Oral Lesion Classification Using Deep Learning with ConvNeXt-Base

Gabriel Bisso Fagundes[1], **Dieine E. B. Schiavon**[1], **Carla D. L. Becker**[1]

[1]Universidade Federal de Ciências da Saúde de Porto Alegre
Porto Alegre, RS, Brasil

{gabriel.fagundes,dieineb,carladiniz}@ufcspa.edu.br

Abstract. *Oral cancer presents high mortality rates when diagnosed late, and the visual similarity between healthy, benign, potentially malignant (OPMD) and malignant (OCA) lesions makes differential diagnosis a significant clinical challenge. This work proposes a ConvNeXt-Base model with transfer learning to classify oral cavity images into these four categories, using 3,000 images from 714 patients with lesion cropping, patient-level stratification and data augmentation. The model achieved 77.69% accuracy, MCC of 0.662 and macro AUC-ROC of 0.918, with AUC of 0.994 for Healthy and 0.943 for OCA. External validation on the SMART-OM dataset [Madan Kumar et al. 2026] yielded 45.65% accuracy, exposing domain shift as a central challenge for deployment in real-world clinical scenarios. The source code is available at (https://github.com/gabrielbisso/Oral-Lesion-Classification-Using-Deep-Learning-with-ConvNeXt-Base)*

Resumo. *O câncer bucal apresenta alta mortalidade quando diagnosticado tardiamente, e a similaridade visual entre lesões saudáveis, benignas, potencialmente malignas (OPMD) e malignas (OCA) torna o diagnóstico diferencial um desafio clínico significativo. Este trabalho propõe um modelo ConvNeXt-Base com transferência de aprendizado para classificar imagens da cavidade oral nessas quatro categorias, utilizando 3.000 imagens de 714 pacientes com recorte de lesões, estratificação por paciente e aumento de dados. O modelo alcançou 77,69% de acurácia, MCC de 0,662 e AUC-ROC macro de 0,918, com AUC de 0,994 para Saudável e 0,943 para OCA. O modelo foi ainda avaliado na base de dados SMART-OM [Madan Kumar et al. 2026], obtendo 45,65% de acurácia, expondo o domain shift como desafio central para implantação em cenários reais. O código fonte está disponível em (https://github.com/gabrielbisso/Oral-Lesion-Classification-Using-Deep-Learning-with-ConvNeXt-Base)*

1. Introdução

O câncer bucal situa-se entre as neoplasias de maior incidência no Brasil e no mundo. Segundo o INCA [Instituto Nacional de Câncer – INCA 2025], estima-se cerca de 17.190 novos casos anuais entre 2026 e 2028, com 7.156 óbitos registrados em 2024, predominantemente entre homens [Instituto Nacional de Câncer – INCA 2026]. A mortalidade permanece elevada em razão do diagnóstico frequentemente tardio: quando identificada em estágios iniciais, a taxa de sobrevida em cinco anos pode superar 80%,

ao passo que casos avançados apresentam prognóstico consideravelmente mais reservado [Dores et al. 2023]. Os principais fatores de risco incluem tabagismo (responsável por aproximadamente 90% dos casos), consumo excessivo de álcool, exposição solar prolongada, infecção pelo HPV e uso de cigarros eletrônicos [Amaral et al. 2022, Sousa et al. 2023, Albino et al. 2025].

Do ponto de vista clínico, os sinais iniciais são frequentemente inespecíficos e indolores, e a semelhança visual entre lesões benignas, potencialmente malignas e malignas dificulta o diagnóstico diferencial mesmo para profissionais experientes [Dores et al. 2023]. Somam-se a isso a escassez de especialistas em regiões periféricas e o acesso limitado a serviços odontológicos [Silva et al. 2023].

Nesse contexto, as Redes Neurais Convolucionais (CNNs) com *transfer learning* representam uma alternativa promissora no auxílio ao diagnóstico clínico: permitem aprender padrões visuais complexos a partir de imagens clínicas, mesmo com conjuntos de dados limitados [Marzouk et al. 2022]. Este trabalho propõe o desenvolvimento e a avaliação de um modelo de classificação capaz de identificar lesões e condições da cavidade oral em quatro categorias clínicas: saudável, benigna, potencialmente maligna (OPMD) e câncer bucal (OCA), utilizando o banco de dados público de [Piyarathne et al. 2024], com 3.000 imagens anotadas de 714 pacientes.

2. Referencial Teórico

2.1. Câncer Bucal

O carcinoma de células escamosas corresponde a aproximadamente 90% dos casos de câncer bucal, originando-se no epitélio oral com potencial invasivo e de disseminação linfática [Silva et al. 2026]. Clinicamente, a doença pode se manifestar como úlceras indolores persistentes, lesões exofíticas ou verrucosas, e alterações de cor e textura da mucosa, destacando-se as placas leucoplásicas e eritroplásicas, ambas consideradas manifestações de alto risco [Pinheiro et al. 2026]. O diagnóstico é frequentemente tardio, reflexo de sinais iniciais inespecíficos, preparo insuficiente dos profissionais e acesso limitado a serviços especializados [Dores et al. 2023]. O processo diagnóstico inclui anamnese, exame intraoral, palpação de linfonodos e confirmação por biópsia [Dhanuthai et al. 2018].

2.2. Pré-processamento de Imagens

O pré-processamento de imagens abrange equalização de histograma, filtragem de ruído, normalização de pixels, conversão de espaço de cor, recorte e redimensionamento, além de ajustes de brilho e contraste. Essa etapa reduz variações decorrentes de diferenças entre equipamentos e protocolos de aquisição, facilitando o aprendizado dos modelos e evitando que heterogeneidades entre imagens comprometam a generalização [Masoudi et al. 2021, Dehbozorgi et al. 2024].

2.3. CNNs e Transfer Learning

CNNs são arquiteturas de aprendizado profundo projetadas para dados com estrutura espacial. Extraem automaticamente características relevantes por operações de convolução, identificando bordas, texturas e formas em diferentes níveis de abstração,

seguidas de camadas de *pooling* e camadas totalmente conectadas para classificação final [Chollet 2021]. O *transfer learning* consiste em aproveitar modelos pré-treinados em grandes bases de dados e adaptá-los a novas tarefas: camadas iniciais são mantidas ou levemente ajustadas (preservando padrões genéricos), enquanto as finais são modificadas para o domínio específico [Shin et al. 2016]. Duas estratégias principais são empregadas: congelamento de camadas e *fine-tuning* [Ribani and Marengoni 2019]. Na área da saúde, onde dados rotulados são escassos, [Marzouk et al. 2022] demonstrou que o *transfer learning* permite alcançar acurácias próximas de 90% mesmo com conjuntos de dados reduzidos.

A eficácia do *transfer learning* neste contexto pode ser atribuída à similaridade entre padrões de textura e cor aprendidos no ImageNet e as características visuais de lesões orais, embora a contribuição relativa de cada fator (padrões espectrais vs. estrutura espacial) permaneça uma questão aberta.

3. Estado da Arte

Os estudos analisados nesta pesquisa foram publicados entre 2024 e 2026 e utilizam exclusivamente o banco de dados de [Piyarathne et al. 2024] com CNNs. A busca foi realizada no Google Acadêmico a partir desse artigo de referência, identificando todos os trabalhos que o citaram. De 35 artigos obtidos, 8 foram selecionados após aplicação dos seguintes critérios de inclusão e exclusão. Como critérios de inclusão, consideraram-se: estudos que utilizaram CNNs no desenvolvimento dos modelos de aprendizado de máquina; trabalhos que empregaram exclusivamente o dataset de referência para o treinamento do modelo; e artigos redigidos em inglês ou português. Foram excluídos: artigos de revisão; estudos que utilizaram metodologias distintas de CNN; e trabalhos que não utilizaram o dataset base ou que o complementaram com outras fontes de dados. Buscas em base de dados como Scopus, PubMed e IEEE Xplore podem ampliar a cobertura da revisão de literatura aqui realizada.

As abordagens multimodais dominam a literatura recente. [Devindi et al. 2024] combinaram MobileNetV3-Large e XGBoost via *early fusion* com metadados de pacientes, alcançando 81% de acurácia e MCC de 0,57, com redução de falsos negativos em pacientes de maior risco. [Wijethunga and Chandrasena 2025] adotaram EfficientNetV2B0 com rede densa para metadados demográficos, aplicando explicabilidade via Grad-CAM, LIME e SHAP, e obtiveram 83,3% de acurácia e AUC de 0,8615. [Mukherjee et al. 2026] integraram análise espectral e dados demográficos com ConvNeXt-v2 e classificação por IHML, alcançando F1 macro de 66,23% e AUC-ROC de 84,45%, com diferenciais de consistência por paciente e estimativa de incerteza. [Dai et al. 2026] propuseram o framework PCL-MFP, combinando imagens (ResNet50), descrições diagnósticas (BERT clínico) e vetores de probabilidade em módulo Transformer, com 79,5% de acurácia e Kappa de 0,71, embora a dependência de descrições geradas por LLM possa limitar a aplicação clínica real.

Abordagens baseadas exclusivamente em CNNs também foram exploradas. [Amien et al. 2025] avaliaram ConvNeXt-Tiny, MobileNetV3-Large e EfficientNet-B0 com *augmentation* sintético e *fine-tuning*, obtendo até 98,21% de acurácia. Contudo, a equalização artificial das classes e a ausência de estratificação por paciente sugerem *data leakage*, comprometendo a confiabilidade das métricas. [Aparna et al. 2025] combi-

naram U-Net e Rede Neural Convolucional Quântica (QCNN), com a U-Net alcançando 95,02% de acurácia e Dice de 0,95, e a QCNN obtendo 74,5%, limitada por ruído quântico e restrições de hardware. Na extração híbrida de características, [Joshi et al. 2025] utilizaram ResNet-50 com SVM e QSVM para classificação binária (câncer oral vs. saudável, 858 imagens), obtendo acurácia de 85,6% e F1 superior a 0,91 com SVM, com o QSVM apresentando desempenho estável porém menor sensibilidade à fusão multimodal.

[Vissamsetti and Shivakanth 2026] utilizam as 3.000 imagens de [Piyarathne et al. 2024], classificando lesões orais em três classes (saudável, OPMD e OSCC, Carcinoma de Células Escamosas Oral, equivalente à classe OCA adotada no presente trabalho), com pré-processamento de redução de ruído, realce de contraste, normalização e aumento de dados. O modelo combina características VGG e ResNet com normalização em lote, *dropout* e pesos pré-treinados, treinado com Adam, *cross-entropy* categórica e *early stopping*, reportando 93,44% de acurácia, 92,67% de *recall* e 96,15% de especificidade, resultados que, como discutido a seguir, devem ser interpretados com cautela dado o contexto metodológico em que foram obtidos.

Esses resultados, contudo, não são diretamente comparáveis aos demais trabalhos revisados. A distribuição reportada (1.200 saudável, 900 OPMD, 900 OSCC) contradiz diretamente a composição original do *dataset*, no qual OCA representa apenas 4,3% das amostras. Como os autores não declaram nenhuma estratégia de rebalanceamento, surgem duas hipóteses igualmente problemáticas: ou o *dataset* foi reamostrado sem transparência metodológica, ou os resultados foram obtidos sobre uma versão modificada que não representa o cenário clínico real. Agrava o quadro a ausência de divisão estratificada por paciente: sem essa separação, imagens do mesmo indivíduo podem coexistir no treino e no teste, configurando *data leakage* e inflacionando artificialmente as métricas. O uso de Grad-CAM como evidência de aprendizado clinicamente relevante, sem qualquer validação quantitativa, constitui generalização metodologicamente inadequada. Já o presente trabalho preserva a distribuição original desbalanceada, realiza balanceamento exclusivamente via pesos de classe, garante divisão estritamente estratificada por paciente, e reporta métricas por classe, expondo explicitamente os desafios de cada categoria.

Tabela 1. Comparativo dos estudos analisados no estado da arte.

Autor(es)	Modelo Principal	Melhor Resultado
Devindi et al. (2024)	MobileNetV3-Large + XGBoost	Ac.: 81%, MCC: 0,57
Aparna et al. (2025)	U-Net + QCNN	Ac.: 95,02%, Dice: 0,95
Amien et al. (2025)	EfficientNet-B0	Ac.: 98,21%*
Mukherjee et al. (2026)	ConvNeXt-v2 + IHML	F1: 66,23%, AUC: 84,45%
Wijethunga e Chandrasena (2025)	EfficientNetV2B0 + metadados	Ac.: 83,3%, AUC: 0,86
Dai et al. (2026)	ResNet50 + BERT + Transformer	Ac.: 79,5%, Kappa: 0,71
Joshi et al. (2025)	ResNet-50 + SVM/QSVM	Ac.: 85,6%, F1: 0,91
Vissamsetti e Shivakanth (2026)	VGG-ResNet customizado	Ac.: 93,44%*

* Resultado possivelmente inflacionado por inconsistências metodológicas.

De modo geral, a literatura demonstra que CNNs com *transfer learning* são eficazes para classificação de lesões bucais. EfficientNet, ResNet e MobileNet foram as arquiteturas mais recorrentes, frequentemente combinadas com *data augmentation* e fusão multimodal. Uma limitação comum é a dificuldade de distinção entre OPMD e lesões benignas, dada a semelhança visual entre essas classes. O desequilíbrio entre classes e a ausência de divisão por paciente permanecem desafios recorrentes, lacunas que o presente trabalho busca endereçar.

4. Materiais e Métodos

4.1. Base de Dados

Foi utilizado o banco de dados de [Piyarathne et al. 2024], composto por 3.000 imagens anotadas de 714 pacientes atendidos no Sri Lanka, capturadas com *smartphones* em condições clínicas reais. As imagens distribuem-se em quatro classes: Saudável (729; 24,3%), Benigno (748; 24,9%), OPMD (1.394; 46,5%) e OCA (129; 4,3%). A distribuição é severamente desbalanceada, com OCA representando apenas 4,3% das amostras. A Figura 1 apresenta exemplos representativos de cada classe, ilustrando a variabilidade visual e as características morfológicas que o modelo foi treinado a distinguir.

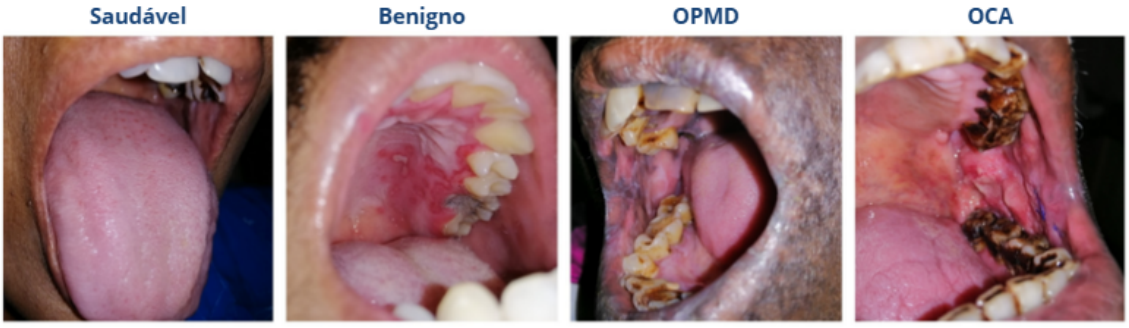


Figura 1. Exemplos de imagens do banco de dados.

4.2. Pré-processamento

Três *pipelines* de pré-processamento foram avaliados: (1) correção de cor e melhora de contraste; (2) recorte da região oral; e (3) recorte exclusivo da região de lesão. Os experimentos indicaram que o terceiro *pipeline* produziu os melhores resultados, concentrando a entrada do modelo nas áreas clinicamente relevantes [Catarino et al. 2025]. A divisão foi realizada de forma estratificada por paciente (70% treino, 15% validação, 15% teste), garantindo que imagens de um mesmo paciente não estivessem em conjuntos distintos, prevenindo *data leakage* e assegurando avaliação realista da capacidade de generalização.

4.3. Arquitetura e Treinamento

Após a avaliação de diferentes configurações de arquitetura e estratégias de *fine-tuning*, a ConvNeXt-Base com *fine-tuning* parcial demonstrou desempenho superior. As camadas iniciais foram congeladas para preservar filtros gerais do pré-treinamento no ImageNet, enquanto as últimas 40 camadas foram ajustadas ao conjunto de dados [Shin et al. 2016].

Sobre a base convolucional, foi adicionada uma *classification head* composta por: GlobalAveragePooling2D → BatchNormalization → Dropout(0,5) → Dense(512, ReLU) → BatchNormalization → Dropout(0,4) → Dense(4, Softmax). O treinamento utilizou *batch size* de 64, limite de 50 épocas e *data augmentation* restrito ao conjunto de treino (rotação $\pm 20°$, espelhamento, zoom $\pm 20\%$, contraste e translação). Para mitigar o desbalanceamento, foram atribuídos pesos proporcionais à frequência de cada classe, com *oversample* $3\times$ da classe OCA no treino. O modelo foi compilado com o otimizador AdamW (taxa de aprendizado: 1×10^{-2}, decaimento de peso: 1×10^{-5}) e função de perda *categorical cross-entropy*. As implementações foram realizadas em Python com Keras/TensorFlow, executadas no Google Colaboratory Pro. O estudo seguiu as recomendações da diretriz CLAIM (*Checklist for Artificial Intelligence in Medical Imaging*) [Tejani et al. 2024], adotada como guia para o relato dos dados, métodos e resultados em pesquisas de inteligência artificial aplicada a imagens médicas.

5. Resultados

5.1. Desempenho Geral

O modelo ConvNeXt-Base foi treinado por 50 épocas sobre 2.842 imagens (treino), com *data augmentation* online aplicado exclusivamente ao conjunto de treino — incluindo *flip* horizontal, rotação de $\pm 20°$, zoom de $\pm 20\%$, variação de contraste de $\pm 20\%$ e translação de $\pm 10\%$. O modelo foi avaliado em 611 imagens (validação) e testado em 632 amostras originais de [Piyarathne et al. 2024], alcançando acurácia geral de 77,69%, MCC de 0,662 e AUC-ROC macro-médio de 0,918. As curvas de treinamento (Figura 2) evidenciam convergência da acurácia de treino em torno de 84,9%, com estabilização da validação em aproximadamente 70% a partir da época 20. O *gap* de aproximadamente 14 pontos percentuais indica *overfitting* moderado, esperado dado o tamanho limitado do banco de imagens, e motiva estratégias adicionais de regularização e busca por mais imagens para esta etapa de treinamento do modelo.

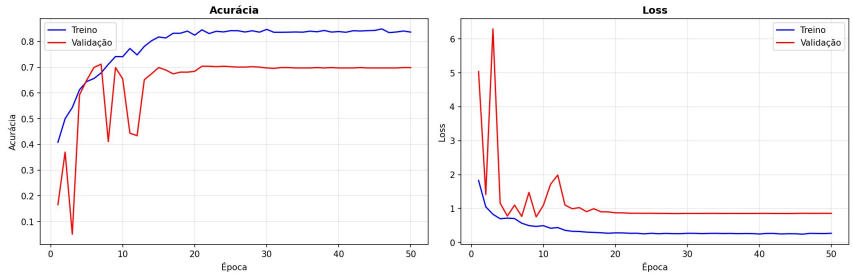


Figura 2. Curvas de acurácia e perda.

5.2. Análise por Classe

A Tabela 2 e a Figura 3 apresentam o desempenho por classe no conjunto de teste.

Tabela 2. Métricas de desempenho por classe (n=632).

Classe	Precisão	Recall	F1	n
Saudável	87,59%	95,24%	91,25%	126
Benigno	70,33%	73,13%	71,71%	201
OPMD	81,44%	73,38%	77,20%	293
OCA	40,91%	75,00%	52,94%	12

A classe Saudável manteve o melhor desempenho, com *recall* de 95,24%, confirmando alta confiabilidade na identificação de pacientes sem lesão. A classe OPMD apresentou *recall* de 73,38% e F1 de 77,20%, com os maiores volumes de erro absoluto: 60 casos classificados como Benigno e 12 como Saudável, refletindo a semelhança visual entre essas categorias. A classe OCA obteve *recall* de 75,00%, com 9 dos 12 casos malignos corretamente identificados, embora a precisão de 40,91% indique falsos positivos relevantes. Para quantificar a incerteza dessas estimativas, foi realizado *bootstrapping* não paramétrico (2.000 reamostragens) que confirmou a estabilidade das métricas globais: acurácia 77,7% (IC 95% [74,5%–81,0%]), MCC 0,663 (IC 95% [0,613–0,711]). Em contraste, observou-se alta incerteza para a classe OCA, com *recall* de 75,4% (IC 95% [50,0%–100,0%]), refletindo o reduzido tamanho amostral (n=12).

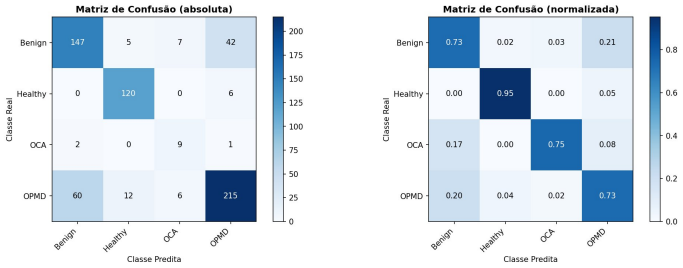


Figura 3. Matriz de confusão.

5.3. Análise ROC e *Precision-Recall*

A curva ROC (Figura 4) evidenciou AUC de 0,994 para Saudável, 0,943 para OCA, 0,872 para OPMD e 0,863 para Benigno, com AUC macro-médio de 0,918. A curva *Precision-Recall* apresentou AP de 0,976 para Saudável, 0,821 para OPMD, 0,789 para Benigno e 0,554 para OCA.

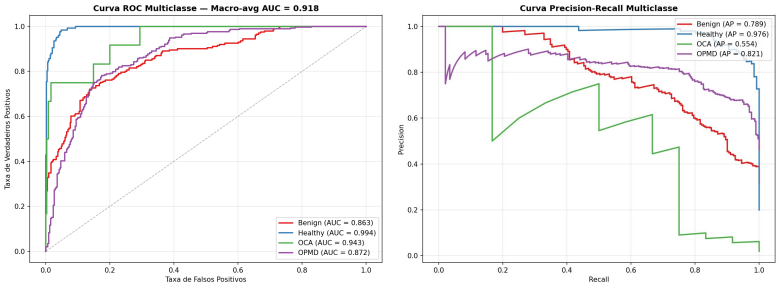


Figura 4. Curva ROC multiclasse e curva *Precision-Recall*.

5.4. Teste de Generalização Externa

O banco SMART-OM [Madan Kumar et al. 2026] é um conjunto de dados público de imagens de mucosa oral capturadas com smartphones Android e iOS em condições clínicas reais, composto por 2.469 imagens RGB de 331 pacientes, sistematicamente organizadas em quatro classes: Normal, Variação do Normal, OPMD e Câncer Oral (OCA), cujos exemplos representativos são ilustrados na Figura 5. Cada imagem foi anotada por cirurgiões-dentistas especialistas utilizando contornos poligonais para delimitar a cavidade oral e as fronteiras das lesões, com metadados clínicos incluindo idade, sexo, diagnóstico clínico e indicadores de risco baseados em estilo de vida. A abordagem de teste externo em banco independente segue a metodologia adotada por [Schiavon et al. 2023], que também avaliou a generalização de modelos CNN em dados de origem distinta.

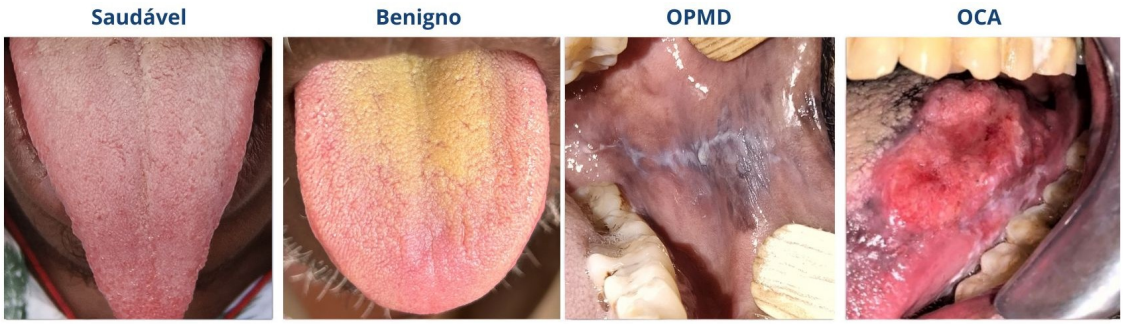


Figura 5. Exemplos de imagens do banco de dados SMART-OM por classe.

No teste externo com o banco SMART-OM [Madan Kumar et al. 2026], as classes foram mapeadas de forma aproximada: *Variation from Normal* corresponde a Benigno, *Normal* a Saudável, *Oral Cancer* a OCA e *OPMD* a OPMD. Esse mapeamento é aproximado, pois os critérios de anotação e as denominações clínicas diferem entre os bancos, podendo introduzir ruído adicional na avaliação. Como a classe OCA dispõe de apenas 20 imagens no banco, utilizar o conjunto completo resultaria em uma avaliação fortemente desbalanceada, penalizando artificialmente as métricas macro. Para garantir uma comparação justa entre classes, optou-se por amostrar 70 imagens por classe — todas as 20 disponíveis para OCA e 70 selecionadas aleatoriamente para as demais —, totalizando 230 imagens.

Tabela 3. Comparação de métricas entre teste interno e teste de fogo (SMART-OM).

Métrica	Teste Interno	Teste de Fogo (SMART-OM)
Loss	0,856	1,471
Acurácia	77,69%	45,65%
Precisão (macro)	0,701	0,496
Recall (macro)	0,792	0,446
Especificidade (macro)	0,914	0,807
AUC-ROC (macro)	0,918	0,672
MCC	0,662	0,244

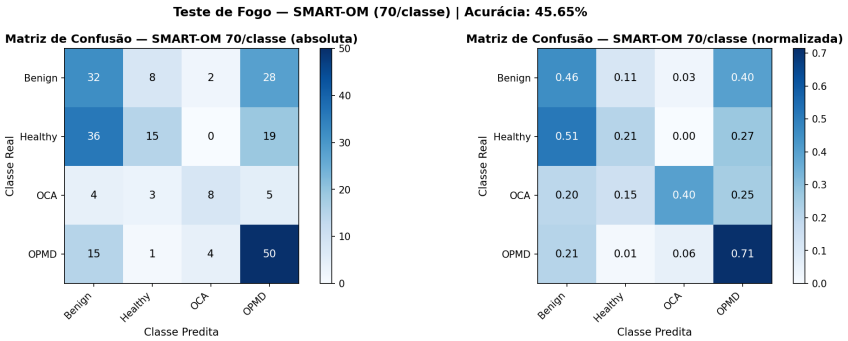


Figura 6. Matriz de confusão — teste externo SMART-OM.

A queda de desempenho entre o teste interno e o externo (77,69% $\rightarrow$ 45,65%) evidencia o impacto do *domain shift* [Schiavon et al. 2023]: diferenças de equipamento (câmeras de smartphone versus câmeras clínicas padronizadas), protocolo de captura, iluminação e critérios de anotação introduzem variabilidade estatística não representada no conjunto de treinamento.

A análise por classe revela um padrão assimétrico. A classe OPMD apresentou a menor degradação de *recall* (73% $\rightarrow$ 71%) e a maior robustez em F1 (0,77 $\rightarrow$ 0,58), sugerindo que suas características visuais — manchas esbranquiçadas, eritroplasia e alterações de textura da mucosa — são suficientemente distintivas para persistir entre domínios. Em contraste, a classe Saudável sofreu a maior queda em *recall* (95% $\rightarrow$ 21%) e em F1 (0,91 $\rightarrow$ 0,31): imagens de mucosa saudável variam amplamente em aparência conforme o equipamento e as condições de iluminação, tornando o modelo sensível a artefatos específicos do domínio de treinamento. OCA manteve *recall* moderado (75% $\rightarrow$ 40%) e apresentou a maior *precision* no teste externo (0,57), o que é relevante dado o caráter maligno da lesão, mas preocupante considerando o baixo número de amostras disponíveis (20 imagens). Para mitigar essas limitações, trabalhos futuros devem explorar técnicas de *domain adaptation*, ajuste fino (*fine-tuning*) com amostras do domínio alvo e harmonização de protocolos de anotação entre bancos.

5.5. Limitações e Encaminhamentos

O modelo confirma a viabilidade da abordagem: acurácia de 77,69%, MCC de 0,662 e AUC-ROC de 0,918 indicam concordância substancial entre predições e rótulos reais, embora o modelo ainda não esteja pronto para aplicação clínica direta. O *overfitting* de 14 pp e a baixa precisão na classe OCA permanecem os principais desafios. O teste externo evidencia a necessidade de estratégias de *domain adaptation* para uso clínico real. Sob os mesmos dados e hiperparâmetros, uma ResNet50 (Figura 7) obteve 69,30% de acurácia e MCC de 0,522 — inferior ao ConvNeXt-Base em todas as métricas, confirmando o ganho específico da arquitetura para este problema.

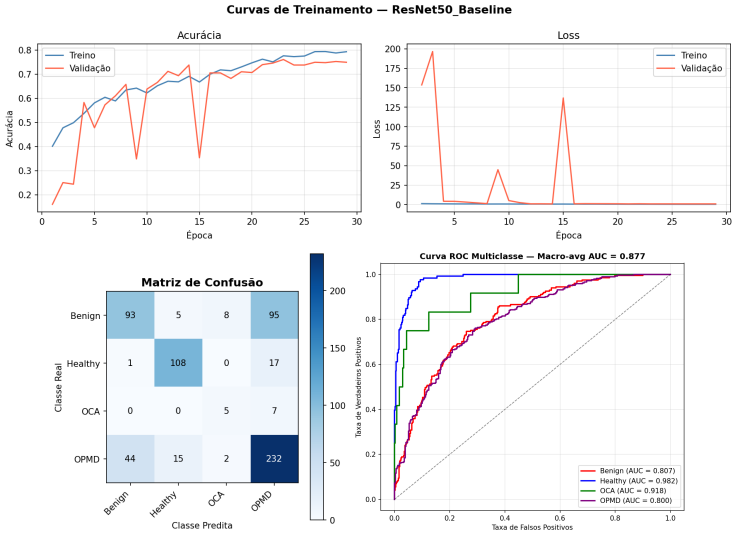


Figura 7. Curvas de treinamento, matriz de confusão e curva ROC do baseline ResNet50.

Cabe notar que o recorte da lesão foi realizado de forma assistida, o que pode superestimar o desempenho em relação a um cenário totalmente automatizado, sem localização prévia.

Como encaminhamentos, pretende-se explorar validação cruzada K-*fold* estratificada por paciente [Burman 1989], ajuste de limiar de decisão para OCA, e incorporação de bancos externos semanticamente compatíveis com as quatro classes do problema. Adicionalmente, encontra-se em desenvolvimento um banco de imagens próprio, composto por imagens de pacientes da região, coletadas com o suporte de uma cirurgiã-dentista especialista, com o objetivo de ampliar a representatividade local do modelo e reduzir o *domain shift* observado no teste externo. O código será disponibilizado publicamente após a publicação.

Referências

Albino, I. L. S., Mendonça, M. B. d., Vieira, C. E. S., Silva, N. F., Silva, D. K. C. d., and Rocha, A. M. (2025). Implicações do uso do cigarro eletrônico na saúde bucal. *Brazilian Journal of Oral and Systemic Health*, 1(2):452–463.

Amaral, R. C. d., Andrade, R. A. R., Couto, G. R., Herrera-Serna, B. Y., Rezende-Silva, E., and Cardoso, M. C. A. C. (2022). Tendências de mortalidade por câncer bucal no Brasil por regiões e principais fatores de risco. *Revista Brasileira de Cancerologia*, 68(2):e081877.

Amien, F. M., Kurniawan, D., Junaidi, A., and Hermanto, B. (2025). A Comparative Study of CNN Architectures: ConvNeXt, MobileNetV3, and EfficientNet for Oral Disease Diagnosis. *Jurnal Pepadun*, 6(1):81–91.

Aparna, M., Sibi, A. L., Kumari, P., Ram, G., Kumar, S. T., Somasundaram, K., and Savithri, V. (2025). Interpretable Analysis of Oral Mucosal Lesions Using U-Net and Quantum Approaches. In *International Conference on Inventive Computation Technologies*, pages 1019–1024. IEEE.

Burman, P. (1989). A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. *Biometrika*, 76(3):503–514.

Catarino, J., Garcia, N. C., Silva, S., and Santinha, J. (2025). The impact of pre-processing techniques on deep learning breast image segmentation. *Scientific Reports*, 16:1095.

Chollet, F. (2021). *Deep Learning with Python*. Manning Publications, New York.

Dai, G., Tian, Z., Xin, C., Dai, D., Wang, C., Zhang, Y., Chen, H., and Hamilton, M. (2026). Prompt-Level Contrastive Learning for Context-Aware Multi-Modal Image Representation in Medical Diagnosis. *Pattern Recognition*, 174:113027.

Dehbozorgi, P., Ryabchykov, O., and Bocklitz, T. (2024). A Systematic Investigation of Image Pre-Processing on Image Classification. *IEEE Access*, 12:64913–64926.

Devindi, G. A. I., Dissanayake, D. M. D. R., Liyanage, S. N., Francis, F. B. A. H., Pavithya, M. B. D., Piyarathne, N. S., Hettiarachchi, P. V. K. S., Rasnayaka, R. M. S. G. K., Jayasinghe, R. D., Ragel, R. G., and Nawinne, I. (2024). Multimodal Deep Convolutional Neural Network Pipeline for AI-Assisted Early Detection of Oral Cancer. *IEEE Access*, 12:124375–124390.

Dhanuthai, K., Rojanawatsirivej, S., Thosaporn, W., Kintarak, S., Subarnbhesaj, A., Darling, M., Kryshtalskyj, E., Chiang, C. P., Shin, H. I., Choi, S. Y., Lee, S. S., and Aminishakib, P. (2018). Oral cancer: A multicenter study. *Medicina Oral, Patologia Oral y Cirugia Bucal*, 23(1):e23–e29.

Dores, L. R. M. d., Neves, A. F. d. S., Oliveira, V. S. F. d., and Dias, A. C. S. (2023). Câncer bucal no Brasil: desafios, prevenção e papel dos profissionais de odontologia para um diagnóstico efetivo. *Libertas Saúde*, 2(2).

Instituto Nacional de Câncer – INCA (2025). Câncer de boca.

Instituto Nacional de Câncer – INCA (2026). Atlas On-line de Mortalidade.

Joshi, N., Baweja, S., Kumar, S., Shokeen, V., Sharma, A., and Gahlawat, V. (2025). A Comparative Study of Classical and Quantum Support Vector Machines for Multimodal Oral Cancer Detection. In *IEEE International Conference on Data Mining Workshops*, pages 493–499. IEEE.

Madan Kumar, P. D., Ranganathan, K., Lavanya, C., Rajeshwari, S., Nayak, A., Kestur, R., Diddigi, R. B., and Behera, S. S. (2026). A Smartphone-based Comprehensive Dataset of Annotated Oral Cavity Images for Enhanced Oral Disease Diagnosis. *Scientific Data*, 13(1):676.

Marzouk, R., Alabdulkreem, E., Dhahbi, S., Nour, M. K., Duhayyim, M. A., Othman, M., Hamza, M. A., Motwakel, A., Yaseen, I., and Rizwanullah, M. (2022). Deep Transfer Learning Driven Oral Cancer Detection and Classification Model. *Computers, Materials & Continua*, 73(2):3905–3920.

Masoudi, S., Harmon, S. A., Mehralivand, S., Walker, S. M., Raviprakash, H., Bagci, U., Choyke, P. L., and Turkbey, B. (2021). Quick guide on radiology image pre-processing for deep learning applications in prostate cancer research. *Journal of Medical Imaging*, 8(1):010901.

Mukherjee, R., Daniel, R., Hazra, S., Dasgupta, S., and Mandal, S. (2026). Patient-Aware Multimodal RGB-HSI Fusion via Incremental Heuristic Meta-Learning for Oral Lesion Classification. arXiv preprint arXiv:2511.12268.

Pinheiro, A. B. L., Sousa, G. A., Gonçalves, K. M. A., Vieira, K. d. S., Sousa, L. J. d., Cunha, F. N. d. L., and Lustosa, K. O. (2026). Desordens orais potencialmente malignas e câncer bucal. *Brazilian Journal of Implantology and Health Sciences*, 8(3):05–22.

Piyarathne, N. S., Liyanage, S. N., Rasnayaka, R. M. S. G. K., Hettiarachchi, P. V. K. S., Devindi, G. A. I., Francis, F. B. A. H., Dissanayake, D. M. D. R., Ranasinghe, R. A. N. S., Pavithya, M. B. D., Nawinne, I. B., Ragel, R. G., and Jayasinghe, R. D. (2024). A Comprehensive Dataset of Annotated Oral Cavity Images for Diagnosis of Oral Cancer and Oral Potentially Malignant Disorders. *Oral Oncology*, 156:106946.

Ribani, R. and Marengoni, M. (2019). A Survey of Transfer Learning for Convolutional Neural Networks. In *SIBGRAPI Conference on Graphics, Patterns and Images Tutorials*, pages 47–57. IEEE.

Schiavon, D., Becker, C., Botelho, V., and Pianoski, T. (2023). Interpreting convolutional neural networks for brain tumor classification: An explainable artificial intelligence approach. In *Anais da XII Brazilian Conference on Intelligent Systems*, pages 77–91, Porto Alegre, RS, Brasil. SBC.

Shin, H.-C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., and Summers, R. M. (2016). Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Transactions on Medical Imaging*, 35(5):1285–1298.

Silva, D. S. d., Santos, A. O. d., and Costa, L. V. (2026). Desordens orais potencialmente malignas (DOPMs). *Facit Business and Technology Journal*, 1(70).

Silva, M. V. I. d., Lima, R. O. d., Monteiro, V. M. d. C., Santos, S. E. N. d., Gaines, A. P. L., Rocha, L. M. B. M., and Santos, E. K. N. d. (2023). Câncer bucal e o papel do Ciru rgião-Dentista no diagnóstico precoce: Revisão integrativa. *Research, Society and Development*, 12(13):e75121344156–e75121344156.

Sousa, A. C. C. d., França, A. A. M. d. S. d., Rodrigues, A. G. L., Ericeira, F. T., Rodrigues, T. A., Silva, V. G. S., and Cunha, L. D. d. (2023). Impactos do uso de cigarro eletrônico na prevalência do câncer bucal: revisão de literatura. *Revista de Estudos Multidisciplinares UNDB*, 3(1).

Tejani, A. S., Klontzas, M. E., Gatti, A. A., Mongan, J. T., Moy, L., Park, S. H., Kahn, C. E., and for the CLAIM 2024 Update Panel (2024). Checklist for Artificial Intelligence in Medical Imaging (CLAIM): 2024 Update. *Radiology: Artificial Intelligence*, 6(4):e240300.

Vissamsetti, N. V. S. S. and Shivakanth, G. (2026). A Deep Learning-Based Image Processing Framework for Oral Lesion Classification. *Engineering, Technology & Applied Science Research*, 16(2):33266–33271.

Wijethunga, M. W. C. J. and Chandrasena, M. (2025). Leveraging Multimodal AI and Explainable AI for Early Detection of Oral Cancer. In *International Conference on Advances in Technology and Computing*, pages 1–6. IEEE.