

## DIAGNÓSTICO CLÍNICO DE AMOSTRAS DE CÂNCER DE MAMA UTILIZANDO SISTEMAS IMUNOLÓGICOS ARTIFICIAIS COM SELEÇÃO NEGATIVA

FERNANDO P. A. LIMA<sup>1</sup>, MARA L. M. LOPES<sup>2</sup>, ANNA D. P. LOTUFO<sup>1</sup>, CARLOS R. MINUSSI<sup>1</sup>

<sup>1</sup> Departamento de Engenharia Elétrica, Faculdade de Engenharia de Ilha Solteira (FEIS),  
Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), Caixa postal 31, 15385-000,  
Ilha Solteira, SP, BRASIL

<sup>2</sup> Departamento de Matemática, Faculdade de Engenharia de Ilha Solteira (FEIS),  
Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), Caixa postal 31, 15385-000,  
Ilha Solteira, SP, BRASIL

E-mails: fernandoparra@aluno.feis.unesp.br, mara@mat.feis.unesp.br,  
annadiva@dee.feis.unesp.br, minuissi@dee.feis.unesp.br

**Abstract**— This paper uses an artificial immune systems applied for diagnosing breast cancer samples. Taking as basis an immunological process, the Negative Selection Algorithm was used to discriminate the samples, attaining a classification for benign or malignant cases. The main application of the methodology is to assist professionals in the breast cancer diagnostic process, thereby providing decision-making agility, efficient treatment planning, reliability and the necessary intervention to save lives. To evaluate this method, the Wisconsin Breast Cancer Diagnosis database (WBCD) was used. This is an actual breast cancer database. The results obtained using the method, when compared with the specialized literature, show accuracy, robustness and efficiency in the breast cancer diagnostic process.

**Keywords**— Filter, Abnormality detection, Electrical distribution systems, Artificial immune systems, Voltage disturbance, Smart Grids.

**Resumo**— Este artigo apresenta uma aplicação dos sistemas imunológicos artificiais para o diagnóstico de amostras de câncer de mama. Tomando-se como base um processo imunológico, utiliza-se o Algoritmo de Seleção Negativa para discriminar as amostras, obtendo uma classificação em casos benignos ou malignos. A principal aplicação para esta metodologia é auxiliar profissionais no processo de diagnóstico de câncer de mama, proporcionando rapidez na tomada de decisão, eficiência no planejamento de tratamentos, confiabilidade e a assistência necessária para salvar vidas. Para avaliar este método, utilizou-se a base de dados Wisconsin Breast Cancer Diagnosis (WBCD). Trata-se de uma base de dados real de câncer de mama. Os resultados obtidos pelo método, quando comparado com a literatura especializada mostram precisão, robustez e eficiência no processo de diagnóstico de câncer de mama.

**Palavras-chave**— Diagnóstico, Câncer de mama, Sistemas Imunológicos Artificiais, Algoritmo de Seleção Negativa, WBCD.

### 1 Introdução

Há algumas décadas observa-se um crescimento significativo na intensidade de novos casos de câncer de mama em todo o mundo, este fato se deve devido ao aumento da expectativa de vida da população. O câncer é uma doença crônica que atinge milhões de pessoas em todo o mundo. Este tipo de câncer ocorre principalmente em países de médio e baixo desenvolvimento, onde a prevalência de câncer se torna um problema de saúde pública, e exige maior vigilância, condutas de prevenção, políticas de conscientização, e tratamentos efetivos.

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 (INCA, 2012).

A taxa de mortalidade por câncer de mama é muito elevada no mundo todo, muito provavelmente porque a doença é diagnosticada em estágios avançados. No Brasil no ano de 2012 foram registradas 12.852 mortes por câncer de mama, sendo 147 homens e 12.705 mulheres (INCA, 2012). Em outra estatística apresentada pela Organização Mundial de Saúde (OMS) observa-se que aproximadamente 39% das mulheres que lutam contra o câncer de mama, vão a óbito (OMS, 2012).

Dados estes fatos torna-se necessário adotar políticas de conscientização, prevenção e, principalmente, buscar realizar o diagnóstico do câncer de mama em estágio inicial. O diagnóstico correto em um estado prematuro da doença pode proporcionar decisões rápidas e auxiliar no planejamento de ações e, evidentemente, a eficiência no tratamento. Porém, diagnosticar corretamente o câncer é um processo complexo e muito difícil, pois envolve muitas variáveis. Para um diagnóstico correto é necessário muita experiência por parte do profissional e, principalmente, que a classificação do estadiamento clínico do tumor (estágio do câncer) esteja correta.

Em (Manikantan et al., 2009) destaca-se que os tradicionais sistemas de classificação utilizados pelos médicos são complexos, e normalmente de difícil utilização, oferecendo limitações aos patologistas, não permitindo rapidez na tomada de decisão. Assim, torna-se necessário o desenvolvimento de sistemas integrados que tenham capacidade de trabalhar com técnicas de processamento e análise de dados, e que, combinados com a experiência dos profissionais, proporcione a assistência necessária para realizar o diagnóstico e planejar o tratamento.

Neste sentido, a aplicação de sistemas inteligentes (SI) se torna 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 literatura encontram-se trabalhos relacionados onde se apresentam métodos inteligentes para auxiliar profissionais a realizar o diagnóstico de doenças, em especial profissionais com pouca experiência. Desta forma na sequência são apresentados alguns trabalhos sobre diagnóstico de câncer de mama que se destacam na literatura.

Na referência (Pena-Reyes e Sipper, 1999), os autores apresentam um método para diagnóstico de amostras de câncer utilizando a lógica fuzzy e o algoritmo genético. Em (Wang e George Lee, 2002) foi usada uma rede ANFIS (Jung, 1993) para realizar o diagnóstico de câncer de mama, os resultados foram considerados satisfatórios (96,30% de acerto). No artigo (Meesad e Yen, 2003) é proposto um sistema híbrido que utiliza uma rede neural em conjunto com um sistema especialista fuzzy. O sistema teve um bom desempenho quando aplicado na base de dados Wisconsin (WBCD, 2012).

Em (Zhao e Davis, 2011) os autores apresentam um método baseado no algoritmo de seleção clonal, onde é adicionada uma modificação através da utilização de um modelo de base radial do tipo mínimos quadráticos parciais para auxiliar no processo de geração de anticorpos. Os autores fizeram uma série de testes e obtiveram bons resultados, com um índice de acerto de 99,58%. Porém, dependendo do valor dado à largura da base radial o processo de classificação se comporta de maneira desordenada.

Os sistemas imunológicos artificiais (SIA) foram inspirados nos sistemas imunológicos biológicos e visam reproduzir, computacionalmente, suas principais características, propriedades e habilidades (Dasgupta, 1998). Os SIA são ferramentas adequadas para realizar diagnósticos, devido a habilidades de detecção de mudanças de comportamento em padrões (de Castro e Timmis, 2002).

Neste artigo, apresenta-se um método para diagnóstico de câncer de mama baseado nos sistemas imunológicos artificiais. A partir das amostras cancerígenas, aplica-se o algoritmo de seleção negativa (ASN) para diferenciar as amostras entre próprias (benignos) e não-próprias (onde há

evidência de malignidade). As amostras classificadas como próprias não representam riscos, isto é, não são prejudiciais ao organismo. As classificadas como não-próprias são amostras que necessitam de uma maior atenção, pois apresentam evidências de malignidade. Para avaliar o desempenho do método, foram realizados testes com uma base de dados real bem explorada na literatura. Sendo a base de dados Wisconsin Breast Cancer Diagnosis (WBCD, 2012).

Este artigo está organizado da seguinte forma. Na seção 2, apresenta-se a base de dados utilizada nos testes. O algoritmo de seleção negativa é abordado na Seção 3. A descrição da metodologia proposta é realizada na Seção 4. Nas Seções 5 e 6 encontram-se os resultados obtidos neste artigo e a conclusão, respectivamente.

## 2 Base de dados Wisconsin Breast Cancer Diagnosis (WBCD)

A base de dados WBCD foi criada pelo Dr. William H. Wolberg. Trata-se de um médico dos hospitais da Universidade de Wisconsin Madison, na cidade de Wisconsin nos Estados Unidos (WBCD, 2012). O Dr. Wolberg, em seu ofício, entre os anos de 1989 e 1991, recebeu diversos casos de tumores na mama para serem analisados. Nas análises realizadas, os tumores foram diagnosticados em benignos e malignos. Com referência a estas informações, foi montada uma base de dados com 9 instâncias representando as características do tumor e, evidentemente, a classificação para estas instâncias, totalizando 10 variáveis (Wolberg et al 1990).

As características armazenadas na base de dados são as seguintes:

- espessura da massa celular (CT);
- uniformidade do tamanho da célula (CS);
- uniformidade do formato da célula (CH);
- adesão marginal (AD);
- Tamanho de uma célula epitelial (EP);
- núcleo vazio (BN);
- cromatina branda (CO);
- nucléolo normal (NN);
- mitose (MM);
- classificação (“benigno” ou “maligno”).

Esta base de dados possui 699 amostras, sendo que 65% representam tumores benignos e 35% representam tumores malignos (Wolberg et al 1990). Neste problema, a classe 2 corresponde a um padrão normal (“benigno”) e a classe 4 corresponde a um padrão anormal (“maligno”).

## 3 Algoritmo de Seleção Negativa (ASN)

O ASN (Forrest et al., 1994) é uma técnica que se baseiam no processo de reconhecimento de padrões

exercido pelo sistema imunológico biológico, sendo elaborado como um modelo computacional. O ASN proposto por (Forrest et al., 1994) é baseado na seleção negativa de linfócitos T dentro do timo, isto é, representa o processo de discriminação das células, entre próprias e não-próprias realizado pelo organismo. Os passos deste algoritmo são apresentados em duas fases, como mostrado a seguir (de Castro, 2001):

- 1) Censoriamento
  - a. Defina o conjunto de cadeias próprias (S) que se deseja proteger;
  - b. Gere cadeias aleatórias e avalie a afinidade (Match) entre cada uma delas e as cadeias próprias. Caso a afinidade seja superior a um limiar estipulado, rejeite a cadeia. Caso contrário, armazene-a em um conjunto de detectores (R).
- 2) Monitoramento
  - a. Dado o conjunto de cadeias que se deseja proteger (cadeias protegidas), avalie a afinidade entre cada uma delas e o conjunto de detectores. Se a afinidade for superior a um limiar preestabelecido, então um elemento não-próprio é identificado.

Nas figuras 1 e 2 apresentam-se os fluxogramas das fases do ASN.

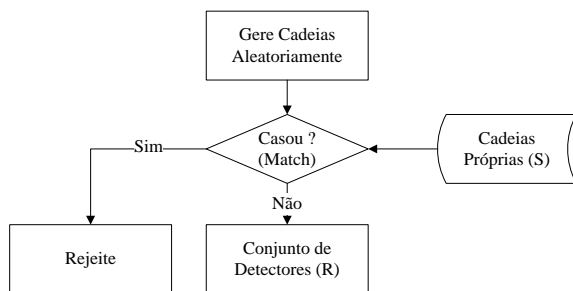


Figura 1. Fluxograma do Censoriamento do ASN.

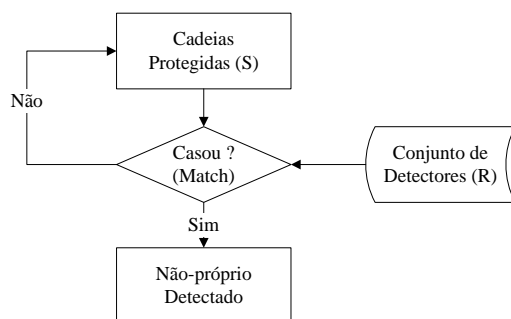


Figura 2. Fluxograma do Monitoramento do ASN.

A fase de censoriamento do ASN consiste-se basicamente em gerar um conjunto de detectores análogos às células tipo T maturadas capazes de reconhecer agentes patogênicos, isto é um padrão não-próprio. A fase de monitoramento consiste-se em monitorar um sistema, visando identificar uma mudança no comportamento do mesmo, e assim

classificar esta mudança utilizando o conjunto de detectores criados na fase de censoriamento. (de Castro e Timmis, 2002; Dasgupta, 1998).

### 3.1 Critério de Afinidade

Para avaliar a afinidade entre as cadeias e comprovar que elas são semelhantes, utiliza-se um critério conhecido como casamento. O casamento pode ser perfeito ou parcial.

O casamento perfeito é quando as duas cadeias, que estão sendo analisadas, possuem todas as variáveis com os mesmos valores, e o casamento parcial, apenas uma quantidade de variáveis entre os padrões deve ter o mesmo valor para se confirmar o casamento, sendo esta quantidade definida previamente. Esta quantidade é conhecida como a taxa de afinidade. A taxa de afinidade representa o grau de semelhança necessário para ocorrer o casamento entre os dois padrões em análise. A taxa de afinidade é definida através da seguinte relação (Bradley e Tyrrell, 2002):

$$TAf = \left( \frac{An}{At} \right) * 100 \quad (1)$$

sendo:

$TAf$  : taxa de afinidade;

$An$  : número de cadeias normais no problema (cadeias próprias);

$At$  : número total de cadeias no problema (cadeias próprias e não-próprias).

Através da equação (1) pode-se calcular de forma correta o valor da taxa de afinidade para o problema proposto, onde a equação (1) propõe uma relação estatística entre todas as amostras do problema.

A expressão (2) representa a forma de quantificar a afinidade total entre os padrões em análise (Bradley e Tyrrell, 2002):

$$Af_T = \frac{\sum_{i=1}^L Vc}{L} * 100 \quad (2)$$

sendo:

$Af_T$  : % de afinidade entre os padrões analisados;

$L$  : quantidade total de variáveis;

$Vc$  : variáveis casada;

$\sum_{i=1}^L Vc$  : somatória (quantidade) de variáveis casadas.

Desta forma se  $Af_T$  for maior ou igual a  $TAf$  ocorre o casamento entre os dois padrões (são considerados semelhantes).

#### 4 Metodologia Proposta

O sistema de diagnóstico de câncer de mama apresentado nesta seção é baseado nos sistemas imunológicos artificiais, em especial no algoritmo de seleção negativa.

O sistema de diagnóstico proposto neste artigo baseia-se na discriminação próprio/não-próprio, sendo que próprio é um padrão benigno e não-próprio um padrão que apresenta evidência de malignidade.

O método proposto se divide em duas fases: sensoriamento e o monitoramento dos dados. Na sequência apresenta-se o conjunto de testes e as fases de sensoriamento e de monitoramento do sistema de 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 uma quantidade menor de atributos para realizar o diagnóstico. Para escolher quais atributos utilizar foi feito um processo de escolha com base no cálculo do desvio padrão das amostras. Foram escolhidos os cinco atributos que apresentaram os índices mais baixos de desvio padrão, ou seja, os dados são mais homogêneos. Foi adotada esta escolha visando proporcionar maior confiabilidade para o sistema.

O desvio padrão foi calculado da seguinte forma:

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (3)$$

As variáveis escolhidas 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, portanto na tabela 1 são apresentados os dados e características da base de dados WBCD que foram utilizados.

Tabela 1. Especificação da base de dados WBCD.

Base de dados	UCI Wisconsin Breast Cancer Data
Tipo	Classificação
Número de dados	699
Número de dados aproveitáveis	683
Dados da classe "benigno"	444
Dados da classe "maligno"	239
Número de atributos	5

##### 4.2 Fase de Sensoriamento

Nesta fase são gerados os detectores para serem utilizados pelo SIA durante o processo de monitoramento. Devido ao problema de diagnóstico clínico de câncer de mama ter duas classes de

padrões, é necessário somente que o algoritmo tenha conhecimento das amostras normais (próprios), para com base nestas informações realizar a discriminação do que é próprio (benigno) e do que é não-próprio (maligno). Assim o funcionamento desta fase é apresentado na figura 3, onde se ilustra como é realizado o processo de geração dos detectores benignos (próprios).

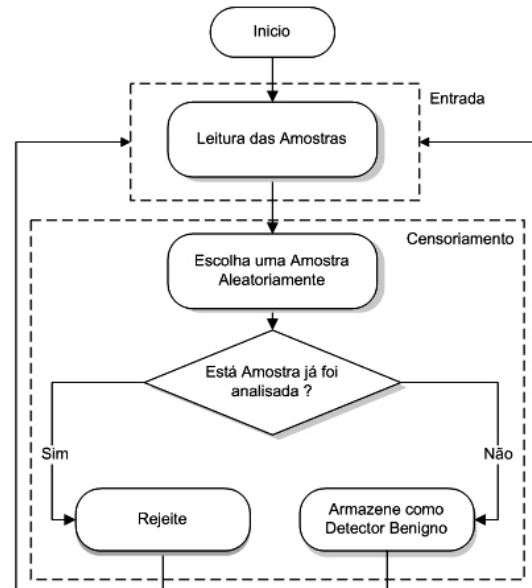


Figura 3. Esquema do sensoriamento do sistema de diagnóstico.

Foi utilizado o critério de casamento parcial (Bradley e Tyrrell, 2002). Utilizando a equação (1) nas especificações da base de dados do problema, calcula-se a taxa de afinidade. Assim tem-se um total de 683 amostras, sendo 444 amostras normais, isto é, sem câncer, e o cálculo é dado da seguinte forma:

$$TAf = \left( \frac{444}{683} \right) * 100 = 65\% \quad (4)$$

O valor da taxa de afinidade é de 65%, e isto significa que para confirmar um casamento entre duas amostras de câncer é necessário que no mínimo 65% das variáveis se casem, desta forma pode-se garantir que as amostras são semelhantes.

##### 4.3 Fase de Monitoramento

A fase de monitoramento é dividida em dois módulos, os quais são responsáveis por fazer a leitura do conjunto de testes e realizar a discriminação próprio/não-próprio.

Na fase de monitoramento faz-se a leitura das amostras que se deseja analisar (conjunto de testes). Essas amostras são comparadas aos detectores benignos (próprios) gerados previamente na fase de sensoriamento. Realizando a comparação da amostra com o conjunto de detectores, avalia-se o casamento e a afinidade entre a amostra e o conjunto de detectores. Caso a afinidade seja satisfeita, ou seja, se

existir semelhança entre as amostras analisadas é considerado um casamento e, então, a amostra é diagnosticada como “benigna”, por ter características próprias. Caso contrário, não existe o casamento entre as amostras em análise, assim a amostra é diagnosticada como “maligna”. Este processo é repetido em cada amostra, e assim é realizado o monitoramento dos dados.

A figura 4 ilustra o funcionamento do processo de monitoramento do sistema de diagnóstico.

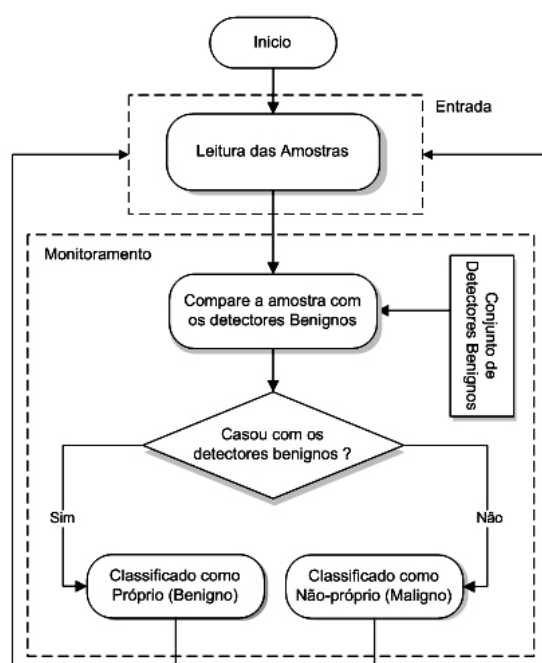


Figura 4: Esquema do monitoramento do sistema de diagnóstico.

## 5 Resultados

Nesta seção apresentam-se os resultados obtidos com a aplicação do método proposto na base de dados WBCD. Todos os testes 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 (Matlab, 2011).

No teste realizado o objetivo foi avaliar o método proposto, verificando a eficiência, precisão e o tempo computacional em relação a diferentes conjuntos de detectores benignos. Utilizou-se como parâmetro uma taxa com um valor fixo de 65%. Foram gerados três conjuntos de detectores benignos, sendo os conjuntos *I*, *II* e *III* que possuem 5, 10 e 20 padrões detectores, respectivamente. Os conjuntos de detectores gerados utilizam 1,12%, 2,25% e 4,50% das amostras benignas, que tem um total de 444 amostras. Vale ressaltar que como o processo é aleatório o conjunto de detectores uma vez gerado, provavelmente não será gerado novamente.

Na tabela 2 apresentam-se os resultados obtidos pelo sistema de diagnóstico, quando aplicado à base de dados WBCD (Wolberg, 1990).

Tabela 2. Resultados obtidos pelo método.

Diagnóstico	Conjunto de detectores		
	<i>I</i>	<i>II</i>	<i>III</i>
Amostras testadas	683	683	683
Classificações Benignos	413	434	443
Classificações Malignos	270	249	240
Acerto (%)	93,01%	97,74%	99,77%
Erro (%)	6,99%	2,26%	0,23%
Tempo (ms)	84,60	81,20	84,41

Neste teste foi possível observar que o sistema de diagnóstico apresenta um bom desempenho (índice de acerto superior a 93%), e que a quantidade de detectores benignos influencia diretamente no diagnóstico final. Comumente utiliza-se até 30% das informações da base de dados para gerar detectores, porém, neste caso, a proposta é utilizar no máximo 5% dos dados, visando proporcionar robustez ao diagnóstico. À medida que a quantidade de detectores aumenta, o diagnóstico se torna mais preciso, isto porque quanto mais conhecimento, mais eficiente é o processo de diagnóstico.

A sensibilidade (*S*) e a especificidade (*E*) do método são dadas pelas seguintes equações:

$$S = \frac{VP}{VP + FP} \quad (5)$$

$$E = \frac{VN}{FN + VN} \quad (6)$$

em que *VP* é verdadeiro positivo, *VN* é verdadeiro negativo, *FP* é falso positivo e *FN* é falso negativo.

Portanto a sensibilidade do método para este problema é de 100%, e isto significa que nenhum indivíduo doente ficaria sem tratamento. E a especificidade é de 99,77 % o que significa que 0,23% (1 indivíduos) receberia o tratamento sem necessidade.

### 5.1 Análise comparativa dos resultados

A melhor configuração do sistema de diagnóstico proposto neste trabalho obteve um índice de acerto de 99,7%, isto comprova que o algoritmo de seleção negativa é eficaz no processo de diagnóstico. Vale ressaltar que para este artigo utilizou-se apenas 5% da informação das amostras para gerar os detectores benignos, o que é uma quantidade muito reduzida de informação. Isto evidencia que o método é robusto e preciso no processo de diagnóstico. O tempo de execução é bem reduzido, o que proporciona rapidez no diagnóstico.

Na tabela 3 apresenta-se um estudo comparativo entre o método proposto e os principais métodos disponibilizados na literatura que utilizam a base de dados WBCD para avaliar o processo de diagnóstico de câncer de mama.

Tabela 3. Análise comparativa de resultados.

Referência	Técnica Utilizada	Acerto (%)
(Meesad e Yen, 2003)	Fuzzy	96,71
(Wang e George Lee, 2002)	ANFIS	96,30
(Camastra, 2006)	Kohonen	96,70
(Wang e George Lee, 2002)	<i>Backpropagation</i>	95,16
(Zhao e Davis, 2011)	SIA e rede neural de base radial	99,58
Este artigo	Algoritmo de seleção negativa	99,70

Na tabela 3 pode-se observar que o método proposto, neste trabalho, apresenta índice de acerto superior às demais técnicas.

## 6 Conclusão

Neste artigo foi apresentado um método de diagnóstico de câncer de mama baseado nos sistemas imunológicos artificiais, em especial, no algoritmo de seleção negativa. Foram descritas as principais etapas e características do ASN e sua aplicação no problema proposto. Como dados de entrada do sistema, o algoritmo precisa apenas de cinco atributos das amostras de câncer de mama. O algoritmo proposto apresentou excelentes resultados obtendo um índice de acerto de 99,77% em todas as amostras testadas. A fase de geração de detectores é a que demanda maior tempo computacional, porém é executada de forma *off-line* não acarretando prejuízo ao algoritmo. Deve-se ressaltar que com uma quantidade mínima de informações, isto é, com um conjunto de detectores bem reduzido, o método foi capaz de realizar o diagnóstico com muita precisão e segurança. A fase de monitoramento do sistema, a partir da leitura dos dados, é realizada rapidamente, com tempo inferior a 90 milésimos de segundo, o que credencia o algoritmo a ser uma ferramenta utilizada em tempo real, já que as decisões precisam ser tomadas rapidamente. Desta forma conclui-se que o sistema de diagnóstico proposto com base nos sistemas imunológicos é bastante eficiente, confiável, robusto e seguro quando aplicado no diagnóstico de amostras câncer de mama.

## Agradecimentos

Os autores agradecem a FAPESP pelo apoio financeiro de pesquisa (Proc. Nº 2011/06394-5) e o Centro de Aprendizado de Máquina e Sistemas Inteligentes da Universidade da Califórnia, (UCI – *Machine Learning Repository*) por disponibilizar a base de dados WBCD.

## Referências Bibliográficas

Bradley, D.W.; Tyrrell, A.M. (2002). Immunotronics - Novel Finite-State-Machine Architectures with Built-In Self-Test Using Self-Nonself

Differentiation. IEEE Transactions on Evolutionary Computation. Vol. 6, pp. 227–238. DOI: [10.1109/TEVC.2002.1011538](https://doi.org/10.1109/TEVC.2002.1011538)

Camastra, F. (2006). Kernel Methods For Clustering. In WIRN/NAIS, volume 3931 of Lecture Notes in Computer Science, pp. 1–9.

Dasgupta, D. (1998). “Artificial Immune Systems and Their Applications”. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

de Castro, L. N. “Engenharia Imunológica: Desenvolvimento e Aplicação de Ferramentas Computacionais Inspiradas em Sistemas Imunológicos Artificiais”. Tese de Doutorado, Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, Campinas, Brasil, 2001.

de Castro, L. N. and Timmis, J. (2002). “Artificial Immune Systems: A New Computational Intelligence Approach”, Springer. 1st edition.

Forrest, S., A. Perelson, Allen, L. & Cherukuri, R. (1994), “Self-Nonself Discrimination in a computer”, Proc. do IEEE Symposium on Research in Security and Privacy, pp. 202-212.

INCA - Instituto Nacional do Câncer (Brasil), disponível em: [www.inca.gov.br](http://www.inca.gov.br), acessado em: 01/09/2012.

Jung, J-S. R. (1993). ANFIS: Adaptive Network-Based Fuzzy Inference System, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 23, No. 3, p. 665-685. DOI: [10.1109/21.256541](https://doi.org/10.1109/21.256541)

Manikantan, K.; Sayed, S.I.; Syrigos, K.N.; Rhys-Evans, P.; Nutting, C.M. Harrington, K.J. and Kazi, R. (2009). Challenges For The Future Modifications of The tnm: Staging System For Head And Neck Cancer. Cancer Treatment Reviews, pg. 35-44. DOI: [10.1016/j.ctrv.2009.04.010](https://doi.org/10.1016/j.ctrv.2009.04.010)

Matlab (2011). 7.8 Version, Mathworks Company.

Meesad, P. e Yen, G. G. (2003). “Combined Numerical and Linguistic Knowledge Representation and Its Application to Medical Diagnosis”, IEEE Transactions on Systems, Man, and Cybernetics. DOI: [10.1109/TSMCA.2003.811290](https://doi.org/10.1109/TSMCA.2003.811290)

OMS - Organização Mundial da Saúde, Disponível em: <http://www.who.int/en/>, acessado em: 01/09/2012.

Pena-Reyes, C. A. e Sipper, M. (1999). “Designing Breast Cancer Diagnostic System via Hybrid Fuzzy-Genetic Methodology”, IEEE International Fuzzy Systems Conference Proceeding, 1999.

Wang, J. e George Lee, C. S. (2002). “Self-Adaptive Neuro-Fuzzy Inference Systems for Classification Applications”, IEEE Transactions on Fuzzy Systems, pp. 790-802. DOI: [10.1109/TFUZZ.2002.805880](https://doi.org/10.1109/TFUZZ.2002.805880)

WBCD, (2012). Wisconsin Breast Cancer Data – UCI Machine Learning Repository, disponível em: [www.archives.ics.uci.edu/ml/](http://www.archives.ics.uci.edu/ml/).

Wolberg, W. H. e Mangasarian, O. L. (1990) “Multisurface Method of Pattern Separation For Medical Diagnosis Applied to Breast Cytology”, Proceedings, da Academia Nacional de Ciências dos EUA, Volume 87, p. 9.193-9.196.

Zhao, W. e Davis, C. E. (2011). A Modified Artificial Immune System Based Pattern Recognition Approach – An Application to Clinical Diagnostics. Artificial Intelligence in medicine, Elsevier, nº 52, pp. 1-9. DOI: [10.1016/j.artmed.2011.03.001](https://doi.org/10.1016/j.artmed.2011.03.001)