

Fusão de Dados Multimodais em Patologia Digital e Texto Clínico: Uma Revisão Sistemática Aplicada à Oncologia Odontológica

Bianca Rocha Nicoletti

biancanicoletti@usp.br

Departamento de Computação e Matemática,

FFCLRP-USP

Ribeirão Preto, SP, Brasil

Victor Hugo da Silva Lembre

victorlembror@usp.br

Departamento de Computação e Matemática,

FFCLRP-USP

Ribeirão Preto, SP, Brasil

Raúl Eduardo Sousa García

raul.eduardo.sousa@usp.br

Departamento de Computação e Matemática,

FFCLRP-USP

Ribeirão Preto, SP, Brasil

Alessandra Alaniz Macedo

ale.alaniz@usp.br

Departamento de Computação e Matemática,

FFCLRP-USP

Ribeirão Preto, SP, Brasil

ABSTRACT

The diagnosis and prognosis of oral cancer fundamentally depend on the analysis of histopathological images and patient clinical history. Despite the advances of artificial intelligence in isolated image and text analysis, the fragmentation of these modalities limits the potential of clinical decision support systems. To investigate the state-of-the-art and map how the technological frontier is addressing this problem, this paper presents a systematic literature review focused on the integration of computer vision and natural language processing models applied to dental oncology. Conducted under the Kitchenham guidelines and the PRISMA protocol, the search covered five scientific databases (ACM, IEEE, Scopus, WoS, and PubMed) limited to the period from 2021 to 2026. The initial screening of 511 records resulted in 90 primary studies of high methodological quality. The findings demonstrate that although advanced architectures such as vision transformers and large language models are widely explored unimodally, multimodal fusion approaches (early, late, or joint fusion) and systems aimed at automated treatment planning are still scarce. We conclude that the development of integrated and explainable multimodal frameworks represents the main gap and research opportunity for digital oral health.

KEYWORDS

Artificial Intelligence, Multimodal Fusion, Oral Cancer, Systematic Review, Clinical Decision Support Systems

1 INTRODUÇÃO

O câncer oral, composto predominantemente pelo carcinoma espinocelular oral, representa um desafio crítico para a saúde pública global. Reconhecido epidemiologicamente como uma

das neoplasias malignas mais comuns do mundo, o carcinoma espinocelular oral registrou quase 380 mil novos casos e 177 mil mortes apenas em 2020 [1]. A elevada taxa de mortalidade está atrelada ao diagnóstico tardio; enquanto a taxa de sobrevivência em cinco anos para estágios iniciais pode atingir 90%, esse índice despenca para menos de 40% quando a doença é identificada em estágios avançados (III e IV) [2–4].

Atualmente, o exame microscópico de biópsias teciduais permanece como o padrão-ouro para o diagnóstico e estadiamento do câncer oral. Contudo, a análise histopatológica tradicional é inerentemente subjetiva, exaustiva e dependente da expertise do patologista. A escassez de especialistas e a alta variabilidade intra e interobservador frequentemente resultam em atrasos ou imprecisões no diagnóstico, comprometendo a janela ideal para a intervenção terapêutica [5, 6].

Diante deste cenário, a inteligência artificial (IA) tem destacado como uma solução tecnológica promissora. Modelos de *deep learning*, especialmente redes neurais convolucionais (CNNs) e, mais recentemente, arquiteturas baseadas em *vision transformers* (ViT), demonstraram desempenho relevante na extração automática de características morfológicas complexas em *whole slide images* (WSI). Estudos recentes relatam que modelos unimodais de visão alcançam acurácias superiores a 95% na detecção e classificação de carcinoma espinocelular oral [2, 6, 7].

Entretanto, o uso exclusivo de dados visuais (IA unimodal) apresenta limitações importantes na prática oncológica real. A dependência restrita a imagens histopatológicas ignora o contexto clínico holístico do paciente, como histórico médico, comportamentos de risco (tabagismo, etilismo) e dados demográficos presentes nos prontuários eletrônicos de saúde (EHR). Pesquisadores apontam que a ausência do histórico clínico é uma limitação relevante que impede a evolução de algoritmos puramente classificatórios para modelos preditivos robustos e clinicamente aplicáveis [6, 8].

Para superar essas limitações, a literatura tem apontado a fusão de dados multimodais como uma promissora abordagem para os sistemas de suporte à decisão clínica. A integração

In: Proceedings of the Brazilian Symposium on Multimedia and the Web (WebMedia'2026). Lavras/MG, Brazil. Porto Alegre: Brazilian Computer Society, 2026.

© 2026 SBC – Brazilian Computing Society.

ISSN 2966-2753

de modelos de visão computacional com técnicas de processamento de linguagem natural (NLP) — e modelos de fundação como os *large language models* (LLMs) — permite que algoritmos processem simultaneamente a complexidade estrutural das imagens e as nuances semânticas dos textos clínicos. Esta sinergia diagnóstica espelha o raciocínio médico humano, sendo apontada como estratégia relevante para transitar do “diagnóstico simples” para o prognóstico e o planejamento automatizado de tratamento personalizado [6, 7, 9].

Apesar desse potencial relevante, a aplicação de IA multimodal na odontologia oncológica ainda é incipiente. A literatura carece de padronização, sofrendo com a escassez de conjuntos de dados curados que combinem imagens de biópsia com metadados clínicos detalhados. Ademais, a falta de transparência algorítmica (*black-box models*) e a ausência de mecanismos de explicabilidade (inteligência artificial explicável - XAI) dificultam a confiança clínica e a adoção dessas tecnologias [5, 9, 10].

Considerando a urgência da aplicação de tecnologias assistivas no combate ao carcinoma espinocelular oral e a fragmentação do atual estado da arte, este artigo apresenta uma revisão sistemática da literatura focada na integração de imagens histopatológicas e texto clínico. O objetivo central é mapear as arquiteturas multimodais existentes, avaliar as métricas de desempenho reportadas e identificar se os modelos propostos limitam-se à detecção ou se já avançam estruturalmente para a recomendação clínica e o planejamento de tratamento, visando nortear o desenvolvimento da próxima geração de sistemas de suporte à decisão clínica em odontologia.

O restante deste artigo está organizado da seguinte forma. A Seção 2 apresenta a metodologia adotada para a condução da revisão sistemática da literatura, incluindo as questões de pesquisa, as estratégias de busca, os critérios de elegibilidade, o processo de seleção dos estudos, a avaliação da qualidade e a extração dos dados. A Seção 3 apresenta e discute os resultados obtidos, abrangendo o panorama dos estudos selecionados, as arquiteturas de IA empregadas, as técnicas de NLP e modelos de linguagem, as estratégias de fusão multimodal e as aplicações voltadas ao suporte à decisão clínica e ao planejamento de tratamento. Por fim, a Seção 4 apresenta as conclusões do estudo, destacando as principais lacunas identificadas e as oportunidades para o desenvolvimento de soluções multimodais aplicadas à oncologia odontológica.

2 METODOLOGIA

Este trabalho apresenta uma revisão sistemática da literatura estruturada e conduzida com base nas diretrizes de Kitchenham *et al.* [11] para a área de engenharia de *software*, complementada pelos critérios de reporte do protocolo PRISMA [12]. Todo o protocolo de pesquisa, incluindo a seleção de estudos, avaliação de qualidade e extração de dados, foi gerenciado utilizando a ferramenta web de planejamento de revisões sistemáticas *Parsifal* [13].

Para garantir a transparência e a reprodutibilidade método-lógica, a execução desta revisão sistemática da literatura foi alinhada às quatro fases do protocolo PRISMA (Identificação, Triagem, Elegibilidade e Inclusão), conforme ilustrado na Figura 1. Todo o fluxo de dados foi gerenciado utilizando a ferramenta *Parsifal* [13].

2.1 Escopo da revisão

Para estruturar o escopo da pesquisa e garantir o alinhamento com os objetivos de planejamento automatizado de tratamento, definiu-se a estratégia PICOC (População, Intervenção, Comparação, Resultados e Contexto) [11]:

P (Population): Pacientes com câncer oral (predominantemente o carcinoma espinocelular oral) cujos dados são representados por imagens histopatológicas isoladas (patologia digital), dados clínicos textuais (prontuários e laudos), ou pela disponibilidade simultânea de ambas as modalidades.

I (Intervention): Modelos avançados de IA, englobando técnicas de visão computacional, NLP e LLMs, abordagens de sistemas multiagentes e arquiteturas de fusão multimodal.

C (Comparison): Métodos tradicionais de diagnóstico realizados por especialistas humanos (patologistas e oncologistas odontológicos), ou a comparação do desempenho de arquiteturas multimodais contra modelos computacionais unimodais (*baselines*).

O (Outcome): Desempenho preditivo na classificação e detecção, viabilidade técnica da integração multimodal, explicabilidade (XAI) e aplicabilidade no planejamento automático de tratamento.

C (Context): Desenvolvimento de sistemas de suporte à decisão clínica no âmbito da saúde bucal digital e odontologia oncológica.

2.2 Questões de pesquisa

Para guiar a investigação e mapear a lacuna existente no planejamento automático de tratamento utilizando fusão de dados multimodais, foram formuladas cinco questões de pesquisa (RQs), apresentadas na Tabela 1.

A RQ1 tem como objetivo mapear as arquiteturas de IA empregadas na análise unimodal de imagens histopatológicas de câncer bucal, identificando os principais modelos utilizados e os desempenhos alcançados na literatura. Por sua vez, a RQ2 busca investigar como técnicas de NLP e LLMs têm sido aplicadas na análise de dados clínicos textuais, especialmente em prontuários eletrônicos e registros médicos relacionados ao câncer oral.

As questões RQ3 e RQ4 concentram-se na investigação das abordagens multimodais propostas na literatura, avaliando como ocorre a fusão entre imagens histopatológicas e dados clínicos textuais, bem como se os modelos desenvolvidos limitam-se às tarefas de diagnóstico e classificação ou avançam estruturalmente para sistemas de suporte à decisão clínica e planejamento automatizado de tratamento.

Table 1: Mapeamento das RQs abordadas neste estudo.

ID	Questão de pesquisa
RQ1	Quais arquiteturas de IA estão sendo utilizadas para análise unimodal de imagens histopatológicas de câncer bucal e quais métricas de desempenho têm sido alcançadas?
RQ2	Como técnicas de NLP e LLMs têm sido aplicadas para classificar dados clínicos textuais relacionados a essa patologia?
RQ3	De que forma a literatura atual realiza a fusão multimodal (<i>early fusion</i> , <i>late fusion</i> ou <i>joint fusion</i>) entre imagens histopatológicas e texto clínico, e quais ganhos de desempenho têm sido reportados?
RQ4	Os modelos propostos limitam-se às tarefas de classificação e diagnóstico ou avançam estruturalmente para recomendação clínica e planejamento automatizado de tratamento?
RQ5	Quais são as principais limitações metodológicas identificadas e como mecanismos de explicabilidade (XAI), sistemas baseados em agentes ou abordagens baseadas em regras são utilizados para garantir a confiabilidade clínica?

Por fim, a RQ5 busca identificar as principais limitações metodológicas presentes nos estudos analisados, além de examinar como mecanismos de explicabilidade (XAI), sistemas baseados em regras ou arquiteturas multiagentes vêm sendo utilizados para aumentar a confiabilidade clínica, interpretabilidade e robustez das soluções propostas.

2.3 Identificação

O que foi feito: A busca primária por evidências científicas foi realizada em cinco bases eletrônicas de alta relevância para as áreas de Computação e Saúde:

- (1) ACM Digital Library
- (2) IEEE Xplore
- (3) Scopus
- (4) PubMed
- (5) Web of Science

As buscas foram realizadas individualmente entre 13 e 23 de abril de 2026, sendo executadas em 13/04/2026 na PubMed, 15/04/2026 na Scopus, 22/04/2026 na Web of Science e IEEE Xplore e 23/04/2026 na ACM Digital Library. Foram recuperados inicialmente 511 registros, distribuídos entre PubMed (n=80), Scopus (n=216), Web of Science (n=102), IEEE Xplore (n=69) e ACM Digital Library (n=44), conforme apresentado na Figura 1.

Como foi feito: A execução baseou-se em três *strings* de busca distintas (Tabela 2): (1) Eixo visual: focado em visão computacional e imagens histopatológicas (WSI); (2) Eixo textual: focado em NLP aplicado a dados clínicos textuais; e (3) Eixo multimodal: focado estritamente na interseção, contendo termos como “*data fusion*” e “*treatment planning*”.

Aplicaram-se filtros predefinidos nas bases para capturar apenas estudos em inglês publicados entre 2021 e 2026. Além disso, adaptações sintáticas pontuais foram documentadas, como a remoção de caracteres especiais ou truncadores na base IEEE Xplore e a utilização do comando de metadados específicos nas bases ACM Digital Library e Web of Science, visando respeitar as limitações dos motores de busca sem comprometer a semântica da consulta.

Por que foi feito: A divisão das *strings* em três eixos (ao invés de uma única *string* global) foi uma decisão arquitetural para evitar a super-restrição booleana. Essa abordagem garantiu a captura tanto de estudos puramente unimodais (*baseline*) quanto do escasso estado da arte em integração multimodal.

2.4 Triagem

O que foi feito: O montante inicial de 511 artigos foi reduzido para 152 estudos potencialmente relevantes.

Como foi feito: Inicialmente, o algoritmo do *Parsifal* [13], aliado à revisão manual, identificou e removeu 217 registros duplicados. Em seguida, procedeu-se à leitura dos títulos e resumos dos 294 artigos únicos, aplicando-se os critérios de inclusão e exclusão rigorosamente definidos na Tabela 3.

Por que foi feito: Esta etapa de filtragem rápida (excluindo 142 artigos) foi necessária para eliminar falsos positivos retornados pelos motores de busca, tais como estudos secundários (n=33), tipos de dados incorretos como imagens radiológicas (n=69), tarefas sem relevância clínica (n=8), tipo de documento inadequado (n=19), patologias fora do escopo da cavidade oral (n=12) e artigo duplicado não identificado pela ferramenta automática do *Parsifal* [13] (n=1), otimizando o esforço analítico para as etapas subsequentes.

Mitigação de viés de seleção: Em conformidade com as diretrizes de *Kitchenham* para engenharia de *software*, as etapas de triagem (leitura de títulos e resumos) e elegibilidade (avaliação do texto completo) foram conduzidas de forma cega e independente por dois pesquisadores. Nos casos em que houve divergência quanto à inclusão, exclusão ou avaliação de qualidade de um manuscrito, as discordâncias foram debatidas para a busca de um consenso. Quando o consenso não foi atingido, um terceiro pesquisador foi consultado para proferir a decisão definitiva de desempate.

2.5 Elegibilidade

O que foi feito: Os 152 artigos elegíveis passaram por uma avaliação aprofundada, resultando na retenção de 90 estudos primários para a síntese final e na separação de 51 artigos para fundamentação metodológica.

Como foi feito: Adotou-se uma estratégia de leitura parcial estendida (introdução, arquitetura e conclusão), bem como a verificação de disponibilidade dos manuscritos. Nesta etapa, 11 artigos foram excluídos por indisponibilidade do texto completo para leitura (**EC6**). Adicionalmente, os artigos que possuíam IA robusta, mas desviavam da aplicação clínica direta (possuindo viés puramente teórico ou exploratório), foram classificados formalmente pelo critério **EC7** (n=51).

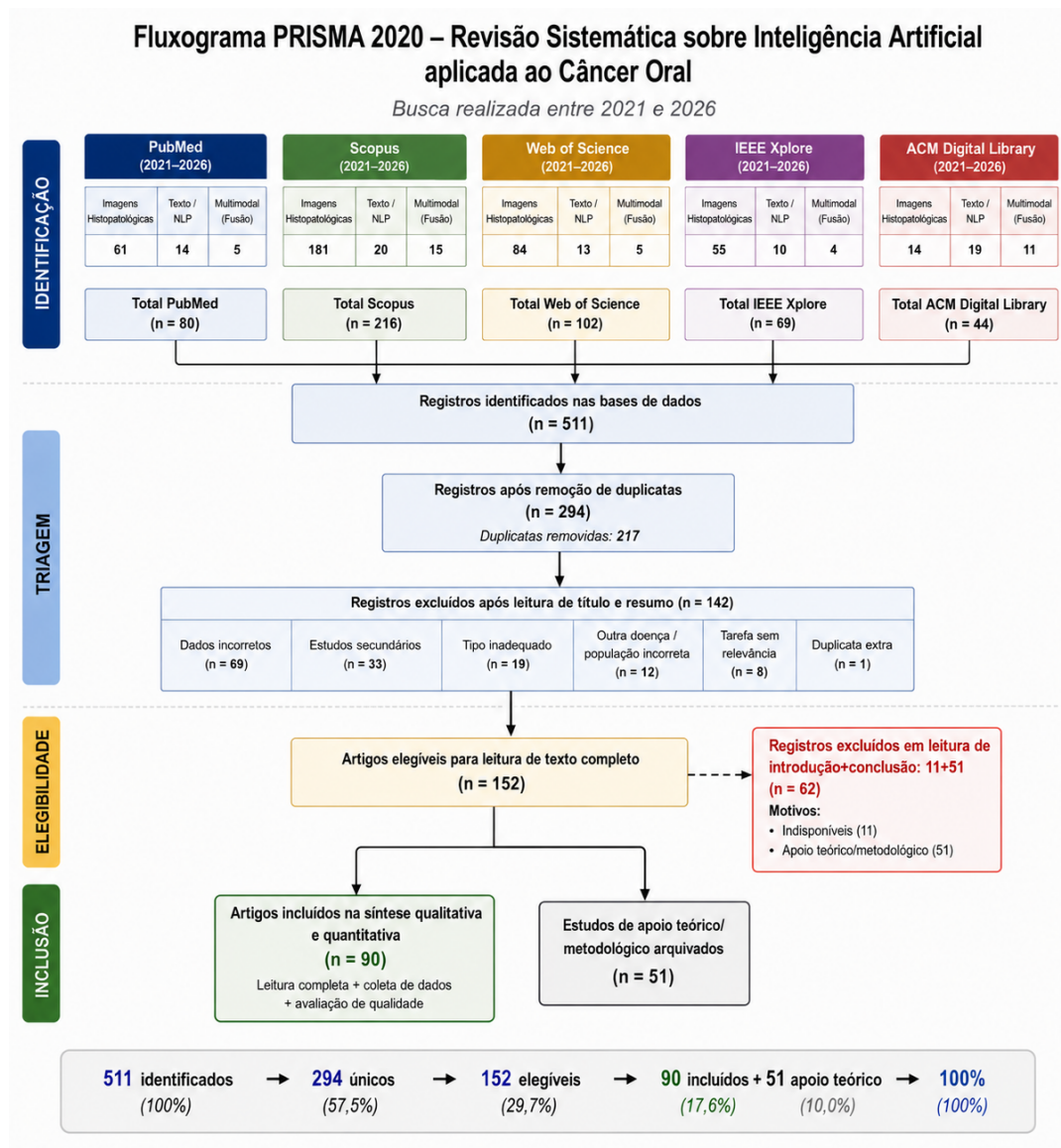


Figure 1: Fluxograma PRISMA detalhando as fases de identificação, triagem, elegibilidade e inclusão dos estudos.

Por que foi feito: Esta triagem intermediária é vital para otimizar o esforço científico. Rejeitar definitivamente os artigos do critério **EC7** resultaria na perda de arquiteturas valiosas; contudo, incluí-los na extração de dados final comprometeria as estatísticas e a resposta às RQs clínicas. Assim, a decisão de arquivá-los separadamente como estudos de apoio garante rigor metodológico a esta revisão, ao mesmo tempo em que consolida um arcabouço tecnológico de ponta. Esse acervo servirá como base diretiva para trabalhos futuros e para a concepção de novos *frameworks* na área.

2.6 Inclusão

O que foi feito: Consolidou-se a amostra final de estudos primários. Estes foram submetidos a uma avaliação rigorosa de vies metodológico e à extração das variáveis necessárias para responder às RQs.

Como foi feito: O processo ocorreu em duas etapas dentro do *Parsifal* [13]. Primeiro, aplicou-se um *checklist* de avaliação de qualidade composto por cinco critérios fundamentais (Tabela 4). A avaliação foi conduzida de forma cega e independente por dois pesquisadores e tabulada em planilhas eletrônicas. Para garantir a confiabilidade do processo e mitigar o viés de seleção, calculou-se a taxa de concordância bruta entre os avaliadores. Considerando o total de 450 aferições

Table 2: Strings de busca utilizadas nas consultas às bases de dados.

Categoria	String de busca
Eixo visual: focado em visão computacional e imagens histopatológicas (WSI).	(“oral cancer” OR “oral squamous cell carcinoma” OR “OSCC” OR “oral cavity cancer” OR “mouth neoplasm”) AND (“histopathology” OR “whole slide image*” OR “WSI” OR “digital pathology” OR “microscopic image*”) AND (“artificial intelligence” OR “machine learning” OR “deep learning” OR “CNN” OR “convolutional neural network*” OR “transformer*” OR “foundation model*” OR “vision transformer*” OR “ViT” OR “large language model*” OR “LLM” OR “vision-language model*” OR “VLM” OR “large multimodal model*” OR “LMM” OR “multi-agent”) AND (“classification” OR “diagnosis” OR “detection”)
Eixo textual: focado em NLP aplicado a dados clínicos textuais.	(“oral cancer” OR “oral squamous cell carcinoma” OR “OSCC” OR “oral cavity cancer” OR “mouth neoplasm”) AND (“clinical record*” OR “electronic health record*” OR “EHR” OR “EMR” OR “medical record*” OR “clinical note*” OR “patient data” OR “textual data”) AND (“natural language processing” OR “NLP” OR “language model*” OR “LLM” OR “transformer*” OR “text mining” OR “machine learning” OR “artificial intelligence”) AND (“classification” OR “diagnosis” OR “prediction” OR “prognosis”)
Eixo multimodal: focado estritamente na interseção, contendo termos como “data fusion” e “treatment planning” (fusão).	(“oral cancer” OR “oral squamous cell carcinoma” OR “OSCC” OR “oral cavity cancer” OR “mouth neoplasm”) AND (“histopathology” OR “whole slide image*” OR “WSI” OR “digital pathology” OR “microscopic image*”) AND (“clinical record*” OR “electronic health record*” OR “EHR” OR “medical record*” OR “clinical note*” OR “patient data” OR “textual data” OR “multimodal” OR “multi-modal” OR “data fusion” OR “early fusion” OR “late fusion” OR “cross-modal”) AND (“artificial intelligence” OR “deep learning” OR “transformer*” OR “foundation model*” OR “vision transformer*” OR “ViT” OR “large language model*” OR “LLM” OR “vision-language model*” OR “VLM” OR “large multimodal model*” OR “LMM” OR “multi-agent”) AND (“classification” OR “diagnosis” OR “treatment planning” OR “decision support”)

realizadas (5 critérios avaliados para cada um dos 90 estudos primários), obteve-se uma expressiva concordância inicial de 89,11% (401 avaliações consensuais). As 49 divergências pontuais, distribuídas em 38 artigos, foram encaminhadas a um terceiro pesquisador, que revisou exclusivamente as questões em conflito para definir a nota definitiva de desempate. A partir dessa validação por pares, cada critério foi pontuado como Sim (1.0), Parcialmente (0.5) ou Não (0.0). Optou-se por não estabelecer uma nota de corte eliminatória; em vez disso, o *score* de qualidade foi utilizado como ferramenta de ponderação para mapear o verdadeiro estado de maturidade da literatura.

Na segunda etapa, aplicou-se o formulário de extração de dados. Além dos metadados básicos de identificação (país, ano de publicação e palavras-chave), extraíram-se 17 variáveis principais de cada estudo, formuladas da seguinte maneira:

- Q1. Ano de publicação do estudo.
- Q2. País de origem da pesquisa.
- Q3. Palavras-chave fornecidas pelos autores.
- Q4. Tipo de câncer ou lesão oral específica abordada.
- Q5. Modalidades de dados utilizadas (apenas imagem, apenas texto ou multimodal).
- Q6. Formato predominante das imagens histopatológicas (ex.: WSI, *patches*).
- Q7. Banco de dados utilizado (público ou institucional privado).
- Q8. Estratégia de fusão multimodal aplicada (*early*, *late* ou *joint fusion*).

Q9. Arquitetura do modelo extrator visual utilizado (ex.: CNN, ViT).

Q10. Arquitetura do modelo de linguagem (NLP/LLM) utilizada.

Q11. Principal tarefa clínica resolvida (ex.: diagnóstico, graduação, prognóstico ou planejamento).

Q12. Técnica de explicabilidade (XAI) implementada (ex.: mapas de atenção ou geração de texto).

Q13. Métricas computacionais reportadas na avaliação de desempenho.

Q14. Estratégia de avaliação humana, usabilidade ou confiança clínica reportada.

Q15. Base estruturada utilizada para injeção de conhecimento externo (ex.: *retrieval-augmented generation*).

Q16. Disponibilização do código-fonte, modelo ou artefatos para reprodutibilidade.

Q17. Principal lacuna ou limitação técnica reportada pelos autores.

Por que foi feito: A abstenção de uma nota de corte eliminatória evita o viés de seleção (positivo) e garante que a revisão exiba o retrato fiel do estado da arte, evidenciando se a área carece de rigor metodológico. Paralelamente, a aplicação exaustiva do formulário (Q1 a Q17) garante a dissecação técnica precisa dos algoritmos mapeados. A síntese destes dados fornece a base empírica e a justificativa arquitetural necessária para evidenciar a lacuna atual na literatura, direcionando o desenvolvimento futuro de *frameworks* multimodais automatizados para planejamento terapêutico em odontologia oncológica.

Table 3: Critérios de inclusão e exclusão utilizados na seleção dos estudos.

ID	Critério
<i>Critérios de inclusão</i>	
IC1	Estudos aplicando IA à análise de carcinoma espinocelular oral.
IC2	Uso de imagens histopatológicas digitais (WSI).
IC3	Uso de NLP, prontuários (EHR/EMR) ou texto clínico.
IC4	Arquiteturas multimodais e fusão entre imagem e texto.
IC5	Uso de <i>deep learning</i> , CNNs, <i>transformers</i> , LLMs ou VLMs.
IC6	Trabalhos contendo métricas quantitativas (acurácia, AUC, F1-score, etc.).
IC7	Foco em diagnóstico, prognóstico ou planejamento terapêutico.
<i>Critérios de exclusão</i>	
EC1	Estudos secundários: revisões de literatura, revisões sistemáticas ou cartas ao editor.
EC2	Tipo de documento inadequado: resumos curtos, anais não revisados ou estudos sem validação empírica clara.
EC3	Outra doença ou população incorreta: foco exclusivo em patologias fora da cavidade oral.
EC4	Dados ou tipos de imagem incorretos: uso exclusivo de radiografias, macroscopia, planilhas tabulares (sem imagem, texto ou ausência de IA treinada).
EC5	Tarefa sem relevância clínica: algoritmos focados em tarefas computacionais genéricas sem foco na decisão oncológica.
EC6	Texto indisponível: artigos cujo manuscrito completo não pôde ser recuperado.
EC7	Apoio teórico e metodológico: estudos de viés puramente exploratório arquivados para fundamentação, não incluídos na síntese final.

Para sistematizar a etapa de resultados e discussão, as variáveis extraídas foram diretamente mapeadas para as RQs correspondentes, conforme sumarizado na Tabela 5. As variáveis Q1, Q2, Q3, Q4 e Q7 foram destinadas à caracterização bibliométrica e de contexto global do corpo de evidências.

Com a aplicação rigorosa dos critérios de elegibilidade, o corpo de evidências final consolidou-se em 90 estudos primários de alta relevância, publicados entre 2021 e 2026. A Tabela 6 apresenta a relação completa dos artigos incluídos nesta síntese. A extração de dados tabulou variáveis bibliométricas, metodológicas e clínicas, cujo mapeamento responde diretamente às RQs estabelecidas.

2.7 Ameaças à validade e limitações da revisão

Embora esta revisão sistemática da literatura tenha seguido rigorosamente as diretrizes de Kitchenham *et al.* [11] e o protocolo PRISMA [12], algumas limitações e ameaças à validade

Table 4: Critérios de avaliação de qualidade (QA) aplicados aos estudos incluídos.

ID	Critério metodológico avaliado
QA1	O estudo descreve detalhadamente o conjunto de dados (tamanho, origem clínica e partição clara de treino e teste)?
QA2	O modelo foi submetido a uma estratégia robusta de validação para evitar <i>overfitting</i> (ex.: <i>cross-validation</i> ou validação externa)?
QA3	A arquitetura da IA, os hiperparâmetros ou a estratégia de fusão estão descritos de forma suficientemente detalhada para permitir reprodutibilidade?
QA4	O estudo utiliza métricas adequadas ao contexto médico (ex.: AUC, F1-score, sensibilidade e especificidade), e não apenas acurácia?
QA5	O modelo desenvolvido é comparado de forma justa com métodos <i>baseline</i> (modelos unimodais ou especialistas humanos)?

Table 5: Mapeamento estrutural entre as variáveis de extração e as RQs.

RQ Alvo	Variáveis de extração (Q) relacionadas
RQ1 (Visão)	Q5 (Imagem), Q6, Q9, Q13
RQ2 (Texto)	Q5 (Texto), Q10, Q13
RQ3 (Fusão)	Q5 (Multimodal), Q8, Q9, Q10, Q13
RQ4 (Planejamento)	Q11, Q15
RQ5 (Limitações)	Q12, Q14, Q16, Q17

Table 6: Artigos incluídos na revisão sistemática.

Nº	Ref.	Nº	Ref.	Nº	Ref.	Nº	Ref.	Nº	Ref.	Nº	Ref.
1	[14]	16	[15]	31	[16]	46	[2]	61	[17]	76	[18]
2	[19]	17	[20]	32	[21]	47	[22]	62	[23]	77	[24]
3	[25]	18	[26]	33	[27]	48	[28]	63	[29]	78	[30]
4	[31]	19	[32]	34	[33]	49	[34]	64	[35]	79	[36]
5	[37]	20	[38]	35	[39]	50	[4]	65	[40]	80	[41]
6	[42]	21	[43]	36	[44]	51	[45]	66	[46]	81	[47]
7	[48]	22	[49]	37	[50]	52	[10]	67	[51]	82	[1]
8	[52]	23	[53]	38	[54]	53	[7]	68	[55]	83	[56]
9	[57]	24	[58]	39	[59]	54	[8]	69	[60]	84	[61]
10	[62]	25	[6]	40	[63]	55	[64]	70	[65]	85	[66]
11	[9]	26	[67]	41	[3]	56	[68]	71	[69]	86	[70]
12	[71]	27	[72]	42	[73]	57	[74]	72	[75]	87	[76]
13	[77]	28	[78]	43	[79]	58	[80]	73	[81]	88	[82]
14	[83]	29	[84]	44	[85]	59	[86]	74	[87]	89	[88]
15	[89]	30	[90]	45	[91]	60	[92]	75	[5]	90	[93]

inerentes ao processo de pesquisa devem ser reconhecidas e mitigadas:

- **Validade de construto:** A escolha das *strings* de busca e das bases de dados pode ter omitido estudos relevantes que utilizam terminologias não mapeadas. Para mitigar esse risco, a busca foi estruturada em três eixos semânticos amplos e executada nas cinco principais bibliotecas digitais da interface computação-saúde. Contudo, a restrição ao idioma inglês e ao recorte temporal (2021–2026) pode ter excluído trabalhos fundacionais mais antigos ou pesquisas regionais publicadas em outros idiomas.
- **Validade interna e viés de seleção:** O processo de triagem e extração de dados está sujeito ao viés interpretativo humano e computacional. Para otimizar o processamento em larga escala desta revisão sistemática da literatura, utilizou-se o modelo *NotebookLM* (LLM *Gemini 1.5 Pro*) como assistente de extração algorítmica. O uso de IA Generativa introduz o risco inerente de alucinação de dados. Para neutralizar essa ameaça, empregou-se a abordagem *human-in-the-loop* (humano no ciclo): a IA operou sob *prompts* de restrição estrita (*zero-hallucination constraint*), e 100% das métricas e classificações extraídas foram submetidas a uma auditoria e verificação cruzada manual com o artigo original. Nenhum dado foi tabulado sem a validação final dos pesquisadores.
- **Validade externa:** Os resultados refletem o panorama tecnológico estrito da oncologia odontológica. Consequentemente, as conclusões sobre a imaturidade da fusão multimodal e escassez de explicabilidade (XAI) podem não ser diretamente generalizáveis para outras áreas mais consolidadas da medicina (como radiologia pulmonar ou câncer de mama), onde a adoção de IA pode estar em estágios metodológicos diferentes.

3 RESULTADOS E DISCUSSÃO

Esta seção apresenta uma visão inicial da distribuição bibliométrica e metodológica dos estudos incluídos na revisão sistemática. As Figuras 2, 3 e 4 sintetizam, respectivamente, a evolução temporal das publicações, as modalidades de dados utilizadas e os países com maior participação no *corpus* analisado.

3.1 Visão geral e bibliometria do estado da arte

A análise da evolução temporal das publicações (Figura 2) revela um crescimento exponencial no interesse pela aplicação de IA na oncologia odontológica. O volume de publicações saltou de apenas 3 estudos em 2021 para um pico de 31 artigos em 2025. O ano de 2026, embora ainda em curso durante a execução desta pesquisa, já apresenta 8 estudos validados, indicando a presença de pesquisas recentes voltadas ao desenvolvimento de modelos preditivos para o carcinoma espinocelular oral.

Do ponto de vista da distribuição geográfica, a Figura 3 demonstra um domínio absoluto da Índia, responsável por

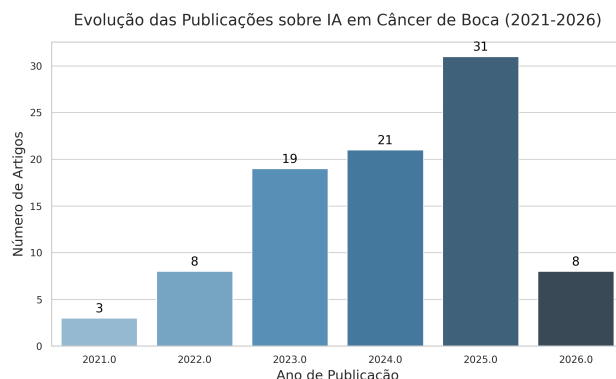


Figure 2: Gráfico com a distribuição anual dos estudos incluídos na revisão sistemática.

quase metade dos estudos incluídos, seguida pela China. Este dado bibliométrico possui forte correlação epidemiológica: o Sul da Ásia concentra aproximadamente um terço de todos os casos mundiais de carcinoma espinocelular oral [58]. A urgência de saúde pública nestes países asiáticos é impulsionada por práticas culturais enraizadas, como o consumo de tabaco de mascar e noz de areca (*betel quid*), somadas à escassez de especialistas patologistas em centros de cuidados primários [15, 50]. Consequentemente, o diagnóstico tardio assola a região — com mais de dois terços dos pacientes buscando auxílio apenas em estágios finais da doença —, o que justifica o alto volume de investimentos estatais e acadêmicos em soluções de triagem automatizada e IA [52].

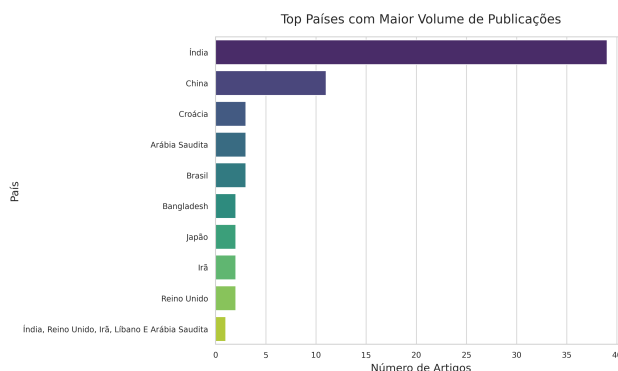


Figure 3: Distribuição geográfica dos estudos por país de origem.

Em contraste ao elevado volume de publicações asiáticas, a contribuição brasileira revela um protagonismo estratégico impulsionado pela relevância epidemiológica do câncer de cavidade oral no país. Segundo estimativas do Instituto Nacional de Câncer (INCA) para o triênio 2026–2028, o Brasil deverá registrar aproximadamente 781 mil novos casos de câncer por ano. Nesse contexto, o câncer de cavidade oral figura como o

quinto tipo mais incidente entre a população masculina, correspondendo a 4,8% dos casos, com maior concentração nas regiões Sul e Sudeste [94]. No Brasil o tabagismo permanece como o principal fator de risco, especialmente quando associado ao consumo crônico de álcool. Diante desse panorama, o INCA reforça a importância do diagnóstico precoce para aumentar as chances de cura, subsidiar o planejamento de políticas públicas e otimizar a alocação de recursos no Sistema Único de Saúde (SUS) [94].

Para mitigar a subjetividade diagnóstica tradicional e a escassez de especialistas, instituições brasileiras têm contribuído de forma relevante para o fornecimento de dados à comunidade científica internacional. Destacam-se bases como o OCDC (Universidade Federal de Uberlândia), utilizada em estudos de segmentação tumoral. O conjunto compreende 15 imagens de lâminas inteiras (WSIs) de pacientes com carcinoma espinocelular oral, a partir das quais foram obtidos 1.050 *patches* histopatológicos anotados. Em um estudo recente, esses *patches* foram organizados em 840 imagens para treinamento e 210 imagens para teste, com resolução de 512 x 512 pixels [40].

Outro recurso relevante é o NDB-UFES (Universidade Federal do Espírito Santo), que se destaca pela associação entre imagens histopatológicas e dados clínicos e sociodemográficos estruturados [10, 40]. A base reúne 237 imagens obtidas a partir de biópsias, sendo 91 de carcinoma espinocelular oral, 89 de leucoplasia com displasia e 57 de leucoplasia sem displasia. As imagens foram coradas com hematoxilina-eosina (H&E), adquiridas com objetivações de 10x e 40x e apresentam resolução de 2048 x 1536 pixels. Além das características histopatológicas, os registros incluem informações como idade, sexo, tabagismo, consumo de álcool, exposição solar e características das lesões [10].

A relevância desta integração de dados no contexto estabeleceu a prova de conceito empírica para a literatura mundial: pesquisadores demonstraram ganhos de até 30,68% na acurácia balanceada ao realizarem a fusão de imagens com o perfil clínico do paciente [10]. Consequentemente, bases nacionais tornaram-se o padrão-ouro exigido para validação cruzada externa de robustez em estudos asiáticos e árabes. Essa excelência na curadoria de dados multimodais justifica e alicerça o fomento contínuo à pesquisa local. O cenário indica que a academia nacional já domina a fase de estruturação de dados; logo, o próximo passo tecnológico urgente é a transição desses experimentos para *frameworks* de IA implementáveis na prática, evoluindo do simples diagnóstico para sistemas automatizados de recomendação e planejamento terapêutico.

3.2 A predominância unimodal e a lacuna multimodal (RQ3)

A categorização das modalidades de dados utilizadas nos estudos primários evidencia uma das principais lacunas tecnológicas atuais da área. Conforme ilustrado na Figura 4, a maioria da literatura (aproximadamente 88% dos estudos) adota uma abordagem estritamente unimodal visual,

focando no processamento exclusivo de imagens histopatológicas. Abordagens unimodais textuais, envolvendo o processamento de informações clínicas ou prontuários por técnicas de NLP, representam uma parcela reduzida da literatura.

Mais criticamente, os modelos verdadeiramente multimodais — que integram imagens histopatológicas e informações clínicas ou textuais na mesma arquitetura de decisão — correspondem a apenas 6 estudos identificados na literatura analisada. Essa discrepância quantitativa não decorre de uma restrição imposta pela presente revisão, mas reflete a própria distribuição da literatura encontrada durante o processo de busca e seleção. A predominância de abordagens unimodais é acompanhada, inclusive, pelo reconhecimento recorrente, nos próprios estudos, de que a integração de informações clínicas e outras modalidades constitui uma direção necessária para pesquisas futuras.

Por exemplo, o estudo [1], entre suas perspectivas futuras, cita a integração de fontes de dados multimodais como estratégia para ampliar a aplicabilidade clínica e a robustez do modelo. De forma semelhante, o estudo [68] destaca que a incorporação de informações contextuais, como dados demográficos dos pacientes, pode contribuir para uma abordagem diagnóstica mais abrangente e sugere a exploração dessa integração em trabalhos futuros.

Dessa forma, os estudos unimodais não representam trabalhos periféricos ao objetivo desta revisão, mas constituem uma referência importante para caracterizar o estágio de maturidade das modalidades individuais e contextualizar a escassez das abordagens integrativas. A elevada concentração de modelos visuais evidencia que a aplicação de IA à patologia digital apresenta maior maturidade no processamento isolado de imagens, enquanto a integração sistemática desses dados com informações clínicas e textuais permanece pouco explorada. Essa lacuna é particularmente relevante para o desenvolvimento de sistemas de suporte à decisão clínica, nos quais a integração de achados histopatológicos com o histórico e o perfil clínico do paciente pode contribuir para o aprimoramento do diagnóstico e subsidiar etapas posteriores, como a estratificação prognóstica e o planejamento terapêutico.

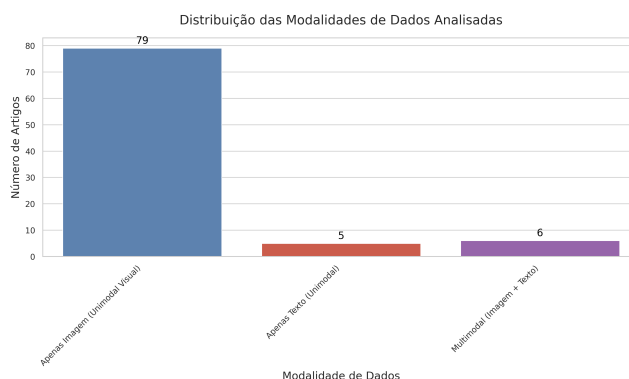


Figure 4: Distribuição dos estudos por modalidade de dados.

3.3 Arquiteturas visuais e a saturação diagnóstica (RQ1)

Para mapear a relação direta entre o esforço computacional e a aplicabilidade clínica das soluções propostas na literatura, os dados extraídos foram consolidados em um gráfico de dispersão ponderada (Figura 5). A análise deste mapeamento fornece a resposta direta à **RQ1**.

Fica evidente a predominância das CNNs tradicionais e a saturação da tarefa de classificação binária. O quadrante superior esquerdo revela que 57 estudos (aproximadamente 64% da amostra) dedicam-se exclusivamente a distinguir tecidos saudáveis de cancerosos utilizando arquiteturas clássicas (como ResNet, VGG e DenseNet).

A literatura demonstra um esforço contínuo para otimizar essa extração de características visuais, evidenciado pelo surgimento de abordagens baseadas em *ensembles* híbridos (7 estudos) e pela adoção recente de ViTs (9 estudos). As métricas de desempenho extraídas confirmam que a área da visão computacional histopatológica atingiu elevada maturidade técnica, com a maioria dos estudos unimodais reportando acurácias e valores de AUC superiores a 95%. Contudo, esse sucesso tecnológico encontra-se concentrado predominantemente em tarefas de diagnóstico do funil oncológico.

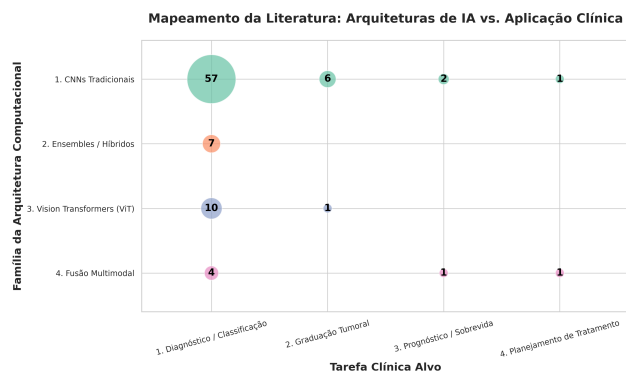


Figure 5: Mapeamento da literatura: cruzamento entre a família de arquitetura computacional adotada e a tarefa clínica resolvida. O tamanho do nó representa o volume de estudos.

3.4 Estratégias de fusão multimodal e modelos textuais (RQ2 e RQ3)

A análise das abordagens multimodais (sintetizadas na Tabela 7) revela uma acentuada heterogeneidade nas estratégias de integração de dados, respondendo diretamente à RQ3. Observou-se a predominância da fusão tardia (*late fusion*), técnica na qual as representações visuais e clínicas são processadas por redes independentes e combinadas apenas na camada final de decisão.

Embora majoritária, a fusão tardia não é a única via explorada. Abordagens de fusão intermediária (*joint fusion*) têm

demonstrado resultados superiores ao integrar as características antes da decisão final. O trabalho [10], por exemplo, obteve seu melhor desempenho (83,24% de acurácia balanceada e AUC de 95%) ao acoplar imagens histopatológicas a dados demográficos estruturados utilizando a estratégia *MetaBlock* (fusão intermediária). De forma contundente, os autores reportaram um salto de 30,68% no desempenho da rede ao incorporar essas variáveis clínicas, comprovando empiricamente o impacto da multimodalidade frente aos modelos estritamente visuais.

Já o estudo [9] adotou uma estratégia distinta ao combinar imagens histopatológicas com informações clínicas textuais. As características visuais foram extraídas por uma CNN, enquanto as informações textuais foram processadas por BERT. A fusão ocorreu posteriormente por concatenação das representações obtidas pelas duas modalidades. O modelo multimodal alcançou sensibilidade de 90%, especificidade de 85% e F1-score de 89%, superando o modelo baseado apenas em CNN, que apresentou F1-score de 81%. Esse resultado evidencia o possível potencial da incorporação de informações clínicas textuais como complemento às características visuais.

A integração multimodal feita pelo estudo [41], para três fontes de informação: imagens histopatológicas, dados transcriptômicos e informações clínicas, utilizou ResNet50 associada ao CLAM para análise visual, enquanto a integração das diferentes fontes ocorreu na etapa prognóstica por meio de um nomograma associado à regressão de Cox. Embora a classificação do grau histológico tenha sido realizada a partir das imagens, a predição prognóstica utilizou conjuntamente características de imagem, transcriptoma e dados clínicos. O nomograma apresentou AUC de 71%, 73% e 69% para a predição de sobrevivência em 1, 3 e 5 anos, respectivamente.

O estudo [39] combinou imagens histopatológicas e informações provenientes de EHR. Diferentes *backbones* visuais foram avaliados, incluindo ViT-Small e Swin-Base, enquanto ClinicalBERT foi empregado para representação das informações clínicas. O modelo baseado em ViT-Small e PathoBench apresentou AUC média de 75,00% nas seis tarefas avaliadas, com destaque para a predição de recorrência, que atingiu AUC de 93,13%. Entretanto, o estudo também evidenciou uma diferença entre as modalidades: o modelo baseado exclusivamente em informações patológicas apresentou AUC média de 73,59%, enquanto o modelo clínico baseado em texto apresentou 61,99%.

Utilizando imagens histopatológicas associadas a informações clínicas em um *pipeline* voltado à estratificação de risco, o estudo [67], realizou o processamento visual empregando EfficientNet para classificação e detecção de invasão vascular e YOLO-NAS para identificação de características histopatológicas, como pólulas de queratina e nucléolos múltiplos. A integração ocorreu em uma etapa posterior por meio de critérios de pontuação clínica, combinando as características detectadas com informações do paciente. O sistema alcançou 99,25% de acurácia na classificação de carcinoma espinocelular oral, 92,5% de precisão na detecção de pólulas de queratina e 99,1% de acurácia na identificação de invasão vascular.

Table 7: Comparação dos estudos multimodais identificados na revisão sistemática.

Estudo	Modalidade visual	Modalidade complementar	Estratégia de fusão	Mecanismo	NLP/LLM	Aplicação
[10]	Imagem histopatológica	Dados clínicos e demográficos	Intermediária e tardia	MetaBlock, MetaNet, MAT e concatenação	Não	Diagnóstico
[9]	Imagem histopatológica	Texto clínico não estruturado	Tardia	CNN + BERT + concatenação	BERT	Diagnóstico
[41]	Imagem histopatológica	Transcriptoma + dados clínicos	Tardia	Nomograma + regressão de Cox	Não	Prognóstico
[39]	Imagens histopatológicas	Prontuários eletrônicos (EMR)	Tardia	Concatenação, LMF, Gated Fusion e Transformer	ClinicalBERT	Diagnóstico e prognóstico
[67]	Imagem histopatológica	Dados clínicos	Tardia	Sistema de pontuação clínica	Não	Estratificação de risco
[40]	Imagem histopatológica	Texto gerado por LLM	Tardia	Grad-CAM + GPT-4	GPT-4	Segmentação e explicabilidade

Por fim, o estudo [40] apresentaram uma configuração multimodal com uma finalidade distinta. O OCANet utiliza imagens histopatológicas para a tarefa principal de segmentação, enquanto o GPT-4 é empregado para gerar descrições textuais associadas às representações visuais e aos mapas de ativação. Dessa forma, a modalidade textual não participa diretamente da predição diagnóstica, mas da explicabilidade do modelo. No dataset ORCA, o método alcançou 90,98% de acurácia, 86,14% de Dice e 77,10% de mIoU, enquanto no OCDC atingiu 98,24% de acurácia, 94,09% de Dice e 88,84% de mIoU.

Em conjunto, os seis estudos demonstram que o conceito de multimodalidade ainda é empregado de maneira heterogênea na literatura. Em alguns trabalhos, como [9, 10, 39, 41], as modalidades complementares participam diretamente da predição ou do prognóstico. Em [67], as informações clínicas são utilizadas posteriormente para a estratificação de risco, enquanto em [40] o componente textual é empregado principalmente para explicabilidade. Portanto, a presença de mais de uma modalidade não implica necessariamente uma integração conjunta no processo de decisão.

Outro aspecto relevante diz respeito ao processamento das informações textuais. Apesar da presença de BERT, ClinicalBERT e GPT-4 em parte dos estudos, a utilização de modelos de linguagem ainda é limitada quando comparada ao volume de trabalhos exclusivamente visuais. Além disso, os estudos analisados não empregam *retrieval-augmented generation* (RAG) para recuperar conhecimento externo durante a inferência. A integração de informações clínicas ocorre predominantemente por meio de variáveis estruturadas ou representações extraídas por modelos específicos, em vez da recuperação dinâmica de documentos clínicos, literatura científica ou diretrizes terapêuticas.

Entre os estudos que incorporam dados clínicos estruturados, foram identificados algoritmos tradicionais de aprendizado de máquina, como *K-Nearest Neighbors*, *Random Forest*, *Support Vector Machines*, XGBoost, regressão logística e AdaBoost [4, 21, 24, 27, 66]. Também foram observadas abordagens baseadas em *deep learning*, incluindo redes neurais densamente conectadas e estratégias que combinam *deep learning* e lógica *fuzzy* para lidar com incertezas presentes nos dados clínicos [24, 27, 66]. Essas abordagens evidenciam que a integração clínica existente na literatura é majoritariamente baseada em características estruturadas, como idade, hábitos de tabagismo e consumo de álcool, estadiamento TNM e informações genômicas [4, 21, 24, 27, 66].

3.5 A lacuna entre: o diagnóstico e planejamento clínico (RQ2, RQ3 e RQ4)

Embora o RAG não tenha sido identificado nos estudos analisados, alguns trabalhos apontam como perspectivas futuras a utilização de NLP para extração de informações a partir de textos clínicos não estruturados e a exploração de arquiteturas baseadas em *transformers* [24, 27]. Esse cenário evidencia uma oportunidade para o desenvolvimento de sistemas capazes de combinar características visuais, dados clínicos estruturados, texto não estruturado e conhecimento externo recuperável.

Do ponto de vista clínico, os resultados também revelam uma importante limitação. A maioria dos estudos multimodais concentra-se em diagnóstico, classificação ou prognóstico, enquanto a utilização integrada dessas modalidades para recomendação terapêutica ou planejamento de tratamento permanece pouco explorada. Assim, embora a literatura já demonstre ganhos de desempenho decorrentes da incorporação de informações complementares, ainda são escassos os sistemas capazes de transformar essa integração em suporte efetivo à tomada de decisão clínica.

Essa lacuna também é observada em trabalhos recentes que exploram a transição de sistemas voltados predominantemente à análise e ao diagnóstico para abordagens orientadas ao planejamento terapêutico. Scaranti e Macedo [95], por exemplo, apresentam uma abordagem de planejamento automático de tratamento odontológico, evidenciando uma perspectiva de aplicação da IA em etapas posteriores ao diagnóstico [95]. Essa direção converge com os resultados desta revisão, que identificou uma representação ainda reduzida de aplicações voltadas ao planejamento clínico e terapêutico.

A Figura 5 expõe visualmente o que pode ser classificado como “lacuna entre diagnóstico e planejamento clínico” da patologia digital em câncer oral.

Quando o foco metodológico avança da simples detecção (tarefa na qual a máquina supera a fadiga visual humana) para a tomada de decisão médica complexa (prognóstico e planejamento), o volume de pesquisas despenca drasticamente. Dos 90 artigos validados, apenas 3 abordam predição de sobrevida e prognóstico [21, 39, 88] e meros 2 estudos ousam propor *frameworks* de planejamento de tratamento ou recomendação clínica [27, 41].

A justificativa para essa lacuna está intrinsecamente ligada à ausência de fusão multimodal (RQ3). O planejamento terapêutico do carcinoma espinocelular oral não pode ser derivado exclusivamente dos pixels de uma lâmina de biópsia. Ele exige o cruzamento do laudo histopatológico com o estadiamento do paciente, histórico de tabagismo, etilismo, comorbidades e diretrizes oncológicas — dados essencialmente não estruturados presentes em prontuários textuais.

O gráfico ilustra que a intersecção ideal para sistemas de suporte à decisão modernos (fusão multimodal aplicada ao planejamento de tratamento) contém **apenas 1 estudo** na literatura mundial recente. Este cenário evidencia que a área está carente de arquiteturas que emulem o raciocínio holístico do comitê médico (tumor *board*). A transição de IA “focada na imagem” para “focada no paciente” através da multimodalidade representa, portanto, a principal fronteira de pesquisa em odontologia oncológica para os próximos anos.

3.6 Explicabilidade e limitações metodológicas para aplicação clínica (RQ5)

Apesar das altas métricas de desempenho computacional reportadas na literatura, a translação destes modelos para a prática clínica enfrenta barreiras de confiabilidade. A análise da **RQ5** revela que a oncologia odontológica computacional ainda sofre da síndrome da “caixa-preta”.

Conforme mapeado na extração de dados (Figura 6), uma parcela considerável dos estudos não implementa nenhum mecanismo de XAI. Entre os trabalhos que buscam mitigar a opacidade algorítmica, a solução quase unânime recai sobre métodos de explicabilidade *post-hoc* visual, com o amplo domínio de mapas de ativação de gradiente (como o Grad-CAM). Embora úteis para indicar em qual região da lâmina a rede neural “focou”, os mapas de calor são insuficientes para justificar decisões clínicas complexas ou gerar laudos textuais que emulem o raciocínio de um patologista. Observou-se a

virtual inexistência de sistemas guiados por regras lógicas ou agentes explicativos em linguagem natural.

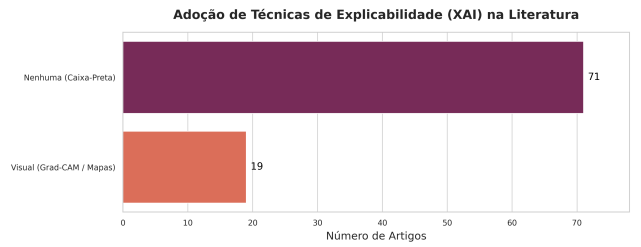


Figure 6: Adoção de técnicas de explicabilidade (XAI) para transparência diagnóstica.

Para consolidar as barreiras que impedem o avanço da área, as limitações reportadas de forma recorrente pelos próprios autores originais foram extraídas qualitativamente e categorizadas através de uma taxonomia temática. A Tabela 8 sintetiza os quatro grandes pilares de deficiência metodológica do atual estado da arte.

Table 8: Taxonomia das principais limitações metodológicas reportadas na literatura de IA para câncer oral.

Categoria de falha	Descrição das limitações reportadas
1. Escassez e desequilíbrio de dados	Amostras reduzidas que elevam o risco de <i>overfitting</i> . Frequente desequilíbrio de classes (ex.: disparidade entre imagens tumorais <i>versus</i> tecido normal, ou entre diferentes graus), o que prejudica a sensibilidade e o <i>recall</i> nas classes minoritárias.
2. Generalização e viés de origem	Modelos desenvolvidos em repositórios de centro único (<i>single-center</i>) ou populações muito restritas. Os algoritmos falham em manter a robustez frente a variações de diferentes <i>scanners</i> , protocolos de coloração de lâminas e diversidade étnica e geográfica global.
3. Opacidade algorítmica	Adoção de arquiteturas profundas (CNNs e ViTs) que atuam como “caixas-pretas”. A falta de interpretabilidade e a dificuldade de traduzir a lógica de decisão para a linguagem dos patologistas clínicos geram desconfiança e barram a adoção na rotina médica.
4. Fragmentação unimodal	Reconhecimento, por parte dos pesquisadores, de que prever riscos e recomendar tratamentos analisando apenas matrizes de pixels é insuficiente sem o cruzamento contextual com o histórico e os dados estruturados e textuais do paciente.

A consolidação dessas limitações fundamenta a premissa de que o campo atingiu um teto de vidro metodológico. Ultrapassá-lo exigirá o abandono de arquiteturas puramente visuais e opacas desenvolvidas em silos de dados pequenos. O avanço clínico real depende da transição para *frameworks* multimodais integrados, dotados de LLMs ou agentes capazes de fornecer explicabilidade textual e recomendações terapêuticas rastreáveis e generalizáveis.

3.7 Avaliação de qualidade metodológica

A aplicação do formulário de avaliação de qualidade revelou nuances críticas sobre o rigor metodológico adotado no estado da arte. A Figura 7 ilustra o percentual de conformidade dos 90 estudos primários em relação aos cinco critérios estabelecidos.

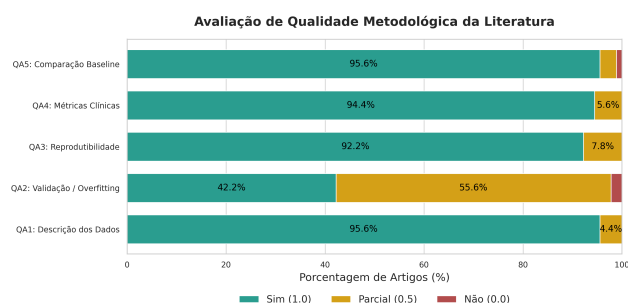


Figure 7: Resultados da avaliação de qualidade metodológica dos estudos primários (QA1–QA5).

Observa-se que a maioria absoluta da literatura obteve alto índice de conformidade nos critérios QA1 (descrição dos dados) e QA4 (métricas clínicas), evidenciando que os autores são transparentes quanto aos *datasets* públicos utilizados e adotam métricas robustas (como AUC e F1-score) para contornar o desbalanceamento de classes, não se limitando apenas à acurácia. A QA5 (comparação com *baselines*) também apresentou resultados positivos, indicando que a comunidade busca comprovar o avanço tecnológico em relação a métodos anteriores.

Em contrapartida, o critério QA2 (estratégia de validação) destacou-se como a principal fragilidade da área. Uma parcela expressiva dos estudos recebeu pontuação “Parcialmente (0,5)” ou “Não (0,0)” neste quesito. Isso ocorre porque a literatura frequentemente restringe a avaliação a divisões *hold-out* ou validações cruzadas (*k-fold*) dentro do mesmo *dataset* (validação interna). A escassez de avaliações em coortes clínicas independentes (validação externa multi-institucional) alerta para um grave risco de *overfitting*, o que compromete a capacidade de generalização e a implantação segura desses algoritmos em ambientes hospitalares do mundo real.

4 CONCLUSÃO

Esta revisão sistemática da literatura mapeou o panorama tecnológico da IA aplicada à oncologia odontológica, analisando 90 estudos primários de alta relevância publicados

na fronteira de 2021 a 2026. Os achados demonstram que a visão computacional atingiu um elevado nível de excelência na extração de características de imagens histopatológicas de carcinoma espinocelular oral. Arquiteturas baseadas em CNNs profundas e ViT consolidaram o estado da arte na tarefa de classificação diagnóstica binária, frequentemente ultrapassando 95% de acurácia.

Contudo, a análise crítica evidenciou lacunas metodológicas e clínicas severas que impedem a translação dessas tecnologias para o ambiente hospitalar real. A literatura atual é predominantemente unimodal (focada em matrizes de pixels) e negligencia o contexto clínico holístico do paciente, presente nos dados não estruturados de EHR. A escassez de abordagens verdadeiramente multimodais e a incipiente aplicação de técnicas de NLP revelam que a transição de um “diagnóstico algorítmico isolado” para um sistema de suporte à decisão clínica ainda é um desafio em aberto. Ademais, a forte predominância de modelos “caixa-preta”, desprovidos de XAI, perpetua a falta de confiança por parte dos patologistas e oncologistas.

Conclui-se, portanto, que o limite tecnológico atual não reside na melhoria marginal da acurácia de imagens isoladas, mas sim na capacidade de integração de dados heterogêneos. Como trabalhos futuros, esta pesquisa fundamenta e justifica a urgência do desenvolvimento de novos *frameworks* multimodais que processem simultaneamente biópsias histopatológicas e laudos clínicos. Tais arquiteturas — preferencialmente guiadas por LLMs, sistemas multiagentes e ancoradas em regras explicáveis — são metodologicamente essenciais para avançar além do diagnóstico simples, viabilizando o planejamento automatizado e personalizado de tratamento e otimizando os fluxos de decisão no SUS e na oncologia global.

CUIDADOS ÉTICOS

Não se aplica.

ACKNOWLEDGMENTS

O presente trabalho foi realizado com apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), nos processos de números 2024/15912-0, 2026/11045-5, 2026/0676 7-1 e 2026/04184-9.

Declaração sobre o uso de Inteligência Artificial: Durante a preparação deste trabalho, os autores utilizaram *NotebookLM (Gemini 1.5 Pro)* como assistente de extração algorítmica de dados dos artigos primários. Adicionalmente, a IA Generativa foi empregada para auxiliar na revisão gramatical, aprimoramento da clareza da escrita acadêmica. Ressalta-se que a IA não foi utilizada para a concepção das ideias originais, análises críticas ou conclusões científicas. Os autores revisaram detalhadamente todo o conteúdo gerado e assumem total e integral responsabilidade pelo artigo final publicado.

REFERENCES

- [1] Francis Rudra D. Cruze, Jeba Wasima, Md. Faruk Hosen, et al. 2026. Oral Cancer Diagnosis Using Histopathology Images: An

- Explainable HybridTransformer Framework. *TECHNOLOGIES* 14 (2026). doi:10.3390/technologies14010039
- [2] Dawit Kiros Redie, Saurabh Bilgaiyan, and Santwana Sagnika. 2023. Oral cancer detection using transfer learning-based framework from histopathology images. *JOURNAL OF ELECTRONIC IMAGING* 32 (2023). doi:10.1117/1.JEI.32.5.053004
 - [3] Yang Xu, Yuan Hong, Xinchun Li, et al. 2024. MedTrans: Intelligent Computing for Medical Diagnosis Using Multiscale Cross-Attention Vision Transformer. *IEEE ACCESS* 12 (2024), 146575–146586. doi:10.1109/ACCESS.2024.3450121
 - [4] John Adeoye, Li-Wu Zheng, Peter Thomson, et al. 2023. Explainable ensemble learning model improves identification of candidates for oral cancer screening. *Oral Oncology* 136 (2023). doi:10.1016/j.oraloncology.2022.106278
 - [5] Gurjot Kaur and Neha Sharma. 2024. Automated Detection of Oral Squamous Cell Carcinoma using Transfer Learning Models from Histopathological Images. *2024 3rd International Conference for Advancement in Technology, ICONAT 2024* (2024). doi:10.1109/ICONAT61936.2024.10775135
 - [6] Jelena Musulin, Daniel Stifanić, Ana Zulijani, et al. 2021. An enhanced histopathology analysis: An ai-based system for multi-class grading of oral squamous cell carcinoma. *Cancers* 13 (2021). doi:10.3390/cancers13081784
 - [7] A. Parkavi, Yash Tiriya, Pragyan Jyoti Borthakur, et al. 2023. Deep Learning Techniques for the Detection and Classification of Oral Cancer Using Histopathological Images. *Proceedings of the International Conference on Circuit Power and Computing Technologies, ICCPCT 2023* (2023), 1625 – 1630. doi:10.1109/ICCPCT58313.2023.10244890
 - [8] Heba M. Afify, Kamel K. Mohammed, and Aboul Ella Hassanien. 2023. Novel prediction model on OSCC histopathological images via deep transfer learning combined with Grad-CAM interpretation. *Biomedical Signal Processing and Control* 83 (2023). doi:10.1016/j.bspc.2023.104704
 - [9] K. Shantha Kumari, S. Udhyashree, and C. Calarany. 2024. Integrating Multimodal AI Models for Enhanced Oral Cancer Detection and Diagnosis. *2nd IEEE International Conference on Integrated Intelligence and Communication Systems* (2024). doi:10.1109/ICIICS63763.2024.10859605
 - [10] Leandro Muniz de Lima, Maria Clara Falcão Ribeiro de Assis, Júlia Pessini Soares, et al. 2023. Importance of complementary data to histopathological image analysis of oral leukoplakia and carcinoma using deep neural networks. *Intelligent Medicine* 3 (2023), 258 – 266. doi:10.1016/j.imed.2023.01.004
 - [11] Barbara A. Kitchenham and Stuart Charters. 2007. *Guidelines for Performing Systematic Literature Reviews in Software Engineering*. Technical Report EBSE-2007-01. Keele University and Durham University. https://www.elsevier.com/___data/promis_misc/525444systematicreviewsguide.pdf Version 2.3, 9 July 2007.
 - [12] Matthew J. Page, Joanne E. McKenzie, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, Cynthia D. Mulrow, Larissa Shamseer, Jennifer M. Tetzlaff, Elie A. Akl, Sue E. Brennan, et al. 2021. The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *BMJ* 372 (2021), n71. doi:10.1136/bmj.n71
 - [13] Parsifal. 2025. Parsifal: An Online Tool for Systematic Literature Reviews. <https://parsifal.ai>. Acesso em: 30 abr. 2026.
 - [14] Eranjoli Nalupurakkal Subhija, Sandhya Tamgadge, and Vaninirappuputhenpurayil Gopalan Reju. 2026. Patch-Based Oral Cancer Detection Using Vision Transformers and Quantitative Patch Scoring. *INTERNATIONAL JOURNAL OF IMAGING SYSTEMS AND TECHNOLOGY* 36 (2026). doi:10.1002/ima.70339
 - [15] Mehran Ahmad, Muhammad Abeer Irfan, Umar Sadique, et al. 2023. Multi-Method Analysis of Histopathological Image for Early Diagnosis of Oral Squamous Cell Carcinoma Using Deep Learning and Hybrid Techniques. *Cancers* 15 (2023). doi:10.3390/cancers15215247
 - [16] P. Misra and A. Chaurasia. 2025. Quantitative Analysis of CNN models for Identifying OSCC. *2025 2nd International Conference on Artificial Intelligence for Innovations in Healthcare Industries* (2025), 1–4. doi:10.1109/ICAIIHI67124.2025.11403121
 - [17] R. Dharani and K. Danesh. 2025. A dual deep learning framework using MaskShiftNet and transformer-augmented CNNs for oral cancer diagnosis from histopathology images. *Intelligence-Based Medicine* 12 (2025). doi:10.1016/j.ibmed.2025.100301
 - [18] R.R. Herdiantoputri, D. Komura, M. Ochi, et al. 2025. Preclinical Evaluation of an Interactive Image Search System of Oral Pathology. *Journal of Dental Research* 104 (2025), 1192 – 1201. doi:10.1177/00220345251329042
 - [19] Kaori Oya, Kazuma Kokomoto, Kazunori Nozaki, and Satoru Toyosawa. 2023. Oral squamous cell carcinoma diagnosis in digitized histological images using convolutional neural network. *JOURNAL OF DENTAL SCIENCES* 18 (2023), 322–329. doi:10.1016/j.jds.2022.08.017
 - [20] Eshat Ahmad Shuvo, Md. Shovon, Md. Nazmul Sarkar, et al. 2025. Optimized Hybrid Cascaded Approach for Accurate Oral Cancer Detection in Histopathology Images Using Deep CNNs. *2025 2nd International Conference on Next-Generation Computing, IoT and Machine Learning* (2025). doi:10.1109/NCIM65934.2025.11160086
 - [21] Sheng-Yao Huang, Ren-Jun Hsu, Dai-Wei Liu, et al. 2025. Using a machine learning algorithm and clinical data to predict the risk factors of disease recurrence after adjuvant treatment of advanced-stage oral cavity cancer. *TZU CHI MEDICAL JOURNAL* 37 (2025), 91–98. doi:10.4103/tcmj.tcmj_56_24
 - [22] Gurjot Kaur, Neha Sharma, and Rupesh Gupta. 2024. Enhanced Detection of Oral Squamous Cell Carcinoma from Histopathological Images using Convolutional Neural Network Architectures. (2024), 1525–1530. doi:10.1109/ICOICI62503.2024.10696099
 - [23] M. Meeyappan, Aniket Verma, Aarush Kaushik, et al. 2024. Oral Cancer detection using Histopathology Images. *2nd IEEE International Conference on Advances in Information Technology* (2024). doi:10.1109/ICAIT61638.2024.10690424
 - [24] Neel Shimpi, Ingrid Glurich, Reihaneh Rostami, et al. 2022. Development and Validation of a Non-Invasive, Chairside Oral Cavity Cancer Risk Assessment Prototype Using Machine Learning Approach. *JOURNAL OF PERSONALIZED MEDICINE* 12 (2022). doi:10.3390/jpm12040614
 - [25] Eranjoli Nalupurakkal Subhija and Vaninirappuputhenpurayil Gopalan Reju. 2026. INTEGRATING HARALICK TEXTURES AND SHAP VALUES FOR INTERPRETABLE PATCH-BASED ORAL CANCER CLASSIFICATION. *BIOMEDICAL ENGINEERING-APPLICATIONS BASIS COMMUNICATIONS* 38 (2026). doi:10.4015/S101623722550053X
 - [26] Vinod Kumar Venkatesan, V. Sujatha, Audithan Sivaraman, et al. 2025. METAHEURISTIC-AI ENHANCED CUSTOM DEEP LEARNING NETWORK OPTIMIZED WITH SAND CAT SWARM ALGORITHM FOR ORAL CANCER DIAGNOSIS. *International Journal Of Advances In Signal And Image Sciences* 11 (2025), 66 – 79. doi:10.29284/ijasis.11.1.2025.66-79
 - [27] Rachasak Somyanonthanakul, Kritsathit Warin, Siththiand Jina-porntham Chaowchuen, et al. 2024. Survival estimation of oral cancer using fuzzy deep learning. *BMC ORAL HEALTH* 24 (2024). doi:10.1186/s12903-024-04279-6
 - [28] Chiagoziem C. Ukwuoma, Qin Zhiguang, Md Belal Bin Heyat, et al. 2022. Detection of Oral Cavity Squamous Cell Carcinoma from Normal Epithelium of the Oral Cavity using Microscopic Images. (2022), 29–36. doi:10.1109/DASA54658.2022.9765023
 - [29] V.G. Janani, S. Vasuki, S.Praveen Kumar, et al. 2025. Deep Learning for Oral Cancer Detection: AFF-U-Net Performance Analysis. *International Conference on Trends in Engineering Systems and Technologies* (2025). doi:10.1109/ICTEST64710.2025.11042274
 - [30] Singaraju Ramya, R. I. Minu, and K. T. Magesh. 2025. Xception Spiking Fractional Neural Network for Oral Squamous Cell Carcinoma Classification Based on Histopathological Images. *IEEE ACCESS* 13 (2025), 98574–98591. doi:10.1109/ACCESS.2025.3574327
 - [31] S. Y. Yang, S. H. Li, J. L. Liu, et al. 2022. Histopathology-Based Diagnosis of Oral Squamous Cell Carcinoma Using Deep Learning. *JOURNAL OF DENTAL RESEARCH* 101 (2022), 1321–1327. doi:10.1177/00220345221089858
 - [32] Arunabha Tarafdar, Atriya Haldar, Susmita Das, et al. 2025. Evaluating Deep Learning Models for Histopathologic Oral Cancer Detection. *2025 5th Asian Conference on Innovation in Technology, ASIANCON 2025* (2025). doi:10.1109/ASIANCON66527.2025.11281308
 - [33] Itisham Ul Haq, Mehran Ahmed, Muhammad Assam, et al. 2023. Unveiling the Future of Oral Squamous Cell Carcinoma Diagnosis: An Innovative Hybrid AI Approach for Accurate Histopathological Image Analysis. *IEEE ACCESS* 11 (2023), 118281–118290. doi:10.1109/ACCESS.2023.3326152
 - [34] Lingyan Zhang, Wuyang Xiong, Daochuan Wu, et al. 2025. Mul2UNet: Deformable Multi-Scale UNet With Decoupled Channel Fusion for Oral Squamous Cell Carcinoma Segmentation.

- IEEE Transactions on Consumer Electronics* 71 (2025), 8477 – 8491. doi:10.1109/TCE.2025.3582774
- [35] Shafayet Rajit and Md. Abdullah Al Sayed. 2024. Federated Learning Based Histopathological Image Classification for Oral Squamous Cell Carcinoma. *Proceedings - 8th IEEE-EMBS Conference on Biomedical Engineering and Sciences* (2024), 339 – 344. doi:10.1109/IECBES61011.2024.10991111
- [36] Bhaswati Singha Deo, Mayukha Pal, Prasanta K. Panigrahi, et al. 2025. Supremacy of attention-based transformer in oral cancer classification using histopathology images. *INTERNATIONAL JOURNAL OF DATA SCIENCE AND ANALYTICS* 20 (2025), 969–987. doi:10.1007/s41060-023-00502-9
- [37] Jelena Stifanic, Daniel Stifanic, Nikola Andelic, et al. 2025. Explainable AI for Oral Cancer Diagnosis: Multiclass Classification of Histopathology Images and Grad-CAM Visualization. *BIOLOGY-BASEL* 14 (2025). doi:10.3390/biology14080909
- [38] A. A. Abd El-Aziz, Mahmood A. Mahmood, and Sameh Abd El-Ghany. 2025. Enhancing Early Detection of Oral Squamous Cell Carcinoma: A Deep Learning Approach with LRT-Enhanced EfficientNet-B3. *Diagnostics* 15 (2025). doi:10.3390/diagnostics15131678
- [39] Jinquan Guan, Junhong Guo, Qi Chen, et al. 2026. A High Magnifications Histopathology Image Dataset for Oral Squamous Cell Carcinoma Diagnosis and Prognosis. *SCIENTIFIC DATA* 13 (2026). doi:10.1038/s41597-026-06736-z
- [40] Syed Jawad Hussain Shah, Ahmed Albishri, Rong Wang, et al. 2025. Integrating local and global attention mechanisms for enhanced oral cancer detection and explainability. *Computers in Biology and Medicine* 189 (2025). doi:10.1016/j.compbiomed.2025.109841
- [41] Tingru Shao, Peirong Ni, Chun Wang, et al. 2025. Prediction of pathological grade of oral squamous cell carcinoma and construction of prognostic model based on deep learning algorithm. *DISCOVER ONCOLOGY* 16 (2025). doi:10.1007/s12672-025-02144-8
- [42] Aradhana Soni, Prabira Kumar Sathy, Amit Kumar Dewangan, et al. 2024. Enhancing oral squamous cell carcinoma detection: a novel approach using improved EfficientNet architecture. *BMC ORAL HEALTH* 24 (2024). doi:10.1186/s12903-024-04307-5
- [43] Suliman Mohamed Fati, Ebrahim Mohammed Senan, and Yasir Javed. 2022. Early Diagnosis of Oral Squamous Cell Carcinoma Based on Histopathological Images Using Deep and Hybrid Learning Approaches. *Diagnostics* 12 (2022). doi:10.3390/diagnostics12081899
- [44] Ziang Xu, Jiakuan Peng, Xin Zeng, et al. 2022. High-Accuracy Oral Squamous Cell Carcinoma Auxiliary Diagnosis System Based on EfficientNet. *FRONTIERS IN ONCOLOGY* 12 (2022). doi:10.3389/fonc.2022.894978
- [45] Mahanty C, Shieh CS, Horng MF, et al. 2026. Explainable vision transformer framework for multi-class classification and prognostic interpretation of oral cancer in histopathology images. (2026). doi:10.1007/s12672-026-04871-y
- [46] Prerna Kulkarni, Nidhi Sarwe, Abhishek Pingale, et al. 2024. Exploring the efficacy of various CNN architectures in diagnosing oral cancer from squamous cell carcinoma. *MethodsX* 13 (2024). doi:10.1016/j.mex.2024.103034
- [47] Ramya Mohan, Arunmozhi Rama, Ramalingam Karthik Raja, et al. 2023. OralNet: Fused Optimal Deep Features Framework for Oral Squamous Cell Carcinoma Detection. *BIOMOLECULES* 13 (2023). doi:10.3390/biom13071090
- [48] Gurjot Kaur, Shefali Gupta, Ashraf Osman Ibrahim, et al. 2024. Enhanced Early Detection of Oral Squamous Cell Carcinoma via Transfer Learning and Ensemble Deep Learning on Histopathological Images. *INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS* 15 (2024), 766–776.
- [49] K. Dhanya, D. Venkata Vara Prasad, and Lokeshwari Y. Venkataramana. 2024. DETECTION OF ORAL SQUAMOUS CELL CARCINOMA USING PRE-TRAINED DEEP LEARNING MODELS. *Experimental Oncology* 46 (2024), 119 – 128. doi:10.15407/exp-oncology.2024.02.119
- [50] V Sumathy and C. Pretty Diana Cyril. 2026. Ensemble deep learning model for early diagnosis of oral Squamous cell Carcinoma from histopathology images. *BIOMEDICAL SIGNAL PROCESSING AND CONTROL* 114 (2026). doi:10.1016/j.bspc.2025.109264
- [51] Trishna Saikia, Ashutosh Dhamaniya, Puneet Gupta, et al. 2025. An Ensemble Technique for Classification of Oral Cancer by Using Histopathological Imaging. *Lecture Notes in Electrical Engineering* 1211 LNEE (2025), 739 – 748. doi:10.1007/978-981-97-4359-9_66
- [52] Mahmoud Ragab and Turkey Omar Asar. 2024. Deep transfer learning with improved crayfish optimization algorithm for oral squamous cell carcinoma cancer recognition using histopathological images. *SCIENTIFIC REPORTS* 14 (2024). doi:10.1038/s41598-024-75330-3
- [53] Alka Kumari, Ashok Kumar Mehta, and Danish Ali Khan. 2026. Concatenated Deep Transfer Learning with Hybridized Optimization Algorithm for Early Prediction and Diagnosis of Oral Cancer Disease. *Journal of Electrical Engineering and Technology* 21 (2026), 1073 – 1082. doi:10.1007/s42835-025-02462-5
- [54] Tuan D. Pham. 2025. Integrating support vector machines and deep learning features for oral cancer histopathology analysis. *BIOLOGY METHODS & PROTOCOLS* 10 (2025). doi:10.1093/biomethods/bpaf034
- [55] Eid Albalawi, Arastu Thakur, Mahesh Thyluru Ramakrishna, et al. 2023. Oral squamous cell carcinoma detection using EfficientNet on histopathological images. *Frontiers in Medicine* 10 (2023). doi:10.3389/fmed.2023.1349336
- [56] Santisudha Panigrahi, Jayshankar Das, and Tripti Swarnkar. 2022. Capsule network based analysis of histopathological images of oral squamous cell carcinoma. *JOURNAL OF KING SAUD UNIVERSITY COMPUTER AND INFORMATION SCIENCES* 34 (2022), 4546–4553. doi:10.1016/j.jksuci.2020.11.003
- [57] Madhusmita Das, Rasmita Dash, Sambit Kumar Mishra, et al. 2024. An Ensemble Deep Learning Model for Oral Squamous Cell Carcinoma Detection Using Histopathological Image Analysis. *IEEE ACCESS* 12 (2024), 127185–127197. doi:10.1109/ACCESS.2024.3450444
- [58] Madhusmita Das, Rasmita Dash, and Sambit Kumar Mishra. 2023. Automatic Detection of Oral Squamous Cell Carcinoma from Histopathological Images of Oral Mucosa Using Deep Convolutional Neural Network. *International Journal of Environmental Research and Public Health* 20 (2023). doi:10.3390/ijerph20032131
- [59] Tuan D. Pham. 2025. CLASS FUSION OF SUPPORT VECTOR MACHINES WITH DEEP LEARNING FEATURES FOR ORAL CANCER HISTOPATHOLOGY CLASSIFICATION. (2025). doi:10.1109/ISBI60581.2025.10980934
- [60] Parthasarathi Pattnayak and Arpeeta Mohanty. 2025. Combining Deep Learning Models to Classify Oral Squamous Cell Carcinoma in Histopathological Pictures. *INDISCON 2025 - IEEE 6th India Council International Subsections Conference* (2025). doi:10.1109/INDISCON66021.2025.11252237
- [61] Akuri Sree Rama Chandra Murthy, Ganta Mercy, Laghuvarapu Jyothi Prakash, et al. 2025. Histopath-DL-OC: Deep Learning for Oral Cancer Prediction from Histopathology Data. *Proceedings of 8th International Conference on Inventive Computation Technologies* (2025), 1190–1197. doi:10.1109/ICICT64420.2025.11004899
- [62] Ashutosh Mohapatra and Rasmita Dash. 2026. A unified framework for oral squamous cell carcinoma detection using machine learning and deep learning techniques. *DISCOVER COMPUTING* 29 (2026). doi:10.1007/s10791-026-10054-8
- [63] Jelena Musulin, Daniel Stifanic, Ana Zulijani, et al. 2021. Automated Grading of Oral Squamous Cell Carcinoma into Multiple Classes Using Deep Learning Methods. (2021). doi:10.1109/BIBE52308.2021.9635261
- [64] Mahamuda Sultana, Suman Bhattacharya, Ananjan Maiti, et al. 2024. A Deep CNN Framework for Oral Cancer Detection Using Histopathology Dataset. *Communications in Computer and Information Science* 1955 CCIS (2024), 239 – 248. doi:10.1007/978-3-031-48876-4_18
- [65] Zijun Guo, Sha Ao, and Bo Ao. 2024. Few-shot learning based oral cancer diagnosis using a dual feature extractor prototypical network. *Journal of Biomedical Informatics* 150 (2024). doi:10.1016/j.jbi.2024.104584
- [66] Rashmi Siddalingappa and Sekar Kanagaraj. 2022. K-nearest-neighbor algorithm to predict the survival time and classification of various stages of oral cancer: a machine learning approach. *F1000Research* 11 (2022), 70. doi:10.12688/f1000research.75469.2
- [67] A. Benagi, D. Narayan, P. K., et al. 2024. Predictive Cloud-Based Application for the Early Detection and Risk Stratification of Oral Cancer. *2024 28th International Computer Science and Engineering Conference (ICSEC)* (2024), 1–6. doi:10.1109/ICSEC62781.2024.10770727

- [68] R. Dharani and K. Danesh. 2025. Optimized deep learning ensemble for accurate oral cancer detection using CNNs and metaheuristic tuning. *Intelligence-Based Medicine* 11 (2025). doi:10.1016/j.ibmed.2025.100258
- [69] T. Saraswathi and V. Murali Bhaskaran. 2023. Classification of Oral Squamous Carcinoma Histopathological images using Alex Net. *Proceedings of the 2023 International Conference on Intelligent Systems for Communication* (2023), 637 – 643. doi:10.1109/ICISCoIS56541.2023.10100510
- [70] Ramya Mohan and Venkatesan Rajinikanth. 2024. Examination of H&E-stained Oral Cavity-Derived Cancer Image: A Study using UDL-Model. *2024 International Conference on System, Computation, Automation and Networking* (2024). doi:10.1109/ICSCAN62807.2024.10893917
- [71] Dalí F. D. dos Santos, Paulo R. de Faria, Bruno A. N. Travençolo, et al. 2023. Influence of Data Augmentation Strategies on the Segmentation of Oral Histological Images Using Fully Convolutional Neural Networks. *Journal of Digital Imaging* 36 (2023), 1608 – 1623. doi:10.1007/s10278-023-00814-z
- [72] P. Fantozzi, S. Alajaji, I. J. Stojanov, et al. 2025. Classifying histological images using cell counting estimators. *2025 IEEE 14th International Conference on Consumer Electronics - Berlin* (2025), 102–107. doi:10.1109/ICCE-Berlin67488.2025.11277969
- [73] Mahmoud Badawy, Hossam Magdy Balaha, Ahmed S. Maklad, et al. 2023. Revolutionizing Oral Cancer Detection: An Approach Using Aquila and Gorilla Algorithms Optimized Transfer Learning-Based CNNs. *BIOMIMETICS* 8 (2023). doi:10.3390/biomimetics8060499
- [74] Marco Acerbis, Nataša Sladoje, and Joakim Lindblad. 2025. A Comparison of Deep Learning Methods for Cell Detection in Digital Cytology. *Lecture Notes in Computer Science* 15726 LNCS (2025), 264 – 277. doi:10.1007/978-3-031-95918-9_19
- [75] Bharanidharan Nagarajan, Sannasi Chakravarthy, Vinod Kumar Venkatesan, et al. 2023. A Deep Learning Framework with an Intermediate Layer Using the Swarm Intelligence Optimizer for Diagnosing Oral Squamous Cell Carcinoma. *Diagnostics* 13 (2023). doi:10.3390/diagnostics13223461
- [76] Jiaxing Li. 2025. Fusion feature-based hybrid methods for diagnosing oral squamous cell carcinoma in histopathological images. *Frontiers in Oncology* 15 (2025). doi:10.3389/fonc.2025.1551876
- [77] S. Senthil Pandi, P. Kumar, R.M. Suchindhar, et al. 2025. Hybrid Deep Learning Framework for Early and Accurate Oral Cancer Detection. *2025 8th International Conference on Circuit, Power and Computing Technologies* (2025), 589–594. doi:10.1109/ICCPCT65132.2025.11176693
- [78] Weibin Liao, Yifan He, Bowen Jiang, et al. 2025. <monospace>HistoMoCo</monospace>: Momentum Contrastive Learning Pre-Training on Unlabeled Histopathological Images for Oral Squamous Cell Carcinoma Detection. *ELECTRONICS* 14 (2025). doi:10.3390/electronics14071252
- [79] Arash Shirani and Niloofar Rastin. 2025. Attention-enhanced DenseNet-121 for histopathological classification of oral cancer. *SIGNAL IMAGE AND VIDEO PROCESSING* 19 (2025). doi:10.1007/s11760-025-05028-0
- [80] Kumar Ankit and Vedit Kumar. 2023. Histopathological Image Based Oral Squamous Cell Carcinoma Classification Using Deep Network Fusion. *2023 10th IEEE Uttar Pradesh Section International Conference* (2023), 1362 – 1367. doi:10.1109/UPCON59197.2023.10434514
- [81] M.E. Paramasivam, B.S. Sriganesh, and S. Sureshkrishna. 2024. Oral Cancer Detection Using Convolutional Neural Network. *4th International Conference on Innovative Practices in Technology and Management 2024* (2024). doi:10.1109/ICIPTM59628.2024.10563232
- [82] Sonam Khattar, Tushar Verma, and Sheenam. 2024. Advancement of Oral Cancer Detection with EfficientResnet: A Hybrid Deep Learning Model. *3rd International Conference on Advances in Computing, Communication and Materials* (2024). doi:10.1109/ICACCM61117.2024.11059208
- [83] Alina Khan, Najiba Yaseen Khan, Punit Kumar, et al. 2025. Oral Cavity Cancer Detection Using Deep Neural Network Models. *2025 7th International Conference on Smart Applications, Communications and Networking* (2025). doi:10.1109/SmartNets65254.2025.11106771
- [84] Dalí F. D. dos Santos, Paulo R. de Faria, Bruno A.N. Travençolo, et al. 2021. Automated detection of tumor regions from oral histological whole slide images using fully convolutional neural networks. *BIOMEDICAL SIGNAL PROCESSING AND CONTROL* 69 (2021). doi:10.1016/j.bspc.2021.102921
- [85] Nigus Wereta Asnake, Aleka Melese Ayalew, and Abebaw Agegne Engda. 2025. Detection of oral squamous cell carcinoma cancer using AlexNet on histopathological images. *DISCOVER APPLIED SCIENCES* 7 (2025). doi:10.1007/s42452-025-06514-3
- [86] Silpiranjan Mishra, Ahmad Danish Rehan, Santhosh Kumar Caliperoumal, et al. 2025. Artificial Intelligence-Assisted Histopathological analysis for the automated diagnosis of oral cancer and precancerous lesions. *2025 International Conference on Meta-verse and Current Trends in Computing* (2025). doi:10.1109/ICMCTC62214.2025.11196422
- [87] Singaraju Ramya, R.I. Minu, and K.T. Magesh. 2023. An Automated Approach for Identification of Oral Squamous Cell Carcinoma based on 2D-ICNN. *Proceedings of the 2023 2nd International Conference on Augmented Intelligence* (2023), 155 – 161. doi:10.1109/ICAISS58487.2023.10250540
- [88] Cai X, Li L, Yu F, et al. 2023. Development of a Pathomics-Based Model for the Prediction of Malignant Transformation in Oral Leukoplakia. (2023). doi:10.1016/j.labinv.2023.100173
- [89] R. Dharani and K. Danesh. 2024. Oral cancer segmentation and identification system based on histopathological images using MaskMeanShiftCNN and SV-UnionNet. *Intelligence-Based Medicine* 10 (2024). doi:10.1016/j.ibmed.2024.100185
- [90] H. J. Gilitwala, R. Sinha, and M. Prasad. 2026. DeepTexBoost OSCC: A Three-Stage Hybrid Classifier Combining GLCM, CNN, and XGBoost for Reliable Oral Cancer Detection. *2026 5th International Conference on Sentiment Analysis and Deep Learning* (2026), 668–675. doi:10.1109/ICSADL67539.2026.11451914
- [91] Santisudha Panigrahi, Bhabani Sankar Nanda, Ruchi Bhuyan, et al. 2023. Classifying histopathological images of oral squamous cell carcinoma using deep transfer learning. *HELIYON* 9 (2023). doi:10.1016/j.heliyon.2023.e13444
- [92] Mahdi Hadilou, Nazanin Mahdavi, Elham Keykha, et al. 2025. Artificial intelligence based vision transformer application for grading histopathological images of oral epithelial dysplasia: a step towards AI-driven diagnosis. *BMC Cancer* 25 (2025). doi:10.1186/s12885-025-14193-x
- [93] S. E. N and R. V. G. 2022. Computer Assisted Recognition of Oral Squamous Cell Carcinoma Using GLCM Features in Haar Wavelet Transformed Space and Random Forest Classifier. *2022 Second International Conference on Next Generation Intelligent Systems* (2022), 1–6. doi:10.1109/ICNGIS54955.2022.10091441
- [94] Instituto Nacional de Câncer. 2026. INCA estima 781 mil novos casos de câncer por ano no Brasil entre 2026 e 2028. <https://www.gov.br/inca/pt-br/assuntos/noticias/2026/inca-estima-781-mil-novos-casos-de-cancer-por-ano-no-brasil-entre-2026-e-2028> Publicado em 23 abr. 2026.
- [95] Alexander Scaranti and Alessandra Macedo. 2025. Planejamento automático de tratamento odontológico. In *Anais Estendidos do XXXI Simpósio Brasileiro de Sistemas Multimídia e Web* (Rio de Janeiro/RJ). SBC, Porto Alegre, RS, Brasil, 7–8. doi:10.5753/webmedia_estendido.2025.16360