

DESEMPENHO COMPARATIVO DE ALGORÍTMOS BASEADOS EM INTELIGÊNCIAS ARTIFICIAIS PARA A DETECÇÃO DE CÁRIE EM ODONTOLOGIA

LUANA PERGHER BOTTEGA¹; VITOR HENRIQUE DIGMAYER ROMERO²;
FAUSTO MEDEIROS MENDES³; MAXIMILIANO SERGIO CENCI⁴; GIANA DA SILVEIRA LIMA⁵; EDUARDO TROTA CHAVES⁶

¹Universidade Federal de Pelotas - luanapergherbottega@gmail.com

²Universidade Federal de Pelotas - vitordigmayer@gmail.com

³Universidade de São Paulo - fmmendes@usp.br

⁴ RadboudUMC - max.cenci@radboudumc.nl

⁵ Universidade Federal de Pelotas - gianalima@gmail.com

⁶ Universidade Federal de Pelotas - eduardo.trota@yahoo.com

1. INTRODUÇÃO

Lesões de cárie secundária, também conhecidas como lesões adjacentes a restaurações ou selantes, são uma das principais causas de falha e substituição de restaurações (KU et al., 2024). A detecção precoce pode ajudar a preservar a estrutura dentária e prevenir tratamentos invasivos. Radiografias interproximais (*bitewings*) são geralmente usadas para auxiliar o diagnóstico, mas sua interpretação é subjetiva e influenciada pela experiência do clínico, sobreposição anatômica, qualidade da imagem e outros fatores associados. A inteligência artificial (IA), especialmente modelos mais aprimorados, oferece análise de imagem consistente e pode reduzir a variabilidade entre os usuários (KU et al.; 2024). No entanto, o desempenho da inteligência artificial também pode ser afetado por artefatos radiográficos, potencialmente causando falsos positivos ou negativos, se os modelos não forem devidamente treinados ou validados. Ferramentas comerciais de inteligência artificial para detecção de cárie já existem, baseadas em estratégias de detecção e segmentação de objetos. Ambas se mostram promissoras para a detecção de cárie, porém poucos estudos compararam modelos comerciais e experimentais, particularmente para lesões secundárias. Este estudo aborda essa questão avaliando a precisão diagnóstica de modelos de inteligência artificial comerciais e baseados em pesquisa nas radiografias interproximais, com ênfase na detecção de cárie secundária.

2. METODOLOGIA

Este estudo transversal de acurácia diagnóstica avaliou o desempenho de modelos de inteligência artificial, comerciais e experimentais (baseados em pesquisas), para detectar lesões de cárie em radiografias interproximais. Foi utilizado como base a Declaração de *Helsinki* e as diretrizes de relatórios específicos para reporte de estudos diagnósticos com IA. Foi obtida aprovação ética (CMO 2025-17935) e o protocolo pré-registrado está disponível publicamente no Open Science Framework. Um total de 200 radiografias anônimas foram selecionadas de um conjunto multicêntrico de 1.261 imagens coletadas na Alemanha, Holanda e Eslováquia.

A inclusão exigiu dentição permanente com pelo menos uma restauração; as exclusões foram baseadas na qualidade técnica, artefatos metálicos ou aparelhos ortodônticos. Um padrão de referência foi estabelecido por meio de consenso de especialistas, a partir de uma processo multifásico com experts em cariologia e

imaginologia. As lesões foram anotadas em dois estágios: (1) presença de cárie, classificadas como superfícies saudáveis, lesões de cárie primárias ou secundárias (com base na associação com restaurações) e (2) estadiamento pela profundidade, classificadas em esmalte ou dentina.

Seis modelos de inteligência artificial foram testados: três comerciais Second Opinion® (SO), Diagnocat® (DC), CranoCatch® (CC) e três algoritmos baseados em pesquisa, Exp_1 (CHAVES et al.; 2024), Exp_2 (VAN NISTELROOIJ et al.; 2025) e Exp_3 (CAO et al.; 2025), desenvolvidos no Radboud AI Lab, usando conjuntos de dados de treinamento independentes, baseadas em redes convolucionais neurais. Para todos os modelos foram aplicados o mesmo conjunto de imagens e os relatórios foram anotados por um mesmo pesquisador. O desempenho diagnóstico (sensibilidade, especificidade, precisão, F1-score e AUC) foi calculado por superfície (oclusal e proximal) (Tabela 1), com sub análises para cárie secundária (SC) e lesões em dentina (DL).

Tabela 1. Características dos Algoritmos de IA Comerciais e Experimentais

Algoritmo & Origem	Arquitetura do Modelo & Segmentação	Distinção Automática de Cárie Secundária?	Estadiamento Automático Esmalte vs Dentina?	Tipo de Dados de Treinamento
Second Opinion® (Pearl Inc., EUA)	Detecção de objetos (caixas delimitadoras)	Não	Sim	Proprietário
Diagnocat (Diagnocat, EUA)	Segmentação (2D/3D)	Sim	Sim	Proprietário
CranioCatch (CranioCatch, Turquia)	Segmentação (U-Net volumétrica)	Não	Não	Proprietário
Exp_1 (Chaves et al., 2024)	Mask R-CNN + Swin Transformer (segmentação por instância)	Sim	Não	Aberto (conjunto usado em publicações)
Exp_2 (van Nistelrooij et al., 2024)	Mask R-CNN + Swin Transformer (segmentação por instância)	Sim	Não	Aberto (conforme protocolos dos estudos)
Exp_3 (Cao et al., 2024)	Mask DINO hierárquico (segmentação multiclasse)	Não	Não	Aberto (conforme protocolos dos estudos)

3. RESULTADOS E DISCUSSÃO

O conjunto de dados final incluiu 200 radiografias interproximais e 1.967 dentes permanentes. Entre todas as superfícies avaliadas, 3.651 (67,2%) foram classificadas como sadias, sem envolvimento de lesão cariosa. Lesões de cárie primária foram encontradas em 1.593 superfícies, sendo 1.139 em esmalte (21,0%) e 454 em dentina (8,4%). Já as lesões de cárie secundária afetaram 537 superfícies, com 483 em esmalte (8,9%) e 54 em dentina (1,0%). Nas superfícies proximais ($n = 2.820$), 2.566 (91,0%) estavam livres de cárries. As lesões de cárie primária envolveram 1.592 superfícies, sendo 1.137 em esmalte (40,3%) e 455 em dentina (16,1%), enquanto as lesões de cárie secundária ocorreram em 518 superfícies, das quais 466 em esmalte (16,5%) e 52 em dentina (1,8%).

Entre os seis modelos de inteligência artificial avaliados, o Second Opinion® apresentou a maior sensibilidade para detecção geral (92,9%), para lesões proximais (93,2%) e em dentina (89,6%), mas também as maiores taxas de falsos positivos, com 8,4% globalmente e 11,4% em superfícies proximais, indicando uma tendência ao sobrediagnóstico. Para cárie secundária, apresentou melhor desempenho equilibrado, com sensibilidade de 74,1% e especificidade de 99,5%.

O Diagnocat®, por sua vez, apresentou um perfil mais equilibrado, com sensibilidade/especificidade de 78,3%/96,0% globalmente, 78,7%/95,1% em superfícies proximais e 73,6%/97,9% para lesões em dentina. Para cárie secundária, atingiu sensibilidade de 67,9% e especificidade de 99,6%, mantendo desempenho semelhante nas superfícies proximais. O CranioCatch® priorizou a especificidade, alcançando 98,6% globalmente e 98,5% para superfícies proximais e lesões em dentina, mas apresentou baixa sensibilidade, de 46,6% globalmente, 46,3% em proximais e aproximadamente 42% para cárie secundária. Apesar disso, exibiu as menores taxas de falsos positivos, de 0,8% para lesões secundárias globais e 1,3% para dentina (Tabela 2).

Tabela 2. Desempenho de detecção de cárie considerando todas as superfícies (cárie global) e apenas superfícies proximais.

	SENS N (% - IC)	ESPEC N (% - IC)	VP (%) CÁRIE GLOBAL	FN (%)	VN (%)	FP (%)	p-valor
SO	92.93%, (0.405–0.931)	90.39% (0.102–0.185)	499 (11.9%)	38 (0.9%)	3296 (78.8%)	350 (8.4%)	<0.001*
CC	46.56%, (0.142–0.269)	97.34% (1.259–2.467)	250 (6.0%)	287 (6.9%)	3552 (84.9%)	97 (2.3%)	<0.001*
DC	78.34%, (1.125–2.191)	96.04% (0.667–1.248)	416 (10.2%)	115 (2.8%)	3425 (83.6%)	141 (3.4%)	0.008*
EXP 1	65.17%, (0.452–0.846)	97.10% (1.097–2.122)	350 (8.4%)	187 (4.5%)	3544 (84.6%)	106 (2.5%)	0.003*
EXP 2	72.26%, (0.175–0.553)	96.28% (1.844–2.081)	388 (9.3%)	149 (3.6%)	3515 (83.9%)	136 (3.2%)	<0.001*
EXP 3	63.13%, (0.396–0.741)	97.51% (1.387–2.746)	339 (8.1%)	198 (4.7%)	3560 (85.0%)	91 (2.2%)	<0.001*
ANÁLISE PROXIMAL							
SO	93.24% (0.980–1.306)	86.04% (0.123–0.228)	483 (17.1%)	35 (1.2%)	1979 (70.1%)	321 (11.4%)	<0.001*
CC	46.33% (0.118–0.230)	96.83% (1.889–3.931)	240 (8.5%)	278 (9.8%)	2227 (78.5%)	73 (2.6%)	<0.001*
DC	78.71% (1.001–1.993)	95.10% (0.998–1.944)	403 (14.8%)	109 (4.0%)	2114 (77.5%)	109 (4.0%)	0.049*
EXP 1	66.80% (0.431–0.826)	95.44% (1.089–2.128)	346 (12.2%)	172 (6.1%)	2196 (77.6%)	105 (3.7%)	0.002*
EXP 2	74.13% (0.652–1.560)	94.18% (1.854–1.937)	384 (13.5%)	134 (4.7%)	2168 (76.4%)	134 (4.7%)	<0.001*
EXP 3	63.90% (0.357–0.682)	96.22% (1.458–2.936)	331 (11.7%)	187 (6.6%)	2215 (78.3%)	87 (3.1%)	<0.001*

A sensibilidade foi calculada como Verdadeiros Positivos / (Verdadeiros Positivos + Falsos Negativos). A especificidade foi calculada como Verdadeiros Negativos / (Verdadeiros Negativos + Falsos Positivos). VP: Verdadeiros Positivos; FN: Falsos Negativos; TN: Verdadeiros Negativos; FP: Falsos Positivos. As diferenças estatisticamente significativas ($p < 0,05$) são indicadas por um asterisco (*).

Entre os modelos experimentais, o Exp_1 mostrou um perfil conservador, com sensibilidade/especificidade de 65,2%/97,1% para lesões globais, 70,4%/99,2% para secundárias e 54,3%/98,8% para dentina. O Exp_2 foi o mais sensível, alcançando 72,3% globalmente, 59,3% para secundária e 73,6% para dentina, sempre com especificidade superior a 97%. O Exp_3 apresentou desempenho intermediário, com sensibilidade de 63,1% para lesões globais, 51,9% para secundária e 64,9% para dentina, mantendo especificidade acima de 96%. Todos os três modelos experimentais apresentaram AUCs superiores a 0,75, refletindo sua calibração conservadora voltada a minimizar falsos positivos.

A maioria dos falsos positivos ocorreu em superfícies restauradas, sobretudo na presença de sobreposição radiográfica, margens defeituosas ou materiais radiopacos que dificultaram a interpretação. Embora falsos negativos possam retardar o tratamento, lesões iniciais costumam progredir lentamente e podem ser manejadas de forma não invasiva. Já os falsos positivos, ao contrário, podem levar a restaurações desnecessárias e à remoção irreversível de tecido sadio, alimentando o ciclo restaurador. Esses resultados reforçam a importância de adequar a seleção de inteligências artificiais aos objetivos clínicos, equilibrando sensibilidade diagnóstica e risco de sobretratamento.

4. CONCLUSÕES

Este estudo confirma uma variabilidade significativa no desempenho das inteligências artificiais para detecção de cárie, influenciada por treinamento, arquitetura e objetivos de design. Os resultados destacam a necessidade de protocolos de validação padronizados e relatórios transparentes de conjuntos de dados.

5. REFERÊNCIAS BIBLIOGRÁFICAS

1. KU, J.C.; LAM, W.Y.; LI, K.Y.; HSUNG, R.T.; CHU, C.H.; YU, O.Y. Accuracy of detection methods for secondary caries around direct restorations: a systematic review and meta-analysis. *J Dent.* 2024;153:105541. doi:10.1016/j.jdent.2024.105541
2. CHAVES, E.T.; VINAYAHALINGAM, S.; VAN NISTELROOIJ, N.; XI, T.; ROMERO, V.H.D.; FLÜGGE, T.; et al. Detection of caries around restorations on bitewings using deep learning. *J Dent.* 2024;143:104886. doi:10.1016/j.jdent.2024.104886
3. VAN NISTELROOIJ, N.; CHAVES, E.T.; CENCI, M.S.; CAO, L.; LOOMANS, B.A.C.; XI, T.; et al. Deep learning-based algorithm for staging secondary caries in bitewings. *Caries Res.* 2025;59(3):163–73. doi:10.1159/000542289
4. CAO, L.; VAN NISTELROOIJ, N.; CHAVES, E.T.; BERGÉ, S.; CENCI, M.S.; XI, T.; et al. Automated chart filing on bitewings using deep learning: enhancing clinical diagnosis in a multi-center study. *J Dent.* 2025;161:105919. doi:10.1016/j.jdent.2025.105919