

CLASSIFICAÇÃO DE MASSAS MAMOGRÁFICAS A PARTIR DE REDES NEURAIS ARTIFICIAIS

Lizandra Régia¹
Alisson Guimarães²

RESUMO

O câncer de mama é a causa mais comum de morte por câncer na população feminina e um sério problema de saúde mundial. A mamografia permite uma detecção precoce do câncer, reduzindo a mortalidade da doença. Sua eficiência tornou-a procedimento padrão para diagnóstico do câncer de mama. Essas razões levaram ao desenvolvimento de sistemas computadorizados para o auxílio à detecção e ao diagnóstico. Para tanto, escolheu-se um modelo de Inteligência Artificial, as Redes Neurais Artificiais tipo regressão logística, fazem parte dos Sistemas Computacionais de Apoio ao Diagnóstico (Computer Aided Diagnosis - CAD), para auxiliar o profissional de saúde no prognóstico para essa anomalia genética. Deve-se ressaltar que o modelo classificatório não objetiva a substituição da opinião do profissional de saúde, mas sim disponibilizar uma opinião complementar como reforço. Portanto, a RNA, ao fim do projeto, será capaz de avaliar e discriminar, a partir dos dados inseridos na rede, massas mamográficas em benignas ou em malignas.

Palavras-chave: Câncer; RNA; classificação.

UNILAB, IED, Discente, lizandraregia@aluno.unilab.edu.br¹
UNILAB, IEDS, Docente, alisson@unilab.edu.br²

INTRODUÇÃO

O câncer de mama é um espectro de anormalidade que se manifesta nos lóbulos e ductos da mama. O sintoma mais comum do câncer de mama é o aparecimento de um nódulo, geralmente indolor, firme e irregular, mas existem tumores moles, flexíveis e bem definidos. A anomalia genética é diagnosticada, em primeiro momento, por meio do autoexame da mama e de exames laboratoriais, como a mamografia e a ressonância magnética.

A partir dos resultados dos exames, as imagens obtidas podem se encontrar dentro de um padrão de classificação, o Sistema de relatórios e dados de imagens de mama, no inglês, Breast Image Reporting and Data System (BI-RADS) elaborado pelo American College of Radiology desde 1992. Os BI-RADS são classificados de acordo com sua magnitude, variando de 0 a 6, considerando os parâmetros de tamanho, densidade, forma e limites das lesões. Dessa forma, ao avaliar individualmente cada caso das pacientes, quando se encontram situações onde o BI-RADS atinge magnitude a partir de 4, considera-se a necessidade de se realizar uma biópsia para avaliar a suspeita de câncer altamente sugestiva.

Mediante à informação, a categoria 4 engloba uma faixa ampla de probabilidade de achados malignos – entre 3% e 94%. Por sua vez, a categoria 5 reúne as lesões com 95% ou mais de chances de serem malignas. Na categoria 6 estão as ressonâncias de mama feitas por pacientes com câncer já diagnosticado. Entretanto, ao analisar o histórico de solicitações de mamografias e posteriores ressonâncias, constatou-se diversos casos onde a classificação do BI-RADS estava equivocada. Isso se torna preocupante à medida que as mulheres são submetidas à biópsia, método invasivo que pode trazer uma melhor avaliação do quadro clínico da paciente.

Nesta perspectiva, desenvolve-se uma Inteligência Artificial (IA), conhecida como Rede Neural Artificial (RNA) capaz de classificar, de acordo com os mesmos parâmetros para BI-RADS, a partir da base de dados do Repositório de Bases de Dados de Aprendizagem de Máquina da Universidade de Irvine, na Califórnia (UCI Repository of Machine Learning Databases, 2018). Portanto, a pesquisa objetiva fornecer uma ferramenta de fácil manuseio, funcionando, inclusive, como uma segunda opinião para o profissional de saúde encarregado acerca do diagnóstico sobre os exames. Desta forma, a base de dados irá atuar como treinamento da RNA, e, posteriormente, ela terá capacidade de classificar as massas mamárias entre malignas e benignas.

METODOLOGIA

- Revisão Bibliográfica sobre Redes Neurais Artificiais;
- Revisão Bibliográfica sobre Regressão Logística;
- Desenvolvimento de uma Rede Neural Artificial (RNA);
- Obtenção de dados a partir do repositório de aprendizagem de máquinas UCI Repository of Machine Learning Databases;
- Análise do comportamento da Inteligência Artificial perante às variáveis;
- Verificação e correção dos dados faltosos;
- Ambientação com o SCILAB, ferramenta computacional utilizada para o desenvolvimento do algoritmo;
- Coleta de dados a partir do Repositório de Bases de Dados de Aprendizagem de Máquina da Universidade de Irvine;
- Pré-processamento de dados: preenchimento de dados com falhas;

RESULTADOS E DISCUSSÃO

Para a realização e desenvolvimento da Rede Neural categórica, utilizou-se o tipo de RNA por Regressão Logística. Os dados obtiveram êxito em seu desempenho, atingindo uma boa porcentagem de aprendizagem na fase de teste, margem de 78% a 86% variando a semente fixada. Analisou-se a coleta de dados a partir da UCI Repository of Machine Learning Databases, Repositório de Bases de Dados de Aprendizagem de Máquinas.

Os dados foram obtidos a partir de 1998.

A modelagem escolhida para essa proposta foi a Regressão Logística, na qual consiste em suas variáveis categóricas via análise binária 0 ou 1, consistindo em 0 a resposta "não" do algoritmo e 1 a resposta "sim" do mesmo. É uma técnica que avalia a probabilidade de obtenção de uma das categorias da variável dependente, portanto, é capaz de obter a probabilidade de ocorrência de determinado evento, assim como a influência de cada variável independente no evento estudado.

A primeira situação analisa o comportamento das épocas de acordo com os 958 perfis de pacientes, consta-se que 172 destes perfis possuíam dados faltosos. Para a normalização dos dados, neste primeiro momento, utilizou-se metodologias advindas da Mineração de Dados, utilizando a média entre as categorias e, analiticamente, inserindo o padrão para preencher as lacunas. Percebe-se que o algoritmo se tornou eficaz à medida em que a curva converge constante e reduz o vetor gradiente da função. Para os dados completos, foi obtida a fase de teste de 82,20%, o melhor dos resultados encontrados nos parâmetros estabelecidos.

A segunda situação ilustra o mesmo comportamento apresentado na Figura 1, difere-se, portanto, na quantidade de dados levados para treinamento, onde foram levados para treinamento 786 dados, ou seja, excluiu-se os dados que continham lacunas. Percebe-se que o comportamento do gráfico foi similar ao da Figura 1, inclusive, a fase de teste obteve 82,17%. Dessa forma, prefere-se utilizar os dados do segundo teste, visto que apresentou comportamento similar e possui dados mais consistentes, com todos os dados perfeitos.

CONCLUSÕES

Dessa forma, a Rede Neural poderia apresentar um melhor desempenho na possibilidade de ter mais dados disponíveis, além disso, deve-se considerar a datação que foram coletados os respectivos perfis, visto que a tecnologia atual, 2022, diferentemente da adotada em 1998 poderia apresentar melhor verossimilhança e consistência nos dados. Outro fato que pode influenciar para aperfeiçoamento dos resultados são as subjetividades dos dados, visto que, por exemplo, cada perfil foi classificado como maligno ou benigno a partir dos dados obtidos à época.

AGRADECIMENTOS

Ao PIBIC-UNILAB pela oportunidade de participar da oportunidade dessa pesquisa como bolsista. Ao meu orientador Prof. Dr. Alisson Guimarães pela excelente condução no projeto, por todo suporte e apoio.

REFERÊNCIAS

D'Orsi CJ, Sickles EA, Mendelson EB, et al. ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System. 5th ed. Reston, VA: American College of Radiology; 2013.

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. Estimativa 2020: A incidência do

Câncer no Brasil. Rio de Janeiro: INCA, 2019. Disponível em: <https://www.inca.gov.br/sites/ufu.sti.inca.local/files//media/document//...> Acesso em: 12 agosto 2022.

CASTRO, Leandro Nunes de; FERRARI, Daniel Gomes. Introdução à Mineração de Dados. Rio de Janeiro: Saraiva, 2016. 682 p.

FERNANDES, Luiz Gustavo Leão. Utilização de Redes Neurais na Análise e Previsão de Séries Temporais. 1995. 77 f. Dissertação (Mestrado) - Curso de Ciências de Computação, UFRGS, Rio Grande do Sul, 1995.

MAXWELL. Redes Neurais Artificiais. 2013. 10 f. Disponível em: https://www.maxwell.vrac.puc-rio.br/16580/16580_4.PDF. Acesso em: 14 maio 2021.

SILVA, Ivan Nunes da; SPATTI, Danilo Hernane; FLAUZINO, Rogério Andrade. Redes Neurais Artificiais para engenharia e ciências aplicadas. 2. ed. Universidade de São Paulo: Artliber, 2016.