



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CAMPUS DE CRATEÚS**  
**CURSO DE GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO**

**FRANCISCO RUBENS DUTRA LIMA**

**ODONTO VISION: SISTEMA CAD PARA SEGMENTAÇÃO AUTOMÁTICA DE  
DENTES EM RADIOGRAFIAS PANORÂMICAS**

**CRATEÚS**

**2026**

FRANCISCO RUBENS DUTRA LIMA

ODONTO VISION: SISTEMA CAD PARA SEGMENTAÇÃO AUTOMÁTICA DE DENTES  
EM RADIOGRAFIAS PANORÂMICAS

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em SISTEMAS DE INFORMAÇÃO do CAMPUS DE CRATEÚS da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em SISTEMAS DE INFORMAÇÃO.

Orientador: Prof. Dr. Bruno Riccelli  
Dos Santos Silva

CRATEÚS

2026

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Sistema de Bibliotecas  
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

- L698o Lima, Francisco Rubens Dutra.  
Odonto Vision: Sistema cad para segmentação automática de dentes em radiografias panorâmicas. /  
Francisco Rubens Dutra Lima. – 2026.  
81 f. : il. color.
- Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Campus de Crateús,  
Curso de Sistemas de Informação, Crateús, 2026.  
Orientação: Prof. Dr. Bruno Riccelli Dos Santos Silva.
1. Segmentação de imagens médicas. 2. Sistema CAD. 3. Modelos de segmentação. 4. Redes Neurais  
Convolucionais. 5. Aprendizado profundo. I. Título.

CDD 005

---

FRANCISCO RUBENS DUTRA LIMA

ODONTO VISION: SISTEMA CAD PARA SEGMENTAÇÃO AUTOMÁTICA DE DENTES  
EM RADIOGRAFIAS PANORÂMICAS

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em SISTEMAS DE INFORMAÇÃO do CAMPUS DE CRATEÚS da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em SISTEMAS DE INFORMAÇÃO.

Aprovada em:

BANCA EXAMINADORA

---

Prof. Dr. Bruno Riccelli Dos Santos  
Silva (Orientador)  
Universidade Federal do Ceará (UFC)

---

Prof. Me. Ítalo Mendes da Silva Ribeiro  
Universidade Federal do Ceará (UFC)

---

Me. Pedro Crosara Motta  
Universidade Federal do Rio de Janeiro (UFRJ)

---

Dra. Débora Ferreira de Assis  
Instituto Atlântico

A Deus, por me conceder força, sabedoria e perseverança ao longo de toda esta caminhada. À minha família, por acreditar em mim e apoiar minhas escolhas, pelo incentivo e apoio ao longo desta trajetória.

## AGRADECIMENTOS

Primeiramente, agradeço a Deus, por toda força, coragem e resiliência que me permitiram chegar até este momento, guiando meus passos e sustentando minha determinação ao longo de toda essa trajetória.

A toda a minha família, pelo amor, apoio e incentivo constantes, que sempre estiveram ao meu lado nos momentos de desafio e nos momentos de conquista. Em especial, à minha mãe Leidiana, meu pai Erivaldo, minha irmã Aparecida, meu avô Antônio Chaga, minha avó Maria José e minha tia Das Dores: vocês, ao longo de todo esse tempo, vivenciaram comigo cada passo da minha caminhada até chegar a este momento. O apoio de vocês foi fundamental, e o carinho e exemplo de vida de cada um me inspiraram a perseverar.

Ao Prof. Dr. Bruno Riccelli dos Santos Silva, por sua orientação magistral ao longo deste trabalho. Sua dedicação, conhecimento e paciência foram essenciais para o desenvolvimento da pesquisa, inspirando-me a superar desafios e a aprimorar minhas habilidades acadêmicas e científicas, bem como a desenvolver competências pessoais, como organização, disciplina, resiliência e pensamento crítico.

Aos membros da banca avaliadora, pelas valiosas contribuições que enriqueceram ainda mais este trabalho.

Agradeço a todo o corpo docente da Universidade Federal do Ceará - Campus de Crateús, pela dedicação, compromisso e excelência no ensino, que proporcionaram uma formação sólida, inspiradora e transformadora. Aos demais funcionários da instituição, registro minha gratidão pelo apoio, atendimento e contribuição para o bom funcionamento da universidade, tornando possível o desenvolvimento acadêmico e pessoal de todos os alunos.

Por fim, agradeço aos meus amigos de graduação pelo apoio, incentivo e companheirismo ao longo desta jornada. Cada conversa, cada momento de descontração e cada palavra de incentivo contribuíram para que eu mantivesse motivação e equilíbrio durante toda a trajetória acadêmica. As amizades feitas durante esse período permanecerão como lembranças valiosas e fonte de inspiração para toda a minha vida.

“ A ferrovia que leva ao sucesso é construída em cima de um solo de humildade com pesados trilhos chamados erros que somente são fixados numa linha reta com maciços pregos de perseverança.”

(Eduardo Siqueira Filho)

## RESUMO

A segmentação de imagens é fundamental em aplicações de análise automática, pois permite a identificação e delimitação precisa de estruturas relevantes. No contexto odontológico, essa tarefa é especialmente importante em radiografias panorâmicas, amplamente utilizadas para fornecer visão completa da estrutura dentária e óssea. Entretanto, a interpretação manual dessas imagens é demorada, sujeita à variabilidade entre profissionais e afetada por ruído e baixo contraste. Avanços recentes em aprendizado profundo, especialmente Redes Neurais Convolucionais (CNNs), têm permitido segmentações mais precisas por meio da extração automática de características relevantes. Assim, este trabalho tem como objetivo desenvolver um sistema CAD para realizar a segmentação automática de dentes a partir de radiografias panorâmicas, com base em uma análise comparativa do desempenho de cinco arquiteturas baseadas na U-Net. O estudo foi conduzido utilizando o conjunto de dados Children's Dental Panoramic Radiographs Dataset, no qual as imagens passaram por uma etapa de pré-processamento e os modelos foram treinados utilizando validação cruzada K-Fold, combinada com Grid Search para a otimização dos hiperparâmetros. Os resultados da etapa de validação externa, realizada em dois subconjuntos do conjunto geral, demonstraram que, no primeiro deles, o Dataset and code, a W-Net apresentou o melhor desempenho, com Sensibilidade 0,809, E-MEASURE 0,925, IOU 0,785 e DSC 0,879, aspectos que indicam maior capacidade de segmentação das regiões dentárias. No segundo subconjunto, Panoramic Radiography Database, a Attention U-Net se sobressaiu, alcançando Sensibilidade 0,965, E-MEASURE 0,975, IOU 0,858 e DSC 0,924, apresentando maior precisão, preservação da forma dentária e menor fragmentação das estruturas. O teste de Wilcoxon indicou que a U-Net++ e a U-Net 3+ apresentaram diferenças significativas em relação à U-Net clássica, enquanto a W-Net e a Attention U-Net tiveram desempenho equivalente. Por fim, considerando conjuntamente os resultados das análises quantitativas e qualitativas das máscaras segmentadas, bem como o tempo de inferência necessário para a geração das segmentações, a Attention U-Net foi selecionada como a arquitetura final a ser integrada ao sistema CAD.

**Palavras-chave:** Segmentação de imagens médicas. Sistema CAD. Modelos de segmentação. Redes Neurais Convolucionais. Aprendizado profundo.

## ABSTRACT

Image segmentation is fundamental in automatic analysis applications, as it allows for the precise identification and delineation of relevant structures. In the dental context, this task is especially important in panoramic radiographs, which are widely used to provide a comprehensive view of the dental and bone structures. However, manual interpretation of these images is time-consuming, subject to inter-professional variability, and affected by noise and low contrast. Recent advances in deep learning, particularly Convolutional Neural Networks (CNNs), have enabled more accurate segmentations through the automatic extraction of relevant features. Thus, this study aims to develop a CAD system to perform automatic tooth segmentation from panoramic radiographs, based on a comparative analysis of the performance of five U-Net-based architectures. The study was conducted using the Children's Dental Panoramic Radiographs Dataset, in which the images underwent a preprocessing stage, and the models were trained using K-Fold cross-validation combined with Grid Search for hyperparameter optimization. The results of the external validation stage, conducted on two subsets of the overall dataset, showed that in the first subset, Dataset and code, the W-Net achieved the best performance, with Sensitivity 0.809, E-MEASURE 0.925, IOU 0.785, and DSC 0.879, indicating a higher capacity for segmenting dental regions. In the second subset, Panoramic Radiography Database, the Attention U-Net stood out, reaching Sensitivity 0.965, E-MEASURE 0.975, IOU 0.858, and DSC 0.924, showing higher accuracy, preservation of tooth shape, and lower fragmentation of structures. The Wilcoxon test indicated that U-Net++ and U-Net 3+ showed significant differences compared to the classical U-Net, while W-Net and Attention U-Net had equivalent performance. Finally, considering the results of both quantitative and qualitative analyses of the segmented masks, as well as the inference time required to generate the segmentations, the Attention U-Net was selected as the final architecture to be integrated into the CAD system.

**Keywords:** Medical image segmentation. CAD system. Segmentation models. Convolutional Neural Networks. Deep learning.

## LISTA DE ILUSTRAÇÕES

Figura 1 – Estrutura da Mandíbula . . . . .	20
Figura 2 – Estrutura do dente . . . . .	22
Figura 3 – Hierarquia de Aprendizado . . . . .	24
Figura 4 – Arquitetura Rede Neural Convolucional (CNN) . . . . .	27
Figura 5 – Operação de convolução . . . . .	28
Figura 6 – Camada totalmente conectada . . . . .	32
Figura 7 – Arquitetura U-Net . . . . .	33
Figura 8 – Arquitetura UNet++ . . . . .	34
Figura 9 – Arquitetura UNet 3+ . . . . .	35
Figura 10 – Arquitetura Attention U-Net . . . . .	36
Figura 11 – Arquitetura W-Net . . . . .	37
Figura 12 – Diagrama da metodologia empregada neste trabalho . . . . .	51
Figura 13 – Estrutura do conjunto de dados . . . . .	52
Figura 14 – Fluxo do Sistema CAD . . . . .	59
Figura 15 – Resultados qualitativos da segmentação: A - Dataset and code; B - Panoramic Radiography Database . . . . .	66
Figura 16 – Página inicial do Sistema CAD . . . . .	71
Figura 17 – Upload da radiografia . . . . .	72
Figura 18 – Resultado da máscara segmentada . . . . .	72
Figura 19 – PDF gerado com a radiografia e a máscara correspondente . . . . .	73

## LISTA DE TABELAS

Tabela 1 – Comparação entre trabalhos relacionados e este trabalho . . . . .	50
Tabela 2 – Melhor configuração de hiperparâmetros de cada arquitetura . . . . .	60
Tabela 3 – Resultados das métricas no conjunto de teste. . . . .	61
Tabela 4 – Resultados das métricas no Dataset and code . . . . .	63
Tabela 5 – Resultados das métricas no Panoramic Radiography Database . . . . .	64
Tabela 6 – Teste de Wilcoxon: comparação entre arquiteturas . . . . .	68
Tabela 7 – Tempo de inferência das redes . . . . .	69

## **LISTA DE ABREVIATURAS E SIGLAS**

AGs	Attention Gates
CNN	Rede Neural Convolutacional
DSC	Coefficiente de Similaridade de Dice
E-MEASURE	Medida de Alinhamento Aprimorada
IA	Inteligência Artificial
IOU	Intersecção sobre União
MAE	Erro Médio Absoluto
ReLU	Unidade Linear Retificada
SADC	Sistemas de Apoio à Decisão Clínica

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>15</b>
<b>1.1</b>	<b>Objetivo Geral</b>	<b>17</b>
<b>1.2</b>	<b>Objetivos Específicos</b>	<b>17</b>
<b>1.3</b>	<b>Contribuições</b>	<b>17</b>
<b>1.4</b>	<b>Organização do Trabalho</b>	<b>18</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>19</b>
<b>2.1</b>	<b>Anatomia Mandibular</b>	<b>19</b>
<b>2.1.1</b>	<i>Estrutura da Mandíbula</i>	<b>19</b>
<b>2.2</b>	<b>Anatomia Dentária</b>	<b>21</b>
<b>2.2.1</b>	<i>Estrutura Dentária</i>	<b>21</b>
<b>2.3</b>	<b>Inteligência Artificial</b>	<b>22</b>
<b>2.4</b>	<b>Aprendizado Profundo</b>	<b>24</b>
<b>2.5</b>	<b>Sistemas de Apoio à Decisão Clínica com IA</b>	<b>25</b>
<b>2.6</b>	<b>Segmentação de Imagens Médicas</b>	<b>26</b>
<b>2.7</b>	<b>Redes Neurais Convolucionais (CNNs)</b>	<b>26</b>
<b>2.8</b>	<b>Camada Convolucional (<i>Convolutional Layer</i>)</b>	<b>28</b>
<b>2.8.0.1</b>	<i>Mapa de Características (Feature Map)</i>	<b>29</b>
<b>2.8.0.2</b>	<i>Passo (Stride)</i>	<b>29</b>
<b>2.8.0.3</b>	<i>Filtros (Kernel)</i>	<b>29</b>
<b>2.8.0.4</b>	<i>Preenchimento (Padding)</i>	<b>30</b>
<b>2.8.0.5</b>	<i>Função de Ativação (Activation Function)</i>	<b>30</b>
<b>2.9</b>	<b>Camada de Agrupamento (<i>Pooling Layer</i>)</b>	<b>30</b>
<b>2.9.0.1</b>	<i>Agrupamento Máximo (Max pooling)</i>	<b>31</b>
<b>2.10</b>	<b>Camada Totalmente Conectada (<i>Fully Connected Layer</i>)</b>	<b>31</b>
<b>2.11</b>	<b>Modelos de Segmentação</b>	<b>32</b>
<b>2.11.1</b>	<i>U-net</i>	<b>32</b>
<b>2.11.2</b>	<i>Unet ++</i>	<b>34</b>
<b>2.11.3</b>	<i>Unet 3+</i>	<b>35</b>
<b>2.11.4</b>	<i>Attention U-Net</i>	<b>36</b>
<b>2.11.5</b>	<i>W-Net</i>	<b>37</b>

2.12	<b>Validação Cruzada</b>	38
2.13	<b>Grid Search</b>	38
2.14	<b>Teste Estatístico de Wilcoxon</b>	39
2.15	<b>Métricas de Avaliação</b>	39
2.15.1	<i>Acurácia</i>	40
2.15.2	<i>Especificidade</i>	40
2.15.3	<i>Sensibilidade</i>	41
2.15.4	<i>Coefficiente de Similaridade de Dice (DSC)</i>	41
2.15.5	<i>Erro Médio Absoluto (MAE)</i>	41
2.15.6	<i>Intersecção sobre União (IOU)</i>	42
2.15.7	<i>Medida de Alinhamento Aprimorada (E-MEASURE)</i>	42
3	<b>TRABALHOS RELACIONADOS</b>	44
4	<b>METODOLOGIA</b>	51
4.1	<b>Visão Geral</b>	51
4.2	<b>Aquisição do Conjunto de Dados</b>	52
4.3	<b>Pré-processamento e Preparação dos Dados</b>	53
4.4	<b>Arquiteturas de Redes Neurais para Segmentação</b>	54
4.5	<b>Treinamento dos modelos de segmentação</b>	54
4.5.1	<i>Aumento de Dados</i>	55
4.5.2	<i>Ajuste de Hiperparâmetros e Validação Cruzada</i>	55
4.6	<b>Validação externa</b>	56
4.7	<b>Avaliação de desempenho</b>	57
4.8	<b>Ambiente de desenvolvimento</b>	58
4.9	<b>Análise Estatística: Teste de Wilcoxon</b>	58
4.10	<b>Desenvolvimento do Sistema CAD</b>	58
5	<b>RESULTADOS E DISCUSSÕES</b>	60
5.1	<b>Resultados da seleção de hiperparâmetros por arquitetura</b>	60
5.2	<b>Resultados da avaliação no conjunto de teste</b>	61
5.3	<b>Resultados no conjunto de dados Dataset and code</b>	62
5.4	<b>Resultados no conjunto de dados Panoramic Radiography Database</b>	64
5.5	<b>Análise qualitativa das máscaras segmentadas</b>	66
5.6	<b>Análise estatística por meio do teste de Wilcoxon</b>	67

<b>5.7</b>	<b>Seleção da melhor arquitetura . . . . .</b>	<b>69</b>
<b>5.8</b>	<b>Sistema CAD . . . . .</b>	<b>70</b>
<b>5.9</b>	<b>Limitações do trabalho . . . . .</b>	<b>74</b>
<b>6</b>	<b>CONCLUSÕES E TRABALHOS FUTUROS . . . . .</b>	<b>75</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>77</b>

## 1 INTRODUÇÃO

A saúde bucal é definida como o estado da boca, dos dentes e das estruturas orofaciais que permite aos indivíduos desempenhar funções essenciais, como comer, respirar e falar, abrangendo também dimensões psicossociais, tais como autoconfiança, bem-estar e a capacidade de socializar e trabalhar sem dor, desconforto ou constrangimento. Além disso, essa condição varia ao longo da vida, desde a infância até a velhice, e exerce papel fundamental na saúde geral, ao auxiliar os indivíduos a participarem da sociedade e a alcançarem seu potencial (WORLD HEALTH ORGANIZATION, 2022a). Assim, uma boa saúde bucal reflete a capacidade do indivíduo de se adaptar às mudanças fisiológicas ao longo da vida, o que lhe permite manter a integridade da boca e dos dentes por meio do autocuidado contínuo e independente (PERES *et al.*, 2019).

Consequentemente, as doenças bucais abrangem diversas condições, como cáries, perda de dentes e traumatismos orodentais. Essas alterações estão entre as doenças crônicas não transmissíveis mais prevalentes no mundo, afetando aproximadamente 3,5 bilhões de pessoas (WORLD HEALTH ORGANIZATION, 2022b). Desse modo, as doenças dentárias tornaram-se uma das condições mais comuns na ciência médica contemporânea e têm apresentado crescimento significativo nos últimos anos. Embora os dentes sejam estruturas altamente resistentes e duráveis, sujeitos a constante uso, continuam vulneráveis a diversas patologias que afetam diretamente sua integridade e função. Portanto, o diagnóstico precoce é essencial para prevenir complicações e preservar a integridade das estruturas dentárias. (ZANNAH *et al.*, 2024).

Para formular o diagnóstico de doenças bucais, as radiografias panorâmicas, conhecidas como ortopantomografia, são ferramentas fundamentais na odontologia. Elas fornecem imagens abrangentes de toda a arcada dentária, como dentes, maxilares e estruturas adjacentes, o que as torna cruciais para a identificação de diversas condições, como cáries, fraturas e alterações ósseas. No entanto, a interpretação manual dessas radiografias pode ser demorada e sujeita a variações entre profissionais da odontologia (ŞAHİN *et al.*, 2024). Além dessas limitações, fatores técnicos como baixa resolução, contraste inadequado e presença de ruído também dificultam o processo, tornando-o ainda mais desafiador e propenso a erros diagnósticos. Nesse contexto, a análise automática dessas imagens surge como uma ferramenta promissora, capaz de aumentar a precisão diagnóstica, reduzir o tempo de triagem e otimizar os custos clínicos. Entre essas soluções, destaca-se a segmentação automática de imagens médicas, que consiste em classificar cada pixel da imagem conforme os objetos de interesse (NADER *et al.*, 2022).

A segmentação de imagens consiste em dividir uma imagem em partes que tenham forte correlação com objetos ou áreas do mundo real contidos na cena, chamados segmentos de pixels conectados, analisando critérios de similaridade (BRAHMI; JDEY, 2024). Esse processo tem como objetivo identificar o objeto ou região ao qual cada pixel pertence, classificando-os em diferentes classes (SAHIN *et al.*, 2024). No contexto de imagens médicas, a segmentação desempenha um papel fundamental no diagnóstico auxiliado por computador e na medicina inteligente, pois visa tornar as alterações de estruturas anatômicas ou patológicas mais claras nas imagens (WANG *et al.*, 2022). Nesse cenário, especificamente no diagnóstico de doenças dentárias, como cárie, lesões periapicais e doença periodontal, a segmentação constitui um elemento-chave ao auxiliar dentistas no enfrentamento das dificuldades para detectar essas condições a olho nu. Ao realçar os limites das estruturas dentárias, ela pode facilitar a identificação e distinção entre os dentes, algo frequentemente dificultado em radiografias devido à sobreposição de estruturas que compromete a interpretação visual (ZANNAH *et al.*, 2024).

Nesse contexto, o aprendizado profundo tem se consolidado como uma das principais abordagens na análise de imagens médicas, com ampla aplicação em tarefas de detecção de objetos, classificação, segmentação e registro (ZHONG *et al.*, 2025). Essa abordagem, que constitui um subconjunto do aprendizado de máquina, utiliza redes neurais artificiais com múltiplas camadas para aprender e reconhecer padrões em dados. Paralelamente, nos últimos anos as Redes Neurais Convolucionais (CNNs) transformaram a análise de imagens médicas, alcançando desempenho notável em tarefas como classificação, segmentação, detecção e reconstrução de imagens (BRAHMI; JDEY, 2024). Sua capacidade de extrair automaticamente informações relevantes elimina a dependência de técnicas manuais e pré-processamento excessivo. Consequentemente, as CNNs têm sido amplamente utilizadas na segmentação de imagens médicas, obtendo resultados expressivos e fortalecendo diagnósticos clínicos e sistemas de apoio à decisão (LIU *et al.*, 2021).

Portanto, este trabalho visa auxiliar no diagnóstico clínico odontológico ao demonstrar o potencial do aprendizado profundo e das CNNs na segmentação automática de estruturas dentárias em radiografias panorâmicas, além de apresentar um sistema CAD que integra o modelo de segmentação, permitindo a avaliação de seu desempenho em um ambiente computacional aplicado.

## 1.1 Objetivo Geral

Desenvolver um sistema CAD para segmentação automática de dentes em radiografias panorâmicas, a partir de uma análise comparativa do desempenho de arquiteturas baseadas na U-Net.

## 1.2 Objetivos Específicos

Para alcançar o objetivo geral, no contexto deste trabalho, foram estabelecidos os seguintes objetivos específicos:

- Realizar o pré-processamento das radiografias panorâmicas e aplicar técnicas de aumento de dados (*Data Augmentation*).
- Implementar e comparar diferentes modelos baseados na arquitetura U-Net para segmentação automática das estruturas dentárias.
- Ajustar os hiperparâmetros dos modelos por meio de busca em grade (*Grid Search*) e validar o desempenho com validação cruzada (*k-fold*).
- Aplicar validação externa a fim de avaliar a capacidade de generalização dos modelos treinados.
- Avaliar o desempenho dos modelos usando métricas quantitativas, como sensibilidade, acurácia, especificidade, DSC, IOU, E-measure e MAE.
- Comparar os resultados dos modelos usando o teste de Wilcoxon.
- Desenvolver um sistema CAD para envio de radiografias panorâmicas e geração automática das segmentações pelo modelo treinado.

## 1.3 Contribuições

Este trabalho apresenta contribuições significativas para a área da odontologia, ao demonstrar o potencial das CNNs baseadas na arquitetura U-Net para a segmentação automática de estruturas dentárias em radiografias panorâmicas. Além disso, contribui com o desenvolvimento de um sistema CAD, que torna o modelo acessível a profissionais da odontologia, permitindo o envio de radiografias e a visualização automática dos resultados da segmentação. Essa ferramenta auxilia na análise das estruturas dentárias, uma vez que algumas delas podem não ser claramente perceptíveis a olho nu, o que promove diagnósticos mais rápidos, precisos e padronizados, o qual oferece suporte direto à tomada de decisões diagnósticas diárias.

## **1.4 Organização do Trabalho**

Este trabalho está organizado de forma que no Capítulo 2 está descrita a fundamentação teórica sobre os temas abordados no trabalho. O Capítulo 3 discute os trabalhos relacionados, com ênfase nas principais abordagens identificadas na literatura. O Capítulo 4 descreve a metodologia adotada com o detalhamento dos experimentos realizados. O Capítulo 5 apresenta e analisa os resultados obtidos, discutindo o desempenho das abordagens avaliadas. Por fim, o Capítulo 6 sintetiza as conclusões, ressalta as contribuições alcançadas e aponta perspectivas para trabalhos futuros.

## 2 FUNDAMENTAÇÃO TEÓRICA

No presente capítulo, são apresentados os elementos teóricos que fundamentam o desenvolvimento deste trabalho. Inicialmente, são abordados conceitos de anatomia mandibular e anatomia dentária, com destaque para suas estruturas. Em seguida, são apresentados os fundamentos de Inteligência Artificial (IA), aprendizado profundo e sistemas de apoio à decisão clínica baseados em inteligência artificial. Posteriormente, discute-se a segmentação de imagens médicas e as Redes Neurais Convolucionais. Na sequência, são descritos os principais modelos de segmentação baseados na arquitetura U-Net, ou seja, U-Net, U-Net++, U-Net 3+, Attention U-Net e W-Net. Por fim, são apresentados os conceitos de validação cruzada, *Grid Search*, teste estatístico de Wilcoxon e as métricas de avaliação empregadas neste trabalho. Os tópicos deste capítulo foram escolhidos para fornecer uma base teórica sólida que apoie a compreensão do tema deste trabalho.

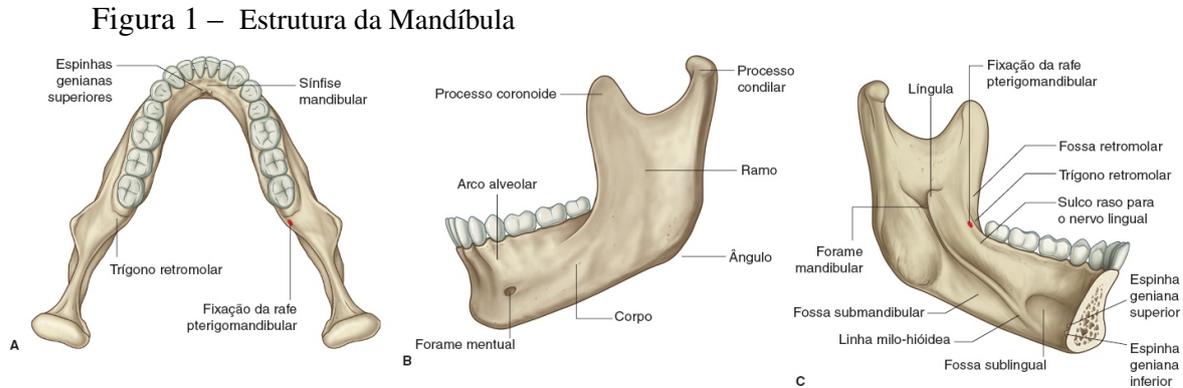
### 2.1 Anatomia Mandibular

A mandíbula é o maior osso do crânio humano, formando a linha inferior e moldando o contorno do terço inferior da face (BREELAND *et al.*, 2019). Além disso, segundo Vollmer *et al.* (2000), trata-se de uma estrutura óssea especializada, em complexa sinergia com músculos, articulações e dentes, cuja forma e função são altamente adaptadas às exigências do sistema mastigatório. Ao contrário dos demais ossos do crânio, ela não se articula com os ossos adjacentes por suturas, mas sim por uma articulação sinovial denominada articulação temporomandibular (CARMO, 2023). Dessa forma, essa articulação, localizada entre a cabeça mandibular e a fossa temporomandibular, permite o livre movimento, contribuindo para a mastigação e a digestão. Além disso, seu formato pode variar conforme os estilos de vida e hábitos de mastigação, de modo que as características morfológicas da mandíbula diferem entre indivíduos de diferentes idades, origens e comportamentos (IWANAGA; TUBBS, 2022).

#### 2.1.1 Estrutura da Mandíbula

A mandíbula é composta por um corpo em forma de “U”, projetado anteroposteriormente, cujas extremidades posteriores formam os ângulos goníacos bilaterais, a partir dos quais os ramos se estendem verticalmente em direção à articulação com a base do crânio (BREELAND *et al.*, 2019). A Figura 1 apresenta a mandíbula em três projeções: superior (A), lateral (B)

e medial (C), evidenciando estruturas como o ramo mandibular, o arco alveolar, os processos coronóide e condilar, além de fossas e forames de relevância clínica.



Fonte: Adaptada de DRAKE et al. (2015)

Conforme ilustrado na Figura 1B, a face superior do corpo da mandíbula sustenta o arco alveolar, responsável pela fixação dos dentes inferiores e fundamental para a mastigação. Nessa mesma face externa localiza-se o forame mental, abertura por onde emerge o nervo mental, responsável pela sensibilidade do lábio inferior e da região do mento. Na região anterior da mandíbula (Figura 1A), observa-se a sínfise mandibular, situada na linha média. Logo após essa região, na face interna do osso, encontram-se as espinhas genianas superiores e inferiores (Figura 1C), que servem como pontos de inserção muscular. As espinhas superiores relacionam-se aos movimentos da língua, enquanto as inferiores conectam músculos ao osso hioide, contribuindo para a sustentação do assoalho da boca (DRAKE *et al.*, 2015).

A partir dessas estruturas, estende-se a linha milo-hióidea (Figura 1C), ao longo da face medial da mandíbula até a região inferior ao último molar. Essa linha separa duas áreas importantes: superiormente localiza-se a fossa sublingual, que abriga a glândula sublingual, e inferiormente a fossa submandibular, onde se aloja a glândula submandibular. Entre a linha milo-hióidea e a região posterior do arco alveolar, observa-se um sulco raso para a passagem do nervo lingual, responsável pela sensibilidade da língua. Posteriormente ao último molar, identifica-se o trígono retromolar, local de fixação da rafe pterigomandibular, que conecta músculos da cavidade oral e da faringe. Por fim, conforme apresentado na Figura 1B, o ramo da mandíbula possui formato aproximadamente quadrangular, destacando-se os processos coronóide e condilar. Em sua face medial encontra-se o forame mandibular, por onde passam o nervo e os vasos alveolares inferiores, responsáveis pela sensibilidade e vascularização dos dentes inferiores (DRAKE *et al.*, 2015).

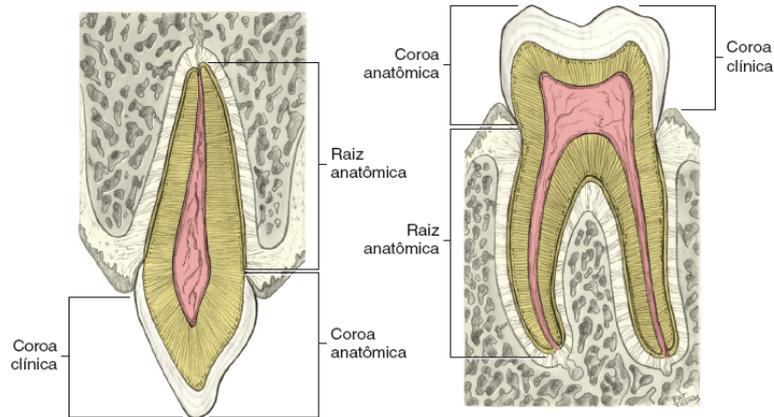
## **2.2 Anatomia Dentária**

A anatomia dentária é um campo especializado que se dedica ao estudo da estrutura, do desenvolvimento e do arranjo dos dentes humanos. Nesse contexto, abrange tanto os dentes primários (decíduos) quanto os permanentes, contemplando suas diferentes classificações, como incisivos, caninos, pré-molares e molares, bem como a análise dos principais componentes que os constituem, como coroa, raiz, esmalte, dentina e polpa. Além disso, envolve o exame das estruturas de suporte, denominadas periodonto, que englobam gengiva, ligamento periodontal, cimento e osso alveolar. Assim, compreender a anatomia dentária contribui para a promoção de melhores cuidados bucais e favorece a valorização da forma e da função dos dentes (ROBERT, 2023).

### **2.2.1 Estrutura Dentária**

Os dentes são apêndices multifuncionais, essenciais para funções humanas básicas, como a alimentação e a fala. Outrossim, são formados por diferentes tecidos, de densidade e dureza variadas, o que lhes confere resistência às forças e ao desgaste característicos da mastigação (MORRIS; TADI, 2023). Por conseguinte, os dentes configuram-se como uma das estruturas anatômicas e histológicas mais singulares e complexas do corpo, cuja composição tecidual é exclusiva da cavidade oral e restrita às estruturas dentárias (CARMO; CHAVES, 2023). No que se refere à anatomia, cada dente divide-se em duas regiões principais: coroa e raiz. A coroa corresponde à porção visível na cavidade oral, enquanto a raiz encontra-se inserida no processo alveolar dos maxilares superior e inferior, sendo fixada por meio do ligamento periodontal (MORRIS; TADI, 2023). A Figura 2 apresenta a estrutura de um dente, evidenciando a coroa e a raiz em suas porções anatômicas e clínicas.

Figura 2 – Estrutura do dente



Fonte: Adaptada de FEHRENBACH E POPOWICS (2020)

Conforme observado na figura, a coroa anatômica corresponde à porção do dente totalmente recoberta por esmalte, independentemente de estar ou não exposta na cavidade oral. Essa região tende a permanecer relativamente constante ao longo da vida, sofrendo alterações principalmente em decorrência de atrição ou outros tipos de desgaste físico. Em contraste, a coroa clínica refere-se à parte da coroa anatômica que se encontra visível na cavidade oral, ou seja, à porção não recoberta pela gengiva. Diferentemente da coroa anatômica, sua extensão pode variar ao longo do tempo, acompanhando mudanças na posição gengival, como processos de retração ou inflamação (FEHRENBACH; POPOWICS, 2020). Além disso, a coloração dessa região exposta pode variar do branco perolado ao amarelo, dependendo de fatores como idade, higiene oral e hábitos de vida do indivíduo (CARMO; CHAVES, 2023).

De maneira semelhante, a raiz anatômica corresponde à porção do dente recoberta por cemento, estrutura responsável por sua fixação ao osso alveolar. Já a raiz clínica designa a parte da raiz anatômica que se encontra exposta na cavidade oral, cuja extensão pode aumentar ao longo do tempo em decorrência da recessão gengival (FEHRENBACH; POPOWICS, 2020). Além dessas distinções, os dentes também apresentam variações quanto ao número de raízes. Em geral, incisivos, caninos e primeiros pré-molares possuem raiz única, enquanto segundos pré-molares e molares podem apresentar duas ou até três raízes, como exemplificado pelo dente posterior ilustrado na figura (CARMO; CHAVES, 2023).

### 2.3 Inteligência Artificial

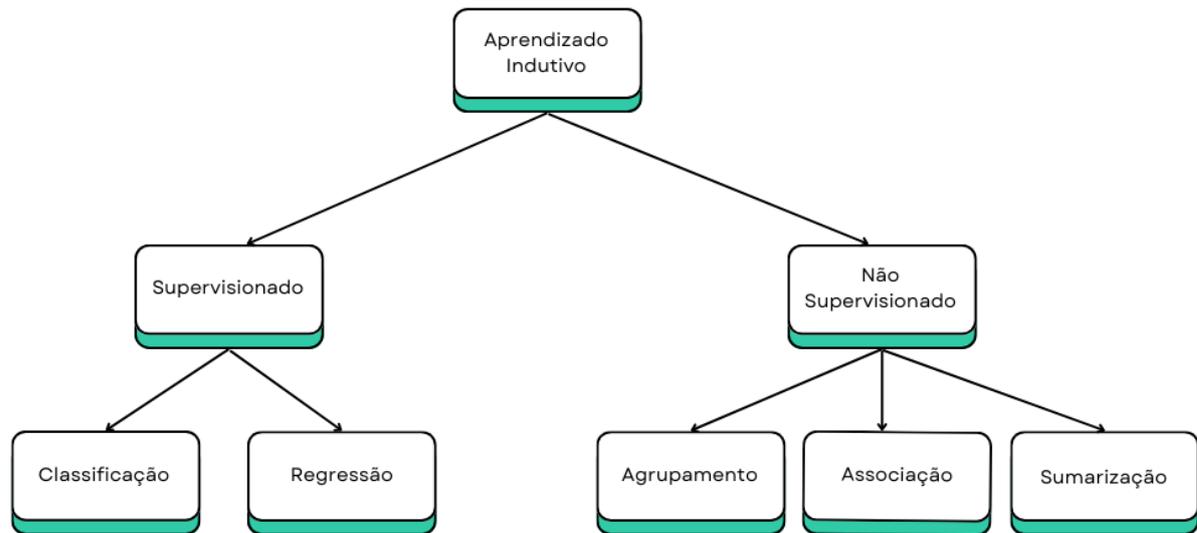
Durante décadas, a área de IA era vista como um campo majoritariamente teórico, aplicado a problemas pequenos e intelectualmente interessantes, mas com pouco impacto prático.

Nesse período, a maior parte das demandas reais de computação era solucionada por meio da programação tradicional, com a definição explícita de cada passo. Esse cenário mudou a partir dos anos 1970, quando técnicas de IA passaram a ser aplicadas a problemas concretos, especialmente por meio da incorporação do conhecimento de especialistas de áreas como a medicina em sistemas computacionais. Desse movimento surgiram os Sistemas Especialistas, ou Sistemas Baseados em Conhecimento, considerados um dos primeiros grandes avanços rumo ao uso prático da IA (FACELI *et al.*, 2011).

Diante desse contexto histórico, a IA refere-se ao desenvolvimento de sistemas computacionais capazes de executar tarefas que normalmente exigiriam habilidades humanas, como raciocínio, tomada de decisão e identificação de padrões. Em essência, trata-se de criar mecanismos capazes de analisar informações, aprender com dados e agir de forma autônoma em diferentes situações (RAO *et al.*, 2024). Além disso, a IA encontra-se hoje amplamente disseminada e integrada a diversos sistemas e tecnologias, pois, mais do que uma ferramenta isolada, seus métodos constituem a base de muitos processos modernos de automação, análise de dados e tomada de decisão (COPELAND, 2015).

Concomitantemente, a IA apresenta aplicações em diversas áreas, como agropecuária, bioinformática, mineração de dados e textos, robótica e saúde. Na agropecuária, ela otimiza o manejo de cultivos, prevê pragas e aprimora a qualidade da produção, já na bioinformática, apoia a análise de genes, proteínas e padrões biológicos complexos, na mineração de dados e textos, identifica padrões relevantes em grandes volumes de informação e, por fim, na saúde, oferece suporte ao diagnóstico, à análise de exames e ao acompanhamento de pacientes (FACELI *et al.*, 2011). Diante desse cenário de ampla aplicação da IA em diferentes domínios, torna-se essencial compreender como esses sistemas aprendem a partir dos dados. A Figura 3 apresenta a hierarquia dos principais tipos de tarefas de aprendizado.

Figura 3 – Hierarquia de Aprendizado



Fonte: Adaptada de FACELI et al. (2011)

Na parte inicial da Figura 3, temos o aprendizado indutivo, cuja função é generalizar padrões a partir dos dados. Nesse sentido, a partir dele emergem duas modalidades principais de aprendizado: o supervisionado e o não supervisionado. O primeiro, de natureza preditiva, utiliza exemplos rotulados para aprender relações que permitam antecipar informações desconhecidas. Nessa modalidade, a distinção ocorre pelo tipo de rótulo, ou seja, problemas de classificação envolvem saídas discretas, enquanto problemas de regressão lidam com valores contínuos. Já o aprendizado não supervisionado, de caráter descritivo, busca identificar estruturas internas nos dados sem o uso de rótulos. Entre suas tarefas mais comuns estão o agrupamento, que organiza instâncias por similaridade, a sumarização, que gera representações compactas mantendo a informação essencial e a associação, que descobre combinações frequentes e relações entre atributos (FACELI *et al.*, 2011).

## 2.4 Aprendizado Profundo

O aprendizado profundo é um subconjunto do aprendizado de máquina e uma tecnologia poderosa construída sobre redes neurais, caracterizada por uma arquitetura de múltiplas camadas que permite aprender características complexas e hierárquicas a partir de grandes volumes de dados, impulsionada por avanços recentes em hardware e pesquisa (SHINDE; SHAH, 2018). O método de aprendizado profundo simula a rede neural humana e, ao combinar múltiplas camadas de processamento não linear, os dados originais são gradualmente abstraídos, camada por camada, de modo que diferentes níveis de características são extraídos e utilizados em tarefas

como detecção, classificação ou segmentação de alvos. Desse modo, a vantagem do aprendizado profundo reside na substituição da aquisição manual de características por um aprendizado não supervisionado ou semi supervisionado, além do uso de algoritmos capazes de realizar uma extração hierárquica eficiente dessas características (CAI *et al.*, 2020).

Nesse contexto, nos últimos anos, as técnicas de aprendizado profundo têm sido amplamente aplicadas na segmentação de imagens médicas, visto que, ao aprender representações semânticas das imagens, os modelos conseguem melhorar a precisão da segmentação e se adaptar de forma flexível a diferentes conjuntos de dados e tarefas (LECUN *et al.*, 2015).

## **2.5 Sistemas de Apoio à Decisão Clínica com IA**

Os Sistemas de Apoio à Decisão Clínica (SADC) são ferramentas computacionais desenvolvidas para auxiliar profissionais de saúde na tomada de decisões médicas. Seu objetivo central é fornecer informações acionáveis, recomendações baseadas em evidências e dados específicos do paciente diretamente no ponto de atendimento. Nos últimos anos, a integração de técnicas de IA impulsionou esses sistemas, tornando-os mais sofisticados. Entre essas técnicas, destacam-se o aprendizado de máquina, o processamento de linguagem natural e o aprendizado profundo, que permitem o processamento e a interpretação de grandes volumes de dados clínicos com maior precisão. Nessa lógica, algoritmos como redes neurais e árvores de decisão identificam padrões relevantes e extraem informações essenciais de conjuntos de dados complexos. Dessa forma, os SADC conseguem oferecer recomendações personalizadas, ajustadas às necessidades de cada paciente, contribuindo para melhores desfechos assistenciais (ELHADDAD; HAMAM, 2024).

Por fim, os SADC integrados à IA apresentam vantagens como a melhoria da qualidade do cuidado e da segurança do paciente, com redução de erros nos diagnósticos, o processamento ágil de informações e o aprimoramento da análise de imagens médicas. Além disso, esses sistemas favorecem a identificação precoce de riscos, otimizam o fluxo de trabalho e automatizam tarefas administrativas, o que possibilita aos profissionais dedicar mais tempo ao atendimento direto ao paciente. Portanto, a integração entre IA e SADC fortalece a confiabilidade em todo o processo de assistência médica (ELHADDAD; HAMAM, 2024).

## 2.6 Segmentação de Imagens Médicas

Nos últimos anos, o amplo sucesso dos modelos de aprendizagem profunda, impulsionaram o desenvolvimento de novas abordagens de segmentação de imagens, capazes de explorar representações complexas aprendidas diretamente a partir dos dados. Como resultado, surgiu uma nova geração de modelos de segmentação que apresenta melhorias de desempenho notáveis, muitas vezes alcançando níveis de precisão superiores aos obtidos por métodos tradicionais (MINAEE *et al.*, 2022). Dessa forma, a segmentação de imagens que é definida como o particionamento de uma imagem em regiões específicas, desempenha um papel essencial em diversas aplicações médicas, especialmente quando automatizada, pois otimiza fluxos clínicos, reduz o tempo de análise e fornece medidas quantitativas relevantes (AZAD *et al.*, 2024).

Nesse contexto, a segmentação de imagens médicas contribui para evidenciar estruturas anatômicas e patológicas, o que a torna fundamental no diagnóstico assistido por computador e na medicina inteligente, devido ao aumento da eficiência e precisão diagnóstica. Atualmente, as tarefas de segmentação são divididas em duas categorias: segmentação semântica, que realiza a classificação em nível de pixel atribuindo uma categoria a cada ponto da imagem, e segmentação de instâncias, que além da classificação pixel a pixel, também diferencia objetos individuais pertencentes à mesma classe (WANG *et al.*, 2022). No presente trabalho, é utilizada a categoria de segmentação semântica.

No âmbito odontológico, segundo Rocha e Endo (2022), a segmentação de imagens dentárias tem ganhado destaque por possibilitar avaliações mais claras, auxiliar na definição de planos de tratamento, favorecer a identificação e a delimitação das estruturas dentárias, além de oferecer suporte à análise clínica e contribuir para diagnósticos mais precisos.

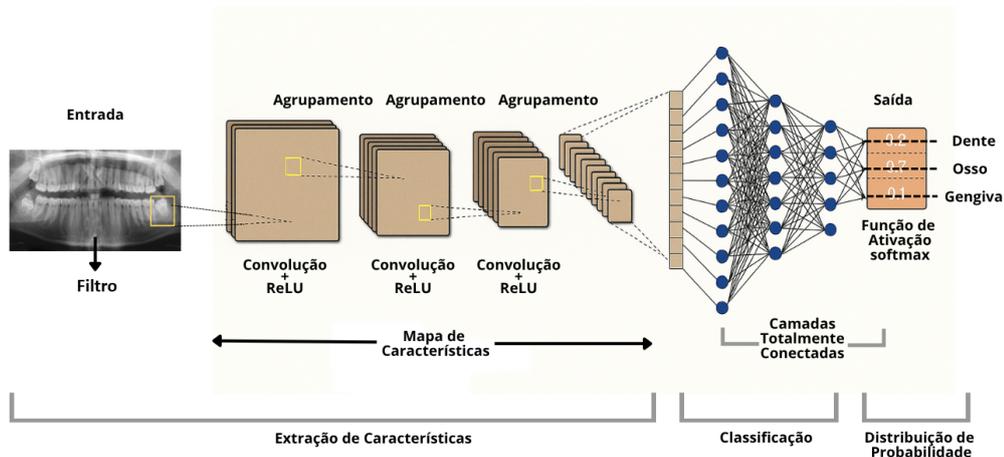
## 2.7 Redes Neurais Convolucionais (CNNs)

As CNNs surgiram como uma evolução das Redes Neurais Artificiais (RNAs), inspiradas no funcionamento do cérebro humano. Desde o modelo de McCulloch e Pitts (1943) e o perceptron de Rosenblatt (1958), o desenvolvimento de redes multicamadas com retropropagação permitiu arquiteturas mais complexas. No final dos anos 1980, LeCun apresentou a LeNet, considerada a primeira CNN aplicada ao reconhecimento de imagens. O avanço decisivo ocorreu em 2012, com a AlexNet, que evidenciou o potencial das redes profundas no ImageNet e impulsionou a adoção das CNNs em aplicações de visão computacional (LI *et al.*, 2021).

Nesse contexto, as CNNs consolidaram-se como um dos modelos mais populares de aprendizado de máquina, tornando-se amplamente utilizadas em diferentes domínios, como visão computacional, reconhecimento de fala e processamento de linguagem natural. No campo da visão computacional, as CNNs destacam-se por alcançar resultados de última geração, desempenho que decorre de sua capacidade de capturar padrões espaciais por meio de uma arquitetura hierárquica de camadas convolucionais, capazes de extrair características em diferentes níveis de abstração (KRICHEN, 2023).

As CNNs são compostas por três tipos de camadas: camada convolucional (*convolutional layer*), camada de agrupamento (*pooling layer*) e camada totalmente conectada (*fully connected layer*). Desse modo, quando essas camadas são empilhadas, uma arquitetura CNN é formada (O'SHEA; NASH, 2015). As duas primeiras camadas realizam a extração de características, enquanto a terceira, mapeia essas características para a saída final. Por conseguinte, à medida que a saída de uma camada é passada para a próxima, as características extraídas organizam-se de forma progressivamente mais complexa (YAMASHITA *et al.*, 2018). A Figura 4 apresenta a ilustração de uma CNN para classificação de imagens.

Figura 4 – Arquitetura Rede Neural Convolucional (CNN)



Fonte: Adaptada de RGUIBI et al. (2022).

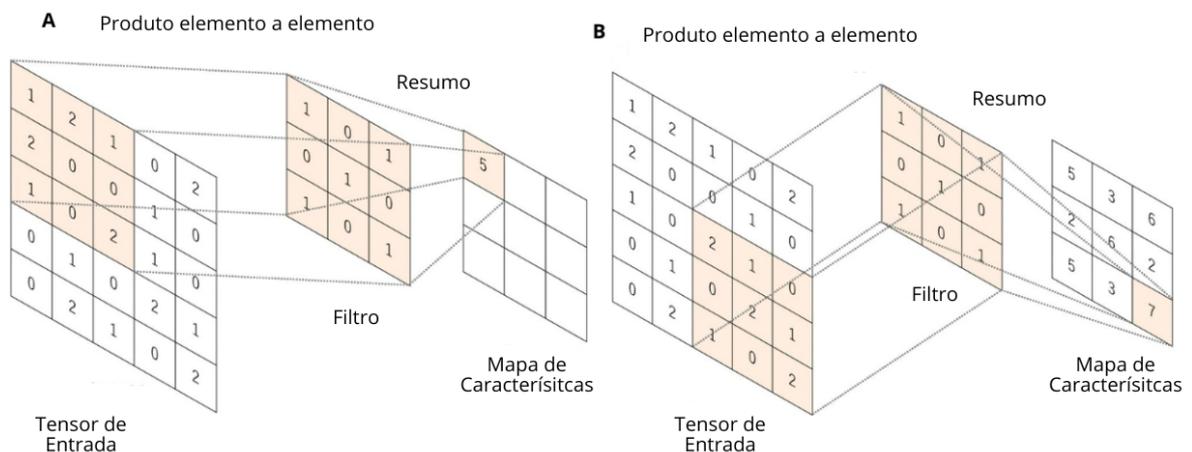
A rede recebe uma radiografia panorâmica, que é processada pela primeira camada convolucional para extrair características iniciais, como contornos e variações de intensidade. Camadas convolucionais posteriores capturam padrões estruturais mais complexos dos dentes e tecidos adjacentes, enquanto camadas subsequentes reduzem a dimensionalidade e preservam as informações mais relevantes. Ao final, as características extraídas são enviadas a camadas totalmente conectadas, que geram uma distribuição de probabilidade para as classes dente, osso

e gengiva, definindo a predição pela classe de maior probabilidade. Os parâmetros do modelo são aprendidos automaticamente por retropropagação, que ajusta os parâmetros para minimizar o erro entre a saída prevista e o rótulo real (RGUIBI *et al.*, 2022).

## 2.8 Camada Convolutiva (*Convolutional Layer*)

A camada convolutiva é um componente fundamental na arquitetura de uma CNN, pois é responsável pela extração de características e, geralmente, combina operações lineares e não lineares, envolvendo a convolução seguida por uma função de ativação. Neste processo, a convolução é uma operação linear usada para extrair características, na qual um filtro, pequenas matrizes aplicadas à imagem de entrada, detecta padrões e extrai informações relevantes. Esse filtro é aplicado sobre a entrada, representada por um tensor, uma estrutura matemática multidimensional que organiza os dados da imagem em altura, largura e canais de intensidade (YAMASHITA *et al.*, 2018). A Figura 5 ilustra uma operação de convolução com um filtro 3×3 que é deslizado sobre o tensor de entrada.

Figura 5 – Operação de convolução



Fonte: Adaptada de YAMASHITA *et al.* (2018).

Nesse contexto, a Figura 5A ilustra que, em cada posição, realiza-se a multiplicação elemento a elemento entre os valores do filtro e a sub-região correspondente do tensor de entrada. Em seguida, a soma desses produtos gera um único valor, o qual é inserido na posição equivalente do tensor de saída, conhecido como mapa de características (*feature map*). Por sua vez, a Figura 5B evidencia a repetição desse procedimento ao longo de toda a entrada, resultando na construção completa do mapa de características. Além disso, o processo é aplicado para cada filtro utilizado,

o que possibilita a geração de múltiplos mapas, cada um enfatizando propriedades distintas da entrada. Dessa forma, diferentes filtros atuam como extratores de características específicos (YAMASHITA *et al.*, 2018).

#### 2.8.0.1 Mapa de Características (Feature Map)

O mapa de características é um componente fundamental das CNNs, pois representa as características extraídas da imagem de entrada pelas camadas convolucionais, o que possibilita a execução de diversas tarefas de reconhecimento visual. Além disso, esses mapas organizam essas características de forma hierárquica, ou seja, nas primeiras camadas, a rede identifica padrões simples, como bordas e texturas, enquanto nas camadas mais profundas combina esses elementos para formar representações cada vez mais complexas. Essa capacidade de transformar gradualmente a imagem em níveis sucessivos de abstração é o que permite às CNNs reconhecerem objetos, estruturas e padrões de maneira eficiente e robusta (KRICHEN, 2023).

#### 2.8.0.2 Passo (Stride)

O passo é a distância entre as posições sucessivas em que o filtro é aplicado na entrada e atua como um hiperparâmetro da camada de convolução. Quando o passo é igual a 1, mais detalhes da imagem são preservados. Por outro lado, valores maiores reduzem a resolução dos mapas gerados, o que reduz tanto seu tamanho quanto a quantidade de dados processados. Além disso, essa redução também pode ser obtida por meio das camadas de agrupamento, que diminuem a altura e a largura dos mapas sem modificar os parâmetros aprendidos. Desse modo, o passo atua diretamente no equilíbrio entre a preservação de detalhes e a eficiência do processamento da rede (YAMASHITA *et al.*, 2018).

#### 2.8.0.3 Filtros (Kernel)

Filtros são pequenas matrizes aplicadas à imagem de entrada para gerar os mapas de características. Esses filtros são aprendidos durante o treinamento, com seus valores ajustados por retropropagação, e seu tamanho determina o alcance e a complexidade das características capturadas, funcionando como extratores de padrões que identificam bordas, texturas e formas à medida que percorrem a imagem. Além disso, cada filtro produz um feature map distinto, o que permite que a CNN aprenda diferentes tipos de características. Por fim, o tamanho do filtro

define a extensão da região analisada em cada convolução, sendo, portanto, um hiperparâmetro essencial da camada convolucional (KRICHEN, 2023).

#### 2.8.0.4 Preenchimento (*Padding*)

O preenchimento consiste na adição de linhas e colunas de zeros ao redor da imagem de entrada antes da aplicação dos filtros. Com isso, a técnica permite preservar as dimensões espaciais da imagem ao longo das camadas convolucionais. Além disso, o uso de preenchimento possibilita controlar o tamanho dos mapas de características, o que garante, por exemplo, que a saída mantenha as mesmas dimensões espaciais da entrada (KRICHEN, 2023).

#### 2.8.0.5 Função de Ativação (*Activation Function*)

As saídas de uma operação linear, como a convolução, são então passadas por uma função de ativação não linear, sendo a Unidade Linear Retificada (ReLU) a mais utilizada atualmente (YAMASHITA *et al.*, 2018). Essa etapa introduz não linearidade ao modelo, para permitir que a rede aprenda padrões mais complexos. Além disso, a popularidade da ReLU em arquiteturas modernas decorre de sua simplicidade computacional e de sua capacidade de mitigar o desaparecimento do gradiente, o que permite um melhor aprendizado e ajuda a evitar a saturação dos neurônios durante o treinamento (KRICHEN, 2023).

A função de ativação ReLU é definida matematicamente como  $f(x) = \max(0, x)$ . Isso significa que, para valores de entrada menores que zero, a saída é igual a zero, resultando em um trecho horizontal no gráfico da função no plano cartesiano à esquerda do eixo vertical. Por outro lado, para valores de entrada maiores ou iguais a zero, a saída cresce linearmente com inclinação igual a 1 (LI *et al.*, 2021).

## 2.9 Camada de Agrupamento (*Pooling Layer*)

A camada de agrupamento é usada para reduzir as dimensões espaciais dos mapas de características produzidos pela camada convolucional. Nesse processo, ela opera de forma independente em cada mapa e realiza uma redução espacial ao selecionar o valor máximo ou médio das regiões não sobrepostas. Assim, o agrupamento não apenas diminui a complexidade computacional da rede, mas também a torna mais robusta a pequenas translações na imagem de entrada (KRICHEN, 2023).

Concomitantemente, essa camada desempenha um papel essencial em uma CNN, visto que reduz a dimensionalidade das representações internas sem adicionar novos parâmetros ao modelo. Dessa forma, essa redução espacial torna a rede mais eficiente e reduz o risco de overfitting, situação em que o modelo memoriza excessivamente o conjunto de treinamento e perde capacidade de generalização para novos dados e amplia o campo recetivo das camadas subsequentes, permitindo que a CNN capture padrões mais amplos e complexos na imagem (WU, 2017).

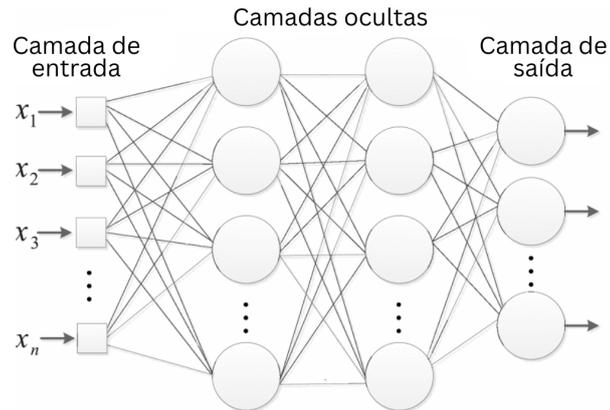
#### 2.9.0.1 Agrupamento Máximo (*Max pooling*)

O agrupamento máximo é um método popular de agrupamento que seleciona o maior valor de cada sub-região do mapa de características, o que resulta em uma matriz reduzida. Esse processo reduz as dimensões espaciais da entrada e, conseqüentemente, a complexidade computacional do modelo. Ao manter apenas um conjunto limitado de informações, o agrupamento máximo preserva os padrões mais relevantes e contribui para o controle do overfitting. Além disso, essa operação reduz a sensibilidade a variações espaciais, o que permite que a CNN reconheça um objeto mesmo quando ele se encontra deslocado ou levemente distorcido. Geralmente, utiliza-se um filtro  $2 \times 2$  com passo 2, configuração que reforça a invariância a translações. Contudo, essa redução dimensional ocorre ao custo da perda de algumas informações (AJIT *et al.*, 2020).

### 2.10 Camada Totalmente Conectada (*Fully Connected Layer*)

A camada totalmente conectada é um componente clássico das CNNs, no qual cada neurônio de uma camada se conecta a todos os neurônios da camada seguinte. Geralmente posicionada nas etapas finais da rede, ela tem como função consolidar as informações extraídas ao longo do processamento para gerar a saída final do modelo (KRICHEN, 2023). De forma complementar, essa camada transforma as ativações obtidas pelas convoluções em pontuações de classe, permitindo que o modelo finalize o processo de classificação. Nesse sentido, sua conectividade total em que cada neurônio se liga aos neurônios das camadas adjacentes, possibilita combinar de maneira global as representações apreendidas ao longo da rede, para que seja concluída a etapa decisória do modelo (O'SHEA; NASH, 2015). A Figura 6 apresenta a estrutura de uma camada totalmente conectada.

Figura 6 – Camada totalmente conectada



Fonte: Liu et al. (2018).

Como mostrado na Figura 6, a primeira é a camada de entrada, responsável por receber os dados processados nas etapas anteriores. Em seguida, encontram-se as camadas intermediárias, conhecidas como camadas ocultas por não manterem ligação direta com os dados de entrada. Por fim, a camada de saída é encarregada de produzir a predição final da rede (LIU *et al.*, 2018).

## 2.11 Modelos de Segmentação

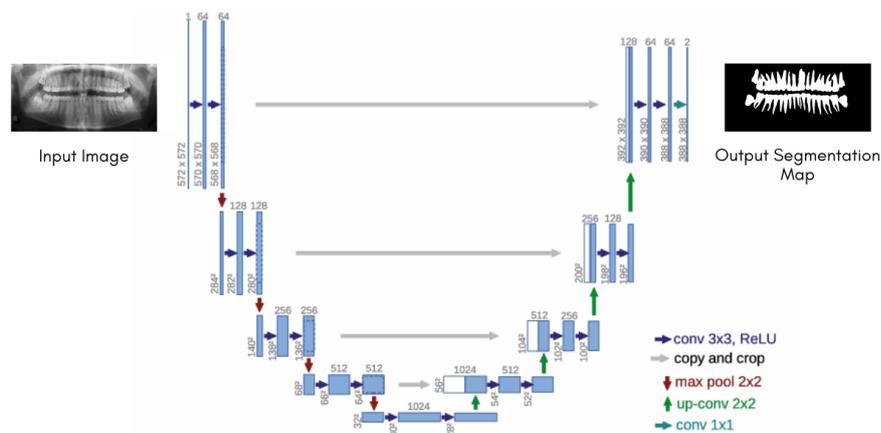
A ampla aceitação e o sucesso da aprendizagem profunda levaram ao desenvolvimento de uma nova geração de modelos de segmentação de imagem, que apresentam melhorias notáveis no desempenho e, muitas vezes, alcançam as taxas de precisão mais altas nos conjuntos de avaliação populares. Nesse contexto, esses modelos de segmentação abrangem um espectro de esforços pioneiros tanto em segmentação semântica quanto em segmentação de instância. Além disso, muitos desses métodos partilham componentes comuns, como codificadores, decodificadores, conexões de salto e arquiteturas multi-escala, os quais contribuem para uma representação mais rica e detalhada das características da imagem (MINAEE *et al.*, 2022). A seguir, apresentam-se os modelos adotados neste trabalho.

### 2.11.1 U-net

A U-Net é uma CNN projetada para segmentação de imagens biomédicas e se destaca por sua estrutura em formato de “U”, composta por um caminho de contração (*encoder*), que extrai o contexto global da imagem, e um caminho de expansão simétrico (*decoder*), responsável

por recuperar a resolução espacial e localizar com precisão as estruturas de interesse. Enquanto o encoder segue o padrão de redes convolucionais tradicionais, o decoder combina informações de diferentes profundidades para refinar contornos e detalhes das regiões segmentadas. Essa arquitetura alcança alto desempenho mesmo com conjuntos reduzidos de imagens anotadas, graças ao uso intensivo de técnicas de aumento de dados e também as skip connections, o que permite superar métodos anteriores em diversos desafios de segmentação médica (RONNEBERGER *et al.*, 2015). A Figura 7 ilustra a arquitetura dessa rede.

Figura 7 – Arquitetura U-Net



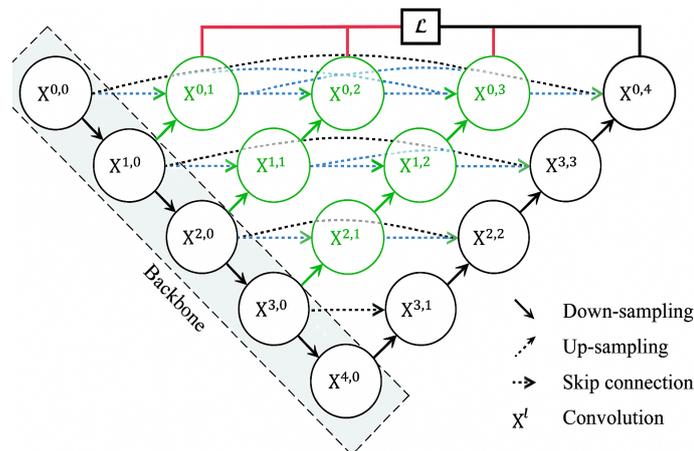
Fonte: Adaptada de RONNEBERGER *et al.* (2015).

Como podemos observar, no lado esquerdo a rede inicia com a *Input Image*, ou seja, a imagem de entrada fornecida ao modelo, e à direita encontra-se o *Output Segmentation Map*, que corresponde ao resultado final produzido após o processo de segmentação. Cada caixa azul na Figura 7 representa um mapa de características, que indica na parte superior a quantidade de canais e, na borda inferior esquerda, as dimensões espaciais desse mapa. As caixas brancas correspondem aos mapas copiados do caminho de contração para o caminho de expansão, o que permite a reutilização das informações de alta resolução durante a reconstrução da imagem segmentada. Além disso, as setas vermelhas indicam as etapas de *max pooling*, as verdes apontam as operações de superamostragem e as setas horizontais cinzas representam as e também as *skip connections* que unem as duas metades da U-Net (RONNEBERGER *et al.*, 2015).

### 2.11.2 Unet ++

A UNet++ é uma arquitetura avançada de CNN voltada à segmentação de imagens médicas, seguindo o modelo *encoder-decoder* com supervisão profunda e introduzindo *skip connections* aninhadas e densas como principal inovação. Essas vias reorganizam a comunicação entre codificador e decodificador por meio de caminhos intermediários que refinam progressivamente os mapas de características. Além disso, a arquitetura incorpora um mecanismo de supervisão profunda que opera em múltiplos níveis da rede, permitindo tanto um modo preciso, baseado na média das saídas das ramificações, quanto um modo rápido, no qual partes do modelo podem ser podadas para acelerar a inferência sem comprometer o desempenho (ZHOU *et al.*, 2018). A arquitetura dessa rede é representada na Figura 8.

Figura 8 – Arquitetura UNet++



Fonte: Zhou et al. (2018).

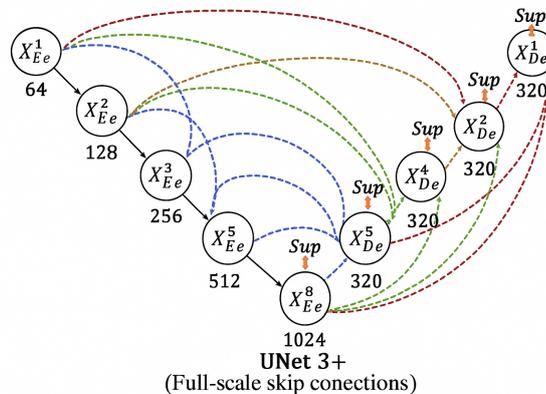
Como ilustrado na Figura 8, a UNet++ é composta por um codificador e um decodificador interligados por *skip connections* redesenhadas. Nessa estrutura, as convoluções presentes nesses caminhos reduzem a lacuna semântica entre os mapas gerados pelo codificador e pelo decodificador, enquanto as conexões densas aprimoram o fluxo de gradiente e a propagação das informações. Além disso, a supervisão profunda favorece um desempenho igual ou superior ao obtido com apenas uma saída de perda, já que as múltiplas saídas participam do processo de aprendizagem de maneira conjunta e consistente durante o fluxo de processamento. No fluxo interno da rede, a etapa de contração (*down-sampling*) reduz a resolução espacial e aumenta o número de canais para extrair características mais abstratas, enquanto a etapa de expansão (*up-sampling*) reconstrói a resolução da máscara, alinhando-a ao tamanho original da imagem. Por fim, as *skip connections* conectam diretamente níveis correspondentes, para recuperar detalhes

finos e garantir segmentações mais precisas (ZHOU *et al.*, 2018).

### 2.11.3 Unet 3+

A UNet 3+ é uma evolução das arquiteturas U-Net e U-Net++, projetada para superar limitações na captura de informações em múltiplas escalas. Diferentemente das versões anteriores, que utilizam conexões de salto simples, aninhadas ou densas, a UNet 3+ emprega *skip connections* em escala completa e supervisão profunda. Essas conexões combinam simultaneamente detalhes de baixo nível e semântica de alto nível, enquanto a supervisão profunda aprende representações hierárquicas dos mapas de características. Como resultado, a UNet 3+ melhora a precisão da segmentação, reduz o número de parâmetros e aumenta a eficiência computacional. Nesse contexto, para explorar plenamente as características multiescala, a arquitetura redesenha as interconexões entre codificador e decodificador, bem como entre os níveis do decodificador, para garantir integração eficaz entre detalhes finos e informações semânticas (HUANG *et al.*, 2020).

Figura 9 – Arquitetura UNet 3+



Fonte: (HUANG *et al.*, 2020).

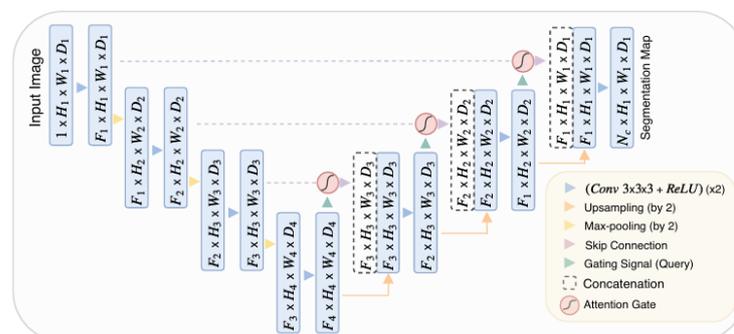
A Figura 9 ilustra que a U-Net 3+ reorganiza as conexões entre codificador e decodificador por meio das *full-scale skip connections*, representadas pelas setas coloridas, que integram simultaneamente detalhes de baixo nível e informações semânticas de múltiplas escalas. Em cada estágio do decodificador, a arquitetura combina mapas provenientes do codificador, de estágios mais profundos e da própria camada do decodificador, permitindo capturar, ao mesmo tempo, detalhes finos e contexto global. As cinco resoluções resultantes dessa fusão são então concatenadas e refinadas por um bloco de agregação composto por uma convolução  $3 \times 3$  com 320 filtros, seguida de *batch normalization* e ativação ReLU. A Figura 9 também evidencia o uso

de supervisão profunda em grande escala, indicado pelos círculos “Sup”, em que cada estágio do decodificador gera uma saída auxiliar processada por convolução  $3 \times 3$ , interpolação até a resolução original e ativação sigmoide. Além disso, cada uma dessas saídas laterais é associada a uma função de perda híbrida, responsável por medir o erro entre a segmentação prevista e a segmentação real e orientar o ajuste dos parâmetros durante o treinamento. Esse mecanismo favorece segmentações mais robustas e precisas, especialmente em estruturas que variam em escala. Com isso, a U-Net 3+ fortalece o aprendizado multiescala e aprimora significativamente a precisão final da segmentação (HUANG *et al.*, 2020).

### 2.11.4 Attention U-Net

A Attention U-Net é outra arquitetura de CNN, desenvolvida para a imagiologia médica, com foco em tarefas de segmentação de imagens. Seu principal objetivo é permitir que o modelo aprenda a destacar automaticamente as estruturas-alvo e suprimir regiões irrelevantes da imagem de entrada, sem a necessidade de módulos externos de localização. A principal inovação da Attention U-Net está na integração dos *Attention Gates (AGs)* que são módulos de atenção auto-contidos que filtram as características propagadas pelas *skip connections* da arquitetura U-Net, suprimindo respostas irrelevantes e mantendo apenas ativações úteis, o que aumenta a sensibilidade e a precisão do modelo. Assim, durante a inferência, a rede realça de forma eficiente as características decisivas e pode ser treinada do zero de maneira convencional (OKTAY *et al.*, 2018). A Figura 10 mostra a arquitetura da Attention U-Net.

Figura 10 – Arquitetura Attention U-Net



Fonte: Oktay et al. (2018).

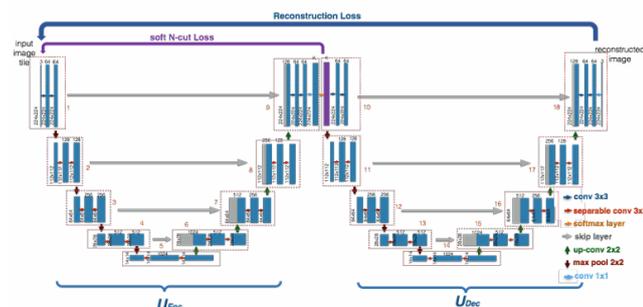
No lado esquerdo da arquitetura, o *encoder* processa a imagem de entrada de forma progressiva por meio de camadas convolucionais com ativação ReLU e operações de *max-pooling*, reduzindo as dimensões espaciais a cada escala. Cada nível, indicado como  $(F_1, H_1, W_1, D_1)$

corresponde a mapas de características definidos pelo número de filtros, altura, largura e profundidade. À medida que a rede avança, as dimensões espaciais diminuem enquanto o número de filtros aumenta, o que permite preservar informações relevantes e capturar características mais abstratas, incluindo padrões importantes, relações locais e globais da imagem. No lado direito da arquitetura, o *decoder* realiza o *upsampling* para recuperar gradualmente a resolução original, sendo cada etapa refinada por convoluções seguidas de ReLU. As *skip connections* transferem mapas do *encoder* para o *decoder*, integrando semântica profunda com detalhes espaciais. Sobre esses mapas atuam os AGs, que usam um sinal de *gating* de escalas mais profundas para suprimir respostas irrelevantes e destacar apenas as características úteis. Após essa filtragem, os mapas são concatenados aos do *decoder* e seguem pelas etapas restantes de *upsampling* até que a rede produza o mapa final de segmentação (OKTAY *et al.*, 2018).

### 2.11.5 W-Net

A W-Net é uma arquitetura de CNN profunda proposta para a segmentação de imagens de forma totalmente não supervisionada, surgindo em um contexto em que a segmentação é um problema central na visão computacional. Sua arquitetura é composta por um autoencoder duplo em formato de “W”, no qual o encoder realiza a segmentação preliminar enquanto o decoder reconstrói a imagem, permitindo um processamento eficiente que combina a extração de detalhes finos com a semântica global. Além disso, todos os módulos utilizam convoluções separáveis em profundidade, que combinam convolução em profundidade com convolução pontual, aumentando a eficiência computacional sem expandir o número de parâmetros da rede (XIA; KULIS, 2017). A arquitetura dessa rede é representada na Figura 11.

Figura 11 – Arquitetura W-Net



Fonte: (XIA; KULIS, 2017).

Como podemos observar na Figura 11, a W-Net apresenta uma estrutura em “W”,

composta por um *encoder* (*UEnc*) à esquerda, responsável pela extração densa de características, e por um *decoder* correspondente (*UDec*) à direita, dedicado à reconstrução da imagem. No *UEnc*, há um caminho de contração que processa as imagens de entrada e conecta os módulos por meio de operações de max-pooling, seguido por um caminho expansivo baseado em convoluções transpostas, que recupera progressivamente a resolução. O *UDec*, por sua vez, mantém uma organização semelhante, pois recebe a saída processada pelo *UEnc* e finaliza com uma convolução  $1 \times 1$  que reconstrói a imagem. Além disso, todos os módulos utilizam convoluções separáveis em profundidade (indicadas por setas vermelhas), que combinam convoluções espaciais e pontuais, para tornar o processamento mais eficiente sem aumentar o número de parâmetros (XIA; KULIS, 2017).

## 2.12 Validação Cruzada

A validação cruzada é uma técnica amplamente utilizada no campo do aprendizado de máquina para avaliar o desempenho de modelos preditivos. Nessa abordagem, o conjunto de dados original é dividido em  $k$  subconjuntos de tamanho aproximadamente igual, denominados *folds*, formados de maneira mutuamente exclusiva, de modo que nenhuma amostra seja compartilhada entre eles. O processo ocorre em  $k$  iterações, ou seja, em cada uma delas, um dos *folds* é utilizado como conjunto de teste, enquanto os  $k - 1$  restantes são combinados para formar o conjunto de treinamento. O modelo é então treinado com os dados de treinamento e avaliado no conjunto de teste. Esse procedimento se repete até que cada *fold* tenha sido utilizado exatamente uma vez como conjunto de teste. Por fim, os resultados obtidos em todas as iterações são agregados, geralmente por meio do cálculo da média, o que fornece uma estimativa mais estável do desempenho do modelo (NTI *et al.*, 2021).

## 2.13 Grid Search

O *Grid Search* ou (busca em grade) é um método tradicionalmente utilizado para a otimização de hiperparâmetros em algoritmos de aprendizado de máquina, cujo objetivo é identificar a configuração mais adequada de uma rede neural. Seu funcionamento baseia-se em uma busca sistemática sobre um subconjunto previamente definido do espaço de hiperparâmetros do modelo, chamado de espaço de busca, no qual se estabelecem antecipadamente os valores ou intervalos possíveis para cada parâmetro. Considerando que alguns hiperparâmetros podem

assumir valores contínuos ou teoricamente ilimitados, a definição desses limites torna a aplicação do método viável na prática e possibilita que o *Grid Search* realize o treinamento e a avaliação de todas as combinações possíveis de hiperparâmetros especificadas. Entre as principais vantagens deste método destacam-se a simplicidade de implementação e a garantia de cobertura do espaço de busca definido, o que assegura a identificação da melhor configuração dentro desse conjunto (LIASHCHYNSKYI, 2019).

## 2.14 Teste Estatístico de Wilcoxon

O teste de Wilcoxon é um método estatístico não paramétrico utilizado para verificar a existência de diferenças significativas entre duas amostras relacionadas ou emparelhadas. Diferentemente dos testes paramétricos, esse método não pressupõe que os dados sigam uma distribuição normal, característica que o torna adequado em situações nas quais essa suposição não pode ser garantida. Por esse motivo, o teste é amplamente empregado na comparação do desempenho de dois modelos quando as métricas de avaliação são obtidas a partir dos mesmos conjuntos de teste ou por meio de procedimentos de validação cruzada, nos quais os resultados são naturalmente emparelhados (RAINIO *et al.*, 2024).

O funcionamento do teste baseia-se na análise das diferenças entre pares de resultados obtidos sob condições experimentais idênticas. Para cada par de observações, calcula-se a diferença  $d_i$  entre os valores correspondentes, as quais são ordenadas de acordo com o valor absoluto  $|d_i|$ , desconsiderando-se o sinal, e recebem postos. Em seguida, os postos associados às diferenças positivas são somados, resultando em  $R^+$ , enquanto os postos correspondentes às diferenças negativas originam  $R^-$ . A estatística do teste é então definida como  $T = \min\{R^+, R^-\}$ . Do ponto de vista inferencial, o teste parte da hipótese nula  $H_0$ , a qual estabelece que a mediana das diferenças  $d_i$  é igual a zero, o que caracteriza a inexistência de diferença sistemática entre os desempenhos comparados. Quando o *p-value* é inferior ao nível de significância adotado, geralmente  $\alpha = 0,05$ , rejeita-se  $H_0$ , o que permite inferir a existência de evidências estatisticamente significativas de diferença entre os modelos (RAINIO *et al.*, 2024).

## 2.15 Métricas de Avaliação

As métricas de avaliação são fundamentais no desenvolvimento e na validação de sistemas, pois permitem quantificar sua qualidade e eficácia, para garantir que o desempenho

esteja alinhado com o propósito real do sistema (FERRER, 2022). Sob essa perspectiva, Müller *et al.* (2022) apontam que essas métricas visam avaliar o quão próxima a segmentação predita está da segmentação real, o que assegura uma análise correta, robusta e padronizada dos algoritmos de segmentação de imagens médicas, aspecto essencial para a confiabilidade de sistemas de apoio à decisão clínica. Nesse contexto, apresentam-se a seguir os termos básicos que fundamentam parte das métricas adotadas neste trabalho:

- **Verdadeiro Positivo (TP):** Quantidade de instâncias positivas corretamente classificadas como positivas pelo modelo.
- **Verdadeiro Negativo (TN):** Quantidade de instâncias negativas corretamente classificadas como negativas pelo modelo.
- **Falso Positivo (FP):** Quantidade de instâncias negativas incorretamente classificadas como positivas pelo modelo.
- **Falso Negativo (FN):** Quantidade de instâncias positivas incorretamente classificadas como negativas pelo modelo.

### 2.15.1 Acurácia

Segundo Hossin e Sulaiman (2015), a acurácia (*accuracy*) é uma das métricas mais utilizadas por pesquisadores para avaliar a capacidade de generalização de classificadores. Ela quantifica a proporção de previsões corretas em relação ao total de instâncias avaliadas, expressando o percentual de exemplos classificados corretamente durante a etapa de teste com dados não vistos. A fórmula a seguir descreve o cálculo da acurácia.

$$\text{Acurácia} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.1)$$

### 2.15.2 Especificidade

A Especificidade (*specificity*) mede a proporção de instâncias negativas que foram corretamente classificadas, correspondendo à taxa de acerto na classe negativa. Uma alta especificidade indica que o modelo é eficaz em identificar corretamente os casos negativos, enquanto uma baixa especificidade sugere que muitos negativos estão sendo incorretamente classificados como positivos (HOSSIN; SULAIMAN, 2015).

$$\text{Especificidade} = \frac{TN}{TN + FP} \quad (2.2)$$

### 2.15.3 Sensibilidade

A Sensibilidade (*sensitivity*) indica a proporção de instâncias positivas corretamente identificadas em relação ao total real de instâncias positivas no conjunto de dados. O denominador corresponde à soma entre verdadeiros positivos e falsos negativos. Essa métrica é especialmente importante em cenários onde o custo de um falso negativo é elevado (TATSAT *et al.*, 2021).

$$\text{Sensibilidade} = \frac{TP}{TP + FN} \quad (2.3)$$

### 2.15.4 Coeficiente de Similaridade de Dice (DSC)

O Coeficiente de Similaridade de Dice (DSC) é uma métrica amplamente utilizada para avaliar o desempenho em tarefas de segmentação de imagens (JOSHI, 2024). Ele mede a similaridade entre os resultados de segmentação e uma "verdade fundamental" (*ground truth*), que são segmentações de referência, geralmente criadas por especialistas. Nesse contexto, o valor do DSC varia de 0 a 1: um valor igual a 0 indica que a imagem segmentada pelo modelo e a imagem rótulo não apresentam sobreposição, ou seja, não compartilham nenhum pixel; já um valor igual a 1 representa correspondência perfeita entre a segmentação prevista e a segmentação de referência. A fórmula a seguir ilustra como o DSC é calculado (ANDREWS; HAMARNEH, 2015):

$$\text{DSC}(y, \hat{y}) = \frac{2TP}{2TP + FP + FN} \quad (2.4)$$

### 2.15.5 Erro Médio Absoluto (MAE)

O Erro Médio Absoluto (MAE) mede a média das diferenças absolutas entre os valores previstos e os valores reais (TERVEN *et al.*, 2025). O MAE é uma medida dependente da escala dos dados, o que significa que seu valor absoluto não possui um limite superior fixo e pode variar amplamente de acordo com a magnitude dos dados previstos. Um MAE igual a 0

indica que todas as previsões corresponderam exatamente aos valores observados (HYNDMAN; KOEHLER, 2006). A fórmula a seguir ilustra como o MAE é calculado (HODSON, 2022):

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2.5)$$

- $n$ : representa o número total de pixels na imagem.
- $y_i$ : valor real (ground truth) do  $i$ -ésimo pixel.
- $\hat{y}_i$ : valor predito pelo modelo para o  $i$ -ésimo pixel.
- $|y_i - \hat{y}_i|$ : erro absoluto entre o valor real e o valor predito para cada pixel.

### 2.15.6 Intersecção sobre União (IOU)

Intersecção sobre União (IOU) é uma métrica de avaliação amplamente utilizada em tarefas de segmentação de imagens. A IOU mede a similaridade entre a máscara de segmentação prevista e a máscara da verdade fundamental e varia de 0 a 1, onde 0 indica nenhuma sobreposição e 1 indica sobreposição perfeita (JOSHI, 2024). A fórmula a seguir ilustra como a IOU é calculada (JOSHI, 2024):

$$\text{IoU} = \frac{\text{Área de Interseção}}{\text{Área de União}} = \frac{|A \cap B|}{|A \cup B|} \quad (2.6)$$

- $A \cap B$ : Corresponde ao número de pixels que são corretamente previstos como pertencentes aos dentes, tanto na máscara prevista quanto na máscara da verdade fundamental.
- $A \cup B$ : Representa o número de pixels classificados como pertencentes aos dentes em ambas as máscaras: a prevista e a da verdade fundamental.

### 2.15.7 Medida de Alinhamento Aprimorada (E-MEASURE)

A Medida de Alinhamento Aprimorada (E-MEASURE), é uma nova medida para avaliar mapas binários de primeiro plano em visão computacional. Esses mapas binários comparam a saída de modelos de segmentação de primeiro plano com um mapa binário de verdade fundamental rotulado por humanos. O objetivo da E-MEASURE é combinar valores de pixel locais com o valor médio de nível de imagem em um único termo, capturando conjuntamente estatísticas de nível de imagem e informações de correspondência de pixel local. A fórmula a seguir ilustra como a E-MEASURE é calculada (FAN *et al.*, 2018):

$$Q_{FM} = \frac{1}{w \times h} \sum_{x=1}^w \sum_{y=1}^h \varphi_{FM}(P_{(x,y)}, G_{(x,y)}) \quad (2.7)$$

- $Q_{FM}$  é a E-measure final, que mede a similaridade global entre o mapa previsto  $P$  e o mapa de verdade  $G$ , combinando informações locais e estatísticas globais.
- $\frac{1}{w \times h}$  é o fator de normalização, que calcula a média dos alinhamentos aprimorados sobre todos os pixels do mapa, onde  $w$  representa a largura e  $h$  a altura da imagem.
- $\sum_{x=1}^w \sum_{y=1}^h$  representa a soma sobre todos os pixels, indicando que o cálculo de  $\varphi_{FM}$  é realizado para cada pixel na posição  $(x,y)$ , percorrendo toda a largura  $w$  e altura  $h$  do mapa.
- $\varphi_{FM}(P_{(x,y)}, G_{(x,y)})$  representa o valor de alinhamento aprimorado para o pixel na posição  $(x,y)$ , avaliando a correspondência entre o pixel previsto  $P(x,y)$  e o da verdade  $G(x,y)$ .

### 3 TRABALHOS RELACIONADOS

Neste capítulo, são apresentados os trabalhos relacionados utilizados como referência para o desenvolvimento desta pesquisa, com foco em abordagens voltadas à segmentação automática de dentes em radiografias panorâmicas.

Nader *et al.* (2022) propõem a Modified U-Net, uma versão da U-Net que incorpora informações de localização via *bounding boxes* nas *skip connections* para melhorar a segmentação dentária em imagens panorâmicas, especialmente em casos com ruído, sobreposição ou dentes ausentes. O método ocorre em duas etapas: primeiro, a Mask R-CNN gera as *bounding boxes* no conjunto *DNS Panoramic*, que contém 543 imagens panorâmicas; depois, essas caixas são usadas como entradas auxiliares no treinamento da Modified U-Net. Testes com *bounding boxes* manuais (Optimal U-Net) e automáticas mostraram um aumento de 5% a 10% no DSC em relação à Modified U-Net. Os resultados mostram que a U-Net ótima alcançou um DSC médio de 94,5%, seguida pela Modified U-Net com 89,5% e pela U-Net original com 85%. O melhor desempenho da Modified U-Net foi especialmente evidente na segmentação de molares. Além disso, ela superou a Mask R-CNN, apresentando segmentações mais precisas e consistentes, comprovando a eficácia da inclusão das informações espaciais.

De forma semelhante, Joshi (2024) propõe uma metodologia baseada na arquitetura U-Net para segmentação automática de dentes em radiografias panorâmicas. O conjunto de dados utilizado neste estudo é o *Tufts Dental Dataset (TDD)*, composto por 1000 radiografias panorâmicas de raios-X e suas respectivas máscaras dentárias segmentadas por especialistas odontológicos. A metodologia envolve etapas de pré-processamento das imagens, como normalização, redimensionamento, remoção de ruído, realce de contraste e aumento de dados. O modelo é treinado com funções de perda como *Binary Cross-Entropy* e *Dice*, aplicando técnicas de regularização para evitar sobreajuste. A avaliação dos resultados, por meio do DSC 88% e do IOU 79%, indica desempenho satisfatório do modelo na segmentação dentária. As segmentações geradas são sobrepostas às radiografias com contornos, cores e medições, destacando o potencial da U-Net para aplicações clínicas e acadêmicas em odontologia.

Buscando uma análise comparativa entre arquiteturas, Zannah *et al.* (2024) propõem uma comparação abrangente de desempenho entre seis variantes da arquitetura U-Net aplicadas à segmentação de imagens odontológicas: Vanilla U-Net, Dense U-Net, Attention U-Net, SE U-Net, Residual U-Net e R2 U-Net, com o objetivo de identificar a variante com melhor desempenho, custo e aplicabilidade clínica. O estudo utilizou 389 radiografias obtidas em uma

clínica odontológica em *Bogura, Bangladesh*, com ampla variabilidade anatômica, faixas etárias e qualidade visual. As imagens foram anotadas com *Label Studio* e pré-processadas por recorte, redimensionamento e divisão em patches de 256×256 pixels, totalizando 6224 amostras. As redes foram treinadas por 100 épocas com *Dice Loss*, otimizador *Adam* e *batch size* 16, seguindo a estrutura clássica da U-Net com quatro blocos e operações padrão como convolução, pooling, normalização e ativação. Os resultados indicaram que variantes com três camadas convolucionais por bloco apresentaram desempenho superior, com destaque para a Dense U-Net que obteve DSC 90,33% e IOU 89,07%, enquanto a R2 U-Net apresentou DSC 90,35%, embora com maior complexidade e tempo de treinamento. Já a Vanilla U-Net com duas camadas demonstrou bom equilíbrio entre simplicidade e desempenho, atingindo 88% em DSC e IOU, sendo considerada mais adequada para aplicações clínicas.

Como desdobramento das investigações sobre variantes da U-Net, *Sahin et al. (2024)* propõem uma arquitetura U-Net 3+, baseada na U-Net, com conexões de salto multi-escala e fusão de características, com o objetivo de desenvolver um modelo de segmentação preciso e confiável de estruturas dentárias em radiografias panorâmicas. O modelo foi treinado com radiografias panorâmicas de crianças (2 a 13 anos), do *Children's Dental Panoramic Radiographs Dataset*, incorporando imagens de adultos para diversificar o conjunto de dados. A arquitetura possui quatro níveis de codificação e decodificação, usando camadas convolucionais, *Batch Normalization*, *ReLU*, *pooling* e *upsampling* com concatenação. Ela integra recursos em múltiplas escalas para melhorar a extração de detalhes. O desempenho foi avaliado por meio de validação cruzada com 5 dobras, utilizando 100 épocas de treinamento em cada divisão, o que garantiu uma análise confiável dos resultados. A avaliação do modelo apresentou métricas robustas, como Acurácia 97,36%, Precisão 93,25%, Sensibilidade 93,38%, DSC 93,31%, Acurácia por pixel 97,36%, AUC 96,11%, Mean IOU 91,82% e Loss 0,0667, destacando seu potencial para aplicações clínicas com segmentação dentária precisa e confiável.

Com uma proposta arquitetural diferenciada, *Zhong et al. (2025)* apresentam a GCNet, uma arquitetura voltada para superar desafios comuns na segmentação de radiografias panorâmicas, como o pequeno tamanho dos datasets, o alto nível de ruído e as fronteiras pouco definidas entre dentes e tecidos periodontais. O modelo incorpora dois módulos centrais: o *Grouped Global Attention (GGA)*, responsável por capturar dependências globais e destacar regiões relevantes, e o *Cross-Layer Fusion (CLF)*, que integra informações de diferentes profundidades para combinar detalhes estruturais e características semânticas. Essa combinação permite melho-

rar a separação entre dentes e tecidos adjacentes, reduzir a sensibilidade ao ruído e preservar contornos finos. O estudo utilizou o *Dental Panoramic Radiographs Dataset*, composto por 106 radiografias pediátricas e mais três conjuntos de adultos, totalizando 3.187 exames com máscaras de segmentação. O treinamento foi realizado por 200 épocas, com batch size 2, otimizador Adam e taxa de aprendizagem de  $1 \times 10^{-3}$ . A GCNet foi comparada com modelos amplamente usados na literatura: U-Net, SegNet, Attention U-Net, Inf-Net, GT-U-Net e Teeth U-Net e apresentou desempenho superior em todas as métricas: DSC 93,38%, Sensibilidade 94,26%, Especificidade 98,21%, E-MEASURE 97,12% e MAE 2,59%. A avaliação qualitativa reforçou esses achados, mostrando contornos mais precisos e contínuos, além de menor variação de desempenho entre radiografias pediátricas e de adultos, evidenciando maior robustez e capacidade de generalização.

Em um contexto complementar, Zhang *et al.* (2023) propõem a criação do primeiro conjunto de dados público internacional de radiografias panorâmicas dentárias de crianças, criado para apoiar tarefas de segmentação de cáries e detecção de doenças odontológicas por meio de deep learning. A proposta surge da falta de bases pediátricas adequadas, já que os datasets existentes concentram-se majoritariamente em adultos, cujas diferenças fisiológicas limitam o uso clínico de modelos treinados exclusivamente nesses dados. O estudo reuniu 193 radiografias de 106 pacientes pediátricos (2 a 13 anos), anotadas com *EISeg* e *LabelMe*, formando o *Children's Dental Panoramic Radiographs Dataset*. Além disso, os autores incorporaram 2692 radiografias de três bases internacionais de adultos, ajustadas para incluir anotações da estrutura dentária, ampliando a abrangência do conjunto. Para avaliar a aplicabilidade do dataset, foram treinadas as arquiteturas U-Net, R2 U-Net, PSPNet e DeepLab V3+, todas com o mesmo protocolo experimental (batch size 4, imagens 512x512 px e loss de entropia cruzada). A U-Net obteve o melhor desempenho geral no conjunto infantil, com IOU de 83,87%, DSC de 91,20%, ACC de 97,10%, Recall de 92,00% e Especificidade de 98,03%. Ainda assim, o DeepLab V3+ apresentou o maior Recall médio 94,86%, e o PSPNet atingiu a maior Especificidade 98,56%, evidenciando a robustez do dataset para diferentes modelos.

Nesse contexto comparativo, Bhat *et al.* (2025) propõem avaliar e comparar a eficácia de quatro arquiteturas avançadas de aprendizado profundo para segmentação dentária em radiografias panorâmicas: U-Net, U-Net++, U-Net com codificador MobileNetV3 e DeepLab. Essas arquiteturas foram selecionadas devido às suas diferentes capacidades de lidar com aspectos essenciais da tarefa de segmentação, como a localização precisa dos contornos, a captura de informações contextuais em múltiplas escalas e a eficiência computacional. Nesse estudo,

foram utilizados dois conjuntos de dados: o da Faculdade de Tecnologia da Universidade Tufts (TUFT), que contém 1000 radiografias panorâmicas dentárias com máscaras rotuladas, e o da Universidade Federal da Bahia–Universidade Estadual de Santa Cruz (UFBA-UESC), disponível publicamente e composto por 1500 radiografias panorâmicas. O processo de avaliação das redes foi conduzido em três etapas. Primeiro, cada arquitetura foi testada separadamente nos datasets TUFT e UFBA, permitindo identificar seu desempenho individual, etapa na qual a U-Net e a U-Net++ apresentaram os melhores desempenhos. Em seguida, realizou-se a avaliação no dataset combinado (TUFT + UFBA), totalizando 2500 radiografias, dividido em 80% para treino e 20% para validação/teste, com as imagens redimensionadas para 512×512 px. Nessa fase, a U-Net++ demonstrou melhor capacidade de generalização e, por esse motivo, foi selecionada para um treinamento adicional com diferentes configurações de hiperparâmetros. Nesse processo, foram testados tamanhos de lote de 4, 8 e 16, sendo o *batch size* 8 o mais eficaz e três otimizadores (*SGD*, *Adam* e *Nadam*), com o *Adam* apresentando o melhor desempenho, atingindo precisão aproximada de 97% durante o treinamento. Após a otimização, a U-Net++ alcançou seus melhores resultados: IOU de 86%, DSC de 90%, Acurácia de 96%, Precisão de 91% e Recall de 90%, confirmando sua robustez e eficácia na segmentação dentária.

Com foco em modificações estruturais da U-Net, Almalki *et al.* (2023) propõe uma abordagem para segmentação dentária em radiografias panorâmicas baseada na U-Net Residual, que incorpora conexões residuais para fortalecer o fluxo de gradientes e melhorar a segmentação. A arquitetura inclui um codificador denoised para reduzir ruídos e extrair características mais limpas. Neste estudo, foi utilizado o conjunto de dados *Tufts Dental*, composto por 1.000 radiográficas panorâmicas, todas anotadas por especialistas. O treinamento seguiu três etapas: pré-processamento, configuração da arquitetura e otimização. No pré-processamento, as imagens foram redimensionadas, valores ausentes tratados, e divisão dos dados em conjuntos de treino e teste, e técnicas de aumento de dados aplicadas para melhorar a generalização. A arquitetura utilizou camadas convolucionais duplas e pesos pré-treinados do ResNet50, e a otimização contou com *softmax* para classificação pixel a pixel e função de perda por entropia cruzada. Para avaliar o desempenho, o modelo foi comparado com outras arquiteturas, incluindo SegNet, U-Net, ResNet, DeepLabV3, DeepLabV3+, U-Net++, PSPNet e FPN. Os resultados demonstram excelente desempenho, com IOU médio de 98%, DSC médio de 98%, distância de *Hausdorff* para o fundo de 1,617 e para a máscara estimada de 1,617. Em relação à segmentação, o algoritmo implementado conseguiu delimitar com precisão as regiões defeituosas dos dentes,

evidenciando o seu potencial área de segmentação dentária. Esses resultados indicam que a abordagem desenvolvida é eficaz, e sua alta precisão e eficiência sugerem potencial para aplicações na indústria odontológica.

Seguindo na mesma linha de aprimoramento arquitetural, Yaswanth *et al.* (2025) propõem um modelo de segmentação dentária baseado em deep learning, utilizando uma U-Net++ modificada integrada ao mecanismo de atenção ECA-Net, que reforça a extração de características relevantes em nível de canal e mantém baixo custo computacional. Neste estudo, utilizou-se o *Tufts Dental Dataset*, composto por 1.000 radiografias panorâmicas dentárias anotadas por especialistas com detalhes sobre dentes e anomalias. Inicialmente, as imagens de entrada foram normalizadas e binarizadas, sendo posteriormente organizadas em grupos para o processo de treinamento. Concomitantemente, para assegurar consistência entre todos os modelos avaliados - U-Net, FCNs, ENet, U-Net++, U-Net3+, SwiftNet e o modelo proposto - foram adotadas as mesmas configurações de parâmetros em todos os experimentos. Dessa forma, utilizou-se a função de perda *Binary Cross-Entropy*, o otimizador Adam com taxa de aprendizagem de 0,0001. Além disso, todos os modelos foram treinados por 100 épocas, com aplicação de *early stopping* para reduzir as perdas e evitar overfitting. Após o treinamento, o desempenho das redes foi avaliado utilizando algumas métricas de segmentação. O modelo proposto apresentou os melhores resultados, alcançando IOU de 83%, DSC de 90%, PSNR de 16,77, mAP de 96% e acurácia de 97%. Na tarefa de segmentação dentária, a abordagem demonstrou elevada eficácia, produzindo contornos consistentes, bem delineados e alinhados às anotações profissionais. Portanto, os resultados obtidos evidenciam que a U-Net++, aliada ao mecanismo de atenção ECA-Net, configura uma solução altamente eficiente para a segmentação automática de dentes.

Por fim, Bouali *et al.* (2024) propõem investigar e comparar o desempenho de quatro arquiteturas de redes neurais convolucionais do tipo UNet - UNet clássica, Attention UNet, UNet3+ e TransUNet - partindo do pressuposto de que essas variantes apresentam desempenho superior em tarefas de segmentação de imagens médicas quando comparadas a CNNs tradicionais. Para a realização deste estudo, foram utilizados dois conjuntos de dados: o *New Tufts Dental Dataset*, composto por 1.000 radiografias panorâmicas digitais obtidas na *Tufts University School of Dental Medicine* e o *UFBA-UESC Dental Images Dataset*, que reúne 1.500 radiografias panorâmicas. No treinamento das quatro redes, os conjuntos de dados foram divididos em 80% para treino e 20% para teste, com imagens redimensionadas e normalizadas

para (256, 256, 3), exceto no TransUNet, que utilizou (64, 64, 3) devido ao maior número de parâmetros. Empregou-se aprendizagem por transferência com VGG19 e ResNet50 pré-treinados no ImageNet, atualizando apenas as camadas adicionais. O processo experimental foi conduzido com o otimizador Adam, taxa de aprendizagem de  $1 \times 10^{-3}$ , *Binary Focal Loss*, *batch size* 8 ou 16 e número de épocas ajustado a cada base: 150 para o *Tufts* e 50 para o UFBA-UESC. No processo de avaliação, a UNet3+ com VGG19 apresentou o melhor desempenho geral, alcançando DSC de 96,23% e IOU de 91,82% no *Tufts*, e DSC de 98,10% e IOU de 94,37% no UFBA-UESC. O AttUnet com ResNet50 também se destacou em DSC, obtendo 96,98% no *Tufts* e 97,16% no UFBA-UESC. O TransUNet-ResNet50 apresentou o melhor valor de IOU no *Tufts*, atingindo 94,19%. Além disso, as quatro variantes da UNet superaram outras redes da literatura, como UNet, UNet++, FCN+UNet, TSASNet e MSLPNet, em ambos os conjuntos de dados. Esses resultados evidenciam a eficácia das arquiteturas UNet e os benefícios do uso de modelos pré-treinados na segmentação de radiografias panorâmicas.

De forma geral, os trabalhos analisados evidenciam a ampla adoção de arquiteturas baseadas na U-Net para a segmentação automática de dentes em radiografias panorâmicas, refletindo avanços relevantes tanto no desenho das arquiteturas quanto nas estratégias de treinamento e avaliação como observado na Tabela 1. Contudo, observa-se a ausência de procedimentos fundamentais na literatura analisada, uma vez que não foi identificado o uso de *Grid Search* para otimização de hiperparâmetros, não houve aplicação do teste estatístico de Wilcoxon para verificar a existência de diferenças de desempenho entre as redes e não se verificou a integração das redes que obtiveram o melhor desempenho a um sistema CAD para uso prático em contextos clínicos. Ademais, no que se refere às estratégias de validação, apenas Şahin *et al.* (2024) adotaram validação cruzada, enquanto somente Zhang *et al.* (2023) avaliou o desempenho das redes em um conjunto externo para verificar a capacidade de generalização dos modelos para dados não vistos, visto que ambos esses procedimentos estiveram ausentes nos demais estudos analisados. Em contraste, o presente trabalho busca suprir essas lacunas ao adotar um protocolo experimental mais robusto, alinhado às limitações identificadas na literatura.

Uma comparação resumida entre os trabalhos relacionados e este estudo encontra-se na Tabela 1.

Tabela 1 – Comparação entre trabalhos relacionados e este trabalho

Trabalho	Base de Dados	Arquiteturas Usadas	Métricas de Avaliação	Teste Estatístico	Sistema CAD	Validação Externa	Validação Cruzada
Nader et al. (2022)	DNS Panorâmico imagens da IvisionLAB	<b>Modified U-Net</b>	DSC	✗	✗	✗	✗
Joshi (2024)	Tufts Dental Dataset (TDD)	<b>U-Net</b>	DSC, IOU	✗	✗	✗	✗
Zannah et al. (2024)	Conjunto de dados exclusivo	<b>Vanilla U-Net</b> , Dense U-Net, Attention U-Net, SE U-Net, Residual U-Net, R2 U-Net	Acurácia, DSC, F1 Score, IOU	✗	✗	✗	✗
SAHIN et al. (2024)	Children's Dental Panoramic Radiographs Dataset	<b>U-Net 3+</b>	Acurácia, Sensibilidade, Precisão, DSC, IOU, Pixel Acc., AUC, Loss	✗	✗	✗	✓
Zhong et al. (2025)	Children's Dental Panoramic Radiographs Dataset	<b>GCNet</b> , U-Net, SegNet, Attention U-Net, Inf-Net, GT-U-Net, Teeth U-Net	DSC, Sensibilidade, Especificidade, E-MEASURE, MAE	✗	✗	✗	✗
Zhang et al. (2023)	Children's Dental Panoramic Radiographs Dataset	<b>U-Net</b> , R2 U-Net, PSPNet, DeepLab V3	Sensibilidade, Especificidade, Acurácia, Precisão, IOU, DSC	✗	✗	✓	✗
Bhat et al. (2025)	UFBA-UESC Dental Images Dataset, Tufts Dental Dataset (TDD)	U-Net, <b>U-Net++</b> , U-Net + MobileNetV3, DeepLab	IOU, DSC, Acurácia, Precisão, Recall	✗	✗	✗	✗
Almalki et al. (2023)	Tufts Dental Dataset (TDD)	SegNet, <b>U-Net Residual</b> , U-Net, ResNet, DeepLabV3, DeepLabV3+, U-Net++, PSPNet, FPN	DSC médio, IoU médio	✗	✗	✗	✗
Yaswanth et al. (2025)	Tufts Dental Dataset (TDD)	U-Net, FCNs, ENet, <b>U-Net++</b> , U-Net 3+, SwiftNet	IOU, DSC, PSNR, mAP, Acurácia	✗	✗	✗	✗
Bouali et al. (2024)	UFBA-UESC Dental Images Dataset, Tufts Dental Dataset (TDD)	U-Net, Attention U-Net, <b>U-Net 3+</b> , TransUNet	IOU, DSC	✗	✗	✗	✗
Este Trabalho	Children's Dental Panoramic Radiographs Dataset	U-Net, U-Net++, U-Net 3+, W-Net, Attention U-Net	Acurácia, Especificidade, Sensibilidade, IOU, MAE, DSC, E-MEASURE	✓	✓	✓	✓

Fonte: Elaborada pelo autor.

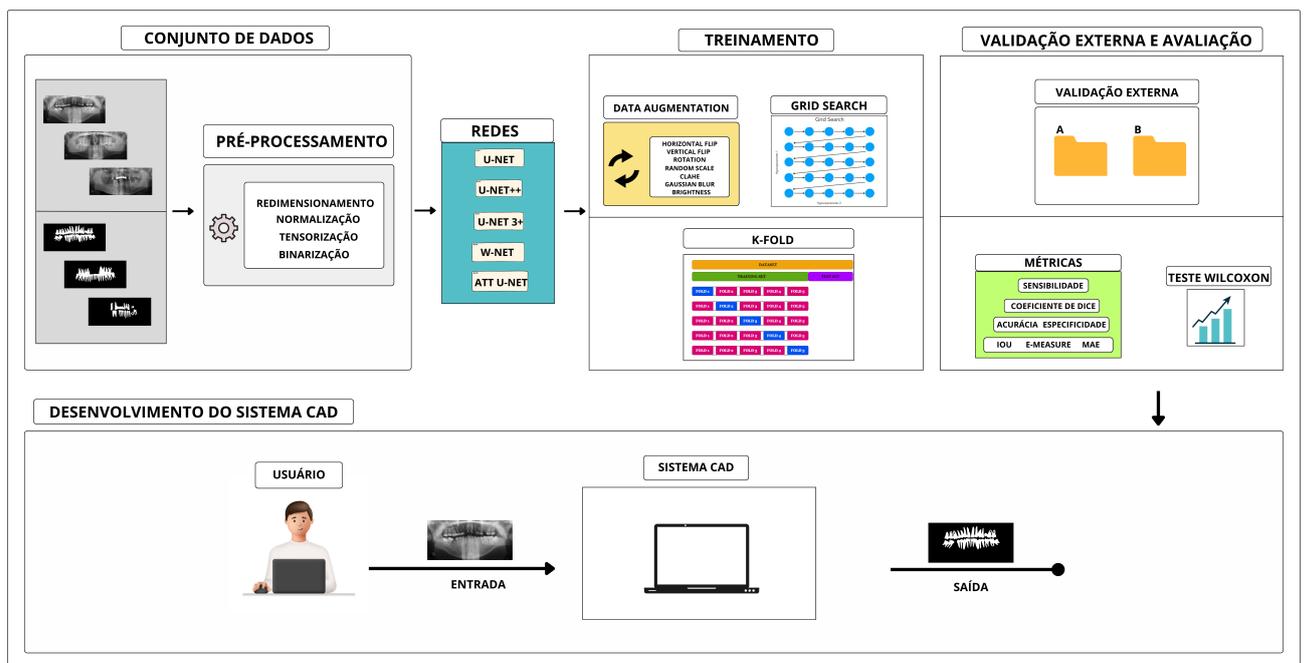
## 4 METODOLOGIA

Neste capítulo, é apresentada a metodologia empregada no desenvolvimento deste trabalho, contemplando todas as etapas necessárias para a construção do modelo de segmentação dental baseado em radiografias panorâmicas. Os experimentos seguem um fluxo sistemático que se inicia com a aquisição do conjunto de dados e finaliza com a implementação do sistema CAD.

### 4.1 Visão Geral

A metodologia empregada neste trabalho, incluindo suas principais etapas, está ilustrada na Figura 12.

Figura 12 – Diagrama da metodologia empregada neste trabalho



Fonte: Elaborado pelo autor.

Primeiramente, estabelece-se o bloco de aquisição do conjunto de dados, composto por radiografias panorâmicas e suas máscaras de segmentação correspondentes. Em seguida, essas imagens passam por uma etapa de pré-processamento, a fim de assegurar maior qualidade aos dados utilizados no treinamento das redes. Após isso, define-se o conjunto de arquiteturas a serem utilizadas: U-Net, U-Net++, U-Net 3+, W-Net e Attention U-Net. Em seguida, inicia-se o bloco de treinamento, no qual se aplicam as técnicas de *Data Augmentation* para ampliar a diversidade do conjunto de dados, bem como a validação cruzada K-Fold, a fim de garantir avaliações mais robustas. Além disso, utiliza-se a técnica de *Grid Search* para otimizar os

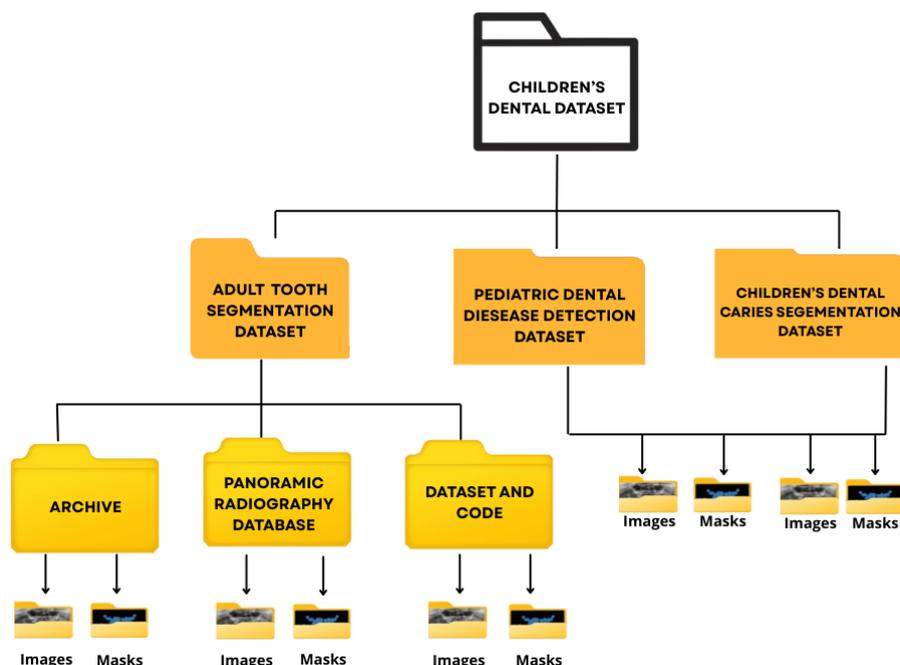
hiperparâmetros de cada arquitetura.

Concluído o treinamento, os modelos são submetidos à validação externa, utilizando dois subdiretórios do conjunto de dados empregados neste trabalho que não participaram das etapas de treino, com o objetivo de avaliar sua capacidade de generalização. Em seguida, o desempenho das arquiteturas é avaliado por meio das seguintes métricas: sensibilidade, especificidade, acurácia, DSC, MAE, E-MEASURE e IOU. Por fim, foi desenvolvido um sistema CAD no qual o modelo com melhor desempenho foi integrado, para permitir o envio de radiografias panorâmicas e a geração automática das segmentações dentárias.

## 4.2 Aquisição do Conjunto de Dados

O conjunto de dados utilizado neste trabalho é o Children's Dental Panoramic Radiographs Dataset, que foi proposto no estudo de Zhang *et al.* (2023) como o primeiro conjunto público internacional de radiografias panorâmicas dentárias de crianças, complementado com imagens de pacientes adultos, para apoiar no processo de segmentação de cáries e a detecção de condições bucais por meio de modelos de aprendizado profundo. Esse material está disponibilizado publicamente na plataforma *Kaggle*, que é uma plataforma online dedicada à ciência de dados e ao aprendizado de máquina e oferece acesso a conjuntos de dados públicos. A estrutura geral do banco de dados é ilustrada na Figura 13.

Figura 13 – Estrutura do conjunto de dados



Fonte: Elaborada pelo autor

No topo, tem-se o *Children's Dental Panoramic Radiographs Dataset*, que corresponde ao conjunto de dados completo e é composto por três subconjuntos distintos. O primeiro é o *Children's Dental Caries Segmentation Dataset*, que reúne 193 radiografias panorâmicas infantis destinadas à segmentação da estrutura dental. O segundo é o *Pediatric Dental Disease Detection Dataset*, composto por 100 radiografias pediátricas, utilizado para a detecção e classificação de dentes afetados por doenças, especialmente cáries. Por fim, tem-se o *Adult Tooth Segmentation Dataset*, estruturado em três diretórios: *Archive* (116 imagens), composto por radiografias panorâmicas dentárias obtidas no Centro de Imagem Médica *Noor (Qom, Irã)*, abrangendo uma ampla gama de condições odontológicas; *Panoramic Radiography Database* (598 imagens), obtido a partir de pacientes selecionados aleatoriamente para um estudo de aprimoramento de imagens; e *Dataset and Code* (1.978 imagens), adquirido em parte no Centro de Diagnóstico por Imagem da UESB e em parte no Hospital Odontológico *Hangzhou Lishui*.

### 4.3 Pré-processamento e Preparação dos Dados

Como observado na Seção 4.2, o conjunto de dados completo é formado por diferentes subconjuntos, contendo radiografias panorâmicas tanto de crianças quanto de adultos. Entretanto, considerando que o escopo deste trabalho é a segmentação da arcada dentária em pacientes adultos, cuja morfologia encontra-se completamente desenvolvida, o que reduz a variabilidade anatômica e favorece a consistência das estruturas a serem segmentadas e, consequentemente, beneficia o treinamento dos modelos, optou-se por selecionar apenas o subconjunto correspondente a essa faixa etária, ou seja, a pasta *Adult Tooth Segmentation Dataset*, que contempla os três diretórios de dados.

A primeira etapa de pré-processamento realizada a partir da escolha do subconjunto a ser utilizado, consistiu em uma análise detalhada dos diretórios correspondentes. Durante essa análise, verificou-se que o diretório *Dataset and Code*, originalmente contendo 1.978 imagens, apresentava 478 imagens duplicadas. A partir dessa identificação, esses dados foram removidos para garantir a consistência do conjunto de dados e evitar redundâncias que poderiam comprometer o treinamento e a avaliação dos modelos. Dessa forma, após essa remoção, o diretório passou a conter 1.500 imagens e, ao se considerar esse quantitativo em conjunto com os demais diretórios disponibilizados, o número total de imagens disponíveis passou a ser de 2.214.

Na sequência, foram realizadas as etapas de pré-processamento das imagens, a fim de padronizar os dados para o treinamento das redes de segmentação. Inicialmente, todas as

imagens foram convertidas para escala de cinza (*grayscale*), uma vez que a segmentação baseia-se exclusivamente na intensidade dos pixels. Em seguida, imagens e máscaras, originalmente com resolução de  $2000 \times 942$  pixels, foram redimensionadas para a resolução de  $256 \times 512$  pixels, o que garante dimensões consistentes para o processamento pelas redes neurais. Os valores de pixel das imagens foram então normalizados para a faixa  $[-1, 1]$ , por meio da aplicação de uma normalização com média 0,5 e desvio padrão 0,5, o que favorece a estabilidade numérica e a convergência do processo de treinamento. Posteriormente, imagens e máscaras foram convertidas em matrizes numéricas e representadas na forma de tensores, os quais são utilizados como entrada pelos modelos de aprendizado profundo. As máscaras de segmentação foram binarizadas, de modo que cada pixel assumisse valor 0 ou 1, o que corresponde às classes de fundo e dente, em conformidade com a formulação da segmentação binária.

#### **4.4 Arquiteturas de Redes Neurais para Segmentação**

Com base no objetivo deste trabalho, que consiste em realizar a segmentação automática de dentes a partir de radiografias panorâmicas, foi realizada uma análise da literatura com o intuito de identificar arquiteturas de CNN adequadas para executar essa tarefa com precisão e eficiência. A partir desse levantamento, verificou-se que arquiteturas do tipo encoder-decoder, especialmente aquelas baseadas na U-Net e suas variações, apresentam desempenho consistente nessa tarefa, sendo, portanto, as selecionadas para implementação neste estudo.

Dentre as arquiteturas identificadas, foram selecionadas para este trabalho as seguintes redes: U-Net, U-Net++, U-Net 3+, W-Net e Attention U-Net. A escolha desse conjunto de redes neurais teve como objetivo possibilitar uma análise comparativa entre essas variações, que se diferenciam principalmente pela forma como exploram conexões entre níveis, fusão multiescala, mecanismos de atenção e estratégias de aprendizado, a fim de avaliar o impacto dessas diferenças na segmentação dentária.

#### **4.5 Treinamento dos modelos de segmentação**

O processo de treinamento dos modelos de segmentação escolhidos foi conduzido de forma sistemática, adotando as mesmas configurações, critérios de avaliação e estratégias de validação para todas as arquiteturas, de modo a garantir uma avaliação consistente do desempenho entre elas.

#### 4.5.1 Aumento de Dados

A técnica de aumento de dados foi aplicada exclusivamente ao conjunto de treinamento por meio da biblioteca Albumentations, em tempo de execução, abrangendo uma série de transformações geométricas e fotométricas. As operações geométricas incluíram inversões horizontais e verticais, rotações aleatórias com limite de até 30 graus e escalonamento aleatório de  $-20\%$  a  $+20\%$  do tamanho original, aplicados com  $50\%$  de probabilidade, permitindo simular variações de orientação, posicionamento e escala das radiografias. Foram empregadas também transformações fotométricas, como ajustes de brilho e contraste, variação controlada de intensidade, aplicação do método CLAHE para realce de contraste local e suavização por desfoque *Gaussiano*, de modo a representar diferentes condições de aquisição e qualidade das imagens. O uso dessa técnica teve como objetivo reduzir o risco de sobreajuste e melhorar a capacidade de generalização dos modelos.

#### 4.5.2 Ajuste de Hiperparâmetros e Validação Cruzada

Ao considerar que o *Adult Tooth Segmentation Dataset*, diretório escolhido para esse estudo, é formado por três subdiretórios distintos, decidiu-se que para as etapas de ajuste de hiperparâmetros (*Grid Search*) e também de treinamento final dos modelos seria utilizado exclusivamente o diretório *Archive* que é composto por 116 imagens panorâmicas. A escolha desse diretório para as etapas descritas anteriormente, esteve relacionada principalmente para analisar o comportamento das arquiteturas selecionadas quando treinadas com um conjunto limitado de dados, bem como à sua capacidade de generalização para dados não vistos.

Para cada arquitetura, foi utilizado o mesmo conjunto de hiperparâmetros, definido com as seguintes configurações: taxas de aprendizado de  $1 \times 10^{-3}$  e  $5 \times 10^{-4}$ , tamanho do lote (*batch size*) igual a 8, e os otimizadores Adam e AdamW. A otimização desses hiperparâmetros foi conduzida por meio da estratégia de *Grid Search*, na qual todas as combinações possíveis dentro do espaço definido são testadas de forma exaustiva. Essa estratégia permitiu verificar de que forma as diferentes combinações de hiperparâmetros afetavam os resultados das redes, para garantir comparações justas entre elas e manter um bom equilíbrio entre estabilidade, aprendizado correto e eficiência. Durante essa etapa, cada configuração foi treinada por 50 épocas, número considerado suficiente para a convergência inicial dos modelos e para permitir comparação consistente entre as combinações avaliadas. Além disso, o treinamento foi realizado de forma

supervisionada, utilizando a função de perda *combined loss*, que combina *BCEWithLogitsLoss* e *Dice Loss* para penalizar discrepâncias entre as máscaras preditas e as máscaras de referência. Os pesos das redes foram ajustados por retropropagação do erro, utilizando os otimizadores Adam ou AdamW, conforme a configuração avaliada.

A avaliação das diferentes configurações foi realizada utilizando validação cruzada *k-fold*, com  $k = 10$ . Em cada *fold*, o conjunto de dados do diretório *Archive* foi dividido em subconjuntos de treinamento e teste, sendo que o subconjunto de treinamento foi, por sua vez, dividido em treinamento e validação. Com isso, o pipeline efetivamente utiliza aproximadamente 72% dos dados para treino, 18% para validação e 10% para teste em cada *fold*. Essa estratégia garante que todas as amostras sejam aproveitadas tanto para treinamento quanto para validação ao longo do processo, isso reduz o viés de uma única divisão e fornece uma estimativa mais confiável da capacidade de generalização dos modelos.

Ao final de cada *fold*, selecionou-se como melhor configuração aquela que apresentou o maior valor de DSC no conjunto de validação, e os pesos do modelo correspondentes, bem como os hiperparâmetros associados, foram armazenados para uso posterior. O conjunto de teste de cada *fold* foi utilizado para a avaliação do desempenho da configuração selecionada. Após a conclusão de todos os folds da validação cruzada, a configuração global final de hiperparâmetros de cada arquitetura foi definida com base no maior valor médio de DSC obtido no conjunto de teste, o que garante uma seleção mais robusta e menos dependente de variações específicas do conjunto de dados.

Após a definição da melhor configuração de cada rede, realizou-se o treinamento final dos modelos no conjunto completo *Archive*, utilizando as configurações ótimas de hiperparâmetros. Nessa etapa, os modelos foram treinados por 50 épocas, para garantir consistência metodológica com a fase de otimização e aplicando as mesmas estratégias adotadas anteriormente, ou seja, aumento de dados e otimização por Adam ou AdamW. Ao término do treinamento, os modelos finais foram armazenados para uma posterior avaliação quantitativa.

#### **4.6 Validação externa**

Após o treinamento final dos modelos utilizando o dataset *Archive* com a melhor configuração de hiperparâmetros obtida via *Grid Search* e validação cruzada *k-fold*, foi utilizada a técnica de validação externa para avaliar a capacidade de generalização dos modelos em dados não vistos durante o treinamento. Para isso, foram utilizados os outros dois subdiretórios

que compõem o conjunto de dados de pacientes adultos: *Dataset and code* e o *Panoramic Radiography Database*. Antes da inferência, todas as imagens e suas respectivas máscaras desses conjuntos passaram pelo mesmo fluxo de pré-processamento adotado na fase de validação, ou seja redimensionamento, normalização e conversão para tensores *PyTorch*, isso visa assegurar a padronização dos dados de entrada e a consistência com o procedimento utilizado durante o treinamento dos modelos. Em seguida, os modelos finais, ajustados com a melhor configuração global de hiperparâmetros, foram carregados para realizar a inferência, o que possibilitou a avaliação individual de todas as imagens dos datasets externos.

Para verificar o desempenho de cada arquitetura no processo de segmentação das estruturas dentárias, cada uma das redes gerou 10 máscaras por dataset, armazenadas em formato PNG. Esse procedimento possibilitou a realização de uma análise qualitativa visual do comportamento das redes e permitiu a comparação entre as arquiteturas quanto à qualidade dos contornos segmentados, à coerência das regiões identificadas e à preservação das estruturas dentárias de interesse. Além disso, a adoção de um número reduzido de máscaras evita o armazenamento excessivo de arquivos e mantém, ainda assim, uma amostra representativa da capacidade de generalização dos modelos em dados externos.

#### **4.7 Avaliação de desempenho**

Para avaliar o desempenho dos modelos no processo de validação externa utilizando os dois conjuntos de dados citados na seção 4.6, selecionamos um conjunto de métricas de avaliação que são amplamente utilizadas em tarefas de segmentação de imagens. Essas métricas foram escolhidas pois quantificam, de forma objetiva, a correspondência entre as regiões segmentadas pelos modelos e as máscaras de referência, isso permite avaliar tanto a sobreposição quanto os erros de segmentação. Dessa forma, após uma busca na literatura, as métricas escolhidas foram: Acurácia, Sensibilidade, Especificidade, Intersecção sobre União (IOU), Coeficiente de Similaridade de Dice (DSC), Erro Médio Absoluto (MAE) e Medida de Alinhamento Aprimorada (E-MEASURE), as quais estão detalhadamente descritas no Capítulo 2.15 da Fundamentação Teórica.

## 4.8 Ambiente de desenvolvimento

Os experimentos foram realizados em um computador com sistema operacional Linux, com distribuição Ubuntu, GPU NVIDIA GeForce RTX 3090, com 24 GB de memória dedicada, processador Intel Core i7 e 32 GB de memória RAM. O ambiente computacional contou com driver NVIDIA versão 580.95.05 e suporte à CUDA 13.0. A implementação foi desenvolvida na linguagem de programação Python, tendo como principal biblioteca o PyTorch, empregado na definição, treinamento e avaliação das arquiteturas de redes neurais. Como suporte ao pipeline experimental, foram utilizadas bibliotecas auxiliares: *Albumentations* para aumento de dados, *NumPy* e *Pandas* para manipulação de dados, *scikit-learn* e *imbalanced-learn* para cálculo das métricas de avaliação e validação experimental.

## 4.9 Análise Estatística: Teste de Wilcoxon

A partir da etapa de avaliação dos modelos com as métricas selecionadas, foi aplicado o teste estatístico não paramétrico de Wilcoxon para amostras pareadas, com o objetivo de verificar se as diferenças observadas no desempenho das arquiteturas avaliadas são estatisticamente significativas. Esse teste compara duas amostras relacionadas a partir das diferenças entre observações correspondentes e foi escolhido por não assumir normalidade dos dados, além de ser adequado à comparação de desempenhos obtidos a partir dos mesmos conjuntos de dados. O teste foi conduzido com base nos valores do Coeficiente de Similaridade de Dice (DSC) obtidos durante o processo de validação cruzada *k-fold*, com comparações pareadas entre todas as combinações de arquiteturas avaliadas.

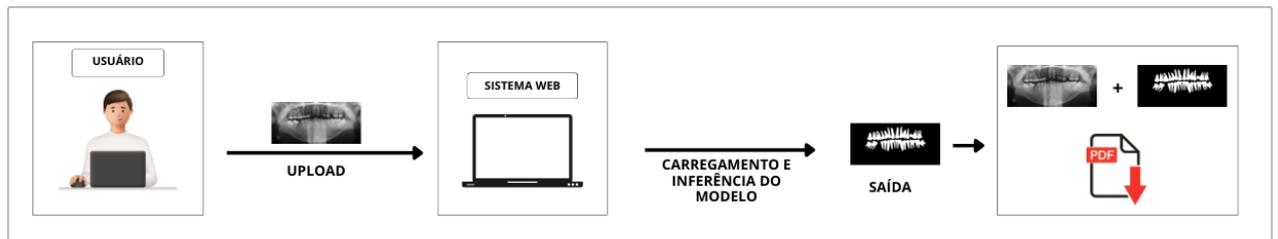
Para cada par de modelos, foram considerados os valores de DSC correspondentes aos mesmos folds, o que assegurou a comparabilidade direta entre os resultados. O nível de significância adotado foi de 5% ( $\alpha = 0,05$ ), sendo consideradas estatisticamente significativas as diferenças com *p-value* inferior ou igual a esse limiar. Dessa forma, o teste de Wilcoxon permitiu identificar se as variações de desempenho observadas entre as arquiteturas refletem diferenças estatisticamente significativas ou não.

## 4.10 Desenvolvimento do Sistema CAD

A partir do processo de avaliação dos modelos implementados e da análise dos resultados gerados por cada um deles, aliado à análise estatística por meio do teste de Wilcoxon,

selecionou-se o modelo de segmentação que apresentou o melhor desempenho no processo de segmentação das estruturas dentárias em comparação com as demais arquiteturas avaliadas. Diante desse resultado, o modelo foi incorporado a um sistema CAD desenvolvido neste trabalho, com o objetivo de viabilizar a aplicação prática do método proposto. Conforme ilustrado na Figura 14, o sistema foi desenvolvido de modo a permitir a interação direta do usuário com a plataforma, de forma a possibilitar a execução automática do modelo previamente treinado.

Figura 14 – Fluxo do Sistema CAD



Fonte: Elaborado pelo autor.

O sistema CAD foi implementado utilizando o *Streamlit*, uma biblioteca *Python* de código aberto que permite o desenvolvimento rápido de aplicações web interativas e visualmente atrativas com o mínimo de codificação, o que possibilita a criação de interfaces dinâmicas e facilita a integração entre modelos de aprendizado profundo e aplicações baseadas na web.

Inicialmente, o usuário envia uma radiografia panorâmica por meio da interface web. Após o envio, o sistema realiza o carregamento do modelo de segmentação selecionado e executa o processo de inferência sobre a imagem submetida. Como resultado, a rede neural gera a máscara de segmentação correspondente às estruturas dentárias. O resultado do processamento é apresentado diretamente ao usuário, com a exibição da máscara de segmentação em conjunto com a radiografia panorâmica original, o que permite a análise comparativa dos resultados. Além disso, o sistema oferece a opção de exportação, permitindo o *download* de um arquivo em formato PDF que contém a radiografia panorâmica e a segmentação gerada lado a lado. Essa integração demonstra que o modelo escolhido pode ser aplicado em um ambiente computacional acessível, o que evidencia a transição de um método experimental para uma solução funcional baseada na web, com potencial de aplicação em ambientes clínicos controlados.

## 5 RESULTADOS E DISCUSSÕES

Neste capítulo, são apresentados e discutidos os resultados obtidos a partir da metodologia proposta no Capítulo 4. Inicialmente, são exibidas as melhores configurações de hiperparâmetros de cada arquitetura, obtidas por meio do procedimento de *Grid Search* e validação cruzada. Na sequência, são exibidos os resultados obtidos na etapa de validação externa, referentes ao *Dataset and code* e ao *Panoramic Radiography Database*, que mostram o desempenho das redes nesses conjuntos de dados e permitem a análise de seu comportamento em ambos. Após essa etapa, é discutida a análise estatística do desempenho das redes, obtida por meio do teste estatístico de Wilcoxon, para verificar se houve diferenças significativas entre elas. Por fim, apresenta-se o sistema CAD desenvolvido, com demonstração de seu funcionamento por meio de exemplos de radiografias panorâmicas e das segmentações geradas, para demonstrar sua aplicabilidade prática.

### 5.1 Resultados da seleção de hiperparâmetros por arquitetura

A Tabela 2 apresenta as melhores configurações de hiperparâmetros obtidas para cada arquitetura. Para cada rede, são exibidos os parâmetros que resultaram no maior valor médio de DSC durante o processo de seleção, incluindo a *learning rate*, o tamanho do *batch* e o otimizador.

Tabela 2 – Melhor configuração de hiperparâmetros de cada arquitetura

Arquitetura	Batch size	Learning rate	Otimizador
Attention U-Net	8	$5 \times 10^{-4}$	Adam
U-Net 3+	8	$5 \times 10^{-4}$	Adam
U-Net++	8	$5 \times 10^{-4}$	AdamW
U-Net	8	$5 \times 10^{-4}$	AdamW
W-Net	8	$5 \times 10^{-4}$	Adam

Fonte: Elaborada pelo autor.

Conforme observado na Tabela 2, as arquiteturas diferiram apenas quanto ao otimizador utilizado, indicando que, embora existam hiperparâmetros comuns que favorecem a convergência, como *batch size* e *learning rate*, a estratégia de atualização dos pesos impacta de forma distinta o desempenho das redes. Em particular, Attention U-Net, U-Net 3+ e W-Net obtiveram melhores resultados com Adam, enquanto U-Net e U-Net++ apresentaram desempenho superior com AdamW. Essa diferença sugere que arquiteturas com blocos de atenção,

conexões densas aninhadas e maiores níveis de profundidade hierárquica se beneficiam de menor regularização explícita, permitindo maior flexibilidade na adaptação dos pesos, o que é relevante em dados clínicos com ruído e variabilidade anatômica elevada.

Por outro lado, redes como U-Net e U-Net++, que dependem fortemente da propagação de características de baixo nível por meio de skip connections e refinamento progressivo das representações, se beneficiam de uma regularização mais explícita, como a introduzida pelo AdamW, o que favorece a estabilidade durante o treinamento e melhor capacidade de generalização. A padronização da *learning rate* e do *batch size* para todas as arquiteturas evidencia que esses hiperparâmetros forneceram equilíbrio adequado entre estabilidade e eficiência no processo de treinamento. Por fim, as configurações indicadas na Tabela 2 foram utilizadas no treinamento final das arquiteturas, de modo a assegurar um treinamento mais consistente e o melhor desempenho possível para cada uma delas.

## 5.2 Resultados da avaliação no conjunto de teste

Os resultados apresentados na Tabela 3 correspondem ao desempenho das arquiteturas avaliadas no conjunto de teste durante a etapa de ajuste de hiperparâmetros e validação cruzada em 10 folds. Em cada *fold*, as métricas foram calculadas a partir do subconjunto de teste, e os valores reportados representam a média e o desvio padrão obtidos ao longo das dez iterações, o que reflete a consistência do desempenho das arquiteturas ao longo do processo de avaliação.

Tabela 3 – Resultados das métricas no conjunto de teste.

ARQUITETURA	ACURÁCIA	ESPECIFICIDADE	SENSIBILIDADE	E-MEASURE	MAE	IOU	DSC
U-Net	0,983 ± 0,002	0,990 ± 0,002	0,950 ± 0,011	0,977 ± 0,002	0,017 ± 0,002	0,905 ± 0,008	0,950 ± 0,004
<b>U-Net++</b>	<b>0,985 ± 0,001</b>	<b>0,991 ± 0,002</b>	<b>0,952 ± 0,010</b>	<b>0,979 ± 0,001</b>	<b>0,015 ± 0,001</b>	<b>0,913 ± 0,006</b>	<b>0,954 ± 0,003</b>
U-Net 3+	0,962 ± 0,002	0,973 ± 0,005	0,906 ± 0,018	0,946 ± 0,004	0,038 ± 0,002	0,801 ± 0,006	0,889 ± 0,004
Attention U-Net	0,983 ± 0,001	0,990 ± 0,002	0,947 ± 0,008	0,977 ± 0,002	0,017 ± 0,001	0,905 ± 0,006	0,950 ± 0,004
W-Net	0,983 ± 0,002	0,990 ± 0,002	0,948 ± 0,010	0,977 ± 0,002	0,017 ± 0,002	0,905 ± 0,008	0,950 ± 0,004

Fonte: Elaborado pelo autor.

Conforme observado na Tabela 3, as arquiteturas U-Net, U-Net++, Attention U-Net e W-Net apresentaram os melhores desempenhos globais no conjunto de teste, com valores médios elevados de acurácia, especificidade, sensibilidade, E-MEASURE, IOU e DSC, além de baixos valores de MAE. Esses resultados indicam elevada capacidade de segmentação das estruturas dentárias, tanto em termos de classificação pixel a pixel quanto de sobreposição entre as máscaras preditas e as máscaras de referência. Entre os modelos avaliados, a U-Net++

destacou-se de forma consistente por apresentar as maiores médias nas métricas de sobreposição espacial (IOU e DSC), sugerindo melhor delineamento das regiões segmentadas. As arquiteturas U-Net, Attention U-Net e W-Net, por sua vez, exibiram desempenhos muito próximos entre si em todas as métricas analisadas, com desvios padrão reduzidos, o que evidencia estabilidade e baixa variabilidade dos resultados ao longo dos *folds*.

Em contraste, a arquitetura U-Net 3+ apresentou desempenho inferior em relação às demais, especialmente nas métricas IOU e DSC, além de maiores valores de MAE, o que indica menor concordância espacial entre as segmentações preditas e as máscaras da verdade fundamental. Ainda assim, os desvios padrão relativamente baixos sugerem que esse comportamento foi consistente ao longo das diferentes partições do conjunto de dados. De forma geral, a combinação de altos valores médios e baixa dispersão das métricas para a maioria das arquiteturas reforça a confiabilidade dos resultados obtidos, isso indica que os modelos apresentaram desempenho consistente na segmentação das estruturas dentárias, com comportamento estável e pouca variação entre as amostras avaliadas.

### **5.3 Resultados no conjunto de dados Dataset and code**

Os resultados referentes ao conjunto *Dataset and code* estão apresentados na Tabela 4. Neste estudo, cada métrica avalia um aspecto específico da segmentação das estruturas dentárias. A acurácia indica o percentual total de pixels corretamente classificados como dentes ou como fundo, a sensibilidade está relacionada à capacidade do modelo de detectar corretamente as regiões de interesse e mostra o quanto ele identifica os verdadeiros positivos e a especificidade avalia o quão bem o modelo reconhece as regiões de fundo, refletindo o controle sobre a geração de falsos positivos. Além disso, o DSC e a IOU quantificam a sobreposição entre a máscara predita e a referência, sendo a IOU mais rigorosa ao penalizar pequenas diferenças espaciais nas bordas dentárias. Por sua vez, o MAE indica o desvio médio pixel a pixel entre a predição e a referência, enquanto a E-MEASURE avalia o alinhamento global da segmentação e analisa a coerência estrutural da dentição.

Tabela 4 – Resultados das métricas no Dataset and code

ARQUITETURAS	ACURÁCIA	ESPECIFICIDADE	SENSIBILIDADE	E-MEASURE	MAE	IOU	DSC
U-Net	0,952	0,995	0,782	0,914	0,048	0,767	0,868
U-Net++	0,953	0,994	0,789	0,917	0,047	0,772	0,871
U-Net 3+	0,946	0,991	0,768	0,906	0,054	0,741	0,851
Attention U-Net	0,951	0,994	0,778	0,912	0,049	0,760	0,864
<b>W-Net</b>	<b>0,955</b>	<b>0,992</b>	<b>0,809</b>	<b>0,925</b>	<b>0,045</b>	<b>0,785</b>	<b>0,879</b>

Fonte: Elaborado pelo autor.

Como mostra a Tabela 4, todas as arquiteturas apresentaram acurácia elevada, evidenciando um desempenho consistente na distinção entre dentes e fundo. De forma complementar, a alta especificidade indica que os modelos foram eficazes em reconhecer corretamente as regiões de fundo, reduzindo a ocorrência de falsos positivos. Esse comportamento é particularmente relevante no contexto clínico, pois evita a identificação indevida de estruturas inexistentes como dentes, minimizando ruídos na segmentação e aumentando a confiabilidade do sistema CAD como ferramenta de apoio ao diagnóstico. No detalhamento dos demais resultados, observa-se que a U-Net apresentou desempenho sólido como modelo base, com IOU de 0,767 e DSC de 0,868. Isso deixa claro que houve uma boa sobreposição entre as máscaras geradas por ela quando comparada a verdade fundamental, ainda com pequenas imprecisões nos contornos dentários. A sensibilidade de 0,782 indica recuperação consistente das regiões de interesse, com baixa perda estrutural. Já em termos de fidelidade local, o MAE de 0,048 e a E-MEASURE de 0,914 sugerem segmentações estáveis e com baixo erro pixel a pixel.

Em paralelo, a U-Net++ apresentou um leve avanço em relação à U-Net, especialmente nos contornos dentários. Isso se refletiu quando observado o aumento do IOU 0,772 e do DSC 0,871, indicando sobreposição mais precisa e melhor preservação dos contornos dentários. Ademais, a sensibilidade superior 0,789 aponta maior recuperação das regiões dentárias, enquanto o menor MAE e o maior valor de E-MEASURE reforçam a redução de erros locais e a maior coerência estrutural, aspecto atribuído às skip connections densas. Por outro lado, a U-Net 3+ apresentou os menores valores globais, com IOU de 0,741 e DSC de 0,851, revelando menor sobreposição e maior fragmentação das máscaras. A sensibilidade de 0,768 sugere maior incidência de falsos negativos, o que resulta em perda de detalhes dentários; o MAE mais elevado e a menor E-MEASURE corroboram o aumento do erro local e a menor consistência estrutural, aspecto que evidencia limitações da arquitetura neste conjunto.

De maneira intermediária, a Attention U-Net obteve desempenho situado entre os modelos anteriores, com IOU de 0,760 e DSC de 0,864, próximos aos da U-Net, indicando

contornos estáveis, porém sem ganhos expressivos em sobreposição. A sensibilidade de 0,778 reforça sua capacidade moderada de recuperar as regiões dentárias. Em complemento, os valores de MAE e E-MEASURE indicam precisão local e coerência estrutural semelhantes às da U-Net.

Por fim, a W-Net destacou-se como a melhor arquitetura, com os maiores valores de IOU 0,785 e DSC 0,879, o que demonstra excelente preservação da forma e dos contornos dentários, com sobreposição mais consistente entre predição e a verdade fundamental. Sua sensibilidade foi a mais alta 0,809, para deixar claro sua capacidade de recuperar integralmente as regiões dentárias. Além disso, apresentou o menor MAE, isso reflete em um menor nível de erro local, e o maior valor de E-MEASURE, para confirmar o alinhamento estrutural superior e consolidando a W-Net como o melhor desempenho entre as arquiteturas avaliadas.

#### 5.4 Resultados no conjunto de dados Panoramic Radiography Database

A Tabela 5 apresenta os resultados das métricas obtidas no segundo conjunto da validação externa *Panoramic Radiography Database*, para garantir a comparabilidade entre as arquiteturas avaliadas.

Tabela 5 – Resultados das métricas no Panoramic Radiography Database

ARQUITETURAS	ACURÁCIA	ESPECIFICIDADE	SENSIBILIDADE	E-MEASURE	MAE	IOU	DSC
U-Net	0,980	0,983	0,956	0,972	0,020	0,844	0,915
U-Net++	0,981	0,984	0,963	0,973	0,019	0,852	0,920
U-Net 3+	0,970	0,978	0,903	0,960	0,030	0,770	0,870
<b>Attention U-Net</b>	<b>0,982</b>	<b>0,984</b>	<b>0,965</b>	<b>0,975</b>	<b>0,018</b>	<b>0,858</b>	<b>0,924</b>
W-Net	0,977	0,980	0,957	0,966	0,023	0,824	0,904

Fonte: Elaborado pelo autor.

Conforme observado na Tabela 5, o desempenho nas métricas de acurácia e especificidade manteve-se elevado, seguindo o mesmo comportamento verificado no conjunto anterior. Esse resultado indica que, em ambos os datasets, as arquiteturas foram capazes de generalizar adequadamente na distinção entre estruturas dentárias e fundo. Além disso, tal comportamento sugere controle consistente de falsos positivos e boa adequação dos modelos às características radiográficas deste conjunto externo. Quanto ao desempenho individual, a U-Net apresentou avanço expressivo em relação ao primeiro conjunto, atingindo IOU de 0,844 e DSC de 0,915, com máscaras mais bem ajustadas aos contornos dentários. Essa melhora se refletiu também na recuperação das regiões de interesse, com sensibilidade de 0,956, e no alinhamento estrutural, dada a redução do MAE e o aumento da E-MEASURE, o que resultou em segmentações mais

estáveis.

Ao analisar a U-Net++, percebe-se que ela manteve as melhorias previamente observadas nas métricas de sobreposição, alcançando IOU de 0,852 e DSC de 0,920, valores que denotam contornos mais refinados e maior aderência espacial. Nesse caso, a capacidade de captura das regiões dentárias foi ainda maior, uma vez que o modelo alcançou sensibilidade de 0,963, enquanto os ganhos em MAE e E-MEASURE consolidaram a coerência global e local das segmentações. Esse conjunto de resultados reforça o papel das skip connections densas no refinamento estrutural das máscaras. No que concerne à U-Net 3+, observou-se uma melhora absoluta em todas as métricas, embora a arquitetura tenha permanecido como a de menor desempenho relativo neste conjunto externo. Com IOU de 0,770 e DSC de 0,870, o modelo apresentou maior fragmentação nas regiões dentárias, com falhas na continuidade dos contornos, o que resulta em uma máscara menos fiel à estrutura real dos dentes, enquanto a sensibilidade de 0,903 ainda denota perda de detalhes dentários; mesmo assim, esses valores mostram-se suficientes para produzir uma segmentação precisa. O MAE mais elevado e o menor E-MEASURE entre as arquiteturas confirmam segmentações menos consistentes e maior incidência de falsos negativos.

Em contraste, a W-Net apresentou desempenho sólido, com maior precisão local, evidenciada pelo menor MAE entre as arquiteturas, e bom desempenho estrutural. O IOU de 0,824 e o DSC de 0,904 indicam que as máscaras preditas mantiveram boa sobreposição com a verdade fundamental, preservando os contornos dentários e a continuidade das regiões segmentadas. A sensibilidade de 0,957 demonstra elevada capacidade de recuperar integralmente as regiões dentárias, reduzindo a perda de detalhes. Além disso, manteve-se bom desempenho na preservação da forma e dos contornos dentários, comportamento já observado no conjunto anterior.

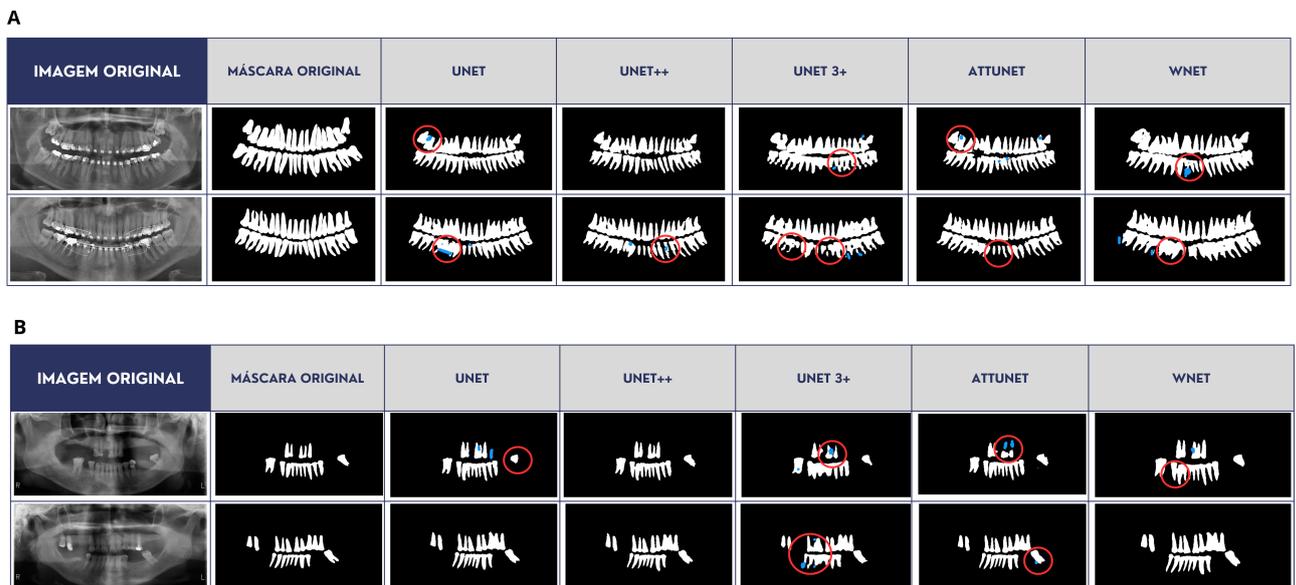
Por fim, a Attention U-Net apresentou o melhor desempenho global no *Panoramic Radiography Database*. Com os maiores valores de IOU 0,858 e DSC 0,924, as máscaras produzidas pelo modelo demonstraram excelente correspondência com a referência, com contornos dentários mais precisos e continuidade das regiões segmentadas. A sensibilidade de 0,965 evidenciou que quase todas as regiões dentárias foram corretamente identificadas, o que reflete elevada capacidade de detecção de áreas críticas. Além disso, com o menor MAE 0,018 e o maior E-MEASURE 0,975, as segmentações apresentaram maior uniformidade e coerência estrutural, aspecto que evidencia a contribuição efetiva do mecanismo de atenção para realçar as regiões de

interesse.

## 5.5 Análise qualitativa das máscaras segmentadas

Como descrito na Seção 4.6, a avaliação qualitativa do processo de segmentação foi realizada a partir da inspeção visual das máscaras geradas pelas arquiteturas na etapa de validação externa. Para exemplificação, foram selecionadas duas amostras de cada rede, mantendo o caráter ilustrativo da comparação entre os modelos. A Figura 15 apresenta os resultados das máscaras nos dois conjuntos de dados.

Figura 15 – Resultados qualitativos da segmentação: A - Dataset and code; B - Panoramic Radiography Database



Fonte: Elaborado pelo autor.

Conforme ilustrado nas Figuras 15 (A) e (B), os resultados qualitativos da segmentação dental evidenciam algumas diferenças entre as arquiteturas avaliadas nos dois conjuntos de dados. Observa-se que a U-Net apresentou limitações recorrentes em ambos os conjuntos de dados, evidenciadas principalmente pela fragmentação das regiões dentárias, pela descontinuidade dos arcos e pela inclusão de pequenas áreas indevidas, conforme indicado nas regiões destacadas. Esses resultados indicam dificuldades do modelo em lidar com variações anatômicas mais complexas e em manter contornos dentários bem definidos, especialmente em regiões de sobreposição e baixo contraste. Em contrapartida, a U-Net++ apresentou desempenho qualitativo consistente nos dois cenários avaliados, pois produziu segmentações mais regulares e visualmente coerentes. Observa-se melhor preservação da forma dentária, com contornos mais contínuos e

próximos às máscaras de referência, inclusive em regiões estruturalmente mais complexas, o que indica maior estabilidade do modelo e menor ocorrência de artefatos.

Por outro lado, a U-Net 3+ demonstrou maior instabilidade visual ao longo das amostras analisadas. As segmentações exibiram falhas recorrentes na definição das estruturas dentárias, com regiões irregulares, variações de espessura e perdas de continuidade, o que sugere dificuldades de generalização quando aplicada a imagens panorâmicas com maior variabilidade anatômica e presença de ruídos. De forma intermediária, a Attention U-Net apresentou apenas um refinamento parcial das regiões de interesse. Embora o mecanismo de atenção contribua para destacar áreas relevantes, ainda foram observadas inconsistências em regiões de sobreposição dentária e pequenas falhas de segmentação, especialmente nas regiões posteriores, resultando em ganhos qualitativos limitados em relação à U-Net padrão. Por fim, a W-Net demonstrou bom desempenho qualitativo, com redução de ruídos e preservação adequada da forma geral dos dentes. As segmentações mostraram-se mais homogêneas e menos fragmentadas quando comparadas à U-Net e à U-Net 3+, porém, em comparação direta com a U-Net++, ainda apresentaram ocorrências pontuais de espessamento excessivo e pequenas imprecisões nos limites segmentados.

No contexto do atendimento clínico, as falhas qualitativas observadas na segmentação apresentada na Figura 15 podem induzir interpretações incorretas da imagem radiográfica. Nesse sentido, a união indevida de dentes adjacentes em uma única estrutura pode levar à interpretação equivocada de fusão dentária ou ausência de espaço interdentário, enquanto a fragmentação ou segmentação incompleta das raízes pode sugerir, de forma incorreta, a presença de fraturas radiculares ou alterações estruturais inexistentes. Assim, esse tipo de erro pode impactar diretamente o diagnóstico, uma vez que o profissional pode considerar como patológica uma condição que não está presente na imagem original.

## **5.6 Análise estatística por meio do teste de Wilcoxon**

Para avaliar se as diferenças de desempenho entre as arquiteturas eram estatisticamente significativas, foi aplicado o teste de Wilcoxon com base nos valores do DSC obtidos durante o processo de validação cruzada. A Tabela 6 sintetiza os *p-value* das comparações pareadas entre as redes, considerando o nível de significância adotado ( $\alpha = 0,05$ ).

Tabela 6 – Teste de Wilcoxon: comparação entre arquiteturas

Comparação	p-value	Diferença estatisticamente significativa
U-Net × AttU-Net	1,00	✗
U-Net × U-Net 3+	0,01	✓
U-Net × U-Net++	0,02	✓
U-Net × W-Net	1,00	✗
U-Net 3+ × AttU-Net	0,01	✓
U-Net++ × AttU-Net	0,01	✓
U-Net++ × U-Net 3+	0,01	✓
W-Net × AttU-Net	0,69	✗
W-Net × U-Net 3+	0,01	✓
W-Net × U-Net++	0,02	✓

Fonte: Elaborado pelo autor.

Com base na Tabela 6, a U-Net não apresentou variações relevantes em comparação à Attention U-Net e à W-Net ( $p$ -value = 1,00 em ambos os casos), indicando, portanto, equivalência de desempenho entre esses modelos no conjunto de dados avaliado. Entretanto, ao comparar a U-Net com arquiteturas mais elaboradas, observaram-se diferenças estatísticas, sendo que a U-Net difere da U-Net 3+ ( $p$ -value = 0,01) e da U-Net++ ( $p$  = 0,02), o que evidencia que as modificações arquiteturais desses modelos impactam de forma relevante os resultados de segmentação.

Adicionalmente, a U-Net 3+ apresentou discrepâncias notáveis na maioria das comparações realizadas, o que reforça seu comportamento distinto em relação às demais arquiteturas e sugerindo que sua estratégia de agregação multiescala e maior complexidade estrutural produzem respostas diferenciadas no contexto avaliado. A U-Net++ também se destacou, apresentando resultados estatisticamente distintos em relação à Attention U-Net e à U-Net 3+ ( $p$ -value = 0,01), isso indica um padrão de desempenho consistente e superior frente a abordagens menos sofisticadas. No que se refere à W-Net, não houve diferença relevante em relação à Attention U-Net ( $p$ -value = 0,69), o que sugere equivalência de desempenho entre essas redes. Contudo, ao compará-la à U-Net 3+ e à U-Net++, observaram-se variações significativas ( $p$  – value < 0,05), isso evidencia que, embora a W-Net apresente desempenho competitivo, ela responde de maneira distinta frente a modelos com estratégias arquiteturais específicas.

De modo geral, o teste de Wilcoxon revela que arquiteturas que empregam estratégias de agregação mais densas, como a U-Net 3+ e a U-Net++, apresentam resultados superiores em relação à U-Net, com base nos valores do DSC. Por outro lado, os resultados obtidos com a W-Net e a Attention U-Net mostram que maior complexidade arquitetural nem sempre se traduz

em melhorias relevantes em relação à U-Net clássica. Esses achados ressaltam a importância da análise estatística para confirmar se alterações estruturais efetivamente proporcionam benefícios práticos no cenário do conjunto de dados analisado.

### 5.7 Seleção da melhor arquitetura

Com base na análise dos resultados quantitativos, observou-se que, no primeiro conjunto de dados *Dataset and code*, a W-Net apresentou o melhor desempenho, enquanto, no segundo conjunto *Panoramic Radiography Database*, a Attention U-Net se destacou como a melhor. Na avaliação estatística por meio do teste de Wilcoxon, as arquiteturas U-Net 3+ e U-Net++ apresentaram diferenças estatisticamente significativas em relação a outras redes, tornando evidente seu desempenho diferenciado. Além disso, a análise qualitativa das máscaras segmentadas mostrou que a U-Net++ apresentou melhor preservação estrutural e contornos mais precisos. Dessa forma, para a seleção da rede mais adequada a ser utilizada no desenvolvimento do sistema CAD, consideraram-se os resultados quantitativos e qualitativos, bem como o custo computacional, avaliado pelo tempo necessário para a segmentação de cada imagem. Em contextos clínicos, a rapidez é um critério essencial, pois o processamento eficiente de grandes volumes de radiografias permite maior agilidade no atendimento, otimiza o tempo de análise e facilita decisões mais rápidas durante o diagnóstico.

Para quantificar o custo computacional mencionado anteriormente, cada rede treinada foi integrada ao sistema e o tempo de predição de cada imagem foi medido em milissegundos usando `time.perf_counter()`, considerando apenas o *forward pass*, ou seja, a etapa do modelo responsável pela geração da máscara segmentada. Para cada arquitetura, foram segmentadas 50 radiografias panorâmicas, registrando-se o tempo individual de predição para cada imagem. A média final do tempo e o desvio padrão obtidos após as 50 execuções de cada rede encontram-se apresentados na Tabela 7.

Tabela 7 – Tempo de inferência das redes

<b>Rede</b>	<b>Média (ms)</b>	<b>Desvio Padrão (ms)</b>
U-Net	≈ 2798,6	≈ 110,4
<b>Attention U-Net</b>	≈ 1527,3	≈ 72,5
W-Net	≈ 3289,8	≈ 115,6
U-Net 3+	≈ 1726,8	≈ 118,9
U-Net++	≈ 1658,7	≈ 130,2

Fonte: Elaborado pelo autor.

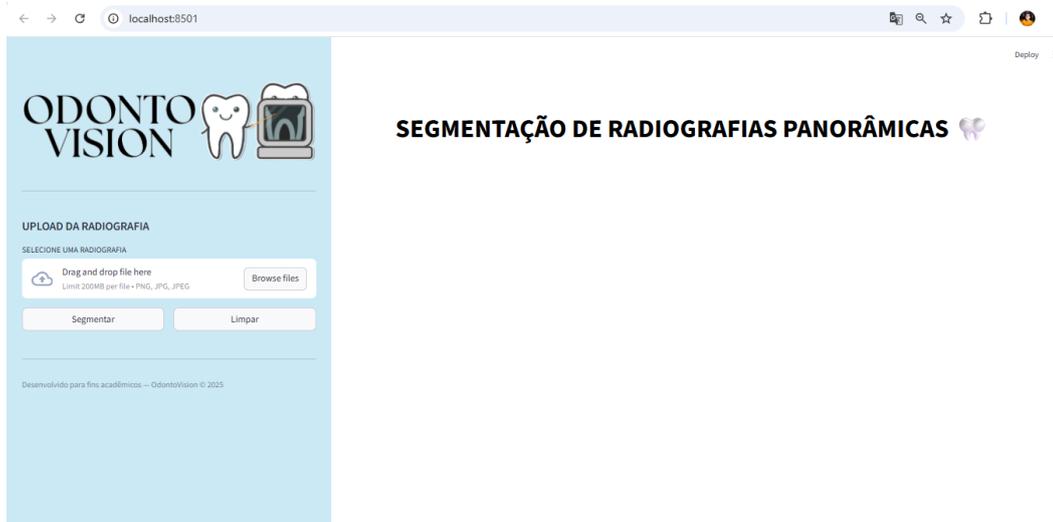
Conforme observado na Tabela 7, a rede Attention U-Net apresentou a menor média de tempo de inferência, aproximadamente 1527,3 ms, indicando que é a arquitetura mais rápida entre as avaliadas para a tarefa de segmentação. O desvio padrão, de cerca de 72,5 ms, evidencia que sua performance é estável e pouco variável entre as execuções. Esses resultados demonstram que a Attention U-Net combina velocidade e consistência, sendo a opção mais adequada para aplicações odontológicas, o que proporciona ao dentista maior agilidade no atendimento, otimiza o tempo de análise das radiografias e favorece decisões mais eficientes durante o diagnóstico. As demais redes apresentaram desempenho satisfatório, mas com tempos de inferência maiores e menor consistência, podendo ser aplicadas em contextos clínicos com maior poder computacional.

Dessa forma, a Attention U-Net foi selecionada como arquitetura final do sistema CAD, uma vez que apresentou desempenho consistente entre as métricas avaliadas, aliado ao menor tempo de inferência entre as redes analisadas. Embora seu melhor resultado tenha sido observado no *Panoramic Radiography Database*, nos demais conjuntos de dados a arquitetura manteve-se entre as melhores colocações, com pouca variação em relação às redes líderes. Esse comportamento evidencia sua robustez e capacidade de generalização, reforçando sua adequação para aplicação em um sistema CAD de apoio ao diagnóstico clínico.

## 5.8 Sistema CAD

A partir da seleção da Attention U-Net como a arquitetura que obteve os melhores resultados com base nos critérios estabelecidos, foi realizada a integração dela ao sistema CAD. A Figura 16 apresenta a interface inicial do sistema, projetado para uso em contextos clínicos, por meio do qual o profissional odontológico pode enviar uma radiografia para que a aplicação gere a máscara segmentada correspondente.

Figura 16 – Página inicial do Sistema CAD



Fonte: Elaborado pelo autor.

Conforme ilustrado na Figura 16, a interface inicial do sistema CAD foi estruturada de modo a favorecer uma interação direta e intuitiva com o usuário. À esquerda, encontra-se a área destinada ao envio da radiografia panorâmica, por meio da qual o dentista pode selecionar aquela que ele deseja ser processada. Esse componente permite tanto a seleção de arquivos a partir do explorador do sistema quanto o uso do recurso de arrastar e soltar, o que torna o procedimento mais ágil e acessível em ambientes clínicos. Além disso, a área de *upload* exibe explicitamente os formatos aceitos (PNG, JPG e JPEG) e o limite máximo do tamanho do arquivo, garantindo clareza quanto às restrições do sistema e evitando erros de submissão. Os botões “Segmentar” e “Limpar”, posicionados logo abaixo, possibilitam, respectivamente, iniciar o processamento da radiografia ou redefinir a área de entrada, permitindo que o usuário revise sua seleção antes da execução da segmentação. Na região central da interface está localizada a área destinada à visualização dos resultados, onde a segmentação da radiografia é exibida após o processamento.

Após o usuário clicar na área de *upload* e selecionar a radiografia panorâmica desejada, o sistema gera imediatamente uma miniatura da imagem no painel esquerdo, conforme ilustrado na Figura 17.

Figura 17 – Upload da radiografia



Fonte: Elaborado pelo autor.

Essa pré-visualização observada na Figura 17, tem a função de confirmar a escolha do arquivo antes do processamento, para permitir ao profissional verificar visualmente se a radiografia selecionada é de fato aquela que se pretende segmentar. Esse mecanismo reduz possíveis erros de entrada e assegura maior controle sobre o fluxo de interação. Em seguida, ao acionar o botão “Segmentar”, o sistema processa a radiografia carregada e exibe o resultado na área central da interface, conforme mostrado na Figura 18.

Figura 18 – Resultado da máscara segmentada



Fonte: Elaborado pelo autor.

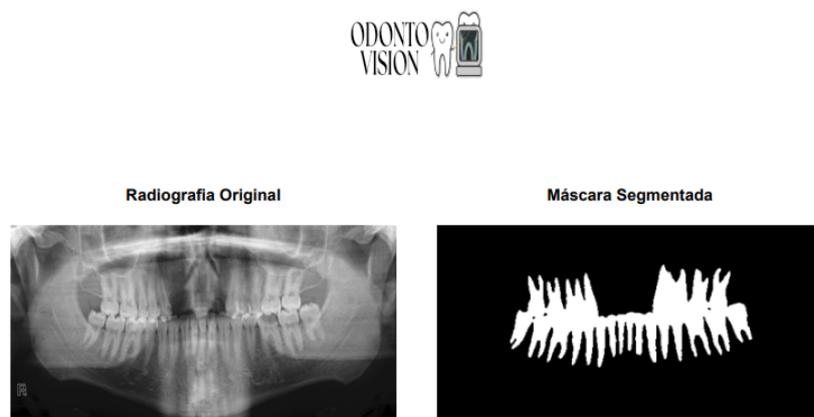
Nessa etapa descrita na Figura 18, são apresentados lado a lado a radiografia original e a máscara segmentada gerada pelo modelo, para que possa ser feita uma análise comparativa

imediate entre ambos os conteúdos. Essa disposição visual facilita a inspeção da estrutura dentária segmentada, pois possibilita verificar se o contorno extraído corresponde adequadamente às regiões de interesse presentes na radiografia de entrada. Além de garantir a coerência entre a imagem original e a máscara inferida pelo modelo, essa visualização também proporciona ao dentista a oportunidade de avaliar a qualidade da segmentação e identificar padrões anatômicos relevantes, como a disposição dentária, o alinhamento e a separação das unidades presentes na arcada. Tal recurso pode auxiliar no suporte ao diagnóstico, uma vez que a segmentação destaca as estruturas dentárias e elimina áreas não relevantes, tornando mais evidente o foco de análise.

Concomitantemente, o profissional dispõe ainda do botão “Limpar”, por meio do qual são removidos da interface tanto a radiografia quanto a máscara segmentada. Essa ação permite reiniciar o processo desde a seleção de uma nova imagem, sem a necessidade de recarregar ou reiniciar o sistema. Tal funcionalidade facilita a submissão de múltiplas radiografias em sequência, possibilitando que diversas segmentações sejam geradas em poucos segundos, o que favorece a análise de um maior volume de exames em um tempo reduzido.

No entanto, caso ele deseje realizar uma análise mais detalhada da segmentação, o sistema disponibiliza o botão “PDF”, localizado abaixo do resultado exibido. Ao acionar essa opção, é gerado automaticamente um relatório contendo a radiografia original e a máscara segmentada lado a lado, conforme ilustrado na Figura 19.

Figura 19 – PDF gerado com a radiografia e a máscara correspondente



Relatório gerado automaticamente — OdontoVision © 2025

Fonte: Elaborado pelo autor.

Esse recurso ilustrado na Figura 19 pode ser útil tanto para fins de documentação quanto para discussão clínica, uma vez que o relatório preserva as informações visuais essenciais

para a avaliação da segmentação. Ao disponibilizar a exportação em PDF, o sistema amplia seu escopo de uso, pois o dentista pode conservar o resultado, compartilhar o arquivo com outros profissionais ou anexá-lo a registros odontológicos. Além disso, a funcionalidade favorece o acompanhamento longitudinal de casos, ao possibilitar a documentação de diferentes etapas de avaliação e tratamento, garantindo consistência entre as imagens observadas e os dados preditos pelo modelo. Dessa forma, a ferramenta não apenas executa a segmentação automática, mas também se integra à prática clínica, ao facilitar a comunicação e o armazenamento das informações geradas.

### **5.9 Limitações do trabalho**

Apesar dos resultados promissores alcançados neste trabalho, algumas limitações devem ser consideradas. A utilização exclusiva de conjuntos de dados compostos por radiografias panorâmicas de pacientes adultos limita a capacidade de generalização do modelo para a segmentação de estruturas dentárias em populações pediátricas, considerando as diferenças morfológicas e fisiológicas associadas às distintas fases de desenvolvimento dentário. Adicionalmente, o sistema CAD proposto deve ser compreendido como uma ferramenta de apoio à decisão clínica, não substituindo a interpretação final do profissional. Em determinadas situações, o sistema pode não segmentar integralmente todas as estruturas dentárias com a precisão desejada, especialmente em casos de sobreposição estrutural ou variações anatômicas acentuadas. Outra limitação relevante refere-se ao fato de que o sistema não fornece, de forma automatizada, informações diagnósticas textuais ou laudos interpretativos associados às radiografias analisadas, restringindo-se à etapa de segmentação das estruturas dentárias.

Por fim, ressalta-se que aspectos de natureza operacional e computacional também devem ser considerados, uma vez que a disponibilização do sistema em um ambiente de produção robusto, estável e escalável, capaz de atender a um número elevado de usuários simultaneamente, impõe desafios relacionados à infraestrutura tecnológica, desempenho e manutenção, o que pode dificultar sua adoção imediata em larga escala.

## 6 CONCLUSÕES E TRABALHOS FUTUROS

Este estudo propôs o desenvolvimento de um sistema CAD para segmentação automática de dentes a partir de radiografias panorâmicas. Para essa tarefa, foram implementadas cinco variantes da arquitetura U-Net, as quais foram treinadas e avaliadas por meio de validação externa, além da análise estatística por meio do teste de Wilcoxon. Os resultados obtidos demonstraram o grande potencial do aprendizado profundo, especialmente das redes CNNs, para capturar padrões complexos em imagens médicas, para permitir a extração precisa de estruturas anatômicas detalhadas e a generalização para diferentes condições clínicas. Além disso, os resultados principais demonstram que as redes analisadas são capazes de identificar corretamente as regiões dentárias, tornando-as aptas para a tarefa de segmentação. A Attention U-Net, modelo com mecanismos de atenção e agregação multiescala, destacou-se com o melhor desempenho geral, combinando maior precisão, preservação da forma dentária, continuidade dos contornos e menor complexidade computacional, resultando em tempos de inferência mais rápidos para a geração das máscaras, sendo ela a escolhida para compor o sistema CAD.

Concomitantemente, observou-se que arquiteturas com maior complexidade computacional, embora possam alcançar bons resultados quantitativos e qualitativos, têm sua complexidade e custo computacional como fatores que devem ser considerados, especialmente em aplicações clínicas que exigem rapidez e estabilidade no processamento, uma vez que o tempo de inferência e a eficiência do modelo são determinantes para o uso prático. Nesse contexto, a análise estatística realizada por meio do teste de Wilcoxon mostrou que algumas redes mais complexas, como a U-Net++ e a U-Net 3+, apresentaram diferenças estatisticamente significativas em relação à U-Net clássica, evidenciando que alterações arquiteturais podem impactar de forma relevante a qualidade das segmentações; por outro lado, redes como a W-Net e a Attention U-Net apresentaram desempenho equivalente à U-Net em alguns casos, reforçando que maior complexidade nem sempre se traduz em ganhos significativos.

Por fim, o sistema CAD desenvolvido, sendo a principal contribuição deste trabalho, trouxe importantes avanços para a prática odontológica, pois permite a segmentação automática dos dentes de forma rápida e precisa, oferecendo visualização comparativa entre a imagem original e a máscara segmentada, além de gerar relatórios em PDF que podem ser utilizados para documentação e acompanhamento clínico. Além disso, a ferramenta auxilia no suporte ao diagnóstico, ao destacar de forma clara as estruturas dentárias e reduzir a necessidade de análise manual detalhada, promovendo maior eficiência, padronização e confiabilidade na interpretação

das radiografias, o que pode contribuir para decisões clínicas mais rápidas e fundamentadas.

Para trabalhos futuros, sugere-se explorar novas arquiteturas de segmentação para avaliar se abordagens recentes podem aprimorar ainda mais a precisão e a preservação estrutural das regiões dentárias, bem como integrar outras bases de dados maiores e mais diversificadas, incluindo radiografias pediátricas, de modo a ampliar sua capacidade de generalização para diferentes faixas etárias e variabilidades anatômicas. Além disso, recomenda-se a integração de funcionalidades adicionais ao CAD, como a identificação e numeração automática dos dentes, para permitir que cada unidade dentária seja reconhecida e rotulada corretamente, a detecção de anomalias ou patologias dentárias, como cáries, fraturas ou outros problemas clínicos, oferecendo suporte ao diagnóstico, e o planejamento de tratamentos odontológicos com base nas segmentações geradas, para fortalecer o caráter clínico da ferramenta.

## REFERÊNCIAS

- AJIT, A.; ACHARYA, K.; SAMANTA, A. A review of convolutional neural networks. In: **2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)**. Bhubaneswar, India: IEEE, 2020. p. 1–5.
- ALMALKI, S. A. *et al.* Denoised encoder-based residual U-Net for precise teeth image segmentation and damage prediction on panoramic radiographs. **Journal of Dentistry**, v. 137, p. 104651, 2023. Disponível em: <<https://doi.org/10.1016/j.jdent.2023.104651>>. Acesso em: 6 nov. 2025.
- ANDREWS, S.; HAMARNEH, G. Multi-region probabilistic dice similarity coefficient using the aitchison distance and bipartite graph matching. 2015. **arXiv**. Disponível em: <<https://arxiv.org/abs/1509.07244>>. Acesso em: 8 out. 2025.
- AZAD, R. *et al.* Medical image segmentation review: The success of u-net. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 46, n. 12, p. 10076–10095, 2024.
- BHAT, S.; BIRAJDAR, G. K.; PATIL, M. D. Tooth segmentation in panoramic dental radiographs using deep convolution neural network – insights from subjective analysis. **Discover Applied Sciences**, v. 7, p. 279, 2025.
- BOUALI, R.; MAHBOUB, O.; LAZAAR, M. Unleashing the potential of applied unet architectures and transfer learning in teeth segmentation on panoramic radiographs. **Intelligenza Artificiale**, v. 18, n. 2, p. 205–217, 2024.
- BRAHMI, W.; JDEY, I. Automatic tooth instance segmentation and identification from panoramic x-ray images using deep cnn. **Multimedia Tools and Applications**, v. 83, n. 18, p. 55565–55585, 2024.
- BREELAND, G.; AKTAR, A.; PATEL, B. C. Anatomy, head and neck, mandible. In: **StatPearls**. Treasure Island (FL): StatPearls Publishing, 2019. Disponível em: <<https://www.ncbi.nlm.nih.gov/books/NBK532292/>>. Acesso em: 3 jun. 2025.
- CAI, L.; GAO, J.; ZHAO, D. A review of the application of deep learning in medical image classification and segmentation. **Annals of Translational Medicine**, v. 8, n. 11, p. 713, 2020.
- CARMO, R. L. d. Mandíbula: Anatomia, estruturas e fraturas. **Kenhub**, 2023. Disponível em: <<https://www.kenhub.com/pt/library/anatomia/a-mandibula>>. Acesso em: 02 jun. 2025.
- CARMO, R. L. d.; CHAVES, C. **Dentes**. 2023. <<https://www.kenhub.com/pt/library/anatomia/anatomia-do-dente>>. Última revisão: 30 out. 2023. Acesso em: 5 jun. 2025.
- COPELAND, J. **Artificial intelligence: A philosophical introduction**. [S.l.]: John Wiley & Sons, 2015.
- DRAKE, R.; VOGL, W.; MITCHELL, A. W. M. **Gray: Anatomia Clínica para Estudantes**. 3. ed. [S.l.]: Elsevier, 2015. 885 p.
- ELHADDAD, M.; HAMAM, S. Ai-driven clinical decision support systems: an ongoing pursuit of potential. **Cureus**, v. 16, n. 4, 2024.
- FACELI, K. *et al.* **Inteligência artificial: uma abordagem de aprendizado de máquina**. Rio de Janeiro: LTC, 2011.

FAN, D.-P. *et al.* Enhanced-alignment measure for binary foreground map evaluation. **arXiv:1805.10421**, 2018.

FEHRENBACH, M. J.; POPOWICS, T. **Anatomia, Histologia e Embriologia dos Dentes e das Estruturas Orofaciais**. 5. ed. Rio de Janeiro: Elsevier, 2020. 216 p.

FERRER, L. Analysis and comparison of classification metrics. **arXiv:2209.05355**, 2022. Disponível em: <<https://arxiv.org/abs/2209.05355>>. Acesso em: 03 out. 2025.

HODSON, T. O. Root mean square error (rmse) or mean absolute error (mae): When to use them or not. **Geoscientific Model Development Discussions**, Göttingen, Germany, v. 2022, p. 1–10, 2022.

HOSSIN, M.; SULAIMAN, M. N. A review on evaluation metrics for data classification evaluations. **International journal of data mining & knowledge management process**, Academy & Industry Research Collaboration Center (AIRCC), v. 5, n. 2, p. 1, 2015.

HUANG, H. *et al.* Unet 3+: A full-scale connected unet for medical image segmentation. In: IEEE. **ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP)**. [S.l.], 2020. p. 1055–1059.

HYNDMAN, R. J.; KOEHLER, A. B. Another look at measures of forecast accuracy. **International journal of forecasting**, Elsevier, v. 22, n. 4, p. 679–688, 2006.

IWANAGA, J.; TUBBS, R. S. **The Mandible: An Atlas of Osteological and Radiological Anatomy**. Springfield, IL, USA: Charles C Thomas, 2022. 456 p. ISBN (e-book): 978-0-398-09379-2. ISBN 978-0-398-09378-5.

JOSHI, R. Segmentation of teeth in panoramic x-ray image using u-net algorithm. In: **2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT)**. [S.l.: s.n.], 2024. p. 1–6.

KRICHEN, M. Convolutional neural networks: a survey. **Computers**, Basel, v. 12, n. 8, p. 151, jul. 2023. Disponível em: <<https://doi.org/10.3390/computers12080151>>. Acesso em: 06 ago. 2025.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, Nature Publishing Group, London, v. 521, n. 7553, p. 436–444, 2015.

LI, Z. *et al.* A survey of convolutional neural networks: analysis, applications, and prospects. **IEEE Transactions on Neural Networks and Learning Systems**, v. 33, n. 12, p. 6999–7019, 2021.

LIASHCHYNSKYI, P. Grid search, random search, genetic algorithm: A big comparison for nas. **arXiv preprint ArXiv:1912.06059**, 2019. Disponível em: <<https://arxiv.org/abs/1912.06059>>. Acesso em: 25 set. 2025.

LIU, K. *et al.* Breast cancer classification based on fully-connected layer first convolutional neural networks. **IEEE Access**, v. 6, p. 23722–23732, 2018.

LIU, X. *et al.* A review of deep-learning-based medical image segmentation methods. **Sustainability**, v. 13, n. 3, p. 1224, 2021.

MINAEE, S. *et al.* Image segmentation using deep learning: A survey. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 44, n. 7, p. 3523–3542, 2022.

MORRIS, A. L.; TADI, P. **Anatomia, Cabeça e Pescoço, Dentes**. 2023. Disponível em: <<https://www.ncbi.nlm.nih.gov/books/NBK557543/>>. Acesso em: 20 ago. 2025.

MÜLLER, D.; SOTO-REY, I.; KRAMER, F. Towards a guideline for evaluation metrics in medical image segmentation. **BMC Research Notes**, v. 15, n. 1, p. 210, 2022.

NADER, R. *et al.* Automatic teeth segmentation on panoramic x-rays using deep neural networks. In: **2022 26th International Conference on Pattern Recognition (ICPR)**. [S.l.: s.n.], 2022. p. 4299–4305.

NTI, I. K. *et al.* Performance of machine learning algorithms with different k values in k-fold cross-validation. **International Journal of Information Technology and Computer Science**, MECS Publisher, v. 13, n. 6, p. 61–71, 2021.

OKTAY, O. *et al.* Attention u-net: Learning where to look for the pancreas. **arXiv preprint arXiv:1804.03999**, 2018. Disponível em: <<https://arxiv.org/abs/1804.03999>>. Acesso em: 05 jul. 2025.

O'SHEA, K.; NASH, R. An introduction to convolutional neural networks. **arXiv preprint arXiv:1511.08458**, 2015. Documento técnico; Introdução a Redes Neurais Convolucionais. Disponível em: <<https://arxiv.org/abs/1511.08458>>. Acesso em: 06 nov. 2025.

PERES, M. A. *et al.* Oral diseases: a global public health challenge. **The Lancet**, v. 394, n. 10194, p. 249–260, 2019.

RAINIO, O.; TEUHO, J.; KLÉN, R. Evaluation metrics and statistical tests for machine learning. **Scientific Reports**, Nature Publishing Group UK London, v. 14, n. 1, p. 6086, 2024.

RAO, V. S.; SATISH, M. A.; PRASAD, M. B. **Artificial intelligence: Principles and applications**. [S.l.]: Leilani Katie Publication, 2024.

RGUIBI, Z. *et al.* Cxai: Explaining convolutional neural networks for medical imaging diagnostic. **Electronics**, v. 11, n. 11, p. 1775, 2022.

ROBERT, J. Anatomia dentária: Compreendendo a estrutura e função dos dentes. **Journal of Interdisciplinary Medicine and Dental Science**, v. 6, n. 3, p. 32–35, 2023. Recebido: 01 mai. 2023; Publicado: 30 mai. 2023.

ROCHA, É. da S.; ENDO, P. T. A comparative study of deep learning models for dental segmentation in panoramic radiograph. **Applied Sciences**, MDPI, v. 12, n. 6, p. 3103, 2022.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. **International Conference on Medical image computing and computer-assisted intervention**. [S.l.], 2015. p. 234–241.

SÄHİN, A. E.; ULUTAŞ, H.; SÜZGEN, E. E. Automated segmentation of dental structures in panoramic radiographs using u-net 3+. In: **2024 8th International Artificial Intelligence and Data Processing Symposium (IDAP)**. [S.l.: s.n.], 2024. p. 1–6.

SHINDE, P. P.; SHAH, S. A review of machine learning and deep learning applications. In: **2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)**. [S.l.: s.n.], 2018. p. 1–6.

TATSAT, H.; PURI, S.; LOOKABAUGH, B. **Blueprints for Machine Learning and Data Science: Financial Use Cases from Trading to Fintech**. Sebastopol, CA: O'Reilly Media, 2021. ISBN 978-1-492-07940-3.

TERVEN, J. *et al.* A comprehensive survey of loss functions and metrics in deep learning. **Artificial Intelligence Review**, v. 58, n. 7, p. 195, 2025.

VOLLMER, D. *et al.* Experimental and finite element study of a human mandible. **Journal of Cranio-Maxillofacial Surgery**, v. 28, n. 4, p. 228–233, 2000.

WANG, R. *et al.* Medical image segmentation using deep learning: A survey. **IET Image Processing**, v. 16, n. 5, p. 1245–1267, 2022. Disponível em: <<https://doi.org/10.1049/ipr2.12419>>. Acesso em: 09 jun. 2025.

WORLD HEALTH ORGANIZATION. **Global strategy on oral health: draft for the 75th World Health Assembly**. Geneva: WORLD HEALTH ORGANIZATION, 2022. Disponível em: <<https://www.who.int/publications/i/item/9789240106031>>. Acesso em: 18 jul. 2025.

WORLD HEALTH ORGANIZATION. **Oral Health**. 2022. Disponível em: <[https://www.who.int/health-topics/oral-health#tab=tab\\_1](https://www.who.int/health-topics/oral-health#tab=tab_1)>. Acesso em: 18 jul. 2025.

WU, J. **Introduction to Convolutional Neural Networks**. Nanjing, China, 2017. LAMDA Group. Self-contained introductory material.

XIA, X.; KULIS, B. W-Net: A deep model for fully unsupervised image segmentation. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. Honolulu, HI, USA: IEEE, 2017. p. 4321–4333.

YAMASHITA, R. *et al.* Convolutional neural networks: an overview and application in radiology. **Insights into Imaging**, v. 9, n. 4, p. 611–629, 2018.

YASWANTH, J. S.; KUMAR, K. J.; SINGH, R. P. Automated tooth segmentation in x-ray images using attention integrated u-net++ model. In: **2025 3rd International Conference on Intelligent Systems, Advanced Computing and Communication (ISACC)**. [S.l.: s.n.], 2025. p. 1317–1322.

ZANNAH, R. *et al.* Semantic segmentation on panoramic dental x-ray images using u-net architectures. **IEEE Access**, v. 12, p. 44598–44612, 2024.

ZHANG, Y. *et al.* Children's dental panoramic radiographs dataset for caries segmentation and dental disease detection. **Scientific Data**, v. 10, n. 1, p. 380, 2023.

ZHONG, W.; REN, X.; ZHANG, H. Automatic x-ray teeth segmentation with grouped attention. **Scientific Reports**, Nature Publishing Group UK London, v. 15, n. 1, p. 64, 2025.

ZHOU, Z. *et al.* Unet++: A nested u-net architecture for medical image segmentation. In: SPRINGER. **International Workshop on Deep Learning in Medical Image Analysis**. [S.l.], 2018. p. 3–11.