

Detecção de Câncer de Mama em Imagens com Inteligência Artificial

Caroline da Silva Grizante
114105

Fundação Hermínio Ometto
Porto Ferreira, Brazil
carolinegrizante105@alunos.fho.edu.br

Emilly Emanuely Ribeiro dos Santos
114095

Fundação Hermínio Ometto
Mogi Guaçu, Brasil
emillyribeiro@alunos.fho.edu.br

Marcela Lovatto
113626

Fundação Hermínio Ometto
Itapira, Brasil
marcela.lovatto@alunos.fho.edu.br

I. INTRODUCTION

Segundo uma análise realizada pelo INCA em 2024, atualmente no Brasil cerca de 40% dos diagnósticos de câncer de mama são feitos em estágios avançados. Além disso, em uma publicação do Instituto Nacional de Câncer (INCA,2024), voltada para a conscientização do câncer de mama, foram apresentados dados que indicam que esse tipo de câncer é a principal causa de óbitos entre as mulheres diagnosticadas com a doença no Brasil. Sendo apontado como o tipo de câncer mais comum depois do câncer de pele.

Com o objetivo de reduzir estes números é incentivado a detecção precoce do câncer, pois ela possui inúmeros benefícios como: redução da mortalidade e diagnósticos em estágio mais precoce, resultando em uma maior chance de tratamentos menos agressivos e mais efetivos (INCA, 2015). Um dos métodos para diagnosticar esta doença é as lesões impalpáveis, que seriam exames de imagens (ultrassonografia, mamografia ou ressonância) que servem como guias para a detecção do câncer (INCA, 2022). Essas análises são feitas por avaliação humana, o que pode gerar variações nos resultados e em diagnósticos tardios. Levando isto em consideração, surge a necessidade de uma ferramenta auxiliar para estes profissionais, em que os ajudem na detecção mais rápida e eficiente.

O processamento automático de imagens de câncer para auxiliar na determinação de um diagnóstico é um tópico que vem sendo pesquisado há mais de 40 anos. Contudo, ele continua evoluindo devido aos avanços da complexidade das imagens a serem analisadas (Stenkvist et al., 1978). Estudos como o de Spagnol et al.(2016), demonstraram o desempenho das Redes Neurais Convolucionais(CNNs) na detecção de câncer. Ao comparar esse método com classificadores tradicionais, ficou comprovado que, em termos de acurácia, as CNNs obtiveram um resultados superiores, apresentando taxas de acertos superiores a 85%. Além disso, os autores destacam a alta capacidade de generalização das CNNS, que apresentaram desempenho consistente mesmo com as variações de pacientes e de resoluções das imagens. Eles também reforçam que o aprendizado por feature learning é mais eficaz do que a extração manual das características.

A. Objetivo Geral

- Desenvolver uma Rede Neural Convolucional (CNNs) como ferramenta auxiliar para profissionais da saúde, com o objetivo de ajudar na detecção de câncer de mama. O sistema será implementado de forma automática e com uma interface amigável, visando tornar os diagnósticos mais simples, rápidos e intuitivos.

B. Objetivos Específicos

- Aquisição dos dados de fontes relevantes, preparação dos dados (realizar os ajustes necessários), classificação das classes e divisão dos dados (treino, teste e validação).
- Definição da modelagem de implementação e dos parâmetros, treinar uma arquitetura de CNN adequada para a classificação.
- Validar e avaliar o desempenho do modelo utilizando métricas apropriadas, como acurácia, precisão, recall e F1-score.
- Analisar a capacidade de generalização do modelo, para ver o seu desempenho com imagens de diferentes pacientes. Com o objetivo de verificar a robustez da solução proposta.

II. TRABALHOS RELACIONADOS

A literatura recente sobre detecção de câncer de mama via aprendizado de máquina e deep learning mostra resultados bastante superiores aos obtidos no presente trabalho, o que sugere que variáveis cruciais de metodologia e dados, e não apenas a escolha da arquitetura, têm papel determinante no desempenho final. Por exemplo, em um estudo seminal, Deep Learning to Improve Breast Cancer Detection on Screening Mammography (Shen et al., 2019) os autores relatam que “um classificador de imagem completa (‘whole image classifier’), treinado com abordagem end-to-end, atinge acurácia de 0,95 em mamografias digitais (FFDM)” e que, ao usar ensemble de quatro modelos, a AUC chega a 0,98.

Esse elevado desempenho foi conseguido mesmo sem depender de anotações manuais de lesões, o rótulo de benigno vs. maligno por imagem foi suficiente. Tal resultado demonstra o grande potencial de redes profundas bem treinadas em dados de mamografia, desde que alimentadas por uma base

de dados ampla, heterogênea e tecnicamente tratada (pré-processamento, padronização, divisão correta entre treino e teste, etc.).

De forma mais ampla, a Diagnostic Accuracy of Machine Learning Models on Mammography in Breast Cancer Classification: A Meta-Analysis (Hanis, Islam Musa, 2022) compila 36 estudos com 68 modelos diferentes, incluindo deep learning, SVM, redes neurais clássicas e outros, e conclui que, de modo geral, os modelos têm desempenho promissor: a AUC combinada ficou em cerca de 0,90, com sensibilidade média de 0,83 e especificidade média de 0,84.

Esses valores servem como uma espécie de “linha de base esperada” para sistemas de detecção binária (benigno ou maligno) em mamografias. Contudo, os autores desse meta-estudo alertam que há heterogeneidade entre trabalhos, em termos de origem dos dados, qualidade das imagens, tipo de classificador e pré-processamento e que muitos resultados podem sofrer viés de publicação (modelos com desempenho ruim tendem a não ser publicados).

Considerando esses estudos de referência, fica evidente que o desempenho relativamente modesto obtido em nosso experimento, com uma rede convolucional simples, mas uma abordagem de fine-tuning sobre VGG16 combinada a heurísticas manuais, resultando em acurácia máxima de 69%, está abaixo das expectativas segundo a literatura.

A discrepância provavelmente decorre de diversas fragilidades no pipeline: primeiramente, a quantidade e diversidade de dados disponíveis podem ser insuficientes para permitir que uma rede profunda aprenda variações relevantes de textura, densidade e padrão presentes em mamografias. Redes como as usadas por Shen et al. obtêm alta performance justamente por terem acesso a grandes bases, possibilitando generalização. Em contraste, redes simples tendem a underfitar e não capturar os padrões sutis que distinguem tecido benigno e maligno.

Outro ponto relevante é a métrica usada. A acurácia simples pode ser enganosa em problemas com classes desbalanceadas ou com alta diferença entre risco de falso-negativo e falso-positivo. Em estudos bem-sucedidos costuma-se usar AUC, sensibilidade e especificidade, métricas que refletem melhor a capacidade discriminativa do modelo. Assim, é possível que um modelo com 69% de acurácia esteja apresentando sensibilidade ou especificidade muito baixa, o que o torna pouco útil clinicamente, mesmo que a acurácia aparente seja razoável.

III. METODOLOGIA

Este trabalho tem como objetivo desenvolver uma ferramenta auxiliar para a detecção de câncer, com foco específico no câncer de mama. Conforme foi apresentado na introdução, muitos casos o câncer é detectado em estágios avançados, o que reduz significativamente as chances de tratamento eficaz. Pensando em uma solução para este problema, o grupo propôs a criação de uma Inteligência Artificial capaz de analisar imagens de mamografia e indicar se há presença de alterações compatíveis com câncer.

A construção dessa solução apresenta diversos desafios, tais como imagens de baixa qualidade, variações no tamanho e formato dos exames, a dificuldade em detectar tumores pequenos, e a necessidade de garantir que a rede neural aprenda padrões relevantes, em vez de memorizar conjuntos de dados. Além disso, imagens médicas frequentemente possuem ruídos, contraste variável e estruturas complexas que dificultam a análise automatizada.

Para resolver estas questões, optou-se pela utilização de uma Rede Neural Convolucional (CNN), porque ela é eficiente em extrair características importantes da imagem, como bordas, texturas, manchas e padrões anormais associados a tumores. As mamografias foram escolhidas como fonte de dados porque é o exame padrão é amplamente utilizado na detecção precoce do câncer de mama, e possuem alta nitidez e grande riqueza de detalhes na imagem. Isto permite a identificação de estruturas internas do tecido e favorece o desempenho da CNN no processo de aprendizado.

A. Dados e Pré-Processamento

As imagens utilizadas para compor o dataset foram adquiridas de duas fontes públicas de mamografias digitais, um foi do Digital Mammography Dataset for Breast Cancer Diagnosis Research (OZA et al., 2023) encontrado no site Figshare. E a outra foi adquirida no repositório RNSA Mamograph T512px 8bit disponível no Kaggle (DELTAECHOV, 2023).

Na curadoria do dataset, cada caso era composto por três tipos de arquivos: a imagem original em formato TIFF, um arquivo de localização da lesão em formato PLA e Reports textuais detalhando a complexidade da doença (e.g., BIRADS, composição mamária). Para a criação dos rótulos de treinamento, foi realizado um processo de curadoria e anotação manual. O problema foi simplificado para uma classificação binária, composta por:

- 1 (Câncer Presente): Imagens com diagnóstico confirmado da doença
- 0 (Câncer Ausente): Imagens sem evidência da doença

Para os casos positivos (câncer presente), foi utilizada a ferramenta de edição de imagem Photoshop para determinar manualmente o centro da lesão (coordenadas X e Y) e o raio aproximado (em pixels). Essa abordagem de anotação manual permitiu a criação de um ground truth mais preciso, que além da classificação binária, forneceu dados de localização cruciais para futuras análises de detecção de objetos.

Além disso, todas as informações obtidas durante a curadoria, incluindo o nome de cada imagem, a classificação se tem câncer ou é normal, e o rótulo atribuído (0 ou 1), foram organizadas em uma planilha CSV criada especificamente para este projeto. Esse arquivo atuou como a base estruturada do dataset, garantindo rastreabilidade, consistência das anotações manuais e uma padronização essencial para o carregamento dos dados durante o treinamento do modelo.

Durante a fase de curadoria e pré-processamento, a equipe de pesquisa identificou quatro desafios críticos que afetaram diretamente a qualidade e a uniformidade do dataset:

- Baixa qualidade das radiografias: Ruídos e a degradação do contraste comprometem a extração de características relevantes pelo modelo de IA. A solução metodológica inicial adotada foi não utilizar estas imagens no treinamento do modelo, para não prejudicar o aprendizado do modelo.
- Banco de dados coloridos: A presença de canais de cor irrelevantes (RGB) em uma modalidade intrinsecamente em escala de cinza aumenta a dimensionalidade e o ruído. A solução inicial foi a conversão manual para escala de cinza (Grayscale).
- Tamanho de imagens muito diferentes: A heterogeneidade no tamanho das imagens exige um redimensionamento forçado, que pode causar distorção ou perda de informações cruciais. A solução inicial foi o redimensionamento para um tamanho fixo 1024x1024, com a ressalva do potencial de perda de resolução.
- Cânceres muito pequenos: A dificuldade na detecção de microcalcificações ou lesões iniciais pode levar a falsos negativos. A solução inicial foi a exclusão dessas imagens do banco de dados.

O dataset final resultante é, portanto, composto por imagens padronizadas, anotadas manualmente e organizadas em duas classes (1, 0), pronto para uso em experimentos de classificação e, futuramente, detecção de lesões.

B. Abordagem

Neste trabalho foi utilizado uma Rede Neural Convolucional (CNN) simples, em que a arquitetura foi construída especificamente para este projeto. Escolhemos esta abordagem de IA devido a alta eficiência das CNNs em extrair características importantes das imagens, e consequentemente permitindo que o modelo aprenda padrões visuais essenciais para a classificação. As CNNs são amplamente aplicadas em problemas de visão computacional, justamente por seu excelente desempenho na detecção de padrões e objetos, além da flexibilidade para ajustar hiper parâmetros para melhorar o desempenho do modelo.

C. Arquitetura

Para a construção do modelo, foi utilizada uma CNN composta por três camadas convolucionais sequenciais, contendo respectivamente 32, 64 e 128 filtros, todos de tamanho 3×3 , ativação ReLU e padding “same”. Essas camadas são responsáveis pela extração de padrões visuais presentes nas imagens, como contornos, texturas e pequenas estruturas relevantes. Após cada camada convolucional, foi aplicada uma camada de MaxPooling2D, com objetivo de reduzir a dimensionalidade e destacar apenas os padrões mais fortes, diminuindo o risco de sobreajuste.

Posteriormente, as ativações foram achatadas (Flatten), para permitir sua entrada em uma camada totalmente conectada com 128 neurônios e ativação ReLU, destinada à combinação das características extraídas. Em seguida de uma camada de Dropout de 40%, aplicada para reduzir o overfitting ao desligar ao desativar aleatoriamente parte dos neurônios durante o

treinamento. Por fim, a camada de saída é composta por um único neurônio com ativação sigmoid, adequado para a tarefa de classificação binária entre imagens normais e imagens com suspeita de câncer.

D. Tecnologias Utilizadas

Este modelo foi desenvolvido na plataforma Kaggle, utilizando um notebook da própria plataforma, onde também foram realizados o carregamento e organização do dataset. O código foi escrito na linguagem Python, devido à sua ampla utilização em aplicações de Inteligência Artificial, e possui diversas bibliotecas específicas que possibilitam a manipulação de imagens, construir redes neurais, e visualizar resultados. As principais bibliotecas utilizadas no projeto foram: os, NumPy, Pandas, PIL (Pillow), Matplotlib, TensorFlow/Keras, e sklearn.metrics.

E. Metodologia e Validação

O desenvolvimento do modelo seguiu o fluxo de um pipeline com cinco etapas principais, que seriam: o carregamento do dataset, pré-processamento das imagens, construção da arquitetura da CNN, treinamento e validação. Inicialmente, as imagens foram relacionadas com os rótulos definidos na planilha criada durante a curadoria, para garantir que os dados estejam classificados nas classes corretas. Logo em seguida, todas as imagens foram pré-processadas, para serem organizadas nos conjuntos de treino, validação e teste. No treinamento, foi aplicado técnicas simples de data augmentation no conjunto de dados, para melhorar a generalização.

A validação foi feita em cada época utilizando o conjunto de validação. Além disto, foi adotado o Early Stopping e Model Checkpoint, que interrompem o treinamento quando não há melhora depois de uma determinada quantidade de épocas, e armazenam o melhor modelo alcançado. Por fim, para avaliar o desempenho final foi utilizado um conjunto de testes, com o cálculo de acurácia, AUC, relatório de classificação e matriz de confusão, garantindo uma avaliação confiável.

F. Resultados

Discussão e interpretação dos resultados em relação aos objetivos do projeto e às métricas de avaliação definidas. O objetivo do projeto era desenvolver um modelo capaz de atingir pelo menos 80% de acurácia. Entretanto, durante os testes, o valor máximo obtido na validação foi de 69%, valor inferior ao esperado. Definimos como métricas de avaliação a acurácia do modelo a matriz de confusão. As curvas de perdas e acurácia mostram um padrão típico de overfitting, e isto é visível ao verificar que a acurácia de treino cresce, enquanto a acurácia de validação se estabiliza e apresenta oscilações, o que indica que o modelo memorizar bem as imagens de treino, mas não generaliza para novas imagens.

A matriz de confusão ilustra que o modelo aprendeu parcialmente a distinguir imagens com câncer e normal. Isto é evidenciado ao analisar a quantidade de imagens que ele classificou corretamente, obteve mais acertos do que erros. Contudo, ele classificou um alto número de imagens de forma

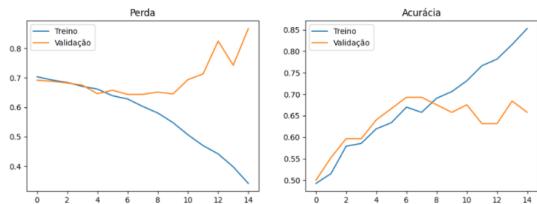


Fig. 1. Curvas de perda e acurácia obtidas durante o treinamento do modelo

incorrecta, são muitos falsos positivos ou negativos. Esses resultados indicam que o modelo não possui um desempenho satisfatório, ele não é viável para aplicações práticas ou para alcançar o objetivo estabelecido.

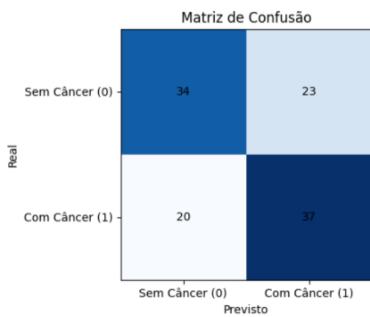


Fig. 2. Matriz de Confusão

Embora o desempenho desejado não tenha sido alcançado, foram realizadas diversas tentativas de treinamentos com o objetivo de melhorar a acurácia do modelo. Em cada nova execução, um parâmetro da arquitetura ou do processo de treinamento foi ajustado, e assim foram analisados os resultados obtidos.

G. Conclusão

O grupo conseguiu reunir um conjunto de imagens com alta resolução de mamografias, e encontrou uma base pública de imagens com cerca de 45 mil imagens. Devido a isto, o grupo conseguiu construir um dataset próprio, realizar a curadoria, padronizar as imagens e produzir anotações manuais que foram salvas em um arquivo csv.

Durante esse processo, alguns desafios foram solucionados, como a presença de imagens em formatos distintos (RGB ou Grayscale). Para resolver esta inconsistência, foi implementado um trecho de código capaz de identificar automaticamente o tipo de imagem e aplicar um padrão. Houve também o redimensionamento das imagens para 1024x1024 pixels, garantindo uniformidade.

Outra decisão importante foi a exclusão de casos extremamente difíceis de rotular, como lesões muito pequenas ou imagens em que não foi possível identificar se havia câncer ou não. E com isso, o grupo conseguiu construir um dataset mais consistente, com imagens nítidas e padronizadas, com um ground truth confiável, reduzindo significativamente as inconsistências entre os dados.

O objetivo principal do projeto, que seria alcançar uma acurácia mínima de 80% não foi atingido, pois a acurácia máxima alcançada foi de 69%. Entretanto, ao analisar outras tarefas desenvolvidas no projeto, pode-se ver que outros objetivos importantes foram cumpridos. Entre eles, a construção completa de um dataset do zero, com a curadoria e as anotações manuais. E o desenvolvimento do projeto proporcionou um entendimento sobre os parâmetros de uma rede neural, e sobre como diferentes ajustes influenciam no desempenho do modelo.

A principal limitação do projeto foi a dificuldade do modelo em generalizar para novos dados, mesmo após vários ajustes dos hiper parâmetros. Foram alterados diversos elementos da arquitetura e do treinamento, um de cada vez, mas o modelo não apresentou uma melhora.

Para trabalhos futuros, recomenda-se ampliar o conjunto de dados, aplicar técnicas mais avançadas de data augmentation e testar arquiteturas especializadas em detecção de regiões de interesse, como modelos da YOLO ou Faster R-CNN.

REFERENCES

- [1] INSTITUTO NACIONAL DE CÂNCER (INCA). A mulher e o câncer de mama no Brasil. Instituto Nacional de Câncer Jose Alencar Gomes da Silva – INCA, 2023. Disponível em: <https://www.gov.br/inca/pt-br/centrais-de-conteudo/exposicoes/a-mulher-e-o-cancer-de-mama-no-brasil>. Acesso em: 10 out. 2025.
- [2] INSTITUTO NACIONAL DE CÂNCER (INCA). Cancer de mama. Instituto Nacional de Cancer Jose Alencar Gomes da Silva – INCA, 2023. Disponível em: <https://www.gov.br/inca/pt-br/assuntos/cancer/tipos/mama>. Acesso em: 10 out. 2025.
- [3] STENKVIST, B.; WESTMAN-NAESER, S.; HOLMQUIST, J. NORDIN, B.; BENGTSSON, E.; VEGELIUS, J.; ERIKSSON, O.; FOX, C. H. Computerized nuclear morphometry as an objective method for characterizing human cancer cell populations. *Cancer Research*, v. 38, n. 12, p. 4688–4697, 1978.
- [4] OLIVEIRA, L. E. S. A machine learning approach for breast cancer diagnosis using digital mammography. In: International Joint Conference on Neural Networks (IJCNN), 2016. Disponível em: <https://www.inf.ufpr.br/lesoliveira/download/IJCNN2016-BC.pdf>. Acesso em: 10 out. 2025.
- [5] Arefan, D., et al. (2019). Deep learning modeling using normal mammograms for breast cancer risk prediction. *BMC Medical Informatics and Decision Making*. [URL não fornecida] Barros, M. E. S., Lins, R. C. (2024).
- [6] Avaliação de redes neurais deep learning para a classificação de Câncer de mama. *Cuadernos de Educación y Desarrollo*. [URL não fornecida] Blog Nucleorad. (2024).
- [7] A Revolução da Inteligência Artificial na Radiologia: Detecção de Microcalcificações em Mamografias. [URL não fornecida] Granado, T. C. (2015).
- [8] Transformação de intensidade para realce de contraste em mamografias digitais. Universidade Federal de Uberlândia. [URL não fornecida] Guerreiro, A. A. P., et al. (2024).
- [9] Integrando inteligência artificial à mamografia: uma abordagem complementar no diagnóstico do câncer de mama. *Revista Eletrônica Acervo Científico*. [URL não fornecida] Makeev, A., et al. (2023).
- [10] Task-based assessment of digital mammography microcalcification detection with deep learning denoising algorithms using in silico and physical phantom studies. *Physics in Medicine Biology*. [URL não fornecida] Pellacani, D. G. (2025).
- [11] Diagnóstico de mamografias utilizando múltiplas vistas e redes neurais convolucionais. Tese de Doutorado, Universidade de São Paulo. [URL não fornecida] Pellenz, A. E., et al. (2024).
- [12] Aplicação da inteligência artificial no diagnóstico e prognóstico do câncer de mama. *Brazilian Journal of Health Review*. [URL não fornecida] Santos, I. (2020).

- [13] Análise comparativa da influência dos espaços de cores na segmentação multi-classe de Whole Slide Imaging do câncer de mama utilizando deep learning. Anais Principais do Simpósio Brasileiro de Computação Aplicada. [URL não fornecida] Silva, M. E. S., Lins, R. C. (2024).
- [14] Avaliação de redes neurais deep learning para a classificação de Câncer de mama. Cuadernos de Educación y Desarrollo. [URL não fornecida]ng in silico and physical phantom studies. Physics in MedicineBiology. [URL não fornecida]
- [15] HE, Kaiming et al. Deep Residual Learning for Image Recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, EUA, p. 770-778, 2016.
- [16] SIMONYAN, Karen; ZISSERMAN, Andrew. Very Deep Convolutional Networks for Large-Scale Image Recognition. In: International Conference on Learning Representations (ICLR), San Diego, EUA, 2015.
- [17] TAN, Mingxing; LE, Quoc. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In: Proceedings of the 36th International Conference on Machine Learning (ICML), Long Beach, EUA, p. 6105-6114, 2019.
- [18] OZA, Parita; OZA, Rajiv; OZA, Urvi; SHARMA, Paawan; PATEL, Samir; KUMAR, Pankaj et al. Digital Mammography Dataset for Breast Cancer Diagnosis Research (DMID). Figshare, 2023. Disponível em: <https://doi.org/10.6084/m9.figshare.24522883.v2>. Acesso em: 30 set. 2025.
- [19] DELTAECHOV. RNSA Mamograph T512px 8bit. Kaggle, 2023. Disponível em: <https://www.kaggle.com/datasets/deltaechov/rnsamamographt512px8bit>. Acesso em: 24 set. 2025
- [20] SHEN, L.; MARGOLIES, L. R.; ROTHSTEIN, J. H.; FLUDER, E.; McBRIDE, R.; SIEH, W. Deep learning to improve breast cancer detection on screening mammography. Scientific Reports, v. 9, n. 12495, 2019. DOI: 10.1038/s41598-019-48995-4. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/31467326/>. Acesso em: 10 out. 2025.
- [21] HANIS, T. M.; ISLAM, M. A.; MUSA, K. I. Diagnostic accuracy of machine learning models on mammography in breast cancer classification: a meta-analysis. Diagnostics, v. 12, n. 1643, 2022. DOI: 10.3390/diagnostics12071643. Disponível em: https://www.mdpi.com/2075-4418/12/7/1643?utm_source=chatgpt.com. Acesso em: 10 out. 2025.