

ÁGUEDA – UMA INTELIGÊNCIA ARTIFICIAL PARA DETECÇÃO PRECOCE DO CÂNCER DE MAMA

Wagner L. Cardozo¹

wagner.cardozo01@fatec.sp.gov.br
Faculdade de Tecnologia de Ourinhos

Rosemeiry C. Prado

rose.prado@fatecourinhos.edu.br
Faculdade de Tecnologia de Ourinhos

Robson P. Bonidia

bonidia@utfpr.edu.br
Universidade Tecnológica Federal do Paraná - UTFPR

1. Introdução

Com base nas estatísticas divulgadas pelo Instituto Nacional do Câncer - INCA, para o próximo triênio, abrangendo 2023 até 2025, é estimado um aumento de 10,5% nos novos casos de câncer de mama no Brasil, representando um risco estimado de 66,54 novos casos a cada 100 mil mulheres, totalizando 73.610 casos de câncer de mama [INCA, 2023].

Sem considerar os tumores de pele não melanoma, o câncer de mama feminina é o mais incidente no país e em todas as Regiões brasileiras. O maior risco estimado é observado na Região Sudeste, de 84,46 por 100 mil mulheres. No mundo, o câncer de mama é a principal causa global de incidência, com 11,7% do total de casos. As maiores taxas de incidência estimadas foram na América do Norte, na Oceania e nos países do Oeste da Europa [FERLAY, 2021; SUNG, 2021].

Consequentemente, profissionais da área da saúde estão buscando novas formas de detecção precoce do câncer, com destaque para a Inteligência Artificial (IA), que têm contribuído significativamente para aumentar a precisão dos diagnósticos, acelerar o processo de exames e possibilitar a prescrição adequada de terapias para tratamento de achados patológicos específicos [PAWAR, 2022; TAYLOR, 2023].

No entanto, a acessibilidade universal às soluções inovadoras tecnológicas formadas por IA é um tópico de crescente importância, especialmente para as classes sociais menos privilegiadas e para os países em desenvolvimento [RUBEIS, 2022]. No contexto da saúde, essa acessibilidade assume uma relevância crucial, pois influencia diretamente a eficácia das medidas de combate a doenças graves, como o câncer de mama [COCCIA, 2019].

Sendo assim, a presente pesquisa visa desenvolver e implementar uma solução de IA, chamada ÁGUEDA, que servirá como uma ferramenta de suporte para os médicos radiologistas e mastologistas durante o processo de tomada de decisões clínicas. O principal foco dessa IA é a capacidade de identificar e classificar, de forma precisa e acurada, achados patológicos nos quadrantes da estrutura anatômica da mama. Essa abordagem visa mitigar a ocorrência de falsos negativos ou positivos, combinando a especialização clínica dos profissionais médicos com a alta acurácia percentual da IA.

A integração da IA na prática clínica não substitui o julgamento dos médicos, mas o complementa, oferecendo percepções que enriquecem a avaliação profissional e ajudam na tomada de decisões mais informadas.

2. Metodologia

Para a construção da IA, foi adotado como método um classificador baseado em aprendizado estatístico, fundamentado na arquitetura de Rede Neural Convolutacional (CNN, do inglês Convolutional Neural Network). O ambiente de desenvolvimento integrado usado foi o Google Colab com recurso de processamento gráfico GPU modelo T4 da Nvidia. O algoritmo foi desenvolvido em linguagem de programação Python.

A arquitetura LeNet-5 foi aplicada na construção da CNN, por ser uma arquitetura robusta e eficiente para a classificação de imagens de mamografia, proporcionando resultados precisos e confiáveis [ZHANG, 2022]. O funcionamento da LeNet-5 é baseado em uma série de camadas convolucionais e de pooling, seguidas por camadas totalmente conectadas, que trabalham em conjunto para extrair e aprender

características discriminativas das imagens mamográficas [ZHANG,2022; ALASUBRAMANIAM, 2023].

Foi utilizado no treinamento, teste e validação da IA, um conjunto de imagens de mamografia rotulados, chamado Mini-DDSM (CBIS-DDSM, do inglês

Curated Breast Imaging Subset of Digital Database for Screening Mammography) e Ultrasound Breast Images for Breast Cancer [LEKAMLAGE, 2020; DHABYANI, 2020]. Este conjunto de imagens de mamografia é público e destinado à pesquisa científica, estando assim conforme a Lei Geral de Proteção de Dados.

Para visualizar e analisar o desempenho do modelo, foram empregadas as bibliotecas Matplotlib e Seaborn. Essas ferramentas permitem a criação de representações gráficas detalhadas e informativas, incluindo gráficos de perda, acurácia, curvas ROC e matrizes de confusão, cruciais para avaliar e interpretar os resultados do modelo.

Para implementar uma interface de usuário amigável e acessível em um ambiente de processamento em nuvem, a biblioteca nativa no ambiente de desenvolvimento integrado do Google, a Google Colab File, foi utilizada para criar um front-end interativo dentro do processamento em nuvem. Isso permite a interação com a IA de forma intuitiva, carregando imagens de mamografia, visualizando resultados e interpretando as previsões do modelo.

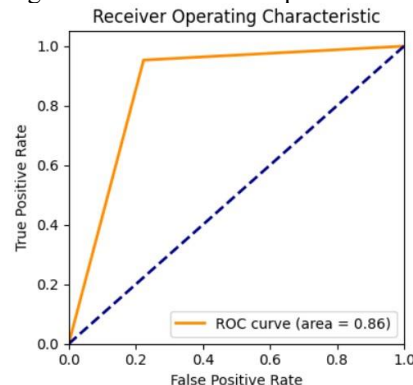
3. Resultados e Discussões

Na pesquisa de arquiteturas de CNNs, foi utilizado o modelo LeNet-5 para classificar imagens, obtendo o melhor resultado com uma acurácia de 92,33%, acurácia balanceada de 90,00%, precisão de 92,85%, F1 score de 89,65% e revocação de 86,66%. No entanto, no início do estudo, ao analisar o desempenho do treinamento, verificou-se um comportamento indesejável, após 20 épocas, em que a perda no treinamento começava a se afastar da perda de validação. Para resolver esse problema, uma abordagem mais avançada e complexa foi adotada.

Optou-se, então, por uma arquitetura mais profunda e robusta, que incluiu o aumento do número de camadas convolucionais e densas. A rede foi projetada com 32, 64 e 128 camadas convolucionais, tornando-a mais capaz de extrair características complexas das imagens,

representando respectivamente 20.187 e 15.279 imagens de mamografia e, para teste, 30% do conjunto de imagens com e sem câncer de mama, representando, respectivamente, 8.650 e 6.548 imagens. O estudo pelo método gráfico da curva ROC apresentou melhor desempenho, bem como o resultado numérico da área AUC sobre a curva ROC (Figura 1).

Figura 01 – Gráfico ROC para o desempenho da IA.



A matriz de confusão também permitiu observar uma robusta precisão na classificação para a generalização de novas imagens de mamografia (Figura 2). A nova arquitetura não apenas resolveu o problema de divergência entre as perdas de treinamento e validação, mas também aumentou a precisão do modelo. Finalmente, a interface da solução é apresentada na Figura 3.

Figura 02 – Representação da matriz de confusão para classificação das imagens de mamografia.

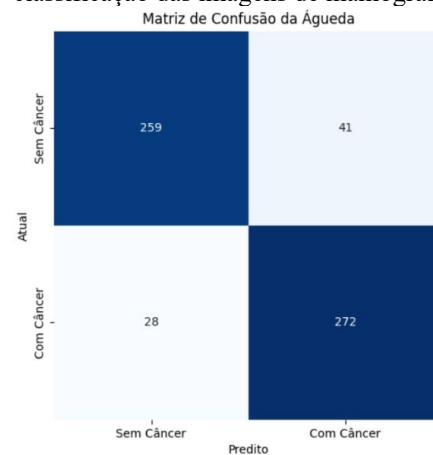
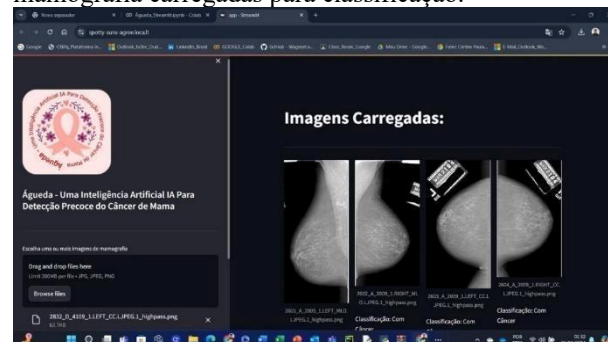


Figura 03 – Tela do ambiente da Águeda com imagens de mamografia carregadas para classificação.



4. Conclusões

O presente trabalho concluiu que é possível fazer uso de soluções inovadoras tecnológicas por meio de IA, que podem auxiliar as práticas profissionais dos médicos radiologistas e mastologistas, aumentando a agilidade na emissão dos laudos e a assertividade na prescrição. Além disso, o uso de soluções de IA acessíveis é essencial para garantir equidade na saúde, promovendo diagnósticos e tratamentos avançados para todas as pacientes, independentemente de origem ou condição socioeconômica. Finalmente, o presente trabalho tem o potencial de ampliar o acesso aos serviços de mamografia, especialmente para populações com menos acesso a tecnologias avançadas para o rastreamento e diagnóstico precoce do câncer de mama.

5. Reconhecimento e Impacto da Pesquisa

O presente trabalho conquistou o 1º lugar nacional, eleito o melhor projeto de Iniciação Científica do Brasil na área de concentração Ciências Exatas e da Terra, subárea Computação e Informática, na categoria "em andamento". A premiação ocorreu no 23º Congresso Nacional de

Iniciação Científica (2023), promovido e organizado pelo SEMESP, concorrendo com mais de 1.200 trabalhos de todo o país. Essa conquista proporcionou grandes oportunidades, incluindo a oferta de uma consultoria para transformar a ideia do projeto em um negócio viável. Esse reconhecimento reforça a relevância e o impacto positivo desta pesquisa na comunidade científica e no setor empresarial, abrindo portas para futuras aplicações práticas.

Referências



Agradecimentos

Agradecemos ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio fundamental a este trabalho, por meio da concessão da Bolsa PIBIT-CNPq.

¹ Aluno de IC com bolsa CNPq-PIBIT.