

# Classificação de defeitos de deiscência em implantes dentários de titânio e zircônio

Antônio B. da Silva Netto, Cleber Zanchettin e Willian F. C. Oliveira

<sup>1</sup>Centro de Informática – Universidade Federal de Pernambuco (UFPE)  
Av. Prof. Moraes Rego, 1235 - Cidade Universitária, Recife - PE, 50670-901

{absn2, cz, wfco}@cin.ufpe.br

**Abstract.** *This article investigates approaches in the field of oral health in conjunction with Deep Learning. The goal is to develop methods for multilabel classification, identifying the presence or absence of periapical lesions or dehiscences caused by titanium and zirconia implants. The study employs an adapted EfficientNet-B0 model in conjunction with the loss function, achieving a 97% accuracy in overall label classification. Moreover, optimized selection and segmentation of cuts in "DICOM" files, sourced from cone-beam computed tomographies, were carried out. This approach contributes to the creation of a new dataset that will aid in the diagnosis made by healthcare professionals. This new dataset attained an average Structural Similarity Index Measure (SSIM) of 0.6 when compared with images selected by expert radiologists, signifying substantial similarity between the generated images. Additionally, analyses on the model's explainability and the treatment of files originating from CBCT (Cone Beam Computed Tomography), or Tomografia Computorizada de Feixe Cônico in Portuguese, are conducted.*

**Resumo.** *Este artigo investiga abordagens na área de saúde bucal em conjunto com Aprendizagem Profunda. O objetivo é desenvolver métodos para a classificação multilabel, identificando a presença ou ausência de lesões periapicais, ou deiscências ocasionadas por implantes de titânio e zircônio. O estudo utiliza um modelo adaptado do EfficientNet-B0 combinado com a função de perda, alcançando uma acurácia de 97% na classificação geral dos rótulos. Além disso, foi realizada a seleção e segmentação otimizada de recortes nos arquivos "DICOM", provenientes de tomografias computadorizadas de feixe cônico. Essa abordagem contribui para a criação de um novo conjunto de dados que auxiliará no diagnóstico realizado por profissionais da saúde. Esse novo conjunto de dados obteve uma média de "SSIM" (Structural Similarity Index Measure) de 0,6, uma vez comparada com imagens selecionadas por radiologistas especialistas, indicando uma ótima semelhança entre as imagens geradas. Adicionalmente, são conduzidas análises de explicabilidade do modelo e do tratamento dos arquivos oriundos do CBCT (Cone Beam Computed Tomography), ou Tomografia Computorizada de Feixe Cônico em português.*

## 1. Introdução

Perda de dentes é um problema bastante comum na nossa sociedade e pode acontecer por meio de algum trauma ou doenças associadas [Gaviria et al. 2014], dessa maneira o

implante dentário é uma técnica extremamente utilizada na odontologia para a substituição desses dentes perdidos, proporcionando tanto benefícios estéticos quanto mecânicos aos pacientes. Esses implantes são principalmente compostos de estruturas de titânio ou zircônio que são inseridas no osso maxilar ou mandibular, fornecendo uma base sólida para a fixação de próteses dentárias de forma permanente.

No entanto, a prática de inserção de implantes dentários pode estar sujeita a complicações ao paciente, como lesões periapicais (inflamações em torno da raiz), que se manifestam em forma de: defeitos de fenestração (abertura inadequada da cortical óssea) e deiscência (retração da gengiva em torno do implante), que podem levar à perda do implante e à infecções generalizadas [Cionca et al. 2017]. Portanto, é de profunda importância diagnosticar precocemente essas condições para evitar problemas futuros.

Assim podemos introduzir o uso da tomografia computadorizada de feixe cônico (Cone Beam Computed Tomography, CBCT), que é uma técnica de imagem bastante utilizada na odontologia para obter imagens tridimensionais detalhadas dos ossos maxilares e mandibulares, fornecendo informações precisas sobre a arcada dentária e os implantes dentários. Essa tecnologia permite uma visualização clara dos defeitos de fenestração e deiscência, possibilitando um diagnóstico mais preciso. Essas tomografias produzem vários recortes em três perspectivas de visualização: axial (recorte visto de cima), sagital (recorte visto de lado), coronal (recorte visto de frente), e cada recorte pode ser armazenado em arquivos do tipo "DICOM" (Digital Imaging and Communications in Medicine), que contêm todos os dados resultantes do processo.

No entanto, a interpretação manual das imagens de "CBCT" pode ser um procedimento demorado e suscetível a erros, devido à grande quantidade de informações, recortes e à complexidade do ruído gerado pelos artefatos na obtenção de exames. Isso requer uma análise não trivial de um radiologista experiente. É nesse contexto que algoritmos de Aprendizagem Profunda, especializados em análise de imagens [Calazans et al. 2022], e algoritmos de segmentação [Kurt Bayrakdar et al. 2021, Hegazy et al. 2019], são acionados desempenhando papéis importantes.

Os algoritmos de segmentação têm como objetivo identificar regiões de interesse em uma imagem, permitindo a delimitação precisa dos implantes e das estruturas adjacentes, como osso e gengiva, facilitando a análise e a detecção de lesões em volta do implante. Um exemplo de ferramenta disponível na indústria é o "Segment-Anything", desenvolvido pela Meta, que oferece a capacidade de segmentar qualquer objeto de interesse em imagens médicas [Kirillov et al. 2023a].

Além disso, as redes neurais convolucionais (CNNs), um tipo de algoritmo de Aprendizagem Profunda, têm se destacado no campo da análise de imagens devido à sua capacidade de aprender padrões complexos e extrair características relevantes automaticamente. As CNNs podem ser treinadas para aprender características distintas que distinguem os ossos dos artefatos metálicos, identificando e localizando os defeitos de fenestração e deiscência em implantes dentários com alta precisão [Minnema et al. 2019].

## **1.1. Objetivos**

O presente trabalho de graduação tem como objetivo desenvolver métodos para a classificação da presença ou não de lesões periapicais ocasionadas por implantes utilizando um modelo adaptado do EfficientNet-B0 disponibilizado pela biblioteca python "PyTorch",

juntamente com a implementação da seleção de melhores recortes nos arquivos "DICOM" obtidos por meio de tomografias computadorizadas de feixe cônico, ajudando na construção de uma base de treinamento e avaliação dos algoritmos.

Essa abordagem permitirá uma análise mais eficiente e precisa das imagens, otimizando o processo de detecção e classificação dos defeitos de deiscência utilizando algoritmos de segmentação como o "SAM" e de Aprendizagem Profunda especializados em imagens. Espera-se que esse trabalho possa contribuir para o avanço no diagnóstico precoce e no tratamento eficaz de complicações relacionadas a implantes dentários, proporcionando maior segurança, planejamento e qualidade nos procedimentos odontológicos.

## **2. Fundamentação Teórica**

Antes de adentrar no desenvolvimento da solução, este artigo vai se aprofundar nos principais conceitos abordados, explicando sobre os termos odontológicos e materiais associados, tecnologias utilizadas na realização das tomografias e um pouco da história da inteligência artificial e algoritmos de visão computacional.

### **2.1. CBCT**

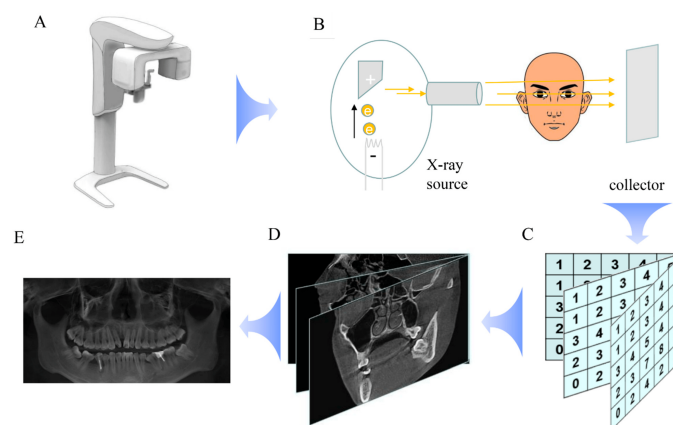
CBCT (a.k.a "Cone Beam Computed Tomography" ou Tomografia Computorizada de Feixe Cônico em português), é uma técnica de imagem bastante utilizada na odontologia para obter imagens tridimensionais precisas e de alta qualidade dos elementos ósseos no complexo maxilofacial [Antunes 2018] foi aprovada para o uso em Medicina Dentária nos EUA no ano de 2000 onde sucedeu de ser adotada rapidamente nos ambientes odontológicos, devido ao seu baixo custo e baixa exposição à radiação quando comparada a outras técnicas [Venkatesh and Venkatesh Elluru 2017]. Suas aplicações variam entre planejamento de implantes dentários, visualização de anomalias dentárias, avaliação da estrutura óssea maxilofacial, diagnóstico de cárie dentária (cavidades) entre outros [US - Food and Drug Administration 2020].

Essa técnica é aplicada em máquinas dos mais diversos estilos e que são conhecidas atualmente como "Máquinas CBCT". Esses engenhos usam um feixe em forma de cone e um detector de painel plano que rotaciona uma vez ao entorno do paciente entre 180-360 graus, abrangendo a área maxilofacial ou alguma área periférica de interesse, capturando em apenas uma rotação entre 180-1024 imagens cefalométricas (Imagens do Crânio Detalhada) [Venkatesh and Venkatesh Elluru 2017]. Essas imagens são reconstruídas utilizando uma versão modificada do algoritmo de "Feldkamp", gerando versões equivalentes distribuídas em 3 planos anatômicos: Axial, Coronal e Sagital; que são salvas no formato de "DICOM", para o uso de softwares de terceiros [Venkatesh and Venkatesh Elluru 2017], como é explicitado na Figura 1.

### **2.2. Planos Anatômicos**

Planos anatômicos, como mostrados na Figura 2, são planos hipotéticos que servem para dividir o corpo humano em seções de maneira que ajudam na localização, descrição e organização espacial de estruturas anatômicas, como órgãos e ossos. Nesses planos especificados estão incluídos: Sagital, Coronal e Axial [Medicine LibreTexts 2023].

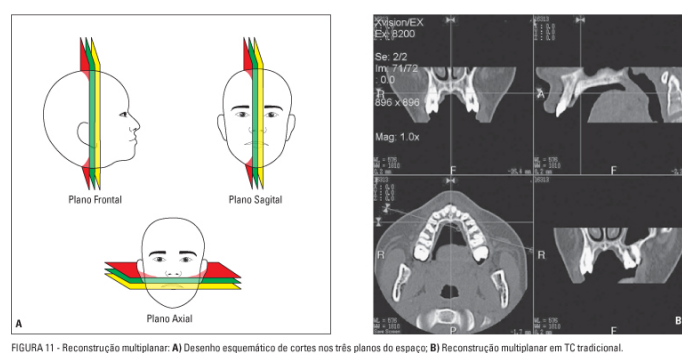
O plano sagital divide o corpo em duas metades: metade do lado esquerdo e metade do lado direito, permitindo observar movimentos e estruturas de perfil dividindo-o em metades exatamente iguais.



**Figura 1. Fluxo da obtenção de imagens CBCT. (A) demonstra a "Máquina CBCT"; (B) demonstra o processo de coleta dos raios-X provenientes da máquina ao crânio do paciente; (C) mostra a reconstrução das imagens em formato digital; (D) geração das imagens distribuídas nos planos anatômicos conhecidos; (E) valores são coletados e salvos em padrão DICOM [Fan et al. 2023]**

O plano coronal, também conhecido como plano frontal, divide o corpo em partes anterior e posterior. Ele é perpendicular ao plano sagital e passa lateralmente, dividindo o corpo em uma metade frontal (anterior) e uma metade posterior. Ao utilizar o plano coronal, é possível observar estruturas como ombros, braços e pernas, permitindo uma visão completa da simetria e assimetria dessas áreas.

O plano axial, também chamado de plano transversal, é perpendicular ao eixo longitudinal do corpo. Ele divide o corpo em partes superior e inferior. Ao utilizar o plano axial, podemos observar secções horizontais do corpo, permitindo visualizar estruturas internas em diferentes níveis. Isso é especialmente útil para a análise de órgãos e tecidos internos, como os cortes em uma tomografia computadorizada ou ressonância magnética.



**FIGURA 11 - Reconstrução multiplanar: (A) Desenho esquemático de cortes nos três planos do espaço; (B) Reconstrução multiplanar em TC tradicional.**

**Figura 2. (A) Representação do recorte dos planos anatômicos; (B) Exemplo de reconstrução dos planos anatômicos em tomografia computadorizada [Garib et al. 2007]**

## 2.3. DICOM

Datado de 1993, DICOM (a.k.a Digital Imaging and Communications in Medicine) é o padrão de comunicação mais utilizado e com suporte para imagens médicas atualmente [Kahn et al. 2007]. É também uma extensão de arquivo que é utilizada por padrão

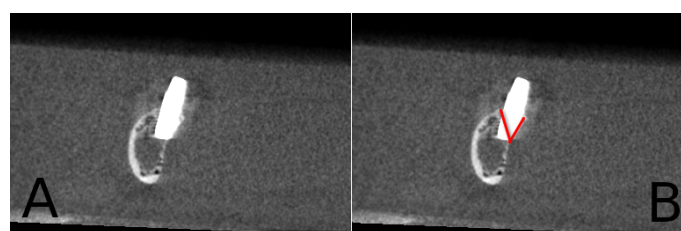
internacionalmente para armazenar, transmitir e processar imagens médicas, como radiografias, tomografias e ressonâncias magnéticas.

O DICOM permite a integração e compartilhamento eficiente de informações médicas entre diferentes sistemas e instituições de saúde. O DICOM é dividido em duas principais estruturas, o DICOMDIR arquivo esse que contém informações sobre o paciente, detalhes sobre a aquisição da imagem; e a segunda estrutura é uma série de imagens codificadas de forma sequencial, onde cada uma delas contém informações detalhadas sobre a anatomia daquela seção específica do corpo em perspectiva axial. Quando essas fatias axiais são organizadas na ordem correta e combinadas, elas formam a representação tridimensional completa da estrutura em estudo, seja um órgão, um osso ou outra área de interesse [Grauer et al. 2009]. Dessa forma, além de podermos ter uma representação tridimensional, poderia também haver a conversão para outros planos anatômicos como sagital e coronal, que são importantes para diagnosticar outros problemas.

## 2.4. Deiscência Óssea

A deiscência, como mostrado na Figura 3 refere-se à abertura ou separação de tecidos, normalmente ossos, que deveriam estar adjacentes ou unidos.

No contexto odontológico, uma deiscência em torno de um implante é caracterizada pela ausência de osso iniciando a partir da parte superior do implante dessa maneira forma-se uma abertura que expõe o dente e áreas circundantes, tornando-os vulneráveis a infecções bacterianas, podendo levar à perda do implante e dor associada [Cionca et al. 2017]. Essa patologia pode levar à recessão gengival (recoo da área gengival), expondo à doenças, interferindo na estética e na higiene bucal, causando inflamação e destruição do osso adjacente, resultando em uma lesão periapical que requer tratamento endodôntico ou cirúrgico para eliminar a infecção. Assim é de importância a correção da deiscência e para promover a regeneração saudável dos tecidos afetados [Chiapasco and Zaniboni 2009].



**Figura 3. (A) Representação de um mandíbula com presença de deiscência em área peri-implantar; (B) Destaque do implante ultrapassando a área limite superior de osso, indicando exposição do implante**

## 2.5. Aprendizagem de Máquina

Aprendizagem de Máquina pode ser definida como a área de estudo que capacita o Aprendizado de Máquina sem ser explicitamente programado [Shinde and Shah 2018]. Dessa maneira a máquina pode melhorar seu desempenho a partir de exemplos, ao receber uma grande quantidade de informação precisa e juntamente com os algoritmos necessários, o computador poderá aprender a partir dos dados, adquirindo novas habilidades, podendo fazer previsões ou decisões dependendo do escopo da tarefa.

Na área de Aprendizagem de Máquina existem três tipos de aprendizado amplamente utilizados: Supervisionado, Não Supervisionado e por Reforço [Ludermir 2021].

No Aprendizado Supervisionado, todos os exemplos são rotulados, ou seja têm uma resposta esperada informando qual é sua classe de pertencimento; esse método é bastante usado em problemas de classificação ou regressão, como por exemplo em um problema de distinguir imagens ou inferir se um paciente é diabético ou não dado um conjunto de informação sobre ele. Esses exemplos de entrada são sempre descritos como uma sequência de dados, em formato de vetor de atributos, onde para cada vetor (exemplo) há um rótulo associado. Nesse tipo de aprendizado, seu objetivo é construir rótulos para exemplos ainda não rotulados.

No Aprendizado Não Supervisionado, exemplos são fornecidos ao algoritmo sem rótulos. Ele agrupa os exemplos com base em similaridades de atributos, buscando formar clusters. Depois, é necessário analisar os clusters para atribuir significado a cada grupo no contexto do problema.

No Aprendizado por Reforço, o algoritmo recebe sinais de reforço (recompensa ou punição) em vez da resposta correta. Ele tenta associar a melhor ação baseado nas suas experiências passadas. É comum em jogos e foi utilizado no AlphaGo [Silver et al. 2016].

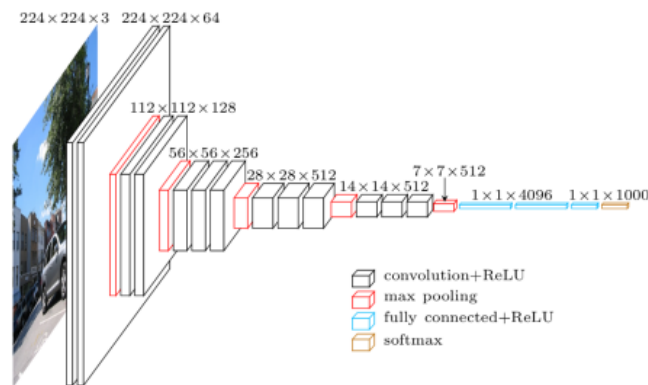
## **2.6. Aprendizagem Profunda**

Aprendizagem Profunda ou "Deep Learning" é um ramo da Aprendizagem de Máquina que tem a capacidade de compreender e processar informações complexas de forma autônoma. Esse método adota uma abordagem mais sofisticada se baseando em redes neurais artificiais profundas, que são estruturas inspiradas no funcionamento do cérebro humano. Essas redes consistem em camadas interconectadas de neurônios artificiais, cada uma delas processando e transformando os dados de entrada, permitindo que modelos computacionais compostos por várias camadas de processamento aprendam representações de dados com múltiplos níveis de abstração, porém com uma diferença, na Aprendizagem Profunda temos a abstração de inúmeras camadas ocultas de processamento, como na Figura 4 onde em cada uma dessas camadas temos uma outra grande quantidade de neurônios de processamento. Assim, esse método é o mais indicado para identificar padrões mais complexos e/ou com um grande custo de poder computacional (volume de dados) [Rusk 2015].

A habilidade de aprender representações de dados em diferentes níveis de abstração é uma das razões pelas quais a Aprendizagem Profunda se destaca em áreas como Visão Computacional, processamento de linguagem natural e muitas outras. Ela permite que as máquinas entendam os dados de maneira mais profunda e, assim, tomem decisões mais informadas e precisas em uma variedade de cenários.

## **2.7. Redes Neurais Convolucionais**

As Redes neurais convolucionais (ou CNN a.k.a "Convolutional Neural Network" do original), são um tipo de algoritmo de inteligência artificial "feed-forward" bastante empregado na área de Visão Computacional, amplamente usado no contexto de Aprendizagem Profunda pela maneira que consegue abstrair atributos complexos, e extrair características relevantes de acordo com a entrada [Alzubaidi et al. 2021].



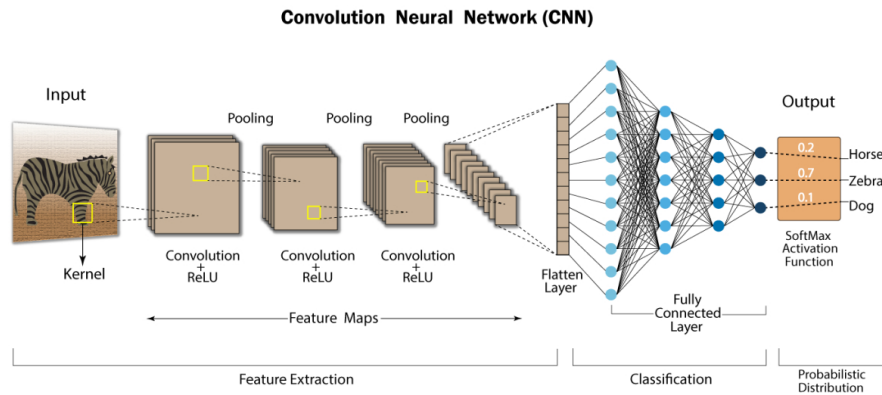
**Figura 4. Arquitetura do VGG-16, modelo de Aprendizagem Profunda**

A estrutura das CNNs foi inspirada nos neurônios do córtex visual do reino animal, de maneira semelhante a uma rede neural convencional, porém com uma diferença: detectar estímulos em regiões específicas. Isso permite atribuir três benefícios-chave às CNNs: representações equivalentes, ou seja, se a entrada for modificada, a saída do algoritmo também será modificada de maneira equivalente; interações esparsas, ou seja, com as constantes diminuições da dimensionalidade da entrada empregadas pelo algoritmo da CNN, já que uma imagem pode conter milhões ou milhares de pixels, o kernel consegue detectar informações significativas que abrangem dezenas ou centenas de pixels, isso implica que precisamos armazenar menos parâmetros, o que não apenas reduz o requisito de memória do modelo, mas também aprimora a eficiência estatística do modelo; por fim, há o compartilhamento de pesos, garantindo que um peso utilizado para gerar um output analisando um ponto qualquer é o mesmo utilizado em um outro na mesma interação, garantindo o ponto das representações equivalentes.

Ao contrário das redes totalmente conectadas (redes FC ou "Fully Connected Networks" do original), pesos compartilhados e conexões locais na CNN são usados para aproveitar plenamente as estruturas de dados de entrada 2D. Essa operação utiliza um número extremamente pequeno de parâmetros, o que simplifica o processo de treinamento e acelera a rede. É importante destacar que apenas pequenas regiões da entrada são captadas por essas células, em vez de dela toda (ou seja, essas células extraem localmente a correlação disponível na entrada, como filtros sobre a entrada).

A estrutura geral de uma CNN se baseia em três partes, como mostrado na Figura 5: camadas de convolução, camadas de "pooling" e as camadas totalmente conectadas (FC) [Yamashita et al. 2018]. As camadas de convolução são responsáveis pela realização de extração de características, utilizando operações lineares e não-lineares, como a convolução, filtros de kernel e a função de ativação associada. As camadas de "pooling" realizam uma redução da dimensionalidade em relação ao que é conhecido como sua entrada. Isso é feito para introduzir invariância a pequenos ruídos, diminuindo parâmetros subsequentes e extraíndo as características mais destacadas. E nas camadas totalmente conectadas é o responsável por receber o vetor das características extraídas pelas camadas passadas, e gerar uma saída desejada no modelo se for uma tarefa de classificação, é esperado que essa camada gere as probabilidades para cada classe. A camada totalmente conectada final geralmente possui o mesmo número de nós de saída que o número de classes. Cada

camada totalmente conectada é seguida por uma função não linear, como a ReLU.



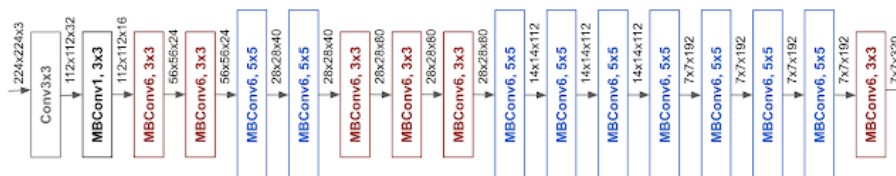
**Figura 5. Arquitetura geral de uma CNN, com as camadas de convolução, "pooling" e totalmente conectadas, que juntas realizam a saída de probabilidades da entrada pertencer a alguma das classes associadas [K E, Swapna 2022]**

## 2.8. Arquitetura EfficientNet

A EfficientNet [Tan and Le 2019], como mostrado na Figura 6, é uma arquitetura de rede neural convolucional (CNN) projetada para alcançar um equilíbrio eficiente entre desempenho e uso de recursos, como computação e memória. Ela alcança altos níveis de desempenho com uma quantidade menor de parâmetros e operações em comparação com muitas arquiteturas convolucionais tradicionais, tornando-o mais eficiente em termos de recursos computacionais, em comparação as variações da ResNet.

O EfficientNet é excelente para "transfer learning", onde um modelo pré-treinado em grandes conjuntos de dados pode ser ajustado para tarefas específicas com conjuntos de dados menores, economizando tempo e esforço no treinamento como é provado no seu artigo.

E isso só foi possível graças ao método de arquitetura "compound scaling", que envolve o dimensionamento proporcional de várias dimensões do modelo, incluindo largura, profundidade, tamanho da convolução e resolução de entrada. Esse escalonamento equilibrado permite otimizar o desempenho da rede neural convolucional para diferentes tarefas de visão computacional, garantindo que todos os componentes estejam ajustados de maneira eficiente e coordenada. Dessa forma, a EfficientNet alcança um equilíbrio entre eficiência e desempenho superior em comparação com abordagens convencionais de escalonamento, permitindo também que novas versões sejam criadas, a partir da arquitetura base da EfficientNet.



**Figura 6. A arquitetura base da EfficientNet-B0 [Tan and Le 2019]**

## 2.9. "Segment Anything Model" por "Meta AI"

"Segment Anything Model" ("SAM") [Kirillov et al. 2023b] é uma abordagem eficiente desenvolvida pela Meta AI para realizar segmentação semântica em imagens. Essa abordagem visa identificar e separar diferentes objetos ou regiões em uma imagem, atribuindo rótulos a cada pixel de acordo com a classe ou categoria à qual pertencem. A segmentação semântica é fundamental em tarefas de visão computacional, como detecção de objetos, navegação de veículos autônomos e análise de imagens médicas. Ela foi treinada em um conjunto de dados próprio para o SA-1B, que continha 11 milhões de imagens e mais de 1 bilhão de máscaras.

O "SAM" é um sistema de segmentação com generalização de "zero-shot", significa que no seu treinamento foi possível atingir um conceito geral sobre o que são objetos, generalizando objetos e imagens desconhecidos sem necessidade de treinamento adicional.

A arquitetura desse modelo é baseada em três frentes: um componente de codificador de imagens, um codificador de prompts e um rápido decodificador de máscara, como mostrado na Figura 7.

O Codificador de Imagens utiliza um "Vision Transformer" (ViT), ajustado minimamente para processar entradas de alta resolução. Essa etapa é crucial, pois fornece representações importantes da imagem que serão usadas durante todo o processo de segmentação.

O Codificador de Prompts considera dois conjuntos de instruções: esparsas (pontos, caixas, texto) e densas (máscaras). A inclusão de instruções é alcançada por meio de convoluções e soma elemento a elemento com a incorporação de imagem. Isso é fundamental para incorporar informações de instruções que servirão de guia para a geração das máscaras da segmentação.

O Decodificador de Máscara, é uma parte central do processo. Ele mapeia eficientemente as representações de imagem, representações de prompt e um token de saída em uma máscara. Esta parte utiliza uma modificação de um bloco decodificador Transformer, e um bloco auxiliador de previsão de máscara. Após executar os dois blocos, a representação da imagem é aumentada e um MLP ("Multilayer Perceptron") mapeia o token de saída para um classificador linear dinâmico, que calcula a probabilidade para cada plano da máscara em cada local da imagem. Isso é essencial para gerar uma segmentação precisa e detalhada, retornando todas as vetor de máscaras associadas, e dados detalhados sobre o mapeamento de cada segmentação feita (taxa de confiabilidade, área segmentada, pontos de seleção).

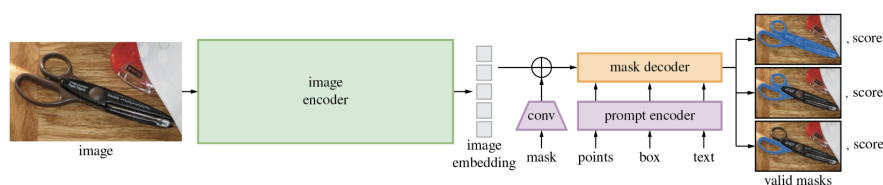


Figura 7. Visão geral do "SAM" [Kirillov et al. 2023b]

### 3. Material

Para a realização desse artigo, o material usado no estudo foi fornecido por parceria com especialistas do Centro de Odontologia da Universidade Federal de Pernambuco.

Foi utilizado três mandíbulas humanas secas distintas como mostrado na Figura 8, combinadas aleatoriamente com um dos quatro implantes dentários causadores de artefatos descritos: dois implantes de titânio (Straumann SLActive 3.3 mm x 8 mm, Institut Straumann AG, Basel, Suíça) e dois implantes de zircônia (PURE Ceramic 3.3 x 8 mm; Institut Straumann AG, Basel, Suíça). E com essa combinação também foram simulados efeitos de deiscência na porção cervical do implante.

Em seguida cada mandíbula foi submetida a exames de CBCT. O exame foi realizado nos aparelhos OP300, Picasso e iCAT. Cada mandíbula foi colocada dentro de um recipiente cilíndrico de plástico e preenchido com água para simulação de tecidos moles. As tomadas de aquisição foram feitas tanto sem auxílio ou com auxílio de uma ferramenta que reduz artefatos (AR ou "Artefact Reduction").

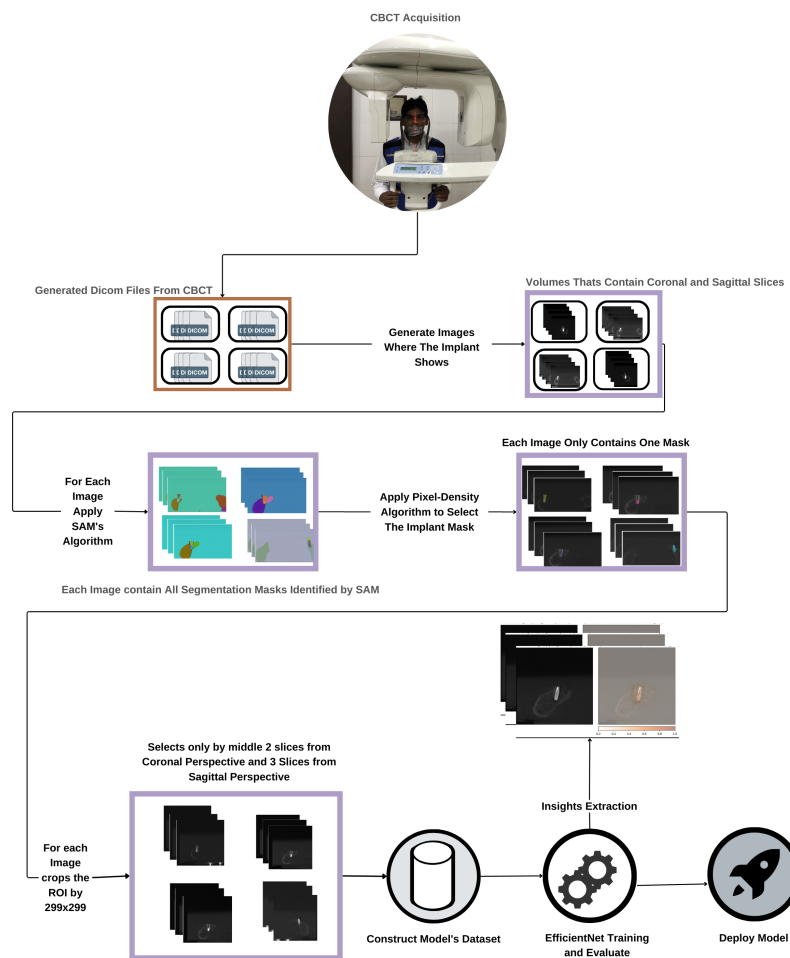
Ao final, foram geradas 440 CBCT exclusivas, cada uma contendo em média 450 arquivos "DICOM", totalizando um armazenamento de 150 Gigabytes. E para cada um dos 440 volumes foi também armazenado se cada um contém deiscência vestibular e/ou deiscência lingual.



**Figura 8. Visão do recipiente e mandíbula usada no conjunto de dados**

### 4. Metodologia

O primeiro passo para iniciar o desenvolvimento do estudo foi fazer uma revisão literária sobre implantes dentários, complicações relacionadas, CBCT, problemas de artefatos nas CBCT e técnicas de análise de imagens odontológicas utilizando Aprendizagem Profunda e outros métodos como segmentação. Nesse passo também foi possível identificar que existem estudos sobre segmentação utilizando CNNs em CBCT afetadas por artefatos metálicos [Minnema et al. 2019], que performam com eficiência, centralizando o implante da mandíbula e conseguindo extrair as máscaras que diferenciam os dentes de ossos. A partir dessa revisão, foi feito um fluxo de trabalho do desenvolvimento do projeto, descrito na Figura 9.



**Figura 9. Visão geral do fluxo do projeto**

Conforme descrito na Figura 9, o fluxo de trabalho construído vai se basear em quatro frentes.

A primeira foi construir as imagens que mostrem o implante aparecendo no conjunto de dados. Essa etapa vai ser concluída apenas com uso de técnicas de processamento de imagens como identificar os padrões luminosos gerados por interferência com os implantes e auxílio da biblioteca "pydicom" da linguagem de programação Python, e ao fim vão ser salvas essas imagens com respeito aos planos coronais e sagitais. Por fim, esses dados servirão de um conjunto de imagens prévios, podendo auxiliar no diagnóstico.

A segunda etapa compreende a segmentação semântica das imagens. Cada imagem é submetida a um processo de processamento pelo "SAM", resultando na geração de várias máscaras. Em sequência, cada uma destas máscaras é submetida a um procedimento de heurística de densidade de pixel, descrito no algoritmo da Figura 10 e da Figura 11. Durante este processo, um novo vetor de valores é criado, e um vetor de threshold com valores que variam entre 0 a 1 com passos entre  $10^{-4}$ , gerando  $10^5$  limitantes, no qual apenas os valores de pixel que excedem um limite mínimo predefinido são inseridos. Ao final deste procedimento, o vetor resultante deve atender a uma porcentagem mínima em relação à quantidade de pixels contida na máscara original. Se o vetor satisfizer os critérios mínimos, ele é selecionado como vetor de saída. Caso contrário, se nenhum vetor

ou dois ou mais vetores forem escolhidos, o algoritmo inicia um procedimento de busca binária para determinar novos limites. Se nenhum vetor for escolhido, isso significa que o recorte contém muitos ruídos, incluindo no próprio implante, dificultando a delimitação da resposta, logo o algoritmo vai diminuir tanto o limitante de valor de pixel e da porcentagem de pixels. E se for maior ou igual a dois vetores de resposta, isso significa que podemos aumentar os limitantes, já que a máscara do implante vai ser a mais luminosa, eliminando as outras máscaras no processo. Essa abordagem é fundamental porque o objetivo primordial do algoritmo é isolar especificamente a máscara com maior intensidade de preenchimento. Essa máscara, por sua vez, representa a região onde o implante está presente. Isso é particularmente crucial, já que, por definição, os implantes tendem a gerar uma intensidade luminosa elevada durante a aquisição CBCT. Essa intensidade pode, em alguns casos, ser confundida com outros artefatos ou ruídos resultantes de interferência metálica. Portanto, a busca binária por valores além de agilizar o processo custoso de operação nessas máscaras esparsas, delimita essa área segmentada que é de extrema importância para a precisão do processo.

A terceira etapa é acionada imediatamente após a seleção da máscara para cada recorte. Neste ponto, é efetuada uma escolha restrita à área do implante, cujas dimensões mínimas são de 299x299 pixels. Caso não seja viável realizar essa separação, a porção remanescente da imagem é preenchida com zeros.

Logo a seguir, para cada volume, a seleção começa a partir do ponto médio, com o objetivo de identificar um conjunto fixo de imagens coronais e sagitais que exibam a melhor qualidade de imagem do implante. Essas imagens são então combinadas para formar o conjunto de dados que irá auxiliar os profissionais de saúde em suas avaliações diagnósticas.

Finalmente, na última etapa, ocorrerá a criação do conjunto de dados, incluindo as divisões de treinamento, validação e teste, utilizando os recortes que foram previamente selecionados. Esses conjuntos serão empregados para treinar e avaliar o desempenho de um modelo modificado EfficientNet-B0, o qual já foi pré-treinado no conjunto de dados do ImageNet e está disponível por meio da biblioteca "Pytorch".

No processo de modificação do modelo importado, algumas alterações serão implementadas. Primeiramente, haverá uma adaptação na quantidade de canais do kernel da primeira camada de convolução, a fim de que esteja compatível com as características do nosso conjunto de dados criado. Além disso, na camada de classificação, a saída do modelo será ajustada para duas classes, a fim de se adequar a uma tarefa de classificação multilabel. Esses rótulos incluem "deiscência vestibular" e/ou "deiscência lingual". Para isso, será inserida uma camada final do tipo Sigmoid. Esta escolha é motivada pela capacidade da função sigmoide de produzir saídas na faixa de 0 a 1, o que representa as probabilidades de cada rótulo estar presente em uma dada entrada. Adicionalmente, a função sigmoid é apropriada para lidar com situações em que múltiplos rótulos podem estar presentes simultaneamente. Cada neurônio de saída associado a um rótulo específico calcula independentemente a probabilidade daquele rótulo estar presente, sem afetar as previsões dos outros rótulos.

Uma vez que o treinamento e a avaliação do modelo tenham sido realizados, a biblioteca "Captum", fornecida pela NVidia na linguagem Python, será empregada. Isso

---

**Algorithm 1: DENSPIX**

---

**Data:** **Threshold** Vetor de Inteiros de tamanho N, **Masks**  
Vetor de Máscaras, **value\_threshold** Int,  
**is\_value\_threshold** Bool

**Result:** Vetor de Máscara  $M'$  ou Erro

```
1  $left \leftarrow 0 \mid right \leftarrow N - 1 \mid M' \leftarrow []$ ;  
2 while  $left \leq right$  do  
3    $middle \leftarrow (right - left)/2 + left$ ;  
4    $masks \leftarrow []$ ;  
5   if  $is\_value\_threshold$  then  
6      $masks \leftarrow$   
        $DENSPIX(Threshold, Masks, middle, False)$ ;  
7   else  
8      $masks \leftarrow$   
        $GENMASK(Masks, value\_threshold, middle)$ ;  
9      $masks.size() > 0 ? M'.append(masks.size())$   
10  end  
11  if  $masks.size() == 1$  then  
12     $\text{return } masks$ ;  
13  else  
14    if  $masks.size() == 0$  then  
15       $right \leftarrow middle - 1$ ;  
16    else  
17       $left \leftarrow middle + 1$ ;  
18    end  
19  end  
20 end  
21  $\text{return } M'$ 
```

---

**Figura 10. Definição do Algoritmo de Densidade de Pixel**

será feito para analisar o comportamento do modelo por meio do Gradiente Integrado, permitindo identificar os pontos de interesse nos quais o modelo está convergindo quando recebe um recorte específico como entrada. Esse processo proporcionará insights valiosos sobre a interpretação e as decisões do modelo, contribuindo para uma análise mais profunda e confiável.

## 5. Resultados e Discussão

Nessa seção o trabalho vai discutir os resultados obtidos utilizando a metodologia explicada anteriormente.

### 5.1. Aquisição de Imagens

Para criar as imagens que destacam o implante, foi necessário primeiramente determinar a métrica adequada para delimitar os recortes. Optamos por usar o parâmetro "HU" (unidade

---

**Algorithm 2: GENMASK**

---

**Data:** *Masks* Vetor de Máscaras, *value\_threshold* Int,  
*percentage\_threshold* Int

**Result:** Vetor de Máscaras *Masks'*

```
1  $M' \leftarrow []$ ;  
2 for mask in Masks do  
3    $pixels \leftarrow mask.size()$ ;  
4    $good\_pixels \leftarrow 0$ ;  
5   for pixel in mask do  
6     if  $pixel \geq value\_threshold$  then  
7        $good\_pixels \leftarrow good\_pixels + 1$   
8     else  
9   end  
10  if  $good\_pixels/pixels \geq percentage\_threshold$  then  
11     $M'.append(mask)$   
12  else  
13 end  
14 return  $M'$ 
```

---

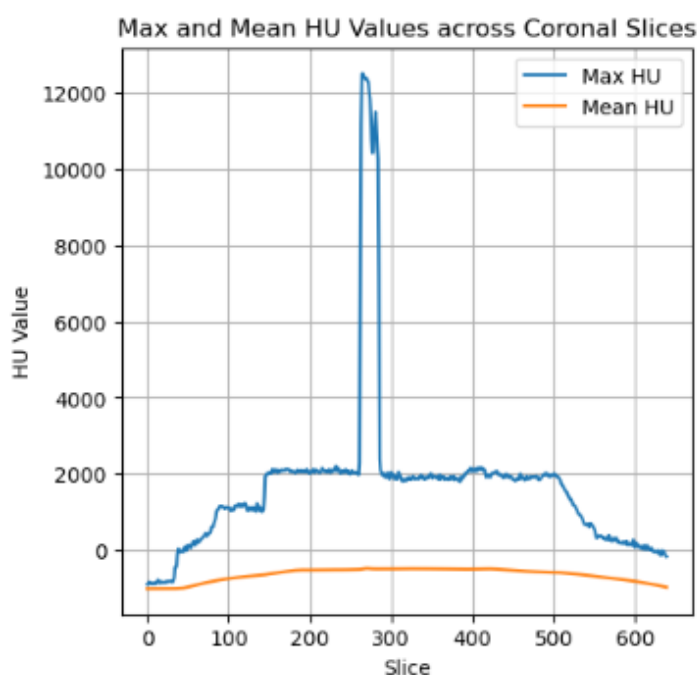
**Figura 11. Definição do Algoritmo de Geração de Máscaras utilizado na Figura 10**

de Hounsfield), que é uma unidade de medida fundamental para descrever a radiodensidade em imagens de CBCT. Essa medida é essencial para diferenciar áreas com densidades variadas nas imagens, conforme ilustrado na Figura 12. No gráfico, o pico mais elevado dos valores máximos de HU indica a presença de um elemento com alta radiodensidade [Katsumata et al. 2006].

Entretanto, esse exemplo nem sempre é absoluto. Artefatos e ruídos podem surgir com valores de HU extremamente elevados, e em recortes onde o implante não está presente verificar apenas o valor máximo não é suficiente. Adicionalmente, a Figura 12 também demonstra que confiar somente na soma dos valores não é viável, já que a média pouco varia. Isso se deve ao fato de que os implantes são consideravelmente pequenos em relação ao recorte, o que resulta em uma diferença mínima na soma total.

Portanto, a abordagem de considerar apenas o valor máximo ou a soma bruta é insuficiente para identificar de forma precisa a presença do implante. É necessário empregar uma função adaptável primeiramente é necessário gerar um ruído multiplicando os valores pelo seu quadrado, já que os valores do implante são altos, penalizando os valores menores e depois implicando uma normalização min-max dos valores para ao fim fazer uma soma, obtendo a Figura 13, que mostra três valores. O surgimento do primeiro pico e do último indicam a presença de vários artefatos e ruídos menores que somados chegam em um alto valor de HU.

Entretanto, é importante notar que esses artefatos e ruídos gerados raramente se aproximam da área total do implante, pois eles tendem a ser muito pequenos em comparação. Portanto, optamos por incorporar outra métrica para análise, baseada na distribuição dos valores de HU. Se esses valores forem maiores do que a metade do valor máximo de HU, eles são incluídos no cálculo da soma. Em seguida, é aplicada uma nova



**Figura 12. Exemplo de Valores máximos e médios do HU em um volume CBCT**

normalização aos dados.

Com base nas curvas resultantes desse processo, focamos naquelas que demonstram preenchimento igual ou superior a 50%, indicando a formação de um objeto retangular. Essas curvas são então consideradas para a seleção dos novos recortes, como podemos ver na Figura 14 o gráfico gerado e na Figura 15 o exemplo de imagens que foram selecionadas.

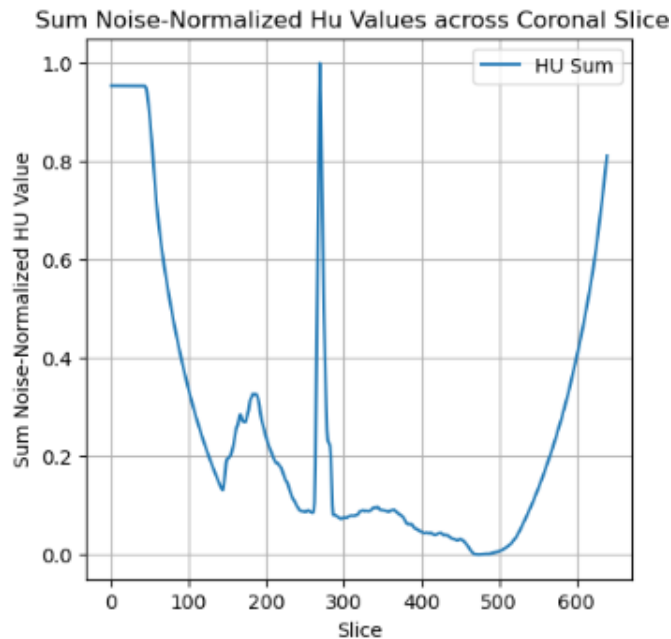
Essa abordagem mais abrangente considera tanto a magnitude dos valores de HU quanto a disposição dos mesmos, permitindo uma análise mais precisa e adaptável. Isso é particularmente relevante para lidar com a presença de artefatos, ruídos e pequenas variações nas imagens, proporcionando maior confiabilidade na seleção dos recortes que contêm informações relevantes sobre o implante.

Após a conclusão desse procedimento, a quantidade original de 150 Gigabytes foi significativamente reduzida, resultando em um novo conjunto de dados de apenas 300 Megabytes. Essa redução equivale a uma economia de **99%** de espaço de armazenamento. Como resultado, a média é de aproximadamente 8 recortes por volume CBCT, totalizando 3520 imagens no conjunto de dados resultante, o qual será utilizado em nossas análises subsequentes.

## 5.2. Segmentação Semântica

Durante esse estágio, foi usado o "SAM" para criar máscaras correspondentes a cada imagem contida no conjunto de dados gerado na etapa anterior. Entretanto, ao utilizá-lo, foi notado que o processo de segmentação também incluía ruídos e artefatos indesejados, como ilustrado na Figura 16. Adicionalmente, como é possível observar na Figura 17, não era possível discernir um padrão claro para selecionar a máscara ideal, uma vez que a área de pixels relacionada aos implantes apresentava considerável variabilidade.

Essa complexidade na segmentação destacou a necessidade de um método mais



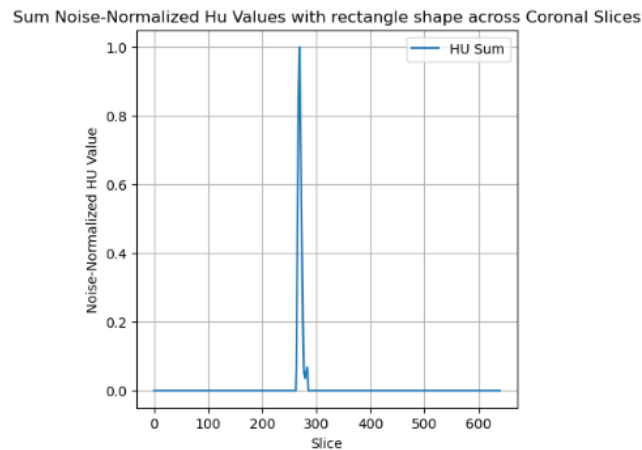
**Figura 13. Exemplo da Soma da Normalização do quadrado do máximo do HU Value em um volume CBCT**

refinado para isolar de forma precisa a área de interesse, ou seja, o implante. Portanto, foi crucial adotar abordagens mais sofisticadas que considerassem as características intrínsecas dos implantes e minimizassem a inclusão de artefatos e ruídos indesejados nas máscaras resultantes.

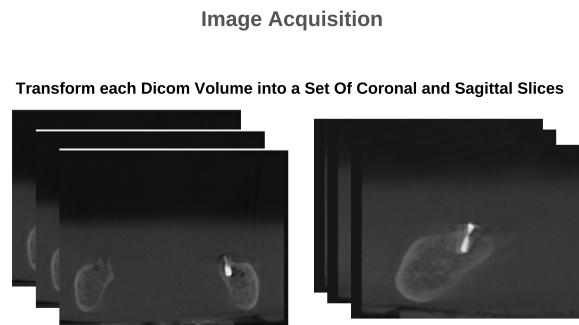
No entanto, baseando-nos na etapa anterior e na própria definição, os implantes feitos de zircônio e titânio possuem valores elevados de HU. Poderia-se então considerar a identificação dos segmentos com os maiores valores de HU. Entretanto, essa abordagem também se mostrou inadequada, conforme evidenciado na Figura 18. Nela, fica claro que artefatos ou ruídos, especialmente quando o implante está presente, tendem a ser maiores e mais frequentes. Porém, esses artefatos ainda exibem uma luminosidade que não segue uma relação linear. Como resultado, optamos por empregar a heurística de densidade de valores de pixels para delimitar todas as máscaras geradas.

No caso em que não fosse viável delimitar uma única máscara, essa situação acarretaria em um erro que interromperia o experimento. Contudo, obtivemos sucesso ao delimitar com precisão todas as máscaras em uma única entidade, o primeiro e segundo quartil dos valores escolhidos como limitantes finais, ficaram entre 50% e 50%, para o limitante de valor e de porcentagem, indicando que a máscara do implante na maioria dos casos é facilmente delimitada e se não fossem a aleatoriedade dos artefatos, nem precisaria da construção do algoritmo. Além disso, todas essas máscaras representavam pixels com valores altos de HU, o que indicava a presença do implante, conforme evidenciado na Figura 19.

Uma vez que a área de interesse foi identificada, procedemos à seleção de um recorte de 299x299 pixels, centrado no implante, gerando assim um novo conjunto de imagens. Em seguida, a partir desse novo conjunto, optamos por escolher 2 recortes



**Figura 14. Exemplo da Soma da Normalização do quadrado do máximo do HU Value com formato retangular em um volume CBCT. Desse gráfico foram selecionados os recortes 266 até o 271**



**Figura 15. Exemplo da criação de um conjunto de imagens sob perspectiva coronal e sagital com o implante em destaque**

de perspectiva coronal e 3 de perspectiva sagital. Essas imagens foram selecionadas de maneira estratégica, sendo escolhidas do centro em direção às extremidades, com o propósito de formar um conjunto de dados que será utilizado como apoio para diagnósticos e pegar os recortes onde o implante aparece mais nitidamente.

Esse processo, conforme ilustrado na Figura 19, permitiu com precisão a identificação da localização dos implantes de zircônio e titânio, mesmo em cenários complexos e variáveis, nos quais artefatos e ruídos estavam presentes.

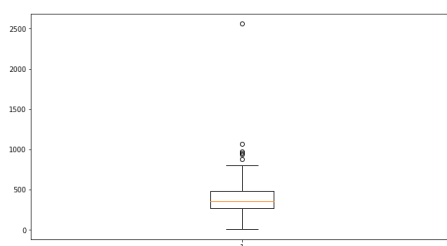
### 5.3. Classificação multilabel

A partir do último estágio, geramos um conjunto de dados final que consiste em 5 imagens por volume CBCT, totalizando 2220 imagens. Todas essas imagens foram padronizadas para o tamanho de 299x299 pixels. Esse conjunto foi projetado para servir como entrada para um modelo de Aprendizagem de Máquina.

Paralelamente, foi realizada uma pesquisa quantitativa e qualitativa com especialistas em odontologia da UFPE, utilizando uma amostra de 20 imagens previamente recortadas da região de interesse, geradas na etapa anterior, em comparação às 20 melhores imagens da mesma amostra gerada pela equipe de especialistas da UFPE. Na



**Figura 16. Exemplo da geração de máscaras pelo "SAM", gerando 7 máscaras distintas para um recorte do conjunto de dados**



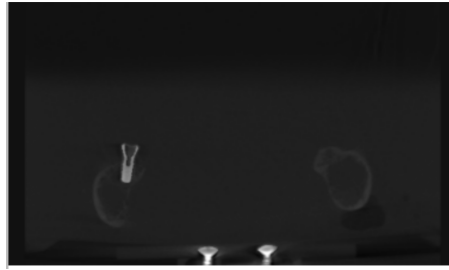
**Figura 17. Exemplo do gráfico de caixa sobre a área das máscaras do implante geradas pelo "SAM"**

pesquisa quantitativa, os especialistas foram questionados se as imagens geradas continham informações suficientes para auxiliar no diagnóstico de deiscência, e todas as 20 imagens geradas pelo algoritmo do artigo receberam a aprovação dos especialistas, indicando a escolha dos melhores recortes para construção do banco de dados. Na pesquisa qualitativa, comparamos as imagens geradas utilizando o Índice de Similaridade Estrutural (SSIM) [Nilsson and Akenine-Möller 2020], que é uma métrica que avalia a semelhança entre duas imagens, levando em consideração a estrutura, textura e informações de luminosidade das imagens para determinar o quão similar elas são, com valores entre -1 e 1, onde 1 indica que as imagens são idênticas e -1 indica que são totalmente diferentes.

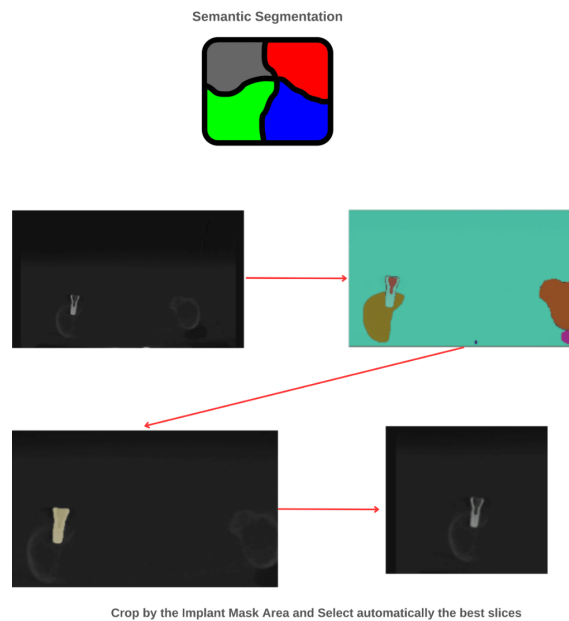
A utilização do SSIM na comparação dos recortes resultou em uma média de 0,6 para as imagens geradas. Isso indica fortes indícios de semelhança entre as imagens, validando a eficácia do algoritmo de geração de imagens, como indicando na Figura 20, que inclusive tem menos artefatos e mais nitidez no destaque da estrutura óssea.

Voltando ao conjunto de entrada gerado, todas as suas 5 imagens por volume foram concatenadas, resultando em um novo dado com dimensões de 5x299x299. Em outras palavras, as 5 imagens são tratadas como se fossem 5 canais distintos de informações. Como saída, temos um vetor de duas posições. Na primeira posição, o valor 0 indica a presença de deiscência vestibular, enquanto o valor 1 indica a ausência da mesma. Na segunda posição, o valor 0 indica a presença de deiscência lingual, enquanto o valor 1 indica a ausência.

Uma análise da distribuição das classes de saída do modelo também foi conduzida, como demonstrado na Figura 21. De maneira geral, não se observou um desequilíbrio acentuado entre as classes. Embora haja mais volumes contendo problemas de deiscência do que aqueles sem, essa disparidade não é tão pronunciada a ponto de requerer a



**Figura 18. Exemplo de recorte do novo conjunto de dados, que demonstra artefatos na área inferior central**



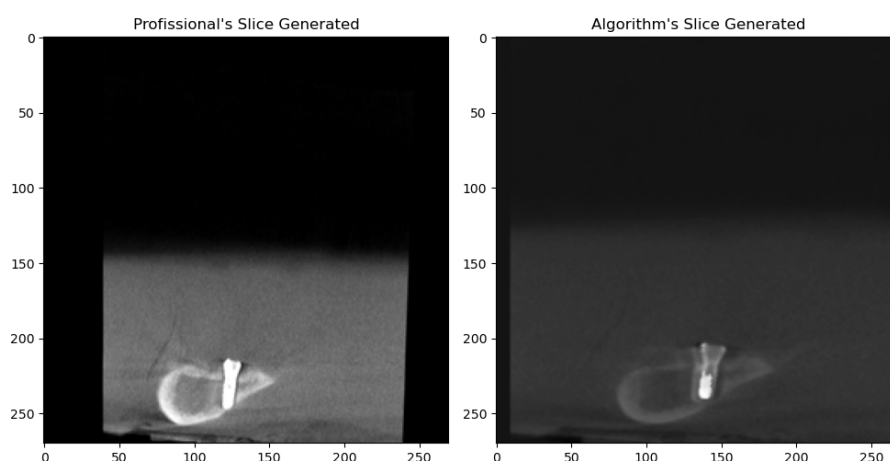
**Figura 19. Fluxo da Segmentação Semântica Gerada**

intervenção de algoritmos de balanceamento de dados. É notável que a distribuição se mantém similar para ambos os tipos de deiscência.

Adicionalmente, ao observar a combinação de rótulos, percebemos que o novo conjunto de dados apresenta a seguinte composição: cerca de 35% das imagens contêm ambos os tipos de deiscência, 40% exibem algum tipo de deiscência e 25% não apresentam qualquer tipo de deiscência. Em geral, a distribuição dos rótulos está bem equilibrada, o que contribui para a robustez e confiabilidade do conjunto de dados e, consequentemente, para o desempenho do modelo de Aprendizagem de Máquina.

Assim, esse conjunto de dados final, composto por informações de imagem concatenadas e vetores de saída com rótulos de presença ou ausência de deiscência vestibular e lingual, está pronto para ser utilizado como entrada em um modelo de Aprendizagem de Máquina.

Utilizando a biblioteca "Pytorch", construímos um DataLoader com as seguintes dimensões: 440x(5x299x299, 2). Na primeira dimensão da tupla, temos as dimensões de cada volume CBCT de entrada, enquanto a segunda dimensão representa os rótulos que serão inferidos durante o treinamento, como é visto na Figura 22.



**Figura 20. Comparação de um exemplo do melhor recorte gerado pelo profissional e pelo algoritmo**

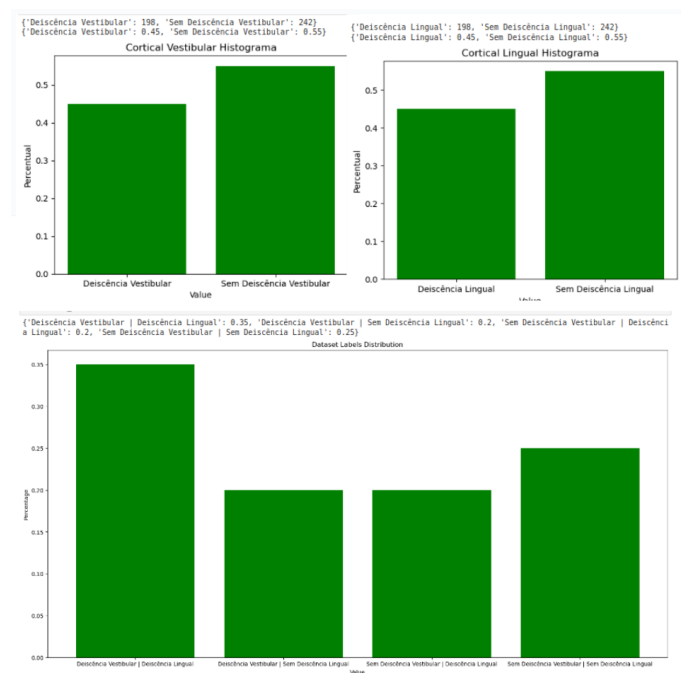
A partir disso, esse DataLoader foi dividido em três conjuntos distintos: treinamento, validação e teste. As proporções foram estabelecidas como 70%, 15% e 15%, respectivamente. A criação do conjunto de validação serve como uma medida preventiva para evitar overfitting do modelo, ou seja, para garantir que o modelo não se ajuste excessivamente ao conjunto de treinamento, mas seja capaz de generalizar o problema. Essa abordagem ajuda a melhorar a capacidade do modelo de lidar com novos dados e de obter resultados mais confiáveis.

Essa divisão foi realizada utilizando a técnica de divisão por rótulo. Isso significa que a separação dos conjuntos foi feita de forma aleatória e garantindo que as proporções dos rótulos fossem respeitadas. Essa abordagem foi empregada para reduzir as chances de criar inadvertidamente um conjunto de dados desbalanceado ou com vazamento de informações entre os diferentes conjuntos. Dessa forma, a integridade e a validade dos conjuntos de treinamento, validação e teste são preservadas.

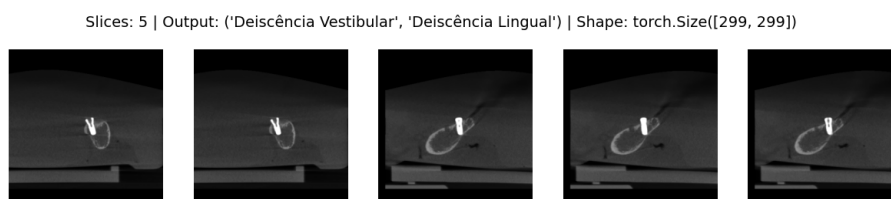
Em seguida, realizamos o treinamento do modelo adaptado do EfficientNet-B0, disponibilizado pelo "Pytorch", utilizando o conjunto de treinamento especificado. Não foi definido um número máximo de épocas; em vez disso, utilizou-se um critério de parada antecipada com uma paciência de 10 em uma versão inicial. A função de perda escolhida foi a "Binary Cross Entropy" para classificação multirrótulo, especificada pela fórmula:  $BCE_{multilabel}(y, p) = - \sum_{i=1}^N (y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i))$ , onde  $N$  é o número de classes,  $y_i$  representa o rótulo verdadeiro para a classe  $i$ , e  $p_i$  é a probabilidade prevista pelo modelo para a classe  $i$ .

A opção pela "BCE" é feita porque o objetivo pode representar simultaneamente múltiplas classes. A entropia cruzada binária é calculada separadamente para cada classe e, em seguida, os valores são somados para obter a perda total. Ao fim de cada época é salvo o melhor modelo em relação a taxa de erro do conjunto de validação, inclusive essa mesma taxa é a utilizada para diminuir ou reiniciar a paciência do modelo.

Ao final do treinamento, o modelo demonstrou uma acurácia de 100% para o rótulo de deiscência vestibular e 98% para o rótulo de deiscência lingual no conjunto de validação, o que resultou em um pequeno número de falsos positivos. Ao avaliarmos o desempenho



**Figura 21. Distribuições das classes de saída**

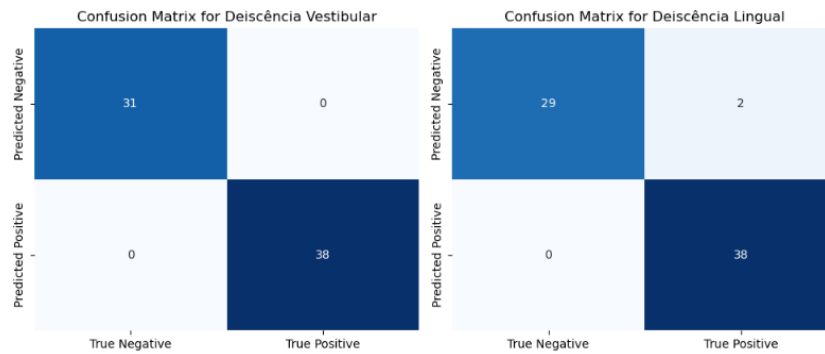


**Figura 22. Exemplo de entrada do DataLoader, onde as duas primeiras imagens são da perspectiva coronal e as outras três são da sagital, e cada uma das imagens representam um canal de imagem, e os rótulos indicam a presença dos dois tipos de deiscência**

do modelo no conjunto de teste, obtivemos uma **acurácia de 100% para o rótulo de deiscência vestibular e 97% para o rótulo de deiscência lingual**, com apenas dois casos de falso positivo. Esses resultados indicam que o modelo, com base no conjunto de dados fornecido, alcançou uma performance excelente, conforme ilustrado na Figura 23.

Para uma análise aprofundada dos resultados obtidos, foi realizada uma projeção da distribuição geral do conjunto de dados usando o UMAP ("Uniform Manifold Approximation and Projection"). Essa projeção permitiu mapear os dados em um espaço de dimensão inferior, criando um gráfico bidimensional que facilita a compreensão e interpretação da distribuição dos dados. O objetivo era identificar agrupamentos, padrões ou anomalias nos dados.

A Figura 24 ilustra a distribuição gerada pelo UMAP para o DataLoader do projeto. Vale lembrar que o conjunto de dados original em formato "DICOM" foi gerado a partir de três mandíbulas diferentes, nas quais foram aleatoriamente dispostos diferentes tipos de implantes, bem como a presença ou ausência de lesões.



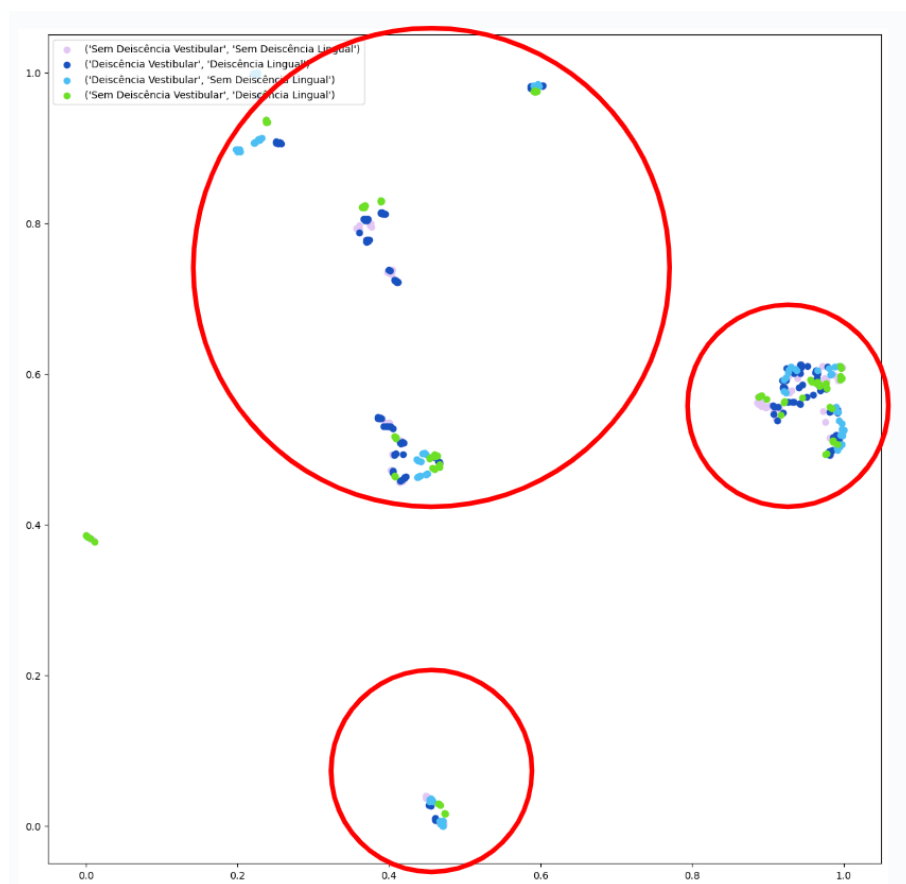
**Figura 23. Matrix de confusão gerada na performance sobre o conjunto de teste**

No entanto, é notável que o modelo foi capaz de generalizar exatamente a estrutura óssea específica dessas três mandíbulas distintas. Isso é evidenciado pelos três clusters bem definidos no gráfico UMAP. O modelo conseguiu delinear padrões de deiscência vestibular e lingual para cada uma das mandíbulas, identificando características específicas que contribuem ou não para a saída do modelo.

Essa capacidade de generalização e discernimento entre diferentes padrões é crucial para a eficácia do modelo na detecção e classificação das condições de interesse. Os resultados do UMAP fornecem insights valiosos sobre como o modelo está aprendendo e interpretando as características anatômicas relevantes para a identificação das deiscências. Porém como falta mais variabilidade de mandíbulas, não fica claro se o modelo apenas conseguiu identificar as três mandíbulas ou se poderia generalizar para uma quantidade maior de mandíbulas.

Aprofundando na explicabilidade do modelo, incorporamos a biblioteca "Captum" e utilizamos a função Gradiente Integrado, a qual tem como propósito entender como o modelo toma decisões com base nas características das entradas. Essa técnica foi desenvolvida para oferecer insights acerca da importância relativa das diferentes partes de uma imagem (ou entrada) na decisão final do modelo. Sua ideia fundamental consiste em calcular o gradiente da saída do modelo em relação à entrada (ou seja, os gradientes parciais em relação a cada pixel da imagem) e, posteriormente, integrar esses gradientes ao longo de um percurso suave entre uma linha de base (imagem inicial, geralmente preta ou branca) e a imagem de entrada real. Tal procedimento é executado para cada pixel, resultando em um conjunto de valores que representam a contribuição de cada pixel para a previsão do modelo. Essa técnica é formalizada pela seguinte fórmula:  $IntegratedGradients(\mathbf{x}, \mathbf{x}') = (\mathbf{x} - \mathbf{x}') \times \int_{\alpha=0}^1 \frac{\partial F(\mathbf{x}' + \alpha \times (\mathbf{x} - \mathbf{x}'))}{\partial \mathbf{x}}, d\alpha$  [Sundararajan et al. 2017].

Ao examinarmos os resultados gerados pelo Gradiente Integrado da biblioteca "Captum", juntamente com a avaliação de uma especialista em odontologia, identificamos um aspecto importante. Além de analisar a parte superior do implante, que é crucial para identificar exposição do implante conforme a definição de deiscência, o modelo também está detectando outros orifícios que foram criados na estrutura óssea da mandíbula, e os orifícios que foram também levados em consideração são da mesma natureza do rótulo previsto, como visto na Figura 25. Isso ocorre porque a mesma mandíbula foi utilizada em múltiplas instâncias. Essa descoberta é um indício promissor de que o modelo está aprendendo a generalizar as características da lesão. Entretanto, devido à limitada



**Figura 24. UMAP do conjunto de dados gerado após a seleção da área de interesse**

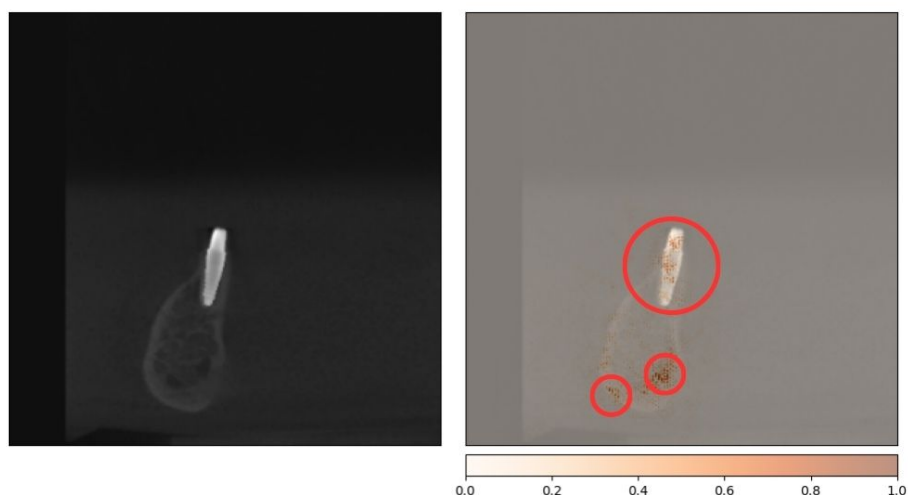
variabilidade no conjunto de dados, essa constatação permanece como uma oportunidade futura e uma hipótese interessante para investigações subsequentes no projeto.

Até o momento, o modelo está demonstrando um desempenho excelente na classificação multilabel do problema. Além disso, o desenvolvimento do novo conjunto de dados para auxiliar no diagnóstico de profissionais de saúde tem se mostrado uma contribuição valiosa. Essa combinação de resultados positivos e a exploração de possibilidades futuras destacam o potencial e a importância desse projeto no contexto da saúde bucal.

## 6. Considerações finais

Este trabalho de graduação representa um esforço significativo e promissor no campo da saúde bucal e diagnóstico de deiscências mandibulares. Ao longo de sua execução, uma série de etapas foi realizada, desde a transformação dos dados no formato "DICOM" para imagens até a adaptação e treinamento do modelo EfficientNet-B0 para a classificação multilabel das deiscências vestibular e lingual. A abordagem inovadora de utilizar o Gradiente Integrado da biblioteca "Captum" para explicar as decisões do modelo acrescenta uma camada de transparência e interpretabilidade ao processo.

A metodologia empregada para seleção e preparação dos dados, incluindo a segmentação semântica, a heurística de densidade de pixel e a geração de conjuntos de treinamento, validação e teste, gerando uma grande diminuição no total de armazenamento



**Figura 25. Aplicação do Gradiente Integrado em uma imagem do conjunto de teste, importante observar os pontos em destaque indicando que o modelo observou para a parte superior do implante e para orifícios da estrutura mandibular**

das tomografias computadorizadas, auxiliando também na transferência desses arquivos, demonstra um compromisso com a qualidade dos resultados e a validade das análises. A projeção via UMAP revela a Aprendizagem Profunda do modelo sobre as características específicas das mandíbulas e das deiscências, apontando para sua capacidade de generalização, mesmo em um conjunto de dados com limitada variação.

As taxas de acurácia alcançadas na validação e teste validam a eficácia do modelo na classificação multilabel. A identificação de padrões além do esperado pelo Gradiente Integrado reforça a importância de uma abordagem constante para a análise e interpretação dos resultados.

## **7. Trabalhos Futuros**

Este projeto aponta para inúmeras oportunidades de investigação e aprimoramento, com o objetivo de expandir ainda mais o impacto e a eficácia da abordagem proposta. Algumas áreas sugeridas para desenvolvimentos futuros incluem:

**Expansão do Conjunto de Dados:** Embora o modelo tenha demonstrado excelentes resultados com o conjunto de dados atual, obtendo um valor de acurácia geral de 97% sob o conjunto de teste, como o atual trabalho apenas usa três mandíbulas distintas, nas quais as três foram reutilizadas durante a construção do primeiro conjunto de dados obtido em parceria com os especialistas de Odontologia da UFPE, se levanta uma hipótese na qual o modelo proposto não performe bem em outros tipos de mandíbulas, logo precisa-se de uma maior diversidade das mesmas, e também de mais tipos de implantes, assim ampliando a capacidade de generalização do modelo e validar ou não a hipótese levantada. A expansão do conjunto de dados pode reforçar a robustez do modelo e sua aplicabilidade em situações mais complexas do mundo real, permitindo a utilização deste modelo como auxílio para diversos profissionais da área de saúde.

**Aplicação em Outras Áreas Médicas:** Agora que se constatou que o projeto atual consegue identificar com precisão uma área importante da saúde bucal, a metodologia

desenvolvida pode ser adaptada e aplicada em outros domínios da medicina e radiologia. A detecção automática de diferentes patologias, como tumores, cistos ou fraturas, poderia ser investigada, ampliando a utilidade da abordagem para além das deiscências.

Desenvolvimento de Pipeline completo para o profissional: Criar uma interface de usuário na qual o profissional pode iniciar um fluxo apenas fazendo o upload dos arquivos "DICOM" e receber as imagens que aparecem o implante, com a explicabilidade do modelo em cada uma delas, aumentando sua acessibilidade e utilidade.

Este projeto, com seus resultados, estabelece uma base sólida para futuros avanços. A convergência entre a tecnologia de Aprendizagem Profunda e o conhecimento médico demonstra ser uma ferramenta poderosa para aprimorar o diagnóstico médico e, potencialmente, transformar a abordagem a diversas condições médicas. Com um olhar voltado para a colaboração contínua e a inovação, este projeto oferece perspectivas empolgantes para o futuro da medicina e radiologia.

## References

- Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., and Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, 8(1).
- Antunes, F. N. R. R. (2018). Uso de cbct (tomografia computadorizada de feixe cônico) em endodontia.
- Calazans, M. A. A., Ferreira, F. A. B. S., Alcoforado, M. d. L. M. G., Santos, A. d., Pontual, A. d. A., and Madeiro, F. (2022). Automatic classification system for periapical lesions in cone-beam computed tomography. *Sensors*, 22(17).
- Chiapasco, M. and Zaniboni, M. (2009). Clinical outcomes of GBR procedures to correct peri-implant dehiscences and fenestrations: a systematic review. *Clinical Oral Implants Research*, 20:113–123.
- Cionca, N., Hashim, D., and Mombelli, A. (2017). Zirconia dental implants: where are we now, and where are we heading? *Periodontol. 2000*, 73(1):241–258.
- Fan, W., Zhang, J., Wang, N., Li, J., and Hu, L. (2023). The application of deep learning on cbct in dentistry. *Diagnostics*, 13(12).
- Garib, D. G., Jr., R. R., Raymundo, M. V., Raymundo, D. V., and Ferreira, S. N. (2007). Tomografia computadorizada de feixe cônico (cone beam): entendendo este novo método de diagnóstico por imagem com promissora aplicabilidade na ortodontia. *Revista Dental Press de Ortodontia e Ortopedia Facial*, 12(2):139–156.
- Gaviria, L., Salcido, J. P., Guda, T., and Ong, J. L. (2014). Current trends in dental implants. *Journal of the Korean Association of Oral and Maxillofacial Surgeons*, 40(2):50.
- Grauer, D., Cevdanes, L. S., and Proffit, W. R. (2009). Working with dicom craniofacial images. *American Journal of Orthodontics and Dentofacial Orthopedics*, 136(3):460–470.
- Hegazy, M. A. A., Cho, M. H., Cho, M. H., and Lee, S. Y. (2019). U-net based metal segmentation on projection domain for metal artifact reduction in dental ct. *Biomedical Engineering Letters*, 9(3):375–385.

- K E, Swapna (2022). Convolutional neural network: Deep learning. <https://developersbreach.com/convolution-neural-network-deep-learning/>, Last accessed on 2023-08-10.
- Kahn, C. E., Carrino, J. A., Flynn, M. J., Peck, D. J., and Horii, S. C. (2007). DICOM and radiology: Past, present, and future. *Journal of the American College of Radiology*, 4(9):652–657.
- Katsumata, A., Hirukawa, A., Noujeim, M., Okumura, S., Naitoh, M., Fujishita, M., Arij, E., and Langlais, R. P. (2006). Image artifact in dental cone-beam CT. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology*, 101(5):652–657.
- Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., and Girshick, R. (2023a). Segment anything.
- Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., and Girshick, R. (2023b). Segment anything.
- Kurt Bayrakdar, S., Orhan, K., Bayrakdar, I. S., Bilgir, E., Ezhov, M., Gusarev, M., and Shumilov, E. (2021). A deep learning approach for dental implant planning in cone-beam computed tomography images. *BMC Medical Imaging*, 21(1):86.
- Ludermir, T. B. (2021). Inteligência artificial e aprendizado de máquina: estado atual e tendências. *Estudos Avançados*, 35(101):85–94.
- Medicine LibreTexts (2023). 1.4d: Body planes and sections. [https://med.libretexts.org/Bookshelves/Anatomy\\_and\\_Physiology/Anatomy\\_and\\_Physiology\\_\(Boundless\)/1%3A\\_Introduction\\_to\\_Anatomy\\_and\\_Physiology/1.4%3A\\_Mapping\\_the\\_Body/1.4D%3A\\_Body\\_Planes\\_and\\_Sections](https://med.libretexts.org/Bookshelves/Anatomy_and_Physiology/Anatomy_and_Physiology_(Boundless)/1%3A_Introduction_to_Anatomy_and_Physiology/1.4%3A_Mapping_the_Body/1.4D%3A_Body_Planes_and_Sections), Last accessed on 2023-08-10.
- Minnema, J., van Eijnatten, M., Hendriksen, A. A., Liberton, N., Pelt, D. M., Batenburg, K. J., Forouzanfar, T., and Wolff, J. (2019). Segmentation of dental cone-beam ct scans affected by metal artifacts using a mixed-scale dense convolutional neural network. *Medical physics*, 46(11):5027–5035.
- Nilsson, J. and Akenine-Möller, T. (2020). Understanding ssim.
- Rusk, N. (2015). Deep learning. *Nature Methods*, 13(1):35–35.
- Shinde, P. P. and Shah, S. (2018). A review of machine learning and deep learning applications. In *2018 Fourth international conference on computing communication control and automation (ICCUBEA)*, pages 1–6. IEEE.
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., and Hassabis, D. (2016). Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489.
- Sundararajan, M., Taly, A., and Yan, Q. (2017). Axiomatic attribution for deep networks.
- Tan, M. and Le, Q. V. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks.

- US - Food and Drug Administration (2020). Dental cone-beam computed tomography. <https://www.fda.gov/radiation-emitting-products/medical-x-ray-imaging/dental-cone-beam-computed-tomography>, Last accessed on 2023-08-05.
- Venkatesh, E. and Venkatesh Elluru, S. (2017). Cone beam computed tomography: Basics and applications in dentistry. *J. Istanbul Univ. Fac. Dent.*, 51(0).
- Yamashita, R., Nishio, M., Do, R. K. G., and Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, 9(4):611–629.