

# Explorando Representações de Recorrência Fractal com CNNs para a Classificação de Displasias Epiteliais Orais

Vitória F. C. Silva<sup>1</sup>, Gustavo C. Miranda<sup>1</sup>, Daniel B. Gonçalves<sup>1</sup>,  
Domingos L. L. de Oliveira<sup>1,2</sup>, Guilherme F. Roberto<sup>3</sup>, Leandro A. Neves<sup>3</sup>,  
Adriano B. Silva<sup>1</sup>, Marcelo Z. do Nascimento<sup>1</sup>

<sup>1</sup> Faculdade de Computação (FACOM) – Universidade Federal de Uberlândia (UFU)  
Uberlândia – MG – Brasil

<sup>2</sup>Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP)  
São Paulo – SP – Brasil

<sup>3</sup>Departamento de Ciência da Computação e Estatística (DCCE) – UNESP  
São José do Rio Preto – SP – Brasil

vitoria.fernandes@ufu.br

**Abstract.** *Early diagnosis of oral epithelial dysplasia is crucial, yet manual histopathological evaluation is influenced by subjective aspects and observer variability. Computer-aided diagnosis (CAD) systems have emerged as a strategy to assist in this analysis. However, traditional neural network approaches often entail high computational costs and fail to explore heterogeneous ensemble strategies to improve generalization. This paper presents a comparative performance analysis of lightweight convolutional neural network (CNN) architectures (MobileNetV2 and EfficientNet-B0) for classifying oral dysplasia using the OralEpitheliumDB dataset. The methodology contrasts the use of original histological images against Recurrence Plot (RP) representations generated from multiscale fractal descriptors. Experimental results demonstrated that the RP transformation yielded substantial improvements across all evaluated metrics. Notably, the heterogeneous ensemble combining MobileNet and EfficientNet, both trained on the RP representation, achieved 100% accuracy, sensitivity, and specificity. In conclusion, recurrence analysis reveals latent texture patterns not captured in the original spatial domain, enabling the implementation of robust and computationally efficient diagnostic screening systems.*

**Resumo.** *O diagnóstico precoce das displasias epiteliais orais é fundamental, mas a avaliação histopatológica manual está sujeita a variabilidades e aspectos subjetivos. Sistemas computacionais de apoio ao diagnóstico (CAD) têm sido propostos para auxiliar nessa análise. No entanto, abordagens baseadas em redes neurais tradicionais frequentemente apresentam alto custo computacional e deixam de explorar estratégias heterogêneas para melhorar a generalização. Este trabalho apresenta uma análise comparativa do desempenho de arquiteturas de redes neurais convolucionais (CNNs) lightweight (MobileNetV2 e EfficientNet-B0) na classificação de displasias orais, utilizando o dataset OralEpitheliumDB. A metodologia contrasta o uso de imagens originais com imagens transformadas em Recurrence Plots (RP), geradas a partir de descritores fractais multiescala. Os experimentos demonstraram que a representação*

*via RP proporcionou melhorias substanciais em todas as métricas avaliadas. A contribuição do estudo foi o ensemble heterogêneo combinando MobileNet e EfficientNet, ambos treinados com a representação RP, que atingiu 100% nas métricas de avaliação. Conclui-se que a análise de recorrência fornece padrões de textura relevantes que não são capturados no domínio espacial original, viabilizando a implementação de sistemas diagnósticos robustos e de baixo custo computacional.*

## **1. Introdução**

Segundo dados publicados pelo Instituto Nacional de Câncer (INCA), estimou-se que, para o período de 2023 a 2025, aproximadamente 704 mil novos diagnósticos de câncer no Brasil, dos quais cerca de 15.100 corresponderam ao câncer da cavidade oral [Santos 2023]. A progressão para o carcinoma geralmente é precedida por lesões potencialmente malignas, que provocam alteração celular desordenada e elevam o risco de carcinogênese. Frequentemente, essas lesões manifestam características de displasias epiteliais orais (DEO) [Warnakulasuriya et al. 2008]. Portanto, o diagnóstico precoce e preciso de DEOs é determinante para aumentar as chances de sucesso no tratamento do paciente [Dabeer et al. 2019].

No diagnóstico das DEOs, a avaliação histopatológica de lâminas microscópicas permanece como o principal critério para a classificação das lesões, sendo a análise das alterações nucleares um dos fatores mais relevantes nesse processo [Cheung et al. 2022]. Características morfológicas como tamanho, densidade e padrão de cromatina dos núcleos estão associadas ao grau de diferenciação celular e, por extensão, à severidade das lesões [Müller 2018]. Entretanto, esse processo é influenciado por aspectos subjetivos, como a experiência e a carga de trabalho do especialista, resultando em variabilidade inter e intraobservador. Essa subjetividade pode ocasionar discrepâncias diagnósticas, com graus semelhantes de DEO sendo atribuídos a padrões morfológicos distintos ou graus diferentes da lesão sendo atribuídos à características semelhantes [Pimenta-Barros et al. 2025]. Embora a estratégia de dupla leitura mitigue esses efeitos, seu uso pode aumentar os custos e o tempo necessário para o fluxo diagnóstico [van Zelst et al. 2020].

Nesse cenário, os sistemas computacionais de apoio ao diagnóstico (do inglês, *computer aided diagnosis* - CAD) surgem como estratégias para reduzir a dependência exclusiva da avaliação subjetiva, fornecendo análises quantitativas dos tecidos histopatológicos e auxiliando no diagnóstico das lesões. [Alajaji et al. 2024]. Um sistema CAD típico envolve diversas etapas, integrando extração de características e classificação para rotular padrões morfológicos e auxiliar nas decisões diagnósticas sobre as lesões. No estágio de extração de características há diversos métodos que são empregados em abordagens clássicas, como engenharia de características, e abordagens baseadas em aprendizagem profunda. Nesta etapa, não há uma abordagem universal que apresente desempenho consistente para todos os tipos de imagens, tornando necessária a investigação de métodos específicos para diferentes aplicações. Nesse contexto, descritores baseados em geometria fractal têm demonstrado resultados relevantes na classificação de imagens histológicas. A geometria fractal permite descrever a textura de superfícies por meio de propriedades invariantes à escala e rotação, sendo a dimensão fractal

(DF), a lacunaridade (LAC) e a percolação (PERC) alguns dos descritores mais utilizados [Roberto et al. 2021, Pereira et al. 2026, Longo et al. 2023].

Tais descritores fornecem vetores de características capazes de representar tanto atributos locais quanto globais das imagens histopatológicas. Para viabilizar sua integração com *convolutional neural network* (CNNs), diferentes estratégias têm sido propostas para converter esses vetores unidimensionais em representações bidimensionais, como o *recurrence plot* (RP) [Pham 2021, Roberto et al. 2024]. Apesar das contribuições, as abordagens da literatura restringem-se ao uso de *ensembles* homogêneos, deixando de explorar o potencial de estratégias heterogêneas que poderiam capturar informações complementares e melhorar a generalização. Além disso, não investigam modelos CNN *lightweight*, que são computacionalmente mais eficientes, ou uma validação mais robusta das abordagens, como amostragem dos dados baseada em *cross-validation*.

Visando preencher essas lacunas, este trabalho apresenta uma análise comparativa do desempenho de arquiteturas CNN *lightweight* na classificação de imagens histológicas de DEOs, utilizando imagens originais e representações RP derivadas de descritores fractais. Para reduzir o custo computacional associado a CNNs convencionais, foram avaliadas as arquiteturas MobileNet e EfficientNet via transferência de aprendizado com pesos pré-treinados no ImageNet. Adicionalmente, foram investigadas estratégias de *ensemble learning*, incluindo configurações homogêneas e heterogêneas. As principais contribuições deste trabalho são as seguintes:

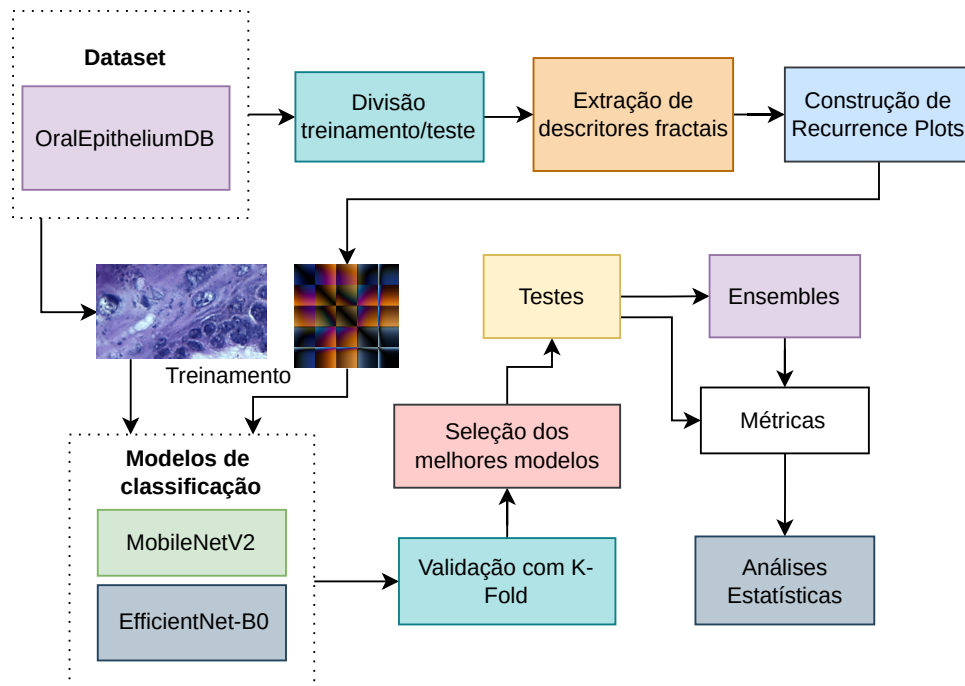
- avaliação comparativa de arquiteturas CNN *lightweight* aplicadas à classificação de imagens histológicas de DEOs, considerando imagens originais e representações RP baseadas em descritores fractais;
- estudo de estratégias de *ensemble learning*, incluindo modelos homogêneos e heterogêneos, para aprimorar a robustez da classificação;
- comparação dos resultados obtidos com métodos do estado da arte reportados na literatura.

## 2. Materiais e Métodos

A metodologia proposta investiga as representações do RP de descritores fractais para classificação de lesões de DEOs. O *pipeline* experimental foi estruturado em três etapas: (i) geração das representações RP, (ii) treinamento individual das arquiteturas CNNs e (iii) exploração de modelos *ensemble learning*. Na Figura 1 são apresentadas as etapas do *pipeline* proposto para classificação das lesões de DEO. Os detalhes de cada etapa do método são descritos nas próximas subseções.

### 2.1. Base de Imagens

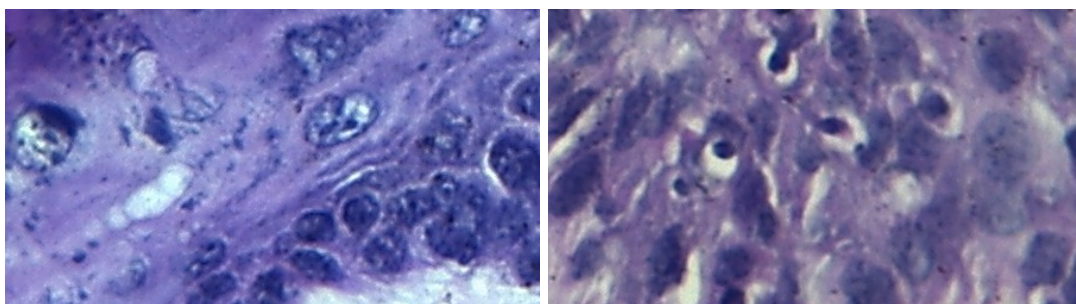
Para a aplicação da metodologia proposta, foi empregada a base de imagens pública OralEpitheliumDB [Silva et al. 2024], composta por imagens histológicas de tecidos com DEO. Esse banco foi obtido a partir de lâminas histológicas provenientes de lesões induzidas experimentalmente na língua de camundongos, por meio da administração controlada de um agente carcinógeno. No protocolo experimental adotado, 30 camundongos tiveram acesso contínuo ao carcinógeno durante um período de 16 semanas. Ao final desse intervalo, 50% dos animais foram sacrificados imediatamente, enquanto os 50%



**Figura 1. Pipeline da metodologia proposta, incluindo as etapas de divisão dos dados, treinamento e avaliação dos modelos de classificação com e sem o uso de *ensemble learning* no dataset.**

restantes foram mantidos por mais 16 semanas antes do sacrifício, permitindo a avaliação da progressão das lesões.

As lâminas histológicas resultantes foram classificadas por especialista nas categorias de tecido saudável e DEO severa. As imagens foram obtidas de regiões de interesse (ROIs), com magnificação de  $400\times$ , resolução de  $250\times 450$  pixels e modelo de cores RGB, sendo 114 ROIs para cada classe, totalizando 228 utilizadas para a avaliação dos métodos propostos, com separação de forma a evitar vazamento de dados entre os conjuntos de treino e o de teste. Na Figura 2 são apresentados exemplos das ROIs obtidas das classes investigadas neste estudo.



**Figura 2. Exemplo de imagens utilizadas no estudo, (à esquerda) um tecido saudável, (à direita) DEO severa.**

## 2.2. Geração das Imagens de *Recurrence Plot*

A representação RP foi construída a partir de descritores fractais multiescala extraídos de cada imagem do *dataset*. Inicialmente, as imagens foram redimensionadas para  $224 \times 224$  *pixels* e processadas para o cálculo das matrizes de probabilidade multiescala por meio do algoritmo *gliding-box*. Nesse procedimento, uma janela de tamanho  $r \times r$  desliza pela imagem, e a matriz registra a frequência de vizinhos cuja distância de cor ao pixel central é menor ou igual a  $r$ . Diferente das abordagens clássicas baseadas em binarização ou tons de cinza, a similaridade é avaliada diretamente no espaço vetorial RGB, preservando a informação cromática original no cálculo da massa fractal. Foi utilizado valores de raios ( $r$ ) variando de 3 a 41, com passo igual a 2 [Roberto et al. 2024]. O passo adotado assegura janelas de dimensões ímpares (com um centro exato), enquanto o limite de 41 resulta em 20 escalas de observação. Essa configuração é estratégica para garantir que a quantidade de atributos extraídos forme uma matriz quadrada para representação RP, evitando distorções por preenchimento (*padding*). A extração considerou três métricas de distância, Minkowski, Euclidiana e Manhattan, a fim de capturar diferentes noções geométricas de vizinhança no espaço de intensidades RGB.

A partir das matrizes, para cada escala, foram extraídos cinco descritores locais específicos: (i) o inverso da massa de pixels, ( $N(r) = \sum \frac{P(m,r)}{m}$ ); (ii) a LAC, calculada a partir do primeiro ( $m_1$ ) e do segundo ( $m_2$ ) momentos da distribuição de probabilidade ( $LAC = \frac{m_2 - m_1^2}{m_1^2}$ ); e três descritores baseados em PERC detalhados em [Roberto et al. 2021], sendo (iii) o número médio de clusters, (iv) a taxa de percolação e (v) o tamanho médio do maior cluster. O conjunto final resultou em um vetor de 300 atributos por imagem, organizados em três blocos de 100 características correspondentes às distâncias obtidas a partir da extração dos cinco descritores locais em 20 escalas de observação distintas. Os atributos foram normalizados por segmentos de 20 elementos utilizando transformação min-max, com o objetivo de reduzir efeitos de escala e preservar a estrutura relativa local dos descritores.

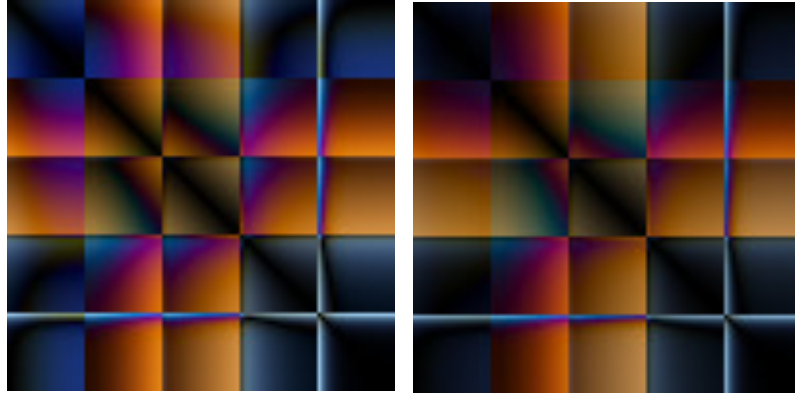
A representação RP é uma técnica proposta por [Eckmann et al. 1995] para a projeção de eventos recorrentes em espaços bidimensionais. A aplicação desta técnica sobre um vetor de características contendo  $N$  atributos gera uma matriz quadrada  $N \times N$ , na qual cada elemento  $R_{i,j}$  da matriz é obtido pela Equação 1:

$$R_{i,j} = \|x_i - x_j\| \quad \forall \quad i, j = 1 \dots N. \quad (1)$$

onde  $x_i$  e  $x_j$  são a  $i$ -ésima e a  $j$ -ésima características no vetor, e  $\|x_i - x_j\|$  é a norma da distância euclidiana. Exemplos das imagens geradas a partir desse procedimento são apresentados na Figura 3. Essa representação converte as informações fractais multiescala em padrões espaciais estruturados, possibilitando que as CNNs explorem dependências não lineares e relações internas entre descritores originalmente definidos no domínio numérico.

## 2.3. Arquiteturas de Classificação

Foram investigadas as arquiteturas MobileNetV2 (MobileNet) e EfficientNet-B0 (EfficientNet), pré-treinadas na base de dados ImageNet. A MobileNet [Sandler et al. 2018] é



**Figura 3. Exemplo de imagens RP geradas, equivalentes às imagens DEO da Figura 2, sendo (esquerda) tecido saudável e (direita) tecido com DEO severa.**

uma arquitetura convolucional projetada para alta eficiência computacional, sendo baseada no uso de convoluções separáveis em profundidade, que dividem a convolução tradicional em uma etapa de filtragem por canal (*depthwise*) seguida de uma combinação entre canais por meio de convoluções  $1 \times 1$  (*pointwise*). Essa estratégia reduz significativamente o número de parâmetros e operações. Além disso, a arquitetura utiliza blocos residuais invertidos com *bottlenecks* lineares, nos quais a expansão dos canais ocorre antes da filtragem espacial, favorecendo melhor preservação de informações com menor custo computacional.

A EfficientNet [Tan and Le 2019] adota o princípio de escalonamento composto, que amplia de forma balanceada a profundidade, a largura e a resolução de entrada da rede, buscando melhor relação entre desempenho e custo computacional. Diferentemente de abordagens que aumentam apenas uma dessas dimensões, esse método promove um crescimento proporcional da arquitetura. Em ambos os modelos, a camada classificadora original foi substituída por uma nova camada totalmente conectada compatível com duas classes, permitindo a adaptação dos modelos pré-treinados ao contexto de DEOs.

Essas arquiteturas foram selecionadas por representarem modelos convolucionais com diferentes estratégias de construção e escalonamento, permitindo uma análise comparativa sob distintas características arquiteturais. Além disso, ambas as arquiteturas foram consideradas por apresentarem configurações com menor número de parâmetros em comparação a modelos mais complexos, o que se alinha ao objetivo de avaliar soluções computacionalmente mais eficientes.

#### **2.4. Estratégias de *Ensemble Learning***

Com base nas probabilidades previstas pelos modelos após aplicação do *softmax*, os resultados foram combinados por meio da soma direta das distribuições de probabilidade para elaboração da estratégia de *ensemble learning*. A classe final foi definida através da equação:

$$\hat{y} = \arg \max (P_1 + P_2), \quad (2)$$

na qual  $P_1$  e  $P_2$  representam os vetores de probabilidade produzidos por cada modelo participante do *ensemble*.

Foram avaliadas as combinações entre os modelos MobileNet e EfficientNet, utilizando imagens originais e RP, e *ensemble*, homogêneo e *ensemble* heterogêneo, conforme apresentado na Tabela 1. Essa organização experimental permitiu analisar o impacto da representação RP dentro da mesma arquitetura, a combinação entre arquiteturas distintas e o efeito conjunto entre múltiplas arquiteturas e representações. A restrição às combinações dois a dois foi adotada com o objetivo de controlar a complexidade experimental e permitir uma análise direta da contribuição de cada modelo no desempenho dos *ensembles*.

**Tabela 1. Cenários avaliados nas estratégias de *ensemble*.**

Cenário	Modelos Combinados
1	MobileNet (Original) + MobileNet (RP)
2	MobileNet (Original) + EfficientNet(Original)
3	MobileNet (Original) + EfficientNet (RP)
4	MobileNet (RP) + EfficientNet (Original)
5	MobileNet (RP) + EfficientNet (RP)
6	EfficientNet (Original) + EfficientNet (RP)

## 2.5. Configuração do Treinamento

O treinamento dos modelos foi conduzido em um computador equipado com processador Intel Core i5 de 12ª geração, GPU NVIDIA GeForce RTX 3050 (4GB VRAM), 16 GB de memória RAM e sistema operacional Windows 11, utilizando a linguagem Python e a biblioteca PyTorch. Os hiperparâmetros de treinamento foram definidos empiricamente a partir de experimentos exploratórios preliminares, utilizando-se de *batch size* de 32, 30 épocas, otimizador Adam com taxa de aprendizado de  $10^{-4}$  e função de perda *Cross-Entropy*.

Considerando o tamanho limitado do conjunto de dados e a variabilidade associada à utilização de uma única partição de treinamento e validação (*holdout*), optou-se pelo uso de validação cruzada estratificada (*Stratified K-Fold*) com cinco divisões, garantindo a preservação da proporção entre as classes em cada partição. Em cada execução, o modelo com menor *validation loss* no conjunto de validação foi selecionado. Além disso, para reduzir a variabilidade associada à inicialização aleatória e à divisão dos dados, os experimentos foram repetidos utilizando cinco *seeds* distintas, sendo posteriormente reportadas as métricas de desempenho no conjunto de teste.

## 2.6. Métricas de Avaliação e Análise Estatística

O desempenho dos modelos individuais e das estratégias de ensemble foi avaliado no conjunto de teste utilizando as métricas acurácia ( $A_C$ ), F1-Score ( $F_1$ ), sensibilidade ( $S_E$ ) e especificidade ( $E_S$ ). A acurácia mede a proporção total de classificações corretas. O F1-Score representa a média harmônica entre precisão e sensibilidade, sendo especialmente relevante em cenários com possível desbalanceamento entre classes. A sensibilidade quantifica a capacidade do modelo de identificar corretamente amostras da classe

positiva, enquanto a especificidade mede a capacidade de identificar corretamente amostras da classe negativa. Como os experimentos foram repetidos com cinco sementes distintas, os resultados obtidos para cada cenário foram submetidos à análise estatística. Para verificar se a combinação de arquiteturas e representações promove ganhos significativos de desempenho, foram aplicados o teste de Friedman e a ANOVA de medidas repetidas, adotando nível de significância de 5%.

### 3. Resultados e Discussão

**Tabela 2. Resultados consolidados dos experimentos com MobileNet, EfficientNet e ensembles.**

Categoria	Cenário	$A_C$ (%)	$F_1$ (%)	$S_E$ (%)	$E_S$ (%)
MobileNet	MobileNet (Original)	90,4545 ± 2,9633	90,3452 ± 3,0577	90,4545 ± 2,9633	90,4545 ± 2,9633
	MobileNet (RP)	<b>99,5455 ± 1,0164</b>	<b>99,5452 ± 1,0169</b>	<b>99,5455 ± 1,0164</b>	<b>99,5455 ± 1,0164</b>
	MobileNets (Original) e (RP)	98,6356 ± 2,0338	98,6342 ± 2,0366	98,6364 ± 2,0328	98,6364 ± 2,0328
EfficientNet	EfficientNet (Original)	82,7273 ± 5,7045	82,0475 ± 6,2696	82,7273 ± 5,7045	82,7273 ± 5,7045
	EfficientNet (RP)	<b>99,0909 ± 1,2448</b>	<b>99,0904 ± 1,2455</b>	<b>99,0909 ± 1,2448</b>	<b>99,0909 ± 1,2448</b>
	EfficientNet (Original) e (RP)	97,7274 ± 2,2727	97,7233 ± 2,2774	97,7273 ± 2,2727	97,7273 ± 2,2727
Heterogêneo	MobileNet_O + EfficientNet_O	90,4545 ± 1,9015	90,3584 ± 1,9521	90,4545 ± 1,9015	90,4545 ± 1,9015
	MobileNet_RP + EfficientNet_RP	<b>100,0000 ± 0,0000</b>	<b>100,0000 ± 0,0000</b>	<b>100,0000 ± 0,0000</b>	<b>100,0000 ± 0,0000</b>
	MobileNet_O + EfficientNet_RP	97,7273 ± 1,6071	97,7247 ± 1,6104	97,7273 ± 1,6071	97,7273 ± 1,6071
	MobileNet_RP + EfficientNet_O	98,1818 ± 1,9015	98,1795 ± 1,9050	98,1818 ± 1,9015	98,1818 ± 1,9015

A categoria MobileNet, da Tabela 2 apresenta os resultados obtidos com a arquitetura MobileNet nos cenários utilizando as imagens originais, as imagens de representação RP e o *ensemble* homogêneo. Para todas as métricas, os valores maiores indicam melhor desempenho, e os valores em negrito representam o melhor resultado geral. Observa-se que a representação RP proporcionou um aumento em todas as métricas em comparação ao uso das imagens originais. O cenário MobileNet RP obteve o melhor desempenho geral, alcançando  $A_C = 99,5455\% \pm 1,0164\%$  e  $F_1 = 99,5455\% \pm 1,0169\%$ , indicando uma classificação robusta e com baixa variabilidade entre as execuções. Em contraste, o modelo MobileNet Original apresentou desempenho significativamente inferior, com  $A_C = 90,4545\% \pm 2,9633\%$  e  $F_1 = 90,3452\% \pm 3,0577\%$ , indicando que as informações contidas nas imagens originais não capturaram padrões discriminativos com a mesma eficácia. O aumento do desvio padrão nesse cenário também sugere maior instabilidade entre diferentes execuções.

Tal comportamento ocorre em razão das propriedades da representação RP. Enquanto as imagens originais estão no domínio espacial e podem apresentar alta variabilidade intra-classe, os descritores fractais capturam características estatísticas e estruturais da textura em múltiplas escalas. A transformação desses descritores em imagens por meio do RP permite mapear relações de similaridade e recorrência entre os padrões extraídos, evidenciando dependências não lineares que não são representadas de maneira explícita no domínio original. Assim, há produção de padrões mais discriminativos, o que facilita a atuação das convoluções, que passam a operar sobre estruturas mais regulares. Além disso, tal abordagem reduz a influência de variações locais irrelevantes, contribuindo para maior robustez e estabilidade dos modelos, conforme observado pelos menores desvios padrão nos experimentos.

O *ensemble* homogêneo entre MobileNet original e RP apresentou desempenho intermediário de  $A_C = 98,6356\% \pm 2,0338\%$  e  $F_1 = 98,6342\% \pm 2,0366\%$ , supe-

rando o modelo baseado exclusivamente em imagens originais, porém sem alcançar o desempenho obtido utilizando o RP. Esse resultado sugere que a representação por RP é descritiva, e que a combinação com a representação original não trouxe ganhos adicionais significativos. De maneira geral, os resultados indicam que a representação RP forneceu informações com maior poder discriminativo para a MobileNet, possivelmente por tornar padrões temporais ou estruturais mais evidentes no domínio espacial, facilitando o aprendizado das convoluções. Além disso, o baixo desvio padrão observado no cenário com a representação RP reforça a robustez dessa abordagem.

De forma semelhante ao observado com a MobileNet, na categoria EfficientNet, a utilização do RP resultou em ganhos expressivos de desempenho. O cenário EfficientNet com RP obteve o melhor desempenho geral, alcançando  $A_C = 99,0909\% \pm 1,2448\%$  e  $F_1 = 99,0904\% \pm 1,2455\%$ , com baixa variabilidade entre as execuções. Em contraste, o modelo EfficientNet original apresentou  $A_C = 82,7273\% \pm 5,7045\%$  e  $F_1 = 82,0475\% \pm 6,2696\%$ , indicando menor estabilidade. O *ensemble* EfficientNet apresentou desempenho intermediário, com  $A_C = 97,7274\% \pm 2,2727\%$  e  $F_1 = 97,7233\% \pm 2,2774\%$ , superando o uso de imagens originais, porém sem ultrapassar o desempenho obtido com a representação RP. Esses resultados reforçam que a representação baseada em RP fornece informações mais discriminativas para a arquitetura EfficientNet, tornando o processo de classificação mais robusto e consistente.

A categoria Heterogênio apresenta os resultados combinando as arquiteturas MobileNet e EfficientNet sob diferentes representações de entrada. Observa-se que o *ensemble* heterogêneo MobileNet\_RP + EfficientNet\_RP apresentou desempenho relevante, atingindo valor de  $100,00\% \pm 0,00\%$  em todas as métricas. Esse resultado indica que, quando ambas as arquiteturas utilizaram a representação baseada em RP, houve complementação entre os modelos, resultando em generalização adequada para o conjunto avaliado. Por outro lado, o *ensemble* heterogêneo MobileNet\_O + EfficientNet\_O apresentou  $A_C = 90,4545\% \pm 1,9015\%$  e  $F_1 = 90,3584\% \pm 1,9521\%$ , reforçando que a limitação principal não está na arquitetura, mas na representação original dos dados.

Os cenários mistos apresentaram desempenho intermediário, ambos com valores acima de 97,5% para todas as métricas, demonstrando que a presença de pelo menos um modelo treinado com RP contribui para elevar significativamente o desempenho do *ensemble learning* heterogêneo. Ainda assim, o melhor resultado foi obtido quando ambas as arquiteturas utilizaram a representação RP.

### 3.1. Análise Estatística

Para avaliar se as diferenças observadas entre os cenários são estatisticamente significativas, foram aplicadas análises de medidas repetidas considerando as diferentes execuções com *seeds* distintas. No experimento com a MobileNet, o teste de Friedman indicou diferença significativa entre os cenários, com  $\chi^2(2) = 8,44$  e  $p = 0,015$ . Comparações múltiplas pelo método de Durbin–Conover mostraram que o modelo MobileNet RP apresentou o desempenho significativamente superior de  $p < 0,001$  para a MobileNet Original, assim como ao *ensemble* MobileNet Original + RP. Não foi observada diferença estatisticamente significativa entre MobileNet RP e o *ensemble* ( $p = 0,471$ ), indicando desempenho estatisticamente equivalente entre esses dois cenários.

Para a EfficientNet, foi aplicada ANOVA de medidas repetidas paramétrica, que

revelou efeito significativo do fator de representação, com  $\rho < 0,001$ . O teste *post hoc* de Tukey indicou que EfficientNet RP apresentou  $\rho = 0,006$ , indicando um desempenho significativamente superior para a EfficientNet Original e a *ensemble* EfficientNet Original + RP. Não houve diferença significativa entre EfficientNet RP e o *ensemble* ( $\rho = 0,143$ ), reforçando a consistência do ganho proporcionado pela representação via RP.

Na comparação entre os *ensembles* heterogêneos com MobileNet + EfficientNet, o teste de Friedman mostrou efeito global significativo entre os cenários avaliados ( $\chi^2(4) = 27,0$ ,  $\rho < 0,001$ ). As comparações múltiplas pelo método de Durbin–Conover revelaram que o *ensemble* formado apenas por representações originais apresentou  $\rho < 0,001$ , sendo um desempenho significativamente inferior aos demais cenários com alguma das arquiteturas treinadas com RP. Esses resultados estatísticos confirmaram que os ganhos observados com a utilização do RP não são decorrentes de variações aleatórias entre execuções, mas representam melhorias estatisticamente significativas em relação ao uso exclusivo das imagens originais.

A Tabela 3 apresenta uma visão geral dos resultados obtidos pela metodologia proposta em comparação com métodos relevantes de visão computacional reportados na literatura para a análise de imagens histopatológicas da cavidade oral. De forma geral, os resultados obtidos mostram-se compatíveis com aqueles reportados na literatura, evidenciando desempenho relevante para a tarefa de classificação, especialmente por meio da associação de descritores fractais, abordagens de *reshape*, modelos baseados em CNNs e *ensemble* heterogêneo.

**Tabela 3. Análise da metodologia proposta com abordagens existentes na literatura para classificação de imagens orais.**

Estudo	Tipo de Lesão	Extração de Características	Classificador	$A_C$ (%)
[Adel et al. 2018]	DEO	ORB	SVM	92,6
[Deif et al. 2022]	Carcinoma	<i>Deep features</i>	XGBoost	96,3
[Silva et al. 2022]	DEO	Textura e morfologia	HOP	92,4
[Longo et al. 2023]	DEO	Fractais, CNN e XAI	<i>ensemble</i>	97,2
<b>Método Proposto</b>	DEO	Fractais, RP e CNNs	<i>ensemble</i>	<b>100,0</b>

#### 4. Conclusão

Este estudo evidenciou que a aplicação de RP em arquiteturas convolucionais leves, como MobileNet e EfficientNet, supera significativamente o uso de imagens originais para o diagnóstico de DEOs. Enquanto os modelos treinados com imagens originais apresentaram limitações de desempenho, a transformação de domínio via RP elevou a acurácia para patamares superiores a 99% em todos os cenários, confirmando a hipótese de que RP mostram padrões de textura que não são capturados pelas convoluções tradicionais no domínio espacial original.

As principais contribuições deste trabalho consistem em: (i) a investigação sistemática do uso de representações baseadas em RP combinadas com imagens originais no contexto de classificação; (ii) a análise comparativa entre estratégias de *ensemble* homogêneo e heterogêneo, avaliando o impacto da diversidade de modelos e representações

no desempenho final; e (iii) a utilização de testes estatísticos para fundamentar a análise dos resultados, garantindo maior robustez.

Os resultados validam a viabilidade técnica de modelos eficientes que, quando combinados em estratégias de *ensemble* heterogêneo, atingiram 100% de acurácia e consistência total entre execuções. Tais descobertas mostram como uma solução positiva para cenários com baixa capacidade computacional, democratizando o acesso à triagem precoce de lesões potencialmente malignas sem a dependência de *hardware* de alto custo.

Em trabalhos futuros, pretende-se investigar a integração com modelos baseados em *Transformers*, bem como avaliar diferentes representações de descritores fractais e seu impacto no desempenho da classificação. Além disso, os experimentos serão ampliados para bases de dados maiores e multi-institucionais, visando analisar a robustez e a capacidade de generalização dos modelos, assim como explorar *ensembles* mais complexos com múltiplos classificadores.

## 5. Agradecimentos

Os autores agradecem o apoio financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (Processos nº 305386/2024-7 e nº 302833/2025-0) e da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) (Processos nº APQ-00727-24 e nº APD-01428-25).

## Referências

- Adel, D., Mounir, J., El-Shafey, M., Eldin, Y. A., El Masry, N., AbdelRaouf, A., and Abd Elhamid, I. S. (2018). Oral epithelial dysplasia computer aided diagnostic approach. In *2018 13th International Conference on Computer Engineering and Systems (ICCES)*, pages 313–318. IEEE.
- Alajaji, S. A., Khoury, Z. H., Jessri, M., Sciubba, J. J., and Sultan, A. S. (2024). An update on the use of artificial intelligence in digital pathology for oral epithelial dysplasia research. *Head and neck pathology*, 18(1):38.
- Cheung, V. K., Hulme, K., Schifter, M., Palme, C., Low, T.-H. H., Clark, J., and Gupta, R. (2022). Oral epithelial dysplasia: a review of diagnostic criteria for anatomic pathologists. *Advances in Anatomic Pathology*, 29(4):227–240.
- Dabeer, S., Khan, M. M., and Islam, S. (2019). Cancer diagnosis in histopathological image: Cnn based approach. *Informatics in Medicine Unlocked*, 16:100231.
- Deif, M. A., Attar, H., Amer, A., Elhaty, I. A., Khosravi, M. R., Solyman, A. A., et al. (2022). Diagnosis of oral squamous cell carcinoma using deep neural networks and binary particle swarm optimization on histopathological images: an aiomt approach. *Computational Intelligence and Neuroscience*, 2022.
- Eckmann, J.-P., Kamphorst, S. O., Ruelle, D., et al. (1995). Recurrence plots of dynamical systems. *World Scientific Series on Nonlinear Science Series A*, 16:441–446.
- Longo, L. H. d. C., Roberto, G. F., Tosta, T. A., de Faria, P. R., Loyola, A. M., Cardoso, S. V., Silva, A. B., do Nascimento, M. Z., and Neves, L. A. (2023). Classification of multiple h&e images via an ensemble computational scheme. *Entropy*, 26(1):34.

- Müller, S. (2018). Oral epithelial dysplasia, atypical verrucous lesions and oral potentially malignant disorders: focus on histopathology. *Oral surgery, oral medicine, oral pathology and oral radiology*, 125(6):591–602.
- Pereira, D. C., Silva, A. B., Longo, L. H., Loyola, A. M., Cardoso, S. V., de Faria, P. R., Tosta, T. A., Neves, L. A., Martins, A. S., and Nascimento, M. Z. (2026). Evaluation of fractal descriptors, deep features and xai representations with lasso-regularized hermite polynomial classifier for h&e histological image classification. *Biomedical Signal Processing and Control*, 112:108393.
- Pham, T. D. (2021). From raw pixels to recurrence image for deep learning of benign and malignant mediastinal lymph nodes on computed tomography. *IEEE Access*, 9:96267–96278.
- Pimenta-Barros, L. A., Ramos-García, P., González-Moles, M. Á., Aguirre-Urizar, J. M., and Warnakulasuriya, S. (2025). Malignant transformation of oral leukoplakia: Systematic review and comprehensive meta-analysis. *Oral diseases*, 31(1):69–80.
- Roberto, G. F. et al. (2021). Associação entre atributos manuais e aprendizado profundo baseada em geometria fractal para classificação de imagens histológicas.
- Roberto, G. F., Neves, L. A., Lumini, A., Martins, A. S., and Nascimento, M. Z. d. (2024). An ensemble of learned features and reshaping of fractal geometry-based descriptors for classification of histological images. *Pattern Analysis and Applications*, 27(1):8.
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Santos, M. d. O. (2023). Estimativa de incidência de câncer no brasil, 2023-2025. *Revista Brasileira de Cancerologia*, 69(1).
- Silva, A. B., Martins, A. S., Tosta, T. A. A., Loyola, A. M., Cardoso, S. V., Neves, L. A., de Faria, P. R., and do Nascimento, M. Z. (2024). Oralepitheliumdb: A dataset for oral epithelial dysplasia image segmentation and classification. *Journal of Imaging Informatics in Medicine*, 37(4):1691–1710.
- Silva, A. B., Martins, A. S., Tosta, T. A. A., Neves, L. A., Servato, J. P. S., de Araújo, M. S., de Faria, P. R., and do Nascimento, M. Z. (2022). Computational analysis of histological images from hematoxylin and eosin-stained oral epithelial dysplasia tissue sections. *Expert Systems with Applications*, 193:116456.
- Tan, M. and Le, Q. V. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the International Conference on Machine Learning (ICML)*.
- van Zelst, J. C., Tan, T., Mann, R. M., and Karssemeijer, N. (2020). Validation of radiologists' findings by computer-aided detection (cad) software in breast cancer detection with automated 3d breast ultrasound: a concept study in implementation of artificial intelligence software. *Acta Radiologica*, 61(3):312–320.
- Warnakulasuriya, S., Reibel, J., Bouquot, J., and Dabelsteen, E. (2008). Oral epithelial dysplasia classification systems: predictive value, utility, weaknesses and scope for improvement. *Journal of oral pathology & medicine*, 37(3):127–133.