

APLICAÇÃO DE APRENDIZADO DE MÁQUINA NO AUXÍLIO A DETECÇÃO DO CÂNCER DE MAMA

¹Luiz Henrique Pinto Marques, ^{1,2}Filipe Loyola Lopes, ¹Maiara Lima Castilho

¹Laboratório de Bionanotecnologia, Universidade do Vale do Paraíba/Instituto de Pesquisa e Desenvolvimento, Avenida Shishima Hifumi, 2911, Urbanova - 12244-000 - São José dos Campos-SP, Brasil, luizhpmarques@gmail.com, mcastilho@univap.br.

²Centro de Tecnologia da Informação Renato Archer/Divisão de Metodologias da Computação, Rodovia Dom Pedro I (SP-65), Km 143,6 - Chácara Campos dos Amarais - 13069-901, Campinas-SP, flopes@cti.gov.br.

Resumo

O câncer de mama é um dos tumores mais comuns entre mulheres, causando grande impacto social e demandando estratégias eficazes de diagnóstico precoce. Apesar dos avanços em técnicas de imagem, como ultrassonografia e mamografia, desafios persistem na detecção inicial da doença. Diante disso, esse trabalho tem como objetivo o desenvolvimento de uma ferramenta baseada em redes neurais convolucionais para auxiliar na detecção precoce do câncer de mama, a partir de imagens de ultrassom mamário. O modelo foi treinado no ambiente Google Colab com Python 3.11, utilizando o banco de dados Breast Ultrasound Images Dataset, composto por 780 imagens de ultrassom categorizadas em classes benignas, malignas e normais, alcançando uma acurácia global de 92%, com destaque para um recall de 95% na identificação de casos malignos e precisão de 100% na classificação de tecidos saudáveis, validado pela matriz de confusão e curva ROC.

Palavras-chave: Câncer de mama. Ultrassonografia. Aprendizado profundo. Inteligência artificial.

Área do Conhecimento: Engenharias – Engenharia Biomédica

Introdução

O câncer de mama é uma das formas mais comuns de câncer entre mulheres em todo o mundo, representando uma significativa preocupação de saúde pública. Segundo o Instituto Nacional de Câncer (INCA), o controle do câncer de mama é uma prioridade no país, onde são estimados 73610 novos casos de 2023 a 2025, o que destaca a necessidade de estratégias eficazes de detecção precoce e tratamento (INCA, 2024). De acordo com a Organização Mundial da Saúde (OMS), estima-se que cerca de 2,3 milhões de novos casos de câncer de mama sejam diagnosticados anualmente, o que representa aproximadamente 11,7% de todos os casos de câncer (BRAY et al., 2018). O impacto emocional, físico e financeiro do câncer de mama é profundo, afetando não apenas os pacientes, mas também suas famílias e comunidades. Neste contexto, torna-se evidente a importância de um diagnóstico precoce e preciso como ferramenta essencial para a redução dos danos associados à doença (LUGLI et al., 2025).

A detecção precoce do câncer de mama é aprimorada por diferentes métodos de imagem. A mamografia, principal método, utiliza raios-X para identificar alterações nos tecidos mamários, como microcalcificações e nódulos. A ultrassonografia mamária complementa a mamografia, especialmente em mulheres com mamas densas, permitindo a visualização e a diferenciação de nódulos e cistos com clareza. A ressonância magnética oferece imagens detalhadas, sendo útil para avaliar casos complexos ou em mulheres com alto risco. A tomossíntese mamária captura imagens em múltiplas camadas, permitindo análise tridimensional do tecido e maior precisão do diagnóstico (ALMEIDA et al., 2023). Embora as técnicas permitam um diagnóstico preciso, ainda existe uma grande dificuldade na detecção deste tipo de tumor em sua fase inicial, no qual apresenta um diâmetro pequeno dificultando sua localização.

Nos últimos anos, a utilização de modelos de aprendizado de máquina (*machine learning*, ML) tem se destacado como uma abordagem inovadora para a detecção de diversas doenças, assim como no diagnóstico do câncer de mama (DALLAGASSA; DE OLIVEIRA, 2024). Esses modelos são capazes

A Ciência do NANO e seu impacto transformador no MACRO

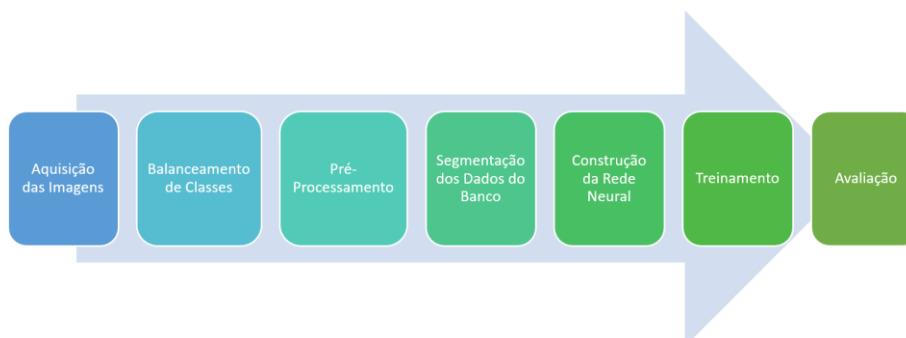
de analisar grandes volumes de dados médicos, como imagens de mamografias, com uma precisão que em alguns casos pode superar a análise humana (LEE; HONG; KIM, 2024). A capacidade de identificar padrões sutis em imagens que podem passar despercebidos a olho nu representa um avanço significativo na luta contra essa doença. Além disso, a implementação de algoritmos de aprendizado profundo (*deep learning*, DL) permitiria a personalização dos diagnósticos, ajustando-se a diferentes perfis de pacientes e melhorando as taxas de detecção (YOON; KIM, 2021, LEE; HONG; KIM, 2024).

A integração da inteligência artificial na prática clínica tem o potencial de reduzir o tempo de diagnóstico e aumentar a precisão na detecção do câncer de mama, especialmente ao analisar imagens de ultrassonografia mamária. Portanto, o objetivo deste trabalho é desenvolver uma ferramenta baseada em aprendizado de máquina para auxiliar na detecção do câncer de mama a partir de imagens de ultrassonografia, promovendo um impacto positivo no prognóstico do paciente.

Metodologia

A metodologia deste trabalho é composta por várias etapas interligadas, que vão desde a construção e busca do banco de dados, passando para o treinamento de modelos de aprendizado profundo baseado nos ultrassons mamários, gerando a avaliação deste modelo (Figura 1). Essas imagens são fundamentais para treinar os modelos de *machine learning*, e o pré-processamento é essencial para garantir que os dados estejam em um formato adequado, balanceando os dados, eliminando ruídos e normalizando as imagens para melhorar a precisão dos algoritmos.

Figura 1 – Metodologia Empregada



Fonte: O autor

O banco de dados utilizado neste trabalho foi o *Breast Ultrasound Images Dataset* (BUSI), disponível no Kaggle (<https://www.kaggle.com/datasets/sabahesaraki/breast-ultrasound-images-dataset>) (Tabela 1). Este conjunto contém um total de 780 imagens, todas com dimensões de 500 x 500 pixels, no formato PNG. As imagens estão categorizadas em três classes: 437 imagens de benigno (0), 210 maligno (1) e 133 do grupo normal (2). Devido ao desbalanceamento natural entre as classes, foi realizado um balanceamento por meio da técnica de *oversampling*, técnica que aumenta o número de amostras das classes minoritárias (maligna e normal) para igualar a quantidade da classe majoritária (benigna), resultando em aproximadamente 437 imagens por classe após o balanceamento e um total de 1311 imagens.

Tabela 1 – Banco de dados

Classe	Descrição	Amostras originais	Amostras com <i>Oversampling</i>
0	Benigno	437	437
1	Maligno	210	437
2	Normal	133	437

Fonte: O autor

A Ciência do NANO e seu impacto transformador no MACRO

A divisão dos dados seguiu a proporção de 80% para treino (1048 imagens), 10% para validação (131 imagens) e 10% para teste (132 imagens), realizada de forma estratificada para preservar a proporção original das classes em ambos os conjuntos. Além disso, foram aplicadas a normalização das imagens, dividindo os pixels da imagem por 255 e a diminuição do tamanho das dimensões das imagens de 500 X 500 pixels para 224 X 224 pixels, para que as redes neurais convolucionais (CNN) possam ter uma performance de treinamento maior e um custo computacional menor (GOODFELLOW et al., 2016). Para o treinamento dos modelos, foram empregadas CNN, utilizando o ambiente de desenvolvimento Google Colab com uma GPU T4 e o kernel Python 3.11. Essa configuração permite a execução eficiente dos algoritmos de aprendizado profundo, aproveitando os recursos computacionais disponíveis na nuvem.

Na sequência, a etapa de treinamento dos modelos de aprendizado profundo é realizada utilizando arquiteturas avançadas, como CNN. O modelo é alimentado com as imagens pré-processadas para aprender a identificar características associadas ao câncer de mama. O treinamento envolve a divisão dos dados em conjuntos de treinamento, validação e teste, permitindo que o modelo aprenda a generalizar suas previsões.

Por fim, para avaliar o desempenho dos modelos, foram utilizadas 3 métricas amplamente reconhecidas para ML, sendo elas a reportagem de classificação, matriz de confusão e a Curva Característica de Operação do Receptor (ROC). A reportagem de classificação é uma ferramenta analítica que sintetiza métricas-chave para avaliar modelos de classificação (SOKOLOVA; LAPALME, 2009). A matriz de confusão é uma ferramenta crucial na avaliação de modelos de classificação em aprendizado de máquina, pois organiza as previsões feitas pelo modelo em relação aos valores reais, permitindo identificar acertos e erros de forma detalhada (KOHÁVI; PROVOSTI, 1998). A curva ROC é uma ferramenta gráfica amplamente utilizada em ML e estatística para avaliar o desempenho de modelos de classificação binária, representando a relação entre a taxa de verdadeiros positivos (sensibilidade) e a taxa de falsos positivos para diferentes valores de limiar de decisão do modelo (FAWCETT, 2006; POWERS, 2011).

Resultados

O modelo de redes neurais convolucionais (CNN) desenvolvido para apoio à detecção de câncer de mama demonstrou boa eficácia em dados de teste, com 92% de acurácia global, indicando robustez na classificação de amostras benignas, malignas e normais (Tabela 2). Destaca-se o *recall* de 95%, ou seja, capacidade do modelo fazer previsões corretas para a classe maligna, fundamental para minimizar falsos negativos em diagnósticos críticos, aliado à precisão de 100% na identificação de tecidos saudáveis (classe 2).

Tabela 2 – Reportagem de classificação

Classe	Precisão	Sensibilidade	Medida F1	Suporte
Benigno (0)	97%	80%	88%	44
Maligno (1)	91%	95%	93%	44
Normal (2)	88%	100%	94%	44

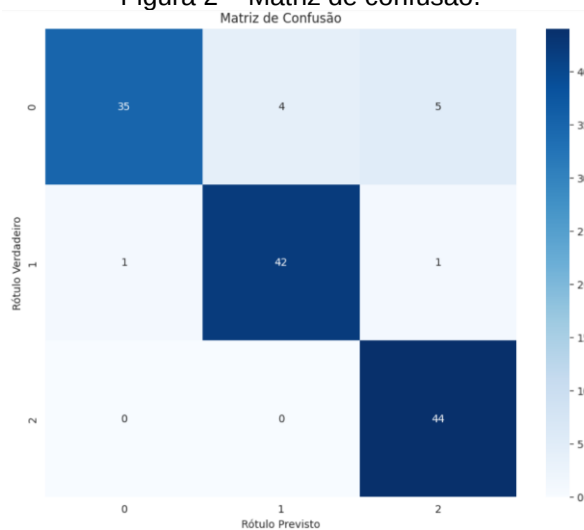
Fonte: O autor

A Figura 2 mostra a matriz de confusão, que é composta por 4 elementos principais, sendo eles, os verdadeiros positivos (VP), que são os casos corretamente classificados como positivos, verdadeiros negativos (VN), que representam os casos corretamente classificados como negativos, falsos positivos (FP), quando o modelo classifica incorretamente um negativo como positivo e falsos negativos (FN), quando um positivo é classificado erroneamente como negativo. A partir desses componentes, derivam-se métricas importantes que ajudam a interpretar a eficácia do modelo, como a acurácia, que mede a proporção total de acertos; a precisão, no qual indica a confiabilidade das previsões positivas; o *recall*, que avalia a capacidade do modelo em identificar todos os casos positivos; a sensibilidade, que é equivalente ao *recall* e mede a capacidade de detectar corretamente os positivos; a

A Ciência do NANO e seu impacto transformador no MACRO

especificidade, que avalia a capacidade do modelo em identificar corretamente os casos negativos e o *F1-score* (medida F1), que combina precisão e *recall* em uma única medida equilibrada. Essas métricas são essenciais para entender as forças e limitações do modelo e orientar melhorias em sua aplicação prática. O modelo obteve apenas 1 falso negativo em 44 casos malignos e 44 acertos em 44 amostras normais, evidenciando a capacidade em evitar erros (Figura 2).

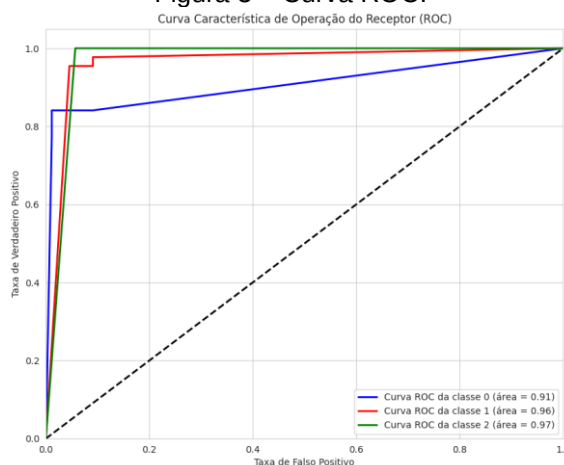
Figura 2 – Matriz de confusão.



Fonte: o autor

A curva ROC representa a relação entre a taxa de verdadeiros positivos (sensibilidade) e a taxa de falsos positivos para diferentes valores de limiar de decisão do modelo. Ao variar esse limiar, a curva ROC mostra como o modelo equilibra a capacidade de detectar corretamente os exemplos positivos e evitar classificações incorretas dos negativos. Uma curva que se aproxima do canto superior esquerdo do gráfico indica um classificador com bom desempenho, enquanto uma linha diagonal sugere um desempenho equivalente ao acaso. A área sob a curva ROC (AUC) é uma métrica que resume essa performance em um único valor, sendo que valores mais próximos de 1 indicam maior capacidade discriminativa, no qual o modelo obteve AUC para as classes 0, 1 e 2 os valores de 0,91; 0,96 e 0,97; respectivamente (Figura 3). Esses valores indicam um excelente desempenho do modelo, especialmente na classe 2, onde a taxa de verdadeiro positivo é alta, sugerindo que o modelo tem uma forte capacidade de discriminação entre as classes.

Figura 3 – Curva ROC.



Fonte: o autor

Discussão

A aplicação de CNN na detecção do câncer de mama demonstra um avanço significativo no auxílio a detecção precoce. O modelo alcançou uma acurácia global de 92% e um recall de 95% para casos malignos, indicando alta sensibilidade na identificação de tumores críticos. Esses resultados são particularmente relevantes em contextos clínicos, onde falsos negativos podem retardar tratamentos essenciais. Além disso, as métricas de avaliação utilizadas, reforçam o desempenho do modelo e já foram utilizados em outros trabalhos como CHO, G. e seus colaboradores, (2023), em que destacaram a eficácia de modelos como Random Forest e XGBoost na detecção precoce de surtos epidêmicos. RIZINDE e colaboradores (2023), evidenciaram a superioridade do Random Forest, comparado a outros modelos, em tarefas clínicas complexas, como a previsão de readmissões hospitalares, evidenciando a relevância de métricas como matriz de confusão, reportagem de classificação e curva ROC para avaliação de desempenho dos modelos.

Apesar do bom desempenho do modelo de CNN, o trabalho enfrenta limitações inerentes ao tamanho reduzido do banco de dados (780 imagens originais), mesmo após a aplicação do *oversampling*, o que permitiu a utilização de 1311 imagens no total. Embora a técnica tenha equilibrado as classes, a quantidade de dados ainda é limitada, necessitando de mais dados para melhorar a capacidade do modelo. Além disso, a precisão de 100% na classe normal, embora impressionante, requer cautela, pois superajustes (*overfitting*) em conjuntos pequenos podem inflar métricas sem refletir a realidade heterogênea dos casos clínicos.

O trabalho abre caminho para inovações futuras, como a integração de múltiplos bancos de dados com imagens de ultrassom mamário, ampliando a quantidade de imagens e a capacidade do modelo na detecção, o uso de técnicas de IA explicável poderiam aumentar a transparência e a confiança dos médicos no modelo e, consequentemente, uma estrutura de rede neural maior para melhorar a capacidade de detecção e integrar IA explicável.

Conclusão

O desenvolvimento desta ferramenta de apoio à detecção precoce de câncer de mama, baseada em CNN, demonstrou ser uma solução tecnologicamente viável e clinicamente relevante para aprimorar a triagem inicial da doença. Com 92% de acurácia global e um recall de 95% para casos malignos, o modelo mostrou-se robusto na identificação de padrões críticos em imagens de ultrassom, reduzindo significativamente falsos negativos e representando um avanço crucial para diagnósticos precoces.

Apesar das limitações inerentes ao tamanho do conjunto de dados (BUSI, com 780 imagens), a aplicação de técnicas de aumento de dados e regularização melhoraram a generalização do modelo, evidenciada pela precisão de 100% na classificação de tecidos saudáveis, pela matriz de confusão, que registrou apenas 1 falso negativo em 44 casos malignos e pela curva ROC que teve AUC de até 0,97.

Referências

- ALMEIDA, H. V. et al. Detecção de câncer de mama: avanços e desafios. **Research, Society and Development**, v. 12, n. 6, p. e9312642091, 2023. DOI: 10.33448/rsd-v12i6.42091.
- BRAY, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. **CA: A Cancer Journal for Clinicians**, v. 68, n. 6, p. 394-424, 2018. DOI: 10.3322/caac.21492.
- CHO, G. et al. Detection of COVID-19 epidemic outbreak using machine learning. **Frontiers in Public Health**, v. 11, artigo 1252357, 2023. DOI: 10.3389/fpubh.2023.1252357.
- DALLAGASSA, Marcelo Rosano; DE OLIVEIRA, Antonio Josenias Cordeiro. Uso de machine learning no diagnóstico de câncer de mama através de ultrassonografia. **Journal of Health Informatics**, Brasil, v. 16, n. Especial, 2024. DOI: 10.59681/2175-4411.v16.iEspecial.2024.1289.

A Ciência do NANO e seu impacto transformador no MACRO

FAWCETT, T. An Introduction to ROC Analysis. **Pattern Recognition Letters**, v. 27, n. 8, p. 861-874, 2006. DOI: 10.1016/j.patrec.2005.10.010.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. 1. ed. Cambridge: MIT Press, 2016. ISBN 978-0-262-03561-3.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. 2. ed. New York: Springer, 2009. ISBN 978-0-387-84857-0.

INSTITUTO NACIONAL DE CÂNCER (Brasil). **Controle do câncer de mama no Brasil: dados e números 2024**. Rio de Janeiro: INCA, 2024. 69 p. ISBN 978-65-88517-69-7. Disponível em: <https://www.inca.gov.br/publicacoes/livros/controle-do-cancer-de-mama-no-brasil-dados-e-numeros-2024>. Acesso em: 8 maio 2025.

KOHÁVI, R.; PROVOST, F. Glossary of Terms: Confusion Matrix. **Machine Learning**, v. 30, n. 2-3, p. 271-274, 1998. DOI: 10.1023/A:1017181826899.

LEE, S. E.; HONG, H.; KIM, E.-K. Diagnostic performance with and without artificial intelligence assistance in real-world screening mammography. **European Journal of Radiology Open**, v. 12, p. 100545, 2024. DOI: 10.1016/j.ejro.2023.100545.

LUGLI, Y. C. et al. Câncer de mama: impactos e desafios enfrentados. **Brazilian Journal of Health Review**, v. 8, n. 2, p. e79226, 2025. DOI: 10.34119/bjhrv8n2-353.

MAIONE, Camila. **Balanceamento de dados com base em oversampling em dados transformados**. 2020. 135 f. Tese (Doutorado em Ciência da Computação em Rede) – Universidade Federal de Goiás, Goiânia, 2020.

POWERS, D. M. W. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. **Journal of Machine Learning Technologies**, v. 2, n. 1, p. 37-63, 2011. DOI: 10.48550/arXiv.2010.16061.

RIZINDE, T.; NGARUYE, I.; CAHILL, N. D. Comparing Machine Learning Classifiers for Predicting Hospital Readmission of Heart Failure Patients in Rwanda. **Journal of Personalized Medicine**, v. 13, n. 9, p. 1393, 2023. DOI: 10.3390/jpm13091393.

SOKOLOVA, M.; LAPALME, G. A systematic analysis of performance measures for classification tasks. **Information Processing & Management**, v. 45, n. 4, p. 427-437, 2009. DOI: 10.1016/j.ipm.2009.03.002.

YOON, Jung Hyun; KIM, Eun-Kyung. Deep Learning-Based Artificial Intelligence for Mammography. **Korean Journal of Radiology**, v. 22, n. 8, p. 1225-1235, 2021. DOI: 10.3348/kjr.2020.1210.