



RELATÓRIO FINAL

**UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA
PRO-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMAÇÃO INSTITUCIONAL DE BOLSAS DE INICIAÇÃO
CIENTÍFICA - PIBIC/FAPESB**

**DIAGNÓSTICO DE MASSAS NODULARES EM IMAGENS
MAMOGRÁFICAS UTILIZANDO REDES DE KOHONEN**

**WEVERSON DOS SANTOS GOMES
MATHEUS GIOVANNI PIRES
INICIAÇÃO CIENTÍFICA – PIBIC/FAPESB**

Agosto/2013



RELATÓRIO FINAL

UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO PROGRAMA INSTITUCIONAL DE BOLSAS DE INICIAÇÃO CIENTÍFICA - PIBIC/FAPESB

DIAGNÓSTICO DE MASSAS NODULARES EM IMAGENS MAMOGRÁFICAS UTILIZANDO REDES DE KOHONEN

Relatório final sobre as atividades desenvolvidas como bolsista do PIBIC apresentado à coordenação do Programa, como parte dos requisitos exigidos na Resolução Normativa do CNPq RN – 017/2006.

**WEVERSON DOS SANTOS GOMES
MATHEUS GIOVANNI PIRES
INICIAÇÃO CIENTÍFICA – PIBIC/FAPESB**

Feira de Santana (BA), Agosto, 2013.

RESUMO

Segundo tipo mais frequente no mundo, o câncer de mama é o mais comum entre as mulheres, respondendo por 22% dos casos novos de câncer a cada ano. Se diagnosticado e tratado oportunamente, o prognóstico é relativamente bom. No Brasil, as taxas de mortalidade por câncer de mama continuam elevadas, muito provavelmente porque a doença ainda é diagnosticada em estágios avançados. Estatísticas indicam aumento de sua incidência tanto nos países desenvolvidos quanto nos em desenvolvimento.

Com o objetivo de auxiliar os especialistas no diagnóstico do câncer de mama, este trabalho apresenta um sistema baseado nos Mapas Auto-Organizáveis de Kohonen, capaz de identificar a presença de nódulos de câncer de mama em imagens mamográficas.

SUMÁRIO

Equação 1: Função Gaussiana.	6
Figura 1: Exemplo de um nódulo mamário (Ribeiro, 2006).	6
Figura 2: Mapa topológico da rede de Kohonen com N=3.	7
Figura 3: Função gaussiana de vizinhança.	7
Tabela 1: Composição dos arquivos de treinamento e teste para a base BCW-Original.	12
Tabela 2: Composição dos arquivos de treinamento e teste para a base BCW-Diagnostic.	12
Tabela 3: Saída do sistema para a base BCW-Original.	13
Tabela 4: Saída do sistema para a base BCW-Diagnostic.	14
RESUMO	3
SUMÁRIO	4
1 INTRODUÇÃO	5
2 REVISÃO DE LITERATURA	6
2.1 Mamografia	6
2.2 Diagnóstico Auxiliado por Computador	7
2.3 Redes Neurais Artificiais	8
2.3.1 Redes de Kohonen	8
3 MATERIAIS E MÉTODOS	10
4 RESULTADOS E DISCUSSÃO	13
5 CONCLUSÃO	15
6 AGRADECIMENTOS	16
7 REFERÊNCIAS BIBLIOGRÁFICAS	16

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 sobrevida 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; Bilska 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.

A interpretação de lesões em uma mamografia é uma tarefa complexa para os especialistas, onde um diagnóstico preciso é diretamente dependente da experiência dos mesmos [Ball *et al.* 2004]. Além da complexidade na interpretação de lesões em uma mamografia, uma série de fatores influencia na interpretação de imagens mamográficas para um correto diagnóstico, tais como, baixo nível de contraste das imagens, nódulos com muitas variações de formas, tamanhos, contornos e densidade, além de possuírem muita semelhança com a densidade do tecido da mama [Menechelli *et al.* 2010a]. Outros fatores que geram problemas para uma análise precisa das imagens mamográficas são a descalibração do aparelho utilizado para fazer as medidas, a má qualidade das imagens geradas pelos aparelhos, as condições de manipulação e conservação do filme e, em imagens digitais, a digitalização da imagem a ser analisada pelo especialista [Ribeiro 2006].

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.* 2005; Porto *et al.* 2010; Marcomini *et al.* 2011]. 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 [Ribeiro *et al.* 2008a; Stemberg 2010]. Por isso, 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, mais

precisamente em mapas auto-organizáveis de Kohonen, para o auxílio no diagnóstico de câncer de mama em imagens mamográficas.

REVISÃO DE LITERATURA

1.1 Mamografia

O rastreamento é a única maneira de detectar e prevenir o câncer de mama precocemente. O instrumento mais efetivo para o diagnóstico precoce é a mamografia [Glingani e Ambrósio 2004; Marcomini *et al.* 2011], a qual pode reduzir em até 23% a taxa de mortalidade, dependendo do estágio em que for descoberta a doença, pois seu principal objetivo é a detecção de lesões não palpáveis ou a verificação de alterações na mama [Borghesan *et al.* 2003; Marcomini *et al.* 2011].

A mamografia constitui uma forma particular de radiografia, que trabalha com níveis de tensões e correntes em intervalos específicos, destinada a registrar imagens da mama a fim de diagnosticar a presença ou ausência de estruturas que possam indicar doenças [Mitchell *et al.* 1988]. O seu principal papel é o rastreamento de mulheres assintomáticas para a possível detecção do câncer de mama em um tamanho e estágio inicial, antes mesmo da própria observação da paciente ou do exame médico rotineiro.

A mamografia possui a vantagem de ser um método não invasivo, com imagens relativamente boas para um rastreamento e que leva um tempo relativamente pequeno para ser feito (Ribeiro 2006). As principais lesões identificadas na mamografia são as microcalcificações, um dos primeiros indícios de formação de tumores com grau elevado de suspeita de malignidade, e os nódulos, que são responsáveis por grande parte dos casos de câncer de mama (Ribeiro 2008a). Segundo Kopans (2000), os nódulos são encontrados em 39% dos casos de câncer não palpáveis. Na Figura 1 é ilustrado um exemplo de nódulo mamário.



Figura 1. Exemplo de um nódulo mamário. Fonte: [Ribeiro 2006]

1.2 Diagnóstico Auxiliado por Computador

Os esquemas CAD (*Computer-Aided Diagnosis*) têm como objetivo fornecer uma “segunda opinião” ao radiologista, auxiliando-o na detecção de lesões suspeitas em uma mamografia. Além de analisar objetivamente essas lesões, em função de suas características de benignidade e malignidade, ele também auxilia na formulação do diagnóstico médico. Desta forma, suas principais finalidades são aumentar a eficiência do diagnóstico e do laudo mamográfico, diminuir o número de erros no

diagnóstico (perdas ou biópsias desnecessárias) e, conseqüentemente, diminuir o número de mortes por câncer de mama [Chan *et al.* 1998, Giger 2004].

Entre as técnicas empregadas em um esquema CAD, um particular interesse remete às que buscam detectar e/ou classificar as microcalcificações [Jiang *et al.* 2007; Halkiotis *et al.* 2007; Jiang *et al.* 1998; Yu *et al.* 2006], nódulos e tumores [Schiabel *et al.* 2008; Ribeiro *et al.* 2007; Brzakovic *et al.* 1990; Younesi *et al.* 2007], além da utilização de técnicas de pré-processamento que visam aprimorar o contraste nas imagens mamográficas para melhorar a detecção das estruturas de interesse [Ram 1982; Nunes *et al.* 2002].

As principais etapas realizadas por um sistema CAD são a segmentação e a classificação. A etapa de segmentação de imagens refere-se ao processo de dividir a imagem em diversas partes elementares (objetos), permitindo a análise destas partes isoladamente [Gonzales e Woods 2000]. Após a segmentação, inicia-se a etapa de classificação dos objetos de interesse para o problema em questão. Para realizar esta classificação é necessária a extração e quantificação das características da imagem dos seus objetos constituintes [Ribeiro *et al.* 2007].

A classificação é formalmente definida como um processo pelo qual um padrão ou sinal recebido é atribuído a uma classe dentre um número predeterminado de classes (categorias) [Haykin 1999]. Muitas técnicas de classificação vêm sendo usadas em diversas aplicações, cada qual se adequando melhor para solucionar diferentes tipos de problemas. Especificamente em sistemas CAD para classificação de massas nodulares, as abordagens baseadas em Redes Neurais Artificiais, Sistemas Fuzzy e Redes Bayesianas vêm sendo utilizadas no diagnóstico do câncer de mama [Rose *et al.* 2010; Xie 2002].

1.3 Redes Neurais Artificiais

As Redes Neurais Artificiais (RNA) constituem-se em modelos computacionais inspirados na anatomia e funcionamento do cérebro humano. Estas redes são compostas por elementos processadores simples, chamados de neurônios artificiais, que imitam as mais elementares funções de um neurônio biológico e dos sistemas nervosos.

Os neurônios utilizados nos modelos de redes neurais realizam funções simples, como coletar os sinais existentes em suas entradas, agregá-los de acordo com sua função de entrada e produzir uma saída por intermédio de sua função de ativação inerente. Como exemplo de propriedades típicas do cérebro, as redes neurais aprendem a partir da experiência e generalizam novos exemplos a partir de exemplos prévios, abstraindo características essenciais de entrada [Haykin 1999].

Uma RNA realiza o reconhecimento de padrões passando inicialmente por uma seção de treinamento, durante a qual se apresenta repetidamente à rede um conjunto de padrões de entrada junto com a categoria à qual cada padrão particular pertence. Mais tarde, apresenta-se à rede um novo padrão que não foi visto antes, mas que pertence à mesma população de padrões utilizada para treinar a rede. A rede é capaz de identificar a classe daquele padrão particular por causa da informação que ela extraiu dos dados de treinamento [Rezende 2003].

1.3.1 Redes de Kohonen

As Redes de Kohonen, também conhecidas como Mapas Auto-Organizáveis de Kohonen (*Self-Organization Maps - SOM*), são redes que possuem uma estrutura reticulada de uma única camada, as quais são treinadas por um algoritmo competitivo [Kohonen 1984 *apud* Silva *et al.* 2010]. O processo de competição é baseado no critério de proximidade, ou seja, aquele neurônio que estiver mais próximo de determinado padrão x vencerá a competição [Silva *et al.* 2010]. Este neurônio é chamado de neurônio vencedor.

Na rede de Kohonen os neurônios se organizam em uma grade ou reticulado, geralmente bidimensional, podendo algumas vezes ser unidimensional [Braga *et al.* 2007]. Esta organização é definida pelo Mapa Topológico. Na Figura 2 é ilustrado um exemplo de mapa topológico bidimensional com $N=3$, onde o número de neurônio é igual a $N.N$.

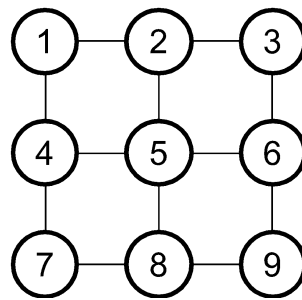


Figura 2. Mapa topológico da rede de Kohonen com $N=3$

Adicionalmente, as Redes de Kohonen possuem conexões laterais de acordo como o Mapa Topológico de Vizinhança. As conexões laterais modelam a competitividade entre os neurônios, onde apenas um será vencedor. A saída do neurônio vencedor pode influenciar ou colaborar com os seus neurônios vizinhos. Para os casos onde os neurônios estão mais próximos aos neurônios vencedores, o potencial dessas conexões laterais são maiores (excitatório). Por outro lado, quanto mais distantes dos neurônios vencedores, menor será o potencial destas conexões [Silva *et al.* 2010].

De acordo com Braga *et al.* (2007), a vizinhança de uma rede SOM pode assumir vários formatos, como por exemplo, quadrado, hexágono ou círculo (gaussiana), e a definição do formato dependerá do problema e da distribuição dos seus dados.

O principal objetivo de um SOM é mapear um conjunto de padrões de entrada, sendo cada um deles N -dimensional, para um mapeamento que pode ser representado em uma ou duas dimensões. A sua estrutura de configuração simples, assim como sua dinâmica de treinamento diferenciada, faz das redes de Kohonen uma ferramenta sofisticada para aplicações em problemas que envolvem classificação de padrões e identificação de agrupamentos de dados (*clustering*) [Silva *et al.* 2010].

MATERIAIS E MÉTODOS

O trabalho desenvolvido é baseado na rede neural auto-organizável proposto pelo finlandês Teuvo Kohonen. Os dados de treinamentos e testes da rede foram obtidos da base de dados Breast Cancer Wisconsin Original (BCW-Original) e Breast Cancer Wisconsin Diagnostic (BCW-Diagnostic), ambas do repositório UCI Machine Learning, da Universidade da Califórnia, criada por Frank e Asuncion (2010).

Essas bases de dados são bastante conhecidas no meio científico e já foram utilizada em outros trabalhos. As amostras, em ambas as bases de dados, foram coletadas a partir de casos clínicos do Dr. Wolberg, da Universidade de Wisconsin.

A base BCW-Original é composta por 699 amostras. Cada amostra representa as informações de um determinado nódulo mamário, a qual possui 11 atributos (características), sendo que um desses atributos é um identificador da amostra, e outro é informa a qual classe a amostra pertence (maligna ou benigna), ambas são desnecessária para a composição dos dados de entrada do classificador neural.

Algumas das amostras da base BCW-Original não foram utilizados, pois não possuem todos os nove atributos necessários, sendo assim o número total de amostras consideradas dessa base é 683, que é o valor equivalente ao número de amostras disponíveis na base de dados menos o número de amostras que falta algum atributo. As amostras dessa base de dados foram divididas para compor os arquivos de treinamento e de teste da rede neural. A maneira como essas amostras foram divididas é mostrada na Tabela 1.

	Número de amostras	Percentual
Total de amostras	683	100% de 683
Total de amostras malignas	239	35% de 683
Total de amostras benignas	444	65% de 683
Total de amostras do arquivo de treinamento	478	70% de 683
Total de amostra malignas no arquivo de treinamento	167	35% de 478
Total de amostra benignas no arquivo de treinamento	311	65% de 478
Total de amostras do arquivo de teste	205	30% de 683
Total de amostra malignas no arquivo de teste	72	35% de 205
Total de amostra benignas no arquivo de teste	133	65% de 205

Tabela 1. Composição dos arquivos de treinamento e teste para a base BCW-Original

A outra base de dados possui 559 amostras e 32 atributos. A maneira como essa base de dados foi dividida para compor os arquivos de treinamento e de teste da rede neural é mostrada na Tabela 2.

	Número de amostras	Percentual
Total de amostras	569	100% de 569
Total de amostras malignas	212	37% de 569
Total de amostras benignas	357	63% de 569
Total de amostras do arquivo de treinamento	398	70% de 569
Total de amostra malignas no arquivo de treinamento	148	37% de 398
Total de amostra benignas no arquivo de treinamento	250	63% de 398
Total de amostras do arquivo de teste	171	30% de 569
Total de amostra malignas no arquivo de teste	64	37% de 171
Total de amostra benignas no arquivo de teste	107	63% de 171

Tabela 2. Composição dos arquivos de treinamento e teste para a base BCW-Diagnostic

Como pode ser observado nas tabelas 1 e 2 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 Vizinhaça usado neste trabalho é uma função gaussiana, a qual é definida pela Equação 1 [Haykin 1999]. Na Figura 3 está ilustrada a função de vizinhaça gaussiana.

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

Equação 1. Função gaussiana

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 vizinhaça topológica, o qual mede o grau que os neurônios excitados na vizinhaça do neurônio vencedor participam do processo de aprendizagem [Haykin 1999].

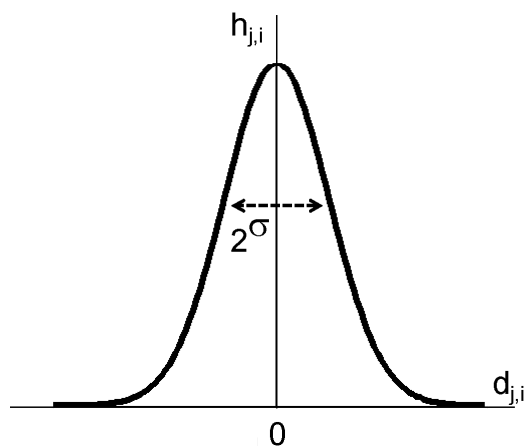


Figura 3. Função gaussiana de vizinhança

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* que, apesar de ser um algoritmo simples, satisfaz o problema, pois como serão utilizadas as bases 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. Sendo o conhecimento da quantidade de grupos uma condição necessária para o funcionamento do K-Means.

A análise utilizando o K-Means foi feita com o uso de recursos da ferramenta Matlab. Os demais recursos foram implementados e adicionados ao sistema desenvolvido.

Na etapa de teste é feita uma classificação das amostras presentes no arquivo de teste, com base nos grupos formados pelo K-Means. Cada resultado apresentado pelo sistema foi comparado à classificação da amostra que consta na base da qual ela foi obtida. Se os resultados são iguais, então o sistema conseguiu classificar corretamente a amostra. O número de acertos e erros são utilizados para calcular as taxas de verdadeiros positivos e de falsos positivos. Os verdadeiros positivos equivalem às amostras cancerígenas que foram classificadas corretamente. Os falsos positivos equivalem às amostras cancerígenas que foram classificadas como benignas.

RESULTADOS E DISCUSSÃO

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 contendo entre 4 e 81 neurônios, para a base BCW-Original, e entre 4 e 400 neurônios, para a base BCW-Diagnostic.

Cada topologia foi testada cinco vezes. Em cada execução, foram calculadas a taxa de verdadeiros positivos e a taxa de falsos positivos. Os valores obtidos com o cálculo de ambas as taxas são utilizados para encontrar a média das taxas de verdadeiros positivos assim como a de falsos positivos.

Para analisar as variações entre os resultados apresentados entre as cinco execuções de cada topologia, também foram calculados os desvios padrões, baseados na média das taxas de verdadeiros positivos e de falsos positivos.

A Tabela 3 mostra os resultados apresentados pela rede utilizando a base de dados BCW-Original.

Topologia	Verdadeiros Positivos	Falsos Positivos	Desvio Padrão (Verdadeiros Positivos)	Desvio Padrão (Falsos Positivos)
2 x 2	0.9722222	0.02255639	0.0	0.0
3 x 3	0.9722222	0.007518797	0.0	0.0
4 x 4	0.9583333	0.0	0.0	0.0
5 x 5	1.0	0.015037594	0.0	0.0
6 x 6	1.0	0.015037594	0.0	0.0
7 x 7	0.98611104	0.0030075188	4.712160915387242E-8	0.002912017626764159
8 x 8	0.99722224	0.0060150377	0.0043920544001236	0.002377652435854208
9 x 9	0.98611104	0.0	4.712160915387242E-8	0.0

Tabela 3. Saída do sistema para a base BCW-Original

Respeitando os critérios adotados para a escolha da melhor topologia, foi definido que a topologia que obteve o melhor desempenho, em relação à base de dados BCW-Original, foi a 5x5, a qual atingiu a pontuação máxima na taxa de verdadeiros positivos, com uma taxa de falsos positivos de apenas 0.015, e desvio padrão igual a zero.

A Tabela 4 mostra os resultados apresentados pela rede utilizando a base de dados BCW-Diagnostic.

Topologia	Verdadeiros Positivos	Falsos Positivos	Desvio Padrão (Verdadeiros Positivos)	Desvio Padrão (Falsos Positivos)
2 x 2	0.9583333	0.41353384	0.0	0.0
3 x 3	0.8194445	0.082706764	3.057653761163813E-8	0.0
4 x 4	0.7777778	0.04511278	0.0	0.0
5 x 5	0.75	0.007518797	0.0	0.0
6 x 6	0.7916667	0.030075189	0.0	0.0
7 x 7	0.7916667	0.0	0.0	0.0
8 x 8	0.8333333	0.0	0.0	0.0
9 x 9	0.86111104	0.007518797	3.057653761163813E-8	0.0
10 x 10	0.8333333	0.0	0.0	0.0

11 x 11	0.8194445	0.015037594	3.057653761163813E-8	0.0
12 x 12	0.8472222	0.007518797	0.0	0.0
13 x 13	0.7916667	0.015037594	0.0	0.0
14 x 14	0.875	0.007518797	0.0	0.0
15 x 15	0.8333333	0.007518797	0.0	0.0
16 x 16	0.88888896	0.02255639	3.057653761163813E-8	0.0
17 x 17	0.8444444	0.0015037594	0.0028499412259194823	0.0015428246283971546
18 x 18	0.8	0.0	0.005699882451724147	0.0
19 x 19	0.86111104	0.037593983	3.057653761163813E-8	0.0
20 x 20	0.8055555	0.004511278	3.057653761163813E-8	0.004628473694088103

Tabela 4. Saída do sistema para a base BCW-Diagnostic

De acordo com os resultados apresentados na Tabela 4, a topologia que apresenta os melhores resultados é a 14x14. A taxa de verdadeiros positivos dessa topologia, apesar de atingir um valor satisfatório, não foi a maior encontrada, entretanto está acompanhada de uma taxa de falsos positivos baixa, além de desvio padrão zero, para as duas taxas, o que faz da topologia 14x14 ideal para a classificação da base BCW-Diagnostic.

CONCLUSÃO

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 das bases Breast Cancer Wisconsin Original (BCW-Original) e Breast Cancer Wisconsin Diagnostic (BCW-Diagnostic), do repositório UCI Machine Learning, da Universidade da Califórnia.

Para a base BCW-Original o sistema chegou a atingir valor 1,0 para taxa de verdadeiros positivos (valor máximo da escala), e uma taxa de falsos positivos inferior a 0.016.

O melhor resultado apresentado pelo sistema, considerando a base BCW-Diagnostic, foi de uma taxa de verdadeiros positivos igual a 0.875, com taxa de falsos positivos inferior a 0.0076.

Os resultados observados demonstram alto grau de confiabilidade e estabilidade do sistema. Sendo assim, o sistema proposto pode ser utilizado como uma segunda opinião para especialistas em análise de imagens mamográficas, contribuindo para uma melhor interpretação das imagens e consequentemente, proporcionar um diagnóstico mais confiável ao paciente, evitando a realização de biópsias desnecessárias e aumentando as chances de identificação precoce do câncer de mama.

AGRADECIMENTOS

Agradecemos ao Programa de Bolsas de Iniciação Científica da PIBIC-UEFS por financiar este trabalho, contribuindo para o desenvolvimento da pesquisa na área de inteligência artificial e permitindo a inserção de alunos de graduação em projetos de pesquisa científica.

REFERÊNCIAS BIBLIOGRÁFICAS

- Ball, J.E., T.W. Butler, and 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.
- Biliska-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.
- Braga, A. P.; Carvalho, A. P. L. F.; Ludermir, T. B. (2007) *Redes Neurais Artificiais: Teoria e Aplicações*, 2ª ed. LTC.
- Brzakovic, D., X.M. Luo, and P. Brzakovic. (1990). An approach to automated detection of tumors in mammograms. *IEEE Transactions on Medical Imaging*. 9(3): p. 233-241.
- Chan, H.P., et al. (1998). Computer-aided detection of microcalcifications in mammograms: methodology and preliminary clinical study. *Investigative Radiology*. 23(9): p. 664-671.
- Frank, A. e 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.
- Giger, M. L. (2004). Computerized analysis of images in the detection and diagnosis of breast cancer. *Seminars in Ultrasound, CT and MRI*. 25: p. 411-418.
- Gonzales, R.C. and R.E. Woods. (2000). *Processamento de imagens digitais*. 1ª ed. Edgard Blücher.
- Halkiotis, S., T. Botsis, e M. Rangoussi. (2007). Automatic detection of clustered microcalcifications in digital mammograms using mathematical morphology and neural networks. *Signal Processing*. 87(7): p. 1559-1568.
- Haykin, S. (1999). *Neural Networks - A Comprehensive Foundation*. 2ª ed. New York: Prentice-Hall.
- 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.
- Jiang, Y., R. M. Nishikawa, and J. Papaioannou. (1998). Requirement of microcalcification detection for computerized classification of malignant and benign clustered microcalcifications. in SPIE.
- Jiang, J., B. Yao, e A.M. Wason. (2007). A genetic algorithm design for microcalcification detection and classification in digital mammograms. *Journal of the Computerized Medical Imaging and Graphics*. 31(1): p. 49-61.
- Kopans, D.B. (1992). The positive predictive value of mammography. *American Journal of Roentgenology*. vol.158, nro.3, p.521-526.
- Kopans, D.B. (2000). *Imagem da Mama*. in MEDSI. Rio de Janeiro.
- 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.
- Mitchell JR., G.W. and L.W. Basset. (1988). *Mastologia Prática*. Revinter.
- Nunes, F.L.S., H. Schiabel, and R.H. Benatti. (2002). Contrast enhancement in dense breast images using the modulation transfer function. *Medical Physics*. 29(12): p. 2925-2936.
- Ram, G. (1982). Optimization of ionizing radiation usage in medical imaging by means of image enhancement techniques. *Medical Physics*. vol.9, p.733-737.
- Rezende, S.O. (2003). *Sistemas Inteligentes - Fundamentos e Aplicações*. 1ª ed. Manoele.
- 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.
- Rose, D.C., et al. (2010). Applying Deep-Layered Clustering to Mammography Image Analytics. in *Biomedical Sciences and Engineering Conference (BSEC)*.
- Schiabel, H., V.T. Santos, and M.F. Angelo. (2008). Segmentation technique for detecting suspect masses in dense breast digitized images as a tool for mammography CAD schemes. in *Applied Computing*.
- Silva, I. N., Spatti, D.H., Flauzino, R.A. (2010). *Redes Neurais Artificiais: para engenharia e ciências aplicadas*. São Paulo: Artliber.
- Tahmoush, D. (2007) Image differencing approaches to medical image classification. In: 36th IEEE Applied Imagery Pattern Recognition Workshop - AIPR, Washington, p.22-27.

- Xie, M. (2002). A method of tumors detection in digital mammography. in International Conference on Communications, Circuits and Systems and West Sino Expositions.
- Younesi, F., et al. (2007). Computer-Aided Mass Detection on Digitized Mammograms Using Adaptive Thresholding and Fuzzy Entropy. in Engineering in Medicine and Biology Society. Lyon, França.
- Yu, S., K. Li, and Y. Huang. (2006). Detection of microcalcifications in digital mammograms using wavelet filter and Markov random field model. Computerized Medical Imaging and Graphics. 30: p. 163-173.

DECLARAÇÃO

Declaro que desenvolvi as atividades constantes do plano de trabalho do projeto **DIAGNÓSTICO DE MASSAS NODULARES EM IMAGENS MAMOGRÁFICAS UTILIZANDO REDES DE KOHONEN** submetido ao PIBIC/UEFS cujos resultados finais constam no presente relatório.

Feira de Santana, 10 de agosto de 2013.

Weverson dos Santos Gomes

DECLARAÇÃO

Declaro que acompanhei as atividades do bolsista WEVERSON DOS SANTOS GOMES do plano de trabalho do projeto DIAGNÓSTICO DE MASSAS NODULARES EM IMAGENS MAMOGRÁFICAS UTILIZANDO REDES DE KOHONEN submetido ao PIBIC/UEFS cujos resultados finais constam no presente relatório.

Feira de Santana, 10 de agosto de 2013.

Matheus Giovanni Pires