ANÁLISE COMPARATIVA DE REDES NEURAIS CONVOLUCIONAIS PARA REDUÇÃO DE RUÍDO EM IMAGENS DE TOMOGRAFIA COMPUTADORIZADA

Gustavo Miguel da Silva[1], Allan Farias Favaro[2]

Resumo: Reduzir a dose de radiação em exames de tomografia computadorizada é uma necessidade clínica. O desafio é que imagens geradas com baixa dose têm qualidade inferior, com ruídos e artefatos que comprometem o diagnóstico. Aumentar a dose melhora a qualidade das imagens, mas aumenta o risco para os pacientes. Por anos, esse foi um *trade-off* sem uma solução plenamente satisfatória. Recentemente, modelos de aprendizado profundo começaram a demonstrar que é possível melhorar essas imagens sem alterar a dose de radiação. A abordagem consiste em treinar redes neurais para aprender a diferença entre uma imagem ruidosa e uma de boa qualidade, e aplicar essa correção após a aquisição da imagem. Vários tipos de arquitetura foram propostos para essa finalidade, e compará-los de forma controlada ainda é uma necessidade apontada pela literatura. Este trabalho realiza essa comparação. Foram avaliadas três arquiteturas de redes neurais convolucionais, RED-CNN, DnCNN e WGAN-VGG, todas treinadas com o mesmo *dataset*, o *LDCT-and-Projection-data da The Cancer Imaging Archive*, que contém imagens abdominais de 50 pacientes com pares de exames em *low-dose* e *full-dose*. As condições de treinamento e avaliação foram padronizadas para que a comparação entre os modelos fosse justa. O desempenho foi medido com as métricas *Peak Signal-to-Noise Ratio*, *Structural Similarity Index Measure* e *Root Mean Square Error*. Os resultados mostram que os modelos RED-CNN e DnCNN foram capazes de realizar a tarefa de suprimir o ruído.

Palavras-chave: tomografia computadorizada; redes neurais convolucionais; redução de ruído.

[1]Curso de Ciência da Computação, Universidade do Extremo Sul Catarinense (Unesc), gustavomiguelsilvafc@gmail.com

[2]Orientador. Curso de Ciência da Computação, Universidade do Extremo Sul Catarinense (Unesc), allanfavaro@gmail.com

ABSTRACT: Reducing the radiation dose in computed tomography examinations is a clinical necessity. The challenge is that images acquired at low dose have lower quality, with noise and artifacts that impair diagnosis. Increasing the dose improves image quality but also increases the risk to patients. For years, this represented a *trade-off* without a fully satisfactory solution. Recently, deep learning models have demonstrated that it is possible to improve these images without changing the radiation dose. The approach consists of training neural networks to learn the difference between a noisy image and a high-quality image, and applying this correction after image acquisition. Several network architectures have been proposed for this purpose, and comparing them under controlled conditions remains a need identified in the literature. This study performs such a comparison. Three convolutional neural network architectures were evaluated: RED-CNN, DnCNN, and WGAN-VGG, all trained using the same *dataset*, the LDCT-and-Projection-data from The Cancer Imaging Archive, which contains abdominal images from 50 patients with paired *low-dose* and *full-dose* examinations. The training and evaluation conditions were standardized to ensure a fair comparison among the models. Performance was assessed using the *Peak Signal-to-Noise Ratio*, *Structural Similarity Index Measure*, and *Root Mean Square Error* metrics. The results show that the RED-CNN and DnCNN models were able to successfully suppress noise.

Keywords: Computed tomography; Convolutional neural networks; Denoising.

1 INTRODUÇÃO

A tomografia computadorizada (TC) tornou-se rotina em serviços de saúde pelo mundo, sendo usada para avaliar desde traumas abdominais até lesões pulmonares de milímetros. O problema é que essa praticidade tem um custo: cada exame expõe o paciente a radiação ionizante, e quanto melhor a qualidade da imagem desejada, maior a dose necessária (Rodrigues; Vitral, 2007). Entretanto, a aquisição de imagens de alta qualidade está diretamente relacionada à dose de radiação ionizante aplicada durante o exame, condição que representa um risco gradual significativo à saúde dos pacientes. Estima-se que cerca de dois terços da exposição à radiação médica da população americana seja proveniente de exames de TC (Kalra et al., 2003), fator que impulsionou o desenvolvimento de estratégias para reduzir as doses, preservando a utilidade diagnóstica das

imagens. Como consequência dessa redução, ocorre a intensificação do ruído e o surgimento de artefatos nas imagens geradas (Bushberg et al., 2020), dificultando a identificação de estruturas e lesões sutis. Diante dessas limitações, as abordagens baseadas em Aprendizado Profundo (*Deep Learning*), em especial as Redes Neurais Convolucionais (RNCs), tem se destacado como uma abordagem que vem produzindo resultados concretos para o pós-processamento de imagens de TC, devido à sua capacidade de aprender padrões complexos dos dados e suprimir ruído sem depender de métodos manuais de extração de características (LeCun; Bengio; Hinton, 2015; Litjens et al., 2017).

A tomografia computadorizada funciona projetando feixes de raios X através do corpo do paciente em múltiplos ângulos, e registrando a atenuação desses feixes em detectores posicionados no lado oposto. A partir dessas medições, algoritmos matemáticos reconstroem fatias transversais da anatomia interna, expressando a atenuação de cada voxel (*Pixel* Tridimensional) em Unidades Hounsfield (HU), uma escala que vai de aproximadamente -1000 HU (ar) até +1000 HU (osso denso), com água calibrada em 0 HU (Rodrigues; Vitral, 2007). Essa escala padronizada é o que permite distinguir tecidos de diferentes densidades com grande precisão, e é exatamente o que pode ser comprometido quando o ruído está presente.

Do ponto de vista físico, o ruído em TC tem origem predominantemente quântica: é consequência direta da natureza estatística dos fótons de raios X. Quando a dose de radiação é reduzida, o número de fótons que atravessa o paciente e chega ao detector diminui, o que aumenta a variabilidade estatística das medições, fenômeno descrito por uma distribuição de Poisson (Bushberg et al., 2020). Na imagem final, esse comportamento se manifesta como variações aleatórias na densidade de *pixels*, produzindo uma textura granular que interfere na visualização de estruturas de baixo contraste, como lesões sutis no fígado ou nódulos pulmonares de pequeno porte.

Além do ruído quântico, outros fatores contribuem para a degradação das imagens em protocolos de baixa dose: ruído eletrônico dos circuitos do detector, artefatos de espalhamento (scatter) e efeitos de endurecimento do feixe, que ocorrem quando fótons de menor energia são preferencialmente absorvidos ao longo do percurso, alterando o espectro efetivo da radiação que chega ao detector (Bushberg et al., 2020). Cada um desses fatores contribui de forma diferente para as distorções visíveis na imagem, e nenhum método de redução de ruído, seja clássico ou base-

ado em aprendizado profundo, lida com todos eles da mesma maneira.

Durante muito tempo, a principal forma de reduzir ruído em imagens de TC era utilizando técnicas como *Filtered Back Projection* (Retroprojeção Filtrada), posteriormente, métodos de reconstrução iterativa. A reconstrução iterativa trouxe melhorias relevantes, porém, seu custo computacional elevado e a tendência de gerar texturas artificiais nas imagens, limitaram sua adoção ampla na prática clínica (Willemink; Noël, 2019). Filtros clássicos como BM3D e os baseados em variação total (*Total Variation*) também foram bastante explorados, porém tendem a suavizar demais as imagens, apagando detalhes estruturais que são fundamentais para um diagnóstico preciso, problema conhecido como *over-smoothing* (Yang et al., 2018). Foi justamente esse impasse que abriu espaço para os modelos de Aprendizado Profundo, treinados com pares de imagens de baixa e alta dose de radiação, eles conseguem aprender diretamente o mapeamento entre imagem ruidosa e imagem de referência, sem depender de regras definidas manualmente, e com melhor preservação de detalhes finos, como bordas de estruturas vasculares e pequenos nódulos (Chen et al., 2017; Litjens et al., 2017).

Diferentemente de redes totalmente conectadas, as RNCs aplicam operações de convolução, filtros de pequena dimensão que percorrem a imagem de entrada e produzem mapas de características que capturam padrões locais como bordas, texturas e gradientes de intensidade (LeCun; Bengio; Hinton, 2015). Como os pesos desses filtros são compartilhados ao longo de toda a imagem, as RNCs são muito mais eficientes em parâmetros do que redes densas, e naturalmente invariantes à posição dos padrões detectados. A profundidade da rede é o que permite o aprendizado hierárquico: camadas iniciais tendem a capturar características simples, enquanto camadas mais profundas combinam essas informações para detectar estruturas progressivamente mais complexas (LeCun; Bengio; Hinton, 2015). No contexto de imagens médicas, essa capacidade de extrair representações em múltiplos níveis de abstração é o que permite às RNCs aprender a diferença entre ruído e sinal diagnóstico de forma automática, sem que o desenvolvedor precise definir manualmente quais características são relevantes (Litjens et al., 2017).

Três arquiteturas concentram grande parte da atenção nos trabalhos recentes sobre *denoising* em *Low-Dose Computed Tomography* (LDCT). A *Residual Encoder-Decoder Convolutional Neural Network* (RED-CNN), proposta por Chen et al. (2017), utiliza uma estrutura *encoder-decoder* com

conexões residuais para recuperar detalhes anatômicos perdidos ao longo do processo de reconstrução. Já a *Denoising Convolutional Neural Network* (DnCNN) proposta por Zhang et al. (2016) trabalha com aprendizado residual, buscando separar o ruído do sinal original diretamente durante o treinamento. Diferentemente dessas abordagens, a *Wasserstein Generative Adversarial Network* com perda perceptual VGG (WGAN-VGG) Yang et al. (2018) emprega aprendizado adversarial e perda perceptual baseada na *Visual Geometry Group 19* (VGG-19), priorizando qualidade visual mesmo quando métricas tradicionais, como *Peak Signal-to-Noise Ratio* (PSNR), não apresentam os melhores resultados. Esses três modelos foram escolhidos por representarem abordagens distintas, regressão direta, aprendizado residual, arquitetura *encoder-decoder* e redes adversariais, o que permite avaliar diferentes estratégias de redução de ruído em imagens de tomografia computadorizada.

Apesar do crescente número de trabalhos propondo novas arquiteturas para redução de ruído em TC, a comparação entre eles ainda é um problema em aberto. Um estudo recente proposto por Eulig, Ommer e Kachelrieß (2024) analisou criticamente oito algoritmos consolidados e concluiu que grande parte dos trabalhos da área utiliza *datasets* distintos, hiperparâmetros inconsistentes e métricas inadequadas, o que torna a reprodução e comparação justa entre os métodos praticamente inviável. Nesse sentido, desenvolver um estudo comparativo com condições controladas e padronizadas, mesmo conjunto de dados, mesmas métricas, mesma infraestrutura de treinamento, contribui diretamente para preencher essa lacuna. Além da importância para a reprodutibilidade dos estudos, há uma consequência prática para o diagnóstico, identificar qual modelo preserva melhor os detalhes anatômicos em imagens de baixa dose pode auxiliar profissionais de saúde a obter diagnósticos mais seguros sem expor o paciente a doses desnecessárias de radiação (Chen et al., 2017; Litjens et al., 2017).

O presente trabalho tem como objetivo geral analisar redes neurais convolucionais aplicados à redução de ruído em imagens de tomografia computadorizada de baixa dose. Foram analisados os modelos RED-CNN, DnCNN e WGAN-VGG, avaliando o desempenho de cada modelo por meio de métricas quantitativas como *PSNR, Structural Similarity Index Measure* (SSIM) e *Root Mean Square Error* (RMSE). Com os resultados obtidos, é possível identificar quais arquiteturas apresentam melhor equilíbrio entre supressão de ruído e preservação de detalhes anatômicos, contribuindo

para a escolha de soluções mais eficazes no contexto clínico.

O trabalho está organizado em cinco seções. A seção 1 apresenta o tema e objetivos do trabalho e contextualiza sobre o problema a ser estudado. A seção 2 aborda os trabalhos correlatos, analisando estudos semelhantes já desenvolvidos na área, destacando suas metodologias e resultados. Na seção 3 estão descritos os materiais e métodos utilizados, incluindo a fundamentação teórica, conjunto de dados, modelos avaliados e os hiperparâmetros escolhidos para o treinamento. A seção 4, dispõe os resultados e discussões, apresenta os resultados obtidos e abre espaço para a discussão, comparando o desempenho dos modelos por meio de métricas quantitativas. Por fim a seção 5 que apresenta a conclusão sobre o trabalho.

2 TRABALHOS CORRELATOS

Alguns trabalhos semelhantes já foram desenvolvidos, utilizando diferentes modelos, configurações e *datasets*.

A dissertação de Almeida (2024) intitulada "Comparações de Modelos Neurais em Redução de Ruído de Imagens de tomografia computadorizada", apresenta a importância de amenizar a exposição de pacientes à radiação ionizante nos exames de tomografia computadorizada (TC). No trabalho é realizado um estudo comparativo entre diferentes modelos de Redes Neurais Convolucionais (RNCs), utilizando as mesmas bases de dados, métricas de avaliação e condições de treinamento, com o objetivo de reduzir ruído em imagens de TC. Os modelos avaliados incluem RED-CNN, U-Net, WGAN e SCANN, treinados com diferentes funções de perda e comparados por meio de métricas PSNR, SSIM e NRMSE. Os resultados indicaram que modelos treinados com funções de perda *pixel-wise* alcançaram métricas quantitativas superiores, porém produziram imagens com borramento e perda de texturas. Em contrapartida, modelos treinados com funções de perda não *pixel-wise*, apesar de métricas menores, conseguiram preservar melhor detalhes visuais diagnósticos. O trabalho reforça a necessidade de múltiplas execuções com variação de hiperparâmetros para comparações mais justas entre as arquiteturas.

O artigo *Low Dose CT Image Denoising: A Comparative Study of Deep Learning Models and Training Strategies* de Zhao et al. (2024) compara sete modelos de *Deep Learning* para redução de ruído em imagens de tomografia computadorizada de baixa dose. O estudo busca identificar os modelos com melhor desempenho para redução de ruído, quais tem a

melhor generalização e os modelos com melhor tempo de processamento. O artigo leva em conta que reduzir a dose de radiação diminui os riscos aos pacientes porem prejudica a qualidade das imagens, gerando muitos artefatos e ruídos. A base de dados do trabalho é composta por imagens humanas simuladas e imagens de porcos reais com diferentes doses de radiação. Os resultados do trabalho apontam que o melhor modelo foi o U-net, com destaques nas métricas PSNR e RMSE, também ressalta que o modelo apresentou a melhor remoção de artefatos e ruídos e uma excelente preservação estrutural das imagens, mantendo detalhes anatômicos mesmo em imagens com pouca dose. O trabalho conclui que modelos baseados em RNCs ainda superam modelos GANs e Transformers no contexto de redução de ruído em imagens de TC.

Segundo o trabalho de Eulig, Ommer e Kachelrieß (2024) Os métodos atuais de remoção de ruído não tem avançado significativamente como dizem outros trabalhos científicos publicados, em seu artigo *Benchmarking deep learning-based low-dose CT image denoising algorithms*, ele analisa criticamente as técnicas de remoção de ruído em imagens de TC de baixa dose. O estudo aponta que as metodologias aplicadas em outros trabalhos são inconsistentes e injustas, o que impossibilita a reprodução delas. O autor especifica que os trabalhos a cerca do tema utilizam *datasets* diferentes, hiperparâmetros injustos, métricas inadequadas e a falta de código aberto para averiguar possíveis implementações e modificações. Por conta dessas variações, o estudo propões uma padronização para avaliar os métodos de redução de ruído em LDCT de forma justa. O trabalho compara um total de oito algoritmos famosos, sendo avaliados com métricas classicas como SSIM, PSNR e VIF. E também com métricas clínicas como reconstrução de lesões, preservação de características radiômicas e bordas, contraste e precisão de números CT (HU). O estudo conclui que, houve pouco ou quase nenhum avanço significativo nos últimos anos, onde arquiteturas modernas performam pior que modelos antigos ou tem desempenho estatisticamente muito semelhante aos modelos antigos. O trabalho aponta que o modelo RED-CNN que foi um dos primeiros métodos, continua sendo extremamente eficiente, superando diversos modelos recentes com excelente preservação estrutural e ótima reconstrução de lesões.

No trabalho *Low-Dose CT with a Residual Encoder-Decoder Convolutional Neural Network* (RED-CNN) de Chen et al. (2017), é proposto um método de aprendizado profundo para redução de ruído em imagens de tomografia computadorizada de baixa dose. O objetivo do trabalho é con-

seguir diminuir o impacto causado pela redução de raios X, aprimorando a detecção de lesões e preservando detalhes anatômicos importantes. Os autores destacam que métodos tradicionais enfrentam algumas limitações para conseguir processar os dados, alguns dependem de acesso direto aos dados dos equipamentos de exame, que em muitas vezes são propriedade privada dos fabricantes. Em relação a isso eles mencionam que técnicas de pós-processamento em imagens constantemente apresentam dificuldade em equilibrar a preservação estrutural e a suavização de ruído. O trabalho utilizou dados simulados do NBIA e os dados clínicos do desafio *NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge*. O modelo RED-CNN foi comparado com os métodos TV-POCS, K-SVD, BM3D, CNN10, KAIST-Net. Os resultados que eles obtiveram, mostravam que em métricas como PNSSR, RMSE e SSIM o RED-CNN era o método com melhor desempenho quantitativo e em relação a parte visual apresentou a melhor preservação de detalhes anatômicos, detecção de lesões, delineamento de vasos sanguíneos e a maior supressão de ruído. A conclusão deles indica a viabilidade de arquiteturas *encoder-decoder* residuais para redução de ruído em LDCT, indicando um equilíbrio entre supressão de ruído e preservação de informação estrutural. O método é indicado para trabalhos futuros voltados a reconstrução 3D, TC espectral e outras modalidades médicas.

Em *Low Dose CT Image Denoising Using a Generative Adversarial Network with Wasserstein Distance and Perceptual Loss* desenvolvido por Yang et al. (2018) propõe um método de redução de ruído em imagens de tomografia computadorizada LDCT utilizando rede adversarial generativa baseada em Wasserstein GAN (WGAN) em conjunto com perda perceptual extraída da rede VGG-19. O trabalho tem como objetivo a redução de ruído das imagens LDCT preservando os detalhes anatômicos importantes, problema frequente em métodos baseados em erro quadrático médio (MSE). Os autores ressaltam que técnicas tradicionais e muitos métodos de *Deep Learning* específicos para reduzir MSE costumam produzir imagens excessivamente suavizadas, ocasionando na perda de detalhes como textura e estruturas importantes para diagnósticos.Com base nessas limitações, utilizaram dois componentes em conjunto, a WGAN que é a responsável por aproximar estatisticamente a distribuição das imagens de tomografia de baixa dose (LDCT) da distribuição das imagens de *Normal-Dose Computed Tomography* (NDCT), proporcionando imagens mais reais e reduzindo artefatos indesejados. Já a perda perceptual VGG atua na preservação de características estruturais e visuais importantes em um espaço

de características profundas. O treinamento do modelo WGAN-VGG, que utiliza uma CNN como um discriminador adversarial e uma rede VGG-19 pré-treinada para cálculo da perda perceptual, foi feito utilizando o *dataset* clínico 2016 *NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge*, onde contem imagens abdominais de CT em dose normal e baixa dose simulada. Os resultados apontam que o método foi capaz de reduzir de maneira significativa o ruído das imagens preservando detalhes anatômicos e bordas quando comparado com abordagens baseadas apenas em MSE. As imagens produzidas apresentam menos efeito de *over-smoothing* e menos artefatos visuais. Ressalta que por mais que métodos baseados em MSE alcancem valores de PSNR maiores, o método WGAN-VGG teve melhor qualidade visual e preservação de estruturas importantes segundo avaliações quantitativas e análise de radiologistas. O trabalho conclui que o método por si só é ineficiente, mas em conjunto com a perda perceptual se torna uma combinação promissora para redução de ruído em LDCT, oferecendo um equilíbrio entre qualidade e redução de ruído comparado a métodos tradicionais e e redes baseadas exclusivamente em MSE.

O conjunto de trabalhos revisados evidenciam que, modelos de aprendizado profundo são claramente superiores a métodos tradicionais na tarefa de *denoising*, mas comparar esses modelos de forma confiável, continua sendo problemático devido a diversidade de *datasets*, métricas e configurações apresentadas na literatura (Eulig; Ommer; Kachelrieß, 2024). Esse trabalho se diferencia por utilizar condições experimentais padronizadas em todos os modelos avaliados, adotando o mesmo *dataset*, métricas e ambiente computacional. Isso permite uma comparação mais justa e reproduzível em relação à maioria dos estudos já publicados.

3 MATERIAIS E MÉTODOS

Este trabalho realiza o treinamento e a validação de três modelos de Redes Neurais Convolucionais aplicados à redução de ruído em imagens de tomografia computadorizada de baixa dose, sendo eles RED-CNN, DnCNN e WGAN-VGG. Todos seguem o paradigma de *denoising* supervisionado com pares de imagens, para cada imagem de baixa dose (LDCT), existe uma imagem correspondente de dose normal (NDCT) que serve como referência (*ground truth*). Durante o treinamento, o modelo recebe a imagem ruidosa como entrada e produz uma estimativa da imagem limpa. A discrepância entre essa estimativa e a referência é quantificada por uma função de perda, cujo gradiente é propagado de volta pela rede para

atualizar os pesos, processo conhecido como retropropagação do erro. O objetivo é comparar o desempenho de cada arquitetura sob as mesmas condições de treinamento e avaliação, identificando pontos fortes e limitações de cada abordagem. Todo o desenvolvimento foi feito em Python 3.8, com a utilização do *framework* PyTorch/torch, com aceleração de processamento por meio de *backends* CUDA e cuDNN. Também foram utilizadas as bibliotecas torchvision, numpy, matplotlib, pydicom, scikit-image, tensorboardX, h5py, opencv-python/cv2, tcia_utils e pandas.

3.1 *DATASET*

O conjunto de dados utilizado foi o *Low Dose CT Image and Projection Data* (LDCT-and-Projection-data)[3] (McCollough et al., 2020), disponível na plataforma The Cancer Imaging Archive (TCIA) (Clark et al., 2013) e descrito por Moen et al. (2021), com tamanho total de 11,4 GB. O *dataset* completo contém dados de 299 pacientes, porém apenas 50 deles possuíam pares de imagens *low-dose* e *full-dose*, os outros pacientes continham somente imagens *full-dose* e por isso foram descartados. Para garantir uma avaliação imparcial, os dados do paciente L299 foram reservados exclusivamente para testes, não sendo utilizados em nenhum outro momento. No total restaram 7.595 pares de treino. Para acessar e baixar os dados foi desenvolvida uma API em Python, e os modelos foram configurados para consumir as imagens diretamente do diretório local gerado por ela. Cada paciente do subconjunto utilizado possui aproximadamente 155 imagens do abdômen humano, sendo cada par composto por uma imagem *low-dose* (25% da dose original) e uma *full-dose* (100% da dose). Esse formato possibilitou o treinamento supervisionado dos modelos, já que as imagens *full-dose* serviram como referência durante o aprendizado.

Antes do treinamento, os dados passaram por um processo de normalização para garantir consistência entre os modelos e assegurar que todos fossem avaliados em condições equivalentes.

3.2 PRÉ-PROCESSAMENTO

Antes do treinamento, as imagens do *dataset* foram convertidas de *Digital Imaging and Communications in Medicine* (DICOM) para arrays NumPy e submetidas a um processo de clipping de Unidades Hounsfield (HU). Os valores de *pixel* foram limitados ao intervalo [-1024, 3072] HU,

[3]Disponível em: <https://www.cancerimagingarchive.net/collection/ldct-and-projection-data/>. Acesso em: 30 jun. 2026.

cobrindo a faixa clinicamente relevante para imagens abdominais, e em seguida normalizados para o intervalo [0, 1] por normalização linear mínimo-máximo. Essa etapa é essencial para manter os gradientes em magnitudes numéricas estáveis durante o treinamento e garantir que todos os modelos recebam entradas comparáveis.

Como as imagens originais têm dimensões variáveis, adotou-se extração de *patches* de 64×64 *pixels*, amostrados aleatoriamente de cada fatia durante o carregamento do *batch*. Essa estratégia aumenta a diversidade de exemplos de treinamento e reduz o consumo de memória por amostra, tornando viável treinar arquiteturas mais exigentes dentro dos 6 GB de VRAM disponíveis. A divisão entre treino, validação e teste foi feita ao nível do paciente, e não ao nível de fatias individuais, para evitar vazamento de dados (*data leakage*). Fatias de um mesmo paciente guardam alta correlação entre si, e dividi-las aleatoriamente faria com que o modelo avaliasse imagens estruturalmente muito próximas às do treinamento, inflando as métricas artificialmente. Dos 50 pacientes disponíveis, 49 foram destinados ao treinamento, e apenas 1 ao teste final. Não foram aplicadas técnicas de aumento de dados como rotação ou espelhamento, pois transformações geométricas em imagens de TC podem alterar a orientação anatômica e comprometer a correspondência *pixel* a *pixel* com as imagens de referência.

Todos os modelos foram treinados com os mesmos hiperparâmetros base para garantir uma comparação justa. O ponto de equilíbrio foi definido a partir dos valores mínimos que o modelo mais exigente conseguia processar sem exceder os 6 GB de VRAM, algumas diferenças entre os modelos como otimizador e função de perda são consequências direta das características de cada modelo, e não introduzem viés na comparação. Os demais hiperparâmetros foram configurados com os mesmos valores de *Batch Size*, *Patch Size*, *Learning Rate* e Épocas.

Tabela 1 – Configuração dos hiperparâmetros dos modelos

Modelo	**Batch size**	**Patch size**	**Épocas**	**LR**
DnCNN	8	64×64	100	1×10^{-4}
RED-CNN	8	64×64	100	1×10^{-4}
WGAN-VGG	8	64×64	100	1×10^{-4}

Fonte: Elaborado pelo autor.

3.3 ARQUITETURAS AVALIADAS

Para os modelos RED-CNN e DnCNN, foram utilizadas implementações públicas disponíveis no GitHub como ponto de partida. A implementação do RED-CNN baseia-se diretamente da arquitetura descrita por Chen et al. (2017), Sinyu (2018), enquanto o modelo DnCNN parte de uma reimplementação em PyTorch do trabalho original proposto por Zhang et al. (2016), sendo utilizada a implementação disponibilizada por Yan (2018). A arquitetura WGAN-VGG foi implementada em PyTorch, como base em Yang et al. (2018), aplicando os princípios da Wasserstein GAN e da perda perceptual baseada em VGG-19.

Em todos os casos, foi necessário adaptar os modelos às restrições do hardware utilizado. A principal limitação foi a memória de vídeo disponível, 6 GB de VRAM, o que exigiu ajustes no *batch size* e na forma de carregamento dos dados. O uso de CUDA para aceleração via GPU também precisou ser configurado explicitamente. Nenhuma dessas mudanças afetou a estrutura interna das redes nem o comportamento das arquiteturas originais.

3.3.1 RED-CNN

A *Residual Encoder-Decoder Convolutional Neural Network* (RED-CNN) foi proposta por Chen et al. (2017) com o objetivo de realizar pós-processamento de imagens de TC sem depender do acesso aos dados brutos do sinograma, o que torna a abordagem compatível com qualquer equipamento, independente do fabricante. O modelo recebe como entrada *patches* 2D de tamanho 64×64 *pixels* extraídos das fatias já normalizadas, e produz como saída uma estimativa da versão limpa desses mesmos *patches*, o treinamento acontece fatia por fatia, e não sobre volumes tridimensionais completos.

A estrutura do modelo segue um padrão *encoder-decoder*, o *encoder* é composto por cinco camadas convolucionais sequenciais com kernel de tamanho fixo, que progressivamente extraem representações da entrada sem recorrer a operações de *pooling* ou redução de resolução espacial; O *decoder*, por sua vez, aplica cinco camadas de convolução transposta para reconstruir a imagem na mesma resolução da entrada. O que diferencia esse modelo de *encoders-decoders* convencionais são as conexões residuais, atalhos que ligam diretamente camadas correspondentes do *encoder* e do *decoder* (He et al., 2016). Essas conexões permitem que informações espaciais mais finas, como bordas e texturas, cheguem ao de-

coder sem precisar ser reaprendidas a partir das representações comprimidas. A função de perda usada no treinamento é o erro quadrático médio (MSE) entre a saída da rede e o *patch* de referência em dose normal.

3.3.2 DNCNN

A *Denoising Convolutional Neural Network* (DnCNN), introduzida por Zhang et al. (2016), parte de uma lógica diferente das abordagens *encoder-decoder*. A premissa do trabalho original é que a rede aprenda o componente de ruído da imagem em vez de reconstruir diretamente a versão limpa, estratégia chamada de aprendizado residual. Na prática, isso significa que a rede é treinada para que sua saída se aproxime do ruído contido na entrada, e a imagem recuperada é obtida pela subtração entre entrada e saída estimada.

A estrutura da rede consiste em 17 camadas convolucionais sequenciais, todas com o mesmo número de filtros de tamanho 3×3, intercaladas com normalização em lote e funções de ativação ReLU. Não há operações de *pooling*, *downsampling* ou *upsampling*, a resolução espacial da entrada é mantida do início ao fim. A normalização em lote reduz o problema do deslocamento interno de covariáveis (*internal covariate shift*), fenômeno em que a distribuição das ativações muda a cada iteração de treinamento, dificultando a convergência. Por ser uma arquitetura compacta em termos de operações por imagem, a DnCNN converge de forma significativamente mais rápida do que modelos *encoder-decoder* (Ioffe; Szegedy, 2015).

3.3.3 WGAN-VGG

A WGAN-VGG foi proposta por Yang et al. (2018) como uma resposta ao problema de *over-smoothing* que afeta modelos treinados exclusivamente com MSE. A ideia central é que minimizar o erro *pixel* a *pixel* não garante que a imagem gerada seja perceptualmente convincente: ela pode ser matematicamente próxima da referência e ainda parecer borrada ou com perda de textura, tanto para o olho humano quanto para fins diagnósticos.

O modelo combina dois componentes. O primeiro é um gerador convolucional que recebe a imagem de baixa dose como entrada e produz diretamente uma estimativa da imagem de dose normal, diferente de GANs generativas clássicas, não há ruído aleatório sendo mapeado para imagens. Esse gerador é treinado dentro do arcabouço da Wasserstein GAN (WGAN): um discriminador adversarial (*critic*) avalia as imagens gera-

das e retorna um escore escalar, sem a transformação sigmoide típica das GANs convencionais, e sem produzir uma probabilidade explícita de real ou falso (Yang et al., 2018). O treinamento adversarial usa o otimizador RMSprop com *clipping* dos pesos do *critic*, seguindo a formulação original da WGAN. Essa abordagem torna o treinamento mais estável do que o GAN original, fornecendo um gradiente mais suave e informativo mesmo quando as distribuições do gerador e dos dados reais têm suporte disjunto.

O segundo componente é a perda perceptual, calculada a partir das ativações de camadas intermediárias da rede VGG-19 pré-treinada no ImageNet. A VGG-19 permanece completamente congelada durante todo o treinamento e serve exclusivamente para computar essa perda, sem atualizar seus pesos. Em vez de penalizar diferenças *pixel* a *pixel*, a rede é penalizada por diferenças em representações de alto nível, como texturas e padrões estruturais. A função de perda total combina três termos: a perda adversarial de Wasserstein, o MSE *pixel* a *pixel* e a perda perceptual baseada na VGG-19. Essa combinação tende a produzir imagens visualmente mais naturais, ainda que os valores de PSNR frequentemente fiquem abaixo dos modelos baseados exclusivamente em MSE, o que reflete uma limitação conhecida das métricas tradicionais para capturar qualidade perceptual.

3.4 MÉTRICAS DE AVALIAÇÃO

O desempenho dos modelos foi avaliado por meio de três métricas quantitativas usadas na literatura de *denoising* em imagens médicas: PSNR, RMSE e SSIM.

O RMSE (*Root Mean Square Error*) representa, em média, o desvio de cada *pixel* da imagem gerada em relação à referência, calculado como (Eulig; Ommer; Kachelrieß, 2024):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$$

onde y_i é o valor do *pixel* na imagem de referência e $\hat{y}_i$ é o valor correspondente na imagem gerada. Por estar na mesma escala dos *pixels*, o RMSE é de fácil interpretação.

O *Peak Signal-to-Noise Ratio* (PSNR) mede a relação entre o valor máximo possível de um *pixel* e a magnitude do erro, expresso em

decibéis (Eulig; Ommer; Kachelrieß, 2024):

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX^2}{MSE}\right)$$

onde MAX é o valor máximo possível de um *pixel* e *Mean Squared Error* (MSE) é a média dos quadrados das diferenças por *pixel*, equivalente a $RMSE^2$ conforme apresentado na equação anterior. Valores mais altos indicam maior semelhança entre as imagens. Como o PSNR é diretamente derivado do MSE, modelos treinados com essa função de perda costumam apresentar melhores valores nessa métrica, o que pode favorecer resultados com *over-smoothing* sem que isso seja detectado pelo indicador (Yang et al., 2018).

O SSIM (*Structural Similarity Index*), proposto por Wang et al. (2004), surgiu como alternativa às métricas baseadas puramente em diferença de *pixels*. O índice avalia três componentes entre a imagem gerada e a referência: luminância, contraste e estrutura, produzindo um valor entre 0 e 1, onde 1 representa identidade perfeita com a *ground truth*. Para imagens de TC, o SSIM é útil justamente por ser sensível a diferenças em bordas e texturas, aspectos com impacto direto na interpretação clínica.

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

As três métricas são complementares, enquanto RMSE e PSNR quantificam o desvio absoluto entre *pixels*, o SSIM aproxima a percepção estrutural humana da qualidade da imagem. Usá-las em conjunto reduz o risco de que um modelo com boas métricas de intensidade apresente degradação estrutural relevante, ou vice-versa.

Após o treinamento, os três modelos foram avaliados sob as mesmas condições, usando o mesmo conjunto de teste e as mesmas métricas. Essa padronização foi uma escolha metodológica, motivada pelas críticas de Eulig, Ommer e Kachelrieß (2024) sobre a falta de reprodutibilidade nos trabalhos da área. O estudo não busca declarar um modelo superior, mas entender o que cada abordagem perde e ganha: mais supressão de ruído geralmente significa menos preservação de detalhes, e as métricas quantitativas deixam isso visível.

3.5 RECURSOS COMPUTACIONAIS

O desenvolvimento foi realizado em um computador pessoal com processador AMD Ryzen 5 3600, placa de vídeo NVIDIA GeForce GTX 1660 Super (6 GB de VRAM), 24 GB de memória RAM (3×8 GB a 2933 MHz) e armazenamento de 2 TB SSD NVMe. O sistema operacional utilizado foi o Windows 11 Pro, com o ambiente de desenvolvimento PyCharm, e aceleração de treinamento via bibliotecas Python CUDA.

4 RESULTADOS E DISCUSSÃO

Os testes dos modelos foram avaliados utilizando dados de um paciente que não esteve presente no conjunto de treinamento, os dados reservados desse paciente contém 155 imagens de tomografia computadorizada. As métricas PSNR, RMSE e SSIM foram calculadas para cada imagem e posteriormente agregadas por meio da média aritmética.

Tabela 2 – Resultados das Métricas

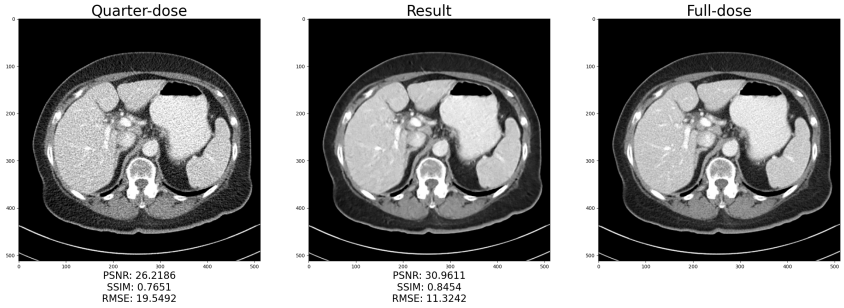
Modelo	**PSNR**	**RMSE**	**SSIM**
RED-CNN	32.4084	9.6992	0.8711
DnCNN	32.0442	10.1124	0.8678
WGAN-VGG	25.8253	20.5594	0.7818

Fonte: Elaborado pelo autor.

Conforme observado na Tabela 2 o modelo RED-CNN apresentou o melhor desempenho, ele obteve os maiores valores de PSNR e SSIM, além do menor valor de RMSE. O modelo DnCNN apresentou resultados próximos aos do RED-CNN, indicando uma capacidade muito semelhante para redução de ruído. Já o modelo WGAN-VGG apresentou um desempenho inferior aos demais, resultando em métricas piores em relação aos outros modelos observados.

Esse padrão de desempenho bate com resultados anteriores. Chen et al. (2017) já demonstravam a superioridade do modelo RED-CNN frente a métodos tradicionais, e Eulig, Ommer e Kachelrieß (2024) , em um *benchmarking* com oito algoritmos, concluíram que nenhum método mais recente supera de forma consistente o RED-CNN em métricas como PSNR e SSIM. Zhao et al. (2024), ao avaliarem sete modelos em condições padronizadas, confirmam que as arquiteturas baseadas em MSE lideram essas métricas, enquanto modelos adversariais como o WGAN ficam consistentemente abaixo.

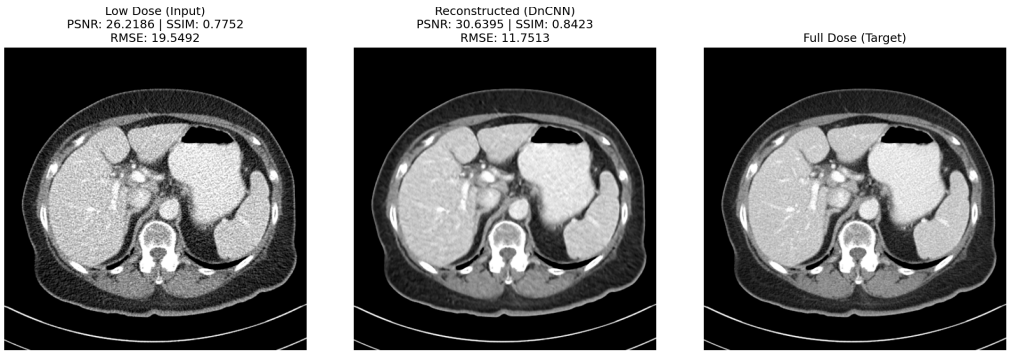
Figura 1 – Resultado obtido pelo RED-CNN



Fonte: Elaborado pelo autor.

A figura 1 apresenta um exemplo representativo do paciente de teste, exibindo a imagem *quarter-dose* (*low-dose*) de entrada, a imagem de saída produzida pelo RED-CNN e a imagem *full-dose* utilizada como referência. Os valores PSNR, SSIM e RMSE presentes na figura ilustram os ganhos identificados na avaliação quantitativa, o modelo elevou o PSNR de 26,22 dB para 30,96 dB e o SSIM de 0,7651 para 0,8454, ao mesmo tempo que conseguiu reduzir o RMSE de 19,55 para 11,32, uma redução de aproximadamente 42% no erro quadrático médio em relação à imagem *full-dose* de referência. Esses resultados indicam que o modelo RED-CNN foi capaz de suprimir o ruído nas imagens de baixa dose, e gerou uma saída próxima da imagem *full-dose*.

Figura 2 – Resultado obtido pelo DnCNN

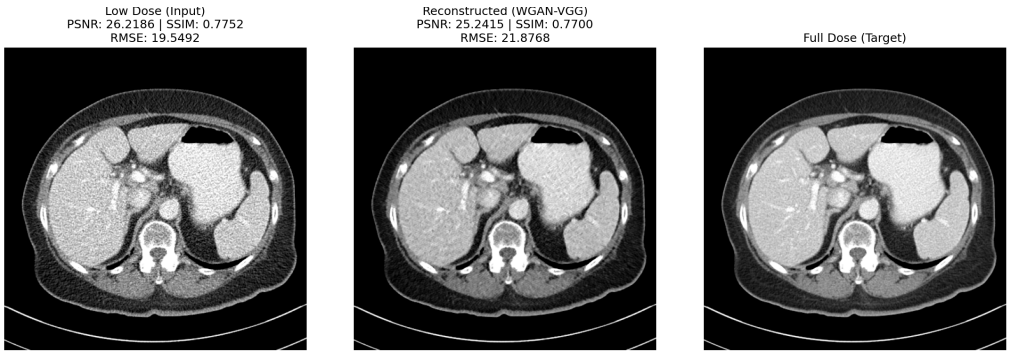


Fonte: Elaborado pelo autor.

A figura 2 apresenta o mesmo recorte representativo do paciente de teste, agora processada pelo DnCNN. Com base nos valores da imagem de entrada com PSNR de 26,22 dB, SSIM de 0,7752 e RMSE de 19,55, o modelo conseguiu produzir uma saída com PSNR de 30,64 dB, SSIM de 0,8423 e RMSE de 11,75, representando uma redução de 40%

no erro quadrático médio em relação à imagem de referência. Esses valores ficaram próximos aos obtidos pelo RED-CNN no mesmo recorte, o que está coerente com os resultados apresentados na Tabela 2, onde o DnCNN atingiu valores de PSNR e SSIM ligeiramente abaixo do RED-CNN que foi o modelo com melhor desempenho.

Figura 3 – Resultado obtido pelo WGAN-VGG



Fonte: Elaborado pelo autor.

A Figura 3 apresenta o mesmo recorte utilizado nos demais modelos, agora processado pelo modelo WGAN-VGG. Diferentemente dos modelos anteriores, os resultados quantitativos indicam que o método não conseguiu melhorar as métricas em relação à imagem de entrada. O PSNR reduziu de 26,22 dB para 25,24 dB, o SSIM diminuiu de 0,7752 para 0,7700 e o RMSE aumentou de 19,55 para 21,88, indicando desempenho inferior em todos os indicadores avaliados. Esses resultados também são consistentes com os valores médios apresentados na Tabela 2, na qual o WGAN-VGG obteve os menores resultados entre os três modelos.

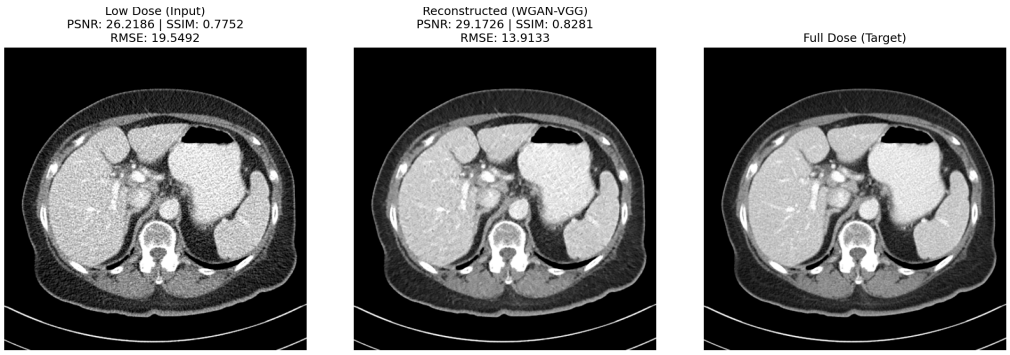
Esse comportamento está de acordo com a literatura. Yang et al. (2018) reconhecem que o método não otimiza diretamente o erro *pixel* a *pixel*, característica que penaliza métricas como PSNR e RMSE, concentrando sua principal vantagem na qualidade visual e na preservação de estruturas. De forma semelhante, Zhao et al. (2024) observaram que modelos baseados em WGAN apresentam valores de PSNR e SSIM consistentemente inferiores aos obtidos por métodos treinados com funções de perda baseadas em MSE.

Enquanto o RED-CNN e o DnCNN buscam minimizar diretamente as diferenças entre a imagem reconstruída e a imagem de referência, o WGAN-VGG adota uma estratégia baseada em aprendizado adversarial e em características perceptuais extraídas pela rede VGG. Como essa abordagem não otimiza diretamente métricas como o RMSE, é esperado que o

modelo apresente valores mais elevados nesse indicador.

Além disso, durante os testes preliminares, observou-se que essa arquitetura foi negativamente influenciada pela taxa de aprendizado adotada na configuração padrão. Em experimentos com taxas de aprendizado mais elevadas, o modelo apresentou resultados quantitativos mais próximos dos obtidos pelos demais métodos. Entretanto, esses testes exploratórios não fizeram parte da configuração final adotada neste trabalho.

Figura 4 – Resultado obtido pelo WGAN-VGG com taxa de aprendizado maior



Fonte: Elaborado pelo autor.

Trocando o hiperparâmetro de *Learning Rate* de 1×10^{-4} para 1×10^{-3} é possível obter valores competitivos, conforme ilustrado na Figura 4. Nessa configuração, o modelo conseguiu elevar o PSNR de 26,22 para 29,20 e o SSIM de 0,7752 para 0,8281, além de reduzir o RMSE de 19,55, para 13,92, representando uma queda de aproximadamente 29% no erro quadrático médio em relação à imagem de referência. Esse desempenho difere do observado na configuração padrão, que chegou a degradar as métricas em relação a própria imagem de entrada. Embora essa avaliação esteja limitada a um recorte representativo e não considere o conjunto completo de teste, os resultados obtidos aproximam o WGAN-VGG dos demais modelos avaliados. Isso sugere que sua principal limitação não está na arquitetura em si, mas na sensibilidade à escolha dos hiperparâmetros, comportamento típico de arquiteturas adversariais.

5 CONCLUSÃO

Este trabalho teve como objetivo comparar três arquiteturas de redes neurais convolucionais, RED-CNN, DnCNN e WGAN-VGG, sendo aplicadas na tarefa de redução de ruído em imagens de tomografia computadorizada de baixa dose. Os três modelos foram treinados e avaliados

nas mesmas condições, o mesmo conjunto de dados, a mesma divisão entre treino e teste, e as mesmas métricas aplicadas sobre o mesmo paciente de teste. Essa padronização foi uma escolha deliberada e motivada por fatores computacionais e principalmente pela dificuldade de comparar resultados entre estudos que utilizam configurações diferentes.

Os resultados mostraram que o modelo RED-CNN obteve o melhor desempenho geral, sendo seguido de perto pelo DnCNN. A margem de diferença entre os dois modelos não foi algo expressivo, o que faz sentido considerando que ambos são treinados para reduzir a diferença direta entre a imagem gerada e a imagem de referência. A DnCNN faz isso aprendendo o ruído presente na entrada para depois subtraí-lo, enquanto a RED-CNN aprende diretamente a imagem limpa. Nas métricas utilizadas, essa diferença de abordagem não produziu uma separação significativa de desempenho.

O WGAN-VGG apresentou um comportamento diferente. Em vez de melhorar os valores em relação à imagem de entrada ruidosa, o modelo piorou todos os indicadores, PSNR caiu, o SSIM diminuiu e o RMSE subiu. Isso não significa necessariamente que essa arquitetura seja inadequada para resolver o problema, mas que o objetivo pelo qual foi treinada, não está alinhado com as métricas *pixel* a *pixel* utilizadas na avaliação. Testes com taxas de aprendizado maiores mostraram resultados quantitativos mais próximos dos outros modelos, o que indica que parte do problema está na configuração dos hiperparâmetros.

O trabalho tem algumas limitações, que abrem espaço para trabalhos futuros. O principal gargalo foi o hardware, com 6 GB de VRAM, foi necessário reduzir o tamanho dos *batches* e a resolução dos *patches*, o que provavelmente comprimiu os resultados do WGAN-VGG. Com a utilização de um ambiente computacional mais robusto, seria possível testar configurações maiores e avaliar se isso impacta os resultados. A avaliação também ficou limitada a um único paciente de teste com 155 imagens, ampliar esse conjunto para mais pacientes tornaria as conclusões mais sólidas. O modelo WGAN-VGG, especificamente, merece uma escolha mais cuidadosa de hiperparâmetros, já que testes preliminares com taxas de aprendizado maiores mostraram resultados mais competitivos. Por fim, incluir métricas voltadas à qualidade perceptual mais recentes seriam extensões naturais para dar continuidade ao trabalho.

Arquiteturas treinadas para minimizar o erro *pixel* a *pixel* em relação à imagem referência tendem a se sair melhor nas métricas conven-

cionais. O WGAN-VGG segue uma lógica diferente, priorizando qualidade visual em vez de fidelidade *pixel* a *pixel*, e por isso os valores de PSNR, SSIM e RSME não refletem bem o que ele tenta fazer. Para avaliar esse tipo de modelo de forma mais justa, seriam necessárias métricas mais adequadas ao seu objetivo de treinamento.

Em trabalhos futuros o passo mais claro seria dar prioridade ao modelo WGAN-VGG: o modelo mostrou sinais de que a taxa de aprendizado padrão estava errada, e testes preliminares com valores maiores já produziram resultados melhores. Antes de descartar a arquitetura, vale ajustar os hiperparâmetros com mais cuidado. Um *dataset* com mais pacientes e uma GPU com mais VRAM tornariam qualquer conclusão mais sólida.

DECLARAÇÃO DE USO DE INTELIGÊNCIA ARTIFICIAL

Os autores declaram que ferramentas de Inteligência Artificial generativa foram utilizadas exclusivamente como apoio auxiliar à elaboração deste trabalho, incluindo assistência na revisão linguística, estruturação textual e organização de ideias. Nenhuma ferramenta de IA foi utilizada para substituição da autoria intelectual, análise científica, interpretação dos resultados ou tomada de decisões acadêmicas. A responsabilidade integral pelo conteúdo, originalidade, veracidade das informações e conformidade com os princípios de integridade científica permanece integralmente atribuída aos autores.

REFERÊNCIAS

ALMEIDA, M. B. d. **Comparações de Modelos Neurais em Redução de Ruído de Imagens de Tomografia Computadorizada**. Dissertação (Dissertação (Mestrado)) — Universidade Federal de Pernambuco (UFPE), Centro de Informática (CIn), Programa de Pós-Graduação em Ciência da Computação, Recife, 2024. Acesso em: 8 jun. 2025.

BUSHBERG, J. T. et al. **The Essential Physics of Medical Imaging**. 4th. ed. [S.l.]: Wolters Kluwer. ISBN 978-1496377215. 2020. Acesso em: 8 jun. 2025.

CHEN, H. et al. **Low-dose ct with a residual encoder-decoder convolutional neural network**. IEEE Transactions on Medical Imaging, IEEE, v. 36, n. 12, p. 2524–2535. 2017, Disponível em: <https://doi.org/10.1109/TMI.2017.2715284>. Acesso em: 8 jun. 2025.

CLARK, K. et al. **The cancer imaging archive (TCIA): Maintaining and operating a public information repository**. Journal of Digital

Imaging, Springer, v. 26, n. 6, p. 1045–1057. 2013, Disponível em: <https://doi.org/10.1007/s10278-013-9622-7>. Acesso em: 8 jun. 2025.

EULIG, E.; OMMER, B.; KACHELRIEß, M. **Benchmarking deep learning-based low-dose ct image denoising algorithms**. Medical Physics, Wiley. 2024, Disponível em: <https://doi.org/10.1002/mp.17379>. Acesso em: 8 jun. 2025.

HE, K. et al. **Deep residual learning for image recognition**. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. Las Vegas, NV, USA: IEEE, 2016. 2016. p. 770–778. Acesso em: 1 jun. 2025.

IOFFE, S.; SZEGEDY, C. **Batch normalization: Accelerating deep network training by reducing internal covariate shift**. In: BACH, F.; BLEI, D. (Ed.). **Proceedings of the 32nd International Conference on Machine Learning**. Lille, France: PMLR, 2015. 2015. (Proceedings of Machine Learning Research, v. 37), p. 448–456. Disponível em: <https://proceedings.mlr.press/v37/ioffe15.html>. Acesso em: 1 jun. 2025.

KALRA, M. K. et al. **Low-dose ct of the abdomen: Evaluation of image improvement with use of noise reduction filters—pilot study**. Radiology, Radiological Society of North America, v. 228, n. 1, p. 251–256. 2003, Disponível em: <https://doi.org/10.1148/radiol.2281020693>. Acesso em: 1 jun. 2025.

LECUN, Y.; BENGIO, Y.; HINTON, G. **Deep learning**. Nature, Nature Publishing Group, v. 521, n. 7553, p. 436–444. 2015, Disponível em: <https://www.nature.com/articles/nature14539>. Acesso em: 1 jun. 2025.

LITJENS, G. et al. **A survey on deep learning in medical image analysis**. Medical Image Analysis, Elsevier, v. 42, p. 60–88. ISSN 1361-8415. 2017, Disponível em: <https://www.sciencedirect.com/science/article/pii/S1361841517301135>. Acesso em: 8 jun. 2025.

MCCOLLOUGH, C. et al. **Low Dose CT Image and Projection Data (LDCT-and-Projection-data)**. 2020. The Cancer Imaging Archive. Disponível em: <https://doi.org/10.7937/9npb-2637>. Acesso em: 12 jun. 2025.

MOEN, T. R. et al. **Low-dose CT image and projection dataset**. Medical Physics, v. 48, n. 2, p. 902–911. 2021, Disponível em: <https://doi.org/10.1002/mp.14594>. Acesso em: 8 jun. 2025.

RODRIGUES, A. F.; VITRAL, R. W. F. **Aplicações da tomografia computadorizada na odontologia**. Pesquisa brasileira em odontopediatria e clinica integrada, Universidade Estadual da Paraíba, v. 7, n. 3, p. 317–324. 2007. Acesso em: 1 jun. 2025.

SINYU, S. **RED_CNN: PyTorch implementation of Low-Dose CT with a Residual Encoder-Decoder Convolutional Neural Network**. 2018. GitHub. Disponível em: <https://github.com/SSinyu/RED_CNN>. Acesso em: 12 jun. 2025.

WANG, Z. et al. **Image quality assessment: from error visibility to structural similarity**. IEEE Transactions on Image Processing, v. 13, n. 4, p. 600–612. 2004, Disponível em: <https://doi.org/10.1109/TIP.2003.819861>. Acesso em: 8 jun. 2025.

WILLEMINK, M. J.; NOËL, P. B. **The evolution of image reconstruction for ct—from filtered back projection to artificial intelligence**. European Radiology, Springer, v. 29, n. 5, p. 2185–2195. ISSN 1432-1084. 2019, Disponível em: <https://doi.org/10.1007/s00330-018-5810-7>. Acesso em: 8 jun. 2025.

YAN, S. **DnCNN-PyTorch: PyTorch implementation of Beyond a Gaussian Denoiser**. 2018. GitHub. Disponível em: <https://github.com/SaoYan/DnCNN-PyTorch>. Acesso em: 12 jun. 2025.

YANG, Q. et al. **Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss**. IEEE Transactions on Medical Imaging, IEEE, v. 37, n. 6, p. 1348–1357. ISSN 0278-0062. 2018, Disponível em: <https://doi.org/10.1109/TMI.2018.2827462>. Acesso em: 8 jun. 2025.

ZHANG, K. et al. **Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising**. arXiv preprint arXiv:1608.03981. 2016, Disponível em: <https://arxiv.org/abs/1608.03981>. Acesso em: 1 jun. 2025.

ZHAO, H. et al. **Low dose ct image denoising: A comparative study of deep learning models and training strategies**. AI Medicine, Scilight, v. 1, n. 1, p. 100005. 2024, Disponível em: <https://www.researchgate.net/publication/385580447_Low_Dose_CT_Image_Denoising_A_Comparative_Study_of_Deep_Learning_Models_and_Training_Strategies>. Acesso em: 8 jun. 2025.