

Predição do Câncer de Mama com Aplicação de Modelos de Inteligência Computacional

R.M. SILVA*, M.R.R.LEAL e F.M. LIMA

Recebido em 10 junho, 2018 / Aceito em 22 janeiro, 2019

RESUMO. O uso de modelos para diagnóstico auxiliado por computador (CAD) tem sido proposto para auxiliar na detecção e classificação do câncer de mama. Neste trabalho, avaliou-se o desempenho dos modelos de rede neural de perceptrons de múltiplas camadas e máquina de vetores de suporte não linear para classificar nódulos de câncer de mama. Dez características morfológicas, do contorno de 569 amostras, foram usadas como entrada nos classificadores. Os resultados médios obtidos no conjunto das 50 simulações realizadas, mostram que os modelos propostos apresentaram bom desempenho (todos ultrapassaram a 90,0%) em termos da acurácia no conjunto de teste. O algoritmo de máquina de vetor de suporte não linear destacou-se quando comparado ao algoritmo de rede neural de perceptrons de múltiplas camadas proposto, com acurácia de $\approx 99,0\%$ e taxa de falso negativo de $\approx 2,0\%$. O modelo de rede neural apresentou desempenho inferior ao classificador de máquina de vetor de suporte não linear. Os resultados médios obtidos, com a aplicação dos modelos propostos, mostram-se promissores, na classificação do câncer de mama.

Palavras-chave: rede neural, máquina de vetor de suporte, câncer de mama.

1 INTRODUÇÃO

Com mais de 14 milhões de novos casos por ano, o câncer vem se tornando uma das doenças mais incidentes no mundo, sendo responsável por 15% das quase 56 milhões de mortes [16]. No Brasil, as estimativas para o ano de 2018 apontam para mais de 600 mil novos casos de câncer, onde o câncer de mama é o que apresenta uma das maiores incidências, após o câncer de pele não melanoma e de próstata [10]. Desde o início das pesquisas sobre o câncer de mama, a melhor maneira para cura da doença é a detecção precoce. A mamografia é uma das melhores técnicas para o rastreamento do câncer de mama disponível atualmente, capaz de registrar imagens da mama com a finalidade de diagnosticar a presença ou ausência de estruturas que possam indicar a doença. Com esse tipo de exame pode-se detectar o tumor antes que ele se torne palpável.

*Corresponding author: Robson Mariano da Silva – E-mail: robsonms@ufrj.br – <https://orcid.org/0000-0002-3768-8449>

Universidade Federal Rural do Rio de Janeiro, Rodovia BR 465, Km 07 - Zona Rural, Seropédica, 23890-000, Rio de Janeiro Brasil. – E-mails: robsonms@ufrj.br, cellerosa@hotmail.com, fabianomarcos1@hotmail.com

No entanto, a avaliação do exame de mamografia e o diagnóstico, realizado por um radiologista, requer bastante habilidade, porém há limitações na predição primária do câncer de mama. Estudos têm revelado que 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, ou seja, entre 65% a 85% das biópsias de mama são realizadas em lesões benignas. 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.

Essa neoplasia vem despertando maior atenção na saúde pública, bem como na comunidade científica, onde pesquisadores estão utilizando técnicas de inteligência computacional no desenvolvimento de sistemas de apoio ao diagnóstico por computador (CAD), visando aumentar a taxa de detecção do câncer de mama [4, 7, 15]. Dentre essas técnicas destacam-se as Redes Neurais Artificiais – RNAs [2, 12, 17, 18] e as Máquinas de Vetores de Suporte – SVMs [3, 9, 11, 14], por serem robustas em um conjunto de dados ruidosos. Apesar dos bons resultados obtidos com RNAs, seus resultados são estocásticos e dependem fortemente da ordem de apresentação dos objetos e dos pesos iniciais atribuídos a suas conexões. Desta forma é recomendável executá-la várias vezes para diferentes configurações dos dados e valores iniciais de pesos, obtendo uma média de desempenho. Frente ao exposto, a contribuição inovadora do presente trabalho é avaliar os modelos de rede neural artificial de *Perceptrons* de Múltiplas Camadas (MLP) e Máquina de Vetores de Suporte (SVMs) não linear em um conjunto de 50 simulações na classificação de malignidade mamária, obtidas em achados mamográficos.

2 REFERENCIAL TEÓRICO

Redes Neurais Artificiais (RNAs) são sistemas paralelos e distribuídos constituídos de unidades simples (neurônios ou nós), que calculam determinadas funções matemáticas (principalmente não-lineares) e apresentam capacidade de generalização, auto-organização e processamento temporal. De maneira análoga ao sistema nervoso de um ser humano, onde os neurônios são dispostos em uma ou mais camadas e interligadas por inúmeras conexões, geralmente unidirecionais, denominadas sinapses. A estas conexões associam-se valores, denominados pesos sinápticos, responsáveis pela ponderação das entradas de cada neurônio como forma de armazenamento do conhecimento de um determinado modelo [8]. Nas redes neurais, a aprendizagem se dá através de um conjunto de unidades simples de processamento chamados de neurônios artificiais ou nós. A Figura 1 mostra uma representação do modelo não linear de um neurônio artificial.

Uma RNA possui a característica de aprender por meio de exemplos, extraindo conhecimento de um determinado conjunto de dados. O conhecimento é adquirido a partir do processo pelo qual os hiperparâmetros livres de uma rede neural, são ajustados por meio de uma forma continuada de estímulo pelo ambiente externo, visando minimizar o valor de uma função de erro. Este processo é definido como aprendizado, que pode ser classificado como supervisionado ou não supervisionado [8]. Dentro do contexto de aprendizado supervisionado, apresentamos à rede as

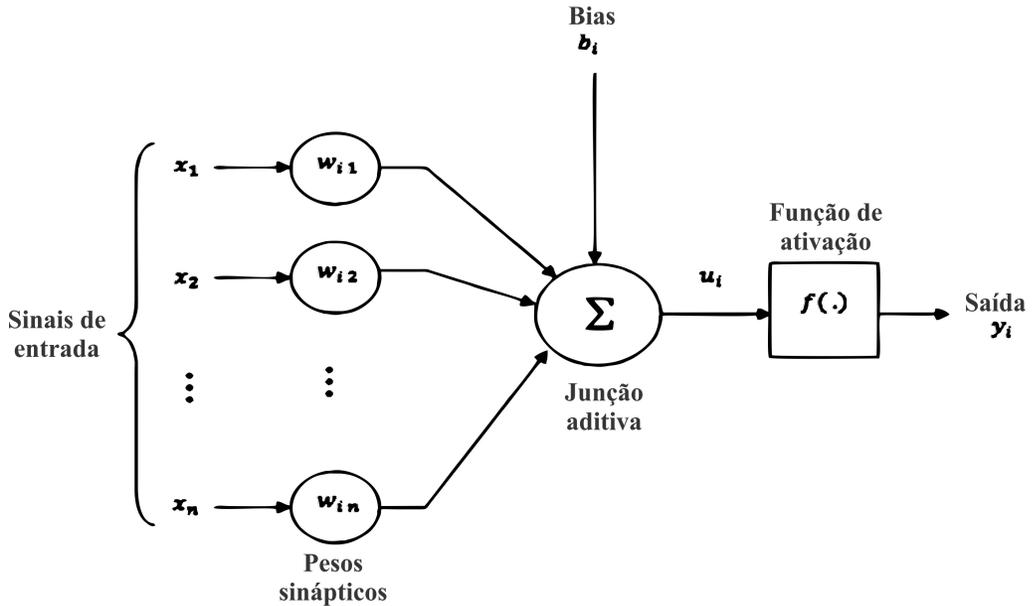


Figura 1: Modelo não linear de neurônio. Fonte adaptado de [8].

entradas disponíveis e a saída desejada, e o algoritmo trabalha no sentido de ajustar os pesos das sinapses por meio do cálculo da diferença entre o valor da saída desejada $y_{di}(t)$ e o valor predito pela RNA $y_{pi}(t)$, no instante t produzindo desta forma um erro $\delta(t)$ equação (2.1) [5].

$$\delta(t) = y_{di}(t) - y_{pi}(t) \quad (2.1)$$

A forma genérica para o ajuste dos pesos, por correção de erros, é apresentada na equação (2.2).

$$w_{ij}(t+1) = w_{ij}(t) + \eta e_i(t)x_i(t) \quad (2.2)$$

onde η é a taxa de aprendizado e $x_i(t)$ é a entrada para o neurônio i no tempo t [5].

Na aprendizagem não supervisionada, não se conhecem os valores das saídas desejadas y_{di} . Logo, a aprendizagem ocorre através da identificação de padrões nas entradas. A escolha da arquitetura de uma RNA está relacionada com os tipos de problemas a serem abordados e é definida por 4 hiperparâmetros principais: número de camadas da rede, número de neurônios em cada camada, tipo de conexão entre neurônios e a topologia da rede. Em relação ao número de camadas, existem redes de camada única, que possuem somente um nó entre as camadas de entrada e saída da rede, sendo restritas a resolver problemas linearmente separáveis.

As redes neurais de múltiplas camadas possuem mais de um neurônio entre uma entrada e uma saída da rede. Dentre as redes de múltiplas camadas temos a do tipo *Perceptrons* de Múltiplas

Camadas (*Multilayer Perceptron* - MLP), que apresentam uma ou mais camadas de neurônios intermediários ou escondidos, sendo considerada uma aproximadora universal. De acordo com o teorema da aproximação universal, qualquer função contínua pode ser uniformemente aproximada por uma rede com pelo menos uma camada de neurônios ocultos, e com uma função de ativação sigmoide [8]. Considere seja $\varphi(\cdot)$ uma função contínua, limitada e monotonamente crescente e I_{m_0} um hiper cubo unitário $[0, 1]^{m_0}$ de dimensão m_0 . O espaço das funções contínuas em I_{m_0} é representado por $C(I_{m_0})$. Então, dada qualquer função $f \in C(I_{m_0})$ e $\varepsilon > 0$, existe um inteiro M e conjuntos de constantes reais α_i, b_i e w_{ij} , onde $i = 1, \dots, m_i$ e $j = 1, \dots, m_0$ tal que podemos definir:

$$F(x_1, \dots, x_{m_0}) = \sum_{i=1}^{m_1} \alpha_i \varphi \left(\sum_{j=1}^{m_0} w_{ij} x_j + b_i \right) \quad (2.3)$$

Como 2.3 uma realização aproximada da função $f(\cdot)$, isto é,

$$|F(x_1, \dots, x_{m_0}) - f(x_1, \dots, x_{m_0})| < \varepsilon$$

para todo x_1, x_2, \dots, x_{m_0} que se encontre no espaço de entrada.

Logo o teorema da aproximação universal é diretamente aplicável aos *perceptrons* de múltiplas camadas. A Figura 2 representa uma rede MLP com três entradas, duas camadas intermediárias com quatro neurônios e uma camada de saída com um neurônio, produzindo uma única informação de saída.

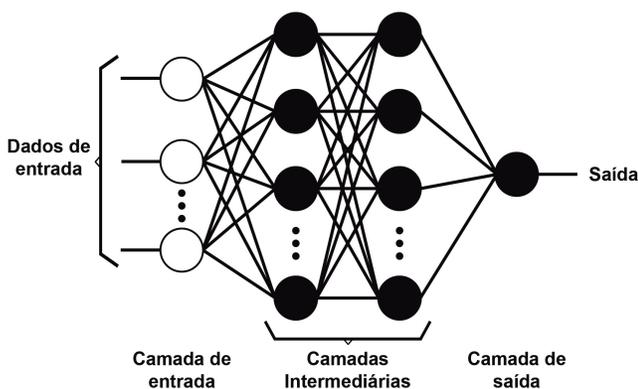


Figura 2: Representação gráfica e uma rede MLP, com duas camadas intermediárias.

As redes MLP têm sido aplicadas com sucesso para resolver diversos problemas, através do seu treinamento de forma supervisionada com a utilização do algoritmo de retropropagação do erro (*backpropagation*), o qual apresenta duas fases distintas. Na primeira fase, ocorre a propagação do sinal funcional (*feedforward*) mantendo-se os pesos fixos de sorte a gerar um valor de saída a partir das entradas fornecidas à rede. Na segunda fase, as saídas são comparadas com os valores desejados, gerando um sinal de erro, que se propaga da saída para a entrada, ajustando-se os

pesos de forma a minimizar o erro. Assim, a forma de calcular o erro depende da camada em que se encontra o neurônio, como mostra a equação 2.4.

$$\delta_l = \begin{cases} f'_a e_l, & \text{se } n_l \in C_{sai} \\ f'_a \sum w_{lk} \delta_k, & \text{se } n_l \in C_{int} \end{cases} \quad (2.4)$$

onde n_l é o l -ésimo neurônio, C_{sai} representa a camada de saída, C_{int} representa uma camada intermediária, f'_a é a derivada parcial da função de ativação do neurônio e e_l é o erro quadrático cometido pelo neurônio de saída quando sua resposta é comparada à desejada, que é definido pela equação 2.5.

$$e = \frac{1}{2} \sum_{q=1}^k (y_q - \hat{y}_q)^2 \quad (2.5)$$

onde \hat{y}_q é a saída produzida pelo neurônio e y_q é a saída desejada.

A derivada parcial define o ajuste dos pesos, utilizando o gradiente descendente da função de ativação. Essa derivada avalia a contribuição de cada peso no erro da rede para a classificação de um dado objeto. Se a derivada para um dado peso for positiva, o peso está provocando um aumento da diferença entre a saída da rede e a saída desejada. Logo, sua magnitude deve ser reduzida de sorte a diminuir o erro. Caso contrário, o peso estará contribuindo para que a saída da rede esteja mais próxima da desejada.

As Máquinas de Vetores de Suporte (*Support Vector Machines - SVMs*) são um conjunto de métodos de aprendizagem supervisionada estruturada nos princípios da teoria da aprendizagem estatística, usadas para classificação e regressão de dados. Após o treinamento de um conjunto de dados, o algoritmo possui propriedades que permitem uma boa generalização de dados, não antes explorados. Possibilitando criar um limite de decisão entre duas classes que permite a previsão de rótulos de um ou mais vetores de características [19]. Este limite de decisão, conhecido como hiperplano, é orientado de modo a ser o mais próximo possível dos pontos de dados na vizinhança de cada uma das classes. Esses pontos mais próximos são denominados vetores de suporte. Considere um conjunto de dados de treinamento rotulado:

$$(x_1, y_1), \dots, (x_n, y_n), \quad x_i \in \mathbb{R}^d \quad e \quad y_i \in (-1, +1) \quad (2.6)$$

onde x_i é uma representação do vetor de característica e y_i o rótulo da classe (negativo ou positivo) de um conjunto de treinamento i . O hiperplano ótimo pode então ser definido como:

$$wx^t + b = 0 \quad (2.7)$$

sendo w o vetor de peso, x o vetor de entrada e b a tendência (ou bias). Onde w e b devem satisfazer as seguintes desigualdades para todos elementos do conjunto de treinamento:

$$\begin{aligned} wx_i^t + b &\geq +1, & \text{se } y_i &= 1 \\ wx_i^t + b &\leq -1, & \text{se } y_i &= -1 \end{aligned} \quad (2.8)$$

O objetivo do treinamento do modelo SVM é encontrar w e b , de modo que o hiperplano separe os dados e maximize a margem $1/\|w\|^2$.

Dessa forma, para um conjunto de dados linearmente separáveis, a SVMs é capaz de realizar a categorização de duas classes através de um hiperplano ótimo, obtendo uma boa generalização na sua classificação. Entretanto, para a classificação binária, onde os dados não são linearmente separáveis no espaço original, faz-se necessário referenciá-lo em um novo espaço de maior dimensão, denominado de espaço de características. Para tal, a utilização de Máquinas de Vetores de Suporte não linear (SVMs-não linear) é necessária.

Máquinas de Vetores de Suporte não linear (SVMs-não linear) são definidas como um algoritmo de aprendizado supervisionado, cujo objetivo é classificar determinado conjunto de dados que são mapeados para um espaço de características multidimensional usando uma função de *kernel*. A função de *kernel* é um artifício matemático, que permite a SVMs fazer um mapeamento dos dados não linearmente separáveis, em um espaço de dimensão maior, onde os dados tornam-se linearmente separáveis [8]. Estas funções mapeiam o conjunto de dados no espaço original, denominado como espaço de entradas, para o espaço de características, ou seja, um *kernel* K recebe dois pontos x_i e x_j do espaço de entradas e retorna o produto escalar desses pontos no espaço de características. Os *kernels* são incorporados ao classificador SVMs por meio da equação (2.9)

$$f(x) = \text{sgn} \left(\sum_{i=1}^{n_{SV}} \alpha_i y_i K(\vec{x}_i, \vec{x}) + b \right) \quad (2.9)$$

onde K denota a função *kernel*, a qual recebe como entrada (\vec{x}_i) o vetor de suporte i e os valores \vec{x} da amostra a ser classificada, α_i os multiplicadores de Lagrange e b o valor do intercepto.

Os métodos baseados na teoria de *kernel* provocaram uma verdadeira revolução nos algoritmos da teoria de aprendizado estatístico, supervisionado e não supervisionado, por possibilitar a criação de versões não lineares dos algoritmos clássicos lineares. Dentre o conjunto de algoritmos encontrado na literatura que utilizam função de *kernel*, o algoritmo de máquinas de vetores de suporte proposto por Vapnik [20] para classificação binária é o mais proeminente. As SVMs possuem funções de *kernel* que caracterizam seu modo de reconhecimento dos padrões, sendo a polinomial, gaussiana e sigmoidal as mais utilizadas (Tabela 1).

Tabela 1: Equações matemáticas referente as funções de *kernel* mais utilizadas.

Tipo de Kernel	Função $k(x_i, x_j)$	Parâmetros
Polinomial	$(\delta(x_i \cdot x_j))^d$	δ, k, d
Gaussiana	$\exp(-\sigma \ x_i - x_j\ ^2)$	σ
Sigmoida	$\tanh(\delta(x_i \cdot x_j) + k)$	δ, k

Na função polinomial, o grau (d) pode ser definido durante o treinamento. Na função gaussiana que corresponde a um espaço de características de dimensão infinita, sua utilização permite que as SVMs apresentem características de uma rede neural de função de base radial (RBF). Já a função sigmoideal possibilita comportamento semelhante ao de uma rede neural MLP. As SVMs realizam o reconhecimento de padrões entre duas classes, encontrando uma função de decisão (hiperplano). Os pontos selecionados a partir dos dados de treinamento são denominados vetores de suporte (SVs). Enquanto as técnicas tradicionais de reconhecimento de padrões são baseadas em minimizar o risco empírico, as SVMs minimizam o risco estrutural.

3 MATERIAIS E MÉTODOS

O conjunto de dados utilizado nesse estudo, consiste de 569 dados, proveniente de pacientes com suspeita de câncer de mama, obtidos junto a base pública do Wisconsin Diagnostic Breast Cancer (WDBC) [21]. Os dados analisados contêm informações referente aos valores médios do raio, textura, perímetro, área, suavidade, compacidade, concavidade, número de pontos côncavo no contorno, simetria e dimensão fractal da lesão. A metodologia visa comparar os modelos computacionais estruturados em Rede Neural MLP e Máquinas de Vetores de Suporte (SVMs-não linear), na classificação de malignidade, referente as características morfológicas do contorno da lesão encontrada em achados mamográficos (Figura 3).

Para avaliar a performance dos modelos propostos nesse estudo, foi utilizada a acurácia ou precisão total (ACC) e a taxa de erro da classe de falso negativo (EFN). Definidos respectivamente por:

$$ACC = \frac{V_P + V_N}{n} \quad (3.1)$$

$$EFN = \frac{F_N}{V_P + F_N} \quad (3.2)$$

onde V_P são as amostras de rótulo positivo (+1) preditas como positivas, V_N são amostras de rótulo negativo (-1) preditas como negativas, F_N são amostras de rótulo positivo (+1) preditas como negativas e n o número total de amostras. Para cada modelo foram realizadas 50 simulações, de sorte a obter uma melhor generalização nos resultados obtidos.

Os modelos computacionais foram implementados utilizando o *software* R e os pacotes *Kernlab* [13] e *Neuralnet* [6] respectivamente, na implementação do modelo SVMs-não linear e no modelo de rede neural MLP. A relação dos hiperparâmetros utilizados nos modelos RN-MLP e

SVMs-não linear está sumarizada na Tabela 2 e na Tabela 3 respectivamente. Os hiperparâmetros utilizados na classificação foram obtidos de forma empírica.

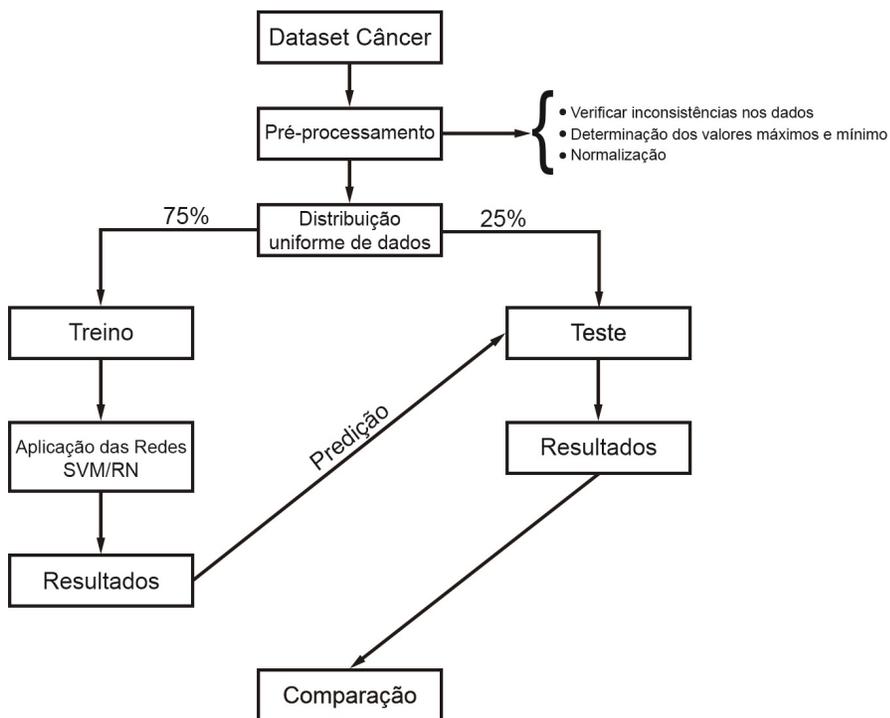


Figura 3: Fluxograma do modelo proposto.

Tabela 2: Parâmetros utilizados no modelo RN-MLP.

Parâmetros	Valor
Nº de simulações	50
Nº máximo de épocas	100000
Função de ativação	logistic
Algoritmo de treinamento	backpropagation
Métrica usada no erro de treinamento	SSE
Nº de camadas ocultas	2
Nº de neurônios da primeira camada oculta	7
Nº de neurônios da segunda camada oculta	4
Inicialização dos pesos	Aleatoriamente com valores [0,1]
Critério de parada	0,001

Tabela 3: Parâmetros utilizados no modelo SVM-não linear.

Parâmetros	Valor
Tipo de Classificador	C-svc
Função de kernel	rbf
Variância da função de kernel (σ)	0,5
Parâmetro de regularização C	1
Critério de parada	0,001

4 RESULTADOS E DISCUSSÕES

Os modelos computacionais propostos nesse trabalho foram avaliados com a incorporação dos atributos referente ao raio, textura, perímetro, área, suavidade, compacidade, concavidade, número de pontos côncavo no contorno, simetria e dimensão fractal da lesão do conjunto dos dados de pacientes portadores de microcalcificação mamárias. Os resultados médios obtidos nas 50 simulações com a aplicação dos modelos estão representados nas tabelas 4 e 5.

Tabela 4: Desempenho médio do modelo RN-MLP na categorização de malignidade no conjunto das 50 simulações.

	Erro treino SSE	ACC (%)	EFN (%)
[<i>média</i> \pm <i>sd</i>]	0,558 \pm 0,50	92,58 \pm 2,03	9,38 \pm 4,39
[<i>mediana</i> \pm <i>irq</i>]	0,507 \pm 0,94	92,25 \pm 3,52	9,49 \pm 6,64
[<i>Max</i> - <i>Min</i>]	1,849 - 0,022	94,37 - 89,44	22,22 - 2,00
Melhor simulação	0,022	94,37	2,00

O modelo RN-MLP, na sua melhor simulação obteve uma acurácia superior a 94%, com valor de falso negativo de 2%. Indicando uma precisão de 98% em termos de sensibilidade no conjunto de teste. No que tange ao erro na detecção de falso negativo, o modelo obteve um valor médio inferior a 10%, no conjunto das 50 simulações realizadas.

Tabela 5: Desempenho médio do modelo SVM-não linear na categorização de malignidade no conjunto das 50 simulações.

	Erro			
	treino (%)	LOOC (%)	ACC (%)	EFN (%)
[<i>média</i> \pm <i>sd</i>]	3,42 \pm 0,54	5,35 \pm 0,70	95,00 \pm 1,78	9,00 \pm 4,45
[<i>mediana</i> \pm <i>irq</i>]	3,51 \pm 0,71	5,42 \pm 0,96	95,07 \pm 2,11	8,33 \pm 5,21
[<i>Max</i> - <i>Min</i>]	4,45 - 2,11	7,25 - 3,25	98,58 - 90,85	21,15 - 0,00
Melhor simulação	3,51	5,60	98,58	1,96

De acordo com a análise dos resultados apresentados na Tabela 5, é possível verificar o desempenho promissor do modelo estruturado em SVM-não linear. Onde se obteve na sua melhor simulação, uma acurácia acima de 98% e taxa de erro de falso negativo inferior a 2% (1,96%). No que tange ao erro de validação cruzada *leave-one-out* (LOOC), podemos verificar que o mesmo

obteve uma amplitude entre o valor máximo e mínimo obtido de 4%, nas 50 simulações realizadas. Os resultados médios obtidos pelos modelos RN-MLP e SVM-não linear, na categorização de malignidade no conjunto das simulações realizadas está representada na Tabela 6.

Tabela 6: Desempenho médio dos modelos propostos no estudo, no conjunto das 50 simulações.

	SVM-não linear		RN-MLP	
	ACC (%)	EFN (%)	ACC (%)	EFN (%)
[<i>média ± sd</i>]	95,00 ± 1,78	9,00 ± 4,45	92,58 ± 2,03	9,38 ± 4,39
Pior simulação	91,55	21,15	90,85	22,22
Melhor simulação	98,58	1,96	94,37	2,00

Para selecionar a melhor e a pior simulação foi utilizado o valor do erro de falso negativo obtidos pelos modelos, visto que este hiperparâmetro é de suma importância na categorização de malignidade. Aplicando o teste de comparação das médias com p-valor $\leq 0,05$, pode-se verificar a existência de diferença estatística significativa entre os resultados, referente à acurácia entre os modelos utilizados no estudo. Indicando que para o hiperparâmetro ACC o modelo SVM-não linear tem melhor desempenho quando comparado com o modelo de RN-MLP.

Apesar do modelo SVM-não linear, apresentar valor médio do erro de falso negativo menor que o obtido pelo modelo RN-MLP, não foi verificada diferença estatística significativa, a nível de 95% (p-valor $\leq 0,05$) entre os resultados obtidos pelos modelos. O modelo SVM-não linear, nas 50 (cinquenta) simulações realizadas no conjunto de teste obteve simulação com valor de 100% de sensibilidade, ou seja, 0% na determinação da taxa de erro de falso negativo. Fato este, que não foi verificado no modelo RN-MLP, onde se obteve valor máximo de 98%.

É importante ressaltar que a acurácia obtida pelos modelos, na classificação da microcalcificação mamária, encontra-se próxima aos valores obtidos na literatura com a utilização de técnicas baseadas em inteligência computacional [20]. Comparando os resultados obtidos por [1], que utilizou o modelo L2-SVM, na classificação no WDBC (ACC $\approx 96,09\%$ e EFN $\approx 2,46\%$), pode-se verificar que o modelo SVM-não linear proposto nesse estudo, apresentou valores em termos de acurácia (ACC $\approx 98,58\%$) e o valor da taxa de erro de falso negativo (EFN $\approx 1,96\%$) superiores.

5 CONSIDERAÇÕES FINAIS

A elevada taxa de incidência e mortes causadas pelo câncer de mama, atualmente no Brasil e no mundo, justifica o desenvolvimento de pesquisas científicas voltadas para estratégias de auxílio na detecção precoce da doença, fator determinante para o sucesso do tratamento. Neste trabalho, propusemos a utilização de modelos computacionais estruturados em RN-MLP e SVM-não linear, na categorização de malignidade em achados mamográficos. A incorporação de informações referente as características morfológicas do contorno da lesão de mama, contribuiu para a performance dos modelos propostos no que tange a determinação da taxa de falso negativo. Logo, essa métrica é de suma importância para os profissionais de saúde, principalmente na detecção de malignidade do nódulo mamário.

Apesar dos resultados obtidos, com a aplicação dos modelos de rede neural de *perceptrons* de múltiplas camada e máquina de vetor de suporte não linear, na classificação de microcalcificação mamárias terem apresentados resultados promissores, percebe-se a necessidade de aprofundar o estudo. Para tal, pretende-se futuramente elaborar um modelo híbrido estruturado na utilização de algoritmos genéticos e rede neural convolucional, de modo a avaliar o desempenho na classificação da lesão mamária, bem como, na otimização dos hiperparâmetros do modelo.

ABSTRACT. The use of computer aided diagnostic (CAD) models has been proposed to aid in the detection and classification of breast cancer. In this work, we evaluated the performance of neural network models of multilayered perceptrons and nonlinear support vector machines to classify breast cancer nodules. Ten morphological characteristics, from the outline of 569 samples, were used as input to the classifiers. The average results obtained in the set of 50 simulations showed that the proposed models presented good performance (all exceeded 90.0 %) in terms of the accuracy of the test set. The nonlinear support vector machine algorithm stands out when compared to the proposed multilevel perceptrons neural network algorithm, with accuracy of $\approx 99,0$ % and false negative rate of *approx* 2.0 %. The neural network model presented inferior performance to the non-linear support vector machine classifier. The average results, with the application of the proposed models, are shown to be promising in the classification of breast cancer.

Keywords: neural network, support vector machine, breast cancer.

REFERÊNCIAS

- [1] A.F.M. Agarap. On breast cancer detection: an application of machine learning algorithms on the wisconsin diagnostic dataset. In “Proceedings of the 2nd International Conference on Machine Learning and Soft Computing”. ACM (2018), pp. 5–9.
- [2] S. Agrawal & J. Agrawal. Neural network techniques for cancer prediction: A survey. *Procedia Computer Science*, **60** (2015), 769–774.
- [3] H. Asri, H. Mousannif, H. Al Moatassime & T. Noel. Using machine learning algorithms for breast cancer risk prediction and diagnosis. *Procedia Computer Science*, **83** (2016), 1064–1069.
- [4] J. Cong, B. Wei, Y. He, Y. Yin & Y. Zheng. A Selective Ensemble Classification Method Combining Mammography Images with Ultrasound Images for Breast Cancer Diagnosis. *Computational and mathematical methods in medicine*, **2017** (2017), 1–7.
- [5] A. de Pádua Braga, A.C.P. de Leon Ferreira & T.B. Ludermir. “Redes neurais artificiais: teoria e aplicações”. LTC Editora (2012).
- [6] S. Fritsch, F. Guenther & M.F. Guenther. Package ‘neuralnet’. *The Comprehensive R Archive Network*, (2016).
- [7] S.J.S. Gardezi, M.M. Eltoukhy & I. Faye. Machine Learning Applications in Breast Cancer Diagnosis. In “Handbook of Research on Machine Learning Innovations and Trends”. IGI Global (2017), pp. 465–490.

- [8] S. Haykin. “Redes neurais: princípios e prática 2 ed”. Bookman Editora (2001).
- [9] M.W. Huang, C.W. Chen, W.C. Lin, S.W. Ke & C.F. Tsai. SVM and SVM ensembles in breast cancer prediction. *PloS one*, **12**(1) (2017), e0161501.
- [10] INCA. “Estimativas 2018. Incidência de Câncer no Brasil”. Instituto Nacional do Câncer (2018).
- [11] H. Jiang, W.K. Ching, W.S. Cheung, W. Hou & H. Yin. Hadamard Kernel SVM with applications for breast cancer outcome predictions. *BMC systems biology*, **11**(7) (2017), 138.
- [12] R. Kala, R. Janghel, R. Tiwari & A. Shukla. Diagnosis of breast cancer by modular evolutionary neural networks. *International Journal of Biomedical Engineering and Technology*, **7**(2) (2011), 194–211.
- [13] A. Karatzoglou, A. Smola, K. Hornik & A. Zeileis. kernlab-an S4 package for kernel methods in R. *Journal of statistical software*, **11**(9) (2014), 1–20.
- [14] K. Menaka & S. Karpagavalli. Breast Cancer Classification using Support Vector Machine and Genetic Programming. *International Journal of Innovative Research in Computer and Communication Engineering*, **1**(7) (2013).
- [15] A. Onan. On the performance of ensemble learning for automated diagnosis of breast cancer. In “Artificial Intelligence Perspectives and Applications”. Springer (2015), pp. 119–129.
- [16] W.H. Organization. “Cancer Report”. International Agency for Research on Cancer. Geneva, Switzerland, who library ed. (2014).
- [17] M. Pérez, M.E. Benalcázar, E. Tusa, W. Rivas & A. Conci. Mammogram classification using back-propagation neural networks and texture feature descriptors. In “Ecuador Technical Chapters Meeting (ETCM), 2017 IEEE”. IEEE (2017), pp. 1–6.
- [18] I. Saritas. Prediction of breast cancer using artificial neural networks. *Journal of Medical Systems*, **36**(5) (2012), 2901–2907.
- [19] A.J. Smola & B. Schölkopf. A tutorial on support vector regression. *Statistics and computing*, **14**(3) (2004), 199–222.
- [20] V. Vapnik. “Statistical learning theory. 1998”. Wiley, New York (1998).
- [21] W.H. Wolberg, W.N. Street & O.L. Mangasarian. Breast cancer Wisconsin (diagnostic) data set. *UCI Machine Learning Repository* [<http://archive.ics.uci.edu/ml/>], (1992).