

INTELIGÊNCIA ARTIFICIAL PARA REDUÇÃO DE FALSO-POSITIVOS E FALSO-NEGATIVOS NO DIAGNÓSTICO DE CÂNCER DE TIREOIDE

Beatriz Tomaz Caparroz¹
Livia Helena Paschoaloto Polo²
Laysa Gabriella Dourado Rugani³
Maria Clara Coimbra Mancini de Sousa⁴
Viktória Soares Lacerda⁵
Luiz Fernando Travain Ferreira⁶
Vitor Arantes Sousa Carvalho⁷
Rebeca Zanella Ruiz⁸
Caroline Fernanda Vitti Silva⁹
Luana Lima Bastos¹⁰
Maíra Rugoni Costa¹¹
Gabrielly Monteiro da Costa¹²
Claudia Pereira Soares Sanchez Lacerda¹³
Jordana Lamoso Collete¹⁴

RESUMO: O câncer de tireoide é a neoplasia endócrina mais comum e apresenta desafios diagnósticos relacionados à alta taxa de falso-positivos e falso-negativos, levando a intervenções desnecessárias ou atrasos no tratamento. Este trabalho revisa a aplicação da inteligência artificial (IA) no diagnóstico dessa doença, abordando métodos que otimizam a análise de imagens de ultrassonografia e citologia. Algoritmos de aprendizado de máquina, redes neurais convolucionais (CNNs) e ferramentas como AIBx e AI-TIRADS demonstraram maior acurácia, especificidade e capacidade de reduzir biópsias e procedimentos invasivos. Estudos compararam a eficácia desses sistemas em relação a métodos tradicionais, como o TIRADS e avaliações clínicas, destacando o potencial da IA em estratificar riscos e identificar nódulos benignos e malignos com maior precisão. Modelos como o MRF-Net alcançaram alta sensibilidade e especificidade, enquanto sistemas semiautomatizados, como o S-Detect, melhoraram o desempenho em contextos com profissionais menos especializados. Apesar dos avanços, limitações incluem a necessidade de maior sensibilidade em casos que demandam decisões cirúrgicas e a adaptação dos modelos a

¹Acadêmica de Medicina. Fundação Dracense de Educação e Cultura (FUNDEC)

²Acadêmica de Medicina. Fundação Dracense de Educação e Cultura (FUNDEC)

³Acadêmica de Medicina. Fundação Dracense de Educação e Cultura (FUNDEC)

⁴Acadêmica de Medicina. Fundação Dracense de Educação e Cultura (FUNDEC)

⁵Acadêmica de Medicina. Fundação Dracense de Educação e Cultura (FUNDEC)

⁶Acadêmico de Medicina. Fundação Dracense de Educação e Cultura (FUNDEC)

⁷Acadêmico de Medicina. Fundação Dracense de Educação e Cultura (FUNDEC)

⁸Acadêmica de Medicina. Fundação Dracense de Educação e Cultura (FUNDEC)

⁹Acadêmica de Medicina. Fundação Dracense de Educação e Cultura (FUNDEC)

¹⁰Acadêmica de Medicina. Fundação Dracense de Educação e Cultura (FUNDEC)

¹¹Acadêmica de Medicina. Fundação Dracense de Educação e Cultura (FUNDEC)

¹²Acadêmica de Medicina. Fundação Dracense de Educação e Cultura (FUNDEC)

¹³Nutricionista. Centro Universitário do Vale do Araguaia (UNIVAR)

¹⁴Acadêmica de Medicina. Fundação Dracense de Educação e Cultura (FUNDEC)

populações específicas. Conclui-se que a IA é uma ferramenta promissora no manejo clínico do câncer de tireoide, mas requer validações adicionais para sua aplicação segura e eficaz em larga escala.

Palavras-chave: Inteligência artificial. Falso-positivo. Falso-negativo. Neoplasias da Glândula Tireoide.

INTRODUÇÃO

O câncer de tireoide é a neoplasia mais comum do sistema endócrino, sendo o terceiro tumor mais frequente em mulheres na Região Sudeste e na Região Nordeste do país. Em relação ao sexo, a maior prevalência acontece em mulheres, em uma proporção de 3:1. (INCA, 2023). Dentre os tipos de neoplasias tireoidianas, os carcinomas bem diferenciados possuem maior incidência, sendo o papilífero em até 80% dos casos, folicular em 15%, pouco diferenciado e anaplásico com menos de 5% dos casos e medular com 5%. Os carcinomas bem diferenciados resultam de células neoplásicas mais semelhantes às do tecido de origem, que possui um melhor prognóstico, com taxas de sobrevivência superiores a 98% (Kumar, Abbas, Aster, 2024). Nos pouco diferenciados e indiferenciados (anaplásicos), o tumor cresce mais rapidamente e possui metástases mais frequentes, porém são carcinomas pouco presentes nos diagnósticos das neoplasias tireoidianas.

A taxa de novos diagnósticos de carcinoma tireoidiano vem aumentando em cerca de 3% ao ano nos Estados Unidos. Na Coreia do Sul, tornou-se o tipo mais comum de câncer recém-diagnosticado em mulheres. Isso ocorre, sobretudo, a facilidade do diagnóstico, por muitas vezes descoberto de forma incidental durante a palpação do pescoço ou ultrassonografia, e geralmente se tratam de pequenas lesões localizadas (na maioria das vezes do tipo papilífero), que têm um prognóstico bastante positivo (Kumar, Abbas, Aster, 2024).

Em relação às manifestações clínicas, a maioria dos pacientes são assintomáticos, mas em estágios mais avançados podem apresentar nódulos palpáveis associados à disfagia, dispneia e globus faríngeos.

Há diversas formas de diagnosticar essas neoplasias, sendo o exame físico da cabeça e pescoço o primeiro instrumento clínico para suspeição de nódulos da tireoide através da palpação de possíveis massas. A ultrassonografia (USG), método não invasivo e de baixo custo, fornece informações a respeito do tamanho, formato e características dos nódulos. A partir desse exame de imagem, o sistema TIRADS (Thyroid Imaging Reporting and Data System) consegue classificar o risco de malignidade e necessidade de biopsiar o tecido. Os

nódulos de tamanho maior ou igual a 1cm só devem ser biopsiados quando apresentarem bordas irregulares, hipoeogenicidade, vascularização aumentada, microcalcificações, entre outras características de malignidade. Dessa forma, a biópsia aspirativa por agulha fina (PAAF), é um procedimento invasivo e muitas vezes necessário para diagnosticar a neoplasia através da análise citológica e orientar a conduta do tratamento. Também há a possibilidade de exames genéticos e moleculares, podendo ser utilizados quando a biópsia for inconclusiva.

A partir da biópsia, o Sistema Bethesda para Laudos Citopatológicos de Tireóide classifica a amostra em I (insatisfatória/não diagnóstica), II (benigna), III (indeterminada), IV (suspeita de neoplasia) e V (suspeita de malignidade) e VI (maligno). Segundo a Associação Brasileira de Câncer de Cabeça e Pescoço, as categorias III e V apresentam uma taxa relativamente alta de incerteza diagnóstica (20 a 30%), sendo necessária uma avaliação adicional para determinar o risco de malignidade. Tais métodos de diagnóstico possuem limitações, como a alta taxa de resultados inconclusivos e a presença de falsos-positivos e falsos-negativos, resultando em diagnósticos equivocados, cirurgias desnecessárias ou atraso no tratamento das lesões malignas.

Atualmente, a inteligência artificial (IA) atua como uma ferramenta inovadora e promissora para contribuir no diagnóstico de neoplasias tireoidianas. Os algoritmos de *machine learning* (aprendizado de máquina) e redes neurais convencionais (CNNs) possuem uma questão promissora na acurácia de diagnósticos ao analisar imagens de USG e amostras histológicas, podendo diminuir a necessidade de procedimentos invasivos, como também permitindo diagnósticos mais fidedignos. Isso porque são algoritmos que identificam padrões específicos através de um conjunto de dados, tornando o diagnóstico mais rápido e com potencial de alta confiabilidade.

Diante disso, o presente artigo busca revisar o uso da inteligência artificial para diagnóstico de câncer de tireoide, com foco na redução de falso-positivos e falso-negativos e discutir as implicações desses promissores métodos de acurácia diagnóstica.

OBJETIVO

Revisar o impacto da inteligência artificial na redução de falso-positivos e falso-negativos no diagnóstico de câncer de tireoide.

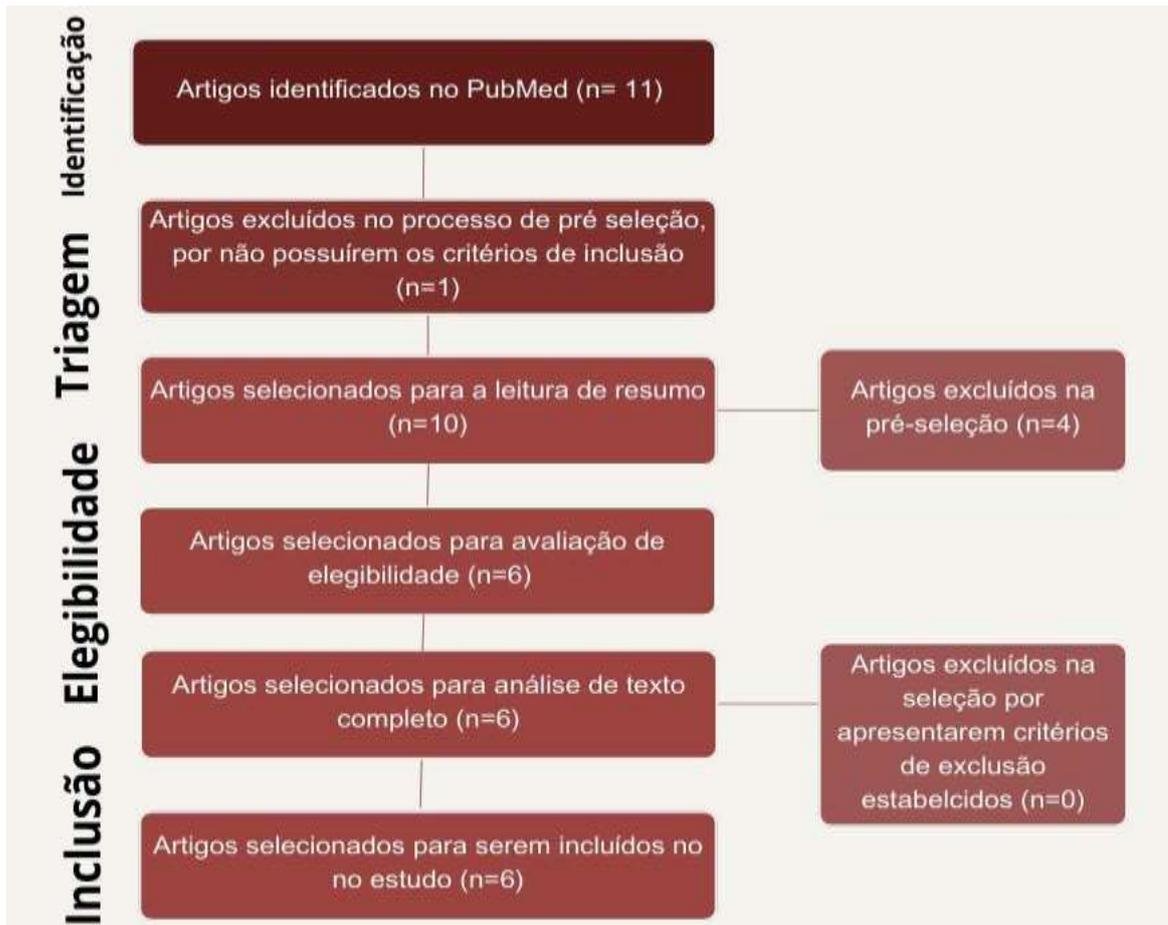
METODOLOGIA

Essa revisão integrativa buscou como base de dados a plataforma PUBMED, utilizando os descritores em inglês: Thyroid cancer, Artificial intelligence, Cancer diagnosis, False positives e False Negatives, a partir dos Descritores em Ciências da Saúde (DeCS).

Os critérios de inclusão utilizados foram: ano de publicação posterior a 2019, limitando o estudo ao período de 2019 a 2024, artigos pertencentes ao idioma inglês e artigos de acesso gratuito. Já os critérios de exclusão foram: artigos de revisão, artigos que não abordavam a inteligência artificial e artigos que não abordavam especificamente o uso da IA no contexto de diagnóstico. Foram encontrados 10 resultados, e a partir dos critérios supracitados, foram analisados 6 artigos.

A fim de demonstrar o caminho realizado de forma sistemática, utilizamos um fluxograma baseado no método PRISMA.

FIGURA 1: Fluxograma para seleção dos artigos encontrados, baseado no PRISMA.



Fonte: Elaborado pelos autores.

RESULTADOS E DISCUSSÃO

A tabela 1 apresenta características dos artigos selecionados (autor, ano de publicação, título, revista publicada, objetivo, amostra, principais resultados). A partir desses achados encontrados foi possível identificar um bom funcionamento de diferentes tipos de IA no diagnóstico de nódulos tireoidianos malignos e benignos. Ademais, há grandes diferenças entre cada uma dessas ferramentas, para melhor compreensão falaremos sobre cada uma delas isoladamente.

TABELA 1: Caracterização dos artigos encontrados para compor o *corpus* da revisão.

| Título do artigo e Revista publicada | Objetivo | Principais resultados |
|---|---|---|
| <p>“External validation of AIBx, an artificial intelligence model for risk stratification, in thyroid nodules”</p> <p><u>Eur Thyroid J</u></p> | <p>Avaliar a eficácia do algoritmo AIBx na estratificação de risco de nódulos tireoidianos utilizando imagens de ultrassom e examinar o desempenho do AIBx quando aplicado a nódulos tireoidianos com resultados citológicos indeterminados.</p> | <p>AIBx, com maior especificidade, tem melhor capacidade de identificar corretamente os nódulos benignos, reduzindo a chance de diagnósticos incorretos de malignidade (reduzindo falso-positivo)</p> |
| <p>Comparison of British Thyroid Association, American College of Radiology TIRADS and Artificial Intelligence TIRADS with histological correlation: diagnostic performance for predicting thyroid malignancy and unnecessary fine needle aspiration rate</p> <p><u>Br J Radiol</u></p> | <p>Comparar o desempenho diagnóstico da British Thyroid Association (BTA), do American College of Radiology Thyroid Imaging Reporting and Data System (ACR-TIRADS) e do Artificial Intelligence TIRADS (AI-TIRADS) na detecção de malignidade em nódulos tireoidianos, avaliando também as taxas comparativas de aspiração por agulha fina (FNA) desnecessária.</p> | <p>Usando o ultrassom como teste (Grau 4-5 positivo, Grau 1-2 negativo), a sensibilidade e especificidade foram, respectivamente, de 98,28% e 42,55% para BTA, 95,24% e 40,57% para ACR-TIRADS e 93,44% e 45,71% para AI-TIRADS. A indicação de FNA foi de 71,6% com BTA, 47,5% com ACR-TIRADS e 46,1% com AI-TIRADS. A taxa de FNA desnecessária foi significativamente maior com BTA (46,3%) em comparação com ACR-TIRADS (30,7%) e AI-TIRADS (29,8%), $p < 0,001$.</p> |

| Título do artigo e Revista publicada | Objetivo | Principais resultados |
|---|--|--|
| <p>“External validation of AIBx, an artificial intelligence model for risk stratification, in thyroid nodules”</p> <p><u>Eur Thyroid J</u></p> | <p>Avaliar a eficácia do algoritmo AIBx na estratificação de risco de nódulos tireoidianos utilizando imagens de ultrassom e examinar o desempenho do AIBx quando aplicado a nódulos tireoidianos com resultados citológicos indeterminados.</p> | <p>AIBx, com maior especificidade, tem melhor capacidade de identificar corretamente os nódulos benignos, reduzindo a chance de diagnósticos incorretos de malignidade (reduzindo falso-positivo)</p> |
| <p>“Investigation of optimal convolutional neural network conditions for thyroid ultrasound image analysis”</p> <p><u>Sci Rep</u></p> | <p>Comparar a eficácia de modelos de algoritmo de “deep learning” baseados em aprendizado “scratch” e aprendizado de transferência na distinção entre imagens benignas e malignas de nódulos tireoidianos nos EUA.</p> | <p>O modelo de aprendizado de transferência superou o modelo zero em todas as métricas. Usar 3902 imagens mostrou desempenho similar ao conjunto completo. O limite de 0,3 deu alta sensibilidade e baixa especificidade, enquanto 0,7 favoreceu a especificidade. Assim, 3900 imagens são suficientes, e o limite pode ser ajustado conforme o perfil populacional.</p> |
| <p>“Predicting Malignancy in Pediatric Thyroid Nodules: Early Experience With Machine Learning for Clinical Decision Support”</p> <p><u>J Clin Endocrinol Metab</u></p> | <p>Desenvolver uma ferramenta de aprendizado de máquina para integrar dados clínicos para a previsão de citologia e histologia não benigna da tireóide.</p> | <p>Embora o presente estudo retrospectivo não tenha gerado um modelo adequado para substituir a prática existente, em grande parte devido à limitação no tamanho da coorte com pontos de dados suficientes, a precisão aprimorada e o AUROC para identificação de biópsia e candidatos cirúrgicos são encorajadores.</p> |
| <p>”The value of a neural network based on multi-scale feature fusion to ultrasound images for the differentiation in thyroid follicular neoplasms”</p> | <p>Criar uma rede de “deep learning” que utilizasse imagens em multiescala para a classificação do carcinoma folicular da tireoide (FTC) e do adenoma folicular da tireoide (FTA) através do pré-operatório dos EUA.</p> | <p>193 foram identificados como FTA e 86 como FTC. O modelo MRF-Net apresentou o melhor desempenho entre os modelos de aprendizado profundo, com precisão de 85,3%, AUC de 84,8%, e valor F1 de 83,08%. Além disso, demonstrou sensibilidade e especificidade superiores. A análise da curva</p> |

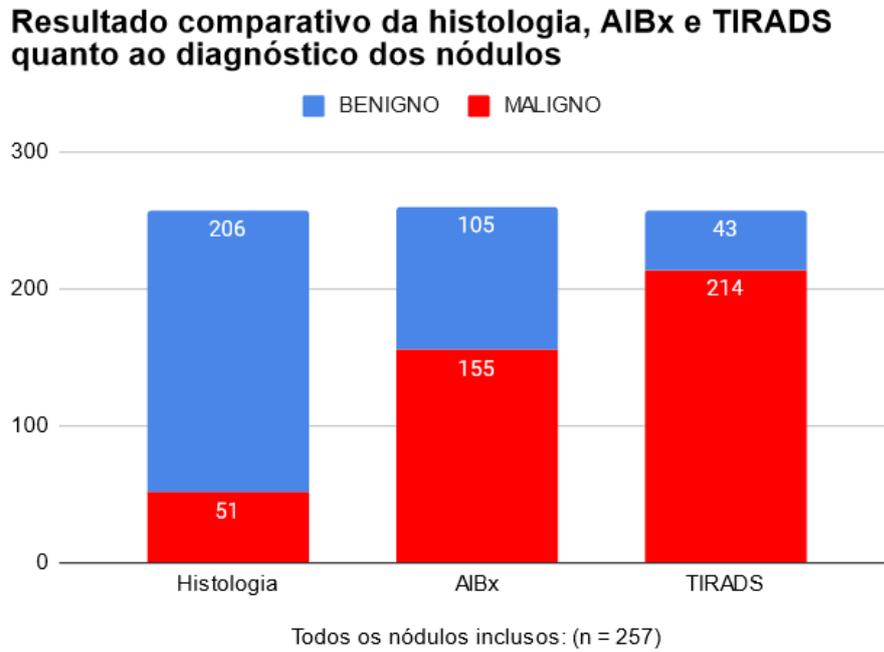
| Título do artigo e Revista publicada | Objetivo | Principais resultados |
|--|--|---|
| <p>“External validation of AIBx, an artificial intelligence model for risk stratification, in thyroid nodules”</p> <p><u>Eur Thyroid J</u></p> | <p>Avaliar a eficácia do algoritmo AIBx na estratificação de risco de nódulos tireoidianos utilizando imagens de ultrassom e examinar o desempenho do AIBx quando aplicado a nódulos tireoidianos com resultados citológicos indeterminados.</p> | <p>AIBx, com maior especificidade, tem melhor capacidade de identificar corretamente os nódulos benignos, reduzindo a chance de diagnósticos incorretos de malignidade (reduzindo falso-positivo)</p> |
| <p><u>BMC Med Imaging</u></p> | | <p>DCA indicou que o MRF-Net ofereceu benefícios líquidos maiores em diversos limites de decisão, superando consistentemente os demais modelos.</p> |
| <p>”Clinical validation of S-Detect™ mode in semi-automated ultrasound classification of thyroid lesions in surgical office”</p> <p><u>Gland Surg.</u></p> | <p>Validar o modo s-Detect™, um sistema de diagnóstico semi-automatizado, para classificar lesões na tireoide durante ultrassons realizados por cirurgiões em consultório.</p> | <p>O sistema CAD melhorou a precisão da avaliação por um cirurgião com habilidades básicas em 6%, aumentando a precisão geral de 76% para 82% ($P < 0,001$), com 3 diagnósticos finais ajustados (1 verdadeiro positivo e 2 verdadeiros negativos). Entretanto, foi inferior à avaliação de cirurgiões com habilidades especializadas, gerando 6 falsos positivos adicionais ($P < 0,001$).</p> |

Fonte: Elaborado pelos autores.

O estudo de coorte de Swan et. al (2022) analisou o algoritmo AIBx, que utilizou imagens de ultrassom de diferentes fabricantes para diagnosticar os nódulos tireoidianos, para verificar se a precisão diagnóstica seria influenciada pelas variadas marcas de fabricantes. O artigo também aplicou o AIBx para avaliar o diagnóstico de citologias indeterminadas, utilizando também as imagens de ultrassom desses nódulos que posteriormente foram biopsiados recebendo classificação de Bethesda indeterminada (III, IV e V).

O gráfico 1 representa os resultados do estudo de Swan et. al (2022), levando a análise histológica como referência do diagnóstico

Gráfico 1: Resultado comparativo da histologia, AIBXs e TIRADS quanto ao diagnóstico dos nódulos



Fonte: Elaborado pelos autores

A tabela 2 ilustra as análises obtidas no estudo, em que evidencia a maior sensibilidade do TIRADS comparada ao AIBx, porém com especificidade reduzida: TIRADS com 19,4% e AIBx com 44,2% de especificidade. Isso explica que o TIRADS tem uma maior capacidade de identificar corretamente os nódulos malignos devido a sua alta sensibilidade. Porém, o TIRADS tem uma maior taxa de falso-positivos devido a sua especificidade ser menor, resultando em casos em que nódulos benignos são classificados como malignos de forma incorreta. Por outro lado, o AIBx, com maior especificidade, tem melhor capacidade de identificar corretamente os nódulos benignos, reduzindo a chance de diagnósticos incorretos de malignidade.

Tabela 2: Comparação de sensibilidade, especificidade, valor preditivo negativo (VPN), valor preditivo positivo (VPP) e acurácia do AIBx e TIRADS.

| Teste | Resultado | | | | |
|--------|-------------------|--------------------|---------|---------|--------------|
| | Sensibilidade (%) | Especificidade (%) | VPN (%) | VPP (%) | Acurácia (%) |
| AIBx | 78,4 | 44,2 | 89,2 | 25,8 | 51 |
| TIRADS | 94,1 | 19,4 | 93 | 22,4 | 34,2 |

Fonte: Elaborado pelos autores

O VPN do TIRADS foi de 93% enquanto do AIBx foi de 89,2%. Esse dado é importante para indicar a probabilidade do teste de identificar a ausência da malignidade, pois quanto maior o VPN, menor o número de falso-negativos, trazendo mais segurança para o paciente que recebe o diagnóstico de que ele realmente não possui um nódulo maligno. O valor do VPN é alto, porém ainda é menor quando comparado ao TIRADS.

Já o VPP do AIBx foi maior comparado ao TIRADS, evidenciando que dentre os nódulos considerados malignos, o AIBx conseguiu se aproximar mais do diagnóstico preciso de malignidade quando comparado ao TIRADS. No geral, ao possuir menos falso-positivos com maior VPP, tem potencial para reduzir o número de biópsias desnecessárias.

A acurácia, proximidade entre o valor obtido experimentalmente e o valor verdadeiro na medição de uma grandeza física, foi de 51% para o AIBx e 34,2% para o TIRADS.

Em relação aos nódulos de citologia indeterminados, a especificidade do AIBx foi de 50% comparada a 12,5% do TIRADS, sendo ainda mais relevante nesses casos.

Portanto, Swan et. al (2022) conclui que a ferramenta AIBx baseada em IA para identificar a malignidade dos nódulos, revelou uma maior precisão diagnóstica quando comparada ao método tradicional TIRADS, com melhor desempenho quando utilizado para casos de citologia indeterminada. Ao comparar o AIBx com a pontuação TIRADS atribuída, o AIBx apresentou maior especificidade e VPP, mas menor sensibilidade e VPN. Os autores explicam que combinar os dois instrumentos pode ser valioso na prática clínica.

Já Watkins et al. (2021) comparou a eficácia dos sistemas de classificação da British Thyroid Association (BTA), American College of Radiology TIRADS (ACR-TIRADS) e Artificial Intelligence TIRADS (AI-TIRADS) na previsão de malignidade. O estudo incluiu 218 nódulos tireoidianos, dos quais 77 (35,3%) eram malignos, e a sensibilidade e especificidade foram avaliadas em relação à histologia como padrão-ouro. Os resultados mostraram que a sensibilidade e especificidade para BTA foi de 98,28 e 42,55%, para ACR-TIRADS: 95,24 e 40,57% e para AI-TIRADS: 93,44 e 45,71%. A PAAF (Punção Aspirativa por Agulha Fina) foi indicada em 101 (71,6%) nódulos benignos utilizando BTA, 67 (47,5%) com ACR-TIRADS e por fim e 65 (46,1%) utilizando AI-TIRADS. A taxa de PAAF desnecessária foi significativamente maior com BTA (46,3%) em comparação com ACR-TIRADS (30,7%) e AI-TIRADS (29,8%) $p < 0,001$.

Concluíram então que BTA, ACR-TIRADS e AI-TIRADS tiveram desempenho semelhante para prever malignidade de nódulo tireoidiano com sensibilidade maior que 93% para todos os sistemas ao classificar os nódulos como malignos (Grau 4-5) ou benignos (Grau 1-2) com base no ultrassom. Entretanto, também foi observado que ACR-TIRADS e AI-TIRADS tiveram uma taxa significativamente menor de PAAF recomendada em nódulos benignos quando comparados com BTA.

Adicionalmente, a pesquisa de Lee et al. (2023) investigou condições ideais para redes neurais convolucionais (CNNs) em imagens de ultrassonografia da tireoide, foram comparados modelos de aprendizado "scratch" e de aprendizado por transferência, além de testes de estresse para determinar a quantidade mínima de dados necessária para treinar o modelo de modo eficaz.

O modelo de aprendizado por transferência apresentou desempenho superior ao modelo "scratch" em todas as validações. O aprendizado por transferência demonstrou o melhor desempenho em comparação com o "scratch", a área sob a curva (AUC) do aprendizado por transferência foi $0,889 \pm 0,007$ nos conjuntos de teste interno e variações entre 0,775 e 0,905 nos externos. A acurácia do aprendizado por transferência foi de 81,6%, mostrando uma alta taxa de acertos, e nos conjuntos externos variou de 58,2% a 69,2%. A sensibilidade do modelo foi de 85,4% no conjunto interno e variou entre 80,3% e 98,4% nos externos, demonstrando alta capacidade do modelo em detectar corretamente os casos positivos, diminuindo os falsos negativos. Por fim, a especificidade foi de 75,1% no conjunto interno e variou entre 24,2% e 56,4% nos conjuntos externos, revelando uma dificuldade em identificar corretamente os casos negativos, podendo haver falsos positivos. Adicionalmente, os autores notaram que, mesmo que a utilização total do conjunto de dados (5575 imagens) levasse a um desempenho ideal, um subconjunto de 70% dessas imagens (aproximadamente 3902) foi o suficiente para reproduzir resultados comparáveis, importante porque maximiza a utilização de recursos ao criar modelos de inteligência artificial com dados restritos. Além disso, destacaram a relevância de adaptar o limiar de probabilidade do modelo de acordo com as particularidades da população-alvo. Por exemplo, ao reduzir o limiar de 0,7 para 0,3, o modelo melhorou sua sensibilidade e valor preditivo negativo.

Já em outro estudo, Radebe et al (2021) buscou analisar como a IA conseguiria auxiliar na previsão da necessidade de biópsias de tireoide e cirurgias em pacientes pediátricos. A respeito das biópsias, dos 67 pacientes estudados, 28 tinham histologia benigna e 39 eram não

benigna. O estudo utilizou um aprendizado de máquina chamado Random Forest (florestas aleatórias) para identificar melhor os pacientes que teriam citologia benigna, com base apenas em características clínicas e ultrassonográficas, e assim evitar biópsias desnecessárias. O modelo alcançou uma acurácia de 77,57% para prever a necessidade de biópsia. A taxa de falsos negativos nessa amostra foi de 10,27%, indicando um risco grande de que os nódulos não recomendados para biópsia seriam, na verdade, malignos, o que não é vantajoso em comparação com a avaliação tradicional.

Nos pacientes com histologia benigna, 14,10% foram classificados pela IA incorretamente, admitindo serem necessários de biópsia por serem malignos, em comparação com a taxa referência de 82,14% mostrando uma melhora de 68,04% na redução de biópsias desnecessárias, diminuindo a taxa de falsos positivos.

A área sob a curva do operador do receptor (AUROC), que mede o desempenho global do modelo, foi de 83,78 frente a 58,93 do modelo clínico tradicional utilizado como referência comparativa do estudo, um aumento de 24,85%. Portanto, o estudo concluiu que o modelo é mais eficaz na discriminação entre nódulos benignos e não benigno.

Em relação a análise da necessidade cirúrgica, 45,83% dos pacientes foram indicados erradamente, pois a IA propôs que seriam indicados para a cirurgia e que na verdade poderiam não precisar dessa intervenção. Essa taxa ainda é alta, mas teve uma redução de 23,22% em comparação à prática clínica histórica. Já quando analisado na AUROC, teve o desempenho de 61,64% comparado a 65,48%, tendo uma queda de 3,84% no que tange ao desempenho de indicações cirúrgicas, sendo então a prática clínica mais efetiva e precisa.

Já Chen et al (2024) coletou 279 imagens de ultrassonografia (USG) para serem analisadas por uma rede de fusão multi-reescala (MRF-Net), comparando o diagnóstico com outros sistemas de IA de aprendizagem profunda: MobileNet V3, ResNet50 e DenseNet121.

Desses nódulos, 193 eram tumores benignos e 86 pacientes foram diagnosticados como malignos. Como resultado, o modelo de IA MRF-Net teve maior sensibilidade (79,41%), especificidade (90,24%), valor preditivo positivo (87,1%) e valor preditivo negativo (84,1%) quando comparado aos outros, ao conseguir identificar os tumores que eram benignos e malignos a partir das imagens de USG.

Comparando as curvas ROC, a AUC do modelo MRF-Net foi maior que a de MobileNET V3 e ResNet50, com valores de $p < 0,05$, correspondentes de 0,0007 e 0,0054, respectivamente, sendo então um modelo de classificação superior, particularmente para

distinguir entre tumores foliculares benignos e malignos da tireoide. A MRF-Net exibiu taxas negativas verdadeiras e verdadeiras mais altas do que os outros três modelos, enquanto suas taxas de falso positivo e falso negativo foram correspondentemente menores.

Por fim, o estudo de Barczyński et al. (2020) buscou analisar a utilidade clínica de um sistema de diagnóstico assistido por computador, o sistema S-Detect em um consultório cirúrgico. A pesquisa comparou diagnósticos de um cirurgião com especialidade básica em análise de imagem e de um operador experiente em diagnóstico de imagem, sem utilizar o CAD (diagnóstico auxiliado pelo computador) e utilizando o CAD.

Tabela 3: Comparação de resultados do estudo de Barczyński et al. (2020)

| Resultado do teste diagnóstico | Especialista sem CAD | Cirurgião sem CAD | Cirurgião com CAD |
|--------------------------------|----------------------|-------------------|-------------------|
| Falso negativo | 1 | 2 | 1 |
| Falso positivo | 2 | 10 | 8 |
| Sensibilidade | 90 | 80 | 90 |
| Especificidade | 95 | 75 | 80 |
| Valor preditivo positivo (%) | 81.82 | 44.44 | 52.94 |
| Valor preditivo negativo (%) | 97.44 | 93.75 | 96.97 |
| Acurácia (%) | 94 | 76 | 82 |

Fonte: Elaborado pelos autores

A respeito dos resultados supracitados na tabela 3, em todos, o valor de p foi $< 0,001$.

A sensibilidade e o VPN utilizando o sistema CAD foi comparável com o de um especialista em diagnóstico de imagem, sendo assim uma ferramenta importante para utilizar quando não possui um profissional com habilidades especializadas para analisar o USG. Já a especificidade e o VPP do sistema foram menores, mas ainda superiores ao julgamento isolado de um cirurgião com habilidades básicas nos EUA.

CONCLUSÃO

A partir desta revisão, podemos concluir que a IA é uma ferramenta promissora na prática médica para melhorar a precisão diagnóstica no câncer de tireóide, reduzindo as taxas de falso-positivos e falso-negativos em comparação com métodos tradicionais. Sua utilização mostrou potencial para otimizar e aprimorar análise de imagens de ultrassonografia e ajudar

na estratificação de risco, especialmente em contextos clínicos onde a experiência com diagnóstico de imagem pode ser limitada. Esses sistemas, como “deep learning”, “machine learning” e modelos por transferência, demonstraram eficácia em identificar corretamente nódulos malignos e benignos a partir dos dados fornecidos, reduzindo a necessidade de biópsias e procedimentos invasivos como a PAAF.

No entanto, alguns desafios permanecem, especialmente em situações que exigem alta sensibilidade, como a recomendação de cirurgias. A combinação de diferentes abordagens e o ajuste dos modelos às características da população-alvo podem ajudar a otimizar os resultados, tornando a IA uma ferramenta complementar valiosa à prática clínica tradicional.

Assim, a IA surge como um apoio importante para melhorar o manejo clínico do câncer de tireoide, sendo essencial continuar o desenvolvimento e a validação desses modelos a partir de estudos para assegurar sua eficácia e aplicação segura em larga escala.

REFERÊNCIAS

BORGES, A. K. DA M. et al.. Câncer de tireoide no Brasil: estudo descritivo dos casos informados pelos registros hospitalares de câncer, 2000-2016 . *Epidemiologia e Serviços de Saúde*, v. 29, n. 4, p. e2019503, 2020.

Câncer de tireoide. Disponível em: <<https://www.gov.br/inca/pt-br/assuntos/cancer/tipos/tireoide>>.

KUMAR, Vinay; ABBAS, Abul K.; ASTER, Jon C. Robbins & Cotran Patologia: Bases Patológicas das Doenças. 10th ed. Rio de Janeiro: GEN Guanabara Koogan, 2023. *E-book*. p.1134. ISBN 9788595159174. Disponível em: <https://integrada.minhabiblioteca.com.br/reader/books/9788595159174/>. Acesso em: 18 out. 2024.

ACBG. Nódulos de Tireoide Indeterminados - Associação Brasileira de Câncer de Cabeça e Pescoço. Disponível em: <https://acbgbrasil.org/nodulos-de-tireoide-indeterminados/?doing_wp_cron=1729296065.8938710689544677734375>. Acesso em: 19 out. 2024.

Swan KZ, Thomas J, Nielsen VE, Jespersen ML, Bonnema SJ. External validation of AIBx, an artificial intelligence model for risk stratification, in thyroid nodules. *Eur Thyroid J*. 2022 Mar 8;11(2):e210129. doi: 10.1530/ETJ-21-0129. PMID: 35113036; PMCID: PMC8963165.

Watkins L, O'Neill G, Young D, McArthur C. Comparison of British Thyroid Association, American College of Radiology TIRADS and Artificial Intelligence TIRADS with histological correlation: diagnostic performance for predicting thyroid malignancy and unnecessary fine needle aspiration rate. *Br J Radiol*. 2021 Jul 1;94(1123):20201444. doi: 10.1259/bjr.20201444. Epub 2021 Jun 9. PMID: 33989038; PMCID: PMC8248201.

Lee JH, Kim YG, Ahn Y, Park S, Kong HJ, Choi JY, Kim K, Nam IC, Lee MC, Masuoka H, Miyauchi A, Kim S, Kim YA, Choe EK, Chai YJ. Investigation of optimal convolutional neural network conditions for thyroid ultrasound image analysis. *Sci Rep.* 2023 Jan 24;13(1):1360. doi: 10.1038/s41598-023-28001-8. PMID: 36693894; PMCID: PMC9873643.

Radebe L, van der Kaay DCM, Wasserman JD, Goldenberg A. Predicting Malignancy in Pediatric Thyroid Nodules: Early Experience With Machine Learning for Clinical Decision Support. *J Clin Endocrinol Metab.* 2021 Nov 19;106(12):e5236-e5246. doi: 10.1210/clinem/dgab435. Erratum in: *J Clin Endocrinol Metab.* 2022 May 17;107(6):e2659. doi: 10.1210/clinem/dgaco27. PMID: 34160618; PMCID: PMC8824766.

Barczyński M, Stopa-Barczyńska M, Wojtczak B, Czarniecka A, Konturek A. Clinical validation of S-Detect™ mode in semi-automated ultrasound classification of thyroid lesions in surgical office. *Gland Surg.* 2020 Feb;9(Suppl 2):S77-S85. doi: 10.21037/gs.2019.12.23. PMID: 32175248; PMCID: PMC7044084.

Chen W, Ni X, Qian C, Yang L, Zhang Z, Li M, Kong F, Huang M, He M, Yin Y. The value of a neural network based on multi-scale feature fusion to ultrasound images for the differentiation in thyroid follicular neoplasms. *BMC Med Imaging.* 2024 Mar 27;24(1):74. doi: 10.1186/s12880-024-01244-1. PMID: 38539143; PMCID: PMC10967122.