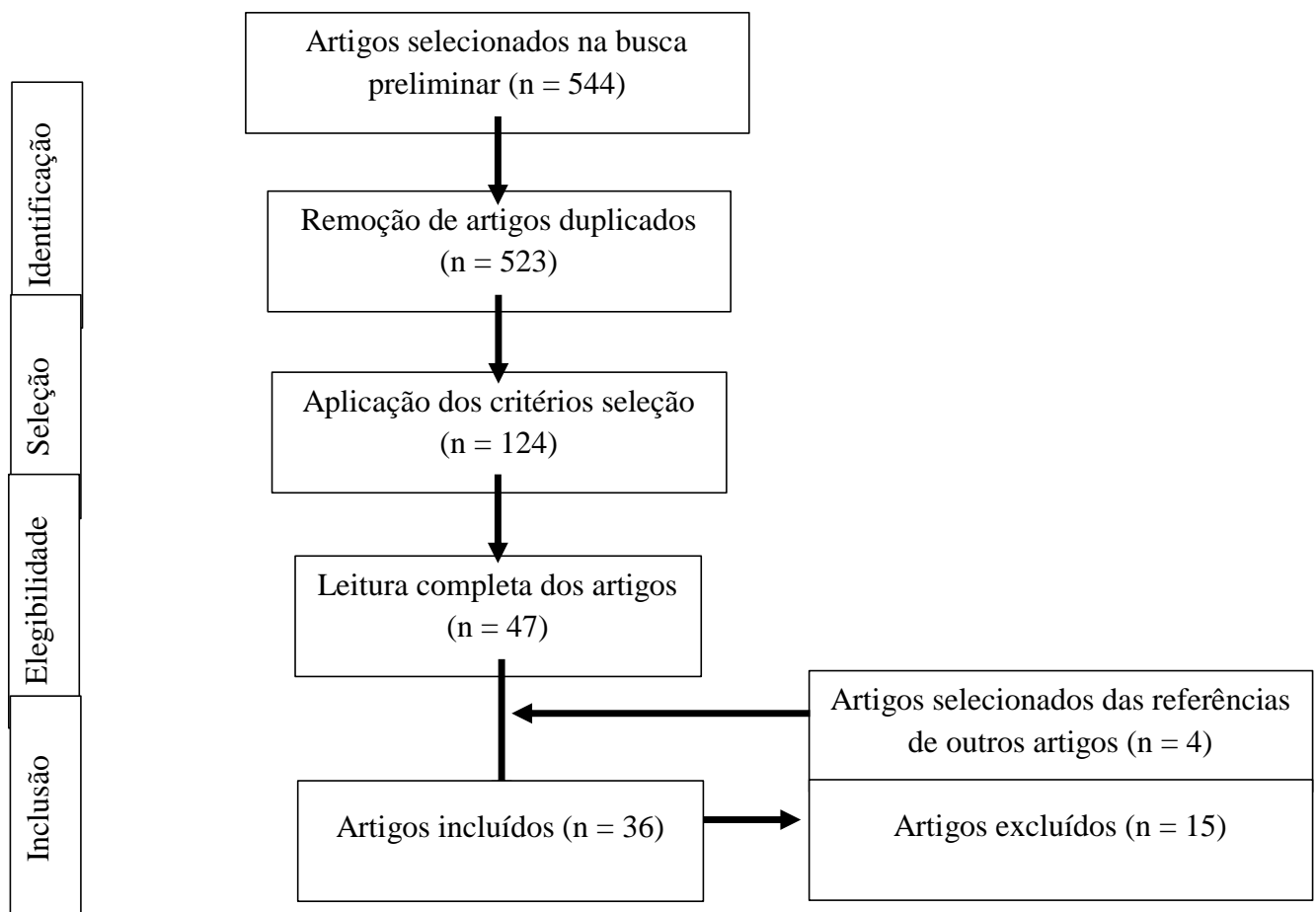
(“computer-aided diagnosis” OR “computer-assisted diagnosis” OR “computer-aided detection” OR CAD) AND (MRI OR magnetic resonance imaging) AND multiparametric AND "prostate cancer"

Aplicando-se a frase resultante da rodada anterior no mecanismo de busca PUBMED CENTRAL (www.ncbi.nlm.nih.gov), foram retornados 209 artigos. Da avaliação preliminar, alguns são claramente inadequados ao estudo e deverão ser excluídos por filtro específico. No mecanismo de busca IEEE (www.ieeexplore.ieee.org) foram retornados 17 artigos, diversos

com alta relevância para esta pesquisa. Porém no mecanismo de busca Cochrane Library (www.cochranelibrary.com), foram retornados 3 artigos, somente um relevante a pesquisa, outro resumo de congresso e outro com tema semelhante, mas não objetivo deste estudo. Por último, no mecanismo de busca do Portal de Periódicos da CAPES, que realiza pesquisa simultaneamente em diversas bibliotecas, foram obtidos 316 artigos, em língua inglesa. Como artigos retornados em uma base de dados nem sempre o foram em outra, sobretudo aqueles selecionados na IEEE, foi optado por utilizar os quatro motores de busca descritos (ScienceDirect, PUBMED, IEEE e Cochrane Library) acrescido da busca final de artigos, porventura não detectados anteriormente, através de múltiplos mecanismos no Portal de Periódicos da Capes.

Dos artigos avaliados como potenciais candidatos a grupo controle, o mais antigo é datado de 2011, sendo este ano definido como data mínima a partir dos quais os artigos foram refinados nos mecanismos de busca. A ilustração abaixo (figura 8) sintetiza a evolução da seleção de dados.

Figura 5 – Evolução do processo de seleção de artigos para análise sistemática.



Para seleccionar os artigos que comporão a base de dados para correlação com o PIRADS, utilizamos a frase

PI-RADS AND VALIDATION AND MRI AND "PROSTATE CANCER"

Aplicando-a nas ferramentas de busca, obtivemos:

- a) Mecanismo de busca do Portal de Periódicos da CAPES: 22 artigos.
- b) ScienceDirect: 477 artigos
- c) PUBMED: 35 artigos
- d) IEEE: 1 artigo
- e) Cochrane Library: 9 artigos

Os mesmos critérios foram utilizados na seleção, resultando em 10 artigos selecionados sobre o PIRADS, sendo que alguns deles eram metanálises e referências na literatura sobre o tema.

4.2. Descrição das variáveis de interesse

Da apreciação da literatura selecionada, diversos dados mostraram-se de interesse para avaliação deste estudo, sobretudo aqueles que relacionam desfechos da aplicabilidade clínica do CAD.

Há, nos últimos anos, um aumento no número de publicações científicas sobre o tema CAD e Ressonância Magnética, sobretudo na área médica, o que corrobora no interesse dos pesquisadores em consolidar o método como ferramenta na prática clínica.

Em uma busca livre com os descritores “CAD” e “RM” no Portal de Periódicos da CAPES foram identificados 10.054 trabalhos entre os anos de 2010 e 2019, sendo 9.625 artigos. Comparado a década anterior, quando foram publicados 3.779 artigos, observa-se um incremento de 2,5 vezes, aproximadamente, no número de trabalhos.

Especificamente na abordagem do CaP, predominam artigos relacionados a técnica e a performance dos sistemas CAD, com várias abordagens diferentes no processamento de imagem e nos métodos de quantificação [17], tais como sequências ponderadas em T2 que são utilizadas na delimitação tumoral através da forma, característica do sinal e da textura; classificadores de tecido normal e tumor; análise estatística de *voxel*; análise de difusibilidade

restrita, mapas de difusão aparente (ADC) e curvas farmacocinéticas do contraste intravenoso empregado; entre outros.

Fei *et al* [17] identificaram quatro características que demonstram a utilidade do CAD para o diagnóstico do CaP através da RM-MP: o aumento da sensibilidade com a redução do tempo de análise; o aumento da especificidade; o aumento da performance de radiologistas menos experientes quando comparado àqueles mais experientes; e a melhor diferenciação entre as lesões indolentes daquelas de alto grau de malignidade.

Em se tratando da performance do sistema CAD comparada ao leitor (o médico radiologista) na interpretação das lesões prostáticas na RM, áreas sob curvas ROC semelhantes foram encontradas, quando se tratando de leitores experientes. Porém, em se tratando de radiologistas pouco experientes, o uso de um sistema CAD aumentou significativamente o desempenho na determinação de lesões, tanto na zona periférica quanto na zona de transição [62]. Com o CAD, radiologistas menos experientes conseguem atingir desempenho similar ao daqueles experientes (área sob a curva ROC de 0,91 e 0,93 respectivamente).

Outro estudo definiu que o uso de um sistema CAD tem o potencial de melhorar a interpretação das imagens da próstata por RM, especificamente nos protocolos multiparamétricos, não somente relacionada aos radiologistas menos experientes, mas também naqueles com maior experiência, sobretudo por melhorar a caracterização de lesões benignas. Nestas, os sistemas CAD têm mostrado performance satisfatória, aumentando a capacidade do radiologista em caracterizar alteração focal na zona periférica da próstata de 19,7% para 40,6% quando aplicado o CAD, com especificidade passando de 79% para 86,2% [63].

Mesmo para os radiologistas experientes, o CAD permitiu detectar maior número de lesões com maior grau de malignidade (aquelas com escore de Gleason maior que 6) [64]. Neste mesmo estudo, foi determinado melhor detecção das lesões com mais de 1,0 cm e redução no tempo total de análise, impactando positivamente o fluxo do serviço e, portanto, reduzindo a carga de trabalho do médico avaliador.

A preocupação em se detectar precocemente aquelas lesões com escore de Gleason elevado, antes mesmo de se tornarem clinicamente significantes, também está presente em estudos. A combinação do mapa de coeficiente de difusão aparente (utilizando-se o 10º percentil do mapa ADC), o coeficiente de difusão aparente médio e a assimetria da intensidade do sinal nas imagens ponderada em T2 (*skewness*) mostraram-se efetivos em identificar o CaP no tecido normal da zona periférica [65]. Nestes, o 10º percentil do ADC, a média do ADC e a constante de transferência de volume (K^{trans}) estão moderadamente correlacionados com o

escore de Gleason da lesão e moderadamente eficazes em separar os CaP de baixo grau (escore de Gleason até 6) dos de alto grau (escore de Gleason maior ou igual a 7).

Outra abordagem interessante de alguns estudos foi discriminar lesões baseadas na categorização PIRADS: através de um sistema CAD de probabilidade de malignidade que, em conjunto com o escore PIRADS determinado pelo radiologista, permitia selecionar quais lesões categorizadas como 3, 4 e 5 requerem biópsia. Aumentando o limiar de biópsia de PIRADS 3 para 4 e utilizando a probabilidade de malignidade fornecida pelo sistema CAD, não há perda significativa da detecção de lesões mais agressivas (sensibilidade 0,99 contra 0,98), ao passo que aumentava a especificidade do método de 0,26 para 0,59, reduzindo, portanto, biópsias desnecessárias [66].

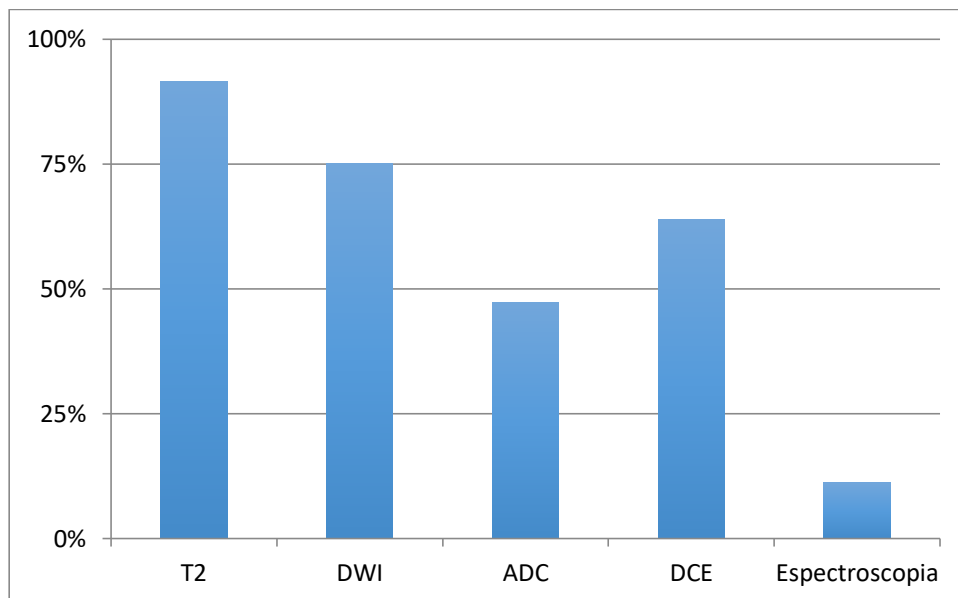
Além da utilização da RM-MP como ferramenta diagnóstica, ao se propor aplicação do método como estratégia de rastreamento, era desejado protocolos de aquisição de imagem rápidos e discriminatórios para seleção daqueles indivíduos que se submeteriam à análise completa. Há sistema CAD empregando apenas duas sequências do protocolo multiparamétrico (T2 e DWI ou T2 e DCE), inclusive outros aplicando somente a ponderação T2. Um estudo obteve resultado favorável utilizando as ponderações T2 e DWI na distinção de CaP das lesões benignas identificadas na RM, com área sob a curva ROC de 0,83; ainda maior (0,89) quando se considerava distinção de lesões benignas presentes ou não na RM, independente da localização (zona de transição ou zona periférica) [67]. Combinando a ponderação T2 com a DCE, houve aumento da acurácia de 0,85 para 0,89 quando comparada com o uso isolado das imagens T2 [68]. Usando um conjunto de descritores discriminantes de textura extraídos das imagens exclusivamente ponderadas em T2 e aplicados a diferentes classificadores, um estudo obteve área sob a curva ROC de 0,90 (\pm 7,6%) [69].

Não há, ainda, padronização quanto a aplicabilidade do CAD no diagnóstico do CaP. Embora o protocolo PIRADS estabelecesse diretrizes para categorização das lesões suspeitas para acometimento neoplásico, a semelhança do que é realizado na mama através do BIRADS (Breast Imaging-Reporting and Data System), uma classificação desenvolvida para ser utilizada com o CAD de próstata também será necessária, sobretudo diante das diversas variáveis passíveis de análise através da máquina e outrora não contempladas na percepção humana. Ademais, a disputa das empresas, que desenvolvem equipamentos de Ressonância Magnética pelo mercado diagnóstico em expansão, propicia pesquisas e surgimento de novas técnicas de aquisição de imagem, que rapidamente são incorporadas a clínica. Com isso, outras variáveis estarão disponíveis ao CAD e farão parte de novas pesquisas na área.

4.3. CAD

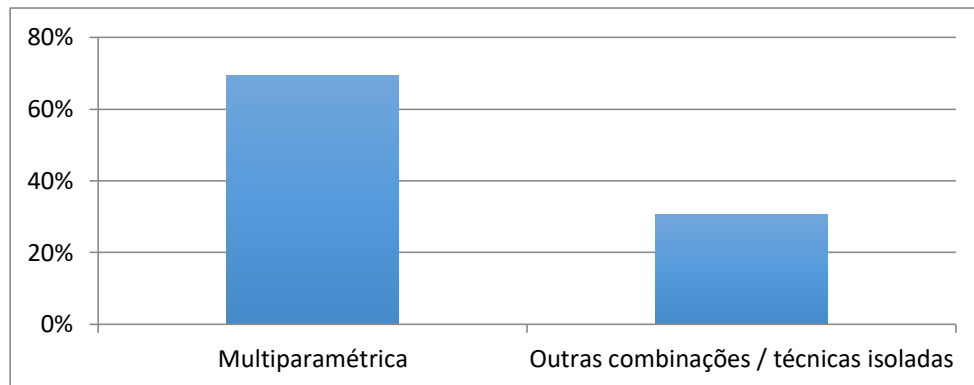
Os 36 artigos selecionados foram categorizados e analisados quanto às suas características, pormenorizadas na tabela 5, do apêndice A. Quase a totalidade dos artigos utilizaram as ponderações T2, e uma grande maioria também empregaram imagens com técnica Difusão (DWI e ADC) e hemodinâmicas (DCE), vide figura 9 abaixo (dados originais na tabela 6, do apêndice A). Disso se pode depreender que as técnicas anatômicas e funcionais (ponderação T2 e DWI / ADC, respectivamente) são as mais preferidas, sobretudo ao aliar tanto a detecção do CaP pela anomalia de sinal (ou até mesmo pela distorção da imagem na estrutura em si), como também pelas alterações metabólicas e celulares que existem no seu interior. Logo, a associação destas técnicas aplicadas no CAD está em consonância com o preconizado nos protocolos multiparamétricos, a exemplo do recomendado no PIRADS, como se observa na figura 10.

Figura 6 – Ponderações de RM utilizadas nos estudos, por percentual de publicações encontradas.



Fonte: o autor, 2021

Figura 7 – Comparação entre os tipos de técnicas utilizadas pelos CAD nos estudos, em percentual.



Fonte: o autor, 2021

A utilização de mais de uma ponderação, conforme já visto, aumenta capacidade do exame na detecção do CaP. Das diversas combinações identificadas nos artigos selecionados, a mais utilizada é a T2, DWI, DCE (aquele que mais emprega as características anatômicas, funcionais e hemodinâmicas, sendo responsável por 60% do conjunto de artigos selecionados), seguido de técnicas sem utilização do meio de contraste paramagnético (28%), vide tabela 7.

Tabela 2 – Técnicas multiparamétricas utilizadas nos estudos

Técnica	Número de estudos	%
T2 e DWI ou T2, DWI, ADC	7	28
T2, DWI, DCE ou T2, ADC, DCE	15	60
T2, DWI, ADC, DCE	3	12
Total multiparamétrico	25	100

Quanto a obtenção do espécime anatomopatológico para confirmação da suspeita do exame de imagem, a maior parte dos estudos utilizou produtos de prostatectomia, seja por seções histológicas ou não, sendo que estas favorecem a análise comparativa com as imagens de RM. Em 10 artigos (27,8%), tal material foi oriundo de biópsias, e em outros 2 (5,6%) foram

utilizados tanto seções histológicas quanto biópsias, e na maioria (24 artigos, 66,7%) somente seções histológicas (tabela 8).

Tabela 3 – Origem do produto para validação do diagnóstico imagiológico, por quantidade de estudos publicados.

Produto	Número de estudos	%
Biópsia	10	27,8%
Seções histológicas	24	66,7%
Biópsia e seções histológicas	2	5,6%
Total	36	100%

Quando se analisa o tipo de validação do CAD pelo número de amostras, há equivalência entre os estudos que usaram somente biópsia (43,1%) e que empregaram seções histológicas (43,8%). A combinação de biópsia e seções histológicas teve número limitado de amostras (13,1%, vide tabela 9).

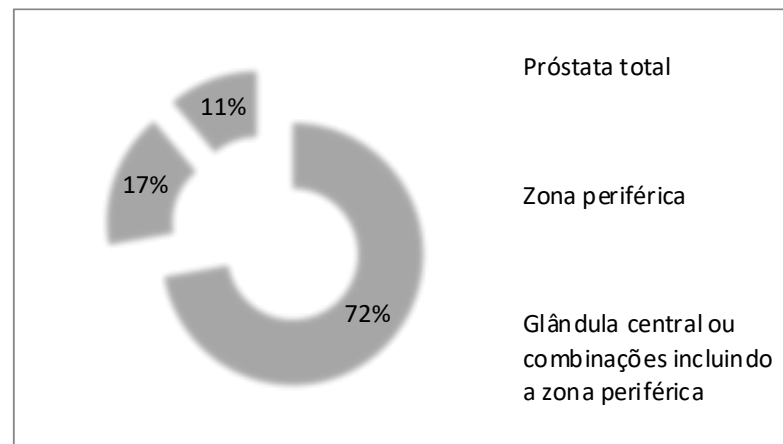
Tabela 4 – Origem do produto para validação do diagnóstico imagiológico, por número de amostras.

Produto de validação	Número de amostras	%
Biópsia	1.164	43,1%
Seções histológicas	1.184	43,8%
Biópsia e seções histológicas	353	13,1%
Total	2.701	100%

É interessante que a análise do CAD englobe a totalidade da glândula prostática (haja visto que o CaP pode ocorrer em qualquer região do órgão, com predomínio na zona periférica), embora nem todos os estudos utilizassem desta abordagem (figura 11): 10 deles se detiveram a

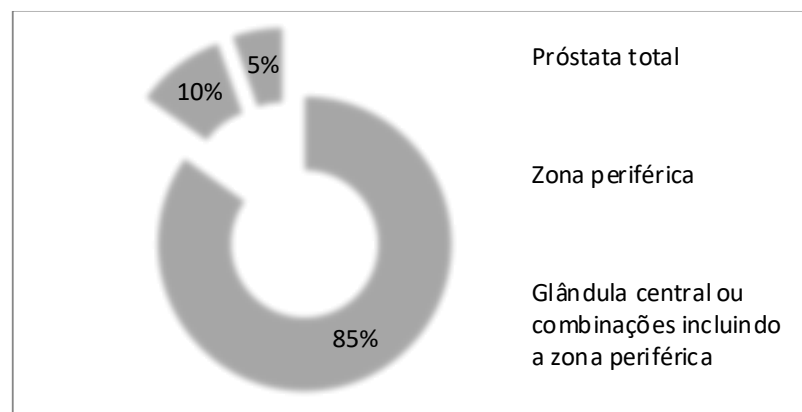
avaliar regiões específicas como a zona periférica somente (6 artigos, 17%) ou combinação de regiões (4 artigos, 11%), porém a maioria das publicações selecionadas (28 artigos, 72%) aplicou o CAD em toda a próstata (dados originais na tabela 10, do apêndice A). Quando se avalia pelo número de amostras (figura 12), grande parte dela, 85%, foi submetida a análise CAD (pormenorizado na tabela 11 do apêndice A).

Figura 8 – Regiões de interesse da aplicação CAD na próstata, em percentual de artigo.



Fonte: o autor, 2021

Figura 9 – Regiões de interesse da aplicação CAD na próstata, por amostras, em percentual.

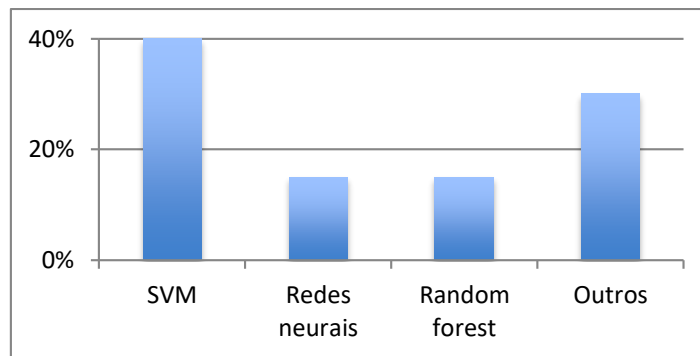


Fonte: o autor, 2021

O tipo de classificador CAD utilizado foi inconstante, predominando as máquinas de vetor de suporte (SVM), seguido das redes neurais e do *random forest*, juntos contribuindo com

28 artigos (70% do conjunto de dados), conforme figura 13. Os demais classificadores (sob a denominação *outros* no gráfico) correspondem a métodos de regressão linear, LDA (*linear discriminant analysis*), QDA (*quadratic discriminant analysis*), além de classificadores próprios e comerciais, inclusive não explicitamente declarados. Há de se salientar que, nos estudos selecionados, a presença de quatro (4) que contemplaram mais de um classificador na análise: dois avaliando SVM e LDA; e outros dois SVM e *random forest* (tabela 12, abaixo).

Figura 10 – Tipos de classificadores CAD, em percentual.



Fonte: o autor, 2021

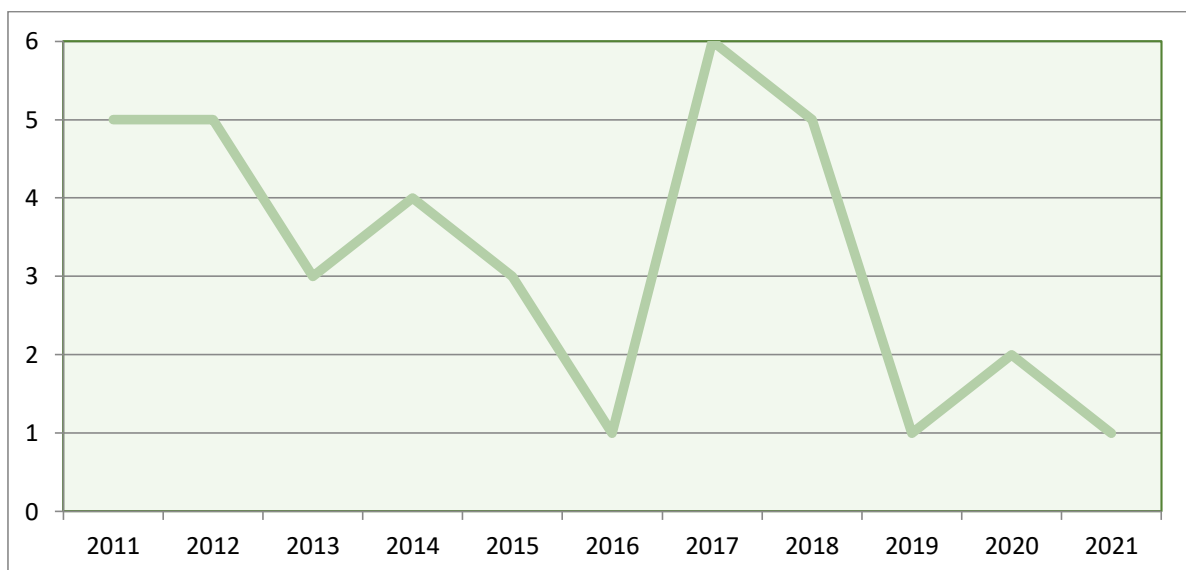
Tabela 5 – Tipo de classificador CAD por quantidade de estudos publicados

Tipo de classificador	Número de estudos	%
SVM	16	40%
Redes neurais	6	15%
<i>Random forest</i>	6	15%
Outros	12	30%
Total	40	100%

O número de amostras utilizadas nos artigos também foi variável, porém se observa incremento em pesquisas publicadas nos anos de 2017 e 2018, conforme a figura 14, abaixo (dados originais na tabela 13, do apêndice A). A média encontrada foi 75 amostras por artigo, totalizando 2.701 neste trabalho (embora, em quatro artigos a mesma amostra foi utilizada em

mais de um tipo de classificador CAD). Em um estudo semelhante, que avaliou características radiômicas na extração de biomarcadores em estudos de imagem [70], com dados obtidos até meados do primeiro semestre de 2017, notou-se crescimento do número de publicações no período de 2010 a 2016. Não há aparente causa para redução do número de publicações específicas de CAD voltados para avaliação de próstata nos anos 2019 e 2020, encontradas nesta revisão. Porém, no processo de seleção e exclusão de artigos, nos anos citados, se observa que alguns já contemplavam estudos voltados para automatização do processo segmentação prostática, que nos anos anteriores era realizado de forma manual, a sugerir tendência dos autores neste sentido.

Figura 11 – Número de artigos publicados por ano de publicação.



Fonte: o autor, 2021

A tabela 14, abaixo, sintetiza a distribuição dos tipos de classificadores quanto ao número de artigos encontrados e número de amostras empregadas. Apesar do maior número de pesquisas que utilizaram o classificador SVM, ao analisar a quantidade de amostras empregadas em cada tipo de abordagem CAD, verifica-se que as redes neurais e as *random forest* são as que mais as demandaram, totalizando 1.540 delas. Ressalva-se que quatro artigos fizeram uso de dois tipos de classificadores, de forma independente (sendo dois aplicando SVM e LDA; e outros dois SVM e *random forest*), e que, para efeito de análise, foram considerados individualmente. Interessante notar na figura 15 que, ao longo da abrangência temporal deste estudo, houve recente predileção pelas redes neurais e *random forest* em detrimento do SVM

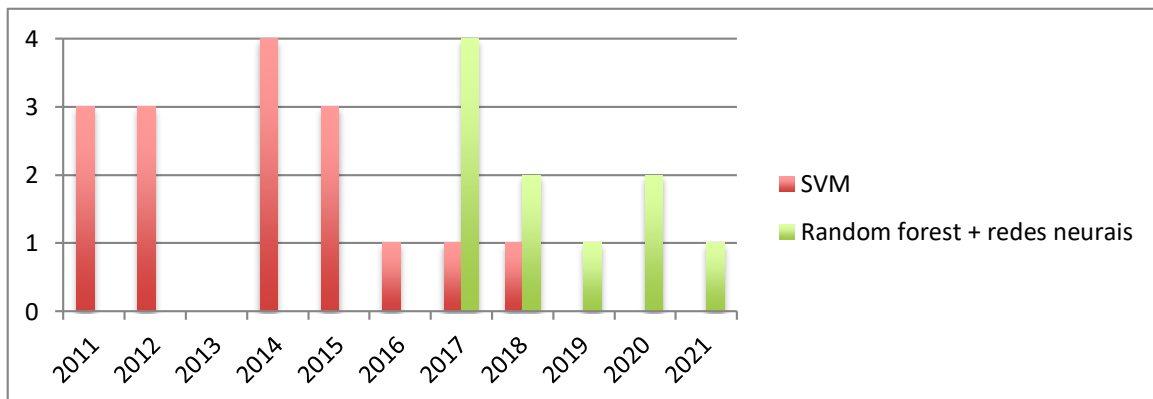
no início da década passada (dados originais na tabela 15 do apêndice A), o que indica tendência do setor em aproveitar modelos de múltiplas saídas.

Tabela 6 – Tipo de classificador por número de amostra e por número de publicações

Tipo de classificador CAD	Número de artigos (%)		Número de amostras (%)	
	Número	Porcentagem	Número	Porcentagem
SVM	16	40%	564	20,0%
Redes Neurais	6	15%	973	34,5%
<i>Random Forest</i>	6	15%	567	20,1%
Outros	12	30%	719	25,5%
Total	40*	100%	2.823*	100%

* Quatro artigos utilizaram a mesma amostra em um segundo tipo de classificador.

Figura 12 – Número de artigos publicados por ano de publicação e por tipo de classificador.



Fonte: o autor, 2021

A tabela 16 do apêndice A descreve os resultados e a performance obtidos de cada artigo publicado. É possível verificar que a maior parte das pesquisas publicadas utilizaram a **área sob a curva ROC** como métrica de análise de desempenho, seguido da estimativa de **sensibilidade** e **especificidade** (tabela 17, abaixo), embora alguns artigos façam uso dos três processos. Quando se avalia exclusivamente o tipo de metodologia empregada quanto ao

número de amostras utilizadas em determinada pesquisa, a área sob a curva ROC também desponta como a mais prevalente.

Tabela 7 – Métricas de desempenho empregadas, por número de amostra e por número de publicações

Métrica	Número de artigos (%)		Número de amostras (%)	
	Área sob a curva ROC	27	42,2%	2.327
Sensibilidade	19	29,7%	942	23,5%
Especificidade	18	28,1%	742	18,5%
Total	64	100%	4.011	100%

Tabela 8 – Número de publicações, por métrica de desempenho empregada e tipo do classificador CAD

Tipo de classificador CAD	Número de artigos, por métrica		
	Área sob a curva ROC	Sensibilidade	Especificidade
SVM	7	10	10
Redes Neurais	6	2	2
<i>Random Forest</i>	3	2	2
Outros	11	5	4
Total	27	19	18

Na tabela 18 nota-se o detalhamento da métrica de desempenho empregada em relação ao tipo de classificador. Novamente há predominância das máquinas de vetor de suporte, porém já se vislumbra a escassez de dados dos classificadores redes neurais e *random forest* para as variáveis sensibilidade e especificidade.

Destes dados gerais obtidos e analisados, depreendeu-se a estatística descritiva apontada na tabela 19. Os testes de normalidade das variáveis área sob a curva ROC, sensibilidade e

especificidade se encontram sumarizados na tabela 20, abaixo, e o histograma e os gráficos Box-plot e Q-Q plot correspondentes a seguir apresentados.

Tabela 9 – Resultados gerais das métricas de desempenho empregadas

Métrica	Mínimo	Máximo	Média	Desvio padrão	Mediana	Coefficiente de variação
Área sob a curva ROC	0,63	0,97	0,849	0,094	0,89	11,1%
Sensibilidade	0,47	1	0,799	0,141	0,83	17,6%
Especificidade	0,39	0,94	0,796	0,142	0,87	17,8%

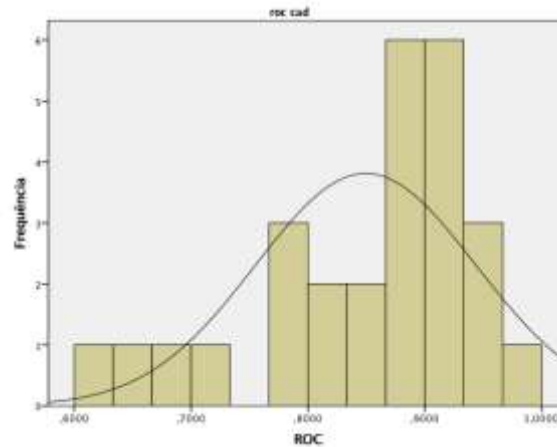
Tabela 10 – Teste de normalidade das métricas de desempenho **área sob a curva ROC, sensibilidade e especificidade.**

Métrica	Testes de Normalidade	
	Kolmogorov-Smirnov	Shapiro-Wilk
Área sob a curva ROC	0,001	0,006
Sensibilidade	0,200	0,154
Especificidade	0,012	0,003

A elevada média da métrica de desempenho área sob a curva ROC (0,852), com pequeno desvio-padrão, demonstra uma boa qualidade do resultado deste teste, e seu pequeno coeficiente de variação (11,3%) evidencia uma baixa dispersão (homogeneidade dos dados), o que é confirmada pela análise visual do seu box-plot (figura 17). Destaca-se, neste, aproximação da mediana ao terceiro quartil e presença de *outlier* isolado logo abaixo do primeiro. Tal *outlier* é referente a artigo que avaliou a detecção do CaP através de um CAD comercial, visando o aperfeiçoamento das biópsias guiadas por USTR, sendo validada por estudo anatomopatológico de produto de prostatectomia [71]. Há de se observar que a *performance* fora da média referida por esta publicação não foi somente com a métrica de desempenho área sob a curva ROC, mas também com as métricas sensibilidade e especificidade. Os testes de normalidade descritos na

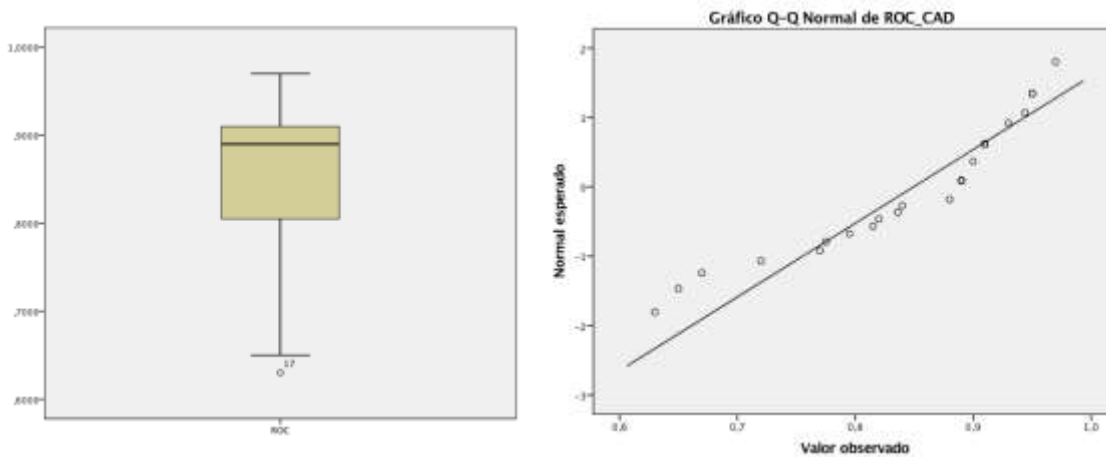
tabela 20 (tanto o de Kolmogorov-Smirnov quanto o de Shapiro-Wilk), o histograma da variável (figura 16) e o padrão gráfico dos valores da amostra, com aproximação da linha de tendência central no Q-Q plot (figura 17), permitem supor que os dados tendem a distribuição normal.

Figura 13– Histograma da métrica área sob a curva ROC (ROC_CAD)



Fonte: o autor, 2021

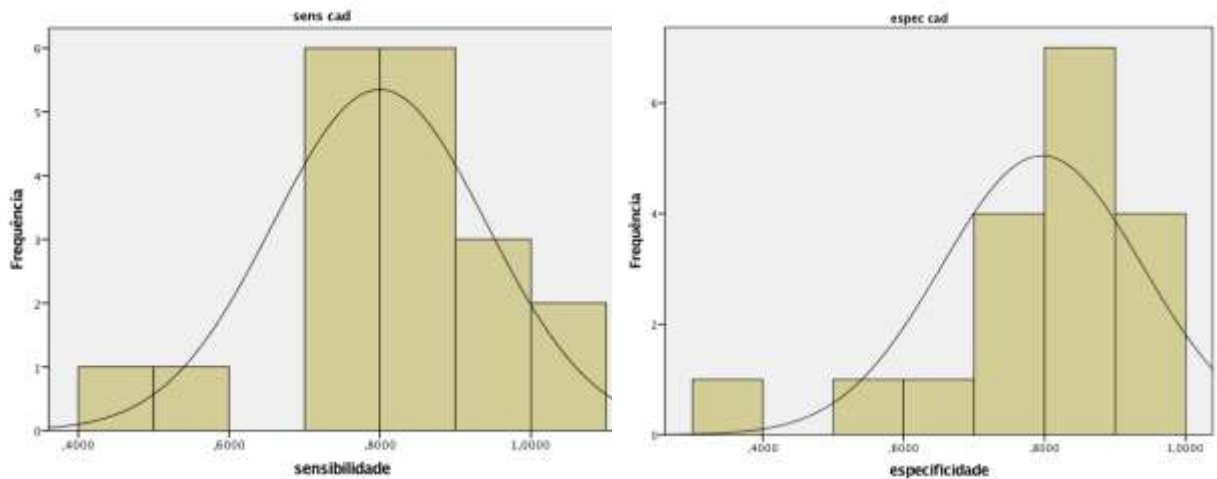
Figura 14 – Box-plot e Q-Q plot da métrica área sob a curva ROC (ROC_CAD)



Fonte: o autor, 2021

Diferente do que foi evidenciado com a métrica de desempenho área sob a curva ROC, a distribuição normal dos dados das métricas sensibilidade e especificidade não foi confirmada, vide o histograma da figura 18 e a tabela 20. A dispersão dos valores, que se afastam da linha de tendência central, pode ser observada na figura 19, que visualmente reforça tal percepção.

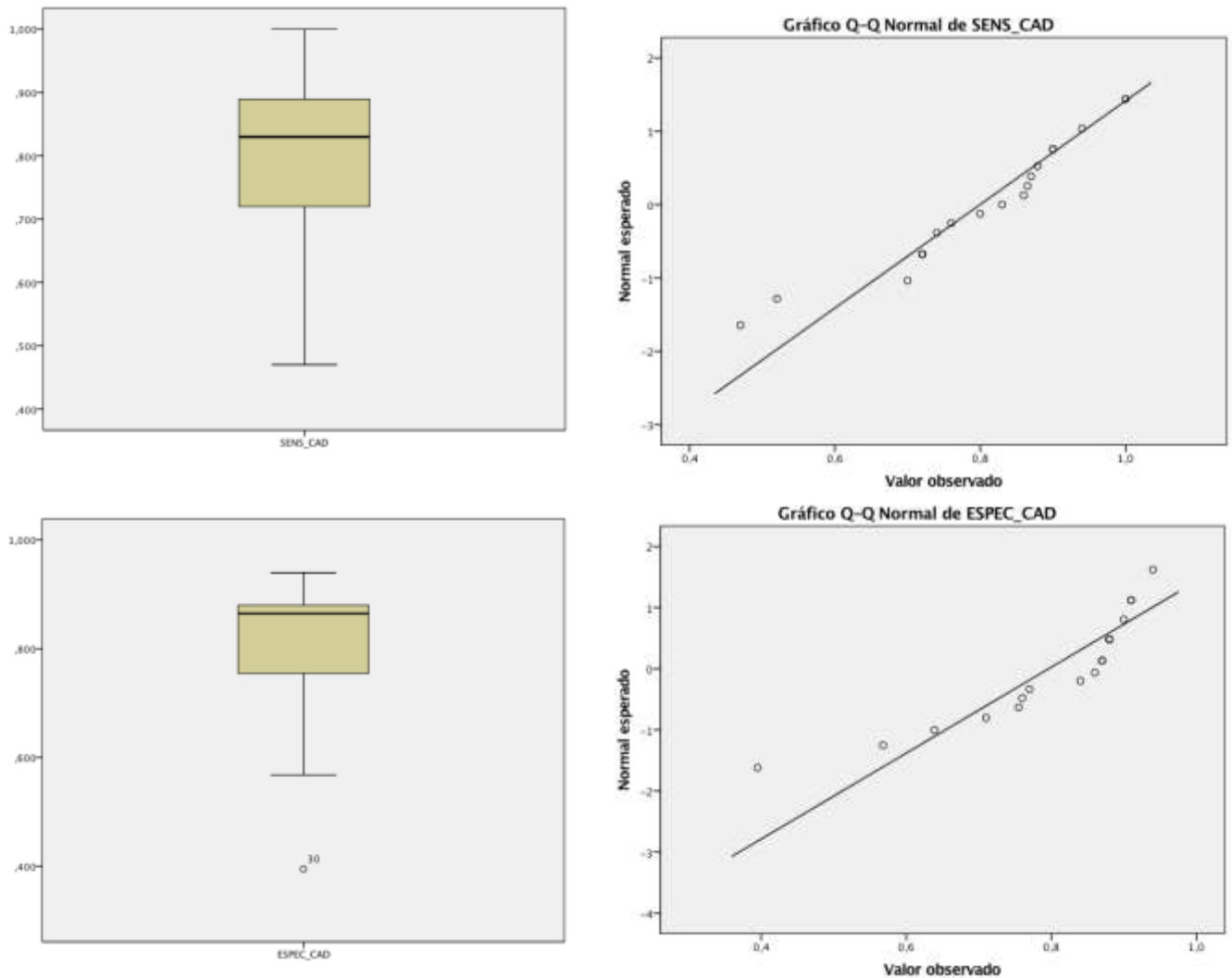
Figura 15 – Histograma das métricas sensibilidade (SENS_CAD) e especificidade (ESPEC_CAD).



Fonte: o autor, 2021

Ainda analisando a tabela 19, embora no ambiente clínico as médias das métricas sensibilidade e especificidade (0,799 e 0,796, respectivamente) sejam interessantes no contexto de tomada de decisão quanto ao emprego do CAD como ferramenta de auxílio diagnóstico do CaP, ao depararmos com as dispersões apontadas pelos desvios-padrão (0,141 e 0,142) e coeficientes de variação (17,6% e 17,8%), a homogeneidade intermediária e a variabilidade de tais dados não fornecem tanta sustentação neste sentido (box-plots da figura 19). Interessante notar que, de todas três métricas, a especificidade foi aquela que apresentou box-plot com mediana mais próxima do 3º percentil, denotando maior assimetria destes dados (que, inclusive, já era perceptível pelo padrão assimétrico da distribuição do histograma da figura 18). A métrica especificidade também apresentou *outlier* bem afastado do primeiro quartil (figura 19). Trata-se de artigo [72] que validou um CAD treinado na caracterização de CaP na zona periférica em imagens de RM-MP, com excelentes resultados de área sob a curva ROC e sensibilidade (0,95 e 1, respectivamente), porém com insatisfatória especificidade, possivelmente atribuído ao uso de escala de Likert com limiar inapropriado.

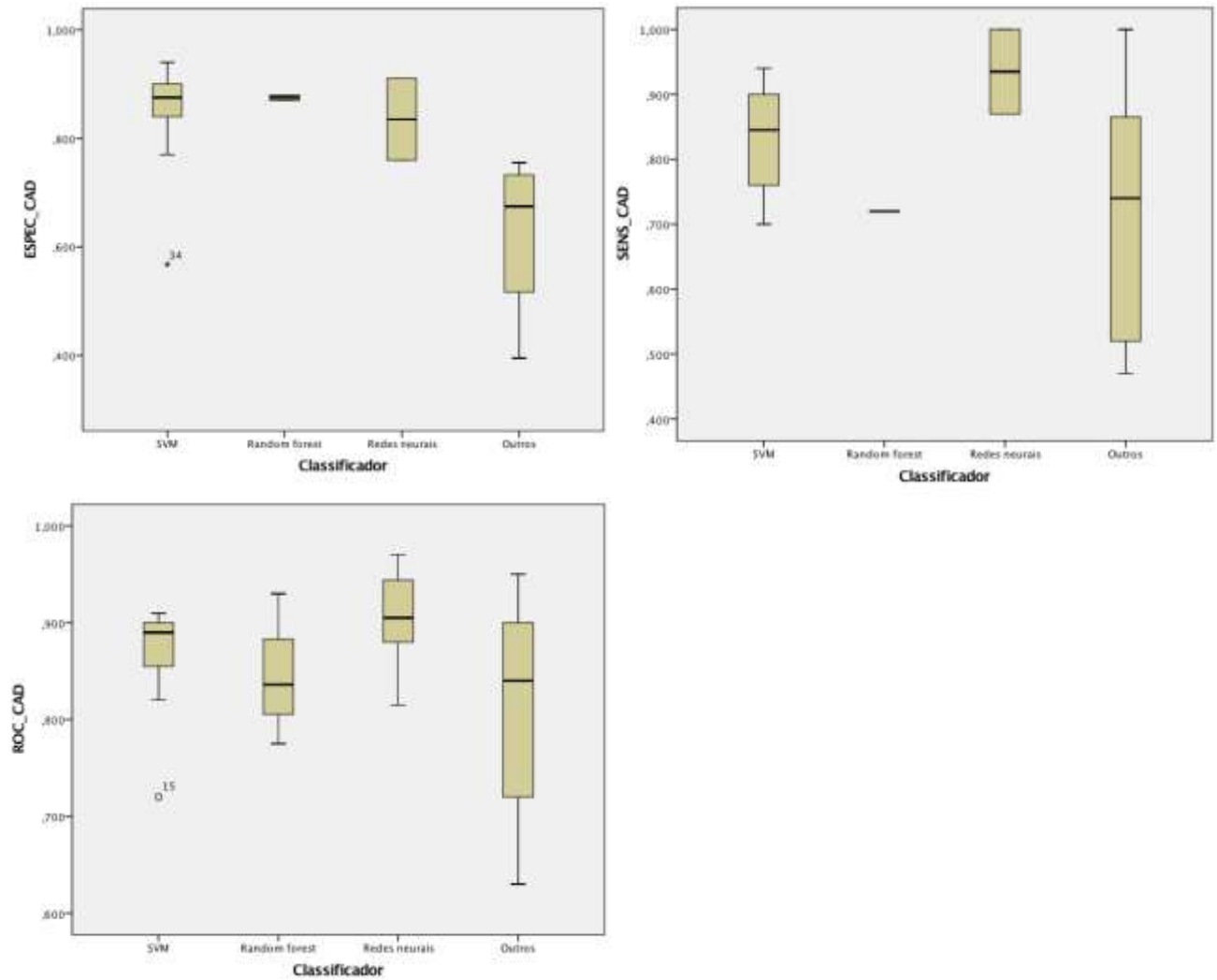
Figura 16 – Box-plot e Q-Q plot das métricas sensibilidade (SENS_CAD) e especificidade (ESPEC_CAD).



Fonte: o autor, 2021

Conforme a tabela 21, do apêndice A, ao categorizarmos os resultados das métricas pelos tipos de classificadores, podemos observar as diferenças de *performance* entre eles, embora a escassez de dados encontrados não permita maiores inferências neste sentido. Como esperado, o detalhamento em “tipo do classificador” melhorou a homogeneidade dos dados (SVM, redes neurais e *random forest* obtiveram coeficiente de variação menor que 15%), ao separá-los daqueles dados mais heterogêneos, pertencentes a categoria *outros*. Por outro lado, reduz ainda mais o conjunto de dados, prejudicando as análises. Os gráficos box-plot, da figura 20, visualmente reforçam esta percepção. Também no processo de agrupar os resultados das métricas pelos tipos de classificadores, nota-se que houve incremento nas médias examinadas, sobretudo da sensibilidade e especificidade, e estreitamento dos desvios-padrão.

Figura 17 – Box-plot das métricas área sob a curva ROC (ROC_CAD), sensibilidade (SENS_CAD) e especificidade (ESPEC_CAD), categorizadas pelo tipo do classificador CAD.



Fonte: o autor, 2021

Na métrica área sob a curva ROC surgiu um *outlier* na caixa referente ao classificador SVM, abaixo do primeiro percentil, referente a um artigo [73] que empregou classificadores supervisionados (comparando SVM e Regressão Logística) para avaliar se a utilização de dicionários de aprendizado na fase de extração de características aumentaria a performance de algum classificador, o que não foi confirmado. O *outlier* afastado do primeiro percentil na variável especificidade, para o classificador SVM, é referente a uma publicação [74] que propõe um sistema CAD baseado em dois mecanismos (LDA e SVM, avaliados independentemente) para estimar o escore de Gleason em MR-MP pré-operatória. Possivelmente o valor da

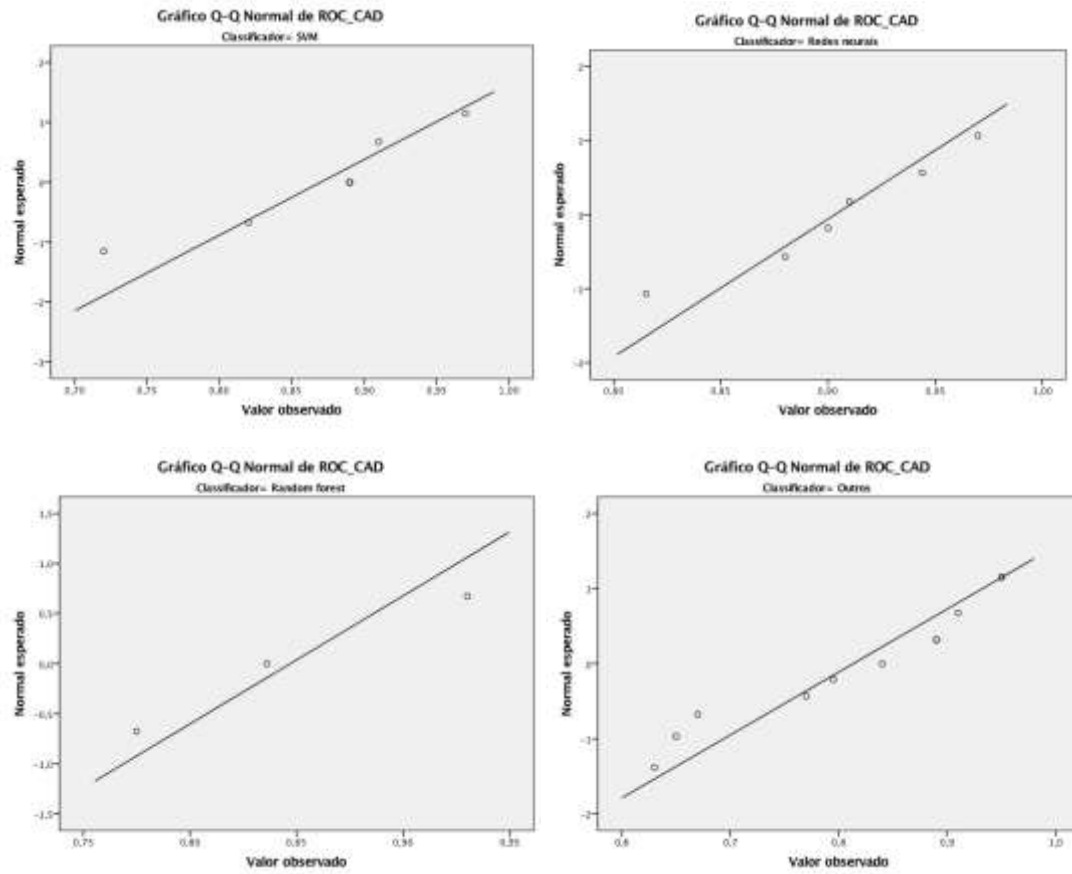
especificidade deste *outlier* é decorrente do efeito do limiar de categorização escolhido em uma escala Likert.

Tabela 11 – Resultados dos testes de normalidade, por tipo do classificador CAD.

Classificador CAD	Métrica	Número de artigos	Testes de normalidade	
			Kolmogorov-Smirnov	Shapiro-Wilk
SVM	Área sob a curva ROC	7	0,004	0,009
	Sensibilidade	10	0,200	0,585
	Especificidade	10	0,015	0,003
Redes Neurais	Área sob a curva ROC	6	0,200	0,882
	Sensibilidade	2	-	-
	Especificidade	2	-	-
<i>Random Forest</i>	Área sob a curva ROC	3	-	0,766
	Sensibilidade	2	-	-
	Especificidade	2	-	-
Outros	Área sob a curva ROC	11	0,200	0,160
	Sensibilidade	5	0,200	0,649
	Especificidade	4	-	0,298

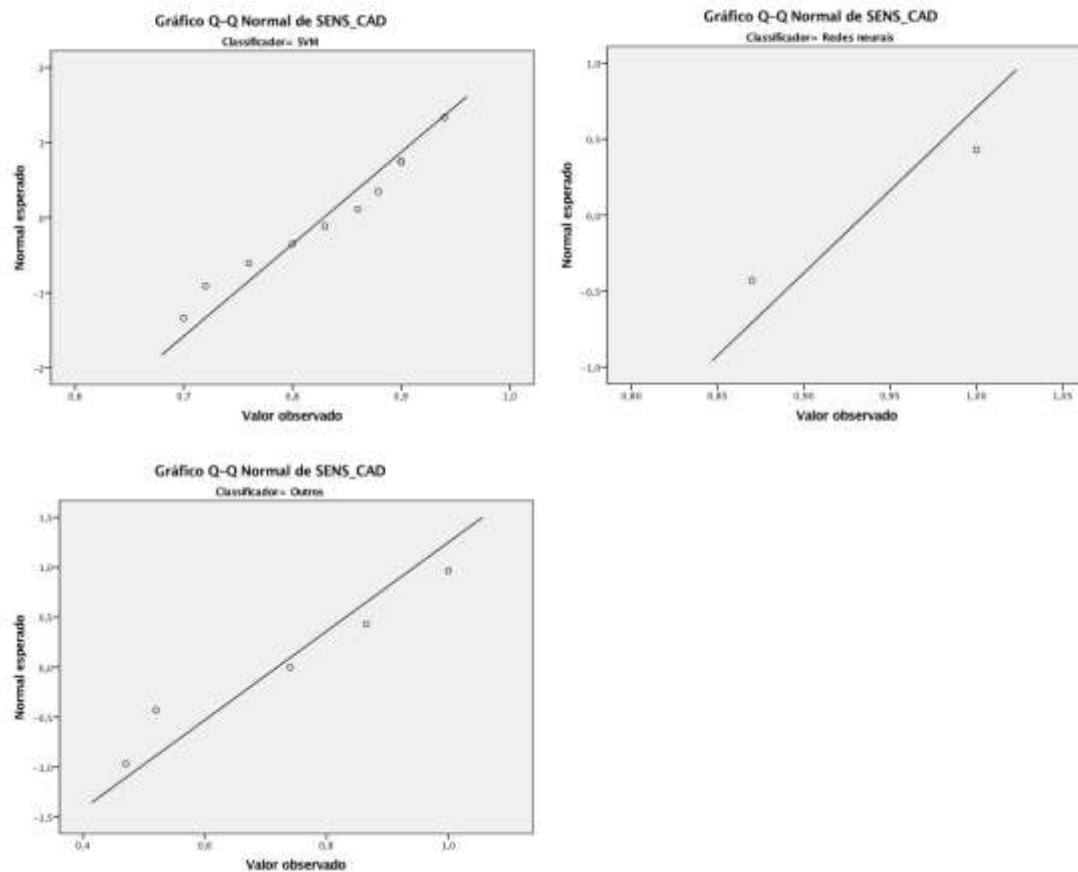
A pequena população categorizada inviabiliza a segurança dos testes de normalidade, disponíveis na tabela 22, acima: os classificadores “redes neurais” e “*random forest*” sequer foram passíveis de avaliação e os demais, mesmo utilizando o teste mais recomendado para estudos com pouco casos (Shapiro-Wilk), obtiveram questionável predomínio de distribuição não-gaussiana, sendo necessário recorrer aos gráficos Q-Q plot (figuras 21, 22 e 23), nos quais se observa pouca aderência aos valores da linha de tendência normal.

Figura 18 – Q-Q plot da métrica área sob a curva ROC (ROC_CAD), categorizado pelo tipo do classificador CAD.



Fonte: o autor, 2021

Figura 19 – Q-Q plot da métrica sensibilidade (SENS_CAD), categorizado pelo tipo do classificador CAD^a.

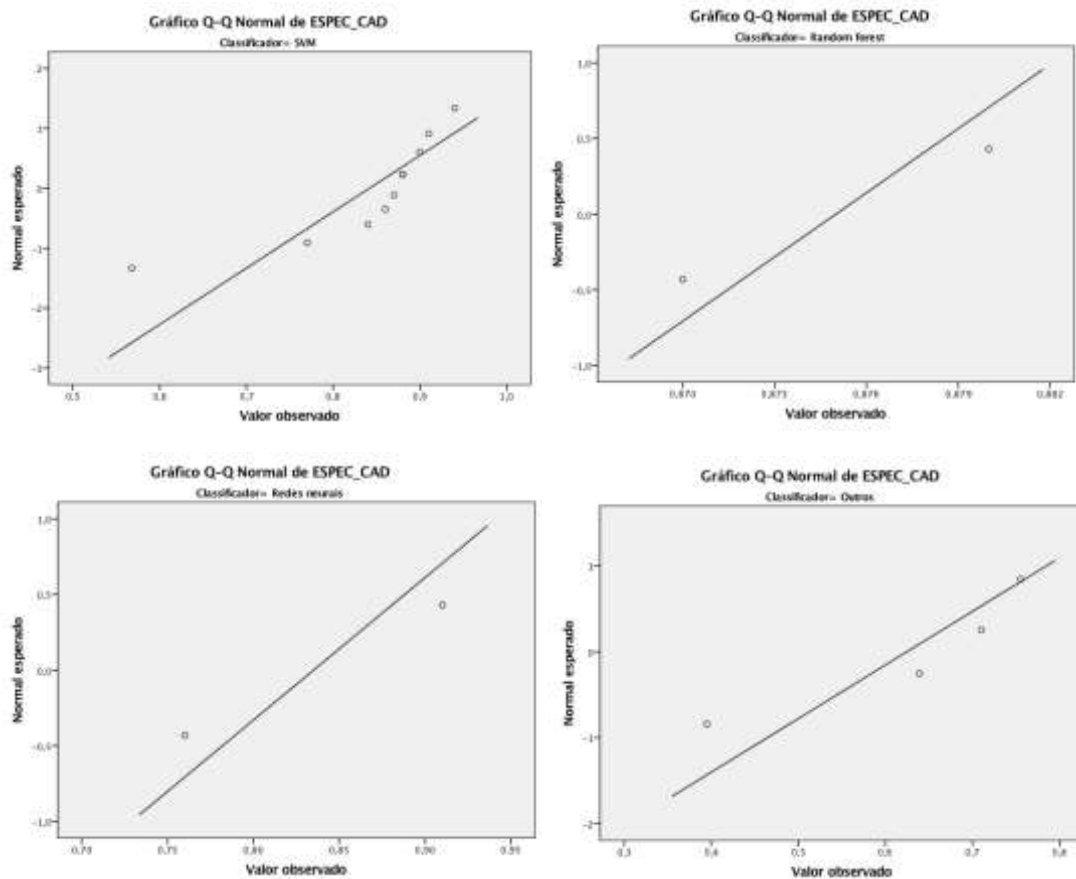


Fonte: o autor, 2021

^a Gráfico de normalidade da métrica sensibilidade (SENS_CAD) para o classificador *random forest* não foi passível de construção devido constância de valor nos dois únicos casos do conjunto de dados.

Logo, analisando as figuras de 21 a 23, assim como a sumarização dos testes de Kolmogorov-Smirnov e Shapiro-Wilk na tabela 22, diante de amostra pequena e ao mesmo tempo heterogênea quanto a normalidade (quer sejam alguns dados de distribuição gaussiana e outros não, quer sejam dados insuficientes para adequada testagem), se fez necessário aplicação de testes não paramétricos visando extrair a relação entre eles, evidente que esperando menor poder estatístico do experimento.

Figura 20 – Q-Q plot da métrica especificidade (ESPEC_CAD), categorizado pelo tipo do classificador CAD.



Fonte: o autor, 2021

4.4. PIRADS

Como anteriormente explicitado, foram escolhidos estudos de validação do protocolo PIRADS, disponíveis nos mesmos mecanismos de busca, como referência na análise do desempenho dos classificadores CAD. A tabela 23 do apêndice A sumariza os resultados obtidos de área sob a curva ROC, sensibilidade e especificidade dos artigos de validação do PIRADS, selecionados. Em uma metanálise conduzida em 2015 [75] na qual era avaliada a performance do PIRADS na detecção do CaP na RM-MP, globalmente foi observado uma sensibilidade de 0,78 (0,70 - 0,84 no intervalo de confiança de 95%) e especificidade de 0,79 (0,68 - 0,86). Outros estudos já foram realizados confrontando a normatização PIRADS versão 2 com estudo anatomopatológico, obtendo-se sensibilidade de 0,75 e especificidade de 0,90, cuja área sob a curva é estimada em 0,89 [76]; até sensibilidade de 0,88 e especificidade de 0,93 [77]. Em outras metanálises mais recentes, de 2017, utilizando-se 21 estudos [78] com o

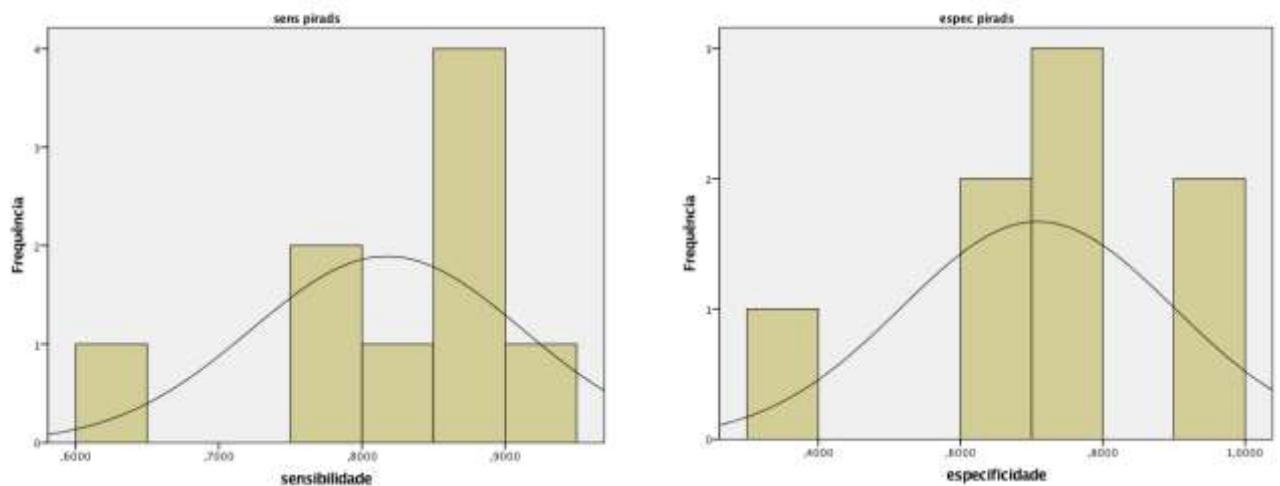
protocolo PIRADS na versão 2, a sensibilidade alcançada foi 0,89 (0,86 - 0,92) e a especificidade de 0,73 (0,60 - 0,83); e com 13 estudos [79] a sensibilidade foi 0,85 (0,78 - 0,91) e especificidade 0,71 (0,60 - 0,80).

Tabela 12 – Resultados gerais das métricas de desempenho dos estudos de validação do PIRADS.

Métrica	Mínimo	Máximo	Média	Desvio padrão	Mediana	Coefficiente de variação
Área sob a curva ROC	0,76	0,89	0,838	0,068	0,865	8,1%
Sensibilidade	0,61	0,92	0,817	0,095	0,85	11,6%
Especificidade	0,32	0,93	0,71	0,191	0,72	26,9%

A tabela 24, acima, pormenoriza tais informações e, na figura 24, é possível observar a distribuição das frequências das métricas de desempenho sensibilidade e especificidade (não foi construído histograma de área sob a curva ROC devido a somente três artigos selecionados).

Figura 21 – Histograma das métricas sensibilidade (SENS_PIRADS) e especificidade (ESPEC_PIRADS) referente aos estudos de validação do PIRADS.

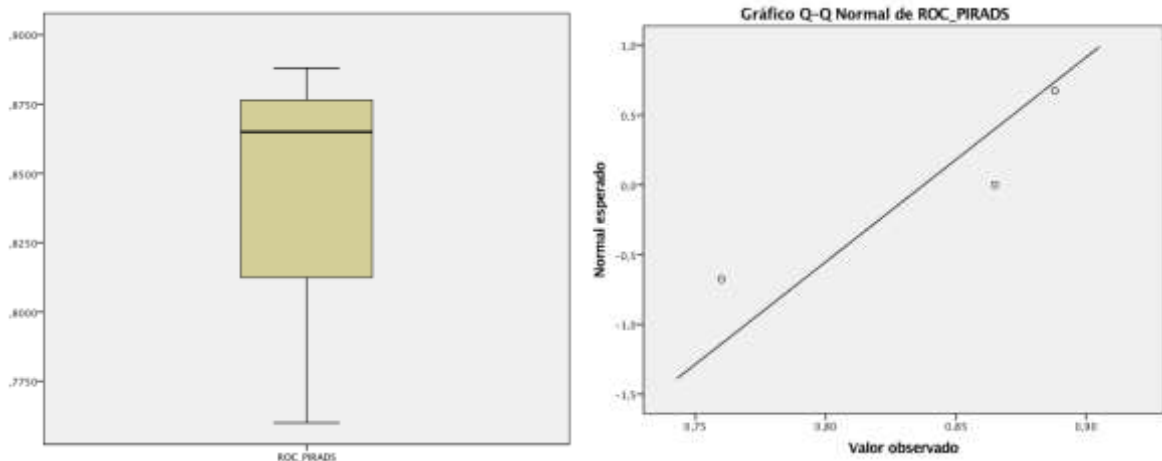


Fonte: o autor, 2021

Dos dados obtidos, o mais escasso é a métrica área sob a curva ROC, somente com 3 estudos, cuja média é 0,838 com desvio padrão de 0,068. Sensibilidade é aquela com maior número de artigos (9 de um total de 10), com média de 0,817 e desvio padrão de 0,095. Especificidade, aquela com 8 artigos disponíveis, apresentou média de 0,71 e desvio padrão de 0,191.

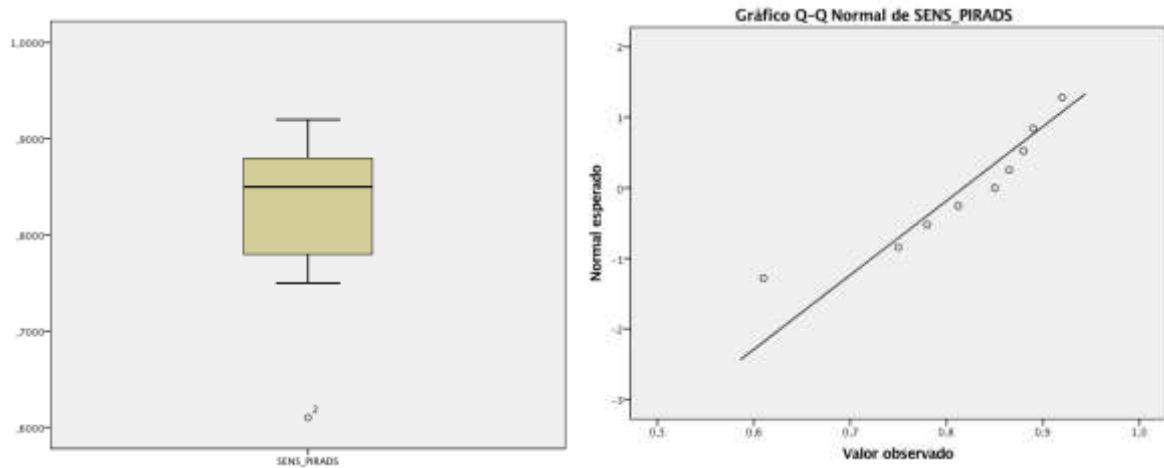
É possível observar que, para efeito de decisão clínica, tanto a média obtida pela área sob a curva ROC quanto a da sensibilidade do PIRADS são interessantes no julgamento do teste diante de um paciente suspeito para a patologia (CaP), aliado a uma dispersão é menor destas variáveis, a primeira à despeito da pequena quantidade de dados. Generalizando, pode-se dizer que os dados do PIRADS são mais homogêneos, menos dispersos, do que aqueles coletados no CAD. Nas figuras de 25 a 27, referente aos *box-plots* e Q-Q plots das métricas área sob a curva ROC, sensibilidade e especificidade, respectivamente, a análise visual reforça tal percepção, embora prejudicada no caso de área sob a curva ROC, pela já citada dificuldade amostral.

Figura 22 – Box-plot e Q-Q plot da métrica área sob a curva ROC dos estudos de validação do PIRADS (ROC_PIRADS).



Fonte: o autor, 2021

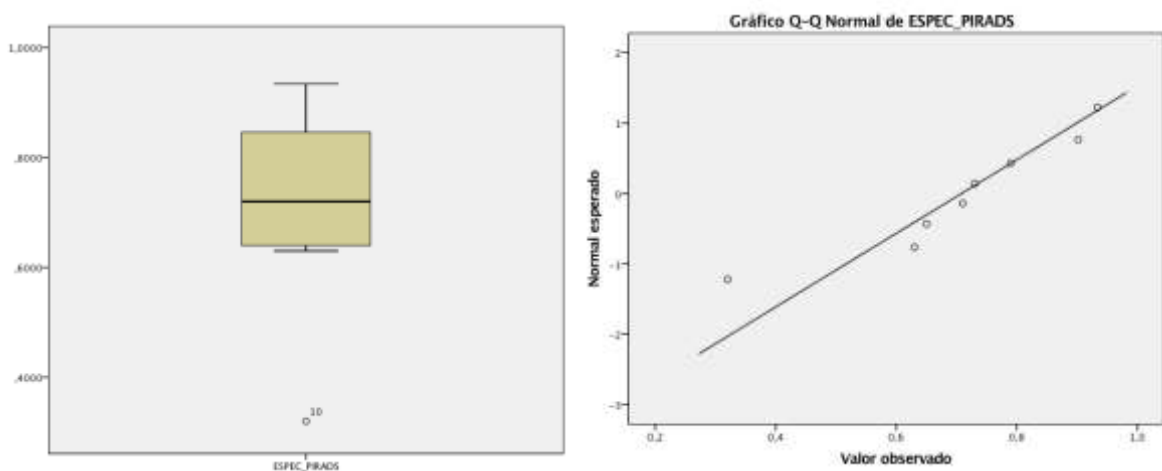
Figura 23 – Box-plot e Q-Q plot da métrica sensibilidade dos estudos de validação do PIRADS (SENS_PIRADS).



Fonte: o autor, 2021

O *outlier* detectado no box-plot da figura 26 corresponde a um estudo [80] que avaliou a concordância entre diversos radiologistas na aplicação do PIRADS quando da implementação da versão 2. Embora os autores tenham observado sensibilidade de até 0,9 na detecção de lesões índices (definidas por eles como aquelas com maior escore de Gleason na anatomia patológica ou maior volume tumoral, caso escore semelhante), a performance geral, para todos os tipos de lesões e independente da experiência do avaliador, foi baixa.

Figura 24 – Box-plot e Q-Q plot da métrica especificidade dos estudos de validação do PIRADS (ESPEC_PIRADS).



Fonte: o autor, 2021

Já o *outlier* identificado no box-plot da figura 27 corresponde a artigo [81] que avaliou a acurácia do PIRADS versão 2 em detectar CaP clinicamente significativo quando o resultado categorizado for igual ou maior que 3 (visando, em última análise, estabelecer na escala de Likert do PIRADS o limiar para assegurar a vigilância ativa ou indicar a biópsia). Dois radiologistas avaliaram o banco imagens de MR-MP de próstata, composto por pacientes com escore de Gleason maior ou igual a 7. Apesar de uma sensibilidade satisfatória (0,81), a especificidade foi muito baixa (0,32).

Mais uma vez, não se observa êxito dos escassos dados ao enfrentarem o teste de normalidade de Kolmogorov-Smirnov, como se observa na tabela 25, tendo que se recorrer ao teste de Shapiro-Wilk e a análise visual dos gráficos Q-Q plot acima ilustrados: a distribuição não gaussiana é também assumida para as variáveis dos estudos de validação do PIRADS.

Tabela 13 – Teste de normalidade das métricas **área sob a curva ROC, sensibilidade e especificidade**, dos estudos de validação do PIRADS.

Métrica	Testes de Normalidade	
	Kolmogorov-Smirnov	Shapiro-Wilk
Área sob a curva ROC	-	0,323
Sensibilidade	0,200	0,185
Especificidade	0,200	0,370

4.5. Resultado do experimento com radiologistas

Do estudo realizado com dois radiologistas com experiência semelhante em RM-MP de próstata, que avaliaram um conjunto de 26 exames, com um total de 46 lesões passíveis de classificação, os dados obtidos foram analisados na tabela de contingência e sintetizados na tabela 26.

As seguintes probabilidades diagnósticas são possíveis:

- o radiologista classificar como CaP uma lesão que realmente é CaP (verdadeiro-positivo);

- o radiologista classificar como CaP uma lesão que não é CaP (neste caso, a lesão verdadeira estaria em outro sítio, portanto um falso-positivo);

- o radiologista deixar de classificar como CaP uma lesão que é CaP (um falso-negativo).

Há que se considerar que o banco de dados utilizado contém o viés de ser composto somente por exames com CaP, o que inviabiliza avaliação da *performance* dos observadores quanto a classificação do exame em normal quando este realmente é normal (verdadeiro-negativo).

Observa-se que o Radiologista A identificou 25 lesões que eram realmente CaP quando confrontadas com a anatomia patológica, enquanto o Radiologista B localizou 31 lesões adequadamente classificadas como câncer. Calculada a sensibilidade, foi obtido o valor de 0,71 para o radiologista A e 0,76 para o B.

Tabela 14 – Experimento com radiologistas: detecção do CaP em imagens de RM-MP com validação pelo padrão-ouro (anatomia patológica).

Padrão-ouro (anatomia patológica)	Detecção CaP - Radiologista A		Total
	Positivo	Negativo	
Presente	25 (54,3%)	10 (21,7%)	35 (76,1%)
Ausente	11 (23,9%)	0 (0%)	11 (23,9%)
Total	36 (78,3%)	10 (21,7%)	46 (100%)
Padrão-ouro (anatomia patológica)	Detecção CaP - Radiologista B		Total
	Positivo	Negativo	
Presente	31 (67,4%)	10 (21,7%)	41 (89,1%)
Ausente	5 (10,9%)	0 (0%)	5 (10,9%)
Total	36 (78,3%)	10 (21,7%)	46 100%)

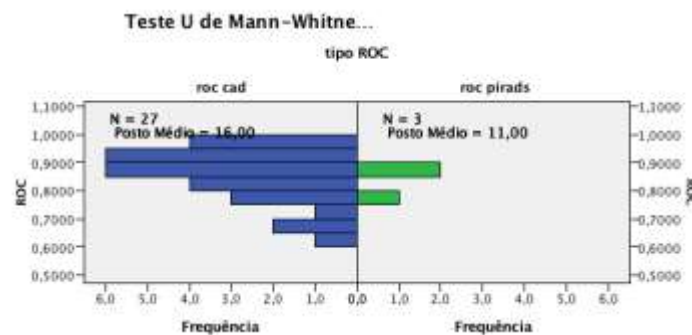
4.6 Análise comparativa entre o cad, pirads e experimento com radiologistas

Inicialmente, para efeito de comparação das métricas área sob a curva ROC, sensibilidade e especificidade entre os grupos CAD e PIRADS, foi utilizado o teste de Mann-Whitney, sendo as hipóteses nula (H_0) e alternativa (H_1) para cada métrica definidas como:

- H_0 : a mediana da métrica do CAD é igual à mediana da métrica do PIRADS
- H_1 : a mediana da métrica do CAD é diferente da mediana da métrica do PIRADS

Todos as métricas foram submetidas ao teste em sua forma bicaudal (não importando, neste momento da análise, qual a diferença entre os grupos, mas se há diferença entre eles). Os seguintes resultados foram obtidos (figuras de 28 a 30):

Figura 25– Resultado do teste de hipótese da métrica área sob a curva ROC.



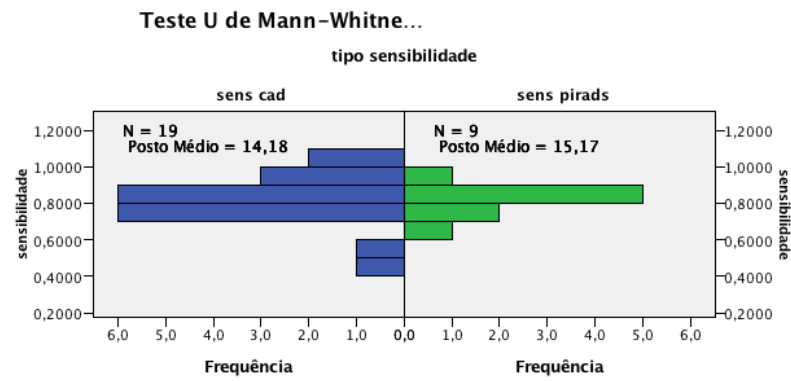
CAD X PIRADS

Área sob a curva ROC

N CAD	27
N PIRADS	3
N total	30
U de Mann-Whitney	27,0
Sig. exata (teste bilateral)	0,387

Fonte: o autor, 2021

Figura 26 – Resultado do teste de hipótese da métrica sensibilidade.



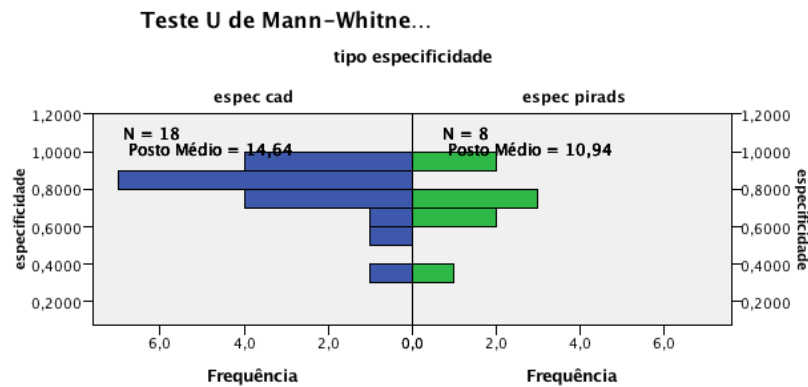
CAD X PIRADS

Sensibilidade

N CAD	19
N PIRADS	9
N total	28
U de Mann-Whitney	91,5
Sig. exata (teste bilateral)	0,772

Fonte: o autor, 2021

Figura 27 – Resultado do teste de hipótese da métrica especificidade.



CAD X PIRADS

Especificidade

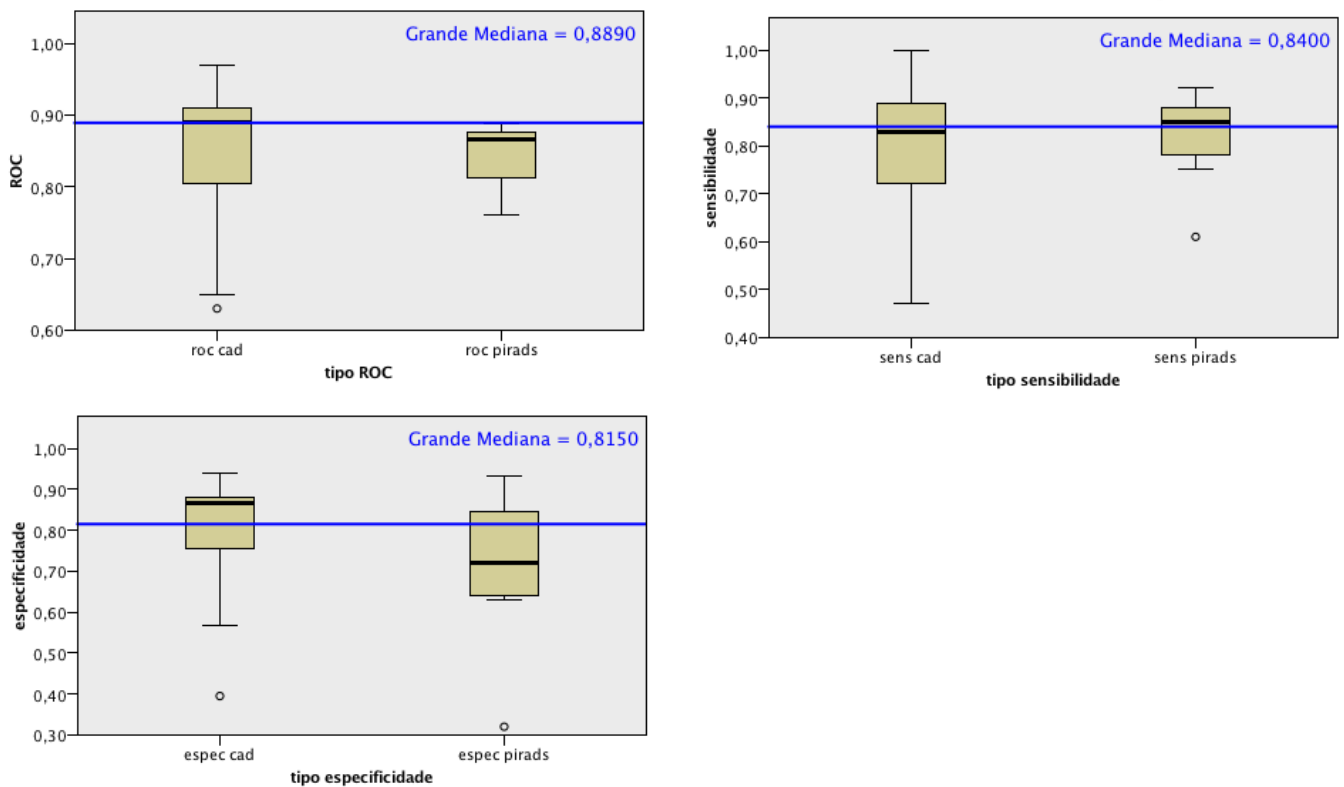
N CAD	18
N PIRADS	8
N total	26
U de Mann-Whitney	51,5
Sig. exata (teste bilateral)	0,261

Fonte: o autor, 2021

Neste teste, as três métricas de desempenho obtiveram resultados maiores que o nível de significância (0,05), sendo área sob a curva ROC = 0,387; sensibilidade = 0,772 e especificidade = 0,261, portanto não rejeitando a hipótese nula. Deve-se considerar que há igualdade entre os grupos, logo equivalência entre o CAD e o PIRADS no diagnóstico do CaP. Porém, segundo SOARES E SIQUEIRA [53], deve-se ter cuidado com esta afirmativa, pois “não se pode dizer que a hipótese nula é verdadeira, mas somente que não há evidência suficiente para que ela seja rejeitada”. Inclusive o erro tipo 2 não pode ser descartado, especialmente quando se depara com casuística pequena, como a deste estudo. Interessante notar que, à análise dos box-plot das respectivas variáveis, a percepção visual acerca da verdade da H_0 é notória, ao vislumbrar equiparação entre as caixas, sobretudo na métrica sensibilidade, na qual até a mediana de um grupo é próxima a do outro (figura 31). No geral, o maior tamanho das caixas e as caudas da distribuição mais longas no grupo CAD em comparação com o

PIRADS, nas três métricas de desempenho, sugerem que o padrão de referência (PIRADS) é mais constante. Tal aspecto pode ser decorrente da variabilidade de classificadores, demais dados e desenhos de estudo dos artigos CAD selecionados, porém possivelmente pode implicar em melhorias no CAD para reprodutibilidade entre suas aplicações.

Figura 28 – Box-plot do teste de medianas de amostras independentes das métricas área sob a curva ROC, sensibilidade e especificidade, por grupos CAD e PIRADS.



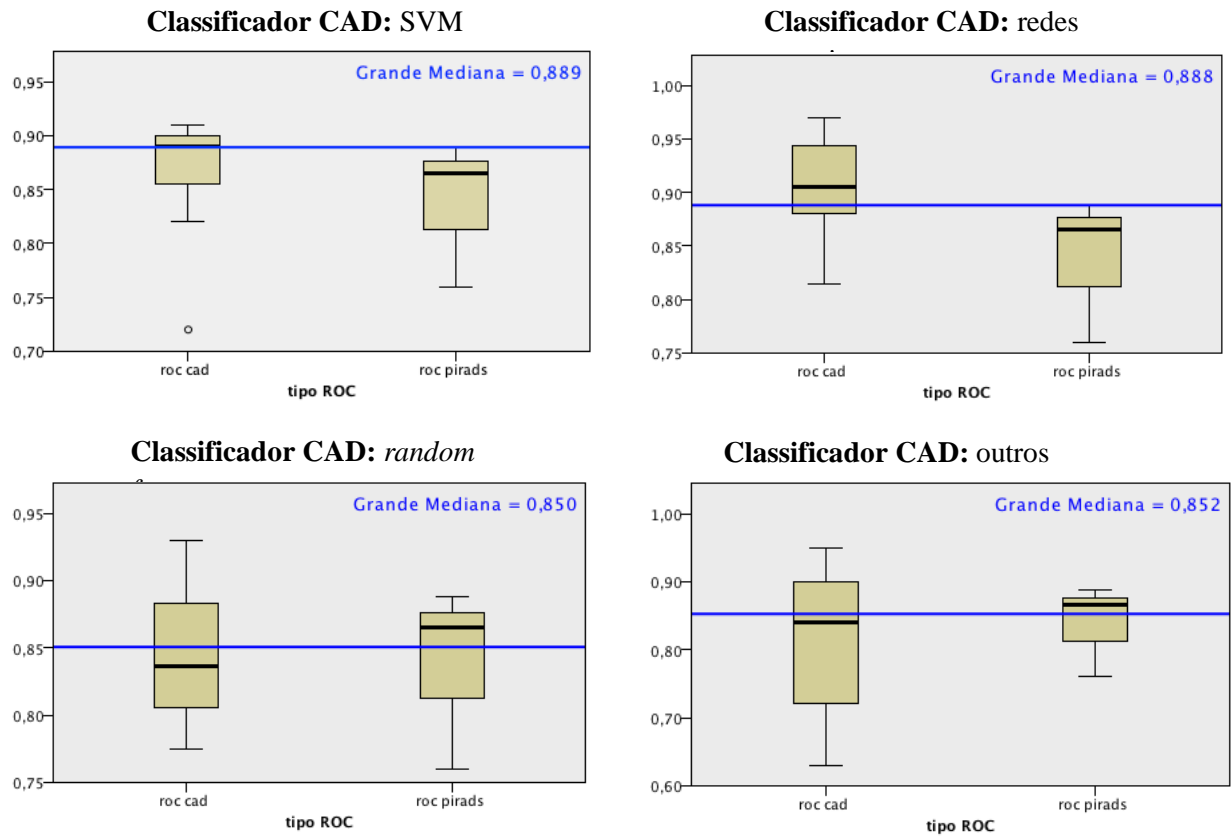
Fonte: o autor, 2021

A seguir, foi aplicado novamente o teste de Mann-Whitney comparando os grupos CAD e PIRADS, para cada métrica, agora categorizando quanto ao tipo de classificador, conforme a tabela 28, abaixo. Como se pode depreender, o mesmo padrão de comportamento foi observado, sempre aceitando a hipótese nula. Diante de tal premissa, erro do tipo 2 novamente não pode ser descartado, sobretudo que, com a categorização, a casuística se tornou ainda mais exígua.

Tabela 15 – Resultado do teste de hipótese das métricas área sob a curva ROC, sensibilidade e especificidade entre os grupos CAD e PIRADS, segundo tipo de classificador.

Teste de Mann-Whitney de amostras independentes			
Tipo do classificador CAD	Métrica	Sig.	Decisão
SVM	Área sob a curva ROC	0,267	Não há evidência suficiente para que a H_0 seja rejeitada
	Sensibilidade	0,905	
	Especificidade	0,146	
Redes Neurais	Área sob a curva ROC	0,167	
	Sensibilidade	0,145	
	Especificidade	0,400	
<i>Random Forest</i>	Área sob a curva ROC	1,000	
	Sensibilidade	0,218	
	Especificidade	0,400	
Outros	Área sob a curva ROC	0,885	
	Sensibilidade	0,364	
	Especificidade	0,461	

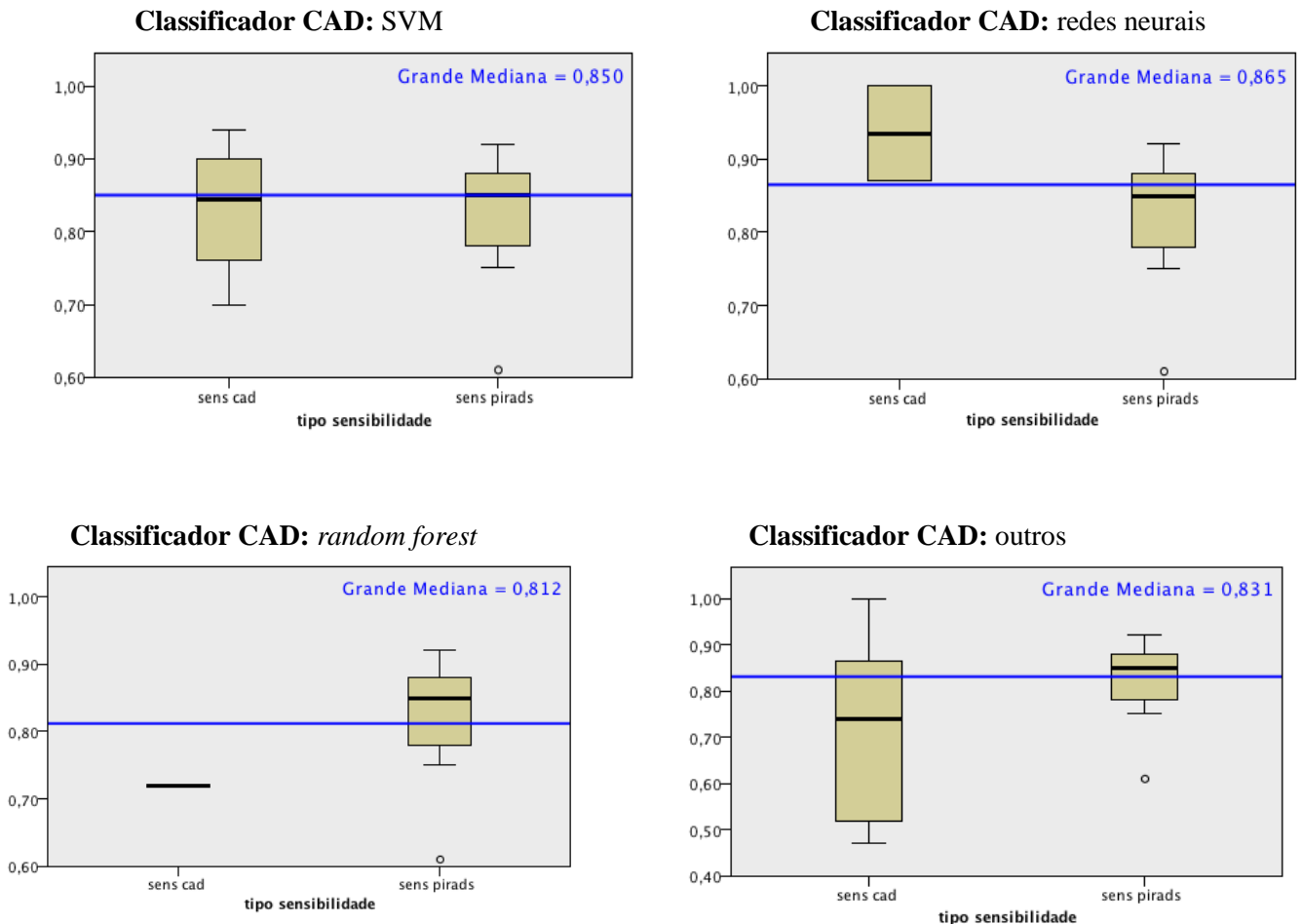
Figura 29 – Box-plot do teste de medianas de amostras independentes da métrica área sob a curva ROC, por grupos CAD e PIRADS, segundo o tipo do classificador CAD.



Fonte: o autor, 2021

Recordando a tabela 18, onde se sumariza a distribuição dos artigos em tipos de classificador, e visualizando a figura 32, o entendimento de que escassez de dados pode comprometer a análise é reforçado, ao demonstrar, por exemplo, diversas sobreposições entre as métricas de desempenho. No caso da variável área sob a curva ROC há, no classificador SVM, um *outlier*, que corresponde ao mesmo artigo [71] identificado na análise não categorizada. Ainda que se exclua uma categoria com distribuição de dados não homogênea, as demais não têm dados suficientes para garantir a confiabilidade da análise. À parte do classificador SVM, a categoria com mais artigos (o classificador redes neurais, que detém seis estudos), é aquela que apresenta o gráfico box-plot com menor sobreposição (possivelmente a categoria na qual se possa observar alguma diferença entre o CAD e o PIRADS, em um possível aumento amostral).

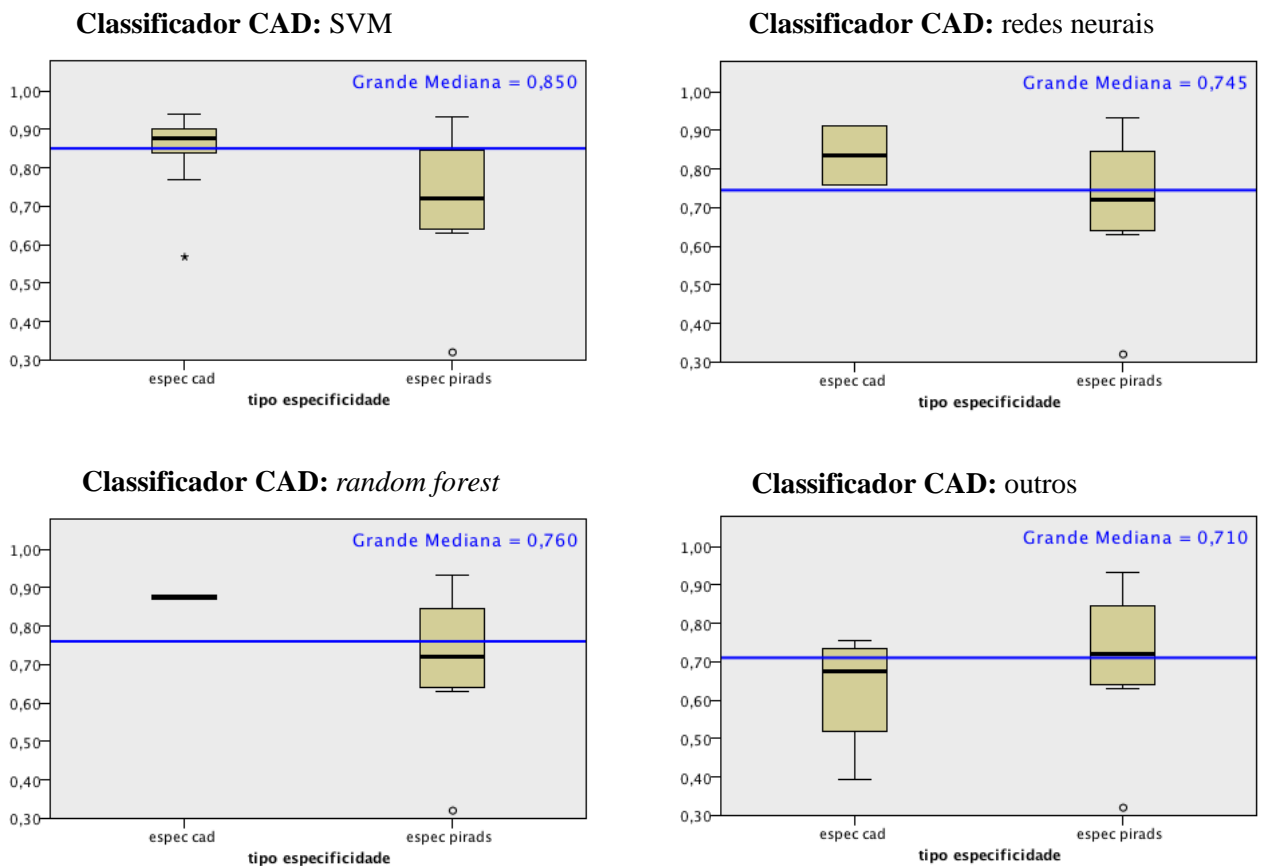
Figura 30 – Box-plot do teste de medianas de amostras independentes da métrica sensibilidade, por grupos CAD e PIRADS, segundo o tipo do classificador CAD.



Fonte: o autor, 2021

De forma semelhante, a métrica sensibilidade (figura 33) também apresentou menor sobreposição das caixas box-plots no classificador redes neurais (embora neste somente dois artigos tenham sido selecionados). O classificador SVM foi o que demonstrou maior equivalência na análise visual entre os grupos CAD e PIRADS, inclusive com a mediana entre os grupos sobrepondo / quase que sobrepondo a mediana de cada grupo. No caso do classificador *random forest*, a escassez de artigos aliada a ocorrência do mesmo valor estatístico encontrado neles impedem inferência gráfica adicional. O *outlier* observado no grupo PIRADS corresponde a aquele artigo [80], anteriormente citado, que verificou a concordância entre radiologistas na aplicação do PIRADS.

Figura 31 – Box-plot do teste de medianas de amostras independentes da métrica especificidade, por grupos CAD e PIRADS, segundo o tipo do classificador CAD.



Fonte: o autor, 2021

No caso da métrica especificidade, nos box-plots da figura 34 é observada coincidência entre as diversas caixas, exceto pelo classificador *random forest*, no qual a pequena quantidade de dados (dois casos, somente) e a equivalência de valores prejudicaram melhor apreciação gráfica. De todo modo, a menor dispersão encontrada foi no classificador SVM, o de maior quantidade de artigos, porém com o *outlier* [73] previamente descrito.

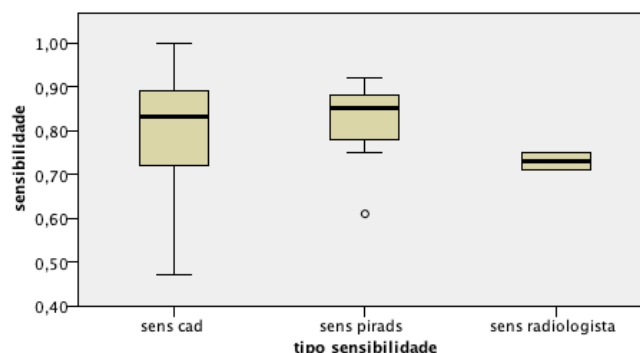
Comparando-se os três grupos, CAD, PIRADS e EXPERIMENTO COM RADIOLOGISTAS, através do teste de hipótese de Kruskal-Wallis de amostras independentes, para a métrica sensibilidade não se observou diferença significativa, conforme a tabela 28. O teste utilizou equação para empates e não são realizadas comparações múltiplas pois o teste inteiro não apresentou diferenças significativas entre os grupos.

Tabela 16 – Resultado do teste de hipótese de Kruskal-Wallis da métrica sensibilidade entre os grupos CAD, PIRADS e EXPERIMENTO COM RADIOLOGISTAS.

CAD X PIRADS X RADIOLOGISTAS	
Sensibilidade	
N CAD	19
N PIRADS	9
N RADIOLOGISTAS	2
N total	30
Estatística do teste de Kruskal-Wallis	1,783
Gaus de liberdade	2
Sig assintótica (teste bilateral)	0,410

No box-plot da figura 35, referente a distribuição das métricas CAD, PIRADS e EXPERIMENTO COM RADIOLOGISTAS no teste de Kruskal-Wallis, nota-se a sobreposição entre os grupos CAD e PIRADS e entre CAD e EXPERIMENTO COM RADIOLOGISTAS, porém não evidente entre os grupos PIRADS e EXPERIMENTO COM RADIOLOGISTAS. Para efeito de esclarecimento, foi realizado o teste de Mann-Whitney de amostras independentes entre os grupos PIRADS e EXPERIMENTO COM RADIOLOGISTAS, não havendo significância estatística entre eles (0,145), à despeito do aspecto visual dos respectivos box-plots.

Figura 32 – Box-plot do teste de hipótese de Kruskal-Wallis de amostras independentes, para a métrica sensibilidade, dos grupos CAD, PIRADS e EXPERIMENTO COM RADIOLOGISTAS.

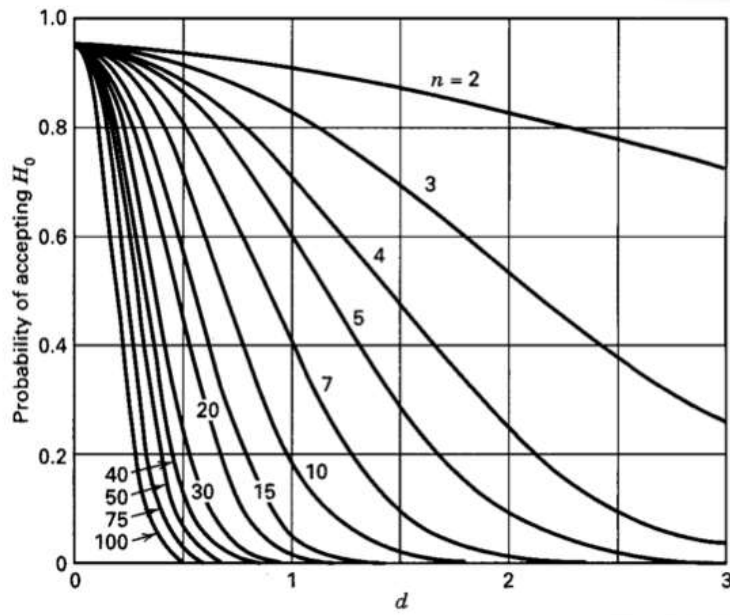


Fonte: o autor, 2021

Do exposto e, considerando as análises preliminares quando do desenho deste estudo, que indicavam alguma superioridade do CAD sobre o PIRADS na detecção do CaP, algumas apreciações devem ser ponderadas, referente a possibilidade de a análise ter cometido erro do tipo 2. Para mitigar tal probabilidade, interessante seria conhecer o poder estatístico de cada teste, no arranjo de combinações comparativas proposto. Esta avaliação permitiria estabelecer maior segurança para aceitar a hipótese nula dos testes. Simulação utilizando o método de Monte Carlo em *software* específico é preconizada neste intuito, embora não seja o propósito deste estudo. Segundo KOMATSU [82], no método de Monte Carlo, a pequena magnitude da diferença entre dois grupos implica em maiores amostras para se conseguir níveis de confiabilidade aceitáveis. Portanto, ainda segundo este autor, não existiria um tamanho amostral padrão que assegure a credibilidade de um teste. Matematicamente, é necessária a estimativa da variância do fator em estudo em uma população e a grandeza do efeito esperado da variável em teste a fim de calcular a amostra requerida. É possível que, ao mensurar o tamanho da amostra necessária neste estudo, no futuro, com novas publicações disponíveis, se verifique novamente as hipóteses escolhidas, com menor erro amostral, visando garantir o resultado estatístico.

Outra maneira de se avaliar o acolhimento da H_0 é utilizar a curva de característica de operação, que retrata o poder discriminatório de um plano de amostragem de aceitação [83]. Admitindo que o poder do teste é $1 - \alpha$ a probabilidade de aceitar H_0 dado que H_1 é verdadeiro, na curva de característica de operação temos um gráfico do erro tipo 2 em relação ao delta referente a diferença entre as médias (ou medianas) das hipóteses testadas. Na figura 36, é representado um gráfico de curva de característica de operação, onde se pode observar que, para um dado valor de delta, o erro tipo 2 decresce conforme o tamanho amostral aumenta; válido para amostras de mesmo tamanho e variância. Portanto, ou a diferença entre os grupos tem que ser grande (um delta largo) ou a amostra tem que ser volumosa para que o poder do teste seja suficiente para não creditarmos uma aceitação de H_0 ao erro tipo 2.

Figura 33 – Exemplo de gráfico de curva de característica de operação.



Fonte: Ferris et al [86]

5. CONCLUSÕES E POSSIBILIDADES DE DESENVOLVIMENTO

Retornando aos objetivos do estudo, os resultados da análise isolada do CAD quando confrontado com o estudo anátomo-patológico (tabela 19) são muito bons do ponto de vista clínico na detecção do CaP, já que as três métricas de desempenho obtiveram valores aplicáveis na prática médica (área sob a curva ROC: 0,849; sensibilidade: 0,799; e especificidade 0,796), principalmente quando comparado com a propeidêutica (toque retal tem sensibilidade de 0,59) e outros métodos diagnósticos como o USTR (sensibilidade de 0,48 e especificidade de até 0,81) e o PSA (entre 4,0 ng/mL e 10,0 ng/mL, tem sensibilidade de 0,79 e especificidade de 0,90). Logo, é factível seu uso como ferramenta adicional no rastreamento médico do CaP, não isoladamente, mas no contexto de técnicas e procedimentos clínicos e laboratoriais. Há, claro, que se realizar outras pesquisas para adequada avaliação das indicações precisas de sua implementação nos já existentes algoritmos e protocolos urológicos de rastreamento e investigação do CaP.

Quanto aos benefícios do CAD, embora não completamente mensuráveis neste estudo, diversas publicações presentes nesta pesquisa robusteceram a percepção que há menor esforço do radiologista na manipulação do grande volume de imagens que compõem um exame de RM-MP, logo implicando em menos tempo de análise e brevidade no resultado do exame, além de melhora no desempenho de radiologistas menos experientes no diagnóstico. Os resultados das variáveis área sob a curva ROC, sensibilidade e especificidade favorecem maior segurança no diagnóstico quanto a redução nas taxas de falso-negativo, ao se aproximarem do observado na prática clínica, embora ainda sejam necessária menor variabilidade entre os estudos, visando constância semelhante ao PIRADS.

A mensuração econômica da implementação do CAD não foi passível de avaliação devido aos estudos selecionados não contemplarem esta variável. Ademais, o uso do CAD comercialmente voltado para aplicação urológica (especificamente no diagnóstico do CaP) ainda é incipiente, ao contrário do CAD para mama e pulmão, nos quais se encontram disponíveis para aquisição, e há como estimar o custo de sua aplicação e o retorno quanto menor tempo dispendido pelo médico radiologista. Serão necessários, ainda, estudos que avaliem o impacto do CAD quanto aos efeitos terminais do CAP, por exemplo, as taxas de hospitalização, morbidade e mortalidade.

Como não houve diferença entre os grupos, não é possível afirmar em qual método estatístico seria observado ganho e, tampouco, qual a técnica de classificação CAD seria a superior, como era um dos objetivos deste estudo.

Confirmada a igualdade entre o CAD e o PIRADS na detecção do CaP, a H_1 deste estudo fica comprovada (é factível o diagnóstico de câncer de próstata através do CAD em RM-MP). Logo, não havendo diferença entre o uso do PIRADS ou do CAD, este poderia ser aplicado na prática médica como ferramenta de serviço, reduzindo a carga de trabalho e o esforço do radiologista na análise das diversas ponderações e imagens do exame de RM-MP, ao passo que o decréscimo da faina permitirá ganho de tempo para, inclusive, análise de mais exames. O resultado desfavorável que este estudo poderia demonstrar, mas que os testes não evidenciaram, seria a hipótese de diferença da análise estatística indicar a inferioridade do CAD frente ao PIRADS, situação na qual seria proscrito o seu uso na rotina médica, pois significaria prática prejudicial ao paciente.

O conhecimento do poder dos testes estatísticos empregados, visando estabelecer não somente maior segurança no consentimento da hipótese nula, mas também a modelagem amostral necessária, careceria de avaliações matemáticas específicas fora do escopo deste estudo, porém, como já foi explicitado, seria interessante para uma futura revisita ao tema, quando da publicação de novos artigos na área. Há perspectiva do surgimento de grande quantidade de estudos nos próximos anos, sobretudo com a iniciativa PROSTATEx, do Centro de Tratamento e Diagnóstico do Câncer, do Instituto Nacional do Câncer, americano, que promove competições entre diversas instituições (acadêmicas ou não), na qual os participantes comparam seus algoritmos computacionais com os de outros, como também da indústria e do governo, de forma estruturada e padronizada, ao utilizar os mesmos bancos de dados. Com esta possibilidade concreta, reforça a candidatura deste estudo ser reconduzido no futuro próximo, quando o aumento de sua amostra, e, portanto, redução do viés amostral, permitirá assegurar com maior confiança a análise apresentada.

No contexto de recursos escassos, como é o caso da saúde pública em diversos países, a implementação CAD na detecção do CaP, ao direcionar a apreciação dos *voxels* de interesse e extração de dados relevantes, otimiza-se o processo de análise, reduzindo seu tempo e, portanto, favorecendo a disponibilidade do exame para uma maior população. Ressalta-se que a RM é um exame patrocinado pelo sistema brasileiro de saúde (SUS – Sistema Único de saúde), assim, sua utilização com o CAD pode permitir que mais indivíduos sejam submetidos ao estudo, logo mais CaP serão descobertos e mais tratamentos precoces poderão ser iniciados,

beneficiando o resultado final dos pacientes. Abre-se, ainda, um novo mercado de *software* voltados nas múltiplas aplicações CAD em CaP.

Ademais, a constante implementação de novas técnicas CAD, não somente na clínica urológica, mas também nas diversas áreas da medicina, vem mostrando promissores resultados, inclusive ao detectar características radiômicas não perceptível ao ser humano ou no emprego de inteligência artificial em algoritmos de tomada de decisão. Por outro lado, nos atuais tempos de medicina baseada em evidência, há possibilidade de a atuação do médico ser cerceada em protocolos que envolvam o CAD, no qual o receio do erro induza a aceitação absoluta do resultado computacional. E, do ponto de vista jurídico, novas consequências surgem referentes a responsabilidade do médico ao aceitar um diagnóstico errado de uma análise de máquina, ou ao não concordar com uma terapia recomendada por computador. E, como já ocorre em vários setores da economia, os espaços ocupados pelas forças de trabalho humana são progressivamente substituídos pelas redes de inteligência artificial. Necessário, portanto, avanços nas pesquisas referente ao CAD, não somente relacionadas ao desenvolvimento e aperfeiçoamento da técnica em si, mas igualmente nas suas diversas implicações socioeconômicas.

REFERÊNCIAS

- AKIN, O. et al. Transition zone prostate cancers: features, detection, localization, and staging at endorectal MR imaging. **Radiology**, v. 239, p. 784-792, 2006.
- ALONZI, R.; PADHANI, A.; ALLEN, C. Dynamic contrast enhanced MRI in prostate cancer. **Eur J Radiol**, v. 63, p. 335–350, 2007.
- AMERICAN COLLEGE OF RADIOLOGY, ACR. Prostate Imaging Reporting and Data System version 2.0, 2015. Disponível em: <<https://www.acr.org/-/media/ACR/Files/RADS/Pi-RADS/PIRADS-V2.pdf>>. Acesso em: 1 nov. 2019.
- BARENTSZ, J. et al. European Society of Urogenital Radiology. ESUR prostate MR guidelines 2012. **Eur Radiol.**, v. 22, p. 746-757, 2012.
- BARROS, A. Tabagismo no Brasil: desigualdades regionais e prevalência segundo características ocupacionais. **Ciência & Saúde Coletiva**, v. 16, n. 9, p. 3707-3716, 2011.
- BARONI, R. et al. Ressonância magnética da próstata: uma visão geral para o radiologista. **Radiologia Brasileira**, v. 42, p. 185–192, 2009.
- BROWN, A. M. et al. Recent advances in image-guided targeted prostate biopsy. **Abdominal Imaging**, v. 40, p. 1788-1799, 2015.
- CAMPA, R. et al. Improvement of prostate cancer detection combining a computer-aided diagnostic system with TRUS-MRI targeted biopsy. **Abdominal Radiology**, v. 44, p. 264-271, 2018.
- CITAK-ER, F. et al. Final Gleason Score Prediction Using Discriminant Analysis and Support Vector Machine Based on Preoperative Multiparametric MR Imaging of Prostate Cancer at 3T. **BioMed Research International**, v. 2014, 2014.
- CONOVER, W. **Practical Nonparametric Statistic**. 3a. ed. [S.l.]: [s.n.], 1999.
- DINH, H.; AL., E. Characterization of prostate cancer with gleason score of at least 7 by using Quantitative Multiparametric MR Imaging: Validation of a Computer-aided Diagnosis System in Patients Referred for Prostate Biopsy. **Radiology**, v. 000, p. 1-9, 2018.
- DOANE, D.; SEWARD, L. **Estatística Aplicada à Administração e Economia**. 4. ed. [S.l.]: [s.n.], 2014.
- DIKAIOS, N. et al. Zone-specific logistic regression models improve classification of prostate cancer on multiparametric MRI. **Eur Radiol**, v. 25, p. 2727–2737, 2015.

DOLAA, E.; NAKHLAB, O.; GENIDIA, E. Assessing the validity of Prostate Imaging Reporting and Data System version 2 (PI-RADS v2) scoring system in diagnosis of peripheral zone prostate cancer. **European Journal of Radiology**, 4, 2017. 19–26.

DOWNING, D.; CLARK, J. **Estatística aplicada**. 2. ed. [S.l.]: [s.n.], 2002.

ERBERSDOBLER, A. Tumour grade, proliferation, apoptosis, microvessel density, p53, and bcl-2 in prostate cancers: differences between tumours located in the transition zone and in the peripheral zone. **Eur. Urol.**, v. 41, p. 40-46, 2002.

EPSTEIN, J. I. et al. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. **Am J Surg Pathol**, v. 29, p. 1228-1242, 2005.

EPSTEIN, J. I. et al. Contemporary Gleason Grading of Prostatic Carcinoma. **Am J Surg Pathol**, v. 41, p. e1-e7, 2017.

FEI, B. Computer-aided diagnosis of prostate cancer with MRI. **Current Opinion in Biomedical Engineering**, 3, 2017. 20-27.

GRÖNBERG, H. Prostate cancer epidemiology. **LANCET**, v. 361, p. 859-64, 2003.

GREER, M. et al. Validation of the Dominant Sequence Paradigm and Role of Dynamic Contrast-enhanced Imaging in PI-RADS Version 2. **Radiology**, 285, 2017. 859-869.

GREER, M. et al. Accuracy and agreement of PIRADSV2 for prostate cancer mpMRI: A multireader study. **J Magn Reson Imaging**, v. 45(2), p. 579-585, 2017.

GIANNINI, V. et al. A fully automatic computer aided diagnosis system for peripheral zone prostate cancer detection using multi-parametric magnetic resonance imaging. **Computerized Medical Imaging and Graphics**, 2015.

GILLIES, R.; KINAHAN, P.; HRICAK, H. Radiomics: Images Are More than Pictures, They Are Data. **Radiology**, v. 278, p. 563–577, 2016.

GIANNINI, V. et al. Multiparametric magnetic resonance imaging of the prostate with computer-aided detection: experienced observer performance study. **Eur Radiol**, 2017. publicado online.

HAMBROCK, T. et al. Prostate cancer: computer-aided diagnosis with multiparametric 3-T MR imaging—effect on observer performance. **Radiology**, v. 266, p. 521–530, 2013.

HAMOEN, E. H. J. et al. Use of the Prostate Imaging Reporting and Data System (PI-RADS) for Prostate Cancer Detection with Multiparametric Magnetic Resonance Imaging: A Diagnostic Meta-analysis. **European Urology**, 67, n. 6, 2015. 1112 - 1121.

HANSFORD, B. et al. Dynamic Contrast-enhanced MR Imaging Curve-type Analysis: Is It Helpful in the Differentiation of Prostate Cancer from Healthy Peripheral Zone? **Radiology**, v. 275, p. 448-457, 2015.

HOWLADER, N. et al. SEER Cancer Statistics Review, 1975-2014, National Cancer Institute. [https](https://seer.cancer.gov/)

INCA - INSTITUTO NACIONAL DO CÂNCER. Câncer de próstata. **INCA - Instituto Nacional do Câncer**, 2020. Disponível em: <<https://www.inca.gov.br/tipos-de-cancer/cancer-de-prostata>>. Acesso em: 21 dezembro 2020.

ILIC, D. et al. Screening for prostate cancer. **Cochrane Database of Systematic Reviews**, n. 1, 2013.

JEMAL, A. et al. Global cancer statistics. **CA Cancer J Clin**, v. 61(2), p. 69–90, 2011.

JOHNSON, R.; WICHERN, D. **Applied Multivariate Statistical Analysis**. 6a. ed. [S.l.]: Pearson, 2007.

KWAK, J. et al. Automated prostate cancer detection using T2-weighted and high-b-value diffusion-weighted magnetic resonance imaging. **Med Phys**, v. 42, p. 2368-2378, 2015.

KOMATSU, A. Comparação dos poderes dos teste t de Student e Mann-Whitney Wilcoxon pelo método de Monte Carlo. **Revista da Estatística UFOP**, v. VI, p. 121-127., 2017.

KUMAR, V.; GU, Y.; BASU, S. Radiomics: the process and the challenges. **Magn Reson Imaging**, v. 30, p. 1234–1248, 2012.

LANGER, D.; HAIDER, M. A. Prostate Cancer Detection With Multi-parametric MRI: Logistic Regression Analysis of Quantitative T2, Diffusion-Weighted Imaging, and Dynamic Contrast-Enhanced. **Magn. Reson. Imaging**, v. 30, p. 327-334, 2009.

LAMBIN, P.; RIOS-VELAZQUEZ, E.; LEIJENAAR, R. Radiomics: extracting more information from medical images using advanced feature analysis. **Eur J Cancer**, v. 48, p. 441–446, 2012.

LAY, N. et al. Detection of prostate cancer in multiparametric MRI using random forest with instance weighting. **J Med Imaging**, v. 4, p. 24506, 2017.

LAWRENCE, E. M. et al. Prostate cancer: performance characteristics of combined T2W and DW-MRI scoring in the setting of template transperineal re-biopsy using MR-TRUS fusion. **European Radiology**, p. 1497–1505, 2014.

LAVERSANNE, M. et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. **CA CANCER J CLIN**, 71, 2021. 209–249.

LEE, K. M.; MCCARRON, C. E. **Guidelines for the Economic Evaluation of Health Technologies: Canada**. Canadian Agency for Drugs and Technologies in Health. [S.l.]. 2017.

LEMAITRE, G. et al. Computer-Aided Detection for Prostate Cancer Detection based on Multi-Parametric Magnetic Resonance Imaging. **Annu Int Conf IEEE Eng Med Biol Soc.**, v. Jul;2017, p. 3138-3141., 2017

LEHAIRE, J. et al. Computer-aided diagnostic system for prostate cancer detection and characterization combining learned dictionaries and supervised classification. **ICIP**, p. 2251-2255, 2014.

LITJENS, G. et al. Clinical evaluation of a computer-aided diagnosis system for determining cancer aggressiveness in prostate MRI. **Eur Radiol**, v. 25, p. 3187–3199, 2015.

LITJENS, G. et al. Automatic computer aided detection of abnormalities in multi-parametric prostate MRI. **Medical Imaging 2011: Computer-Aided Diagnosis**, v. 79630T, 2011.

LITJENS, G. et al. Computer-aided detection of prostate cancer in MRI. **IEEE Trans Med Imaging**, v. 33(5), p. 1083-1092, 2014.

LIU, L. et al. Computer-aided Detection of Prostate Cancer with MRI. **Academic Radiology**, 23, n. 8, 2016. 1024-1046.

MONTGOMERY, D. 2.4.3 Choice of Sample Size. In: _____ **Design and analysis of experiments**. [S.l.]: John Wiley & Sons, 2013. p. 46-48.

MONTGOMERY, D.; RUNGER, G. **Applied Statistics and Probability for Engineers**. [S.l.]: John Wiley & Sons, 2002.

MINISTÉRIO DA SAÚDE/SVS/DASIS/CGIAE/SISTEMA DE INFORMAÇÃO SOBRE MORTALIDADE - SIM. Atlas On-line de Mortalidade. Disponível em: <<https://mortalidade.inca.gov.br/>>. Acesso em: maio 2021.

MUSSI, T. A ressonância magnética multiparamétrica de próstata nos dias atuais. **einstein (São Paulo)**, v. 16, p. eMD4408, 2018.

NARDI, A. C. et al. **Diretrizes urologia AMB**. Rio de Janeiro: [s.n.], 2014.

NAM, R. et al. Assessing individual risk for prostate cancer. **J Clin Oncol.**, v. 25(24), p. 3582-8, 2007.

NATIONAL Institutes of Health. Disponível em:
<<http://training.seer.cancer.gov/prostate/abstract-code-stage/morphology.html>>. Acesso em: 28 out. 2019.

NIU, X. et al. Developing a nomogram based on multiparametric magnetic resonance imaging for forecasting high-grade prostate cancer to reduce unnecessary biopsies within the prostate-specific antigen gray-zone. **BMC Med Imaging** **17**, v. 17, p. 11–19, 2017.

PARKIN, M. et al. Global cancer statistics, 2002. **A Cancer Journal for Clinicians**, v. 55, n. 2, p. 74-108, 2005.

PATEL, N. et al. Assessment of PI-RADS v2 categories ≥ 3 for diagnosis of clinically significant prostate cancer. **Abdom Radiol**, v. 44(2), p. 705-712, 2019.

PENG, Y. et al. Quantitative analysis of multiparametric prostate MR images: differentiation between prostate cancer and normal tissue and correlation with Gleason score—a computer-aided diagnosis development study. **Radiology**, v. 267, p. 787–796, 2013.

PIERORAZIO, P. M. et al. Prognostic Gleason grade grouping: data based on the modified Gleason scoring system. **BJU International**, v. 111, 2013. ISSN 753-760.

POLO, T.; MIOT, H. Aplicações da curva ROC em estudos clínicos e experimentais. **J Vasc Bras**, 19, 2020. e20200186.

PUECH, P. et al. Prostate MRI: can we do without DCE sequences in 2013? **Diagn Interv Imaging**, v. 94, p. 1299–1311, 2013.

RAWLA, P. Epidemiology of Prostate Cancer. **World J Oncol.**, 10(2), 2019. 63-89.

QIAN, C. et al. In vivo MRI based prostate cancer identification with random forests and auto-context model. **Int Workshop Mach Learn Med Imaging**, v. 8679, p. 314–322, 2014.

RAZALI, N.; WAH, Y. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. **Journal of Statistical Modeling and Analytics**, 2, n. 1, 2011. 21-33.

ROSS, S. **Introduction to Probability and Statistics of Engineers and Scientists**. 3^a ed. ed. [S.l.]: Elsevier, 2004. 624 p.

SANDHU, S. et al. Prostate Cancer. **The Lancet**, ago. 2021.

RESEARCH, W. C. R. F. I. F. C. **Continuous Update Project Expert Report 2018. Diet, nutrition, physical activity and prostate cancer.** [S.l.]. 2018.

SERPA NETO, A. et al. Estudo descritivo da mortalidade por câncer de próstata no Estado de São Paulo no período de 1980-2007. **Einstein (São Paulo)**, v. 8, n. 4, p. 433-436, out 2010.

SIEGEL, R. L.; MILLER, K. D.; JEMAL, A. Cancer Statistics. **Cancer J Clin**, 2017. 7-30.

SOARES, J.; SIQUEIRA, A. **Introdução à estatística médica.** 1. ed. [S.l.]: [s.n.], 1999.

SUNG, H. et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. **CA CANCER J CLIN**, 71, 2021. 209–249.

SUN, Y. et al. Multiparametric MRI and radiomics in prostate cancer: a review. **Australasian Physical & Engineering Sciences in Medicine**, v. 42, p. 3-25, 2019.

TRAVERSO, A. et al. Repeatability and Reproducibility of Radiomic Features: A Systematic Review. **International Journal of Radiation Oncology**, v. 102, 2018.

TROGDON, J. et al. Total Medicare Costs Associated With Diagnosis and Treatment of Prostate Cancer in Elderly Men. **JAMA Oncol**, v. 5(1), p. 60-66, 2019.

ÜTTERER, J. et al. Prostate cancer localization with dynamic contrast-enhanced MR imaging and proton MR spectroscopic imaging. **Radiology**, v. 241, p. 449-458, 2006.

KUMAR, V.; GU, Y.; BASU, S. Radiomics: the process and the challenges. **Magn Reson Imaging**, v. 30, p. 1234–1248, 2012.

NIAF, E. et al. Prostate Focal Peripheral Zone lesions: Characterization at Multiparametric MR Imaging - nfluence of a Computer-aided Diagnosis System. **Radiology: Volume : Number 3—June** , v. 271, n. 3, p. 761-769, 2014.

VOS, P. et al. Computer-assisted analysis of peripheral zone prostate lesions using T2-weighted and dynamic contrast enhanced T1-weighted MRI. **Phys Med Biol**, v. 55, p. 1719-1734, 2010.

WANG, X. et al. The diagnostic value of PI-RADS V1 and V2 using multiparametric MRI in transition zone prostate clinical cancer. **Oncology Letters**, 16, n. 3, 2018. 3201-3206.

WOO, S. et al. Diagnostic Performance of Prostate Imaging Reporting and Data System Version 2 for Detection of Prostate Cancer: A Systematic Review and Diagnostic Meta-analysis. **European Urology**, 72, n. 2, 2017. 177–188.

ZHANG, L. et al. A meta-analysis of use of Prostate Imaging Reporting and Data System Version 2 (PI-RADS V2) with multiparametric MR imaging for the detection of prostate cancer. **European Radiology**, 27, n. 12, 2017. 5204–5214.

APÊNDICES

APÊNDICE A – TABELAS

Tabela 17 – Características dos artigos selecionados

#	Artigo	Técnica	Validação	Região	Classificador	Amostra
1	Viswanath 2011	T2, DWI, DCE	seções histológicas	próstata total	outro	12
2	Lopes 2011	T2	seções histológicas	próstata total	SVM	17
3	Liu 2011	T2, DWI, DCE	seções histológicas	próstata total	SVM	20
4	Sung 2011	DCE	seções histológicas	zona periférica zona de transição	SVM	42
5	Tiwari 2011	T2, espectroscopia	seções histológicas	próstata total	<i>random forest</i>	36
6	Viswanath 2012	T2	seções histológicas	zona periférica glândula central	quadratic discriminant analysis	22
7	Vos 2012	T1, T2, ADC, DCE	biópsia	próstata total	LDA	200
8	Niaf 2012	T2, DWI, DCE	seções histológicas	zona periférica	SVM	30
9	Artan 2012	T2, ADC, T1 pós contraste	seções histológicas	próstata total	SVM	15
10	Shah 2012	T2, ADC, DCE	seções histológicas	zona periférica	SVM	31
11	Hambrock 2013	T2, DWI, DCE	seções histológicas	zona periférica zona de transição	outro	34
12	Tiwari 2013	T2, espectroscopia	seções histológicas	próstata total	outro	29
13	Cameron 2015	T2, DWI, ADC, DCE	seções histológicas	próstata total	SVM	13
14	Jurek 2017	T2, DWI, ADC, DCE	seções histológicas	zona periférica	outro	10

#	Artigo	Técnica	Validação	Região	Classificador	Amostra
15	Lehaire 2014	T2, DWI, DCE	seções histológicas	zona periférica	SVM Regressão linear	35
16	Lemaitre 2017	T2, DWI, DCE	biópsia	próstata total	<i>random forest</i>	17
17	Niaf 2014	T2, DWI, DCE	seções histológicas	zona periférica	SVM	30
18	Rezaeilouyeh 2014	T2, ADC, DCE	seções histológicas	próstata total	SVM	4
19	Trigui 2016	T2, DWI, DCE, espectroscopia	biópsia	próstata total	SVM <i>random forest</i>	20
20	Tsehay 2017	T2, DWI, ADC	biópsia	próstata total	redes neurais	196
21	Giannini 2015	T2, DWI, ADC	seções histológicas	próstata total	SVM	56
22	Li 2018	T2, DWI, DCE	biópsia	glândula central	SVM	48
23	Trigui 2017	T2, DWI, DCE, espectroscopia	biópsia	próstata total	SVM <i>random forest</i>	34
24	Sumathipala 2018	T2, DWI, ADC	seções histológicas	próstata total	redes neurais	186
25	Campa 2018	T2, DWI, DCE	biópsia	próstata total	outro (comercial)	63
26	Song 2018	T2, DWI, ADC	seções histológicas	próstata total	redes neurais	55
27	Dinh 2018	ADC, DCE	biópsia e seções histológicas	zona periférica	não declarado	129
28	Thon 2017	T2, ADC, DCE	biópsia	próstata total	outro (comercial)	104
29	Lay 2017	T2, DWI, ADC	biópsia e seções histológicas	próstata total	<i>random forest</i>	224

#	Artigo	Técnica	Validação	Região	Classificador	Amostra
30	Kwak 2015	T2, DWI	biópsia	próstata total	SVM	136
31	Citak-er 2014	T2, DWI, DCE	seções histológicas	próstata total	LDA SVM	33
32	Peng 2013	T2, ADC, DCE	seções histológicas	próstata total	LDA	48
33	Xu 2019	T2, DWI, ADC	biópsia	próstata total	redes neurais	346
34	Mehralivand 2020	T2, DWI, DCE	seções histológicas	próstata total	<i>random forest</i>	236
35	Pellicer-valero 2021	T2, DWI, ADC, DCE	seções histológicas	próstata total	redes neurais	75
36	Hiremath 2020	ADC	seções histológicas	próstata total	redes neurais	115

Tabela 18 – Ponderações de RM utilizadas nos estudos

Ponderação	Quantidade de estudos	%
T2	33	91,6
DWI	27	75,0
ADC	17	47,2
DCE	23	63,8
Espectroscopia	4	11,1

Tabela 19 – Região de análise do CAD por quantidade de estudos publicados.

Região	Quantidade de estudos	%
Próstata total	26	72%
Zona periférica	6	17%
Glândula central ou combinações incluindo a zona periférica	4	11%
Total	36	100%

Tabela 20 – Região de análise do CAD por quantidade de estudos publicados

Região	Número de amostras	%
Próstata total	2.290	85%
Zona periférica	265	10%
Glândula central ou combinações incluindo a zona periférica	146	5%
Total	2.701	100%

Tabela 21 – Número de artigos (estudos) publicados e número de amostras por ano de publicação do estudo, e seus correspondentes percentuais.

Ano de publicação do estudo	Número de artigos (estudos)	(%)	Número de amostras	(%)
2011	5	13,9%	127	4,7%
2012	5	13,9%	298	11,0%
2013	3	8,3%	111	4,1%
2014	4	11,1%	102	3,8%
2015	3	8,3%	205	7,6%
2016	1	2,8%	20	0,7%
2017	6	16,7%	585	21,7%
2018	5	13,9%	481	17,8%
2019	1	2,8%	346	12,8%
2020	2	5,6%	351	13,0%
2021	1	2,8%	75	2,8%
Total	36	100,0%	2701	100%

Tabela 22 – Número de artigos publicados por ano de publicação e por tipo de classificador

Ano de publicação do estudo	SVM	Random Forest	Redes neurais
2011	3	1	
2012	3		
2013			
2014	2		
2015	3		
2016			
2017		2	1
2018	1		3
2019			1
2020		1	2
2021			1
Total	16	5	6

Tabela 23 – Resultados dos classificadores na detecção do CaP

Ano da publicação	Autor	Síntese do estudo
2011	Satish Viswanath	Área sob a curva ROC: 0,77
2011	R. Lopes	Sensibilidade 0,83 Especificidade 0,91
2011	Xin Liu	Área sob a curva ROC: 0,89 Sensibilidade 0,8 Especificidade 0,84
2011	Yu Sub Sung	Zona periférica: sensibilidade 0,89, especificidade 0,89 Zona de transição: sensibilidade 0,91, especificidade 0,64 Média: sensibilidade 0,90, especificidade 0,77
2011	P. Tiwari	Área sob a curva ROC: 0,89
2012	Satish Viswanath	Zona periférica: área sob a curva ROC: 0,73 Glândula central: área sob a curva ROC: 0,86 Média: área sob a curva ROC: 0,795
2012	P. C. Vos	Sensibilidade 0,74
2012	Emilie Niaf	Área sob a curva ROC: 0,89
2012	Yusuf Artan	Sensibilidade 0,76 Especificidade 0,86
2012	Vijay Shah	Sensibilidade 0,9 Especificidade 0,9
2013	Thomas Hambrock	Área sob a curva ROC: 0,91

Ano da publicação	Autor	Síntese do estudo
2013	Pallavi Tiwari	Área sob a curva ROC: 0,89
2015	Andrew Cameron	Sensibilidade 0,86 Especificidade 0,88
2017	Jakub Jurek	Área sob a curva ROC: 0,67
2014	Jerome Lehaire	SVM: área sob a curva ROC: 0,72 RL: área sob a curva ROC: 0,84
2017	Guillaume Lemaitre	Área sob a curva ROC: 0,836
2014	Emilie Niaf	Área sob a curva ROC: 0,82
2014	Hadi Rezaeilouyeh	T2 sensibilidade: 0,92 especificidade: 0,83 ADC sensibilidade: 0,89 especificidade: 1,0 DCE sensibilidade: 1,0 especificidade: 1,0 Média: sensibilidade: 0,94 especificidade: 0,94
2016	Rania Trigui	SVM sensibilidade: 0,72 especificidade: 0,87 Random forest sensibilidade: 0,72 especificidade: 0,87
2017	Yohannes Tsehay	Área sob a curva ROC: 0,9
2015	Valentina Giannini	Área sob a curva ROC: 0,91
2018	Jiance Li	Área sob a curva ROC: 0,91
2017	Rania Trigui	SVM sensibilidade: 0,72 especificidade: 0,88 Random forest sensibilidade: 0,72 especificidade: 0,88

Ano da publicação	Autor	Síntese do estudo
2018	Yohan Sumathipala	Área sob a curva ROC: 0,88
2018	Riccardo Campa	Sensibilidade: 0,52 Especificidade: 0,71 Área sob a curva ROC: 0,63
2018	Yang Song	Sensibilidade: 0,87 Especificidade: 0,91 Área sob a curva ROC: 0,944
2018	Au Hoang Dinh	Sensibilidade: 1,0 Especificidade: 0,395 Área sob a curva ROC: 0,95
2017	Anika Thon	Sensibilidade: 0,47 Especificidade: 0,755 Área sob a curva ROC: 0,65
2017	Nathan Lay	Área sob a curva ROC: 0,93
2015	Jin Tae Kwak	Área sob a curva ROC: 0,89
2014	Fusun Citak-Er	LDA sensibilidade: 0,865 especificidade: 0,639 SVM sensibilidade: 0,879 especificidade: 0,568
2013	Yahui Peng	Área sob a curva ROC: 0,95
2019	Helen Xu	Área sob a curva ROC: 0,97
2020	Sherif Mehralivand	Área sob a curva ROC: 0,775

Ano da publicação	Autor	Síntese do estudo
2021	Oscar Pellicer-Valero	Sensibilidade: 1,0 Especificidade: 0,76 Área sob a curva ROC: 0,91
2020	Amogh Hiremath	Área sob a curva ROC: 0,78 - 0,85 Média: 0,815

Tabela 24 – Resultados das métricas de desempenho empregadas, por tipo do classificador CAD

Tipo do classificador CAD	Métricas de desempenho	Número de artigos	Mínimo	Máximo	Média	Desvio padrão	Mediana	Coefficiente de variação
SVM	Área sob a curva ROC	7	0,72	0,91	0,86	0,069	0,89	8,0%
	Sensibilidade	10	0,7	0,94	0,83	0,081	0,845	9,8%
	Especificidade	10	0,57	0,94	0,84	0,106	0,875	12,6%
Redes Neurais	Área sob a curva ROC	6	0,82	0,97	0,9	0,054	0,935	6,0%
	Sensibilidade	2	0,87	1	0,935	0,092	0,90	9,8%
	Especificidade	2	0,76	0,91	0,835	0,106	0,835	12,7%
<i>Random Forest</i>	Área sob a curva ROC	3	0,78	0,93	0,85	0,078	0,836	9,2%
	Sensibilidade ^a	2	-	-	-	-	-	-
	Especificidade	2	0,87	0,88	0,88	0,007	0,875	0,8%
Outros	Área sob a curva ROC	11	0,63	0,95	0,81	0,119	0,84	14,7%
	Sensibilidade	5	0,47	1	0,72	0,225	0,74	31,3%
	Especificidade	4	0,62	0,39	0,76	0,16	0,675	21,1%

^a A variável sensibilidade foi constante (0,72) nos dois únicos casos da amostra.

Tabela 25 – Resultados de estudos e metanálises para validação do PIRADS.

#	Estudo	Área sob a curva ROC	Sensibilidade	Especificidade
1	Dola 2017		0,88	0,934
2	Greer 2016		0,61	
3	Muller 2015	0,865	0,865	0,63
4	Wang 2018	0,888	0,75	0,902
5	Hamoen 2015		0,78	0,79
6	Woo 2017		0,89	0,73
7	Zhang 2017		0,85	0,71
8	Greer 2019		0,92	0,65
9	Benndorf 2017	0,76		
10	Patel 2018		0,812	0,32

APÊNDICE B – OUTRAS FIGURAS


Figura 34 – Ferramenta de extração de dados

Ferramenta para extração de dados

Estudo	
Modalidades	<input type="checkbox"/> T2 <input type="checkbox"/> ADC <input type="checkbox"/> Contraste dinâmico <input type="checkbox"/> Espectroscopia
Validação	<input type="checkbox"/> Biópsia <input type="checkbox"/> Prostatectomia <input type="checkbox"/> Secções histológicas totais <input type="checkbox"/> Seguimento (<i>follow-up</i>)
Região	<input type="checkbox"/> Toda próstata <input type="checkbox"/> Zona periférica <input type="checkbox"/> Zona de transição <input type="checkbox"/> Glândula central
Classificadores	<input type="checkbox"/> SVM (<i>support vector machine</i>) <input type="checkbox"/> AAN (<i>artificial neural network</i>) <input type="checkbox"/> LDA (<i>linear discriminant analysis</i>) <input type="checkbox"/> QDA (<i>quadratic discriminant analysis</i>) <input type="checkbox"/> Outros (descrever): _____ _____
Tamanho da amostra	
Resultados	<input type="checkbox"/> Sensibilidade: _____ <input type="checkbox"/> Especificidade: _____ <input type="checkbox"/> Curva ROC: _____ <input type="checkbox"/> Outros: _____

Fonte: o autor, 2020

Figura 35 – Ficha de avaliação radiológica das imagens de MR-MP da próstata utilizada no EXPERIMENTO COM RADIOLOGISTAS.


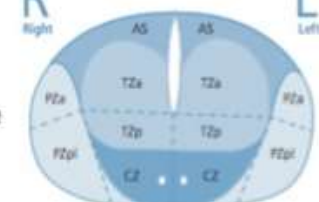
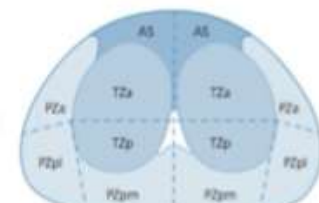
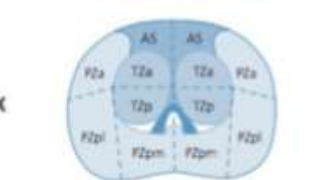



PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS E TECNOLOGIA EM SAÚDE

REVISÃO SISTEMÁTICA DO COMPUTER-AIDED DIAGNOSIS (CAD) NA AVALIAÇÃO DE NÓDULOS PROSTÁTICOS ATRAVÉS DA RESSONÂNCIA MAGNÉTICA MULTIPARAMÉTRICA

FORMULÁRIO EXTRAÇÃO DE DADOS:

DETECÇÃO E CLASSIFICAÇÃO DE CaP EM RM-MP PELO RADIOLOGISTA

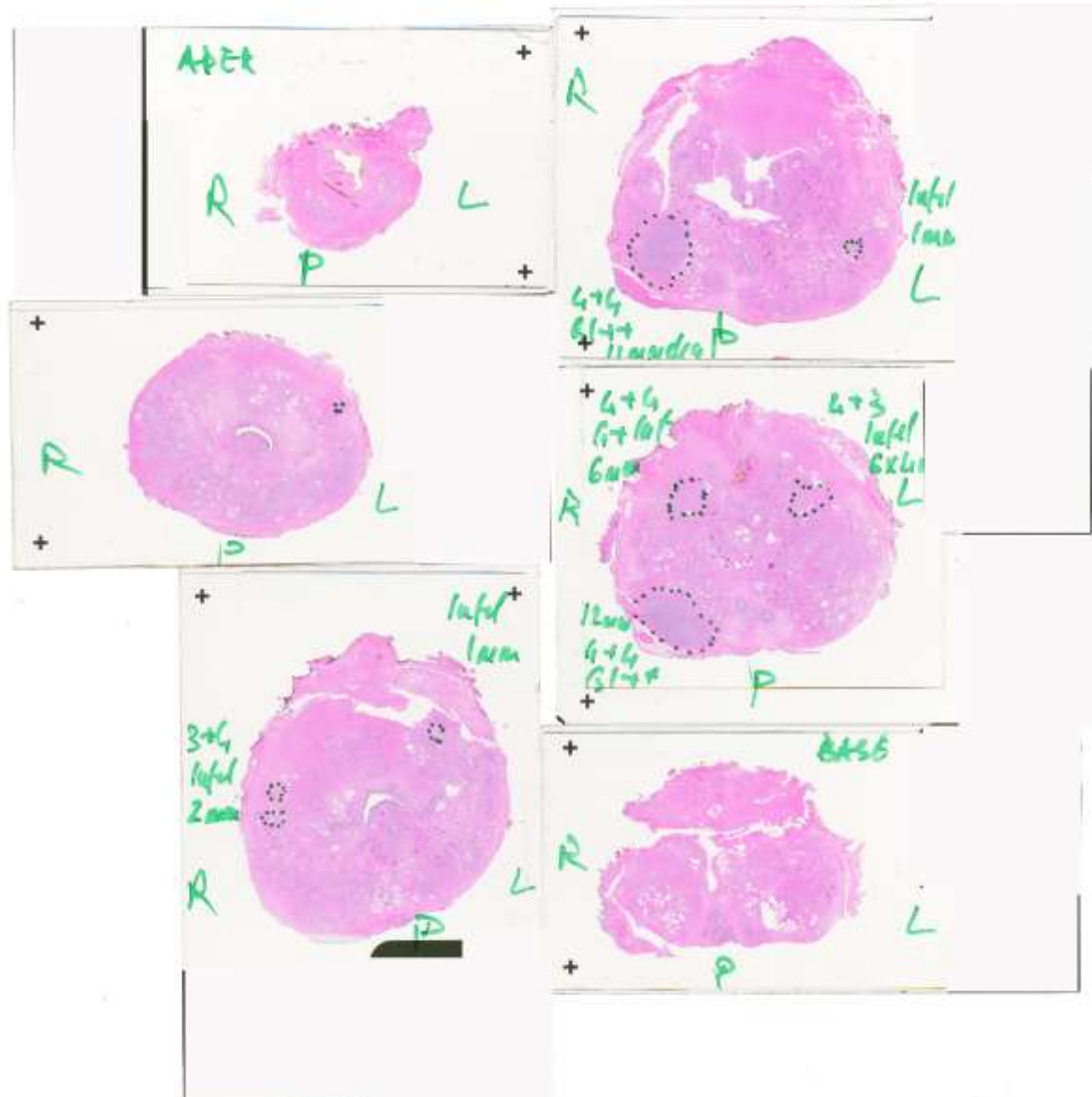
RADIOLOGISTA	B	CASO	MIP-PROSTATE-01-0001
<p>SV</p>  <p>Base</p>  <p>Mid</p>  <p>Apex</p>  <p>Urethra</p> 			<p style="text-align: center;">INSTRUÇÕES</p> <p>1. UTILIZE O PINCEL VERMELHO PARA IDENTIFICAR A LESÃO PRINCIPAL NO GRÁFICO AO LADO. CASO HAJA OUTRA LESÃO, MARQUE COM PINCEL VERDE. SE HOVER MAIS DE DUAS LESÕES, IGNORE AS DEMAIS.</p> <p>2. CLASSIFIQUE A LESÃO PRINCIPAL (E A SECUNDÁRIA, SE HOVER) DE ACORDO COM A CATEGORIZAÇÃO PIRADS.</p> <p>LESÃO PRINCIPAL PIRADS: _____ (VERMELHO)</p> <p>LESÃO SECUNDÁRIA PIRADS: _____ (VERDE)</p> <p style="text-align: center;">OBRIGADO PELA SUA PARTICIPAÇÃO!</p>

Fonte: o autor, 2021

ANEXOS

ANEXO A – OUTRAS FIGURAS

Figura 36 – Exemplo de anatomia patológica de espécime de próstata oriundo de prostatectomia total na coloração hematoxilina-eosina, identificando-se lesões compatíveis com CaP.



Fonte: disponível em <http://i2cvb.github.io/>

ANEXO B – CHECKLIST REVISÃO SISTEMÁTICA (baseado em PRISMA -
www.prisma-statement.org)

CHECKLIST REVISÃO SISTEMÁTICA		
Tópico	Item	Localização
Título	Identifique o artigo como uma revisão sistemática, metanálise ou ambas.	Capa
Resumo estruturado	Apresente um resumo estruturado incluindo, se aplicável: referencial teórico; objetivos; fonte de dados; critérios de elegibilidade; participantes e intervenções; avaliação do estudo e síntese dos métodos; resultados; limitações; conclusões e implicações dos achados principais; número de registro da revisão sistemática (revisão ainda não registrada).	Resumo e <i>abstract</i>
Racional	Descreva a justificativa da revisão no contexto do que já é conhecido.	Item 1.3
Objetivos	Apresente uma afirmação explícita sobre as questões abordadas com referência a participantes, intervenções, comparações, resultados e desenho de estudos (PICOS).	Item 1.4
Protocolo e registro	Indique se existe um protocolo de revisão, se e onde pode ser acessado (ex.: endereço eletrônico), e, se disponíveis, forneça informações sobre o registro da revisão, incluindo o número de registro.	Em registro no PROSPERO, aguardando resposta.
Critérios de elegibilidade	Especifique características do estudo (ex.: PICOS, extensão do seguimento) e características dos relatos (ex.: anos considerados, idioma, se é publicado) usadas	Item 3.1

	como critérios de elegibilidade, apresentando justificativa.	
Fontes de informação	Descreva todas as fontes de informação na busca.	Item 3.1
Busca	Apresente a estratégia completa de busca eletrônica para pelo menos uma base de dados, incluindo os limites utilizados, de forma que possa ser repetida.	Item 3.1
Seleção dos estudos	Apresente o processo de seleção dos estudos.	Item 3.1
Processo de coleta de dados	Descreva o método de extração de dados dos artigos.	Item 3.1
Lista de dados	Liste e defina todas as variáveis obtidas dos dados e quaisquer suposições ou simplificações realizadas.	Item 3.1
Risco de viés em cada estudo	Descreva os métodos usados para avaliar o risco de viés em cada estudo (incluindo a especificação, se foi feita durante o estudo ou no nível de resultados), e como essa informação foi usada na análise de dados.	Item 3.2
Medidas de sumarização	Defina as principais medidas de sumarização dos resultados.	Item 3.4
Síntese dos resultados	Descreva os métodos de análise dos dados e combinação de resultados dos estudos, se realizados, incluindo medidas de consistência para cada metanálise.	Item 4.2
Risco de viés entre estudos	Especifique qualquer avaliação do risco de viés que possa influenciar a evidência cumulativa (ex.: viés de publicação, relato seletivo nos estudos).	Item 3.2

Análises adicionais	Descreva métodos de análise adicional, se realizados, indicando quais foram pré-especificados.	Não aplicável
Seleção de estudos	Apresente números dos estudos rastreados, avaliados para elegibilidade e incluídos na revisão, razões para exclusão em cada estágio, preferencialmente por meio de gráfico de fluxo.	Figura 8
Características de estudos	Para cada estudo, apresente características para extração dos dados e apresente as citações.	Tabela 5
Risco de viés em cada estudo	Apresente dados sobre o risco de viés em cada estudo e, se disponível, alguma avaliação em resultados.	Quando disponíveis, avaliados os <i>outliers</i>
Resultados de estudos individuais	Para todos os desfechos considerados (benefícios ou riscos), apresente, para cada estudo: (a) sumário simples de dados para cada grupo de intervenção; e (b) efeitos estimados e intervalos de confiança.	Tabelas 16 e 23
Síntese dos resultados	Apresente resultados para cada metanálise feita, incluindo intervalos de confiança e medidas de consistência.	Figuras 28, 29 e 30. Tabela 27
Risco de viés entre os estudos	Apresente resultados da avaliação de risco de viés entre os estudos	Não avaliável
Análises adicionais	Apresente resultados de análises adicionais, se realizadas (ex.: análise de sensibilidade ou subgrupos, metarregressão)	Experimento com radiologistas (item 4.5)
Sumário de evidência	Sumarize os resultados principais, incluindo a força de evidência para cada resultado; considere sua relevância	Item 4.6

	para grupos-chave (ex.: profissionais da saúde, usuários e formuladores de políticas).	
Limitações	Discuta limitações no nível dos estudos e dos desfechos.	Item 4.6
Conclusões	Apresente a interpretação geral dos resultados no contexto de outras evidências e implicações para futuras pesquisas.	Item 5
Financiamento	Descreva fontes de financiamento para revisão sistemática e outros suportes.	Não há financiamento.