

DIAGNÓSTICO DE CÂNCER DE MAMA BASEADO NA REDE NEURAL DE KOHONEN

Weverson dos Santos Gomes¹; Matheus Giovanni Pires²

1. Bolsista FAPESB, Graduando em Engenharia de Computação, Universidade Estadual de Feira de Santana, e-mail: kiko.ecomp@gmail.com
2. Orientador, Departamento de Ciências Exatas, Universidade Estadual de Feira de Santana, e-mail: mgpires@ecomp.uefs.br

PALAVRAS-CHAVE: Massas nodulares, câncer de mama, Redes de Kohonen.

INTRODUÇÃO

O câncer de mama é uma das doenças mais temidas entre as mulheres devido à sua frequência, alta taxa de mortalidade feminina em todo o mundo (Tahmoush, 2007) e, sobretudo, pelos seus efeitos psicológicos, que afetam a percepção da sexualidade e a própria imagem pessoal. Só no Brasil, segundo o INCA (Instituto Nacional de Câncer), o número de mortes em 2010 foi de 12.852, sendo 147 homens e 12.705 mulheres. A estimativa de novos casos para 2012 foi de 52.680 (INCA, 2013).

De acordo com INCA (2013), "No Brasil, as taxas de mortalidade por câncer de mama continuam elevadas, muito provavelmente porque a doença ainda é diagnosticada em estádios avançados. Na população mundial, a sobrevivência média após cinco anos é de 61%". Como suas causas ainda não são bem conhecidas, a prevenção é a única maneira de reduzir a mortalidade de mulheres com câncer de mama, pois a detecção precoce aumenta as chances de cura substancialmente (Ball *et al.* 2004; Mavroforakis, 2005).

A mamografia é considerada como o exame mais eficaz para a detecção do câncer de mama ainda impalpável devido a sua alta sensibilidade. No entanto, segundo Giger (2000), de 10% a 30% das mulheres que apresentaram câncer de mama tiveram resultados negativos quando submetidas à mamografia, o que leva a crer que houve uma má interpretação dos exames. Distorções na interpretação e classificação de lesões por especialistas implicam um número maior de biópsias desnecessárias (Mavroforakis, 2005; Bilka e Wolak, 2005), ou seja, entre 65% a 85% das biópsias de mama são realizadas em lesões benignas (Kopans, 1992). Com isso, há uma redução na relação custo-benefício dos exames e, no pior caso, a não detecção da doença, caracterizando um diagnóstico falso-negativo.

Buscando minimizar esses problemas e auxiliar na detecção precoce do câncer de mama, vêm sendo desenvolvidos esquemas denominados CAD, Diagnóstico Auxiliado por Computador (do Inglês, *Computer-Aided Diagnosis*), com o objetivo de auxiliar o radiologista na interpretação das imagens mamográficas (Ribeiro *et al.*, 2007). Estes esquemas são sistemas computacionais que auxiliam no diagnóstico de pacientes e podem servir como uma segunda opinião aos radiologistas quanto à localização e à natureza dos nódulos mamários.

Portanto, com o objetivo de auxiliar especialistas da área de análise de imagens mamográficas, este trabalho propõe o desenvolvimento de um esquema CAD baseado em Redes Neurais Artificiais (RNA), mais precisamente em redes de Kohonen, para o auxílio no diagnóstico de câncer de mama.

METODOLOGIA

O trabalho desenvolvido é baseado na rede neural proposto pelo finlandês Teuvo Kohonen, conhecida como Rede de Kohonen ou Mapas Auto-Organizáveis (Haykin, 1999). Os dados de treinamentos e testes da rede foram obtidos da base de dados "*Breast Cancer*

Wisconsin”, do repositório UCI Machine Learning, da Universidade da Califórnia, criada por Frank e Asuncion (2010).

Essa base de dados é bastante conhecida no meio científico e já foi utilizada em diversos outros trabalhos. Ela possui 699 amostras, as quais foram coletadas a partir de casos clínicos do Dr. Wolberg, da Universidade de Wisconsin. Cada amostra representa as informações de um determinado nódulo mamário, a qual possui 9 atributos (características), e mais dois atributos de identificação, isto é, o código da amostra e a classificação do nódulo, ou seja, se é um nódulo maligno ou benigno. Cada neurônio da rede possui nove entradas, que são os atributos das amostras.

Das 699 amostras disponíveis na base de dados, 16 não puderam ser utilizadas, pois faltavam informações referentes a algum atributo. Das 683 amostras restantes, 478 (equivalente a 70% do número total de amostras) foram utilizadas no processo de treinamento da rede e 205 (equivalente a 30% do número total de amostras) foram usadas nos testes da rede. Dessa forma, nenhuma amostra utilizada nos testes participou do processo de treinamento. O número total de amostras malignas é de 239 e de benignas é de 444, o que equivale a 35% e 65%, respectivamente, do conjunto de 683 amostras.

O conjunto de amostras usado no treinamento da rede contém 65% de amostras do tipo benigno e 35% do tipo maligno. Para o conjunto de teste se manteve a mesma proporção. A representatividade de cada classe (benigno e maligno) nos conjuntos de treinamento e teste é a mesma em relação ao conjunto total de amostras. Como há uma maior quantidade de amostras do tipo benigno, buscou-se preservar esta diferença nos conjuntos de treinamento e teste para evitar uma possível redução na representatividade da classe maligna, prejudicando o aprendizado da rede para esta classe.

O Mapa Topológico de Vizinhança¹ usado neste trabalho é uma função gaussiana, a qual é definida pela Equação 1 (Haykin, 1999). Na Figura 1 está ilustrada a função de vizinhança gaussiana.

$$h_{j,i(x)} = \exp\left(\frac{d_{j,i}^2}{2\sigma^2}\right) \quad (1)$$

A função $h_{j,i}$ é a distância do neurônio i para o neurônio vencedor j . $d_{j,i}$ é a distância entre os neurônios j e i . O parâmetro σ é a largura da vizinhança topológica, o qual mede o grau que os neurônios excitados na vizinhança do neurônio vencedor participam do processo de aprendizagem (Haykin, 1999).

Após o treinamento da rede de Kohonen, tem-se uma lista de neurônios vencedores, ou seja, para cada amostra utilizada no processo de treinamento da rede, há um neurônio vencedor associado a esta amostra. No entanto, falta ainda realizar a *clusterização*, ou seja, definir os agrupamentos e atribuir um rótulo a todos os neurônios pertencentes a um mesmo agrupamento. Neste trabalho foi utilizado o algoritmo de agrupamento *K-Means*, por ser um algoritmo simples e de fácil implementação. Além disso, como será usada a base de dados “*Breast Cancer Wisconsin*” para treinar e testar a rede neural, já se sabe de antemão quantos grupos devem ser formados, que neste caso são dois, benigno e maligno.

¹ O Mapa Topológico de Vizinhança define quais são os neurônios vizinhos para cada neurônio da Rede de Kohonen.

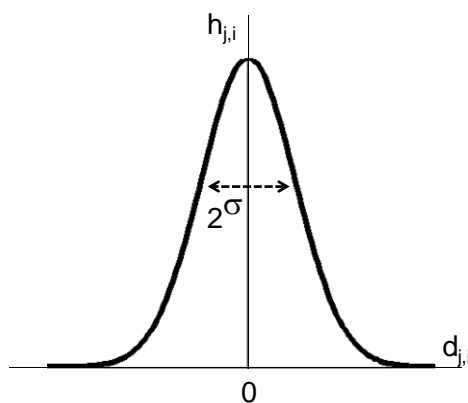


Figura 1. Função gaussiana de vizinhança.

RESULTADOS

O bom desempenho de uma RNA, dentre outros fatores, está relacionado com a configuração de sua topologia. A topologia de uma RNA consiste em definir qual o melhor número de neurônios que a rede deve possuir. No intuito de encontrar a topologia mais adequada para compor o classificador neural, foram realizados experimentos utilizando quantidades diferentes de neurônios, mais precisamente foram analisadas as topologias com 4, 9, 25, 100, 225, 400 e 625 neurônios.

Cada topologia foi testada cinco vezes e ao final das cinco execuções foram calculadas a média de acerto, o desvio padrão, a melhor e a pior taxa de acerto das cinco execuções. Os resultados estão descritos na Tabela 1.

Tabela 1. Resultados obtidos para cada topologia testada.

Topologia	Média de acerto (%)	Desvio padrão	Melhor taxa (%)	Pior taxa (%)
2x2 (4 neurônios)	98,53	0	98,53	98,53
3x3 (9 neurônios)	98,53	0	98,53	98,53
5x5 (25 neurônios)	98,53	0	98,53	98,53
10x10 (100 neurônios)	97,07	0	97,07	97,07
15x15 (225 neurônios)	96,78	1.14	97,56	96,09
20x20 (400 neurônios)	95,02	3.03	96,58	93,65
25x25 (625 neurônios)	94,73	3.34	96,58	92,19

Conforme mostrado na Tabela 1, as topologias com 4, 9 e 25 neurônios obtiveram as mais altas taxas de acerto. Além disso, não foram identificadas variações no percentual de acerto das topologias, por isso o desvio padrão para essas três topologias foi zero. É possível observar que a partir de 100 neurônios, a taxa de acerto começa a cair, e a partir de 225 neurônios, o desvio padrão tende a aumentar.

De acordo com os resultados da Tabela 1, a melhor topologia é a 2x2 (4 neurônios). Apesar de as topologias 3x3 e 5x5 apresentarem a mesma taxa de acerto da topologia 2x2 e terem desvio padrão igual à zero, elas possuem um número maior de neurônios. Nestas situações, onde há topologias com o mesmo desempenho, recomenda-se sempre optar pela topologia mais otimizada, ou seja, com a menor quantidade de neurônios, pois estas demandam menos tempo para serem treinadas.

CONSIDERAÇÕES FINAIS

Neste trabalho foram apresentados os resultados de diagnósticos de câncer de mama realizados por um sistema baseado na rede neural de Kohonen. As informações para treinamento e teste da rede foram obtidas da base “*Breast Cancer Wisconsin*”, do repositório UCI Machine Learning, da Universidade da Califórnia.

O sistema apresentou uma taxa de acerto acima de 98%, com desvio padrão igual à zero, o que demonstra confiabilidade e estabilidade nos resultados do sistema. Sendo assim, o sistema proposto pode contribuir, como uma segunda opinião, aos especialistas em análise de imagens mamográficas, para uma melhor interpretação das imagens e conseqüentemente, proporcionando um diagnóstico mais confiável ao paciente e evitando a realização de biópsias desnecessárias.

REFERÊNCIAS

- BALL, J.E.; T.W. BUTLER; L.M. BRUCE. (2004). Towards Automated Segmentation and Classification of Masses in Digital Mammograms. *In: International Conference of the IEEE EMBS*. San Francisco, CA, USA.
- BILSKA-WOLAK, A.O. *et al.* (2005). Computer aid for decision to biopsy breast masses on mammography: validation on new cases. *Academic Radiology*. 12(6): p. 669-670.
- FRANK, A.; ASUNCION, A. 2010. *UCI Machine Learning Repository*, <http://archive.ics.uci.edu/ml>, Irvine, CA: University of California, School of Information and Computer Science.
- GIGER, M. L. (2000). Computer-aided diagnosis of breast lesions in medical images. *Computing in Science & Engineering*. 2(5): p. 39-45.
- INCA (2013). *Tipos de Câncer: Mama*. Instituto Nacional de Câncer. URL: <http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/mama>. Acesso em 21/02/2013.
- HAYKIN, S. (1999). *Neural Networks - A Comprehensive Foundation*. 2ª ed. New York: Prentice-Hall.
- MAVROFORAKIS, M. *et al.* (2005). Significance analysis of qualitative mammographic features, using linear classifiers, neural networks and support vector machines. *European Journal of Radiology*. 54(1): p. 80-89.
- KOPANS, D.B. (1992). The positive predictive value of mammography. *American Journal of Roentgenology*. vol.158, nro.3, p.521-526.
- RIBEIRO, P.B. *et al.* (2007). Selection of the Best Features for the Classification of Contour of Mammary Nodules Using Artificial Neural Networks. *In: Society for Imaging Informatics in Medicine*. Providence, Rhode Island.
- TAHMOUSH, D. (2007) Image differencing approaches to medical image classification. *In: 36th IEEE Applied Imagery Pattern Recognition Workshop - AIPR*, Washington, p.22-27.