

**INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE
SANTA CATARINA - CÂMPUS FLORIANÓPOLIS
DEPARTAMENTO ACADÊMICO DE ELETRÔNICA
CURSO DE GRADUAÇÃO EM ENGENHARIA ELETRÔNICA**

MARCELO ZAMPIERI PEREIRA DA SILVA

**AVALIAÇÃO DA INFLUÊNCIA DE MÉTODOS DE PRÉ-PROCESSAMENTO DE
IMAGENS MAMOGRAFICAS NO DESEMPENHO DA EFFICIENTNET-B0 PARA
CLASSIFICAÇÃO DE LESÕES MAMÁRIAS**

FLORIANÓPOLIS, 2026.

**INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE
SANTA CATARINA - CÂMPUS FLORIANÓPOLIS
DEPARTAMENTO ACADÊMICO DE ELETRÔNICA
CURSO DE GRADUAÇÃO EM ENGENHARIA ELETRÔNICA**

MARCELO ZAMPIERI PEREIRA DA SILVA

**AVALIAÇÃO DA INFLUÊNCIA DE MÉTODOS DE PRÉ-PROCESSAMENTO DE
IMAGENS MAMOGRAFICAS NO DESEMPENHO DA EFFICIENTNET-B0 PARA
CLASSIFICAÇÃO DE LESÕES MAMÁRIAS**

Trabalho de Conclusão de Curso submetido ao Instituto Federal de Educação, Ciência e Tecnologia de Santa Catarina como parte dos requisitos para obtenção do título de Bacharel em Engenharia Eletrônica.

Orientador:
Prof. Robinson Pizzio, Dr. Eng.

FLORIANÓPOLIS, 2026.

Ficha de identificação da obra elaborada pelo autor.

Silva, Marcelo

AVALIAÇÃO DA INFLUÊNCIA DE MÉTODOS DE PRÉ-PROCESSAMENTO DE IMAGENS MAMOGRÁFICAS NO DESEMPENHO DA EFFICIENTNET-B0 PARA CLASSIFICAÇÃO DE LESÕES MAMÁRIAS / Marcelo Silva; orientação de Robinson Pizzio. - Florianópolis, SC, 2026.

63 p.

Trabalho de Conclusão de Curso (TCC) - Instituto Federal de Santa Catarina, Câmpus Florianópolis. Bacharelado em Engenharia Eletrônica. Departamento Acadêmico de Eletrônica.

Inclui Referências.

1. Câncer de mama. 2. Mamografia digital. 3. Pré-processamento de imagens. 4. Redes Neurais Convolucionais. 5. EfficientNet-B0. I. Pizzio, Robinson. II. Instituto Federal de Santa Catarina. III. AVALIAÇÃO DA INFLUÊNCIA DE MÉTODOS DE PRÉ-PROCESSAMENTO DE IMAGENS MAMOGRÁFICAS NO DESEMPENHO DA EFFICIENTNET-B0 PARA CLASSIFICAÇÃO

AVALIAÇÃO DA INFLUÊNCIA DE MÉTODOS DE PRÉ-PROCESSAMENTO DE
IMAGENS MAMOGRÁFICAS NO DESEMPENHO DA EFFICIENTNET-B0 PARA
CLASSIFICAÇÃO DE LESÕES MAMÁRIAS

MARCELO ZAMPIERI PEREIRA DA SILVA

Este trabalho foi julgado adequado para obtenção do título de Bacharel em Engenharia Eletrônica e aprovado na sua forma final pela banca examinadora do Curso de Engenharia Eletrônica do Instituto Federal de Educação, Ciência e Tecnologia de Santa Catarina.

Florianópolis, 11 de agosto, 2026.

Banca Examinadora:

Documento assinado digitalmente
 **ROBINSON PIZZIO**
Data: 12/08/2026 10:48:51-0300
Verifique em <https://validar.iti.gov.br>

Robinson Pizzio, Dr. Eng.

Documento assinado digitalmente
 **PEDRO GIASSI JUNIOR**
Data: 12/08/2026 14:17:37-0300
Verifique em <https://validar.iti.gov.br>

Pedro Giassi Junior, Dr. Eng.

Documento assinado digitalmente
 **CARLOS GONTARSKI SPERANZA**
Data: 12/08/2026 15:14:46-0300
Verifique em <https://validar.iti.gov.br>

Carlos Gontarski Esperanza, Dr. Eng.

Dedico este trabalho aos meus pais,
pelo incentivo constante,
pelas cobranças necessárias e
por todo o esforço dedicado à minha formação.

AGRADECIMENTOS

Agradeço, primeiramente, aos meus pais, por todo o apoio, incentivo e esforço dedicados à minha formação. Suas cobranças, conselhos e confiança foram essenciais para que eu chegasse até este momento.

Ao meu orientador, Prof. Robinson Pizzio, agradeço pela orientação, pela disponibilidade e pelas contribuições realizadas durante o desenvolvimento deste trabalho.

Aos meus amigos Caio Meira, Guilherme Franco e Rafael Aquino, agradeço pela amizade, pelo companheirismo e pela capacidade de tornar os dias difíceis mais leves, transformar situações negativas em boas risadas e fazer dos momentos mais simples boas lembranças. A presença de vocês, assim como a dos demais amigos e colegas de curso, tornou essa trajetória muito mais leve, agradável e significativa.

Aos professores que fizeram parte da minha formação, agradeço pelos conhecimentos compartilhados, pelos ensinamentos e pelas contribuições que foram fundamentais para o meu desenvolvimento acadêmico e profissional ao longo da graduação.

RESUMO

Este trabalho avalia a influência de métodos de pré-processamento no desempenho da arquitetura EfficientNet-B0 para a classificação de lesões mamárias benignas e malignas em imagens de mamografia. Foram selecionadas 127 imagens pertencentes ao *Digital Mammography Image Dataset* (DMID), sendo 80 associadas a lesões benignas e 47 a lesões malignas. Para realizar uma comparação controlada, as mesmas imagens foram organizadas em três condições experimentais pareadas. A condição Original foi formada pelas imagens do subconjunto DMID antes da aplicação dos métodos avaliados, submetidas apenas às etapas comuns de extração da região de interesse, inclusão de margem de 10% e redimensionamento para 768×768 pixels. O Método A correspondeu às versões das mesmas imagens processadas pelo procedimento de padronização proposto no trabalho Mammo-Bench. O Método B foi produzido a partir das imagens da condição Original por meio da implementação do método apresentado no artigo *Noise Removal for Mammogram Images Using Gaussian Based Unsharp Masking*, baseado na combinação de filtragem Gaussiana e mascaramento de nitidez. A correspondência entre as três condições foi mantida por meio dos identificadores das imagens, preservando as mesmas lesões, rótulos e divisões de treinamento e validação. A classificação foi realizada com a EfficientNet-B0 inicializada com pesos pré-treinados no ImageNet. O protocolo experimental empregou validação cruzada estratificada com cinco divisões e três repetições, totalizando 15 execuções por condição e 45 treinamentos. O desempenho foi avaliado por meio das métricas AUC, acurácia, sensibilidade, especificidade, precisão, F1-score e Coeficiente de Correlação de Matthews, além do teste não paramétrico de Friedman. A condição Original e o Método A apresentaram AUC média de 0,930, enquanto o Método B apresentou AUC média de 0,914. O Método A obteve os maiores valores médios de acurácia, sensibilidade, F1-score e MCC. O Método B apresentou maior precisão e especificidade, porém menor sensibilidade. Apesar das diferenças numéricas, não foram identificadas diferenças estatisticamente significativas entre as três condições. Conclui-se que, no conjunto e no protocolo avaliados, os métodos de pré-processamento produziram diferenças numéricas no comportamento do classificador, mas não foram encontradas evidências estatisticamente significativas de superioridade em relação à condição Original.

Palavras-chave: Câncer de mama. Mamografia digital. Pré-processamento de imagens. Redes Neurais Convolucionais. EfficientNet-B0. Aprendizado por transferência.

ABSTRACT

This study evaluates the influence of image preprocessing methods on the performance of the EfficientNet-B0 architecture for the classification of benign and malignant breast lesions in mammographic images. A total of 127 images belonging to the *Digital Mammography Image Dataset* (DMID) were selected, including 80 benign and 47 malignant cases. To provide a controlled comparison, the same images were organized into three paired experimental conditions. The Original condition comprised the DMID images before the application of the evaluated preprocessing methods and underwent only the common steps of region-of-interest extraction, addition of a 10% margin, and resizing to 768×768 pixels. Method A consisted of the corresponding images processed according to the standardization procedure proposed in the Mammo-Bench study. Method B was generated from the Original images by implementing the procedure presented in the article *Noise Removal for Mammogram Images Using Gaussian Based Unsharp Masking*, which combines Gaussian filtering and unsharp masking. Image identifiers were used to preserve the correspondence among the three conditions, ensuring that the same lesions, labels, and training and validation partitions were used throughout the experiments. Classification was performed using EfficientNet-B0 initialized with ImageNet pretrained weights. The experimental protocol employed stratified five-fold cross-validation with three repetitions, resulting in 15 runs per condition and 45 training procedures. Performance was evaluated using AUC, accuracy, sensitivity, specificity, precision, F1-score, and Matthews Correlation Coefficient, in addition to the nonparametric Friedman test. The Original condition and Method A achieved a mean AUC of 0.930, whereas Method B achieved a mean AUC of 0.914. Method A obtained the highest mean accuracy, sensitivity, F1-score, and MCC values. Method B achieved higher precision and specificity but lower sensitivity. Despite the numerical differences, no statistically significant differences were identified among the three conditions. It was therefore concluded that, under the evaluated dataset and experimental protocol, the preprocessing methods produced numerical differences in classifier performance, but no statistically significant evidence of superiority over the Original condition was found.

Keywords: Breast cancer. Digital mammography. Image preprocessing. Convolutional Neural Networks. EfficientNet-B0. Transfer learning.

LISTA DE FIGURAS

Figura 1 – Exemplos de projeções mamográficas presentes no conjunto de dados DMID	19
Figura 2 – Estrutura simplificada de uma Rede Neural Convolucional	22
Figura 3 – Aplicação do aprendizado por transferência na classificação de imagens mamográficas	25
Figura 4 – Fluxograma simplificado da filtragem Gaussiana e do mascaramento de nitidez	28
Figura 5 – Representação simplificada da curva ROC	31
Figura 6 – Exemplo do procedimento de extração da região de interesse	38
Figura 7 – Comparação visual entre as condições experimentais	42
Figura 8 – Curvas ROC médias das condições experimentais	50
Figura 9 – Boxplots dos valores de sensibilidade e especificidade obtidos nas 15 execuções	51
Figura 10 – Matrizes de confusão agregadas das três repetições da validação cruzada	52

LISTA DE TABELAS

Tabela 1 – Estrutura da matriz de confusão para classificação binária	29
Tabela 2 – Resumo das métricas utilizadas na avaliação do modelo	32
Tabela 3 – Distribuição das imagens utilizadas por classe	35
Tabela 4 – Origem e definição das condições experimentais	40
Tabela 5 – Parâmetros utilizados no Método B	41
Tabela 6 – Configuração da arquitetura utilizada	43
Tabela 7 – Parâmetros utilizados durante o treinamento	44
Tabela 8 – Transformações utilizadas no aumento de dados	45
Tabela 9 – Organização da validação experimental	46
Tabela 10 – Desempenho médio das condições experimentais	48
Tabela 11 – Métricas médias relacionadas à classificação das classes	49
Tabela 12 – Medianas, intervalos interquartis e resultados do teste de Friedman	53

LISTA DE ABREVIATURAS E SIGLAS

AUC	<i>Area Under the Curve</i> (Área sob a Curva ROC)
BI-RADS	<i>Breast Imaging Reporting and Data System</i>
CAD	<i>Computer-Aided Diagnosis</i> (Diagnóstico Assistido por Computador)
CC	Craniocaudal
CNN	<i>Convolutional Neural Network</i> (Rede Neural Convolucional)
CUDA	<i>Compute Unified Device Architecture</i>
DMID	<i>Digital Mammography Image Dataset</i>
FDA	<i>Food and Drug Administration</i>
FN	Falso Negativo
FP	Falso Positivo
IA	Inteligência Artificial
IFSC	Instituto Federal de Educação, Ciência e Tecnologia de Santa Catarina
INCA	Instituto Nacional de Câncer
MBConv	<i>Mobile Inverted Bottleneck Convolution</i>
MCC	<i>Matthews Correlation Coefficient</i> (Coeficiente de Correlação de Matthews)
MLO	Mediolateral Oblíqua
NCI	<i>National Cancer Institute</i>
ReLU	<i>Rectified Linear Unit</i> (Unidade Linear Retificada)
ROI	<i>Region of Interest</i> (Região de Interesse)
ROC	<i>Receiver Operating Characteristic</i>
SVM	<i>Support Vector Machine</i> (Máquina de Vetores de Suporte)
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo
WHO	<i>World Health Organization</i>

SUMÁRIO

1	INTRODUÇÃO	13
1.1	Justificativa	14
1.2	Definição do Problema	15
1.3	Objetivo Geral	15
1.4	Objetivos Específicos	15
1.5	Estrutura do Trabalho	15
2	FUNDAMENTAÇÃO TEÓRICA	17
2.1	Câncer de Mama	17
2.2	Mamografia Digital	18
2.3	Inteligência Artificial Aplicada ao Diagnóstico Médico	20
2.4	Redes Neurais Convolucionais	21
2.5	Aprendizado por Transferência	24
2.6	Arquitetura EfficientNet	26
2.7	Pré-processamento de Imagens Mamográficas	27
2.8	Métricas de Avaliação	29
2.9	Teste de Friedman	32
2.10	Resumo do Capítulo	33
3	METODOLOGIA	34
3.1	Caracterização da Pesquisa	34
3.2	Conjunto de Dados	35
3.3	Preparação das Imagens	36
3.3.1	Seleção e organização das amostras	36
3.3.2	Extração das regiões de interesse	37
3.3.3	Redimensionamento e padronização	38
3.4	Origem e Aplicação dos Métodos de Pré-processamento	39
3.4.1	Condição Original	40
3.4.2	Método A: pré-processamento Mammo-Bench	40
3.4.3	Método B: redução de ruído por filtragem Gaussiana e mascaramento de nitidez	41
3.5	Arquitetura do Modelo	42
3.6	Configuração do Treinamento	43
3.6.1	Aumento de dados	44
3.7	Validação Experimental	45
3.8	Métricas e Análise Estatística	46
3.9	Resumo do Capítulo	47
4	APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS	48
4.1	Desempenho Geral das Condições Experimentais	48
4.2	Análise da Curva ROC e da AUC	49
4.3	Sensibilidade e Especificidade	50
4.4	Análise das Matrizes de Confusão	51
4.5	Análise Estatística	52
4.6	Discussão dos Resultados	53
4.7	Resumo do Capítulo	55
5	CONSIDERAÇÕES FINAIS	56
5.1	Sugestões para Trabalhos Futuros	58

REFERÊNCIAS 60

1 INTRODUÇÃO

O câncer de mama constitui um importante problema de saúde pública e permanece entre as neoplasias de maior incidência na população feminina. No Brasil, são estimados 78.610 novos casos de câncer de mama feminino por ano para o triênio de 2026 a 2028, correspondendo a aproximadamente 30% dos casos de câncer estimados entre as mulheres, desconsiderando-se os tumores de pele não melanoma (INSTITUTO NACIONAL DE CÂNCER – INCA, 2026). A detecção precoce desempenha papel fundamental na redução dos impactos da doença, sendo a mamografia um dos principais métodos empregados no rastreamento e na identificação de alterações suspeitas (INSTITUTO NACIONAL DE CÂNCER – INCA, 2025). Entretanto, a interpretação desses exames pode ser dificultada por fatores como densidade mamária, sobreposição dos tecidos e características sutis das lesões (NATIONAL CANCER INSTITUTE – NCI, 2025).

Paralelamente, a Inteligência Artificial (IA) vem sendo aplicada em diferentes áreas da medicina, destacando-se as aplicações relacionadas ao processamento e à análise de imagens médicas. Entre as técnicas utilizadas encontram-se as Redes Neurais Convolucionais (*Convolutional Neural Networks* – CNNs), capazes de aprender automaticamente características relevantes das imagens durante o treinamento (LECUN *et al.*, 1998). O avanço dessas arquiteturas, aliado ao uso de modelos pré-treinados por meio do *Transfer Learning*, possibilitou o desenvolvimento de sistemas de diagnóstico assistido por computador (*Computer-Aided Diagnosis* – CAD), investigados como ferramentas de apoio à análise de exames médicos (YOSINSKI *et al.*, 2014; MCKINNEY *et al.*, 2020).

Apesar dos avanços obtidos, o desempenho desses modelos também é influenciado pelas características do conjunto de dados utilizado durante o treinamento. Em mamografia, diferentes estratégias de pré-processamento são empregadas para padronização das imagens, redução de ruído, ajuste de contraste, segmentação e extração de regiões de interesse. Revisões da literatura demonstram ampla variedade de procedimentos e configurações de pré-processamento, bem como diferenças entre conjuntos de dados e protocolos de avaliação (SOLTANI *et al.*, 2024). Dessa forma, os benefícios obtidos com essas operações podem variar de acordo com a técnica, o conjunto de dados e a arquitetura utilizada (BHOLE; SUSEELA; PAREKH, 2025; INGLE; VIDHATE; CHAUDHARI, 2024).

Diante desse cenário, este trabalho propõe a avaliação da influência de diferentes métodos de pré-processamento no desempenho da arquitetura EfficientNet-B0 para classificação de lesões mamárias em exames de mamografia. Para isso, foram utilizadas imagens mamográficas do conjunto de dados *Digital Mammography Image*

Dataset (DMID), previamente classificadas como benignas ou malignas, permitindo comparar o desempenho do modelo quando submetido a diferentes estratégias de processamento das imagens. Os resultados obtidos foram analisados por meio de métricas de classificação e análise estatística, buscando verificar se os métodos de pré-processamento avaliados produzem diferenças significativas no desempenho da rede neural.

1.1 Justificativa

O desenvolvimento de sistemas de diagnóstico assistido por computador baseados em Inteligência Artificial tem despertado crescente interesse da comunidade científica, especialmente em razão dos avanços das Redes Neurais Convolucionais e do aprendizado por transferência. Estudos demonstram o potencial dessas técnicas na análise de mamografias, embora seu desempenho dependa de fatores como características do conjunto de dados, estratégia de validação e processamento das imagens (MCKINNEY *et al.*, 2020; SOLTANI *et al.*, 2024).

Nesse contexto, o pré-processamento de imagens mamográficas constitui uma etapa frequentemente empregada antes do treinamento dos modelos. Entre as operações descritas na literatura encontram-se redução de ruído, realce de contraste, segmentação da região mamária, remoção de estruturas não relacionadas à análise e padronização das dimensões (SOLTANI *et al.*, 2024). Métodos específicos podem apresentar resultados distintos: enquanto técnicas de redução de ruído foram associadas a melhorias de desempenho em determinados classificadores (INGLE; VIDHATE; CHAUDHARI, 2024), outras abordagens concentram-se principalmente na padronização das imagens provenientes de diferentes conjuntos de dados (BHOLE; SUSEELA; PAREKH, 2025).

A diversidade de métodos, bases de dados, arquiteturas e procedimentos de avaliação encontrada na literatura dificulta estabelecer uma conclusão única sobre o efeito do pré-processamento (SOLTANI *et al.*, 2024). Dessa forma, torna-se relevante avaliar essas estratégias de maneira controlada, mantendo constantes as demais condições experimentais. Esse procedimento permite verificar se as alterações introduzidas nas imagens resultam em diferenças mensuráveis no desempenho do modelo e se os possíveis ganhos justificam o aumento da complexidade do fluxo de processamento.

Neste trabalho, foram selecionados dois métodos de pré-processamento representativos de estratégias distintas encontradas na literatura. O Método A, baseado no Mammo-Bench, concentra-se na padronização das imagens por meio de operações como segmentação da região mamária, remoção do músculo peitoral e recorte da imagem (BHOLE; SUSEELA; PAREKH, 2025). O Método B, baseado no trabalho

de INGLE; VIDHATE; CHAUDHARI (2024), utiliza uma abordagem voltada à redução de ruído e ao realce de características, combinando filtragem Gaussiana e mascaramento de nitidez. A escolha de abordagens com objetivos distintos permitiu investigar, sob o mesmo protocolo experimental, se diferentes formas de preparação das mesmas imagens produziram alterações mensuráveis no desempenho da rede.

1.2 Definição do Problema

Em que medida diferentes técnicas de pré-processamento de imagens mamográficas influenciam o desempenho de uma arquitetura de aprendizado profundo na classificação de lesões benignas e malignas?

1.3 Objetivo Geral

Avaliar a influência de diferentes técnicas de pré-processamento de imagens mamográficas no desempenho da arquitetura EfficientNet-B0 para a classificação de lesões benignas e malignas.

1.4 Objetivos Específicos

- a) Preparar as imagens do conjunto DMID por meio da extração das regiões de interesse e da padronização das dimensões, organizando-as em três condições experimentais pareadas: Original, Método A, baseado no pré-processamento Mammo-Bench, e Método B, baseado na combinação de filtragem Gaussiana e mascaramento de nitidez;
- b) Treinar a arquitetura EfficientNet-B0 utilizando as três condições experimentais sob o mesmo protocolo experimental;
- c) Comparar o desempenho dos modelos por meio das métricas acurácia, sensibilidade, especificidade, precisão, F1-score, Coeficiente de Correlação de Matthews (MCC) e área sob a curva *Receiver Operating Characteristic* (ROC), denominada AUC (*Area Under the Curve*);
- d) Avaliar estatisticamente os resultados obtidos, investigando se os métodos de pré-processamento apresentam diferenças de desempenho nas condições experimentais adotadas.

1.5 Estrutura do Trabalho

Este trabalho está organizado em cinco capítulos. O primeiro capítulo apresenta a contextualização do tema, a definição do problema de pesquisa, a justificativa para o desenvolvimento do trabalho, bem como os objetivos geral e específicos.

O segundo capítulo aborda a fundamentação teórica necessária para a compreensão do trabalho, contemplando os principais conceitos relacionados ao câncer de mama, à mamografia digital, às Redes Neurais Convolucionais, ao aprendizado por transferência, à arquitetura EfficientNet-B0, às técnicas de pré-processamento de imagens e às métricas utilizadas na avaliação de modelos de classificação.

O terceiro capítulo descreve a metodologia adotada, apresentando o conjunto de dados utilizado, as etapas de preparação das imagens, os métodos de pré-processamento empregados, a configuração do ambiente de treinamento, a arquitetura da rede neural utilizada e os procedimentos adotados para avaliação do desempenho do modelo.

O quarto capítulo apresenta os resultados obtidos durante os experimentos, realizando a comparação entre os diferentes métodos de pré-processamento por meio das métricas de desempenho e da análise estatística, bem como a discussão dos resultados observados.

Por fim, o quinto capítulo apresenta as conclusões do trabalho, destacando as principais contribuições obtidas, as limitações encontradas durante o desenvolvimento da pesquisa e sugestões para trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta os conceitos fundamentais para a compreensão e o desenvolvimento deste trabalho, com foco na aplicação de técnicas de Inteligência Artificial ao diagnóstico assistido por computador em exames de mamografia. Inicialmente, são abordados aspectos relacionados ao câncer de mama e à mamografia digital, destacando sua importância para o diagnóstico precoce da doença. Em seguida, apresenta-se uma revisão dos principais conceitos de Inteligência Artificial aplicados ao processamento de imagens médicas, incluindo Redes Neurais Convolucionais, aprendizado por transferência e a arquitetura EfficientNet-B0, utilizada neste trabalho. Posteriormente, são discutidas as técnicas de pré-processamento de imagens mamográficas empregadas na literatura e as principais métricas utilizadas para avaliação do desempenho de modelos de classificação. Os conceitos apresentados neste capítulo fornecem a fundamentação teórica necessária para a compreensão da metodologia adotada e para a análise dos resultados experimentais apresentados nos capítulos seguintes.

2.1 Câncer de Mama

O câncer de mama é uma doença caracterizada pela multiplicação desordenada de células do tecido mamário, podendo originar tumores capazes de invadir estruturas próximas e se disseminar para outras regiões do organismo. Essas alterações geralmente se iniciam nos ductos responsáveis pelo transporte do leite ou nos lóbulos mamários, estruturas relacionadas à sua produção. A doença apresenta diferentes características biológicas e formas de evolução, podendo permanecer restrita ao local de origem ou desenvolver comportamento invasivo, com possibilidade de atingir os linfonodos e outros órgãos (WORLD HEALTH ORGANIZATION – WHO, 2026).

O desenvolvimento do câncer de mama está associado a diferentes fatores de risco, entre os quais se destacam o avanço da idade, o histórico familiar, determinadas alterações genéticas, a exposição hormonal, a obesidade, o consumo de bebidas alcoólicas e alguns hábitos relacionados ao estilo de vida. Entretanto, a presença de um ou mais desses fatores não determina necessariamente o desenvolvimento da doença, assim como a ausência de histórico familiar não elimina sua possibilidade. Dessa forma, o câncer de mama deve ser compreendido como uma doença multifatorial, resultante da interação entre características biológicas, genéticas, ambientais e comportamentais (WORLD HEALTH ORGANIZATION – WHO, 2026).

As alterações mamárias podem apresentar características benignas ou malignas. Lesões benignas não possuem comportamento canceroso e, geralmente, permanecem restritas ao tecido de origem. As lesões malignas, por sua vez, apresentam

crescimento desordenado e capacidade de invasão, podendo comprometer tecidos próximos e produzir metástases. A diferenciação entre essas condições exige investigação clínica e exames complementares, uma vez que determinadas alterações podem apresentar características semelhantes durante as etapas iniciais de avaliação (NATIONAL CANCER INSTITUTE – NCI, 2025).

A detecção precoce desempenha papel fundamental no controle da doença, pois permite identificar alterações em estágios iniciais, aumentando a possibilidade de tratamentos menos agressivos e de resultados clínicos mais favoráveis. As principais estratégias empregadas são o diagnóstico precoce, direcionado às pessoas que apresentam sinais ou sintomas suspeitos, e o rastreamento de indivíduos sem manifestações aparentes. Nesse contexto, a mamografia constitui um dos principais recursos para a identificação de alterações mamárias antes que se tornem clinicamente perceptíveis (INSTITUTO NACIONAL DE CÂNCER – INCA, 2025).

A interpretação dos exames requer a análise de características como formato, margens, densidade e distribuição das alterações encontradas. Entretanto, a sobreposição dos tecidos, a baixa diferença de contraste entre determinadas estruturas e a diversidade de apresentação das lesões podem dificultar sua avaliação. Por esse motivo, a qualidade das imagens e os métodos utilizados para auxiliar sua análise possuem relevância direta no processo de diagnóstico. As características da mamografia digital e sua aplicação na detecção de lesões mamárias serão apresentadas na seção seguinte (NEWELL *et al.*, 2025).

2.2 Mamografia Digital

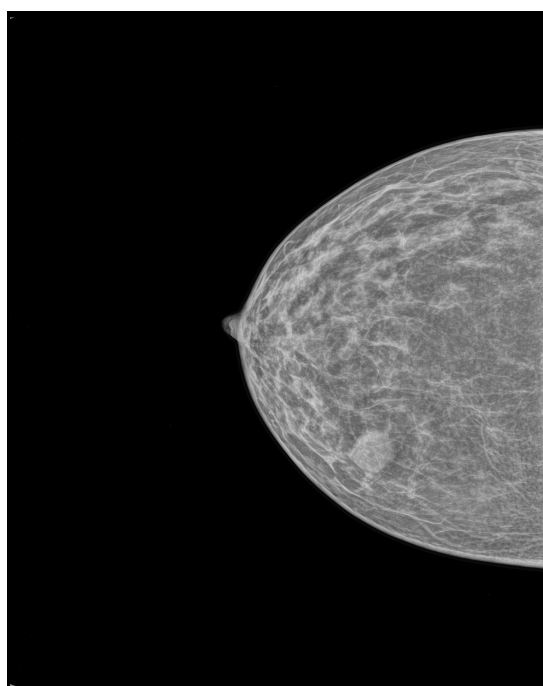
A mamografia é uma técnica de diagnóstico por imagem que utiliza raios X de baixa dose para produzir representações das estruturas internas da mama. O exame pode ser empregado tanto no rastreamento de indivíduos sem sinais ou sintomas aparentes quanto na investigação de alterações identificadas durante a avaliação clínica. Sua capacidade de revelar lesões ainda não perceptíveis por palpação torna essa modalidade uma ferramenta relevante para a detecção precoce do câncer de mama (UNITED STATES FOOD AND DRUG ADMINISTRATION – FDA, 2026).

Nos sistemas convencionais, a imagem mamográfica é registrada em filmes radiográficos que precisam ser revelados antes da análise. Na mamografia digital, o filme é substituído por detectores de estado sólido, capazes de converter os raios X recebidos em sinais elétricos utilizados para a formação da imagem. Essa representação digital facilita o armazenamento, a transmissão e a visualização dos exames, além de permitir a aplicação de ajustes de contraste, ampliação de regiões específicas e técnicas computacionais de processamento de imagens (UNITED STATES FOOD AND DRUG ADMINISTRATION – FDA, 2026).

Durante a aquisição do exame, a mama é posicionada e comprimida entre duas superfícies. Esse procedimento reduz a espessura do tecido, limita os movimentos durante a exposição e diminui a sobreposição das estruturas anatômicas. Como a mamografia produz uma representação bidimensional de uma estrutura tridimensional, a aquisição de imagens em diferentes projeções permite observar a mama a partir de perspectivas complementares (NEWELL *et al.*, 2025).

Entre as projeções utilizadas na mamografia, destacam-se a craniocaudal (CC) e a mediolateral oblíqua (MLO). Na projeção CC, a mama é comprimida horizontalmente e examinada em uma direção aproximadamente superior-inferior. Essa projeção permite observar principalmente as regiões medial e lateral do tecido mamário. Na projeção MLO, a aquisição é realizada em direção oblíqua, possibilitando a inclusão de uma porção maior do tecido mamário, da região axilar e do músculo peitoral. A utilização conjunta dessas projeções contribui para a localização dos achados e para a diferenciação entre alterações reais e possíveis sobreposições de tecidos (NEWELL *et al.*, 2025).

A Figura 1 apresenta exemplos das projeções CC e MLO extraídos do conjunto de dados *Digital Mammography Image Dataset* (DMID). O conjunto disponibiliza imagens mamográficas acompanhadas de informações clínicas e anotações relacionadas às regiões que apresentam alterações (OZA *et al.*, 2024). Embora as imagens mostradas correspondam a exames distintos, sua apresentação conjunta permite visualizar as diferenças de posicionamento e orientação entre as duas projeções.



(a) Projeção craniocaudal (CC).



(b) Projeção mediolateral oblíqua (MLO).

Figura 1 – Exemplos de projeções mamográficas presentes no conjunto de dados DMID

Fonte: Adaptado de OZA *et al.* (2024).

Na Figura 1a, observa-se uma projeção craniocaudal, na qual a mama apresenta orientação predominantemente horizontal. Na Figura 1b, observa-se uma projeção mediolateral oblíqua, caracterizada pela orientação inclinada da mama e pela presença de estruturas próximas à parede torácica. Essas diferenças modificam a distribuição espacial dos tecidos e são relevantes para o processamento computacional das imagens.

Entre os principais achados analisados em uma mamografia encontram-se massas, calcificações, assimetrias e distorções da arquitetura mamária. A descrição dessas alterações considera características como forma, margem, densidade, distribuição e localização. Para padronizar a interpretação e a comunicação dos resultados, o *American College of Radiology* desenvolveu o sistema *Breast Imaging Reporting and Data System* (BI-RADS). Esse sistema estabelece uma terminologia comum, organiza os relatórios e classifica os achados de acordo com o grau de suspeita e a conduta recomendada (NEWELL *et al.*, 2025).

Apesar de sua relevância, a mamografia apresenta limitações associadas às características do tecido mamário e à forma como as estruturas são projetadas na imagem. Em mamas densas, os tecidos fibroglandulares e determinadas lesões podem apresentar intensidades semelhantes, reduzindo o contraste e dificultando a identificação de alterações suspeitas. Além da densidade mamária, fatores como tamanho e localização da lesão, qualidade da aquisição e experiência do profissional responsável pela interpretação podem influenciar a sensibilidade do exame (NATIONAL CANCER INSTITUTE – NCI, 2025).

A interpretação das imagens mamográficas exige, portanto, a análise conjunta de diferentes características visuais, muitas vezes sutis e sujeitas à sobreposição dos tecidos. A disponibilidade dos exames em formato digital possibilitou o desenvolvimento de sistemas de diagnóstico assistido por computador, capazes de processar as imagens e identificar padrões potencialmente relevantes. Nesse contexto, técnicas de Inteligência Artificial passaram a ser investigadas como ferramentas de apoio à classificação e à tomada de decisão, conforme apresentado na seção seguinte.

2.3 Inteligência Artificial Aplicada ao Diagnóstico Médico

A Inteligência Artificial (IA) compreende um conjunto de técnicas computacionais desenvolvidas para executar tarefas que normalmente dependem de capacidades humanas, como reconhecimento de padrões, interpretação de informações e tomada de decisão. No contexto médico, essas técnicas podem ser utilizadas para analisar dados clínicos, laboratoriais e imagens de exames, auxiliando na identificação de alterações que possam estar associadas a determinadas doenças (LITJENS *et al.*, 2017).

Uma das principais áreas da Inteligência Artificial é o aprendizado de máquina, ou *Machine Learning*, no qual modelos computacionais aprendem relações a partir de exemplos previamente apresentados. Em vez de depender exclusivamente de regras definidas manualmente, o algoritmo ajusta seus parâmetros com base nos dados utilizados durante o treinamento. Quando são empregadas redes neurais compostas por diversas camadas de processamento, essa abordagem é denominada aprendizado profundo, ou *Deep Learning* (LECUN; BENGIO; HINTON, 2015).

Na análise de imagens médicas, os modelos de aprendizado profundo podem ser empregados em tarefas como classificação, detecção e segmentação. A classificação atribui uma categoria à imagem ou à região analisada, enquanto a detecção busca identificar a presença e a localização de uma determinada alteração. A segmentação, por sua vez, delimita os pixels pertencentes a uma estrutura anatômica ou lesão (LITJENS *et al.*, 2017).

Em mamografia, sistemas baseados em Inteligência Artificial podem analisar características relacionadas à forma, textura, densidade e distribuição das estruturas presentes no exame. Estudos realizados com conjuntos de dados de diferentes populações demonstram o potencial desses sistemas para apoiar a interpretação das imagens e reduzir determinados erros de classificação em condições experimentais específicas (MCKINNEY *et al.*, 2020). Entretanto, esses modelos não devem ser considerados substitutos dos profissionais da saúde, uma vez que seu desempenho depende da representatividade dos dados, da qualidade das anotações e das condições nas quais foram desenvolvidos e avaliados.

A utilização clínica de modelos de Inteligência Artificial também exige atenção a aspectos como capacidade de generalização, transparência, segurança e possibilidade de reprodução dos resultados. Um modelo pode apresentar elevado desempenho no conjunto de dados utilizado durante seu desenvolvimento e, ainda assim, obter resultados inferiores quando aplicado a imagens provenientes de outros equipamentos, instituições ou grupos populacionais. Dessa forma, a avaliação deve empregar procedimentos que reduzam o risco de sobreajuste e permitam analisar a capacidade de generalização do modelo (LITJENS *et al.*, 2017).

Entre as técnicas de aprendizado profundo utilizadas na análise de imagens, destacam-se as Redes Neurais Convolucionais, desenvolvidas especificamente para extrair padrões espaciais e visuais. Suas características e seu funcionamento são apresentados na seção seguinte.

2.4 Redes Neurais Convolucionais

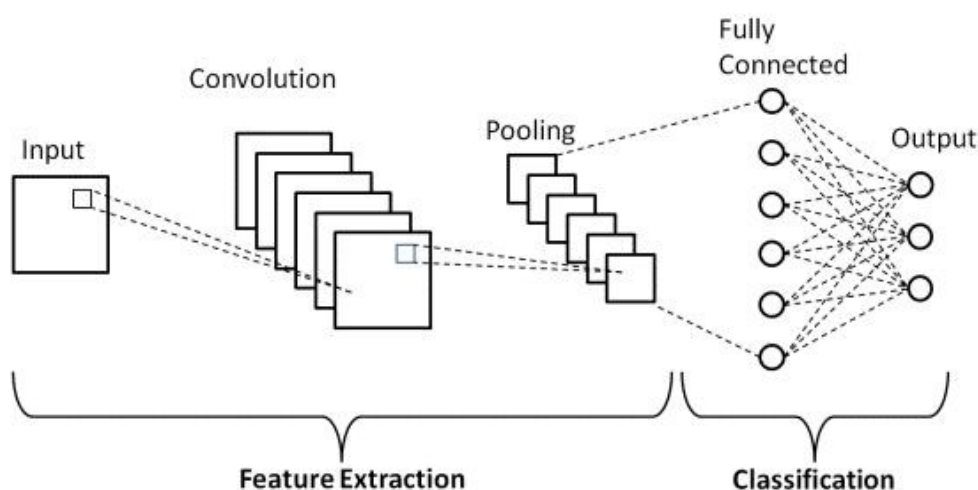
As Redes Neurais Convolucionais, ou *Convolutional Neural Networks* (CNNs), constituem uma classe de redes neurais desenvolvida para processar dados organiza-

dos espacialmente, especialmente imagens. Diferentemente de abordagens que dependem da definição manual das características relevantes, as CNNs são capazes de aprender filtros diretamente a partir dos exemplos apresentados durante o treinamento (LECUN *et al.*, 1998).

De maneira geral, o funcionamento de uma CNN pode ser dividido em duas etapas principais: a extração de características e a classificação. Na primeira etapa, a imagem de entrada é processada por camadas convolucionais e operações de redução espacial, responsáveis por identificar e organizar os padrões visuais presentes nos dados. Na segunda etapa, as características extraídas são fornecidas às camadas classificadoras, que produzem a saída correspondente às classes do problema analisado.

A Figura 2 apresenta uma representação simplificada desse processo. Inicialmente, a imagem é fornecida à rede como entrada. Em seguida, as camadas de convolução geram diferentes mapas de características, enquanto as operações de *pooling* reduzem suas dimensões espaciais. Por fim, as informações extraídas são encaminhadas às camadas totalmente conectadas, responsáveis pela classificação e pela geração da saída da rede (LECUN *et al.*, 1998). É importante destacar que a Figura 2 representa uma CNN genérica e simplificada. No experimento desenvolvido neste trabalho, o pooling não foi tratado como um hiperparâmetro a ser modificado ou comparado. As operações de redução espacial seguiram a estrutura definida pela própria implementação da EfficientNet-B0, mantida constante nas três condições experimentais.

Figura 2 – Estrutura simplificada de uma Rede Neural Convolucional



Fonte: COUTO (2023).

Conforme ilustrado na Figura 2, o principal componente da etapa de extração de características é a camada de convolução. Nessa operação, pequenos filtros,

também denominados núcleos ou *kernels*, percorrem a imagem de entrada e produzem mapas de características. Cada filtro possui pesos ajustados durante o treinamento e passa a responder a determinados padrões visuais.

Nas primeiras camadas da rede, os filtros geralmente identificam características simples, como bordas, linhas e variações de intensidade. À medida que as informações avançam pelas camadas, essas características são combinadas e utilizadas para representar padrões progressivamente mais complexos, como formas, texturas e estruturas presentes na imagem (LECUN *et al.*, 1998).

Após as operações de convolução, geralmente são aplicadas funções de ativação não lineares, permitindo que a rede represente relações mais complexas entre os dados. Uma função amplamente utilizada é a Unidade Linear Retificada, conhecida como ReLU, que mantém os valores positivos e substitui os valores negativos por zero. Essa operação introduz não linearidade no modelo e contribui para o ajuste dos parâmetros durante o treinamento (LECUN; BENGIO; HINTON, 2015).

As CNNs também podem empregar operações de redução espacial, como o *pooling*, representado na região central da Figura 2. Essas operações diminuem as dimensões dos mapas de características, reduzindo a quantidade de informações que precisam ser processadas pelas camadas seguintes. Além de limitar o custo computacional, essa redução pode tornar a representação menos sensível a pequenas variações na posição dos elementos da imagem.

Entretanto, reduções espaciais excessivas podem eliminar detalhes relevantes. Essa limitação deve ser considerada principalmente na análise de imagens médicas, nas quais estruturas pequenas, bordas sutis e diferenças reduzidas de contraste podem possuir importância para a classificação.

Após a etapa de extração de características, os mapas produzidos pela rede são convertidos em uma representação utilizada pelas camadas classificadoras. Na arquitetura simplificada da Figura 2, essa etapa é representada pelas camadas totalmente conectadas, nas quais cada neurônio recebe informações provenientes da camada anterior.

Ao final da rede, a camada de saída produz valores associados às classes analisadas. Em uma classificação binária de lesões mamárias, por exemplo, a saída pode representar as probabilidades de a região processada pertencer às classes benigna ou maligna. Durante o treinamento, os pesos das diferentes camadas são ajustados de acordo com a diferença entre as previsões produzidas e os rótulos corretos das imagens.

A estrutura apresentada na Figura 2 representa uma CNN genérica e simplificada. Arquiteturas modernas podem incorporar outros componentes, como normalização, conexões residuais, mecanismos de atenção e diferentes formas de agrega-

ção das características. Apesar dessas variações, o princípio de utilizar camadas para extrair padrões visuais e posteriormente realizar a classificação permanece presente.

O treinamento completo de uma CNN pode exigir grande quantidade de dados e recursos computacionais. Essa limitação é especialmente relevante em aplicações médicas, nas quais os conjuntos de imagens são geralmente menores e dependem de anotações produzidas por especialistas. Uma estratégia utilizada para reduzir essa dificuldade é o aprendizado por transferência, apresentado na seção seguinte.

2.5 Aprendizado por Transferência

O aprendizado por transferência, ou *Transfer Learning*, é uma estratégia na qual o conhecimento obtido durante o treinamento de um modelo em uma tarefa de origem é reutilizado em uma nova tarefa. Em aplicações de visão computacional, é comum utilizar redes previamente treinadas em grandes conjuntos de imagens e adaptá-las para problemas com menor disponibilidade de dados.

Um dos conjuntos mais utilizados para esse treinamento inicial é o ImageNet, que reúne imagens organizadas em diferentes categorias e foi desenvolvido para apoiar pesquisas relacionadas ao reconhecimento visual em larga escala (DENG *et al.*, 2009). O treinamento em uma base extensa permite que as primeiras camadas das redes aprendam representações visuais genéricas, como bordas, contornos, cores e texturas.

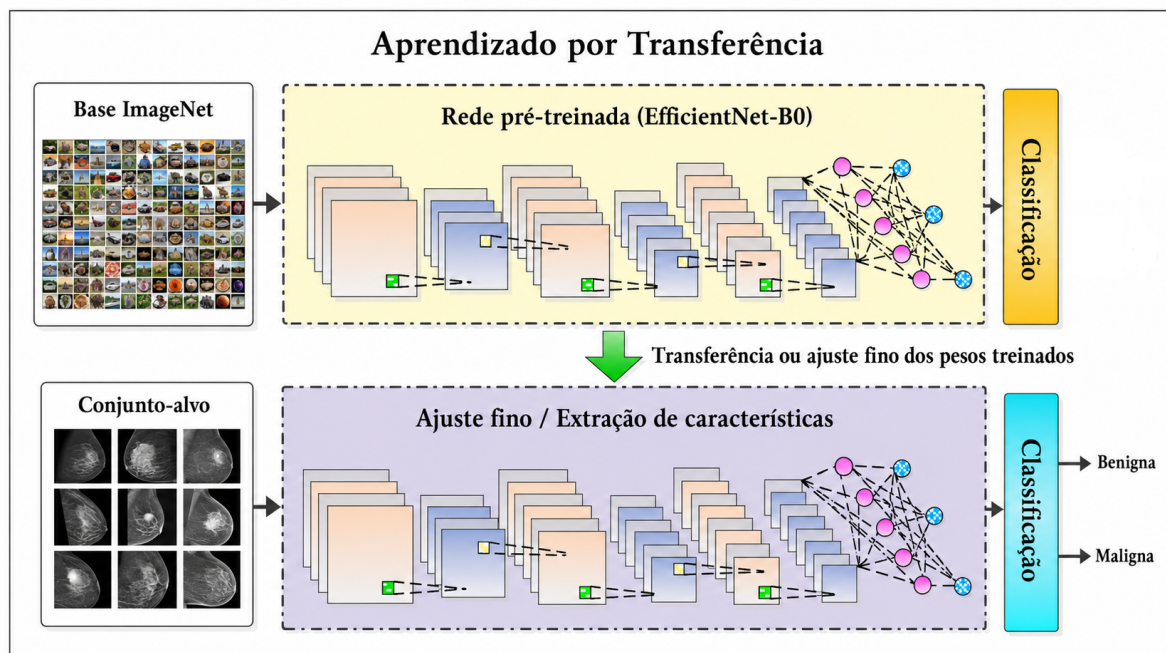
As características aprendidas nas camadas iniciais tendem a ser mais gerais, enquanto as camadas finais se tornam progressivamente especializadas na tarefa de origem. Estudos sobre a transferibilidade das representações demonstram que o reaproveitamento dos pesos pode facilitar a otimização do modelo e melhorar sua capacidade de generalização, mesmo quando existem diferenças entre o conjunto de origem e a aplicação de destino (YOSINSKI *et al.*, 2014).

Para adaptar um modelo pré-treinado, a camada de classificação original é substituída por uma nova camada correspondente às classes do problema estudado. Posteriormente, pode-se manter parte da rede congelada, atualizando apenas o novo classificador, ou realizar o ajuste fino, denominado *fine-tuning*, no qual os pesos de outras camadas também são atualizados. A escolha entre essas abordagens depende da quantidade de dados disponível, da similaridade entre as tarefas e do risco de sobreajuste.

A Figura 3 apresenta uma representação simplificada da aplicação do aprendizado por transferência à classificação de imagens mamográficas. Na parte superior, a arquitetura EfficientNet-B0 é inicialmente treinada no conjunto ImageNet, permitindo o aprendizado de características visuais genéricas. Em seguida, os pesos obtidos nessa etapa são transferidos para a nova tarefa, na qual a rede é ajustada utilizando

imagens mamográficas.

Figura 3 – Aplicação do aprendizado por transferência na classificação de imagens mamográficas



Fonte: Elaborado pelo autor com base em ZHANG *et al.* (2019).

Conforme ilustrado na Figura 3, o processo possui uma tarefa de origem e uma tarefa de destino. Na tarefa de origem, o treinamento realizado no ImageNet permite que a rede aprenda filtros capazes de identificar características presentes em diferentes tipos de imagens. Na tarefa de destino, esses filtros são reutilizados como ponto de partida para a extração de características das mamografias.

Durante a adaptação, a camada classificadora utilizada na tarefa de origem é substituída por uma nova camada correspondente às classes benigna e maligna. Os parâmetros previamente aprendidos podem ser mantidos congelados ou atualizados durante o treinamento com o conjunto de imagens mamográficas. No ajuste fino, geralmente é empregada uma taxa de aprendizado reduzida, evitando alterações abruptas nos pesos já aprendidos.

Embora imagens naturais e mamografias apresentem características distintas, os filtros genéricos aprendidos nas primeiras camadas ainda podem fornecer uma inicialização mais adequada do que pesos aleatórios. As camadas posteriores são então ajustadas para reconhecer padrões específicos das imagens médicas, como texturas, margens, formas e diferenças de densidade relacionadas às lesões.

No presente trabalho, o aprendizado por transferência foi empregado por meio da arquitetura EfficientNet-B0, inicializada com pesos obtidos no ImageNet (TAN; LE, 2019). A camada de classificação original foi substituída por uma nova estrutura

destinada à classificação binária de lesões benignas e malignas. Essa abordagem permite aproveitar representações previamente aprendidas e reduz a necessidade de treinar a rede a partir de uma inicialização completamente aleatória.

2.6 Arquitetura EfficientNet

A EfficientNet corresponde a uma família de Redes Neurais Convolucionais desenvolvida com o objetivo de melhorar a relação entre desempenho, número de parâmetros e custo computacional. Sua principal contribuição é uma estratégia de escalonamento composto, na qual profundidade, largura e resolução de entrada são ampliadas de maneira coordenada (TAN; LE, 2019).

A profundidade está relacionada à quantidade de camadas da rede, enquanto a largura representa o número de canais utilizados em cada estágio. A resolução determina as dimensões espaciais da imagem processada. Aumentar apenas um desses elementos pode melhorar a capacidade do modelo, mas também pode produzir uma arquitetura desequilibrada ou elevar excessivamente o custo computacional. O escalonamento composto busca estabelecer uma proporção entre essas dimensões.

A EfficientNet-B0 constitui o modelo de referência da família. A partir dela foram desenvolvidas versões progressivamente maiores, denominadas EfficientNet-B1 até EfficientNet-B7. Modelos superiores apresentam maior quantidade de camadas, canais e resolução de referência, resultando em maior capacidade de representação, mas também em maior demanda de memória e processamento.

A arquitetura utiliza blocos convolucionais móveis invertidos, conhecidos como MBConv, derivados da estrutura proposta na MobileNetV2. Esses blocos empregam convoluções separáveis em profundidade e conexões residuais com o objetivo de reduzir o número de operações sem eliminar a capacidade de extração de características (SANDLER *et al.*, 2018).

Os blocos também incorporam o mecanismo *Squeeze-and-Excitation*, responsável por recalibrar a importância dos canais dos mapas de características. Esse mecanismo analisa globalmente as ativações produzidas e atribui pesos diferentes aos canais, permitindo que a rede enfatize informações consideradas mais relevantes para a tarefa (HU; SHEN; SUN, 2018).

A EfficientNet-B0 foi selecionada neste trabalho por oferecer equilíbrio entre capacidade de classificação e custo computacional. Essa característica é relevante diante da resolução das imagens mamográficas e das limitações de processamento do ambiente utilizado. A utilização de pesos pré-treinados também possibilita sua aplicação em conjuntos de dados menores por meio do aprendizado por transferência.

Antes de serem fornecidas à arquitetura, as imagens precisam ser organizadas em dimensões compatíveis e podem ser submetidas a procedimentos de prepa-

ração e aprimoramento. Essas operações compõem a etapa de pré-processamento, discutida na próxima seção.

2.7 Pré-processamento de Imagens Mamográficas

O pré-processamento compreende um conjunto de operações aplicadas às imagens antes de sua utilização pelos algoritmos de classificação. Seus objetivos podem incluir padronização das dimensões, remoção de regiões sem informação, redução de ruído, correção de intensidade e realce de características consideradas relevantes (SOLTANI *et al.*, 2024).

Imagens mamográficas podem apresentar diferenças relacionadas ao equipamento de aquisição, formato dos arquivos, resolução, contraste e presença de elementos externos à região mamária. Essas variações podem dificultar o treinamento do modelo e favorecer o aprendizado de características relacionadas à origem das imagens em vez de padrões associados às alterações de interesse (SOLTANI *et al.*, 2024).

Uma das operações utilizadas consiste na extração da Região de Interesse, ou ROI, que limita a imagem à região que contém a alteração previamente identificada. O recorte reduz a quantidade de fundo e mantém a lesão em maior proporção na entrada da rede. A inclusão de uma margem ao redor da ROI foi adotada neste trabalho para preservar parte do tecido adjacente à alteração.

Outra etapa importante é a padronização das dimensões. As redes neurais processam imagens agrupadas em lotes, exigindo que as amostras possuam dimensões iguais. O redimensionamento possibilita essa uniformização, mas deve ser aplicado com cautela, pois reduções acentuadas podem eliminar estruturas pequenas e ampliações excessivas não acrescentam novas informações. Métodos recentes de preparação de bases mamográficas empregam operações como segmentação da mama, remoção do músculo peitoral e recorte inteligente para aumentar a consistência entre imagens provenientes de diferentes bases (BHOLE; SUSEELA; PAREKH, 2025).

A redução de ruído busca atenuar variações indesejadas de intensidade que possam dificultar a visualização das estruturas. Entre os métodos existentes encontra-se o filtro Gaussiano, que realiza uma suavização baseada na média ponderada dos pixels vizinhos. Apesar de reduzir componentes de alta frequência, uma suavização excessiva pode diminuir a visibilidade de bordas e detalhes relevantes (INGLE; VIDHATE; CHAUDHARI, 2024).

Matematicamente, um núcleo Gaussiano bidimensional pode ser representado por:

$$G(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right), \quad (1)$$

em que σ determina a dispersão da função. A imagem suavizada $I_G(x, y)$ é obtida pela aplicação desse núcleo sobre a imagem de entrada $I(x, y)$.

No mascaramento de nitidez, uma máscara pode ser obtida a partir da diferença entre a imagem de entrada e sua versão suavizada:

$$M(x, y) = I(x, y) - I_G(x, y). \quad (2)$$

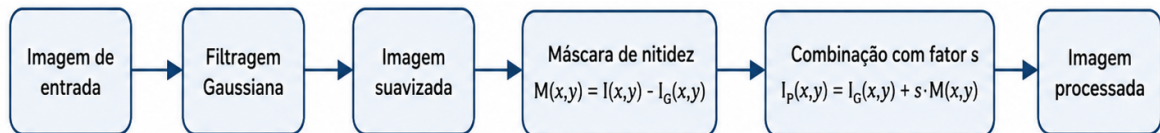
A máscara contém principalmente componentes modificados pela suavização e pode ser posteriormente utilizada para recuperar ou realçar detalhes da imagem. No procedimento avaliado neste trabalho, a imagem processada é obtida por:

$$I_P(x, y) = I_G(x, y) + s \cdot M(x, y), \quad (3)$$

em que s controla a intensidade aplicada ao mascaramento de nitidez (INGLE; VIDHATE; CHAUDHARI, 2024).

A Figura 4 apresenta de forma simplificada as etapas envolvidas nesse procedimento. Inicialmente, a imagem de entrada é suavizada por meio da filtragem Gaussiana. Em seguida, a diferença entre a imagem original e a imagem suavizada é utilizada para obter a máscara de nitidez, que é posteriormente combinada à imagem suavizada para gerar a imagem processada.

Figura 4 – Fluxograma simplificado da filtragem Gaussiana e do mascaramento de nitidez



Fonte: Elaborado pelo autor com base em INGLE; VIDHATE; CHAUDHARI (2024).

Embora o pré-processamento possa aumentar a uniformidade e melhorar visualmente as imagens, seus benefícios para a classificação não são garantidos. Uma operação pode destacar determinadas características, mas também modificar padrões originalmente presentes nos exames. Por esse motivo, a influência dos métodos deve ser avaliada experimentalmente, mantendo constantes as demais condições de treinamento.

A comparação entre diferentes versões das imagens exige a utilização de métricas capazes de representar tanto o desempenho geral quanto os diferentes tipos de erro produzidos pelo classificador. Essas métricas são apresentadas na seção seguinte.

2.8 Métricas de Avaliação

A avaliação de um modelo de classificação binária pode ser realizada a partir da matriz de confusão, que organiza as previsões de acordo com os rótulos reais. Considerando a classe maligna como positiva, os resultados são divididos em verdadeiros positivos (VP), verdadeiros negativos (VN), falsos positivos (FP) e falsos negativos (FN).

A Tabela 1 apresenta a organização dos quatro resultados possíveis em uma classificação binária.

Tabela 1 – Estrutura da matriz de confusão para classificação binária

Classe real	Predito maligno	Predito benigno
Maligno	Verdadeiro positivo (VP)	Falso negativo (FN)
Benigno	Falso positivo (FP)	Verdadeiro negativo (VN)

Fonte: Elaborado pelo autor (2026).

Os verdadeiros positivos correspondem às lesões malignas corretamente classificadas, enquanto os verdadeiros negativos representam as lesões benignas identificadas corretamente. Um falso positivo ocorre quando uma lesão benigna é classificada como maligna. O falso negativo ocorre quando uma lesão maligna é classificada como benigna, situação especialmente relevante em aplicações de apoio ao diagnóstico.

A acurácia representa a proporção total de previsões corretas e pode ser calculada por:

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN} \quad (4)$$

Embora seja de fácil interpretação, a acurácia pode produzir uma visão incompleta quando existe desequilíbrio entre as classes. Um modelo pode alcançar valor elevado ao favorecer a classe mais frequente, mesmo apresentando desempenho insuficiente na classe de maior interesse (SOKOLOVA; LAPALME, 2009).

A sensibilidade, também denominada revocação ou *recall*, representa a proporção de casos positivos corretamente identificados:

$$\text{Sensibilidade} = \frac{VP}{VP + FN} \quad (5)$$

Na classificação de lesões mamárias, essa métrica indica a capacidade do modelo de reconhecer os casos malignos. Valores baixos de sensibilidade estão associados a uma quantidade maior de falsos negativos.

A especificidade mede a proporção de casos negativos corretamente classificados:

$$\text{Especificidade} = \frac{VN}{VN + FP} \quad (6)$$

Essa métrica representa a capacidade do modelo de reconhecer lesões benignas. Uma especificidade reduzida indica maior ocorrência de falsos positivos.

A precisão indica a proporção de previsões positivas que realmente pertencem à classe positiva:

$$\text{Precisão} = \frac{VP}{VP + FP} \quad (7)$$

A precisão deve ser analisada em conjunto com a sensibilidade, pois um modelo pode reduzir os falsos positivos adotando uma decisão mais restritiva, ao mesmo tempo que aumenta a quantidade de falsos negativos.

O F1-score corresponde à média harmônica entre precisão e sensibilidade:

$$F1 = 2 \cdot \frac{\text{Precisão} \cdot \text{Sensibilidade}}{\text{Precisão} + \text{Sensibilidade}} \quad (8)$$

Essa métrica busca equilibrar a capacidade de identificar os casos positivos e a confiabilidade das previsões positivas. Entretanto, o F1-score não considera diretamente os verdadeiros negativos.

O Coeficiente de Correlação de Matthews, ou MCC, utiliza os quatro componentes da matriz de confusão:

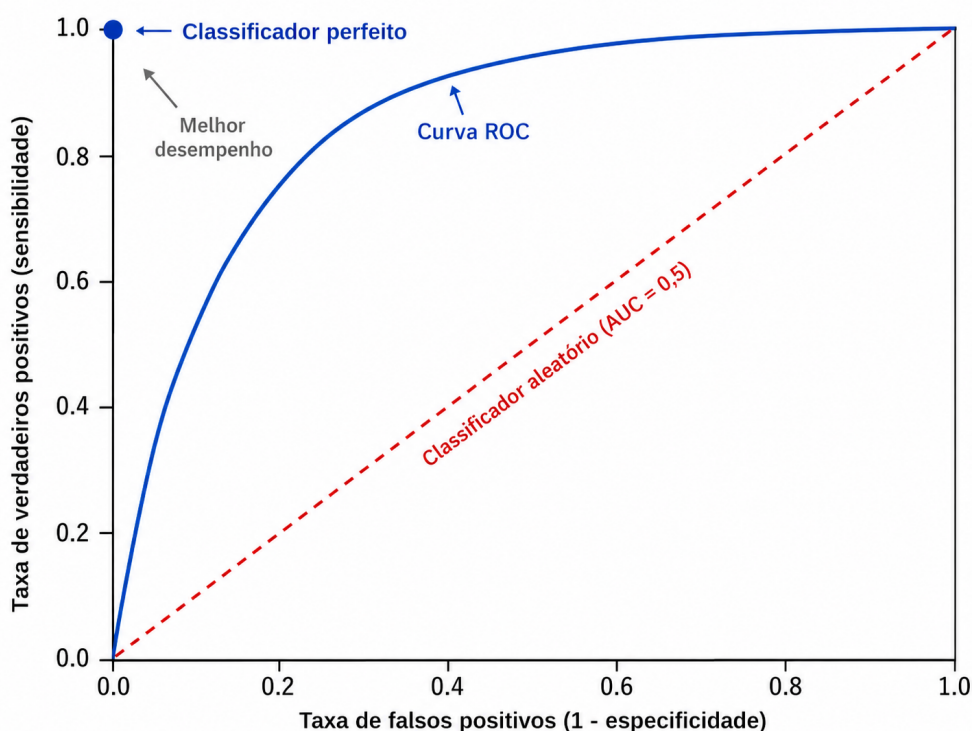
$$MCC = \frac{VP \cdot VN - FP \cdot FN}{\sqrt{(VP + FP)(VP + FN)(VN + FP)(VN + FN)}} \quad (9)$$

Seu valor varia entre -1 e 1 . Um resultado igual a 1 indica classificação perfeita, zero representa desempenho equivalente a uma previsão sem correlação e -1 corresponde à discordância completa entre as previsões e os rótulos. Por considerar todos os elementos da matriz de confusão, o MCC é especialmente útil na avaliação de problemas binários com distribuição desigual entre as classes (CHICCO; JURMAN, 2020).

A curva ROC (*Receiver Operating Characteristic*) representa a relação entre a taxa de verdadeiros positivos e a taxa de falsos positivos para diferentes limiares de decisão. A área sob essa curva, denominada AUC (*Area Under the Curve*), resume a capacidade de discriminação do modelo em um único valor. Quanto mais próxima de 1, maior é a capacidade do classificador de atribuir probabilidades superiores aos casos positivos em relação aos negativos (FAWCETT, 2006).

A Figura 5 apresenta uma representação simplificada da curva ROC. A linha diagonal representa o desempenho esperado de um classificador sem capacidade discriminativa, correspondente a uma AUC igual a 0,5. À medida que a curva se aproxima do canto superior esquerdo, aumenta a capacidade do modelo de obter elevada sensibilidade com menor taxa de falsos positivos. Consequentemente, valores de AUC mais próximos de 1 indicam maior capacidade de discriminação entre as classes (FAWCETT, 2006).

Figura 5 – Representação simplificada da curva ROC



Fonte: Elaborado pelo autor com base em FAWCETT (2006).

Nenhuma métrica isolada descreve completamente o comportamento de um classificador. Por esse motivo, a utilização conjunta de acurácia, sensibilidade, especificidade, precisão, F1-score, MCC e AUC permite analisar diferentes aspectos do desempenho e identificar possíveis desequilíbrios entre os tipos de erro.

A Tabela 2 sintetiza a interpretação das métricas utilizadas neste trabalho.

Tabela 2 – Resumo das métricas utilizadas na avaliação do modelo

Métrica	Interpretação
Acurácia	Proporção total de amostras classificadas corretamente.
Sensibilidade	Capacidade do modelo de identificar corretamente as lesões malignas.
Especificidade	Capacidade do modelo de identificar corretamente as lesões benignas.
Precisão	Proporção das classificações malignas que correspondem efetivamente a lesões malignas.
F1-score	Média harmônica entre precisão e sensibilidade.
MCC	Medida de correlação que considera todos os elementos da matriz de confusão.
AUC	Capacidade do modelo de distinguir as classes considerando diferentes limiares de decisão.

Fonte: Elaboração própria com base em SOKOLOVA; LAPALME (2009), FAWCETT (2006) e CHICCO; JURMAN (2020).

2.9 Teste de Friedman

O teste de Friedman é um procedimento estatístico não paramétrico desenvolvido para comparar condições relacionadas. Ele é empregado quando as mesmas unidades experimentais são avaliadas sob diferentes tratamentos ou métodos e quando não se deseja assumir que os resultados seguem uma distribuição normal (FRIEDMAN, 1937; CONOVER, 1999).

O teste organiza as observações em blocos. Cada bloco contém um resultado para cada uma das condições avaliadas. Os valores são convertidos em postos dentro de cada bloco, de modo que a análise seja baseada na ordenação relativa dos resultados, e não diretamente em suas magnitudes. Em caso de empate, são atribuídos postos médios (CONOVER, 1999).

Considerando n blocos e k condições, seja R_j a soma dos postos atribuídos à condição j . A estatística do teste pode ser expressa por:

$$\chi_F^2 = \frac{12}{nk(k+1)} \sum_{j=1}^k R_j^2 - 3n(k+1), \quad (10)$$

em que χ_F^2 representa a estatística de Friedman. Quando existem empates, pode ser aplicada uma correção à estatística calculada (CONOVER, 1999).

A hipótese nula estabelece que as condições comparadas apresentam o mesmo comportamento. A hipótese alternativa considera que pelo menos uma delas apresenta comportamento distinto. Para amostras de tamanho adequado, a estatística pode ser aproximada por uma distribuição qui-quadrado com $k - 1$ graus de liberdade.

A decisão é realizada por meio da comparação entre o p-valor e o nível de significância adotado. Quando $p < \alpha$, a hipótese nula é rejeitada. Quando $p \geq \alpha$, não existem evidências estatísticas suficientes para rejeitá-la.

O teste de Friedman indica a presença de uma diferença global, mas não determina diretamente entre quais condições essa diferença ocorre. Quando o resultado global é significativo, podem ser realizadas comparações posteriores entre pares de condições. O teste também é empregado na comparação de métodos de aprendizado de máquina quando os desempenhos são obtidos sob condições relacionadas (DEMŠAR, 2006).

2.10 Resumo do Capítulo

Este capítulo apresentou os principais conceitos necessários para a compreensão do trabalho. Inicialmente, foram discutidos os fundamentos relacionados ao câncer de mama e à mamografia digital, destacando a importância desses exames para a identificação de alterações mamárias.

Em seguida, foram apresentados os conceitos de Inteligência Artificial aplicada ao diagnóstico médico, Redes Neurais Convolucionais e aprendizado por transferência. Também foram descritas as principais características da arquitetura EfficientNet-B0 e os mecanismos utilizados para equilibrar capacidade de representação e custo computacional.

Posteriormente, foram abordadas as etapas de pré-processamento de imagens mamográficas e sua possível influência no treinamento dos modelos. Por fim, foram descritas as métricas utilizadas para avaliar o desempenho da classificação e os fundamentos do teste não paramétrico de Friedman, empregado na comparação estatística das condições experimentais. A fundamentação apresentada fornece a base necessária para a descrição da metodologia e para a interpretação dos resultados.

3 METODOLOGIA

Este capítulo apresenta os procedimentos adotados para avaliar a influência de diferentes métodos de pré-processamento no desempenho de uma Rede Neural Convolucional aplicada à classificação de lesões mamárias. Inicialmente, são descritas a natureza da pesquisa e as características do conjunto de dados utilizado. Em seguida, são detalhadas as etapas de seleção das imagens, extração das regiões de interesse e padronização das dimensões.

Posteriormente, são apresentados os métodos de pré-processamento empregados para gerar as diferentes versões do conjunto de dados. Também são descritas a arquitetura EfficientNet-B0, as configurações utilizadas durante o treinamento e o procedimento de validação experimental. Por fim, são apresentadas as métricas de desempenho e a análise estatística utilizada para comparar os resultados obtidos.

A metodologia foi estruturada de forma que as condições de treinamento permanecessem constantes entre os experimentos. Dessa forma, foram mantidos constantes entre os experimentos a arquitetura, os hiperparâmetros de treinamento e as divisões dos dados, permitindo realizar uma comparação controlada da influência dos métodos de pré-processamento sobre o desempenho do modelo.

3.1 Caracterização da Pesquisa

Este trabalho caracteriza-se como uma pesquisa aplicada, de abordagem quantitativa e natureza experimental. Seu caráter aplicado está relacionado ao emprego de técnicas de processamento de imagens e aprendizado profundo na investigação da classificação computacional de lesões mamárias. A abordagem quantitativa decorre da utilização de métricas numéricas e de análise estatística para comparar o desempenho obtido nas diferentes condições experimentais.

O experimento foi estruturado como uma comparação pareada entre três versões das mesmas imagens provenientes do DMID. A condição Original foi utilizada como referência, o Método A correspondeu ao pré-processamento proposto no Mammo-Bench (Bhole; SUSEELA; PAREKH, 2025), e o Método B foi baseado no procedimento de redução de ruído apresentado por INGLE; VIDHATE; CHAUDHARI (2024). A origem e a aplicação de cada condição são detalhadas na Seção 3.4.

A variável modificada entre os experimentos foi, portanto, a condição de pré-processamento aplicada às imagens. Foram mantidos constantes a composição do conjunto, as regiões de interesse, os rótulos, as dimensões das imagens, as divisões de treinamento e validação, a arquitetura EfficientNet-B0 e os hiperparâmetros de treinamento. Os métodos empregados e a origem de cada condição experimental são detalhados na Seção 3.4.

3.2 Conjunto de Dados

A base de origem utilizada neste trabalho foi o *Digital Mammography Image Dataset* (DMID), que disponibiliza imagens de mamografia digital acompanhadas de informações clínicas, classificações diagnósticas e coordenadas relacionadas às regiões que contêm as alterações mamárias (OZA *et al.*, 2024).

Os arquivos de imagem empregados nos experimentos foram obtidos a partir do subconjunto DMID disponibilizado no repositório do Mammo-Bench. O Mammo-Bench reúne e padroniza imagens provenientes de diferentes bases públicas de mamografia, entre as quais está o DMID, organizando-as em pastas correspondentes às imagens originais e às imagens resultantes do procedimento de pré-processamento proposto pelos autores (BHOLE; SUSEELA; PAREKH, 2025).

Neste trabalho, não foram utilizadas imagens provenientes das demais bases que compõem o Mammo-Bench. Foram selecionadas exclusivamente as amostras identificadas como pertencentes ao DMID. Dessa forma, o Mammo-Bench não constituiu uma segunda base de dados do experimento, mas o repositório a partir do qual foram obtidas duas versões correspondentes das imagens do DMID: uma identificada como original e outra submetida ao fluxo de pré-processamento proposto no artigo.

O repositório do Mammo-Bench associa identificadores únicos às imagens e disponibiliza arquivos de metadados destinados ao relacionamento entre as imagens padronizadas e suas respectivas bases de origem (BHOLE; SUSEELA; PAREKH, 2025). Esses identificadores foram utilizados para localizar as versões original e pré-processada correspondentes a cada amostra do DMID selecionada para o experimento.

Foram mantidas somente as imagens associadas aos diagnósticos benigno e maligno. Amostras pertencentes a outras categorias foram excluídas, resultando em um problema de classificação binária. A classe benigna foi representada pelo rótulo 0, enquanto a classe maligna foi representada pelo rótulo 1.

A Tabela 3 apresenta a distribuição das 127 imagens selecionadas após a aplicação dos critérios de inclusão e exclusão. O conjunto final foi composto por 80 imagens benignas e 47 imagens malignas, correspondendo, respectivamente, a 63% e 37% das amostras utilizadas.

Tabela 3 – Distribuição das imagens utilizadas por classe

Classe	Rótulo	Número de imagens	Percentual
Benigna	0	80	63%
Maligna	1	47	37%
Total	—	127	100%

Fonte: Elaborado pelo autor (2026).

Cada amostra selecionada foi representada nas três condições experimentais definidas neste trabalho. As condições preservaram o mesmo identificador, o mesmo rótulo diagnóstico e a correspondência com a mesma lesão. Assim, as diferenças entre as condições experimentais foram associadas ao processamento aplicado às imagens, e não a alterações na composição das amostras.

3.3 Preparação das Imagens

A preparação das imagens compreendeu as etapas comuns de seleção das amostras, extração das regiões de interesse e padronização das dimensões. Essas etapas foram aplicadas às versões correspondentes das condições Original e Método A. Posteriormente, o Método B foi gerado a partir das regiões de interesse preparadas da condição Original.

3.3.1 Seleção e organização das amostras

Inicialmente, os registros pertencentes ao subconjunto DMID foram filtrados a partir dos metadados disponibilizados com as imagens. Foram mantidas somente as amostras pertencentes às classes benigna e maligna, conforme os critérios definidos na Seção 3.2.

A lista resultante continha o identificador de cada imagem e o respectivo rótulo diagnóstico. Esses identificadores foram utilizados para localizar, no repositório do Mammo-Bench, a versão original e a versão pré-processada correspondente a cada amostra do DMID.

Foi verificada a presença de cada identificador nas duas versões do repositório. Somente foram mantidas as imagens para as quais foi possível estabelecer correspondência entre:

- a versão identificada como original;
- a versão submetida ao procedimento Mammo-Bench;
- o rótulo diagnóstico disponibilizado pelo DMID.

A terceira condição experimental foi posteriormente produzida neste trabalho pela aplicação do método de redução de ruído às imagens da condição Original. Portanto, essa versão também preservou os mesmos identificadores e rótulos das outras duas condições.

Ao final da organização, cada lesão possuía três representações associadas ao mesmo identificador: Original, Método A e Método B. Essa estrutura permitiu

realizar uma comparação pareada, evitando que diferenças na quantidade de imagens, na distribuição das classes ou na seleção das lesões influenciassem os resultados.

3.3.2 Extração das regiões de interesse

As regiões de interesse, ou ROIs, foram extraídas utilizando as coordenadas disponibilizadas nas anotações do DMID. A caixa delimitadora original foi expandida em 10% de sua largura e de sua altura, preservando parte do tecido localizado ao redor da lesão. Essa margem foi incluída porque as características das regiões adjacentes, como contorno, transição de intensidade e relação com o tecido próximo, podem fornecer informações relevantes para a classificação.

Considerando uma caixa delimitadora definida pelos limites $x_{\min}$, $x_{\max}$, $y_{\min}$ e $y_{\max}$, sua largura w e sua altura h foram calculadas por:

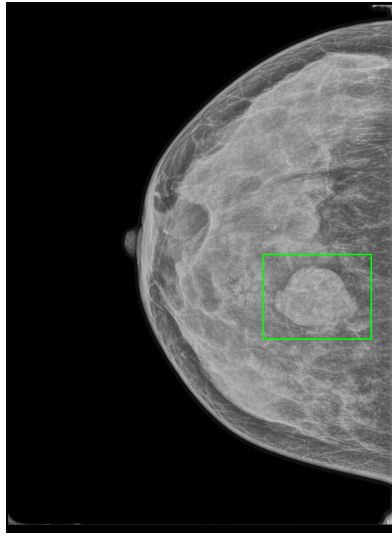
$$w = x_{\max} - x_{\min}, \quad h = y_{\max} - y_{\min}. \quad (11)$$

As margens horizontal e vertical foram determinadas por:

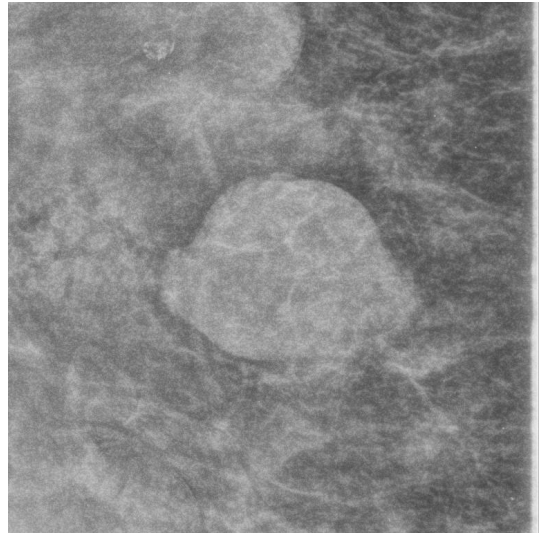
$$\Delta x = 0,10 \cdot w, \quad \Delta y = 0,10 \cdot h. \quad (12)$$

A nova região de recorte foi obtida pela expansão dos quatro limites da caixa original. Quando a margem ultrapassava os limites da imagem, as coordenadas eram restringidas às dimensões válidas do arquivo, evitando índices externos à área disponível.

A Figura 6 apresenta um exemplo da mamografia original com a caixa delimitadora da lesão e a região obtida após a inclusão da margem adicional de 10%. A utilização da mesma imagem permite visualizar diretamente a área preservada ao redor da lesão.



(a) Mamografia original com a delimitação da região de interesse.



(b) Região de interesse recortada com margem adicional de 10%.

Figura 6 – Exemplo do procedimento de extração da região de interesse

Fonte: Elaborado pelo autor (2026).

Conforme representado na Figura 6, o recorte reduz a quantidade de fundo e de tecido distante da alteração, ao mesmo tempo que mantém informações próximas às margens da lesão. Esse procedimento também aumenta a proporção ocupada pela região de interesse na entrada da rede neural.

3.3.3 Redimensionamento e padronização

Após a extração, todas as ROIs foram redimensionadas para 768×768 pixels. A utilização de dimensões uniformes foi necessária para organizar as imagens em lotes e fornecê-las à arquitetura de aprendizado profundo. O mesmo tamanho foi mantido nas três condições experimentais, evitando que diferenças de resolução influenciassem a comparação.

As imagens foram armazenadas em escala de cinza durante as etapas de preparação. No momento do carregamento para o treinamento, cada imagem foi convertida para uma representação com três canais idênticos, tornando sua estrutura compatível com a EfficientNet-B0 inicializada com pesos do ImageNet.

Antes de serem fornecidas à rede neural, as imagens foram normalizadas canal a canal utilizando as médias (0,485, 0,456, 0,406) e os desvios-padrão (0,229, 0,224, 0,225), correspondentes à normalização adotada para as entradas associadas aos pesos pré-treinados IMAGENET1K_V1.

$$I'_c = \frac{I_c - \mu_c}{\sigma_c}, \quad (13)$$

em que I_c representa o valor de entrada do canal, μ_c a média correspondente e σ_c o respectivo desvio-padrão. Essa etapa foi aplicada imediatamente antes da entrada das imagens na EfficientNet-B0, após as transformações de aumento de dados quando estas eram utilizadas. Nas amostras de validação, que não receberam transformações aleatórias, foram realizadas apenas as operações necessárias à preparação da entrada, incluindo a normalização.

A mesma normalização foi aplicada às três condições experimentais.

3.4 Origem e Aplicação dos Métodos de Pré-processamento

A avaliação foi realizada utilizando três condições experimentais formadas a partir das mesmas 127 imagens do subconjunto DMID. As condições não representam diferentes bases de dados, mas versões correspondentes das mesmas lesões, diferenciadas pelo processamento aplicado às imagens.

Os arquivos utilizados nas condições Original e Método A foram obtidos a partir do subconjunto DMID disponibilizado no repositório do Mammo-Bench. Embora o Mammo-Bench reúna e padronize imagens provenientes de diferentes conjuntos de mamografia, neste trabalho foram selecionadas exclusivamente as amostras identificadas como pertencentes ao DMID (BHOLE; SUSEELA; PAREKH, 2025).

Para cada imagem selecionada, foram localizadas no repositório sua versão anterior ao procedimento de pré-processamento do Mammo-Bench e sua versão correspondente após esse procedimento. A associação foi realizada por meio dos identificadores dos arquivos, permitindo preservar a correspondência entre a imagem, a lesão e o respectivo rótulo diagnóstico.

A condição Original foi formada pelas versões anteriores à aplicação dos dois métodos avaliados. O Método A utilizou as versões pré-processadas segundo o procedimento publicado no trabalho Mammo-Bench (BHOLE; SUSEELA; PAREKH, 2025). O Método B foi gerado neste trabalho a partir das imagens da condição Original, por meio da implementação do procedimento publicado por INGLE; VIDHATE; CHAUDHARI (2024), que combina filtragem Gaussiana e mascaramento de nitidez.

A mesma seleção de identificadores foi utilizada nas condições Original e Método A, às quais foram aplicados os mesmos critérios de extração das regiões de interesse, inclusão da margem adicional e redimensionamento. Em seguida, o Método B foi gerado a partir das regiões de interesse da condição Original. Dessa forma, cada identificador do DMID possuía uma representação correspondente nas três condições experimentais.

A Tabela 4 sintetiza a origem e o processamento associado a cada condição.

Tabela 4 – Origem e definição das condições experimentais

Condição	Origem das imagens	Definição experimental
Original	Versões originais do subconjunto DMID disponibilizadas no repositório Mammo-Bench	ROI extraída da versão original, com margem adicional de 10% e redimensionamento para 768×768 pixels.
Método A	Versões pré-processadas das mesmas imagens do subconjunto DMID disponibilizadas no repositório Mammo-Bench	ROI correspondente à mesma lesão após o pré-processamento Mammo-Bench, utilizando os mesmos critérios de margem e redimensionamento da condição Original.
Método B	Geradas neste trabalho a partir das imagens da condição Original	Gerado a partir da ROI preparada da condição Original, após o redimensionamento para 768×768 pixels, pela aplicação de filtragem Gaussiana e mascaramento de nitidez.

Fonte: Elaborado pelo autor (2026).

3.4.1 Condição Original

A condição Original foi utilizada como referência para a comparação. Ela foi formada pelas imagens identificadas como originais no subconjunto DMID disponibilizado pelo repositório do Mammo-Bench.

Essas imagens não foram submetidas aos dois métodos de pré-processamento avaliados no trabalho. Foram aplicadas somente as etapas comuns necessárias à preparação das entradas da rede neural: seleção das amostras, extração da região de interesse, inclusão da margem adicional de 10% e redimensionamento para 768×768 pixels.

A condição Original representa, portanto, a versão com menor quantidade de processamento adicional. Sua inclusão permitiu verificar se os procedimentos publicados na literatura produziram ganhos em relação às imagens utilizadas apenas com as etapas comuns de preparação.

3.4.2 Método A: pré-processamento Mammo-Bench

O Método A corresponde ao procedimento de pré-processamento proposto no trabalho Mammo-Bench (Bhole; SUSEELA; PAREKH, 2025). O objetivo desse procedimento é aumentar a uniformidade entre imagens provenientes de diferentes bases, formatos e condições de aquisição.

O fluxo descrito pelos autores inclui a padronização do formato e do fundo

das imagens, a segmentação da região mamária, a remoção do músculo peitoral nas projeções em que essa estrutura está presente e o recorte da região mamária (Bhole; Suseela; Parekh, 2025). Após essas operações, as imagens são organizadas em um conjunto padronizado para aplicações de aprendizado de máquina.

Neste trabalho, o procedimento completo do Mammo-Bench não foi reimplementado. Foram utilizadas as versões já pré-processadas disponibilizadas no repositório pelos autores. Entre todas as imagens que integram o Mammo-Bench, foram selecionadas exclusivamente aquelas correspondentes aos mesmos identificadores do subconjunto DMID presente na condição Original.

Consequentemente, o Método A não representa a utilização de uma base de dados diferente. Ele corresponde às versões pré-processadas das mesmas imagens e lesões utilizadas na condição Original.

Após a seleção das amostras correspondentes, foram aplicados os mesmos critérios de extração das regiões relacionadas às lesões, inclusão da margem e redimensionamento utilizados nas demais condições experimentais. Dessa forma, as imagens fornecidas à rede apresentaram a mesma resolução e estavam associadas aos mesmos rótulos diagnósticos.

3.4.3 Método B: redução de ruído por filtragem Gaussiana e mascaramento de nitidez

O Método B foi implementado conforme o procedimento de filtragem Gaussiana e mascaramento de nitidez apresentado na Seção 2.7 a partir das imagens da condição Original. Os parâmetros empregados na implementação são apresentados na Tabela 5.

Tabela 5 – Parâmetros utilizados no Método B

Parâmetro	Descrição	Valor utilizado
k	Dimensões do núcleo utilizado na filtragem Gaussiana	5×5
σ	Desvio-padrão utilizado pelo filtro Gaussiano	1,0
s	Intensidade aplicada ao mascaramento de nitidez	1,5

Fonte: Elaborado pelo autor (2026).

O núcleo de 5×5 foi utilizado para realizar a suavização local, enquanto $\sigma = 1,0$ controlou a distribuição dos pesos do filtro. O fator $s = 1,5$ determinou a intensidade com que a máscara de nitidez foi adicionada à imagem suavizada.

O Método B foi aplicado após a extração e o redimensionamento das regiões de interesse da condição Original. Assim, não houve alteração na seleção das

lesões, nos rótulos ou nas dimensões das imagens.

A Figura 7 apresenta a mesma região de interesse nas três condições experimentais, permitindo comparar visualmente as modificações introduzidas pelos métodos.

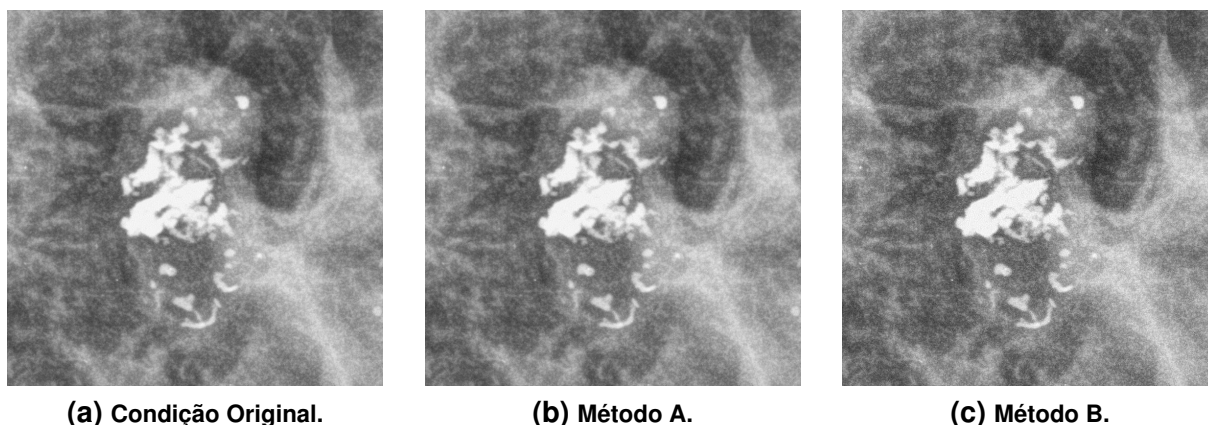


Figura 7 – Comparação visual entre as condições experimentais

Fonte: Elaborado pelo autor (2026).

Na Figura 7, a primeira imagem corresponde à condição Original, submetida apenas às etapas comuns de preparação. A segunda apresenta a versão da mesma região após o procedimento Mammo-Bench, denominado Método A. A terceira corresponde ao Método B, baseado na combinação de filtragem Gaussiana e mascaramento de nitidez.

As diferenças visuais entre as condições são sutis e não permitem, por si só, determinar qual processamento é mais adequado para a classificação. Por esse motivo, a influência dos métodos foi avaliada quantitativamente a partir do desempenho da EfficientNet-B0 sob o mesmo protocolo experimental.

3.5 Arquitetura do Modelo

A arquitetura utilizada foi a EfficientNet-B0, inicializada com pesos pré-treinados no conjunto ImageNet. Essa configuração permite aplicar o aprendizado por transferência ao problema de classificação de imagens mamográficas (DENG *et al.*, 2009; TAN; LE, 2019).

As imagens fornecidas ao modelo foram preparadas conforme descrito na Seção 3.3.3, incluindo a conversão para três canais e a normalização compatível com os pesos pré-treinados utilizados.

A camada classificadora original da EfficientNet-B0 foi substituída por uma sequência composta por uma camada de *dropout*, com probabilidade de 0,30, e uma camada linear com duas saídas. Essas saídas correspondem às classes benigna e maligna. A Tabela 6 resume as principais características do modelo.

Tabela 6 – Configuração da arquitetura utilizada

Componente	Configuração
Arquitetura	EfficientNet-B0
Inicialização	Pesos IMAGENET1K_V1
Dimensão de entrada	768 × 768 pixels
Canais de entrada	Três canais idênticos obtidos da imagem em escala de cinza
Normalização	Média e desvio-padrão do ImageNet
Regularização	<i>Dropout</i> com probabilidade de 0,30
Camada de saída	Camada linear com duas saídas
Classes	Benigna e maligna

Fonte: Elaborado pelo autor (2026).

Durante a validação, os valores produzidos pela camada de saída foram convertidos em probabilidades por meio da função *softmax*. A probabilidade associada à classe maligna foi armazenada para o cálculo da AUC e das demais métricas de avaliação.

3.6 Configuração do Treinamento

Os experimentos foram implementados em Python utilizando as bibliotecas PyTorch, Torchvision, OpenCV, Albumentations, Pandas, NumPy e Scikit-learn. O processamento foi executado em uma unidade gráfica AMD por meio do DirectML. Como o ambiente não utilizou a plataforma *Compute Unified Device Architecture* (CUDA), o treinamento com precisão mista automática permaneceu desativado.

Para controlar a aleatoriedade, foi definida a semente base 42. A partir dela, foram derivados os estados aleatórios utilizados nas diferentes repetições. As divisões dos dados também foram armazenadas em arquivo, permitindo reutilizar exatamente os mesmos índices nas três condições experimentais.

A função de perda utilizada foi a entropia cruzada com *label smoothing* igual a 0,02. A otimização foi realizada pelo AdamW, que aplica o decaimento dos pesos de forma desacoplada da atualização adaptativa dos gradientes (LOSHCHILOV; HUTTER, 2019). A taxa de aprendizado inicial foi definida em 2×10^{-4} , com *weight decay* igual a 1×10^{-4} .

A taxa de aprendizado foi controlada pelo agendador *ReduceLROnPlateau*, utilizando a AUC de validação como referência. Quando não ocorria melhora por três épocas, a taxa era multiplicada por 0,5, respeitando o limite mínimo de 1×10^{-6} . Também foi aplicado recorte da norma dos gradientes com valor máximo igual a 1,0.

O treinamento foi limitado a 40 épocas e utilizou interrupção antecipada com paciência igual a 10. O modelo correspondente à maior AUC de validação foi ar-

mazenado e restaurado ao final de cada execução. A Tabela 7 apresenta os principais parâmetros.

Tabela 7 – Parâmetros utilizados durante o treinamento

Parâmetro	Valor
Número máximo de épocas	40
Tamanho do lote	4
Passos de acumulação de gradiente	2
Tamanho efetivo do lote	8
Taxa de aprendizado inicial	2×10^{-4}
<i>Weight decay</i>	1×10^{-4}
Função de perda	Entropia cruzada
<i>Label smoothing</i>	0,02
Otimizador	AdamW
Fator do agendador	0,5
Paciência do agendador	3 épocas
Taxa de aprendizado mínima	1×10^{-6}
Norma máxima dos gradientes	1,0
Paciência da interrupção antecipada	10 épocas
Critério de seleção do modelo	Maior AUC de validação
Semente base	42

Fonte: Elaborado pelo autor (2026).

3.6.1 Aumento de dados

O aumento de dados foi aplicado somente às amostras utilizadas no treinamento. As imagens de validação receberam apenas a normalização e a conversão para tensor, evitando a introdução de variações aleatórias durante a avaliação.

As transformações foram definidas com baixa intensidade para aumentar a variabilidade das amostras sem modificar de forma excessiva a aparência das lesões. A Tabela 8 apresenta as operações e suas probabilidades.

Tabela 8 – Transformações utilizadas no aumento de dados

Transformação	Configuração	Probabilidade
Espelhamento horizontal	Inversão no eixo horizontal	0,50
Transformação afim	Escala entre 0,95 e 1,05; translação de até 3%; rotação entre -10° e 10°	0,70
Brilho e contraste	Variação máxima de 0,08 em brilho e contraste	0,30
Ruído Gaussiano	Desvio-padrão no intervalo de 0,01 a 0,03	0,15

Fonte: Elaborado pelo autor (2026).

3.7 Validação Experimental

A validação foi realizada por meio de validação cruzada estratificada com cinco divisões e três repetições. A estratificação manteve aproximadamente a proporção das classes benigna e maligna em cada divisão. Em cada *fold*, quatro partes foram utilizadas no treinamento e uma parte foi reservada para validação, resultando em aproximadamente 80% das amostras para treinamento e 20% para validação. Não foi separado um conjunto de teste independente. Como o procedimento empregou validação cruzada, as amostras que compõem o conjunto de validação em uma determinada divisão podem integrar o conjunto de treinamento em outras divisões. As métricas finais foram calculadas a partir das previsões obtidas nas amostras de validação de cada *fold*.

As três repetições utilizaram estados aleatórios diferentes, derivados da semente base 42. Esse procedimento gerou 15 execuções para cada condição experimental. Como foram avaliadas três condições, o experimento completo totalizou 45 treinamentos.

As divisões foram geradas uma única vez e armazenadas em arquivo. Posteriormente, os mesmos índices foram empregados nas condições Original, Método A e Método B. Essa decisão garantiu uma comparação pareada, na qual cada condição foi avaliada utilizando exatamente as mesmas amostras de treinamento e validação.

A Tabela 9 resume a organização do procedimento de validação adotado nos experimentos.

Tabela 9 – Organização da validação experimental

Característica	Configuração
Método de validação	Validação cruzada estratificada
Número de divisões	5
Número de repetições	3
Execuções por condição	15
Condições avaliadas	3
Total de treinamentos	45
Proporção aproximada por <i>fold</i>	80% treinamento e 20% validação
Reutilização das divisões	Mesmos índices para as três condições

Fonte: Elaborado pelo autor (2026).

Para cada execução, foram armazenados o modelo correspondente à melhor AUC, a época em que esse resultado ocorreu e as probabilidades produzidas para as amostras do conjunto de validação. Essas previsões fora da amostra de treinamento foram posteriormente utilizadas para calcular as métricas e realizar a análise estatística.

3.8 Métricas e Análise Estatística

O desempenho foi avaliado pelas métricas AUC, acurácia, sensibilidade, especificidade, precisão, F1-score e Coeficiente de Correlação de Matthews. A AUC foi calculada diretamente a partir das probabilidades atribuídas à classe maligna. Para as métricas derivadas da matriz de confusão, as probabilidades foram convertidas em rótulos utilizando limiar de decisão igual a 0,5.

As métricas foram obtidas separadamente para cada combinação de repetição e fold. Para a descrição geral dos resultados, foram calculadas a média e o desvio-padrão das 15 execuções de cada condição. Como complemento à análise não paramétrica, também foram calculadas a mediana, o primeiro quartil e o terceiro quartil.

A comparação estatística foi realizada de forma pareada, uma vez que as três condições experimentais foram avaliadas utilizando as mesmas divisões dos dados. Cada combinação entre repetição e fold foi considerada um bloco, enquanto Original, Método A e Método B foram considerados as três condições relacionadas.

Dessa forma, a análise foi composta por $n = 15$ blocos e $k = 3$ condições. Para cada métrica, foi aplicado o teste não paramétrico de Friedman, apresentado na Seção 2.9 (CONOVER, 1999; DEMŠAR, 2006).

Como foram comparadas três condições, a distribuição de referência utilizada na interpretação da estatística apresentou $k - 1 = 2$ graus de liberdade. Foi adotado nível de significância de $\alpha = 0,05$.

As hipóteses consideradas foram:

- H_0 : as três condições apresentam o mesmo comportamento para a métrica analisada;
- H_1 : pelo menos uma das condições apresenta comportamento diferente.

Como os conjuntos de treinamento dos diferentes folds apresentam sobreposição de amostras, os resultados das execuções não constituem observações completamente independentes. Por esse motivo, o teste de Friedman foi empregado como uma análise comparativa exploratória, devendo seus resultados ser interpretados dentro das condições do protocolo experimental adotado.

Caso o teste global apresentasse $p < 0,05$, seriam necessárias comparações posteriores para identificar quais condições diferiam. Como nenhuma das métricas apresentou resultado significativo, não foram realizados testes pós-hoc.

A interpretação dos resultados considerou tanto as diferenças numéricas entre as métricas quanto a análise estatística. Dessa forma, uma variação observada nas médias não foi automaticamente tratada como evidência de superioridade de um método quando não acompanhada de significância estatística.

3.9 Resumo do Capítulo

Este capítulo apresentou a metodologia empregada para avaliar a influência dos métodos de pré-processamento na classificação de lesões mamárias. Foi demonstrado que as três condições experimentais foram formadas por versões correspondentes das mesmas 127 imagens do subconjunto DMID.

A condição Original foi composta pelas imagens anteriores à aplicação dos métodos avaliados. O Método A utilizou as versões pré-processadas segundo o procedimento Mammo-Bench, enquanto o Método B foi produzido neste trabalho pela aplicação de filtragem Gaussiana e mascaramento de nitidez.

Também foram descritas a extração das regiões de interesse, a padronização das dimensões, a arquitetura EfficientNet-B0, as configurações de treinamento e o aumento de dados. Por fim, foram apresentados o protocolo de validação cruzada e o teste de Friedman utilizado na análise estatística.

4 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS

Este capítulo apresenta e discute os resultados obtidos na classificação de lesões mamárias utilizando a arquitetura EfficientNet-B0. As condições experimentais são identificadas conforme a nomenclatura definida na Seção 3.4: Original, Método A e Método B.

O Método A corresponde às imagens processadas segundo o procedimento Mammo-Bench, enquanto o Método B corresponde às imagens submetidas à combinação de filtragem Gaussiana e mascaramento de nitidez. Todas as condições foram formadas pelas mesmas 127 amostras do DMID e avaliadas com as mesmas divisões de treinamento e validação.

4.1 Desempenho Geral das Condições Experimentais

A Tabela 10 apresenta a média e o desvio-padrão das principais métricas de desempenho obtidas nas 15 execuções realizadas para cada condição. A utilização do desvio-padrão permite avaliar não apenas o desempenho médio, mas também a variação dos resultados entre as diferentes divisões da validação cruzada.

Tabela 10 – Desempenho médio das condições experimentais

Condição	AUC	Acurácia	F1-score	MCC
Original	$0,930 \pm 0,038$	$0,830 \pm 0,052$	$0,747 \pm 0,102$	$0,639 \pm 0,119$
Método A	$0,930 \pm 0,049$	$0,842 \pm 0,055$	$0,769 \pm 0,103$	$0,671 \pm 0,119$
Método B	$0,914 \pm 0,068$	$0,814 \pm 0,074$	$0,697 \pm 0,159$	$0,603 \pm 0,177$

Fonte: Elaborado pelo autor (2026).

Os resultados mostram que a condição Original e o Método A apresentaram a mesma AUC média, aproximadamente 0,930. O Método A apresentou os maiores valores médios de acurácia, F1-score e MCC, enquanto a condição Original apresentou menor dispersão na AUC. O Método B apresentou médias inferiores de AUC, acurácia, F1-score e MCC, além de maior variabilidade entre as execuções.

Como a classificação envolve classes com quantidades distintas de amostras, a análise não deve ser limitada à acurácia. Métricas como sensibilidade, especificidade e MCC permitem avaliar de maneira mais completa o comportamento do classificador (SOKOLOVA; LAPALME, 2009; CHICCO; JURMAN, 2020). A Tabela 11 apresenta as métricas relacionadas ao comportamento do modelo em cada classe.

Tabela 11 – Métricas médias relacionadas à classificação das classes

Condição	Precisão	Sensibilidade	Especificidade
Original	$0,823 \pm 0,104$	$0,711 \pm 0,168$	$0,900 \pm 0,074$
Método A	$0,843 \pm 0,127$	$0,741 \pm 0,167$	$0,900 \pm 0,091$
Método B	$0,845 \pm 0,143$	$0,632 \pm 0,220$	$0,921 \pm 0,083$

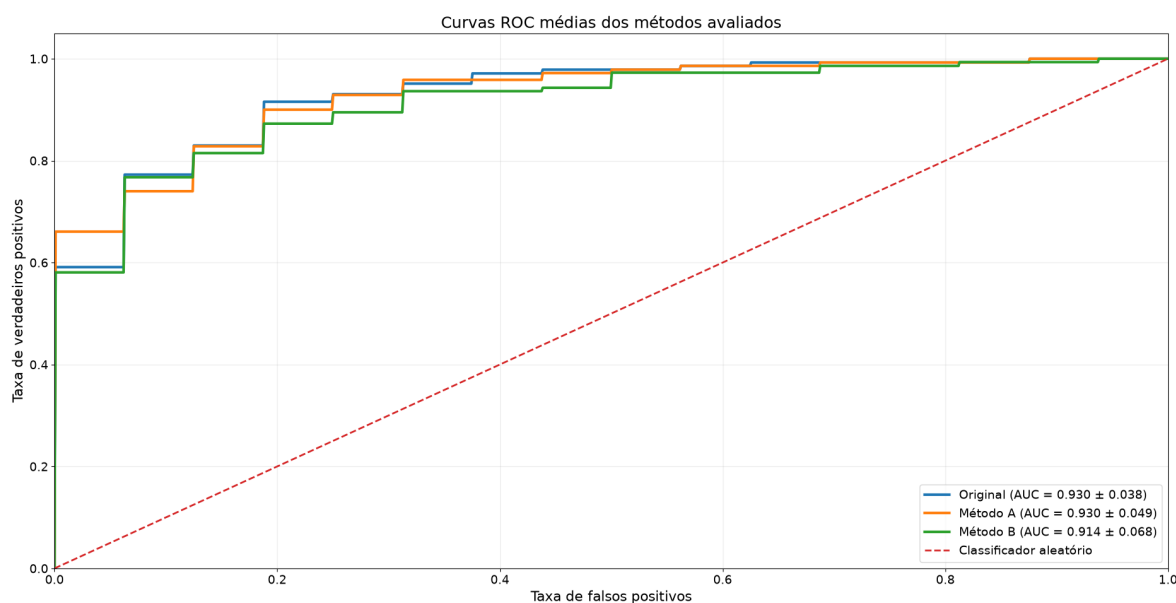
Fonte: Elaborado pelo autor (2026).

Observa-se um comportamento diferente entre as condições quando as classes são analisadas individualmente. O Método A apresentou a maior sensibilidade média, igual a 0,741, indicando maior capacidade de identificar corretamente as amostras malignas. O Método B, por outro lado, apresentou a maior especificidade, igual a 0,921, e a maior precisão, igual a 0,845. Esse resultado indica que o Método B reduziu a ocorrência de falsos positivos, porém apresentou maior dificuldade em reconhecer corretamente parte das lesões malignas.

4.2 Análise da Curva ROC e da AUC

A curva ROC (*Receiver Operating Characteristic*) permite avaliar o comportamento do classificador considerando diferentes limiares de decisão, relacionando a taxa de verdadeiros positivos à taxa de falsos positivos. A área sob essa curva (AUC) resume a capacidade de discriminação entre as classes, sendo que valores mais próximos de 1 indicam maior capacidade de separação entre amostras benignas e malignas (FAWCETT, 2006).

Para representar o comportamento observado durante todo o procedimento de validação, foram calculadas as curvas ROC das 15 execuções realizadas para cada condição, correspondentes às cinco divisões da validação cruzada em três repetições. As curvas foram interpoladas em um mesmo conjunto de valores de taxa de falsos positivos e, posteriormente, foi calculada a curva média de cada condição. A Figura 8 apresenta esse resultado.

Figura 8 – Curvas ROC médias das condições experimentais

Fonte: Elaborado pelo autor (2026).

As curvas apresentadas na Figura 8 demonstram comportamento semelhante entre as três condições experimentais. A condição Original e o Método A apresentaram AUC média de aproximadamente 0,930, enquanto o Método B apresentou valor médio de 0,914.

A condição Original e o Método A apresentaram valores médios de AUC praticamente idênticos. Entretanto, a condição Original apresentou menor desvio-padrão entre as execuções, indicando menor variabilidade dessa métrica. O Método B, por sua vez, apresentou maior desvio-padrão da AUC.

As três curvas permaneceram consideravelmente acima da linha diagonal, indicando que a EfficientNet-B0 apresentou capacidade elevada de discriminar lesões benignas e malignas nas três condições analisadas.

4.3 Sensibilidade e Especificidade

A sensibilidade e a especificidade dependem do limiar utilizado para converter as probabilidades produzidas pelo modelo em classes. Neste trabalho, foi mantido o limiar de decisão de 0,5 definido na metodologia. Para esse ponto de corte, a sensibilidade representa a proporção de lesões malignas corretamente classificadas, enquanto a especificidade corresponde à proporção de lesões benignas corretamente classificadas.

As Figuras 9a e 9b apresentam a distribuição dessas métricas ao longo das 15 execuções de cada condição.

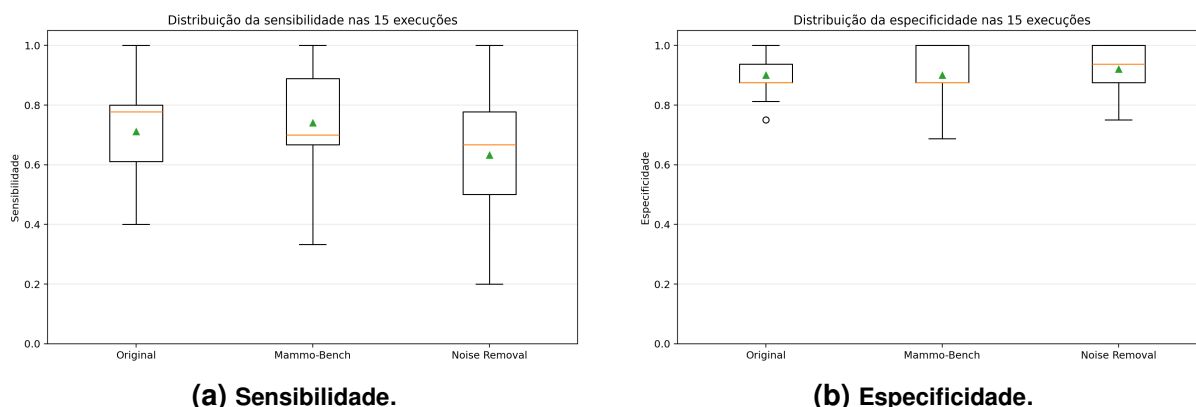


Figura 9 – Boxplots dos valores de sensibilidade e especificidade obtidos nas 15 execuções

Fonte: Elaborado pelo autor (2026).

Nos boxplots apresentados na Figura 9, a linha central representa a mediana das execuções, enquanto o triângulo indica a média. As caixas representam o intervalo entre o primeiro e o terceiro quartis.

Na sensibilidade, o Método A apresentou o maior valor médio, com 0,741, seguido pela condição Original, com 0,711, e pelo Método B, com 0,632. A diferença numérica entre os Métodos A e B foi maior nessa métrica do que nas demais, indicando comportamentos distintos principalmente na identificação das lesões malignas.

Em contrapartida, o Método B apresentou especificidade média de 0,921, superior ao valor de 0,900 observado tanto na condição Original quanto no Método A. Esse comportamento caracteriza uma troca entre sensibilidade e especificidade: o Método B classificou uma proporção maior das amostras benignas corretamente, porém deixou de reconhecer uma proporção maior das amostras malignas.

Essa característica também explica a precisão relativamente elevada do Método B. Como essa condição produziu menos falsos positivos, uma parcela maior das predições realizadas como malignas estava correta. Entretanto, essa vantagem ocorreu simultaneamente à redução da sensibilidade.

4.4 Análise das Matrizes de Confusão

Para complementar a análise das métricas, foram construídas matrizes de confusão a partir das predições realizadas fora das amostras utilizadas no treinamento. Como cada uma das 127 imagens foi avaliada uma vez em cada repetição da validação cruzada, as matrizes apresentadas na Figura 10 agregam 381 predições para cada condição.

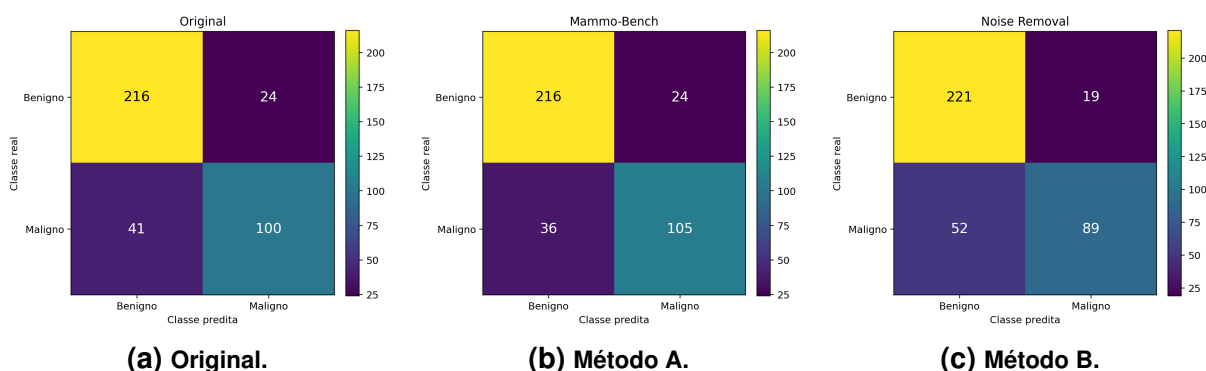


Figura 10 – Matrizes de confusão agregadas das três repetições da validação cruzada

Fonte: Elaborado pelo autor (2026).

Na Figura 10a, a condição Original apresentou 216 classificações corretas de amostras benignas e 100 classificações corretas de amostras malignas. Foram observados 24 falsos positivos e 41 falsos negativos.

O Método A manteve o mesmo número de verdadeiros negativos e falsos positivos, com 216 e 24 ocorrências, respectivamente. Entretanto, apresentou 105 verdadeiros positivos e 36 falsos negativos. Essa redução no número de falsos negativos é coerente com a maior sensibilidade observada na Tabela 11.

O Método B apresentou 221 verdadeiros negativos e apenas 19 falsos positivos, resultado que explica sua maior especificidade. Em contrapartida, foram registrados 89 verdadeiros positivos e 52 falsos negativos.

As matrizes de confusão apresentadas são agregações das predições das três repetições, enquanto as métricas das Tabelas 10 e 11 correspondem à média das métricas calculadas separadamente em cada *fold*. Por esse motivo, pequenas diferenças podem ocorrer ao recalcular as métricas diretamente a partir dos valores agregados das matrizes.

4.5 Análise Estatística

Embora diferenças numéricas tenham sido observadas entre as condições, foi necessário avaliar se essas diferenças se mantiveram de forma consistente ao longo das execuções. Para isso, foi aplicado o teste não paramétrico de Friedman, considerando Original, Método A e Método B como condições relacionadas (CONOVER, 1999; DEMŠAR, 2006).

Como complemento às médias e aos desvios-padrão apresentados nas Tabelas 10 e 11, a Tabela 12 apresenta a mediana e o intervalo entre o primeiro e o terceiro quartis das 15 execuções de cada condição, juntamente com o p-valor obtido pelo teste de Friedman.

Tabela 12 – Medianas, intervalos interquartis e resultados do teste de Friedman

Métrica	Original	Método A	Método B	p-valor
AUC	0,924 [0,894; 0,962]	0,944 [0,910; 0,960]	0,924 [0,889; 0,963]	0,627
Acurácia	0,840 [0,800; 0,863]	0,880 [0,804; 0,885]	0,840 [0,780; 0,863]	0,600
Sensibilidade	0,778 [0,611; 0,800]	0,700 [0,667; 0,889]	0,667 [0,500; 0,778]	0,077
Especificidade	0,875 [0,875; 0,938]	0,875 [0,875; 1,000]	0,938 [0,875; 1,000]	0,763
Precisão	0,800 [0,764; 0,861]	0,818 [0,760; 1,000]	0,800 [0,739; 1,000]	0,472
F1-score	0,778 [0,710; 0,821]	0,800 [0,732; 0,850]	0,750 [0,658; 0,800]	0,201
MCC	0,653 [0,568; 0,735]	0,748 [0,573; 0,763]	0,667 [0,546; 0,714]	0,650

Fonte: Elaborado pelo autor (2026).

Foi adotado nível de significância de $\alpha = 0,05$. Para todas as métricas analisadas, o teste de Friedman apresentou $p > 0,05$. Dessa forma, não foram encontradas evidências estatísticas suficientes para rejeitar a hipótese de ausência de diferenças entre as três condições.

A sensibilidade apresentou o menor p-valor, $p = 0,077$, porém esse resultado permaneceu acima do nível de significância adotado. Observa-se também que a ordenação das medianas não coincide necessariamente com a das médias em todas as métricas, refletindo a distribuição e a variabilidade dos resultados entre as execuções.

É importante destacar que a ausência de significância estatística não demonstra que as condições sejam equivalentes. O resultado indica apenas que, considerando o conjunto de dados e o protocolo experimental utilizados, não foi encontrada evidência estatística suficiente para demonstrar diferença entre as condições avaliadas.

4.6 Discussão dos Resultados

A análise conjunta dos resultados indica que nenhuma das condições experimentais apresentou superioridade estatisticamente comprovada. O Método A, correspondente ao pré-processamento Mammo-Bench, obteve os maiores valores médios de acurácia, sensibilidade, F1-score e MCC. A condição Original apresentou a mesma AUC média e menor variação dessa métrica. O Método B, baseado em filtragem Gaussiana e mascaramento de nitidez, apresentou maior precisão e especificidade, porém desempenho inferior principalmente em sensibilidade.

O desempenho da condição Original é um resultado relevante, pois demonstra que a EfficientNet-B0 foi capaz de obter elevada capacidade de discriminação mesmo sem a aplicação de um método adicional de aprimoramento após a extração e a padronização das ROIs. Uma possível interpretação é que a combinação entre aprendizado por transferência e aumento de dados tenha fornecido robustez sufici-

ente para que a rede aprendesse características relevantes diretamente das imagens originais.

O Método A apresentou uma melhora numérica nas métricas relacionadas ao equilíbrio entre as classes. O F1-score médio aumentou de 0,747 para 0,769 e o MCC de 0,639 para 0,671 quando comparado às imagens originais. Além disso, a sensibilidade passou de 0,711 para 0,741. Apesar desses ganhos, a AUC média permaneceu praticamente inalterada e o teste de Friedman não indicou diferença estatisticamente significativa.

No Método B, a filtragem Gaussiana combinada ao mascaramento de nitidez produziu um comportamento mais conservador na classificação da classe maligna. O aumento da especificidade e a redução dos falsos positivos indicam melhor reconhecimento das amostras benignas, porém a sensibilidade foi reduzida e houve aumento do número de falsos negativos.

Uma possível explicação é que as modificações introduzidas pela suavização e pelo realce tenham alterado características sutis relevantes para parte das lesões malignas. Essa interpretação deve ser considerada como hipótese, uma vez que o experimento realizado avaliou o desempenho final do classificador e não a contribuição individual de cada característica visual.

Os resultados diferem parcialmente dos apresentados por INGLE; VIDHATE; CHAUDHARI (2024), que observaram melhora média na acurácia de classificadores tradicionais após a aplicação do método de remoção de ruído em mamografias. Entretanto, os experimentos não são diretamente equivalentes, pois foram realizados com conjuntos de dados, classificadores e protocolos experimentais distintos. No presente trabalho, o método foi avaliado no DMID em conjunto com uma EfficientNet-B0 pré-treinada, enquanto o estudo de referência avaliou classificadores como SVM, Random Forest e Gradient Boosting.

Outro fator importante é o tamanho do conjunto experimental. Foram utilizadas 127 imagens, sendo 80 benignas e 47 malignas. Embora a validação cruzada repetida permita utilizar as amostras de forma eficiente e avaliar a variação entre diferentes divisões, o número limitado de imagens aumenta a incerteza das estimativas e reduz a capacidade de detectar diferenças pequenas entre os métodos.

Dessa forma, os resultados não sustentam a afirmação de que um pré-processamento específico seja universalmente superior. Dentro das condições avaliadas, o Método A apresentou o melhor equilíbrio numérico entre as métricas, enquanto as imagens originais alcançaram desempenho bastante próximo com menor variabilidade de AUC. O Método B apresentou maior especificidade, porém com redução de sensibilidade. A escolha entre os métodos deve, portanto, considerar o objetivo da aplicação e o custo associado a falsos positivos e falsos negativos.

4.7 Resumo do Capítulo

Este capítulo apresentou os resultados obtidos nas condições Original, Método A e Método B. A condição Original e o Método A apresentaram AUC média de aproximadamente 0,930, enquanto o Método B apresentou AUC média de aproximadamente 0,914.

O Método A apresentou os maiores valores médios de acurácia, sensibilidade, F1-score e MCC. O Método B apresentou maior precisão e especificidade, porém menor sensibilidade e maior número de falsos negativos.

O teste de Friedman não identificou diferenças estatisticamente significativas entre as três condições. Assim, embora os métodos tenham produzido diferenças numéricas no comportamento do classificador, não foram encontradas evidências suficientes para demonstrar a superioridade de uma das condições no protocolo experimental adotado.

5 CONSIDERAÇÕES FINAIS

Este trabalho teve como objetivo avaliar a influência de métodos de pré-processamento no desempenho da arquitetura EfficientNet-B0 para a classificação de lesões mamárias benignas e malignas.

Para isso, as mesmas 127 imagens do subconjunto DMID foram organizadas em três condições experimentais pareadas. A condição Original utilizou as imagens anteriores à aplicação dos métodos avaliados. O Método A correspondeu às versões processadas segundo o procedimento proposto no Mammo-Bench, enquanto o Método B foi produzido pela aplicação de filtragem Gaussiana e mascaramento de nitidez com base no artigo *Noise Removal for Mammogram Images Using Gaussian Based Unsharp Masking*.

A metodologia foi estruturada de modo a manter constantes a composição das amostras, os rótulos, as regiões de interesse, as divisões de treinamento e validação, a arquitetura e as configurações de treinamento. Dessa forma, a comparação foi concentrada nas modificações introduzidas pelos métodos de pré-processamento.

Os resultados demonstraram que as três condições possibilitaram desempenho elevado na classificação das lesões. A condição Original e o Método A apresentaram AUC média de aproximadamente 0,930, enquanto o Método B apresentou AUC média de aproximadamente 0,914. Em termos de acurácia, F1-score e Coeficiente de Correlação de Matthews, o Método A apresentou os maiores valores médios, indicando o melhor equilíbrio numérico entre as métricas avaliadas.

A análise das classes individualmente mostrou, entretanto, diferenças importantes no comportamento das condições. O Método A apresentou a maior sensibilidade média, indicando melhor capacidade de identificar corretamente as lesões malignas. O Método B apresentou a maior especificidade e precisão, porém acompanhado de redução da sensibilidade e aumento do número de falsos negativos.

As matrizes de confusão reforçaram essa interpretação. O Método A apresentou menor quantidade de falsos negativos quando comparado às demais condições, enquanto o Método B apresentou menor quantidade de falsos positivos, porém maior número de lesões malignas classificadas incorretamente como benignas.

Apesar das diferenças numéricas observadas, o teste de Friedman não identificou diferenças estatisticamente significativas entre as três condições para nenhuma das métricas analisadas. Portanto, considerando o conjunto de dados e o protocolo experimental utilizados, não foi possível demonstrar estatisticamente a superioridade de um método específico de pré-processamento.

Esse resultado é relevante porque indica que a utilização de etapas adicionais de processamento não necessariamente produz melhoria no desempenho de

uma Rede Neural Convolucional. As imagens originais, submetidas apenas à extração da região de interesse e à padronização das dimensões, apresentaram desempenho muito próximo ao obtido com o Método A e superior, em algumas métricas, ao observado com o Método B. Assim, o aumento da complexidade do processamento deve ser justificado por ganhos consistentes e mensuráveis, e não apenas por alterações visuais nas imagens.

Também foi possível observar que a EfficientNet-B0, associada ao aprendizado por transferência, apresentou capacidade de extrair características relevantes mesmo em um conjunto de dados relativamente reduzido. A utilização dos pesos pré-treinados, do aumento de dados e da validação cruzada permitiu estabelecer uma metodologia controlada para a comparação entre as diferentes condições experimentais.

Como principal limitação, destaca-se a quantidade restrita de imagens utilizadas nos experimentos. O conjunto final foi composto por 127 amostras, sendo 80 benignas e 47 malignas. Embora a validação cruzada repetida permita aproveitar melhor as amostras disponíveis, a dimensão do conjunto aumenta a variabilidade dos resultados e limita a capacidade da análise estatística de identificar diferenças pequenas entre as condições.

Também deve ser considerada a utilização de um único conjunto de dados. Os experimentos foram realizados exclusivamente com imagens do DMID e, portanto, os resultados obtidos refletem as características dessa base. Dessa forma, o comportamento observado não deve ser generalizado diretamente para outros conjuntos mamográficos sem validação externa.

Adicionalmente, a análise estatística das execuções obtidas pela validação cruzada repetida apresenta uma limitação relacionada à dependência entre os resultados. Embora cada amostra de validação não tenha sido utilizada no treinamento do modelo avaliado naquela execução, os conjuntos de treinamento dos diferentes folds apresentam sobreposição de amostras. Consequentemente, os resultados das execuções não constituem observações completamente independentes. Por esse motivo, o teste de Friedman foi interpretado como uma análise comparativa exploratória, devendo seus resultados ser considerados dentro das condições do protocolo experimental adotado.

Diante dos resultados obtidos, conclui-se que o objetivo proposto foi alcançado. Foi possível implementar e comparar, sob o mesmo protocolo experimental, diferentes abordagens de pré-processamento aplicadas à classificação de lesões mamárias. O Método A apresentou o melhor desempenho numérico geral, enquanto as imagens originais alcançaram resultados muito próximos com menor quantidade de processamento. O Método B apresentou maior especificidade, porém menor sensibi-

lidade. Entretanto, nenhuma dessas diferenças foi estatisticamente significativa, não sendo possível afirmar que uma das condições avaliadas seja superior às demais no protocolo experimental adotado.

5.1 Sugestões para Trabalhos Futuros

Como continuidade deste trabalho, recomenda-se inicialmente ampliar o número de imagens utilizadas nos experimentos. A inclusão de um conjunto maior de amostras benignas e malignas permitiria reduzir a variabilidade das métricas e aumentar o poder estatístico das comparações entre os métodos de pré-processamento.

Também é recomendável realizar validação externa utilizando outros conjuntos públicos de mamografia. A aplicação do mesmo modelo em bases diferentes permitiria avaliar sua capacidade de generalização e verificar se o comportamento observado no DMID se mantém diante de diferenças de aquisição, equipamentos, resolução e características da população analisada.

Quando identificadores de pacientes estiverem disponíveis, uma evolução importante consiste em realizar a divisão dos dados no nível do paciente, garantindo que imagens pertencentes ao mesmo indivíduo não sejam distribuídas entre treinamento e validação. Esse procedimento permite uma avaliação mais rigorosa da capacidade de generalização do modelo para pacientes não observados durante o treinamento.

Outra possibilidade consiste em avaliar diferentes arquiteturas de Redes Neurais Convolucionais e modelos de visão computacional. Arquiteturas como outras variantes da família EfficientNet, ResNet e modelos baseados em *Vision Transformers* podem ser comparadas utilizando o mesmo protocolo experimental, permitindo verificar se o efeito do pré-processamento depende da arquitetura empregada.

Sugere-se também investigar de forma mais detalhada os parâmetros utilizados nos métodos de pré-processamento. No caso do Método B, diferentes tamanhos de núcleo Gaussiano, valores de desvio-padrão e intensidades do mascaramento de nitidez podem ser avaliados. Dessa maneira, seria possível verificar se a configuração empregada neste trabalho representa a melhor combinação para o conjunto DMID ou se outros parâmetros permitem preservar melhor as características associadas às lesões malignas.

Outra linha de investigação consiste em avaliar diferentes estratégias para o desequilíbrio entre as classes, como ponderação da função de perda, amostragem balanceada ou funções de custo específicas. Essas técnicas podem contribuir para aumentar a sensibilidade sem comprometer excessivamente a especificidade.

Também pode ser realizada a otimização do limiar de decisão. Neste trabalho, o valor de 0,5 foi utilizado para converter as probabilidades em classes. Em

aplicações futuras, esse limiar pode ser definido a partir da curva ROC ou de critérios relacionados ao custo dos erros de classificação, permitindo priorizar, por exemplo, maior sensibilidade para a detecção das lesões malignas.

A utilização de técnicas de interpretabilidade representa outra possibilidade relevante. Métodos como mapas de ativação podem ser empregados para verificar quais regiões da imagem contribuíram para a decisão da rede. Esse tipo de análise permitiria investigar se o modelo está efetivamente utilizando características associadas à lesão ou informações externas à região de interesse.

Por fim, recomenda-se avaliar a calibração das probabilidades produzidas pelo modelo e realizar estudos com conjuntos independentes de teste. A combinação dessas etapas com uma base de dados maior, validação externa e análise por paciente pode fornecer evidências mais robustas sobre a influência real do pré-processamento no desempenho de sistemas de apoio à classificação de mamografias.

REFERÊNCIAS

- BHOLE, G.; SUSEELA, S.; PAREKH, N. **Mammo-Bench: A Large-Scale Benchmark Dataset of Mammography Images**. 2025. Preprint. Disponível em: <https://www.medrxiv.org/content/10.1101/2025.01.31.25321510>. Acesso em: 26 jul. 2026. 13, 14, 27, 34, 35, 39, 40, 41
- CHICCO, D.; JURMAN, G. **The advantages of the Matthews correlation coefficient over F1 score and accuracy in binary classification evaluation**. *BMC Genomics*, v. 21, n. 1, p. 6, 2020. 30, 32, 48
- CONOVER, W. J. **Practical Nonparametric Statistics**. 3. ed. New York: John Wiley & Sons, 1999. ISBN 9780471160687. 32, 46, 52
- COUTO, G. **O poder das Redes Neurais Convolucionais: um guia prático para construir sua CNN básica**. 2023. Disponível em: <https://pt.linkedin.com/pulse/o-poder-das-redes-neurais-convolucionais-um-guia-pr%C3%A1tico-couto-j7omf>. Acesso em: 26 jul. 2026. 22
- DEMŠAR, J. **Statistical Comparisons of Classifiers over Multiple Data Sets**. *Journal of Machine Learning Research*, v. 7, p. 1–30, 2006. 33, 46, 52
- DENG, J. *et al.* **ImageNet: A Large-Scale Hierarchical Image Database**. *In: 2009 IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.]: IEEE, 2009. p. 248–255. 24, 42
- FAWCETT, T. **An introduction to ROC analysis**. *Pattern Recognition Letters*, v. 27, n. 8, p. 861–874, 2006. 31, 32, 49
- FRIEDMAN, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, v. 32, n. 200, p. 675–701, 1937. 32
- HU, J.; SHEN, L.; SUN, G. **Squeeze-and-Excitation Networks**. *In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.]: IEEE, 2018. p. 7132–7141. 26
- INGLE, S.; VIDHATE, A.; CHAUDHARI, S. **Noise Removal for Mammogram Images Using Gaussian Based Unsharp Masking**. *In: 2024 8th International Conference on Computing, Communication, Control and Automation (ICCUBEA)*. [S.l.]: IEEE, 2024. p. 1–6. 13, 14, 15, 27, 28, 34, 39, 54
- INSTITUTO NACIONAL DE CÂNCER – INCA. **Detecção precoce**. 2025. Disponível em: <https://www.gov.br/inca/pt-br/assuntos/gestor-e-profissional-de-saude/controlado-cancer-de-mama/acoes/deteccao-precoce>. Acesso em: 26 jul. 2026. 13, 18
- INSTITUTO NACIONAL DE CÂNCER – INCA. **Brasil – estimativa dos casos novos**. 2026. Disponível em: <https://www.gov.br/inca/pt-br/assuntos/cancer/numeros/estimativa/estado-capital/brasil>. Acesso em: 30 jul. 2026. 13
- LECUN, Y.; BENGIO, Y.; HINTON, G. **Deep Learning**. *Nature*, v. 521, p. 436–444, 2015. 21, 23

- LECUN, Y. *et al.* **Gradient-Based Learning Applied to Document Recognition.** *Proceedings of the IEEE*, v. 86, n. 11, p. 2278–2324, 1998. 13, 22, 23
- LITJENS, G. *et al.* **A Survey on Deep Learning in Medical Image Analysis.** *Medical Image Analysis*, v. 42, p. 60–88, 2017. 20, 21
- LOSHCHILOV, I.; HUTTER, F. **Decoupled Weight Decay Regularization.** *In: International Conference on Learning Representations.* [S.l.: s.n.], 2019. 43
- MCKINNEY, S. M. *et al.* **International evaluation of an AI system for breast cancer screening.** *Nature*, v. 577, p. 89–94, 2020. 13, 14, 21
- NATIONAL CANCER INSTITUTE – NCI. **Breast Cancer Screening (PDQ®) – Health Professional Version.** 2025. Disponível em: <https://www.cancer.gov/types/breast/hp/breast-screening-pdq>. Acesso em: 26 jul. 2026. 13, 18, 20
- NEWELL, M. S. *et al.* **ACR BI-RADS® v2025 Manual.** Reston, Virginia: American College of Radiology, 2025. 18, 19, 20
- OZA, P. *et al.* **Digital Mammography Dataset for Breast Cancer Diagnosis Research (DMID) with Breast Mass Segmentation Analysis.** *Biomedical Engineering Letters*, v. 14, n. 2, p. 317–330, 2024. 19, 35
- SANDLER, M. *et al.* **MobileNetV2: Inverted Residuals and Linear Bottlenecks.** *In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.* [S.l.]: IEEE, 2018. p. 4510–4520. 26
- SOKOLOVA, M.; LAPALME, G. **A systematic analysis of performance measures for classification tasks.** *Information Processing and Management*, v. 45, n. 4, p. 427–437, 2009. 29, 32, 48
- SOLTANI, H. *et al.* **Advancements in Breast Cancer Diagnosis: A Comprehensive Review of Mammography Datasets, Preprocessing and Classification Techniques.** *Acta Informatica Pragensia*, v. 13, n. 2, p. 308–326, 2024. 13, 14, 27
- TAN, M.; LE, Q. V. **EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks.** *In: Proceedings of the 36th International Conference on Machine Learning.* [S.l.]: PMLR, 2019. (Proceedings of Machine Learning Research, v. 97), p. 6105–6114. 25, 26, 42
- UNITED STATES FOOD AND DRUG ADMINISTRATION – FDA. **Frequently Asked Questions About Digital Mammography.** 2026. Disponível em: <https://www.fda.gov/radiation-emitting-products/mammography-information-patients/frequently-asked-questions-about-digital-mammography>. Acesso em: 26 jul. 2026. 18
- WORLD HEALTH ORGANIZATION – WHO. **Breast cancer.** 2026. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/breast-cancer>. Acesso em: 26 jul. 2026. 17
- YOSINSKI, J. *et al.* **How transferable are features in deep neural networks?** *In: Advances in Neural Information Processing Systems 27.* [S.l.]: Curran Associates, 2014. p. 3320–3328. 13, 24

ZHANG, X. *et al.* **Research on Scene Classification Method of High-Resolution Remote Sensing Images Based on RFPNet.** *Applied Sciences*, v. 9, n. 10, p. 2028, 2019. 25