



UNIVERSIDADE FEDERAL DO MARANHÃO
Programa de Pós-Graduação em Ciência da Computação

Alison Corrêa Mendes

***Deteccção de Dentes e Classificação de Patologias Dentárias em
Imagens de Radiografias Panorâmicas Utilizando Aprendizado
Profundo***

**São Luís
2025**

UNIVERSIDADE FEDERAL DO MARANHÃO – UFMA
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

ALISON CORRÊA MENDES

**DETECÇÃO DE DENTES E CLASSIFICAÇÃO DE PATOLOGIAS
DENTÁRIAS EM IMAGENS DE RADIOGRAFIAS PANORÂMICAS
UTILIZANDO APRENDIZADO PROFUNDO**

SÃO LUÍS – MA

2025

ALISON CORRÊA MENDES

**DETECÇÃO DE DENTES E CLASSIFICAÇÃO DE PATOLOGIAS
DENTÁRIAS EM IMAGENS DE RADIOGRAFIAS PANORÂMICAS
UTILIZANDO APRENDIZADO PROFUNDO**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da UFMA como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Orientador: Prof. Dr. Darlan Bruno Pontes Quintanilha

Coorientador: Prof. Dr. Anselmo Cardoso de Paiva

SÃO LUÍS – MA

2025

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Diretoria Integrada de Bibliotecas/UFMA

Mendes, Alison Corrêa.

Detecção de Dentes e Classificação de Patologias Dentárias em Imagens de Radiografias Panorâmicas Utilizando Aprendizado Profundo / Alison Corrêa Mendes. - 2025.

84 p.

Coorientador(a) 1: Anselmo Cardoso de Paiva.

Orientador(a): Darlan Bruno Pontes Quintanilha.

Dissertação (Mestrado) - Programa de Pós-graduação em Ciência da Computação/ccet, Universidade Federal do Maranhão, São Luís - Ma, 2025.

1. Radiografia Panorâmica. 2. Aprendizado Profundo. 3. Detecção de Dentes. 4. Classificação de Anomalias Dentárias. 5. Redes Neurais Convolucionais. I. Paiva, Anselmo Cardoso de. II. Quintanilha, Darlan Bruno Pontes.

ALISON CORRÊA MENDES

**DETECÇÃO DE DENTES E CLASSIFICAÇÃO DE PATOLOGIAS
DENTÁRIAS EM IMAGENS DE RADIOGRAFIAS PANORÂMICAS
UTILIZANDO APRENDIZADO PROFUNDO**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da UFMA como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Dissertação aprovada em ____ de _____ de _____.

**Prof. Dr. Darlan Bruno Pontes
Quintanilha**
Orientador

Prof. Dr. Anselmo Cardoso de Paiva
Coorientador

**Prof. Dr. João Dallyson Sousa de
Almeida**
Examinador Interno 1

**Prof. Dra. Luciana Salles Branco de
Almeida**
Examinador Interno 2

Prof. Dra. Luana Batista da Cruz
Examinador Externo

SÃO LUÍS – MA
2025

À minha família e aos verdadeiros amigos

AGRADECIMENTOS

A Deus, primeiramente, por sempre me guiar e sustentar ao longo desta jornada. Até aqui me ajudou o Senhor

À minha família, pelo amor incondicional, carinho e dedicação, que foram fundamentais para que eu chegasse até aqui.

Ao meu orientador, Darlan Bruno Pontes Quintanilha, pelo apoio, paciência e orientação ao longo do mestrado.

Ao meu coorientador, Anselmo Cardoso de Paiva, por seu apoio fundamental nos meus primeiros passos na pesquisa, pela orientação e pelo suporte constante.

Ao professor Geraldo Braz, pelo suporte, pelos valiosos conselhos e pelo conhecimento compartilhado, que tanto contribuíram para minha formação.

Aos professores do Núcleo de Computação Aplicada, incluindo os já citados, por todo o ensinamento, dedicação e incentivo ao longo do mestrado.

Ao Alexandre Pessoa, pelo auxílio na escrita dos artigos e desta dissertação.

Ao Wesley Kelson, pelo apoio no desenvolvimento e na elaboração das figuras desta dissertação.

À CAPES, pelo apoio financeiro durante o mestrado.

À Kerllyane, pelo suporte, apoio e companheirismo em todos os momentos, tornando essa caminhada mais leve.

Aos amigos, incluindo os já citados, pela amizade, pelo companheirismo, pelos ensinamentos e por todos os momentos de descontração e apoio ao longo dessa trajetória.

E, por fim, a todos que, de alguma forma, contribuíram para que eu chegasse até aqui. Meu sincero agradecimento.

*“E não nos cansemos de fazer o bem,
pois no tempo próprio colheremos, se não desanimarmos.”*

Gálatas 6:9

RESUMO

A radiografia panorâmica desempenha um papel fundamental no diagnóstico e planejamento odontológico. No entanto, sua interpretação está sujeita a erros devido a fatores como fadiga, estresse e experiência do profissional. Este trabalho propõe um método baseado em aprendizado profundo para a detecção automática de dentes e classificação de anomalias dentárias em imagens de radiografias panorâmicas. Diferentes arquiteturas de redes neurais convolucionais foram avaliadas para ambas as tarefas. Na detecção de dentes, a arquitetura YOLOv8x apresentou os melhores resultados, com mAP50 de 0,945 e mAP50-95 de 0,578. Na classificação de patologias, a DenseNet121 obteve um F1 Score de 0,854, demonstrando uma boa relação entre identificação correta e redução de falsos positivos. Os resultados obtidos sugerem que o método proposto, que divide o processo em duas etapas utilizando a YOLOv8x para a detecção de dentes e a DenseNet121 para a classificação de patologias, apresenta desempenho superior em relação à abordagem em que a própria YOLOv8x é utilizada para realizar as duas tarefas. Enquanto o método proposto alcançou um F1 Score de 0,6014, a abordagem única baseada apenas na YOLOv8x obteve um F1 Score de 0,433. O método foi treinado e avaliado utilizando o conjunto de dados DENTEX, composto por radiografias panorâmicas anotadas por especialistas, incluindo informações sobre a numeração dos dentes e a presença de patologias como dentes impactados, cáries e lesões periapicais. Esses resultados reforçam o potencial da abordagem proposta como ferramenta de apoio ao diagnóstico odontológico, contribuindo para a redução de erros e para a melhoria da eficiência na análise de radiografias panorâmicas.

Palavras-chave: Radiografia Panorâmica; Aprendizado Profundo; Detecção de Dentes; Classificação de Anomalias Dentárias; Redes Neurais Convolucionais.

ABSTRACT

Panoramic radiography plays a fundamental role in dental diagnosis and treatment planning. However, its interpretation is subject to errors due to factors such as fatigue, stress, and the experience of the professional. This work proposes a deep learning-based method for the automatic detection of teeth and the classification of dental anomalies in panoramic radiographic images. Different convolutional neural network architectures were evaluated for both tasks. In tooth detection, the YOLOv8x architecture achieved the best results, with a mAP50 of 0.945 and a mAP50-95 of 0.578. In pathology classification, DenseNet121 achieved an F1 Score of 0.854, demonstrating a good balance between correct identification and the reduction of false positives. The results suggest that the proposed method, which separates the process into two stages using YOLOv8x for tooth detection and DenseNet121 for pathology classification, outperforms the approach where YOLOv8x alone is used to perform both tasks. While the proposed method achieved an F1 Score of 0.6014, the single-stage approach based solely on YOLOv8x obtained an F1 Score of 0.433. The method was trained and evaluated using the DENTEX dataset, which consists of panoramic radiographs annotated by specialists, including information on tooth numbering and the presence of pathologies such as impacted teeth, caries, and periapical lesions. These results reinforce the potential of the proposed approach as a decision-support tool in dental diagnostics, contributing to error reduction and improving the efficiency of panoramic radiograph analysis.

Keywords: Panoramic Radiography; Deep Learning; Tooth Detection; Classification of Dental Anomalies; Convolutional Neural Networks.

LISTA DE ILUSTRAÇÕES

Figura 1 – Esquema representativo de um sistema de imagem por raios X.	24
Figura 2 – Exemplo de diferentes tipos de radiografias intraorais e extraorais. . .	25
Figura 3 – Exemplo de imagem de RP.	27
Figura 4 – Esquema ilustrativo do processo de obtenção de uma RP.	27
Figura 5 – Exemplo de alterações dentárias. (a) Lesão periapical; (b) Dente im- pactado; (c) Dente com cárie; (d) Cárie profunda.	28
Figura 6 – Classificação para estágios mais profundos de cárie. (a) Cárie profunda atingindo o quarto pulpar com uma zona de dentina separando a lesão da polpa (b) e extremamente profunda penetrando toda a espessura da dentina.	30
Figura 7 – Notação FDI World Dental Federation.	31
Figura 8 – Exemplo de operação de convolução com um <i>kernel</i> 3x3 aplicado a uma imagem 5x5 com deslocamento de 2 pixels por unidade.	32
Figura 9 – Exemplo do antes e depois de uma imagem após equalização de histo- grama.	34
Figura 10 – Modelo de neurônio McCulloch-Pitts.	36
Figura 11 – Exemplo de ANN.	36
Figura 12 – Uma arquitetura CNN com cinco camadas.	38
Figura 13 – Visualização do comportamento da FL para determinados valores de γ	41
Figura 14 – A YOLOv1, com 24 camadas convolucionais e 2 totalmente conectadas, usando filtros 1×1 para reduzir a dimensionalidade.	42
Figura 15 – Arquitetura YOLOv8.	43
Figura 16 – Exemplo de um Bloco Denso com 5 camadas.	45
Figura 17 – Arquitetura da DenseNet-121.	46
Figura 18 – Fluxo principal do método proposto.	53
Figura 19 – Exemplo dos diferentes tipos de anotação disponíveis do conjunto de dados DENTEX.	54
Figura 20 – Distribuição dos rótulos para a categoria b da base de dados DENTEX.	54
Figura 21 – Distribuição dos rótulos para a categoria c da base de dados DENTEX.	55
Figura 22 – Formato de anotações para cada rótulo.	56

Figura 23 – Exemplo do antes e depois do pré-processamento de uma imagem do conjunto de dados.	57
Figura 24 – Fluxo completo do processo de detecção e numeração de dentes. . . .	58
Figura 25 – Frequência de conflito de <i>bounding boxes</i> no novo conjunto de dados.	59
Figura 26 – Fluxo completo do processo de treinamento e avaliação para a classificação de anomalias dentárias.	60
Figura 27 – Validação Cruzada <i>k</i> -fold.	62
Figura 28 – Exemplo de detecções corretas pelo modelo YOLOv8x.	65
Figura 29 – Exemplo de falso negativo em que o modelo YOLOv8x perdeu detecções.	65
Figura 30 – Histórico das rodadas (<i>trials</i>) da otimização dos hiperparâmetros dos modelos.	67
Figura 31 – Exemplo de classificação correta e incorreta de Lesão Periapical. (A) Mapa de calor gerado pela DenseNet121, destacando a área patológica; (B) Mapa de calor gerado pela InceptionV3, com ativação em região fora do dente.	70
Figura 32 – Exemplo de classificação correta de dente Impactado. (A) Mapa de calor gerado pela DenseNet121, destacando a região do dente e áreas do osso ao redor com alta ativação; (B) Mapa de calor gerado pela InceptionV3, com regiões de ativação concentradas fora do dente. . .	71
Figura 33 – Exemplo de classificação correta de dente com cárie. (A) Mapa de calor gerado pela DenseNet121, destacando regiões no dente com alta ativação; (B) Mapa de calor gerado pela Inception V3, com regiões de ativação mais difusas e menos concentradas em pontos específicos. . .	72
Figura 34 – Exemplo de falso positivo para dente com Lesão Periapical.	72
Figura 35 – Exemplo de falso positivo para dente com Cárie.	73
Figura 36 – Exemplo de falso positivo para dente Impactado.	74

LISTA DE TABELAS

Tabela 1	– Resumo de trabalhos que fizeram uso de IA em imagens odontológicas.	23
Tabela 2	– Distribuição das imagens por classe no conjunto de dados.	59
Tabela 3	– Distribuição final das imagens por classe no conjunto de dados após os ajustes.	60
Tabela 4	– Os valores são apresentados como médias seguidas de desvios-padrão entre parênteses, obtidos por validação cruzada <i>k-fold</i>	63
Tabela 5	– Os valores são apresentados como médias seguidas de desvios-padrão entre parênteses, obtidos de um conjunto de teste separado.	64
Tabela 6	– Valores dos hiperparâmetros ótimos para arquitetura avaliada.	66
Tabela 7	– Métricas de avaliação obtidas na validação cruzada utilizando 5 <i>folds</i> .	68
Tabela 8	– Métricas de desempenho obtidas no conjunto de teste separado.	69
Tabela 9	– Comparação entre a YOLOv8x realizando localização e classificação diretamente e a abordagem YOLOv8x + DenseNet121.	69
Tabela 10	– Artigos que possuem relação com o método proposto.	76
Tabela 11	– Artigos que não possuem relação com o método proposto.	76

LISTA DE ABREVIATURAS E SIGLAS

ANN	<i>Artificial Neural Network</i>
AUC	<i>Area Under the Curve</i>
CBCT	<i>Cone Beam computer Tomography</i>
CE	<i>Cross Entropy Loss</i>
CNN	<i>Convolutional Neural Network</i>
EI	<i>Expected Improvement</i>
FL	<i>Focal Loss</i>
HITL	<i>Human-in-the-Loop</i>
IA	Inteligência Artificial
IoU	<i>Intersection over Union</i>
mAP	<i>Mean Average Precision</i>
O2PR	<i>OdontoAI Open Panoramic Radiographs</i>
OPG	Ortopantomografia
ROC	<i>Receiver Operating Characteristic</i>
SMBO	<i>Sequential Model-Based Global Optimization</i>
TPE	<i>Tree-structured Parzen Estimator</i>
YOLO	<i>You Only Look Once</i>

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Objetivos	17
1.1.1	Objetivos Específicos	17
1.2	Contribuições	17
1.3	Organização do Trabalho	18
2	TRABALHOS RELACIONADOS	20
3	FUNDAMENTAÇÃO TEÓRICA	24
3.1	Imagens Radiográficas na Odontologia	24
3.1.1	Radiografia Panorâmica	26
3.1.2	Anomalias Dentárias	27
3.1.3	Numeração de Dentes	30
3.2	Processamento de Imagens Digitais	30
3.2.1	Convolução	32
3.2.2	Equalização de Histograma	33
3.3	Aprendizado de Máquina	34
3.3.1	Redes Neurais Artificiais	35
3.3.2	Redes Neurais Convolucionais	36
3.4	Função de Custo	39
3.4.1	Focal Loss	39
3.5	<i>You Only Look Once</i> (YOLO)	41
3.5.1	YOLOv8	42
3.6	DenseNet	44
3.6.1	DenseNet-121	44
3.7	Otimização de Hiperparâmetros	46
3.7.1	Otimização Bayesiana	47
3.7.2	Sequential Model-Based Global Optimization	48
3.7.3	Tree-structured Parzen Estimator	48
3.8	Métricas de Avaliação	49
3.8.1	Precisão, Recall e F1 Score	50
3.8.2	Média da Precisão Média (mAP)	51

3.9	Considerações Finais	51
4	MATERIAIS E MÉTODO	53
4.1	Aquisição da Base	53
4.2	Detecção e Numeração de Dentes	55
4.2.1	Preparação e Pré-processamento dos Dados	56
4.3	Classificação de Patologias	58
4.3.1	Preparação e Pré-processamento dos Dados	58
4.3.2	Otimização de Hiperparâmetros	60
5	RESULTADOS E DISCUSSÃO	62
5.1	Validação Cruzada k -fold	62
5.2	Resultados da Detecção de Dentes	63
5.2.1	Resultados Quantitativos	63
5.2.2	Estudos de Caso	64
5.3	Resultados da Classificação de Patologias	66
5.3.1	Otimização de Hiperparâmetros	66
5.3.2	Resultados Quantitativos	68
5.3.3	Estudos de Caso	69
5.4	Considerações Finais	74
6	CONCLUSÃO	75
6.1	Trabalhos Futuros	75
6.2	Produções científicas	76
	REFERÊNCIAS	77

1 INTRODUÇÃO

Imagens radiográficas são amplamente utilizadas na odontologia para fins de diagnóstico e planejamento de tratamento (RONDON; PEREIRA; NASCIMENTO, 2014). Devido a vantagens como baixa dose de radiação, rápida aquisição da imagem e mínimo desconforto para o paciente, a radiografia panorâmica (RP) possui grande importância na prática odontológica (BILGIR et al., 2021). Além disso, a RP também se destaca por oferecer uma visão abrangente tanto dos dentes maxilares quanto mandibulares, incluindo suas estruturas de suporte. Contudo, a RP possui sobreposições e distorções que podem, juntamente com as estruturas anatômicas externas, dificultar a interpretação dessas imagens (PERSCHBACHER, 2012).

Exames clínicos e radiológicos desempenham um papel fundamental no diagnóstico e no planejamento de tratamentos odontológicos. Entre eles, as radiografias dentárias são ferramentas indispensáveis para identificar e tratar condições como cáries, doenças periodontais e outras patologias orais (AFFAIRS et al., 2006). No entanto, a interpretação dessas imagens pode ser desafiadora, pois é uma tarefa que exige tempo e está sujeita a erros devido a fatores como fadiga, estresse e experiência do profissional (YILMAZ et al., 2024).

As radiografias dentárias apresentam diferentes níveis de exposição à radiação. Uma série de 18 radiografias intraorais, por exemplo, equivale a até 21 dias de exposição à radiação de fundo, enquanto a RP corresponde a um período entre 1,7 e 3 dias. (LUDLOW; DAVIES-LUDLOW; WHITE, 2008). Por essa razão, a RP se destaca como uma alternativa mais vantajosa em termos de dose de radiação quando comparada às múltiplas radiografias intraorais (CHOI, 2011).

Fatores como fadiga, estresse e experiência do dentista estão entre as principais causas de diagnósticos incorretos e extrações desnecessárias. Um levantamento realizado em Jeddah, Arábia Saudita, por Jan et al. (2019), analisou 186 questionários respondidos por dentistas de quatro clínicas odontológicas universitárias (taxa de resposta de 37% de um total de 486 distribuídos). O estudo identificou uma prevalência de extrações equivocadas de 21,1%. As principais causas relatadas foram a comunicação inadequada (31,6%), encaminhamentos imprecisos (28,9%) e a exaustão de dentistas sobrecarregados (28,9%). Esses erros podem gerar impactos negativos significativos para os pacientes,

como dor, perda funcional, necessidade de tratamentos adicionais, custos financeiros e consequências psicológicas associadas à perda de dentes saudáveis.

Um outro estudo, realizado na Nigéria, estimou uma prevalência de extrações erradas de 13% (ADEYEMO et al., 2011). Já em um estudo conduzido em Israel, foi observado que a maior parte dos casos envolveu a extração de um dente adjacente ao alvo (67%), enquanto 15% dos casos ocorreram por confusão entre quadrantes, 13% por confusão entre dentes decíduos e permanentes, e 14% devido à marcação incorreta do dente a ser extraído (TAICHER, 2010).

Considerando os desafios relacionados à interpretação de radiografias, a inteligência artificial (IA) surge como uma abordagem promissora. Métodos baseados em aprendizado de máquina, especialmente o aprendizado profundo, têm demonstrado eficácia na resolução de uma ampla gama de problemas em diversos setores, incluindo a medicina e a odontologia (HWANG et al., 2019). Entre as abordagens de IA mais relevantes, destacam-se a detecção de objetos e a classificação de imagens. Detecção de objetos envolve a identificação de instâncias de objetos pertencentes a uma categoria específica em uma imagem, fornecendo não apenas sua presença, mas também sua localização espacial precisa e extensão de cada instância encontrada (LIU et al., 2020). Por sua vez, a tarefa de classificação de imagens consiste em atribuir um rótulo ou categoria a uma imagem com base em seu conteúdo visual (FANG et al., 2020).

Estudos recentes têm empregado o uso de IA na radiologia oral e maxilofacial para diferentes tarefas, como classificação, detecção e segmentação (HEO et al., 2021). Além de aliviar a carga dos profissionais, esses métodos podem aumentar a confiabilidade e a eficiência do diagnóstico, contribuindo para melhores desfechos clínicos e redução de procedimentos desnecessários.

Este trabalho propõe um método para detecção de dentes e classificação de anomalias dentárias por meio de imagens de radiografias panorâmicas. A partir de técnicas avançadas de processamento de imagens e aprendizado profundo, o trabalho visa identificar de forma automática e robusta estruturas dentárias e detectar sinais indicativos de condições patológicas, como cáries, lesões periapicais e dentes impactados. Dessa forma, pretende-se contribuir para o aprimoramento dos processos de diagnóstico assistido por computador na odontologia, promovendo um método que auxilie profissionais da saúde a identificar problemas bucais de maneira rápida, confiável e eficiente.

1.1 Objetivos

O objetivo geral deste trabalho é desenvolver um método baseado em aprendizado profundo capaz de: localizar dentes em radiografias panorâmicas; numerar os dentes localizados; e classificar a presença ou ausência de anomalias associadas a cada dente. O método proposto visa auxiliar profissionais da odontologia no diagnóstico e no planejamento de tratamentos.

1.1.1 Objetivos Específicos

Para que o objetivo geral deste trabalho seja possível, os seguintes objetivos específicos precisam ser alcançados:

- Desenvolver um método baseado em aprendizado profundo para detecção automática de dentes em radiografias panorâmicas, compreendendo a localização e a identificação dos dentes;
- Construir um conjunto de dados anotado a partir de caixas delimitadoras extraídas de radiografias panorâmicas, contendo dentes individuais rotulados para quatro anomalias dentárias;
- Avaliar diferentes arquiteturas de redes neurais convolucionais (*convolutional neural network*, CNNs) para a tarefa de classificação de anomalias dentárias;
- Realizar a otimização de hiperparâmetros para melhorar o desempenho do modelo de classificação.

1.2 Contribuições

As contribuições deste trabalho concentram-se na aplicação e avaliação de um fluxo em duas etapas para a análise automatizada de radiografias panorâmicas, envolvendo detecção de dentes e posterior classificação de anomalias dentárias. Este trabalho se diferencia pela comparação entre o desempenho de modelos que executam ambas as tarefas de forma conjunta e modelos especializados em cada etapa separadamente. As principais contribuições incluem:

- Avaliação comparativa entre diferentes arquiteturas de redes neurais convolucionais (CNNs) para a classificação de anomalias dentárias, incluindo cáries, lesões periapicais e dentes impactados;
- Análise do impacto da separação entre as etapas de detecção e classificação, demonstrando as vantagens e limitações da abordagem dividida em relação à estratégia unificada;
- Construção de um conjunto de dados anotado com *bounding boxes* de dentes rotulados com patologias, utilizada no treinamento e validação dos modelos;
- Integração dos modelos de detecção e classificação em um pipeline unificado, com potencial de aplicação prática no apoio ao diagnóstico odontológico.

1.3 Organização do Trabalho

Esta dissertação é composta por capítulos, organizados da seguinte maneira: Trabalhos Relacionados (Capítulo 2), Fundamentação Teórica (Capítulo 3), Materiais e Métodos (Capítulo 4), Resultados e Discussão (Capítulo 5) e Conclusão (Capítulo 6).

O Capítulo 2 apresenta pesquisas que utilizaram IA aplicada a imagens médicas odontológicas, destacando os avanços recentes na área. Também permite identificar as principais lacunas existentes na literatura, motivando o desenvolvimento do presente estudo.

O Capítulo 3 descreve os conceitos essenciais para a compreensão do método proposto, abordando temas como processamento de imagens, aprendizado de máquina, funções de custo, arquiteturas de redes neurais e otimização de hiperparâmetros. Além disso, são detalhadas as principais anomalias dentárias consideradas neste trabalho, permitindo uma melhor compreensão dos desafios abordados.

O Capítulo 4 apresenta, de forma detalhada, todas as etapas do método proposto para detecção e numeração de dentes, além da classificação de patologias em radiografias panorâmicas. Também são descritos o conjunto de dados utilizado, pré-processamentos aplicados, configurações dos modelos empregados e as métricas de avaliação escolhidas para validar os resultados.

O Capítulo 5 descreve os resultados obtidos para cada etapa do método, incluindo a análise quantitativa da detecção de dentes e da classificação das patologias, bem como a

otimização de hiperparâmetros. Além da análise das métricas aplicadas, são apresentados estudos de caso que fornecem exemplos visuais dos acertos e erros do modelo.

Por fim, o Capítulo 6 reúne as considerações finais do trabalho, discute as limitações da abordagem e aponta sugestões para pesquisas futuras. Também são apresentadas as produções científicas desenvolvidas durante o período.

2 TRABALHOS RELACIONADOS

Nos últimos anos, a pesquisa sobre o uso da IA na odontologia tem avançado consideravelmente, especialmente em aplicações de visão computacional. Em relação a imagens radiográficas, Silva, Oliveira e Pithon (2018) observaram que a maioria dos estudos anteriores não priorizava o uso de RPs.

Silva, Oliveira e Pithon (2018) apresentaram uma revisão abrangente acerca dos métodos de segmentação de imagens de raio-X odontológicas, com o objetivo de identificar tendências, avaliar conjuntos de dados existentes e determinar o método mais adequado segmentar os dentes em relação às estruturas não dentárias presentes na RP. Um novo conjunto de dados anotados contendo 1.500 imagens de RPs foi apresentado, permitindo uma avaliação de algoritmos de segmentação. Os resultados indicaram que os métodos existentes tiveram dificuldade em segmentar os dentes com precisão devido aos desafios impostos pelas estruturas ósseas, levando a altas taxas de falsos positivos. O algoritmo com melhor desempenho em termos de F1 Score atingiu 61,83%, enquanto o de pior desempenho alcançou apenas 17,51%, evidenciando as limitações dessas técnicas.

Panetta et al. (2021) apresentaram um conjunto de dados multimodal composto por 1.000 RPs anotadas para segmentação dentária e detecção de anomalias, com o objetivo de aprimorar as aplicações de IA no diagnóstico odontológico. O conjunto de dados incorpora a expertise de radiologistas por meio de protocolos de rastreamento ocular e think-aloud, uma técnica em que especialistas verbalizam seus pensamentos durante a análise, permitindo capturar seu raciocínio e integrá-lo ao desenvolvimento de sistemas de IA. O método inclui uma análise abrangente do desempenho de algoritmos para aprimoramento e segmentação de imagens, demonstrando o potencial do conjunto de dados em melhorar a precisão diagnóstica.

Nader et al. (2022) concentraram-se na segmentação automática de dentes em RPs usando redes neurais profundas, especificamente uma arquitetura U-Net modificada. O método consistiu em um processo de duas etapas: treinar um modelo Mask R-CNN para localizar dentes e extrair caixas delimitadoras e, em seguida, treinar uma rede U-Net modificada usando essas caixas delimitadoras previstas. Os resultados demonstraram a eficácia da abordagem na segmentação precisa das posições dos dentes, com potenciais aplicações no diagnóstico e tratamento odontológico, alcançando um coeficiente de Dice

médio de 90% para a U-Net modificada.

Park et al. (2022) desenvolveram um modelo baseado em aprendizado profundo para detectar regiões onde há ausência de dentes em imagens de raio-X panorâmicas, com o objetivo de aprimorar o planejamento de implantes dentários. O método envolveu a segmentação de instâncias dentárias para melhorar o desempenho de detecção de regiões dentárias perdidas. Os resultados indicaram um alto desempenho na segmentação de instâncias dentárias, com uma média da precisão média (*mean average precision*, mAP) de 92,14% considerando um limiar de *intersection over union* (IoU) de 0,5. No entanto, o desempenho na detecção de regiões dentárias ausentes foi inferior (59,09%), possivelmente devido a variações nas estruturas orais. Dados sintéticos foram utilizados para melhorar o desempenho do modelo, mas foram observadas limitações na diversidade.

Karaoglu et al. (2023) propuseram uma abordagem híbrida para a segmentação e numeração de dentes em RPs, combinando o uso do algoritmo Mask R-CNN com um método heurístico. Com um conjunto de dados composto por 2.702 imagens divididas em treino, validação e teste, os autores realizaram validação cruzada com 5 *folds*. O método heurístico, desenvolvido para complementar o desempenho do Mask R-CNN, melhorou significativamente os resultados em métricas como precisão (96,08%) e F1 Score (95,87%). Este trabalho destaca a eficácia de métodos híbridos na redução da carga de trabalho de radiologistas e no aprimoramento de diagnósticos assistidos por IA.

Silva et al. (2023) abordaram o desenvolvimento de um grande conjunto de dados rotulado, o OdontoAI Open Panoramic Radiographs (O2PR), voltado para pesquisas em RPs odontológicas. Utilizando a abordagem de rotulagem manual aprimorada pelo conceito Human-in-the-Loop (HITL), estratégia que combina a interação humana com modelos de IA para aprimorar a qualidade da rotulagem e a precisão dos sistemas automatizados, e arquiteturas de aprendizado profundo para segmentação de instâncias, o estudo conseguiu otimizar significativamente o processo de rotulagem. Como resultado, foi criado um conjunto de dados substancialmente maior do que os anteriores, incluindo 2.000 radiografias rotuladas disponíveis publicamente.

Abdulkreem et al. (2024) investigaram o uso de IA no diagnóstico automatizado de caninos maxilares impactados em RPs. O estudo utilizou uma CNN baseada na arquitetura SqueezeNet para classificar caninos impactados e não impactados. Inicialmente, a área sob a curva (*area under the curve*, AUC) da curva característica de operação do receptor (*receiver operating characteristic*, ROC), conhecida como AUC-ROC, apresentou

um valor de 84%. No entanto, a introdução de detectores treinados para identificar marcos anatômicos e orientar o corte automático das imagens resultou em uma melhora significativa na classificação, elevando a AUC-ROC para 96%. Esse aprimoramento também impactou a acurácia e a precisão do modelo, reforçando o potencial da IA para otimizar o pré-processamento de RPs. As análises demonstraram sua relevância na melhoria da identificação e classificação de caninos impactados, com possíveis aplicações nas práticas clínicas de ortodontia.

Kazimierczak et al. (2024) avaliaram a eficácia de um software de IA chamado Diagnocat na detecção de lesões periapicais em diferentes tipos de imagens odontológicas. O estudo mostrou que, enquanto a sensibilidade do software foi baixa (33,33%) em imagens de ortopantomografia (OPG), como também são chamadas as RPs, os resultados para tomografia computadorizada de feixe cônico (*cone beam computer tomography*, CBCT), foram muito mais robustos, com sensibilidade de 77,78% e F1 Score de 84,00%. Apesar das limitações em imagens bidimensionais e de menor qualidade, o sistema demonstrou alta especificidade em ambos os casos, sugerindo seu potencial como ferramenta diagnóstica, especialmente para imagens tridimensionais de CBCT.

Apesar dos avanços, os trabalhos apresentam algumas limitações. A segmentação dos dentes é difícil devido aos desafios das estruturas ósseas, o que resulta em falsos positivos. A detecção de regiões dentárias onde há ausência de dentes é prejudicada por variações nas estruturas orais. Além disso, identificar anomalias dentárias específicas é desafiador, com a maioria dos métodos se concentrando apenas na segmentação e identificação de dentes normais.

Nesse sentido, este trabalho é justificado pela crescente demanda por soluções assistidas por computador na odontologia, especialmente em cenários em que a interpretação radiográfica apresenta desafios significativos. Ao propor um método em duas etapas para a detecção de dentes e classificação de anomalias dentárias, o objetivo deste trabalho é desenvolver um método que auxilie os profissionais da saúde na identificação de condições patológicas de forma mais eficiente e confiável, promovendo avanços no diagnóstico clínico e na qualidade do atendimento odontológico.

A Tabela 1 apresenta um resumo dos trabalhos supracitados, que empregaram IA na análise de imagens odontológicas, destacando o objetivo principal de cada pesquisa, o número de imagens utilizadas e o tipo de imagem explorada.

Tabela 1 – Resumo de trabalhos que fizeram uso de IA em imagens odontológicas.

Trabalho	Objetivo	Imagens	Tipo de imagem
Silva, Oliveira e Pithon (2018)	Revisão de métodos e conjunto de dados para segmentação de dentes	1.500	RP
Panetta et al. (2021)	Conjunto de dados multimodal para segmentação dentária e detecção de anomalias	1.000	RP
Nader et al. (2022)	Segmentação automática com U-Net e Mask R-CNN	1.500	RP
Park et al. (2022)	Detecção de ausência dentária usando aprendizado profundo	455	RP
Silva et al. (2023)	Grande conjunto de dados rotulado para pesquisas em RPs	2.000	RP
Karaoglu et al. (2023)	Segmentação e numeração com abordagem híbrida	2.702	RP
Abdulkreem et al. (2024)	Diagnóstico automatizado de caninos maxilares impactados	1.490	RP
Kazimierczak et al. (2024)	Detecção de lesões periapicais com Diagnocat	-	RP e CBCT

3 FUNDAMENTAÇÃO TEÓRICA

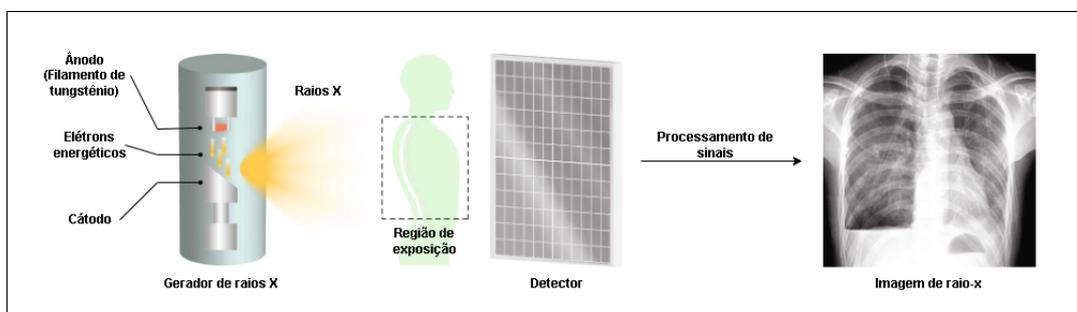
Este capítulo apresenta os conceitos teóricos essenciais para a compreensão do método proposto. São descritos temas relacionados ao processamento de imagens, aprendizado de máquina, funções de custo, arquiteturas de redes neurais e métodos de otimização de hiperparâmetros, além das anomalias dentárias consideradas neste trabalho. Esses conceitos são fundamentais para o desenvolvimento do método proposto.

3.1 Imagens Radiográficas na Odontologia

Os raios X são um tipo de radiação ionizante com comprimento de onda entre 0,01 e 10 nm. Ao viajar através de uma matéria, podem ser transmitidos, absorvidos ou espalhados. Sua excelente capacidade de penetração fez da imagem radiográfica uma modalidade amplamente utilizada na medicina (OU et al., 2021).

Uma imagem radiográfica representa a projeção em raios X das estruturas internas dos pacientes. Para sua obtenção, um sistema de imagem radiográfica é geralmente composto por um gerador de raios X, responsável pela emissão da radiação, e um detector, que capta os raios transmitidos e os converte em uma imagem, como ilustrado na Figura 1. Na radiografia convencional, esse papel é desempenhado pelo filme radiográfico, que armazenava e exibia as informações obtidas. Durante décadas, o filme foi o principal meio para aquisição e arquivamento de imagens diagnósticas (BERKHOUT; SANDERINK; STELT, 2003).

Figura 1 – Esquema representativo de um sistema de imagem por raios X.



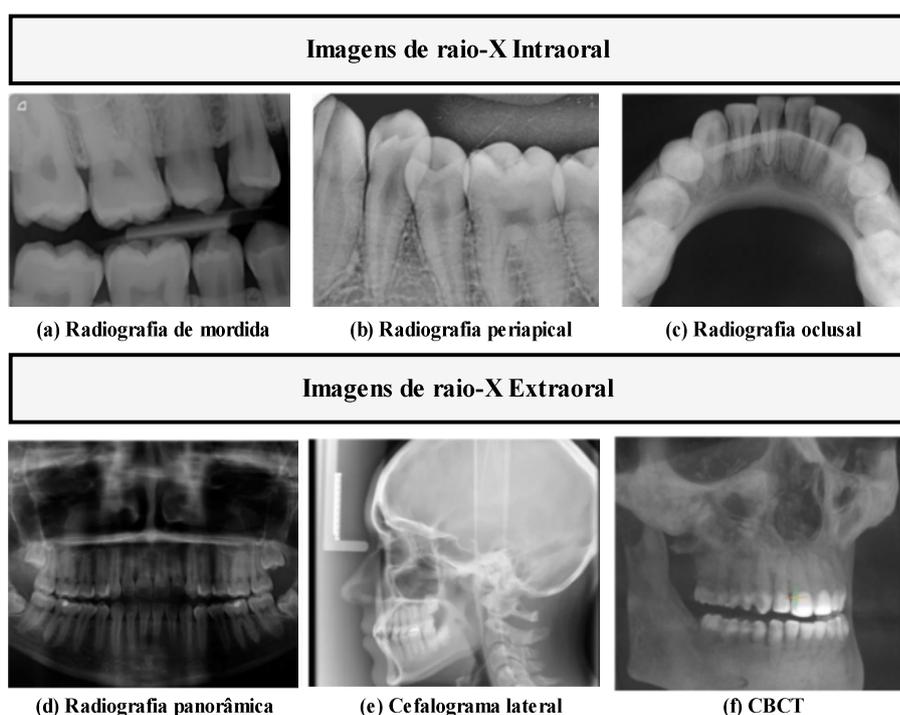
Fonte: Adaptado de (OU et al., 2021).

No século XXI a odontologia tem testemunhado uma transição significativa para o uso de imagens digitais, assim como o século XX foi marcado pela predominância de imagens em filme (WHITE; YOON; TETRADIS, 1999). Essa mudança reflete os avanços

tecnológicos e os benefícios das tecnologias digitais, como maior rapidez nos processos, menor exposição à radiação, menor impacto ambiental e maior facilidade para educar os pacientes sobre seus diagnósticos e tratamentos (WHITE; YOON; TETRADIS, 1999; MOLANDER; GRONDAHL; EKESTUBBE, 2004).

A radiografia odontológica, divide-se em dois tipos principais: intraoral e extraoral (SHAFI et al., 2023). A radiografia intraoral, realizada com o sensor dentro da boca, oferece alta resolução para identificar cáries, doenças dentárias e anomalias estruturais, sendo subdividida em três tipos: de mordida, que avalia dentes superiores e inferiores; periapical, que mostra dentes inteiros, incluindo raízes e ossos circundantes; e oclusal, que examina o posicionamento e desenvolvimento dos dentes nas arcadas. Já a radiografia extraoral, feita com o sensor fora da boca, é ideal para avaliar mandíbula, crânio e ossos faciais. Entre seus principais tipos, destacam-se os raios X panorâmicos, que oferecem uma visão geral da cavidade oral e ajudam na identificação de dentes impactados e tumores; os cefalogramas laterais, utilizados em ortodontia para analisar a relação entre dentes e estruturas faciais; e a CBCT, que permite reconstruções tridimensionais detalhadas da anatomia bucomaxilofacial, sendo essencial para implantodontia e planejamento cirúrgico. A Figura 2 ilustra os diferentes tipos de radiografias mencionados.

Figura 2 – Exemplo de diferentes tipos de radiografias intraorais e extraorais.



Fonte: Adaptado de (SHAFI et al., 2023).

A evolução das tecnologias radiográficas, que vai das imagens em filme às digitais, transformou a prática odontológica ao oferecer diagnósticos mais rápidos e precisos. Esse avanço continua a impulsionar melhorias na qualidade do planejamento e diagnóstico odontológico. Dentre os diversos tipos de imagens radiográficas disponíveis na odontologia, este trabalho focará especificamente na RP, por ser amplamente utilizada na prática clínica, permitindo a visualização de toda a arcada dentária em uma única imagem, e baixa exposição à radiação.

3.1.1 Radiografia Panorâmica

Atualmente, diversas técnicas radiográficas odontológicas são amplamente utilizadas para complementar o exame clínico, sendo as modalidades mais comuns aquelas específicas da odontologia, como a radiografia intraoral e a panorâmica. Dentre as citadas, a RP destaca-se pela capacidade de capturar, em uma única imagem, todos os dentes da maxila e da mandíbula (PAUWELS, 2020).

Embora a invenção dessa tecnologia seja amplamente atribuída a Tammissalo (1975), a ideia por trás desse conceito é mais antiga. Frank (1922) foi o primeiro a propor uma técnica de varredura com feixe estreito para os maxilares, chegando a obter uma patente e a cunhar o termo “radiografia panorâmica”. Anos depois, Numata (1933) publicou resultados clínicos relacionados a um protótipo de “radiografia parabólica”. Posteriormente, Heckmann (1939) desenvolveu um método de radiografia com feixe estreito para objetos curvos, contribuindo para o avanço dessa técnica.

As RPs são um tipo de técnica de imagem extraoral usada em odontologia para capturar uma visão ampla dos dentes, maxilares e estruturas circundantes em uma única imagem, como mostra a Figura 3. Esta modalidade de imagem permite que os dentistas examinem todos os dentes de uma vez, incluindo aqueles que ainda estão abaixo da linha da gengiva, e diagnostiquem várias condições, como cáries, fraturas dentárias, infecções e doenças dos ossos faciais (WAFSA, 2005). A técnica envolve o uso de uma máquina de raio-X rotativa, conforme ilustrado na Figura 4, que captura uma imagem bidimensional da área maxilofacial, fornecendo uma visão geral abrangente que auxilia nos exames clínicos e no planejamento do tratamento (WAFSA, 2005).

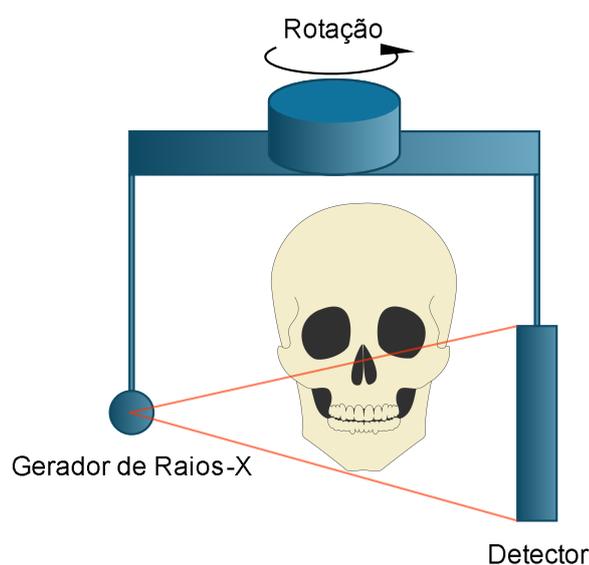
Devido a sua capacidade de capturar uma visão panorâmica dos dentes e estruturas adjacentes, a RP tornou-se uma importante ferramenta no diagnóstico odontológico. Neste

Figura 3 – Exemplo de imagem de RP.



Fonte: Base de dados DENTEX (HAMAMCI et al., 2023).

Figura 4 – Esquema ilustrativo do processo de obtenção de uma RP.



Fonte: Acervo do Autor.

trabalho, são aplicados métodos baseados em aprendizado profundo, para detectar e identificar dentes, e classificar determinadas anomalias dentárias em imagens de RP.

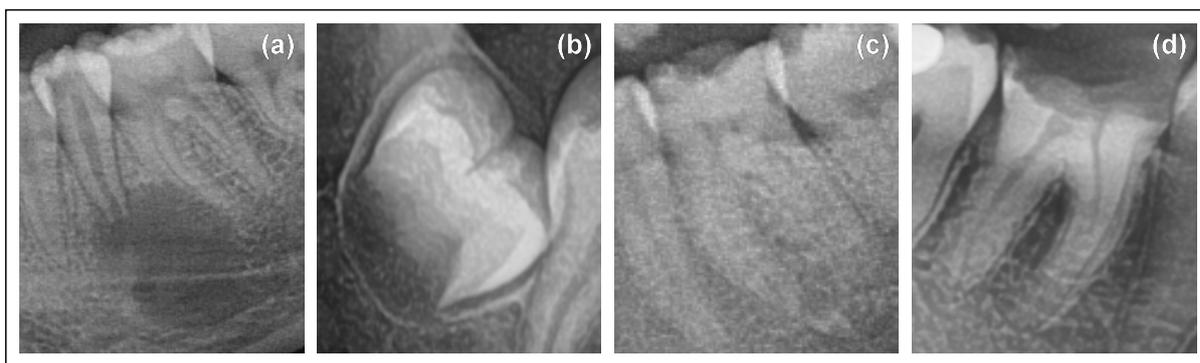
3.1.2 Anomalias Dentárias

A saúde bucal é um componente essencial da qualidade de vida, estando intimamente ligada ao bem-estar geral. As anomalias dentárias referem-se a alterações estruturais ou funcionais nos dentes, que podem comprometer a saúde, a função mastigatória e a estética do paciente. Neste trabalho, são abordadas as principais anomalias dentárias avaliadas, com ênfase em suas características clínicas e radiográficas relevantes

para os métodos de detecção de dentes e classificação de anomalias. Entre as condições consideradas, destacam-se: lesões periapicais, dentes impactados, cáries dentárias e cáries profundas. A seguir, cada uma dessas anomalias é descrita em detalhes, ressaltando os desafios diagnósticos.

As lesões periapicais são condições patológicas comuns que afetam os tecidos perirradiculares. Essas lesões ocorrem no ápice da raiz do dente como resposta a infecções ou inflamações da polpa dentária, frequentemente resultantes de cáries não tratadas, traumas ou outros fatores que levam à necrose pulpar (KARUNAKARAN et al., 2017). Elas podem se manifestar de diferentes formas, sendo as mais comuns o granuloma periapical, o cisto radicular e o abscesso periapical (SHTEYER; ROZOVSKY, 1972; LALONDE; LUEBKE, 1968). O granuloma periapical é uma inflamação crônica assintomática na maioria dos casos, mas que pode causar dor ou inchaço. O cisto radicular, geralmente maior, origina-se de restos epiteliais em resposta à inflamação e está associado a dentes não vitais. Já o abscesso periapical é o acúmulo de pus que pode provocar dor intensa e edema. Radiograficamente, essas lesões são radiolúcidas devido à reabsorção óssea ao redor do ápice radicular, sendo comumente diagnosticadas por meio de exames de imagem (STOJMENOVA; GJUROVSKI, 2022). A Figura 5 (a) apresenta uma lesão periapical observada em uma radiografia.

Figura 5 – Exemplo de alterações dentárias. (a) Lesão periapical; (b) Dente impactado; (c) Dente com cárie; (d) Cárie profunda.



Fonte: Acervo do autor.

Um dente impactado é aquele que permanece total ou parcialmente recoberto por mucosa e osso por um período superior a dois anos após a idade esperada para sua erupção fisiológica (AL-ZOUBI et al., 2017). A impaction dentária é uma condição frequentemente observada na prática clínica, com prevalência variável de acordo com a população estudada, a localização geográfica e o tipo de dente envolvido (MANJUNATHA

et al., 2014). Os terceiros molares são os dentes mais comumente impactados, seguidos pelos caninos maxilares. A RP é a principal modalidade de imagem utilizada para a detecção de dentes impactados e a avaliação das lesões associadas (SARICA et al., 2019). A Figura 5 (b) ilustra um caso de impactação dentária.

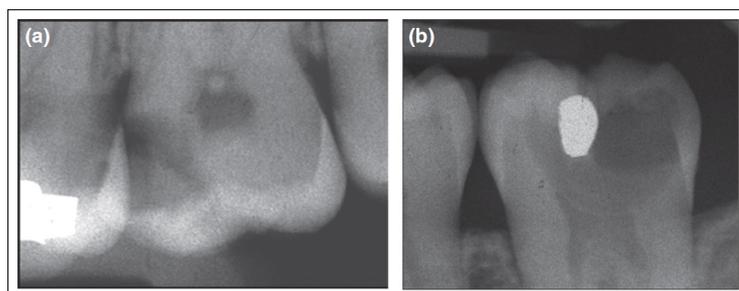
A cárie dentária é a destruição localizada dos tecidos duros dos dentes, causada por ácidos provenientes da fermentação bacteriana de carboidratos da dieta. O processo da doença começa dentro do biofilme bacteriano (placa dentária), que se forma nas superfícies dentais, provocando a desmineralização do esmalte e podendo atingir camadas mais profundas se não tratado. Multifatorial, a cárie é influenciada por fatores como alterações microbiológicas no biofilme, saliva, flúor, dieta rica em açúcar e hábitos de higiene (SELWITZ; ISMAIL; PITTS, 2007). A Figura 5.c apresenta uma radiografia de um dente com cárie.

A cárie profunda ocorre quando a lesão cariosa progride significativamente na dentina, atingindo o terço ou quarto interno dessa camada, com risco elevado de exposição da polpa dentária (INNES et al., 2016). Esse processo é intensificado em áreas de estagnação, onde o biofilme microbiano metaboliza carboidratos fermentáveis, gerando ácidos que promovem a desmineralização do esmalte e desencadeiam respostas defensivas, como o aumento da dentina intratubular e inflamação inicial na polpa (BJØRNDAL et al., 2019). Com a progressão, a dentina é invadida por bactérias, resultando em maior destruição estrutural e eventual cavitação. A Figura 5 (d) apresenta uma cárie profunda visualizada em uma radiografia.

Radiograficamente, as cáries profundas apresentam-se como lesões que se aproximam do tecido pulpar e são classificadas em dois tipos principais: cáries profundas, onde há uma camada de dentina radiopaca separando o tecido cariado da polpa, e cáries extremamente profundas, nas quais a desmineralização alcança a polpa, aumentando o risco de pulpite e outras complicações (BJØRNDAL et al., 2019). A Figura 6 ilustra os estágios mais avançados da cárie profunda.

Dessa forma, as patologias descritas representam condições comuns e clinicamente relevantes na prática odontológica, sendo de grande importância sua identificação precisa em exames de RP. Ao longo desta dissertação, serão discutidas as abordagens e técnicas utilizadas para a identificação automatizada dessas anomalias dentárias.

Figura 6 – Classificação para estágios mais profundos de cárie. (a) Cárie profunda atingindo o quarto pulpar com uma zona de dentina separando a lesão da polpa (b) e extremamente profunda penetrando toda a espessura da dentina.



Fonte: (BJØRNDAL et al., 2019)

3.1.3 Numeração de Dentes

O sistema de numeração de dentes mais adotado pelos dentistas é a notação FDI World Dental Federation (FDI, do francês *Fédération Dentaire Internationale*), uma organização que representa a odontologia globalmente. Esse sistema padronizado facilita a comunicação entre profissionais de diferentes regiões e idiomas, contribuindo para registros clínicos consistentes e trocas de informações mais precisas (BLINKHORN; CHOI; PAGET, 1998).

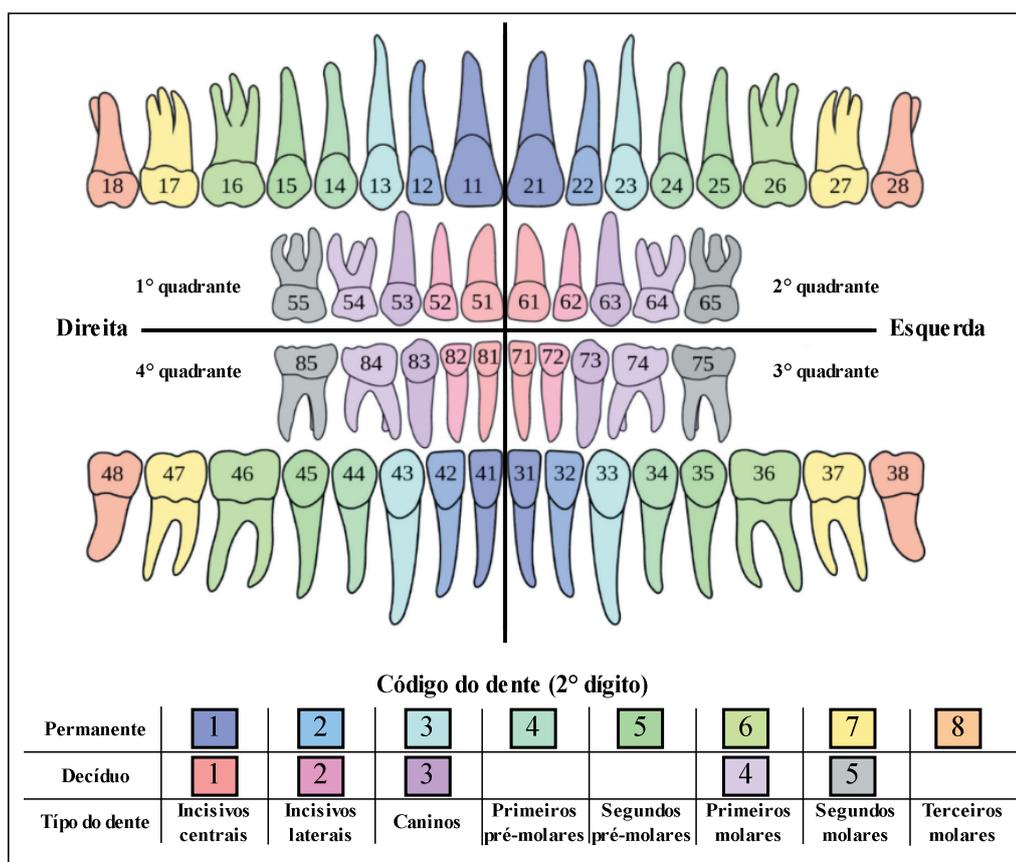
Na notação FDI, cada dente é representado por um código de dois dígitos, em que o primeiro dígito indica o quadrante e o tipo de dentição, diferenciando dentes permanentes (quadrantes 1 a 4) e decíduos ou de leite (quadrantes 5 a 8). O segundo dígito identifica a posição do dente dentro do quadrante, numerado de 1 a 8 para dentes permanentes e de 1 a 5 para dentes decíduos. A Figura 7 ilustra essa notação, destacando a lógica numérica que facilita o reconhecimento rápido e preciso da posição do dente na arcada dentária.

Neste trabalho, adota-se a notação FDI para identificação dos dentes, devido às suas diversas vantagens. Entre elas, destacam-se a redução de erros na diferenciação entre os lados da boca (direito/esquerdo) e as arcadas dentárias (superior/inferior) (YURDUKORU, 1989). Além disso, essa notação se mostra prática na digitação e é amplamente compatível com softwares de gestão de clínicas dentárias e sistemas de radiologia, reforçando sua aplicabilidade em contextos clínicos e tecnológicos.

3.2 Processamento de Imagens Digitais

Uma imagem digital é uma forma de representação numérica de uma imagem bidimensional. Essa representação é estruturada como uma matriz composta por elementos

Figura 7 – Notação FDI World Dental Federation.



Fonte: Adaptado de (SILVA et al., 2023).

chamados pixels (KRUTSCH; TENORIO, 2011). O processamento de imagens digitais é uma área de grande relevância, pois melhora a qualidade visual para interpretação humana e viabiliza o armazenamento, transmissão e análise automatizada das imagens para a percepção por máquinas (CHITRADEVI; SRIMATHI, 2014; GONZALEZ; WOODS, 2000). Essa tecnologia tem ampla aplicação em diversas áreas, auxiliando na extração detalhada de informações sobre os objetos estudados (LYUBCHENKO et al., 2016). Historicamente, uma das primeiras aplicações notáveis ocorreu no início do século passado, com o aprimoramento da qualidade de imagens transmitidas pelo sistema Bartlane, que conectava Londres e Nova Iorque por cabos submarinos (FILHO; NETO, 1999).

No contexto técnico, o processamento de imagens digitais consiste em transformar sucessivamente uma imagem para facilitar a extração de informações relevantes (ALBUQUERQUE; ALBUQUERQUE, 2000). Esses procedimentos, normalmente expressos de forma algorítmica, permitem a análise, armazenamento, transmissão e representação de imagens, atendendo tanto à percepção humana quanto às necessidades de processamento automatizado (FILHO; NETO, 1999). A versatilidade e aplicabilidade do processamento

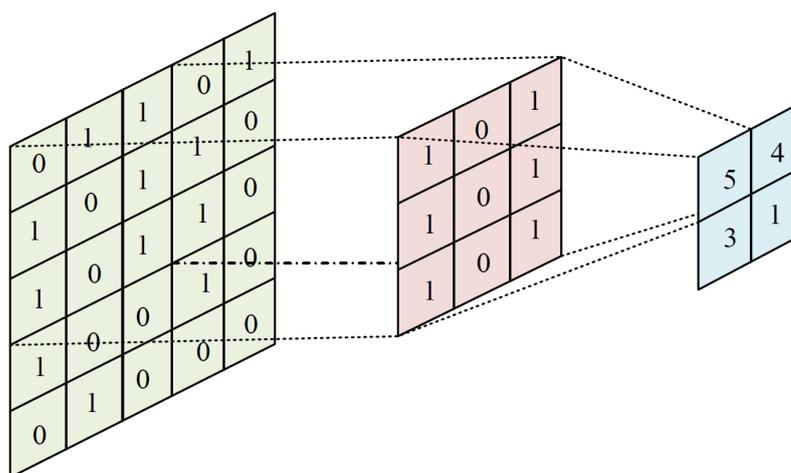
de imagens tornam-no essencial para o avanço de tecnologias em diversos domínios.

3.2.1 Convolução

A convolução é uma operação matemática fundamental em sistemas lineares e no processamento de imagens digitais, sendo amplamente utilizada em diversas aplicações de engenharia e matemática (PODLOZHNYUK, 2007). Essencialmente, a convolução aplica um filtro, representado por uma matriz chamada *kernel*, sobre uma imagem para realizar transformações que destacam características específicas. Esse processo envolve o movimento do *kernel* sobre a imagem, calculando-se a soma dos produtos entre os elementos do *kernel* e os pixels correspondentes da matriz de origem (GONZALEZ; WOODS, 2000).

No contexto bidimensional, a convolução pode ser entendida como uma matriz que “desliza” sobre outra matriz, unidade por unidade, gerando como resultado a soma dos produtos elemento a elemento (KIM; CASPER, 2013). Quando a matriz de origem é uma imagem e o *kernel* estabelece um filtro, o resultado da convolução é uma versão filtrada dessa imagem. Esse procedimento permite aplicar filtros espaciais, como suavização ou detecção de bordas, e é repetido até que o *kernel* percorra todos os pixels possíveis da matriz de entrada, transformando a imagem original de acordo com os objetivos do processamento. A Figura 8 ilustra um exemplo de operação de convolução.

Figura 8 – Exemplo de operação de convolução com um *kernel* 3x3 aplicado a uma imagem 5x5 com deslocamento de 2 pixels por unidade.



Fonte: (SONG, 2023).

A operação de convolução entre um *kernel* $w(x, y)$ e uma imagem $f(x, y)$, denotada por $w(x, y) \star f(x, y)$, é formalmente definida pela expressão,

$$g(x, y) = \sum_{s=-a}^a \sum_{t=-b}^b w(s, t) f(x - s, y - t), \quad (3.1)$$

onde $g(x, y)$ é a imagem resultante da convolução, $f(x, y)$ representa a imagem original e $w(x, y)$ é o *kernel* aplicado. Os parâmetros x e y indicam as coordenadas do pixel atual na imagem de saída, enquanto $a = \frac{m-1}{2}$ e $b = \frac{n-1}{2}$ correspondem aos deslocamentos baseados nas dimensões $m \times n$ do *kernel* (GONZALEZ; WOODS, 2000). Essa formulação assegura que o *kernel* percorra todos os pixels da imagem original, proporcionando uma transformação consistente e controlada.

3.2.2 Equalização de Histograma

O histograma de uma imagem digital é a representação gráfica que descreve a distribuição dos valores de intensidade dos pixels, fornecendo informações sobre a frequência de ocorrência de diferentes níveis de cinza (KRUTSCH; TENORIO, 2011). Ele é amplamente utilizado para avaliar o contraste da imagem e facilitar a identificação de detalhes visuais, sendo uma etapa importante no pré-processamento de imagens em diversas aplicações.

Para uma imagem com níveis de intensidade definidos no intervalo $[0, L - 1]$, o histograma pode ser descrito matematicamente por uma função discreta,

$$h(r_k) = n_k, \quad (3.2)$$

onde r_k representa o k -ésimo nível de intensidade e n_k corresponde ao número de pixels com essa intensidade (GONZALEZ; WOODS, 2000).

A normalização do histograma é realizada dividindo n_k pelo número total de pixels da imagem, denotado por MN . Esse procedimento resulta em uma estimativa da probabilidade de ocorrência de cada nível de intensidade, garantindo que a soma de todas as probabilidades seja igual a 1.

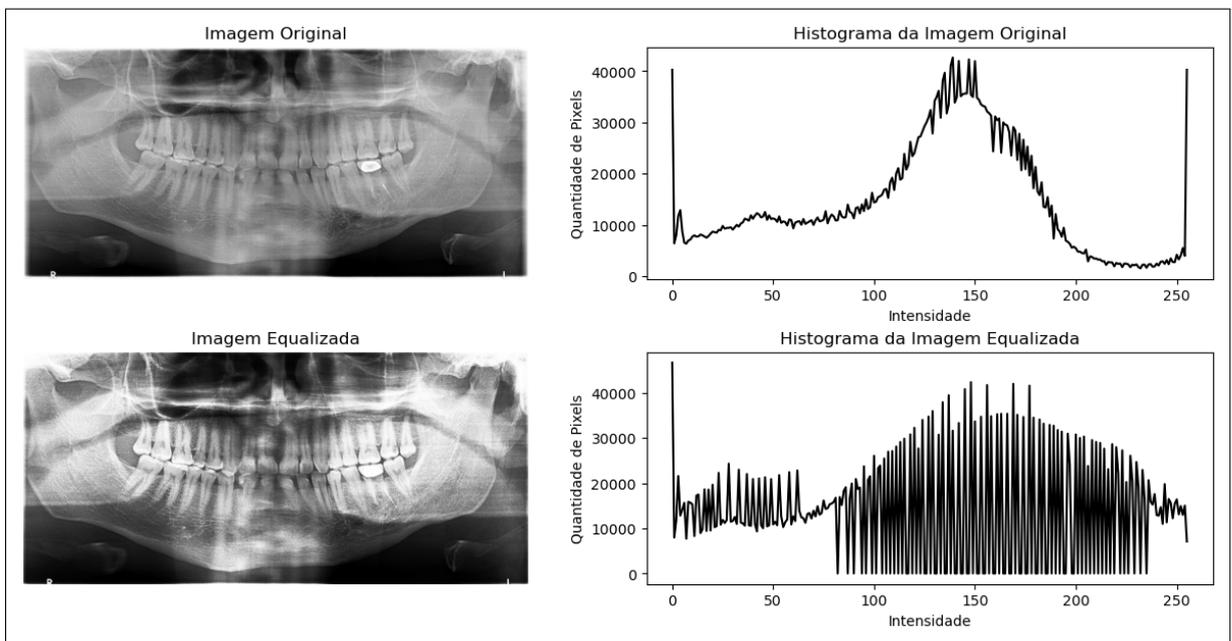
Um processo fundamental associado ao histograma é a equalização, que redistribui os níveis de intensidade para aumentar o contraste da imagem. A Figura 9 ilustra esse efeito, comparando a imagem original com a versão equalizada, na qual se observa uma

melhora na visibilidade dos detalhes. A transformação utilizada para calcular os novos níveis de intensidade pode ser expressa como,

$$s_k = T(r_k) = (L - 1) \sum_{j=0}^k p_r(r_j) = \frac{(L - 1)}{MN} \sum_{j=0}^k n_j, \quad k = 0, 1, 2, \dots, L - 1, \quad (3.3)$$

onde $p_r(r_j)$ representa a probabilidade normalizada do j -ésimo nível de intensidade. Essa transformação promove uma distribuição mais uniforme dos níveis de intensidade na imagem, melhorando a visibilidade dos detalhes e ampliando o intervalo dinâmico.

Figura 9 – Exemplo do antes e depois de uma imagem após equalização de histograma.



Fonte: Acervo do autor.

3.3 Aprendizado de Máquina

O aprendizado de máquina é uma abordagem que permite aos sistemas aprenderem com dados de treinamento específicos para resolver tarefas analíticas de forma automatizada (JANIESCH; ZSCHECH; HEINRICH, 2021). Por meio da identificação automática de relacionamentos e padrões significativos em exemplos e observações, os algoritmos de aprendizado de máquina iterativamente extraem padrões complexos e ocultos, eliminando a necessidade de programação explícita para cada tarefa (BISHOP;

NASRABADI, 2006). Essa característica torna o aprendizado de máquina uma ferramenta útil para problemas que envolvem grande variabilidade e complexidade nos dados.

Entre os avanços mais significativos nesse campo, destaca-se o desenvolvimento de redes neurais artificiais (*artificial neural networks*, ANNs) em arquiteturas cada vez mais profundas, caracterizadas como aprendizado profundo (GOODFELLOW, 2016; LECUN; BENGIO; HINTON, 2015). O aprendizado profundo se mostra especialmente eficaz em domínios que envolvem grandes volumes de dados complexos, como texto, imagem, vídeo, fala e áudio. Nessas áreas, redes neurais profundas frequentemente superam algoritmos mais rasos de aprendizado de máquina, graças à sua capacidade de processar dados complexos e extrair padrões detalhados e contextuais (LECUN; BENGIO; HINTON, 2015)

3.3.1 Redes Neurais Artificiais

As redes neurais artificiais, ou simplesmente redes neurais, são inspiradas nos primeiros modelos de processamento sensorial do cérebro, de forma a ser possível criar uma ANN simulando uma rede de neurônios em um computador (KROGH, 2008). Aplicando algoritmos que imitam os processos de neurônios reais, é possível fazer com que uma rede “aprenda” a resolver determinados problemas.

O modelo de neurônio de McCulloch-Pitts (ou unidade de limiar), representado na Figura 10, recebe entrada de N outras unidades ou fontes externas, numeradas de 1 a N (MCCULLOCH; PITTS, 1943). O total de entradas para uma unidade é a soma ponderada de todas as entradas,

$$\sum_{i=1}^N w_i x_i = w_1 x_1 + w_2 x_2 + \dots + w_N x_N, \quad (3.4)$$

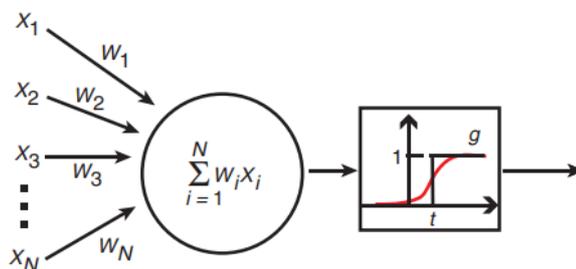
onde x_i representa uma entrada i e w_i representa um peso associado a ela. Dessa forma, estando acima de um limiar t a saída é igual a 1, caso contrário, a saída será 0. Seja σ uma função degrau que assume o valor 0 para argumentos negativos e 1 para argumentos não negativos. Assim, a saída pode ser expressa como:

$$\sigma\left(\sum_{i=1}^N w_i x_i - t\right). \quad (3.5)$$

Uma rede neural é constituída principalmente por um grande número de neurônios (ou nós) computacionais interconectados, que são distribuídos de forma a ter uma camada

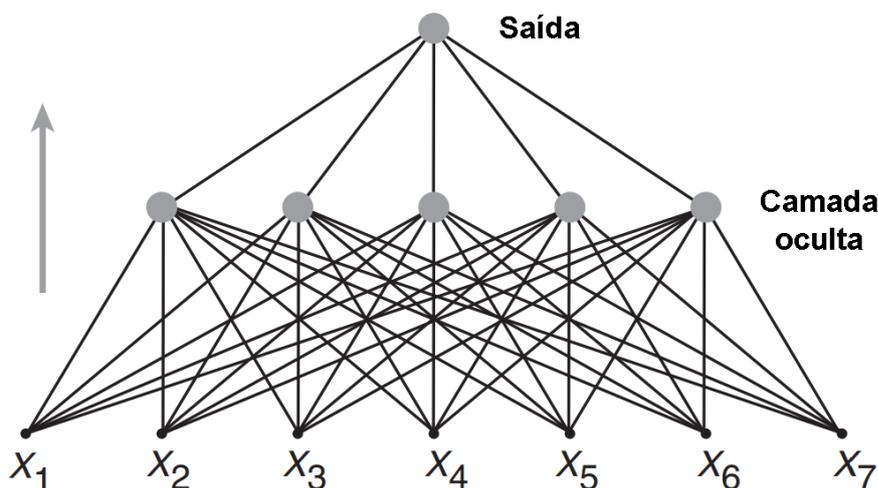
de neurônios de entrada, seguida de camadas de neurônios ocultas, e uma camada de neurônios final na saída, como é representado na Figura 11.

Figura 10 – Modelo de neurônio McCulloch-Pitts.



Fonte: (KROGH, 2008).

Figura 11 – Exemplo de ANN.



Fonte: Adaptado de (KROGH, 2008).

3.3.2 Redes Neurais Convolucionais

As CNNs representam uma das redes neurais profundas mais utilizadas. Assim como as ANNs, CNNs são compostas por neurônios que se ajustam por meio do aprendizado (O'SHEA; NASH, 2015).

As CNNs são uma subclasse de ANNs que possuem pelo menos uma camada de convolução. Elas se destacam por sua organização tridimensional, onde os neurônios de cada camada são distribuídos em altura, largura e profundidade. Enquanto altura e largura correspondem às dimensões espaciais da entrada, a profundidade representa o número de mapas de características (*feature maps*) gerados pelos filtros aplicados, e não

a quantidade total de camadas na rede. Essa estrutura permite que as CNNs capturem padrões espaciais de forma mais eficiente do que redes totalmente conectadas.

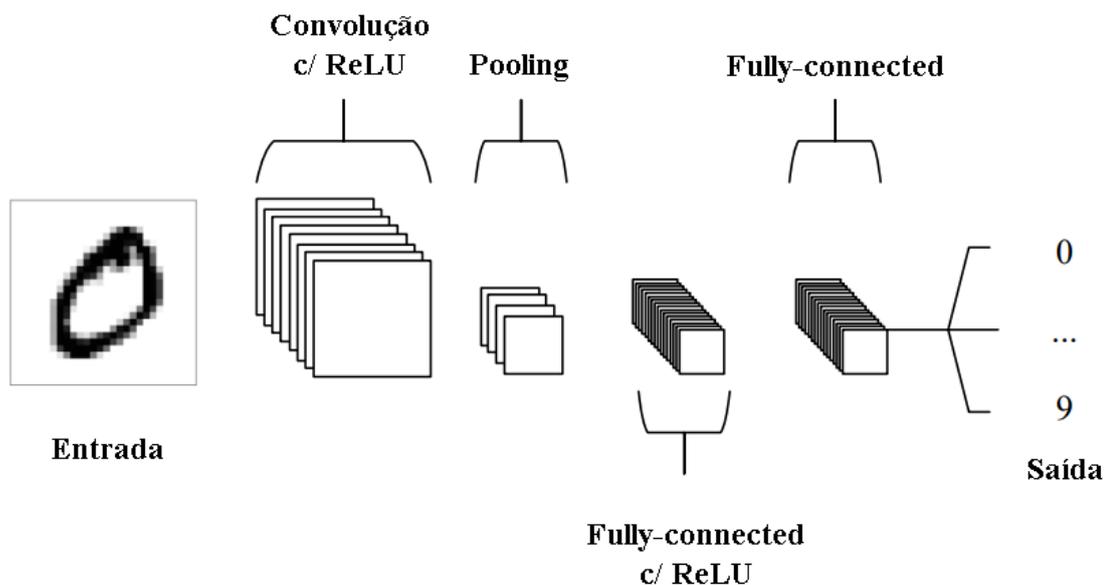
Diferentemente das ANNs tradicionais, nas quais os neurônios de uma camada costumam estar totalmente conectados à camada anterior, nas CNNs, cada neurônio processa apenas uma pequena região da entrada anterior, preservando a estrutura espacial dos dados. Essa característica torna as CNNs particularmente eficazes para o reconhecimento de padrões em imagens, pois permite a extração automática de características hierárquicas, reduzindo o número de parâmetros e melhorando a generalização do modelo (O'SHEA; NASH, 2015).

No que diz respeito à arquitetura, CNNs geralmente possuem três tipos principais de camadas: convolucional (*convolutional*), de agrupamento (*pooling*) e totalmente conectada (*fully-connected*). No entanto, algumas arquiteturas modernas incluem camadas adicionais, como normalização por lote (*batch normalization*) e funções de ativação. A Figura 12 apresenta um exemplo dessa estrutura. Uma das principais vantagens das CNNs sobre ANNs tradicionais é a capacidade de extrair automaticamente características espaciais dos dados, reduzindo a necessidade de extração manual de características. Além disso, o compartilhamento de pesos nas camadas convolucionais reduz significativamente o número de parâmetros quando comparado a redes totalmente conectadas, tornando o treinamento mais eficiente e viabilizando modelos mais profundos para tarefas complexas (ALBAWI; MOHAMMED; AL-ZAWI, 2017).

A camada convolucional processa a entrada aplicando filtros que realizam uma soma ponderada dos valores da imagem em pequenas regiões locais, gerando um mapa de características (O'SHEA; NASH, 2015). Conforme descrito na Seção 3.2.1, a convolução em redes neurais convolucionais utiliza filtros que deslizam sobre a imagem de entrada, multiplicando seus pesos pelos valores dos pixels correspondentes e somando os resultados. Esse processo permite extrair representações relevantes, como bordas e texturas nas camadas iniciais, e características mais abstratas em camadas mais profundas. O mapa de características gerado serve como entrada para as próximas camadas da rede.

A camada de *pooling* reduz a dimensionalidade espacial da entrada ao aplicar operações como *max pooling*, que seleciona o maior valor de cada região, ou *average pooling*, que calcula a média dos valores. Diferentemente da convolução, essa operação não utiliza pesos aprendíveis, apenas seleciona estatisticamente os valores mais representativos dentro de regiões da entrada. A principal função do *pooling* é reduzir o número de parâmetros

Figura 12 – Uma arquitetura CNN com cinco camadas.



Fonte: Adaptado de (O'SHEA; NASH, 2015).

e a complexidade computacional da rede, preservando as informações mais importantes para a tarefa de reconhecimento (O'SHEA; NASH, 2015).

A função de ativação ReLU (*Rectified Linear Unit*) é uma das mais utilizadas para introduzir não linearidade na rede ao transformar a saída da camada anterior, zerando valores negativos e mantendo valores positivos inalterados. Essa propriedade impede que os neurônios fiquem saturados em faixas negativas, além de facilitar a propagação do gradiente durante o treinamento. Dessa forma, a ReLU contribui para que a rede aprenda representações mais complexas e profundas (BROWNLEE, 2019).

Na camada totalmente conectada, ocorre a classificação das imagens na CNN com base nas *features* extraídas nas camadas anteriores. Essas camadas geram pontuações de classe a partir das ativações, permitindo a tomada de decisão na rede. Diferentemente das camadas convolucionais e de *pooling*, os neurônios de uma camada totalmente conectada estão ligados a todos os neurônios da camada anterior, permitindo uma combinação abrangente das informações extraídas. A camada seguinte pode ter uma estrutura diferente, como uma camada de saída com um número reduzido de neurônios correspondendo às classes do problema.

3.4 Função de Custo

No contexto da otimização matemática e da teoria da decisão, uma função de custo (também chamada de função de perda ou função de erro) é uma ferramenta que atribui um valor real a um evento ou a variáveis, representando o “custo” associado ao desvio de um resultado desejado. A função de perda é fundamental para traduzir a performance do modelo em um valor numérico que pode ser minimizado por métodos de otimização. Em estatística, essas funções são amplamente utilizadas para estimar parâmetros, onde o “custo” reflete a discrepância entre valores estimados e verdadeiros (HASTIE et al., 2005; WALD, 1949).

O aprendizado de máquina pode ser interpretado como um problema de minimização de risco estrutural (*structural risk minimization*), no qual o objetivo é encontrar os parâmetros do modelo que minimizem uma função de custo, balanceando o ajuste ao conjunto de treinamento e a complexidade do modelo. A formulação matemática do problema de otimização é geralmente expressa como,

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N L_{\theta}(f(x_i)) + \lambda R(f), \quad (3.6)$$

onde N é o número de amostras, $L(\cdot)$ a função de perda, θ os parâmetros do modelo, $R(f)$ o termo de regularização e $\lambda \geq 0$ o fator de equilíbrio entre risco empírico e regularização. A função de perda define o risco empírico, enquanto a regularização evita o sobreajuste (*overfitting*), fenômeno em que o modelo se ajusta excessivamente aos dados de treinamento, perdendo capacidade de generalização para novos dados (WANG et al., 2020)

A escolha da função de perda depende do problema, dos dados e dos objetivos do modelo. Neste trabalho, foi adotado a perda focal (*focal loss*, FL) como função de perda durante o treinamento dos modelos de classificação.

3.4.1 Focal Loss

A FL é uma função de perda proposta por (LIN et al., 2017) para abordar o problema de desequilíbrio de classe, reformulando a perda de entropia cruzada (*cross entropy loss*, CE) (GOOD, 1992), de forma a reduzir o impacto dos exemplos bem classificados no cálculo da perda.

$$CE(p, y) = \begin{cases} -\log(p); & \text{se } y = 1, \\ -\log(1 - p); & \text{caso contrário,} \end{cases} \quad (3.7)$$

onde $y \in \{\pm 1\}$, sendo p a probabilidade para a classe com $y = 1$ e $p \in [0, 1]$. p_t é definido como,

$$p_t = \begin{cases} p; & \text{se } y = 1, \\ 1 - p; & \text{caso contrário,} \end{cases} \quad (3.8)$$

assim, temos que $CE(p, y) = CE(p_t) = -\log(p_t)$.

A FL concentra o treinamento em um conjunto esparsos de exemplos difíceis evitando que um grande número de exemplos fáceis (ou seja, aqueles já classificados corretamente com alta confiança) sobrecarreguem o modelo durante o treinamento (LIN et al., 2017). Ela propõe a adição de um fator modulador $(1 - p_t)^\gamma$ para CE, com parâmetro de foco ajustável $\gamma \geq 0$. Formalmente a FL é definida da seguinte forma:

$$FL(p_t) = -(1 - p_t)^\gamma \log(p_t). \quad (3.9)$$

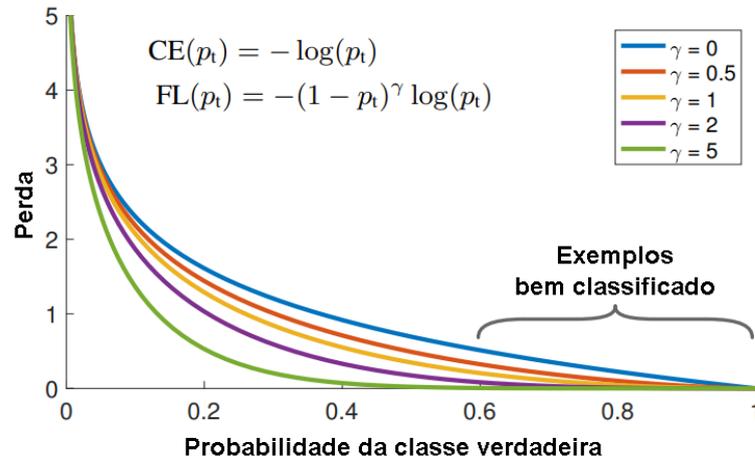
Uma variante α -balanceada dessa equação pode ser utilizada. Nessa abordagem, um fator α é introduzido para ajustar a influência das classes no cálculo da perda, sendo especialmente útil em problemas de classificação com dados desbalanceados. De acordo com os experimentos de Lin et al. (2017), essa variante resulta em uma precisão ligeiramente melhorada em relação à forma não α -balanceada,

$$FL(\rho_t) = -\alpha_t(1 - \rho_t)^\gamma \log(\rho_t), \quad (3.10)$$

pode-se também observar que quando $\gamma = 0$ e $\alpha_t = 1$, FL é igual a CE.

A Figura 13 apresenta uma visualização da FL para diferentes valores de γ no intervalo $[0, 5]$. Nota-se que, à medida que γ aumenta, a função de perda passa a atribuir menor peso aos exemplos mais fáceis, ou seja, aqueles com perda baixa. No gráfico, a CE é representada pela curva azul ($\gamma = 0$). Observa-se que, com o aumento de γ , a FL ajusta a penalização dos exemplos, potencialmente melhorando o desempenho do modelo em problemas com desbalanceamento de classes.

Figura 13 – Visualização do comportamento da FL para determinados valores de γ .



Fonte: Adaptado de (LIN et al., 2017).

3.5 *You Only Look Once* (YOLO)

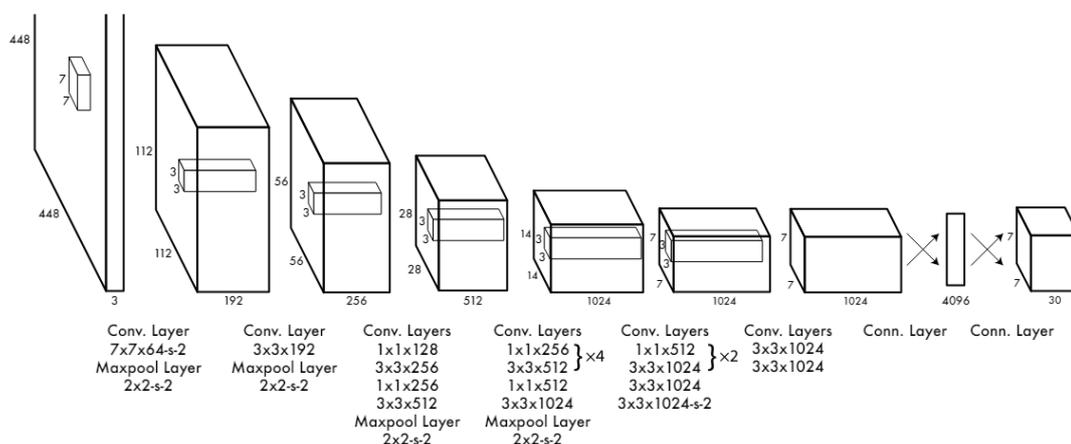
Entre os diversos algoritmos de detecção de objetos, o *You Only Look Once* (YOLO) destacou-se como uma abordagem inovadora e eficiente, unificando as etapas de detecção em uma única passagem pela rede. Desenvolvido por Redmon (2016), o YOLO marcou uma ruptura significativa em relação às técnicas anteriores, que geralmente envolviam processos mais lentos e fragmentados, como janelas deslizantes ou abordagens de duas etapas, como o Fast R-CNN (GIRSHICK, 2015).

Ao adotar uma estratégia baseada em regressão direta, o YOLO conseguiu prever simultaneamente as classes e as coordenadas das caixas delimitadoras, reduzindo drasticamente o tempo de processamento enquanto mantinha alta precisão. A primeira versão do YOLO, conhecida como YOLOv1, divide a imagem de entrada em uma grade $S \times S$, com cada célula da grade prevendo B caixas delimitadoras, cada uma associada a C classes. A previsão incluiu cinco parâmetros principais para cada caixa: P_c (confiança), b_x e b_y (coordenadas do centro), b_h (altura) e b_w (largura).

No artigo original, os autores utilizaram o conjunto de dados PASCAL VOC (EVERINGHAM et al., 2010) com $C = 20$, $S = 7$ e $B = 2$, resultando em um tensor de saída $7 \times 7 \times 30$. A arquitetura YOLOv1 consistia em 24 camadas convolucionais, seguidas por duas camadas totalmente conectadas, utilizando ativações *Leaky Rectified Linear Unit* (MAAS et al., 2013), e camadas convolucionais 1×1 inspiradas em GoogLeNet (SZEGEDY et al., 2015) para manter a eficiência do modelo, como mostra a Figura 14.

Essa arquitetura foi posteriormente complementada pelo modelo mais leve Fast YOLO, que reduziu a complexidade computacional ao usar apenas nove camadas convolucionais.

Figura 14 – A YOLOv1, com 24 camadas convolucionais e 2 totalmente conectadas, usando filtros 1×1 para reduzir a dimensionalidade.



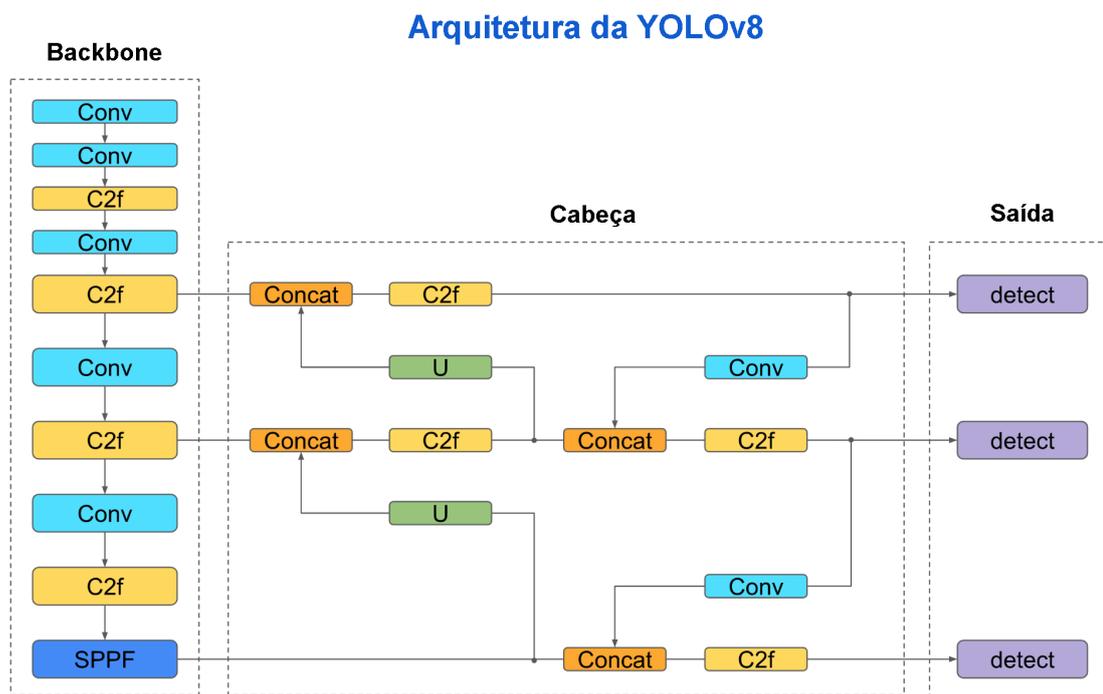
Fonte: (REDMON, 2016).

3.5.1 YOLOv8

YOLOv8 é uma versão do modelo de detecção de objetos YOLO, lançada pela Ultralytics em janeiro de 2023 (JOCHER; QIU; CHAURASIA, 2023). Atualmente, o YOLOv8 oferece cinco variantes com diferentes níveis de complexidade e exigência computacional: YOLOv8n (nano), YOLOv8s (small), YOLOv8m (medium), YOLOv8l (large) e YOLOv8x (extra large) (TERVEN; CÓRDOVA-ESPARZA; ROMERO-GONZÁLEZ, 2023). Essas versões se diferenciam principalmente pela profundidade e largura da rede, o que afeta diretamente seu desempenho geral e o tempo de processamento. Modelos menores, como o YOLOv8n, são mais rápidos e consomem menos recursos, sendo adequados para dispositivos com capacidade limitada, enquanto os maiores, como o YOLOv8x, entregam resultados mais robustos, embora demandem maior poder de processamento.

A arquitetura do YOLOv8 é composta por duas partes principais: o *Backbone* e a Cabeça, ambas implementadas como redes neurais totalmente convolucionais (SOHAN et al., 2024). Essa estrutura modular foi projetada para proporcionar maior eficiência na extração de características e na geração das previsões durante o processamento das imagens. A Figura 15 ilustra a arquitetura do YOLOv8, destacando seus principais componentes estruturais.

Figura 15 – Arquitetura YOLOv8.



Fonte: (SOHAN et al., 2024).

O *Backbone* do YOLOv8 apresenta uma rede modificada baseada na CSPDarknet53 (BOCHKOVSKIY; WANG; LIAO, 2020), composta por 53 camadas convolucionais. Ele utiliza conexões parciais entre estágios para melhorar a transmissão de informações entre diferentes níveis da rede. Essa estrutura aprimora a extração de características relevantes da imagem de entrada. O módulo C2f integra características de alto nível com informações contextuais, aumentando a precisão da detecção. Além disso, o modelo emprega o *partial pyramid pooling fast* (HE et al., 2015), que processa características em múltiplas escalas de forma eficiente, aprimorando o desempenho.

A Cabeça do YOLOv8 processa os mapas de características produzidos pelo *Backbone*, gerando a saída final do modelo na forma de caixas delimitadoras e classes de objetos. Um dos principais diferenciais é seu design modular e destacável, o que permite que as tarefas de classificação, pontuação de objetividade e regressão sejam gerenciadas de maneira independente (JOCHER; CHAURASIA; QIU, 2023). Esse design melhora a precisão geral do modelo e facilita o foco de cada ramificação em sua tarefa específica.

Além disso, as camadas U (*Upsample*) são usadas para aumentar a resolução dos mapas de características, enquanto uma sequência de camadas convolucionais e uma camada linear preveem as caixas delimitadoras e probabilidades de classe. O design da

Cabeça foi otimizado para equilibrar velocidade e precisão, com uma seleção cuidadosa do número de canais e tamanhos de *kernel*.

O módulo de Detecção completa a estrutura do YOLOv8, utilizando camadas convolucionais e lineares para mapear as características extraídas pelo *Backbone* em caixas delimitadoras e classes de objetos. Todo o design foi projetado para ser rápido, eficiente e preciso, garantindo alta performance em detecção de objetos. Essa otimização torna o YOLOv8 adequado para uma ampla gama de aplicações em tempo real.

3.6 DenseNet

A DenseNet (*dense convolutional network*) é uma arquitetura de rede neural que se destaca por conectar cada uma de suas camadas a todas as demais de maneira *feedforward*. Em uma rede neural *feedforward* (*feed forward network*), as conexões entre os nós seguem sempre em frente, sem formar ciclos (ZELL, 1994, p. 73). Enquanto redes convolucionais tradicionais com L camadas estabelecem apenas L conexões diretas, cada camada conectando-se apenas à sua sucessora imediata, a DenseNet expande essa conectividade ao estabelecer $\frac{L(L+1)}{2}$ conexões diretas (HUANG et al., 2017).

Preservando a natureza *feedforward* da rede, Huang et al. (2017) propõe um padrão de conectividade entre as camadas de forma que a l^{th} camada recebe os mapas de características de todas as camadas anteriores, x_0, \dots, x_{l-1} , como entrada:

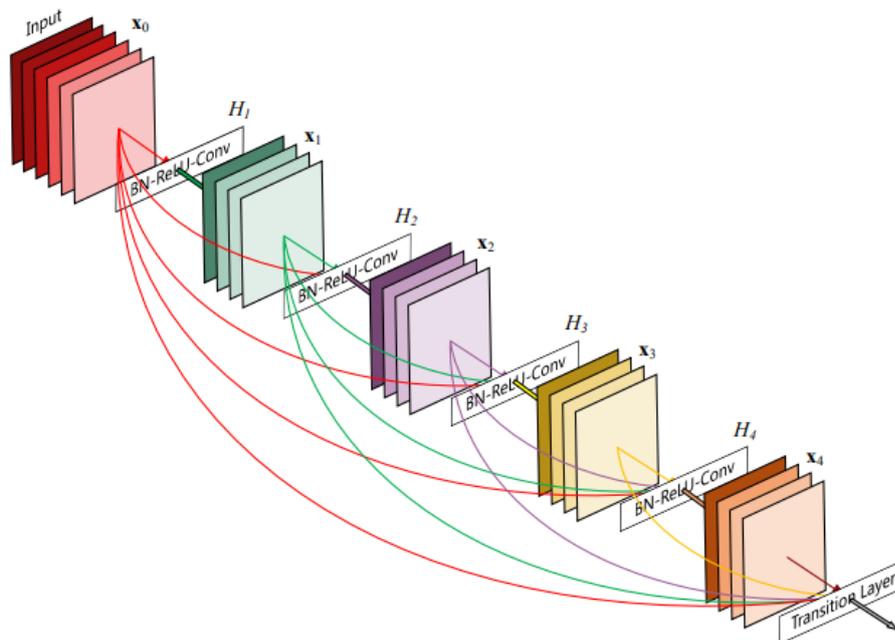
$$X_l = H_l([X_0, X_1, \dots, X_{l-1}]), \quad (3.11)$$

onde $[X_0, X_1, \dots, X_{l-1}]$ é referente à concatenação dos mapas de características produzidos nas camadas $0, \dots, l-1$. Essa característica de conectividade densa dá origem ao nome DenseNet. O H_l nessa equação representa as múltiplas entradas em um único tensor. A Figura 16 mostra essa conectividade entre as camadas.

3.6.1 DenseNet-121

A DenseNet-121 consiste em quatro blocos densos, onde cada bloco contém um determinado número de camadas convolucionais. O bloco denso 1 possui 6 camadas convolucionais, o bloco denso 2 possui 12, o bloco denso 3 possui 24, e o bloco denso 4 possui 16. Cada camada convolucional é seguida por uma operação de *batch normalization*

Figura 16 – Exemplo de um Bloco Denso com 5 camadas.



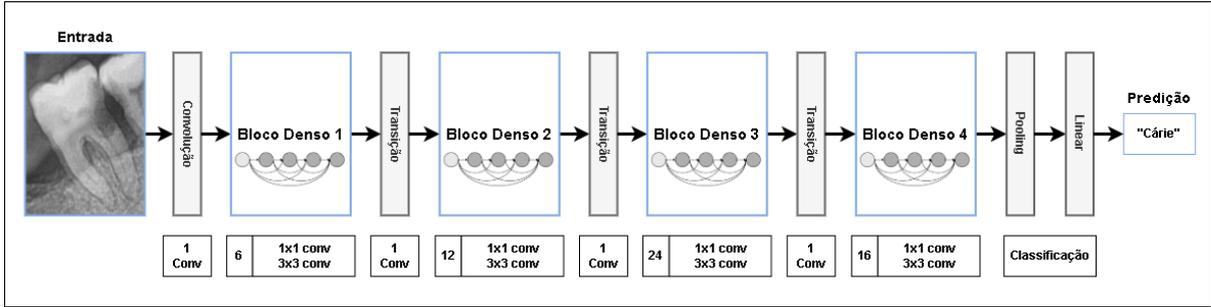
Fonte: (HUANG et al., 2017).

e uma função de ativação ReLU, antes da próxima convolução ser aplicada para extração de características.

A *batch normalization* tem como objetivo acelerar o treinamento de redes neurais profundas, por meio de uma etapa de normalização que ajusta as médias e variações das entradas da camada. Entre seus benefícios estão a redução do deslocamento interno de covariáveis (IOFFE; SZEGEDY, 2015), a diminuição da dependência dos gradientes em relação à escala dos parâmetros ou aos seus valores iniciais (IOFFE; SZEGEDY, 2015), além de atuar como uma forma de regularização do modelo, reduzindo a necessidade de técnicas adicionais como o *Dropout*, usado para mitigar problemas de *overfitting* (SRIVASTAVA et al., 2014)."

A DenseNet-121 utiliza convoluções 1x1 seguidas de *average pooling* 2x2 (responsável por calcular a média de cada região do mapa de características) como camadas de transição entre blocos adjacentes. Ao término do último bloco denso, há uma camada de *pooling* conectada a um classificador, conforme ilustrado na Figura 17.

Figura 17 – Arquitetura da DenseNet-121.



Fonte: Acervo do autor.

3.7 Otimização de Hiperparâmetros

A otimização de hiperparâmetros é o processo de busca pelo conjunto mais adequado de hiperparâmetros de um modelo, visando maximizar seu desempenho em um conjunto de validação (ANDONIE, 2019). Os hiperparâmetros diferem dos parâmetros internos do modelo, pois não são ajustados diretamente pelo algoritmo de aprendizado, mas definidos antes do treinamento. Por exemplo, em redes neurais, os hiperparâmetros incluem aspectos como a arquitetura da rede (número de camadas e neurônios) e a taxa de aprendizado, enquanto os pesos da rede são parâmetros aprendidos automaticamente. Esse processo de otimização pode ser formalmente representado pela seguinte equação,

$$x^* = \arg \min_{x \in \mathcal{N}} f(x), \quad (3.12)$$

onde $f(x)$ é uma função objetivo a ser minimizada, avaliada no conjunto de validação, como o erro quadrático médio (*root mean squared error*) ou a taxa de erro. Nessa formulação, x^* representa o conjunto de hiperparâmetros que minimiza $f(x)$, enquanto x pode assumir qualquer valor dentro do espaço de busca \mathcal{N} .

Essa equação pode ser expandida para considerar a otimização interna dos parâmetros aprendidos, conforme,

$$x^* = \arg \min_{x \in \mathcal{N}} f(x, y^*; S_{\text{validação}}), \quad (3.13)$$

onde a otimização interna encontra y^* , o conjunto ótimo de parâmetros aprendidos para um dado x :

$$y^* = \arg \min_{y \in \mathcal{T}} f(x, y; S_{\text{treino}}), \quad (3.14)$$

sendo $S_{validação}$ e S_{treino} os conjuntos de dados de validação e treinamento, respectivamente, e y os parâmetros ajustáveis do modelo, pertencentes ao espaço de busca τ (ANDONIE, 2019).

3.7.1 Otimização Bayesiana

A otimização bayesiana é uma abordagem eficiente para a escolha de hiperparâmetros em problemas nos quais a avaliação da função objetivo é computacionalmente custosa (FRAZIER, 2018). Diferentemente de métodos como *Random Search* e *Grid Search*, que exploram o espaço de busca sem considerar avaliações anteriores, a otimização bayesiana utiliza um modelo probabilístico para estimar quais configurações de hiperparâmetros têm maior probabilidade de resultar em um bom desempenho (ANDONIE, 2019).

Esse modelo probabilístico, chamado de modelo substituto (*surrogate model*), busca aproximar a função objetivo a partir dos dados observados. A cada nova avaliação da função objetivo, o modelo substituto é atualizado para refinar sua estimativa da distribuição posterior. O objetivo é balancear a exploração de regiões pouco testadas e a exploração intensiva em áreas promissoras, permitindo encontrar boas configurações de hiperparâmetros com menos avaliações diretas da função objetivo (KOEHRSEN, 2018).

A otimização bayesiana pode ser descrita formalmente como a modelagem da probabilidade de uma pontuação condicional à escolha dos hiperparâmetros:

$$p(\text{pontuação}|\text{hiperparâmetros}). \quad (3.15)$$

Esse processo ocorre de forma iterativa, seguindo as seguintes etapas (KOEHRSEN, 2018):

1. Criar um modelo probabilístico substituto da função objetivo;
2. Determinar os hiperparâmetros mais promissores com base no modelo substituto;
3. Avaliar esses hiperparâmetros na função objetivo real;
4. Atualizar o modelo substituto com os novos resultados;
5. Repetir os passos anteriores até atingir o limite de iterações ou tempo.

Com essa abordagem, a otimização bayesiana melhora a eficiência da busca por hiperparâmetros ideais e serve como base para metodologias avançadas, como *sequential model-based optimization* (SMBO), abordado a seguir.

3.7.2 Sequential Model-Based Global Optimization

O SMBO é uma formalização da otimização bayesiana, desenvolvida para resolver problemas de otimização nos quais a avaliação da função objetivo é computacionalmente custosa (ANDONIE, 2019). O termo *sequential* refere-se à execução iterativa de testes, onde cada nova avaliação busca aprimorar a seleção dos hiperparâmetros com base na atualização de um modelo probabilístico substituto.

O diferencial do SMBO em relação a abordagens como *random search* e *grid search* está na modelagem da relação entre hiperparâmetros e desempenho do modelo. Para isso, diferentes técnicas podem ser utilizadas na construção do modelo substituto, como: *gaussian processes*, *random forest regressions* e *tree-structured parzen estimators* (TPE).

3.7.3 Tree-structured Parzen Estimator

O TPE é uma variante do SMBO que se distingue pela forma como modela a função de busca por novos hiperparâmetros. Em vez de utilizar um modelo substituto tradicional, o TPE reformula a modelagem do espaço de busca por meio da distribuição condicional $p(y|x)$ e dos critérios de seleção dos próximos hiperparâmetros.

O TPE otimiza o critério de Melhoria Esperada (*Expected Improvement*, EI), que estima a melhoria esperada ao amostrar um novo ponto (JONES, 2001),

$$EI_{y^*}(x) := \int_{-\infty}^{\infty} \max(y^* - y, 0) p_M(y|x) dy. \quad (3.16)$$

No TPE, a distribuição condicional $p(y|x)$ é reformulada usando a regra de Bayes,

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)}, \quad (3.17)$$

A distribuição condicional $p(x|y)$ é modelada separadamente para diferentes intervalos de y :

$$p(x|y) = \begin{cases} l(x) & \text{se } y < y^* \\ g(x) & \text{se } y \geq y^*, \end{cases} \quad (3.18)$$

onde $l(x)$ representa a distribuição dos hiperparâmetros associados a menores valores da função objetivo e $g(x)$ àqueles associados a valores mais altos. Assim, a equação de EI pode ser reformulada como:

$$EI_{y^*}(x) = \frac{\gamma y^* l(x) - l(x) \int_{-\infty}^{y^*} p(y) dy}{\gamma l(x) + (1 - \gamma) g(x)} \propto \left(\gamma + \frac{g(x)}{l(x)} (1 - \gamma) \right)^{-1}. \quad (3.19)$$

Isso sugere que maximizar a melhoria esperada equivale a maximizar a razão entre as distribuições $l(x)$ e $g(x)$, tornando o TPE uma abordagem eficaz para a seleção de hiperparâmetros (KOEHRSEN, 2018).

3.8 Métricas de Avaliação

As métricas de avaliação são fundamentais para analisar o desempenho de um modelo, permitindo refinamentos contínuos. Na detecção de objetos, essas métricas avaliam a precisão das caixas delimitadoras previstas em comparação com as caixas do *ground truth*, de forma independente para cada classe, com base na sobreposição (PADILLA et al., 2021). Já na classificação de imagens, as métricas mensuram a capacidade do modelo de atribuir corretamente categorias às imagens, considerando aspectos como precisão, *recall* e F1 Score. Termos-chave incluem:

- **True Positive (TP)**: Na detecção de objetos, refere-se à detecção e localização corretas de um dente, incluindo sua respectiva numeração. Na classificação de imagens, é a atribuição correta de uma imagem à sua classe verdadeira.
- **False Positive (FP)**: Na detecção, refere-se à detecção incorreta ou posicionamento incorreto. Na classificação, é a atribuição incorreta de uma imagem a uma classe à qual ela não pertence.
- **True Negative (TN)**: Não aplicável na detecção de objetos devido ao número infinito de caixas não detectáveis em qualquer imagem (PADILLA; NETTO; SILVA, 2020). Na classificação de imagens, refere-se às imagens corretamente excluídas de uma classe à qual não pertencem.

- **False Negative (FN):** Na detecção, representa a falha em detectar um dente presente no *ground truth*. Na classificação, é o caso em que uma imagem não é atribuída à sua classe verdadeira.

Nas próximas subseções serão abordadas as métricas de precisão, *recall*, F1 Score, mAP50 e mAP50-95, considerando sua importância para a avaliação geral do desempenho do modelo. Essas métricas são calculadas com base nos valores de TP, FP, TN e FN apresentados anteriormente.

3.8.1 Precisão, Recall e F1 Score

Neste trabalho, as métricas precisão, *recall* e F1 Score são utilizadas para avaliar o desempenho dos modelos em ambas as etapas: de detecção e numeração de dentes e de classificação de patologias dentárias.

A precisão mede a capacidade do modelo de realizar previsões corretas em relação às detecções ou classificações realizadas. Na detecção de objetos, indica a proporção de dentes detectados corretamente entre todas as detecções realizadas. Na classificação, reflete a proporção de imagens corretamente atribuídas às suas classes em relação a todas as previsões feitas. O cálculo da precisão é expresso na Equação 3.20:

$$\text{Precisão} = \frac{TP}{TP + FP} \quad (3.20)$$

O *recall* mede a capacidade do modelo de identificar todas as instâncias relevantes. Na detecção, corresponde à proporção de dentes presentes no *ground truth* que foram corretamente detectados. Na classificação, refere-se à proporção de imagens de uma classe que foram corretamente atribuídas a essa classe. O cálculo do *recall* é apresentado na Equação 3.21:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3.21)$$

O F1 Score mede o equilíbrio entre precisão e *recall*, sendo particularmente útil quando há uma distribuição desigual entre classes ou quando tanto erros de detecção quanto de omissão são críticos. O cálculo do F1 Score é apresentado na Equação 3.22:

$$\text{F1 Score} = 2 \cdot \frac{\text{Precisão} \cdot \text{Recall}}{\text{Precisão} + \text{Recall}} \quad (3.22)$$

3.8.2 Média da Precisão Média (mAP)

A precisão média (*average precision*, AP) é uma métrica amplamente utilizada na avaliação de modelos de detecção de objetos. Para uma determinada classe, a AP corresponde à área sob a curva *precision-recall* (P-R), que é construída com base nas variações da precisão e do *recall* obtidas ao se alterar os limiares de confiança das detecções (ZAIDI et al., 2022). Quanto maior a AP, melhor o desempenho do modelo naquela classe, pois ela reflete a capacidade de manter alta precisão mesmo com maior cobertura de detecções.

O cálculo da AP depende da métrica IoU, que mede a sobreposição entre a caixa predita pelo modelo e a caixa de referência (*ground truth*). O IoU é definido como a razão entre a área de interseção e a área da união dessas caixas:

$$IoU = \frac{\text{Área de sobreposição}}{\text{Área de união}} \quad (3.23)$$

A média da precisão média (*mean average precision*, mAP) é então calculada pela média das APs obtidas para todas as classes avaliadas:

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (3.24)$$

Existem variações da mAP baseadas nos limiares de IoU utilizados para definir verdadeiros positivos:

- **mAP50**: Calcula a AP considerando um limiar fixo de IoU igual a 0,5. Detecções com $IoU \geq 0.5$ são consideradas corretas.
- **mAP50-95**: Corresponde à média das APs obtidas para múltiplos limiares de IoU, variando de 0.5 a 0.95, com incremento de 0.05 (ex.: 0.5, 0.55, ..., 0.95).

3.9 Considerações Finais

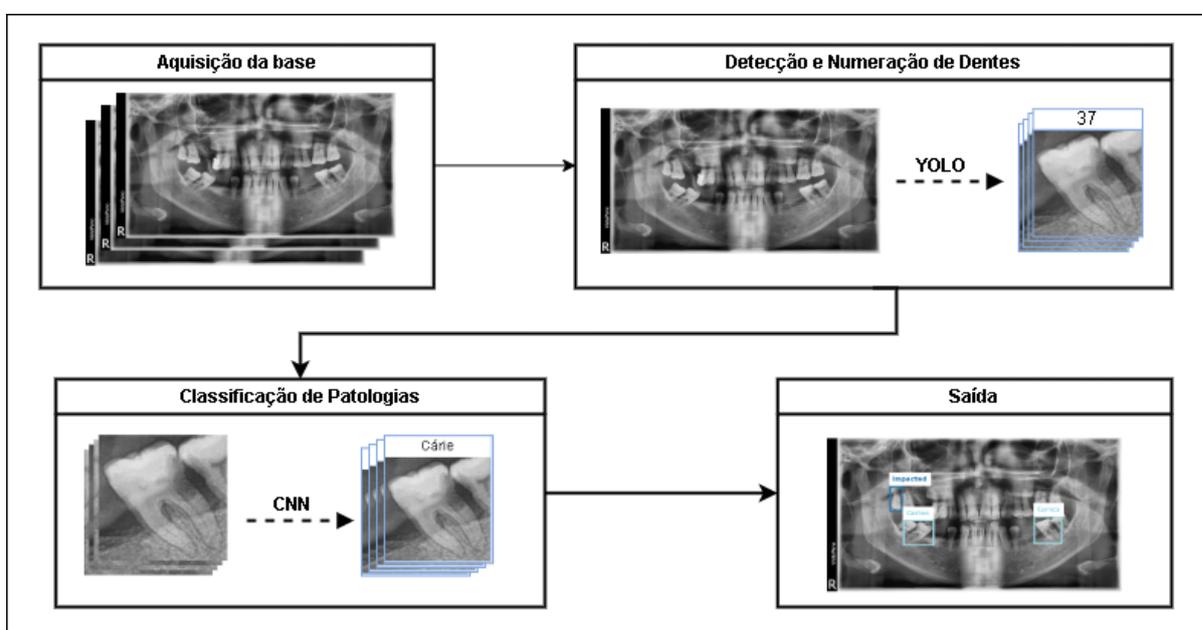
Neste capítulo, apresentamos os principais conceitos teóricos que fundamentam este trabalho, incluindo técnicas de processamento de imagens, aprendizado de máquina, funções de custo, arquiteturas de redes neurais e métodos de otimização de hiperparâmetros. Além disso, discutimos as anomalias dentárias consideradas, estabelecendo a base necessária para a compreensão dos métodos empregados.

No próximo capítulo, são descritos os materiais e o método utilizado na construção e avaliação da abordagem proposta.

4 MATERIAIS E MÉTODO

Este capítulo descreve detalhadamente os procedimentos e técnicas utilizados na condução deste trabalho, que tem como objetivo a detecção e numeração de dentes e a classificação de patologias em imagens de radiografias panorâmicas. O método proposto é dividido em duas etapas principais: a detecção de dentes nas imagens e a subsequente classificação das patologias associadas a cada dente. A Figura 18 apresenta o fluxo principal do método proposto. Ao decorrer deste capítulo serão descritos em detalhes cada etapa que compõe o método.

Figura 18 – Fluxo principal do método proposto.



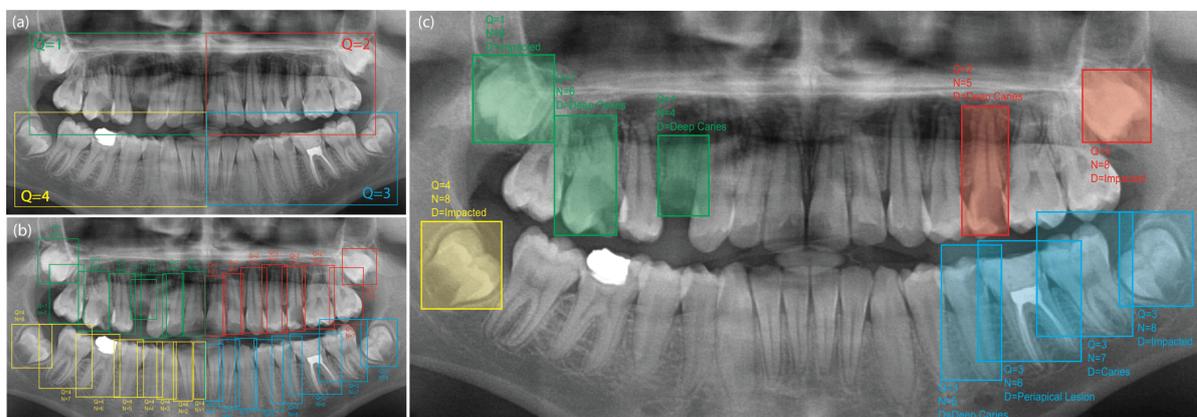
Fonte: Acervo do autor.

4.1 Aquisição da Base

Neste trabalho, foi utilizado o conjunto de dados DENTEX, que consiste em anotações hierárquicas para a detecção de dentes e anomalias em radiografias panorâmicas (HAMAMCI et al., 2023). As anotações foram realizadas por estudantes de odontologia e verificadas por dentistas, com o objetivo de minimizar vieses. O conjunto de dados é dividido em três categorias: a primeira contém dados parcialmente anotados para a numeração dos quadrantes anatômicos (Figura 19.a); a segunda contém dados parcialmente anotados para a numeração dos quadrantes e para a numeração e localização de todos

os dentes visíveis (Figura 19.b); e a terceira contém dados totalmente anotados com informações completas sobre quadrantes, numeração e localização de dentes anômalos (Figura 19.c).

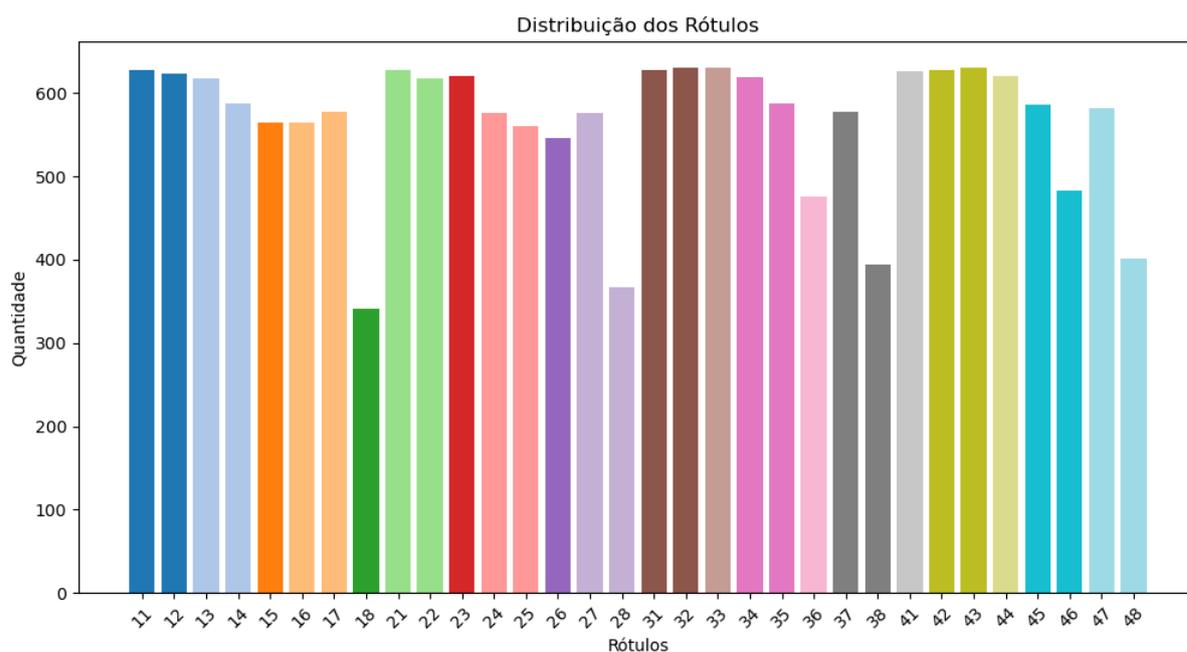
Figura 19 – Exemplo dos diferentes tipos de anotação disponíveis do conjunto de dados DENTEX.



Fonte: (HAMAMCI et al., 2023).

Para este trabalho, foram utilizadas as categorias (b) e (c) do conjunto de dados. A categoria (b) foi utilizada para a etapa de detecção e numeração de dentes. Esta categoria contém 637 imagens radiográficas, cujos rótulos estão distribuídas conforme o gráfico da Figura 20.

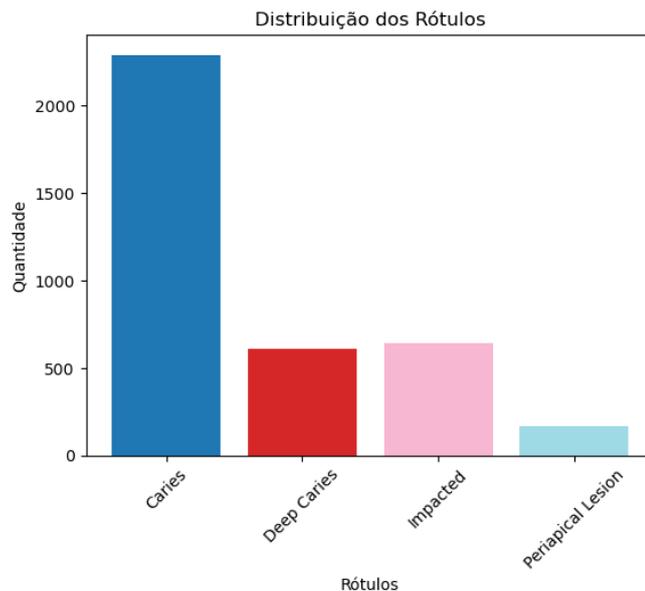
Figura 20 – Distribuição dos rótulos para a categoria b da base de dados DENTEX.



Fonte: Acervo do autor.

Já a categoria (c) deste conjunto de dados foi utilizada para gerar a base de treinamento para classificação de patologias orais, sendo elas: *Caries* (Cáries), *Deep Caries* (Cáries Profundas), *Impacted* (Impactado) e *Periapical Lesion* (Lesão Periapical). Esta categoria contém 755 imagens radiográficas, cujos rótulos estão distribuídas conforme o gráfico da Figura 21.

Figura 21 – Distribuição dos rótulos para a categoria c da base de dados DENTEX.



Fonte: Acervo do autor.

4.2 Detecção e Numeração de Dentes

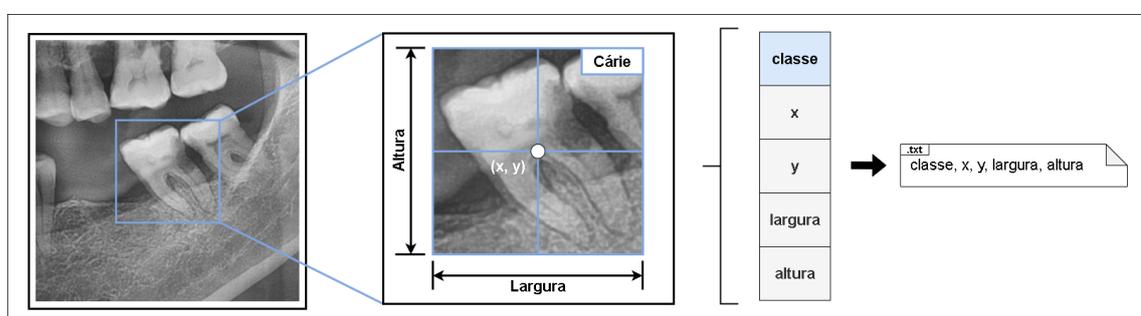
A detecção e numeração de dentes é a primeira etapa do método proposto, na qual os dentes presentes nas radiografias panorâmicas são localizados e numerados usando a notação FDI. Inicialmente, foi realizado um pré-processamento específico do conjunto de dados, com o objetivo de garantir maior qualidade e consistência nos resultados.

Na etapa de detecção, foi utilizado o modelo YOLOv8, cuja arquitetura inclui melhorias em relação a versões anteriores, como o bloco *Spatial Pyramid Pooling - Fast* (SPPF), que permite a extração eficiente de características em múltiplas escalas. Essa capacidade favorece a detecção de objetos pequenos, como os dentes em imagens panorâmicas.

4.2.1 Preparação e Pré-processamento dos Dados

Para garantir a compatibilidade do conjunto de dados DENTEX com o modelo YOLOv8, as informações foram convertidas para o formato apropriado. Esse formato exige que cada rótulo inclua a classe do objeto, representada pelo número do dente segundo a notação FDI, e as coordenadas normalizadas da caixa delimitadora, conforme ilustrado na Figura 22. A normalização das coordenadas é fundamental para que as informações sejam interpretadas corretamente pelo modelo, independentemente das dimensões originais da imagem.

Figura 22 – Formato de anotações para cada rótulo.



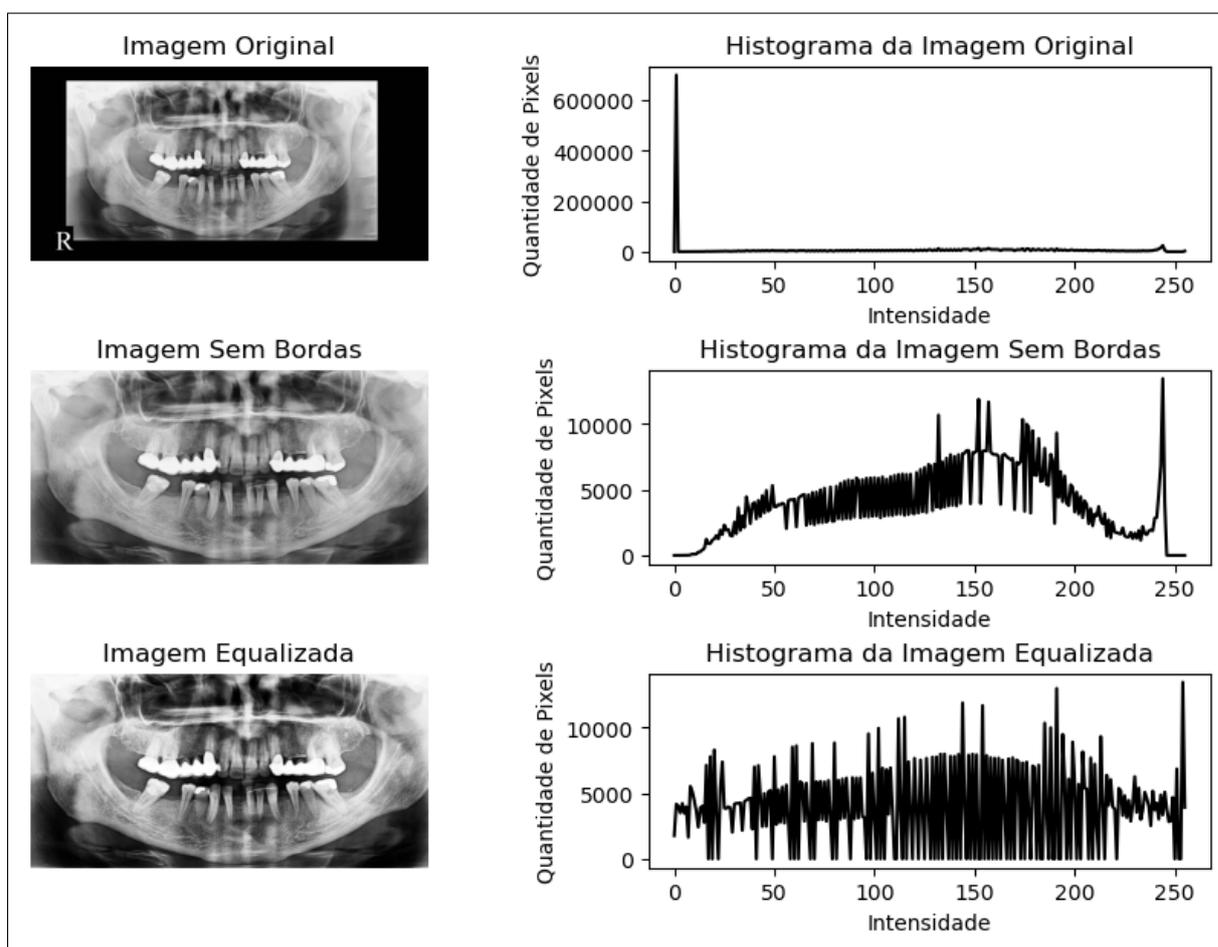
Fonte: Acervo do autor.

Durante a inspeção do conjunto de dados, identificou-se a presença de bordas pretas em algumas imagens, assim como marcas indicativas de dispositivos utilizados no processo de aquisição das radiografias. Essas características podem introduzir ruído, prejudicando o treinamento do modelo. Para solucionar esse problema, foi implementado um algoritmo de detecção de bordas que percorreu a matriz de pixels da imagem de cima para baixo, de baixo para cima, da esquerda para a direita e da direita para a esquerda. A cada iteração, os pixels foram analisados em busca do primeiro valor cuja intensidade ultrapassasse um limiar predefinido, definido empiricamente com base na aplicação do algoritmo ao conjunto de treinamento e na avaliação qualitativa das saídas. Esse limiar permitiu diferenciar regiões relevantes da imagem de possíveis fundos pretos. Os primeiros pixels detectados em cada uma dessas direções foram registrados como limites de corte, permitindo a remoção precisa das bordas indesejadas.

Com a remoção das bordas, tornou-se necessário ajustar as coordenadas das caixas delimitadoras para que permanecessem alinhadas com as novas dimensões das imagens. Esse ajuste foi realizado recalculando os valores das coordenadas com base nas proporções da imagem após o recorte.

Além do recorte, foi aplicada a equalização do histograma em todas as imagens. A equalização do histograma é especialmente útil em imagens radiográficas, que frequentemente apresentam variações de contraste que dificultam a identificação de estruturas específicas (KRUTSCH; TENORIO, 2011). Esse método foi escolhido para melhorar o contraste e a distribuição da intensidade, tornando os detalhes mais evidentes para o modelo durante o treinamento. Esse efeito pode ser observado na Figura 23, que apresenta um exemplo de imagem antes e depois da aplicação do pré-processamento.

Figura 23 – Exemplo do antes e depois do pré-processamento de uma imagem do conjunto de dados.



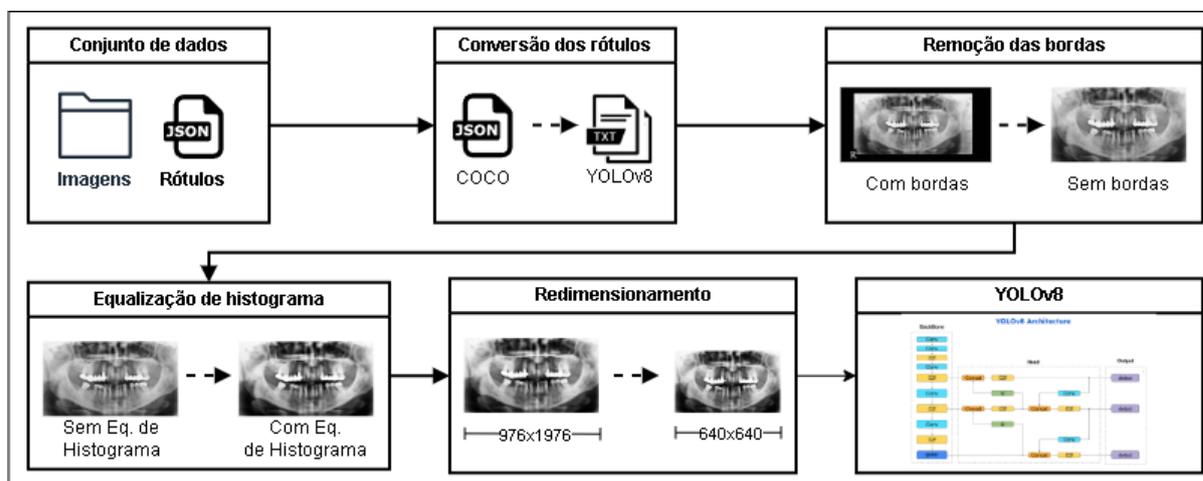
Fonte: Acervo do autor.

Por fim, todas as imagens foram redimensionadas para 640x640 pixels, um requisito para a entrada no YOLOv8. Esse redimensionamento garante uniformidade no conjunto de dados e facilita o processamento em lotes (*batches*) durante o treinamento. Essa padronização também permite que o modelo mantenha consistência na interpretação das proporções e tamanhos dos objetos dentro das imagens.

A Figura 24 apresenta um diagrama ilustrando o fluxo completo de preparação

e pré-processamento aplicado ao conjunto de dados. Essas etapas foram essenciais para assegurar que as imagens estivessem adequadamente formatadas e livres de artefatos que pudessem interferir na precisão da detecção e numeração dos dentes.

Figura 24 – Fluxo completo do processo de detecção e numeração de dentes.



Fonte: Acervo do autor.

4.3 Classificação de Patologias

A classificação de patologias orais é a etapa final do método proposto, responsável por identificar e categorizar as condições presentes nos dentes detectados previamente. Nesta fase, as imagens individuais dos dentes, recortadas a partir das *bounding boxes* geradas na etapa de detecção, foram utilizadas como entrada para o treinamento e a avaliação das arquiteturas CNNs. A abordagem foca na detecção de quatro principais condições: dentes impactados, cárie (incluindo cáries profundas, que foram unificadas em uma única classe), e lesões periapicais.

4.3.1 Preparação e Pré-processamento dos Dados

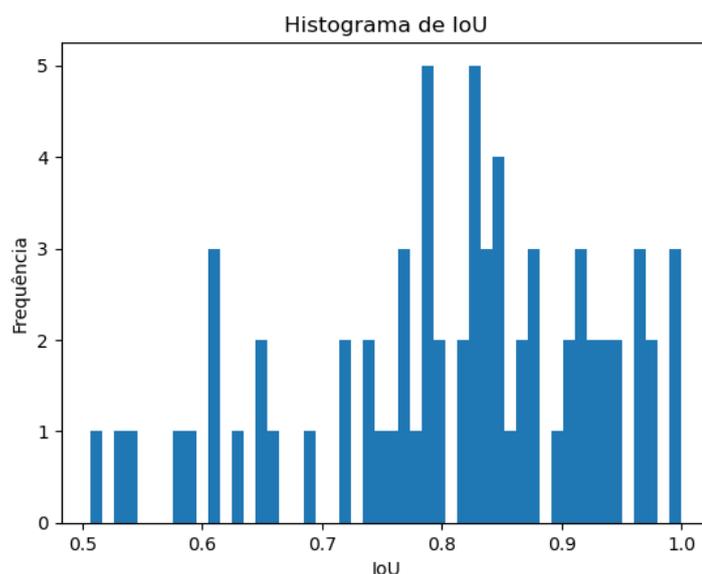
Para a etapa de classificação de patologias orais, foi criado um novo conjunto de dados com base nas *bounding boxes* e nas classes da categoria (c) do conjunto DENTEX. A partir das 755 imagens originalmente disponíveis, foi gerada uma nova base contendo 3.711 imagens de dentes anotadas para quatro diferentes patologias. A Tabela 2 mostra a distribuição de imagens para cada patologia do conjunto de dados gerado.

Após a criação do novo conjunto de dados, observou-se a ocorrência de alguns casos em que mais de uma *bounding box* era associada ao mesmo dente, porém com rótulos

Tabela 2 – Distribuição das imagens por classe no conjunto de dados.

Patologia	Instâncias
Impactado	644
Cárie	2.290
Lesão Periapical	167
Cárie Profunda	610
Total	3.711

diferentes. Para garantir a consistência do problema como multiclasse, as imagens que apresentavam esses conflitos foram descartadas. A identificação desses casos foi realizada utilizando o índice de IoU para cada par de *bounding boxes*, considerando como conflito os pares com IoU superior a 0,5. A Figura 25 apresenta um gráfico de frequência desses casos em função do IoU.

Figura 25 – Frequência de conflito de *bounding boxes* no novo conjunto de dados.

Fonte: Acervo do autor.

Além disso, as classes Cárie e Cárie Profunda foram unificadas em um único rótulo denominado Cárie. Essa decisão foi baseada no entendimento de que ambas representam o mesmo processo patológico de desmineralização do tecido dentário, diferenciando-se apenas pela profundidade da lesão (BJØRNDAL, 2008). Essa unificação também contribuiu para simplificar o problema de classificação, especialmente em situações onde as categorias menos representadas poderiam impactar negativamente o desempenho do modelo. Do ponto de vista clínico, essa abordagem não compromete a tomada de decisão do especialista, uma vez que a classificação automática serve como apoio inicial, cabendo ao profissional

avaliar a extensão da lesão e definir a conduta adequada a partir da imagem original e do contexto do paciente.

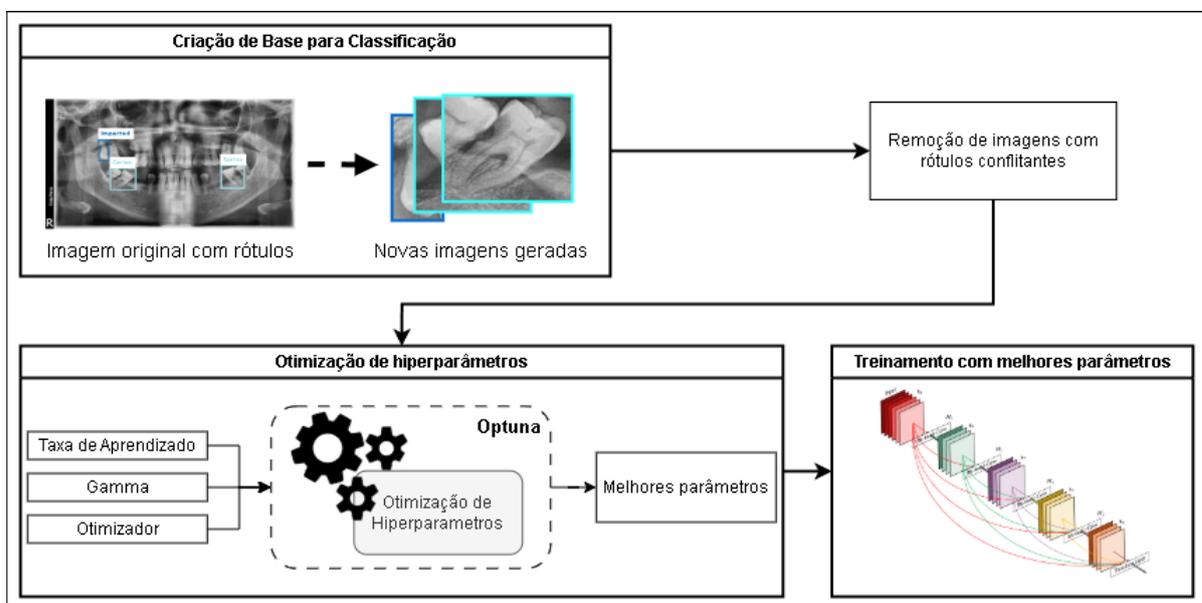
A Tabela 3 apresenta a distribuição final das imagens no conjunto de dados após a remoção de imagens conflituosas e a junção das classes Cárie e Cárie Profunda.

Tabela 3 – Distribuição final das imagens por classe no conjunto de dados após os ajustes.

Patologia	Instâncias
Impactado	583
Cárie	2.551
Lesão Periapical	107
Total	3.241

A Figura 26 apresenta um diagrama ilustrando o fluxo completo de preparação e pré-processamento aplicado ao conjunto de dados para esta etapa de classificação de patologias. Esses ajustes no conjunto de dados foram essenciais para assegurar a consistência das anotações e aumentar a robustez do modelo na classificação das patologias orais.

Figura 26 – Fluxo completo do processo de treinamento e avaliação para a classificação de anomalias dentárias.



Fonte: Acervo do autor.

4.3.2 Otimização de Hiperparâmetros

A otimização de hiperparâmetros foi realizada utilizando o algoritmo TPE, implementado em Akiba et al. (2019), com o objetivo de maximizar a métrica F1 Score

ponderada por classe. Esse processo foi aplicado a diversas arquiteturas de redes neurais profundas, incluindo MobileNetV2 (SANDLER et al., 2018), VGG16 (SIMONYAN, 2014), DenseNet121 (HUANG et al., 2017), ResNet50 (HE et al., 2016), InceptionV3 (SZEGEDY et al., 2016), Xception (CHOLLET, 2017), EfficientNet (B5 - B7) (TAN; LE, 2019). A seleção dessas arquiteturas foi baseada em sua recorrência em estudos anteriores envolvendo imagens médicas, incluindo aplicações específicas em radiografias odontológicas. Três espaços de busca de hiperparâmetros foram definidos:

1. **Otimização do Otimizador:** Um espaço categórico foi utilizado, permitindo a seleção entre Adam (KINGMA, 2014), RMSprop (HINTON; SRIVASTAVA; SWERSKY, 2012) e SGD (RUDER, 2016).
2. **Taxa de Aprendizado (*Learning Rate*):** Um espaço contínuo foi configurado para valores no intervalo $[0,00001; 0,01]$, com uma escala logarítmica para a sugestão de valores.
3. **Parâmetro Gamma da FL:** Valores contínuos no intervalo $[0,5; 5,0]$ foram explorados, seguindo um padrão logarítmico na escolha.

Para cada arquitetura, foram realizados 50 *trials*, nos quais os hiperparâmetros sugeridos foram avaliados com base no F1 Score ponderada por classe. Em cada *trial*, os modelos foram treinados por 50 épocas, com *Early Stopping* monitorando a F1 Score ponderada, utilizando uma paciência de 5 épocas, e um tamanho de *batch* de 16. Ao final do treinamento, o melhor peso foi recuperado e salvo para análises e reprodutibilidade. Os resultados de cada *trial*, incluindo os valores de F1 Score e os hiperparâmetros utilizados, foram registrados em arquivos CSV e no Weight & Biases (W&B) (BIEWALD, 2020) para organização e posterior exploração. Esse processo garantiu uma busca eficiente pelos melhores hiperparâmetros, maximizando o desempenho dos modelos nas tarefas de classificação de patologias orais.

5 RESULTADOS E DISCUSSÃO

Neste capítulo, serão apresentados os resultados obtidos em cada um dos experimentos realizados neste trabalho. Todos os experimentos foram conduzidos utilizando a linguagem de programação Python (ROSSUM; JR, 1995), versão 3.8. A plataforma CUDA (NVIDIA; VINGELMANN; FITZEK, 2020) foi adotada para dar suporte à execução do treinamento em uma GPU NVIDIA GeForce GTX 1660 Super equipada com aproximadamente 6GB de VRAM.

5.1 Validação Cruzada k -fold

A validação cruzada é um método de reamostragem de dados utilizado para avaliar a capacidade de generalização dos modelos e evitar o *overfitting*. É um dos métodos mais amplamente utilizadas para seleção e avaliação de modelos em aprendizado de máquina (BERRAR et al., 2019). Existem diferentes tipos de validação cruzada. Neste trabalho, optou-se por usar a validação cruzada k -fold.

O método de validação cruzada k -fold envolve a divisão cíclica do conjunto de amostras em uma série de partes de tamanho igual, onde k indica o número de partições (ou *folds*) em que o conjunto de dados é dividido (RAMEZAN; WARNER; MAXWELL, 2019). O modelo é treinado em $k - 1$ dessas partes e validado na parte restante. Esse processo é repetido k vezes, alternando a parte de validação, garantindo que cada parte do conjunto de dados seja usada uma vez como validação e $k - 1$ vezes como treinamento. Por fim, a média dos resultados é calculada (STONE, 1974). A Figura 27 apresenta um exemplo de fluxo validação cruzada k -fold.

Figura 27 – Validação Cruzada k -fold.



Fonte: Acervo do autor.

5.2 Resultados da Detecção de Dentes

Nesta seção, são apresentados os resultados quantitativos e estudos de caso relacionados à etapa de detecção e numeração de dentes em imagens de radiografias panorâmicas.

5.2.1 Resultados Quantitativos

Foram conduzidos experimentos com o objetivo de analisar o desempenho do modelo YOLOv8 aplicado à tarefa de detecção de dentes em imagens de radiografias panorâmicas. Este modelo foi treinado e avaliado no conjunto de dados DENTEX. As métricas padrão adotadas pelo YOLOv8 utilizadas para avaliar o desempenho do modelo foram a precisão, *recall*, mAP50 e mAP50-95.

Neste experimento, foram utilizadas 637 imagens, das quais 507 foram distribuídas em cinco conjuntos (*folds*) para os processos de treinamento e validação do modelo. As 127 imagens restantes foram separadas exclusivamente para compor o conjunto de teste, garantindo que esse conjunto fosse independente dos dados utilizados no treinamento e validação. Essa abordagem foi aplicada a quatro versões diferentes do YOLOv8. Os modelos foram treinados por 100 épocas com um tamanho de lote de 16, removendo os *augmentations* automáticos e restaurando os melhores pesos ao final de cada treinamento. A Tabela 4 apresenta os resultados obtidos na validação cruzada com *k*-fold.

Tabela 4 – Os valores são apresentados como médias seguidas de desvios-padrão entre parênteses, obtidos por validação cruzada *k*-fold.

Modelo	Precision	Recall	mAP50	mAP50-95
YOLOv8s	0,8876 ($\pm 0,0105$)	0,8930 ($\pm 0,0105$)	0,9288 ($\pm 0,0106$)	0,5245 ($\pm 0,0161$)
YOLOv8m	0,9095 ($\pm 0,0153$)	0,9160 ($\pm 0,0109$)	0,9422 ($\pm 0,0113$)	0,5527 ($\pm 0,0157$)
YOLOv8l	0,9174 ($\pm 0,0095$)	0,9283 ($\pm 0,0141$)	0,9465 ($\pm 0,0098$)	0,5644 ($\pm 0,0156$)
YOLOv8x	0,9212 ($\pm 0,0117$)	0,9296 ($\pm 0,0125$)	0,9468 ($\pm 0,0116$)	0,5657 ($\pm 0,0154$)

A análise dos resultados na Tabela 4 demonstra um desempenho crescente entre as versões YOLOv8s, YOLOv8m, YOLOv8l e YOLOv8x na tarefa de detecção de dentes em imagens de radiografias panorâmicas. A precisão aumenta de 0,8876 ($\pm 0,0105$) no YOLOv8s para 0,9212 ($\pm 0,0117$) no YOLOv8x, indicando uma melhoria na identificação correta dos dentes. O *recall* também aumenta de 0,8930 ($\pm 0,0105$) para 0,9296 ($\pm 0,0125$), sugerindo que as versões mais avançadas recuperam mais instâncias relevantes. A métrica mAP50 segue essa tendência, subindo de 0,9288 ($\pm 0,0106$) para 0,9468 ($\pm 0,0116$). A

métrica mAP50-95, que avalia uma faixa mais ampla de valores de IoU, melhora de 0,5245 ($\pm 0,0161$) para 0,5657 ($\pm 0,0154$). Esses resultados sugerem que as versões maiores do YOLOv8 oferecem maior precisão, *recall* e desempenho mAP, refletindo maior robustez e eficácia na detecção e numeração de dentes em imagens de radiografias panorâmicas.

Após a análise desses resultados, os modelos obtidos em cada um dos cinco folds da validação cruzada foram testados individualmente no conjunto de teste separado, que não foi incluído no treinamento. Essa avaliação adicional teve como objetivo medir a capacidade de generalização dos modelos YOLOv8 na detecção de dentes em imagens de radiografias panorâmicas.

Os resultados apresentados na Tabela 5 mostram que a versão YOLOv8x mantém o melhor desempenho geral, com os valores mais altos para precisão ($0,9283 \pm 0,0035$), *recall* ($0,9327 \pm 0,0021$), mAP50 ($0,9450 \pm 0,0024$) e mAP50-95 ($0,5781 \pm 0,0040$). Comparando as versões, nota-se que o YOLOv8l e o YOLOv8m também demonstram desempenho robusto, superando a versão YOLOv8s em todas as avaliações realizadas. A consistência dos baixos desvios padrão entre os modelos reforça a confiabilidade da precisão. Além disso, os modelos YOLOv8, especialmente as versões maiores, apresentam uma forte capacidade de generalização na tarefa de detecção de dentes.

Tabela 5 – Os valores são apresentados como médias seguidas de desvios-padrão entre parênteses, obtidos de um conjunto de teste separado.

Modelo	Precision	Recall	mAP50	mAP50-95
YOLOv8s	0,8884 ($\pm 0,0062$)	0,9016 ($\pm 0,0059$)	0,9294 ($\pm 0,0032$)	0,5370 ($\pm 0,0024$)
YOLOv8m	0,9174 ($\pm 0,0038$)	0,9221 ($\pm 0,0047$)	0,9422 ($\pm 0,0027$)	0,5663 ($\pm 0,0028$)
YOLOv8l	0,9260 ($\pm 0,0033$)	0,9286 ($\pm 0,0041$)	0,9432 ($\pm 0,0024$)	0,5763 ($\pm 0,0020$)
YOLOv8x	0,9283 ($\pm 0,0035$)	0,9327 ($\pm 0,0021$)	0,9450 ($\pm 0,0024$)	0,5781 ($\pm 0,0040$)

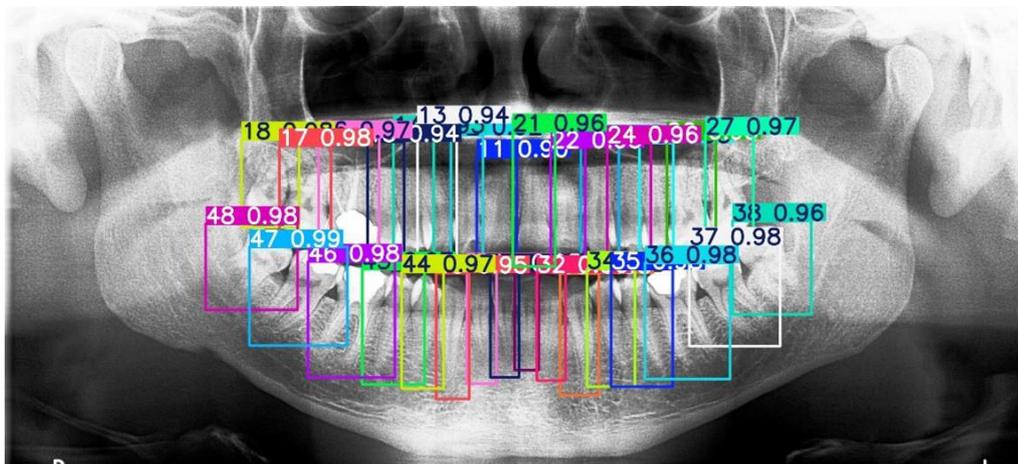
5.2.2 Estudos de Caso

Foram realizados alguns estudos de caso para avaliar qualitativamente o modelo YOLOv8x na detecção de dentes em imagens de radiografias panorâmicas. Esses estudos fornecem uma análise visual do desempenho do modelo, destacando casos específicos de detecções corretas, falsos positivos e falsos negativos.

Um exemplo do modelo YOLOv8x detectando com sucesso todos os dentes em uma radiografia panorâmica é apresentado na Figura 28. A figura exhibe as caixas delimitadoras correspondentes, juntamente com o número do dente na notação FDI e a confiança, que representa a probabilidade de a detecção estar correta. As caixas delimitadoras encapsulam

com precisão cada dente, demonstrando a capacidade do modelo de localizar e identificar as estruturas dentárias com precisão.

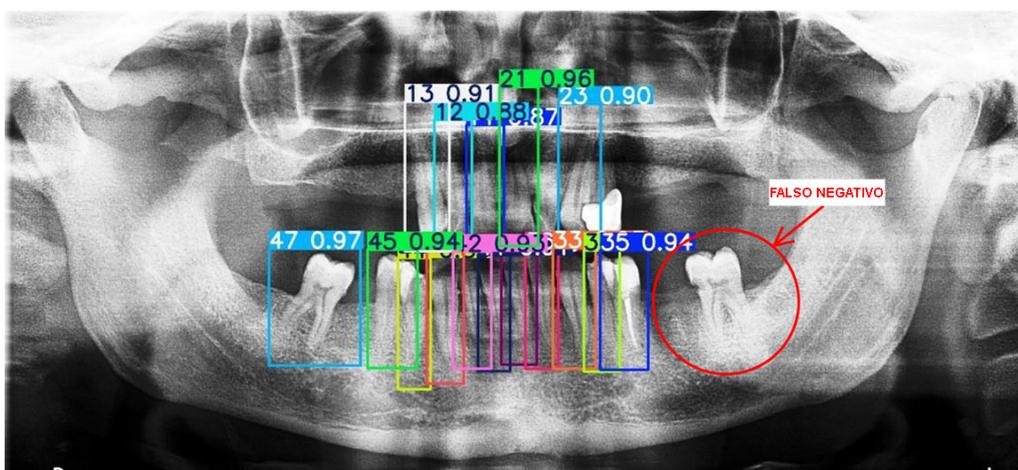
Figura 28 – Exemplo de detecções corretas pelo modelo YOLOv8x.



Fonte: Acervo do autor.

Exemplos de falsos negativos, nos quais o modelo YOLOv8x falhou em detectar determinados dentes, é apresentado na Figura 29. Esses casos frequentemente envolvem dentes com menor contraste ou parcialmente obscurecidos por outras estruturas, indicando desafios potenciais para o modelo em cenários específicos.

Figura 29 – Exemplo de falso negativo em que o modelo YOLOv8x perdeu detecções.



Fonte: Acervo do autor.

Esses resultados qualitativos obtidos nos estudos de caso apoiam os achados quantitativos, demonstrando a eficácia dos modelos YOLOv8 na detecção de dentes em imagens de radiografias panorâmicas. Exemplos visuais ilustram a alta precisão do modelo na identificação de estruturas dentárias, evidenciada por dentes corretamente detectados

com caixas delimitadoras precisas e altos índices de confiança. Além disso, os falsos negativos ressaltam a importância de um refinamento contínuo do modelo.

5.3 Resultados da Classificação de Patologias

Nesta seção, são apresentados os resultados da otimização de hiperparâmetros, seguidos dos resultados quantitativos e estudos de caso dos experimentos relacionados à etapa de classificação de patologias dentárias em imagens de radiografias panorâmicas.

5.3.1 Otimização de Hiperparâmetros

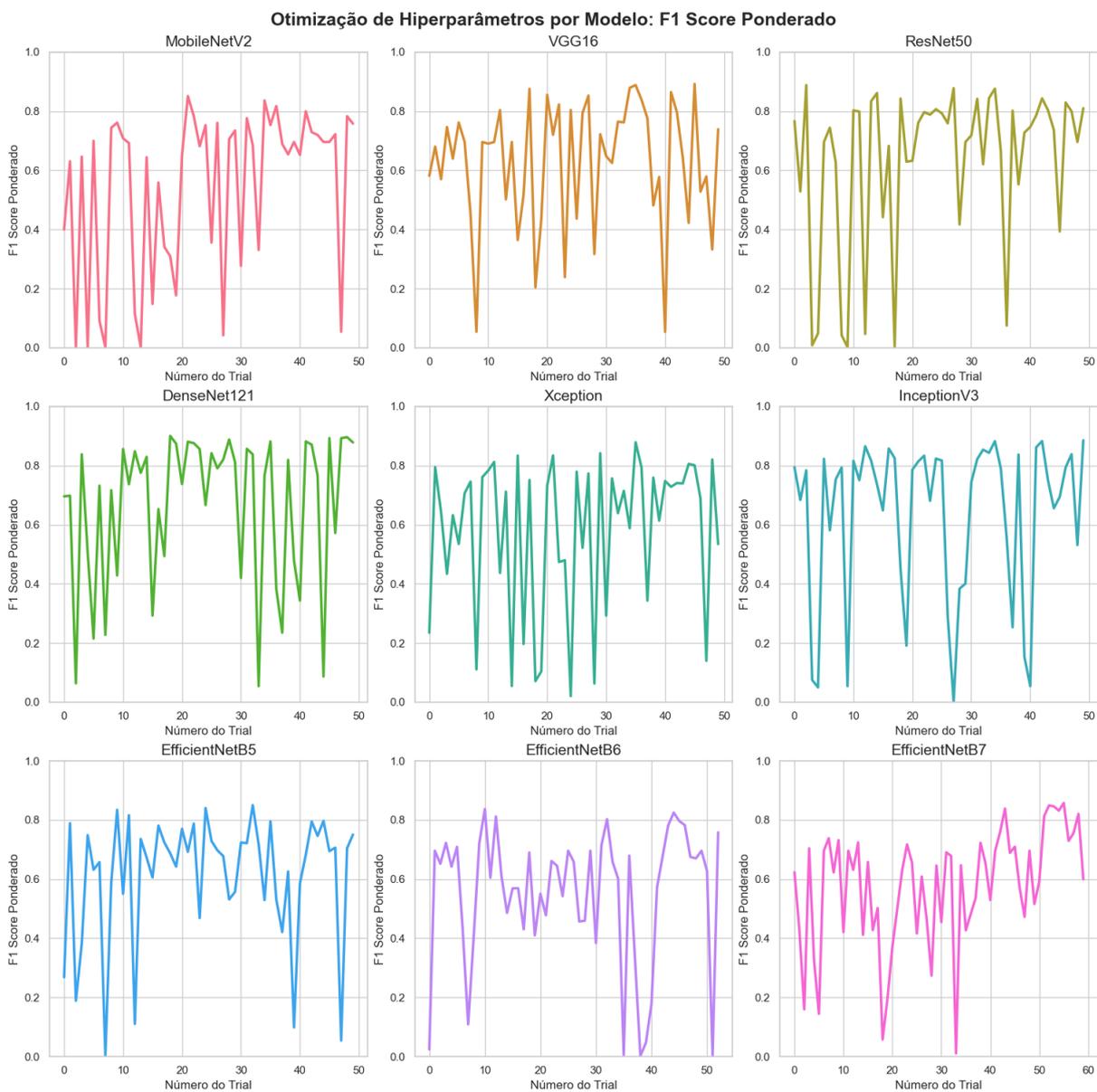
Para a etapa de classificação de patologias dentárias, foi realizada uma otimização de hiperparâmetros para diferentes arquiteturas CNNs. A otimização foi conduzida utilizando o método TPE, com o objetivo de maximizar a métrica F1 Score ponderada por classe. Os gráficos da Figura 30 ilustram a evolução da métrica F1 Score ponderada ao longo dos 50 *trials* para cada arquitetura, permitindo uma análise visual da convergência e desempenho do processo de otimização.

A Tabela 6 destaca as diferenças nos hiperparâmetros ótimos para cada arquitetura de rede neural. Por exemplo, o modelo DenseNet121 apresentou melhor desempenho com uma taxa de aprendizado de 0,00011 e utilizou o otimizador Adam, enquanto a EfficientNetB5 obteve melhores resultados com uma taxa de aprendizado significativamente maior de 0,00965, utilizando o otimizador SGD. Essas variações nos hiperparâmetros ótimos reforçam a importância de personalizar a configuração para cada modelo, pois as características intrínsecas das arquiteturas influenciam suas necessidades de configuração.

Tabela 6 – Valores dos hiperparâmetros ótimos para arquitetura avaliada.

Modelo	Gamma	Taxa de Aprendizado	Otimizador
MobileNetV2	0,58461	0,00001	Adam
VGG16	0,92059	0,00001	RMSprop
DenseNet121	2,21633	0,00011	Adam
ResNet50	1,09426	0,00002	Adam
InceptionV3	1,08632	0,00004	Adam
Xception	0,57655	0,00010	Adam
EfficientNetB5	2,84346	0,00965	SGD
EfficientNetB6	1,10834	0,00877	SGD
EfficientNetB7	2,20676	0,00006	Adam

Figura 30 – Histórico das rodadas (*trials*) da otimização dos hiperparâmetros dos modelos.



Fonte: Acervo do autor.

5.3.2 Resultados Quantitativos

A avaliação quantitativa da classificação foi realizada por meio de validação cruzada. Para isso, um total de 3.078 imagens foi distribuído entre cinco conjuntos (*folds*), que foram utilizados nos processos de treinamento e validação do modelo. As 163 imagens restantes foram separadas de forma exclusiva para compor o conjunto de teste, garantindo que esse conjunto fosse independente dos dados utilizados no treinamento e validação. A Tabela 7 apresenta as principais métricas de desempenho, incluindo F1 Score, precisão e *recall* ponderados por classe positiva, obtidos no conjunto de validação de cada *fold* para cada uma das arquiteturas avaliadas.

Tabela 7 – Métricas de avaliação obtidas na validação cruzada utilizando 5 *folds*.

Modelo	Precision	Recall	F1 Score
MobileNetV2	0,8209 ($\pm 0,0337$)	0,7583 ($\pm 0,0864$)	0,7582 ($\pm 0,0563$)
VGG16	0,8694 ($\pm 0,0260$)	0,7677 ($\pm 0,0744$)	0,7879 ($\pm 0,0687$)
DenseNet121	0,8749 ($\pm 0,0265$)	0,8625 ($\pm 0,0241$)	0,8546 ($\pm 0,0257$)
ResNet50	0,8535 ($\pm 0,0182$)	0,8622 ($\pm 0,0214$)	0,8462 ($\pm 0,0291$)
InceptionV3	0,8623 ($\pm 0,0265$)	0,8437 ($\pm 0,0385$)	0,8422 ($\pm 0,0319$)
Xception	0,8861 ($\pm 0,0334$)	0,8281 ($\pm 0,0447$)	0,8330 ($\pm 0,0484$)
EfficientNetB5	0,8806 ($\pm 0,0119$)	0,7621 ($\pm 0,0611$)	0,8037 ($\pm 0,0415$)
EfficientNetB6	0,8738 ($\pm 0,0274$)	0,7079 ($\pm 0,0786$)	0,7609 ($\pm 0,0619$)
EfficientNetB7	0,8878 ($\pm 0,0197$)	0,6160 ($\pm 0,1401$)	0,6889 ($\pm 0,1224$)

Os resultados apresentados na Tabela 7 indicam que dentre os modelos comparados, o DenseNet121 destacou-se como o mais equilibrado, atingindo os melhores resultados em *recall* ($0,8625 \pm 0,0241$) e F1 Score ($0,8546 \pm 0,0257$), o que demonstra sua eficiência tanto na identificação correta das classes positivas quanto no balanceamento entre precisão e *recall*. Em termos de precisão, o EfficientNetB7 apresentou o maior valor ($0,8878 \pm 0,0197$), entretanto, seu *recall* ($0,6160 \pm 0,1401$) e F1 Score ($0,6889 \pm 0,1224$) ficaram significativamente abaixo de outros modelos, indicando que sua alta precisão foi obtida às custas de uma menor cobertura. Modelos como VGG16, ResNet50 e InceptionV3 também apresentaram desempenho competitivo, mas com F1 Score inferior ao DenseNet121. A consistência dos seus resultados é refletida pelos baixos desvios padrão, sugerindo que este modelo mantém um desempenho estável em diferentes divisões dos dados.

Para avaliar o desempenho geral e a capacidade de generalização dos modelos, cada um dos cinco modelos treinados durante a validação cruzada foi testado no mesmo conjunto de testes separado, contendo 163 imagens. Essa análise permitiu verificar a

estabilidade e consistência das arquiteturas em diferentes divisões dos dados.

A Tabela 8 evidencia que o modelo EfficientNetB7 atingiu a maior precisão ($0,8812 \pm 0,0244$) no conjunto de teste, enquanto DenseNet 121 apresentou o melhor *recall* ($0,8380 \pm 0,0490$). InceptionV3 destacou-se com o F1 Score mais elevado ($0,8307 \pm 0,0337$), sugerindo um equilíbrio entre precisão e *recall*. A análise dos desvios padrão indica uma consistência robusta nos resultados, reforçando a confiabilidade dos modelos na tarefa de classificação em diferentes cenários de teste.

Tabela 8 – Métricas de desempenho obtidas no conjunto de teste separado.

Modelo	Precisão	Recall	F1 Score
MobileNetV2	0,8142 ($\pm 0,0319$)	0,7668 ($\pm 0,0572$)	0,7503 ($\pm 0,0457$)
VGG16	0,8592 ($\pm 0,0441$)	0,7742 ($\pm 0,0981$)	0,7872 ($\pm 0,1027$)
DenseNet 121	0,8505 ($\pm 0,0385$)	0,8380 ($\pm 0,0490$)	0,8263 ($\pm 0,0531$)
ResNet 50	0,8375 ($\pm 0,0254$)	0,8319 ($\pm 0,0320$)	0,8099 ($\pm 0,0457$)
InceptionV3	0,8477 ($\pm 0,0285$)	0,8343 ($\pm 0,0385$)	0,8307 ($\pm 0,0337$)
Xception	0,8665 ($\pm 0,0304$)	0,8184 ($\pm 0,0345$)	0,8152 ($\pm 0,0438$)
EfficientNetB5	0,8716 ($\pm 0,0184$)	0,7926 ($\pm 0,0441$)	0,8226 ($\pm 0,0302$)
EfficientNetB6	0,8619 ($\pm 0,0087$)	0,7435 ($\pm 0,0593$)	0,7834 ($\pm 0,0457$)
EfficientNetB7	0,8812 ($\pm 0,0244$)	0,6625 ($\pm 0,1091$)	0,7276 ($\pm 0,0904$)

A Tabela 9 compara o desempenho da YOLOv8x, que executa a detecção e classificação das patologias dentárias, com o método de duas etapas proposto neste trabalho, no qual a YOLOv8x detecta os dentes e a DenseNet121 realiza a classificação das patologias. Os resultados mostram que o método de duas etapas superou a YOLOv8x em todas as métricas, alcançando F1 Score ponderado de 0,6014, em comparação com 0,433 quando a YOLOv8x executa ambas as tarefas diretamente. Esses resultados sugerem que dividir o método em duas etapas pode melhorar significativamente a identificação de patologias em imagens de radiografia panorâmica.

Tabela 9 – Comparação entre a YOLOv8x realizando localização e classificação diretamente e a abordagem YOLOv8x + DenseNet121.

Modelo	Precisão	Recall	F1 Score
YOLOv8x	0,448	0,451	0,433
YOLOv8x + DenseNet121	0,6109	0,6049	0,6014

5.3.3 Estudos de Caso

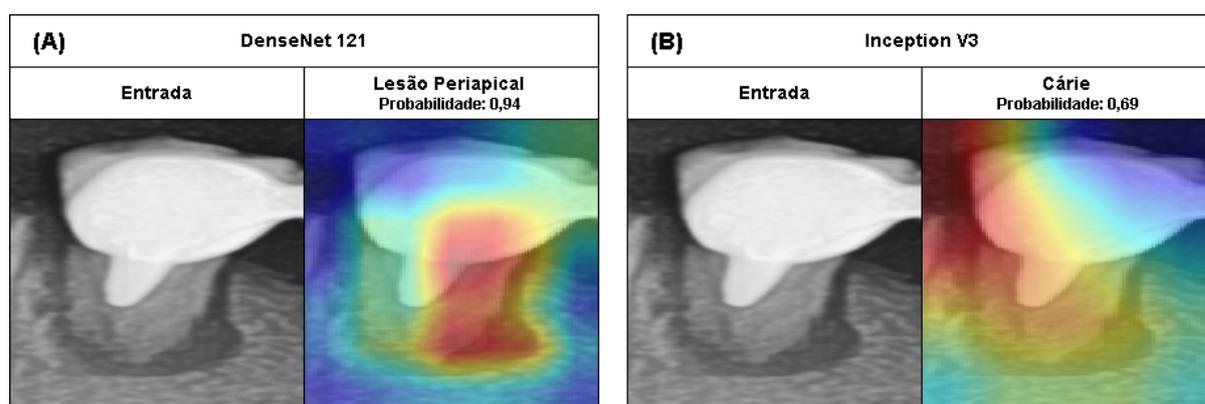
Para avaliar qualitativamente a etapa de classificação e a qualidade das previsões realizadas pelos modelos treinados, foram realizados estudos de caso utilizando a técnica

Gradient-weighted Class Activation Mapping (Grad-CAM) (SELVARAJU et al., 2017). Essa técnica gera mapas de calor que destacam as regiões da imagem consideradas mais relevantes pelo modelo durante o processo de tomada de decisão. Os mapas permitem identificar visualmente as áreas de maior influência na classificação, proporcionando uma compreensão detalhada do desempenho e das limitações de cada arquitetura analisada.

Os modelos selecionados para esta avaliação foram a DenseNet-121 e a Inception V3, escolhidos com base nos melhores valores de F1 Score ponderado no conjunto de teste. A seleção dos dentes do estudo de caso foi realizada com base na DenseNet-121, considerando as imagens com maior probabilidade para a classe analisada.

A Figura 31 apresenta um exemplo de Lesão Periapical no dente 44, corretamente classificada pela DenseNet121 e incorretamente classificada pela InceptionV3. A DenseNet121 atribuiu uma probabilidade de 0,94 para a classe correta, com seu Grad-CAM (Figura 31 (A)) evidenciando a área patológica, indicando que o modelo concentrou sua atenção na região de interesse ao tomar a decisão. Em contraste, a InceptionV3 classificou erroneamente a imagem como Cárie, com probabilidade de 0,69. Seu Grad-CAM (Figura 31 (B)) apresentou a maior ativação em uma região fora do dente, com menor intensidade na área da lesão, o que sugere que o modelo não focou adequadamente a região patológica para a tomada de decisão.

Figura 31 – Exemplo de classificação correta e incorreta de Lesão Periapical. (A) Mapa de calor gerado pela DenseNet121, destacando a área patológica; (B) Mapa de calor gerado pela InceptionV3, com ativação em região fora do dente.

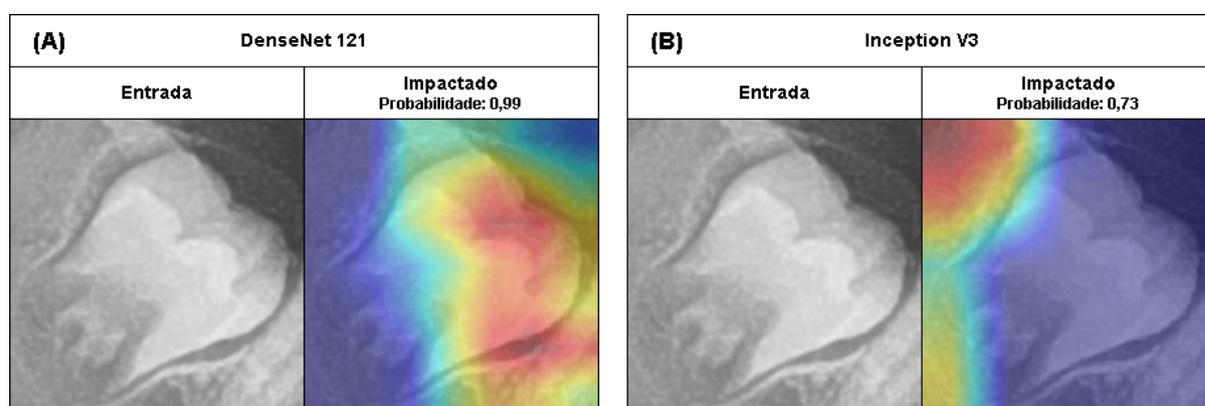


Fonte: Acervo do autor.

Na Figura 32, apresenta a comparação dos mapas de calor para a classificação de um dente Impactado. O Grad-CAM gerado pela DenseNet121 (Figura 32 (A)) destaca tanto a região do dente quanto áreas adjacentes do osso com alta ativação, indicando

um foco preciso na área de interesse, o que resultou em uma probabilidade de 0,99 para a classe correta. Por outro lado, a InceptionV3 (Figura 32 (B)) também classificou a imagem corretamente com probabilidade de 0,73. No entanto, suas regiões de ativação estão concentradas fora do dente, indicando um foco menos preciso na região de interesse. Essas diferenças sugerem que, embora ambas as arquiteturas tenham conseguido classificar corretamente, a DenseNet121 demonstrou maior confiança e capacidade de identificar e priorizar as regiões relevantes.

Figura 32 – Exemplo de classificação correta de dente Impactado. (A) Mapa de calor gerado pela DenseNet121, destacando a região do dente e áreas do osso ao redor com alta ativação; (B) Mapa de calor gerado pela InceptionV3, com regiões de ativação concentradas fora do dente.



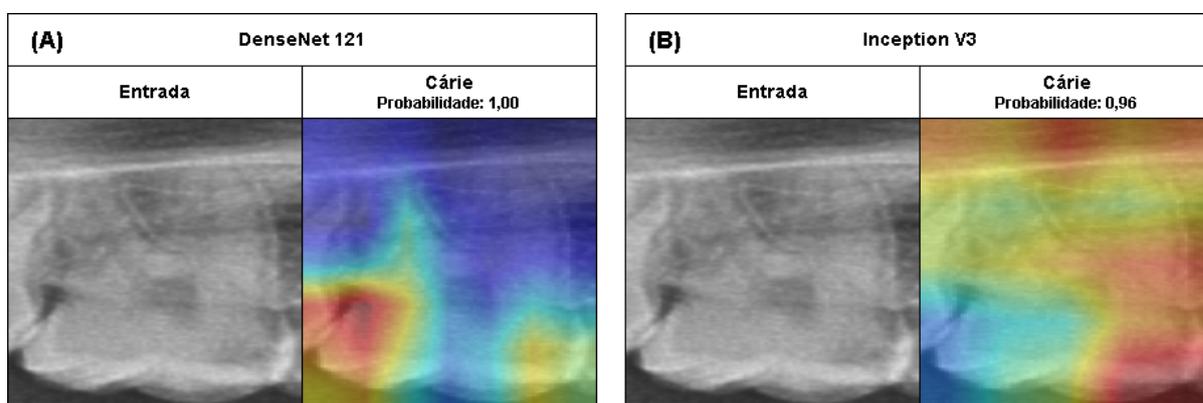
Fonte: Acervo do autor.

A Figura 33 apresenta os mapas de calor gerados para a classificação de um dente com cárie. A DenseNet121 (Figura 33 (A)) evidencia majoritariamente regiões no dente, com ativação mais concentrada em dois pontos específicos, resultando em uma classificação correta com probabilidade de 1,00. Já a InceptionV3 (Figura 33 (B)) também classificou corretamente a imagem, com probabilidade de 0,96. Contudo, seu Grad-CAM apresenta ativação mais difusa, abrangendo quase toda a imagem sem focar em regiões específicas, indicando menor precisão no foco da lesão.

A fim de avaliar qualitativamente os erros de classificação realizados pelos modelos treinados, foram analisados exemplos de falso positivo com maior probabilidade para cada uma das três classes principais: Lesão Periapical, Cárie e Impactado. Utilizou-se novamente a técnica Grad-CAM, que permite identificar visualmente as regiões mais relevantes para a tomada de decisão do modelo, destacando potenciais limitações e falhas no processo de aprendizado.

Na Figura 34, é observado um exemplo de falso positivo para a classe Lesão

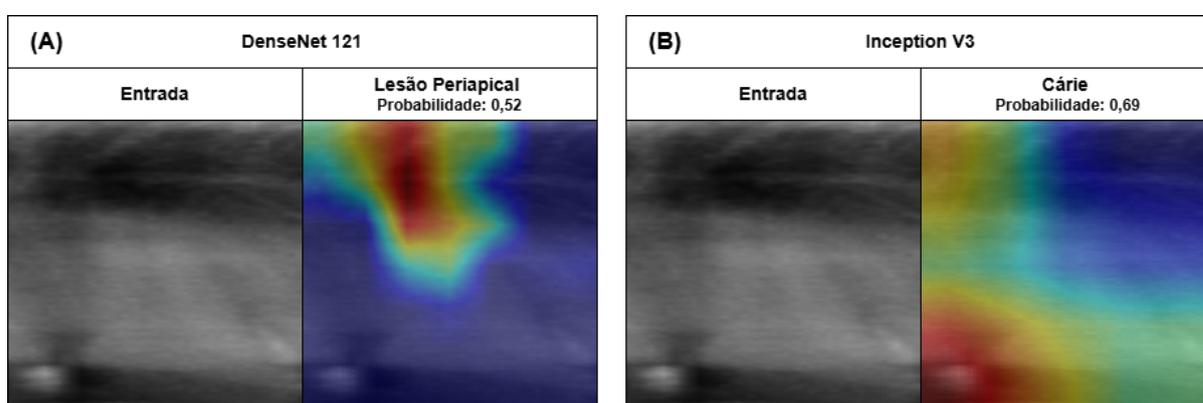
Figura 33 – Exemplo de classificação correta de dente com cárie. (A) Mapa de calor gerado pela DenseNet121, destacando regiões no dente com alta ativação; (B) Mapa de calor gerado pela Inception V3, com regiões de ativação mais difusas e menos concentradas em pontos específicos.



Fonte: Acervo do autor.

Periapical. O Grad-CAM gerado pela DenseNet121 apresentou alta ativação concentrada na região da raiz do dente, resultando em uma classificação incorreta como Lesão Periapical, com probabilidade de 0,52. No entanto, o *ground truth* indica que a condição correta é Cárie. Esse resultado sugere que o modelo interpretou erroneamente características da raiz como indicativas de lesão periapical. Por outro lado, a InceptionV3 classificou corretamente o dente como Cárie, com probabilidade de 0,69, embora sua ativação estivesse distribuída tanto na região do dente alvo quanto em áreas externas, incluindo um dente adjacente. Isso indica que, apesar da classificação correta, o modelo não focou exclusivamente na região da lesão.

Figura 34 – Exemplo de falso positivo para dente com Lesão Periapical.

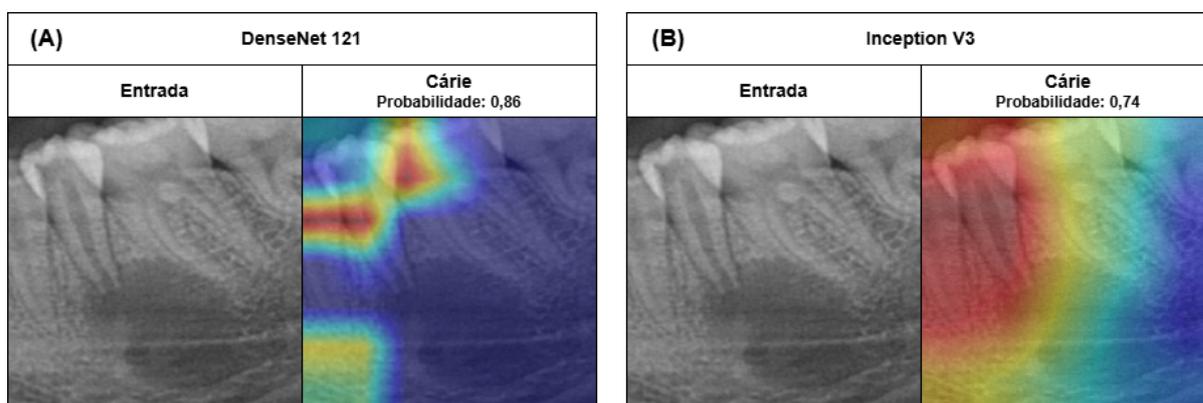


Fonte: Acervo do autor.

A Figura 35 apresenta os mapas de calor para exemplos de falso positivo na classe Cárie. A DenseNet121 (Figura 35 (A)) destacou regiões específicas de alta ativação no dente

analisado e também nos dentes adjacentes, atribuindo incorretamente uma probabilidade de 0,86 para a classe Cárie, enquanto a classe verdadeira era Lesão Periapical. Já a InceptionV3 (Figura 35 (B)) também cometeu um erro de classificação, atribuindo uma probabilidade de 0,74 para a classe Cárie. No entanto, sua ativação esteve menos concentrada em regiões específicas e se distribuiu principalmente fora da área da lesão, abrangendo tanto o dente alvo quanto dentes adjacentes. Esses resultados refletem limitações em ambas as arquiteturas no reconhecimento de características específicas da classe.

Figura 35 – Exemplo de falso positivo para dente com Cárie.

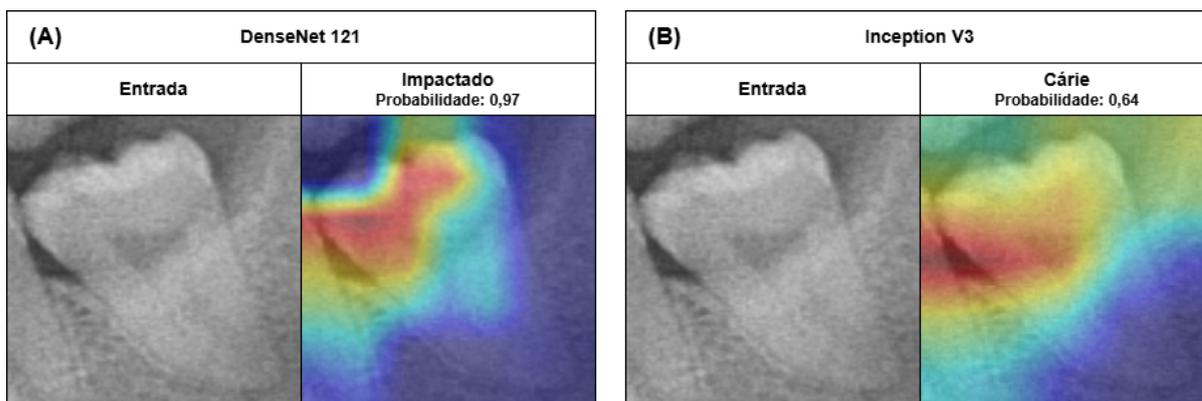


Fonte: Acervo do autor.

Por fim, na Figura 36, são analisados falsos positivos para a classe Impactado. A DenseNet121 (Figura 36 (A)) apresentou regiões de alta ativação no dente e áreas adjacentes do osso, atribuindo incorretamente uma probabilidade de 0,97 para a classe Impactado, enquanto a classe correta era Cárie. Já a InceptionV3 (Figura 36 (B)) classificou corretamente o dente como Cárie, com uma probabilidade de 0,64. No entanto, sua ativação foi mais difusa, cobrindo boa parte do dente alvo, um dente adjacente e outras áreas fora do dente, indicando uma menor especificidade no foco da lesão.

Essas análises qualitativas sugerem que, embora os modelos alcancem probabilidades relativamente altas de falsos positivos em algumas imagens de dentes, suas estratégias de atenção diferem substancialmente. A DenseNet121 tende a concentrar-se em regiões mais específicas, ainda que com interpretações equivocadas, enquanto a InceptionV3 frequentemente apresenta ativações mais dispersas, o que, em alguns casos, contribuiu para uma menor precisão na identificação das características das classes verdadeiras.

Figura 36 – Exemplo de falso positivo para dente Impactado.



Fonte: Acervo do autor.

5.4 Considerações Finais

Clinicamente, a detecção automatizada e precisa de dentes, aliada à classificação de anomalias dentárias, pode acelerar processos manuais, reduzir erros humanos e padronizar diagnósticos na prática odontológica. Essa abordagem não apenas otimiza a identificação e numeração dentária, mas também aprimora a eficiência clínica ao permitir relatórios em larga escala com maior rapidez e consistência. Dessa forma, contribui para um fluxo de trabalho mais preciso e confiável, beneficiando tanto profissionais quanto pacientes.

6 CONCLUSÃO

Este trabalho apresentou um método baseado em aprendizado profundo para a detecção automática de dentes e classificação de anomalias dentárias em imagens de radiografias panorâmicas. A partir do desenvolvimento e avaliação de diferentes modelos CNN, foi possível demonstrar a eficácia da abordagem proposta na identificação de estruturas dentárias e na categorização de condições patológicas relevantes para a prática odontológica.

Os resultados obtidos evidenciaram que modelos da família YOLOv8, especialmente a versão YOLOv8x, apresentaram um alto desempenho na detecção de dentes, com métricas de precisão e *recall* superiores a 0,92. Além disso, a análise da classificação de patologias revelou que a arquitetura DenseNet121 obteve os melhores índices de equilíbrio entre precisão e *recall*, com F1 Score de 0,8546, indicando uma forte capacidade do modelo em identificar corretamente patologias em radiografias panorâmicas. Essas análises reforçam a importância da escolha adequada de arquiteturas e da otimização de hiperparâmetros para melhorar a eficácia dos modelos de aprendizado profundo aplicados à odontologia.

A automação do processo de detecção e classificação de patologias dentárias pode trazer impactos significativos para o diagnóstico odontológico, permitindo maior agilidade e padronização na análise de exames radiográficos. A utilização de métodos de IA reduz a dependência da interpretação manual, minimizando erros e proporcionando um suporte adicional aos profissionais da área.

Com os resultados obtidos, nota-se que o método proposto consegue resultados promissores com potencial para auxílio aos profissionais da odontologia. Todavia, o método ainda pode ser aprimorado para aumentar sua robustez e aplicabilidade em cenários clínicos reais.

6.1 Trabalhos Futuros

Com o objetivo de superar as limitações identificadas, são listadas a seguir algumas sugestões para trabalhos futuros:

- Ampliar a base de dados utilizada para treinar os modelos, incluindo um maior número de amostras e patologias;

- Investigar o uso de arquiteturas mais recentes baseadas em *Transformers*, como Vision Transformer (ViT) (DOSOVITSKIY et al., 2020), Swin Transformer (LIU et al., 2021) e ConvNeXt (LIU et al., 2022), que têm demonstrado resultados promissores em tarefas de visão computacional;
- Implementar uma etapa de segmentação dos dentes para reduzir erros de classificação decorrentes da influência de regiões de dentes adjacentes;
- Validar a abordagem proposta em ambientes clínicos para avaliar sua eficácia na prática odontológica.

6.2 Produções científicas

A Tabela 10 apresenta os artigos científicos publicados que possuem relação com o método proposto neste trabalho.

Tabela 10 – Artigos que possuem relação com o método proposto.

Título	Conferência/Periódico	Qualis	Status
Automated Tooth Detection and Numbering in Panoramic Radiographs Using YOLO	Procedia Computer Science	A3	Publicado

A Tabela 11 elenca os artigos científicos que não possuem relação com o método proposto pelo presente trabalho, entretanto foram publicados durante o mestrado.

Tabela 11 – Artigos que não possuem relação com o método proposto.

Título	Conferência/Periódico	Qualis	Status
Multi-label Classification of Pathologies in Chest Radiograph Images Using DenseNet.	Brazilian Conference on Intelligent Systems	A4	Publicado

REFERÊNCIAS

- ABDULKREEM, A.; BHATTACHARJEE, T.; ALZAABI, H.; ALALI, K.; GONZALEZ, A.; CHAUDHRY, J.; PRASAD, S. Artificial intelligence-based automated preprocessing and classification of impacted maxillary canines in panoramic radiographs. **Dentomaxillofacial Radiology**, Oxford University Press, v. 53, n. 3, p. 173–177, 2024.
- ADEYEMO, W. L.; ODERINU, O. H.; OLOJEDE, A. C.; FASHINA, A. A.; AYODELE, A. O. Experience of wrong-site tooth extraction among nigerian dentists. **The Saudi Dental Journal**, Elsevier, v. 23, n. 3, p. 153–156, 2011.
- AFFAIRS, A. D. A. C. on S. et al. The use of dental radiographs: update and recommendations. **The Journal of the American Dental Association**, Elsevier, v. 137, n. 9, p. 1304–1312, 2006.
- AKIBA, T.; SANO, S.; YANASE, T.; OHTA, T.; KOYAMA, M. Optuna: A next-generation hyperparameter optimization framework. In: **Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining**. New York, NY, USA: Association for Computing Machinery, 2019. p. 2623–2631.
- AL-ZOUBI, H.; ALHARBI, A. A.; FERGUSON, D. J.; ZAFAR, M. S. Frequency of impacted teeth and categorization of impacted canines: A retrospective radiographic study using orthopantomograms. **European journal of dentistry**, Thieme Medical and Scientific Publishers Private Ltd., v. 11, n. 01, p. 117–121, 2017.
- ALBAWI, S.; MOHAMMED, T. A.; AL-ZAWI, S. Understanding of a convolutional neural network. In: IEEE. **2017 International Conference on Engineering and Technology (ICET)**. Antalya, Turkey, 2017. p. 1–6.
- ALBUQUERQUE, M. P. de; ALBUQUERQUE, M. P. de. Processamento de imagens: métodos e análises. **Centro Brasileiro de Pesquisas Físicas MCT**, 2000.
- ANDONIE, R. Hyperparameter optimization in learning systems. **Journal of Membrane Computing**, Springer, v. 1, n. 4, p. 279–291, 2019.
- BERKHOUT, E.; SANDERINK, G.; STELT, P. van der. Digital intra-oral radiography in dentistry. diagnostic efficacy and dose considerations. **Oral radiology**, Springer, v. 19, p. 1–13, 2003.
- BERRAR, D. et al. **Cross-validation**. 2019.
- BIEWALD, L. **Experiment Tracking with Weights and Biases**. 2020. Software available from wandb.com. Disponível em: <<https://www.wandb.com/>>.
- BILGIR, E.; BAYRAKDAR, İ. Ş.; ÇELİK, Ö.; ORHAN, K.; AKKOCA, F.; SAĞLAM, H.; ODABAŞ, A.; ASLAN, A. F.; OZCETIN, C.; KILLI, M. et al. An artificial intelligence approach to automatic tooth detection and numbering in panoramic radiographs. **BMC medical imaging**, Springer, v. 21, p. 1–9, 2021.
- BISHOP, C. M.; NASRABADI, N. M. **Pattern recognition and machine learning**. New York: Springer, 2006. v. 4.
- BJØRNDAL, L. The caries process and its effect on the pulp: the science is changing and so is our understanding. **Journal of endodontics**, Elsevier, v. 34, n. 7, p. S2–S5, 2008.

- BJØRNDAL, L.; SIMON, S.; TOMSON, P.; DUNCAN, H. Management of deep caries and the exposed pulp. **International endodontic journal**, Wiley Online Library, v. 52, n. 7, p. 949–973, 2019.
- BLINKHORN, A. S.; CHOI, C. L.; PAGET, H. E. An investigation into the use of the fdi tooth notation system by dental schools in the uk. **European Journal of Dental Education**, Wiley Online Library, v. 2, n. 1, p. 39–41, 1998.
- BOCHKOVSKIY, A.; WANG, C.-Y.; LIAO, H.-Y. M. Yolov4: Optimal speed and accuracy of object detection. **arXiv preprint arXiv:2004.10934**, 2020.
- BROWNLEE, J. A gentle introduction to the rectified linear unit (relu). **Machine learning mastery**, v. 6, 2019.
- CHITRADEVI, B.; SRIMATHI, P. An overview on image processing techniques. **International Journal of Innovative Research in Computer and Communication Engineering**, Citeseer, v. 2, n. 11, p. 6466–6472, 2014.
- CHOI, J.-W. Assessment of panoramic radiography as a national oral examination tool: review of the literature. **Imaging science in dentistry**, Korean Academy of Oral and Maxillofacial Radiology, v. 41, n. 1, p. 1–6, 2011.
- CHOLLET, F. Xception: Deep learning with depthwise separable convolutions. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. Honolulu, HI, USA: IEEE, 2017. p. 1251–1258.
- DOSOVITSKIY, A.; BEYER, L.; KOLESNIKOV, A.; WEISSENBORN, D.; ZHAI, X.; UNTERTHINER, T.; DEGHANI, M.; MINDERER, M.; HEIGOLD, G.; GELLY, S. et al. An image is worth 16x16 words: Transformers for image recognition at scale. **arXiv preprint arXiv:2010.11929**, 2020.
- EVERINGHAM, M.; GOOL, L. V.; WILLIAMS, C. K.; WINN, J.; ZISSERMAN, A. The pascal visual object classes (voc) challenge. **International journal of computer vision**, Springer, v. 88, p. 303–338, 2010.
- FANG, W.; LOVE, P. E.; LUO, H.; DING, L. Computer vision for behaviour-based safety in construction: A review and future directions. **Advanced Engineering Informatics**, Elsevier, v. 43, p. 100980, 2020.
- FILHO, O. M.; NETO, H. V. **Processamento digital de imagens**. Rio de Janeiro: Brasport, 1999.
- FRANK, Z. A. **Panoramic x-ray apparatus**. [S.l.]: Google Patents, 1922. US Patent 1,408,559.
- FRAZIER, P. I. Bayesian optimization. In: **Recent advances in optimization and modeling of contemporary problems**. Catonsville, MD, USA: Informs, 2018. p. 255–278.
- GIRSHICK, R. Fast r-cnn in proceedings of the iee international conference on computer vision (pp. 1440–1448). **Piscataway, NJ: IEEE.[Google Scholar]**, v. 2, 2015.
- GONZALEZ, R. C.; WOODS, R. E. **Processamento de imagens digitais**. São Paulo, Brasil: Editora Blucher, 2000.

GOOD, I. J. Rational decisions. In: **Breakthroughs in statistics**. New York, NY, USA: Springer, 1992. p. 365–377.

GOODFELLOW, I. **Deep learning**. Cambridge, MA, USA: MIT press, 2016.

HAMAMCI, I. E.; ER, S.; SIMSAR, E.; YUKSEL, A. E.; GULTEKIN, S.; OZDEMIR, S. D.; YANG, K.; LI, H. B.; PATI, S.; STADLINGER, B. et al. Dentex: an abnormal tooth detection with dental enumeration and diagnosis benchmark for panoramic x-rays. **arXiv preprint arXiv:2305.19112**, 2023.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J.; FRANKLIN, J. The elements of statistical learning: data mining, inference and prediction. **The Mathematical Intelligencer**, Springer, v. 27, n. 2, p. 83–85, 2005.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 37, n. 9, p. 1904–1916, 2015.

_____. Deep residual learning for image recognition. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. Las Vegas, NV, USA: IEEE, 2016. p. 770–778.

HECKMANN, K. Die roentgenperspektive undiere umwand-lung durch eine neue aufnahmetechnik. **Roentgenstrahlen**, v. 60, p. 144–157, 1939.

HEO, M.-S.; KIM, J.-E.; HWANG, J.-J.; HAN, S.-S.; KIM, J.-S.; YI, W.-J.; PARK, I.-W. Artificial intelligence in oral and maxillofacial radiology: what is currently possible? **Dentomaxillofacial Radiology**, Oxford University Press, v. 50, n. 3, p. 20200375, 2021.

HINTON, G.; SRIVASTAVA, N.; SWERSKY, K. Neural networks for machine learning lecture 6a overview of mini-batch gradient descent. **Cited on**, v. 14, n. 8, p. 2, 2012.

HUANG, G.; LIU, Z.; MAATEN, L. V. D.; WEINBERGER, K. Q. Densely connected convolutional networks. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. Honolulu, HI, USA: IEEE, 2017. p. 4700–4708.

HWANG, J.-J.; JUNG, Y.-H.; CHO, B.-H.; HEO, M.-S. An overview of deep learning in the field of dentistry. **Imaging science in dentistry**, Korean Academy of Oral and Maxillofacial Radiology, v. 49, n. 1, p. 1–7, 2019.

INNES, N.; FRENCKEN, J. E.; BJØRNDAL, L.; MALTZ, M.; MANTON, D. J.; RICKETTS, D.; LANDUYT, K. V.; BANERJEE, A.; CAMPUS, G.; DOMÉJEAN, S. et al. Managing carious lesions: consensus recommendations on terminology. **Advances in dental research**, SAGE Publications Sage CA: Los Angeles, CA, v. 28, n. 2, p. 49–57, 2016.

IOFFE, S.; SZEGEDY, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: **International conference on machine learning**. Lille, France: PMLR, 2015. p. 448–456.

JAN, A.; ALBENAYAN, R.; ALSHARKAWI, D.; JADU, F. The prevalence and causes of wrong tooth extraction. **Nigerian Journal of Clinical Practice**, Medknow, v. 22, n. 12, p. 1706–1714, 2019.

- JANIESCH, C.; ZSCHECH, P.; HEINRICH, K. Machine learning and deep learning. **Electronic Markets**, Springer, v. 31, n. 3, p. 685–695, 2021.
- JOCHER, G.; CHAURASIA, A.; QIU, J. Yolo by ultralytics, 8.0. 0. **Publish at <https://github.com/ultralytics/ultralytics>**, 2023.
- JOCHER, G.; QIU, J.; CHAURASIA, A. **Ultralytics YOLO**. 2023. Disponível em: <<https://github.com/ultralytics/ultralytics>>.
- JONES, D. R. A taxonomy of global optimization methods based on response surfaces. **Journal of global optimization**, Springer, v. 21, n. 4, p. 345–383, 2001.
- KARAOGLU, A.; OZCAN, C.; PEKINCE, A.; YASA, Y. Numbering teeth in panoramic images: A novel method based on deep learning and heuristic algorithm. **Engineering Science and Technology, an International Journal**, Elsevier, v. 37, p. 101316, 2023.
- KARUNAKARAN, J.; ABRAHAM, C. S.; KARTHIK, A. K.; JAYAPRAKASH, N. Successful nonsurgical management of periapical lesions of endodontic origin: A conservative orthograde approach. **Journal of pharmacy & bioallied sciences**, Wolters Kluwer–Medknow Publications, v. 9, n. Suppl 1, p. S246, 2017.
- KAZIMIERCZAK, W.; WAJER, R.; WAJER, A.; KIIAN, V.; KLOSKA, A.; KAZIMIERCZAK, N.; JANISZEWSKA-OLSZOWSKA, J.; SERAFIN, Z. Periapical lesions in panoramic radiography and cbct imaging—assessment of ai’s diagnostic accuracy. **Journal of Clinical Medicine**, MDPI, v. 13, n. 9, p. 2709, 2024.
- KIM, S.; CASPER, R. Applications of convolution in image processing with matlab. **University of Washington**, p. 1–20, 2013.
- KINGMA, D. P. Adam: A method for stochastic optimization. **arXiv preprint arXiv:1412.6980**, 2014.
- KOEHRSEN, W. A conceptual explanation of bayesian hyperparameter optimization for machine learning. **Towards Data Science**, 2018.
- KROGH, A. What are artificial neural networks? **Nature biotechnology**, Nature Publishing Group, v. 26, n. 2, p. 195–197, 2008.
- KRUTSCH, R.; TENORIO, D. Histogram equalization. **Freescale Semiconductor, Document Number AN4318, Application Note**, Freescale Semiconductor Austin, TX, USA, v. 30, 2011.
- LALONDE, E. R.; LUEBKE, R. G. The frequency and distribution of periapical cysts and granulomas: an evaluation of 800 specimens. **Oral Surgery, Oral Medicine, Oral Pathology**, Elsevier, v. 25, n. 6, p. 861–868, 1968.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **nature**, Nature Publishing Group UK London, v. 521, n. 7553, p. 436–444, 2015.
- LIN, T.-Y.; GOYAL, P.; GIRSHICK, R.; HE, K.; DOLLÁR, P. Focal loss for dense object detection. In: **Proceedings of the IEEE international conference on computer vision**. Venice, Italy: IEEE, 2017. p. 2980–2988.

- LIU, L.; OUYANG, W.; WANG, X.; FIEGUTH, P.; CHEN, J.; LIU, X.; PIETIKÄINEN, M. Deep learning for generic object detection: A survey. **International journal of computer vision**, Springer, v. 128, p. 261–318, 2020.
- LIU, Z.; LIN, Y.; CAO, Y.; HU, H.; WEI, Y.; ZHANG, Z.; LIN, S.; GUO, B. Swin transformer: Hierarchical vision transformer using shifted windows. In: **Proceedings of the IEEE/CVF international conference on computer vision**. Montreal, QC, Canada: IEEE, 2021. p. 10012–10022.
- LIU, Z.; MAO, H.; WU, C.-Y.; FEICHTENHOFER, C.; DARRELL, T.; XIE, S. A convnet for the 2020s. In: **Proceedings of the IEEE/CVF conference on computer vision and pattern recognition**. New Orleans, LA, USA: IEEE, 2022. p. 11976–11986.
- LUDLOW, J. B.; DAVIES-LUDLOW, L. E.; WHITE, S. C. Patient risk related to common dental radiographic examinations: the impact of 2007 international commission on radiological protection recommendations regarding dose calculation. **The journal of the American Dental association**, Elsevier, v. 139, n. 9, p. 1237–1243, 2008.
- LYUBCHENKO, V.; MATARNEH, R.; KOBYLIN, O.; LYASHENKO, V. Digital image processing techniques for detection and diagnosis of fish diseases. **IJARCSSE**, 2016.
- MAAS, A. L.; HANNUN, A. Y.; NG, A. Y. et al. Rectifier nonlinearities improve neural network acoustic models. In: ATLANTA, GA. **Proc. icml**. Atlanta, GA, USA: JMLR, 2013. v. 30, n. 1, p. 3.
- MANJUNATHA, B. S.; CHIKKARAMAIAH, S.; PANJA, P.; KORATAGERE, N. Impacted maxillary second premolars: a report of four cases. **Case Reports**, BMJ Publishing Group, v. 2014, p. bcr2014205206, 2014.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, Springer, v. 5, n. 4, p. 115–133, 1943.
- MOLANDER, B.; GRONDAHL, H.; EKESTUBBE, A. Quality of film-based and digital panoramic radiography. **Dentomaxillofacial Radiology**, Oxford University Press, v. 33, n. 1, p. 32–36, 2004.
- NADER, R.; SMORODIN, A.; FOURNIERE, N. D. L.; AMOURIQ, Y.; AUTRUSSEAU, F. Automatic teeth segmentation on panoramic x-rays using deep neural networks. In: **2022 26th International Conference on Pattern Recognition (ICPR)**. Montreal, QC, Canada: IEEE, 2022. p. 4299–4305.
- NUMATA, H. Consideration of the parabolic radiography of the dental arch. **J Shimazu Stud**, v. 10, p. 13, 1933.
- NVIDIA; VINGELMANN, P.; FITZEK, F. H. **CUDA, release: 10.2.89**. 2020. Disponível em: <<https://developer.nvidia.com/cuda-toolkit>>.
- O'SHEA, K.; NASH, R. An introduction to convolutional neural networks. **arXiv preprint arXiv:1511.08458**, 2015.
- OU, X.; CHEN, X.; XU, X.; XIE, L.; CHEN, X.; HONG, Z.; BAI, H.; LIU, X.; CHEN, Q.; LI, L. et al. Recent development in x-ray imaging technology: Future and challenges. **Research**, AAAS, 2021.

- PADILLA, R.; NETTO, S. L.; SILVA, E. A. D. A survey on performance metrics for object-detection algorithms. In: **2020 international conference on systems, signals and image processing (IWSSIP)**. Niteroi, Brazil: IEEE, 2020. p. 237–242.
- PADILLA, R.; PASSOS, W. L.; DIAS, T. L.; NETTO, S. L.; SILVA, E. A. D. A comparative analysis of object detection metrics with a companion open-source toolkit. **Electronics**, MDPI, v. 10, n. 3, p. 279, 2021.
- PANETTA, K.; RAJENDRAN, R.; RAMESH, A.; RAO, S. P.; AGAIAN, S. Tufts dental database: a multimodal panoramic x-ray dataset for benchmarking diagnostic systems. **IEEE journal of biomedical and health informatics**, IEEE, v. 26, n. 4, p. 1650–1659, 2021.
- PARK, J.; LEE, J.; MOON, S.; LEE, K. Deep learning based detection of missing tooth regions for dental implant planning in panoramic radiographic images. **Applied Sciences**, MDPI, v. 12, n. 3, p. 1595, 2022.
- PAUWELS, R. History of dental radiography: Evolution of 2d and 3d imaging modalities. **Med Phys Int**, v. 8, n. 1, p. 235–77, 2020.
- PERSCHBACHER, S. Interpretation of panoramic radiographs. **Australian dental journal**, Wiley Online Library, v. 57, p. 40–45, 2012.
- PODLOZHNYUK, V. Image convolution with cuda. **NVIDIA Corporation white paper, June**, v. 2097, n. 3, 2007.
- RAMEZAN, C. A.; WARNER, T. A.; MAXWELL, A. E. Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification. **Remote Sensing**, MDPI, v. 11, n. 2, p. 185, 2019.
- REDMON, J. You only look once: Unified, real-time object detection. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. Las Vegas, NV, USA: IEEE, 2016.
- RONDON, R. H. N.; PEREIRA, Y. C. L.; NASCIMENTO, G. C. do. Common positioning errors in panoramic radiography: A review. **Imaging science in dentistry**, Korean Academy of Oral and Maxillofacial Radiology, v. 44, n. 1, p. 1–6, 2014.
- ROSSUM, G. V.; JR, F. L. D. **Python reference manual**. Amsterdam, NLD: Centrum voor Wiskunde en Informatica Amsterdam, 1995.
- RUDER, S. An overview of gradient descent optimization algorithms. **arXiv preprint arXiv:1609.04747**, 2016.
- SANDLER, M.; HOWARD, A.; ZHU, M.; ZHMOGINOV, A.; CHEN, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. Salt Lake City, UT, USA: IEEE, 2018. p. 4510–4520.
- SARICA, I.; DERINDAĞ, G.; KURTULDU, E.; NARALAN, M.; ÇAĞLAYAN, F. A retrospective study: do all impacted teeth cause pathology? **Nigerian journal of clinical practice**, Medknow, v. 22, n. 4, p. 527–533, 2019.

- SELVARAJU, R. R.; COGSWELL, M.; DAS, A.; VEDANTAM, R.; PARIKH, D.; BATRA, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: **Proceedings of the IEEE international conference on computer vision**. Venice, Italy: IEEE, 2017. p. 618–626.
- SELWITZ, R. H.; ISMAIL, A. I.; PITTS, N. B. Dental caries. **The Lancet**, Elsevier, v. 369, n. 9555, p. 51–59, 2007.
- SHAFI, I.; FATIMA, A.; AFZAL, H.; DÍEZ, I. d. l. T.; LIPARI, V.; BREÑOSA, J.; ASHRAF, I. A comprehensive review of recent advances in artificial intelligence for dentistry e-health. **Diagnostics**, MDPI, v. 13, n. 13, p. 2196, 2023.
- SHTEYER, A.; ROZOVSKY, E. Periapical lesions–types, incidence and clinical features. **Refu'at ha-peh vеха-shinayim (Tel Aviv, Israel: 1969)**, v. 21, p. 100–103, 1972.
- SILVA, B. P. M.; PINHEIRO, L. B.; SOBRINHO, B. P. P.; LIMA, F. P.; SOBRINHO, B. P. P.; LIMA, K. A. B.; PITHON, M. M.; CURY, P. R.; OLIVEIRA, L. R. d. Boosting research on dental panoramic radiographs: a challenging data set, baselines, and a task central online platform for benchmark. **Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization**, Taylor & Francis, v. 11, n. 4, p. 1327–1347, 2023.
- SILVA, G.; OLIVEIRA, L.; PITHON, M. Automatic segmenting teeth in x-ray images: Trends, a novel data set, benchmarking and future perspectives. **Expert Systems with Applications**, Elsevier, v. 107, p. 15–31, 2018.
- SIMONYAN, K. Very deep convolutional networks for large-scale image recognition. **arXiv preprint arXiv:1409.1556**, 2014.
- SOHAN, M.; RAM, T. S.; REDDY, R.; VENKATA, C. A review on yolov8 and its advancements. In: **International Conference on Data Intelligence and Cognitive Informatics**. Singapore: Springer, 2024. p. 529–545.
- SONG, S. The use of color elements in graphic design based on convolutional neural network model. **Applied Mathematics and Nonlinear Sciences**, 2023.
- SRIVASTAVA, N.; HINTON, G.; KRIZHEVSKY, A.; SUTSKEVER, I.; SALAKHUT-DINOV, R. Dropout: a simple way to prevent neural networks from overfitting. **The journal of machine learning research**, JMLR. org, v. 15, n. 1, p. 1929–1958, 2014.
- STOJMENOVA, V. T.; GJUROVSKI, S. R. Periapical lesions: current modalities. **MEDIS-International Journal of Medical Sciences and Research**, Institute of Management and Knowledge, v. 1, n. 1, p. 25–28, 2022.
- STONE, M. Cross-validatory choice and assessment of statistical predictions. **Journal of the royal statistical society: Series B (Methodological)**, Wiley Online Library, v. 36, n. 2, p. 111–133, 1974.
- SZEGEDY, C.; LIU, W.; JIA, Y.; SERMANET, P.; REED, S.; ANGUELOV, D.; ERHAN, D.; VANHOUCHE, V.; RABINOVICH, A. Going deeper with convolutions. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. Boston, MA, USA: IEEE, 2015. p. 1–9.

SZEGEDY, C.; VANHOUCKE, V.; IOFFE, S.; SHLENS, J.; WOJNA, Z. Rethinking the inception architecture for computer vision. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. Las Vegas, NV, USA: IEEE, 2016. p. 2818–2826.

TAICHER, S. Wrong tooth extraction: root cause analysis. **Quintessence Int**, v. 41, p. 869–72, 2010.

TAMMISALO, E. H. Professor yrjö v. paatero-the pioneer of panoramic oral tomography. **Dentomaxillofacial Radiology**, Oxford University Press, v. 4, n. 1, p. 53–56, 1975.

TAN, M.; LE, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In: **Proceedings of the International Conference on Machine Learning**. Long Beach, CA, USA: PMLR, 2019. p. 6105–6114.

TERVEN, J.; CÓRDOVA-ESPARZA, D.-M.; ROMERO-GONZÁLEZ, J.-A. A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. **Machine Learning and Knowledge Extraction**, MDPI, v. 5, n. 4, p. 1680–1716, 2023.

WAFI'A, A. F. Common positioning errors in panoramic radiography. 2005.

WALD, A. Statistical decision functions. **The Annals of Mathematical Statistics**, JSTOR, p. 165–205, 1949.

WANG, Q.; MA, Y.; ZHAO, K.; TIAN, Y. A comprehensive survey of loss functions in machine learning. **Annals of Data Science**, Springer, p. 1–26, 2020.

WHITE, S. C.; YOON, D. C.; TETRADIS, S. Digital radiography in dentistry: what it should do for you. **Journal of the California Dental Association**, Taylor & Francis, v. 27, n. 12, p. 942–952, 1999.

YILMAZ, S.; TASYUREK, M.; AMUK, M.; CELIK, M.; CANGER, E. M. Developing deep learning methods for classification of teeth in dental panoramic radiography. **Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology**, Elsevier, v. 138, n. 1, p. 118–127, 2024.

YURDUKORU, B. [standardization of the tooth numbering systems]. **Ankara Universitesi Dis Hekimligi Fakultesi dergisi = The Journal of the Dental Faculty of Ankara University**, v. 16, n. 3, p. 527–531, September 1989. ISSN 1016-8710. Disponível em: <<http://europepmc.org/abstract/MED/2489508>>.

ZAIDI, S. S. A.; ANSARI, M. S.; ASLAM, A.; KANWAL, N.; ASGHAR, M.; LEE, B. A survey of modern deep learning based object detection models. **Digital Signal Processing**, Elsevier, v. 126, p. 103514, 2022.

ZELL, A. **Simulation neuronaler netze**. Bonn, Germany: Addison-Wesley Bonn, 1994. v. 1.