Classificação Ordinal de Lesões de Cárie Cavitadas e Não Cavitadas em Fotografias da Superfície Oclusal Baseada no ICDAS com Transfer Learning

Ana Larissa Teixeira Dantas[1], Jadiel Silva da Cunha[1], Julyana Raab Pereira[2], Beatriz Gonçalves Neves[3], Bruno Riccelli dos Santos Silva[1], Adriana Pigozzo Manso[5], Wellington Franco[1], Lidiany Karla Azevedo Rodrigues[2]

[1]Universidade Federal do Ceará, Campus Crateús – Crateús, CE – Brasil

[2]Universidade Federal do Ceará, Faculdade de Farmácia, Odontologia e Enfermagem – Fortaleza, CE – Brasil

[3]Universidade Federal do Ceará, Campus Sobral – Sobral, CE – Brasil

[4]University at Buffalo School of Dental Medicine – Buffalo, NY – USA

[5]Faculty of Dentistry, University of British Columbia – Vancouver, BC – Canada

{larissa.teixeira, jadielsilva}@alu.ufc.br

{bruno.silva, wellington}@crateus.ufc.br

Abstract. *The subjectivity of visual caries diagnosis compromises clinical reproducibility. This study evaluated ordinal classification of severity (ICDAS) on 356 images using Transfer Learning with 10 convolutional architectures pretrained on ImageNet and Frank and Hall decomposition. Static feature extraction associated with classical models (SVM, MLP, and RF) was compared with partial fine-tuning. Predictive behavior varied according to architectural complexity: deep networks, such as EfficientNetV2B3, required the hybrid approach, while classical ones, such as VGG19, were optimized with fine-tuning (both with QWK $\approx$ 0.84; F1-Score $>$ 81%). The predictive convergence of these strategies attests to the feasibility of dental telediagnosis.*

Resumo. *A subjetividade do diagnóstico visual da cárie compromete a reprodutibilidade clínica. Este estudo avaliou a classificação ordinal da severidade (ICDAS) em 356 imagens, utilizando Transfer Learning com dez arquiteturas convolucionais pré-treinadas no ImageNet e decomposição de Frank e Hall. Comparou-se a extração estática associada a modelos clássicos (SVM, MLP e RF) com o ajuste fino parcial. O comportamento preditivo variou conforme a complexidade arquitetural: redes profundas, como EfficientNetV2B3, exigiram a abordagem híbrida, enquanto as clássicas, como VGG19, otimizaram-se com ajuste fino (ambas com QWK $\approx$ 0,84; F1-Score $>$ 81%). A convergência preditiva dessas estratégias atesta a viabilidade do telediagnóstico odontológico.*

1. Introdução

A cárie dentária é uma doença multifatorial cuja severidade varia de desmineralizações iniciais do esmalte a cavitações dentinárias [Frencken 2017]. Segundo o *Global Burden of*

Disease (GBD), 3,69 bilhões de pessoas foram afetadas por condições orais em 2021, com impacto desproporcional em regiões de baixo Índice Sociodemográfico, onde a escassez de especialistas limita o acesso ao tratamento [Bernabe et al. 2025].

A detecção precoce permite intervenções não invasivas e reduz o ciclo restaurador repetitivo, caracterizado por reintervenções sucessivas que culminam na perda irreversível da estrutura dental [Bader and Shugars 2006]. Nesse contexto, o *International Caries Detection and Assessment System* (ICDAS) constitui um critério clínico amplamente validado para classificar a severidade da doença [Ismail et al. 2007]. Todavia, sua aplicação em triagens populacionais é limitada pela necessidade de calibração profissional e pela presença física do profissional. Paralelamente, estudos prévios consolidaram a fotografia intraoral via *smartphones* como uma ferramenta promissora para o telediagnóstico, apresentando desempenho não invasivo e satisfatório na distinção binária entre superfícies hígidas e cavitadas [Kohara et al. 2018]. O mesmo estudo, contudo, evidencia uma limitação importante: a acurácia visual humana em imagens bidimensionais reduz-se na distinção de lesões iniciais, devido à sutileza das alterações cromáticas e texturais. Estudos comparativos indicam que a sensibilidade e reprodutibilidade da inspeção visual são limitadas pela variabilidade interexaminador [Gimenez et al. 2015].

Abordagens com *Deep Learning* têm sido investigadas como alternativas para mitigar tais limitações [Krothapalli and Cherukumalli Kapalavayi 2025]. Entre essas abordagens, as Redes Neurais Convolucionais (CNNs) destacam-se pela capacidade de aprender representações hierárquicas diretamente a partir das imagens, dispensando a extração manual de características [LeCun et al. 2015]. No entanto, o treinamento fim-a-fim tradicional é computacionalmente custoso e altamente suscetível ao sobreajuste em cenários médicos de escassez amostral, comuns na área médica [Litjens et al. 2017]. Como alternativa, estratégias de aprendizado por transferência (*Transfer Learning*) baseadas na extração de características (*Feature Extraction*) utilizam CNNs pré-treinadas como extratoras fixas, delegando a etapa decisória a modelos supervisionados externos [Razavian et al. 2014]. Outra via é o ajuste fino parcial (*partial fine-tuning*), que adapta apenas as camadas convolucionais superiores ao novo domínio, preservando as representações de baixo nível previamente aprendidas [Lee et al. 2020].

Diante desse contexto, este trabalho avalia dez arquiteturas de CNN de variadas complexidades na classificação ordinal de lesões de cárie, empregando *Transfer Learning* sob dois paradigmas: abordagens híbridas (CNNs congeladas associadas a SVM, MLP e RF) e ajuste fino parcial. Investiga-se como a interação entre a complexidade arquitetural e a estratégia de adaptação influencia a generalização e a eficácia diagnóstica em cenários de dados limitados, contribuindo com evidências metodológicas para o telediagnóstico.

Este artigo está organizado da seguinte forma: a Seção 2 discute os trabalhos relacionados; a Seção 3 descreve os procedimentos metodológicos; a Seção 4 apresenta os resultados e a respectiva discussão; e, por fim, a Seção 5 expõe as conclusões e as perspectivas de trabalhos futuros.

2. Trabalhos Relacionados

Técnicas de *Machine Learning* e *Deep Learning* têm sido aplicadas para mitigar as limitações da inspeção visual no diagnóstico odontológico. O estudo de [Alabd-Aljabar et al. 2024] propôs uma abordagem híbrida de *Transfer Learning* para ra-

diografias panorâmicas (OPG), combinando extratores profundos (AlexNet) e um SVM, alcançando acurácia de 94%. Embora a arquitetura demonstre a viabilidade de integrar representações profundas a modelos tradicionais com menor custo computacional, o estudo depende de exames radiográficos, o que exige infraestrutura clínica especializada. Em contraste, o presente estudo adapta a integração entre representações profundas e classificadores tradicionais para fotografias intraorais obtidas por *smartphones*, um formato não invasivo que amplia consideravelmente a acessibilidade ao telediagnóstico.

Em [Duong et al. 2021], os autores utilizaram 620 fotografias móveis de dentes extraídos para classificar a severidade da cárie segundo o ICDAS. Para lidar com a complexidade multiclasse, o problema foi convertido em uma cascata binária utilizando o classificador SVM alimentado por atributos cromáticos extraídos manualmente (RGB), atingindo acurácia de 92,3% em lesões cavitadas e 83,3% nas iniciais. Apesar dos bons resultados, os próprios autores destacam que os atributos extraídos manualmente (*handcrafted features*) são sensíveis a reflexos e a variações de iluminação. Por outro lado, a abordagem proposta dispensa a extração manual, empregando redes convolucionais pré-treinadas para aprender representações diretamente das imagens clínicas.

O estudo de [Tareq et al. 2023] avançou na análise de fotografias móveis não padronizadas utilizando um *ensemble* da arquitetura YOLO aliado ao *Transfer Learning* (VGG16), alcançando 86,96% de acurácia global. Mais recentemente, [Adnan et al. 2024] validaram o uso de arquiteturas pré-treinadas (YOLOv5s) em um conjunto de 7.465 imagens de *smartphones*, atingindo um *F1-Score* de 88,0% na detecção de lesões. Contudo, tais abordagens permanecem ancoradas nos paradigmas de detecção binária ou de classificação multiclasse nominal, sem modelar a estrutura ordinal da progressão da doença. Diferentemente, este trabalho incorpora explicitamente a progressão ordinal da cárie (Hígido $<$ Inicial $<$ Cavitado) por meio da Decomposição de Frank e Hall e adota um protocolo com 20 repetições independentes e testes não paramétricos, visando aumentar a robustez estatística em cenários de poucas amostras.

3. Metodologia

Esta seção descreve o fluxo experimental para a classificação ordinal da severidade das lesões de cárie, cujo *pipeline* computacional é ilustrado na Figura 1.

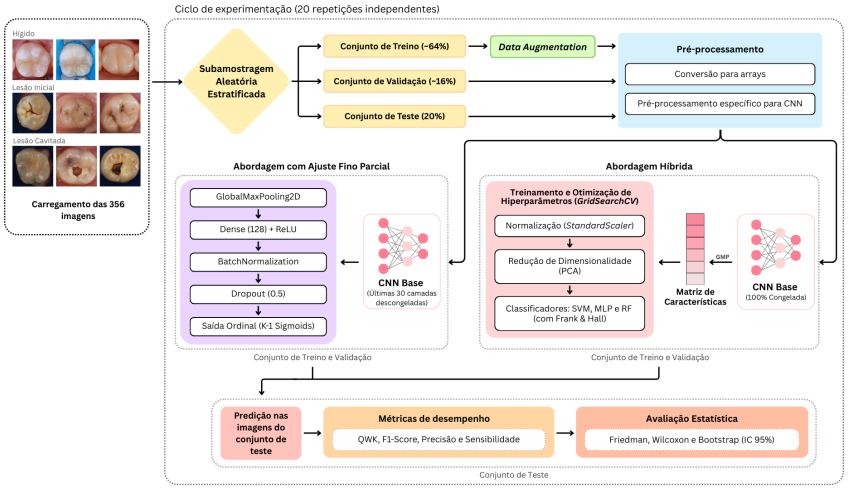

Figura 1. Fluxograma do pipeline experimental.

O experimento foi conduzido em 20 repetições independentes. O banco de fotografias oclusais foi particionado por meio de subamostragem aleatória estratificada, aplicando-se um aumento dinâmico de dados exclusivamente ao subconjunto de treino. Em seguida, os dados foram encaminhados para dois paradigmas de avaliação: extração estática de características com CNNs congeladas, alimentando classificadores clássicos (SVM, MLP, RF), na abordagem híbrida; e ajuste fino parcial das camadas superiores das CNNs. Ambas as vias empregaram a formulação ordinal de Frank e Hall para a classificação de severidade, com avaliação estatística subsequente. Para assegurar o controle da variabilidade estocástica e da independência entre repetições, o particionamento e o aumento de dados ancoraram-se em sementes pseudoaleatórias dinâmicas (*seed* = 42 + i, onde i representa a iteração corrente).

3.1. Ambiente Experimental

O ambiente computacional foi estruturado em Python (v3.12.3) sob Ubuntu 24.04.3 LTS, operando com processador Intel Core i9-13900F, 32 GB de RAM e GPU NVIDIA RTX 4070 (12 GB VRAM). Empregou-se TensorFlow/Keras (v2.19.1) para a extração de características e Scikit-Learn (v1.8.0) para a otimização dos classificadores tradicionais.

3.2. Aquisição de Imagens e Definição das Classes

O conjunto de dados compreende fotografias intraorais e de dentes extraídos, com foco nas superfícies oclusais dos molares permanentes. As imagens institucionais foram obtidas junto ao Programa de Pós-Graduação em Odontologia da Universidade Federal do Ceará, após aprovação pelo Comitê de Ética em Pesquisa (CEP) sob o parecer nº 8.241.809 (CAAE: 94514425.9.0000.5045), possuindo restrição de compartilhamento público. Adicionalmente, o acervo foi complementado com imagens de repositórios públicos. As imagens foram selecionadas seguindo critérios de inclusão que exigiam foco adequado e iluminação suficiente para a visualização clara das alterações estruturais.

A rotulagem baseou-se nos critérios visuais do ICDAS [Pitts and Ekstrand 2013]. Para alinhar o diagnóstico à conduta terapêutica (intervenção não invasiva *versus* invasiva), os códigos originais foram estratificados em três classes de severidade, compondo um conjunto total de 356 imagens: *Hígido* (ICDAS 0; $N = 102$), caracterizada por superfícies sem alterações estruturais; *Lesão Inicial* (ICDAS 1–2; $N = 80$), com alterações restritas ao esmalte, sem cavitação clínica; e *Lesão Cavitada* (ICDAS 3–6; $N = 174$), abrangendo desde microcavitação até exposição evidente de dentina. Superfícies oclusais restauradas com resina composta ou selantes suspeitos de recidiva de cárie foram incluídas e classificadas conforme a severidade da lesão adjacente. A anotação (*ground truth*) ocorreu sob supervisão de especialistas seniores (L.K.A.R. e B.G.N.), com divergências resolvidas por consenso.

3.3. Particionamento e Aumento de Dados

A validação experimental foi conduzida por meio de subamostragem aleatória estratificada repetida (*Repeated Stratified Random Sub-sampling*). O conjunto de dados total foi particionado em subconjuntos de treinamento ($\approx$64%), validação ($\approx$16%) e teste (20%).

Para mitigar o desbalanceamento da distribuição amostral e ampliar a diversidade visual, aplicou-se um protocolo de aumento de dados com balanceamento dinâmico exclusivamente no subconjunto de treinamento em cada repetição da validação. Definiu-se um Alvo de Diversidade correspondente a 2,9 vezes a contagem da classe majoritária

$(2,9 \times N_{C_{max}})$. O fator multiplicativo foi estabelecido após avaliação exploratória de diferentes escalas de expansão (2,0–3,5). Valores inferiores a 2,8 levaram a uma compensação insuficiente, enquanto valores superiores a 3,0 resultaram em degradação progressiva do desempenho, possivelmente associada à redundância sintética. O fator 2,9 apresentou melhor equilíbrio entre estabilidade preditiva e diversidade amostral. O número de amostras geradas por classe correspondeu à diferença entre este alvo e a contagem original.

As transformações (Figura 2) incluíram operações geométricas como rotação aleatória ($\theta \in [-20°, +20°]$) e zoom com recorte central (escala de $1,05\times$ a $1,15\times$), além de variações fotométricas de brilho/contraste ($\pm 10\%$) e nitidez ($1,0\times$ a $1,5\times$).

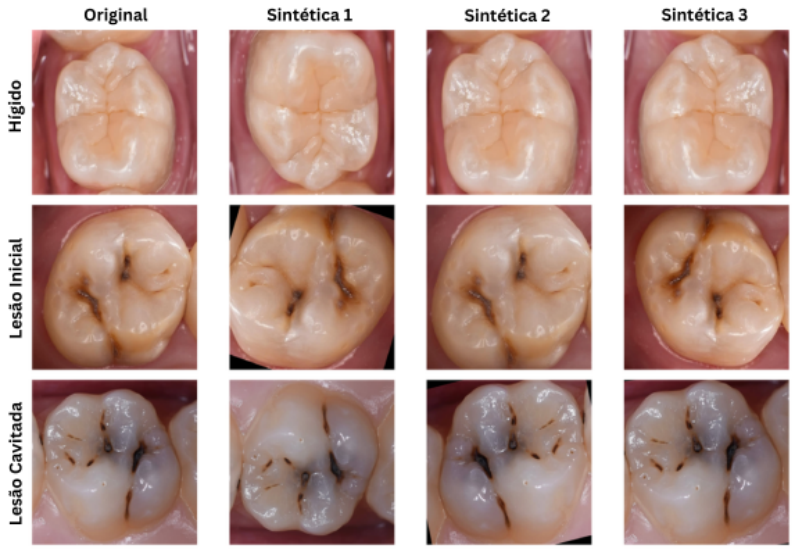

Figura 2. Amostras do protocolo de aumento de dados.

Esta abordagem resultou em um conjunto de treinamento final perfeitamente balanceado, com 321 imagens por classe (Total $N = 963$), mitigando o viés de classe e aumentando a robustez do modelo frente a variações de posicionamento e de iluminação.

3.4. Pré-processamento das Imagens

Após a etapa de aumento de dados, todas as imagens foram submetidas a um fluxo de pré-processamento padronizado. Inicialmente, os arquivos foram carregados e convertidos para *float32* ($[0, 255]$). Os três canais de cor (RGB) foram preservados, considerando a relevância das informações cromáticas para a caracterização visual das lesões de cárie. Na sequência, as imagens foram redimensionadas por interpolação bilinear para os formatos de entrada específicos de cada arquitetura, seguindo as diretrizes de pré-processamento da biblioteca Keras[1] para modelos pré-treinados no ImageNet, conjunto de imagens utilizado como base de pré-treinamento das redes convolucionais empregadas neste estudo.

3.5. Arquiteturas de CNN e Classificadores

Para a análise da severidade das lesões de cárie, foram avaliadas 10 arquiteturas de Redes Neurais Convolucionais (CNNs) pré-treinadas. O critério de seleção

[1] `https://keras.io/api/applications/`

visou contemplar um amplo espectro de paradigmas topológicos e de complexidades paramétricas, abrangendo modelos clássicos de alta capacidade representativa (VGG16 e VGG19 [Simonyan and Zisserman 2015], ResNet50 [He et al. 2016a], ResNet50V2 [He et al. 2016b], InceptionV3 [Szegedy et al. 2016], InceptionResNetV2 [Szegedy et al. 2017]) e variantes modernas otimizadas para eficiência computacional (MobileNetV3Small [Howard et al. 2019], ConvNeXtSmall [Liu et al. 2022], EfficientNetV2B0 e EfficientNetV2B3 [Tan and Le 2021]). Sob o paradigma de *Transfer Learning*, as bases convolucionais foram utilizadas como extratores de representações profundas. Em ambos os paradigmas avaliados, o tensor de saída das camadas convolucionais foi convertido em um vetor de características por meio de *Global Max Pooling*.

Dada a progressão patológica da cárie (Hígido $<$ Inicial $<$ Cavitado), adotou-se a Regressão Ordinal via Decomposição Binária de Frank e Hall [Frank and Hall 2001]. Nesta abordagem, um problema com K classes ordenadas ($V_1 < \cdots < V_K$) é decomposto em $K-1$ classificadores binários, onde o i-ésimo modelo estima $P(y > V_i)$. A probabilidade de cada classe é reconstruída por:

$$P(y = V_i) = P(y > V_{i-1}) - P(y > V_i) \tag{1}$$

Assumindo-se $P(y > V_0) = 1$ e $P(y > V_K) = 0$. A escolha da Decomposição de Frank e Hall justifica-se por sua universalidade, permitindo a adaptação de qualquer classificador tradicional binário ao escopo ordinal, sem a necessidade de projetar funções de perda personalizadas. Essa estrutura matemática foi aplicada transversalmente a ambas as estratégias preditivas avaliadas.

Na abordagem híbrida, os vetores extraídos foram normalizados (*StandardScaler*) e integrados a um *pipeline* contendo redução de dimensionalidade por Análise de Componentes Principais (PCA) [Jolliffe 2002], considerando retenção de 95% e 99% da variância, ou ausência de redução (*passthrough*). Os vetores resultantes alimentaram algoritmos clássicos de *Machine Learning*, nomeadamente *Support Vector Machine* (SVM) [Cortes and Vapnik 1995], *Multilayer Perceptron* (MLP) [Rumelhart et al. 1986] e *Random Forest* (RF) [Breiman 2001], todos adaptados à formulação ordinal de Frank e Hall. A otimização conjunta dos componentes do *pipeline* foi conduzida por meio de busca em grade (*GridSearch*) com partição de validação rigorosa (*PredefinedSplit*), evitando o vazamento de dados. O espaço de busca explorado, com destaque para as configurações mais frequentemente selecionadas (moda), está detalhado na Tabela 1. O coeficiente *Quadratic Weighted Kappa* (QWK) foi utilizado como métrica de seleção.

Para a abordagem de ajuste fino parcial, acoplou-se uma cabeça de decisão à CNN. Visando mitigar o sobreajuste, optou-se pelo descongelamento empírico das 30 camadas finais de cada arquitetura. O tensor de saída das bases convolucionais alimentou uma camada densa de 128 neurônios com ativação ReLU, seguida de *Batch Normalization* e *Dropout* (0,5). Em consonância com a decomposição de Frank e Hall adaptada para redes neurais [Cheng et al. 2008], a camada de saída foi configurada com $K-1$ neurônios (2 unidades) com função de ativação sigmoide independente, onde cada neurônio i estima diretamente $P(y > V_i)$. O modelo foi treinado com o otimizador Adam (1×10^{-5}) e função de perda *Binary Crossentropy*, utilizando *Early Stopping* (paciência de 15, limite de 80 épocas) e *ReduceLROnPlateau*. Na inferência, a consistência monotônica das saídas foi garantida pela restrição $P(y > V_i) = \min(P(y > V_{i-1}), P(y > V_i))$, permitindo a classificação final pela classe de maior probabilidade reconstruída.

Tabela 1. Espaço de busca de hiperparâmetros dos classificadores.

Modelo	Hiperparâmetros
SVM	$C \in \{0.1, 1, \mathbf{10}, 10^2, 10^3\}$, *kernel* $\in \{\text{linear}, \mathbf{rbf}\}$, $\gamma \in \{\text{scale}, \text{auto}, 10^{-2}, 10^{-3}, \mathbf{10^{-4}}\}$
MLP	*hidden_layer_sizes* $\in \{(\mathbf{128}), (256), (128, 64), (256, 128)\}$, *activation* $\in \{\mathbf{relu}, \text{tanh}\}$, $\alpha \in \{10^{-2}, 10^{-3}, \mathbf{10^{-4}}\}$, *learning_rate* $\in \{\mathbf{constant}, \text{adaptive}\}$
RF	*n_estimators* $\in \{100, \mathbf{200}, 500\}$, *max_depth* $\in \{5, \mathbf{10}\}$, *min_samples_leaf* $\in \{\mathbf{1}, 4\}$, *min_samples_split* $\in \{2, \mathbf{5}\}$

Nota: Valores em negrito indicam as configurações mais frequentemente selecionadas pelo *GridSearch* ao longo das repetições experimentais.

3.6. Métricas de Avaliação e Testes Estatísticos

A eficácia diagnóstica foi quantificada primariamente pelo *Kappa* de Cohen Ponderado Quadrático (QWK) [Cohen 1968]. Dada a natureza ordinal da cárie e o desbalanceamento de classes típico de cenários clínicos, a acurácia tradicional é metodologicamente limitada [Sokolova and Lapalme 2009]. O QWK contorna essa limitação ao mensurar a concordância entre as predições e o padrão-ouro, corrigindo os acertos aleatoriamente. Sua ponderação penaliza os erros proporcionalmente ao quadrado da distância entre as predições, atribuindo maior peso a discrepâncias diagnósticas extremas (e.g., dente hígido diagnosticado como cavitado). Complementarmente, avaliaram-se F1-Score, Sensibilidade (*Recall*) e Precisão em regime *macro-average*. A adoção dessa média aritmética não ponderada assegura um impacto idêntico em todas as categorias, impedindo que o alto índice de acertos na classe majoritária (Lesão Cavitada) ofusque o desempenho nas classes patológicas minoritárias.

A avaliação estatística de desempenho foi conduzida em três etapas. Inicialmente, o teste não paramétrico de Friedman comparou globalmente as 40 abordagens. Constatada a diferença significativa, testes pareados de Wilcoxon *post-hoc* isolaram o impacto da escolha do classificador. Estas comparações foram realizadas intra-arquitetura, aplicando-se a correção de Bonferroni para múltiplas hipóteses ($\alpha \approx 0,0083$). Complementarmente, a estabilidade preditiva de todas as abordagens avaliadas foi estimada via *Bootstrap* (10.000 iterações), com extração de IC 95% para todas as métricas. Para a análise de viabilidade operacional, a latência de inferência foi quantificada isolando-se o tempo de *forward pass* da rede base (extração profunda) e somando-o ao tempo de predição do classificador final, reportando-se o custo médio por imagem em milissegundos.

4. Resultados e Discussão

A Tabela 2 consolida o desempenho das 40 abordagens avaliadas, apresentando as métricas em termos de média e seu respectivo Intervalo de Confiança (IC 95%) obtidos ao longo das 20 repetições experimentais.

Tabela 2. Desempenho comparativo (Média, IC 95%) das diferentes arquiteturas CNN e classificadores avaliados no conjunto de teste.

Modelo	QWK	F1-Score (%)	Recall (%)	Precisão (%)
VGG16 + SVM	0,78 (0,75 - 0,81)	77,12 (74,98 - 79,33)	77,47 (75,47 - 79,53)	77,77 (75,50 - 80,08)
VGG16 + MLP	0,77 (0,74 - 0,80)	76,06 (74,17 - 78,03)	76,84 (75,01 - 78,66)	76,37 (74,35 - 78,46)
VGG16 + RF	0,64 (0,59 - 0,68)	65,73 (62,92 - 68,54)	67,41 (65,00 - 69,81)	73,45 (70,53 - 76,29)

VGG16 + Ajuste fino	**0,83 (0,80 - 0,86)**	**78,14 (75,83 - 80,53)**	**78,82 (76,67 - 80,99)**	**80,19 (78,04 - 82,34)**
VGG19 + SVM	0,78 (0,75 - 0,81)	77,60 (75,80 - 79,53)	78,08 (76,22 - 80,12)	77,99 (76,16 - 79,88)
VGG19 + MLP	0,75 (0,72 - 0,78)	74,84 (72,42 - 77,30)	76,01 (73,61 - 78,45)	75,37 (72,92 - 77,90)
VGG19 + RF	0,63 (0,60 - 0,66)	67,20 (64,22 - 70,12)	68,88 (66,33 - 71,37)	73,48 (70,61 - 75,97)
VGG19 + Ajuste fino	**0,84 (0,82 - 0,87)**	**81,22 (78,91 - 83,38)**	**81,46 (79,39 - 83,48)**	**82,43 (80,22 - 84,66)**
ConvNeXtSmall + SVM	0,73 (0,71 - 0,76)	73,20 (70,02 - 76,01)	73,30 (70,16 - 76,16)	74,00 (70,86 - 76,86)
ConvNeXtSmall + MLP	0,69 (0,66 - 0,72)	70,03 (67,28 - 72,94)	70,74 (67,92 - 73,80)	70,57 (67,73 - 73,53)
ConvNeXtSmall + RF	0,57 (0,52 - 0,62)	60,02 (56,79 - 63,39)	62,40 (59,85 - 65,22)	65,90 (62,07 - 69,52)
ConvNeXtSmall + Ajuste fino	**0,74 (0,70 - 0,77)**	**73,94 (70,58 - 77,15)**	**74,26 (71,24 - 77,21)**	**77,01 (73,69 - 80,19)**
ResNet50 + SVM	**0,81 (0,79 - 0,83)**	**79,71 (78,08 - 81,37)**	**80,04 (78,29 - 81,79)**	**80,19 (78,52 - 81,90)**
ResNet50 + MLP	0,80 (0,78 - 0,82)	78,67 (76,51 - 80,77)	79,42 (77,37 - 81,51)	79,38 (77,32 - 81,58)
ResNet50 + RF	0,66 (0,61 - 0,70)	70,62 (67,21 - 73,90)	73,24 (70,53 - 76,01)	72,12 (68,21 - 75,56)
ResNet50 + Ajuste fino	0,73 (0,69 - 0,77)	74,89 (72,44 - 77,20)	75,68 (73,52 - 77,66)	78,15 (75,59 - 80,54)
ResNet50V2 + SVM	**0,83 (0,80 - 0,85)**	**82,11 (80,56 - 83,77)**	**82,82 (81,34 - 84,30)**	**82,10 (80,43 - 83,88)**
ResNet50V2 + MLP	0,80 (0,77 - 0,82)	79,51 (77,44 - 81,65)	80,77 (78,61 - 82,93)	79,43 (77,32 - 81,64)
ResNet50V2 + RF	0,61 (0,57 - 0,65)	64,51 (61,14 - 67,49)	67,34 (64,85 - 69,70)	71,29 (67,06 - 75,19)
ResNet50V2 + Ajuste fino	0,72 (0,68 - 0,75)	74,34 (72,08 - 76,76)	75,35 (73,39 - 77,44)	77,70 (75,12 - 80,38)
InceptionV3 + SVM	**0,75 (0,72 - 0,78)**	**78,48 (76,73 - 80,27)**	**78,93 (77,00 - 80,76)**	**78,99 (77,42 - 80,66)**
InceptionV3 + MLP	0,72 (0,69 - 0,75)	77,57 (75,85 - 79,41)	78,54 (76,66 - 80,41)	77,94 (76,17 - 79,80)
InceptionV3 + RF	0,52 (0,48 - 0,56)	55,84 (50,64 - 61,15)	61,18 (57,45 - 65,10)	59,09 (52,42 - 65,79)
InceptionV3 + Ajuste fino	0,66 (0,61 - 0,70)	73,08 (70,27 - 75,69)	73,93 (71,46 - 76,35)	76,54 (74,03 - 79,15)
InceptionResNetV2 + SVM	**0,75 (0,71 - 0,78)**	75,66 (72,91 - 78,25)	76,03 (73,46 - 78,37)	**76,77 (73,91 - 79,60)**
InceptionResNetV2 + MLP	0,74 (0,71 - 0,77)	**76,31 (73,96 - 78,72)**	**77,39 (75,19 - 79,51)**	76,75 (74,02 - 79,69)
InceptionResNetV2 + RF	0,60 (0,56 - 0,65)	65,09 (60,58 - 69,23)	67,62 (64,24 - 70,89)	67,33 (61,52 - 72,59)
InceptionResNetV2 + Ajuste fino	0,69 (0,66 - 0,73)	71,81 (68,66 - 75,03)	72,90 (70,00 - 75,92)	74,84 (71,61 - 78,21)
MobileNetV3Small + SVM	0,83 (0,80 - 0,85)	80,09 (77,81 - 82,17)	80,11 (78,09 - 82,00)	81,04 (78,53 - 83,37)
MobileNetV3Small + MLP	**0,83 (0,81 - 0,86)**	**80,57 (77,95 - 82,97)**	**81,01 (78,61 - 83,30)**	**81,67 (79,23 - 83,89)**
MobileNetV3Small + RF	0,67 (0,64 - 0,71)	66,52 (62,54 - 70,10)	68,91 (66,18 - 71,38)	69,25 (63,94 - 73,78)
MobileNetV3Small + Ajuste fino	0,56 (0,47 - 0,63)	61,56 (55,49 - 67,30)	63,76 (57,74 - 69,34)	65,30 (59,19 - 70,94)
EfficientNetV2B0 + SVM	**0,83 (0,80 - 0,86)**	80,40 (78,28 - 82,63)	80,50 (78,48 - 82,59)	**81,24 (78,86 - 83,71)**
EfficientNetV2B0 + MLP	0,82 (0,79 - 0,84)	**80,52 (78,62 - 82,36)**	**81,28 (79,48 - 82,98)**	80,77 (78,88 - 82,67)
EfficientNetV2B0 + RF	0,67 (0,63 - 0,71)	67,75 (64,44 - 70,90)	69,79 (67,14 - 72,50)	69,00 (64,64 - 72,90)
EfficientNetV2B0 + Ajuste fino	0,69 (0,65 - 0,73)	73,02 (70,81 - 75,29)	74,42 (72,37 - 76,52)	74,90 (72,67 - 77,07)
EfficientNetV2B3 + SVM	**0,84 (0,82 - 0,87)**	**82,58 (80,33 - 84,74)**	82,54 (80,27 - 84,74)	**83,41 (81,14 - 85,64)**
EfficientNetV2B3 + MLP	**0,84 (0,81 - 0,87)**	82,36 (80,12 - 84,60)	**83,33 (81,12 - 85,43)**	82,40 (80,09 - 84,71)
EfficientNetV2B3 + RF	0,68 (0,64 - 0,72)	69,97 (65,48 - 74,15)	72,57 (68,86 - 76,03)	72,13 (67,26 - 76,37)
EfficientNetV2B3 + Ajuste fino	0,72 (0,68 - 0,75)	72,87 (69,88 - 75,78)	74,50 (71,73 - 77,16)	74,72 (71,84 - 77,54)

Os resultados evidenciam que a extração estática de características, acoplada a classificadores híbridos (SVM e MLP), constitui uma alternativa robusta e competitiva. Os classificadores SVM e MLP acoplados à EfficientNetV2B3, bem como a VGG19 com ajuste fino parcial, obtiveram os maiores valores de QWK (≈0,84), com F1-Score, Sensibilidade e Precisão superiores a 81%. O teste de Friedman confirmou diferenças globais significativas ($p < 0,001$), legitimando as análises *post-hoc* de Wilcoxon.

A análise estatística par a par, cujos valores-p estão consolidados na Tabela 3, indicou dependência entre a complexidade arquitetural da base convolucional e a estratégia de aprendizado adotada. Em arquiteturas como VGG e ConvNeXt, o ajuste fino parcial apresentou desempenho competitivo, com superioridade estatisticamente significativa observada na VGG19. Em contraste, em arquiteturas como EfficientNet, ResNet, Inception e MobileNet, observou-se uma tendência de melhor desempenho das abordagens híbridas, com SVM e MLP superando o ajuste fino em diversas comparações com significância estatística. Esse padrão sugere maior sensibilidade de determinadas arquiteturas à atualização parcial de pesos em cenários de amostras restritas, possivelmente associada ao risco de sobreajuste, reforçando o *Transfer Learning* híbrido como estratégia estável em regimes de escassez amostral.

Tabela 3. Valores-p do teste pareado de Wilcoxon intra-arquitetura.

Arquitetura	SVM *vs* MLP	SVM *vs* RF	SVM *vs* AF	MLP *vs* RF	MLP *vs* AF	RF *vs* AF
VGG16	0,7012	**<0,001**	0,0107	**<0,001**	**0,0049**	**<0,001**
VGG19	0,0441	**<0,001**	**0,0010**	**<0,001**	**<0,001**	**<0,001**
ConvNeXtSmall	0,0484	**<0,001**	0,6742	**<0,001**	0,0637	**<0,001**
ResNet50	0,0532	**<0,001**	**0,0006**	**<0,001**	**0,0027**	**0,0049**
ResNet50V2	0,0121	**<0,001**	**<0,001**	**<0,001**	**<0,001**	**<0,001**
InceptionV3	0,0240	**<0,001**	**<0,001**	**<0,001**	0,0094	**<0,001**
InceptionResNetV2	0,6009	**<0,001**	**0,0020**	**<0,001**	**0,0049**	**0,0017**
MobileNetV3Small	0,8124	**<0,001**	**<0,001**	**<0,001**	**<0,001**	0,0266
EfficientNetV2B0	0,4980	**<0,001**	**<0,001**	**<0,001**	**<0,001**	0,3300
EfficientNetV2B3	0,8695	**<0,001**	**<0,001**	**<0,001**	**<0,001**	0,0240

Nota: Valores em negrito indicam diferença estatística significativa após correção de Bonferroni ($\alpha \approx 0,0083$). AF = Ajuste Fino.

Paralelamente, observou-se um subdesempenho sistemático do RF, que se mostrou estatisticamente inferior ao SVM e ao MLP em todas as configurações. Esse comportamento pode estar relacionado à limitação estrutural de métodos baseados em particionamento quando aplicados a vetores profundos de alta dimensionalidade e com correlação latente. Em contrapartida, não foram identificadas diferenças significativas entre SVM e MLP, sugerindo eficácia comparável na separação dos vetores profundos. A análise de robustez via *Bootstrap* (IC 95%) indicou uma ampla sobreposição entre os intervalos de confiança da métrica QWK nas arquiteturas de melhor desempenho. Essa convergência é observada, por exemplo, nas combinações EfficientNetV2B3 + SVM [0,82; 0,87], VGG19 + ajuste fino [0,82; 0,87] e MobileNetV3Small + MLP [0,81 – 0,86]. Em contraste, arquiteturas com desempenho global inferior, como InceptionV3, InceptionResNetV2 e ConvNeXtSmall, apresentaram intervalos isolados, com limites superiores abaixo de 0,79. Dada essa convergência no topo da distribuição, a escolha da topologia clínica final pode ser estrategicamente guiada por restrições operacionais de *hardware*.

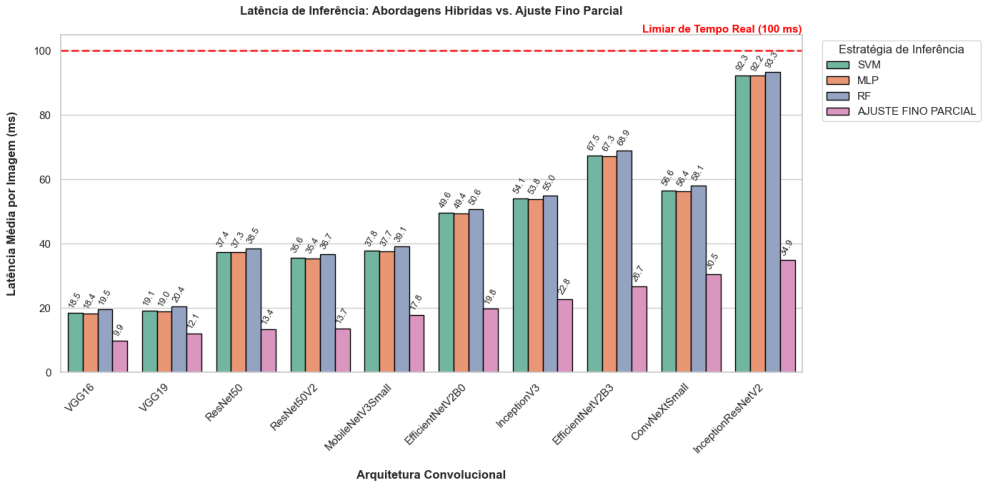

Figura 3. Latência de inferência em abordagens híbridas e ajuste fino parcial.

Por fim, a viabilidade do telediagnóstico foi corroborada pela análise da latência de inferência (Figura 3). Nas abordagens híbridas, o tempo foi dominado pela extração convolucional, com variação mínima entre classificadores (≈1–3 ms). O ajuste fino parcial apresentou latências menores, com reduções típicas de 40–60%. Essa diferença decorre do *overhead* adicional das abordagens híbridas, que envolvem etapas intermediárias

de processamento e execução desacoplada, ao passo que o ajuste fino opera de ponta a ponta na GPU. Observou-se ainda que arquiteturas mais complexas implicam maior latência absoluta. Notavelmente, a maioria das configurações permaneceu abaixo do limiar de tempo real ($\approx$100 ms), evidenciando viabilidade prática e a influência conjunta da arquitetura e da estratégia de inferência. Embora latências em dispositivos móveis tendam a ser superiores, arquiteturas leves como MobileNetV3Small permanecem promissoras para aplicações em tempo real.

5. Conclusão e Trabalhos Futuros

Este estudo avaliou sistematicamente a classificação ordinal da severidade de lesões de cárie, comparando abordagens híbridas de extração de características com estratégias de ajuste fino parcial em dez arquiteturas convolucionais. Os resultados demonstraram a viabilidade empírica do telediagnóstico mesmo sob restrições amostrais, evidenciando que o desempenho preditivo depende do alinhamento entre a complexidade arquitetural da rede e a estratégia de aprendizado empregada. Como fundamento metodológico, a decomposição de Frank e Hall assegurou coerência matemática às decisões do modelo, respeitando a progressão biológica natural da doença.

A sobreposição dos intervalos de confiança no topo do desempenho preditivo, aliada à análise de latência, sugere que múltiplas configurações apresentam viabilidade técnica para futura aplicação clínica em tempo real. Ressalta-se, porém, que o presente estudo concentra sua avaliação em métricas quantitativas laboratoriais. Como desdobramentos futuros, recomenda-se a expansão do conjunto amostral e a investigação de funções de perda explicitamente ordinais. Adicionalmente, faz-se necessária a validação externa dos *pipelines* de alto desempenho em cenários clínicos distintos e coortes multicêntricas. Essa etapa é indispensável para avaliar o comportamento do modelo sob condições heterogêneas de uso no mundo real (como imagens de *smartphones*) e discutir potenciais riscos diagnósticos. Por fim, reconhece-se que a natureza *black box* das arquiteturas profundas avaliadas representa uma barreira significativa à adoção clínica. Assim, a integração de métodos de Inteligência Artificial Explicável (XAI) constitui um passo fundamental para trabalhos subsequentes.

6. Declaração de uso da IA Generativa

Não antecipamos quaisquer preocupações sociais ou éticas imediatas decorrentes deste trabalho. Além disso, reconhecemos o uso de LLMs para auxiliar na verificação de gramática e estilo em partes do manuscrito.

Referências

Adnan, N., Ahmed, S. M. F., Das, J. K., Aijaz, S., Sukhia, R. H., Hoodbhoy, Z., and Umer, F. (2024). Developing an ai-based application for caries index detection on intraoral photographs. *Scientific Reports*, 14:26752.

Alabd-Aljabar, A., Raisan, Z., Adnan, M., and Dhou, S. (2024). A hybrid transfer learning approach to teeth diagnosis using orthopantomogram radiographs. *IEEE Access*, 12:178142–178152.

Bader, J. D. and Shugars, D. A. (2006). The evidence supporting alternative management strategies for early occlusal caries and suspected occlusal dentinal caries. *Journal of Evidence Based Dental Practice*, 6(1):91–100.

Bernabe, E., Marcenes, W., Abdulkader, R. S., et al. (2025). Trends in the global, regional, and national burden of oral conditions from 1990 to 2021: a systematic analysis for the global burden of disease study 2021. *The Lancet*, 405(10482):897–910.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Cheng, J., Wang, Z., and Pollastri, G. (2008). A neural network approach to ordinal regression. In *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN)*, pages 1279–1284.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4):213–220.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297.

Duong, D. L., Kabir, M. H., and Kuo, R. F. (2021). Automated caries detection with smartphone color photography using machine learning. *Health Informatics Journal*, 27(2):14604582211007530.

Frank, E. and Hall, M. (2001). A simple approach to ordinal classification. In De Raedt, L. and Flach, P., editors, *Machine Learning: ECML 2001*, pages 145–156, Berlin, Heidelberg. Springer Berlin Heidelberg.

Frencken, J. E. (2017). Atraumatic restorative treatment and minimal intervention dentistry. *British Dental Journal*, 223(3):183–189.

Gimenez, T., Piovesan, Carmel e Braga, M. M., Ricketts, D. N., and Mendes, F. M. (2015). Visual inspection for caries detection: a systematic review and meta-analysis. *Journal of Dental Research*, 94(7):895–904.

He, K., Zhang, X., Ren, S., and Sun, J. (2016a). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778.

He, K., Zhang, X., Ren, S., and Sun, J. (2016b). Identity mappings in deep residual networks. In *European Conference on Computer Vision*, pages 630–645. Springer.

Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q. V., and Adam, H. (2019). Searching for mobilenetv3. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1314–1324.

Ismail, A. I., Sohn, W., Tellez, M., Amaya, A., Sen, A., Hasson, H., and Pitts, N. B. (2007). The international caries detection and assessment system (icdas): an integrated system for measuring dental caries. *Community Dentistry and Oral Epidemiology*, 35(3):170–178.

Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer, New York, NY, 2nd edition.

Kohara, E. K., Abdala, C. G., Novaes, T. F., Braga, M. M., Haddad, A. E., and Mendes, F. M. (2018). Is it feasible to use smartphone images to perform telediagnosis of different stages of occlusal caries lesions? *PLoS ONE*, 13(9):e0202116.

Krothapalli, N. and Cherukumalli Kapalavayi, N. (2025). Deep learning in dental diagnostics: Caries detection through smartphone photographs – a systematic review. *Journal of Global Oral Health*, 8:91–97.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

Lee, D.-H., Li, Y., and Shin, B.-S. (2020). Mid-level feature extraction method based transfer learning to small-scale dataset of medical images with visualizing analysis. *Journal of Information Processing Systems*, 16(6):1293–1308.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., and Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88.

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., and Xie, S. (2022). A convnet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11966–11976.

Pitts, N. B. and Ekstrand, K. R. (2013). International caries detection and assessment system (icdas) and its international caries classification and management system (iccms): methods for staging of the caries process and enabling dentists to manage caries. *Community Dentistry and Oral Epidemiology*, 41:e41–e52.

Razavian, A. S., Azizpour, H., Sullivan, J., and Carlsson, S. (2014). Cnn features off-the-shelf: An astounding baseline for recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 512–519.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088):533–536.

Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations (ICLR)*.

Sokolova, M. and Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information processing & management*, 45(4):427–437.

Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2818–2826.

Tan, M. and Le, Q. V. (2021). Efficientnetv2: Smaller models and faster training. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 10096–10106.

Tareq, A., Faisal, M. I., Islam, M. S., Rafa, N. S., Chowdhury, T., Ahmed, S., Farook, T. H., Mohammed, N., and Dudley, J. (2023). Visual diagnostics of dental caries through deep learning of non-standardised photographs using a hybrid yolo ensemble and transfer learning model. *International Journal of Environmental Research and Public Health*, 20(7):5351.