

Utilizando uma Rede Neural Artificial ARTMAP-Fuzzy para Realizar o Diagnóstico Clínico de Amostras de Câncer de Mama

James C. Silva[†] Fernando P. A. Lima[†] Mara L. M. Lopes^{*} Carlos R. Minussi[†]

[†]Departamento de Engenharia Elétrica, Faculdade de Engenharia de Ilha Solteira (FEIS)

^{*}Departamento de Matemática, Faculdade de Engenharia de Ilha Solteira (FEIS)

Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), Ilha Solteira, SP, BRASIL

E-mails: jamesclauton@gmail.com, engfernandoparra@gmail.com,

mara@mat.feis.unesp.br, minussi@dee.feis.unesp.br

Palavras-chave: *Diagnóstico, Câncer de Mama, Redes Neurais Artificiais, ARTMAP-Fuzzy, WBCD.*

Resumo: *Neste artigo apresenta-se um método para o diagnóstico de amostras de câncer de mama utilizando uma rede neural artificial ARTMAP-Fuzzy, a qual é responsável por discriminar as amostras, obtendo uma classificação em casos benignos ou malignos. A principal aplicação desta metodologia é auxiliar os profissionais no processo de diagnóstico de câncer de mama, visando rapidez na tomada de decisão, confiabilidade e a assistência necessária para salvar vidas. Para avaliar este método, utilizou-se a base de dados Wisconsin Breast Cancer Diagnosis (WBCD), que é composta por dados reais de câncer de mama. Os resultados obtidos pelo método são considerados satisfatórios comparados com a literatura.*

1 Introdução

O câncer é uma doença crônica que atinge milhões de pessoas em todo o mundo. Segundo um levantamento realizado pelo Instituto Nacional de Câncer (INCA) no ano de 2012, evidenciou-se que o câncer de mama é o segundo mais frequente no mundo e o mais comum entre as mulheres, respondendo por 22% dos casos novos a cada ano. No ano de 2012 também foram diagnosticados 52.680 novos casos de câncer de mama no Brasil [3]. O câncer de mama é o que mais leva os pacientes a óbito, aproximadamente 39% das mulheres que lutam contra o câncer de mama, morrem [9].

O diagnóstico correto em um estado prematuro do câncer de mama pode auxiliar na tomada de decisões, no planejamento de ações e, evidentemente, na eficiência do tratamento. No entanto, diagnosticar corretamente o câncer é um processo complexo e muito difícil, em consequência das diversas variáveis envolvidas. Para um diagnóstico preciso, exige-se muita experiência e, principalmente, que a classificação do estadiamento clínico do tumor (estágio do câncer) esteja correta.

Os tradicionais sistemas de classificação de patologias utilizados pelos médicos são complexos, e normalmente de difícil utilização, oferecendo muitas limitações [6]. Desta forma, torna-se necessário o desenvolvimento de sistemas integrados que trabalhem com técnicas de processamento e análise de dados, e que, combinados com a experiência dos profissionais, proporcione a assistência necessária para realizar o diagnóstico preciso da doença.

Neste sentido, a aplicação de métodos baseados em computação inteligente (CI) torna-se uma possível solução para o problema de diagnóstico. Os métodos inteligentes são capazes de extrair informações e conhecimento de problemas complexos, e são de fácil aplicação. Na sequência apresentam-se os trabalhos mais relevantes sobre diagnóstico de câncer de mama disponível na literatura.

No artigo [10], os autores apresentam um método para diagnóstico de amostras de câncer utilizando a lógica *fuzzy* e o algoritmo genético. Em [11] foi usada uma rede ANFIS para realizar o diagnóstico de câncer de mama. No artigo [8] foi utilizado um sistema híbrido com uma rede neural e um sistema especialista *fuzzy*. Em [4] os autores apresentam um método

baseado no algoritmo de seleção negativa para o diagnóstico de câncer de mama. Os autores citados fizeram uma série de testes e obtiveram bons resultados.

Neste artigo, apresenta-se um método para diagnóstico de câncer de mama utilizando uma rede neural artificial baseada na teoria da ressonância adaptativa, denominada rede ARTMAP-Fuzzy. Optou-se por empregar tal arquitetura pela sua estabilidade, plasticidade e eficiência em relação as demais redes, pois além de obter um treinamento rápido em relação a maioria das outras redes neurais. O processo de treinamento é realizado instantaneamente. Esta arquitetura permite a inclusão do módulo de treinamento continuado, ou seja, podem-se inserir novos padrões sem a necessidade de reiniciar o treinamento. Desta forma, o principal objetivo é aplicar a rede ARTMAP-Fuzzy nas amostras cancerígenas da base de dados Wisconsin Breast Cancer Diagnosis[12] [13], que contém dados reais sobre o câncer de mama.

2 Base de Dados Wisconsin Breast Cancer Diagnosis (WBCD)

A base de dados WBCD foi criada pelo Dr. William H. Wolberg em seu escritório, entre os anos de 1989 e 1991 no hospital escola da Universidade de Wisconsin Madison. A base de dados contém 9 instâncias representando as características do tumor e, evidentemente, a classificação para estas instâncias, totalizando 10 variáveis [12].

As características armazenadas na base de dados são as seguintes [5]:

1. espessura da massa celular (CT);
2. uniformidade do tamanho da célula (CS);
3. uniformidade do formato da célula (CH);
4. adesão marginal (AD);
5. tamanho de uma célula epitelial (EP);
6. núcleo vazio (BN);
7. cromatina branda (CO);
8. nucléolo normal (NN);
9. mitose (MM);
10. classificação (“benigno” ou “maligno”).

Esta base de dados possui 699 amostras, sendo que 65% representam tumores benignos e 35% representam tumores malignos [13].

3 ARTMAP-Fuzzy

A rede neural ARTMAP-Fuzzy corresponde a um sistema de aprendizado supervisionado composto por um par de módulos da teoria da ressonância adaptativa, ARTa-Fuzzy e ARTb-Fuzzy, conectados pelo módulo de memória associativa inter-ART. Esta rede incorpora a teoria dos conjuntos fuzzy, operador fuzzy AND (\wedge), capacitando o aprendizado do sistema neural em resposta a padrões de entrada binários e analógicos. A Figura 1 ilustra a arquitetura da rede neural ARTMAP-Fuzzy [2].

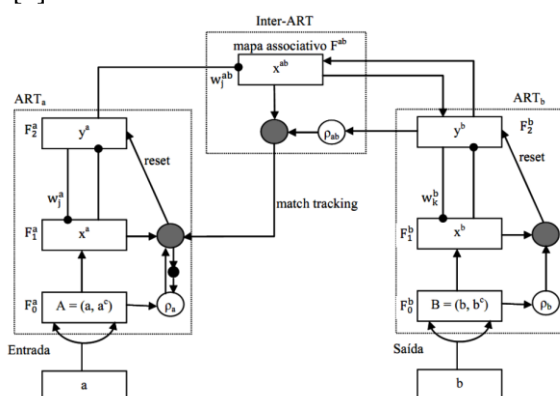


Figura 1 – Arquitetura da rede neural ARTMAP-Fuzzy [2].

Um mecanismo interno denominado *match tracking* é responsável pelo processo autorregulador da rede, no qual é maximizada a generalização e minimizado o erro. Sempre que a rede realiza um prognóstico errado, através de uma conexão associativa instruída, o parâmetro

de vigilância do módulo ARTa-*Fuzzy* é incrementado em uma quantidade mínima necessária para corrigir o erro no módulo ARTb-*Fuzzy*.

Existem três parâmetros fundamentais no desenvolvimento da rede neural ARTMAP-*Fuzzy* [2], são eles:

- Parâmetro de escolha α ($\alpha > 0$) : atua na seleção de categorias;
- Taxa de treinamento β ($\beta \in [0 1]$) : controla a velocidade da adaptação da rede;
- Parâmetro de vigilância ρ ($\rho \in [0 1]$) : controla a ressonância da rede. Trata-se do parâmetro responsável pelo número de categorias criadas.

Se ρ possui um valor grande, a rede torna-se mais seletiva reduzindo a capacidade de generalização. Caso ρ tenha um valor pequeno, reduz-se o número de categorias criadas, aumentando a capacidade de generalização.

4 Metodologia Proposta

O sistema de diagnóstico de câncer de mama é baseado em uma rede neural artificial ARTMAP-*Fuzzy*. Este sistema é o responsável por realizar a discriminação das amostras de câncer, entre benigno e maligno.

O método proposto é executado em duas fases: treinamento e o diagnóstico dos dados. Na sequência apresenta-se o conjunto de testes e o esquema de funcionamento da rede ARTMAP-*Fuzzy* no diagnóstico de câncer de mama.

4.1 Conjunto de testes

A base WBCD possui 10 atributos, no entanto, para este artigo optou-se por utilizar 5 atributos. Para escolher os atributos, utilizou-se o cálculo do desvio padrão das amostras. Foram escolhidos os 5 atributos que apresentaram os índices mais baixos de desvio padrão, isto é, os dados mais homogêneos. Esta estratégia foi adotada visando maior confiabilidade.

O desvio padrão foi calculado utilizando a Equação (1):

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (z_i - \bar{z})^2} \quad (1)$$

em que n é o número total de amostras, z são as amostras envolvidas e \bar{z} é a média amostral.

Os atributos selecionados pela análise de menor desvio são: a espessura da massa celular (CT), a uniformidade do tamanho da célula (CS), a uniformidade do formato da célula (CH), o núcleo vazio (BN) e o nucléolo normal (NN). Durante a separação destas variáveis observou-se que alguns dados não são aproveitáveis, ou seja, o desvio padrão foi muito próximo de zero.

Tabela 1 – Especificação da base de dados WBCD.

Base de dados WBCD	
Tipo	Classificação
Número de dados	699
Número de dados aproveitáveis	682
Dados da classe “benigno”	444
Dados da classe “maligno”	238
Número de atributos	5

Na Tabela 1 ilustra-se a especificação da base de dados WBCD usada no treinamento e diagnóstico da rede neural ARTMAP-*Fuzzy*.

4.2 Sistema de Diagnóstico

O sistema de diagnóstico de câncer de mama proposto neste artigo é baseado na rede ARTMAP-*Fuzzy*. O esquema de funcionamento deste sistema é ilustrado no diagrama de blocos apresentado na Figura 2.

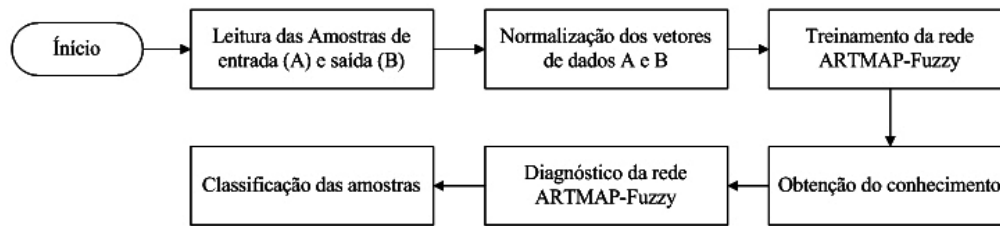


Figura 2 – Esquema de funcionamento do sistema de diagnóstico.

O módulo de leitura das amostras é responsável por fazer a aquisição dos dados WBCD. O conjunto de entrada de dados da rede neural ARTMAP-Fuzzy é composto pelos atributos da amostra de câncer de mama da base de dados WBCD (espessura da massa celular, uniformidade do tamanho da célula, uniformidade do formato da célula, o núcleo vazio e nucléolo normal (NN)). O conjunto de saída é formado pela classificação do conjunto de entrada (amostras cancerígenas) em benigno ou maligno.

O módulo de treinamento da rede neural ARTMAP-Fuzzy é realizado de forma *off-line* e serve para ajustar os pesos e criar as categorias (conhecimento) que serão utilizadas no diagnóstico dos dados. Neste módulo, a rede é treinada considerando-se um conjunto de dados e será gerada uma quantidade de categorias, no máximo, igual ao número de amostras de entrada.

Desta forma a rede ARTMAP-FUZZY utiliza o parâmetro α para escolha da categoria, sendo assim responsável por determinar que a rede busque a menor categoria criada para esse padrão, já o β é importante para determinar a velocidade do treinamento, fazendo assim que a rede utilize mais tempo treinando um determinado padrão a ser utilizado na categoria. E o parâmetro ρ será utilizado para a criação das categorias, ou seja, à partir dele é que se determina a quantidade de categorias e se um padrão é semelhante o bastante para que seja incluso em uma categoria já existe.

O módulo de diagnóstico é realizado de forma *on-line*. Neste fase, um novo conjunto de amostras, diferente do conjunto de dados utilizados na fase de treinamento, deve ser apresentado à rede. Cada amostra será analisada individualmente. Todas as amostras são comparadas com o conhecimento (categorias criadas na fase de treinamento) extraído pela rede neural ARTMAP-Fuzzy. Assim, quando houver um casamento entre os padrões classifica-se em categoria a amostra apresentada a rede ARTMAP-Fuzzy. Obtida a categoria correspondente a um determinado padrão é, portanto, identificado o diagnóstico, que é a classificação da amostra cancerígena como sendo benigna ou maligna.

5 Resultados

Os testes (simulações) foram realizados utilizando um PC Intel Core 2 Duo 1.9 GHz, 2 GB de Memória RAM, e sistema operacional Windows 7 Ultimate 32 bits. O algoritmo foi desenvolvido em MATLAB® [7].

Os parâmetros da rede neural ARTMAP-Fuzzy utilizados no processo de treinamento e diagnóstico foram: $\alpha = 0,1$; $\beta = 1$ (treinamento rápido); $\rho_a = 0,8$; $\rho_b = 1$ e $\rho_{ab} = 1$.

Na fase de treinamento foi utilizado um conjunto de dados com 341 amostras, sendo 222 amostras em condições normais (benigno) e 119 amostras cancerígenas (maligno). Este conjunto de dados corresponde a 50% dos dados disponíveis. Para avaliar a rede neural ARTMAP-Fuzzy na fase de diagnóstico foi utilizado um conjunto de dados com 341 amostras, sendo 222 amostras benignas e 119 amostras malignas. Estes dados utilizados nos testes representam 50% dos dados disponíveis. A base de dados simulada conta com 682 amostras, resumidas em 2 padrões diferentes, benigno e maligno.

A Equação (2) fornece o cálculo para a análise do percentual de acerto obtido pelo diagnóstico da rede neural ARTMAP-Fuzzy.

$$acertos = \frac{t * 100}{t + f} \tag{2}$$

em que t representa o total de acertos e f o total de erros.

Na Tabela 2 apresentam-se os resultados obtidos pelo sistema de diagnóstico de câncer de mama aplicados à base de dados WBCD.

Tabela 2 - Resultados obtidos pelo método.

	Fase de Treinamento	Fase de Diagnóstico
Amostras Utilizadas	341	341
Classificações Corretas	341	322
Classificações Erradas	0	19
Acerto (%)	100,00	94,43
Tempo (s)	4,6	1,07

Para obtenção destes resultados, o sistema foi submetido a uma fase de testes e ajustes de parâmetros. Os resultados apresentados foram obtidos com a melhor configuração da rede neural ARTMAP-Fuzzy. Foram analisados dois parâmetros, a sensibilidade (S) e a especificidade (E) do método que são dadas pelas seguintes equações:

$$S = \frac{VP}{VP+FP} \tag{3}$$

$$E = \frac{VN}{FN+VN} \tag{4}$$

sendo:

VP : verdadeiro positivo; VN : verdadeiro negativo; FP : falso positivo; FN : falso negativo.

Foram realizados 25 testes com variações nos valores de cada parâmetro, obtendo, assim, sua melhor configuração. Ao variar o parâmetro α em 0,2 entre 0,1 e 1, $\rho_a, \rho_b, \rho_{ab}$ em 0,05 entre 0,8 e 1 não se obteve mudança nos valores de sensibilidade e especificidade. Porém, quando o parâmetro β é variado ocorre uma diferença muito grande no valor de especificidade, ficando entre 35% a 70%. Como consequência desta análise tem-se que algumas pessoas sem a doença estariam sendo indicadas para a realização de exames mais específicos.

Para β com valor igual a 1, a rede neural artificial ARTMAP-Fuzzy obtêm seu melhor resultado, atingindo um valor de especificidade de 91,44%, neste caso, somente 8,56% dos indivíduos sem a doença estariam sendo recomendados para um exame mais detalhado. Para os testes realizados, o valor de sensibilidade obtido foi de 100%, indicando que todos os indivíduos que possuem a doença seriam submetidos a exames mais detalhados, e consequentemente, ao tratamento.

Neste sentido, a rede neural ARTMAP-Fuzzy apresentou um bom desempenho, ou seja, nenhum paciente ficou sem assistência medica, caso fosse indicado, por este método, a suspeita de malignidade.

6 Conclusão

Neste artigo foi apresentado um método de diagnóstico de câncer de mama utilizando uma rede neural artificial ARTMAP-Fuzzy. O método proposto apresentou bons resultados obtendo um índice de acerto superior a 94% para a melhor configuração da rede. O diagnóstico obtido através da rede neural ARTMAP-Fuzzy não deixou de detectar nenhum caso de câncer, isto é muito importante, do ponto de vista de salvar vidas. A rede ARTMAP-Fuzzy apresentou um processamento rápido, com tempo inferior a 100 milésimos de segundo, o que credencia o algoritmo a ser empregado em tempo real, na tomada de decisões. Desta forma, conclui-se que

a rede neural artificial ARTMAP-*Fuzzy* usada neste artigo é bastante eficiente, confiável, robusta e precisa, quando aplicada no diagnóstico de amostras câncer de mama. Vale ressaltar que o método apresentado neste trabalho é puro, não utilizando outras técnicas em conjunto, diferentemente da maioria dos trabalhos disponíveis na literatura.

Como trabalhos futuros, pretende-se analisar outros tipos de doenças, empregando a mesma metodologia, baseada na rede neural ARTMAP-Fuzzy; pretende-se também aprimorar esta técnica, compondo uma metodologia híbrida, utilizando sistemas imunológicos artificiais e lógica fuzzy.

Agradecimentos

Os autores agradecem a CAPES e a FAPESP (Proc. N° 2011/06394-5) pelo apoio financeiro de pesquisa e ao Centro de Aprendizado de Máquina e Sistemas Inteligentes da Universidade da Califórnia, (UCI – Machine Learning Repository) por disponibilizar a base de dados WBCD.

Referências

- [1] F. Camastra. Kernel Methods For Clustering. In WIRN/NAIS, volume 3931 of Lecture Notes in Computer Science, pp. 1–9 (2006).
- [2] G. A. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds, D. B. Rosen, “Fuzzy ARTMAP: A Neural Network Architecture for Incremental Supervised Learning of Analog Multidimensional Maps”, IEEE Trans. on Neural Networks, Vol. 3, No. 5, pp. 698-713 (Sept. 1992).
- [3] INCA - Instituto Nacional do Câncer (Brasil), disponível em: www.inca.gov.br.
- [4] F. P. A. Lima; M. L. M. Lopes; A. D. P. Lotufo e C. R. Minussi. Diagnóstico Clínico de Amostras de Câncer de Mama utilizando Sistemas Imunológicos Artificiais com Seleção Negativa, SBAI – Simpósio Brasileiro de Automação Inteligente, Fortaleza, pp. 1-6 (2013).
- [5] O. L. Mangasarian; R. Setiono and W. H. Wolberg. Pattern recognition via linear programming: Theory and application to medical diagnosis, in: "Large-scale numerical optimization", Thomas F. Coleman e Yuying Li, editores, Publicações SIAM, Philadelphia 1990, pp. 22-30 (1990).
- [6] K. Manikantan; S. I. Sayed; K. N. Syrigos; P. Rhys-Evans; C. M. Nutting; K. J. Harrington and R. Kazi. Challenges For The Future Modifications of The TNM: Staging System For Head And Neck Cancer. Cancer Treatment Reviews, pp. 35-44 (2009).
- [7] Matlab (2011). 7.8 Version, Mathworks Company.
- [8] P. Meesad and G. G. Yen. Combined Numerical and Linguistic Knowledge Representation and Its Application to Medical Diagnosis, IEEE Transactions on Systems, Man, and Cybernetics- Part A: Systems And Humans, Vol. 33, No. 2, pp206-222 (March 2003).
- [9] OMS - Organização Mundial da Saúde, Disponível em: <http://www.who.int/en/>.
- [10] C. A. Peña-Reyes and M. Sipper. Designing Breast Cancer Diagnostic System via Hybrid Fuzzy-Genetic Methodology, 1999 IEEE International Fuzzy Systems Conference Proceeding, Seoul-Korea, pp. I135-I139.(August 1999).
- [11] J. Wang and C. S. George Lee. Self-Adaptive Neuro-Fuzzy Inference Systems for Classification Applications, IEEE Transactions on Fuzzy Systems, pp. 790-802 (2002).
- [12] WBCD, (2012). Wisconsin Breast Cancer Data – UCI Machine Learning Repository, disponível em: www.arquivos.ics.uci.edu/ml/
- [13] W. H. Wolberg and O. L. Mangasarian. Multisurface Method of Pattern Separation For Medical Diagnosis Applied to Breast Cytology, Proceedings of National Academy of Sciences (USA), Vol. 87, pp. 9.193-9.196 (1990).