

# Utilização de redes neurais multicamadas para previsão do câncer de mama

LUCAS AGOSTINHO FERNANDES<sup>1</sup>, VITOR BRYAN MELO DE LIMA<sup>2</sup>,  
PAULO HENRIQUE ARAUJO NOBRE<sup>3</sup>, IAGO DA SILVA ALVES<sup>4</sup>

<sup>1</sup>Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)

<>

**Resumo.** Este artigo traz a proposta da conscientização e importância do conhecimento e presença da inteligência artificial no cotidiano da população brasileira, fazendo uso de tais conhecimentos e tecnologias para solução e detecção de quadros patológicos, e, neste caso específico, a respeito do câncer de mama. O uso de redes neurais para diagnosticar doenças com auxílio de computadores (CAD) tem sido feito para colaborar na detecção e classificação do câncer de mama. Neste projeto, apresentamos resultados e desempenho do uso de perceptrons de múltiplas camadas para classificar nódulos de câncer de mama. Dez características morfológicas, do contorno de 569 amostras, foram usadas como entrada nos classificadores, na qual a rede classifica os dados de entrada em duas classes: tumores malignos e tumores benignos. A base de dados utilizada foi a de Wisconsin Breast Cancer Database (WBCD). A WBCD é uma base de dados pública que agrega informações de tumores de câncer de mama analisados através de FNA (Fine Needle Aspiration).

**Palavras-chaves:** Câncer de mama; inteligência artificial; redes neurais; perceptron de multicamadas; classificação.

**Abstract.** This article presents the proposal of awareness and the importance of knowledge and presence of artificial intelligence in the daily life of the Brazilian population, applying this knowledge and technologies for solution and detection of pathological conditions, and, in this specific case, breast cancer. The use of neural networks to diagnose computer aided diseases (CAD) has been done to collaborate in the detection and classification of breast cancer. In this project, we present results and performance of using multilayer perceptrons to classify breast cancer nodules. Ten morphological characteristics, from the contour of 569 samples, were used as input to the classifiers, in which the network classifies the input data into two classes: malignant tumors and benign tumors. The database used was the Wisconsin Breast Cancer Database (WBCD). WBCD is a public database that aggregates information on breast cancer tumors analyzed through Fine Needle Aspiration (FNA).

**Keywords:** breast cancer; artificial intelligence; neural networks; multilayer perceptron; classiy.

## 1 INTRODUÇÃO

O câncer de mama é uma doença causada por um tumor maligno, em que ocorre a multiplicação desordenada de células da mama e o desenvolvimento de um ou mais nódulos na mesma. Esse tipo de câncer é o segundo tipo mais comum no mundo, e o de maior ocorrência em pessoas do sexo feminino, correspondendo a cerca de 25% de novos casos a cada ano, e a 29% de novos casos no Brasil. Além de ser a espécie mais comum da doença no sexo feminino, a doença também é a responsável pela maior quantidade de mortes, em que, em grande parte isso se deve ao tardio diagnóstico da doença, a qual se descoberta em estágios avançados, dificulta cada vez mais o tratamento da própria.

Uma das formas de detecção da doença é a mamografia, esse exame é caracterizado pelo registro de imagens da mama, com a finalidade de diagnosticar a presença ou ausência de estruturas que possam indicar a doença. Esse tipo de exame é uma forma de detecção do tumor antes que ele atinja um estágio avançado. Contudo, a avaliação do exame e o diagnóstico deste, necessita de ser realizado por um radiologista e muita habilidade do mesmo, havendo casos de resultados negativos na detecção primária de tumores malignos, justamente por haver limitações na predição inicial da doença. Estudos revelam que 10% a 30% das mulheres que apresentaram câncer de mama obtiveram resultados negativos quando foram submetidas à mamografia, ocorrendo a possibilidade de uma má interpretação dos exames. Distorções na interpretação e classificação de lesões por especialistas implicam um número maior de biópsias desnecessárias, isto é, entre 65% a 85% das biópsias de mama são realizadas em lesões benignas. Com isso, há uma redução na relação custo-benefício dos exames, provocando uma menor credibilidade por parte dos pacientes nesses exames, e, no pior caso, a não detecção da doença, gerando um diagnóstico falso negativo.

Esse tipo de câncer promove diversos estudos na comunidade científica e médica, na qual pesquisadores estão fazendo uso de técnicas de inteligência computacional no desenvolvimento de sistemas de apoio ao diagnóstico por computador (CAD), com o objetivo de aumentar a taxa de detecção, e consequentemente prevenção dessa patologia. É válido ressaltar que a presença de nódulos nas mamas não indica necessariamente a presença de um tumor maligno, visto que os tumores também podem ser benignos.

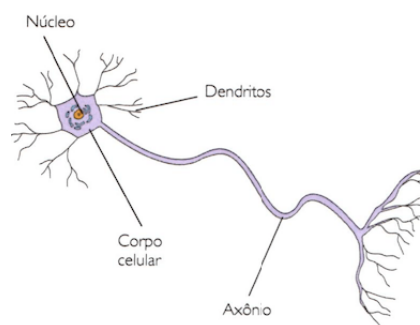
Devido a necessidade de se obter uma maior eficiência na detecção e diagnóstico do câncer de mama, em meados de 1980, pesquisadores criaram bases de dados visando agregar informações relacionadas a esse tipo

de câncer, e, dentre elas, a mais influente e reconhecida base de dados é a Wisconsin Breast Cancer Database (WBCD), sendo esta composta por 699 casos e cada um deles com 11 parâmetros, em que 10 deles são padrões obtidos por exames, e a décimo primeiro é um valor binário que indica se o tumor é maligno (0) ou benigno (1).

Visto isso, o objetivo deste trabalho é simular a detecção de casos de câncer de mama e avaliar o seu desempenho utilizando de redes neurais artificiais (RNAs), e mais especificamente uma rede Perceptron de multicamadas. O tipo de RNA utilizado possui como característica um aprendizado supervisionado, assim, a vantagem de usá-las como classificador de padrões é que elas podem, por meio de padrões mostrados com antecedência à rede durante o treinamento, encontrar soluções não lineares para novas entradas do problema descrito. É válido ressaltar que os métodos descritos e uso da computação para obter diagnósticos médicos de modo algum substitui o profissional médico responsável da área, mas esses métodos CAD (diagnóstico auxiliado por computador) possuem o objetivo principal de contribuir e auxiliar esses processos médicos.

## 2 FUNDAMENTAÇÃO

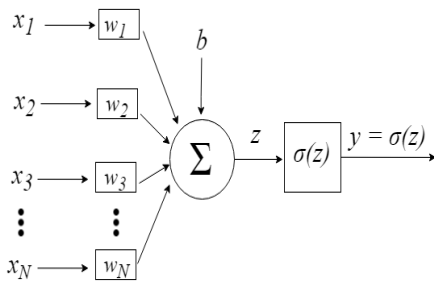
Redes Neurais Artificiais (RNAs) são sistemas paralelos e distribuídos constituídos de unidades simples (neurônios ou nós), que possuem a função de calcular determinadas funções matemáticas (lineares e não-lineares). Essas unidades são distribuídas em uma ou mais camadas e interligadas por um grande número de conexões. Na grande maioria dos modelos estas conexões estão associadas a pesos, os quais armazenam o conhecimento representado no modelo e servem para ponderar a entrada recebida por cada neurônio da rede. A inspiração principal para esse modelo e nome de redes neurais artificiais é o cérebro humano.



**Figura 1:** Imagem representando um neurônio humano. Fonte: (<https://www.coladaweb.com/biologia/histologia/neuronios>).

Essa área surgiu com o objetivo principal do homem por meio do computador, simular um cérebro e fazer a máquina aprender e generalizar informações recebidas, em que isso ocorre de maneira análoga ao sistema nervoso de um ser humano, no qual os neurônios são dispostos em uma ou mais camadas e interligadas por inúmeras conexões, geralmente unidirecionais, denominadas sinapses. Essas conexões são associadas a valores denominados pesos sinápticos, em que estes são responsáveis pela ponderação das entradas de cada neurônio como forma de armazenamento do conhecimento de um determinado modelo. De uma maneira mais simplificada, uma rede neural biológica funciona por meio de uma entrada, a qual a rede processa essa informação e retorna uma resposta, com isso, o neurônio só é ativado caso o valor seja maior que o seu limiar de excitação (threshold).

As entradas do neurônio são representadas pelo vetor  $x = [x_1, x_2, x_3, \dots, x_N]$ . A essas conexões associam-se valores, denominados pesos sinápticos (ou weights, também representados por  $w$ ), responsáveis pela ponderação das entradas de cada neurônio como forma de armazenamento do conhecimento de um determinado modelo. Ao chegarem ao neurônio, são multiplicados pelos respectivos pesos sinápticos, que são os elementos do vetor  $w = [w_1, w_2, w_3, \dots, w_N]$ , os valores desses pesos podem ser negativos ou positivos, dependendo de as conexões serem inibitórias ou excitatórias. Essa operação do produto escalar entre as entradas e os seus respectivos pesos resulta em  $z$ .  $z$  é uma função dada pela soma dos valores  $x_i \times w_i$  de todas as conexões, e o valor resultante é enviado para a função de ativação, que define a saída ( $y$ ) do neurônio.



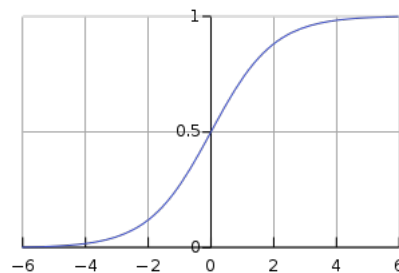
**Figura 2:** Imagem representando um neurônio artificial. Fonte: ([https://www.researchgate.net/figure/Representacao-de-um-neuronio-artificial\\_fig3\\_264735169](https://www.researchgate.net/figure/Representacao-de-um-neuronio-artificial_fig3_264735169)).

O termo adicional  $b$  (bias ou viés) provê um grau de liberdade a mais, que não é afetado pela entrada nessa expressão. O valor  $z$  passa então por uma função de ativação, sendo esta não linear e responsável por limitar

$$z = \sum_{i=1}^N x_i w_i + b$$

**Figura 3:** Representação da equação da função soma com bias.

tal valor a um certo intervalo, produzindo o valor final de saída  $y$  do neurônio. Algumas funções de ativação usadas são a degrau, sigmoide e tangente hiperbólica, em que a utilizada no treinamento da rede do projeto foi a função sigmoide.



**Figura 4:** Representação gráfica da função sigmoide. Fonte: ([https://pt.wikipedia.org/wiki/Fun%C3%A7%C3%A3o\\_log%C3%ADstica](https://pt.wikipedia.org/wiki/Fun%C3%A7%C3%A3o_log%C3%ADstica)).

A função de ativação sigmoide é utilizada por meio da fórmula:

$$Y = 1 / (1 + \exp(-x)) \quad (1)$$

As RNAs possuem quatro maneiras de aprendizado, sendo elas o aprendizado supervisionado, aprendizado não-supervisionado, aprendizado por reforço e aprendizado por competição. Dentro do contexto utilizado de aprendizado supervisionado, apresentamos à rede as entradas disponíveis e a saída desejada, e o algoritmo opera no sentido de ajustar os pesos das sinapses por meio do cálculo da diferença entre o valor da saída desejada  $y_d(t)$  e o valor predito pela RNA  $y_p(t)$ , no instante  $t$ , produzindo desta forma um erro  $e(t)$ . Caso a saída obtida pela após o cálculo da função de ativação seja diferente da saída esperada, é necessário que haja o ajuste dos pesos sinápticos e uma nova iteração será realizada. A fórmula do erro pode ser observada a seguir:

$$e(t) = y_d(t) - y_p(t) \quad (2)$$

A forma genérica para o ajuste dos pesos, por correção de erros, é apresentada na equação a seguir:

$$w_{ij}(t+1) = w_{ij}(t) + e_i(t)x_i(t) \quad (3)$$

em que  $\alpha$  é a taxa de aprendizado e  $x_i(t)$  é a entrada para o neurônio  $i$  no tempo  $t$ .

Ao realizar a combinação de diversos desses neurônios em uma estrutura com camadas, é formada uma rede neural com múltiplas camadas chamada de Perceptron Multicamadas (Multilayer perceptron ou MLP). O MLP é composto de uma camada de entrada, uma ou mais camadas intermediárias (chamadas de camadas escondidas ou hidden layer) e uma camada de saída.

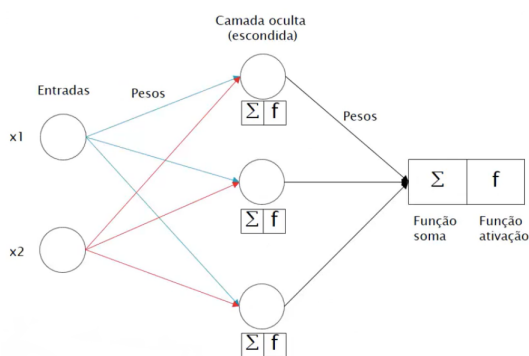


Figura 5: Representação gráfica da função sigmoide.

Nesse modelo, os valores de entrada  $x$  passam pela camada inicial, há o processo de multiplicação por seus respectivos pesos e é obtido uma saída para cada camada com base na função soma e função de ativação, logo após os valores de saída são ligados às entradas da camada seguinte, e assim por diante, até ser ter como resultado um valor na camada de saída. Todo esse processo descrito, da rede fornecer um valor de entrada, processar e retornar uma resposta é chamada de feed-forward. Como consequência, a entrada se propagava ao longo da rede, camada a camada em um sentido progressivo. É válido ressaltar que uma maior quantidade de camadas ocultas resulta em uma maior complexidade da rede, tornando-a mais apta a resolução de problemas mais complexos.

O responsável por tornar MLP extremamente presente nos dias atuais é o algoritmo de retropropagação do erro, ou backpropagation. Como foi visto que as RNAs aprendem com base no erro e em sua atualização, esse algoritmo foi extremamente revolucionário pois apresentou uma maneira de atualizar não apenas os pesos da camada de saída, mas também os pesos das camadas intermediárias. De maneira resumida, o erro é calculado por meio do que a rede tinha como saída esperada e o que de fato foi obtido pela rede, com isso, é feito novamente o cálculo de todos os pesos, começando da última camada e indo até a primeira, sempre com o objetivo de diminuir o erro.

O algoritmo backpropagation, de modo resumido consiste de dois passos, sendo um para frente e outro para trás. Esse passo para frente é o processo de feedforward já descrito, na segunda fase, as saídas são comparadas com os valores desejados, produzindo um sinal de erro que se propaga da saída para a entrada, ajustando-se os pesos de forma a minimizar o erro. Desse modo, a forma de calcular o erro depende da camada em que se encontra o neurônio. Como há o objetivo de diminuir o valor da função de erro, é efetuado o cálculo dos valores dos gradientes para cada peso da rede, e para minimizar o erro, é necessário que a rede tome o caminho oposto ao gradiente em relação aos pesos de cada camada. A atualização começa na camada de saída e os erros são propagados para trás, sendo fornecidos os sinais de erro que são correspondentes as sinapses da camada anterior, de modo que ele atualize os seus pesos.

A equação a seguir mostra como é feita a atualização pesos na camada de saída:

$$w(n+1) = w(n) + \alpha \cdot f' \cdot e \cdot y \quad (4)$$

- $w(n+1)$  representa o valor atual do peso sináptico.
- $w(n)$  representa o valor anterior do peso sináptico.
- $\alpha$  é a taxa de aprendizagem.
- $f'$  é a derivada da função de ativação (no caso, a função sigmoide) do neurônio da camada de saída cujo peso sináptico está sendo atualizado.
- $e$  representa o erro obtido (o erro é calculado de acordo com a diferença entre a saída desejada e o valor da função de ativação do neurônio o qual peso está sendo atualizado).
- $y$  representa o valor obtido pelo neurônio da camada anterior ligado ao peso sináptico que está sendo atualizado.

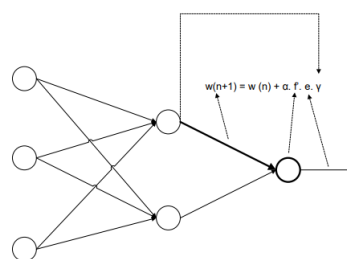


Figura 6: Imagem representando a atualização de pesos nas camadas de saída.

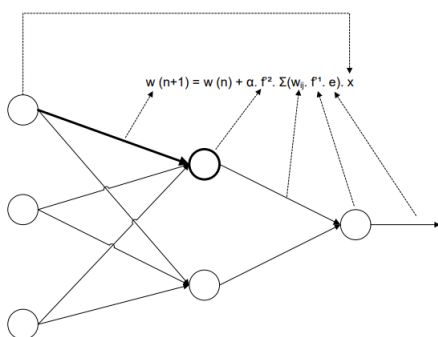
A seguir é tratado sobre a atualização de pesos nas camadas anteriores, essa forma ocorre de maneira diferente pois essas camadas não possuem saídas desejadas, desse modo, a atualização ocorre com a consideração proporcional do erro obtido em cada neurônio da saída.

$$W(n+1) = w(n) + .f'.\Sigma(wij.f'.e).x \quad (5)$$

é a taxa de aprendizagem.

$f'$  é a primeira derivada da função de ativação do neurônio da camada intermediária onde termina o peso sináptico

$\Sigma$  é o somatório do produto de cada peso sináptico (que liga o neurônio onde termina o peso sináptico que está sendo atualizado aos neurônios posteriores) multiplicado pela derivada da função de ativação do neurônio seguinte e pelo erro obtido no neurônio seguinte.



**Figura 7:** Imagem representando a atualização de pesos nas camadas ocultas.

Todas essas iterações serão realizadas durante o período de treinamento da rede, de modo que o erro seja minimizado, ou até um certo objetivo, ou até um determinado número de iterações disposto no código (também chamado de épocas). É válido ressaltar que é impossível uma rede complexa obter 100% de precisão e sempre acertar em casos tanto de treino quanto reais.

### 3 MATERIAIS E MÉTODOS

O banco de dados utilizado nesse projeto, consiste em 569 dados provenientes de pacientes com suspeita de câncer de mama, obtidos junto a base pública do Wisconsin Diagnostic Breast Cancer (WDBC). Os dados analisados possuem informações que se referem a: raio

(média das distâncias do centro aos pontos no perímetro), textura (desvio padrão dos valores da escala de cinza), perímetro, área, suavidade (variação local nos comprimentos do raio), espessura (perímetro / área - 1,0), concavidade (rigor das porções côncavas do contorno), número de pontos côncavos (número de partes côncavas do contorno), simetria e dimensão fractal da lesão ("aproximação da costa- 1).

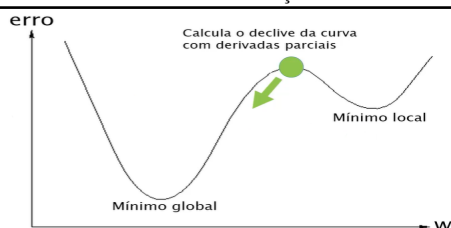
A simulação foi feita utilizando da linguagem de programação Python, e como descrito, foi implementado uma rede MLP com 30 atributos e 569 instâncias, representando então os neurônios de entrada por uma matriz 569x30. Os valores alvo ou saídas desejadas são representados por uma matriz 569x1, na qual toda essa matriz é preenchida por 0 ou 1, indicando o tipo de tumor desejado pela base de dados, sendo 0 representado por um tumor benigno, e 1 por um tumor maligno. A tabela a seguir mostra todas as informações utilizadas na implementação da rede:

**Tabela 1:** Estilos a serem utilizados.

Parâmetros	Valor
Nº de simulações	20
Nº máximo de épocas	100000
Função de ativação	Sigmóide
Algoritmo de treinamento	Backpropagation
Nº de camadas ocultas	2
Nº de neurônios da 1ª camada oculta	7
Nº de neurônios da 2ª camada oculta	4
Inicialização dos pesos	Aleatórios entre 0 e 1
Critério de parada (erro)	0,001

O algoritmo utilizado para treino da rede foi o backpropagation, com o objetivo de realizar a descida do gradiente com o objetivo de minimizar o erro. O gráfico a seguir mostra de maneira simplificada a atuação de descida do gradiente de acordo com os resultados obtidos durante o treinamento de uma RNA:

A descida do gradiente funciona de acordo com dois parâmetros: os pesos (representados por  $w$ ) e o erro obtido durante o treinamento, no qual os pesos são alterados de acordo com a aplicação da derivada parcial na ativação de neurônios para que a rede busque minimizar o erro por meio da busca de um mínimo global. É necessário ter cuidado na aplicação desse algoritmo, pois a rede pode encontrar mínimos locais, que dependendo de casos e do algoritmo utilizado, a rede pode se tornar custosa e deixar de atingir seu potencial ótimo devido a



**Figura 8:** Gráfico demonstrando a descida do gradiente em busca do mínimo global.

esses acontecimentos.

#### 4 RESULTADOS E DISCUSSÃO

A avaliação da rede neural foi realizada por meio de matrizes de resultados obtidos de casos teste gerada com a linguagem Python, essa matriz foi obtida após o treinamento de cada rede. Para se obter um padrão e medir a precisão da rede, foram feitas 20 simulações, a análise dos resultados foi feita considerando o acerto final da rede na classificação dos nódulos, sendo a classe 0 representada por nódulos benignos e 1 por nódulos malignos.

**Tabela 2:** Estilos a serem utilizados.

Parâmetros	Valor
Nº de simulações	20
Precisão máxima	91,52%
Precisão mínima	84,58%
Precisão média	87,22%
Erro (melhor simulação)	0,041

O modelo obteve como melhor resultado em uma de suas simulações uma precisão de aproximadamente 91,52%, demonstrando uma precisão alta para um modelo com algoritmos básicos de otimização e de atualização de erros. O pior resultado da rede foi em uma simulação na qual resultou em uma precisão de 84,58%. Um set com mais neurônios na camada escondida e melhores algoritmos de otimização pode contribuir e trazer resultados melhores em termo de precisão, foi observado que há outras redes com a mesma base de dados que obtiveram até 98,53% de precisão, o que demonstra a importância do uso de tal rede no âmbito profissional.

#### 5 CONCLUSÃO

Esse projeto tem como objetivo obter resultados e a simulação da detecção de exames de câncer de mama,

usando de uma base de dados real para diferenciar os tipos tumores (benignos ou malignos) identificados em um exame de mamografia. Esse artigo ressalta a importância do uso de novas tecnologias e adesão das mesmas, fruto de pesquisas e desenvolvimento científico, para a evolução e obtenção de melhores resultados no âmbito médico.

Apesar dos resultados obtidos, com a aplicação dos modelos de rede neural de perceptrons de múltiplas camada na classificação de tumores nas mamas terem apresentados resultados promissores, é preciso percebermos a necessidade de aprofundar o estudo, pois pouco se sabe a respeito do uso de tal conhecimento na área médica nos dias atuais, principalmente no Brasil. Para tal, pretende-se futuramente elaborar um modelo híbrido estruturado na utilização de algoritmos de redes neurais artificiais, em conjunto com diversas técnicas para melhora na performance das mesmas, além de técnicas de visão computacional, sendo este, outro campo que vem sendo bastante desenvolvido, fruto de diversas pesquisas e investimento na área, tudo isso de modo para a avaliação do desempenho na classificação da lesão mamária, bem como na otimização dos hiperparâmetros do modelo. Ressalta-se que a técnica de inteligência artificial descrita neste artigo é extremamente genérica, apenas para fruto de simulação e estímulo de busca por conhecimento na área, essa mesma rede implementada pode ser utilizada para diversas outras bases de dados de caráter médico, apenas com mudança de alguns parâmetros específicos como a quantidade de camadas e algoritmo descrito, além do conjunto de amostras para entrada da rede.

#### REFERÊNCIAS

- BRAGA, A. P.; CARVALHO, A. C. P. L. F.; LUDEMIR, T. B. Redes neurais artificiais: teoria e aplicações. Rio de Janeiro: LTC, 2007.
- STUART J. RUSSEL. Inteligência Artificial: Uma Abordagem Moderna. 3ª edição, 2009.
- MCCULLOCH, W. S.; PITTS, W. H. A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, n. 5, p.115-133, 1943.
- ROSENBLATT, F. The perceptron: a probabilistic model for information storage and retrieval in the brain. Psychological Review, 1958.
- S. Haykin. "Redes neurais: princípios e prática 2 ed". Bookman Editora (2001).
- INCA. "Estimativas 2018. Incidência de Câncer no Brasil". Instituto Nacional do Câncer (2018).
- A.F.M. Agarap. On breast cancer detection: an application of machine learning algorithms on the wiscon-

sin diagnostic dataset. In “Proceedings of the 2nd International Conference on Machine Learning and Soft Computing”. ACM (2018), pp. 5-9.

S. Agrawal J. Agrawal. Neural network techniques for cancer prediction: A survey. *Procedia Computer Science*, 60 (2015), 769-774.

Andrioni, V.; Guingo, B. C.; Pereira, W. C. A.; Infantosi, A. F. C. (2010) “Rede neural artificial para redução de falso-positivos em mamografia detectados por diferença de gaussiana”. *Anais do XII Congresso Brasileiro de Engenharia Biomédica*. Tiradentes: v.1