**DIAGNÓSTICO RÁPIDO MAMOGRAFIA**

Amanda Rezende dos Santos <amanda.rezende0202@gmail.com>

Everton Mendonça Lima <everton.lima07@gmail.com>

Layla Joana Santos <layla.joana@gmail.com>

Universidade Federal de Sergipe (UFS) – Curso de Sistemas de Informação – Campus Itabaiana

Av. Vereador Olímpio Grande, S/N – Bairro Centro – CEP 49500-000 – Itabaiana – SE

**Resumo**

O presente artigo tem como objetivo relatar o desenvolvimento de uma aplicação para diagnosticar, através do uso de Redes Neurais Artificiais, um caso de câncer de mama. Será explanado inicialmente sobre a Rede Neural Artificial, logo após sobre a problemática que levou a escolha deste tema. Por fim, será relatado sobre o desenvolvimento do sistema através de imagens dos principais trechos do código.

Palavras-Chaves: Redes Neurais Artificiais, Câncer de mama, diagnóstico.

***Abstract***

*Title: Quick Diagnostic Mammography*

*The present article aims to report the development of a system to diagnose, through the use of Artificial Neural Networks, a case of breast cancer. It will be explained initially about the Artificial Neural Network, soon after about the problematic that led to the choice of this theme. Finally, it will be reported on the development of the system through images of the main passages of the code.  
  
Keywords: Artificial Neural Networks, Breast Cancer, diagnosis.*

1. **INTRODUÇÃO**

O presente artigo tem como finalidade descrever o desenvolvimento de um sistema para diagnóstico de câncer de mana. O objetivo do sistema é poder, com base no treinamento obtido através de uma base de dados, informar se o câncer é maligno ou benigno, tendo como dados de entrada a avaliação do BI – RADS, a forma, a margem e a densidade de uma lesão observada em uma mamografia, auxiliando.

Para o desenvolvimento do referido sistema, foi utilizado conceitos de Redes Neurais Artificiais (RNA’s) com múltiplas camadas de neurônios. Inicialmente será explanado o que são as RNA’s e como elas funcionam. Posteriormente, será relatado sobre o câncer de mama, e por fim, sobre o desenvolvimento do sistema e as suas funcionalidades.

1. **REDES NEURAIS ARTIFICIAIS (RNA’s)**

As Redes Neurais Artificiais são um sistema computacional constituído por unidades conhecidas como neurônio. Os neurônios são elementos processadores interligados, representando o sistema nervoso humano, ou seja, sua forma de pensar e tomar decisão sobre determinado assunto. Ela faz uso de *Perceptron* – algoritmo de aprendizado para resolver problemas linearmente separáveis e classificar as entradas em grupos diferentes. A imagem abaixo, representa o modelo de um neurônio artificial.

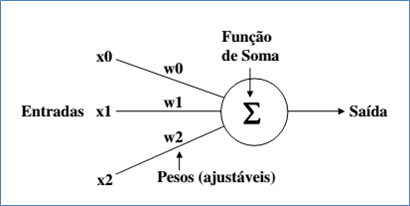


Figura 1 - Representação de um neurônio artificial.

Onde:

* x0, x1, x2 representam as minhas entradas;
* w0, w1, w2 representam os pesos sinápticos;
* A Função de Soma, soma todos os sinais de entrada ponderados pelos pesos das conexões.
* A Saída, também representada por Y, é dada da seguinte forma:

A maioria dos modelos de redes neurais possui alguma regra de treinamento, onde os pesos de suas conexões são ajustados de acordo com os padrões apresentados. Em outras palavras, elas aprendem através de exemplos.

O Perceptron apresenta as seguintes características:

* **A Camada de Entrada**: onde os padrões são apresentados a rede;
* **A Camada de Saída**: onde o resultado final é apresentado;
* **Neurônios da Camada Intermediária ou Oculta**: onde é feita a maior parte do processo de classificação, através das conexões ponderadas.

Para o desenvolvimento do sistema, foi utilizado a rede MLP – *Multi Layer Perceptron*, que é semelhante ao perceptron simples, porém contém mais de uma camada intermediária. Como pode ser observado na figura abaixo:

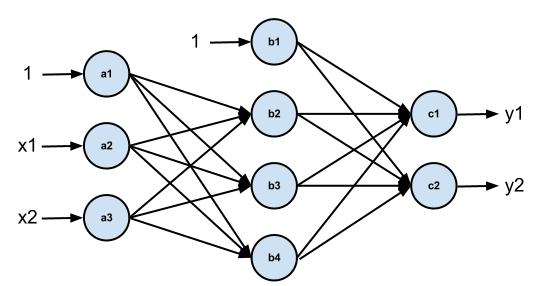


Figura - Representação da MLP

Sua rede de treinamento é feita pelo algoritmo de retro propagação do erro (*Back propagation*). Ele inicializa os pesos da rede com valores aleatórios para cada padrão de entrada. A partir desta camada, as unidades calculam sua resposta que é produzida na camada de saída, o erro é calculado e se o valor obtido for diferente do resultado, os pesos dos neurônios são atualizados até o erro ser aceitável.

Após a rede estiver treinada, estará pronta para ser usada como um classificador, onde, será inserido os dados da camada de entrada, que serão processados pelas camadas intermediárias, e o resultado exposto pela camada de saída, sem a necessidade de repropagar o erro.

1. **PROBLEMATICA DO TEMA**

O câncer de mama, segundo o Instituto Nacional de Câncer (INCA), é o segundo tipo de câncer mais comum no mundo e no Brasil entre as mulheres, perdendo apenas para o câncer de pele. Ele representa 28% dos casos novos a cada ano. O câncer de mama é a maior causa de morte por câncer nas mulheres em todo o mundo, com cerca de 522 mil mortes estimadas por ano. É a segunda causa de morte por câncer nos países desenvolvidos, atrás somente do câncer de pulmão, e a maior causa de morte por câncer nos países em desenvolvimento.

No entanto, apesar do câncer de mama ser mais comumente ocorrido em mulheres, os homens também podem desenvolvê-lo. Porém, é mais raro, representando apenas 1% do total de casos da doença. Em homens, a doença tende a ser diagnosticada em estágios mais avançados do que em mulheres, devido ao baixo índice de suspeita e a menor dimensão do tecido mamário.

O exame para a detecção do câncer é realizado pela mamografia – uma radiografia das mamas realizada em um equipamento específico, o “mamógrafo”, que permite detectar lesões ainda muito pequenas, na fase inicial do câncer, em que o tumor não é palpável.

Entretanto, para a confirmação do diagnóstico, normalmente é realizado uma biopsia, sendo que em 70% dos casos essa biopsia se faz desnecessária. Para amenizar essa situação e auxiliar ao médico na sua tomada de decisão sobre a realização de uma biopsia ou um tratamento com acompanhamento, foi-se sugerido um sistema que ao ser inseridas informações obtidas na mamografia como: BI-RADS, forma, margem e densidade, o sistema dará um possível diagnóstico do tumor, no caso, se o mesmo é maligno ou benigno.

1. **DESENVOLVIMENTO DO SISTEMA**

Como citado anteriormente, o sistema tem como propósito informar, dado um conjunto de entradas, se o tumor é benigno ou maligno. Para o desenvolvimento do sistema foi utilizado uma base de dados fornecida pela Universidade da Califórnia sobre mamografias que, como já foi dito, compõe os seguintes campos: BI-RADS, forma, margem, densidade e o resultado, ou seja, benigno ou maligno. A base foi usada para efetuar o treinamento da rede.

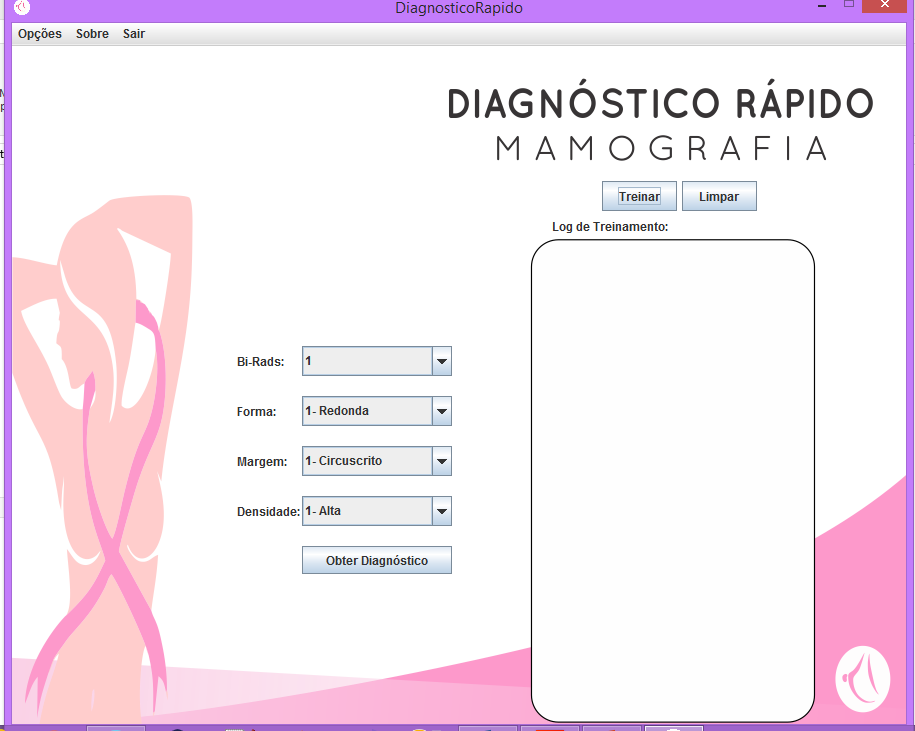


Figura - Tela do sistema.

Na imagem acima, mostra as quatro entradas a serem avaliadas pelo sistema. O BI-RAIDS, que é uma forma padronizada em todo o mundo de relatar os achados radiológicos da mamografia com o intuito de diminuir a má interpretação dos laudos. Na base utilizada, o BI-RAIDS está classificado de 1 à 5, onde, o BI\_RAIDS 1 significa que a mamografia não apresenta nenhuma alteração, ou seja, exame normal. O BI-RAIDS 2 significa que foi encontrado alguma alteração na mamografia, mas as características da lesão permitem afirmar que ela é benigna. O BI-RAIDS 3 significa que foi encontrado algumas alterações na mamografia, que provavelmente é benigna, no entanto, o médico não tem 100% de certeza. O BI-RAIDS 4 significa que foi encontrado alguma alteração na mamografia que pode ou não ser um câncer, o que resulta a uma submissão de uma biópsia. O BI-RAIDS 5 significa que foi encontrado alterações na mamografia que possivelmente é derivada de um câncer de mama.

A segunda entrada é a forma da massa que está dividida em quatro: 1 – redonda, 2 – oval, 3 – lobular, 4 – irregular.

A terceira entrada são as margens da massa que por sua vez estão divididos em 5: 1- circunscrito, 2 – microlobulado, 3 – obscurecido, 4 – mal definido e 5 – espiculado.

E a última entrada é a densidade da massa, que pode ser 1 – alta, 2 – padrão, 3 – baixa ou 4 – contendo gordura.

No Log de Treinamento, é onde será mostrado o treinamento da rede com base na MLP. Onde, para exibi-lo, deverá inicialmente informar o número de épocas, que representa o número de vezes que a rede irá treinar. A taxa de Aprendizado, é o valor que será usado para alterar os pesos após não consegui convergir e a taxa de erro, que é a margem de erro aceito pelo sistema.

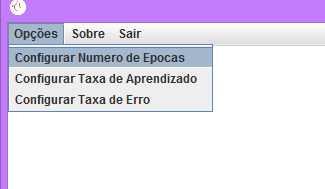
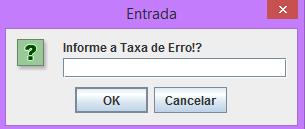
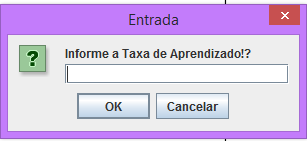
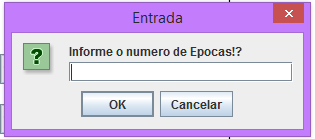
**

Figura - Configurações da Rede.

**

Após realizar as dadas configurações da rede, será exibido o treinamento da rede, e exibida no Log, como mostra a imagem abaixo.

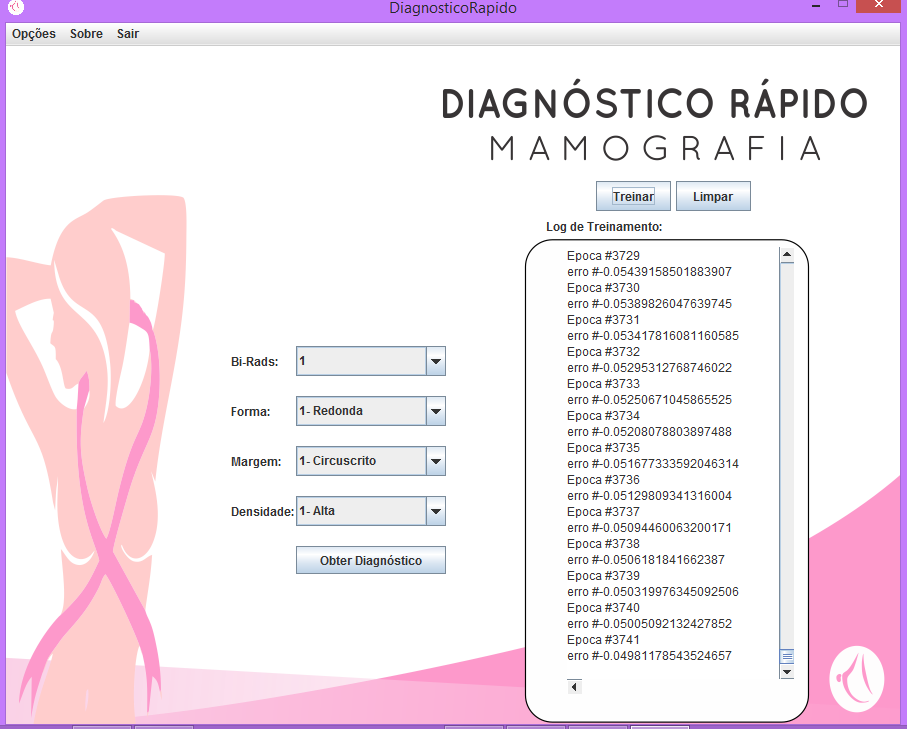


Figura - exibição do treinamento da rede.

Feito isso, é só inserir os valores das entradas e clicar no botão “*Obter Diagnóstico”* que será exibido uma caixa com o resultado.

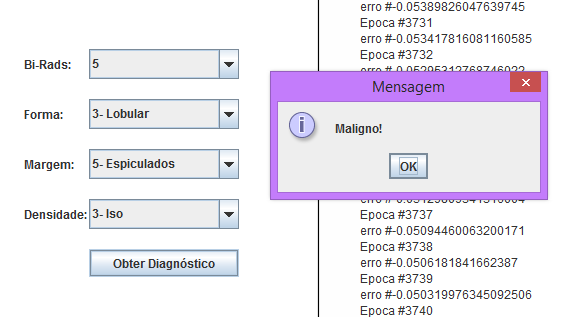


Figura - Exibição do diagnóstico com base nos dados de entrada.

Os principais métodos para a realização do sistema foram o “*Treinar”* e o “*Classificar”*.

O método de treinar tem como parâmetro a base de dados e o valor esperado. Este método é repetido em quanto o valor do erro em percentual for maior que o valor definido e o número de épocas for menor que o estipulado. Como pode ser observado na figura abaixo o retorno deste método é o número de épocas que foram necessárias para convergir e o valor do erro que foi aceito.

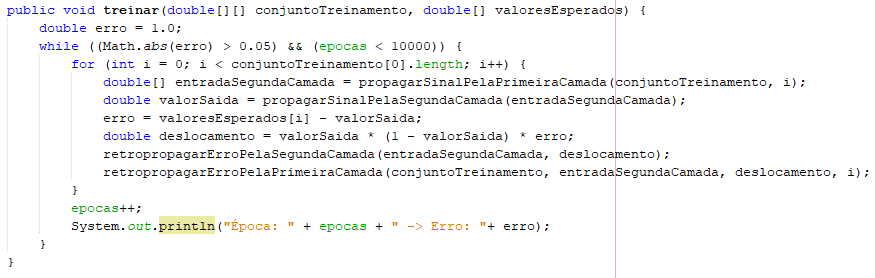


Figura - Método de Treinamento da rede.

A propagação do sinal é responsável por fazer a comunicação entre a camada atual e a camada seguinte, o resultado da propagação será a saída da primeira camada e a mesma é dada pela seguinte formula.

Que é representada pelo seguinte trecho de código:

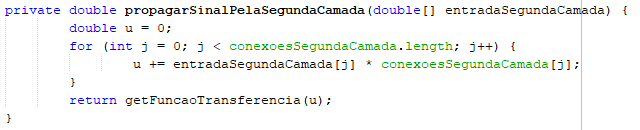


Figura - Método de Propagação do Sinal.

Após ter encontrado o valor de saída é calculado o erro, o mesmo é usado para ajustar o valor do gradiente, que tem como função minimizar o erro entre a saída da rede e a saída desejada.

O aprendizado da rede neural artificial é dado após a retro propagação do sinal, ou como também é conhecido o Back Propagation, o processo de aprendizado que volta da segunda cama para a primeira fazendo as devidas correções do peso de acordo com as entradas da segunda camada e o gradiente. Como pode ser observado na formula a seguir e no trecho de código abaixo:

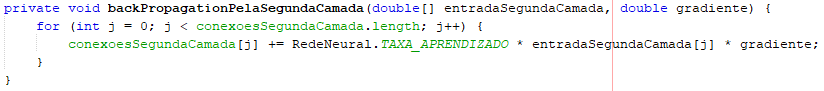


Figura - BackPropagation

O método classificar é o responsável por distinguir se as entradas refletem em um diagnóstico. Mas para que se possa classificar é necessário primeiro verificar se o número de épocas é maior que o limite máximo menos um.

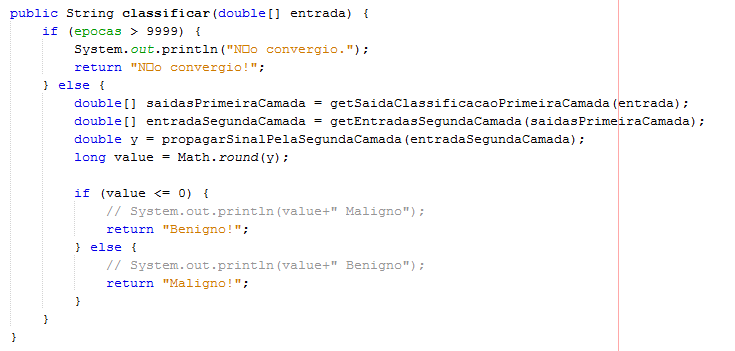


Figura - Método que Classifica a rede.

O número de épocas deve ser definido no processo de configuração da rede neural artificial, se o número de épocas não for maior que o limite definido, o mesmo será classificado com base no valor que é propagado através do método de propagação de sinal pela segunda camada.

Como os valores possui muitas casas decimais para que não tenha tanta similaridade entre os resultados no método de classificação é necessário arredonda o valor encontrado para que se classifique de forma adequada. Se o valor encontrado for maior ou igual a 0 (zero) para as entradas dadas, em nossa aplicação, é classificada como benigno, já se o valor for menor que zero é classificado como maligno.

**CONCLUSÃO**

Foi observado que na implementação de uma rede neural artificial com uso de perceptron de múltiplas camadas (MLP), faz uso de um algoritmo chamado de back propagation que foi citado anteriormente. Este algoritmo consiste basicamente em dois passos, o primeiro é a propagação positiva do sinal, onde todos os pesos da rede são mantidos fixos e o segundo é a retro propagação do erro que é o processo que ajusta os pesos tomando como base o erro encontrado entre o valor de saída e o valor esperado.

Em problemas que não são linearmente separáveis a MLP pode ser vista como uma ótima ferramenta de classificação, visto que cada camada intermediaria pode ser entendida como uma dimensão, ou seja, aumentando-se a quantidade de neurônios na camada intermediaria, aumenta-se a capacidade de mapeamento não linear da rede. Deve-se tomar cuidado para que não haja um overfitting/superação que seria o treinamento excessivo. Como também em um rede com um número muito pequeno de camadas intermediarias pode ocorrer o underfitting/inadequado, gerado valores que não condiz com a realidade.

A qualidade e a eficiência da rede neural artificial de múltiplas camadas dependem fortemente da especificação da arquitetura da rede, número de camadas intermediaria, função de ativação do neurônio, da regra de aprendizado, back propagation, taxa de aprendizado, valor aceitável do erro e dos valores iniciais nos pesos.

**REFERÊNCIAS**

ELTER, Matthias. Mammographic Mass Data Set . Disponível em: <http://archive.ics.uci.edu/ml/datasets/Mammographic+Mass>. Acesso em: 16 ago. 2017.

ANDRADE, Wesley Pereira. Câncer de mama: Entenda a classificação BIRADS. Disponível em: <http://www.minhavida.com.br/saude/materias/18470-cancer-de-mama-entenda-a-classificacao-birads>. Acesso em: 06 set. 2017.

TATIBANA, Cassia Yuri; KAETSU, Deisi Yuki. Redes Neurais. Disponível em: <http://www.din.uem.br/ia/neurais/>. Acesso em: 06 set. 2017.

ZUBEN, Profs Fernando J. Von; ATTUX, Romis R. F. . Perceptron de Múltiplas Camadas\*. Disponível em: <http://ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/ia353\_1s07/topico5\_07.pdf>. Acesso em: 08 set. 2017.

MINISTÉRIO DA SAÚDE, INCA. MAMA. Disponível em: <http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/mama/cancer\_mama>. Acesso em: 09 set. 2017.

PINHEIRO, Dr Pedro. ENTENDA A CLASSIFICAÇÃO BI-RADS DA MAMOGRAFIA. Disponível em: <https://www.mdsaude.com/2016/12/classificacao-bi-rads.html>. Acesso em: 09 set. 2017.

ICMC, Andre. Perceptron Multi-Camadas (MLP). Disponível em: <http://conteudo.icmc.usp.br/pessoas/andre/research/neural/MLP.htm>. Acesso em: 11 set. 2017.

ZUBEN, Profs Fernando J. Von; CASTRO, Prof. Leandro N. de . Redes Neurais Artificiais. Disponível em: <http://ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/ia006\_03/topico5\_03.pdf>. Acesso em: 14 set. 2017.