

M

MESTRADO

GESTÃO DAS ORGANIZAÇÕES, RAMO: GESTÃO DE UNIDADES DE SAÚDE

Utilização da Inteligência Artificial na otimização do diagnóstico do cancro da próstata

Gabriela Maria da Costa Araújo Basaloco

Utilização da Inteligência Artificial na otimização do diagnóstico do cancro da próstata

Autor

Gabriela Maria da Costa Araújo Basaloco

Orientador

Professor Doutor Rui Pimenta Escola Superior de Saúde do Instituto Politécnico do Porto

Dissertação de Estágio apresentada(o) para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Gestão das Organizações– Ramo: Unidades de Saúde pela Escola Superior de Saúde do Instituto Politécnico do Porto

Agradecimentos

Gostaria de expressar a minha profunda gratidão a todos aqueles que, de forma direta ou indireta, contribuíram para a realização deste projeto de mestrado sobre a utilização da inteligência artificial na otimização do diagnóstico do cancro da próstata.

Em primeiro lugar, agradeço ao meu orientador professor Doutor Rui Pimenta, pelo apoio incansável, orientação e partilha de conhecimentos, que foram essenciais para o desenvolvimento deste trabalho. O incentivo constante foi imperativo para enfrentar os desafios e ultrapassar as barreiras ao longo deste percurso.

À minha família, pela compreensão e apoio incondicional em todos os momentos, especialmente durante as fases mais intensas do projeto. Sem dúvida que o seu suporte foi um pilar essencial para alcançar este objetivo.

Aos meus amigos e colegas, pela partilha de momentos de estudo e de motivação, e por me acompanharem nesta jornada académica. A sua presença foi inestimável para manter a motivação e o foco.

Finalmente, agradeço aos investigadores e profissionais que contribuem diariamente para o avanço da inteligência artificial na área da saúde, cujo trabalho me inspirou a explorar o potencial desta tecnologia na otimização do diagnóstico do cancro da próstata.

A todos, o meu sincero agradecimento. Este projeto é o resultado de uma caminhada conjunta e de um esforço coletivo que possibilitou a sua concretização.

Resumo

Introdução: O cancro da próstata é uma das neoplasias mais prevalentes entre os homens, exigindo diagnósticos cada vez mais precisos e padronizados. A inteligência artificial (IA) surge como uma ferramenta emergente com potencial para complementar e otimizar os métodos atualmente disponíveis.

Metodologia: Esta dissertação teve como objetivo avaliar o contributo da IA no diagnóstico do cancro da próstata, através de uma revisão sistemática da literatura e de uma análise de dados. Foram incluídos 14 estudos na revisão, destacando-se modelos baseados em redes neurais treinados com imagens histológicas e dados de imagem médica.

Resultados: Os estudos analisados demonstraram desempenho elevado dos algoritmos, nalguns casos comparáveis ao de patologistas experientes. A análise de dados, que integrou 8 estudos com métricas padronizadas, confirmou a eficácia dos modelos de IA, revelando uma sensibilidade média de 99,18%, especificidade de 87,20% e um índice de Kappa de 0,80, sugerindo concordância quase perfeita com os padrões de referência.

Conclusão: Apesar dos resultados promissores, foram identificadas limitações, como a heterogeneidade entre os estudos e a escassez de validação externa. Ainda assim, conclui-se que a IA representa uma ferramenta com potencial real para apoiar a decisão clínica, desde que integrada de forma ética, segura e sustentada em evidência científica.

Palavras-chave: Inteligência Artificial; Cancro da Próstata; Diagnóstico Médico; Concordância de Diagnóstico.

Abstract

Introduction: Prostate cancer is one of the most prevalent neoplasms among men, requiring increasingly accurate and standardized diagnostic. Artificial intelligence (AI) emerges as an innovative tool with the potential to complement and optimize currently available methods.

Methodology: This dissertation aimed to assess the contribution of AI to the diagnosis of prostate cancer through a systematic literature review and a quantitative statistic analysis. Fourteen studies were included in the review, with emphasis on models based on neural networks trained with histological images and medical imaging data.

Results: The analyzed studies demonstrated a high performance of the algorithms in some cases comparable to that of experienced pathologists.

The statistic analysis, which integrated 8 studies with standardized metrics, confirmed the effectiveness of AI models, revealing a mean sensitivity of 99,18%, specificity of 87,29% and a Kappa index of 0,80, suggesting an almost perfect agreement with reference standards.

Conclusion: Despite these promising results, some limitations were identified, such as methodological heterogeneity and the scarcity of external validation. Nevertheless, it is concluded that AI represents a tool with real clinical potential to support decision-making, provided it is integrated in an ethical, safe and supported by scientific evidence.

Keywords: Artificial Intelligence; Prostate Cancer; Medical Diagnosis; Diagnostic Agreement.

Índice

1.	Introdução.....	1
2.	Enquadramento Teórico	2
2.1.	Inteligência Artificial.....	2
2.1.1.	Aplicação de IA na saúde.....	3
2.1.2.	Implicações éticas e legais da IA.....	6
2.2.	Cancro da Próstata.....	8
2.2.1.	Custos do cancro da próstata na União Europeia e em Portugal.....	9
2.2.2.	Epidemiologia do cancro da próstata na União Europeia e em Portugal.....	9
2.2.3.	Etiologia.....	15
2.2.4.	Fatores de Risco	15
2.2.5.	Deteção e Diagnóstico.....	18
2.2.6.	Estadiamento.....	20
2.2.7.	Tratamento	21
2.2.8.	Otimização no diagnóstico do cancro da próstata.....	21
3.	Metodologia.....	22
3.1.	Definição da pesquisa e identificação dos estudos.....	23
3.2.	Análise de dados	24
3.3.	Tipos de concordância.....	24
3.4.	Métodos estatísticos para avaliar a concordância em variáveis contínuas.....	25
3.5.	Métodos estatísticos para avaliar concordância em variáveis categóricas.....	25
3.6.	Avaliação da qualidade.....	26
3.7.	Discussão dos Resultados.....	34
4.	Análise estatística.....	Erro! Marcador não definido.
4.1.	Caracterização geral dos estudos.....	37
4.2.	Análise de Métricas Contínuas.....	38
4.2.1.	Análise descritiva de AUC	39
4.2.2.	Análise de Métricas:Sensibilidade e Especificidade.....	44
4.2.3.	Análise de Métrica: Índice de Kappa.....	54
4.2.4.	Análise de métrica: Acurácia.....	57
5.	Discussão de resultados da análise estatística.....	58

6. Conclusão.....	60
Referências Bibliográficas.....	62

Abreviaturas

ADN – Ácido Desoxirribonucleico
APAH – Associação Portuguesa de Administradores Hospitalares
ATM – Ataxia Telangiectasia Mutada
AUC – Area Under the Curve
bpMRI – Biparametric Magnetic Resonance Imaging
BRCA – Breast Cancer Gene
DL – Deep Learning
DWI – Diffusion-Weighted Imaging
ECIS – European Cancer Information System
ELAC2 – ELA Cystein-Rich Protein 2
EUA – Estados Unidos da América
FRA – Fundamental Rights Agency
GCO – Global Cancer Observatory
GPAN – Grupo de peritos independente Criado pela Comissão Europeia
GRADE – Grading of Recommendations Assessment, Development and Evaluation
HBP – Hiperplasia Benigna da Próstata
HOX – Homeobox gene
IA – Inteligência Artificial
IARC – International Agency for Research on Cancer
JI – Joanna Briggs Institute
ML – Machine Learning
mpMRI – Multiparametric Magnetic Resonance Imaging
MSR1 – Macrophage Scavenger Receptor 1
OCDE – Organização para a Cooperação e Desenvolvimento Económico
PET-PSMA – Positron Emission Tomography-Prostate Specific Membrane Antigen
PET-TC – Positron Emission Tomography-Computed Tomography
PI-RADS – Prostate Imaging-Reporting and Data System
PRISMA – Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSA – Antígeno Específico da Próstata
RGPD – Regulamento Geral sobre a Proteção de Dados
RMN – Ressonância Magnética Nuclear
RNaseL – Ribonuclease L enzyme
ROC – Receiver Operating Characteristic Curve
SPMS – Serviços Partilhados do Ministério de Saúde
SNS – Serviço Nacional de Saúde
TAC – Tomografia Axial Computorizada
UE – União Europeia

Índice de Figuras

Figura 1 – Custo estimado do cancro per capita na UE, relativos a 2018	9
Figura 2 – Percentagem da incidência dos cancros na UE, relativos a 2022	10
Figura 3 – Incidência do cancro da próstata nos países da UE, relativos a 2022.....	11
Figura 4 – Percentagem da incidência dos cancros em Portugal, relativo a 2022	11
Figura 5 – Taxa de mortalidade por cancro na UE, relativo a 2022	12
Figura 6 – Taxa de mortalidade por cancro da próstata na UE, relativo a 2022.....	13
Figura 7 – Taxa de mortalidade por cancro em Portugal, relativo a 2022	14
Figura 8 – Prevalência do cancro da próstata nos países da UE durante 1 ano, relativo a 2022	15
Figura 9 – Número de casos de cancro da próstata no mundo dos 0 aos +85 anos, relativos a 2022.....	16
Figura 10 – Número de casos de cancro da próstata no mundo dos 50 aos 84 anos, relativos a 2022.....	16
Figura 11 – Diagrama da aplicação da metodologia PRISMA.....	27
Figura 12 – Representação gráfica da AUC dos estudos incluídos na análise estatística, com barras de erro correspondentes a +/- 2 erros padrão (SE), aproximando-se de um intervalo de confiança de 95% (fonte: Excel)	40
Figura 13 – Representação gráfica dos valores de sensibilidade obtidos nos estudos incluídos (fonte: SPSS v29.0.2.0).....	45
Figura 14 – Representação gráfica dos valores de especificidade obtidos nos estudos incluídos (fonte: SPSS v29.0.2.0).....	45
Figura 15 – Representação gráfica de dispersão com linha de tendência linear, entre sensibilidade e especificidade (fonte: SPSS v29.0.2.0)	47
Figura 16 – Representação gráfica da métrica sensibilidade dos estudos incluídos na análise estatística, com barras de erro correspondentes a +/- 2 erros padrão (SE), aproximando-se de um intervalo de confiança de 95% (fonte: Excel)	50
Figura 17 – Representação gráfica da métrica especificidade dos estudos incluídos na análise estatística, com barras de erro correspondentes a +/- 2 erros padrão (SE), aproximando-se de um intervalo de confiança de 95% (fonte:Excel).....	50
Figura 18 – Representação gráfica dos valores de índice de Kappa obtidos nos estudos (fonte: SPSS v29.0.2.0).....	56

Índice de Tabelas

Tabela 1 – Características e resultados da análise metodológica dos estudos.....	27
Tabela 2 – Estudos incluídos na análise estatística com métricas quantitativas de desempenho da IA.....	37
Tabela 3 – Análise estatística descritiva dos estudos (fonte: SPSS v 29.0.2.0)	38
Tabela 4 – IC95%, Erro padrão, Pesos e Variâncias da métrica AUC (fonte: SPSS v29.0.2.0)	41
Tabela 5 – Valores da heterogeneidade da métrica AUC	42
Tabela 6 – Correlação de Pearson entre sensibilidade e especificidade (fonte: SPSS v29.0.2.0)	46
Tabela 7 – IC95%, Erro padrão, Pesos e Variâncias da métrica sensibilidade (fonte: SPSS v29.0.2.0)	48
Tabela 8 – IC95%, Erro padrão, Pesos e Variâncias da métrica especificidade (fonte: SPSS v29.0.2.0)	49
Tabela 9 – Valores da heterogeneidade da métrica sensibilidade.....	51
Tabela 10 – Valores da heterogeneidade da métrica especificidade	53
Tabela 11 – Valores do Índice de Kappa dos estudos incluídos.....	55
Tabela 12 – Valores de acurácia dos estudos incluídos	57

1. Introdução

O cancro da próstata é a neoplasia maligna mais comum entre os homens e a segunda principal causa de morte por cancro neste grupo populacional (Bray et al., 2018). O diagnóstico é fundamental para aumentar as taxas de sobrevivência e melhorar a qualidade de vida dos doentes. No entanto, os métodos tradicionais de diagnóstico, como o exame do antigénio específico da próstata (PSA) e a biópsia trans retal guiada por ecografia, apresentam limitações em termos de sensibilidade e especificidade, conduzindo a potenciais diagnósticos tardios ou a procedimentos invasivos desnecessários (Mottet et al., 2021).

A inteligência artificial (IA), em particular o ML e as redes neurais profundas, emergiu como uma ferramenta promissora na área médica, que oferecem soluções inovadoras para diagnósticos complexos (Esteva et al., 2019). No contexto do cancro da próstata, a IA tem sido aplicada na análise de imagens de ressonância magnética multiparamétrica (mpMRI), o que permite uma deteção mais precisa de lesões suspeitas e reduz a variabilidade do interobservador entre radiologistas (Harmon et al., 2019).

Estudos recentes demonstram que algoritmos de IA podem igualar ou até ultrapassar o desempenho de especialistas na deteção e classificação de tumores prostáticos (Schelb et al., 2019; Bulten et al., 2022). Por exemplo, a utilização de modelos de DL na análise de imagens histopatológicas tem demonstrado um potencial para melhorar a precisão na avaliação do grau de Gleason, sendo este um fator de prognóstico crucial (Campanella et al., 2019).

Apesar dos avanços significativos, a integração da IA na prática clínica enfrenta desafios, incluindo questões éticas e legais, e a necessidade de validação clínica robusta (Topol, 2019). Além disso, há preocupações relativas à interpretação dos modelos de IA e à fiabilidade dos resultados em diferentes populações e contextos clínicos (Nagendran et al., 2020).

Esta dissertação foi realizada no âmbito do Mestrado em Gestão das Organizações, pela Escola Superior de Saúde do Instituto Politécnico do Porto.

O projeto tem como objetivo explorar de forma abrangente a utilização da IA no diagnóstico do cancro da próstata. Pretende-se avaliar o estado atual da tecnologia, explorar os seus benefícios e discutir os desafios associados à sua implementação clínica. Através de uma revisão sistemática da literatura e análise estatística dos estudos existentes, este trabalho contribuirá para o entendimento mais aprofundado do papel da IA na melhoria dos cuidados de saúde nesta área.

2. Enquadramento Teórico

2.1. Inteligência Artificial

O termo IA contém uma referência explícita ao conceito de inteligência, sendo este um conceito vago tanto nas máquinas como nos seres humanos. Os investigadores de IA utilizam sobretudo o conceito de racionalidade, isto é, a capacidade de escolher a melhor ação a adotar para atingir um determinado objetivo, tendo em conta certos critérios que se devem otimizar e os recursos disponíveis.

Para Russel e Norvig (2020), a IA abrange diversas áreas, incluindo raciocínio lógico, perceção e interação com o ambiente, realçando o facto de a IA não só imitar o pensamento humano como também ter a capacidade de criar agentes que realizam ações eficazes em diversas circunstâncias.

Em 1950, Alan Turing, o criador do modelo matemático em que assentam todos os dispositivos computacionais atuais, publicou um artigo que partia da seguinte questão: “podem as máquinas pensar?”. Para poder responder a esta questão, seria necessário definir os conceitos “máquina” e “pensar”, o que não é fácil. Turing, foi uma das primeiras pessoas a colocar a possibilidade de as máquinas poderem aprender, algo que até então era visto como possível apenas aos seres humanos e a alguns animais (Turing, 1950).

Segundo Vinagre e Moniz (2020), a definição de IA é complexa, devido à panóplia de interpretações que o conceito pode assumir. IA refere-se a sistemas que simulam aspetos do comportamento humano inteligente, como learning e tomada de decisões, mas o termo “inteligência” é frequentemente usado de forma imprecisa, especialmente pelos media.

A falta de transparência, leva muitas vezes a equívocos quanto ao verdadeiro propósito da IA, que na realidade não é substituir a capacidade humana, mas sim otimizá-la.

Sichman (2021) define IA como um ramo da ciência/engenharia de computação, que visa desenvolver sistemas computacionais capazes de realizar tarefas que, normalmente, exigem inteligência humana, como o reconhecimento de padrões, tomadas de decisão e learning. A IA procura criar modelos e algoritmos que fazem com que as máquinas executem tarefas complexas de forma autónoma, ajustando-se e aprimorando-se com a experiência.

Existem três capacidades principais dos sistemas de IA: perceção (utilização de sensores para captar dados do ambiente, sejam estruturados ou não estruturados como bases de dados ou

imagens e texto), raciocínio e tomada de decisão (análise dos dados recolhidos para formular ações incluindo técnicas de aprendizagem automática, redes neurais e algoritmos de otimização) e atuação (execução de ações através de atuadores como comandos para robôs ou geração de respostas em sistemas de conversação) (GPAN IA, 2019).

No campo da IA, existe uma subárea chamada de Machine Learning (ML), que desenvolve algoritmos que permitem que os sistemas “aprendam” a partir de dados e melhorem as suas respostas com o tempo. Ao invés de programar sistemas para cada tarefa específica, os sistemas de ML aprendem a criar regras e padrões com base em conjuntos de dados, o que permite realizar tarefas como classificações ou previsões (Russel e Norvig, 2020).

No campo do ML, há uma área designada de Deep Learning (DL) que utiliza redes neurais artificiais profundas, inspiradas no cérebro humano, para realizar análises complexas e extrair padrões a partir de grandes volumes de dados. Estas redes neurais profundas são especialmente úteis para o reconhecimento de imagens e o processamento de linguagem natural, onde os dados são extensos e complexos. (Russel e Norvig, 2020).

Atualmente, tanto os produtores como os fornecedores concordam, cada vez mais, na necessidade de regular a IA. A discussão concentra-se em como deveria ser regulamentado o uso desta tecnologia.

Os próprios cidadãos estão a tornar-se cada vez mais conscientes e a aceitar a importância de regulamentar a IA. Esta necessidade levou à recente aprovação do Artificial Intelligence Act pela Comissão Europeia e pelo Parlamento Europeu. Este ato tem como objetivo abordar as questões levantadas pela utilização de sistemas de IA, adotando uma abordagem baseada no risco, isto significa, encontrar um equilíbrio entre incentivar o desenvolvimento tecnológico e garantir a segurança necessária, bem como estabelecer diretrizes para o uso da IA no setor de saúde (Cansado, 2024).

2.1.1. Aplicação de IA na saúde

A IA está a provocar mudanças significativas no setor da saúde, especialmente ao possibilitar diagnósticos mais rápidos e precisos. De acordo com Hamet e Tremblay (2017), a IA na medicina divide-se em duas vertentes principais: a vertente virtual, que inclui algoritmos de ML para a gestão de dados e apoio nas tomadas de decisão, e a vertente física, que envolve o uso de robôs para auxiliar em cirurgias e na monitorização de pacientes. Adicionalmente, Obasa e Palk (2023) sublinham a importância da IA na medicina de precisão, onde a tecnologia é

empregue para identificar padrões em vastas quantidades de dados, contribuindo para diagnósticos e tratamentos personalizados.

O uso da IA no diagnóstico e na medicina de precisão permite que os sistemas de saúde antecipem a evolução das doenças e recomendem tratamentos mais eficazes. Por exemplo, na área de oncologia, a IA tem sido uma grande aliada, dado que os algoritmos de DL são usados para identificar lesões em exames de imagem, conforme mencionam Obasa e Palk (2023). O recurso a sistemas como redes neurais profundas, que processam dados complexos, melhora a acurácia e reduz o erro humano em diagnósticos visuais.

Apesar dos benefícios prometidos, ambos os autores alertam para desafios éticos e de segurança associados ao uso de IA na saúde. Hamet e Tremblay (2017) destacam que a privacidade dos dados é uma das principais preocupações éticas nesta área. Com o aumento do uso de big data, onde enormes volumes de informações médicas e pessoais são processados para análise e diagnóstico, há um risco significativo de violações de privacidade. Este risco é ainda maior devido à natureza sensível dos dados de saúde que podem revelar detalhes críticos sobre a condição de cada paciente. Portanto, a segurança dos dados deve ser uma prioridade para proteger os pacientes de usos indevidos ou invasivos das suas informações pessoais.

Outro desafio ético importante é o potencial viés nos algoritmos de IA, o que pode levar a discriminações e desigualdades no acesso aos cuidados médicos. Obasa e Palk (2023) explicam que, se os dados utilizados para treinar um sistema de IA não representarem adequadamente todas as populações, o algoritmo pode desenvolver preconceitos que desfavorecem certos grupos. Por exemplo, se um algoritmo de diagnóstico é treinado principalmente com dados de pacientes de uma determinada etnia ou região, ele pode não ser tão eficaz para pacientes de outras origens. Este viés algorítmico pode perpetuar ou até agravar desigualdades preexistentes na saúde.

A transparência e a capacidade de explicar o funcionamento dos algoritmos de IA são outra preocupação importante. Hamlet e Tremblay (2017) observam que muitos modelos avançados, especialmente redes neurais profundas, funcionam como “caixas-pretas”, onde o processo de tomada de decisão é opaco. Isso significa que, em muitos casos, nem os próprios desenvolvedores entendem completamente como o sistema chegou a uma determinada conclusão. Esta falta de transparência pode ser problemática na área da saúde, onde as decisões têm um impacto direto na vida dos pacientes.

Outra questão ética importante é o consentimento informado. Quando a IA é utilizada para analisar dados de saúde, é essencial que os pacientes estejam cientes de como os seus dados serão utilizados e que autorizem esse uso de forma consciente (Obasa e Palk, 2023). O processo de consentimento informado deve ser adaptado para garantir que os pacientes compreendam plenamente as implicações do uso de IA nos seus cuidados médicos. Sem esta transparência, a confiança dos pacientes na tecnologia e nos profissionais de saúde pode ser comprometida.

Hamet e Tremblay (2017) também exploram como a IA pode afetar a autonomia dos profissionais de saúde. Com a introdução de sistemas automatizados de apoio à decisão clínica, existe o risco dos médicos se tornarem excessivamente dependentes da tecnologia, limitando a sua própria capacidade de julgamento. Este é um dilema ético pois a tecnologia deve servir como um complemento à experiência e conhecimento médico e não um substituto, para a experiência e o conhecimento médico. O uso responsável da IA requer que os sistemas respeitem a autonomia dos profissionais e que sejam utilizados para complementar, e não substituir, as decisões humanas.

Obasa e Palk (2023) defendem que é necessário implementar políticas rigorosas para assegurar que a IA seja utilizada de forma justa e transparente, respeitando os direitos dos pacientes e minimizando os riscos. Sugerem que as regulamentações sejam desenvolvidas como uma abordagem colaborativa, envolvendo profissionais de saúde, desenvolvedores de IA e legisladores, para garantir que as normas reflitam os interesses de todas as partes envolvidas.

2.2.2.1. Aplicação de IA na saúde em Portugal

Portugal tem demonstrado interesse crescente na adoção de IA na área da saúde, em parte impulsionado pela necessidade de modernizar o Serviço Nacional de Saúde (SNS) e melhorar a eficiência e qualidade dos serviços de saúde. Segundo o relatório da Comissão Europeia, a IA aplicada à saúde no país foca-se em projetos de pesquisa e inovação, sendo, no entanto, ainda limitado o uso dessa tecnologia de forma generalizada nos hospitais e clínicas nacionais.

A IA tem o potencial de revolucionar o setor da saúde, oferecendo melhorias substanciais na qualidade dos serviços prestados. Segundo Fernandes (2024), a IA pode acelerar o desenvolvimento de novos tratamentos, facilitando o diagnóstico precoce de doenças e otimizar o uso de recursos em hospitais, reduzindo tanto o tempo de espera como os custos associados aos tratamentos.

De acordo com um estudo da Accenture (2023), intitulado de “Reinvent care delivery to solve clinical shortage”, cerca de 40% das horas de trabalho nos cuidados de saúde poderiam ser apoiadas por tecnologias de IA. Este suporte é possível através de aplicações que monitorizam o estado de saúde dos pacientes, recomendam tratamentos e ajudam nos diagnósticos.

A IA também promove a chamada “medicina P4” (preditiva, preventiva, personalizada e participativa), onde os cuidados são adaptados às características únicas de cada paciente, melhorando a sua experiência (Fernandes, 2024).

As potencialidades da IA poderiam ser expandidas de forma a permitir o acesso dos pacientes aos seus dados de saúde, de uma forma mais célere e significativa, permitindo um seguimento mais regular e um maior controlo e conhecimento do seu bem-estar (Machado, 2019; Fernandes, 2024).

Embora sejam reconhecidos os potenciais da IA, os profissionais enfatizam a necessidade de um investimento adequado e de um quadro regulatório que assegure a proteção de dados e a transparência nos processos automatizados de tomada de decisão. A inclusão de IA nos sistemas de saúde exige não só uma infraestrutura tecnológica robusta, mas também formação para os profissionais de saúde (APAH, 2022).

Em suma, a IA e a digitalização são partes essenciais para a evolução do sistema de saúde em Portugal. Há uma forte procura para que os programas eleitorais incluam compromissos claros sobre a implementação destas tecnologias, com o objetivo de melhorar a eficiência, personalização e acessibilidade dos cuidados de saúde.

2.1.2. Implicações éticas e legais da IA

O crescente uso da IA em diversos setores como a saúde, a justiça e a educação, trouxe inúmeros benefícios, mas também apresentou preocupações éticas e legais significativas.

O viés algorítmico é uma das principais preocupações éticas identificadas. De acordo com Lopes e Mendes (2023), os algoritmos de ML, ao serem treinados com dados oblíquos, perpetuam preconceitos existentes o que pode resultar em decisões injustas e discriminatórias.

Por exemplo, a Amazon teve de interromper o uso de um sistema de recrutamento por IA porque o programa aprendeu a preferir candidatos masculinos com base em dados históricos (Dastin, 2018).

A falta de transparência dos algoritmos de IA, frequentemente referida como “caixa-preta”, dificulta a compreensão de como decisões são tomadas. Winfield e Jirotko (2018) alertaram que a opacidade pode levar a uma falta de confiança nos sistemas de IA, especialmente nas áreas da saúde e justiça.

Já a Agência dos Direitos Fundamentais da União Europeia destaca que a falta de esclarecimento nos sistemas de IA pode resultar em decisões que não podem ser contestadas ou compreendidas, afetando os direitos fundamentais dos indivíduos (FRA, 2021).

Outra das implicações éticas, refere-se à privacidade e proteção de dados. Os sistemas de IA dependem de grandes quantidades de dados pessoais. Segundo o relatório da FRA (2021), a utilização massiva de dados pessoais por sistemas de IA aumenta o risco de violações de privacidade, o que exige uma regulação mais robusta para proteger os direitos dos cidadãos.

A nível das implicações legais da utilização de IA, refere-se a responsabilidade e regulamentação. Segundo HFW (2025), os desafios legais residem na identificação de quem é o responsável pelas decisões automatizadas quando estas causam danos, seja o programador, o fornecedor de dados ou a empresa que utiliza o sistema. A criação de um quadro regulatório robusto na UE é essencial para garantir que os sistemas de IA estejam alinhados com os direitos fundamentais, abordando questões de responsabilidade e supervisão regulatória (FRA, 2021).

A UE desenvolveu diretrizes éticas para a utilização de IA, conhecidas como “Ethics Guidelines for Trustworthy AI”. Estas diretrizes incluem três requisitos fundamentais para uma IA fiável:

- Legalidade: o sistema deve estar em conformidade com todas as leis e regulamentações aplicáveis;
- Ética: o sistema deve ser desenvolvido e utilizado de forma a respeitar os princípios éticos como a justiça, transparência e privacidade;
- Robustez: o sistema deve ser seguro, tanto do ponto de vista técnico, como social, minimizando riscos de erros e impactos negativos inesperados.

Estas diretrizes estabelecem princípios como a transparência que requer a explicabilidade dos sistemas de IA para que as suas decisões possam ser compreendidas e contestadas, e a justiça, que visa garantir que os algoritmos não reforcem preconceitos ou discriminações existentes. Além disso, as diretrizes enfatizam a importância da supervisão humana, especialmente em contextos de alto risco, para evitar tomadas de decisão automatizadas que possam ter consequências adversas para os indivíduos e a sociedade (GPAN, 2019).

A SPMS enfatiza que a implementação do RGPD, que se encontra em vigor na UE desde maio de 2018, é crucial para a proteção de dados pessoais dos utentes e para garantir a conformidade das entidades do setor de saúde. O conhecimento das regras que serão aplicáveis ao tratamento dos dados pessoais torna-se imprescindível e as entidades devem adaptar os seus processos e metodologias para estarem em conformidade. (SPMS, 2021)

Concluindo, enquanto a IA promete inovações significativas na saúde, a sua implementação deve ser acompanhada por rigorosos mecanismos de proteção de dados. O cumprimento do RGPD é fundamental para garantir que a aplicação de IA não comprometa a segurança e privacidade dos dados dos pacientes. A regulação deve equilibrar a inovação tecnológica com a proteção dos direitos dos cidadãos, promovendo uma utilização ética e legal da IA no setor da saúde (Cansado, 2024).

2.2. Cancro da Próstata

Segundo o SNS (2023), "cancro ou tumor maligno é um termo médico utilizado para denominar um vasto conjunto de doenças caracterizadas por um crescimento anormal e descontrolado das células e que, na maioria das vezes, formam uma massa chamada tumor. No entanto, alguns cancros, como os que ocorrem no sangue, não formam massas tumorais."

A próstata é uma glândula acessória, constituída por tecidos glandular e muscular e tem a forma e aproximadamente a mesma dimensão que uma noz, com cerca de 4cm de comprimento e 2cm de largura. A glândula prostática localiza-se na face posterior da sínfise púbica, na base da bexiga, onde envolve a uretra prostática e os dois canais ejaculadores. A glândula prostática pode aumentar de tamanho por várias razões, entre as quais as infeções, os tumores e o envelhecimento (Seeley et al., 2007).

A IARC (2024), estimou que em 2022, o cancro da próstata afetou 1467854 homens em todo o mundo, sendo a segunda maior causa de incidência a seguir ao cancro do pulmão e a quinta maior causa de morte (397430 homens).

A Europa representa menos de 10% da população mundial, mas representa 25% de todos os casos de cancro (GCO, 2020).

2.2.1. Custos do cancro da próstata na União Europeia e em Portugal

De acordo com o a OCDE, em 2018, os custos do cancro na UE foram de 326 euros per capita. Estima-se que os custos do cancro na UE foram de 167,4 mil milhões de euros (Fórmula: Gasto total = População total x gasto per capita).

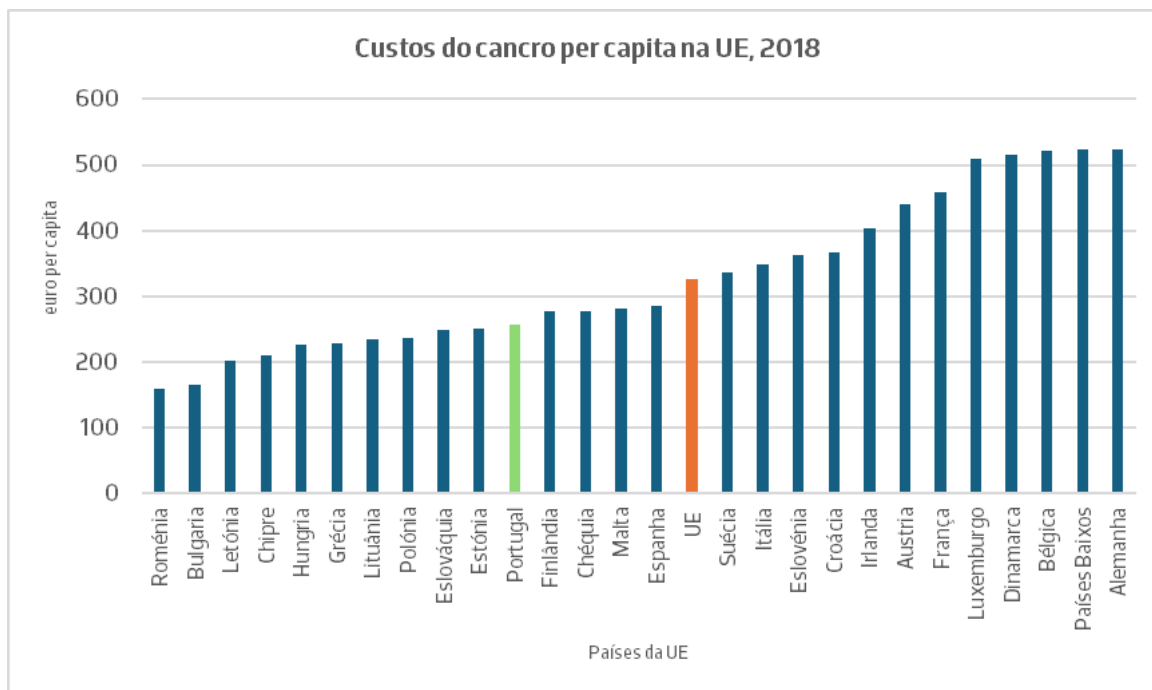


Figura 1 – Custo estimado do cancro per capita na UE, relativos a 2018

Os custos do cancro na UE variam entre 160 euros per capita, correspondendo a Roménia, e 524 euros per capita, correspondendo aos Países Baixos e Alemanha (figura 1) (OCDE, 2023). Portugal encontra-se em 11º lugar nos países com menos custos no cancro per capita, tendo um gasto de 256 euros, sendo que a média da UE é de 326 euros per capita (OCDE, 2023).

O cancro é a segunda maior causa de morte na União Europeia, após as doenças cardiovasculares. Segundo a Joint Research Centre (2020), a incidência e mortalidade do cancro é maior nos homens do que nas mulheres. Em todos os países europeus, cerca de 706000 homens e 555000 mulheres morreram com um cancro em 2020.

2.2.2. Epidemiologia do cancro da próstata na União Europeia e em Portugal

Segundo Last (2008) epidemiologia é definida como “o estudo da distribuição e dos determinantes de estados ou eventos relacionados à saúde em populações específicas, e a sua aplicação na prevenção e controlo dos problemas de saúde”.

Incidência, prevalência e mortalidade são três métricas usadas em epidemiologia, que contribuem para a compreensão da distribuição e dos determinantes das doenças nas populações.

De acordo com Bonita et al (2006), incidência refere-se ao número de novos casos de uma doença que ocorrem numa determinada população durante um determinado período de tempo; prevalência refere-se ao número de casos (novos e antigos) de uma doença numa população num determinado momento (reflete a carga da doença na população); mortalidade diz respeito ao número de óbitos numa determinada população devido a uma doença específica durante um período específico (é um indicador da gravidade da doença e da eficácia dos cuidados de saúde).

2.2.2.2. Incidência

Na Europa, o número de casos de cancro tem aumentado ao longo dos últimos anos, sendo que em 2022, a incidência foi de 2,7 milhões de pessoas onde 1,3 milhões resultaram em morte. (Comissão Europeia, 2024)

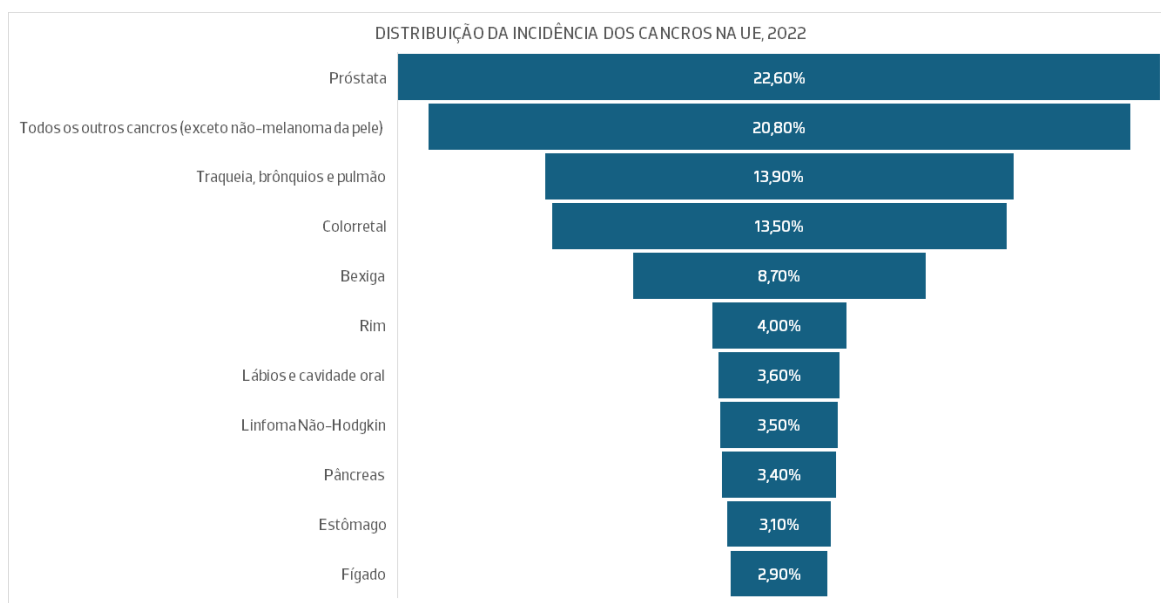


Figura 2 – Percentagem da incidência dos cancros na UE, relativos a 2022

Segundo o ECIS, em 2022, o número de novos casos de cancro, nos homens, na UE foi de 1 465 846, sendo que 337 145 (23%) casos eram relativos ao cancro da próstata (ECIS, 2022).

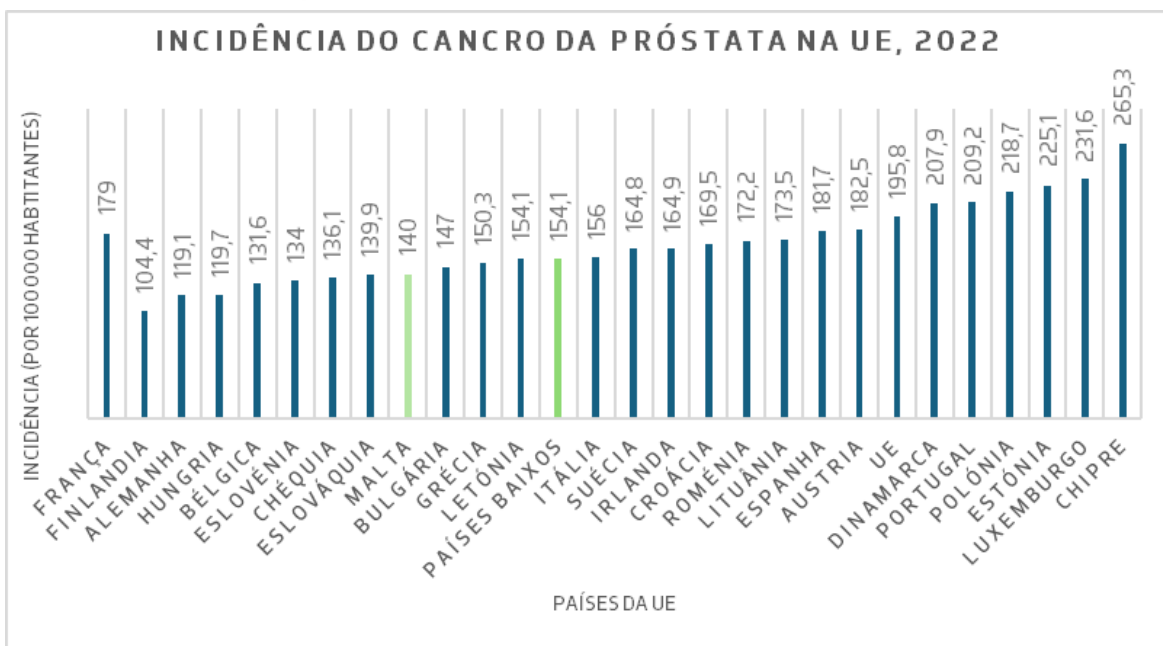


Figura 3 - Incidência do cancro da próstata nos países da UE, relativos a 2022

Observa-se uma variação significativa na incidência entre os países pertencentes a UE. A média na UE situa-se em torno de 154 casos por 100 000 habitantes.

No norte da Europa, a incidência do cancro da próstata é superior a qualquer outro, sendo a Lituânia o país com um maior número de casos (265 por 100000 habitantes), seguido da Suécia com 232 casos por 100000 habitantes. Já o país com menor número de incidência do cancro da próstata é a Bulgária com 104 casos por 100000 habitantes. (ECIS, 2022).

Na UE, o cancro da próstata afeta 18,7% dos homens com idade compreendida entre 45 a 65 anos e 25,7% em homens com idade igual ou superior a 65 anos (ECIS, 2022).

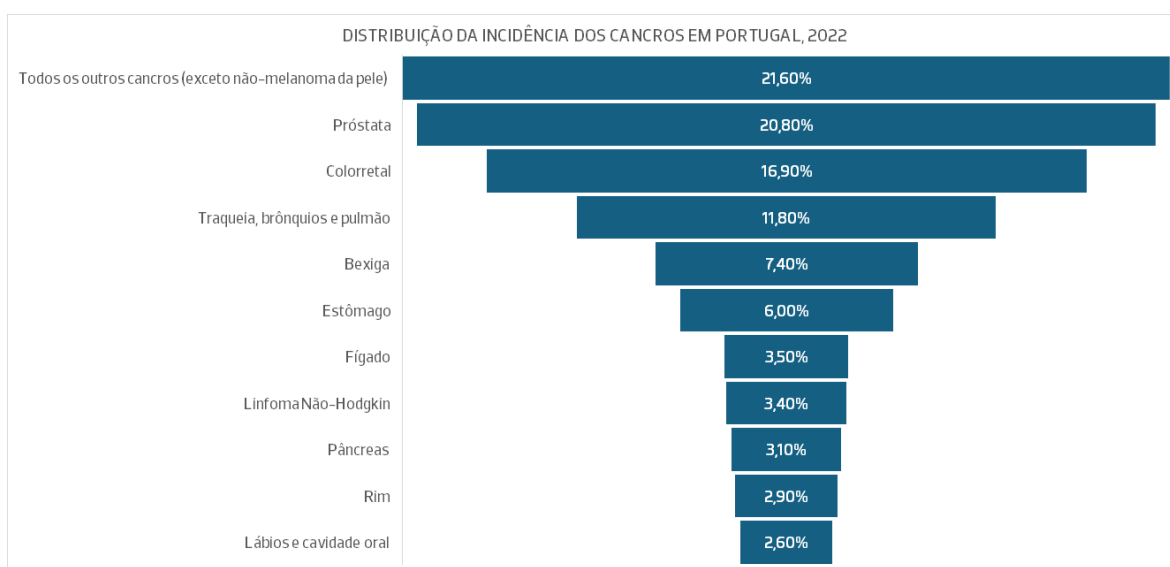


Figura 4 - Percentagem da incidência dos cancros em Portugal, relativo a 2022

Em 2022, houve cerca de 36152 novos casos de cancro em Portugal, sendo que 7529 correspondem a casos de cancro da próstata.

De acordo com os dados do ECIS (2022), em Portugal, o cancro da próstata foi o cancro com maior incidência, representando 21% do total de novos casos.

A estimativa de incidência do cancro da próstata em Portugal foi de cerca de 18,9% em homens com a idade compreendida entre 45 e 64 anos, e de 23% em homens com idade igual ou superior a 65 anos (ECIS, 2022).

2.2.2.3. Mortalidade

Segundo o ECIS, em 2022, o número de óbitos provocados por cancro, nos homens, na UE foi de 717274, tendo sido registados 76772 óbitos por cancro da próstata (ECIS,2022).

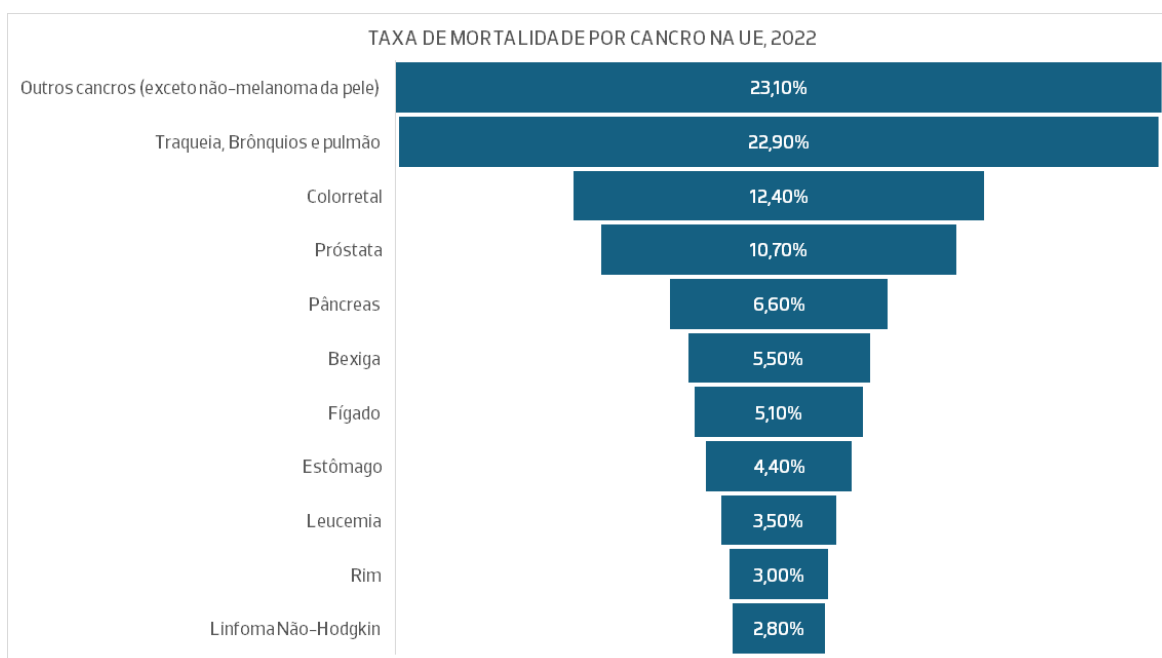


Figura 5 - Taxa de mortalidade por cancro na UE, relativo a 2022

O cancro da próstata apresenta uma percentagem de 10,7% de taxa de mortalidade na UE, ocupando o terceiro lugar, antecedendo o cancro da traqueia, brônquios e pulmão com 22,9% e o cancro colorretal com 12,4% (ECIS, 2022).

Na figura 6, encontra-se representado um gráfico relativo a taxa de mortalidade por cancro da próstata na UE no ano de 2022.

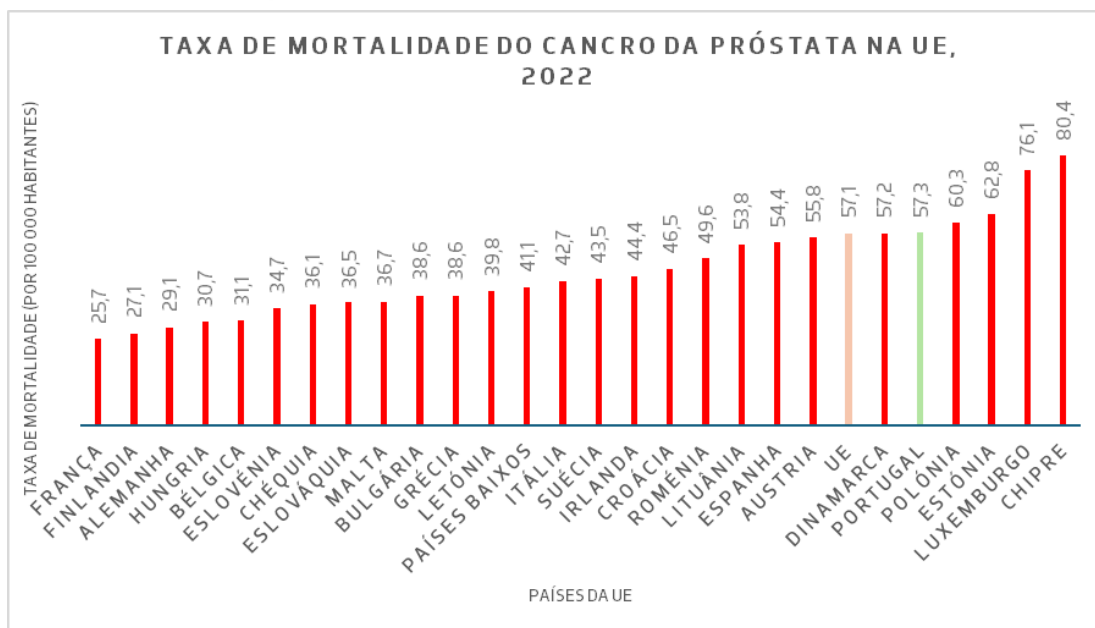


Figura 6 - Taxa de mortalidade por cancro da próstata na UE, relativo a 2022

Observa-se uma variação na taxa de mortalidade entre os países pertencentes à UE. A média na UE é de 57,1% equivalendo acerca de 97 óbitos por 100000 habitantes.

O Chipre é o país que apresenta a taxa de mortalidade do cancro da próstata mais elevada com uma taxa de 80,4%, o que equivale a cerca de 176 óbitos por 100000 habitantes, seguido do Luxemburgo com uma taxa de mortalidade de 76,1% (171 óbitos por 100000 habitantes). Já o país com menor taxa de mortalidade provocada por cancro da próstata é França com 25,7% (31 óbitos por 100000 habitantes) (ECIS, 2022).

Em Portugal, a taxa de mortalidade é ligeiramente superior à média da UE, fixando-se nos 57,3%, o que corresponde a cerca de 99 óbitos por 100000 habitantes (ECIS, 2022).

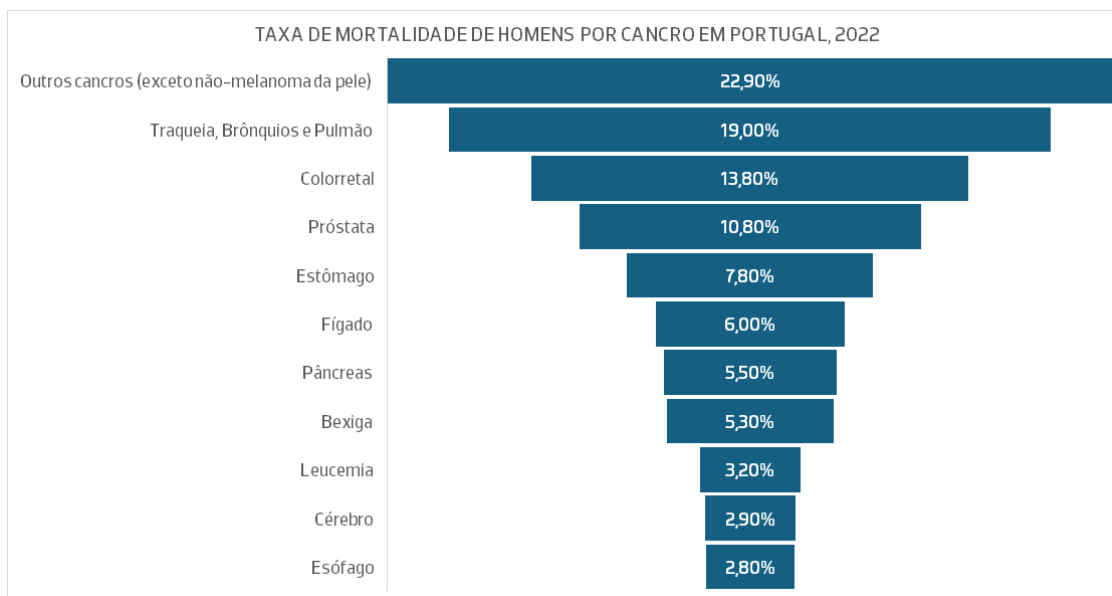


Figura 7 – Taxa de mortalidade por cancro em Portugal, relativo a 2022

Relativamente a taxa de mortalidade em homens provocada por cancro, em Portugal, o cancro da próstata apresenta uma percentagem de 10,8% de taxa de mortalidade ocupando o terceiro lugar, antecedendo o cancro da traqueia, brônquios e pulmão com 19% e o cancro colorretal com 13,8% (ECIS,2022).

No ano de 2022, houve cerca de 19382 óbitos em Portugal, provocados por cancro, sendo que 2083 provocados por cancro da próstata.

A estimativa da taxa de mortalidade por cancro da próstata em Portugal foi de 13,3% em homens com idade igual ou superior a 65 anos (ECIS, 2022).

2.2.2.4. Prevalência

Conforme apresentado na figura 8, a prevalência do cancro da próstata varia entre 78 e 238 homens por cada 100000 habitantes.

Os países como a Lituânia e Suécia apresentam valores mais altos, enquanto a Bulgária e a Roménia apresentam valores mais baixos.

Nos países do norte da Europa, a maior prevalência pode estar relacionada com práticas de rastreio mais eficazes e maior acesso a diagnósticos precoces.

Altas taxas de prevalência podem ser influenciadas por uma maior sobrevivência dos pacientes, indicando acesso a tratamentos eficazes.

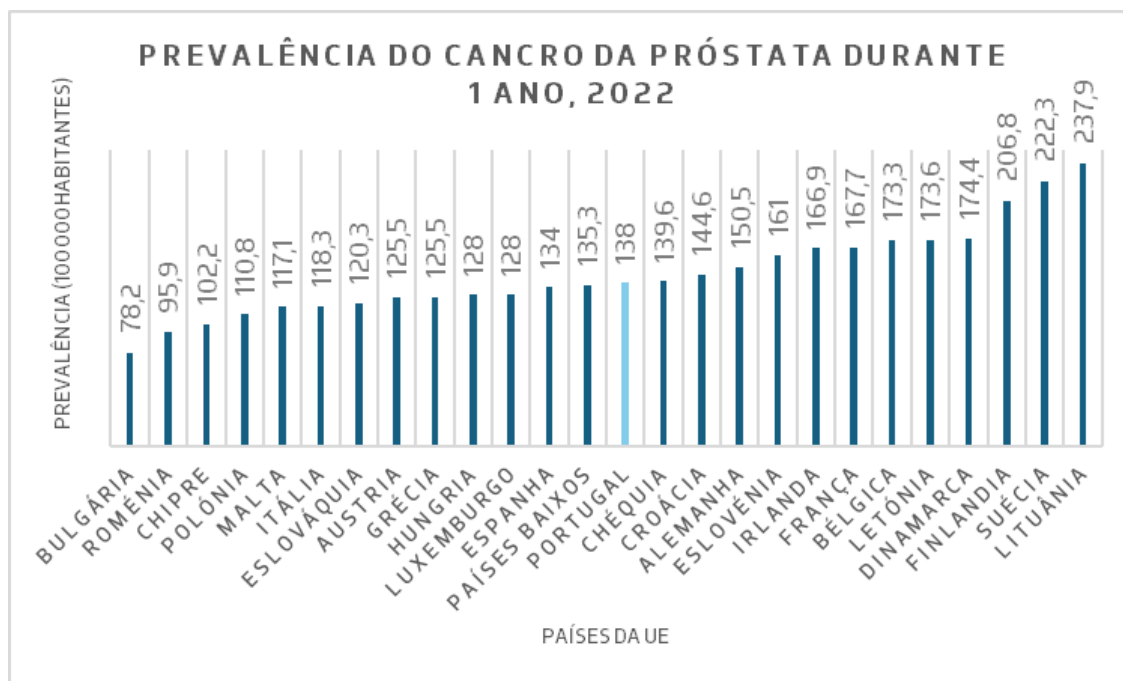


Figura 8 – Prevalência do cancro da próstata nos países da UE durante 1 ano, relativo a 2022

2.2.3. Etiologia

Etiologia é o estudo das causas e fatores de uma doença. No caso do cancro da próstata, a etiologia é multifatorial e resulta da interação entre os fatores genéticos, hormonais, ambientais e de estilo de vida.

O cancro da próstata, normalmente surge após os 40 anos sendo que a sua incidência aumenta progressivamente a partir dos 50anos (Silva, 2019).

2.2.4. Fatores de Risco

2.2.4.1. Idade

De acordo com Grönberg et al. (2015), a idade é o principal fator de risco do cancro da próstata. Segundo a IARC (2024), em 2022, o número total de cancros da próstata no mundo equivalia a 1 467854 homens, sendo que a maior incidência é na Europa com 473011 casos.

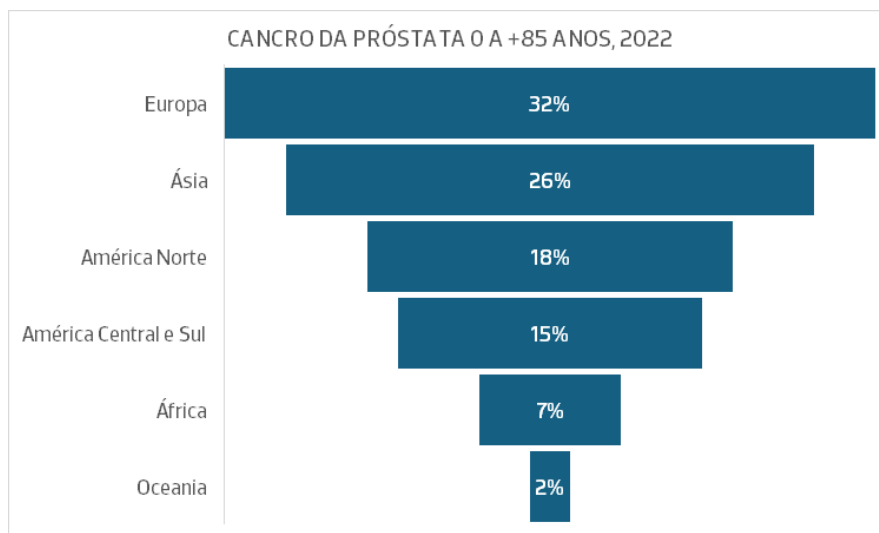


Figura 9 – Número de casos de cancro da próstata no mundo dos 0 aos +85 anos, relativos a 2022

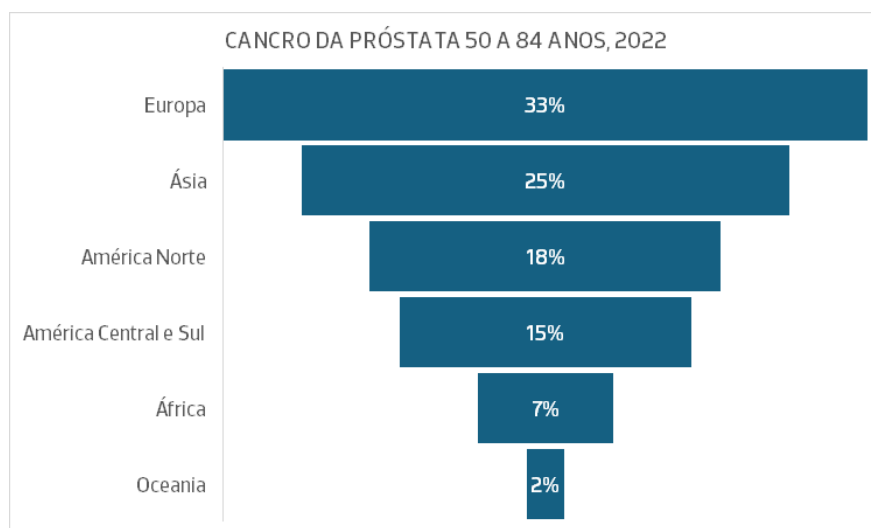


Figura 10 – Número de casos de cancro da próstata no mundo dos 50 aos 84 anos, relativos a 2022

Em 2022, o número de casos de cancro da próstata incide-se maioritariamente em homens com mais de 49 anos, sendo que a Europa lidera com um total de 441646 casos, num total de 1350245 casos em todo o mundo (IARC, 2024).

Com o envelhecimento, as alterações hormonais masculinas sofrem mudanças, ou seja, a produção de testosterona começa a diminuir o que leva ao aumento de DHT, e consequentemente ao risco de cancro (Silva, 2019).

2.2.4.2. Antecedentes Familiares

De acordo com Hemminki (2012), homens com um pai diagnosticado com cancro da próstata, apresentam um risco 2 vezes maior de desenvolver a doença, sendo que este risco aumenta significativamente consoante o número de familiares diagnosticados.

2.2.4.3. Fatores Genéticos

Existem quatro genes identificados como responsáveis pelo cancro: os genes supressores de tumores, os proto-oncogenes, os genes reparadores do ADN e os genes reguladores dos processos de apoptose. Os genes supressores de tumores, são genes responsáveis pela codificação de proteínas reguladoras de mecanismo de inibição tumoral (Stephens e Aigner, 2009). Os proto-oncogenes são genes responsáveis pela proliferação celular (Stephens e Aigner, 2009). Os genes reparadores do ADN são genes que em caso de mutação, aceleram o processo de acumulação dos proto-oncogenes resultando no desenvolvimento de tumores (Ciccia e Elledge, 2015). Os genes reguladores dos processos de apoptose regulam a morte celular programada de forma a prevenir a progressão do cancro.

Segundo Sekhoacha et al. (2022), a predisposição genética é um fator crítico na etiologia do cancro da próstata, ou seja, homens com familiares de primeiro grau diagnosticados têm maior probabilidade de desenvolver a doença.

As mutações hereditárias nos genes BRCA, HOX, ATM, RNase L, MSR1 e ELAC2, podem aumentar o risco de cancro da próstata (Sekhoacha et al., 2022).

2.2.4.4. Etnia

De acordo com a American Cancer Society (2020), o cancro da próstata ocorre com maior frequência em homens de raça negra e caribenhos do que em homens de outra raça, no entanto não existem razões específicas para essas diferenças raciais e étnicas.

2.2.4.5. Estilo de Vida

A nível do estilo de vida, existem alguns fatores que tem impacto no desenvolvimento do cancro da próstata, tais como: a alimentação, a obesidade e a falta de exercício físico (American Cancer Society, 2020).

Ainda não é claro o papel exato da dieta no desenvolvimento do cancro da próstata, no entanto, alguns estudos sugerem que o consumo excessivo de laticínios pode influenciar ligeiramente no aparecimento do cancro da próstata, assim como a ingestão de grandes quantidades de cálcio, que apesar de ser benéfico para a saúde, pode causar um desenvolvimento do cancro quando consumido em demasia (American Cancer Society, 2020).

Relativamente ao IMC, se estiver elevado (obesidade), os homens têm uma maior probabilidade de desenvolver cancro da próstata, com um crescimento mais rápido e agressivo (Uehara et al., 2018). A falta de exercício aumenta o risco de doença mesmo nos indivíduos normoponderais (Vidal et al., 2014).

2.2.5. Detecção e Diagnóstico

Num mundo ideal, o cancro seria totalmente prevenido e para que isso fosse possível, seria importante a deteção precoce de todos os cancros, aumentando assim a taxa de sobrevivência.

Em 2021, foi apresentado um Plano Europeu de Luta contra o Cancro, onde foram delineadas estratégias abrangentes para a prevenção, deteção, tratamento e qualidade de vida dos pacientes com cancro na UE (Comissão Europeia, 2021).

Nesse mesmo plano, estão contemplados os seguintes tópicos: garantir que menos de 5% da população use tabaco até 2040, reduzir o consumo nocivo de álcool e reduzir a exposição dos jovens à comercialização de bebidas alcoólicas em pelo menos 10% até 2025, reduzir a poluição ambiental de forma a diminuir a exposição a substâncias cancerígenas e radiações e promover a literacia em saúde (Comissão Europeia, 2021).

Outra das iniciativas seria a deteção precoce, no entanto, existem desigualdades de acesso entre os Estados-Membros. Todos os cidadãos, devem ter o mesmo direito aos cuidados, diagnóstico e tratamento de alta qualidade e o mesmo acesso aos medicamentos contra o cancro, independentemente da localização da sua habitação (Comissão Europeia, 2021).

De acordo com a National Cancer Institute (2019), os homens podem fazer um rastreio para detetar o cancro da próstata, antes de desenvolverem quaisquer sintomas. Estes sintomas podem ser: problemas urinários, incapacidade de urinar ou dificuldade em iniciar ou parar o fluxo de urina, necessidade frequente de urinar principalmente à noite, dor ou ardor durante a micção, dificuldade em ter uma ereção, sangue na urina ou no sémen, dor frequente na zona inferior das costas, nas ancas ou na zona superior das coxas. Apesar de tudo, estes sintomas podem ser provocados por tumores benignos ou outros problemas.

Os exames de diagnóstico permitem identificar a presença do tumor antes do aparecimento de sintomas, avaliar o grau de progressão e caracterização do tumor. Existem dois exames essenciais ao diagnóstico: o toque retal e as análises ao sangue para determinar o PSA. Só mais tarde e em caso de necessidade de métodos de diagnóstico complementares, são realizados uma ecografia transretal da próstata, uma RMN e uma TC (European Association of Urology, 2024).

2.2.5.1. Exame do Toque Retal

O toque retal é um exame de primeira linha no diagnóstico e despiste deste tumor, ou de outra patologia que afeta a próstata tal como a HBP. Este exame é realizado de forma a identificar precocemente a presença do cancro para que o tratamento seja mais eficaz e potencialmente

curativo. No entanto, só poderá ser feito consoante a vontade e motivação do paciente. Mesmo após uma intervenção cirúrgica ou outro tipo de tratamento associado a este cancro, mantém-se sempre como um exame de eleição para verificar a normalidade e o sucesso terapêutico (European Association of Urology, 2024).

O cancro da próstata é detetado graças ao exame do toque retal, independentemente dos níveis de PSA, em 18% dos casos (European Association of Urology, 2024).

2.2.5.2. PSA

O PSA é uma glicoproteína com função de enzima, produzida pelas células epiteliais da próstata e cuja função consiste em tornar o esperma líquido após a formação do “coágulo” no esperma ejaculado. Tem um peso molecular de 34000 Daltons, produzida tanto por células prostáticas normais como em células alteradas (Silva, 2019).

Valores normais de PSA:

- Até aos 50 anos: até 2,5 ng/ml
- Entre os 50 e os 60 anos: até 3,5 ng/ml
- Entre os 60 e os 70 anos: até 4,5 ng/ml
- Acima de 70 anos: até 6,5 ng/ml (National Cancer Institute, 2019).

2.2.5.3. Outros Exames de Diagnóstico

A ecografia trans-retal, a cistoscopia e a biópsia são outros exames complementares para diagnosticar o cancro da próstata.

A biópsia consiste na remoção de tecido, para procurar células cancerígenas. É um procedimento realizado pelo médico, que pode ser guiado por uma ecografia. Após a recolha do tecido, o patologista analisa-o e deteta a presença ou não de cancro da próstata (National Cancer Institute, 2019).

O score de Gleason avalia a agressividade e o prognóstico do cancro da próstata, baseando-se no aspeto microscópico das células tumorais da próstata após a análise de uma amostra obtida por biópsia. Este score mede a diferença entre o tecido tumoral e o tecido prostático normal.

O objetivo é identificar os dois padrões celulares mais comuns no tumor. Cada padrão celular é classificado numa escala de 1 a 5 (1 – células semelhantes ao tecido prostático normal bem organizadas; 2 a 4 – células progressivamente menos organizadas; 5 – células completamente desorganizadas e sem estrutura) (Humphrey, 2004).

O cálculo do score de Gleason é igual a soma do padrão primário com o padrão secundário. Os scores variam entre 6 a 10. Score ≤ 6 – Tumor de baixo grau (cresce lentamente e a probabilidade de metastizar é baixa); Score 7 – Tumor de grau intermédio ; Score 8 a 10 – Tumor de alto grau (cresce rapidamente e a probabilidade de metastizar é alta) (Humphrey, 2004).

2.2.6. Estadiamento

Classificar ou estadiar um cancro é essencial para determinar o prognóstico (a sua provável evolução) e para escolher o tratamento mais adequado. O estadio baseia-se no estado do tumor primário (T), na disseminação do tumor para os gânglios linfáticos (N) e na sua metastização para outras partes do corpo (M) (Fundação Champalimaud, 2023).

O sistema internacional TNM varia entre TX a T4, NX a N4 e MX a M4.

Classificação do tumor Primário (T):

TX – não existe informação sobre o tumor primário;

T0 – não há sinal de tumor primário

T1 a T4 – descrições condensadas do tamanho e do nível de disseminação do tumor (quanto mais elevado o número, maior o tumor) (Fundação Champalimaud, 2023).

Classificação do envolvimento dos gânglios linfáticos (N):

NX – Não existe informação sobre os gânglios linfáticos próximos

N0 – Gânglios linfáticos próximos sem atingimento pelo cancro

N1 a N4 – descrição do tamanho, localização e número de gânglios linfáticos afetados pelo cancro (quanto maior o número da classe, maior a disseminação do cancro para os gânglios linfáticos próximos) (Fundação Champalimaud, 2023).

Classificação de presença de metástases (M):

MX – presença de metástases não pode ser avaliada;

M0 – O cancro não criou metástases;

M1 – O cancro criou metástases (Fundação Champalimaud, 2023).

Através de esta classificação pelo sistema internacional TNM, é possível obter um estadio do cancro que varia entre Estadio I e Estadio IV. O Estadio I significa que o cancro ainda não penetrou profundamente os tecidos próximos e não houve disseminação para os gânglios linfáticos ou outros locais no corpo; Estadio II e III significa que os cancros já cresceram mais profundamente dentro dos tecidos circundantes e podem ter alastrado para os gânglios

linfáticos; Estadio IV o cancro já invadiu outras partes do corpo (metastização). Existe ainda o Estadio 0 que significa cancro in situ, ou localizado (Fundação Champalimaud, 2023).

Para definir o estadio do tumor, é necessário realizar exames como a TAC (tomografia axial computadorizada), PET-PSMA (que avalia o metabolismo através de um marcador radioativo), PET-TC (que combina imagens metabólicas e imagens anatómicas) e RMN (ressonância magnética nuclear, que fornece imagens detalhadas através de campos magnéticos) (Instituto da Próstata, 2023).

2.2.7. Tratamento

De acordo com o Instituto da Próstata (2023), o tratamento do cancro da próstata depende do estadio em que o tumor se encontra. O tratamento pode ser a braquiterapia (colocação de implantes radioativos no interior da próstata), a cirurgia (remoção da glândula prostática e tecidos envolventes, por via laparoscópica, robótica ou cirurgia convencional), a radioterapia externa (aplicação externa de raios de alta energia em várias sessões), a crioterapia (inserção de sondas por via perineal que congelam o tecido tumoral), ou a hormonoterapia (bloqueio hormonal com fármacos).

2.2.8. Otimização no diagnóstico do cancro da próstata

A aplicação da IA no diagnóstico do cancro da próstata tem tido avanços significativos, destacando-se na área de imagiologia, reduzindo falsos positivos e melhorando a precisão diagnóstica.

Rodrigues et al. (2023), destacam que os modelos de segmentação baseados em DL, ao analisarem a ressonância magnética multiparamétrica (mpMRI), podem melhorar a precisão na deteção de cancro da próstata. Este estudo demonstra que a IA, especialmente quando combinada com técnicas avançadas de segmentação, como o modelo nnU-Net, atinge elevada precisão na segmentação glandular, com um Dice score médio de 0,9289 (métrica utilizada para avaliar a sobreposição entre duas amostras de dados, sendo amplamente usada nas tarefas de segmentação de imagens médicas), superior a outros modelos tradicionais.

Em 2022, segundo a International Agency for Research on Cancer, o cancro da próstata foi o segundo cancro mais frequente nos homens a nível mundial e o quinto em termos de mortalidade. O diagnóstico do cancro da próstata baseia-se em medidas específicas, como os níveis de PSA (antigénio específico da próstata) e o exame do toque retal, seguido de uma

biópsia, em que a avaliação da agressividade da doença baseia-se numa escala qualitativa (PIRADS – Prostate Imaging-Reporting and Data System), cuja interpretação pode variar consoante o radiologista, introduzindo subjetividade. A medição de PSA e o exame retal podem conduzir a um sobrediagnóstico e a um tratamento excessivo.

Um estudo internacional comparativo (PI-CAI) demonstrou que um sistema de IA pode atingir uma precisão comparável à de radiologistas experientes ao identificar o cancro da próstata clinicamente significativo através da ressonância magnética (MRI). Este sistema mostrou uma área sob a curva ROC (AUROC) de 0,91, superior aos radiologistas (AUROC = 0,86). Desta forma, a IA revelou-se mais eficaz em evitar diagnósticos excessivos e desnecessários, reduzindo em 50,4% os falsos positivos e em 20% a deteção de cancros indolentes em relação aos especialistas humanos (Saha et al, 2024).

A radiómica, uma técnica que extrai características quantitativas de imagens médicas, também tem sido utilizada para eliminar a subjetividade presente na interpretação das imagens. Segundo Rodrigues et al. (2023), ao transformar essas características em dados estruturados, os algoritmos de IA conseguem identificar padrões que ajudam a prever a agressividade do cancro, o que é crucial para a tomada de decisões clínicas mais personalizadas.

Além disso, a abordagem semi-supervisionada de IA (método que utiliza dados rotulados (informação correta) e dados não rotulados para treinar modelos de IA) introduzida por Bosma et al. (2023), permite reduzir o número de notas manuais necessárias, mantendo uma alta performance na deteção de cancros significativos da próstata, o que demonstra que a IA torna o diagnóstico mais eficiente e acessível.

Estes avanços destacam que a IA não só melhora a precisão e a eficiência dos diagnósticos, como também auxilia na estratificação de risco, o que pode levar a um tratamento direcionado a cada paciente.

3. Metodologia

O presente trabalho baseia-se numa Revisão Sistemática da Literatura, uma vez que é um tipo de estudo que utiliza métodos rigorosos para sintetizar evidências disponíveis sobre um determinado tópico ou questão de pesquisa sendo caracterizada por uma abordagem sistemática na procura, seleção e análise crítica de estudos relevantes. A sua principal característica é a aplicação de um protocolo de pesquisa detalhado e pré-definido, que define

os critérios de inclusão e exclusão dos estudos, os métodos de busca e seleção dos dados, além das estratégias de análise e síntese dos resultados de forma a garantir transparência, replicabilidade e minimização de viés no processo de revisão.

A revisão sistemática decorreu entre 12 de novembro de 2024 e 1 de janeiro de 2025, e iniciou-se pela formulação da questão de investigação segundo o anagrama PICO – População-alvo/Problema, Intervenção, Comparação e Outcome de forma a orientar na seleção de termos-chave que poderiam ser utilizados na pesquisa de artigos de investigação em bases de dados bibliográficas (Gallagher-Ford & Melnyk, 2019).

Deste modo, define-se a questão de investigação, por via do critério PICO: P de População-alvo/Problema (homens elegíveis para diagnóstico do cancro da próstata); Intervenção (Implementação de sistemas de IA para otimizar o diagnóstico do cancro da próstata); C de Comparação (cenário atual de diagnóstico do cancro da próstata vs com uso de IA) e, O de Outcome (redução de tempo de diagnóstico e melhoria na eficiência e precisão do diagnóstico). Assim, a questão problema seria: “Em homens com risco de cancro da próstata, a utilização da inteligência artificial otimiza os métodos já existentes para o diagnóstico?”.

A definição dos critérios de inclusão teve como finalidade orientar a pesquisa e selecionar a literatura em função dos resultados pretendidos e da questão formulada. São estes: (1) artigos com ano de publicação entre 2019 e 2024; (2) artigos com texto disponível na íntegra e de forma gratuita; (3) artigos disponíveis na língua portuguesa ou inglesa e (4) artigos que abordem aplicações de IA no diagnóstico do cancro da próstata. Os critérios de exclusão incluíram estudos não relacionados com o diagnóstico, estudos que abordem outras doenças, estudos que abordem outros tipos de cancro, revisões sistemáticas da literatura.

Apesar do foco em estudos indexados em bases científicas formais, foi aberta exceção para um estudo técnico relevante, desde que cumprisse critérios de clareza metodológica e contributo quantitativo direto.

3.1. Definição da pesquisa e identificação dos estudos

A revisão sistemática iniciou-se dia 12 de novembro de 2024 e foi efetuada eletronicamente na base de dados PubMed® com as palavras-chave: inteligência artificial (artificial intelligence), cancro da próstata (prostate cancer), diagnóstico médico (medical diagnosis), biópsia (biopsy) e concordância de diagnóstico (diagnostic agreement) tendo por base a

seguinte equação de pesquisa: ((artificial intelligence) AND (prostate cancer) AND (medical diagnosis) AND (biopsy) AND (diagnostic agreement)).

Paralelamente, numa fase inicial da identificação dos estudos, obtiveram-se 70 registos. Posteriormente, afinou-se a pesquisa utilizando os filtros de pesquisa e aplicaram-se os quatro critérios de inclusão e ainda os critérios de exclusão. Em seguida, realizou-se primeiramente a triagem dos documentos pelo título e, posteriormente, pelo resumo. Averiguou-se ainda, se estas fontes possuíam referências bibliográficas fidedignas.

Numa fase final, para os artigos selecionados tendo em conta o título e o resumo, procedeu-se à leitura do texto integral e utilizou-se como último critério de inclusão a sua relevância face à temática em investigação, resultando, no final, em 13 artigos.

3.2. Análise de dados

A presente revisão sistemática foi conduzida de acordo com as recomendações Preferred Reporting Items for Systematic Reviews and Meta-Analyses 2020 (PRISMA) sendo que o PRISMA correspondente aos artigos em análise encontra-se na figura 11.

A declaração PRISMA 2020 foi elaborada principalmente para revisões sistemáticas de estudos que avaliam os efeitos de intervenções em saúde, independentemente do delineamento dos estudos incluídos. Esta metodologia consiste num conjunto de diretrizes baseadas em evidências para relatar estudos em revisões sistemáticas e meta-análises, compreendendo 27 itens numa checklist que deve ser incluída no relatório de uma revisão sistemática ou meta-análise (Page et al., 2022).

3.3. Tipos de concordância

A qualidade das medições clínicas é essencial tanto para a prestação de cuidados como para a investigação. De acordo com Bland (2000) e Shoukri (2005), existe sempre algum grau de erro inerente às medições, causado por fatores como o instrumento utilizado, a experiência do observador e a variabilidade natural do fenómeno biológico.

Para Bland (2000) e Shoukri (2005) os tipos de concordância podem ser descritos da seguinte forma:

- Concordância interobservador (quando dois ou mais observadores diferentes realizam medições do mesmo fenómeno);

- Concordância intraobservador (quando o mesmo observador repete a medição em dois momentos distintos);
- Concordância entre métodos (quando se comparam os resultados de dois métodos distintos de medição do mesmo fenómeno);
- Validade (quando se compara o resultado da medição com um valor considerado verdadeiro ou de referência).

A ausência de concordância pode indicar que uma medição não é fiável ou que existe variabilidade clínica significativa.

Conforme referem Fleiss et al. (2003), "a utilidade de uma classificação depende da sua consistência; se não houver concordância entre observadores, então a classificação perde a sua relevância clínica e científica".

3.4. Métodos estatísticos para avaliar a concordância em variáveis contínuas

Existem vários métodos estatísticos para avaliar a concordância entre observações, quando estas são expressas como variáveis contínuas. No entanto, nenhum método é considerado universalmente o melhor, pois todos apresentam vantagens e limitações (Shoukri, 2005).

Um dos métodos mais simples consiste em utilizar a análise de variância (ANOVA) de uma via para estimar o desvio padrão intraindivíduo, ou seja, o erro de medição, e calcular o coeficiente de repetibilidade que define a diferença máxima esperada entre duas medições do mesmo indivíduo com 95% de confiança.

Outro método amplamente utilizado é o Coeficiente de Correlação Intraclasse (ICC) que mede a proporção da variabilidade total que se deve a diferenças reais entre indivíduos em oposição a erros de medição. O ICC é bastante intuitivo, assumindo valores entre 0 (nenhuma concordância) e 1 (concordância perfeita). Contudo, o seu valor é influenciado pela heterogeneidade da amostra (Müller e Büttner, 1994).

Uma alternativa é o Coeficiente de Correlação de Concordância (CCC) de Lin (1989) que ajusta o coeficiente de Correlação de Pearson, tendo em conta o desvio em relação à linha de identidade. Este método é mais sensível ao viés sistemático entre observadores, mas, tal como o ICC, também pode ser errático quando a variância entre sujeitos é pequena (Nickerson, 1997).

3.5. Métodos estatísticos para avaliar concordância em variáveis categóricas

Quando as variáveis em análise são categóricas, os métodos estatísticos utilizados para avaliar a concordância entre observações ou entre métodos de medição são diferentes daqueles aplicáveis a variáveis contínuas.

O método estatístico mais conhecido e amplamente utilizado neste contexto é o Coeficiente Kappa de Cohen (Cohen, 1960). O Kappa mede a concordância observada além da concordância que seria esperada apenas pelo acaso. Este coeficiente varia entre -1 e 1, sendo que os valores positivos indicam maior concordância do que o acaso, valores negativos indicam discordância sistemática e valor de 0 indica que a concordância equivale à que seria esperada por puro acaso.

No entanto, o Kappa apresenta limitações, tais como, a sensibilidade à prevalência das categorias (se quase todos os casos forem classificados como “negativo” o Kappa pode ser artificialmente baixo mesmo existindo elevada concordância) e é afetado pelo viés entre observadores, ou seja, se um observador tende a usar mais determinadas categorias do que o outro (Byrt, Bishop & Carlin, 1993).

Para lidar com estas limitações, foi proposto o Kappa Ponderado, que é muito útil para variáveis ordinais. Neste caso, diferentes tipos de desacordo recebem pesos distintos, sendo considerado menos grave discordar entre categorias próximas do que entre categorias distantes (Cohen, 1968).

Quando o objetivo é avaliar a validade de um teste diagnóstico, utilizam-se indicadores como a Sensibilidade (proporção de verdadeiros positivos identificados corretamente), Especificidade (proporção de verdadeiros negativos), Valor preditivo positivo e negativo e Acurácia (proporção de classificações corretas face ao total de casos).

Desta forma, a escolha do método estatístico deve ter em conta o tipo de variável, o objetivo da análise e as características da amostra.

3.6. Avaliação da qualidade

Num momento prévio à realização do estudo, foi definido por todos os revisores que só se incluíam os estudos que demonstrassem qualidade metodológica, tendo em consideração os instrumentos preconizados pela JBI para estudos quantitativos e qualitativos.

A implementação do modelo JBI é essencial para a integração de práticas baseadas em evidências nos cuidados de saúde. Apresenta uma abordagem estruturada e adaptável o que facilita a adoção de mudanças e garante a melhoria contínua da qualidade e eficiência. A sua

importância apresenta-se pela sua capacidade de transformar a prática clínica e promover cuidados de excelência fundamentados por evidências (Munn et al., 2023).

Para cada estudo, a qualidade metodológica, foi avaliada separadamente (Anexo A).

Profissionais de saúde, gestores de políticas em saúde e pacientes necessitam de saber o grau de certeza que têm nos resultados do conjunto de evidências utilizadas para sustentar as recomendações em saúde. Portanto, identificar a certeza que temos numa evidência não é uma tarefa simples e requer métodos rigorosos, estruturados e transparentes como a GRADE. A abordagem GRADE foi desenvolvida por um grupo de pesquisadores com o objetivo de avaliar e graduar a certeza do conjunto de evidências numa revisão sistemática, ou seja, a confiança que depositamos sobre a eficácia e segurança de um diagnóstico, por exemplo, todavia este sistema é mais utilizado em estudos randomizados (Anexo B).

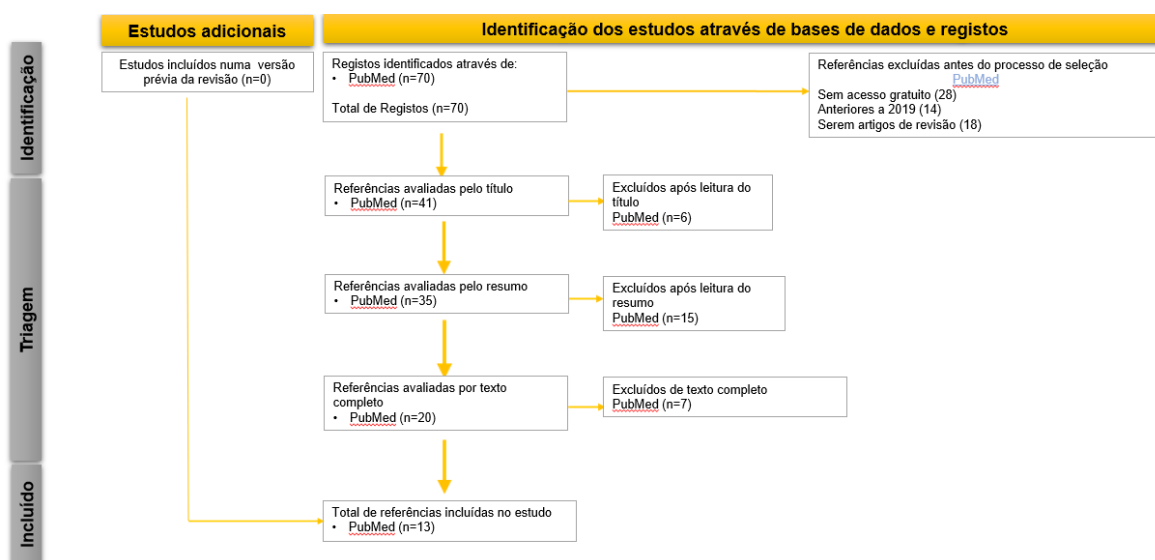


Figura 11 – Diagrama da aplicação da metodologia PRISMA

Tabela 1 – Características e resultados da análise metodológica dos estudos

Base de dados	Autores/ Ano/ País	Título do artigo	Amostra	Objetivo	Resultado
---------------	-----------------------	------------------	---------	----------	-----------

PubMed®	Ström P. et al. (2020)	Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study	8980	Desenvolver um sistema de IA para deteção, localização e graduação (Gleason score) do cancro da próstata.	AUC:0,997 Sensibilidade: 99,6% Especificidade: 86,6% Kappa (média) – 0,62 Kappa (vs patologistas) – 0,83
PubMed®	Bulten W. et al. (2022)	Artificial intelligence assistance significantly improves Gleason Grading of prostate biopsies by pathologists	160	Avaliar o impacto da assistência por IA na precisão dos patologistas na graduação de Gleason em biópsias prostáticas.	Kappa (sem IA) – 0,799 Kappa (com IA) – 0,872 Mairo benefício para patologistas com <15 anos de experiência.
PubMed®	Schelb P. et al. (2019)	Classification of cancer at prostate MRI: DL vs Clinical PI-RADS assessment	312	Comparar o desempenho de uma modelo de DL (rede neural U-Net) treinado com imagens de ressonância magnética prostática ponderadas em T2 e por difusão, na tarefa de deteção e segmentação de lesões suspeitas de cancro da próstata clinicamente significativo.	<u>U-Net</u> <u>Sensibilidade: 92%</u> <u>Especificidade: 47%</u> <u>Acurácia: 66%</u> <u>PI-RADS</u> <u>Sensibilidade: 88%</u> <u>Especificidade: 50%</u>

PubMed®	Yoo S. et al. (2019) Canadá	Prostate Cancer detection using deep convolutional neural networks e (175 com cancro da próstata clinicament e + 252 sem cancro da próstata)	427	Desenvolver e testar uma pipeline automatizada baseada em redes neurais convulcionais (CNN) para deteção de cancro clinicamente significativo da próstata, tanto ao nível do corte (slice) como ao nível de paciente, com base em imagens de RMN	<u>AUC</u> Slice: 0,87 Paciente: 0,84
	Hiremath A. et al. (2021) EUA e Países Baixos	An integrated nomogram combining deep learning, prostate imaging - reportong and data system (PI-RADS) scoring, and clinical variables for identification of clinically significant prostate cancer on biparametric MRI: a retrospective multicentre study	592 pacientes com 823 lesões analisadas	Construir um nomograma integrado (ClaD) que combina predições de IA (DL em bpMRI), score PI-RADS e variáveis clínicas (PSA, volume da próstata e lesão) para detetar o cancro clinicamente significativo	<u>AUC</u> ClaD: 0,81 IA + clínica: 0,76 PI-RADS + clínica: 0,76 <u>Acurácia</u> ClaD: 77,92% IA + clínica: 75,7% PI-RADS + clínica: 76,1% <u>Sensibilidade</u> ClaD: 83,23% IA+clínica: 81,5% PI-RADS+clínica: 86,7% <u>Especificidade</u> ClaD: 59,18% IA+clínica: 55,1% PI-RADS+clínica: 38,8% ClaD teve um aumento de 5-7% na AUC (p<0,001)

PubMed®

Artificial Intelligence system shows performance at the lever or uropathologist (2022)	593 digitais de biópsias da próstata (130 normais e 463 com adenocarcinoma)	Validar externamente o desempenho do sistema de IA DeepDx Prostate na deteção e graduação de cancro da próstata, comparando com patologistas especializados e analisar o impacto da IA na prática de um patologista geral.	Sensibilidade: 98,49%
Jung M. et al. (2022)	Coreia do Sul e Chipre	Coreia do Sul e Chipre	Índice de Kappa (Gleason Grade Group -GG): Kappa – 0,713 Kappa Ponderado – 0,922
Coreia do Sul e Chipre	Coreia do Sul e Chipre	Coreia do Sul e Chipre	Índice de Kappa (Gleason Score-GS) Kappa – 0,654 Kappa Ponderado – 0,904

PubMed®

Pantanowitz L. et al. (2020) EUA

An artificial intelligence algorithm for prostate cancer diagnosis in whole slide images of core needle biopsies: a blinded clinical validation and deployment study

2501 lâminas de 210 casos (amostra interna) 1627 lâminas de 100 casos (amostra externa)

Validar clinicamente um algoritmo de IA para diagnóstico de cancro da próstata com base em imagens digitalizadas de biópsias com agulha, incluindo detecção de cancro, Gleason score, invasão perineural e percentagem tumoral, e avaliar a sua implementação clínica.

Deteção de cancro:
- Sensibilidade: 98,46%
- Especificidade: 97,33%
- AUC: 0,991

Gleason 7-10:
- Sensibilidade: 85,9%
- Especificidade: 90,41%
- AUC: 0,941

Gleason padrão 5:
- Sensibilidade: 85%
- Especificidade: 90,84%
- AUC: 0,971

Invasão perineural
- Sensibilidade: 86,96%
- Especificidade: 90,74%
- AUC: 0,857

PubMed®	Singhal N. et al. (2022)	Índia	A deep learning system for prostate cancer diagnosis and grading in whole slice images of core needle biopsies	6670 imagens digitais de lâminas de biópsias com agulha	Desenvolver um sistema de DL para detetar áreas de cancro e atribuir grupos de Gleason em imagens de biópsias prostáticas.	<p><u>Teste interno</u></p> <p>Acurácia: 89,4%</p> <p>Kappa – 0,92</p> <p>AUC – 0,997</p> <p><u>Validação externa</u></p> <p>Acurácia: 85,3%</p> <p>Kappa – 0,96</p> <p>AUC – 0,991</p> <p><u>Teste cego</u></p> <p>Acurácia: 83,1%</p> <p>Kappa – 0,93</p> <p>AUC – 0,930</p>
	Campanella G. et al. (2019)	EUA	Clinical-grade computational pathology using weakly supervised deep learning on whole slide images	44732 lâminas de 15187 pacientes e 3 instituições	Desenvolver um sistema de apoio à decisão clínica baseado em DL, para classificar imagens inteiras de lâminas histológicas	<p>AUC:0,991</p> <p>Sensibilidade: 100%</p> <p>Especificidade: 70-80%</p>
PubMed®	Korevaar S. et al. (2021)	Austrália	Incidental detection of prostate cancer with computed tomography scans	571 abdomino-pélvico - 139 com cancro da próstata clinicament e significativ o e 432 controlo (sem cancro conhecido)	Verificar se algoritmos DL aplicados a TACS (não otimizados para a próstata) podem identificar cancro de forma incidental e com que precisão	<p>AUC: 0,88</p> <p>Sensibilidade: 55,6% (limiar 0,5) podendo subir para cerca de 78% com limiar ajustado</p> <p>Especificidade: 98,8%</p> <p><u>Comparação com humanos</u></p> <p>AUC radiologista 1: 0,61</p> <p>AUC radiologista 2: 0,70</p> <p>Kappa (IA vs gold standard): 0,63</p>

PubMed®

Yet another Conjunto Desenvolver e Acurácia no conjunto
 automated de treino: validar um sistema de validação interna:
 Gleason 6664 de IA (YAAGGS) que 77,5%;
 Grading imagens de possa realizar a Acurácia na validação
 System biópsias graduação de externa: 54,5%.
 (YAAGGS) by prostáticas Gleason Kappa(conjunto
 weakly ; (Validação automaticamente, interno): 0,650
 supervised externa: melhorar a precisão Kappa Ponderado
 Mun Y., et deep learning 167; e reduzir a (validação interna):
 al. Validação variabilidade entre 0,897
 (2021) interna: patologistas e ser Kappa (validação
 134) aplicado em grande externa): 0,389
 escala sem Kappa Ponderado
 Coreia do necessidade de (validação externa):
 Sul rotulagem manual 0,880
 excessiva.

Os resultados mostram uma forte concordância entre o sistema de IA e o gold standard estabelecido por consenso de especialistas.

PubMed®	Li J. et al. (2023) EUA	Interobserver Reproducibility of Quantifyinf Gleason Pattern 4 cancer in prostate biopsy: implications for clinical practice	55 biópsias de próstata com adenocarcinoma. Analisadas por 12 patologistas	Avaliar a reprodutibilidade interobservador na quantificação do Gleason Pattern 4 em biópsias prostáticas considerando variáveis como o tamanho do tumor e os subtipos histológicos.	<p><u>Reprodutibilidade global foi moderada:</u> Kappa – 0,57</p> <p>A reprodutibilidade variou consoante o tamanho da lesão:</p> <ul style="list-style-type: none"> . Tumores ≤ 2mm: Kappa – 0,51 . Tumores 2.1 a 5mm: Kappa – 0,50 . Tumores > 5mm: Kappa – 0,66 <p>A reprodutibilidade variou com o subtipo histológico:</p> <ul style="list-style-type: none"> . Glândulas malformadas (P): Kappa – 0,43 . Glândulas fundidas (F): Kappa – 0,57 . Cribiformes/Glomerulóides: Kappa – 0,74
	PubMed®	Sherafatmandjoo H. et al. (2024) Irão	Prostate cancer diagnosis based on multi-parametric MRI, clinical and pathological factors using deep learning	343 pacientes	Investigar o impacto da integração de dados clínicos/ patológicos (ex: PSA, volume da próstata) no desempenho de modelos de DL que analisam mpMRI para detetar cancro da próstata.

3.7. Discussão dos Resultados

Foram analisados 13 estudos publicados que investigaram a aplicação da IA no diagnóstico do cancro da próstata, abrangendo diferentes metodologias, populações e tipos de tecnologia. Estes estudos utilizaram algoritmos de DL aplicados quer a imagens médicas (MRI, TC), quer a lâminas histológicas digitalizadas, com o objetivo de melhorar a precisão diagnóstica, reduzir o erro humano e aumentar a eficiência na prática clínica.

A utilização de IA em patologia digital revelou-se particularmente promissora. Ström et al. (2020), por exemplo, demonstraram que o sistema de IA desenvolvido atingiu uma AUC de 0,997, sensibilidade de 99,6% e especificidade de 86,6%. Resultados semelhantes foram obtidos por Campanella et al. (2019), num estudo multicêntrico de larga escala, com uma AUC de 0,991, sensibilidade de 100% e especificidade entre 70% e 80%. Estes dados confirmaram a capacidade da IA para detetar e classificar lesões prostáticas com um nível de precisão comparável ao dos especialistas humanos, reforçando o seu valor como ferramenta de apoio no diagnóstico primário.

No que respeita à graduação do score de Gleason, um dos principais desafios diagnósticos, dada a sua natureza subjetiva, os resultados também foram encorajadores. O estudo de Bulten et al. (2022) mostrou que a assistência por IA aumentou a concordância entre patologistas, com o índice de Kappa a aumentar de 0,799 para 0,872. Este efeito foi especialmente visível entre profissionais com menos de 15 anos de experiência, sugerindo que a IA pode atuar como um mecanismo de compensação da variabilidade interobservador. De forma semelhante, Mun et al. (2021) reportaram um Kappa ponderado de 0,897 em validação interna e de 0,880 em validação externa com o sistema YAAGGS, utilizando rotulagem fraca no treino de algoritmos – uma abordagem inovadora que poderá facilitar a escalabilidade destes sistemas.

A robustez da IA foi igualmente evidenciada por Jung et al. (2022), cujo sistema DeepDX Prostate demonstrou uma concordância superior à dos relatórios clínicos originais, validando a fiabilidade do modelo não apenas em termos de desempenho, mas também de consistência. Já Singhal et al. (2023) apresentaram resultados notáveis em diferentes contextos (interno, externo e teste cego), com acurácia sempre superior a 83% e índices de Kappa acima de 0,92, confirmando a estabilidade dos modelos em múltiplos cenários clínicos.

Na área da imagem médica, a aplicação de IA à mpMRI também mostrou avanços importantes. Yoo et al. (2019) obtiveram AUC de 0,87 por imagem (slice) e de 0,84 por paciente na deteção de cancro clinicamente significativo, ultrapassando o desempenho habitual do sistema PI-RADS, que tende a variar entre 0,69 e 0,81. Estes dados confirmam o valor de IA como alternativa mais objetiva e menos dependente da interpretação individual.

Em modelos integrativos, Hiremath et al. (2021) combinaram variáveis clínicas com DL e PI-RADS num nomograma (ClAD), alcançando uma AUC de 0,81 e uma acurácia de quase 78%. O modelo combinatório demonstrou melhor desempenho do que qualquer uma das abordagens isoladas. De forma semelhante, Sherafatmandjoo et al. (2024) demonstraram que a

integração de dados clínicos e patológicos com imagens de mpMRI aumentou significativamente a performance dos modelos, atingindo valores de sensibilidade (94,24%), especificidade (98,62%) e acurácia média (95%) notavelmente altos.

Uma perspetiva particularmente interessante é trazida por Korevaar et al. (2021), que exploraram a capacidade da IA para detetar incidentemente cancro da próstata em TAC's abdominopélvicos não otimizados para esse fim. Com uma AUC de 0,88 e especificidade de 98,8%, o modelo demonstrou potencial como ferramenta de triagem oportunista. No entanto, a sensibilidade inicial de 55,6% sugere que o desempenho depende fortemente do limiar aplicado, o que poderá exigir ajustes consoante o contexto clínico.

Por fim, Li et al. (2023) analisaram a reprodutibilidade interobservador na quantificação do padrão de Gleason 4 e encontraram uma concordância moderada (Kappa - 0,57), com variações significativas consoante o tamanho tumoral e o subtipo histológico. Este estudo evidencia uma das fragilidades da prática clínica atual – a variabilidade diagnóstica – e justifica a aposta crescente na IA como mecanismo de padronização e melhoria da qualidade.

Em síntese, os estudos analisados apontam de forma consistente para o potencial da IA enquanto tecnologia complementar no diagnóstico do cancro da próstata. As vantagens mais evidentes incluem o aumento da precisão, a redução do tempo de análise, a mitigação da variabilidade interobservador e a possibilidade de aplicação em diferentes contextos clínicos. No entanto, a generalização destes sistemas à prática clínica dependerá da validação contínua, da aceitação pelos profissionais de saúde e da integração com as infraestruturas já existentes nos sistemas de saúde.

4. Análise estatística

Após a análise qualitativa dos estudos incluídos na Revisão Sistemática da Literatura, procede-se agora à realização de uma análise estatística com o objetivo de sintetizar, de forma quantitativa, a performance diagnóstica dos modelos de IA aplicados ao diagnóstico do cancro da próstata.

Foram incluídos nesta análise os estudos que reportaram métricas objetivas de desempenho, nomeadamente: AUC, sensibilidade e especificidade, e que utilizaram um método de referência validado. A seleção dos estudos foi baseada nos seguintes critérios: estudos originais com desenho observacional ou ensaios clínicos e comparação da IA com o gold standard.

A análise estatística foi conduzida com recurso ao software SPSS versão 29.0.2.0, para estimar a média ponderada das métricas diagnósticas. Esta abordagem foi escolhida tendo em conta a heterogeneidade metodológica entre os estudos (em termos de algoritmos utilizados, população-alvo, desenho do estudo e modalidades diagnósticas)

Sempre que possível, foram calculadas medidas agregadas de AUC, sensibilidade e especificidade, com intervalos de confiança a 95% e representadas em gráficos ilustrativos do tipo forest plot que permitirão visualizar o efeito conjunto da aplicação da IA no diagnóstico do cancro da próstata.

A qualidade metodológica dos estudos foi avaliada com base na escala GRADE e o viés de publicação.

4.1. Caracterização geral dos estudos

Tabela 2 - Estudos incluídos na análise estatística com métricas quantitativas de desempenho da IA

Autor (Ano)	Modalidade	AUC	Sensibilidade (%)	Especificidade (%)	Índice de Kappa	Amostra (validação externa)
Schelb P. et al. (2019)	mpMRI	0,890	91,30	76,9	0,89	52
Campanella G. et al. (2019)	Histologia digital	0,989	100	75,00	0,87	1824
Ström P. et al. (2020)	Histologia digital	0,997	99,87	99,86	0,86	6682
Korevaar S. et al. (2021)	TAC	0,880	55,60	98,80	0,63	115
Yoo S. et al. (2019)	mpMRI	0,870	-----	-----	-----	427
Hiremath A. et al. (2021)	mpMRI + clínica	0,850	89,20	66,70	0,74	281
Sherafatmandjoo H. et al. (2024)	mpMRI + clínica	-----	94,24	98,62	-----	343
Pantanowitz L. et al. (2020)	Histologia digital	0,997	100	97,30	0,86	1521

Os oito estudos incluídos na presente análise estatística apresentam uma considerável diversidade metodológica, abrangendo diferentes modalidades de diagnóstico assitido por IA

no contexto do cancro da próstata. Nos estudos com valores não reportados diretamente, os dados foram excluídos da análise quantitativa dessas métricas específicas, mas mantidos na caracterização geral.

Três dos estudos basearam-se na análise de histologia digital, através de lâminas digitalizadas de biópsias prostáticas como é o caso de Campanella G. et al. (2019), Ström P. et al. (2020) e Pantanowitz L. et al. (2020), quatro dos estudos incidiram sobre exames de imagem médica nomeadamente mpMRI – Schelb P. et al. (2019), Yoo S. et al. (2020), Hiremmath A. et al. (2023) e Sherafatmandjoo H. et al. (2024).

As amostras analisadas variaram significativamente, desde 52 a 6682 pacientes, refletindo a diversidade dos contextos clínicos e metodologias utilizadas.

A área sob a curva (AUC) apresentou valores elevados na maioria dos estudos, variando entre 0,850 e 0,997, o que evidencia uma forte capacidade discriminativa dos algoritmos de IA. As sensibilidades reportadas situaram-se entre 55,6% e 100%, enquanto as especificidades variaram entre 66,7% e 98,86%.

Estes resultados reforçam o potencial da IA como ferramenta complementar no diagnóstico precoce do cancro da próstata, embora a heterogeneidade justifique a realização da análise estatística para integrar e comparar os dados de forma sistemática.

Tabela 3 – Análise estatística descritiva dos estudos (fonte: SPSS v 29.0.2.0)

	Média	Desvio Padrão	Mínimo	Máximo
AUC (7 estudos)	0,92129	0,072194	0,840	0,997
Sensibilidade (7 estudos)	0,9103	0,15954	0,556	1,000
Especidificade (7 estudos)	0,8886	0,13095	0,670	0,998
Índice de Kappa (5 estudos)	0,7320	0,15865	0,600	0,920

4.2. Análise de Métricas Contínuas

As métricas de AUC, sensibilidade e especificidade, acurácia e índice de Kappa foram agregadas por média ponderada com base no tamanho amostral de cada estudo, com cálculo de intervalos de confiança a 95%.

4.2.1. Análise descritiva de AUC

A análise descritiva das AUC dos estudos incluídos, foi realizada no SPSS v29.0.2.0. A AUC é uma métrica estatística utilizada para avaliar a capacidade discriminativa de modelos preditivos, sendo particularmente relevante em contextos de diagnóstico médico (Mandrekar, 2010).

No conjunto dos 7 estudos analisados, a AUC apresentou uma média de 0,92129 (\pm 0,072), com valores mínimos e máximos de 0,840 e 0,997 respetivamente. A amplitude total foi de 0,157 o que demonstra alguma variabilidade no desempenho dos algoritmos avaliados. Esta dispersão pode estar relacionada com diferenças metodológicas, amostras utilizadas, tipo de validação empregue ou ainda o tipo de modalidades (histologia digital vs. imagem médica).

Estes resultados demonstram uma elevada capacidade discriminativa dos sistemas de IA utilizados nos estudos, sendo que a maioria apresenta AUCs próximas da unidade, o que indica uma excelente performance na distinção entre casos positivos e negativos de cancro da próstata.

4.2.1.1. Média ponderada e intervalo de confiança da AUC dos estudos incluídos

Com o objetivo de sintetizar a performance diagnóstica dos modelos de IA aplicados ao diagnóstico do cancro da próstata, foi calculada a média simples da AUC dos estudos incluídos na presente análise estatística. A AUC é uma métrica estatística utilizada para avaliar a capacidade discriminativa de modelos preditivos, sendo particularmente relevante em contextos de diagnóstico médico (Mandrekar, 2010).

Para este cálculo, consideraram-se os valores de AUC reportados nos 7 estudos elegíveis, bem como os respetivos tamanhos da amostra. Foi calculado o erro padrão (SE) estimado de cada estudo através da seguinte fórmula:

$$SE = \frac{ICS_{sup} - IC_{inf}}{2 \times 1,96}$$

Posteriormente, foi atribuída a cada estudo uma ponderação inversamente proporcional ao quadrado do seu erro padrão:

$$VAR = SE^2 \quad \text{Peso} = \frac{1}{VAR}$$

Estes pesos foram utilizados para calcular a média ponderada da AUC, o que permite reduzir a influência de estudos com menor precisão estatística. O cálculo da média foi realizado no software SPSS versão 29.0.2.0, utilizando a função “Weight cases” e o módulo de estatísticas descritivas.

O valor obtido para a AUC média foi de 0,994, com um intervalo de confiança 95% de [0,8663–1,000], indicando uma performance global excelente dos modelos de IA no diagnóstico do cancro da próstata. Para o cálculo do intervalo de confiança, foi utilizada a seguinte fórmula:

$$IC95\% = Média \pm 1,96 * \frac{SD}{\sqrt{k}}$$

SD representa o desvio padrão ponderado e K o número de estudos incluídos.

A seguinte figura (12) permite visualizar de forma clara e comparativa, o desempenho discriminativo de cada modelo de IA avaliado, bem como a sua precisão. As barras de erro mais curtas indicam estimativas mais estáveis e confiáveis enquanto as barras mais longas refletem maior incerteza estatística. O gráfico apresentado complementa a média simples previamente apresentada, proporcionando uma visão mais completa sobre a dispersão dos resultados e as diferenças entre os estudos.

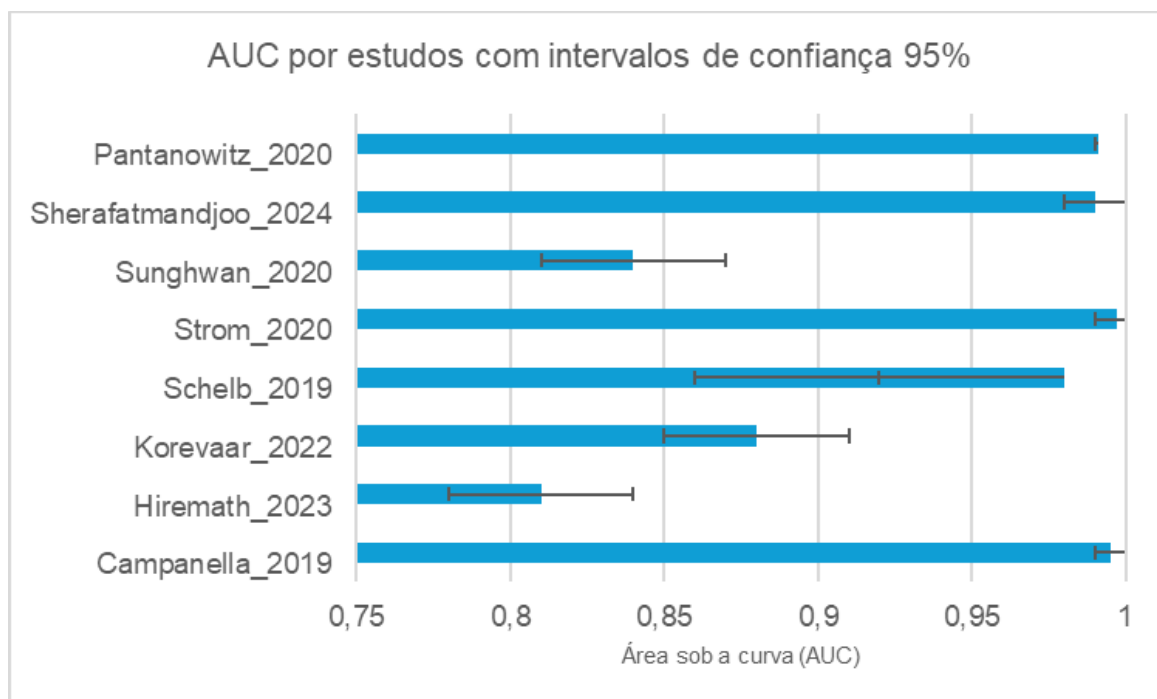


Figura 12 - Representação gráfica da AUC dos estudos incluídos na análise estatística, com barras de erro correspondentes a +/- 2 erros padrão (SE), aproximando-se de um intervalo de confiança de 95% (fonte: Excel)

Tabela 4 – IC95%, Erro padrão, Pesos e Variâncias da métrica AUC (fonte: SPSS v29.0.2.0)

Estudos	AUC	IC95% Inferior	IC95% Superior	Erro Padrão (SE)	Peso	Variância
Schelb P. et al (2019)	0,890	0,810	0,970	0,4082	600,25	0,001666
Campanella G. et al. (2019)	0,995	0,988	0,990	0,00051	3841600	0,00000026
Ström P. et al. (2020)	0,997	0,996	0,998	0,000510	3841600	0,00000026
Korevaar S. et al. (2021)	0,880	0,860	0,900	0,10204	9604,00	0,000104
Hiremath A. et al. (2021)	0,850	0,820	0,880	0,015306	4268,44	0,000234
Yoo S. et al. (2019)	0,840	0,760	0,910	0,038265	682,95	0,001464
Pantanowitz L. et al. (2020)	0,997	0,993	0,999	0,001531	426844,44	0,000002

A tabela 4 apresenta os valores de AUC dos estudos incluídos na análise estatística, acompanhados dos respetivos intervalos de confiança a 95%, erro padrão, variâncias e pesos estatísticos. Estes dados foram utilizados para o cálculo da média ponderada da AUC global, conferindo maior robustez estatística ao modelo de síntese quantitativa.

Foi utilizada a seguinte fórmula para calcular a média ponderada de AUC:

$$AUC_{ponderada} = \frac{\sum((AUC_i \times w_i))}{\sum w_i}$$

Onde:

AUC_i = AUC do estudo i

W_i = peso do estudo i (calculado como $\frac{1}{variância}$)

Para calcular o Erro padrão (SE) da média ponderada, foi usada a seguinte fórmula:

$$SEAUC = \sqrt{\frac{1}{\sum w_i}}$$

Por último para calcular o IC95%, foi usada a seguinte fórmula:

$$IC95\% = AUC_{ponderada} \pm Z \times SE$$

Onde:

Z = 1,96 (valor crítico para 95%)

Com base nos pesos atribuídos pela inversa da variância de cada estudo, obteve-se uma média ponderada de AUC de 0,9962. O intervalo de confiança a 95% foi calculado em [0,9956;0,9968] com um erro padrão estimado de 0,00029. Estes resultados indica uma elevada precisão diagnóstica global dos modelos de IA aplicados ao cancro da próstata nos estudos analisados.

4.2.1.2. Heterogeneidade das AUC

A heterogeneidade estatística é uma componente crítica na interpretação dos resultados de uma análise estatística, porque reflete a variabilidade dos efeitos observados entre os estudos incluídos. De acordo com Higgins et al. (2003), a presença de heterogeneidade significativa pode comprometer a validade da média combinada dos efeitos, sendo essencial avaliá-la antes da síntese quantitativa final. Na presente análise estatística, foi utilizada a estatística Q de Cochran e o índice I² para quantificar a heterogeneidade da AUC dos modelos de IA aplicados ao diagnóstico do cancro da próstata.

Para avaliar a heterogeneidade da AUC serão utilizadas diferentes fórmulas, tais como:

Cálculo do Q (estatística de Cochran):

$$Q = \sum Qi = \sum_{i=1}^k wi \times (AUCi - AUCponderada)^2$$

Onde:

- Q é a estatística total de heterogeneidade
- $Wi = \frac{1}{variância_i}$
- AUC_i = valor da AUC do estudo
- AUC_{ponderada} = média ponderada (0,9962)

Tabela 5 - Valores da heterogeneidade da métrica AUC

Estudo	AUC _i	Média (AUC _{ponderada})	(AUC _i - AUC _{ponderada}) ²	Q _i
Schelb P. et al., (2019)	0,890	0,9962	0,01129	6,77
Campanella G. et al., (2019)	0,995	0,9962	0,00000144	5,53
Strom P. et al., (2020)	0,997	0,9962	0,00000064	2,46

Korevaar S. et al., (2021)	0,880	0,9962	0,01346	129,28
Hiremath A. et al., (2021)	0,850	0,9962	0,02142	91,43
Yoo S. et al. (2019)	0,840	0,9962	0,02436	16,64
Pantanowitz L. et al., (2020)	0,997	0,9962	0,00000064	2,96

Consoante os resultados apresentados na tabela 5, podemos calcular o valor de Q.

$$Q = \sum Qi = 255,07$$

Posteriormente, serão calculados os graus de liberdade (df). De acordo com Fisher R. (1934), os graus de liberdade referem-se ao número de valores independentes que podem variar num cálculo estatístico depois de se aplicar uma restrição (por exemplo a média). Fórmula:

$$df = k - 1$$

Onde:

- K = número de estudos incluídos

Neste caso, df = 6 (sabendo que são usados 7 estudos)

O índice de heterogeneidade I², proposto por Higgins et al. (2003), quantifica a proporção da variabilidade entre estudos que é atribuída a heterogeneidade real e não ao erro amostral. É calculado com base na estatística Q de Cochran e nos graus de liberdade e assume valores entre 0% (sem heterogeneidade) e 100% (heterogeneidade total). É calculado pela seguinte fórmula:

$$I^2 = \max\left(0, \frac{Q - df}{Q}\right) \times 100$$

Onde:

- Q = 255,07
- Df = 6

Desta forma, I² = 97,65%, o que indica uma heterogeneidade muito elevada entre os estudos analisados, o que sugere que a variabilidade dos resultados não pode ser atribuída apenas ao erro aleatório. Assim, justifica-se plenamente a utilização de um modelo de efeitos aleatórios na síntese de dados, de forma a acomodar a variabilidade real existente entre os estudos.

4.2.2. Análise de Métricas: Sensibilidade e Especificidade

A sensibilidade (true positive rate) e a especificidade (true negative rate) são métricas essenciais para a avaliação da performance de testes de diagnóstico binários, especialmente relevantes em contexto clínico.

Segundo Altman D. (1994), sensibilidade refere-se à capacidade do teste identificar corretamente os verdadeiros positivos, ou seja, os indivíduos que efetivamente têm uma doença.

A fórmula para calcular a sensibilidade é:

$$\text{Sensibilidade} = \frac{\text{Verdadeiros Positivos}}{\text{Verdadeiros Positivos} + \text{Falsos Negativos}}$$

A especificidade refere-se à capacidade do teste identificar corretamente os casos verdadeiros negativos, ou seja, os indivíduos que não têm uma doença (Altman, 1994), e é calculada pela seguinte fórmula:

$$\text{Especificidade} = \frac{\text{Verdadeiros Negativos}}{\text{Verdadeiros Negativos} + \text{Falsos Positivos}}$$

Em contexto de diagnóstico do cancro da próstata, estas métricas assumem uma grande importância, porque uma sensibilidade elevada reduz o risco de não detetar o cancro clinicamente significativo (falsos negativos) e uma especificidade elevada reduz o risco de alarmes falsos, evitando biópsias ou tratamentos desnecessários (falsos positivos).

Para estas métricas foi excluído o estudo de Yoo S. et al. (2019) por não apresentar estes dados relativos a esta variável. A métrica sensibilidade (%) que traduz a capacidade dos modelos de IA em identificar corretamente os casos positivos obteve uma média de 0,9103 com um IC95% de [0,598;1,000] e um desvio padrão de 0,1595. A existência de uma elevada média, IC95% amplo e desvio padrão elevado indicam uma variabilidade entre os estudos possivelmente relacionada com a heterogeneidade das metodologias utilizadas. Relativamente à métrica especificidade (%), a média foi de 0,8886, com um IC95% de [0,632;1,000] e desvio padrão de 0,1309. Estes resultados evidenciam uma maior variabilidade nos valores de especificidade entre os estudos analisados, evidenciando que a capacidade dos modelos de IA em identificar corretamente os casos negativos não foi uniforme entre os estudos avaliados.

A figura 13 e 14 representam os valores de sensibilidade (%) e especificidade obtidos nos estudos incluídos. Verifica-se uma elevada sensibilidade global dos modelos de IA, com a exceção do estudo de Korevaar (2021), que reporta um valor mais reduzido (0,556). A variação entre os estudos reflete diferenças metodológicas e populacionais, bem como a

heterogeneidade na modalidade de imagem utilizada. No que diz respeito a especificidade (%) o estudo de Hiremath (2021) apresenta o valor mais baixo (0,667) o que pode sugerir diferenças nos critérios de avaliação e modalidades diagnósticas usadas.

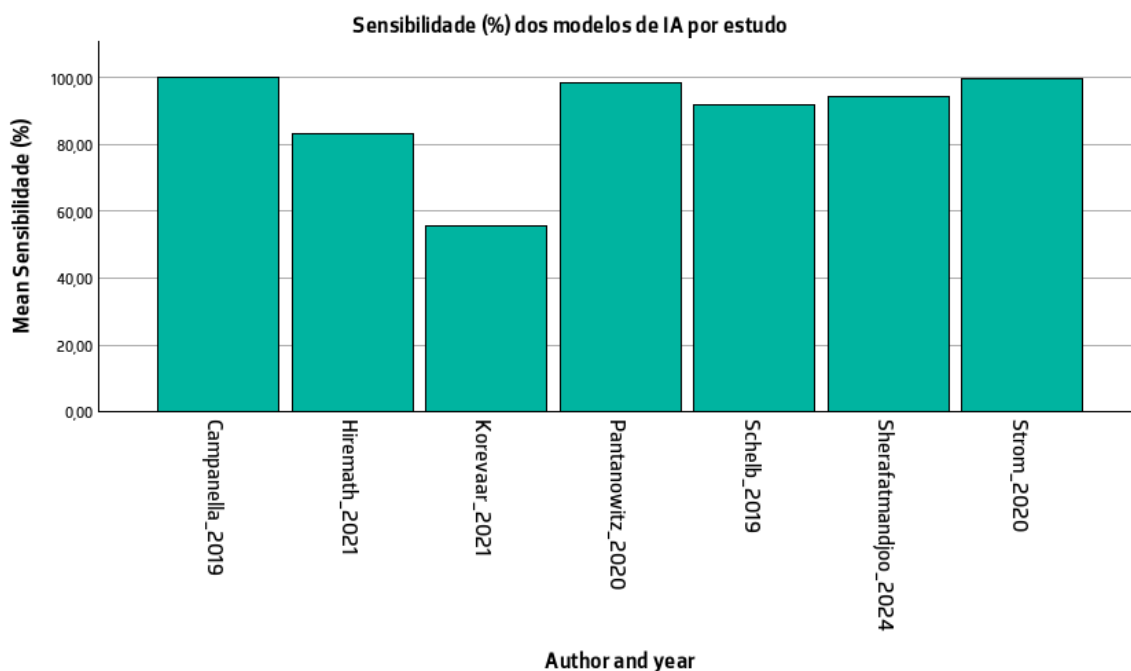


Figura 13 - Representação gráfica dos valores de sensibilidade obtidos nos estudos incluídos (fonte: SPSS v29.0.2.0)

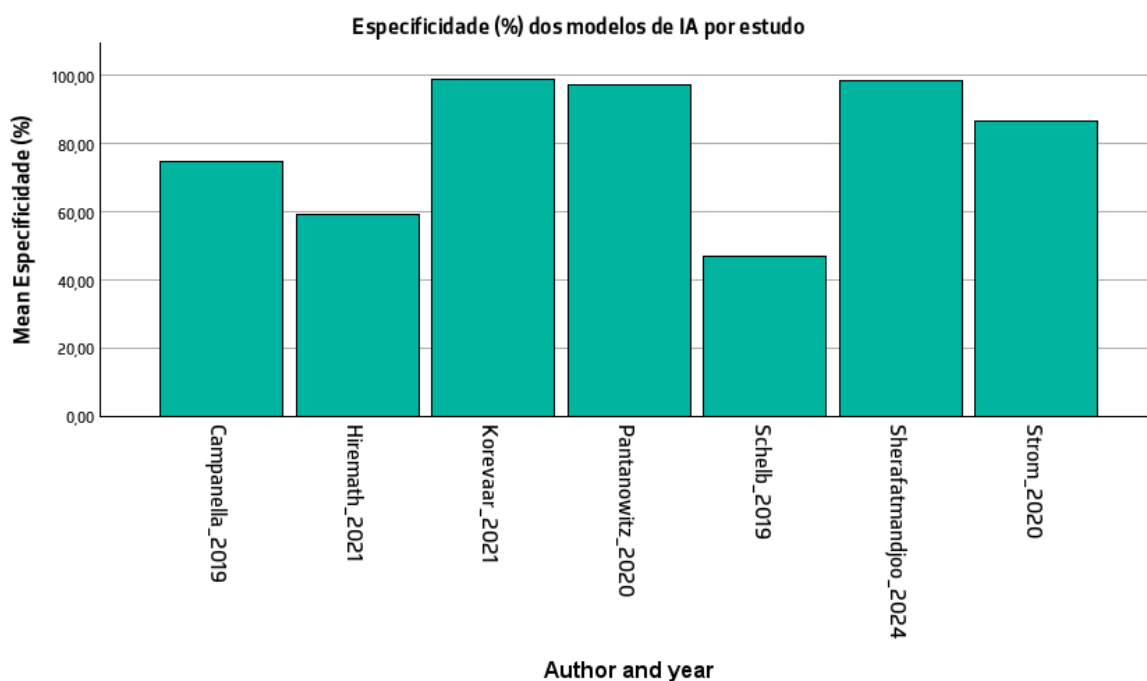


Figura 14 - Representação gráfica dos valores de especificidade obtidos nos estudos incluídos (fonte: SPSS v29.0.2.0)

Tabela 6 – Correlação de Pearson entre sensibilidade e especificidade (fonte: SPSS v29.0.2.0)

			Sensibilidade	Especificidade
Sensibilidade	Correlação de Pearson	de	1	-0,167
	Sig. (2-tailed)			0,720
	N		7	7
Especificidade	Correlação de Pearson	de	-0,167	1
	Sig. (2-tailed)		0,720	
	N		7	7

Para avaliar a relação linear entre as métricas de sensibilidade e especificidade dos modelos de IA reportados nos estudos incluídos, foi realizada uma análise de correlação de Pearson, uma vez que são variáveis contínuas. O coeficiente de correlação obtido foi de $r = -0,167$, com um valor de significância bilateral de $p = 0,720$.

Estes resultados indicam uma correlação negativa muito fraca e estatisticamente não significativa entre as duas variáveis, dado que $p > 0,05$, o que não permite rejeitar a hipótese nula de ausência de correlação. Em termos práticos, isto sugere que os estudos analisados não apresentam uma tendência consistente entre os níveis de sensibilidade e especificidade, ou seja, modelos com elevada sensibilidade não demonstram necessariamente níveis proporcionais de especificidade e vice-versa.

A reduzida dimensão amostra ($N = 7$) poderá ter influenciado a ausência de significância estatística, limitando o poder estatístico da análise. Ainda assim, esta análise permite observar que não existe uma relação linear evidente entre estas métricas de desempenho no conjunto de modelos analisados.

A seguinte figura(15), representa um gráfico de dispersão com linha de tendência linear, para uma melhor visualização da relação entre sensibilidade e especificidade. Esta representação mostra a distribuição dos valores de cada estudo e evidencia a ausência de uma relação linear clara entre as duas métricas o que está de acordo com o resultado da correlação de Pearson obtida. A inclinação praticamente nula da linha de tendência reforça a inexistência de um padrão consistente entre a sensibilidade e a especificidade dos modelos de IA avaliados.

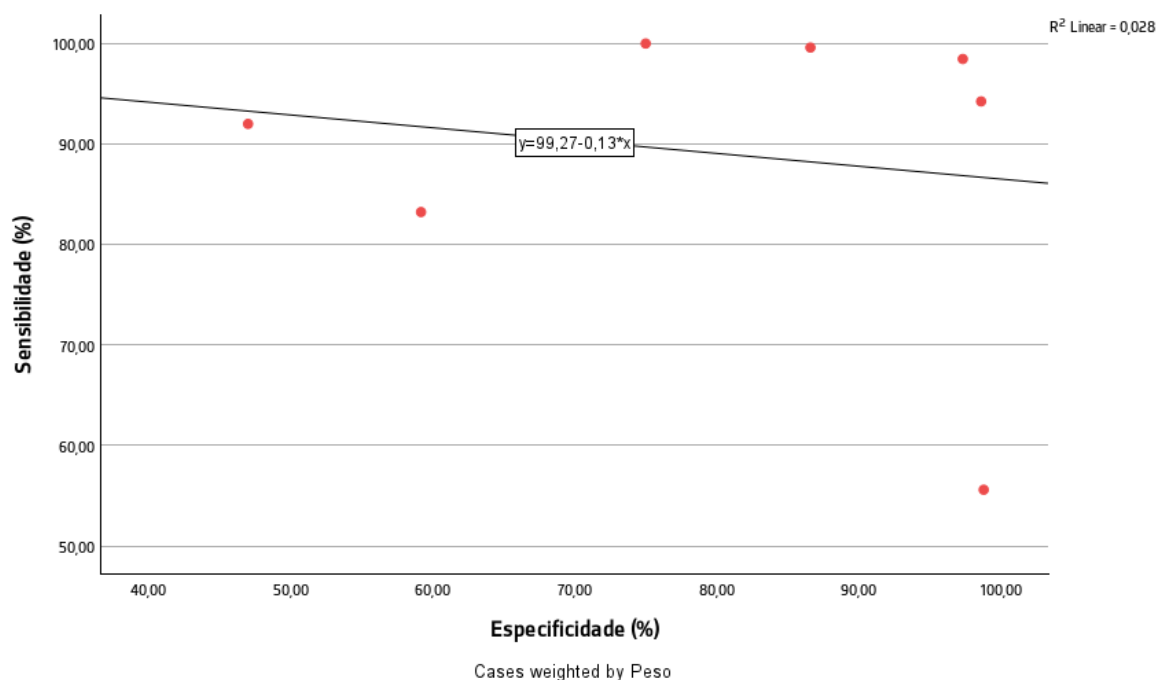


Figura 15 - Representação gráfica de dispersão com linha de tendência linear, entre sensibilidade e especificidade (fonte: SPSS v29.0.2.0)

4.2.2.1. Média ponderada e intervalo de confiança da Sensibilidade e Especificidade dos estudos incluídos

Para sintetizar a performance diagnóstica dos modelos de IA, foram também calculadas as médias simples da sensibilidade e especificidade dos estudos incluídos na presente análise estatística. Estas métricas são fundamentais na avaliação de testes diagnósticos: a sensibilidade expressa a capacidade do modelo em detetar corretamente os casos positivos, enquanto a especificidade representa a capacidade de identificar corretamente os casos negativos.

Ao contrário da AUC, cujas estimativas apresentavam variâncias muito reduzidas e alta homogeneidade entre os estudos, a sensibilidade e especificidade revelaram maior dispersão. Por esse motivo, optou-se por calcular os intervalos de confiança com base no desvio padrão ponderado e no número de estudos, utilizando uma abordagem mais conservadora.

Foram considerados os valores de sensibilidade e especificidade reportados em 7 estudos elegíveis. Para cada métrica foi calculado o erro padrão (SE) estimado de cada estudo com base nos respetivos intervalos de confiança 95%, utilizando a seguinte fórmula:

$$SE = \frac{IC_{sup} - IC_{inf}}{2 \times 1,96}$$

Com base no erro padrão, calculou-se a variância de cada estudo:

$$VAR = SE^2$$

Seguidamente, foi atribuído a cada estudo o peso inversamente proporcional à variância:

$$\text{Peso} = \frac{1}{\text{VAR}}$$

Estes pesos permitiram calcular a média ponderada da sensibilidade e da especificidade, reduzindo o impacto de estudos com menor precisão estatística. Os cálculos foram realizados no software SPSS versão 29.0.2.0, utilizando a função “Weight Cases” e o módulo de estatísticas descritivas.

A média ponderada sensibilidade foi de 0,9918, com um intervalo de confiança 95% de [0,916;1,000] e desvio padrão de 0,1027, indicando uma capacidade excecional dos modelos de IA para detetar corretamente os casos positivos.

A média ponderada de especificidade foi de 0,8720, com um intervalo de confiança 95% de [0,734;1,000] e um desvio padrão de 0,1860., refletindo uma boa capacidade para excluir corretamente os casos negativos.

Este intervalo de confiança foi calculado da seguinte forma:

$$\text{IC95\%} = \text{Média} \pm 1,96 \times \frac{SD}{\sqrt{k}}$$

Onde:

- SD representa o desvio padrão ponderado
- K o número de estudos incluídos.

Tabela 7 - IC95%, Erro padrão, Pesos e Variâncias da métrica sensibilidade (fonte: SPSS v29.0.2.0)

Estudos	Sensibilidad e	IC95% Inferior	IC95% Superior	Erro Padrão (SE)	Peso	Variância
Schelb P. et al., (2019)	0,913	0,732	0,976	0,6224 5	258,10	0,00387 4
Campanella G. et al., (2019)	1,000	0,987	1,000	0,0033 16	90925,4 4	0,000011
Strom P. et al., (2020)	0,998	0,993	0,999	0,0016 33	375156, 25	0,00000 3
Korevaar S. et al., (2021)	0,560	0,350	0,750	0,1020 41	96,04	0,010514
Hiremath A. et al., (2021)	0,892	0,845	0,933	0,0224 49	1984,30	0,00050 4
Sherafatmandjo o H. et al., (2024)	0,940	0,901	0,984	0,02117 3	2230,57	0,00044 8

Pantanowitz L. et al., (2020)	1,000	0,979	1,000	0,0053 57	34844, 44	0,00002 9
--------------------------------------	-------	-------	-------	--------------	--------------	--------------

Tabela 8 - IC95%, Erro padrão, Pesos e Variâncias da métrica especificidade (fonte: SPSS v29.0.2.0)

Estudos	Especificidade	IC95% Inferior	IC95% Superior	Erro Padrão (SE)	Peso	Variância
Schelb P. et al., (2019)	0,769	0,607	0,880	0,0696 43	206,18	0,004850
Campanella G. et al., (2019)	0,750	0,730	0,770	0,0102 04	9604,0 0	0,000104
Strom P. et al., (2020)	0,870	0,803	0,941	0,0352 04	806,69	0,001239
Korevaar S. et al., (2021)	0,990	0,930	1,000	0,0178 57	3136,0 0	0,000319
Hiremath A. et al., (2021)	0,670	0,600	0,733	0,0339 29	868,70 0,98	0,001151
Sherafatmandjo o H. et al., (2024)	0,980	0,971	1,000	0,0073 98	18271,5 8	0,000055
Pantanowitz L. et al., (2020)	0,970	0,960	0,982	0,0056 12	31748, 76	0,000031

A figura 16 e 17 representam o intervalo de confiança 95% relativamente à métrica sensibilidade e especificidade, respetivamente.

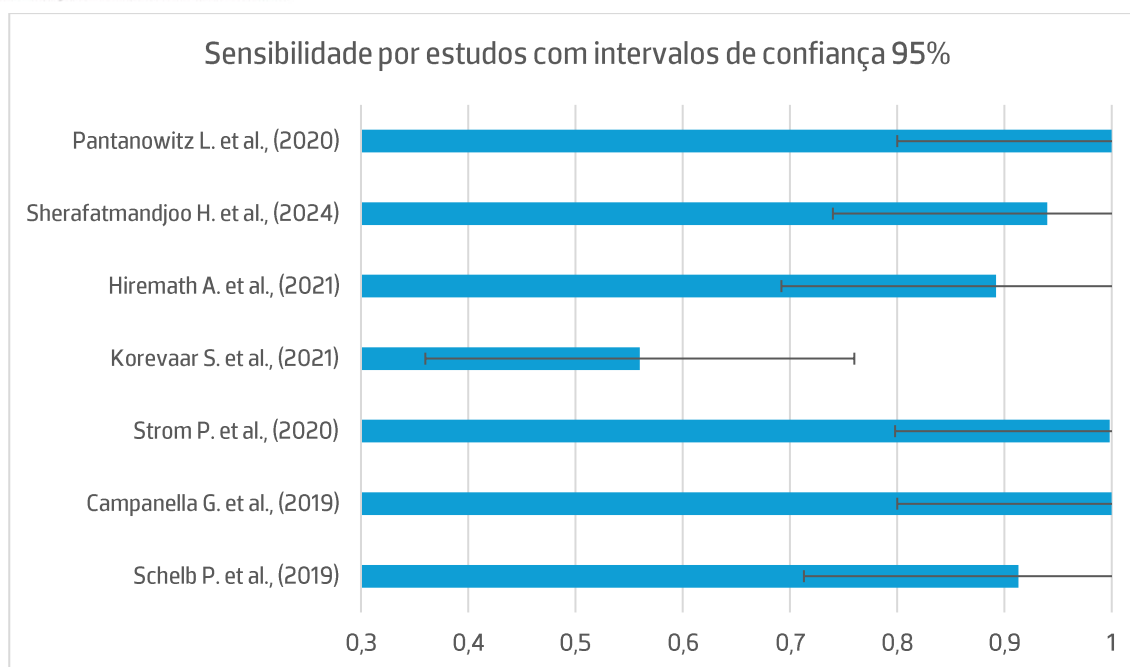


Figura 16 - Representação gráfica da métrica sensibilidade dos estudos incluídos na análise estatística, com barras de erro correspondentes a ± 2 erros padrão (SE), aproximando-se de um intervalo de confiança de 95% (fonte: Excel)

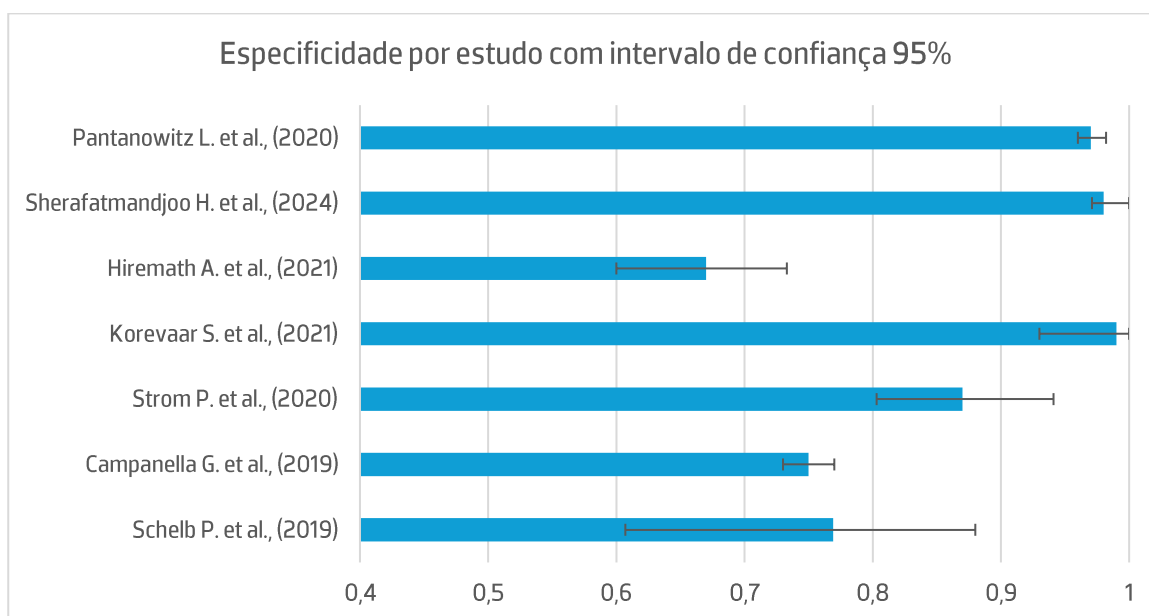


Figura 17 - Representação gráfica da métrica especificidade dos estudos incluídos na análise estatística, com barras de erro correspondentes a ± 2 erros padrão (SE), aproximando-se de um intervalo de confiança de 95% (fonte: Excel)

4.2.2.2. Heterogeneidade das métricas Sensibilidade e Especificidade

Para avaliar a consistência dos resultados entre os estudos incluídos, foi realizada uma análise de heterogeneidade separada para a sensibilidade e especificidade. O grau de heterogeneidade foi medido através do teste de Cochran (Q) e o índice I^2 , que estima a

percentagem de variação total entre os estudos que se deve à heterogeneidade real e não ao acaso.

Para avaliar a heterogeneidade da sensibilidade serão utilizadas diferentes fórmulas, tais como:

Cálculo do Q (estatística de Cochran):

$$Q = \sum Q_i = \sum_{i=1}^k w_i \times (\text{Sensibilidade}_i - \text{Sensibilidade ponderada})^2$$

Onde:

- Q é a estatística total de heterogeneidade
- $W_i = \frac{1}{\text{variância}_i}$
- Sensibilidade_i = valor da Sensibilidade do estudo
- Sensibilidade ponderada = média ponderada (0,9918)

Tabela 9 - Valores da heterogeneidade da métrica sensibilidade

Estudo	Sensibilidade _i	Média (Sensibilidade ponderada)	(Sensibilidade _i – Sensibilidade ponderada) ²	Q _i
Schelb P. et al., (2019)	0,99	0,9918	0,00000324	0,84
Campanella G. et al., (2019)	1,00	0,9918	0,00006724	6,11
Strom P. et al., (2020)	0,99	0,9918	0,00000324	1,22
Korevaar S. et al., (2021)	0,56	0,9918	0,18735364	17,99
Hiremath A. et al., (2021)	0,89	0,9918	0,01040864	20,66
Sherafatmandjoo H. et al., (2024)	0,94	0,9918	0,00270564	6,03
Pantanowitz L. et al., (2020)	1,00	0,9918	0,00006724	2,34

Consoante os resultados apresentados na tabela 5, podemos calcular o valor de Q.

$$Q = \sum Q_i = 55,19$$

Posteriormente, serão calculados os graus de liberdade (df). De acordo com Fisher R. (1934), os graus de liberdade referem-se ao número de valores independentes que podem variar num cálculo estatístico depois de se aplicar uma restrição (por exemplo a média). Fórmula:

$$df = k - 1$$

Onde:

- K = número de estudos incluídos

Neste caso, df = 6 (sabendo que são usados 7 estudos)

O índice de heterogeneidade I^2 , proposto por Higgins et al. (2003), quantifica a proporção da variabilidade entre estudos que é atribuída a heterogeneidade real e não ao erro amostral. É calculado com base na estatística Q de Cochran e nos graus de liberdade e assume valores entre 0% (sem heterogeneidade) e 100% (heterogeneidade total). É calculado pela seguinte fórmula:

$$I^2 = \max\left(0, \frac{Q - df}{Q}\right) \times 100$$

Onde:

- Q = 55,19
- Df = 6

Desta forma, $I^2 = 89,13\%$. Este valor sugere que aproximadamente 89% da variabilidade observada entre os estudos deve-se a diferenças reais nos efeitos estimados (ex. características populacionais, algoritmos de IA, modalidades de diagnóstico) e não apenas ao erro aleatório.

Assim, justifica-se a adoção de um modelo de efeitos aleatórios para a síntese dos dados de sensibilidade, de modo a acomodar a variabilidade real existente e obter estimativas mais robustas e gerais.

Para avaliar a heterogeneidade da especificidade serão utilizadas diferentes fórmulas, tais como:

Cálculo do Q (estatística de Cochran):

$$Q = \sum Qi = \sum_{i=1}^k wi \times (\text{Especificidade}_i - \text{Especificidade ponderada})^2$$

Onde:

- Q é a estatística total de heterogeneidade
- $Wi = \frac{1}{\text{variância}_i}$
- Especificidade_i = valor da Especificidade do estudo

- Especificidade ponderada = média ponderada (0,8720)

Tabela 10 – Valores da heterogeneidade da métrica especificidade

Estudo	Especificidade _i	Média (Especificidade ponderada)	(Especificidade _i – Especificidade ponderada) ²	Q _i
Schelb P. et al., (2019)	0,99	0,8720	0,0139240	2,87
Campanella G. et al., (2019)	0,75	0,8720	0,0148840	142,95
Strom P. et al., (2020)	0,87	0,8720	0,0000040	0,00
Korevaar S. et al., (2021)	0,99	0,8720	0,939240	43,67
Hiremath A. et al., (2021)	0,67	0,8720	0,0408040	35446,47
Sherafatmandjoo H. et al., (2024)	0,98	0,8720	0,0116640	213,12
Pantanowitz L. et al., (2020)	0,97	0,8720	0,0096040	304,92

Consoante os resultados apresentados na tabela 5, podemos calcular o valor de Q.

$$Q = \sum Q_i = 36672,04$$

Posteriormente, serão calculados os graus de liberdade (df). De acordo com Fisher R. (1934), os graus de liberdade referem-se ao número de valores independentes que podem variar num cálculo estatístico depois de se aplicar uma restrição (por exemplo a média). Fórmula:

$$df = k - 1$$

Onde:

- K = número de estudos incluídos

Neste caso, df = 6 (sabendo que são usados 7 estudos)

O índice de heterogeneidade I², proposto por Higgins et al. (2003), quantifica a proporção da variabilidade entre estudos que é atribuída a heterogeneidade real e não ao erro amostral. É calculado com base na estatística Q de Cochran e nos graus de liberdade e assume valores

entre 0% (sem heterogeneidade) e 100% (heterogeneidade total). É calculado pela seguinte fórmula:

$$I^2 = \max\left(0, \frac{Q - df}{Q}\right) \times 100$$

Onde:

- $Q = 36672,04$
- $Df = 6$

Desta forma, $I^2 = 98,99\%$. Este valor indica uma heterogeneidade extremamente elevada entre os estudos relativamente à especificidade, o que demonstra que a variabilidade observada não pode ser atribuída apenas ao erro aleatório. Assim, justifica-se de forma clara a aplicação de um modelo de efeitos aleatórios, que permite acomodar as diferenças metodológicas e clínicas reais existentes entre os estudos.

4.2.3. Análise de Métrica: Índice de Kappa

O índice de Kappa (K) foi utilizado em diversos estudos incluídos nesta dissertação com o objetivo de avaliar o grau de concordância entre os sistemas de IA e o diagnóstico de referência. Ao contrário das métricas sensibilidade e especificidade, o índice de Kappa tem a particularidade de corrigir a concordância observada tendo em conta a concordância esperada ao acaso.

De acordo com a escala proposta por Landis e Koch (1977), os valores de K podem ser interpretados da seguinte forma:

- $K \leq 0$ – concordância inferior ao acaso
- $0,00 < K < 0,20$ – concordância ligeira
- $0,21 - 0,40$ – concordância razoável
- $0,41 - 0,60$ – concordância moderada
- $0,61 - 0,80$ – concordância substancial
- $0,81 - 1$ – concordância quase perfeita

Nos estudos analisados o índice de Kappa foi reportado em 6 estudos, com valores que variam entre 0,63 e 0,89. Esta métrica revelou-se útil para quantificar o grau de acordo entre os outputs dos algoritmos de IA e o diagnóstico final, especialmente em contextos de imagem médica e interpretação histológica.

Tabela 11 – Valores do Índice de Kappa dos estudos incluídos

Estudo	Comparação	Tipo de Kappa	Valor de Kappa	Interpretação de concordância
Schelb P. et al. (2019)	IA vs Gold Standard	Simple	0,89	Quase perfeita
Campanella G. et al., (2019)	IA vs Diagnóstico Clínico	Simple	0,87	Quase perfeita
Strom P. et al., (2020)	IA vs Diagnóstico Clínico	Ponderado	0,86	Quase perfeita
Korevaar S. et al., (2021)	IA vs Diagnóstico Clínico	Simple	0,63	Substancial
Hiremath A. et al., (2021)	Radiologistas com IA (interobservador)	Ponderado	0,74	Substancial
Pantanowitz L. et al., (2020)	IA vs Diagnóstico Clínico	Simple	0,86	Quase perfeita

Os estudos incluídos na presente análise estatística utilizaram diferentes versões do Índice de Kappa para avaliar a concordância entre os sistemas de IA e os métodos de referência. De acordo com a natureza das tarefas diagnósticas e dos dados em análise, foram aplicadas duas variants principais: Índice de Kappa Simple (adequado para classificações binárias) e o Índice de Kappa Ponderado (adequado para escalas ordinais).

O Índice de Kappa simple, foi utilizados em estudos nos quais a classificação baseava-se em categorias dicotómicas, nomeadamente presença ou ausência de cancro da próstata.

Os valores reportados foram elevados, indicando uma concordância significative entre a IA e o Gold Standard. Apesar de todos os estudos utilizarem o mesmo tipo de Kappa, a variabilidade nos resultados pode estar relacionada com a experiência dos observadores, o tipo de imagem e a complexidade da tarefa diagnóstica.

Nos estudos que envolveram escalas ordinais para avaliação do risco ou grau de malignidade, foi utilizado o Índice de Kappa ponderado, que penaliza as discordâncias consoante a sua gravidade. Nestes casos, os resultados também evidenciaram níveis elevados de concordância. É importante destacar que no caso de Hiremath A. et al., (2021), a comparação não foi entre IA e gold standard, mas sim entre observadores humanos que utilizaram a IA

como ferramenta de suporte. Assim, o valor reportado representa a concordância interobservador com o uso de IA, refletindo a capacidade da tecnologia em aumentar a uniformidade entre os profissionais.

De forma complementar, calculou-se a média simples dos valores de Kappa reportados nos estudos incluídos. A média dos valores de Kappa simples foi de 0,81, e dos Kappa ponderados foi 0,80. Estes valores, apesar de não terem validade inferencial, permitem observar uma tendência geral para a concordância substancial a quase perfeita entre os métodos analisados, segundo Landis & Koch (1977).

Estas médias descritivas foram calculadas através da seguinte fórmula:

$$\text{Média} = \frac{\sum_{i=1}^n x_i}{n}$$

Onde:

- X_i = valor individual do conjunto de dados
- n = número total de estudos.

Apesar de existirem fórmulas para o cálculo do desvio padrão e do intervalo de confiança do Índice de Kappa, a sua aplicação requer o acesso a dados de concordância individuais, como a matriz de classificação entre métodos. Como essa informação não foi disponibilizada na maioria dos estudos incluídos, optou-se por não proceder ao cálculo estatístico desses parâmetros, limitando-se esta análise aos valores reportados pelos autores.

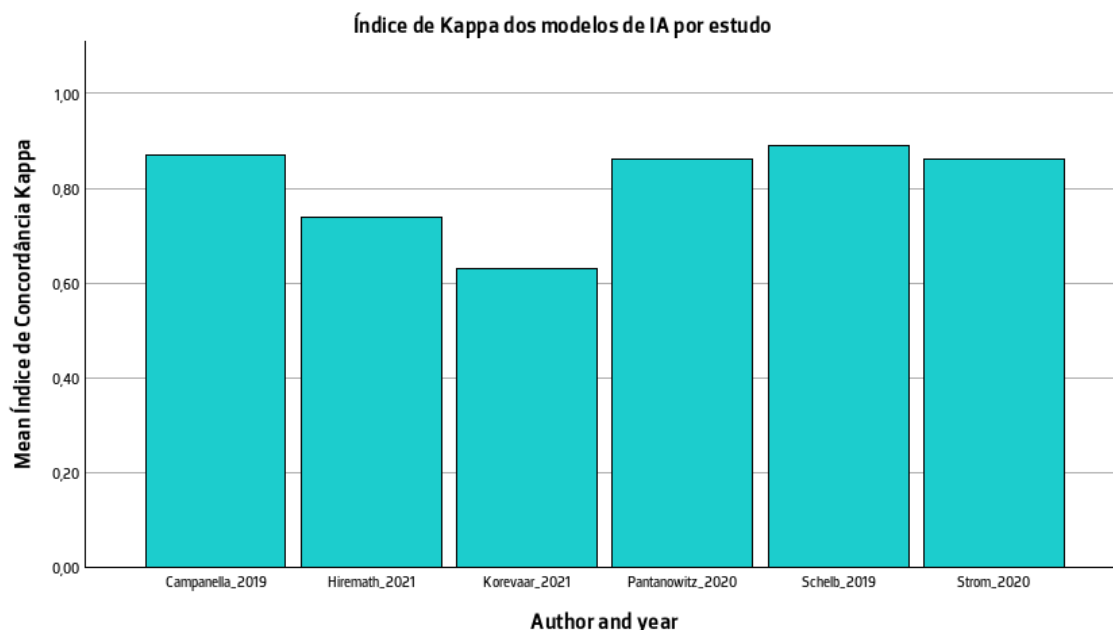


Figura 18 - Representação gráfica dos valores de índice de Kappa obtidos nos estudos (fonte: SPSS v29.0.2.0)

4.2.4. Análise de métrica: Acurácia

A acurácia é uma métrica contínua que reflete a proporção de todas as classificações corretas efetuadas por um teste diagnóstico, incluindo tanto verdadeiros positivos como verdadeiros negativos (Santini et al., 2021).

Apesar da acurácia ser usada frequentemente como métrica de desempenho global, apresenta limitações importantes. Deeks & Altman (2004) salientam que testes com elevada acurácia podem ser clinicamente pouco úteis quando a doença é muito rara ou muito comum, pois a acurácia é influenciada pela prevalência e pode ocultar uma fraca capacidade discriminativa do modelo.

Tabela 12 – Valores de acurácia dos estudos incluídos

Estudo	Comparação	Acurácia (%)
Schelb P. et al., (2019)	IA vs Gold Standard	76
Campanella G. et al., (2019)	IA vs Diagnóstico Clínico	91–99
Sherafatmandjoo H. et al., (2024)	IA vs Gold Standard	85–96
Yoo S. et al., (2021)	IA vs Diagnóstico Clínico	91–95

Schelb P. et al. (2019), reportaram uma acurácia de 76% para um modelo de DL baseado na arquitetura U-Net. Campanella et al. (2019), num estudo de larga escala com mais de 44000 lâminas histológicas, obtiveram valores de acurácia entre 91% e 99%, avaliando a concordância entre a IA e os diagnósticos de patologistas de referência.

Sherafatmandjoo et al. (2024) verificaram um aumento substancial da acurácia (85% para 96%) quando o algoritmo integrava dados clínicos e patológicos, o que reforça a utilidade da IA quando complementada por variáveis contextuais. De forma semelhante, Yoo et al. (2021) obtiveram valores de acurácia entre 91,2% e 95,3%, com o melhor desempenho registado num modelo do tipo EfficientNet-B0.

Estes valores, embora limitados a apenas quatro estudos, sugerem um desempenho global elevado da IA na tarefa de classificação de casos oncológicos da próstata, quando comparada com métodos de referência. A ausência desta métrica nos restantes estudos pode dever-se à

ênfase na utilização de medidas como a AUC, sensibilidade e especificidade, mais comuns em contextos de avaliação diagnóstica.

5. Discussão de resultados da análise estatística

A presente análise estatística procurou sintetizar, os resultados dos estudos que aplicaram IA ao diagnóstico do cancro da próstata, comparando o desempenho dos sistemas automáticos com o de especialistas humanos e métodos clínicos de referência. Foram incluídos 8 estudos, selecionados com base na disponibilidade de métricas padronizadas como sensibilidade, especificidade, AUC (área sob a curva ROC) e coeficientes de concordância (Kappa), o que permitiu uma análise comparativa robusta.

De forma global, os resultados confirmaram o elevado desempenho da IA no apoio ao diagnóstico, com valores de AUC frequentemente superiores a 0,90, indicando excelente capacidade discriminativa. Em particular, algoritmos treinados em imagens histológicas de biópsias ou em exames de imagem médica (como mpMRI) demonstraram uma acuidade diagnóstica equiparável ou superior à dos métodos tradicionais. Estes dados não só reforçam a fiabilidade técnica da IA, como também sugerem que o seu uso poderá contribuir para diagnósticos mais precoces e decisões clínicas melhor fundamentadas.

A análise dos coeficientes de concordância Kappa, utilizados para avaliar a consistência entre observadores ou entre métodos, revelou valores especialmente promissores. Em vários estudos, a aplicação de IA conduziu a um Kappa $\geq 0,80$, o que, segundo a literatura, é interpretado como uma concordância quase perfeita. Este nível de acordo tem implicações práticas significativas, sobretudo em contextos clínicos onde a variabilidade interobservador pode comprometer a uniformidade do diagnóstico, como acontece frequentemente na avaliação do score de Gleason.

Um aspeto metodológico importante a considerar nesta análise estatística é o paradoxo do Kappa, que pode afetar a interpretação dos resultados de concordância. Este fenómeno ocorre quando há uma distribuição desigual das categorias (por exemplo, grande predominância de diagnósticos negativos), o que tende a reduzir artificialmente os valores do coeficiente Kappa, mesmo quando existe elevada concordância observada. Alguns dos estudos incluídos na análise estatística apresentaram distribuições assimétricas que, embora não comprometam os resultados, exigem uma leitura crítica. A utilização de IA pode ajudar a contornar este desafio, ao aplicar critérios uniformes e objetivos de avaliação, o que pode contribuir para uma

melhoria na fiabilidade dos diagnósticos, especialmente em cenários onde a subjetividade humana tem impacto significativo.

Outro aspeto digno de nota é o contributo da IA para a redução da variabilidade interobservador, sobretudo entre profissionais com níveis de experiência distintos. Em vários estudos, a introdução de IA como apoio à decisão aumentou a concordância entre observadores, o que pode representar uma vantagem significativa na prática clínica. Este reforço da uniformização diagnóstica poderá ter um papel importante na melhoria da qualidade dos cuidados, na redução de erros e na maior previsibilidade dos tratamentos recomendados.

Ainda assim, a análise estatística revelou alguma heterogeneidade entre os estudos incluídos, tanto a nível de métodos utilizados como na população analisada e nas definições operacionais das métricas avaliadas. Esta diversidade metodológica constitui uma limitação à comparabilidade direta dos resultados, mas reflete também a versatilidade da IA em diferentes contextos e fases do processo diagnóstico. Importa referir que a maioria dos estudos incluídos tem natureza observacional ou retrospectiva, o que poderá introduzir um viés de seleção ou limitar a validade externa dos resultados. A escassez de ensaios clínicos prospetivos ou randomizados neste campo é uma lacuna ainda por colmatar.

Do ponto de vista da aplicabilidade, os resultados da análise estatística demonstraram que a IA pode desempenhar um papel crucial na otimização do diagnóstico do cancro da próstata. Contudo, a sua adoção generalizada dependerá de fatores como a validação contínua dos modelos, a sua integração com sistemas informáticos hospitalares e a confiança por parte dos profissionais de saúde. Além disso, será necessário garantir que os algoritmos utilizados são transparentes, auditáveis e treinados com dados de qualidade, para que não se tornem fontes de viés ou desigualdade no acesso aos cuidados.

Em suma, os dados desta análise estatística permitem uma leitura cautelosamente otimista quanto ao potencial da IA no reforço da precisão diagnóstica e da consistência entre observadores. Contudo, a evidência disponível ainda carece de amadurecimento metodológico e de maior uniformização entre estudos para que seja possível fazer recomendações mais definitivas.

6. Conclusão

A presente dissertação teve como objetivo responder à seguinte questão: “Em homens com risco de cancro da próstata, a utilização da IA otimiza os métodos já existentes para o diagnóstico?”. Com base na evidência reunida através de uma revisão sistemática da literatura e de uma análise estatística, conclui-se que a IA apresenta uma capacidade real de complementar os métodos diagnósticos tradicionais, oferecendo ganhos em precisão, padronização e eficiência.

A revisão sistemática identificou múltiplos modelos de IA, maioritariamente baseados em redes neurais convolucionais, com desempenho elevado na deteção e classificação de lesões prostáticas. Particular destaque foi dado aos algoritmos treinados com grandes volumes de imagens histológicas, que demonstraram resultados equiparáveis – e por vezes superiores – aos de patologistas experientes, especialmente na identificação precoce de adenocarcinomas. A análise estatística reforçou estas evidências, revelando uma média ponderada de sensibilidade de 99,18% e especificidade de 87,20%, bem como um índice médio de Kappa de 0,80, traduzindo concordância quase perfeita entre os sistemas de IA e os padrões de referência.

Ainda assim, a elevada heterogeneidade entre os estudos, a diversidade metodológica e a escassez de validação externa constituem limitações importantes, que dificultam a generalização dos resultados. A falta de padronização no desenvolvimento, treino e aplicação dos modelos de IA continua a ser um obstáculo à sua adoção generalizada nos serviços de saúde.

Do ponto de vista da gestão em saúde, este trabalho sublinha a importância de integrar a IA de forma gradual, ética e segura, valorizando-a como uma ferramenta de apoio à decisão clínica – e nunca como substituto do raciocínio médico. Para isso, será essencial investir na formação dos profissionais, na definição de protocolos baseados em evidência e na criação de estruturas técnicas e regulatórias que sustentem uma implementação responsável.

Como linhas futuras de investigação, recomenda-se o desenvolvimento de ensaios prospetivos multicêntricos com validação externa robusta, maior uniformização das métricas de avaliação e a definição de guidelines internacionais específicas para o uso da IA em oncologia. Paralelamente, o reforço da literacia digital entre os profissionais de saúde será determinante para garantir uma adoção segura, eficaz e sustentável desta tecnologia. O caminho da IA na medicina está ainda a ser traçado, mas os dados disponíveis mostram que o

seu contributo no diagnóstico oncológico é não só possível, como desejável, desde que feito com prudência, ciência e sentido humano.

Referências Bibliográficas

Accenture. (2023). Reinventing care delivery to solve clinician shortage. *Accenture*.
<https://www.accenture.com/us-en/insights/consulting/total-enterprise-reinvention>

Agência dos Direitos Fundamentais da União Europeia (FRA). (2021). Preparar o futuro: Inteligência artificial e direitos fundamentais. *Serviço das Publicações da União Europeia*.
<https://doi.org/10.2811/990836>

Altman, D. G., & Bland, J. M. (1994). *Diagnostic tests 1: Sensitivity and specificity*. *BMJ*, 308(6943), 1552. <https://doi.org/10.1136/bmj.308.6943.1552>

American Cancer Society. (2020). *Fatores de risco para o câncer de próstata*.
<https://www.cancer.org/es/cancer/tipos/cancer-de-prostata/causas-riesgos-prevencion/factores-de-riesgo.html>

Associação Portuguesa de Administradores Hospitalares (APAH) & EY. (2022). Que prioridades para a Saúde em Portugal? O que os principais agentes do setor da saúde esperam dos programas eleitorais que irão a votos nas próximas eleições legislativas. *EY*.
https://assets.ey.com/content/dam/ey-sites/ey-com/pt_pt/topics/health/pdf/ey-estudo-ey-apah-sns.pdf

Bland, M. (2000). *An Introduction to Medical Statistics* (3.^a ed.). *Oxford University Press*. ISBN: 0-19-263269-8

Bonita, R., Beaglehole, R., & Kjellström, T. (2006). *Basic epidemiology* (2^a ed., p. 3). *World Health Organization*. ISBN 9241547073.

Bosma, J. S., Saha, A., Hosseinzadeh, M., Slootweg, I., de Rooij, M., & Huisman, H. (2023). Semisupervised learning with report-guided pseudo labels for deep learning-based prostate cancer detection using biparametric MRI. *Radiology: Artificial Intelligence*, 5(5), e230031.
<https://doi.org/10.1148/ryai.230031>

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 68(6), 394-424.
<https://doi.org/10.3322/caac.21492>

Bulten, W., Kartasalo, K., Chen, P. H. C., Ström, P., Pinckaers, H., Nagpal, K., ... & Eklund, M. (2022). Artificial intelligence for diagnosis and Gleason grading of prostate cancer: The

PANDA challenge. *Nature Medicine*, 28(1), 154–163. <https://doi.org/10.1038/s41591-021-01620-2>

Byrt, T., Bishop, J., & Carlin, J. B. (1993). Bias, prevalence and kappa. *Journal of Clinical Epidemiology*, 46(5), 423–429. [https://doi.org/10.1016/0895-4356\(93\)90018-V](https://doi.org/10.1016/0895-4356(93)90018-V)

Campanella, G., Hanna, M. G., Geneslaw, L., Miraflor, A., Werneck Krauss Silva, V., Busam, K. J., ... & Fuchs, T. J. (2019). Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature Medicine*, 25(8), 1301–1309. <https://doi.org/10.1038/s41591-019-0508-1>

Cansado, M. da P. (2024). Inteligência artificial no setor da saúde: Desafios jurídicos e regulação. *Gabinete de Estratégia e Estudos do Ministério da Economia*. <https://www.gee.gov.pt/pt/estudos-e-seminarios/gee-papers-category/33911-inteligencia-artificial-no-setor-da-saude-desafios-juridicos-e-regulacao>

Ciccia, A., & Elledge, S. J. (2015). DNA repair and cancer development: When erroneous DNA repair leads to mutations or chromosomal aberrations affecting oncogenes or inactivating tumor suppressor genes, cells undergo malignant transformation resulting in cancerous growth. *Frontiers in Genetics*, 6, 157. <https://doi.org/10.3389/fgene.2015.00157>

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics*, 10(1), 101–129.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>

Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4), 213–220. <https://doi.org/10.1037/h0026256>

Comissão Europeia. (s.d.). European Cancer Information System Explorer. [https://ecis.jrc.ec.europa.eu/explorer.php?\\$0-4\\$1-AE28E\\$4-1\\$3-All\\$6-0,85\\$5-2022,2030\\$7-7\\$21-0\\$CLongtermChart2_4\\$X0_14-](https://ecis.jrc.ec.europa.eu/explorer.php?$0-4$1-AE28E$4-1$3-All$6-0,85$5-2022,2030$7-7$21-0$CLongtermChart2_4$X0_14-)

Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. <https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/>

Deeks, J. J., & Altman, D. G. (2004). Diagnostic tests 4: Likelihood ratios. *BMJ*, 329(7458), 168–169. <https://doi.org/10.1136/bmj.329.7458.168>

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., ... & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25(1), 24–29. <https://doi.org/10.1038/s41591-018-0316-z>

European Association of Urology. (2024). EAU guidelines on prostate cancer. <https://uroweb.org/guidelines/prostate-cancer>

European Commission, Joint Research Centre. (2020). Cancer incidence and mortality estimates – EU-27, 2020. In *The European cancer burden in 2020. ECIS JRC*. <https://publications.jrc.ec.europa.eu/repository/handle/JRC125850>

European Commission, Joint Research Centre. (2022). Cancer estimates 2022 factsheet. *European Commission*. https://ecis.jrc.ec.europa.eu/sites/default/files/2024-01/jrc_CancerEstimates2022_factsheet.pdf

European Commission, Joint Research Centre. (2022). Cancer incidence and mortality estimates – ECIS Explorer. *European Cancer Information System*. <https://ecis.jrc.ec.europa.eu/explorer.php>

Fernandes, M. S., & Goldim, J. R. (2024). Artificial intelligence and decision making in health: Risks and opportunities. In H. S. Antunes, P. M. Freitas, A. L. Oliveira, C. M. Pereira, E. V. de Sequeira, & L. B. Xavier (Eds.), *Multidisciplinary perspectives on artificial intelligence and the law* (Vol. 58, pp. 187–205). *Springer*. https://doi.org/10.1007/978-3-031-41264-6_10

Fisher, R. A. (1934). *Statistical methods for research workers* (5^a ed., pp. 263–266). Oliver & Boyd. <https://archive.org/details/in.ernet.dli.2015.205971>

Fisher, R. A. (1928). *Statistical methods for research workers* (2nd ed., pp. 263–266). Oliver and Boyd. <https://archive.org/details/statisticalmethoe2fish>

Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical methods for rates and proportions* (3.^a ed.). John Wiley & Sons. <https://books.google.pt/books?id=9Vef07a8GeAC>

Fundação Champalimaud. (n.d.). Check-up #12: Estadiamento do cancro. <https://www.fchampalimaud.org/pt-pt/news/check-up-12-estadiamento-do-cancro>

Gallagher Ford, L., & Melnyk, B. M. (2019). The underappreciated and misunderstood PICOT question: A critical step in the EBP process. *Worldviews on Evidence-Based Nursing*, 16(6), 422–423. <https://doi.org/10.1111/wvn.12408>

Gorelick, N., Ayache, N., Vakalopoulou, M., & Paragios, N. (2019). Prostate cancer detection and grading using multimodal deep learning techniques. In *2019 IEEE 16th*

International Symposium on Biomedical Imaging (ISBI 2019) (pp. 1709–1713). *IEEE*.
<https://doi.org/10.1109/ISBI.2019.8759468>

Grönberg, H., Adolfsson, J., Aly, M., Nordström, T., Wiklund, P., & Eklund, M. (2015). Prostate cancer diagnostics 2015: state of the art and future perspectives. *The Lancet Oncology*, 16(e85–e92). [https://doi.org/10.1016/S1470-2045\(15\)00361-7](https://doi.org/10.1016/S1470-2045(15)00361-7)

Grupo de Peritos de Alto Nível sobre a Inteligência Artificial. (2019). Uma definição de IA: Principais capacidades e disciplinas científicas. *Comissão Europeia*.
<https://ec.europa.eu/digital-strategy/en/high-level-expert-group-artificial-intelligence>

Hamet, P., & Tremblay, J. (2017). Artificial intelligence in medicine. *Metabolism: Clinical and Experimental*, 69, S36–S40. <https://doi.org/10.1016/j.metabol.2017.01.011>

Harmon, S. A., Tuncer, S., Sanford, T., Choyke, P. L., & Türkbey, B. (2019). Artificial intelligence at the intersection of pathology and radiology in prostate cancer. *Diagnostic and Interventional Radiology*, 25(3), 183–188. <https://doi.org/10.5152/dir.2019.19125>

Hemminki, K. (2012). Familial risk and familial survival in prostate cancer. *World Journal of Urology*, 30(2), 143–148. <https://doi.org/10.1007/s00345-011-0801-1>

HFW. (2025). Legal liability for AI-driven decisions – When AI gets it wrong, who can you turn to? HFW. <https://www.hfw.com/insights/legal-liability-for-ai-driven-decisions-when-ai-gets-it-wrong-who-can-you-turn-to/>

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, 327(7414), 557–560.
<https://doi.org/10.1136/bmj.327.7414.557>

Hiremath, A., Shiradkar, R., Fu, P., Mahran, A., Rastinehad, A. R., Tewari, A., Tirumani, S. H., Purysko, A., Ponsky, L., & Madabhushi, A. (2021). An integrated nomogram combining deep learning, Prostate Imaging–Reporting and Data System (PI-RADS) scoring, and clinical variables for identification of clinically significant prostate cancer on biparametric MRI: A retrospective multicentre study. *The Lancet Digital Health*, 3(7), e445–e454.
[https://doi.org/10.1016/S2589-7500\(21\)00082-0](https://doi.org/10.1016/S2589-7500(21)00082-0)

Humphrey, P. A. (2004). Gleason grading and prognostic factors in carcinoma of the prostate. *Modern Pathology*, 17(3), 292–306. <https://doi.org/10.1038/modpathol.3800054>

IBM Corp. (2023). IBM SPSS Statistics for Windows (Versão 29.0.2.0) [Software]. *IBM Corp*. <https://www.ibm.com/products/spss-statistics>

Instituto da Próstata. (2023). Tratamentos para cancro da próstata. <https://www.institutodaprostata.com/pt/problemas/cancro-da-prostata>

International Agency for Research on Cancer. (2020). World Fact Sheet. <https://gco.iarc.who.int/media/globocan/factsheets/populations/900-world-fact-sheet.pdf>

International Agency for Research on Cancer (2024). Global Cancer Observatory: Cancer Today. *International Agency for Research on Cancer*. https://gco.iarc.fr/today/en/dataviz/bars-compare-populations?mode=cancer&sexes=1&cancers=27&multiple_populations=1&populations=900&group_populations=1&sort_by=value0&key=total

International Agency for Research on Cancer (IARC). (s.d). Global Cancer Observatory: Cancer Tomorrow. *International Agency for Research on Cancer*. https://gco.iarc.fr/tomorrow/en/dataviz/bars?sexes=1&key=percent&show_bar_mode_prop=0&populations=8_40_56_70_100_112_191_196_203_208_233_246_250_276_300_348_352_372_380_428_440_442_470_498_499_528_578_616_620_642_643_688_703_705_724_752_756_804_807_826&cancers=27

Joint Research Centre. (n.d.). European Cancer Information System (ECIS). *European Commission*. <https://ecis.jrc.ec.europa.eu/en->

Joint Research Centre. (n.d.). Country cancer profiles. *European Commission*. <https://cancer-inequalities.jrc.ec.europa.eu/country-cancer-profiles>

Korevaar, S., Tennakoon, R., Page, M., Brotchie, P., Thangarajah, J., Florescu, C., Sutherland, T., Mao Kam, N., & Bab-Hadiashar, A. (2021). Incidental detection of prostate cancer with computed tomography scans. *Scientific Reports*, 11(1), 7956. <https://doi.org/10.1038/s41598-021-86972-y>

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>

Last, J. M. (Ed.). (2001). Epidemiology. In *A dictionary of epidemiology* (4th ed., p. 61). *Oxford University Press*.

Li, J., Ettl, M., Amin, A., Bhalla, R., Das, K., Deng, F.-M., Lee, P., Matoso, A., Melamed, J., Mendrinós, S., Tian, W., Yaskiv, O., Shah, R. B., & Zhou, M. (2023). Interobserver reproducibility of quantifying Gleason pattern 4 cancer in prostate biopsy: Implications for clinical practice.

<https://doi.org/10.14218/JCTP.2022.00026>

Lin, L. I.-K. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, 45(1), 255–268. <https://doi.org/10.2307/2532051>

Lopes, C. M. N., & Mendes, J. C. (2023). Ética e inteligência artificial: Desafios e melhores práticas. *Revista da UFMG*, 30, Artigo e47673. <https://doi.org/10.35699/2965-6931.2023.47673>

Machado, L. S. (2019). Médico robô: responsabilidade civil por danos praticados por atos autônomos de sistemas informáticos dotados de inteligência artificial. *Lex Medicinae*, 16(31–32), 101–114. <https://www.centrodedireitobiomedico.org/data/publicacoes/PUB2024826192049.pdf>

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology*, 5(9), 1315–1316. <https://doi.org/10.1097/JTO.0b013e3181ec173d>

McKenzie, J. E., Brennan, S. E., Ryan, R. E., Thomson, H. J., Johnston, R. V., & Thomas, J. (2012). Chapter 3: Defining the criteria for including studies and how they will be grouped for the synthesis. In J. P. T. Higgins & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions* (Version 5.1.0). *Systematic Reviews*, 2(1), 80. <https://doi.org/10.1186/2046-4053-2-80>

Jung, M., Jin, M.-S., Kim, C., Lee, C., Nikas, I. P., Park, J. H., & Ryu, H. S. (2022). Artificial intelligence system shows performance at the level of uropathologists for the detection and grading of prostate cancer in core needle biopsy: An independent external validation study. *Modern Pathology*, 35(7), 1449–1457. <https://doi.org/10.1038/s41379-022-01077-9>

Mottet, N., van den Bergh, R. C. N., Briers, E., Van den Broeck, T., Cumberbatch, M. G., De Santis, M., ... & Cornford, P. (2021). EAU-EANM-ESTRO-ESUR-SIOG Guidelines on Prostate Cancer 2021. *European Urology*, 79(2), 243–262. <https://doi.org/10.1016/j.eururo.2020.09.046>

Mun, Y., Paik, I., Shin, S.-J., Kwak, T.-Y., & Chang, H. (2021). Yet another automated Gleason grading system (YAAGGS) by weakly supervised deep learning. *NPJ Digital Medicine*, 4, Article 104. <https://doi.org/10.1038/s41746-021-00469-6>

Munn, Z., McArthur, A., Klugar, M., Stannard, D., Cooper, A. S., Enuameh, Y., & Porritt, K. (2023). Providing a scaffold for considering theoretical frameworks in evidence

implementation projects: The JBI approach to evidence implementation. *JBI Evidence Implementation*, 21(4), 386–393. <https://doi.org/10.1097/XEB.0000000000000397>

Müller, R., & Büttner, P. (1994). A critical discussion of intraclass correlation coefficients. *Statistics in Medicine*, 13(23-24), 2465–2476. <https://doi.org/10.1002/sim.4780132310>

Nagendran, M., Chen, Y., Lovejoy, C. A., Carson, G., Berry, K., Chowkwanyun, M., ... & Maruthappu, M. (2020). Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *BMJ*, 368, m689. <https://doi.org/10.1136/bmj.m689>

National Cancer Institute. (2019). Prostate cancer screening (PDQ®) – Patient version. <https://www.cancer.gov/types/prostate/hp/prostate-screening-pdq>

Nickerson, C. A. E. (1997). A note on “A concordance correlation coefficient to evaluate reproducibility.” *Biometrics*, 53(4), 1503–1507. <https://doi.org/10.2307/2533516>

Obasa, A. E., & Palk, A. C. (2023). Responsible application of artificial intelligence in health care. *South African Journal of Science*, 119(5-6), Article 14889. <https://doi.org/10.17159/sajs.2023/14889>

OECD & European Commission. (2023). Austria: Country cancer profile 2023. *OECD Publishing*. https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/eu-country-cancer-profile-austria-2023_a49c8b3a/798b21c9-en.pdf

OECD & European Commission. (2023). EU Country Cancer Profile: Belgium 2023. *OECD Publishing*. https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/eu-country-cancer-profile-belgium-2023_50d52108/9a976db3-en.pdf

OECD & European Commission. (2023). EU Country Cancer Profile: Bulgaria 2023 [PDF]. EU Country Cancer Profiles, *OECD Publishing*. https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/eu-country-cancer-profile-bulgaria-2023_c9a57aa0/f6915046-en.pdf

OECD & European Commission. (2023). EU country cancer profile: Croatia 2023 [PDF]. EU Country Cancer Profiles. *OECD Publishing*. https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/eu-country-cancer-profile-croatia-2023_23acfeaa/372db8b8-en.pdf

OECD & European Commission. (2023). EU Country Cancer Profile: Cyprus 2023 [PDF].
EU Country Cancer Profiles. *OECD Publishing*.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/eu-country-cancer-profile-cyprus-2023_0dff1cc/86732eb6-en.pdf

OECD & European Commission. (2023). EU Country Cancer Profile: Czech Republic 2023 [PDF]. EU Country Cancer Profiles. *OECD Publishing*.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/eu-country-cancer-profile-czech-republic-2023_42c64a42/c37fd099-en.pdf

OECD & European Commission. (2023). EU country cancer profile: Denmark 2023 [PDF]. EU Country Cancer Profiles. *OECD Publishing*.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/eu-country-cancer-profile-denmark-2023_89208faa/ee590fe8-en.pdf

OECD & European Commission. (2023). EU Country Cancer Profile: Estonia 2023 [PDF]. EU Country Cancer Profiles. *OECD Publishing*.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/eu-country-cancer-profile-estonia-2023_b38a083f/8434f41c-en.pdf

OECD & European Commission. (2023). EU Country Cancer Profile: Finland 2023 [PDF]. EU Country Cancer Profiles. *OECD Publishing*.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/eu-country-cancer-profile-finland-2023_91adbe7f/427186d4-en.pdf

OECD & European Commission. (2023). EU Country Cancer Profile: France 2023 [PDF]. EU Country Cancer Profiles. *OECD Publishing*.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/eu-country-cancer-profile-france-2023_bf3c2485/7abfbfad-en.pdf

OECD & European Commission. (2023). EU Country Cancer Profile: Germany 2023 [PDF]. EU Country Cancer Profiles. *OECD Publishing*.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/eu-country-cancer-profile-germany-2023_0d340a8e/fcc8586e-en.pdf

OECD & European Commission. (2023). EU Country Cancer Profile: Greece 2023 [PDF]. EU Country Cancer Profiles. *OECD Publishing*.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/eu-country-cancer-profile-greece-2023_e4e73f2c/30b7e1f9-en.pdf

OECD & European Commission. (2023). EU Country Cancer Profile: Hungary 2023 [PDF]. EU Country Cancer Profiles. *OECD Publishing*.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/eu-country-cancer-profile-hungary-2023_fc16eb00/ccaf0398-en.pdf

OECD & European Commission. (2023). EU Country Cancer Profile: Ireland 2023 [PDF]. EU Country Cancer Profiles. *OECD Publishing*.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/eu-country-cancer-profile-ireland-2023_5a11a2ef/f31164f8-en.pdf

OECD & European Commission. (2023). EU Country Cancer Profile: Italy 2023 [PDF]. EU Country Cancer Profiles. *OECD Publishing*.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/eu-country-cancer-profile-italy-2023_771850d8/a0a66c1d-en.pdf

OECD & European Commission. (2023). EU Country Cancer Profile: Latvia 2023 [PDF]. EU Country Cancer Profiles. *OECD Publishing*.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/eu-country-cancer-profile-latvia-2023_d8b53c97/3b2c7642-en.pdf

OECD & European Commission. (2023). EU Country Cancer Profile: Lithuania 2023 [PDF]. EU Country Cancer Profiles. *OECD Publishing*.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/eu-country-cancer-profile-lithuania-2023_c00feb52/31316f70-en.pdf

OECD & European Commission. (2023). EU Country Cancer Profile: Luxembourg 2023 [PDF]. EU Country Cancer Profiles. *OECD Publishing*.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/eu-country-cancer-profile-luxembourg-2023_abfcf9c1/6c6cdb7d-en.pdf

OECD & European Commission. (2023). EU Country Cancer Profile: Malta 2023 [PDF]. EU Country Cancer Profiles. *OECD Publishing*.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/eu-country-cancer-profile-malta-2023_412959ac/cff97a9a-en.pdf

OECD & European Commission. (2023). EU Country Cancer Profile: Netherlands 2023 [PDF]. EU Country Cancer Profiles. *OECD Publishing*.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/eu-country-cancer-profile-netherlands-2023_eee73169/89b32870-en.pdf

OECD & European Commission. (2023). EU Country Cancer Profile: Poland 2023 [PDF].

EU Country Cancer Profiles. *OECD Publishing*.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/eu-country-cancer-profile-poland-2023_6949bf84/04cfc3ee-en.pdf

OECD & European Commission. (2023). EU Country Cancer Profile: Portugal 2023 [PDF]. EU Country Cancer Profiles. *OECD Publishing*.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/eu-country-cancer-profile-portugal-2023_590cac0d/2f84bf79-en.pdf

OECD & European Commission. (2023). EU Country Cancer Profile: Romania 2023 [PDF]. EU Country Cancer Profiles. *OECD Publishing*.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/eu-country-cancer-profile-romania-2023_b7601b86/267467c6-en.pdf

OECD & European Commission. (2023). EU Country Cancer Profile: Slovak Republic 2023 [PDF]. EU Country Cancer Profiles. *OECD Publishing*.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/eu-country-cancer-profile-slovak-republic-2023_da001f61/01a8d361-en.pdf

OECD & European Commission. (2023). EU Country Cancer Profile: Slovenia 2023 [PDF]. EU Country Cancer Profiles. *OECD Publishing*.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/eu-country-cancer-profile-slovenia-2023_a9958054/cd631ba0-en.pdf

OECD & European Commission. (2023). EU Country Cancer Profile: Spain 2023 [PDF]. EU Country Cancer Profiles. *OECD Publishing*.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/eu-country-cancer-profile-spain-2023_2c613050/260cd4e0-en.pdf

OECD & European Commission. (2023). EU Country Cancer Profile: Sweden 2023 [PDF]. EU Country Cancer Profiles. *OECD Publishing*.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/eu-country-cancer-profile-sweden-2023_12fd6141/7b5ff594-en.pdf

OECD & European Union (2020), Health at a Glance: Europe 2020: State of Health in the EU Cycle, *OECD Publishing*, Paris, <https://doi.org/10.1787/82129230-en>.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., & Moher, D. (2022). A declaração PRISMA 2020: Diretriz atualizada para relatar revisões

<https://doi.org/10.26633/RPSP.2022.112>

Parlamento Europeu. (2020). *Lutar contra o cancro na EU: estatísticas e ação (infografias)*.

<https://www.europarl.europa.eu/topics/pt/article/20200131ST071517/lutar-contra-o-cancro-na-ue-estatisticas-e-acao-infografias>

Pantanowitz, L., Quiroga-Garza, G. M., Bien, L., Heled, R., Laifenfeld, D., Linhart, C., Sandbank, J., Sucar, B., Vecsler, M., & Hazan, G. (2020). An artificial intelligence algorithm for prostate cancer diagnosis in whole slide images of core needle biopsies: A blinded clinical validation and deployment study. *The Lancet Digital Health*, 2(8), e407–e416. [https://doi.org/10.1016/S2589-7500\(20\)30159-X](https://doi.org/10.1016/S2589-7500(20)30159-X)

Rodrigues, N. M., Silva, S., Vanneschi, L., & Papanikolaou, N. (2023). A comparative study of automated deep learning segmentation models for prostate MRI. *Cancers*, 15(5), 1467. <https://doi.org/10.3390/cancers15051467>

Russell, S. J., & Norvig, P. (2020). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

Saha, A., Bosma, J. S., Twilt, J. J., van Ginneken, B., Bjartell, A., Padhani, A. R., ... de Rooij, M. (2024). Artificial intelligence and radiologists in prostate cancer detection on MRI (PI-CAI): an international, paired, non-inferiority, confirmatory study. *The Lancet Oncology*, 25(7), 879–887. [https://doi.org/10.1016/S1470-2045\(24\)00220-1](https://doi.org/10.1016/S1470-2045(24)00220-1)

Santini, A., Man, A., & Voidăzan, S. (2021). Accuracy of diagnostic tests. *The Journal of Critical Care Medicine*, 7(3), 241–248. <https://doi.org/10.2478/jccm-2021-0022>

Schelb, P., Kohl, S., Radtke, J. P., Wiesenfarth, M., Kickingereider, P., Bickelhaupt, S., ... & Maier-Hein, K. H. (2019). Classification of cancer at prostate MRI: deep learning versus clinical PI-RADS assessment. *Radiology*, 293(3), 607–617. <https://doi.org/10.1148/radiol.2019190273>

Sekhoacha, M., Riet, K., Motlounge, P., Gumenku, L., Adegoke, A., & Mashele, S. (2022). Prostate cancer review: genetics, diagnosis, treatment options, and alternative approaches. *Molecules*, 27(17), 5730. <https://doi.org/10.3390/molecules27175730>

Seeley, R.R, Stephens T.D., & Tate, P. (2007). *Anatomia e Fisiologia* (pp 1039) (6ªed). Lusociência

Serviço Nacional de Saúde. (s.d.). O que é o cancro? SNS 24. <https://www.sns24.gov.pt/tema/doencas-oncologicas/cancro/#o-que-e-o-cancro>

Serviços Partilhados do Ministério da Saúde (SPMS). (2017). Guia de privacidade e proteção de dados: RGPD na era digital (Versão 2). https://spms.min-saude.pt/wp-content/uploads/2017/03/Guia-Privacidade-SMPS_RGPD_digital_20.03.172-v.2.pdf

Sherafatmandjoo, H., Safaei, A. A., Ghaderi, F., & Allameh, F. (2024). Prostate cancer diagnosis based on multi-parametric MRI, clinical and pathological factors using deep learning. *Scientific Reports*, 14, Article 14951. <https://doi.org/10.1038/s41598-024-65354-0>

Shoukri, M. M. (2004). Measures of interobserver agreement and reliability (2nd ed.). Chapman & Hall/CRC. <https://archive.org/details/measuresofintero0000shou>

Sichman, J. S. (2021). Inteligência Artificial e sociedade: avanços e riscos. *Estudos Avançados*, 35(101), 37-50. <https://doi.org/10.1590/s0103-4014.2021.35101.004>

Silva, M.F. (2019). Implicações Clínicas entre Obesidade e Carcinoma da Próstata [Tese de Doutoramento, Universidade do Porto]. *Repositório Aberto da Universidade do Porto*. <https://hdl.handle.net/10216/12105910216/121059>

Singhal, N., Soni, S., Bonthu, S., Chattopadhyay, N., Samanta, P., Joshi, U., Jojera, A., Chharchhodawala, T., Agarwal, A., Desai, M., & Ganpule, A. (2022). A deep learning system for prostate cancer diagnosis and grading in whole slide images of core needle biopsies. *Scientific Reports*, 12, Article 3383. <https://doi.org/10.1038/s41598-022-07217-0>

Stephens, F. O., & Aigner, K. R. (2009). Basics of Oncology (2nd ed.). <https://archive.org/details/basicsofoncology0000step/mode/2up>

Ström, S. S., Kamat, A. M., & Gruschkus, S. K. (2023). Prostate cancer: Epidemiology, prevention, and screening. In J. E. Niederhuber, J. O. Armitage, J. H. Doroshow, J. E. Kastan, & M. M. Tepper (Eds.), *Abeloff's Clinical Oncology* (6th ed., Vol. 1, pp. 1449–1460). Elsevier. [https://doi.org/10.1016/S1470-2045\(19\)30738-7](https://doi.org/10.1016/S1470-2045(19)30738-7)

Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56. <https://doi.org/10.1038/s41591-018-0300-7>

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433–460. <https://doi.org/10.1093/mind/LIX.236.433>

Uehara, H., Kobayashi, T., Matsumoto, M., Watanabe, S., Yoneda, A., & Bando, Y. (2018). Adipose tissue: Critical contributor to the development of prostate cancer. *The Journal of Medical Investigation*, 65(1–2), 9–17. <https://doi.org/10.2152/jmi.65.9>

Vidal, A. C., Howard, L. E., Moreira, D. M., Castro-Santamaria, R., Andriole, G. L., & Freedland, S. J. (2014). Obesity increases the risk for high-grade prostate cancer: Results from the REDUCE study. *Cancer Epidemiology, Biomarkers & Prevention*, 23(12), 2936–2942. <https://doi.org/10.1158/1055-9965.EPI-14-0795>

Vinagre, J., & Moniz, N. (2020). Inteligência Artificial: riscos e promessas. *Revista de Ciência Elementar*, 8(04), 052. <https://doi.org/10.24927/rce2020.052>

Winfield, Alen F., & Jirotko, Marina (2018). Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133), 20180085. <https://doi.org/10.1098/rsta.2018.0085>

Yoo, S., Gujrathi, I., Haider, M. A., & Khalvati, F. (2019). Prostate cancer detection using deep convolutional neurais networks. *Scientific Reports*, 9(1), 19518. <https://doi.org/10.1038/s41598-019-55972-4>


Anexos


Anexo A


Anexo I –Exemplo de representação gráfica da avaliação da qualidade metodológica dos julgamentos dos autores para cada estudo incluído, a partir da avaliação da ferramenta para estudos de acurácia diagnóstica JBI.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Avaliação de Qualidade
Strom P. et al., (2019)	●	●	●	●	●	●	●	●	●	●	●
Bulten W. et al., (2021)	●	●	●	●	●	●	●	●	●	●	●
Schelb et al., (2019)	●	●	●	●	●	●	●	●	●	●	●
Yoo S. et al., (2019)	●	●	●	●	●	●	●	●	●	●	●
Hiremath A. et al., (2023)	●	●	●	●	●	●	●	●	●	●	●
Karjalainen M. et al., (2023)	●	●	●	●	●	●	●	●	●	●	●
Jung M. et al., (2022)	●	●	●	●	●	●	●	●	●	●	●
Pantanowitz L. et al., (2020)	●	●	●	●	●	●	●	●	●	●	●
Singhal N. et al., (2022)	●	●	●	●	●	●	●	●	●	●	●
Campanella G. et al., (2019)	●	●	●	●	●	●	●	●	●	●	●
Mun Y. et al., (2021)	●	●	●	●	●	●	●	●	●	●	●

Legenda

 = Sim

 = Não

 = Não está claro

N/A = Não aplicável

Questões

Q1 – Foram claramente definidos os critérios de inclusão dos participantes?

Q2 – O ambiente clínico e os participantes foram descritos em detalhe?

Q3 – O teste diagnóstico foi realizado de forma independente do padrão de referência?

Q4 – O padrão de referência foi adequado (gold standard válido)?

Q5 – O tempo entre o teste e o padrão de referência foi curto o suficiente?

Q6 – Os resultados do padrão de referência foram interpretados sem conhecer os do padrão de referência?









Q7 – Os resultados do padrão de referência foram interpretados sem conhecer os do teste?

Q8 – Houve concordância entre o teste e o padrão de referência?

Q9 – Todas as amostras receberam ambos os testes (index + referência)?


Q10 – Foram apresentadas medidas de acurácia diagnóstica apropriadas?


Anexo II – Exemplo de representação gráfica da avaliação da qualidade metodológica dos julgamentos dos autores para cada estudo incluído, a partir da avaliação da ferramenta para estudos retrospectivos do JBI.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Avaliação da qualidade
Korevaar S. et al., (2021)								N/A	

Legenda

 = Sim

 = Não

 = Não está claro

N/A = Não aplicável

Questões

Q1 – Os critérios de inclusão dos participantes no estudo foram claramente definidos?

Q2 – Os participantes e o cenário foram descritos em detalhe?

Q3 – A exposição (CT + IA) foi medida de forma válida e fiável?

Q4 – Os desfechos (presença de CaP clinicamente significativo) foram medidos de forma válida e fiável?

Q5 – Houve identificação clara e gestão de potenciais fatores de confusão?

Q6 – Foram utilizadas estratégias apropriadas de análise estatística?

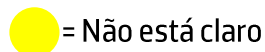
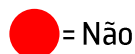
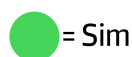
Q7 – O tempo entre a exposição e o desfecho foi apropriado?

Q8 – Foi estabelecida uma relação temporal clara entre exposição e desfecho?

Anexo III - Exemplo de representação gráfica da avaliação da qualidade metodológica dos julgamentos dos autores para cada estudo incluído, a partir da avaliação da ferramenta para estudos de coorte retrospectivos do JBI.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Avaliação de qualidade
Li J. et al., (2020)	●	●	●	●	●	●	N/A	●	●	●	●	●

Legenda



N/A = Não aplicável

Questões

Q1 – Grupos comparáveis e bem definidos?

Q2 – Participantes recrutados de forma apropriada?

Q3 – Exposição medida de forma válida e fiável?

Q4 – Condição medida objetivamente?

Q5 – Fatores de confusão identificados?

Q6 – Estratégias para lidar com fatores de confusão?

Q7 – Participantes livres do desfecho no início do estudo?











Q8 – Seguimento suficiente e completo?

Q9 – Métodos estatísticos apropriados?


Q10 – Resultados fiáveis com base nos dados apresentados?

Q11 – Conclusões justificadas pelos resultados?


Anexo IV – Exemplo de representação gráfica da avaliação da qualidade metodológica dos julgamentos dos autores para cada estudo incluído, a partir da avaliação da ferramenta para estudos de testes diagnósticos do JBI.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Avaliação de qualidade
Sherafatmandjoo H. et al., (2024)							N/A				

Legenda

 = Sim

 = Não

 = Não está claro

N/A = Não aplicável

Questões

Q1 – Critério de inclusão bem definido?

Q2 – População representativa da prática clínica?

Q3 – Padrão de referência adequado e aplicado a todos?

Q4 – Teste índice claramente descrito e aplicável?

Q5 – Comparação cega entre teste índice e padrão de referência?

Q6 – Todos os participantes receberam os mesmos testes de referência?

Q7 – Tempo aceitável entre teste índice e referência?

Q8 – Métodos estatísticos apropriados?

Q9 – Resultados aplicáveis à prática clínica?

Q10 – Baixo risco de viés na interpretação de resultados?

Anexo B

Avaliação da Qualidade da evidência com a ferramenta GRADE

Resultado	Nº Estudos	Qualidade Inicial	Risco de Viés	Inconsistência	Imprecisão	Indireção	Qualidade Final
Sensibilidade	8	Alta	Baixo	Moderada	Baixa	Nenhuma	Moderada
Especificidade	8	Alta	Baixo	Moderada	Moderada	Nenhuma	Moderada
Índice de Kappa	6	Alta	Baixo	Alta	Moderada	Nenhuma	Baixa
Área sob a Curva (AUC)	5	Alta	Baixo	Baixa	Baixa	Nenhuma	Alta

P.PORTO

ESCOLA
SUPERIOR
DE SAÚDE



M

MESTRADO

GESTÃO DAS ORGANIZAÇÕES

RAMO: Unidades de Saúde