

ESTUDO SOBRE A UTILIZAÇÃO DE REDES NEURAIS ARTIFICIAIS MULTICAMADAS PARA A DETECÇÃO DE CÂNCER DE MAMA

STUDY ON USING MULTILAYER ARTIFICIAL NEURAL NETWORKS FOR BREAST CANCER DETECTION

Anna M. Uehara¹, Mariane A. Silva², Jefferson A. R. Passerini³

¹Faculdade de Tecnologia Professor José Camargo – Fatec Jales, anna.uehara@fatec.sp.gov.br

²Faculdade de Tecnologia Professor José Camargo – Fatec Jales, mariane.silva14@fatec.sp.gov.br

³Faculdade de Tecnologia Professor José Camargo – Fatec Jales, jefferson.passerini@fatec.sp.gov.br

Informação e Comunicação

Subárea: Matemática e Inteligência Computacional

RESUMO

O câncer de mama representa uma das principais causas de morbidade e mortalidade entre as mulheres em todo o mundo, sendo o diagnóstico precoce a principal estratégia para reduzir sua letalidade. Assim, a aplicação de modelos computacionais no auxílio aos profissionais médicos no diagnóstico da neoplasia de mama é um tema relevante para a sociedade. Neste estudo, realizou-se um levantamento bibliográfico de trabalhos científicos que abordaram redes neurais do tipo MLP (Multilayer Perceptron) e se aplicou uma rede neural multicamadas com variações na configuração da função de ativação e na arquitetura das camadas ocultas para auxiliar no diagnóstico do câncer de mama. Os testes realizados revelaram que a MLP com função de ativação hiperbólica tangente e uma arquitetura de duas camadas (com 100 neurônios na primeira camada e 50 neurônios na segunda camada) obteve o melhor desempenho nos testes de validação cruzada utilizando a base de dados WDBC, alcançando uma acurácia de 97,8884%. Esse desempenho demonstrou ser equivalente ao observado em outras publicações e está em consonância com estudos que apontam a função de ativação hiperbólica tangente como adequada para esse tipo de pesquisa. Portanto, os resultados deste estudo reforçam a importância da aplicação de redes neurais MLP na detecção precoce do câncer de mama e destacam a relevância da configuração específica adotada, que se mostrou eficaz na melhoria do diagnóstico da doença.

Palavras-chave: neoplasia; câncer de mama; rede neural perceptron multicamadas; inteligência artificial; aprendizado supervisionado.

ABSTRACT

Breast cancer is one of the leading causes of morbidity and mortality among women worldwide, and early diagnosis is the primary strategy to reduce its lethality. Therefore, the application of computational models to assist medical professionals in the diagnosis of breast neoplasm is a relevant topic for society. In this study, a literature review of scientific papers that addressed Multilayer Perceptron (MLP) neural networks was conducted, and a multilayer neural network with variations in activation function configuration and hidden layer architecture was applied to aid in the diagnosis of breast cancer. The conducted tests revealed that the MLP with a hyperbolic tangent activation function and a two-layer architecture (with 100 neurons in the first layer and 50 neurons in the second layer) achieved the best performance in cross-validation tests using the WDBC database, achieving an accuracy of 97.8884%. This performance was found to be equivalent to that observed in other publications and is consistent with studies indicating that the hyperbolic tangent activation function is suitable for this type of research. Therefore, the results of this study reinforce the importance of applying MLP neural networks in the early detection of breast cancer and highlight the relevance of the

adopted specific configuration, which proved to be effective in improving the diagnosis of the disease.

Keywords: neoplasm; breast cancer; multilayer perceptron neural network; artificial intelligence; supervised learning.

1 INTRODUÇÃO

O câncer de mama é uma das principais causas de morbidade e mortalidade entre as mulheres em todo o mundo. A detecção precoce desempenha um papel fundamental na melhoria da taxa de sobrevivência e na eficácia do tratamento. Exames de mamografia são amplamente reconhecidos como um dos métodos mais eficazes para o diagnóstico precoce deste tipo de câncer.

O câncer de mama é a neoplasia mais prevalente entre as mulheres em todo o mundo, com aproximadamente 2,3 milhões de novos casos estimados em 2020, representando cerca de 24,5% de todos os tipos de câncer diagnosticados nas mulheres. A incidência varia entre regiões, sendo mais elevada nos países desenvolvidos. No Brasil, em 2023, foram estimados 73.610 novos casos, com uma taxa de incidência de 66,54 por 100 mil mulheres, tornando-o o câncer mais comum no país. Além disso, o câncer de mama lidera as estatísticas de mortalidade por câncer entre as mulheres brasileiras, com uma taxa de mortalidade de 11,71 por 100 mil, especialmente nas regiões Sul e Sudeste (INCA, 2023b).

A pesquisa médica e de saúde desempenha um papel crucial na sobrevivência de todas as espécies, abrangendo investigações e informações relacionadas a diversas doenças, medicamentos, riscos e, especialmente, diagnóstico e tratamento. A integração e expansão da tecnologia no campo da medicina são de importância fundamental para aprimorar a capacidade e precisão no diagnóstico de doenças, análise de tendências e outros fatores relevantes, incluindo o tratamento. A World Health Organization (2023) relatou que, até 2020, 2,3 milhões de mulheres foram diagnosticadas com câncer de mama e houve 685 mil mortes relacionadas. Até o final de 2020, existiam 7,8 milhões de mulheres vivas que haviam sido diagnosticadas com a doença nos últimos 5 anos, tornando a moléstia mais prevalente no mundo (World Health Organization, 2023).

A World Health Organization (2023) ainda afirma que houve pouca variação na mortalidade entre os anos de 1930 e 1970, mas, a partir de 1990, com programas de detecção precoce em vários países, incluindo terapias médicas eficazes, houve aumento da perspectiva de sobrevida à doença.

A mamografia é o exame indicado para o rastreamento da neoplasia, sendo recomendado que mulheres, a partir dos 50 anos até os 69 anos, façam o exame a cada dois anos. A mamografia deve estar sempre associada ao exame clínico das mamas, mas a confirmação somente é feita por meio do exame histopatológico (biópsia), que analisa uma pequena porção de tecido mamário (INCA, 2023a).

No entanto, a análise minuciosa de mamografias é uma tarefa desafiadora que requer um alto grau de especialização e experiência por parte dos radiologistas. Nesse contexto, a inteligência artificial, em particular as Redes Neurais Artificiais (MLP - Multilayer Perceptron), emergem como uma ferramenta promissora para aprimorar a detecção e a classificação de anomalias em mamografias, oferecendo a possibilidade de uma abordagem mais precisa, rápida e escalável (Ting; Sim, 2017; Cedeño; Quintanilla; Andina, 2011; Isa *et al.*, 2010).

Esse artigo científico explora o potencial das redes neurais MLP na detecção de câncer de mamografias, apresentando uma revisão da literatura, metodologias de implementação e resultados obtidos, com ênfase nos benefícios e nos desafios associados à aplicação dessa tecnologia no contexto da saúde da mulher.

2 REFERENCIAL TEÓRICO

O câncer de mama que pode apresentar uma variedade de sintomas, especialmente em estágios mais avançados da enfermidade. No entanto, é importante notar que a maioria das pessoas não manifesta sintomas, quando a doença ainda está em estágios iniciais. Os sintomas mais comuns incluem a presença de um nódulo ou espessamento na mama, frequentemente sem causar dor. Além disso, mudanças no tamanho, forma ou aparência da mama podem ser indicativos de uma condição preocupante (World Health Organization, 2023).

Outros sinais a serem observados são a retração da pele, vermelhidão, covinhas ou outras alterações na pele da mama, bem como mudanças na aparência do mamilo ou da pele ao seu redor, conhecida como aréola. Além disso, a presença de secreção anormal ou sanguinolenta no mamilo também pode ser um sintoma preocupante. Mesmo quando um nódulo na mama não causa dor, é de suma importância procurar cuidados médicos para uma avaliação adequada (World Health Organization, 2023).

Vale ressaltar que a detecção precoce desempenha um papel crucial para o tratamento eficaz. Tumores pequenos, que não se espalharam para os gânglios linfáticos próximos, têm maiores chances de serem tratados com sucesso. No entanto, o câncer de mama pode se disseminar para outras áreas do corpo ao longo do tempo, frequentemente afetando os gânglios linfáticos debaixo do braço como primeiro local de disseminação. À medida que as células cancerosas se espalham para outros órgãos, como pulmões, fígado, cérebro e ossos, novos sintomas relacionados ao câncer, como dor óssea ou dores de cabeça, podem surgir. Desta forma, a compreensão desses sintomas e a busca de atendimento médico adequado desempenham um papel fundamental na detecção precoce e no tratamento exitoso do câncer de mama (World Health Organization, 2023).

A rede neural perceptron de múltiplas Camadas (MLP) é uma rede neural feed-forward de retropropagação, que é a técnica de rede neural mais frequentemente utilizada em reconhecimento de padrões (Bishop, 1999; Duda, Hart, 1973).

As MLPs são classificadores de aprendizado supervisionado que consistem em uma camada de entrada, uma camada de saída e uma ou mais camadas ocultas que extraem informações úteis durante o aprendizado e atribuem coeficientes de ponderação modificáveis aos componentes das camadas de entrada. Na primeira etapa, os pesos atribuídos às unidades de entrada e aos nós nas camadas ocultas e entre os nós na camada oculta e a saída determinam a saída. Essa saída é comparada com a saída desejada. Um sinal de erro é então retropropagado, e os pesos de conexão são ajustados correspondentemente (Kathija; Nisha; Sathik, 2017).

Em um estudo realizado por Ayer, Chen e Burnside (2013), foi discutido o uso de redes neurais artificiais (ANNs) de três camadas para analisar 256 imagens de mamografias e classificá-las. Foram utilizadas amostras de microcalcificações em intervalos de tamanho, que variavam de 50 a 250 μm , 100 a 500 μm , 200 a 1.000 μm e 400 a 2.000 μm para a formulação da rede. A classificação com base no tamanho das malignidades foi realizada por meio de quatro ANNs. A rede neural apresentou uma sensibilidade de 84% a uma especificidade de 75%.

Ting e Sim (2017) desenvolveram uma rede neural multicamadas para classificar câncer de mama, com base em 170 mamografias previamente testadas para malignidades. A rede classificou com sucesso as malignidades em categorias de benigno, maligno e normal, alcançando uma precisão de 90,59% e uma sensibilidade de 90,53% quando comparada com os resultados conhecidos. Isso demonstra a eficácia do modelo na assistência aos examinadores médicos na detecção de câncer de mama.

Em um estudo conduzido por Fogel, Wasson e Boughton (1995), foram comparadas duas configurações de redes neurais multilayer perceptron (ML-NN) para o diagnóstico de câncer de mama. Eles realizaram dois experimentos, utilizando conjuntos de imagens de mamografias de pacientes para treinamento e teste das redes. Uma configuração (9-2-1) obteve uma taxa de

corretude de até 98,6%, enquanto a outra (9-9-1) alcançou até 97,9% de corretude. Concluíram que o ML-NN 9-2-1 apresenta maior precisão e é uma configuração otimizada para essa finalidade.

Floyd *et al.* (1994) descrevem uma abordagem envolvendo redes neurais multilayer perceptron (ML-NN) foi utilizada para prever a malignidade do câncer de mama. Foram empregados oito parâmetros de mamografias mamárias como entradas para a ML-NN, incluindo tamanho do tumor, distorção assimétrica, margem do tumor e outros. A rede foi treinada e testada com 260 casos, classificando a saída em "benigno" ou "maligno" para biópsia. O modelo alcançou um desempenho notável, identificando 99 dos 168 casos benignos e 92 dos 92 casos malignos, destacando a utilidade da ML-NN na predição da malignidade do câncer de mama.

Isa *et al.* (2010) avaliaram diferentes funções de ativação em uma rede neural multilayer (MLP) para a detecção e classificação de células mamárias. O uso da função de ativação hiperbólica tangente resultou na maior precisão, atingindo 97,0% na detecção e classificação de células mamárias, superando outras funções como a sigmoide, senoidal, exponencial e logarítmica neuronal. Portanto, a função hiperbólica tangente foi identificada como a opção mais eficaz e adequada para ser empregada em MLPs em aplicações de detecção de câncer de mama.

Meinel *et al.* (2007) conduziu um estudo por meio de uma rede neural de retropropagação foi empregada para criar um sistema de Diagnóstico Auxiliado por Computador (CAD) que aprimora imagens de Ressonância Magnética das mamas. O sistema foi treinado com 80 lesões, das quais 43 eram malignas e 37 benignas, utilizando 13 características de entrada selecionadas dentre 42 possíveis. Os resultados demonstraram que os sistemas CAD melhoram a qualidade das imagens de Ressonância Magnética, facilitando a detecção eficiente de distúrbios mamários.

Cedeño, Quintanilla e Andina (2011) propuseram uma rede neural chamada "Artificial Metaplasticity Multi-Layer Perceptron" (AMMLP) com o objetivo de detectar malignidades em imagens de mama. Utilizando a metaplasticidade artificial para minimizar erros, a rede foi treinada com 410 amostras, das quais 144 eram malignas e 266 benignas, e testada com 233 amostras, consistindo em 95 casos malignos e 178 benignos. Os resultados revelaram uma notável precisão de 99,26%, uma especificidade de 97,89% e uma sensibilidade de 100%, superando os métodos convencionais de CAD que utilizam redes neurais de retropropagação. Isso destaca a eficácia do AMMLP na detecção de malignidades mamárias. Os autores utilizaram a base de dados Winsconsin Diagnostic Breast Cancer - WDBC (Wolberg *et al.*, 1995).

Kathija, Nisha e Sathik (2017) conduziram um estudo utilizando a base de dados WDBC utilizando 10% das amostras para teste e 90%, utilizando validação cruzada para treinamento de uma rede neural multicamadas (MLP), obteve como resultado uma acurácia de 98,99%.

A abordagem de classificação do diagnóstico de exames para detecção de câncer de mama utilizando redes neurais multicamadas é bastante prevalente na literatura, atualmente vem sendo substituída por redes neurais profundas (*deep learning*), como as redes neurais convolucionais – CNN do inglês: *Convolutional Neural Network*, que são mais complexas e assim geram um treinamento mais custoso.

Durante o processo de avaliação de um classificador, é fundamental compreender a capacidade de generalização do teste a ser realizado, obtida por meio da validação cruzada dos dados analisados. A validação cruzada é aplicada com o propósito de estimar a precisão do modelo para um novo conjunto de dados, sendo sua abordagem principal a divisão do conjunto de dados em subconjuntos mutuamente exclusivos, nos quais alguns são utilizados para treinamento e outros para validação ou teste. A divisão dos dados pode ser realizada utilizando

as abordagens holdout, leave-one-out ou k-fold, sendo esta última a mais comumente empregada (Kohavi, 1995; Laureano; Caetano; Cortez, 2014).

A validação cruzada K-fold utiliza múltiplas divisões dos dados em conjuntos de treinamento e teste. Neste método, o classificador utiliza um desses conjuntos para teste e os demais para treinamento, repetindo o processo para cada conjunto. Após várias iterações, a média da taxa de acerto do classificador pode ser calculada usando métricas de desempenho, como a acurácia. Essa técnica é aplicável a diversos classificadores, como Support Vector Machines (SVM), Decision Tree, Rotation Forest, entre outros (Hornung *et al.*, 2014; Triba *et al.*, 2015).

As métricas de avaliação adotadas nas tarefas de classificação têm papel fundamental na obtenção dos resultados da otimização de um classificador. Assim a seleção de métricas de avaliação adequadas é uma chave importante para discriminar e obter o classificador ideal (Hossin; Sulaiman, 2015). Métricas são aplicadas na área de processamento de imagem, identificação de padrões e classificação, essas métricas são extraídas a partir de uma rotulagem dos dados investigados.

Para o problema de classificação binária, a avaliação de uma melhor solução para uma determinada classificação pode ser definida com base na matriz de confusão, a partir do qual rótulos são atribuídos para as amostras classificadas, em que:

- Verdadeiro positivo (*vp*): se uma instância é positiva e é classificada como positiva;
- Verdadeiro negativo (*vn*): se uma instância é negativa e é classificada como negativa;
- Falso positivo (*fp*): se uma instância é negativa e é classificada como positiva;
- Falso negativo (*fn*): se uma instância é positiva e é classificada como negativa.

A partir da matriz de confusão podemos definir indicadores de desempenho como a acurácia que mede a proporção de previsões corretas sobre o total de instâncias avaliadas e é definida pela fórmula $acc = \frac{vp+vn}{vp+fp+vn+fn}$; a sensibilidade mede a fração de padrões positivos que são classificados corretamente e é dada pela equação $acc = \frac{vp}{vp+fn}$; a especificidade mede a fração de padrões negativos que são classificados corretamente e é determinada pela equação $acc = \frac{vn}{vn+fp}$; e a precisão mede a relação entre as previsões positivas realizadas corretamente e todas as previsões positivas (incluindo as falsas), sendo definida pela fórmula $acc = \frac{vp}{vp+fp}$ (Hossin; Sulaiman, 2015).

3 METODOLOGIA

Para elaboração deste trabalho, foram levantados trabalhos científicos que envolveram a aplicação de redes neurais multicamadas em aplicações de detecção de câncer de mama sem limite de tempo. Outra abordagem deste trabalho é a aplicação prática da MLP na base de dados WDBC (Wolberg *et al.*, 1995).

A base de dados WDBC é composta por 569, em que se observa uma distribuição de classes: 357 benignas, 212 malignas. Cada amostra é composta por 32 características, das quais a primeira característica é o número de identificação da amostra o qual foi descartada para efeito de classificação, e a segunda característica é o campo de diagnóstico composto pelas letras M, para casos malignos e B para casos benignos; este campo será o dado predictor.

As características de 2-32 compreendem 10 informações que representam valores reais calculadas para cada núcleo celular:

- a) raio (média das distâncias do centro para os pontos na periferia)
- b) textura (desvio padrão dos valores em escala de cinza)
- c) perímetro
- d) área

- e) suavidade (variação local nos comprimentos dos raios)
- f) compacidade ($\text{perímetro}^2 / \text{área} - 1,0$)
- g) concavidade (gravidade das porções côncavas do contorno)
- h) pontos côncavos (número de porções côncavas do contorno)
- i) simetria
- j) dimensão fractal ("aproximação da linha costeira" - 1).

Para cada uma das 10 características são apresentados a média, o erro padrão e o "pior" ou maior (média dos três maiores valores) dessas características. Esses cálculos foram realizados para cada imagem, resultando em 30 características. Por exemplo, o campo 3 representa o raio (médio), o campo 13 representa o raio (desvio padrão), e o campo 23 representa o raio (pior).

Para a implementação do modelo, foi empregada a linguagem Python, fazendo uso das bibliotecas Numpy (2023), Matplotlib (2023), Pandas (2023) e Scikit-learn (2023). Para a condução dos testes, foi utilizado o ambiente do Google Colab, que possibilita a criação de um ambiente virtualizado de testes na infraestrutura em nuvem da empresa Google.

A partir da base de dados, as características foram separadas em dois conjuntos distintos: os dados previsores (para diagnóstico) e os classificadores (correspondentes aos atributos dos campos de 3 a 32 da base de dados). Em seguida, foram aplicados procedimentos de pré-processamento aos dados previsores e classificadores.

Os dados previsores, originalmente apresentados em formato alfanumérico (M ou B), foram convertidos em valores numéricos. Para essa finalidade, utilizou-se a classe LabelEncoder, disponível na biblioteca Scikit-learn (2023), a fim de efetuar a transformação dos dados, atribuindo o valor 1 (um) à representação "M" e o valor 0 (zero) à representação "B".

No contexto dos dados classificatórios, é evidente a presença de significativas disparidades nas magnitudes das informações apresentadas. Por exemplo, o campo de característica 3 apresenta uma gama de valores variando desde o mínimo de 6,9810 até o máximo de 28,1100, enquanto o campo de característica 6 possui um valor mínimo de 143,5000 e um valor máximo de 2501,0000. A considerável disparidade nas faixas de valores entre as características da base de dados não favorece o desempenho dos classificadores.

Portanto, a padronização de um conjunto de dados se torna um requisito essencial para muitos classificadores, uma vez que seu desempenho pode ser comprometido, caso as características individuais não sigam aproximadamente uma distribuição normal. Para alcançar esse objetivo, foi empregada a classe "StandardScaler" disponível na biblioteca Scikit-learn (2023), a qual realiza a padronização dos valores das características presentes na base de dados. Essa padronização é calculada por meio de um escore, que é obtido a partir da relação entre a média da característica e o seu desvio padrão, sendo particularmente apropriada quando há a presença de valores atípicos nos dados (Scikit-learn, 2023).

Após a padronização das características, o modelo emprega a técnica de validação cruzada (*cross-validation*) para criar 10 diferentes conjuntos de treinamento e teste. Em cada combinação de dados, a proporção é mantida em 10% para teste e 90% para treinamento. Além disso, o parâmetro "random_state" foi configurado de modo que a seleção das amostras em cada grupo ocorra de maneira repetitiva a cada execução do modelo, garantindo a consistência das análises nas diversas configurações do algoritmo.

Dentro de cada "split" da validação cruzada, o classificador MLP é treinado usando os dados de treinamento, seguido pelo teste do modelo para determinar a acurácia do "split". Ao final, é calculada a média das acurácias nas diferentes execuções.

Para os testes de execução da MLP, foram estabelecidos os parâmetros de parada no treinamento da rede neural, com "max_iter" definido como 2000, o que determina o número

máximo de épocas de treinamento, e "tol" configurado em 0.0000010, estipulando a tolerância mínima para a melhoria no erro acumulado durante o processo de treinamento da rede. É importante ressaltar que os parâmetros "max_iter" e "tol" foram aplicados em todas as configurações investigadas no estudo.

Os parâmetros "activation" foram definidos para especificar a função de ativação dos neurônios nas camadas ocultas da rede neural, com as opções relu (padrão), tanh e logística. Além disso, a arquitetura das camadas ocultas da rede neural foi configurada através do parâmetro "hidden_layer_sizes". Nesse contexto, foram exploradas as seguintes opções: a configuração padrão, que consiste em uma camada oculta com 100 neurônios; uma configuração com duas camadas ocultas, a primeira com 100 neurônios e a segunda com 50 neurônios; e, por fim, uma terceira configuração com três camadas ocultas, nas quais tínhamos 100 neurônios na primeira camada, 100 na segunda camada e 50 na terceira camada da rede.

Deste modo, tem-se as seguintes configurações de testes identificadas de modo a facilitar a apresentação dos resultados (Quadro 1).

Quadro 1 – Configuração de execução das redes neurais (MLP)

Id	“Activation” (função de ativação)	“Hidden_layer_sizes” (Arquitetura da rede)
A	Relu	100 neurônios (1 camada)
B	Tanh	100 neurônios (1 camada)
C	Logistic	100 neurônios (1 camada)
D	Relu	100 neurônios camada 1; 50 neurônios camada 2
E	Tanh	100 neurônios camada 1; 50 neurônios camada 2
F	Logistic	100 neurônios camada 1; 50 neurônios camada 2
G	Relu	100 neurônios camada 1; 100 neurônios camada 2 e 50 neurônios camada 3
H	Tanh	100 neurônios camada 1; 100 neurônios camada 2 e 50 neurônios camada 3
I	Logistic	100 neurônios camada 1; 100 neurônios camada 2 e 50 neurônios camada 3

Fonte: Elaborado pelos autores.

4 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Os resultados deste trabalho acadêmico se dividem na revisão da bibliografia e nos testes de modelo realizados. O levantamento bibliográfico pode ser resumido no Quadro 2.

Quadro 2 – Revisão Bibliográfica de trabalhos científicos

Autor	Descrição	Base de Dados	Resultados
Ayer, Chen e Burnside (2013)	Analizou 256 imagens de mamografias através de 4 redes neurais	Não definido	Sensibilidade de 84% e especificidade de 75%
Ting e Sim (2017)	Analizou 170 imagens de mamografias previamente testas para malignidades pela rede neural	Não definido	Precisão de 90,59% e sensibilidade de 90,53%
Fogel, Wasson e Boughton (1995)	Testou redes MLP com arquiteturas (9-2-1), (9-9-1) com imagens de mamografias	Não definido	Acurácia de 98,6% (9-2-1) e 97,9% (9-9-1)
Floyd <i>et al.</i> (1994)	Utilizou MLP em imagens de mamográficas com 260 amostras onde 168 casos eram benignos e 92 malignos	Não definido	Acertou 99 dos 168 casos benignos e 92 dos 92 casos malignos
Isa <i>et al.</i> (2010)	Avaliou diferentes funções de ativação em uma MLP.	Não definido	A função de ativação hiperbólica tangente obteve precisão de 97%
Meinel <i>et al.</i> (2007)	Testou uma rede neural com retropropagação em uma base de 80 imagens de ressonância magnética das mamas onde tem-se 43 amostras malignas e 37 benignas.	Não definido	Curva ROC – 90,7%
Cedeño, Quintanilla e Andina (2011)	Propuseram uma “ <i>Artificial metaplasticity Multi-Layer Perceptron – AMMLP</i> ” para detectar malignidades em imagens de mama	WDBC	Obteve precisão de 99,26%, especificidade de 97,89% e sensibilidade de 100%
Kathija, Nisha e Sathik (2017)	Treinou uma rede neural MLP com distribuição de 10% das amostras para testes e 90% para treinamento utilizando validação cruzada	WDBC	Obteve acurácia de 98,99%

Fonte: Elaborado pelos autores.

Como se observa, o trabalho de Cedeño, Quintanilla e Andina (2011) e Kathija, Nisha e Sathik (2017) utilizaram a mesma base de dados – WDBC (Wolberg *et al.*, 1995), muitos dos trabalhos pesquisados não informavam as bases de dados utilizadas, e foram selecionados que utilizavam redes neurais do tipo MLP, trabalhos mais recentes utilizam redes neurais profundas (deep learning) como as redes neurais convolucionais. Redes neurais profundas exigem mais capacidade de processamento para serem treinadas o que inviabilizaria a utilização do Google Colab.

O modelo desenvolvido neste trabalho foi testado no Google Colab com a base de dados WDBC (Wolberg *et al.*, 1995), de acordo com as configurações definidas na metodologia (Quadro 1). Cada uma das configurações propostas foi testada em 10 combinações diferentes de dados de teste e treinamento, e a média da acurácia das execuções foi calculada. Os resultados dos testes realizados podem ser visualizados na Tabela 1.

Tabela 1 – Resultado – Testes MLP base de dados WDBC

Identificação	Acurácia	Especificidade	Sensibilidade	Precisão
A	97,1835%	98,0172%	95,8440%	96,6330%
B	97,5344%	98,3581%	96,4953%	96,9343%
C	97,5344%	98,6706%	96,1107%	96,3343%
D	97,3621%	98,6075%	95,4594%	97,5164%
E	97,8884%	98,9484%	96,5107%	97,8343%
F	96,8295%	97,6042%	96,1260%	95,3694%
G	97,0050%	98,3531%	95,1095%	97,0741%
H	97,1867%	98,3417%	95,7260%	96,9996%
I	96,4786%	98,0640%	94,2705%	96,4566%

Fonte: Elaborado pelos autores.

Observando-se a Tabela 1, pode-se verificar que as amostras G, H e I que são redes neurais com três camadas ocultas não performaram melhor do que as configurações com uma camada e duas camadas. Nas configurações de uma camada e de duas camadas, observam-se os melhores resultados de acurácia, amostras E (97,8884%), B (97,5344%) e C (97,5344%); de especificidade amostras E (98,9484%), C (98,6706%) e D (98,6075%); de sensibilidade com as amostras E (96,5107%), F (96,1260%) e B (96,4953%).

Em relação à precisão, observaram-se as amostras com melhor desempenho, sendo a amostra E (97,8343%) com 2 camadas, a amostra D (97,5164%) com duas camadas e a amostra G (97,0741%) com três camadas. A precisão é o único quesito avaliado no qual uma MLP de três camadas está entre as melhores avaliadas.

Quando se observa pelo aspecto da função de ativação em relação a acurácia, temos a função “relu” nas amostras A, D e G em que a melhor performance foi na amostra D (97,3621%) que é uma rede neural de 2 camadas; na função de ativação “tanh” temos as amostras B, E e H onde a melhor amostra E (97,8884%) com duas camadas; com a função de ativação “logistic” temos as amostras C, F e I sendo a melhor a amostra C (97,5344%) com uma camada.

Nos testes realizados, a amostra E, configurada com a função de ativação "tanh" e duas camadas em sua arquitetura (uma com 100 neurônios e outra com 50 neurônios), apresentou os melhores resultados em termos de avaliação, dentre todas as configurações estudadas. Este desempenho se alinha com as descobertas de Isa *et al.* (2010), que, em seu estudo, investigou várias funções de ativação em uma MLP e concluiu que a função de ativação hiperbólica tangente, a qual retorna valores dentro do intervalo de -1 a 1, proporcionou o melhor desempenho. Essas conclusões corroboram com os resultados observados no presente trabalho.

Em relação aos trabalhos correlatos pesquisados que realizaram testes com a mesma base de dados, temos que Kathija, Nisha e Sathik (2017) obteve uma acurácia de 98,99% comparativamente nos testes realizados neste trabalho temos a amostra E com uma acurácia de 97,8884%, o que se pode considerar dentro do desvio padrão entre as análises.

Quando se compara com Cedeño, Quintanilla e Andina (2011) que propôs uma MLP com modificações, temos: precisão de 99,26%, especificidade de 97,89% e sensibilidade de 100%; em relação a amostra E temos uma precisão 97,8343%, especificidade de 98,9484% e sensibilidade de 96,5107%. Assim os resultados, apesar de aplicarem-se apenas configurações básicas na MLP, continuam próximos aos trabalhos correlatos estudados.

5 CONSIDERAÇÕES FINAIS

Como observado, o câncer de mama é a neoplasia mais prevalente em todo o mundo, e a cura da doença depende, em grande parte, do diagnóstico precoce. Portanto, a aplicação de modelos computacionais para auxiliar os profissionais médicos é de extrema relevância para a sociedade.

Neste trabalho, foram revisados estudos relacionados e implementados modelos de Redes Neurais Perceptron Multicamadas (MLP) com diversas configurações. Por meio de testes

conduzidos com a base de dados WDBC, empregando a técnica de validação cruzada, identificou-se que, no contexto abordado, a MLP com a função de ativação hiperbólica tangente (tanh) e uma arquitetura composta por duas camadas ocultas (com 100 neurônios na primeira camada e 50 neurônios na segunda camada) obteve o melhor desempenho, com uma acurácia média de 97,8884%. Esses resultados estão em consonância com descobertas em estudos correlatos e reforçam a eficácia da função de ativação "tanh" nesse tipo de análise.

As configurações propostas foram definidas de forma empírica, o que limita a obtenção de correlações mais substanciais. Logo, é imperativo aprofundar a pesquisa bibliográfica em busca de elementos que possam guiar a definição de configurações de arquitetura mais aprimoradas para as redes neurais. Além disso, é aconselhável expandir o estudo para incluir a aplicação de redes neurais profundas, que apresentam um desempenho superior em relação às redes neurais tradicionais, embora demandem recursos computacionais mais significativos.

REFERÊNCIAS

AYER, T.; CHEN, O.; BURNSIDE, E. S. Artificial neural networks in mammography interpretation and diagnostic decision making. **Computational and Mathematical Methods in Medicine**, v. 2013.

BISHOP, C. M. **Neural networks for pattern recognition**. New York: Oxford University, 1999.

CEDEÑO, A. M.; QUINTANILLA, J. D.; ANDINA, D. WBCD breast cancer database classification applying artificial metaplasticity neural network. **Expert Systems with Applications**, v. 38, n. 8, p. 9573-9579, 2011.

DUDA, R. O.; HART, P. E. Pattern Classification and Scene Analysis. **Wiley Interscience Publication**, New York, v. 7, p. 370, 1973.

FLOYD, C. E. *et al.* Prediction of breast cancer malignancy using an artificial neural network. **Cancer. ACS Journals**, v. 74, n. 1, p. 2944-2948, 1994.

FOGEL, D. B.; WASSON, E. C.; BOUGHTON, E. M. Evolving neural networks for detecting breast cancer. **Cancer Letters**, v. 96, n. 1, p. 49-53, 1995.

HORNUNG, R. *et al.* Full versus incomplete cross-validation: measuring the impact of imperfect separation between training and test sets in prediction error estimation. **Technical Report**, Munich, n. 159, abr. 2014.

HOSSIN, M.; SULAIMAN, M. N. A review on evaluation metrics for data classification evaluations. **International Journal of Data Mining & Knowledge Management Process**, v. 5, n. 2, p. 1-11, mar. 2015.

INSTITUTO NACIONAL DO CÂNCER – INCA. **Câncer de mama: vamos falar sobre isso?** 8. ed. Rio de Janeiro: INCA, 2023a.

INSTITUTO NACIONAL DO CÂNCER – INCA. **Outubro Rosa 2023**. Disponível em: <https://www.gov.br/inca/pt-br/assuntos/campanhas/2023/outubro-rosa>. Acesso em: 25 out. 2023b.

ISA, I. S. *et al.* Suitable MLP Network Activation Functions for Breast Cancer and Thyroid Disease Detection. *In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL INTELLIGENCE, MODELLING AND SIMULATION, 2.*, 2010, Bali. **Anais [...]**.Bali, 2010. p. 39-44.

KATHIJA, A.; NISHA, S. S.; SATHIK, M. M. Breast cancer data classification using neural network approach of MLP algorithm. **International journal of trend in research and development**, v. 4, n. 3, p. 275-279, 2017.

KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. **Ijcai**, v. 14, n. 2, p. 1137-1145, ago. 1995.

LAUREANO, R. M. S.; CAETANO, N.; CORTEZ, P. Previsão de tempos de internamento num hospital português: aplicação da metodologia CRISP-DM. **RISTI: Revista Ibérica de Sistemas e Tecnologias da Informação**, n. 13, p. 581-591, jun. 2014.

MATPLOTLIB. **Matplotlib documentation**. Disponível em: <https://matplotlib.org/>. Acesso em: 15 set. 2023.

MEINEL, L. A. *et al.* Breast MRI lesion classification: improved performance of human readers with a backpropagation neural network computer-aided diagnosis (CAD) system. **Journal of Magnetic Resonance Imaging**, v. 25, n. 1, p.89-95, 2007.

NUMPY. **Numpy documentation**. Disponível em: <https://numpy.org/doc/>. Acesso em: 15 set. 2023.

PANDAS. **Pandas documentation**. Disponível em: <https://pandas.pydata.org/>. Acesso em: 14 set. 2023.

SCIKIT-LEARN. **Scikit-learn documentation**. Disponível em: <https://scikit-learn.org/stable/>. Acesso em: 14 set. 2023

TING, F. F.; SIM, K. S. Self-regulated multilayer perceptron neural network for breast cancer classification, *In: INTERNATIONAL CONFERENCE ON ROBOTICS, AUTOMATION AND SCIENCES – ICORAS, 2017, Melaka*. **Anais [...]**. Melaka: IEEE, 2017.

TRIBA, M. N. *et al.* PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the k-fold cross-validation quality parameters. **Mol. BioSyst**, v. 11, n. 1, p. 13-19, 2015.

WOLBERG, W. *et al.* Breast cancer wisconsin: diagnostic. **UCI Machine Learning Repository**, 1995. Disponível em: <https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic>. Acesso em: 01 set. 2023.

WORLD HEALTH ORGANIZATION. **Breast cancer**. 2023. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/breast-cancer>. Acesso em: 15 out. 2023.