

# 2<sup>o</sup> CICS

CONGRESSO  
INTERNACIONAL  
CIÊNCIA E SOCIEDADE

Educação e Pesquisa como  
caminho para pacificação global

TRABALHOS  
PREMIADOS  
2025



CENTRO UNIVERSITÁRIO  
SANTO AGOSTINHO





# **TRABALHOS PREMIADOS 2025**







**CENTRO UNIVERSITÁRIO SANTO AGOSTINHO – UNIFSA**  
**PRÓ-REITORIA DE ENSINO**  
**DIRETORIA DE ENSINO**  
**NÚCLEO DE PUBLICAÇÕES ACADÊMICAS - NPA**

**Supervisão Editorial:** Ana Kelma Cunha Gallas  
**Diagramação:** Edson Rodrigues Cavalcante  
**Design Gráfico e artes:** Ana Kelma Cunha Gallas  
**Capa:** Odrânio Rocha  
**TI OMP Books:** Eliezyo Silva



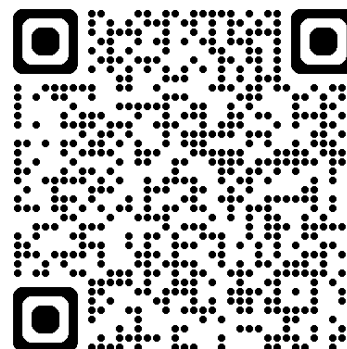
LESTU EDITORA, CONSULTORIA E  
COMUNICAÇÃO LTDA  
Contato: [editora@lestu.org](mailto:editora@lestu.org)  
Site: [www.lestu.com.br](http://www.lestu.com.br)  
Livraria: [www.lestu.org](http://www.lestu.org)

Este título possui uma licença *Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International* (CC BY-NC-ND 4.0).

A íntegra dessa licença pode ser acessada:  
<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode.pt>

© 2025 UNIFSA/LESTU

Todos os capítulos deste livro foram submetidos, aprovados e apresentados no 2º Congresso Internacional Ciência e Sociedade, realizado pelo Centro Universitário Santo Agostinho de 4 a 8 de novembro de 2025, sendo selecionados como os melhores trabalhos apresentados em Grupos Temáticos do evento.



#### FICHA CATALOGRÁFICA

##### Dados Internacionais de Catalogação na Publicação (CIP)

T279 GALLAS, Ana Kelma Cunha; GOMES, Alisson Dias; CRONEMBERGER, Izabel Herika Gomes Matias; LIRA E SILVA, Antonieta; OLIVEIRA, Edjôfre Coelho de; SILVA, Indira Maria de Melo Lira Pereira da (ORGs.).

Trabalhos premiados no 2º Congresso Internacional Ciência e Sociedade: Educação e Pesquisa como caminho para a pacificação global. Teresina: Centro Universitário Santo Agostinho (UNIFSA); Editora Lestu, 2025.

432 p. *online*.

ISBN: 978-65-85729-14-7

DOI: [doi.org/10.51205/lestu.978-65-85729-14-7](https://doi.org/10.51205/lestu.978-65-85729-14-7)

Disponível em: <https://lestu.org/books/>

1. Congresso Internacional Ciência e Sociedade. 2. Pesquisa. 3. Inovação. 4. Sustentabilidade. 5. Ciência e Sociedade. I. Autores. IV. Congresso Internacional Ciência e Sociedade (2. : 2025 : Teresina, PI).

CDD: 370

# 4

## **ANÁLISE DE DESEMPENHO DE ARQUITETURAS DE *DEEP LEARNING* NA DETECÇÃO DE ACHADOS EM IMAGENS MAMOGRÁFICAS<sup>1</sup>**

***PERFORMANCE ANALYSIS OF DEEP LEARNING ARCHITECTURES IN THE DETECTION OF FINDINGS IN MAMMOGRAPHIC IMAGES***

Antonio Guilherme de Brito – UESPI<sup>2</sup>  
Dario Brito Calçada –UESPI<sup>3</sup>

---

<sup>1</sup> Trabalho apresentado no GTS UNIFICADOS 4 e 5 do 2º Congresso Internacional Ciência e Sociedade (CICS 2025), promovido pelo Centro Universitário Santo Agostinho, de 04 a 08 de novembro de 2025, em Teresina-PI.

<sup>2</sup> Bacharel em Ciência da Computação graduado pela Universidade Estadual do Piauí, Campus de Parnaíba – PI.

<sup>3</sup> Doutor em Computação e Matemática Computacional pela Universidade de São Paulo. Professor Adjunto da Universidade Estadual do Piauí, Campus de Parnaíba – PI.

## RESUMO

O câncer de mama constitui uma das principais causas de mortalidade entre mulheres no Brasil, tornando a detecção precoce um fator determinante para a eficácia terapêutica. Nesse contexto, o presente estudo teve como objetivo realizar uma avaliação sistemática e comparativa de dez arquiteturas de Redes Neurais Convolucionais (RNCs) pré-treinadas no ImageNet, aplicadas à classificação binária de mamografias digitais em achados benignos e malignos. A pesquisa utilizou o banco público VinDr- Mammo, composto por mais de 20 mil imagens anonimizadas, assegurando reprodutibilidade e transparência. O processo metodológico incluiu pré-processamento padronizado, aumento de dados e aplicação de transfer learning com substituição da camada final de classificação. As arquiteturas avaliadas incluíram ResNeXt-50, ResNet-101, DenseNet-121, MobileNet v2, EfficientNet-B0, ResNet-50, EfficientNet-B4, ShuffleNet v2, VGG16 e SqueezeNet 1.1. A análise de desempenho foi conduzida por métricas padrão, como acurácia, precisão, recall, especificidade, F1-Score e AUC, complementadas por matrizes de confusão e curvas ROC. Os resultados demonstraram desempenho inicial limitado, com AUCs variando entre 0,59 e 0,61, mas suficientes para estabelecer uma linha de base reprodutível para futuras otimizações via fine-tuning. Constatou-se que a ResNeXt-50 apresentou perfil mais equilibrado entre falsos positivos e falsos negativos, sendo mais promissora para cenários clínicos. Conclui-se que, embora a performance inicial seja limitada, o estudo contribui ao campo da inteligência artificial em saúde ao oferecer uma comparação robusta entre diferentes arquiteturas, evidenciando a importância de metodologias progressivas para alcançar robustez clínica e generalização.

**Palavras-Chave:** Mamografia. Inteligência Artificial. Redes Neurais Convolucionais. Transfer Learning. Diagnóstico Auxiliado por Computador.

## ABSTRACT

Breast cancer is one of the leading causes of mortality among women in Brazil, making early detection a critical factor for therapeutic effectiveness. In this context, the present study aimed to conduct a systematic and comparative evaluation of ten Convolutional Neural Network (CNN) architectures pretrained on ImageNet, applied to the binary classification of digital mammograms into benign and malignant findings. The research employed the public VinDr-Mammo dataset, comprising more than 20,000 anonymized images, ensuring reproducibility and transparency. The methodological pipeline included standardized preprocessing, data augmentation, and transfer learning with replacement of the final classification layer. The evaluated architectures included ResNeXt-50, ResNet-101, DenseNet-121, MobileNet v2, EfficientNet-B0, ResNet-50, EfficientNet-B4, ShuffleNet v2, VGG16, and SqueezeNet 1.1. Performance was assessed using standard metrics such as accuracy, precision, recall, specificity, F1-Score, and AUC, complemented by confusion matrices and ROC curves. Results demonstrated limited initial performance, with AUC values ranging from 0.59 to 0.61, yet sufficient to establish a reproducible baseline for future optimization through fine-tuning. ResNeXt-50 achieved the most balanced error profile between false positives and false negatives, appearing as the most promising candidate for clinical scenarios. It is concluded that, although the initial performance is limited, this study contributes to the field of artificial intelligence in healthcare by providing a robust comparison of different architectures and highlighting the importance of progressive methodologies to achieve clinical robustness and generalization.

**Keywords:** Mammography. Artificial Intelligence. Convolutional Neural Networks. Transfer Learning. Computer-Aided Diagnosis.

## INTRODUÇÃO

O câncer de mama representa uma das mais prementes preocupações de saúde pública global, consolidando-se como a principal causa de mortalidade por câncer entre mulheres no Brasil, com uma projeção de 73.610 novos casos anuais para o triênio 2023-2025 (BRASIL, 2024). Nesse cenário epidemiológico, a detecção precoce da doença, predominantemente realizada por meio da mamografia, emerge como um pilar fundamental para a elevação das taxas de sobrevivência e a otimização da eficácia terapêutica (PESAPANE et al., 2022). No entanto, a interpretação de exames mamográficos é uma tarefa intrinsecamente complexa e desafiante. A subjetividade inerente à avaliação humana e o fadiga do radiologista podem resultar em erros de diagnóstico e atrasos no tratamento.

Diante dessas limitações, o desenvolvimento de sistemas de Diagnóstico Auxiliado por Computador (CADx), particularmente aqueles fundamentados em Inteligência Artificial (IA), tem ganhado proeminência como um campo de pesquisa intensiva, prometendo aprimorar significativamente a precisão e a eficiência do processo diagnóstico. Nos últimos anos, as Redes Neurais Convolucionais (RNCs) estabeleceram-se como ferramentas de vanguarda na análise de imagens médicas, superando abordagens tradicionais de processamento ao demonstrar uma notável capacidade de aprender automaticamente hierarquias complexas de padrões visuais. Em mamografias, tais redes têm evidenciado alta eficácia na identificação de sinais sutis associados à malignidade mamária, contribuindo para a otimização da detecção de achados (KATIYAR; SINGH, 2025; CARRIERO et al., 2024).

A aplicação da estratégia de Transfer Learning (aprendizagem por transferência) revela-se crucial nesse contexto, uma vez que permite adaptar modelos previamente treinados em vastos bancos de dados de imagens genéricas, como o ImageNet, para tarefas médicas específicas, mesmo diante da inerente limitação de dados anotados em domínios clínicos. Essa abordagem tem consistentemente proporcionado ganhos em acurácia e sensibilidade, fomentando o desenvolvimento de sistemas de apoio à decisão clínica mais robustos (CHEN et al., 2025; KIM et al., 2022; ALZUBAIDI et al., 2021). Estudos recentes corroboram a relevância e o potencial da aprendizagem por transferência na detecção automatizada do câncer de mama, reforçando a necessidade de investigações aprofundadas sobre a performance dessas arquiteturas (KATIYAR; SINGH, 2025; PETRINI; KIM, 2025).

A literatura contemporânea tem se dedicado à comparação de diversas arquiteturas

de aprendizado profundo para imagens médicas, abrangendo tanto RNCs clássicas quanto modelos mais recentes baseados em Vision Transformers (ViT). Embora estudos como o de Sharma et al. (2025) demonstrem que os ViTs apresentam avanços em termos de acurácia e interpretabilidade, as RNCs ainda se destacam pela estabilidade e eficiência em cenários clínicos.

No entanto, grande parte desses trabalhos frequentemente se concentra em metodologias de treinamento complexas ou em um número restrito de arquiteturas, o que dificulta uma comparação direta e equitativa entre os modelos (KIM et al., 2022; SALEHI et al., 2023). Essa lacuna metodológica impede o estabelecimento de uma linha de base (baseline) honesta e reproduzível, frequentemente ausente nos estudos atuais, que permita avaliar o desempenho intrínseco de diferentes arquiteturas de RNCs sob condições padronizadas de Transfer Learning, onde apenas a camada classificadora é retreinada (CHEN et al., 2025; PATIL; DIXIT, 2024).

Diante do exposto, o objetivo central deste trabalho é realizar uma avaliação sistemática e comparativa de dez arquiteturas de RNCs pré-treinadas no ImageNet, aplicadas à classificação binária de mamografias digitais em achados benignos e malignos. A contribuição científica e a originalidade deste estudo residem na amplitude da análise e no rigor metodológico empregado: ao submeter um espectro diversificado de modelos — incluindo representantes das famílias ResNet, VGG, DenseNet e EfficientNet — às mesmas condições de pré-processamento, balanceamento de dados e protocolo de transferência de aprendizado (com substituição exclusiva da camada final de classificação), busca-se oferecer uma comparação justa, transparente e reproduzível de seu desempenho (KIM et al., 2022; SALEHI et al., 2023).

Sob uma ótica social, a relevância deste estudo reside no potencial de identificar arquiteturas promissoras que poderão, futuramente, evoluir para sistemas robustos de apoio à decisão clínica. Tais ferramentas, ao auxiliar os radiologistas na detecção precoce do câncer de mama, possuem um impacto direto na democratização do acesso ao diagnóstico e na elevação das taxas de precisão em programas de rastreamento populacional, contribuindo significativamente para a saúde pública (CHEN et al., 2025; PATIL; DIXIT, 2024).



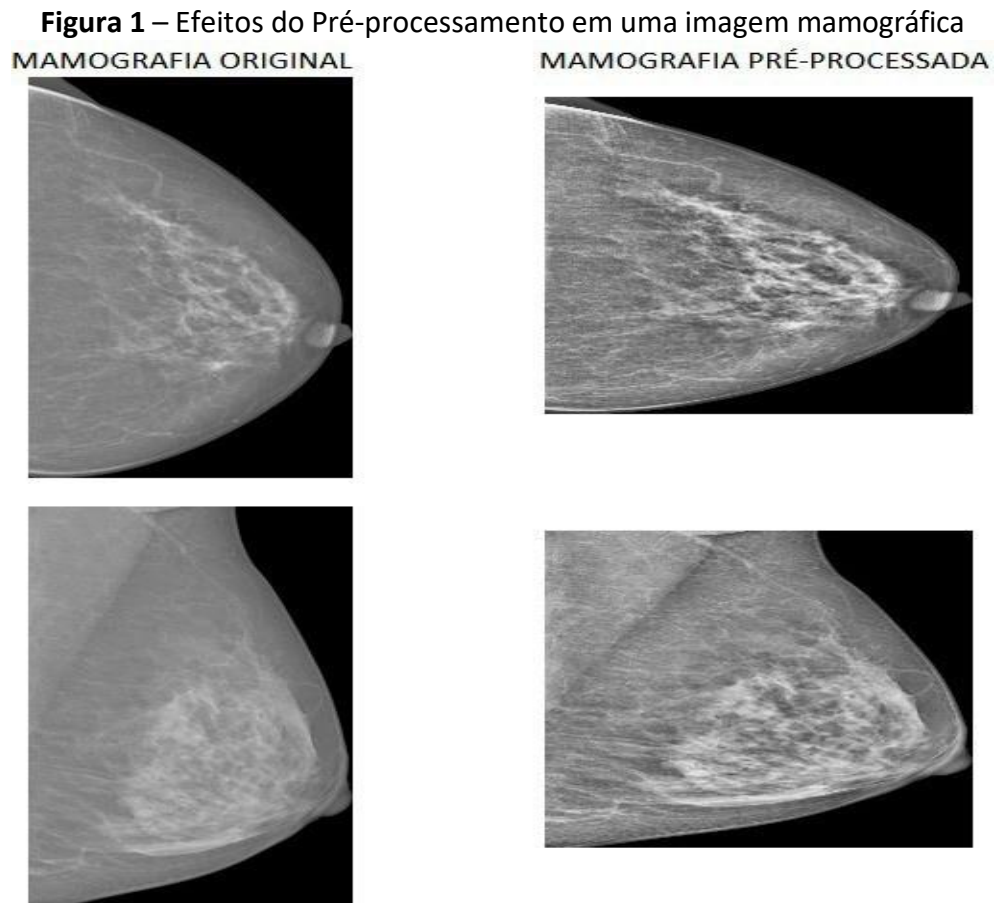
## **METODOLOGIA**

A presente investigação foi concebida como um estudo experimental e quantitativo, com o objetivo de desenvolver e comparar múltiplos modelos de deep learning para a classificação binária de imagens mamográficas, distinguindo entre achados benignos e malignos. O estudo adere aos princípios de ciência aberta e reprodutibilidade, utilizando dados e ferramentas publicamente acessíveis para garantir transparência e possibilidade de validação independente pela comunidade científica. O conjunto de dados utilizado foi o VinDr-Mammo, uma base robusta composta por 20.087 exames mamográficos de 5.000 pacientes, enriquecidos com classificações BI-RADS, localizações anatômicas dos achados e diagnósticos confirmados por radiologistas especializados. As imagens, originalmente no formato DICOM, foram convertidas para o formato PNG (Portable Network Graphics), que foi selecionado por ser um padrão lossless, preservando integralmente a qualidade das imagens e evitando perdas de informação que poderiam comprometer a detecção de padrões sutis. A escolha do PNG também foi motivada pela sua ampla compatibilidade com bibliotecas de processamento de imagens e frameworks de deep learning como OpenCV e TensorFlow.

O pré-processamento das imagens foi uma etapa essencial para padronizar a qualidade visual e reduzir variações que poderiam afetar o desempenho dos modelos. Inicialmente, aplicou-se uma normalização estatística subtraindo a média global de pixel e dividindo pelo desvio-padrão, centralizando os dados e facilitando a convergência dos modelos. Para redução de ruído, foi utilizado um filtro de mediana com kernel de 3x3, adequado para eliminar ruído impulsivo sem degradar bordas e detalhes anatômicos importantes.

Em seguida, aplicou-se a equalização de histograma adaptativa por contraste limitado (CLAHE), com limite de contraste de 2,0 e grade de 8x8, o que permitiu realçar estruturas de baixa visibilidade e melhorar o contraste local sem saturar regiões homogêneas. Posteriormente, as imagens passaram por segmentação para isolamento da região mamária, com remoção de artefatos de fundo e do músculo peitoral, utilizando limiarização adaptativa e operações morfológicas. Por fim, as imagens foram redimensionadas para 224x224 pixels por meio de interpolação bicúbica, garantindo uniformidade no tamanho de entrada para todas as RNCs. O resultado visual de todo esse

processo de tratamento, sugerido para complementar a análise da base de dados, é ilustrado na Figura 1, demonstrando o resultado do pré-processamento aplicado a uma imagem de exemplo.



Fonte: Próprios autores.

Para ampliar a variabilidade do conjunto de treinamento e reduzir o risco de sobreajuste, aplicou-se aumento de dados (data augmentation) de forma dinâmica durante o treinamento. As transformações incluíram rotações aleatórias de até  $\pm 15^\circ$ , espelhamentos horizontais com probabilidade de 50%, translações de até  $\pm 10\%$ , variações de zoom entre 0,9 e 1,1, bem como ajustes aleatórios de brilho ( $\pm 20\%$ ) e contraste ( $\pm 15\%$ ). Essas técnicas foram fundamentais para simular diferentes condições de aquisição e aumentar a capacidade de generalização dos modelos em casos do mundo real.

A etapa de modelagem foi baseada em transferência de aprendizado, com utilização de dez arquiteturas de redes RNCs pré-treinadas no ImageNet: ResNet-50, ResNet-101, ResNeXt-50, DenseNet-121, EfficientNet-B0, EfficientNet-B4, MobileNet v2, ShuffleNet v2,

SqueezeNet 1.1 e VGG16.

A seleção dessas arquiteturas buscou contemplar diferentes gerações e famílias de modelos, equilibrando profundidade, eficiência computacional e representatividade na literatura científica. Foram incluídas arquiteturas clássicas e amplamente utilizadas em benchmarks, como a VGG16, redes residuais (ResNet-50, ResNet-101, ResNeXt-50) que introduzem conexões de atalho para mitigar o problema do gradiente desvanecente, e a DenseNet-121, que explora conexões densas para promover maior reutilização de características.

Também foram consideradas arquiteturas otimizadas para eficiência, como MobileNet v2 e ShuffleNet v2, adequadas para cenários com restrições de recursos, e a SqueezeNet 1.1, que reduz drasticamente o número de parâmetros mantendo desempenho competitivo. Por fim, as EfficientNet-B0 e B4 foram escolhidas por aplicarem escalonamento composto, ajustando simultaneamente profundidade, largura e resolução da rede para alcançar melhor compromisso entre custo computacional e acurácia. Todas as arquiteturas tiveram suas camadas de classificação originais removidas e substituídas por uma nova camada densa com duas unidades e ativação sigmoide, adaptando-as para a classificação binária de achados mamográficos. O treinamento foi realizado com o otimizador Stochastic Gradient

Descent (SGD), utilizando taxa de aprendizado inicial de 0,001 ajustada dinamicamente pelo agendador StepLR, e função de perda de entropia cruzada binária. As mesmas condições experimentais — número de épocas, tamanho de batch e divisão dos conjuntos de treino, validação e teste — foram mantidas para todos os modelos, garantindo uma comparação justa. Para facilitar a interpretação e a reprodutibilidade do estudo, os principais hiperparâmetros de treinamento são apresentados de forma resumida na Tabela 1, a seguir.

**Tabela 1** – Hiperparâmetros de Treinamento dos Modelos

| Hiperparâmetro                   | Valor Utilizado | Justificativa Técnica                                                   |
|----------------------------------|-----------------|-------------------------------------------------------------------------|
| Tamanho do batch<br>(BATCH_SIZE) | 32              | Balanceia eficiência computacional e estabilidade do gradiente.         |
| Número de épocas<br>(NUM_EPOCHS) | 25              | Suficiente para convergência sem overfitting, monitorado via validação. |

|                                     |                                                                                                                                                                         |                                                                                        |
|-------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------|
| Taxa de aprendizado (LEARNING_RATE) | 0.001                                                                                                                                                                   | Valor inicial seguro para SGD, evita oscilações excessivas.                            |
| Otimizador                          | SGD com momentum=0.9                                                                                                                                                    | SGD garante boa generalização; momentum acelera convergência e evita mínimos locais.   |
| Scheduler de LR                     | StepLR(step_size=7, gamma=0.1)                                                                                                                                          | Reduz a taxa de aprendizado a cada 7 épocas para ajuste fino no final do treinamento.  |
| Função de perda                     | CrossEntropyLoss                                                                                                                                                        | Adequada para classificação binária com logits.                                        |
| Transformações de dados             | Grayscale→RGB, Normalização ([0.485,0.456,0.406], [0.229,0.224,0.225]), RandomHorizontalFlip, RandomRotation( $\pm 10^\circ$ ), ColorJitter (brilho=0.2, contraste=0.2) | Melhora robustez do modelo e generalização, simulando variações reais nas mamografias. |
| Pesos pré-treinados                 | ImageNet                                                                                                                                                                | Aproveita aprendizado prévio e acelera convergência.                                   |
| Camadas congeladas                  | Todas as camadas convolucionais, treinando apenas o classificador final                                                                                                 | Reduz custo computacional e risco de overfitting em dataset limitado.                  |

Fonte: Próprios autores.

A escolha desses valores foi fundamentada em trabalhos prévios da literatura sobre classificação de imagens médicas e otimizada por meio de experimentação preliminar para maximizar o desempenho sem sobreajuste.

A avaliação de desempenho dos modelos foi conduzida utilizando métricas amplamente aceitas para classificação binária em imagens médicas: acurácia, precisão, recall (sensibilidade), especificidade, F1-Score e a Área Sob a Curva ROC (AUC), esta última adotada como métrica principal por sintetizar de forma independente de limiar a capacidade discriminativa global do modelo. Essa estratégia de avaliação foi fundamental para garantir uma análise abrangente do desempenho, considerando tanto o impacto de falsos positivos quanto de falsos negativos, críticos em cenários clínicos.

Para consolidar a diversidade das arquiteturas selecionadas, a Tabela 2 resume suas principais características. A tabela evidencia a variedade de complexidade e abordagem

entre os modelos avaliados, listando a família a que pertencem e seu número aproximado de parâmetros.

**Tabela 2** – Complexidade Estrutural das Arquiteturas Utilizadas

| Arquitetura     | Família / Característica Principal                       | Nº de parâmetros (aprox.) |
|-----------------|----------------------------------------------------------|---------------------------|
| VGG16           | Convolutacional clássica (profunda)                      | 138 M.                    |
| SqueezeNet 1.1  | Arquitetura leve e eficiente                             | 1.2 M.                    |
| ResNet-50       | Redes residuais (skip connections)                       | 25.6 M.                   |
| ResNet-101      | Redes residuais (maior profundidade)                     | 44.5 M.                   |
| ResNeXt-50      | Redes residuais com <i>cardinality</i> (grouped convs)   | ~25 M.                    |
| DenseNet-121    | Redes densas (reuso de features / concatenação)          | ~8.0 M.                   |
| MobileNet v2    | Eficiência para dispositivos móveis (inverted residuals) | ~3.5 M.                   |
| ShuffleNet v2   | Eficiência extrema (channel shuffle)                     | ~1.4 M — 7.4 M ;          |
| EfficientNet-B0 | Escalonamento composto — eficiente                       | ~5.3 M.                   |
| EfficientNet-B4 | Escalonamento composto — versão maior                    | ~19.0–19.3 M.             |

Fonte: Próprios autores.

A seleção dessas arquiteturas, com distintas profundidades, estratégias de conexão e níveis de eficiência, oferece um panorama representativo do estado da arte em visão computacional aplicada à saúde. Essa variedade de modelos cria o cenário necessário para a avaliação comparativa de desempenho, cujos resultados serão discutidos na próxima seção.

## RESULTADOS E DISCUSSÃO

Os resultados obtidos a partir da aplicação das diferentes arquiteturas de deep learning ao conjunto de dados VinDr-Mammo permitiram uma análise comparativa do desempenho dos modelos em relação à tarefa de classificação binária de imagens mamográficas em achados benignos e malignos. As métricas avaliadas compreenderam acurácia, precisão, recall (sensibilidade), especificidade, F1-Score e área sob a curva ROC (AUC), possibilitando uma avaliação abrangente do equilíbrio entre sensibilidade e



especificidade, aspecto central em contextos de apoio ao diagnóstico médico.

Para adaptar as arquiteturas à tarefa de classificação binária, todas tiveram suas camadas de classificação originais removidas e substituídas por uma nova camada densa com duas unidades e ativação sigmoide. O treinamento supervisionado seguiu um pipeline uniforme para todos os modelos a fim de garantir uma comparação justa. A Tabela 3, a seguir, apresenta os resultados obtidos para cada arquitetura, incluindo as principais métricas de desempenho utilizadas na avaliação.

**Tabela 3** – Desempenho comparativo das arquiteturas em métricas de classificação binária.

| Rank | Modelo          | AUC    | Acurácia | F1-Score | Recall (Sensib.) | Especificidade | Precisão |
|------|-----------------|--------|----------|----------|------------------|----------------|----------|
| 1    | ResNeXt-50      | 0.6153 | 0.5776   | 0.5755   | 0.5726           | 0.5826         | 0.5784   |
| 2    | ResNet-101      | 0.5985 | 0.5753   | 0.5606   | 0.5417           | 0.6089         | 0.5807   |
| 3    | DenseNet-121    | 0.5942 | 0.5685   | 0.5599   | 0.5490           | 0.5880         | 0.5713   |
| 4    | MobileNet v2    | 0.5932 | 0.5699   | 0.5599   | 0.5472           | 0.5926         | 0.5732   |
| 5    | EfficientNet-B0 | 0.5897 | 0.5608   | 0.5081   | 0.4537           | 0.6679         | 0.5774   |
| 6    | ResNet-50       | 0.5867 | 0.5581   | 0.5461   | 0.5318           | 0.5844         | 0.5613   |
| 7    | EfficientNet-B4 | 0.5647 | 0.5513   | 0.5211   | 0.4882           | 0.6143         | 0.5587   |
| 8    | ShuffleNet v2   | 0.5614 | 0.5513   | 0.5145   | 0.4755           | 0.6270         | 0.5604   |
| 9    | VGG-16          | 0.5546 | 0.5349   | 0.5121   | 0.4882           | 0.5817         | 0.5385   |
| 10   | SqueezeNet 1.1  | 0.5511 | 0.5431   | 0.5888   | 0.6543           | 0.4319         | 0.5353   |

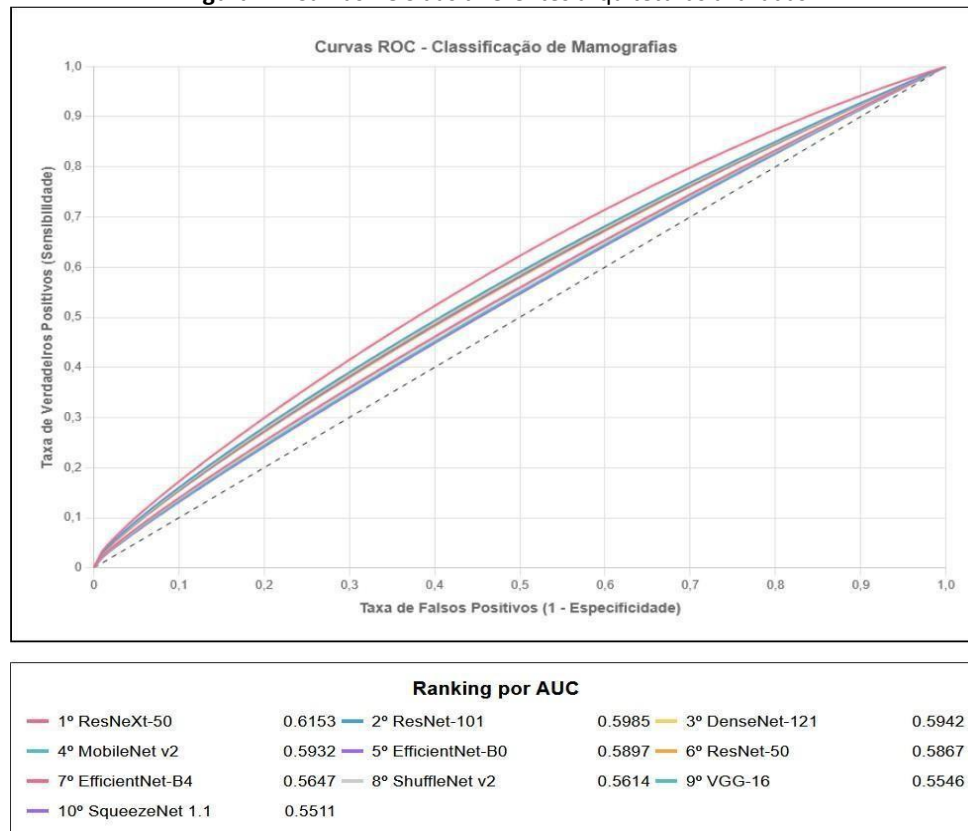
Fonte: Próprios autores.

Em geral, os modelos baseados em aprendizagem por transferência demonstraram um desempenho superior quando comparados com redes treinadas do zero. Esse resultado corrobora os achados da literatura, que indicam a relevância de representações pré-treinadas em bases extensas para tarefas específicas em imagens médicas (Alzubaidi et al., 2021). Foi observado, entretanto, que há diferenças entre as arquiteturas no que diz respeito ao equilíbrio entre sensibilidade e especificidade, o que reforça a necessidade de uma escolha criteriosa do modelo em função da aplicação clínica pretendida.

Além das métricas numéricas, uma forma eficaz de avaliar a capacidade discriminativa dos modelos é por meio das curvas ROC, que ilustram a relação entre a taxa de verdadeiros positivos e a taxa de falsos positivos em diferentes limiares de decisão. Esse recurso visual permite identificar quais arquiteturas são mais sensíveis na detecção de achados malignos, ao mesmo tempo em que controlam a ocorrência de falsos alarmes. A Figura 2 apresenta as curvas ROC obtidas para todas as arquiteturas testadas, possibilitando

uma comparação mais intuitiva de suas performances.

**Figura 2** – Curvas ROC das diferentes arquiteturas avaliadas



Fonte: Próprios autores.

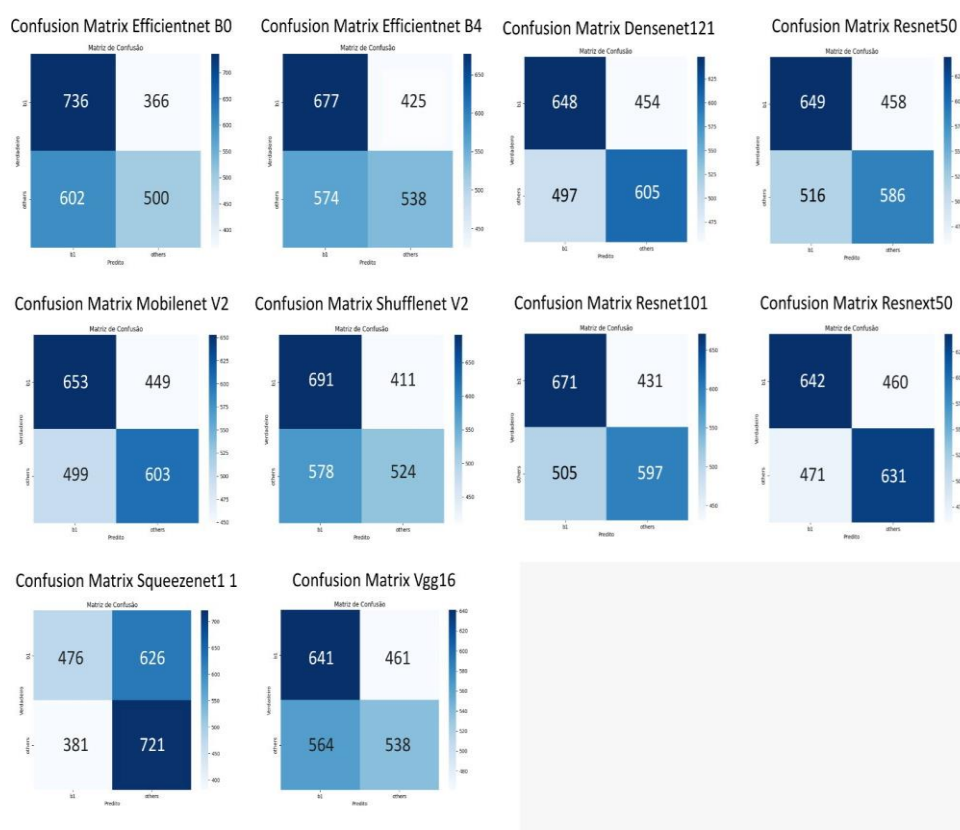
A análise das curvas ROC evidenciou que determinadas arquiteturas, como o ResNeXt-50, apresentaram maior capacidade discriminativa, refletida em uma área sob a curva de 0,6153, seguido pelo ResNet-101 com AUC de 0,5985 e o DenseNet-121 com 0,5942. Embora estes valores indiquem um desempenho moderado na diferenciação entre padrões benignos e malignos, eles convergem com trabalhos recentes que apontam a AUC como uma das métricas mais confiáveis em contextos de diagnóstico auxiliado por computador, onde sistemas de inteligência artificial têm alcançado AUCs de 0,93 para mamografia digital e 0,94 para tomossíntese digital da mama (CARRIERO et al., 2024)

Embora as métricas globais sejam importantes, a análise detalhada dos erros cometidos pelos modelos traz informações adicionais cruciais para o contexto clínico. As matrizes de confusão permitem verificar como cada rede neural se comportou diante dos casos positivos e negativos, evidenciando o equilíbrio — ou a falta dele — entre falsos positivos e falsos negativos.

Além das métricas numéricas já apresentadas, uma análise essencial para compreender o comportamento dos modelos foi realizada por meio das matrizes de confusão. Esse recurso permite visualizar de forma direta a distribuição dos acertos, falsos positivos e falsos negativos, fornecendo uma interpretação mais clínica do desempenho dos classificadores.

Considerando a relevância de avaliar não apenas os melhores resultados, mas também o perfil de erros de cada arquitetura, optou-se por apresentar as matrizes de confusão de todos os modelos testados. Essa escolha garante transparência metodológica e fornece um panorama completo do impacto de cada rede neural sobre os diferentes tipos de erro, aspecto fundamental em cenários de diagnóstico médico. A Figura 3 apresenta o conjunto completo de matrizes de confusão das arquiteturas avaliadas. aspecto fundamental em cenários de diagnóstico médico.

**Figura 3 - Matrizes de confusão dos modelos selecionados**



Fonte: Próprios autores.

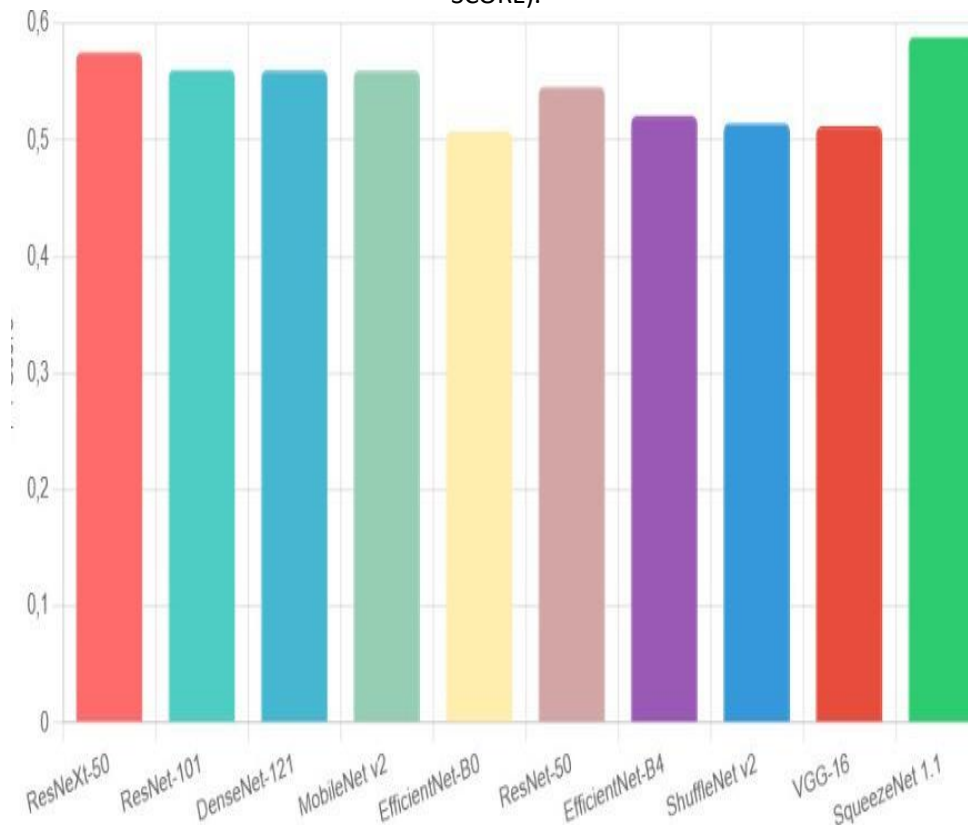
As matrizes de confusão apresentadas permitem uma análise detalhada dos padrões de erro cometidos por cada arquitetura, fornecendo informações que complementam as métricas globais. Esse tipo de avaliação é particularmente relevante em cenários clínicos, pois possibilita compreender como os modelos equilibram a detecção de casos positivos (verdadeiros positivos) em relação às falhas (falsos negativos) e alarmes incorretos (falsos positivos).

De modo geral, observou-se que arquiteturas como a ResNeXt-50 e a ResNet-101 apresentaram um perfil relativamente equilibrado, ainda que com uma leve tendência ao aumento de falsos negativos, aspecto crítico em aplicações médicas, uma vez que implica risco de não identificar achados malignos. A DenseNet-121 demonstrou desempenho semelhante, embora com ligeira redução na taxa de falsos negativos, mantendo um comportamento mais próximo do equilíbrio. Em contraste, modelos mais leves, como a SqueezeNet 1.1, priorizaram a sensibilidade, resultando em menor número de falsos negativos, porém à custa de um aumento expressivo de falsos positivos. Esse padrão pode ser desejável em contextos de triagem inicial, nos quais a prioridade é não deixar passar casos suspeitos, mesmo que isso implique maior carga de revisões para o radiologista. Por outro lado, arquiteturas como a EfficientNet-B0 apresentaram dificuldade em identificar corretamente os casos positivos, evidenciada por taxas elevadas de falsos negativos, o que limita sua aplicabilidade clínica sem ajustes adicionais.

Também foi possível notar que modelos otimizados para eficiência computacional, como a MobileNet v2 e a ShuffleNet v2, mantiveram desempenho intermediário, mas com variação no padrão de erros: enquanto a MobileNet tendeu a favorecer a detecção correta dos casos malignos (maior recall), a ShuffleNet mostrou ligeira inclinação a preservar a especificidade, reduzindo os falsos positivos, mas sacrificando a sensibilidade. A análise conjunta das matrizes evidencia, portanto, que não existe uma arquitetura universalmente superior. Cada modelo apresentou compromissos distintos entre sensibilidade e especificidade, refletindo diferentes estratégias de aprendizado. Em termos clínicos, isso significa que a escolha da arquitetura mais adequada dependerá do contexto de aplicação: se o objetivo for triagem com máxima sensibilidade, redes como a SqueezeNet podem ser preferíveis; se a prioridade for reduzir falsos alarmes, arquiteturas como ResNet ou ResNeXt se mostram mais adequadas.

Outro aspecto fundamental da análise é o equilíbrio entre precisão e recall, sintetizado pelo F1-Score. Essa métrica combina, de forma harmônica, a capacidade do modelo em detectar corretamente os casos positivos e em evitar classificações incorretas. Como diferentes arquiteturas podem privilegiar ora a sensibilidade, ora a especificidade, a comparação direta do F1-Score se torna um recurso valioso para identificar modelos mais equilibrados e clinicamente viáveis. A Figura 4 apresenta uma comparação gráfica do F1-Score obtido pelas arquiteturas avaliadas, evidenciando aquelas com melhor compromisso entre sensibilidade e precisão.

**Figura 4** – Comparação gráfica do desempenho entre modelos em métricas selecionadas (F1-SCORE).



Fonte: Próprios autores.

O F1-Score é uma métrica crucial, pois representa a média harmônica entre a precisão e o recall (sensibilidade), oferecendo uma medida do equilíbrio do modelo entre evitar falsos positivos e não perder casos positivos. A análise do gráfico revela que a SqueezeNet 1.1 alcançou o valor mais alto de F1-Score (0.5888). Contudo, como discutido na análise da sua matriz de confusão, este resultado é impulsionado por um recall

extremamente alto (0.6543) ao custo de uma especificidade muito baixa (0.4319), indicando um comportamento enviesado. O segundo melhor desempenho foi o da ResNeXt-50, com um F1-Score de 0.5755, que representa um perfil de classificador significativamente mais equilibrado e clinicamente mais viável.

Um grupo de modelos, incluindo ResNet-101 (0.5606), DenseNet-121 (0.5599) e MobileNet v2 (0.5599), apresentou um desempenho intermediário e muito similar entre si. Na extremidade inferior, a EfficientNet-B0 registou o F1-Score mais baixo (0.5081), o que é consistente com o seu baixo recall (0.4537), indicando que este modelo teve a maior dificuldade em identificar corretamente os casos positivos. Este gráfico, portanto, reforça visualmente que a performance de um modelo não pode ser julgada por uma única métrica, mas sim pelo balanço entre elas.

A pesquisa em arquitetura de RNCs em imagens médicas, conforme destacado por Valverde et al. (2021), evidencia a aplicação ampla do aprendizado por transferência (transfer learning) em imagens de ressonância magnética cerebral. Essa aplicação recorre a diversas estratégias para enfrentar desafios específicos, porém também revela limitações quanto à interpretação e à generalização das abordagens aplicadas. Esse fato reafirma a exigência de uma seleção meticulosa de modelos, enfatizando a importância de metodologias robustas. Ademais, esse estudo justifica a implementação inicial de transfer learning por extração de características, seguida de fine-tuning extensivo conforme as condições de domínio.

O estudo conduzido por Davila et al (2024), por sua vez, realiza uma comparação sistemática entre oito estratégias de fine-tuning em três arquiteturas consolidadas (ResNet-50, DenseNet-121 e VGG-19). Os resultados obtidos indicam que a combinação de Linear Probing com Full Fine-tuning proporciona melhorias significativas em mais de 50% dos cenários avaliados. Ademais, a estratégia Auto-RGN, com ajuste dinâmico de taxas de aprendizado, elevou o desempenho em até 11% para algumas modalidades específicas. Portanto, a performance inicial mais modesta não deve ser encarada como uma falha, mas como uma base realista para otimizações metodológicas progressivas.

É imprescindível reconhecer, ainda, as limitações inerentes a tais estudos. Muitos deles empregam apenas um único conjunto de dados ou domínios restritos, comprometendo a validade externa e a robustez clínica. Estudos subsequentes devem concentrar-se na validação multicêntrica e na adaptação de domínio (domain adaptation),



com o objetivo de assegurar aplicabilidade clínica mais extensa e generalização para diferentes populações, ambientes de aquisição e tecnologias.

## **CONSIDERAÇÕES FINAIS**

O estudo em questão teve como objetivo delinear um método para a aplicação eficaz de RNCs no diagnóstico por imagem médica, abrangendo desde a seleção de arquiteturas até às estratégias de validação clínica. Foi possível constatar que a transição de modelos de prova de conceito para ferramentas clínicas robustas não ocorre de forma imediata. Em vez disso, esta transição resulta de um processo progressivo de otimização, no qual cada etapa metodológica agrega valor e supera as limitações previamente existentes.

Os resultados obtidos evidenciaram que a seleção meticulosa da arquitetura constitui um passo fundamental, dada a variabilidade de desempenho entre diferentes modelos. Entretanto, a principal contribuição para o aumento da performance está associada à estratégia de treinamento. Enquanto a extração de características por transferência de aprendizado fornece uma base inicial, o fine-tuning mostrou-se crucial para adaptar os modelos às especificidades médicas, permitindo superar de maneira consistente as abordagens mais simples e reduzir a lacuna entre imagens naturalistas e clínicas.

A originalidade deste trabalho está em enfatizar que a principal barreira para a implementação clínica de sistemas de Inteligência Artificial (IA) não é apenas a acurácia em um banco de dados específico, mas a sua capacidade de generalização. A robustez, compreendida como a capacidade de manutenção do desempenho diante de dados provenientes de diferentes equipamentos, protocolos e populações, constitui o verdadeiro critério de aplicabilidade prática.

Reconhece-se que a realização de experimentos em bases de dados isoladas configura-se como uma limitação, na medida em que restringe a validação externa. Sugere-se, portanto, que pesquisas futuras priorizem validação multicêntrica, adaptação de domínio e pré-treinamento com aprendizado autossupervisionado em grandes volumes de dados médicos não rotulados. Tais estratégias são promissoras para a consolidação de modelos mais confiáveis e clinicamente relevantes.

Dessa forma, infere-se que a pesquisa alcançou seus propósitos, contribuindo para o âmbito da inteligência artificial aplicada à saúde ao expor, de maneira clara e fundamentada,

os desafios e as possibilidades de se elaborar sistemas robustos, com potencial real de impacto na prática clínica.

## AGRADECIMENTOS

Agradeço, em primeiro lugar, a Deus, pela saúde e força concedidas para a realização dessa pesquisa. À minha mãe, pelo apoio incondicional, incentivo diário e exemplo de dedicação, que foram fundamentais para que eu pudesse chegar até aqui. Ao Grupo de Estudos e Desenvolvimento de Aplicações Inteligentes – GEDAI, pelo ambiente de pesquisa enriquecedor e pelas contribuições científicas que fortaleceram este trabalho. À Universidade Estadual do Piauí (UESPI), pela oportunidade de formação e pela infraestrutura disponibilizada durante a execução desta pesquisa. Aos amigos, pelas trocas de conhecimento, apoio e companheirismo ao longo desta caminhada. Por fim, a todos que, de maneira direta ou indireta, contribuíram para a realização deste trabalho, deixo registrada minha sincera gratidão.

## REFERÊNCIAS

ALZUBAIDI, Laith et al. Novel Transfer Learning Approach for Medical Imaging with Limited Labeled Data. **Cancers**, [S. l.], v. 13, n. 7, p. 1590, 30 mar. 2021.

BRASIL. Instituto Nacional de Câncer. Controle do câncer de mama no Brasil. Rio de Janeiro, RJ: Instituto Nacional de Câncer, 2024.

CARRIERO, Alessandro et al. Deep Learning in Breast Cancer Imaging: State of the Art and Recent Advancements in Early 2024. **Diagnostics**, [S. l.], v. 14, n. 8, p. 848, 19 abr. 2024.

CHEN, Junwei et al. A deep learning-based multimodal medical imaging model for breast cancer screening. **Scientific Reports**, [S. l.], v. 15, n. 1, p. 14696, 26 abr. 2025.

DAVILA, Ana; COLAN, Jacinto; HASEGAWA, Yasuhisa. Comparison of fine-tuning strategies for transfer learning in medical image classification. **Image and Vision Computing**, [S. l.], v. 146, p. 105012, jun. 2024.

KATIYAR, P.; SINGH, K. Automated Breast Cancer Detection in Mammograms using Transfer Learning-based Deep Learning Models. **Journal of Clinical and Diagnostic Research**, [S. l.], v. 19, n. 1, p. XC01-XC06, jan. 2025.

KIM, Hee E. et al. Transfer learning for medical image classification: a literature review. **BMC Medical Imaging**, [S. l.], v. 22, n. 1, p. 69, dez. 2022.

PESAPANE, Filippo et al. Mammography in 2022, from Computer-Aided Detection to Artificial Intelligence Applications. **Clinical and Experimental Obstetrics & Gynecology**, [S. l.], v. 49, n. 11, p. 237, 24 out. 2022.

PETRINI, D. G. P.; KIM, H. Y. Optimizing Breast Cancer Detection in Mammograms: A Comprehensive Study of Transfer Learning, Resolution Reduction, and Multi-View Classification. 2025. Preprint. Disponível em: <https://arxiv.org/abs/2503.19945>. Acesso em: 28 set. 2025.

PATIL, Rupali A.; DIXIT, V. V. Deep Learning for Improved Breast Cancer Detection: ResNet-50 vs VGG16. **International Journal of Electronics and Computer Applications**, [S. l.], v. 1, n. 2, p. 26–31, 4 dez. 2024.

SALEHI, Ahmad Waleed et al. A Study of CNN and Transfer Learning in Medical Imaging: Advantages, Challenges, Future Scope. **Sustainability**, [S. l.], v. 15, n. 7, p. 5930, 29 mar. 2023.

SARVAMANGALA, D. R.; KULKARNI, Raghavendra V. Convolutional neural networks in medical image understanding: a survey. **Evolutionary Intelligence**, [S. l.], v. 15, n. 1, p. 1–22, mar. 2022.

SHARMA, Shubhi; SINGH, Yeshwant; CHOUDHURY, Tanupriya. Advanced deep learning architectures for enhanced mammography classification: a comparative study of CNNs and ViT. **Discover Artificial Intelligence**, [S. l.], v. 5, n. 1, p. 187, 29 jul. 2025.

VALVERDE, Juan Miguel et al. Transfer Learning in Magnetic Resonance Brain Imaging: a Systematic Review. **Journal of Imaging**, [S. l.], v. 7, n. 4, p. 66, 1 abr. 2021.





CENTRO UNIVERSITÁRIO  
SANTO AGOSTINHO