

# Redes Neurais Profundas para classificar imagens de radiografias odontológicas panorâmicas na perspectiva de grupos etários: Um estudo exploratório

Michele Faria de Oliveira<sup>1</sup>, Fischer Jônatas Ferreira<sup>2</sup>, Iális Cavalcante de Paula Júnior<sup>1</sup>

<sup>1</sup>Universidade Federal do Ceará (UFC) – Sobral, CE – Brazil

<sup>2</sup>Universidade Federal de Itajubá (UNIFEI) – Itabira, MG – Brazil

mfariaoliveira08@gmail.com, fischer@unifei.br, ialis@sobral.ufc.br

**Abstract.** *Age estimation from panoramic dental radiographs has important applications in forensic, clinical, and legal contexts. Traditional methods based on dental development analysis present limitations related to subjectivity and difficulties in generalizing across populations. In this context, deep neural networks have been investigated as an alternative to automate this process. This study evaluates the performance of different convolutional neural network architectures for age group classification using panoramic radiographs. Experiments were conducted with nine architectures using preprocessing techniques, data augmentation, and transfer learning. The results indicate that the EfficientNetB5 architecture achieved the best performance, followed by ResNet50 and InceptionV3, highlighting the potential of these approaches for applications in forensic dentistry and decision support systems.*

**Resumo.** *A estimativa de idade a partir de radiografias panorâmicas odontológicas tem aplicações relevantes em contextos forenses, clínicos e legais. Métodos tradicionais baseados na análise do desenvolvimento dentário apresentam limitações relacionadas à subjetividade e à dificuldade de generalização entre populações. Nesse cenário, redes neurais profundas têm sido investigadas como alternativa para automatizar esse processo. Este trabalho avalia o desempenho de diferentes arquiteturas de redes neurais convolucionais na classificação de faixas etárias a partir de radiografias panorâmicas. Foram conduzidos experimentos com nove arquiteturas, utilizando técnicas de pré-processamento, aumento de dados e transfer learning. Os resultados indicaram melhor desempenho da arquitetura EfficientNetB5, seguida por ResNet50 e InceptionV3, evidenciando o potencial dessas abordagens para aplicações em odontologia forense e sistemas de apoio à decisão.*

## 1. Introdução

A estimativa da idade humana a partir da arcada dentária é uma tarefa relevante no contexto da odontologia forense Ozlu Ucan et al. (2025). Em cenários que exigem identificação humana, como em casos de vítimas de desastres, na análise de maioridade em contextos legais ou no planejamento de tratamentos odontológicos, a avaliação do desenvolvimento dental pode fornecer informações importantes sobre a idade do indivíduo Nino-Sandoval et al. (2024); Gracea et al. (2025). Métodos tradicionais, como

os propostos por *Demirjian, Nolla e Willems*, são amplamente utilizados na literatura e estabelecem uma relação entre os estágios de desenvolvimento dos dentes e a idade cronológica Sivri et al. (2024); Baylis et al. (2026). No entanto, estudos apontam limitações nesses métodos, uma vez que sua aplicação depende, em grande parte, da interpretação visual realizada por especialistas. Além disso, eles foram desenvolvidos a partir de populações específicas, o que dificulta sua generalização para diferentes contextos populacionais e compromete sua precisão Singh et al. (2026).

Nesse contexto, abordagens baseadas em *Deep Learning* têm sido investigadas como alternativas para automatizar e aprimorar o processo de estimativa de idade a partir de radiografias panorâmicas odontológicas Ong et al. (2024); Oliveira et al. (2024). Essas técnicas têm apresentado resultados promissores, especialmente em tarefas de visão computacional aplicadas à área da saúde, devido à sua capacidade de extrair automaticamente padrões complexos das imagens Moreira et al. (2024); Oliveira et al. (2024). Conhecer as redes neurais utilizadas no problema é fundamental para garantir a eficácia do modelo preditivo, especialmente em aplicações sensíveis como a estimativa de idade Sivri et al. (2024); Tokinaga et al. (2025). Cada arquitetura apresenta características distintas em termos de profundidade, capacidade de generalização e tempo de processamento Koch et al. (2025). A escolha inadequada pode resultar em modelos com alto índice de falsos positivos, comprometendo a confiabilidade dos resultados.

Para que o uso dessas redes seja eficaz, é indispensável que apresentem um bom nível de acerto, especialmente em diferentes faixas etárias e conjuntos de dados Nushi et al. (2025). Assim, compreender o comportamento de cada rede em relação ao problema proposto permite selecionar aquelas com melhor desempenho, garantindo maior precisão nas estimativas e reduzindo o risco de decisões equivocadas em contextos clínicos, forenses ou judiciais Sivri et al. (2024). Diferentes arquiteturas de redes neurais profundas apresentam características distintas quanto à capacidade de generalização, ao desempenho e ao custo computacional Khanagar et al. (2024). No entanto, faltam estudos para uma compreensão mais aprofundada do comportamento dessas arquiteturas no contexto da estimativa de idade e da identificação de modelos mais adequados e confiáveis para aplicações práticas.

Diante disso, este trabalho investiga o desempenho de diferentes arquiteturas de redes neurais profundas na classificação de faixas etárias a partir de radiografias panorâmicas odontológicas, buscando identificar modelos mais eficazes para essa tarefa. Para atingir esse objetivo, foi montado um banco de imagens composto por radiografias panorâmicas odontológicas de indivíduos brasileiros. As arquiteturas avaliadas foram: *ResNet101*, *ResNet50*, *AlexNet*, *VGG-16*, *DenseNet201*, *EfficientNetB5*, *EfficientNetB3*, *MobileNetV2* e *InceptionV3*. Por fim, o desempenho dos modelos foi avaliado com base em métricas, como acurácia, precisão, *recall*, *F1-Score* e especificidade.

Logo, as principais contribuições deste trabalho são listadas a seguir: (I) Investigação do desempenho de diferentes arquiteturas de redes neurais profundas para a classificação de faixas etárias a partir de radiografias panorâmicas odontológicas. (II) Disponibilização pública de um pacote de replicação contendo dataset<sup>1</sup>. (III) Os experimentos foram conduzidos utilizando uma combinação de conjuntos de dados.

---

<sup>1</sup><https://ageestimationsbcas2026.github.io/Age-estimation-from-panoramic-dental-radiographs/>

O restante do trabalho está organizado da seguinte forma: a Seção 2 discute alguns trabalhos relacionados. A Seção 3 apresenta a metodologia utilizada para a condução do trabalho. A Seção 4 apresenta e discute os resultados alcançados. A Seção 5 apresenta as implicações deste estudo. Finalmente, as Seções 6 e 7 apresentam as ameaças à validade e a conclusão do estudo, respectivamente.

## 2. Trabalhos Relacionados

Esta seção discute alguns trabalhos relacionados sobre os seguintes aspectos: estimativa da idade automatizada usando aprendizado profundo, estado do desenvolvimento dental usando aprendizado profundo e comparação de métodos tradicionais (manuais) e automatizados de estimativa de idade.

*Estimativa Automatizada da Idade Dental Usando Aprendizado Profundo:* Estudos recentes têm explorado o uso de aprendizado profundo para estimativa automatizada da idade a partir de radiografias. O trabalho de Mualla et al., 2020 utiliza redes pré-treinadas (*AlexNet* e *ResNet101*) para extração de características e classificadores, destacando o melhor desempenho da combinação *AlexNet* com KNN. De forma semelhante, Ataş et al., 2022 avalia arquiteturas de aprendizado por transferência em radiografias panorâmicas, com melhor desempenho da *InceptionV3* (MAE de 3,13 anos). No contexto brasileiro, Oliveira et al., 2024 utiliza uma adaptação da *InceptionV4* com aumento de dados e alcança MAE de 3,83 anos. Por fim, Wang et al., 2023 compara *VGG16* e *ResNet101* em uma população chinesa, observando melhor desempenho da *VGG16*, especialmente em indivíduos mais jovens.

*Comparação de Métodos Manuais e Automatizados de Estimativa da Idade:* Alguns estudos comparam métodos tradicionais da odontologia e abordagens baseadas em aprendizado profundo para estimativa da idade. O trabalho de Guo et al., 2021 compara um método manual com um modelo automatizado aplicado a radiografias panorâmicas, considerando os limiares etários de 14, 16 e 18 anos. Os resultados indicam desempenho superior dos modelos automáticos. De forma semelhante, Sivri et al., 2024 analisa o método de *Demirjian* em comparação com arquiteturas, como *DenseNet201* e *InceptionV3*. O estudo aponta limitações do método *Demirjian*, como subjetividade e maior tempo de análise. Entretanto, indica boa correspondência entre as abordagens.

*Estado do Desenvolvimento Dental Usando Aprendizado Profundo:* Alguns estudos têm investigado o uso de aprendizado profundo para a análise do desenvolvimento dental em radiografias panorâmicas. O trabalho de Ong et al., 2024 propõe um sistema totalmente automatizado baseado em três etapas: detecção com *YOLOv5*, segmentação com *U-Net* e classificação com *EfficientNet*, obtendo bons resultados na identificação dos estágios de desenvolvimento dentário. De forma semelhante, Shi et al., 2024 apresenta um *framework* em três etapas que utiliza *YOLOv3* para detecção e numeração dos dentes, seguido da rede *SOS-Net* para estimativa do estágio de desenvolvimento e de uma meta-análise estatística para inferir a idade do indivíduo.

## 3. Configuração do Estudo

O estudo foi conduzido em duas etapas. Na primeira, realizou-se a classificação de radiografias panorâmicas em três faixas etárias (1–14, 15–25 e 26–91 anos) por meio do treinamento e avaliação de diferentes arquiteturas de redes neurais convolucionais. Na

segunda etapa, foi realizada uma validação externa, na qual os modelos com melhor desempenho foram treinados em um conjunto de dados independente e testados com o conjunto de dados utilizados na primeira etapa, a fim de avaliar a capacidade de generalização dos modelos. O fluxo metodológico compreendeu as seguintes etapas: ARTEFATOS DO ESTUDO, CLASSIFICAÇÃO, VALIDAÇÃO EXTERNA e MÉTRICAS DE AVALIAÇÃO.

### 3.1. Artefatos do Estudo

*Base de Dados:* O estudo utilizou dois conjuntos de imagens de radiografias panorâmicas no formato JPEG. O primeiro, construído para este projeto, é composto por 1.545 imagens de indivíduos com idades entre 1 e 91 anos. As imagens foram organizadas nas faixas etárias de 1–14, 15–25 e 26–91 anos, divisão fundamentada nos marcos fisiológicos do desenvolvimento dentário, a fim de favorecer a presença de características morfológicas relevantes nas imagens para a distinção entre os grupos etários. Para a etapa de validação externa, utilizou-se um segundo conjunto de dados composto por 10.036 imagens, com indivíduos de idades de 3 a 96 anos Oliveira et al. (2024). Esse conjunto também foi previamente organizado em três faixas etárias 1–14, 15–25 e 26–99 anos, respeitando a proposta metodológica.

*Ambiente de Desenvolvimento:* Os experimentos foram desenvolvidos na plataforma *Google Colab Pro*, ambiente de computação em nuvem que oferece acesso a unidades de processamento gráfico (GPU), essenciais para o treinamento de modelos de aprendizado profundo. A integração com o *Google Drive* também facilitou o armazenamento das imagens e dos modelos treinados. As implementações foram realizadas em *Python*, utilizando bibliotecas amplamente empregadas em aprendizado profundo. O pré-processamento das imagens foi realizado com *OpenCV*, enquanto o treinamento e a construção dos modelos utilizaram *PyTorch* e *PyTorch Lightning*. Para manipulação de dados e visualização de resultados foram utilizadas as bibliotecas *NumPy*, *Matplotlib* e *Seaborn*.

*Arquitetura de Redes Neurais Profundas:* As arquiteturas utilizadas neste estudo foram selecionadas com base em resultados de estudos anteriores de Oliveira et al. (2025). Foram escolhidos modelos que apresentaram bom desempenho em tarefas de classificação de imagens médicas, considerando também as limitações dos recursos computacionais disponíveis. Assim, foram avaliadas as arquiteturas *ResNet101*, *ResNet50*, *AlexNet*, *VGG-16*, *DenseNet201*, *EfficientNetB5*, *EfficientNetB3*, *MobileNetV2* e *InceptionV3*.

### 3.2. Classificação

A etapa de classificação teve como objetivo avaliar o desempenho de diferentes arquiteturas de redes neurais convolucionais na identificação de faixas etárias a partir de radiografias panorâmicas. Para isso, foram utilizados modelos pré-treinados. Foram avaliadas nove arquiteturas de redes neurais profundas: *ResNet101*, *ResNet50*, *AlexNet*, *VGG-16*, *DenseNet201*, *EfficientNetB5*, *EfficientNetB3*, *MobileNetV2* e *InceptionV3*. O treinamento foi realizado por até 50 épocas utilizando a função de perda entropia cruzada, com otimização baseada em Adam ou Nadam e estratégia de parada antecipada (*early stopping*) com paciência de 10 épocas.

*Validação Cruzada:* Foi utilizada validação cruzada do tipo *Holdout*, com divisão estratificada do conjunto de dados em 80% para treinamento e 20% para avaliação, sendo esta última subdividida em 10% para validação durante o treinamento e 10% para teste.

*Pré-Processamento dos Dados:* As imagens foram redimensionadas para  $224 \times 224$  pixels e submetidas a uma máscara para delimitação da região de interesse (ROI). Em seguida, foram convertidas para escala de cinza, filtradas por meio de filtragem bilateral para redução de ruído com preservação de bordas e, por fim, teve-se aplicado o CLAHE para aprimoramento do contraste local.

*Aumento de Dados:* Foram aplicadas técnicas de *data augmentation* nas imagens de treinamento para aumentar a variabilidade dos dados e reduzir o risco de *overfitting*. As transformações incluíram variações de brilho e contraste, inversão horizontal, além de operações de deslocamento, escala e rotação, simulando variações comuns em imagens clínicas e contribuindo para melhorar a capacidade de generalização dos modelos.

*Transfer Learning:* Para reduzir o custo computacional e o tempo de treinamento, adotou-se a técnica de *Transfer Learning*. Nessa abordagem, as arquiteturas utilizadas foram inicializadas com pesos previamente treinados no conjunto de dados *ImageNet*, permitindo aproveitar características já aprendidas em tarefas de classificação de imagens.

*Validação Externa:* A validação externa teve como objetivo avaliar a capacidade de generalização dos modelos em dados distintos dos utilizados no treinamento inicial. As três arquiteturas com melhor desempenho foram treinadas com o conjunto de dados independente descrito por Oliveira et al., 2024 e posteriormente testadas no banco de dados deste estudo, sem pré-processamento.

*Métricas de Avaliação:* O desempenho dos modelos foi avaliado por meio de métricas amplamente utilizadas em problemas de classificação multiclasse. Foram consideradas as métricas de acurácia, precisão, *recall*, *F1-score* e especificidade. Essas métricas foram calculadas com base na matriz de confusão, que organiza as previsões em verdadeiros positivos, falsos positivos, verdadeiros negativos e falsos negativos. Todas as avaliações foram realizadas exclusivamente com o conjunto de teste, garantindo uma análise imparcial da capacidade de generalização dos modelos.

## 4. Resultados e Discussão

Esta seção apresenta os resultados das nove arquiteturas avaliadas. O desempenho dos modelos na classificação etária foi analisado por meio das métricas de acurácia, precisão, *recall*, *F1-score* e especificidade, além do tempo de treinamento de cada arquitetura.

### 4.1. Resultados Encontrados para as Métricas Essenciais

A Tabela 1 apresenta os resultados das métricas para as nove arquiteturas avaliadas. De modo geral, os resultados indicam desempenho satisfatório na tarefa de classificação das faixas etárias a partir de radiografias panorâmicas. Entre os modelos avaliados, a *Efficient-NetB5* apresentou o melhor desempenho global, alcançando acurácia de 86,36%, precisão de 87,47%, *recall* de 86,36%, *F1-score* de 86,59% e especificidade de 93,05%. Esse resultado sugere maior capacidade do modelo em distinguir corretamente as diferentes classes etárias, mantendo bom equilíbrio entre falsos positivos e falsos negativos.

As arquiteturas *InceptionV3* e *ResNet50* também apresentaram desempenho competitivo, ambas com acurácia de 83,77%. A *InceptionV3* destacou-se pelo menor tempo de treinamento, enquanto a *ResNet50* apresentou valores ligeiramente superiores de precisão e especificidade, indicando bom desempenho geral.

**Tabela 1. Comparação de desempenho entre diferentes arquiteturas de redes**

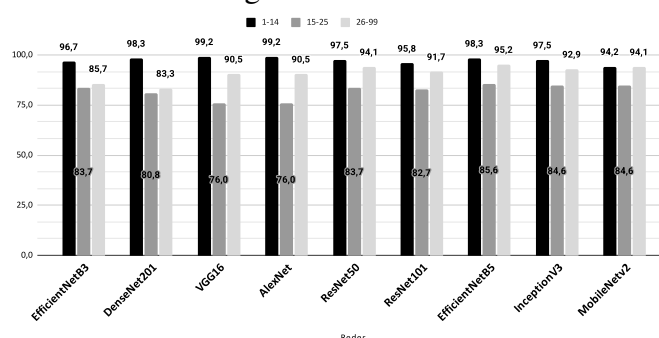
Rede	Acurácia (%)	Precisão (%)	Recall (%)	F1-Score (%)	Especificidade(%)	Tempo (s)
ResNet101	80,52	81,35	80,52	80,78	90,06	1.219
ResNet50	83,77	84,97	83,77	84,01	91,73	2.911
AlexNet	77,92	81,19	77,92	78,41	88,53	3.536
VGG-16	77,92	81,05	77,92	78,51	88,53	3.575
DenseNet201	76,62	77,72	76,62	76,62	87,48	3.029
EfficientNetB5	86,36	87,47	86,36	86,59	93,05	2.122
EfficientNetB3	78,57	78,67	78,57	78,62	88,68	1.915
MobileNetV2	81,82	82,65	81,82	81,99	90,64	1.280
InceptionV3	83,77	84,62	83,77	84,01	91,66	1.506

Arquiteturas como *MobileNetV2* e *ResNet101* apresentaram resultados intermediários, combinando desempenho razoável com tempos de treinamento reduzidos. Em contraste, modelos mais antigos, como *AlexNet* e *VGG-16*, obtiveram métricas inferiores, sugerindo menor capacidade de captura de padrões relevantes para a tarefa de classificação etária.

A arquitetura *DenseNet201* apresentou o menor desempenho entre os modelos avaliados, possivelmente devido à maior complexidade estrutural, que pode dificultar a generalização em conjuntos de dados específicos como radiografias panorâmicas. De forma geral, os resultados indicam que arquiteturas mais recentes e otimizadas, como a família *EfficientNet*, apresentam maior potencial para aplicações de classificação etária em imagens radiográficas.

## 4.2. Resultados Encontrados para Especificidade

A Figura 1 apresenta a comparação da especificidade por classe etária entre as arquiteturas avaliadas. De forma geral, observou-se alto desempenho na classe 1–14, indicando que os modelos foram eficazes em evitar falsos positivos nessa faixa etária, possivelmente devido à presença de características morfológicas mais distintivas em indivíduos mais jovens.



**Figura 1. Especificidade por faixa etária**

Por outro lado, a classe 15–25 apresentou os menores valores de especificidade entre as arquiteturas analisadas, indicando maior dificuldade na distinção dessa faixa etária em relação às classes adjacentes, possivelmente devido à sobreposição de características morfológicas no período de transição entre a adolescência e a idade adulta. Ainda assim, a arquitetura *EfficientNetB5* manteve especificidade superior a 90%, destacando-se como um dos modelos mais equilibrados. Para a classe 26–99, observou-se desempenho

consistente entre as arquiteturas, com a maioria dos modelos apresentando especificidade acima de 90%, possivelmente associado à maior estabilidade das estruturas anatômicas em indivíduos adultos.

### 4.3. Resultados Encontrados para Matriz de Confusão

A Figura 2 apresenta as matrizes de confusão obtidas para as arquiteturas avaliadas na tarefa de classificação das três faixas etárias consideradas neste estudo (1–14, 15–25 e 26–99 anos). Cada subfigura corresponde ao desempenho de uma arquitetura específica, permitindo comparar visualmente o comportamento dos modelos em relação aos acertos e às confusões entre as classes etárias.

De forma geral, os modelos apresentaram melhor desempenho nas classes extremas (1–14 e 26–99 anos), com menores taxas de confusão entre essas categorias. Esse comportamento sugere que características morfológicas mais evidentes nessas faixas etárias facilitam a distinção entre os grupos. Em contrapartida, a classe intermediária (15–25 anos) concentrou a maior parte dos erros, indicando maior dificuldade das redes em diferenciar indivíduos em fase de transição entre a adolescência e a idade adulta.

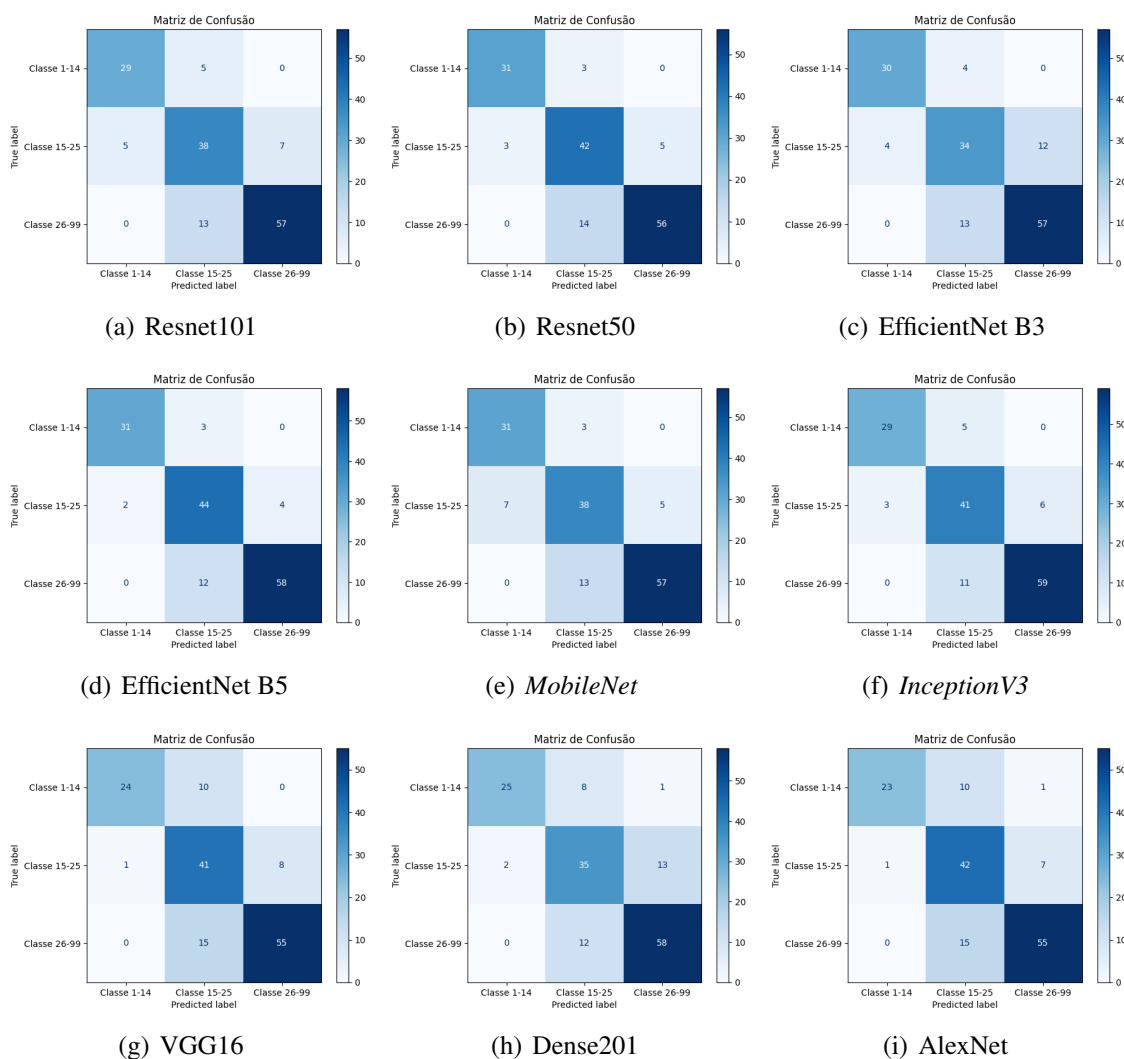
Entre os modelos analisados, a *ResNet50* apresentou desempenho superior à *ResNet101* (Figuras 2(a) e 2(b)), com maior acurácia nas classes 1–14 e 15–25, além de menor dispersão dos erros. Resultados semelhantes foram observados entre as arquiteturas da família *EfficientNet*, nas quais a *EfficientNetB5* superou a *EfficientNetB3* em todas as faixas etárias, evidenciando maior capacidade de extração de características morfológicas relevantes.

As arquiteturas *InceptionV3* e *MobileNetV2* apresentaram desempenho competitivo, com destaque para a *InceptionV3* na classificação da classe intermediária. Por outro lado, modelos mais antigos, como *AlexNet*, *VGG-16* e *DenseNet201*, apresentaram desempenho inferior, especialmente na faixa etária de 1–14 anos. De modo geral, observa-se que nenhuma arquitetura apresentou confusão significativa entre as classes extremas, reforçando a robustez dos modelos para distinguir indivíduos muito jovens de adultos. Entre os modelos avaliados, destacam-se a *EfficientNetB5* e a *ResNet50*, que apresentaram os resultados mais consistentes na tarefa de classificação etária.

Ainda, a Figura 2 apresenta as matrizes de confusão para a rede *EfficientNet B3* (Figura 2(c)) e a *EfficientNet B5* (Figura 2(d)). Ambas apresentaram elevada acurácia na faixa 1–14 anos, com 30 acertos para a *B3* e 31 para a *B5*, sendo os erros concentrados na classe 15–25. Essa concentração indica que a rede aprendeu as transições naturais entre faixas adjacentes. Dessa forma, a classe intermediária foi novamente a mais desafiadora, com 68% de acurácia na *B3* (34 de 50) e 88% na *B5* (44 de 50), evidenciando a maior robustez da *B5* na identificação de padrões associados ao final da adolescência e início da vida adulta. Na faixa 26–99, a *B3* obteve 57 acertos (81,4%) e a *B5* 58 acertos (82,9%), confirmando a vantagem geral da *B5*. De forma geral, a *EfficientNet-B5* superou a *B3* em todas as faixas etárias e métricas avaliadas. Sua arquitetura mais robusta permite a extração mais eficaz de características morfológicas sutis, refletindo em menor dispersão dos erros e maior estabilidade preditiva. Portanto, embora a *EfficientNet-B3* consiga realizar a tarefa de classificação, ainda há margem significativa para novos testes e aprimoramentos.

Além disso, a Figura 2 apresenta as matrizes de confusão para a rede *MobileNet*

(Figura 2(e)) e *InceptionV3* (Figura 2(f)). A *MobileNetV2* alcançou excelente desempenho na classe 1–14 anos (91,2%), com erros restritos à classe 15–25 e nenhuma confusão com a classe 26–99. Para a faixa intermediária, o modelo apresentou 76% de acerto (38 de 50), com 7 erros para a classe inferior e 5 para a superior. Na classe 26–99, obteve 81,4% de acurácia (57 de 70), com erros apenas para a faixa adjacente. Já a *InceptionV3* demonstrou 85,3% de acurácia na classe 1–14, 82% na 15–25 e 84,3% na 26–99. Sua maior na faixa 15–25 em relação a *MobileNetV2*, indica uma melhor separação dos padrões morfológicos intermediários.



**Figura 2. Matrizes de confusão das arquiteturas avaliadas**

Ainda, a Figura 2 apresenta as matrizes de confusão para a rede *VGG16* (Figura 2(g)), a *Dense201* (Figura 2(h)) e *AlexNet* (Figura 2(i)). De maneira geral, essas arquiteturas apresentaram desempenho inferior em relação às redes supracitadas, especialmente na classe 1–14 anos. A *AlexNet* obteve apenas 67,7% de acurácia nessa faixa, a *VGG-16* alcançou 70,6% e a *DenseNet201* 73,5%. Os erros estiveram majoritariamente associados à confusão com a faixa 15–25, indicando limitações na capacidade dessas redes em capturar padrões morfológicos associados a infância. Todavia, a *AlexNet* e a *VGG-*

16 apresentaram desempenho razoável na classe 15–25 (84% e 82%, respectivamente), e a *DenseNet201* destacou-se na faixa 26–99 com acurácia de 82,9%. Os resultados mostram que todas as redes tiveram melhor desempenho nas classes 1–14 e 26–99, com menor taxa de confusão entre os extremos etários.

A faixa 15–25, por outro lado, concentrou a maior parte dos erros, o que é justificável, dada sua posição como zona de transição morfológica. Esse padrão foi recorrente em todas as arquiteturas, com acurácias variando de 68% (*EfficientNet-B3*) a 88% (*EfficientNet-B5*), e sugere que ajustes na forma de agrupar as faixas ou a utilização de abordagens como aprendizado focalizado podem ser estratégias promissoras. Ressalta-se, que nenhum dos modelos cometeu erros críticos entre as classes 1–14 e 26–99, reforçando a robustez geral das redes para classificações mais extremas. Tal característica é particularmente relevante em aplicações clínicas ou forenses, nas quais equívocos graves entre crianças e adultos podem comprometer a validade das análises. Dessa forma, embora existam margens para aperfeiçoamentos, os modelos avaliados demonstram bom potencial na tarefa proposta, com destaque para a *EfficientNet-B5* e a *ResNet50*.

#### 4.4. Resultados para Validação Externa

A Tabela 2 apresenta os resultados das métricas de avaliação das três arquiteturas selecionadas para a etapa de validação externa: *EfficientNetB5*, *ResNet50* e *InceptionV3*. Entre os modelos avaliados, a *EfficientNetB5* obteve o melhor desempenho global, alcançando acurácia de 94,50%, precisão de 91,55%, *recall* de 94,14%, *F1-score* de 92,83% e especificidade de 96,92%, embora tenha apresentado o maior tempo de treinamento (384 minutos).

**Tabela 2. Resultados de desempenho dos modelos na avaliação externa**

Rede	Acurácia (%)	Precisão (%)	Recall (%)	F1-Score (%)	Especificidade (%)	Tempo (min)
EfficientNetB5	94,50	91,55	94,14	92,83	96,92	384
ResNet50	92,88	92,91	92,88	92,89	95,14	350
InceptionV3	91,91	91,86	91,91	91,87	94,16	300

A arquitetura *ResNet50* também demonstrou desempenho elevado, com acurácia de 92,88%, precisão de 92,91%, *recall* de 92,88%, *F1-score* de 92,89% e especificidade de 95,14%, com tempo de treinamento de 350 minutos. Já a *InceptionV3* apresentou resultados ligeiramente inferiores em termos de métricas, porém destacou-se pelo menor tempo de treinamento (300 minutos).

Ressalta-se que o treinamento dos modelos com um conjunto de dados numericamente maior, utilizado na etapa de validação externa, contribuiu para o aumento das métricas de desempenho. A maior variabilidade dos dados de treinamento provavelmente favoreceu a capacidade de generalização das redes neurais, permitindo melhor identificação das características relevantes para a classificação etária.

De forma geral, os resultados obtidos na validação externa confirmam a robustez dos modelos, evidenciando sua capacidade de adaptação e generalização frente a dados distintos daqueles utilizados no treinamento inicial. A análise conjunta das métricas indica superioridade da *EfficientNetB5* em desempenho global, enquanto a *InceptionV3* se destaca pela maior eficiência computacional. A Figura 3 apresenta as matrizes de confusão obtidas para cada uma das arquiteturas avaliadas na validação externa.

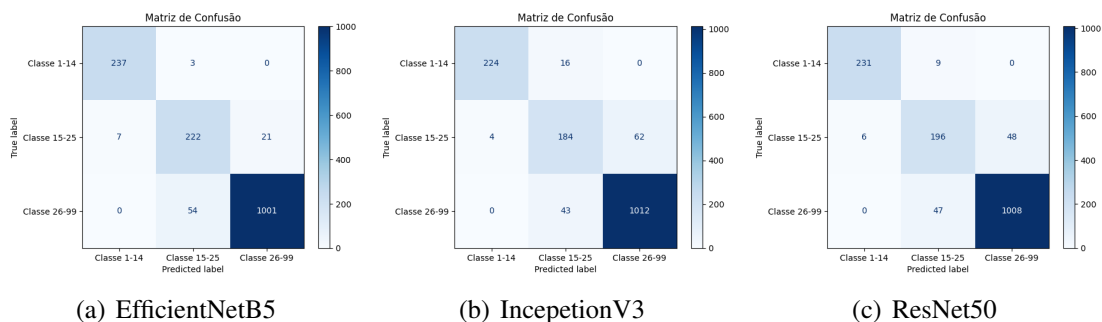


Figura 3. Matriz de confusão *EfficientNetB5*, *InceptionV3* e *ResNet50*

## 5. Implicações do Estudo

Os resultados obtidos neste estudo apresentam implicações relevantes para pesquisadores, desenvolvedores de ferramentas e profissionais que atuam com aprendizado profundo aplicado à saúde e à visão computacional.

*Implicações para Pesquisadores:* Este estudo evidencia a importância de protocolos experimentais bem definidos para a avaliação de arquiteturas de redes neurais profundas na classificação etária a partir de radiografias panorâmicas. A padronização das etapas de pré-processamento, aumento de dados e das métricas de avaliação contribui para a transparência metodológica e facilita a replicação e comparação com estudos futuros. A disponibilização do dataset também favorece práticas de ciência aberta, permitindo que outros pesquisadores avaliem a robustez dos modelos em diferentes contextos e populações. Além disso, os resultados apontam oportunidades para pesquisas futuras, incluindo a investigação de arquiteturas mais recentes, como *Vision Transformers*, técnicas de balanceamento de classes, aprendizado autossupervisionado e abordagens voltadas à estimativa contínua de idade.

*Implicações para Desenvolvedores de Ferramentas:* O *pipeline* experimental proposto, que abrange desde o pré-processamento até a validação dos modelos, pode servir como base para o desenvolvimento de ferramentas automatizadas de classificação de imagens médicas. A utilização de ambientes em nuvem e bibliotecas amplamente difundidas favorece a criação de soluções reprodutíveis e portáteis. Além disso, os resultados obtidos podem orientar a escolha de arquiteturas adequadas para aplicações práticas, considerando critérios como desempenho, custo computacional e tempo de treinamento. Essas arquiteturas podem ser incorporadas a sistemas de apoio à decisão voltados a profissionais da área forense ou odontológica.

## 6. Ameaças à Validade

Embora o estudo exploratório tenha sido cuidadosamente planejado e conduzido, algumas ameaças à validade podem ter influenciado os resultados. Essas ameaças foram analisadas considerando as quatro categorias: validade de construção, validade interna, validade externa e validade de conclusão.

*Validade de Construção:* Uma possível ameaça está relacionada à seleção das arquiteturas de redes neurais avaliadas. Como não foi viável analisar todas as arquiteturas existentes na literatura, a escolha foi fundamentada em uma revisão sistemática, garantindo alinha-

mento com o estado da arte e relevância para o problema investigado. Ainda, o protocolo experimental foi definido com base em estudos anteriores e na literatura especializada.

*Validade Interna:* A validade interna pode ser afetada por fatores não controlados durante os experimentos, como variações de hardware ou software. Para reduzir esse risco, todos os experimentos foram realizados no ambiente *Google Colab*, que oferece infraestrutura padronizada e suporte a GPUs, contribuindo para maior estabilidade e reprodutibilidade dos resultados.

*Validade Externa:* A generalização dos resultados para outros contextos representa uma limitação potencial. Embora dois conjuntos de dados distintos tenham sido utilizados, não se pode garantir desempenho equivalente em bases provenientes de outras populações ou equipamentos. A disponibilização pública dos dados e dos artefatos do estudo permite que novos trabalhos realizem validações adicionais.

*Validade de Conclusão:* Para reduzir riscos de interpretações equivocadas, os resultados foram apresentados por meio de diferentes métricas amplamente utilizadas na literatura, incluindo acurácia, precisão, *recall*, especificidade e *F1-score*. Além disso, o pré-processamento padronizado das imagens e a disponibilização dos códigos e modelos contribuem para a transparência e replicabilidade do estudo.

## 7. Conclusão

Este trabalho apresentou um estudo exploratório para classificação de faixas etárias a partir de radiografias panorâmicas odontológicas utilizando nove arquiteturas de redes neurais convolucionais. Os experimentos foram conduzidos com protocolos padronizados de pré-processamento, aumento de dados e *transfer learning*. Os resultados indicaram desempenho satisfatório para todas as arquiteturas avaliadas, com destaque para a *EfficientNetB5*, seguida por *ResNet50* e *InceptionV3*. Observou-se maior dificuldade na classificação da faixa etária intermediária (15–25 anos). De modo geral, os resultados reforçam o potencial das redes neurais profundas para apoiar a estimativa de idade a partir de radiografias panorâmicas e orientar pesquisas futuras na área.

## Referências

- Ataş, İ., Özdemir, C., Ataş, M., and Doğan, Y. (2022). Forensic dental age estimation using modified deep learning neural network. *Balkan Journal of Electrical and Computer Engineering*, 11(4):298–305.
- Baylis, S., Dipnall, J. F., and Basset, R. (2026). Dental age estimation: a scoping review comparing the manual application of the demirjian method and artificial intelligence modalities. *International journal of legal medicine*.
- de Oliveira, M. F., Ferreira, F. J., and de Paula Júnior, I. C. (2025). Estimativa de idade por meio redes neurais profundas: Uma revisão sistemática da literatura.
- Gracea, R. S., Winderickx, N., Vanheers, M., Hendrickx, J., Preda, F., Shujaat, S., de Llano-Pérula, M. C., and Jacobs, R. (2025). Artificial intelligence for orthodontic diagnosis and treatment planning: A scoping review. *Journal of dentistry*, 152:105442.
- Guo, Y.-C., Han, M., Chi, Y., Long, H., Zhang, D., Yang, J., Yang, Y., Chen, T., and Du, S. (2021). Accurate age classification using manual method and deep convolutional neural network based on orthopantomogram images. *International journal of legal medicine*, 135:1589–1597.

- Khanagar, S. B., Albalawi, F., Alshehri, A., Awawdeh, M., Iyer, K., Alsomaie, B., Aldhebaib, A., Singh, O. G., and Alfadley, A. (2024). Performance of artificial intelligence models designed for automated estimation of age using dento-maxillofacial radiographs—a systematic review. *Diagnostics*, 14(11):1079.
- Koch, R. M., Mentzel, H.-J., and Heinrich, A. (2025). Deep learning for forensic age estimation using orthopantomograms in children, adolescents, and young adults. *European Radiology*, 35(7):4191–4202.
- Moreira, A. G., dos Santos, S. A., de Oliveira, M. F., de Paula Júnior, I. C., and de Assis, D. F. (2024). Classificação de tumores cerebrais em imagens de ressonância magnética. In *Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*, pages 424–435. SBC.
- Mualla, N., Houssein, E. H., and Hassan, M. (2020). Dental age estimation based on x-ray images. *Computers, Materials & Continua*, 62(2).
- Nino-Sandoval, T. C., Doria-Martinez, A. M., Escobar, R. A. V., Sánchez, E. L., Rojas, I. B., Álvarez, L. C. V., Mc Cann, D. S. F., and Tamara-Patino, L. M. (2024). Efficacy of the methods of age determination using artificial intelligence in panoramic radiographs—a systematic review. *International Journal of Legal Medicine*, 138(4):1459–1496.
- Nushi, V., Santos, R., Brkić, H., Oliveira, A., Francisco, A., and Palmela Pereira, C. (2025). Pre-trained vgg16 model for forensic dental age estimation. *Egyptian journal of forensic sciences*, 15(1):1–12.
- Oliveira, W., Albuquerque Santos, M., Burgardt, C. A. P., Anjos Pontual, M. L., and Zanchettin, C. (2024). Estimation of human age using machine learning on panoramic radiographs for brazilian patients. *Scientific Reports*, 14(1):19689.
- Ong, S.-H., Kim, H., Song, J.-S., Shin, T. J., Hyun, H.-K., Jang, K.-T., and Kim, Y.-J. (2024). Fully automated deep learning approach to dental development assessment in panoramic radiographs. *BMC Oral Health*, 24(1):426.
- Ozlu Ucan, G., Gwassi, O. A. H., Apaydin, B. K., and Ucan, B. (2025). Automated age estimation from opg images and patient records using deep feature extraction and modified genetic–random forest. *Diagnostics*, 15(3):314.
- Shi, Y., Ye, Z., Guo, J., Tang, Y., Dong, W., Dai, J., Miao, Y., and You, M. (2024). Deep learning methods for fully automated dental age estimation on orthopantomograms. *Clinical Oral Investigations*, 28(3):198.
- Singh, K., Kumar, P., and Urs, A. B. (2026). Digital methods of dental age. *Dental Age Assessment: A Global Perspective*, page 111.
- Sivri, M. B., Taheri, S., Ercan, R. G. K., Yağcı, Ü., and Golrizkhatami, Z. (2024). Dental age estimation: A comparative study of convolutional neural network and demirjian’s method. *Journal of Forensic and Legal Medicine*, 103:102679.
- Tokinaga, R., Mine, Y., Yoshimi, Y., Okazaki, S., Ito, S., Takeda, S., Ogawa, S., Peng, T.-Y., Kakimoto, N., Tanimoto, K., et al. (2025). Age estimation from lateral cephalograms using deep learning: A pilot study from early childhood to older adults. *Journal of Clinical Medicine*, 14(19):7084.
- Wang, J., Dou, J., Han, J., Li, G., and Tao, J. (2023). A population-based study to assess two convolutional neural networks for dental age estimation. *BMC Oral Health*, 23(1):109.