

UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ENGENHARIA DE SÃO CARLOS

Pedro Cavallini

**Estudo Comparativo de Métodos Baseados em
Modelo e em Dados para Redução de Ruído em
Imagens Mamográficas Digitais**

São Carlos

2025

Pedro Cavallini

**Estudo Comparativo de Métodos Baseados em
Modelo e em Dados para Redução de Ruído em
Imagens Mamográficas Digitais**

Monografia apresentada ao Curso de Engenharia Elétrica com ênfase em Eletrônica, da Escola de Engenharia de São Carlos da Universidade de São Paulo, como parte dos requisitos para obtenção do Título de Engenheiro(a) Eletricista.

Orientador: Prof. Dr. Marcelo Andrade da Costa Vieira

**São Carlos
2025**

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTE TRABALHO,
POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS
DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pela Biblioteca Prof. Dr. Sérgio Rodrigues Fontes da
EESC/USP com os dados inseridos pelo(a) autor(a).

C377e Cavallini, Pedro
Estudo comparativo de métodos baseados em modelo e
em dados para redução de ruído em imagens mamográficas
digitais / Pedro Cavallini; orientador Marcelo Andrade
da Costa Vieira. São Carlos, 2025.

Monografia (Graduação em Engenharia Elétrica com
ênfase em Eletrônica) -- Escola de Engenharia de São
Carlos da Universidade de São Paulo, 2025.

1. Mamografia Digital. 2. Redução de Ruído. 3.
Métodos Baseados em Modelo. 4. Aprendizado Profundo. 5.
Auto-Supervisionado. 6. Qualidade de Imagem. 7.
Avaliação de Imagem Médica. I. Título.

FOLHA DE APROVAÇÃO

Nome: Pedro Cavallini

Título: "Estudo Comparativo de Métodos Baseados em Modelo e em Dados para Redução de Ruído em Imagens Mamográficas Digitais"

Trabalho de Conclusão de Curso defendido e aprovado
em 08/12/2025,

com NOTA 9,0 (Nove, Zero)), pela Comissão
Julgadora:

Prof. Associado Marcelo Andrade da Costa Vieira - Orientador
SEL/EESC/USP

Prof. Dr. Ricardo Augusto Souza Fernandes - SEL/EESC/USP

Mestre Renann de Faria Brandão - Doutorando EESC/USP

Coordenador da CoC-Engenharia Elétrica - EESC/USP:
Professor Associado José Carlos de Melo Vieira Júnior

RESUMO

CAVALLINI, P. **Estudo Comparativo de Métodos Baseados em Modelo e em Dados para Redução de Ruído em Imagens Mamográficas Digitais.** 2025. 57 p. Monografia (Trabalho de Conclusão de Curso) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2025.

Este trabalho apresenta um estudo comparativo de métodos baseados em modelo (*model-based*) e métodos baseados em dados (*data-based*) para redução de ruído em imagens de mamografia digital de campo total (FFDM). A motivação central da pesquisa reside no dilema fundamental em imagens médicas adquiridas por raios-X: a qualidade da imagem é intrinsecamente vinculada à dose de radiação, tornando crítico o desenvolvimento de técnicas de filtragem de ruído que permitam otimizar protocolos de baixa dose sem comprometer a detectabilidade de estruturas anatômicas, como microcalcificações. O trabalho implementou e avaliou sistematicamente métodos analíticos — filtro Gaussiano e BM3D — bem como métodos de aprendizado profundo supervisionados e auto-supervisionados — Noise2Clean, Noise2Void e Noise2Sim — utilizando duas bases de dados: imagens sintéticas geradas pelos ensaios clínicos virtuais (VICTRE) e imagens clínicas reais de mamografia. Os resultados demonstraram que o BM3D alcançou o melhor desempenho geral (MNSE: 2,419%), apresentando excelente balanço entre supressão de ruído e preservação de estruturas (\mathcal{R}_N : 0,881%; \mathcal{B}^2 : 1,539%). Entre os métodos de aprendizado profundo, o Noise2Sim destacou-se por apresentar o menor viés quadrático entre os métodos *data-based* (\mathcal{B}^2 : 2,357%), preservando melhor as estruturas de alta frequência críticas para o diagnóstico, apesar de maior ruído residual. O Noise2Void alcançou redução eficiente de ruído (\mathcal{R}_N : 0,613%) mas apresentou borramento excessivo (\mathcal{B}^2 : 5,369%), limitação reconhecida do método. Uma observação crítica foi que métricas tradicionais como PSNR e SSIM priorizam redução de ruído sem adequadamente ponderar o borramento, tornando a decomposição do MNSE (*Mean Normalized Squared Error*) em componentes de ruído residual (\mathcal{R}_N) e viés quadrático (\mathcal{B}^2) mais informativa para avaliação em contextos médicos. Conclui-se que, embora métodos *model-based* como BM3D permaneçam como referência quando os parâmetros de aquisição são bem caracterizados, os métodos *data-based* auto-supervisionados, particularmente o Noise2Sim, demonstram potencial significativo para aplicações clínicas em cenários com múltiplos equipamentos e protocolos heterogêneos, abrindo caminho para otimização de dose mantendo qualidade diagnóstica em mamografia digital.

Palavras-chave: Mamografia Digital; Redução de Ruído; Métodos Baseados em Modelo; Aprendizado Profundo; Auto-Supervisionado; Qualidade de Imagem; Avaliação de Imagem Médica.

ABSTRACT

CAVALLINI, P. **Comparative Study of Model-Based and Data-Based Methods for Noise Reduction in Digital Mammographic Images.** 2025. 57 p. Monograph (Course Conclusion Paper) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2025.

This work presents a comparative study of model-based and data-based methods for noise reduction in full-field digital mammography (FFDM) images. The central motivation of the research lies in the fundamental dilemma in medical imaging acquired by X-rays: image quality is intrinsically linked to radiation dose, making it critical to develop filtering techniques that allow optimization of low-dose protocols without compromising the detectability of anatomical structures, such as microcalcifications. The work systematically implemented and evaluated analytical methods — Gaussian filter and BM3D — as well as supervised and self-supervised deep learning methods — Noise2Clean, Noise2Void, and Noise2Sim — using two databases: synthetic images generated by virtual clinical trials (VICTRE) and real clinical mammography images. Results demonstrated that BM3D achieved the best overall performance (MNSE: 2.419%), presenting excellent balance between noise suppression and structure preservation (\mathcal{R}_N : 0.881%; \mathcal{B}^2 : 1.539%). Among deep learning methods, Noise2Sim stood out by presenting the lowest squared bias among data-based methods (\mathcal{B}^2 : 2.357%), better preserving high-frequency structures critical for diagnosis, despite higher residual noise. Noise2Void achieved efficient noise reduction (\mathcal{R}_N : 0.613%) but exhibited excessive blurring (\mathcal{B}^2 : 5.369%), a recognized limitation of the method. A critical observation was that traditional metrics such as PSNR and SSIM prioritize noise reduction without adequately weighting blurring, making the decomposition of MNSE (Mean Normalized Squared Error) into residual noise (\mathcal{R}_N) and squared bias (\mathcal{B}^2) components more informative for evaluation in medical contexts. It is concluded that while model-based methods like BM3D remain as reference when acquisition parameters are well characterized, data-based self-supervised methods, particularly Noise2Sim, demonstrate significant potential for clinical applications in scenarios with multiple heterogeneous equipment and protocols, opening avenues for dose optimization while maintaining diagnostic quality in digital mammography.

Keywords: Digital Mammography; Noise Reduction; Model-Based Methods; Deep Learning; Self-Supervised Learning; Image Quality; Medical Image Assessment.

LISTA DE FIGURAS

Figura 1 – Fluxograma do algoritmo BM3D proposto por Dabov <i>et al.</i> (2007).	21
Figura 2 – Fluxo de processamento do pipeline de <i>denoising</i> de Borges <i>et al.</i> (2017) aplicado às imagens de mamografia digital.	21
Figura 3 – Esquema de mascaramento do pixel durante treinamento do Noise2Void. (a) uma imagem ruidosa; (b) <i>patch</i> extraído da imagem, mostrando o mascaramento do pixel central, selecionando um pixel aleatório da vizinhança (quadrado azul). (c) <i>target</i> correspondente do patch de (b). Imagen retirada de Krull, Buchholz e Jug (2019).	24
Figura 4 – Processo de construção dos k <i>patches</i> mais similares (Step-1) e construção dos pares de treinamento (Step-2). Imagem retirada de Niu <i>et al.</i> (2022).	25
Figura 5 – Exemplos de imagens VCT <i>raw</i> geradas pelo VICTRE.	30
Figura 6 – Exemplos de imagens clínicas <i>raw</i> nas projeções CC e MLO.	30
Figura 7 – Comparação das curvas de perda de treinamento e validação para os três métodos de <i>denoising</i> ao longo das épocas.	35
Figura 8 – Imagem VCT <i>raw</i> com o ROI utilizado na Figura 9.	36
Figura 9 – Comparação visual dos resultados de <i>denoising</i> de diferentes métodos em relação à imagem ruidosa e ao <i>ground truth</i> (imagem limpa sintética). .	37
Figura 10 – Comparação das Métricas de Qualidade de Imagem (SSIM, QILV e HaarPSI).	38
Figura 11 – Comparação da métrica PSNR para os Métodos de Remoção de Ruído. .	38
Figura 12 – Comparação do <i>denoising</i> do filtro Gaussiano para diferentes tamanhos de <i>kernel</i>	39
Figura 13 – Comparação das Métricas de Qualidade de Imagem (SSIM, QILV e HaarPSI) para diferentes tamanhos de Kernel do Filtro Gaussiano. . .	40
Figura 14 – Comparação da métrica PSNR para diferentes tamanhos de Kernel do Filtro Gaussiano.	40
Figura 15 – Comparação da métrica MNSE e suas decomposições $\mathcal{R}_{\mathcal{N}}$ e \mathcal{B}^2 para os métodos de remoção de ruído. Valores em percentual (%).	41
Figura 16 – Imagem Clínica <i>raw</i> com o ROI utilizado na Figura 17.	43
Figura 17 – Comparação visual dos resultados de <i>denoising</i> de diferentes métodos para ROI de imagem clínica.	44
Figura 18 – Comparação da métrica NAQI para os métodos de remoção de ruído. Valores mais altos indicam melhor qualidade percebida.	45
Figura 19 – Imagem VCT <i>raw</i> com o ROI de transição utilizado na Figura 21. . .	45
Figura 20 – Imagem Clínica <i>raw</i> com o ROI de transição utilizado na Figura 22. .	46

Figura 21 – Comparação visual dos resultados de <i>denoising</i> do ROI de transição para diferentes métodos em relação à imagem ruidosa e ao <i>ground truth</i> (imagem limpa sintética).	47
Figura 22 – Comparação visual dos resultados de <i>denoising</i> de diferentes métodos para ROI de transição de imagem clínica.	48
Figura 23 – Distribuição da métrica SSIM pela região da mama para os métodos BM3D com VST e Filtro Gaussiano 5×5	48

LISTA DE TABELAS

Tabela 1 – Métricas de Desempenho de Métodos de Remoção de Ruído	38
Tabela 2 – Métricas de Desempenho de Filtros Gaussianos com <i>Kernels</i> de Diferentes Tamanhos.	40
Tabela 3 – MNSE, $\mathcal{R}_{\mathcal{N}}$ e \mathcal{B}^2 dos Métodos de Remoção de Ruído.	41
Tabela 4 – NAQI dos Métodos de Remoção de Ruído.	44

LISTA DE ABREVIATURAS E SIGLAS

3D	3-Dimensões
AWGN	Additive White Gaussian Noise
BM3D	Block-Matching and 3D Filtering
BSD68	Berkeley Segmentation Dataset 68
CC	Craniocaudal
CNN	Convolutional Neural Network
cryo-TEM	Cryogenic Transmission Electron Microscopy
DB	Data-Based
DBT	Digital Breast Tomosynthesis
DCT	Discrete Cosine Transform
DICOM	Digital Imaging and Communications in Medicine
DL	Deep Learning
DM	Digital Mammography
DnCNN	Deep Denoising CNN
DWT	Discrete Wavelet Transform
FFDM	Full-Field Digital Mammography
FR-IQA	Full Reference Image Quality Assessment
GPU	Graphics Processing Unit
HaarPSI	Haar Wavelet-based Perceptual Similarity Index
IQA	Image Quality Assessment
LDCT	Low-Dose Computed Tomography
MB	Model-Based
MLO	Médio Lateral Oblíqua
MNSE	Mean Normalized Squared Error

MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
MTF	Modulation Transfer Function
N2G	Noise2Grad
N2N	Noise2Noise
N2Sim	Noise2Sim
N2V	Noise2Void
NAQI	Normalized Anisotropic Quality Index
NLM	Non-Local Means
N-RMSE	Normalized Root Mean Squared Error
NR-IQA	No Reference Image Quality Assessment
PCCT	Photon-Counting Computed Tomography
PSNR	Peak Signal-to-Noise Ratio
QILV	Quality Index based on Local Variance
RED-CNN	Residual Encoder-Decoder CNN
ReLU	Rectified Linear Unit
ROI	Region of Interest
SNR	Signal-to-Noise Ratio
SSIM	Structural Similarity Index Measure
UPS	Uniform Pixel Selection
VCT	Virtual Clinical Trials
VICTRE	Virtual Imaging Clinical Trial for Regulatory Evaluation
VST	Variance Stabilizing Transformation
ZCD	Zero Conditional Discrepancy Mean
ZCN	Zero Conditional Noise Mean

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Objetivo	18
2	REVISÃO BIBLIOGRÁFICA	19
2.1	Métodos Baseados em Modelo	19
2.2	Métodos Baseados em Dados	21
2.3	Métricas de Avaliação da Qualidade de Imagem	26
3	MATERIAIS & MÉTODOS	29
3.1	Base de Dados	29
3.2	Métodos Baseados em Modelo	31
3.3	Métodos Baseados em Dados	31
3.3.1	Arquitetura e Hiperparâmetros	32
3.3.2	Parâmetros Específicos dos Métodos Auto-Supervisionados	33
3.4	Avaliação dos Métodos	33
4	RESULTADOS & DISCUSSÕES	35
5	LIMITAÇÕES DO TRABALHO	49
6	TRABALHOS FUTUROS	51
7	CONCLUSÃO	53
	REFERÊNCIAS	55

1 INTRODUÇÃO

O câncer de mama representa uma das principais causas de mortalidade entre mulheres em todo o mundo (WHO, 2025), constituindo-se como um desafio significativo para a saúde pública global. A detecção precoce deste tipo de câncer é fundamental para aumentar as chances de sucesso do tratamento, identificando a doença ainda em estágios iniciais e localizados. Além disso, a participação em programas de rastreamento mamográfico tem demonstrado reduzir significativamente a mortalidade relacionada ao câncer de mama, com taxas de sobrevivência em pacientes participantes variando entre 80% e 86%, comparadas a 53-74% em pacientes não participantes (Tabár *et al.*, 2021).

Neste contexto, a mamografia digital de campo total (FFDM - *Full-Field Digital Mammography*) desempenha um papel fundamental como um dos métodos mais comuns para o rastreamento e diagnóstico precoce do câncer de mama (Michell; Batohi, 2018). A qualidade das imagens mamográficas é essencial para a identificação de lesões sutis, especialmente microcalcificações, que podem indicar a presença de neoplasias em estágios iniciais (Borges *et al.*, 2018a). Contudo, a qualidade da imagem em mamografia está intrinsecamente relacionada à dose de radiação ionizante administrada à paciente durante o exame (Yaffe, 2000).

A aquisição de imagens médicas por raios-X, incluindo a mamografia, envolve um compromisso fundamental entre a qualidade da imagem obtida e a dose de radiação aplicada ao paciente. A natureza quântica dos raios-X resulta em um tipo específico de ruído conhecido como ruído quântico, que é inerente aos sistemas de imagem baseados em integradores de energia. Este ruído surge devido às flutuações estatísticas na quantidade de fôtons detectados e segue uma distribuição de Poisson, onde a variância do ruído é dependente do sinal (Bushberg *et al.*, 2012).

A relação entre dose de radiação e ruído na imagem é inversamente proporcional à raiz quadrada do número de fôtons detectados. Consequentemente, para obter imagens com menor nível de ruído e maior relação sinal-ruído (SNR - *Signal-to-Noise Ratio*), é necessário aumentar a dose de radiação (Bushberg *et al.*, 2012). Por outro lado, a redução da dose resulta em maior magnitude de ruído, o que pode comprometer a visualização de estruturas anatômicas e lesões, potencialmente afetando negativamente o desempenho diagnóstico (Borges *et al.*, 2018a).

Nesse cenário, métodos de filtragem de ruído assumem papel central ao buscar melhorar a qualidade das imagens sem necessidade de aumento da dose de radiação. Técnicas de *denoising* adequadamente projetadas podem atenuar o ruído quântico e eletrônico preservando, tanto quanto possível, bordas e detalhes anatômicos relevantes,

contribuindo para manter a performance diagnóstica mesmo em condições de exposição reduzida (Borges *et al.*, 2018b). Essa estratégia é particularmente atraente em contextos de otimização de protocolos de baixa dose, nos quais se pretende minimizar o risco radiogênico sem comprometer a capacidade de detecção de câncer. Entretanto, uma preocupação latente no contexto de restauração de imagens por meio de técnicas de filtragem de ruído é a preservação das estruturas originais e da nitidez da imagem, que podem ter impactos substanciais na detecção de microcalcificações e lesões mamárias (Soares, 2025).

Na literatura de processamento de imagens, o conceito de restauração é abrangente, referindo-se ao processo de recuperar uma imagem original a partir de uma versão degradada, o que pode incluir a reversão de borramento (*deblurring*), distorções geométricas e outros artefatos. A filtragem de ruído (*denoising*), por sua vez, é um subconjunto específico da restauração, focado exclusivamente na mitigação de flutuações estocásticas do sinal. Para fins de delimitação deste trabalho, adota-se o termo 'restauração' em seu sentido estrito de filtragem de ruído; portanto, todas as referências à restauração de imagens ao longo do texto dizem respeito especificamente às técnicas e processos aplicados para a supressão de ruído nas imagens mamográficas.

Historicamente, a atenuação de ruído em imagens médicas baseou-se em filtros analíticos, ou métodos baseados em modelo (MB - *Model-Based*), que incorporam modelos explícitos do sinal e do ruído. Dentre as abordagens clássicas destacam-se filtros lineares de suavização, como o filtro Gaussiano, filtros adaptativos no sentido de mínima variância, como o filtro de Wiener, e métodos multiescala baseados em transformada wavelet, que exploram a esparsidade do sinal em domínios transformados para separar componentes de ruído e de estrutura (Gonzalez; Woods, 2018). Mais recentemente, algoritmos não locais de última geração, como o BM3D (Block-Matching and 3D Filtering), têm sido amplamente empregados como referência em denoising, utilizando busca de blocos similares e filtragem colaborativa no domínio de transformadas 3D para alcançar elevado grau de supressão de ruído com preservação de detalhes finos (Dabov *et al.*, 2007).

Os filtros analíticos baseados em modelos apresentam, em geral, resultados robustos e previsíveis, especialmente quando os parâmetros que caracterizam o sistema de aquisição e o ruído são bem conhecidos. Entretanto, essa dependência de conhecimento prévio detalhado do domínio limita sua aplicabilidade em cenários mais heterogêneos. Em prática clínica, aspectos como modelo e fabricante do equipamento, protocolo de exposição, faixa de espessuras mamárias e variações institucionais tornam complexa a tarefa de parametrizar manualmente um método analítico de forma ideal para diferentes máquinas e tipos de exame. Consequentemente, embora extremamente úteis em ambientes bem controlados, esses métodos podem carecer da flexibilidade necessária quando se busca uma solução mais generalista, que mantenha desempenho consistente em múltiplos cenários de aquisição (Xia *et al.*, 2023).

Em paralelo a essas abordagens, o avanço da capacidade computacional e a crescente disponibilidade de grandes volumes de dados clínicos e sintéticos impulsionaram o desenvolvimento de métodos baseados em dados (DB - *Data-Based*) para restauração de imagens médicas. Esses métodos utilizam os próprios dados para aprender representações internas do sinal e do ruído, reduzindo a necessidade de modelagem analítica explícita e possibilitando que a estrutura estatística seja capturada pela própria técnica de aprendizado (Xia *et al.*, 2023). Nesse contexto, redes neurais profundas (*deep learning*), especialmente arquiteturas convolucionais, têm demonstrado grande capacidade de aprender relações complexas entre imagens ruidosas e suas representações limpas, generalizando esse aprendizado para novas amostras que não foram vistas durante o treinamento (Lee *et al.*, 2017; Sarvamangala; Kulkarni, 2021).

A abordagem clássica de *deep learning* para restauração de imagens é o aprendizado supervisionado, em que a rede é treinada com pares de imagens limpas e ruidosas correspondentes. Nesses métodos, o modelo aprende um mapeamento direto do espaço de imagens degradadas para o espaço de imagens de referência, minimizando uma função de perda definida entre a saída predita e um *ground-truth* assumidamente livre de ruído. No entanto, em domínios como mamografia digital, a obtenção de imagens verdadeiramente limpas é inviável ou até mesmo impossível, pois o próprio processo de aquisição sempre envolve degradações pelo ruído quântico e eletrônico (Bushberg *et al.*, 2012).

Uma solução amplamente adotada para contornar essa limitação é o uso de imagens sintéticas geradas por ensaios clínicos virtuais (VCT – *Virtual Clinical Trials*). Por meio de *phantoms* antropomórficos digitais e simulações virtuais, é possível produzir pares de imagens limpas e com ruído controlado, ajustando parâmetros de dose, espectro de raios X e características do detector de forma precisa. Esses dados sintéticos fornecem um *ground-truth* idealizado para treinar redes supervisionadas em condições bem definidas, incluindo cenários que seriam impraticáveis em estudos clínicos reais, como variações sistemáticas de dose e densidade mamária (Sharma *et al.*, 2019; Borges *et al.*, 2019).

Apesar das vantagens, o uso de imagens sintéticas introduz outro desafio: discrepâncias entre os domínios sintético e clínico podem prejudicar a generalização dos modelos treinados. Diferenças sutis – ou mesmo substanciais – na textura do tecido, distribuição de ruído, presença de artefatos e variações anatômicas não capturadas pelos *phantoms* podem levar a queda de desempenho quando o modelo é aplicado em dados reais. Nesse cenário, surgem os métodos *data-based* auto-supervisionados, que dispensam a necessidade de pares limpo-ruidoso e aprendem diretamente a partir das próprias imagens clínicas ruidosas, explorando hipóteses sobre independência do ruído e redundância estrutural para guiar o processo de aprendizado (Lehtinen *et al.*, 2018; Krull; Buchholz; Jug, 2019; Niu *et al.*, 2022; Sander; Pock; Cremers, 2021).

Inicialmente, o Noise2Noise introduziu a possibilidade de treinar redes neurais para

restauração utilizando apenas pares de realizações da imagem com ruídos independentes (Lehtinen *et al.*, 2018). Posteriormente, avançaram-se abordagens ainda mais flexíveis, como o Noise2Void (Krull; Buchholz; Jug, 2019), que utiliza redes com regiões cegas (*blind-spot*) para predizer o valor de um pixel a partir dos pixels vizinhos na própria imagem ruidosa, e o Noise2Sim (Niu *et al.*, 2022), que explora a similaridade local entre regiões da imagem para construir pares de treinamento a partir de uma única imagem ruidosa. Essas técnicas ampliam a aplicabilidade do *deep learning* em contextos clínicos reais, nos quais a obtenção de imagens de referência é inviável, favorecendo a generalização e a robustez dos modelos para diferentes equipamentos, doses e características dos exames.

1.1 Objetivo

Diante desse panorama, o presente trabalho tem como objetivo realizar um estudo comparativo, no contexto específico da mamografia digital de campo total (FFDM), sobre os métodos de filtragem de ruído, considerando o equilíbrio entre desempenho, robustez e aplicabilidade clínica: métodos *model-based* analíticos (como filtros clássicos e BM3D) e métodos *data-based* baseados em aprendizado profundo, incluindo tanto estratégias supervisionadas (treinadas com dados sintéticos de VCT) quanto auto-supervisionadas (treinadas diretamente em imagens clínicas ruidosas). Ao comparar sistematicamente essas abordagens em termos de métricas de qualidade de imagem e potencial impacto diagnóstico, busca-se fornecer evidências que auxiliem na escolha de estratégias de processamento mais adequadas para otimização de dose e melhoria da qualidade de imagens de mamografia digital.

2 REVISÃO BIBLIOGRÁFICA

A restauração de imagens médicas é uma etapa técnica fundamental para potencializar o valor diagnóstico dos exames, promovendo a correção de artefatos, degradação de sinal e ruído oriundos do próprio processo de aquisição (Maia *et al.*, 2019). O objetivo do processamento de restauração é recuperar a imagem o mais próximo possível da representação anatômica verdadeira, o que se mostra particularmente relevante no contexto da mamografia digital, devido ao impacto direto da qualidade de imagem na acurácia da detecção e caracterização de lesões (Bushberg *et al.*, 2012). Técnicas de restauração são historicamente aplicadas como pré-processamento ou pós-processamento para corrigir pixels defeituosos, compensar perdas ou distorções, aumentar o contraste e, principalmente, suprimir ruído sem comprometer a preservação de detalhes de interesse clínico (Maia *et al.*, 2019).

A restauração de imagens mamográficas degradadas — especialmente aquelas obtidas sob protocolos de baixa dose para redução do risco radiogênico — tem relação direta e mensurável com o desempenho diagnóstico, impactando tanto a percepção visual do radiologista quanto a acurácia de tarefas específicas de detecção. Borges *et al.* (2018a) conduziram um estudo perceptivo que avaliou a influência da restauração de imagens de tomossíntese mamária adquiridas em baixas doses, demonstrando que a aplicação de técnicas avançadas de restauração possibilitou a recuperação do contraste e de características anatômicas relevantes sem introdução de artefatos significativos, resultando em melhora substancial da performance dos observadores humanos. De forma complementar, Borges *et al.* (2018b) apresentaram evidências quantitativas de que métodos de restauração, ao suprimir o ruído de forma controlada, restauram a visibilidade de lesões e detalhes finos, e contribuem para elevar a qualidade subjetiva e objetiva das imagens, reforçando o potencial dessas técnicas como ferramentas de suporte ao diagnóstico em ambientes clínicos sistematicamente expostos à necessidade de otimização da dose.

As abordagens para restauração podem ser agrupadas em duas grandes vertentes: métodos baseados em modelos (*model-based* - MB) e métodos baseados em dados (*data-based* - DB).

2.1 Métodos Baseados em Modelo

Os métodos MB partem da premissa de modelar matematicamente a relação entre o sinal ideal e a imagem degradada, incorporando hipóteses sobre o sistema de aquisição, propriedades físicas, estatísticas do ruído e características anatômicas. Entre estes, destacam-se os seguintes filtros:

- Filtros Lineares: O filtro Gaussiano, que realiza uma média ponderada dos pixels vizinhos utilizando uma função gaussiana como kernel de convolução, é um dos métodos mais simples e computacionalmente eficientes. Embora efetivo na redução de ruído, este filtro tende a suavizar também as bordas e estruturas finas da imagem, resultando em perda de nitidez. O filtro de média opera de forma similar, mas utilizando pesos uniformes (Gonzalez; Woods, 2018).
- Filtro de Wiener: Diferentemente dos filtros lineares convencionais, o filtro de Wiener adapta-se à variância local da imagem, realizando maior suavização em regiões homogêneas (baixa variância) e preservando melhor as regiões de alta variância, correspondentes a bordas e estruturas. Este filtro é particularmente efetivo quando o ruído pode ser modelado como ruído aditivo branco gaussiano (AWGN - *Additive White Gaussian Noise*) e minimiza o erro quadrático médio (MSE - *Mean Squared Error*) entre a imagem estimada e a imagem verdadeira (Gonzalez; Woods, 2018).
- Filtro de Mediana: Pertencente à classe de filtros não-lineares baseados em ordem estatística, o filtro de mediana substitui cada pixel pelo valor mediano de sua vizinhança. É especialmente eficaz na remoção de ruído impulsivo (*salt-and-pepper noise*), preservando razoavelmente as bordas da imagem (Gonzalez; Woods, 2018).
- Transformada Wavelet: Métodos baseados em transformada *wavelet* exploram a representação multiescala dos sinais, decompondo a imagem em diferentes níveis de resolução e orientação. A atenuação de ruído é realizada através de técnicas de limiarização dos coeficientes *wavelet*, permitindo separar componentes de sinal de componentes de ruído. A transformada *wavelet* discreta (DWT - *Discrete Wavelet Transform*) tem sido amplamente utilizada em processamento de imagens médicas devido à sua capacidade de representação esparsa e preservação de bordas (Gonzalez; Woods, 2018).
- *Block-Matching and 3D Filtering* (BM3D): O algoritmo BM3D (Dabov *et al.*, 2007) representa um dos métodos estado-da-arte em atenuação de ruído analítico (Goyal *et al.*, 2020). Este método combina busca por blocos similares com filtragem colaborativa no domínio transformado em dois estágios principais (conforme Figura 1): (1) *matching e filtragem básica* — identifica blocos similares através de busca não-local, agrupa-os em estruturas 3D e aplica transformadas decorrelacionantes (wavelet ou DCT) seguidas de encolhimento colaborativo; (2) *filtragem refinada e agregação* — refina a estimativa inicial através de filtragem de Wiener e combina os blocos processados via média ponderada. O BM3D explora efetivamente a redundância não-local presente em imagens, alcançando excelente balanço entre supressão de ruído e preservação de estruturas finas, sendo frequentemente utilizado como método de referência para comparação de novos algoritmos de *denoising*.

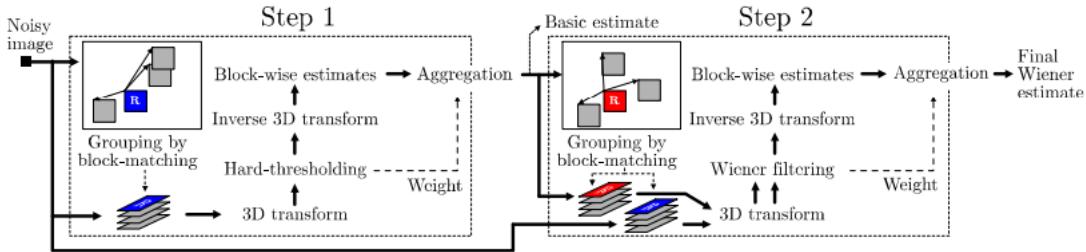


Figura 1 – Fluxograma do algoritmo BM3D proposto por Dabov *et al.* (2007).

Além dos filtros clássicos e do BM3D em sua versão tradicional, destaca-se na literatura o pipeline proposto por Borges *et al.* (2017), representado na Figura 2, desenvolvido especificamente para otimizar o denoising em imagens de mamografia digital e tomosíntese considerando a natureza mista (Gaussiano-Poisson) do ruído nessas modalidades. Esse pipeline promove inicialmente a remoção dos parâmetros de calibração aplicados pelo detector, incluindo ganho espacial e offset, a partir da análise de imagens uniformes adquiridas com o mesmo equipamento clínico. Em seguida, utiliza uma transformação de estabilização de variância (*Variance Stabilizing Transformation* - VST), baseada na transformação de Anscombe generalizada, para converter o ruído sinal-dependente em um ruído aproximadamente Gaussiano de variância constante. Após essa normalização, métodos de *denoising* projetados para ruído Gaussiano independente, como o BM3D, podem ser aplicados com maior eficácia. Por fim, uma etapa reversa de transformação e restauração dos parâmetros de calibração recupera a imagem ao seu domínio original. Borges *et al.* (2017) demonstraram que o pipeline aprimora substancialmente a eficiência do *denoising*, com melhorias de até 20% na N-RMSE espacial e 15% em termos de SNR de frequência, sem perda significativa de nitidez segundo a MTF, tornando-se assim uma solução robusta e adaptável para processamento de imagens em diferentes cenários clínicos de mamografia digital.



Figura 2 – Fluxo de processamento do pipeline de *denoising* de Borges *et al.* (2017) aplicado às imagens de mamografia digital.

2.2 Métodos Baseados em Dados

Com o avanço da capacidade computacional e o crescimento dos volumes de dados disponíveis, métodos baseados em dados (*data-based* - DB) emergiram como alternativa para superar as limitações dos MB. Ao invés de depender de modelagem explícita dos processos físicos, os DB utilizam estatísticas extraídas do próprio conjunto de imagens para aprender representações internas do sinal e do ruído, tornando possível adaptar automaticamente o processamento sem intervenção manual sobre a parametrização (Lehtinen *et al.*, 2018).

O surgimento das redes neurais convolucionais (CNNs) marcou um ponto de inflexão significativo na história da análise de imagens, particularmente após o sucesso apresentado no desafio ImageNet Large Scale Visual Recognition Competition (ILSVRC) em 2012, que catalisou o interesse acadêmico e industrial no aprendizado profundo (Lee *et al.*, 2017). Desde sua introdução nos primeiros anos 2000, as CNNs foram progressivamente adaptadas para aplicações médicas, inicialmente focando em tarefas de segmentação e classificação. Sarvamangala e Kulkarni (2021) destacam que o ponto de virada crítico ocorreu quando o GoogleNet alcançou 89% de acurácia na detecção de câncer em imagens patológicas, superando o desempenho de patologistas humanos.

As aplicações iniciais de CNNs em imagens médicas abrangeram diagnóstico de retinopatia diabética, classificação de lesões de pele, análise de radiografias de tórax, e detecção de tuberculose (Huang *et al.*, 2020). Lee *et al.* (2017) documentam que as primeiras aplicações bem-sucedidas de deep learning em radiologia incluíram segmentação de pulmões, detecção de tumores cerebrais e segmentação de estruturas biológicas. A crescente disponibilidade de dados, aumento da capacidade computacional através de processadores GPU e desenvolvimento de algoritmos mais sofisticados de treinamento criaram o ambiente propício para a revolução digital na medicina (Lee *et al.*, 2017).

A aplicação de redes neurais profundas especificamente para tarefas de remoção de ruído em imagens médicas emergiu como resposta aos desafios de aquisição de imagens com doses reduzidas de radiação. Uma das primeiras propostas influentes foi o RED-CNN (Residual Encoder-Decoder CNN), desenvolvido por Chen *et al.* (2017) para *denoising* de tomografia computadorizada de baixa dose (LDCT). O RED-CNN combinou conceitos de autoencoders, redes de deconvolução e conexões de atalho (*skip connections*), demonstrando capacidade de suprimir efetivamente ruído mantendo preservação de detalhes anatômicos. O trabalho de Chen *et al.* (2017) alcançou melhoria significativa em relações sinal-ruído (SNR) e demonstrou melhor preservação de detalhes comparado aos métodos clássicos como ASD-POCS.

Contemporaneamente, Zhang *et al.* (2016) apresentaram o DnCNN (Deep Denoising CNN), um modelo revolucionário que explorou aprendizado residual para Gaussian denoising. O DnCNN foi notável por sua capacidade de lidar com denoising gaussiano cego (*blind Gaussian denoising*) em níveis de ruído desconhecidos, utilizando uma única rede treinada. O modelo prediz o componente residual (diferença entre imagem ruidosa e limpa) em vez de predizer diretamente a imagem limpa, um paradigma que mostrou-se particularmente efetivo.

Reconhecendo as dificuldades práticas em obter grandes quantidades de pares de imagens limpas e ruidosas em domínios médicos, a utilização de dados sintéticos gerados computacionalmente emergiu como solução viável. Os ensaios clínicos virtuais (VCT - *Virtual Clinical Trials*) representam uma metodologia robusta para geração de dados

sintéticos realísticos. Badano *et al.* (2018) introduziram o VICTRE (*Virtual Imaging Clinical Trial for Regulatory Evaluation*), uma plataforma de código aberto que simula sistemas de mamografia digital (DM) e tomossíntese digital da mama (DBT) utilizando transporte de raios-X por Monte Carlo.

Outra solução que surge, a fim de utilizar dados reais no treinamento de modelos de *Deep Learning* e reconhecendo que métodos supervisionados tradicionais enfrentam limitações práticas significativas em domínios de imagem médica, foram desenvolvidas abordagens que permitem treinamento sem necessidade de pares limpas-ruidosas.

Lehtinen *et al.* (2018) propuseram Noise2Noise (N2N), um paradigma fundamentado em raciocínio estatístico básico. A ideia central é que, se duas realizações independentes de ruído n_1 e n_2 de uma mesma cena limpa estão disponíveis, de modo que $x_1 = s + n_1$ e $x_2 = s + n_2$, uma rede neural pode ser treinada para mapear $x_1 \rightarrow x_2$. Sob a hipótese de ruído com média zero e independente do sinal, Lehtinen *et al.* (2018) demonstraram que treinar com pares ruidosos é estatisticamente equivalente a treinar com dados limpos, a menos de uma constante aditiva relacionada à variância do ruído.

Lehtinen *et al.* (2018) validaram Noise2Noise em múltiplos domínios: remoção de ruído fotográfico, *denoising* de imagens sintéticas de Monte Carlo, e reconstrução de varreduras de MRI subamostradas. Em aplicações fotográficas, o modelo Noise2Noise alcançou desempenho comparável ou superior ao treinamento supervisionado com dados limpos. Particularmente, para imagens sintéticas de Monte Carlo, o método provou ser vantajoso pois elimina a necessidade de renderizações de alta exposição (*clean data*) que são 2000 vezes mais lentas que renderizações de baixa exposição. Contudo, a principal limitação do Noise2Noise em domínios como mamografia digital é que requer múltiplas realizações independentes da mesma cena, o que não é viável em protocolos clínicos onde tipicamente apenas uma única aquisição por projeção é realizada.

Krull, Buchholz e Jug (2019) propuseram Noise2Void (N2V), representando um avanço significativo ao possibilitar treinamento a partir de imagens ruidosas individuais, sem necessidade de múltiplas realizações ou imagens limpas. A estratégia fundamental é a construção de uma "*blind-spot network*", onde a rede é impedida de ter acesso direto ao pixel que está sendo predito. Durante treinamento, pixels específicos da imagem são mascarados (tipicamente através de amostragem uniforme de pixels - UPS), e a rede é treinada para predizer o valor desses pixels mascarados utilizando apenas informações de sua vizinhança local (janela de contexto). A Figura 3 demonstra esse processo de mascaramento do pixel para realizar o treinamento.

A fundamentação teórica de Noise2Void baseia-se em duas suposições estatísticas: (1) o sinal é não pixel-wise independente (possui correlação espacial), e (2) o ruído é condicionalmente pixel-wise independente dado o sinal. Sob estas hipóteses, a rede aprende a predição que corresponde à estimativa de mínimo erro quadrático médio do

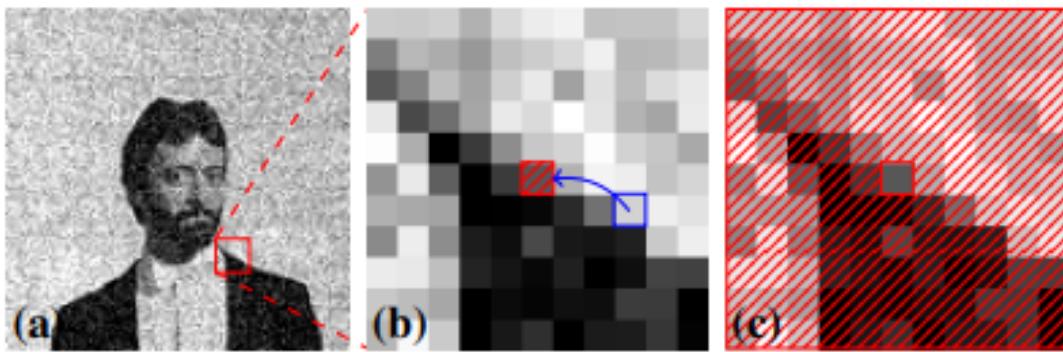


Figura 3 – Esquema de mascaramento do pixel durante treinamento do Noise2Void. (a) uma imagem ruidosa; (b) *patch* extraído da imagem, mostrando o mascaramento do pixel central, selecionando um pixel aleatório da vizinhança (quadrado azul). (c) *target* correspondente do patch de (b). Imagem retirada de Krull, Buchholz e Jug (2019).

signal verdadeiro.

Krull, Buchholz e Jug (2019) demonstraram experimentalmente que Noise2Void alcança desempenho comparável a métodos supervisionados em domínios como microscopia de fluorescência, imagens naturais (BSD68) e dados de microscopia de transmissão de elétrons (cryo-TEM). Contudo, o Noise2Void apresenta limitações reconhecidas: (1) tendência ao borramento excessivo, afetando preservação de detalhes finos; (2) surgimento de artefatos em padrão de tabuleiro (*checkerboard artifacts*) devido à estratégia de mascaramento de pixels; (3) dependência exclusiva de vizinhança local imediata, que pode ser insuficiente para recuperar estruturas de maior escala.

Sander, Pock e Cremers (2021) propuseram Noise2Grad (N2G), um método que baseia sua estratégia de treinamento na observação de que o gradiente de uma imagem ruidosa é dominado por componentes de ruído. O algoritmo opera em duas etapas: (1) um módulo atua removendo aproximadamente o ruído; (2) um módulo de aproximação de ruído extrai componentes de ruído através de síntese de novos pares ruidosos-limos. Esses pares sintetizados retroalimentam o módulo de remoção de ruído para guiar melhor remoção.

O core da estratégia de Noise2Grad é usar pares de imagens não-emparelhadas (ruidosa e limpa) para treinar uma rede de extração de ruído, restrito pela informação de gradiente da imagem.

Sander, Pock e Cremers (2021) reportaram que Noise2Grad alcança desempenho próximo aos métodos supervisionados e auto-supervisionados (Noise2Noise) em *datasets* como BSD300 para ruído gaussiano, speckle e Poisson. Contudo, a aplicabilidade de Noise2Grad em mamografia é limitada pela sua dependência de imagens limpas, mesmo que não pareadas.

Niu *et al.* (2022) propuseram Noise2Sim (N2Sim), representando uma abordagem

de *denoising* auto-supervisionado baseada em similaridade. Diferentemente do Noise2Void que explora apenas vizinhanças locais imediatas, o Noise2Sim explora de forma não-local a redundância estrutural em imagens através da busca de *patches* similares.

O algoritmo do Noise2Sim procede em três etapas: (1) para cada *patch* de referência, identifica os k *patches* mais similares na imagem através de busca não-local; (2) constrói pares de treinamento de forma aleatória e independente pixel a pixel a partir dos $k + 1$ *patches* similares; (3) treina a rede de *denoising* para mapear os *inputs* construídos para os *targets* correspondentes. Esse processo é representado na Figura 4.

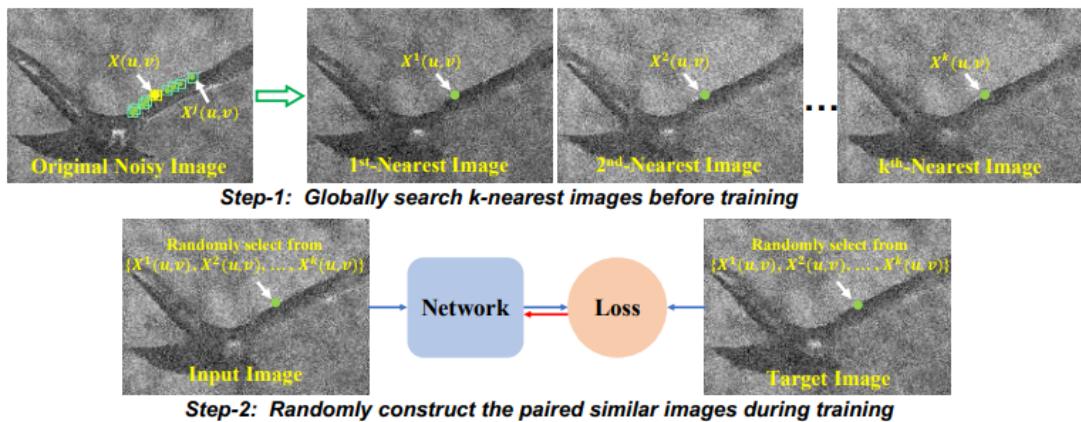


Figura 4 – Processo de construção dos k *patches* mais similares (Step-1) e construção dos pares de treinamento (Step-2). Imagem retirada de Niu *et al.* (2022).

Teoricamente, Niu *et al.* (2022) demonstraram que, sob duas condições brandas – *Zero Conditional Noise mean* (ZCN) e *Zero Conditional Discrepancy mean* (ZCD) – o aprendizado com Noise2Sim é assintoticamente equivalente (no limite do número de amostras de treinamento) ao aprendizado supervisionado com alvos limpos verdadeiros. Diferentemente do Noise2Void que assume independência estrita de ruído entre pixels vizinhos, Noise2Sim é teoricamente capaz de suprimir tanto ruídos independentes quanto ruídos correlacionados.

Niu *et al.* (2022) validaram extensivamente Noise2Sim em imagens de tomografia computadorizada de baixa dose (LDCT) e tomografia computadorizada de contagem de fótons (PCCT), demonstrando desempenho comparável ou superior a métodos supervisionados em métricas PSNR, SSIM e avaliação estatística. As análises visuais e quantitativas confirmaram superior preservação de estruturas anatômicas comparado a BM3D e outros métodos auto-supervisionados. Crucialmente, Noise2Sim opera de forma não-linear através de redes neurais profundas e colaborativa, onde todos os conjuntos de *patches* similares em todas as imagens de treinamento contribuem coletivamente para o aprendizado, uma vantagem significativa sobre métodos lineares como *Non-Local Means* (NLM).

A escolha entre abordagens supervisionadas e auto-supervisionadas em mamografia digital envolve considerações importantes. Métodos supervisionados (Noise2Clean baseado

em dados VICTRE) oferecem a vantagem de um *ground truth* gerado computacionalmente, permitindo otimização direta para mapeamento *noisy-to-clean*. Contudo, enfrentam o desafio de generalização potencial quando há discrepâncias entre características de ruído sintético e ruído real em imagens clínicas.

Métodos auto-supervisionados como Noise2Void e Noise2Sim não requerem dados limpos, permitindo treinamento diretamente em imagens clínicas ruidosas. Essa exposição a ruído real potencialmente melhora robustez, mas introduz desafios em que a rede deve aprender a restauração sem supervisão explícita. Noise2Void, embora revolucionário, apresenta as limitações reconhecidas de possível borramento excessivo e dependência de independência pixel-wise de ruído, violada em imagens com ruído correlacionado. Noise2Sim, por sua vez, explora redundância não-local de forma mais abrangente e tem fundamentação teórica para lidar com ruído correlacionado, potencialmente oferecendo vantagens em contextos de mamografia onde textura e estrutura são críticas.

2.3 Métricas de Avaliação da Qualidade de Imagem

A avaliação da qualidade de imagens médicas, especialmente na mamografia digital, é crucial para monitorar a influência dos algoritmos de restauração, protocolos de aquisição e níveis de dose sobre o desempenho diagnóstico. Dentre as metodologias de *Image Quality Assessment* (IQA), destacam-se as métricas objetivas, classificadas em *full-reference* (FR-IQA), quando existe uma imagem de referência ideal, e *no-reference* (NR-IQA), que não dependem de uma imagem original não degradada.

As métricas *full-reference* são amplamente utilizadas em ensaios clínicos virtuais ou quando o sinal original ideal está disponível. Exemplos incluem:

- PSNR (*Peak Signal-to-Noise Ratio*) – mede a razão entre a potência máxima do sinal e a potência do ruído (descrito pela MSE), sendo uma das métricas mais tradicionais em processamento de imagens. (Wang; Bovik, 2009)
- SSIM (*Structural Similarity Index Measure*) – avalia a similaridade estrutural entre duas imagens considerando aspectos de luminância, contraste e estrutura. Diferentemente do PSNR, o SSIM incorpora características da percepção visual humana, demonstrando maior correlação com a qualidade percebida de imagens. (Wang *et al.*, 2004)
- QILV (*Quality Index based on Local Variance*) – compara a distribuição de variância local entre a imagem restaurada e a referência, sendo sensível a mudanças nas estruturas locais. O QILV é particularmente útil para avaliar distorções estruturais diferentes do simples borramento uniforme. (Aja-Fernandez; Vegas-Sanchez-Ferrero; Alberola-Lopez, 2006)

- HaarPSI (*Haar Wavelet-based Perceptual Similarity Index*) – utiliza decomposição *wavelet* de Haar para avaliar a similaridade perceptual entre imagens, incorporando características da percepção visual humana em múltiplas escalas e orientações (Reisenhofer *et al.*, 2018).
- MNSE (*Mean Normalized Squared Error*) – permite a decomposição entre *Residual Noise* (\mathcal{R}_N) e *Bias* (\mathcal{B}_2), oferecendo uma análise mais completa do *trade-off* entre remoção de ruído e preservação de sinal (Borges *et al.*, 2018b). Esta decomposição permite avaliar separadamente o ruído residual (efetividade da filtragem) e o viés quadrático (borramento ou artefatos introduzidos). O MNSE é especialmente relevante para aplicações em imagens médicas, onde ambos os componentes impactam a qualidade diagnóstica (Vimieiro, 2023).

Por outro lado, em cenários clínicos reais, onde o sinal de referência raramente está disponível, utilizam-se métricas *no-reference*, que avaliam a qualidade exclusivamente com base na imagem avaliada. Dentre essas métricas, destaca-se a NAQI (*Normalized Anisotropic Quality Index*), que se baseia na entropia Rényi anisotrópica da distribuição espaço-frequência para quantificar a organização estrutural e a anisotropia da imagem (Oliveira *et al.*, 2016). Valores mais altos indicam melhor preservação da organização anatômica e qualidade superior; permite rápida avaliação em imagens clínicas sem *ground-truth*.

Limitações importantes emergem na comparação e interpretação dessas métricas, especialmente quando utilizadas fora do domínio de imagens naturais. O trabalho de Soares (2025) destaca que muitos índices tradicionais de distorção apresentam baixa correlação com tarefas diagnósticas clínicas, principalmente devido à divergência entre o compromisso percepção-distorção e a variação estatística do conteúdo anatômico. Métricas como SSIM e PSNR, por exemplo, podem não refletir adequadamente o impacto diagnóstico de ruído ou borramento em imagens médicas, pois foram desenvolvidas e validadas para imagens naturais.

Assim, recomenda-se cautela na interpretação dos resultados das métricas objetivas. Ideais para monitoramento prático e otimização de algoritmos, essas métricas devem ser complementadas por validações baseadas em tarefa ou estudos com observadores humanos/computacionais, principalmente em contextos de avaliação clínica da mamografia digital.

3 MATERIAIS & MÉTODOS

Neste trabalho, foi realizada uma comparação entre métodos *model-based* e *data-based* para supressão de ruído em imagens de mamografia digital, investigando as vantagens e limitações de cada abordagem. A implementação dos métodos de restauração está disponível em https://github.com/Pedroc890/denoising_methods_FFDM.

Para os métodos *model-based*, buscou-se implementar o filtro Gaussiano e o BM3D (*Block-Matching and 3D Filtering*) implementado com uma transformação de estabilização da variância (VST - *Variance-Stabilizing Transformation*), explorando, respectivamente, um modelo de baixa complexidade em comparação com estado-da-arte no campo de *denoising* (Goyal *et al.*, 2020).

Se por um lado, os métodos *model-based* necessitam ajuste de parâmetros que se relacionam com um conhecimento *a priori* das características do domínio trabalhado, as abordagens *data-based* buscam aprender, a partir de um grande volume de dados, as características desse domínio para realizar a tarefa almejada. Dessa forma, explorando os diferentes tipos de conjuntos de dados disponíveis na área de imagens médicas, foi implementado um modelo de *deep learning* supervisionado (Noise2Clean) e dois modelos auto-supervisionados (Noise2Void e Noise2Sim).

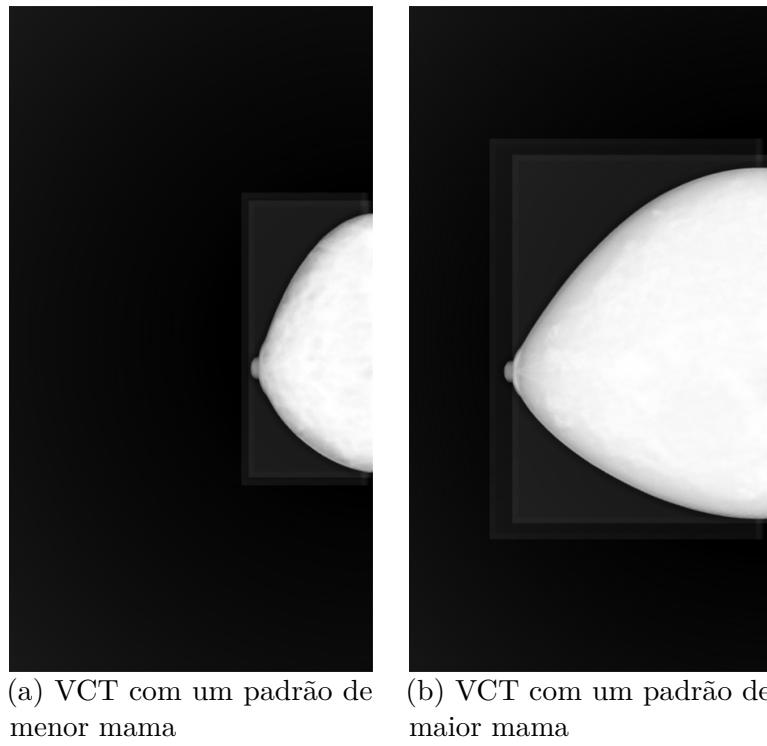
3.1 Base de Dados

Neste trabalho foram utilizadas duas bases principais de imagens de mamografia digital de campo total (FFDM): (i) um conjunto sintético gerado por ensaios clínicos virtuais (VCT), empregado como referência controlada para avaliação e treinamento de métodos supervisionados, e (ii) um conjunto de imagens clínicas de FFDM, utilizado para treinamento dos métodos auto-supervisionados e para avaliação em condições reais de aquisição.

Para a base de dados de imagens sintéticas, imagens VCT de FFDM foram geradas através do *software* de ensaios clínicos virtuais VICTRE (*Virtual Clinical Trial for Regulatory Evaluation*) (Sharma *et al.*, 2019) – utilizadas como *Ground Truth* – e posteriormente inserido ruído quântico e eletrônico para simular a aquisição da imagem utilizando 100% da dose padrão de mamografia. A inserção de ruído foi realizada utilizando parâmetros do equipamento *Selenia Dimensions Mammography System* (Hologic Inc., EUA) e seguindo o modelo de ruído definido em Borges *et al.* (2019). As imagens das Figuras 5a e 5b são exemplos de VCTs geradas por meio do VICTRE, simulando diferentes estruturas associadas ao exame de FFDM.

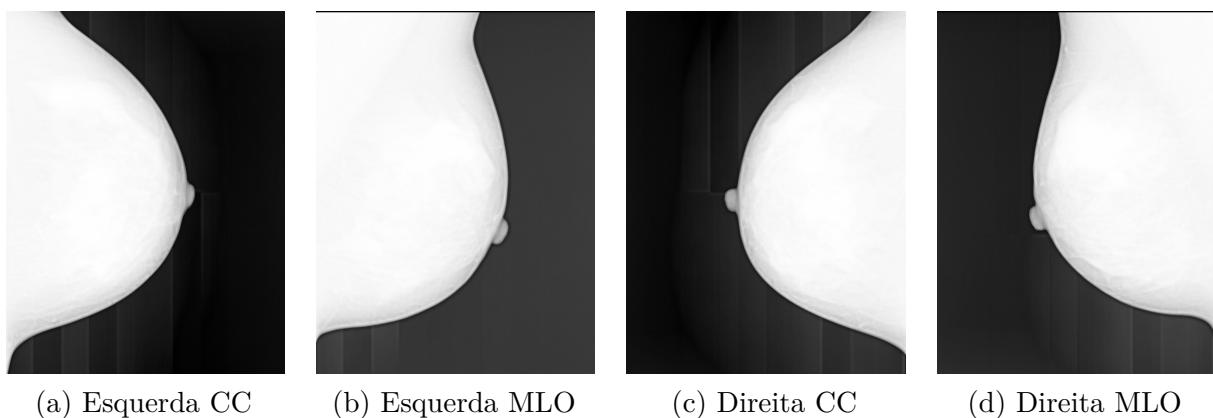
Já para a base de dados de imagens clínicas, foram utilizadas imagens reais de FFDM

Figura 5 – Exemplos de imagens VCT *raw* geradas pelo VICTRE.



do Hospital de Câncer de Barretos (Brasil), contendo tanto as projeções de craniocaudal (CC) e médio lateral oblíqua (MLO) das mamas esquerda e direita, salvas no formato *raw* (ou também DICOM *"for processing"*). Todos os dados clínicos foram devidamente anonimizados para preservar os registros médicos dos pacientes. As imagens da Figura 6 são exemplos de imagens clínicas utilizadas no trabalho, com realizações de ambas as mamas em projeções CC e MLO.

Figura 6 – Exemplos de imagens clínicas *raw* nas projeções CC e MLO.



É importante ressaltar que as bases de dados utilizadas neste estudo, tanto o conjunto de imagens sintéticas (VCT) quanto as imagens clínicas reais, são compostas por exames considerados normais ou regiões de interesse livres de lesões patológicas específicas, como nódulos e microcalcificações. Consequentemente, o escopo deste trabalho delimita-se

à análise técnica da eficiência dos algoritmos na filtragem de ruído e na preservação da integridade do sinal. Não obstante, ainda que a validação de diagnóstico não seja o objeto direto de estudo, as análises quantitativas e qualitativas aqui conduzidas consideram o impacto potencial dos métodos na preservação de estruturas finas, fator determinante para a detectabilidade clínica dessas lesões em um contexto de diagnóstico real.

3.2 Métodos Baseados em Modelo

Inicialmente, foi implementado o filtro Gaussiano com *kernel* de tamanho 5×5 pixels, ponderando entre a capacidade de filtragem de ruído e a perda de nitidez. A fim de demonstrar também o comportamento das métricas utilizadas em relação ao ruído residual e borramento, experimentou-se kernels de tamanhos 3×3 e 11×11 .

Para a implementação do método BM3D neste trabalho, foi adotado o pipeline proposto por Borges *et al.* (2017), conforme a Figura 2, desenvolvido especificamente para imagens de mamografia digital (FFDM) e tomossíntese (DBT). Inicialmente, as imagens foram processadas para remoção dos parâmetros de calibração do detector, como *offset* e ganho espacial, de forma a aproximar o sinal bruto do modelo físico dos processos de aquisição. Em sequência, empregou-se uma transformação de estabilização da variância (VST - *Variance-Stabilizing Transformation*), baseada na transformação de Anscombe generalizada, permitindo converter o ruído Poisson-Gaussiano dependente do sinal para um ruído aproximadamente Gaussiano de variância constante. Após esta normalização estatística, o algoritmo BM3D foi executado conforme as diretrizes do pipeline, fazendo uso de *patch-based estimation*, análise de auto-similaridade não-local e filtragem colaborativa no domínio das transformadas, garantindo a preservação de detalhes anatômicos e eficiente supressão do ruído. Finalizada esta etapa, aplicou-se a transformação inversa da VST e a reinserção dos parâmetros de calibração originais, obtendo-se as imagens restauradas compatíveis com o padrão clínico. Os valores de ganho do detector, *offset* do pixel e ruído eletrônico utilizados nesta implementação foram extraídos e ajustados conforme especificações do sistema *Selenia Dimensions Mammography System* (Hologic Inc., EUA), garantindo maior realismo e aderência ao cenário clínico estudado.

Com esses dois métodos, foi possível explorar neste trabalho os impactos de aplicar um modelo com pouquíssimos parâmetros e necessidade de pouco conhecimento prévio a respeito do domínio de atuação (filtro Gaussiano) em comparação com um modelo onde é necessário o conhecimento de diversos parâmetros a respeito do processo de aquisição de imagem, analisando as vantagens e desvantagens dos dois métodos.

3.3 Métodos Baseados em Dados

Para comparação dos diferentes métodos *data-based* no contexto de restauração de imagens FFDM, foram implementadas abordagens supervisionadas e auto-supervisionadas

de aprendizado profundo, investigando as diferenças de treinamentos com imagens sintéticas e clínicas.

Na abordagem supervisionada (Noise2Clean), onde é oferecido à rede um par de imagens com uma realização limpa (sem ruído) e uma realização ruidosa da mesma cena, foram utilizadas 120 imagens sintéticas VCT com inserção de ruído por meio de simulação, a fim de obter a imagem ruidosa (*input*) e sua correspondente imagem limpa (*ground truth*). O ruído inserido simula a realização de um exame de mamografia digital com 100% da dose típica. O *dataset* de 120 imagens foi dividido em 108 imagens para treinamento e 12 imagens para validação durante o treinamento (proporção de 90%/10%).

Com os 120 pares de imagens sintéticas (limpas e ruidosas), divididas em treinamento e validação, foi extraído *patches* de 60×60 sem sobreposição (considerando o tamanho da imagem de 3300×1800). Utilizando o método de Otsu (Otsu, 1979), selecionou-se apenas *patches* da região da mama, tanto para o treinamento quanto para validação.

Já na abordagem auto-supervisionada, foram utilizadas 250 imagens clínicas de FFDM para treinamento dos modelos, uma vez que esses métodos não necessitam de *ground truth* para ser usado como *target* do aprendizado. Os métodos escolhidos para avaliar essa abordagem foram o Noise2Void (Krull; Buchholz; Jug, 2019) e o Noise2Sim (Niu *et al.*, 2022), por necessitarem apenas uma realização da imagem com ruído. Abordagens como Noise2Noise (Lehtinen *et al.*, 2018), onde é necessário duas realizações da mesma imagem com ruídos independentes, se tornam difíceis de serem aplicadas com imagens clínicas pela forma com que os exames são conduzidos, justificando a escolha dos dois métodos mencionados. O *dataset* de imagens clínicas foi dividido em 225 imagens para treinamento dos modelos e 25 imagens para validação durante o treinamento (proporção de 90%/10%).

Da seleção de 250 imagens clínicas, divididas em treinamento e validação, foi extraído *patches* de 64×64 (considerando agora imagens de 3328×2560 e 4096×3328) apenas da região da mama, semelhante o processo utilizado para imagens sintéticas.

3.3.1 Arquitetura e Hiperparâmetros

Para todos os três métodos *data-based* implementados, foi escolhida a rede U-Net (Ronneberger; Fischer; Brox, 2015) como arquitetura base, composta por 3 blocos de *encoder* e 3 blocos de *decoder*. Cada *encoder* e *decoder* possuí duas camadas convolucionais, função de ativação *ReLU* e *Batch Normalization* antes de cada função de ativação, com o *decoder* possuindo também uma camada de convolução transposta adicional. Essa arquitetura foi escolhida por ser comumente utilizada por modelos de restauração de imagens baseado em dados, tal qual as implementações de Niu *et al.* (2022) e Krull, Buchholz e Jug (2019).

A função de perda foi padronizada como MSE (*Mean Squared Error*) para todos os treinamentos, utilizando a técnica de *early stopping* após 5 épocas sem melhoria na perda de validação, a fim de prevenir o *overfitting* da rede.

Learning rate dos modelos foi fixado em 1×10^{-4} , utilizando o otimizador *Adam* com os outros hiperparâmetros padrão.

3.3.2 Parâmetros Específicos dos Métodos Auto-Supervisionados

Replicando as configurações com melhor desempenho do Noise2Void, foi utilizada uma Janela de Contexto de 5×5 pixels. Esta janela define a vizinhança de um pixel mascarado que será utilizada pela rede para predizer seu valor. O Método de Mascaramento adotado foi a Amostragem Uniforme de Pixel (UPS - *Uniform Pixel Selection*), e a Taxa de Amostragem foi fixada em 0,198%. Considerando o tamanho dos *patches* de entrada de 64×64 pixels utilizados no treinamento, essa taxa de amostragem resulta no mascaramento de 8 pixels por *input* em cada iteração. É importante ressaltar que a função de perda MSE é aplicada apenas aos pixels que foram "cegados", garantindo que a rede aprenda a restaurar o valor do pixel central apenas a partir de seus vizinhos ruidosos. (Krull; Buchholz; Jug, 2019)

No caso do Noise2Sim, prezando também pela implementação dos parâmetros que apresentaram melhores resultados em Niu *et al.* (2022), o tamanho da janela utilizada para geração dos vizinhos semelhantes foi de 3×3 pixels, enquanto a quantidade de pixels similares (k) foi fixada como $k = 8$. Dado o conjunto de $k + 1$ imagens similares, o método para obtenção dos dados de treinamento foi selecionar um par de forma aleatória e construído independente pixel a pixel, tal qual o método 4 da Seção S1.3.3 de Niu *et al.* (2022).

3.4 Avaliação dos Métodos

Para avaliação dos resultados obtidos com os modelos *model-based* e *data-based* já treinados, foram usadas novas imagens clínicas e sintéticas, isto é, que não fizeram parte do conjunto de treinamento dos métodos *data-based*, obtidas tal qual a Seção 3.3.

Tanto para os métodos *model-based* quanto *data-based*, os processos de *denoising* foram aplicados para 10 imagens VCT e 10 imagens clínicas. No caso dos métodos *data-based*, a inferência se deu por meio da extração de *patches* das imagens VCT e clínicas com tamanhos de 180×180 e 192×192 , respectivamente, apresentando sobreposição em todas as direções. Após a inferência dos modelos supervisionado e auto-supervisionado, o *patch* central de 60×60 e 64×64 , respectivamente, foi utilizado para reconstrução da imagem final restaurada, a fim de evitar efeitos de borda pela rede convolucional.

Para a avaliação quantitativa dos métodos, foram utilizadas cinco métricas *full-*

reference, isto é, que carecem de referência (*ground-truth*) para serem aplicadas, e uma métrica *no-reference*, que é aplicada diretamente na imagem a ser avaliada. As métricas *full-reference*, aplicadas às imagens VCT, foram PSNR (*Peak Signal-to-Noise Ratio*), SSIM (*Structural Similarity Index Measure*), QILV (*Quality Index based on Local Variance*), HaarPSI (*Haar Wavelet-based Perceptual Similarity Index*) e MNSE (Mean Normalized Squared Error) com suas decomposições em *Residual Noise* (\mathcal{R}_N) e *Bias* (\mathcal{B}_2). Já métrica *no-reference* escolhida foi a NAQI (*Normalized Anisotropic Quality Index*), aplicada a imagem clínica.

Todas as métricas foram calculadas dentro da região de interesse da mama para garantir uma avaliação consistente entre os diferentes métodos.

4 RESULTADOS & DISCUSSÕES

Especificamente para os métodos *data-based*, o treinamento de todos os modelos atingiu um ponto ótimo (*early stopping*) antes do máximo de épocas definidas, convergindo para minimização da função de perda. O melhor modelo para Noise2Clean e Noise2Sim foi na 12^a época (com MSE de validação igual a 3.65×10^{-6} e 0.29×10^{-6} , respectivamente), enquanto para Noise2Void foi na 19^a época (com MSE de validação igual a 1.49×10^{-6}). Essa minimização pode ser percebida nos gráficos da perda de treinamento e validação das Figuras 7a, 7b e 7c, para os métodos Noise2Clean, Noise2Void e Noise2Sim, respectivamente.

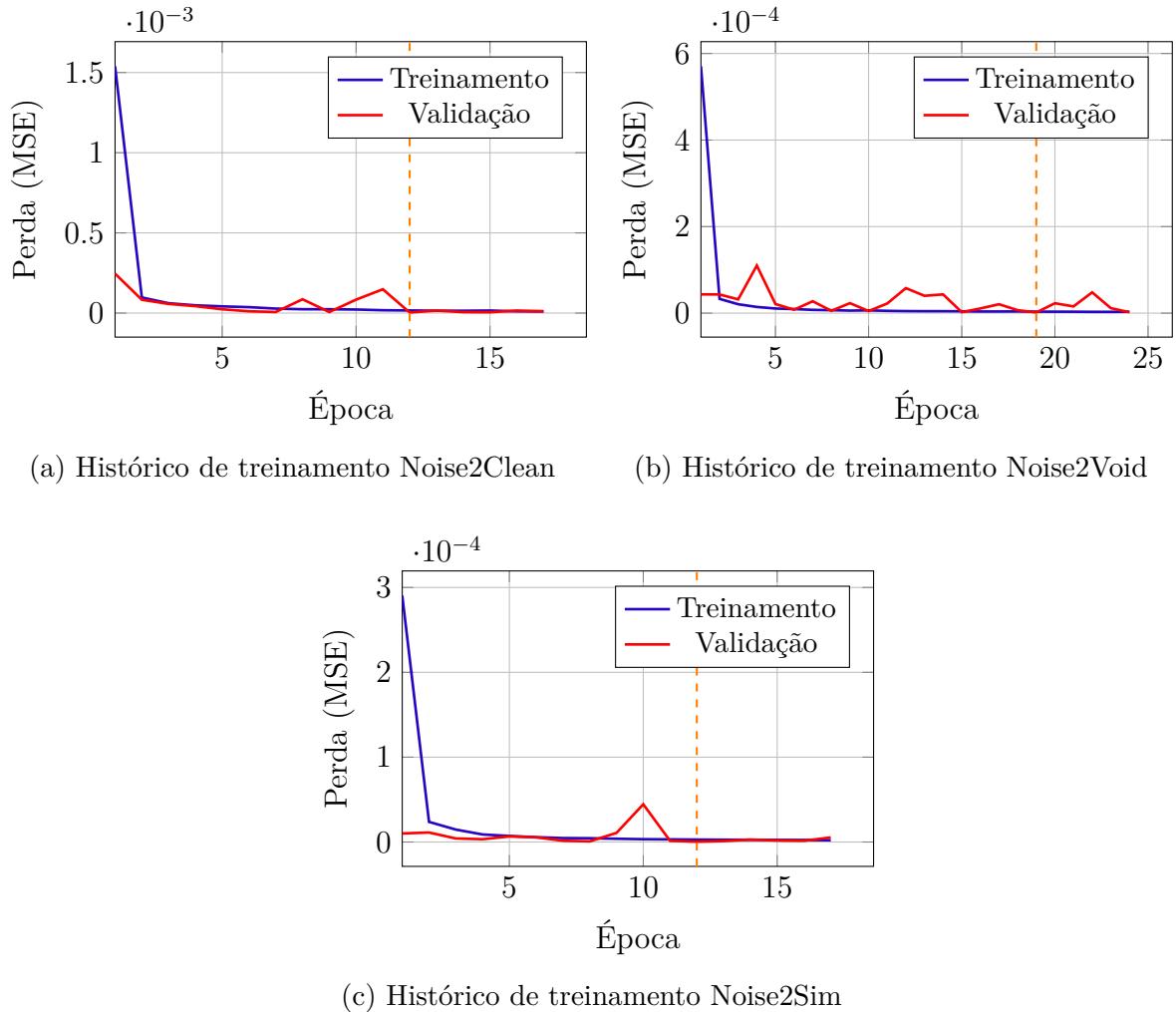


Figura 7 – Comparação das curvas de perda de treinamento e validação para os três métodos de *denoising* ao longo das épocas.

O resultado qualitativo das inferências dos modelos estão apresentados na Figura 9. A Figura 8 apresenta a região de interesse (ROI) do qual os *patches* de comparação foram retirados. É possível perceber visualmente a diferença entre os métodos de restauração, tendo o filtro Gaussiano e Noise2Void com respostas similares em termos de filtragem de

ruído e borramento; Noise2Clean com uma certa perda de contraste da imagem; Noise2Sim com um maior ruído residual mas menor borramento que os três últimos; e por fim BM3D se destacando em melhor supressão de ruído e menor borramento.

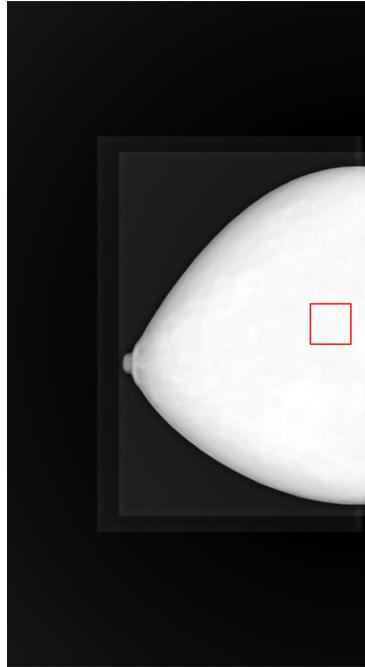


Figura 8 – Imagem VCT *raw* com o ROI utilizado na Figura 9.

A Tabela 1 apresenta as métricas de desempenho PSNR, SSIM, QILV e HarPSI para os métodos implementados baseados em aprendizado profundo (Noise2Clean, Noise2Void e Noise2Sim), além dos métodos analíticos (filtro Gaussiano 5×5 e BM3D). As Figuras 10 e 11 demonstram os mesmos resultados em formato de gráfico, para melhor visualização. Todos os métodos de filtragem demonstraram melhoria substancial em relação às imagens ruidosas originais. O BM3D e o filtro Gaussiano 5×5 alcançaram os maiores valores de PSNR (45,34 e 45,33, respectivamente) e SSIM (0,934 e 0,923, respectivamente), seguidos pelos métodos baseados em aprendizado profundo. Entre os métodos de *deep learning*, o Noise2Void apresentou desempenho superior (PSNR: 43,32; SSIM: 0,862), seguido pelo Noise2Sim (PSNR: 42,27; SSIM: 0,854) e Noise2Clean (PSNR: 42,12; SSIM: 0,842).

A métrica QILV, que avalia a qualidade da imagem baseada na variância local, mostrou resultados expressivos de todos os métodos na tarefa de *denoising*. O QILV compara a distribuição de variância local entre a imagem restaurada e a referência, sendo particularmente sensível a mudanças na estrutura local da imagem. Todos os métodos apresentaram valores de QILV superiores a 0,989, indicando alta similaridade estrutural local com a imagem de referência. O Noise2Clean obteve o melhor desempenho entre os métodos de *deep learning* (0,996), aproximando-se dos valores dos métodos analíticos BM3D (0,990) e filtro Gaussiano (0,989).

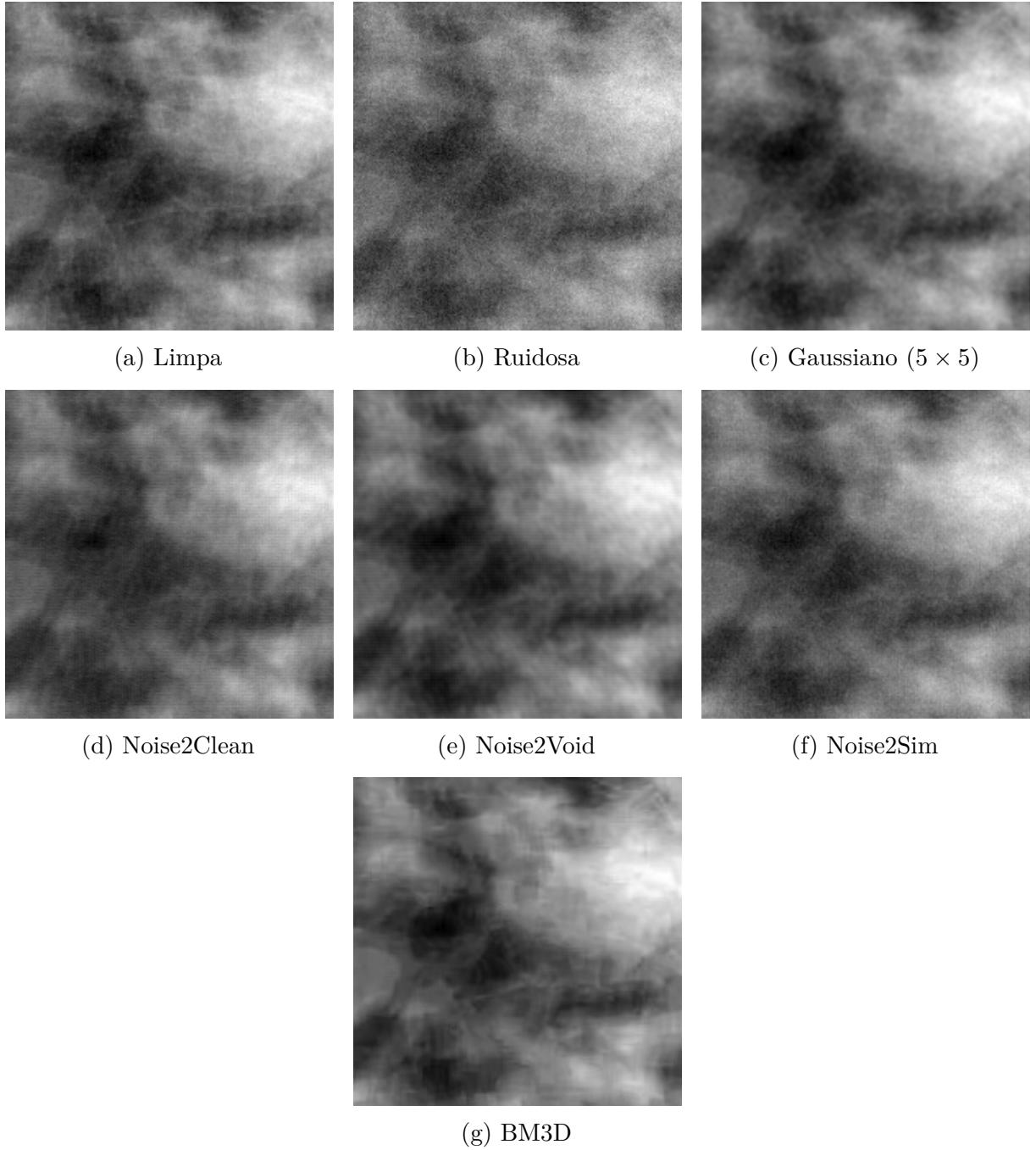


Figura 9 – Comparação visual dos resultados de *denoising* de diferentes métodos em relação à imagem ruidosa e ao *ground truth* (imagem limpa sintética).

A métrica HaarPSI, que avalia a similaridade perceptual baseada em decomposição *wavelet* de Haar, apresentou menor variação relativa entre os métodos comparada à SSIM. Enquanto a SSIM variou de 0,842 a 0,934, a HaarPSI apresentou valores entre 0,851 e 0,884. Esta menor variação sugere que, apesar das diferenças na similaridade estrutural capturadas pela SSIM, a qualidade perceptual das imagens restauradas pelos diferentes métodos permaneceu relativamente próxima. Especificamente, a diferença entre o melhor método (filtro Gaussiano: 0,884) e o Noise2Sim (0,851) foi de apenas 3,7%, comparada a uma diferença de 9,8% observada na SSIM.

Tabela 1 – Métricas de Desempenho de Métodos de Remoção de Ruído

Método	PSNR	SSIM	QILV	HaarPSI
Noisy	38.87 [38.71 39.02]	0.763 [0.740 0.787]	0.802 [0.776 0.828]	0.816 [0.805 0.827]
N2Clean	42.12 [41.76; 42.49]	0.842 [0.818; 0.865]	0.996 [0.995; 0.998]	0.862 [0.851; 0.874]
N2Void	43.32 [42.75; 43.88]	0.862 [0.840; 0.885]	0.994 [0.993; 0.996]	0.860 [0.847; 0.874]
N2Sim	42.27 [42.01; 42.52]	0.854 [0.831; 0.878]	0.992 [0.987; 0.998]	0.851 [0.842; 0.860]
Filtro 5 × 5	45.33 [44.99; 45.67]	0.923 [0.913; 0.932]	0.989 [0.987; 0.992]	0.884 [0.876; 0.892]
BM3D	45.34 [44.97; 45.70]	0.934 [0.929; 0.940]	0.990 [0.989; 0.991]	0.883 [0.875; 0.891]

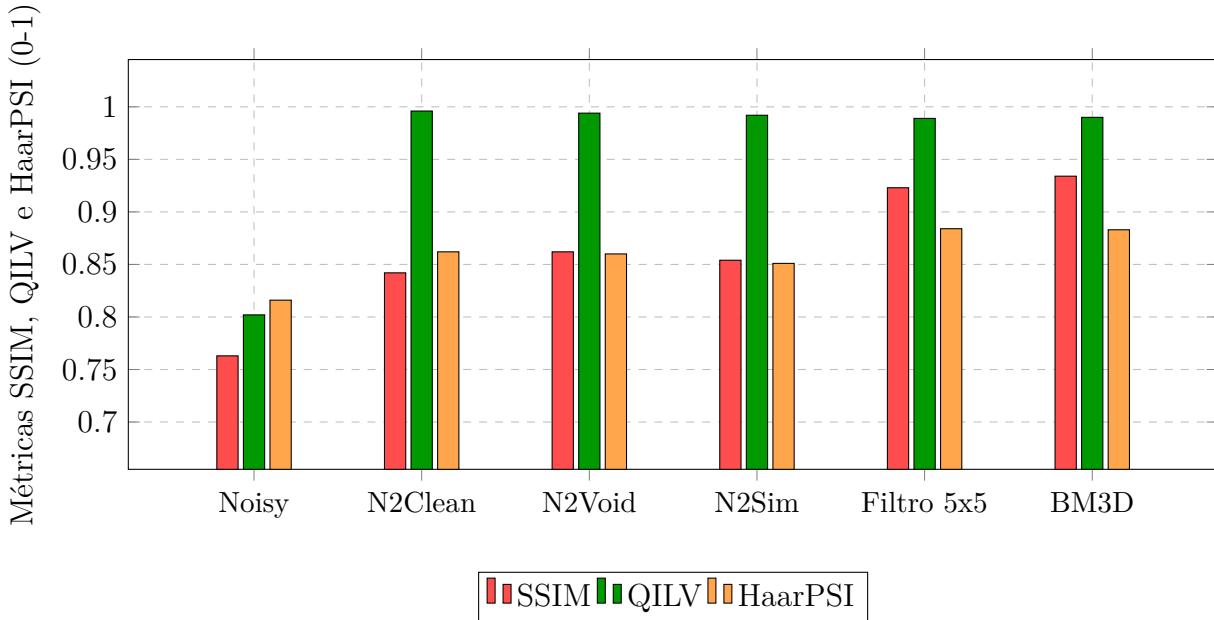


Figura 10 – Comparação das Métricas de Qualidade de Imagem (SSIM, QILV e HaarPSI).

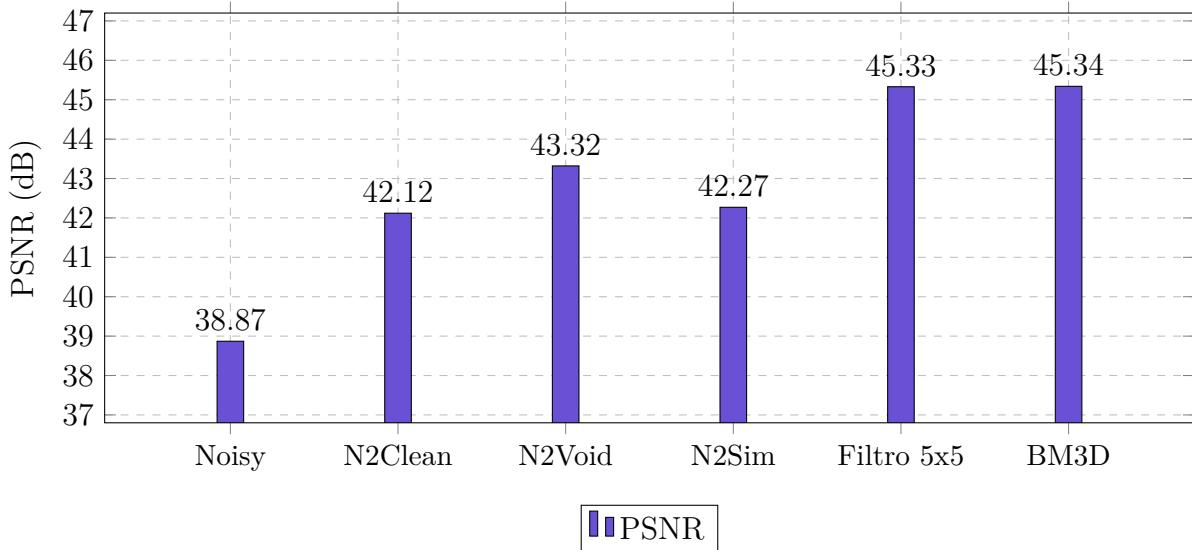


Figura 11 – Comparação da métrica PSNR para os Métodos de Remoção de Ruído.

Embora os métodos baseados em *deep learning* tenham apresentado diferenças significativas, principalmente de PSNR e SSIM, em relação os modelos analíticos, é importante ressaltar algumas limitações das métricas tradicionais para avaliação de qualidade em imagens médicas. A Figura 12 mostra uma comparação da filtragem de ruído realizada com diferentes tamanhos do filtro Gaussiano. Quanto maior o filtro, mais componentes de alta frequência da imagem são removidos, deixando a mesma com a aparência mais borrosa. Esse efeito pode ser indesejado em se tratando da utilização desses exames para detecção de microcalcificações e realizações de diagnósticos médicos (Soares, 2025).

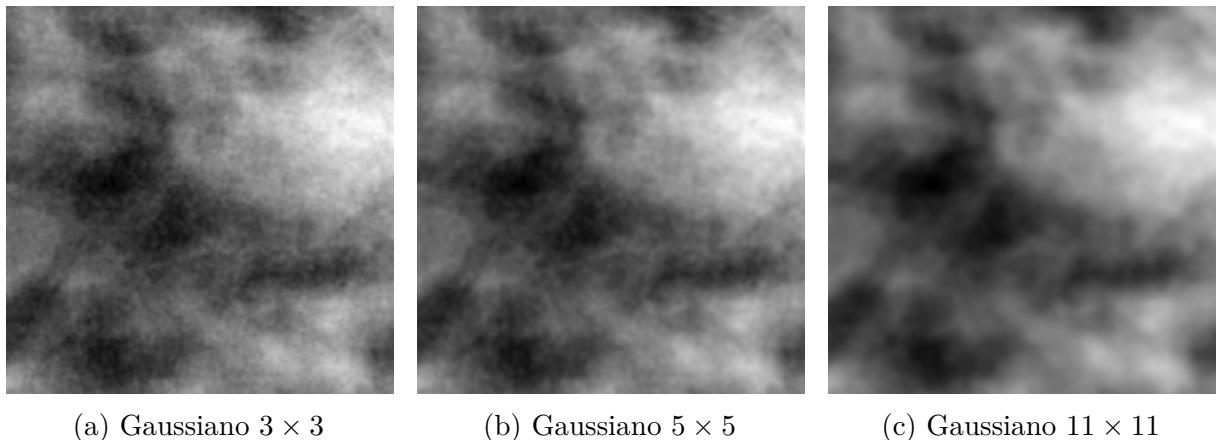


Figura 12 – Comparação do *denoising* do filtro Gaussiano para diferentes tamanhos de *kernel*

Entretanto, analisando as mesmas métricas aplicadas na Tabela 1 para os diferentes filtros Gaussianos, é possível perceber na Tabela 2 que o filtro Gaussiano 11×11 , que causa borramento excessivo e elimina grande parte das componentes de alta frequência da imagem, ainda mantém valores elevados de PSNR (44,13), SSIM (0,897) e HaarPSI (0,854). Esta observação confirma que tais métricas priorizam principalmente a redução de ruído, atribuindo peso insuficiente ao borramento introduzido no processo de filtragem.

A métrica QILV mostrou maior sensibilidade ao borramento, com redução de 0,997 (filtro 3×3) para 0,966 (filtro 11×11). No entanto, mesmo com esta variação, os valores permanecem elevados, não refletindo adequadamente a degradação perceptual e funcional causada pelo borramento excessivo.

Nas Figuras 13 e 14 estão também esses resultados dos Filtros Gaussianos em formato de gráfico.

Tendo isso em vista, a decomposição do MNSE (*Mean Normalized Squared Error*) oferece uma perspectiva mais abrangente para o contexto de imagens médicas, permitindo ponderar separadamente o ruído residual (\mathcal{R}_N) e o borramento do sinal (\mathcal{B}^2). Esta característica é particularmente importante para mamografia, onde tanto o ruído quanto o borramento podem impactar negativamente a detectabilidade de estruturas sutis.

A Tabela 3 apresenta a decomposição do MNSE em suas componentes de ruído

Tabela 2 – Métricas de Desempenho de Filtros Gaussianos com *Kernels* de Diferentes Tamanhos.

Método	PSNR	SSIM	QILV	HaarPSI
Filtro 3×3	44.94 [44.70; 45.18]	0.918 [0.907; 0.929]	0.997 [0.997; 0.998]	0.871 [0.863; 0.880]
Filtro 5×5	45.33 [44.99; 45.67]	0.923 [0.913; 0.932]	0.989 [0.987; 0.992]	0.884 [0.876; 0.892]
Filtro 11×11	44.13 [43.65; 44.60]	0.897 [0.885; 0.908]	0.966 [0.960; 0.971]	0.854 [0.842; 0.867]

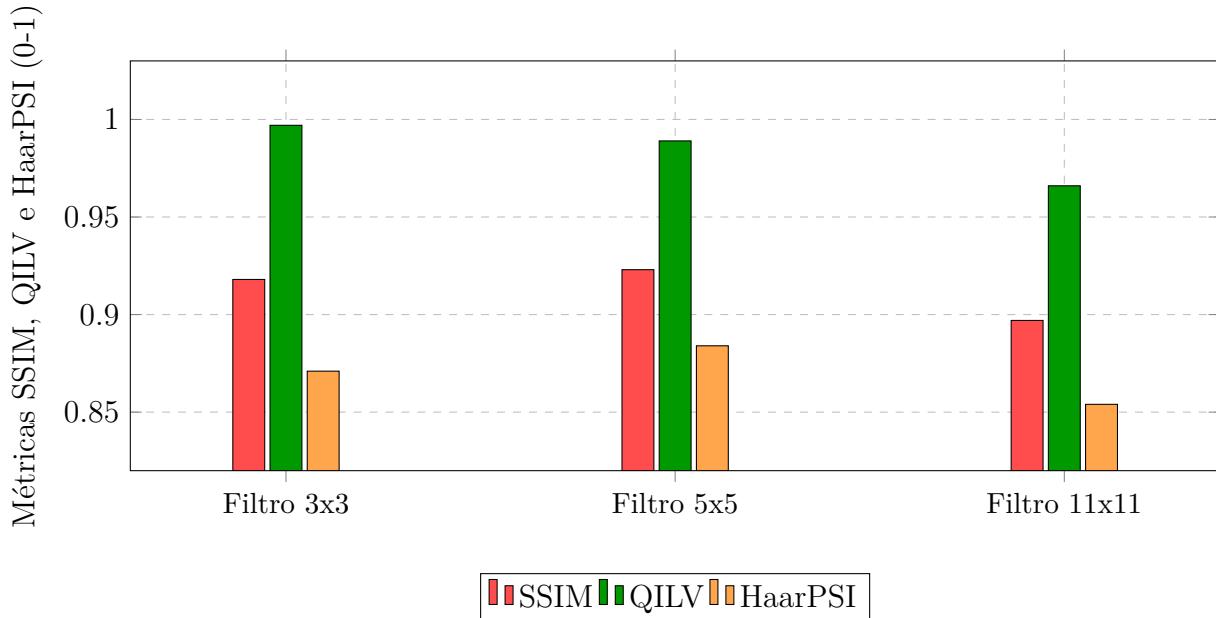


Figura 13 – Comparaçāo das Métricas de Qualidade de Imagem (SSIM, QILV e HaarPSI) para diferentes tamanhos de Kernel do Filtro Gaussiano.

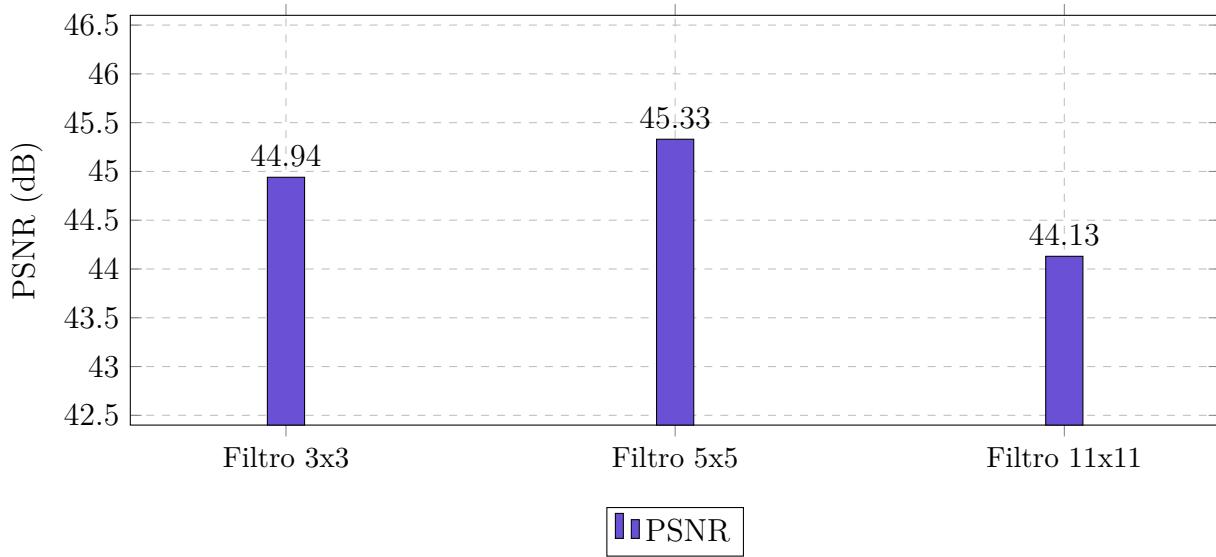


Figura 14 – Comparaçāo da métrica PSNR para diferentes tamanhos de Kernel do Filtro Gaussiano.

residual (\mathcal{R}_N) e bias quadrático (\mathcal{B}^2) para todos os métodos avaliados. Além da tabela, a Figura 15 apresenta os mesmos dados em formato de gráfico para melhor visualização.

Esta análise permite compreender o compromisso entre remoção de ruído e preservação de sinal nos diferentes métodos.

Tabela 3 – MNSE, $\mathcal{R}_{\mathcal{N}}$ e \mathcal{B}^2 dos Métodos de Remoção de Ruído.

Método	MNSE (%)	$\mathcal{R}_{\mathcal{N}}$ (%)	\mathcal{B}^2 (%)
Noisy	10.47 [10.45; 10.48]	10.19 [10.17; 10.20]	0.278 [0.272; 0.284]
N2Clean	8.059 [8.045; 8.072]	1.800 [1.797; 1.803]	6.258 [6.167; 6.35]
N2Void	5.983 [5.975; 5.99]	0.613 [0.612; 0.614]	5.369 [5.258; 5.481]
N2Sim	5.452 [5.442; 5.463]	3.096 [3.091; 3.100]	2.357 [2.331; 2.382]
Filtro 3×3	3.530 [3.525; 3.535]	1.525 [1.523; 1.527]	2.005 [1.946; 2.063]
Filtro 5×5	4.039 [4.034; 4.044]	0.823 [0.822; 0.824]	3.216 [3.116; 3.316]
Filtro 11×11	6.537 [6.532; 6.541]	0.237 [0.237; 0.238]	6.299 [6.100; 6.498]
BM3D	2.419 [2.415; 2.424]	0.881 [0.879; 0.882]	1.539 [1.531; 1.546]

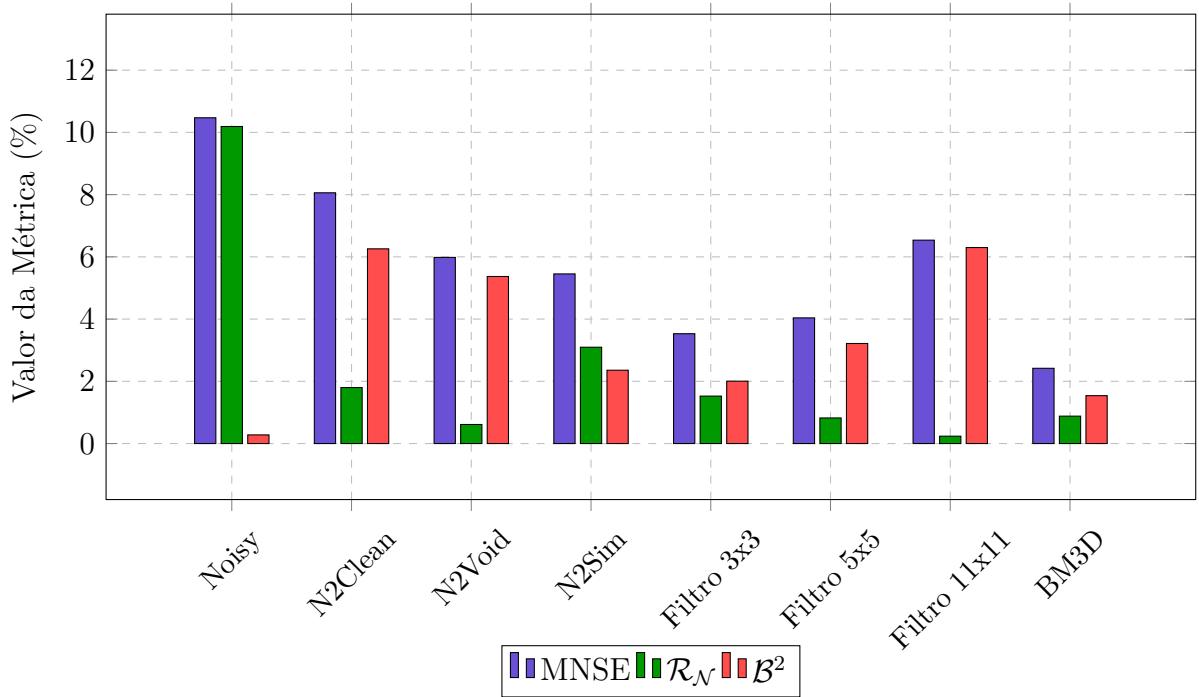


Figura 15 – Comparação da métrica MNSE e suas decomposições $\mathcal{R}_{\mathcal{N}}$ e \mathcal{B}^2 para os métodos de remoção de ruído. Valores em percentual (%).

Foram também adicionados a essa visão (Figura 3) os resultados do filtro Gaussiano 3×3 e 11×11 , a fim de evidenciar o comportamento das métricas a respeito do borramento, elaborado anteriormente. É possível perceber que, de fato, o filtro Gaussiano 11×11 apresenta o maior \mathcal{B}^2 e um dos maiores MNSE totais, indicando um borramento excessivo no processo de *denoising*. Entretanto, principalmente para métricas como PSNR e SSIM, figura entre os melhores resultados, superando métodos com MNSE total e \mathcal{B}^2 menores.

O Noise2Void apresentou o menor ruído residual entre os métodos de *deep learning* ($\mathcal{R}_{\mathcal{N}}$: 0,613%), aproximando-se do desempenho dos métodos analíticos. Contudo, este

resultado foi acompanhado de um \mathcal{B}^2 elevado (5,369%), indicando maior suavização do sinal. Este comportamento sugere que a rede aprendeu a remover ruído de forma agressiva, mas ao custo de maior borramento das estruturas finas da imagem. Esta tendência ao borramento excessivo é uma limitação conhecida do Noise2Void, relacionada ao seu mecanismo de *blind-spot training*. Krull, Buchholz e Jug (2019) observaram que o Noise2Void pode gerar artefatos de tabuleiro (*checkerboard artifacts*) e borramento devido à estratégia de mascaramento de pixels e à dependência exclusiva de informações de vizinhança local. Estudos subsequentes propuseram modificações arquiteturais, como o N2V2, especificamente para mitigar esses artefatos e reduzir o borramento (Höck *et al.*, 2022).

O Noise2Clean apresentou desempenho intermediário em termos de filtragem de ruído ($\mathcal{R}_{\mathcal{N}}$: 1,8%), porém com \mathcal{B}^2 de 6,258% — o maior entre todos os métodos avaliados. Este resultado pode estar relacionado à limitação na quantidade e diversidade de imagens VCT utilizadas no treinamento, impedindo que a rede generalizasse adequadamente para as variações anatômicas presentes nas imagens de teste. Huang *et al.* (2023) reforçam que métodos de *self-supervised learning*, ao aprenderem diretamente de dados clínicos diversos, tendem a apresentar melhor generalização em comparação com métodos supervisionados treinados em conjuntos de dados limitados ou exclusivamente sintéticos.

O Noise2Sim, embora tenha apresentado maior ruído residual ($\mathcal{R}_{\mathcal{N}}$: 3,096%), obteve o menor \mathcal{B}^2 entre os métodos de *deep learning* (2,357%). Esta característica é particularmente relevante para imagens médicas, onde a preservação de estruturas de alta frequência, como microcalcificações, é crítica para o diagnóstico. O menor borramento observado no Noise2Sim pode ser vantajoso em tarefas clínicas que dependem da nitidez de detalhes finos (Soares, 2025).

Os métodos analíticos — BM3D e filtro Gaussiano 5×5 — demonstraram desempenho superior na métrica MNSE global. O BM3D apresentou o melhor equilíbrio geral (MNSE: 2,419%; $\mathcal{R}_{\mathcal{N}}$: 0,881%; \mathcal{B}^2 : 1,539%), confirmando sua posição como método estado-da-arte em filtragem de ruído. O filtro Gaussiano 5×5 obteve desempenho comparável em termos de MNSE (4,039%), porém com maior \mathcal{B}^2 (3,216%), evidenciando maior borramento em relação ao BM3D.

Sumarizando, todos os métodos baseados em *deep learning* avaliados neste trabalho melhoraram significativamente a qualidade da imagem em comparação com as imagens ruidosas originais, como evidenciado pela redução substancial do MNSE de 10,47% (imagem ruidosa) para 5,452%–8,059% (métodos de *deep learning*). Esta melhoria, embora ainda inferior ao BM3D (MNSE: 2,419%) e filtro Gaussiano 5×5 (MNSE: 4,039%), demonstra o potencial desses métodos *data-based*, especialmente considerando que se trata de implementações iniciais com espaço considerável para otimização.

A análise através do MNSE e suas decomposições revela *trade-offs* distintos entre

os métodos: o Noise2Void prioriza a remoção de ruído; o Noise2Sim preserva melhor as estruturas de alta frequência; o Noise2Clean apresenta limitações relacionadas à generalização; o filtro Gaussiano revela a necessidade de ponderação entre restauração e borramento, através da escolha do tamanho do *kernel*; e o BM3D revela excelente performance quando conhecidos os parâmetros do problema de restauração. Do ponto de vista dos métodos *data-based*, os resultados preliminares indicam aplicação promissora desses métodos auto-supervisionados para mamografia digital, especialmente considerando que foram treinados exclusivamente com imagens clínicas ruidosas, sem necessidade de *ground truth* limpo e informações de contexto do domínio. Já em uma aplicação onde se sabe com precisão as características do processo de aquisição das imagens e estrutura fundamental de sua degradação, os métodos *model-based* oferecem uma solução robusta e confiável, principalmente em implementações mais sofisticadas como BM3D.

Na Figura 17, é apresentado também o resultado dos modelos aplicados a imagens clínicas de FFDM, onde os patches foram retirados da região de interesse indicada na Figura 16. Percebe-se visualmente um comportamento muito parecido com o que foi demonstrado com imagens sintéticas, onde Noise2Void e o filtro Gaussiano apresentam boa supressão de ruído mas com um maior borramento; Noise2Clean também demonstrando filtragem do ruído se comparado com a imagem original, porém com limitações na reconstrução das estruturas; Noise2Sim priorizando uma imagem com menor borramento e estruturas mais nítidas em detrimento do ruído residual; e por fim o método BM3D apresentando o melhor balanço entre *denoising* e borramento.

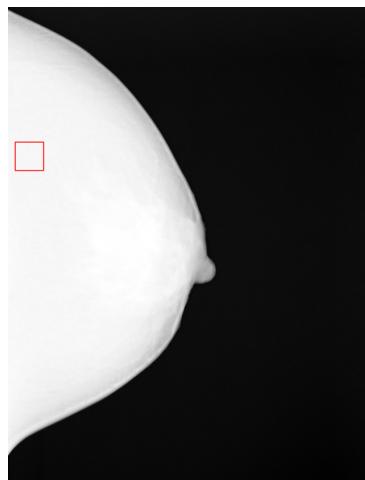


Figura 16 – Imagem Clínica *raw* com o ROI utilizado na Figura 17.

Na Tabela 4, estão os resultados compilados da métrica *no-reference* NAQI, que foi medida nas imagens clínicas após o processo de *denoising*. A Figura 18 também apresenta os mesmos dados em formato de gráfico. Verifica-se que todos os métodos melhoraram a métrica, destacando os métodos Noise2Void, filtro Gaussiano e BM3D como os melhores (sem diferenças significativas entre eles). Entretanto, conforme demonstrado por Soares (2025), métricas *no-reference* como NAQI apresentam limitações importantes na represen-

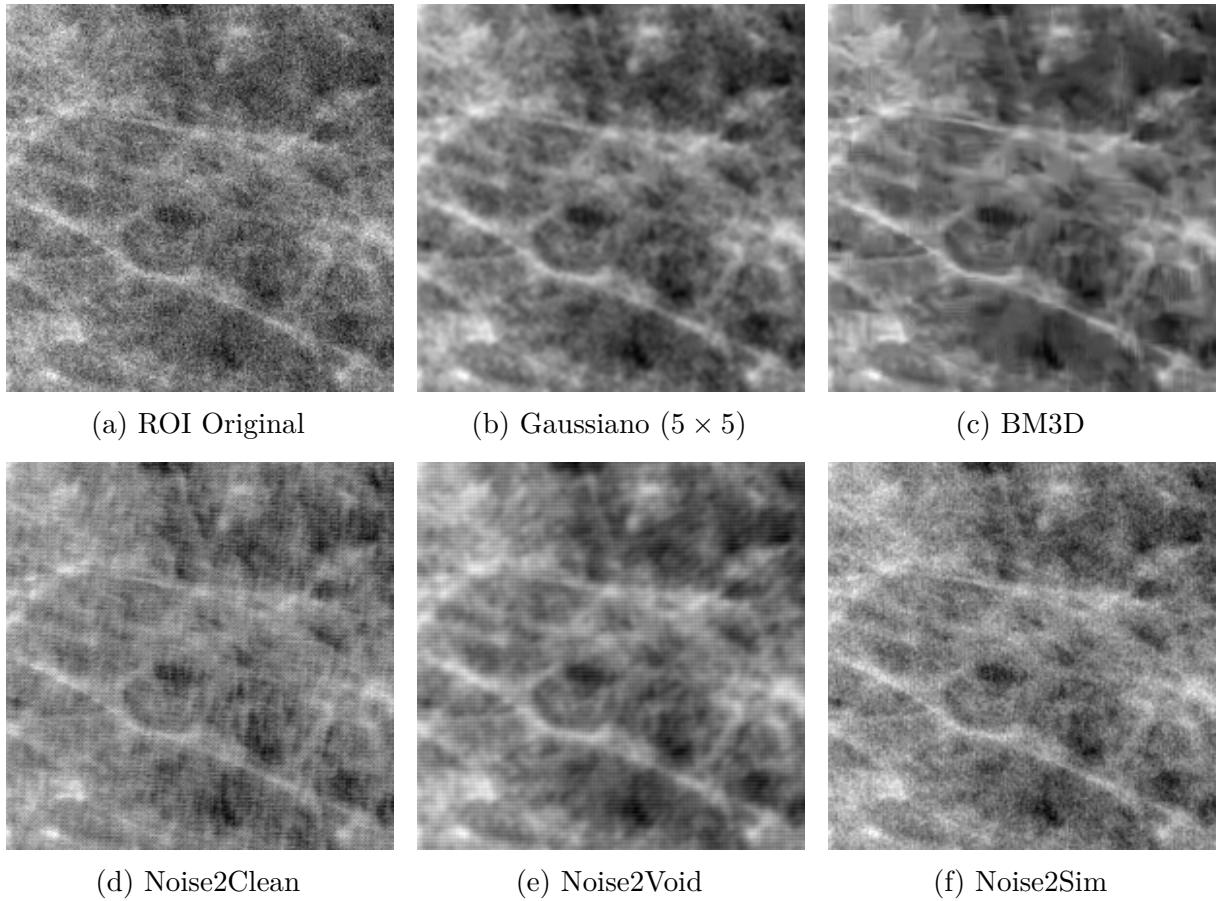


Figura 17 – Comparação visual dos resultados de *denoising* de diferentes métodos para ROI de imagem clínica.

tação da qualidade da imagem, principalmente com relação ao grau de detectabilidade de microcalcificações.

Tabela 4 – NAQI dos Métodos de Remoção de Ruído.

Método	NAQI
Noisy	0.096 [0.077; 0.114]
N2Clean	0.210 [0.172; 0.247]
N2Void	0.306 [0.277; 0.334]
N2Sim	0.178 [0.152; 0.204]
Filtro 5 × 5	0.276 [0.242; 0.310]
BM3D	0.283 [0.254; 0.313]

É importante ressaltar que o domínio analisado neste trabalho — mamografia digital de campo total (FFDM) — difere dos domínios explorados nos trabalhos originais de Noise2Sim e, principalmente, Noise2Void. O método Noise2Void foi originalmente proposto e validado em imagens de microscopia de fluorescência e imagens naturais, enquanto o Noise2Sim foi aplicado principalmente em tomografia computadorizada. A aplicação desses métodos em mamografia digital representa uma extensão para um domínio



Figura 18 – Comparação da métrica NAQI para os métodos de remoção de ruído. Valores mais altos indicam melhor qualidade percebida.

com características distintas de ruído e estrutura de sinal, o que pode ter um efeito na performance dos métodos.

A seguir, é apresentado o resultado obtido para outras regiões de interesse (ROI) das imagens VCT e clínicas, agora de uma região de transição (parte mais próxima a borda da mama). Essa região pode ser identificada nas Figuras 19 e 20 para as imagens VCT e clínica, respectivamente. Os *patches* resultantes do processo de restauração de cada uma dessas regiões estão disponíveis nas Figuras 21 (VCT) e 22 (clínica).

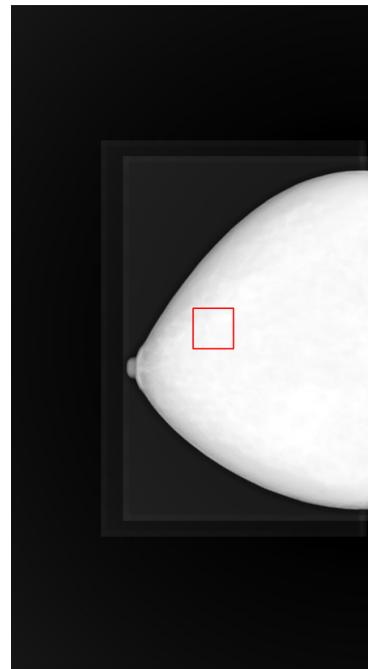


Figura 19 – Imagem VCT *raw* com o ROI de transição utilizado na Figura 21.

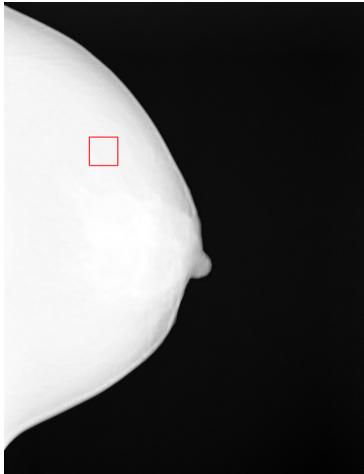


Figura 20 – Imagem Clínica *raw* com o ROI de transição utilizado na Figura 22.

É possível perceber que, para as regiões de transição da imagem sintética, não houve grandes diferenças na restauração pelos diferentes métodos, por se tratar de uma região onde a VCT não exibe estruturas tão bem definidas como no interior da mama simulada. Porém, para as regiões de transição da imagem clínica, os métodos apresentaram claras diferenças entre si, em termos de nitidez e preservação de estruturas. O Noise2Clean e, principalmente, o Noise2Void, não representaram bem a composição da imagem original após a restauração, muito provavelmente pela similaridade do nível de cinza dos pixels nessa região. Já o filtro Gaussiano, BM3D e Noise2Sim, mantiveram o comportamento observado nos outros resultados, demonstrando a robustez desses métodos.

Os resultados para as regiões de transição deixam ainda mais evidente a necessidade de avaliações baseadas em tarefas e testes clínicos, uma vez que imagens sintéticas podem não representar de forma fidedigna todos os cenários, mascarando o resultado de algumas métricas utilizadas para validação dos métodos de restauração.

Sob uma perspectiva teórica, a aplicação direta de filtros lineares invariantes no espaço, como o filtro Gaussiano, desconsidera a natureza heterocedástica do ruído em imagens de raios-X, cuja variância é dependente da intensidade do sinal (distribuição Poisson-Gaussiana). Diferentemente do algoritmo BM3D, que normaliza essa dependência através da Transformada de Estabilização de Variância (VST) garantindo uma tratativa uniforme, o filtro Gaussiano aplica um núcleo de suavização fixo, resultando inevitavelmente em uma filtragem desigual: o ruído é suprimido de forma inconsistente entre regiões de alta e baixa densidade. Contudo, essa disparidade teórica e local é mascarada nas métricas de avaliação global (como o PSNR médio). Observa-se um efeito de compensação estatística, onde o desempenho insuficiente em determinadas regiões da imagem é contrabalançado pelo desempenho em outras no cálculo da média, fazendo com que a diferença numérica entre os métodos pareça menos significativa do que a análise da estabilidade do ruído sugere. Isso pode ser percebido nas figuras 23a e 23b, que ilustra a distribuição do valor

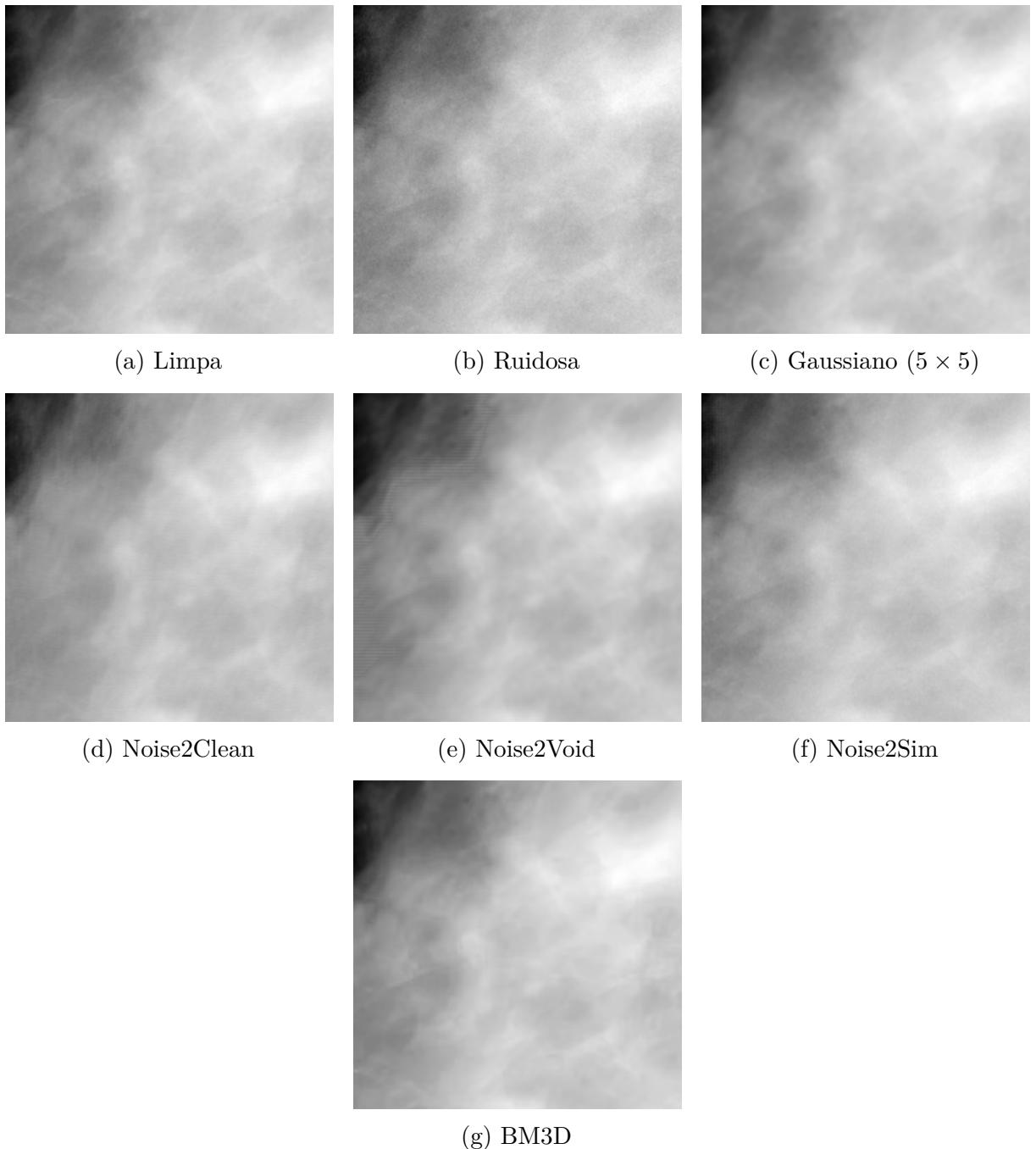


Figura 21 – Comparaçāo visual dos resultados de *denoising* do ROI de transição para diferentes métodos em relação à imagem ruidosa e ao *ground truth* (imagem limpa sintética).

da métrica SSIM ao longo da região da mama para os métodos BM3D e Filtro Gaussiano 5×5 , respectivamente.

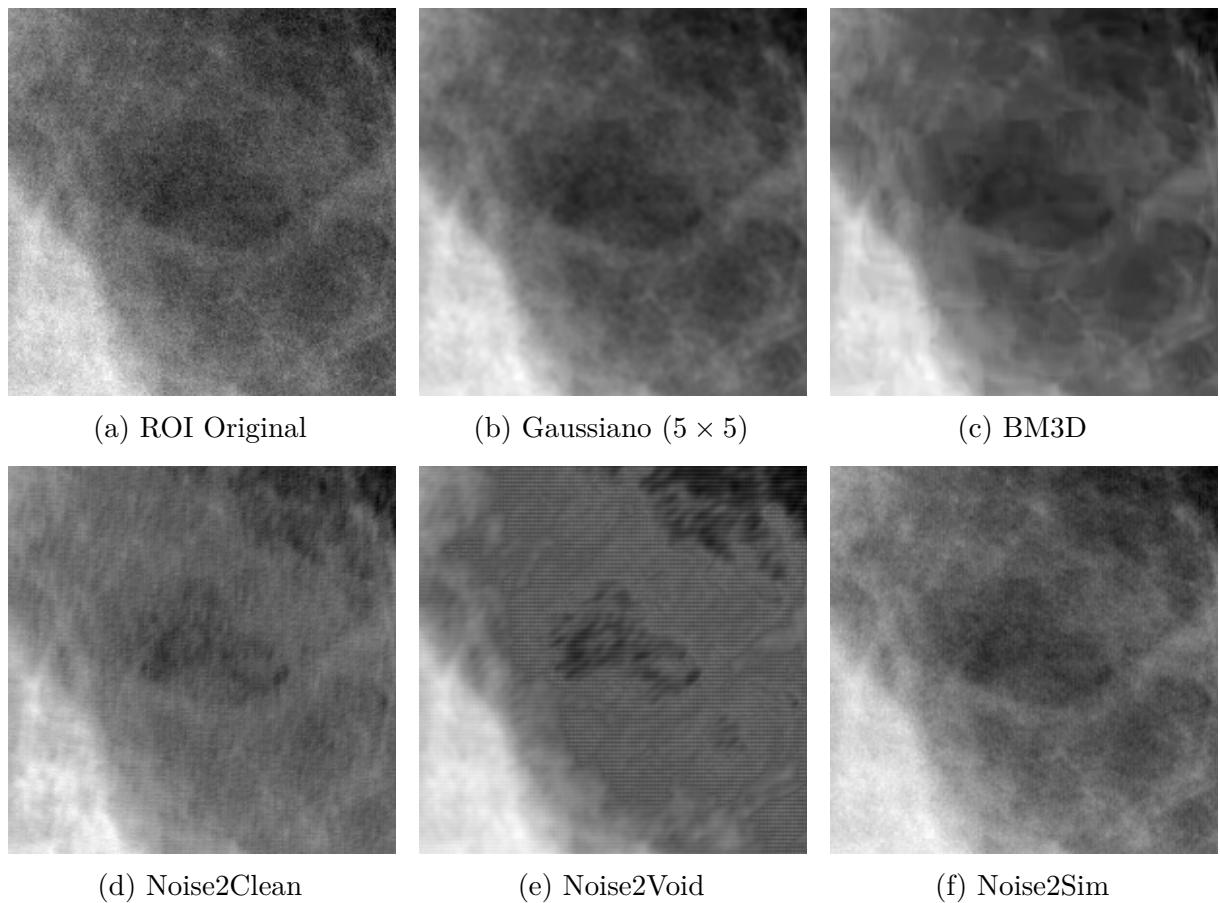
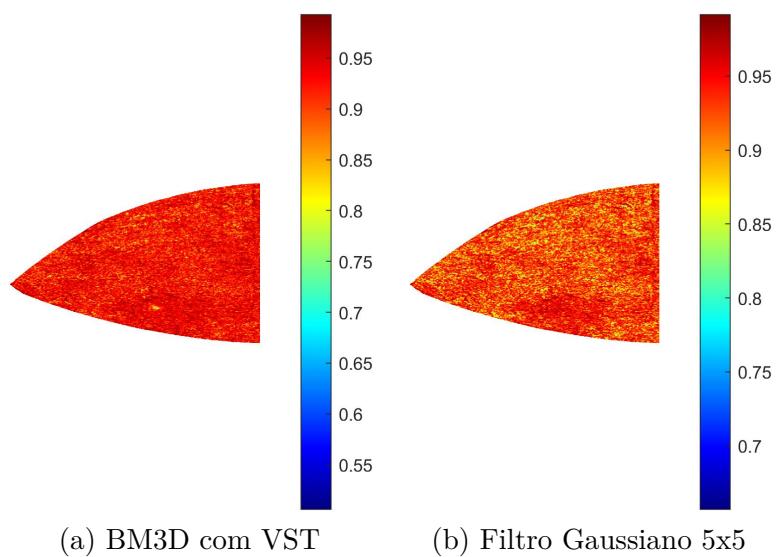


Figura 22 – Comparação visual dos resultados de *denoising* de diferentes métodos para ROI de transição de imagem clínica.

Figura 23 – Distribuição da métrica SSIM pela região da mama para os métodos BM3D com VST e Filtro Gaussiano 5×5 .



5 LIMITAÇÕES DO TRABALHO

Apesar dos resultados promissores obtidos na comparação entre métodos baseados em modelo e baseados em dados, este trabalho apresenta limitações que devem ser consideradas na interpretação dos resultados e que motivam as propostas de trabalhos futuros.

Uma limitação primordial refere-se à quantidade restrita de imagens utilizadas para o treinamento dos modelos de aprendizado profundo. A base de dados, composta por um número limitado de imagens clínicas e sintéticas (VCT), pode restringir a capacidade de generalização dos modelos, como observado nas dificuldades do método supervisionado Noise2Clean em lidar com variações de aquisição não vistas durante o treinamento. A diversidade em termos de densidade mamária, fabricantes de equipamentos e protocolos de aquisição também foi limitada, o que é um fator crítico para a robustez clínica de métodos *data-based*.

Do ponto de vista arquitetural, o estudo limitou-se à utilização de uma rede U-Net padrão com 3 blocos para todos os métodos de *deep learning*. Embora suficiente para uma comparação *baseline*, essa arquitetura pode não ser capaz de capturar complexidades mais finas e estruturas de alta frequência tão eficientemente quanto arquiteturas mais modernas ou profundas. Além disso, foram observados artefatos específicos, como o efeito de *checkerboard* no método Noise2Void, que degradaram o desempenho em métricas de viés quadrático e impactaram a qualidade visual final.

Outra limitação importante é a ausência de estudos clínicos com radiologistas e validações baseadas em tarefas (*task-based assessment*). O trabalho baseou-se majoritariamente em métricas de qualidade de imagem técnicas (como PSNR, SSIM e MNSE). Embora informativas, essas métricas não garantem necessariamente que a detectabilidade de lesões sutis, como microcalcificações e nódulos, foi preservada ou aprimorada. A falta de uma validação clínica impede a confirmação definitiva de que os métodos implementados mantêm a eficácia diagnóstica em cenários de redução de dose.

6 TRABALHOS FUTUROS

Os resultados preliminares deste trabalho indicam diversas oportunidades para investigações futuras, principalmente para os métodos *data-based*, permitindo consolidar e ampliar a compreensão sobre a aplicação de métodos de aprendizado profundo auto-supervisionado e supervisionado para *denoising* em mamografia digital (FFDM). A limitação de generalização observada no Noise2Clean sugere que aumentar o conjunto de imagens VCT para treinamento e expandir o conjunto de imagens clínicas poderia melhorar significativamente a capacidade de generalização dos modelos supervisionados e auto-supervisionados, respectivamente, permitindo maior diversidade em termos de densidade mamária, equipamentos e protocolos de aquisição.

Além da U-Net de 3 blocos utilizada neste trabalho pelos métodos de *deep learning*, seria recomendável avaliar arquiteturas mais avançadas, como U-Net com mecanismos de *attention*, redes com conexões densas (DenseNet) (Huang *et al.*, 2018) e arquiteturas híbridas que combinam CNNs com *Transformers*, uma vez que essas abordagens poderiam melhor capturar estruturas de alta frequência importantes para mamografia. A investigação do método N2V2 (Höck *et al.*, 2022) seria particularmente relevante para mitigar os artefatos de *checkerboard* observados no Noise2Void, potencialmente reduzindo o viés quadrático elevado reportado neste trabalho.

Para avaliar mais rigorosamente a qualidade das imagens restauradas em tarefas específicas de detecção de microcalcificações, recomenda-se implementar avaliação baseada em *observer models*, particularmente o *Channelized Hotelling Observer* (CHO), também implementado no pacote de VCT do VICTRE. Esta abordagem permitiria quantificar a detectabilidade de estruturas finas através do índice de detectabilidade (d') e métricas de SNR *task-specific*, fornecendo uma medida objetiva mais diretamente relacionada ao desempenho diagnóstico em comparação com métricas tradicionais. Adicionalmente, seria prudente implementar funções de perda alternativas para os modelos *data-based*, particularmente combinações de MSE com *perceptual loss* ou *loss functions* baseadas na métrica MNSE, similares à abordagem de Vimieiro (2023), que poderiam melhorar o balanceamento entre supressão de ruído e preservação de estruturas. Otimização dos hiperparâmetros dos métodos auto-supervisionados — taxa de mascaramento para Noise2Void, tamanho da janela de similaridade para Noise2Sim — poderia otimizar o desempenho neste domínio específico.

Paralelamente às otimizações técnicas, é possível também implementar um estudo piloto com radiologistas para avaliar a qualidade visual das imagens restauradas e a confiança diagnóstica em tarefas de detecção de microcalcificações. Seria também relevante conduzir análise espectral preliminar, para verificar se os métodos preservam características

espectrais importantes e não introduzem correlação artificial de ruído nas imagens restauradas, permitindo compreender melhor as diferenças observadas nas métricas PSNR/SSIM versus HaarPSI/MNSE. Os estudos futuros deveriam focar em validar rigorosamente os resultados promissores observados, particularmente nos métodos *data-based* a vantagem do Noise2Sim em manter menor borramento enquanto filtra ruído de forma significativa. A validação mediante *observer models* e estudos clínicos seria essencial para confirmar se essa característica realmente se traduziria em melhor performance diagnóstica, validando portanto o potencial clínico desses métodos para redução de dose de radiação em mamografia digital mantendo qualidade diagnóstica.

7 CONCLUSÃO

A proposta deste trabalho foi investigar e comparar o desempenho de métodos *model-based* e *data-based* para remoção de ruído em mamografia digital de campo total (FFDM), considerando tanto imagens sintéticas geradas por ensaios clínicos virtuais (VCT) quanto imagens clínicas reais. Foram avaliados métodos analíticos clássicos (filtro Gaussiano e BM3D) e métodos baseados em aprendizado profundo, em abordagens supervisionada (Noise2Clean) e auto-supervisionadas (Noise2Void e Noise2Sim). A análise foi conduzida por meio de um conjunto abrangente de métricas de qualidade *full-reference* e *no-reference*, associado a uma discussão crítica sobre suas limitações no contexto da imagem médica.

Os resultados demonstraram que o BM3D apresentou o melhor desempenho geral, alcançando um excelente balanço entre remoção de ruído (MNSE: 2,419%, \mathcal{R}_N : 0,881%) e preservação de detalhes finos (\mathcal{B}^2 : 1,539%), consolidando sua posição como método estado-da-arte em *denoising* analítico. Dentre os métodos baseados em dados (*data-based*), o Noise2Sim apresentou resultados particularmente promissores no contexto de mamografia digital, obtendo o menor viés quadrático entre os métodos de *deep learning* (\mathcal{B}^2 : 2,357%), o que o torna especialmente relevante para aplicações clínicas onde a nitidez de microcalcificações é crítica. O Noise2Void, embora tenha alcançado a menor quantidade de ruído residual entre os métodos de aprendizado profundo (\mathcal{R}_N : 0,613%), apresentou borramento acentuado (\mathcal{B}^2 : 5,369%), evidenciando as limitações reconhecidas de artefatos de tabuleiro decorrentes de seu mecanismo de *blind-spot training*. O Noise2Clean, apesar de ter demonstrado potencial, sofreu limitações relacionadas à quantidade e diversidade do conjunto de imagens VCT de treinamento, resultando no maior viés quadrático (\mathcal{B}^2 : 6,258%) entre todos os métodos avaliados.

Uma observação crítica emergiu durante a análise: as métricas tradicionais de qualidade de imagem, como PSNR e SSIM, apresentam limitações significativas para avaliação de *denoising* em contextos médicos, uma vez que priorizam a redução de ruído sem ponderar adequadamente os efeitos prejudiciais do borramento excessivo. A decomposição do MNSE em componentes de ruído residual (\mathcal{R}_N) e viés quadrático (\mathcal{B}^2) mostrou-se muito mais informativa para mamografia, permitindo uma avaliação mais rigorosa do compromisso entre filtragem de ruído e preservação de estruturas de alta frequência.

Embora os métodos *data-based* tenham apresentado desempenho quantitativo inferior aos métodos analíticos em métricas globais, destacam-se como alternativas viáveis e promissoras, particularmente os métodos auto-supervisionados, que não requerem *ground truth* ou conhecimento prévio detalhado do domínio. A capacidade de treinar redes neurais profundas diretamente em imagens clínicas ruidosas, sem necessidade de pares limpas-ruidosas ou simulações sintéticas, representa uma vantagem significativa para aplicações

em cenários clínicos reais com múltiplos equipamentos, protocolos de aquisição distintos e variações anatômicas heterogêneas.

As investigações futuras devem focar em (1) expansão e diversificação dos conjuntos de dados (imagens VCT e clínicas), permitindo que os modelos aprendam com maior variedade de cenários clínicos; (2) exploração de arquiteturas mais avançadas, como U-Net com mecanismos de *attention* e redes híbridas que combinam CNNs com *Transformers*, que possam capturar estruturas de alta frequência de forma mais eficaz; (3) implementação do Noise2Void versão 2 (N2V2) para mitigar artefatos de *checkerboard* e reduzir o borramento observado na versão original; (4) otimização de funções de perda alternativas, particularmente combinações que priorizam a métrica MNSE ou incorporem *perceptual loss*, melhorando o balanço entre remoção de ruído e preservação de estruturas; (5) avaliação através de observer models para quantificar a detectabilidade de microcalcificações de forma mais diretamente relacionada ao desempenho diagnóstico; e (6) validação clínica com radiologistas, através de estudos piloto de qualidade visual e confiança diagnóstica em tarefas de detecção.

Conclui-se que, embora métodos *model-based* como o BM3D permaneçam como referência em *denoising* quando os parâmetros do sistema de aquisição são bem caracterizados, os métodos *data-based* auto-supervisionados, especialmente o Noise2Sim, demonstram potencial significativo para aplicações clínicas em mamografia digital. Esses métodos oferecem flexibilidade e robustez para generalizar a múltiplos cenários de aquisição, posicionando-se como ferramentas promissoras para otimização de dose de radiação mantendo a qualidade diagnóstica. Recomenda-se que futuras investigações validem rigorosamente esses resultados preliminares através de métricas *task-specific* e estudos com observadores, confirmando se as características de menor borramento do Noise2Sim realmente se traduzem em superior performance diagnóstica, consolidando assim seu potencial clínico para mamografia de baixa dose.

REFERÊNCIAS

- AJA-FERNANDEZ, S.; VEGAS-SANCHEZ-FERRERO, G.; ALBEROLA-LOPEZ, C. Image quality assessment based on local variance. *In: IEEE. 2006 International Conference of the IEEE Engineering in Medicine and Biology Society*. New York, 2006. p. 4815–4818.
- BADANO, A. *et al.* Evaluation of digital breast tomosynthesis as replacement of full-field digital mammography using an in silico imaging trial. **JAMA Oncology**, v. 4, n. 12, p. 1752–1760, 2018.
- BORGES, L. R. *et al.* Pipeline for effective denoising of digital mammography and digital breast tomosynthesis. *In: FLOHR, T. G.; LO, J. Y.; SCHMIDT, T. G. (ed.). Medical Imaging 2017: Physics of Medical Imaging*. [S.l.: s.n.], 2017. v. 10132, p. 1013206.
- BORGES, L. R. *et al.* Technical Note: Noise models for virtual clinical trials of digital breast tomosynthesis. **Medical Physics**, v. 46, n. 6, p. 2683–2689, jun. 2019.
- BORGES, L. R. *et al.* Restored low-dose digital breast tomosynthesis: a perception study. *In: NISHIKAWA, R. M.; SAMUELSON, F. W. (ed.). Medical Imaging 2018: Image Perception, Observer Performance, and Technology Assessment*. Bellingham: [S.l.: s.n.], 2018a. (Proc. SPIE, v. 10577), p. 51–6.
- BORGES, L. R. *et al.* Restoration of low-dose digital breast tomosynthesis. **Measurement Science and Technology**, v. 29, n. 6, p. 064003, 2018b.
- BUSHBERG, J. T. *et al.* **The Essential Physics of Medical Imaging**. 3rd. ed. Philadelphia: Lippincott Williams Wilkins, 2012.
- CHEN, H. *et al.* Low-dose ct with a residual encoder-decoder convolutional neural network. **IEEE Transactions on Medical Imaging**, v. 36, n. 12, p. 2524–2535, 2017.
- DABOV, K. *et al.* Image denoising by sparse 3-d transform-domain collaborative filtering. **IEEE Transactions on Image Processing**, v. 16, n. 8, p. 2080–2095, 2007.
- GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing**. 4. ed. [S.l.: s.n.]: Pearson, 2018.
- GOYAL, B. *et al.* Image denoising review: From classical to state-of-the-art approaches. **Information Fusion**, v. 55, p. 220–244, 2020. ISSN 1566-2535.
- HUANG, G. *et al.* **Densely Connected Convolutional Networks**. 2018. Disponível em: <https://arxiv.org/abs/1608.06993>.
- HUANG, S.-C. *et al.* Self-supervised learning for medical image classification: a systematic review and implementation guidelines. **npj Digit. Med.**, v. 6, p. 74, 2023. Disponível em: <https://doi.org/10.1038/s41746-023-00811-0>.
- HUANG, S.-C. *et al.* Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation protocols. **NPJ Digital Medicine**, v. 3, n. 1, p. 136, 2020.

HöCK, E. *et al.* **N2V2 – Fixing Noise2Void Checkerboard Artifacts with Modified Sampling Strategies and a Tweaked Network Architecture.** 2022. Disponível em: <https://arxiv.org/abs/2211.08512>.

KRULL, A.; BUCHHOLZ, T.-O.; JUG, F. Noise2void-learning denoising from single noisy images. *In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* [S.l.: s.n.], 2019. p. 2129–2137.

LEE, J. G. *et al.* Deep learning in medical imaging: general overview. **Korean Journal of Radiology**, v. 18, n. 4, p. 570–584, 2017.

LEHTINEN, J. *et al.* Noise2Noise: Learning Image Restoration without Clean Data. **International Conference on Machine Learning (ICML)**, 2018.

MAIA, T. F. *et al.* Processamento e análise de imagens médicas. **Revista Brasileira de Física Médica**, v. 13, n. 1, p. 34–48, 2019.

MICHELL, M. J.; BATOHY, B. Role of tomosynthesis in breast imaging going forward. **Clin Radiol**, v. 73, n. 4, p. 358–371, abr. 2018.

NIU, C. *et al.* **Suppression of Correlated Noises with Similarity-based Unsupervised Deep Learning.** 2022. Version 6, submitted 5 Jan 2022. Disponível em: <https://arxiv.org/abs/2011.03384>.

OLIVEIRA, H. C. R. de *et al.* Validation of no-reference image quality index for the assessment of clinical mammograms. *In: Medical Imaging 2016: Physics of Medical Imaging.* SPIE, 2016. v. 9787, p. 978713. Disponível em: <https://www.imagephysics.com/189.pdf>.

OTSU, N. A threshold selection method from gray-level histograms. **IEEE Transactions on Systems, Man, and Cybernetics**, v. 9, n. 1, p. 62–66, 1979.

REISENHOFER, R. *et al.* A haar wavelet-based perceptual similarity index for image quality assessment. **Signal Processing: Image Communication**, Elsevier BV, v. 61, p. 33–43, fev. 2018. ISSN 0923-5965. Disponível em: <http://dx.doi.org/10.1016/j.image.2017.11.001>.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI).* [S.l.: s.n.], 2015.

SANDER, R.; POCK, T.; CREMERS, D. Noise2grad: Extract image noise to denoise. *In: International Joint Conference on Artificial Intelligence (IJCAI).* [S.l.: s.n.], 2021. p. 826–832.

SARVAMANGALA, D. R.; KULKARNI, R. V. Convolutional neural networks in medical image analysis. **Emerging Trends in Electrical, Electronics & Instrumentation Engineering: An International Journal**, v. 8, n. 2, p. 24–46, 2021.

SHARMA, D. *et al.* Technical Note: In silico imaging tools from the VICTRE clinical trial. **Medical Physics**, v. 46, n. 9, p. 3924–3928, set. 2019.

SOARES, L. E. **Predição da qualidade de mamografias digitais sob a perspectiva da detectabilidade de microcalcificações: uma abordagem radiômica.** 2025. Tese (Tese de Doutorado em Processamento de Sinais e Instrumentação) — Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2025.

TABÁR, L. *et al.* Early detection of breast cancer rectifies inequality of breast cancer outcomes. **J Med Screen**, v. 28, n. 1, p. 34–38, mar. 2021.

VIMIEIRO, R. d. B. **Model-based deep learning to restore low-dose digital breast tomosynthesis images.** 2023. Tese (Tese de Doutorado em Processamento de Sinais e Instrumentação) — Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2023.

WANG, Z.; BOVIK, A. C. Mean squared error: Love it or leave it? a new look at signal fidelity measures. **IEEE Signal Processing Magazine**, v. 26, n. 1, p. 98–117, 2009.

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.

WHO. **World Health Organization - Breast Cancer.** 2025. <https://www.who.int/>. Acesso em: 23 Nov. 2025.

XIA, W. *et al.* Physics-/model-based and data-driven methods for low-dose computed tomography: A survey. **IEEE Signal Processing Magazine**, v. 40, n. 2, p. 89–100, March 2023.

YAFFE, M. J. Digital mammography. In: METTER, R. V.; BEUTEL, J.; KUNDEL, H. (ed.). **Handbook of Medical Imaging: Physics and Psychophysics.** Bellingham: SPIE Press, 2000. v. 1, cap. 5, p. 329–372.

ZHANG, K. *et al.* Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. **IEEE Transactions on Image Processing**, v. 26, n. 7, p. 3142–3155, 2016.