

UMA ABORDAGEM BASEADA EM REDES PERCEPTRON MULTICAMADAS PARA A CLASSIFICAÇÃO DE MASSAS NODULARES EM IMAGENS MAMOGRÁFICAS

Luan de Oliveira Moreira¹; Matheus Giovanni Pires²

1. Bolsista PROBIC, Graduando em Engenharia de Computação, Universidade Estadual de Feira de Santana, e-mail: luanecomp@yahoo.com.br
2. Orientador, Departamento de Ciências Exatas, Universidade Estadual de Feira de Santana, e-mail: mgpires@ecom.ufes.br

PALAVRAS-CHAVE: Massas Nodulares, Classificação de Padrões, Redes *Multi-Layer Perceptron*.

INTRODUÇÃO

O câncer de mama tem sido uma das principais causas de mortalidade entre as mulheres em toda a parte do mundo (TAHMOUSH, 2007). A mamografia é o método mais eficaz para detecção precoce da doença. De cada 100 mulheres submetidas a exames para detecção da presença de massas suspeitas do câncer de mama, 80% dos casos apresentam imagens mamográficas que, ao serem analisadas por um radiologista, geram dúvidas para um diagnóstico preciso (GLINGANI & AMBRÓSIO, 2004).

Distorções na interpretação e classificação de lesões suspeitas por especialistas implicam um número maior de biopsias desnecessárias. Ou seja, entre 65% a 85% das biopsias de mama são realizadas em lesões benignas (RIBEIRO *et al.*, 2008).

Buscando auxiliar os especialistas na detecção precoce do câncer de mama, vêm sendo desenvolvidos os esquemas *Computer-Aided Detection* (CAD) com o objetivo de melhorar a acurácia do diagnóstico, de maneira a auxiliar o radiologista na interpretação e avaliação de estruturas de interesse presentes na imagem (RIBEIRO *et al.*, 2008).

As *Redes Neurais Artificiais* (RNAs) são frequentemente usadas como classificadores em esquemas CAD (RIBEIRO *et al.*, 2008). Assim, com o objetivo de auxiliar especialistas da área de análise de imagens mamográficas no diagnóstico de massas nodulares, este trabalho propõe o desenvolvimento de um sistema para classificação de nódulos através de mamografias digitalizadas utilizando RNAs, mais especificamente as redes *Multi-Layer Perceptron* (MLP) (HAYKIN, 2001).

METODOLOGIA

A proposta do trabalho é executar o treinamento e testes na rede MLP com diferentes números de atributos e topologias distintas. O objetivo é encontrar a configuração da rede que mais de adequa à classificação de nódulos em imagens mamográficas, tanto em número de atributos utilizados como entrada na RNA utilizada, quanto na sua topologia.

O desempenho do classificador, neste caso, as RNAs, depende de dois estágios anteriores ao processo de classificação: a aquisição do padrão e a extração e seleção de atributos que melhor representem cada classe (RIBEIRO, 2006). Tomando como base essa afirmação, serão utilizadas técnicas para selecionar os melhores atributos para servirem como entradas no classificador.

Para a extração de atributos das regiões de Interesse (RIs) das imagens mamográficas, serão utilizadas as técnicas de extração conhecidas como Descritores de Intensidade, Descritores de Textura de Haralick e Descritores Geométricos, a fim de obter uma quantidade significativa de atributos, que serão utilizados para uma melhor classificação

das imagens (PATROCINIO *et al*, 2004). A etapa de extração dos atributos ainda não foi concluída e, por esse motivo, os testes realizados foram feitos utilizando um banco de dados de imagens mamográficas existente na internet chamado *Breast Cancer Wisconsin Database*, que será detalhado posteriormente.

Durante os testes do classificador CAD será apresentado ao sistema implementado um arquivo texto, com um número previamente definido de padrões de entradas. Cada padrão de entrada possuirá, inicialmente, todos os atributos extraídos, totalizando 29 atributos, que serão utilizados como entradas para o classificador. A ideia é utilizar todos os atributos, a fim de testar a eficiência da classificação de nódulos em imagens mamográficas. A partir daí serão extraídos os melhores atributos dentre os 29 existentes para servir como entrada para novos testes na rede. Essa etapa depende da extração dos atributos citado anteriormente, e que ainda não foi implementado.

Para a classificação foi utilizada a Rede Neural Artificial MLP. Para configurar a rede e visualizar melhor os resultados dos testes, foi desenvolvida uma interface gráfica. A tela inicial da interface desenvolvida pode ser visualizada na Figura 1.



Figura 1. Tela inicial da interface desenvolvida.

O programa ainda dá ao usuário a opção de exibir o resultado calculado pela rede, a partir dos padrões de entrada provenientes do arquivo de testes. O resultado calculado é exibido em uma tabela, também mostrada pela interface. As saídas do classificador implementado indicam o número entre zero e seis equivalentes na padronização *Breast Imaging Reporting and Data System* (BI-RADS) que trata da padronização dos laudos médicos de exames mamográficos (RIBEIRO, 2006). A interface gráfica, bem como a lógica utilizada para a classificação dos padrões (Rede MLP) já se encontram implementadas.

As imagens mamográficas utilizadas para a extração dos atributos utilizados como entradas para o classificador CAD serão obtidas através de banco de dados já existentes. Para a seleção dos melhores atributos dentre os extraídos das imagens será utilizada o método de Distribuições Gaussianas, onde quanto menor a sobreposição das curvas, melhor o atributo poderá representar cada padrão (RIBEIRO, 2006). A lógica responsável por fazer a distribuição gaussiana ainda não foi implementada, mas deve ser uma atualização do programa já feito (interface gráfica e classificador), não sendo necessário criar um novo programa para isso.

A Figura 2 mostra um exemplo de dois atributos que não possuem sobreposição de curvas, onde o eixo vertical indica os resultados de uma função gaussiana $G(y)$ qualquer utilizada, em relação aos valores de y atribuídos à função que são mostradas no eixo horizontal, sendo considerados assim bons atributos para servirem como entrada em um classificador, e um exemplo de várias curvas de Distribuição Gaussiana com sobreposição, o que indica que os atributos selecionados provavelmente não farão uma boa classificação dos nódulos se oferecidos como entrada para um classificador CAD que utiliza rede MLP. A leitura dos eixos na Figura 2-B é análoga à descrita para a Figura 2-A.

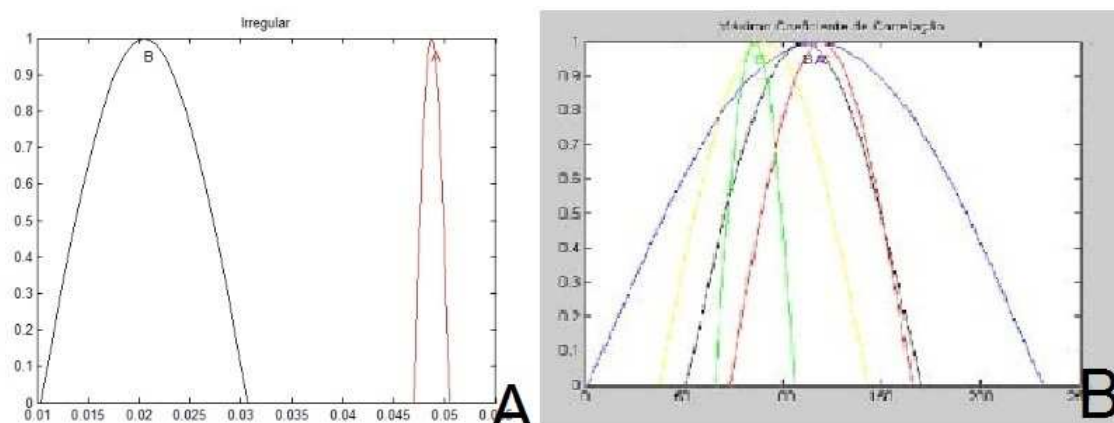


Figura 2. Distribuições Gaussianas. A - Distribuição Gaussiana sem sobreposição de curvas; B - Distribuição Gaussiana com sobreposição de curvas.

RESULTADOS

Como a etapa de extração dos descritores das imagens mamográficas ainda está sendo implementada, para os testes foi utilizado um banco de dados de imagens mamográficas existente na internet chamado *Breast Cancer Wisconsin Database*, cujo arquivo de entrada encontra-se no formato *.data*, possuindo 699 padrões de entradas, com 10 atributos e a saída esperada da rede para cada padrão. A saída esperada pelo banco de dados utilizado pode ser definida como sendo de forma binária, classificando os padrões de entrada como sendo com câncer ou sem câncer.

Após o processamento do arquivo de entradas *Breast Cancer Wisconsin Database* pelo algoritmo desenvolvido, o arquivo foi dividido em dois outros arquivos, onde um é utilizado para o treinamento da rede e possui 618 padrões de entradas e o outro é utilizado para teste da rede, e possui 81 padrões de entradas. A Tabela 1 mostra os parâmetros utilizados no teste da rede, e a taxa de acerto para cada topologia testada.

Tabela 1. Testes realizados pelo classificador CAD.

Neurônios na Camada1	Neurônios na Camada2	Neurônios na Camada3	Taxa de Aprendizagem	Erro	Épocas de Treinamento	Taxa de Acerto (%)
5	5	0	10^{-2}	10^{-5}	758	97.53
5	10	0	10^{-2}	10^{-5}	752	97.53
10	10	0	10^{-2}	10^{-5}	753	97.53
15	10	0	10^{-2}	10^{-5}	905	97.53

5	5	5	10^{-2}	10^{-5}	1045	97.53
5	10	5	10^{-2}	10^{-5}	813	94.57
10	10	5	10^{-2}	10^{-5}	1041	97.53
15	10	5	10^{-2}	10^{-5}	1099	97.53
5	5	0	5×10^{-2}	10^{-5}	298	97.53
5	10	0	5×10^{-2}	10^{-5}	305	97.53
3	2	0	10^{-2}	10^{-5}	784	97.53
3	2	0	5×10^{-2}	10^{-5}	265	97.53

Para cada configuração mostrada na Tabela 1 foram executados cinco testes. Os valores de Ciclos de Treinamento e Taxa de Acerto são a média dos valores obtidos de cada um. Podemos perceber que mesmo variando o número de camadas intermediárias do classificador, a taxa de acertos continuou a mesma em todos os testes. O diferencial de cada configuração testada está na quantidade de épocas de treinamento que o classificador demora para treinar a rede. A melhor configuração encontrada para a rede dentre os testes realizados foi a com três neurônios na primeira camada escondida, dois neurônios na segunda camada escondida, taxa de aprendizagem de 0.05, erro (critério de convergência) com 265 épocas para a rede convergir e taxa de acerto de 97.53%.

Durante a análise dos resultados dos testes, foi percebido que dois padrões de entrada no arquivo dos testes nunca eram classificados corretamente. A explicação pode vir do fato de que esses padrões estão em uma região do espaço que faz fronteira entre as duas classes de nódulos detectadas pela RNA, e que por esse motivo possui difícil classificação, ou que os padrões apresentam dados (atributos) errados, o que faz com que eles sejam classificados como sendo de uma classe as quais não pertencem.

CONSIDERAÇÕES FINAIS

A utilização das redes MLP para classificação de nódulos mamários se mostrou eficiente, obtendo resultados satisfatórios em comparação com outras técnicas já aplicadas presentes na literatura (RIBEIRO, 2006). Os resultados mostram a importância deste trabalho e que o objetivo principal do projeto tende a ser alcançado, tendo em vista que o objetivo futuro é utilizar banco de imagens mamográficas com as saídas esperadas formatadas no padrão BI-RADS, além de testar o Classificador utilizando os melhores atributos extraídos das imagens mamográficas.

REFERÊNCIAS

- GLINGANI, F. A.; AMBRÓSIO, P. E. 2004. Sistema de análise computadorizada para auxílio à detecção de lesões de mama baseado em redes neurais artificiais, *In: Anais do IX Congresso Brasileiro de informática em Saúde*, Ribeirão Preto.
- HAYKIN, S. 2001. *Redes Neurais Artificiais: Princípios e Práticas*. Porto Alegre, Bookman, 900p.
- PATROCINIO, A. C. 2004. Classificador automático de achados mamográficos em imagens digitais de mamas densas utilizando técnicas híbridas. Universidade de São Paulo, Tese.
- RIBEIRO, P. B. 2006. Classificação por Análise de Contornos de Nódulos Mamários Utilizando Redes Neurais Artificiais. Universidade de São Paulo, Tese.

RIBEIRO, P. B.; SCHIABEL, H.; PATROCINIO, A. C.; ROMERO, R. A. F. 2008. Análise da Variação de Textura em Imagens Mamográficas para Classificação de Massas Suspeitas. *In: III Workshop de Visão Computacional, Bauru.*

TAHMOUSH, D. 2007. Image differencing approaches to medical image classification. *In: 36th IEEE Applied Imagery Pattern Recognition Workshop , Washington, p.22-27.*