

---

## **Redes neurais apoiando a tomada de decisão na análise de crédito bancário e detecção do câncer de mama**

Lidio Mauro Lima de Campos

Professor Adjunto III da Universidade Federal do Pará - UFPA na Faculdade de Computação / Instituto de Ciências Exatas e Naturais (ICEN). Doutor em Engenharia Elétrica com ênfase em Computação Aplicada pelo Programa de Pós-Graduação em Engenharia Elétrica - PPGGE da Universidade Federal do Pará  
limadecampos@gmail.com

Jonathan Cris Pinheiro Garcia

Graduação em Geofísica pela Universidade Federal do Pará. Graduando em Sistemas de Informação pela Universidade Federal do Pará  
jonathan.garcia00@hotmail.com

Editor Científico: José Edson Lara  
Organização Comitê Científico  
Double Blind Review pelo SEER/OJS  
Recebido em 03.07.2018  
Aprovado em 24.09.2018



Este trabalho foi licenciado com uma Licença Creative Commons - Atribuição – Não Comercial 3.0 Brasil

## Resumo

As Redes Neurais Artificiais (RNAs) tem sido utilizadas nas soluções de variados problemas, dentre eles, os que envolvem tomada de decisões. Neste escopo, o objetivo desta pesquisa é apresentar uma ferramenta que dê suporte ao processo de decisão médica no diagnóstico do câncer de mama e da concessão de crédito bancário a clientes de um banco, por meio da utilização de uma RNA Multilayer Perceptron. Verificando-se sua eficácia e a melhor convergência, por meio do Modelo de Validação Cruzada ( $k=5$ ), para o treinamento e testes de RNA's. Os testes foram realizados com diferentes números de neurônios na camada escondida, utilizando-se como comparativo o erro quadrático médio, visando otimizar o desempenho da RNA. Uma taxa de aprendizado em torno de 0,9 foi adequada nas simulações para os dois problemas. Os resultados elencados indicam a eficiência da técnica, para ambos os problemas, haja vista que a capacidade de generalização das RNAs testadas foi em média de 97,31% para o diagnóstico do câncer de mama e 99,1% para a concessão de crédito bancário.

**Palavras-chave:** Redes Neurais Artificiais; Tomada de Decisão; Validação Cruzada.

## Neural networks supporting decision-making in bank credit analysis and breast cancer detection

### Abstract

Artificial Neural Networks (ANNs) have been used in the solutions of several problems, among them, those involving decision making. In this scope, the objective of this research is to present a tool that supports the medical decision process in the diagnosis of breast cancer and the granting of bank credit to clients of a bank, through the use of a Multilayer Perceptron RNA. Verifying its effectiveness and the best convergence, through the Cross Validation Model ( $k = 5$ ), for the training and testing of ANNs. The tests were performed with different numbers of neurons in the hidden layer, using as comparison the average quadratic error, in order to optimize the RNA performance. A learning rate of around 0.9 was adequate in the simulations for the two problems. The results indicated the efficiency of the technique for both problems, given that the generalization capacity of the ANNs tested was on average 97.31% for the diagnosis of breast cancer and 99.1% for the granting of bank credit.

**Keywords:** Artificial Neural Networks; Decision Making; Cross Validation.

## Redes neurales apoyando la toma de decisión en el análisis de crédito bancario y detección del cáncer de mama

### Resumen

Las redes neuronales artificiales (ARNs) se han utilizado en las soluciones de variados problemas, entre ellos, los que involucran la toma de decisiones. En este ámbito, el objetivo de esta investigación es presentar una herramienta que apoye el proceso de decisión médica en el diagnóstico del cáncer de mama y la concesión de crédito bancario a clientes de un banco, por medio de la utilización de una RNA Multilayer Perceptron. Se verificó su eficacia y la mejor convergencia, por medio del Modelo de Validación cruzada ( $k = 5$ ), para el entrenamiento y pruebas de ARN's. Las pruebas se realizaron con diferentes números de neuronas en la capa oculta, utilizando como comparativo el error cuadrático medio, buscando optimizar el desempeño de la RNA. Una tasa de aprendizaje en torno a 0,9 fue adecuada en las simulaciones para los dos problemas. Los resultados señalados indican la eficiencia de la técnica para ambos problemas, teniendo en cuenta que la capacidad de generalización de las ARNs

probadas fue en promedio del 97,31% para el diagnóstico del cáncer de mama y el 99,1% para la concesión de crédito bancario 99%.

**Palabras clave:** Redes Neuronales Artificiales; Toma de decisión; Validación Cruzada.

## 1 Introdução

As Redes Neurais Artificiais (RNAs) são modelos matemático/computacionais que possuem unidades de processamento, interligadas entre si por conexões que representam pesos, que executam operações em paralelo e de forma distribuída (Haykin, 1994, 2001). As mesmas são ferramentas com a capacidade de implementar métodos estatísticos para reconhecimento de padrões (Nunes, 2012). Devido à sua generalidade, a escolha da arquitetura da RNA, que será utilizada na solução de um dado problema, depende do tipo de tarefa que o RNA irá executar.

As RNAs tem sido utilizadas de forma crescente para em diversos ramos de negócios: predição de séries temporais, (Szegedy *et al.*, 2014) reconhecimento humano usando a medida biométrica da orelha (Sánchez & Melin, 2014), (Sánchez, Melin, & Castillo, 2015) predição de radiação da energia solar (Chatziagorakis *et al.*, 2014), visão por computador (Vikas *et al.*, 2018), o reconhecimento de voz (Sak, Senior, & Beaufays, 2014), tradução automática (Sutskever, Vinyals, & Le, 2014), tomada de decisão no mercado financeiro (Gambogi, 2013), biometria (Karishma & Swati, 2018), (Nitin, Pranshu, & Upasana 2018) sistema de reconhecimento de vídeo (Ajay, Shubham, & Manan 2018), dentre outras (Skrizhevsky, Sutskever, & Hinton, 2012) e (Wang & Kosinski, 2017).

Conforme menciona (Gambogi, 2013) a previsão do comportamento do mercado financeiro continua sendo do interesse dos investidores, pois avaliações acertadas se traduzem em elevado retornos. Embora o acesso a informações financeiras tenha aumentado significativamente com o advento da internet, a tarefa de interpretá-las, de avaliar o seu impacto no mercado, continua sendo uma árdua tarefa, inclusive para as instituições que contam com recursos matérias e humanos especializados. Dessa forma, a utilização das redes neurais, tem sido amplamente estudadas visando oferecer suporte no processo de tomada de decisão sobre investimentos, devido a capacidade de mapear funções complexas.

As RNAs podem aprender a partir de exemplos, reconhecer um padrão escondido em observações históricas e usá-los para prever valores futuros. Além disso, são capazes de lidar com informação incompleta ou dados ruidosos e podem ser muito eficazes, especialmente em

situações onde não é possível definir as regras ou etapas que levam à solução de um problema. Possuem a capacidade de aprender ao interagir com seu ambiente e, dessa forma, extrair conhecimento do mesmo, o que é realizado por meio de um processo de treinamento, que consiste de um processo iterativo, todas essas características justificam a sua utilização, como ferramenta de apoio ao processo de tomada de decisão complexo.

Assim sendo, o ajuste, correto, dos parâmetros de uma RNA *Multi Layer Perceptron* (número de neurônios na camada intermediária, taxa de aprendizado, *momentum*) é essencial para um bom treinamento e para a validação da capacidade de generalização da mesma. Por isso, antes de se usar o modelo em produção, precisa-se avaliar o seu desempenho, ou seja, é necessário medir quanto o mesmo é eficiente em prever dados que ainda não conheceu, ou seja, testá-lo de acordo com sua capacidade de generalização.

Um dos métodos mais comuns de avaliação do modelo é chamado *hold out*. No qual separam-se os dados originais entre treino e teste. Os dados de treino são submetidos a RNA e produzem o modelo. Após isso, os dados de teste são submetidos ao mesmo para que seja feita a previsão. Normalmente, esse modelo usa 70% dos dados para treino e 30% para testes. Se o desempenho foi satisfatório o modelo pode ser colocado em produção para classificar novos dados. Uma vantagem do método de *hold out* é que os dados são totalmente independentes; além disso, só precisa ser executado uma vez, por isso tem custos computacionais mais baixos. Uma desvantagem é que a avaliação do desempenho está sujeita a uma maior variância, dado o menor tamanho dos dados.

Na técnica *k-fold cross validation* o conjunto de dados de treinamento, é dividido em  $k$  subconjuntos. Desses  $k$  subconjuntos, um é selecionado para ser utilizado para validação e os  $k-1$  restantes são utilizados no treinamento. O processo de validação é repetido  $k$  vezes, de modo que cada um dos  $k$  subconjuntos sejam utilizados pelo menos uma vez como dados de validação para o modelo. Após isso avalia-se o desempenho médio do modelo nos  $k$  testes. O objetivo desses testes diversas vezes é aumentar a confiabilidade da estimativa do modelo. Essa teoria de validação foi apresentada por (Geisser, 1975) e foi demonstrando que a precisão dos testes possui uma relação inversa com  $k$ . Uma vantagem do método é que o mesmo é sujeito a menor variância quando  $k$  aumenta. Como desvantagens tem-se que: os custos computacionais são maiores e o modelo precisa ser treinado  $k$  vezes a cada passo de validação (mais um a mais na etapa de teste).

O objetivo desse trabalho é utilizar a tecnologia de Redes Neurais Artificiais em dois problemas de classificação, no primeiro a rede classifica clientes em adimplentes ou inadimplentes, após a fase de reconhecimento dos padrões, a técnica é utilizada na análise de novos clientes, servindo como ferramenta auxiliar na tomada de decisão quanto ao crédito bancário. No segundo, auxilia na descoberta de padrões e informações úteis para o diagnóstico de tumores de câncer de mama. Para isso, nesse artigo demonstra-se a eficiência da técnica *k-fold cross validation* para validação de RNAs treinadas nesses dois problemas de classificação não triviais *Credit German e Breast Cancer*. Conclui-se que em ambos os problemas de classificação estudados o desempenho dos classificadores neurais foi acima de 96%.

Para que os objetivos da pesquisa sejam alcançados, na seção 2 faz-se a revisão da literatura, na seção 3 apresentam-se conceitos sobre Redes Neurais Artificiais, na seção 4 detalha-se técnica de *k-fold cross validation*. Na seção 5, apresentam-se a execução das etapas da metodologia e os resultados de simulação obtidos e finalmente nas seções 6 e 7 as conclusões e as referências bibliográficas consultadas.

## 2 Revisão de Literatura

O trabalho de (Bonini, 2016) descreve experimentos utilizando o classificador árvore de decisão, algoritmo J.48, para extração de conhecimento de *dataset* com dados sobre câncer de mama. No mesmo detalham-se as características dos dados que compõem o *dataset* e as estratégias utilizadas para realização dos experimentos. O algoritmo classificou mais de 96% de amostras corretamente.

Holsbach, Fogliatto e Anzanello (2014) apresentam um método para a seleção de variáveis para classificação dos casos em duas classes de resultado, benigno ou maligno, baseado na análise citopatológica de amostras de célula da mama de pacientes. As variáveis são ordenadas de acordo com um novo índice de importância de variáveis que combina os pesos de importância da Análise de Componentes Principais e a variância explicada a partir de cada componente retido. Observações da amostra de treino são categorizadas em duas classes através das ferramentas *k-vizinhos mais próximos* e Análise Discriminante, seguida pela eliminação da variável com o menor índice de importância. Usa-se o subconjunto com a máxima acurácia para classificar as observações na amostra de teste. Aplicando ao *Wisconsin Breast Cancer Database*, o método proposto apresentou uma média de 97,77% de acurácia de classificação, restando uma média de 5,8 variáveis.

Arora *et al.* (2008) analisam um software de CAD térmico que possui três modos de operação: triagem, clínico e rede neural artificial. Os autores não detalham as características extraídas e nem como o programa funciona, as acurácias obtidas nos três modos: triagem, 66,70%; clínico, 71,40%; e rede neural artificial, 81,80%.

Steiner *et al.* (2004) abordam, comparativamente, duas técnicas (Redes Neurais e Árvores de decisão) para a análise de concessão de crédito bancário, por uma instituição financeira, utilizando o dataset *credit german*. O trabalho apresenta uma proposta, através de análises quantitativas, para o reconhecimento da qualidade do perfil de crédito de determinados clientes, classificando os adimplentes dos inadimplentes. A pesquisa utiliza duas técnicas, Redes Neurais e Árvores de Decisão. Como variáveis de entrada, foram utilizadas sete informações de 2.855 clientes com respostas binárias (sim ou não). Além disso, a informação se houve ou não cumprimento sobre o crédito.

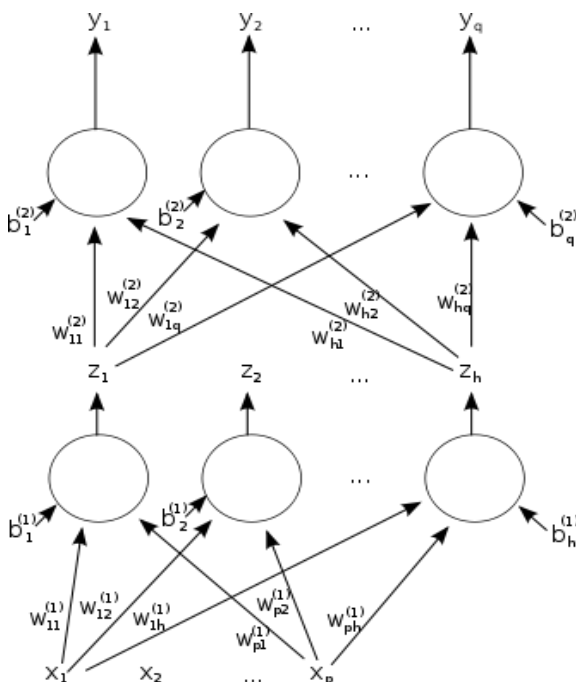
O trabalho de Gambogi (2013) apresenta um sistema de *trading* que toma decisões de compra e de venda do índice *Standard & Poors 500*, na modalidade seguidor de tendência, mediante o emprego de Redes Neurais Artificiais multicamadas com propagação para frente, no período de 5 anos, encerrado na última semana do primeiro semestre de 2012. Geralmente o critério usual de escolha de redes neurais nas estimativas de preços de ativos financeiros é o do menor erro quadrático médio entre as estimativas e os valores observados. Na seleção das redes neurais foi empregado o critério do menor erro quadrático médio na amostra de teste, entre as redes neurais que apresentaram taxas de acertos nas previsões das oscilações semanais do índice *Standard & Poors 500* acima de 60% nessas amostras de teste. Esse critério possibilitou ao sistema de *trading* superar a taxa anual de retorno das redes neurais selecionadas pelo critério usual e, por larga margem, a estratégia de compre e segura no período. A escolha das variáveis de entrada das redes neurais recaiu entre as que capturaram o efeito da anomalia do momento dos preços do mercado de ações no curto prazo, fenômeno amplamente reconhecido na literatura financeira.

### 3 Redes Neurais Artificiais

Redes Neurais Artificiais são inspiradas biologicamente no funcionamento do cérebro, assim uma rede neural típica é constituída por um conjunto de neurônios interligados, influenciando uns aos outros formando um sistema maior capaz de reconhecer padrões por meio

de treinamento. O algoritmo de aprendizado generaliza os dados e memoriza o conhecimento dentro dos parâmetros adaptáveis da rede, denominado de pesos.

Uma **Rede Direta**, como ilustrado na Figura 1, consiste em várias camadas compostas por nós (neurônios) onde cada neurônio de uma camada possui ligação com todos os neurônios da camada seguinte. Conforme visto na Figura 1, a rede direta, denominada *perceptron* multicamada, possui uma camada de entrada, que atua como um conjunto de sensores da rede captando os estímulos do ambiente e pode ter uma ou mais camadas intermediárias que é onde a maior parte do processamento é realizada através das conexões e seus pesos respectivos, podem ser considerados como extratoras de características e uma camada de saída onde o resultado final é concluído e apresentado. Ainda na Figura 1 não existem conexões entre a saída de um neurônio e algum outro neurônio localizado em uma camada anterior ao primeiro, ou seja, não possui ciclos, fato que caracteriza uma rede *feedforward*.



**Figura 1.** Rede Perceptron Multicamada  
Fonte: Wikipedia.

A configuração pela qual os neurônios de uma RNA e estruturada depende do algoritmo de aprendizagem a ser utilizado para o treinamento. No presente trabalho e usado uma rede neural multicamadas compostas de camadas alinhadas e totalmente conectadas diretamente com uma camada de neurônios. Neste tipo de rede, as entradas são apresentadas na primeira camada, que é chamada camada de entrada. Esta distribui as informações de entrada para a

camada oculta da rede, que por sua vez pode possuir uma ou mais camadas ocultas e uma camada de saída (Kovács, 2006). O valor de saída é obtido através da sequência de funções de ativação definido na camada oculta.

A aprendizagem é uma das características notáveis das RNA's. No aprendizado supervisionado, o treinamento possui um conhecimento sobre os dados, uma especie de instrutor que confere o quanto a rede está próxima de uma solução aceitável, este conhecimento, está representado sob forma de um conjunto de amostras de *entrada-saída*. O processo de treinamento modifica os pesos das RNAs com a finalidade de melhorar um determinado critério de desempenho, de tal forma que, para o conjunto de entrada informado, a rede seja capaz de calcular uma saída o mais próximo possível da saída desejada.

A aprendizagem de Redes Multilayer Perceptron (MLP) é um processo iterativo, conhecido como aprendizagem por experiência, no qual padrões de treinamento (exemplos) são apresentados a rede e com base nos erros obtidos, são realizados ajustes nos pesos sinápticos, com o intuito de diminuir o erro nas próximas iterações. Sua estrutura é formada pela camada de entrada, que atua como os sensores da rede captando os estímulos do ambiente e pode ter uma ou mais camadas intermediárias que é onde a maior parte do processamento é realizada através das conexões e seus pesos respectivos, podem ser considerados como extratoras de características e uma camada de saída onde o resultado final é concluído e apresentado.

O treinamento de um Perceptron de Múltiplas Camadas (MLP) consiste em ajustar os pesos e os *thresholds* (bias) de suas unidades para que a classificação desejada seja obtida. Quando um padrão é inicialmente apresentado à rede, ela produz uma saída e, após medir a diferença entre a resposta atual e a desejada, são aplicados ajustes apropriados nos pesos de modo a reduzir esta distância. Este procedimento é conhecido como Regra Delta. O algoritmo mais utilizado para o treinamento destas redes MLP é uma generalização da Regra Delta denominada de Backpropagation.

O **Backpropagation** é baseado na regra de aprendizagem por correção de erro. Durante o treinamento com o algoritmo Backpropagation, a rede opera em uma sequência de dois passos. No primeiro, um padrão é apresentado à camada de entrada da rede. O sinal resultante se propaga através da rede, camada por camada, até que a resposta seja produzida pela camada de saída. No segundo passo, a saída obtida é comparada à saída desejada para esse padrão particular. Se esta não estiver correta, o erro é calculado. Este erro é propagado a partir da camada de saída até a camada de entrada, e os pesos das conexões das unidades das camadas



internas vão sendo modificados à medida que o erro é retropropagado. Ao utilizar a retro propagação o algoritmo aprende de forma supervisionada, em tempo discreto e utiliza um método de gradiente descendente para correção de erro, assim a codificação executa um mapeamento entrada-saída, através da minimização de uma função custo. Está por sua vez, realiza iterativamente ajustes nos pesos sinápticos de acordo com o erro quadrático acumulado para o conjunto de treinamento (Fabro, 2001). O processo de evolução da redução gradativa de erro pode levar a convergência. A medida que a rede aprende, o valor do erro converge para um valor estável, assim o processo prossegue, até o critério de um valor mínimo de erro global seja atingido ou o número de épocas (interações) seja alcançado. Os passos do algoritmo de retro propagação são descritos abaixo.

- 1) Inicializar os pesos da rede. Cada peso deve ser ajustado aleatoriamente para um número entre  $-0.1$  e  $0.1$ .

$w_{ij}^1 = \text{aleatório}(-0.1;0.1)$ ,  $w_{ij}^2 = \text{aleatório}(-0.1;0.1)$  para todo  $i=0, \dots, A$ ;  $j=1, \dots, B$ ; para todo  $i=0, \dots, B$ ;  $j=0, \dots, C$ .

- 2) Inicializar as ativações das unidades limite. Seus valores nunca mudam.

$x_o = 1.0$  e  $h_o = 1.0$

- 3) Escolher um par de padrão de entrada-saída. Supondo que o vetor de entrada seja  $x_i$  e o vetor de saída seja  $y_j$ . Atribuem-se níveis de ativação às unidades de entrada.

- 4) Propagar a ativação das unidades da camada de entrada para as unidades da camada oculta, usando a função de ativação.

$$h_j = \frac{1}{1 + e^{-\sum_{i=0}^A w_{ij} \cdot x_i}}, \text{ para todo } j=1, \dots, B \quad (1)$$

- 5) Propagam-se as ativações das unidades da camada oculta para as unidades da camada de saída, usando a função de ativação.

$$o_j = \frac{1}{1 + e^{-\sum_{i=0}^B w_{2ij} \cdot x_i}} \text{ para todo } j=1, \dots, C \quad (2)$$

- 6) Computar os erros das unidades da camada de saída, denotados por  $\delta 2_j$ . Os erros baseiam-se na saída real da rede ( $o_j$ ) e na saída ( $y_j$ ).

$$\delta 2_j = o_j(1 - o_j)(y_j - o_j) \text{ para todo } j=1, \dots, C \quad (3)$$

- 7) Computar os erros das unidades da camada oculta, denotados por  $\delta 1_j$ .

$$\delta 1_j = h_j(1 - h_j) \sum_{i=0}^C \delta 2_i \cdot w_{2ij} \text{ para todo } j=1, \dots, B \quad (4)$$

- 8) Ajuste dos pesos entre a camada oculta e a camada de saída. O coeficiente de aprendizagem é denotado por  $\eta$ , sua função é a mesma de na aprendizagem por perceptrons.

$$\Delta w_{2ij}(t+1) = \Delta w_{2ij}(t) + \eta \delta 2_j \cdot h_i \text{ para todo } i=0, \dots, B; j=1, \dots, C \quad (5)$$

Ajuste os pesos entre a camada de entrada e a camada oculta

$$\Delta w_{1ij}(t+1) = \Delta w_{1ij}(t) + \eta \delta 1_j \cdot x_i \text{ para todo } i=0, \dots, A; j=1, \dots, B \quad (6)$$

Vá para a etapa 4 e repita. Quando todos os pares entrada-saída tiverem sido apresentados à rede, uma época terá sido completada. Repita as etapas de 4 a 10 para tantas épocas quantas forem desejadas.

As equações (7) e (8) mostram as fórmulas para atualização dos pesos, da camada de saída e da camada intermediária, respectivamente.

$$\Delta w_{2_{km}}(t+1) = \Delta w_{2_{km}}(t) + \eta [(y_{nm} - o_m)(1 - o_m) \cdot o_m] \cdot h_k \quad (7)$$

$$\Delta w_{1_{jk}}(t+1) = \Delta w_{1_{jk}}(t) + \eta [(1 - h_k) h_k \cdot \sum_{m=0}^C (y_{nm} - o_m)(1 - o_m) \cdot o_m \cdot w_{2_{km}}] x_{nj} \quad (8)$$

#### 4 Técnica De Validação K-Fold Cross Validation

Uma das etapas mais importantes em um projeto de classificadores é a etapa de validação dos resultados. Muitas das técnicas mais poderosas de aprendizado apresentam uma grande quantidade de parâmetros e quanto menos restrições colocarmos no nosso modelo maior a probabilidade de encontrarmos um super ajustamento, ou como é mais conhecido, o *overfitting*.

*Overfitting* ocorre quando o método que estimamos consegue bons resultados apenas nos dados que eles foram treinados. Já na presença de novas observações, percebemos uma grande piora na qualidade da predição. Ou seja, o *overfitting* ocorre quando o método de aprendizado não consegue generalizar os resultados para dados que não foram utilizados no processo de treino.

A técnica *k-fold Cross Validation* estima o erro do método de aprendizado em observações não utilizadas no treino, ou seja, estima como o modelo construído irá se comportar em novos dados, claro que isto é válido apenas se mantivermos a mesma probabilidade conjunta das variáveis explicativas e da variável resposta utilizada durante o treino. O *k-fold Cross Validation* consiste em dividir a base em *k* pedaços. Para cada pedaço, estimamos o método sem a presença desta parte e verificamos o erro médio no pedaço não utilizado durante o treino. Abaixo, na Figura 2, apresenta-se a descrição desse algoritmo.

1. Arranjar os exemplos de treinamento em ordem randômica
2. Dividir os exemplos de treinamento em “*k*” conjuntos não sobrepostos  $D_1, D_2, \dots, D_k$ . (*K* pedaços de aproximadamente  $D/k$  exemplos cada).
3. Para  $i=1 \dots k$ ;  
Treine o classificador usando todos os exemplos que não pertencem ao conjunto  $i$  ( $D \setminus D_i$ ).  
Teste classificador em todos os exemplos no conjunto  $i$  ( $D_i$ ).  
Computar  $n_i$ , o número de exemplos no conjunto  $i$  que foram classificados erradamente.
4. Retorne a seguinte estimativa de erro para o classificador:  
$$E = \frac{\sum_{i=1}^k n_i}{D} * 100\%$$
  
OBS: Para  $t$  execuções do algoritmo:  $E = \frac{\sum_{i=1}^t E_i}{t}$

**Figura 2.** Algoritmo *k-fold cross-validation*

Fonte: elaborado pelos autores (2016)

## 5 Experimentos Realizados

Nessa secção apresentam-se os a descrição dos experimentos e os resultados de simulação para os dois problemas estudados “German Credit” e “Breast Cancer”.

### 5.1 German Credit Data

O primeiro “dataset” escolhido para análise foi o *German Credit Data* acessado no seguinte repositório: [https://archive.ics.uci.edu/ml/datasets/statlog+\(german+credit+data\)](https://archive.ics.uci.edu/ml/datasets/statlog+(german+credit+data)).

O conjunto de dados contém observações sobre 20 variáveis para 1000 candidatos a crédito de uma instituição financeira. Cada um deve ser classificado como "bom crédito" (700 casos) ou "crédito ruim" (300 casos). Novos requerentes de crédito também podem ser avaliados com base nessas 20 variáveis preditoras. Desejou-se desenvolver uma regra de classificação de crédito que pudesse ser usada para determinar se um novo candidato representa um baixo ou alto risco para pagar um empréstimo.

Assim, quando um banco recebe um pedido de empréstimo, com base no perfil do candidato, o banco deve tomar uma decisão sobre se deve prosseguir com a aprovação do empréstimo ou não. Dois tipos de riscos estão associados à decisão do banco. Se o candidato tem um bom risco de crédito, ou seja, é provável que pague o empréstimo, então não aprovar o empréstimo para a pessoa resulta em perda de negócios para o banco. Ou se o candidato tem um risco de crédito ruim, ou seja, não é provável que pague o empréstimo, então, aprovar o empréstimo para a pessoa resulta em uma perda financeira para o banco.

Para melhor compreensão do trabalho desenvolvido é importante descrever os dados que compõem o *dataset credit german*. A Figura 3 apresenta os mesmos e seus significados.

#### 5.1.1 Treinamento e Testes das Redes Neurais

Para o treinamento e validação dos resultados obtidos pela Rede Neural utilizou-se a técnica de validação cruzada *K-fold*. Nesse método, o conjunto de dados original é dividido em *K* subconjuntos. Destes *K* subconjuntos, um subconjunto é retido para ser utilizado na validação do modelo e os *K-1* subconjuntos restantes são utilizados no treinamento da rede neural. O processo de validação cruzada é, então, repetido *K* vezes, de modo que cada um dos *K*

subconjuntos sejam utilizados exatamente uma vez como dados de teste para validação do modelo. O resultado final desse processo mede o desempenho médio do classificador nos K testes.

O conjunto original dos dados do *dataset credit german* foi dividido em cinco conjunto de dados ( $k = 5$ ), ou seja, cada partição de dados, denominados (D1, D2, D3, D4, D5) ficou com duzentos registros. A partir disso, utilizou-se uma partição de dados fixa para testes alternada com as demais para treino da rede neural. Como pode-se observar na Tabela 1, os primeiros 200 registros foram denominados de D1 (dados destacados em amarelo), esse conjunto foi mantido fixo para testar as RNAs e os demais subconjuntos D2, D3, D4 e D5 foram usados para treinar as redes neurais, ou seja, na primeira iteração, foram utilizadas as seguintes partições de dados para {treino, teste} = {{D2, D1}, {D3, D1}, {D4, D1}, {D5, D1}}. Sendo esse processo repetido, para as demais partições de dados. A primeira coluna da Tabela 1, ilustra as partições fixas usadas nos testes das redes neurais e a segunda coluna, as partições utilizadas para os treinos das redes neurais.

Em cada iteração do método de validação cruzada, ilustrada em cada linha da Tabela 1, utilizaram-se para treino de cada rede neural os parâmetros que estão destacados na coluna 5 (número de neurônios na camada intermediária), na coluna 6 (taxa de aprendizado), na coluna 7 (número de épocas de treinamento da rede neural). As colunas 3,4,6,8 e 9 apresentam os resultados dos testes das redes neurais. Na coluna 3 apresentam-se número de instâncias classificadas incorretamente, na coluna 4 a quantidade de instâncias classificadas corretamente, na coluna 8 apresentam-se as taxas de acertos, na coluna 9 a média total de acertos.

Os testes foram feitos estimando parâmetros de forma aleatoriamente, a medida que a rede respondia a cada alteração dos parâmetros avaliavam-se os resultados até se obter a melhor combinação de parâmetros para os quais a rede neural apresentava a maior taxa de acertos, sendo esses registrados na Tabela 1. Por exemplo, variou-se, por exemplo, a taxa de aprendizado e os demais parâmetros foram mantidos constantes, o que se observou foi que o treinamento da rede convergia mais rápido, quando se estipulou o valor de 0.9991, para a taxa de aprendizado. Esse mesmo procedimento foi utilizado para ajustar o número de neurônios na camada intermediária, variou-se esse parâmetro, mantendo-se fixos os demais, na Tabela 1 ilustram-se os melhores resultados obtidos. Observou-se que se aumentado a quantidade de neurônios, a partir de 7, não necessariamente produzia uma melhor acurácia na classificação por parte da rede neural, pois alguns resultados apresentaram acurácias equivalentes.

1 - (qualitativo)	Status da conta corrente
2 - (Numérico)	Duração do mês
3 - (qualitativo)	Histórico de credito
4 - (qualitativo)	Objetivo do empréstimo
5 - (numérico)	Quantidade de credito
6 - (qualitativo)	Conta poupança
7 - (qualitativo)	tempo de emprego atual
8 - (numérico)	Taxa de parcelamento em percentagem do rendimento disponível
9 - (qualitativo)	Status pessoal e sexo
10 - (qualitativo)	Outros devedores / garantias
11 - (numérico)	Tempo na Residência atual
12 - (qualitativo)	Propriedades
13 - (numérico)	Idade em anos
14 - (qualitativo)	Outros planos de parcelamento
15 - (qualitativo)	Habitação
16 - (numérico)	Número de créditos existentes neste banco
17 - (qualitativo)	Trabalho
18 - (numérico)	Número de pessoas sujeitas a manutenção
19 - (qualitativo)	Telefone
20 - (qualitativo)	Trabalhador estrangeiro

**Figura 3.** Descrição dos Atributos para a base de dados sobre Banco de credito alemão  
Fonte: elaborado pelos autores (2018).

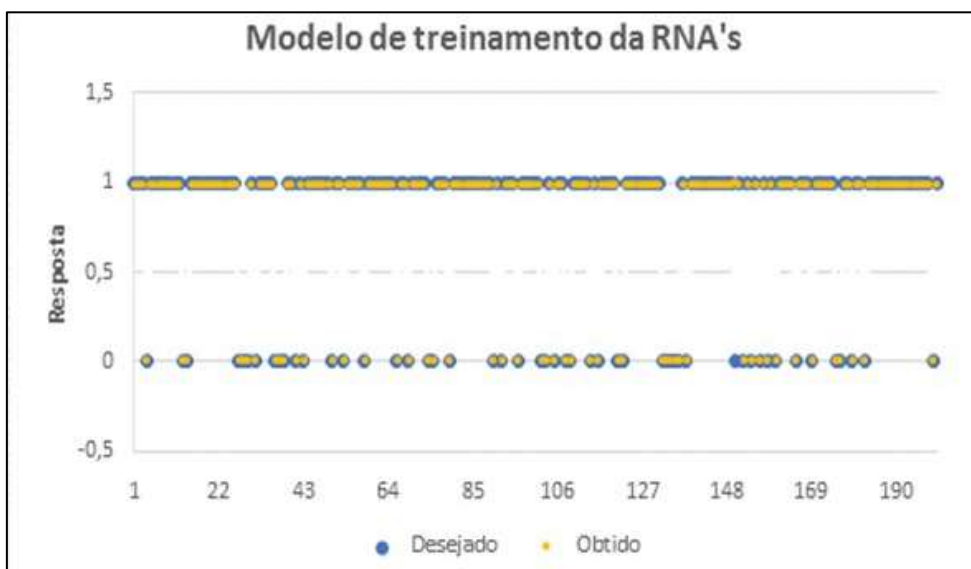
**Tabela 1**

Tabela de testes para a base de dados sobre Banco de credito alemão

Teste	Treino	Erros	Acertos	Neurônios Camada Intermediária	Taxa de Aprendizizado	Épocas	% Acertos	Media
D1	D2	1	199	7	0.9991	70000	99,5	99
	D3	2	198	7	0.9991	70000	99	
	D4	0	200	7	0.9991	70000	100	
	D5	5	195	7	0.9991	70000	97,5	
D2	D1	0	200	7	0.9991	70000	100	99,125
	D3	2	198	7	0.9991	70000	99	
	D4	0	200	7	0.9991	70000	100	
	D5	5	195	7	0.9991	70000	97,5	
D3	D1	0	200	7	0.9991	70000	100	99,5
	D2	1	199	7	0.9991	70000	99,5	
	D4	0	200	7	0.9991	70000	100	
	D5	3	197	7	0.9991	70000	98,5	
D4	D1	0	200	7	0.9991	70000	100	98,25
	D2	4	196	7	0.9991	70000	98	
	D3	5	195	7	0.9991	70000	97,5	
	D5	5	195	7	0.9991	70000	97,5	
D5	D1	0	200	7	0.9991	70000	100	99,625
	D2	1	199	7	0.9991	70000	99,5	
	D3	2	198	7	0.9991	70000	99	
	D4	0	200	7	0.9991	70000	100	

Fonte: elaborado pelos autores (2018).

Na Figura 4, apresenta-se uma comparação entre os valores obtidos e desejados, para uma única etapa de treino-teste de uma rede neural. No eixo Y, a resposta varia entre 0 e 1, no eixo X em azul são destacados os valores reais do modelo do *dataset credit german*, e em amarelo as respostas da rede após o treino. Para o treinamento da rede neural, foram utilizados 200 registros do dataset, sendo que a rede conseguiu predizer corretamente 199 com a taxa de acerto de 99,5 %.



**Figura 4.** Exemplo de Teste de uma Rede Neural

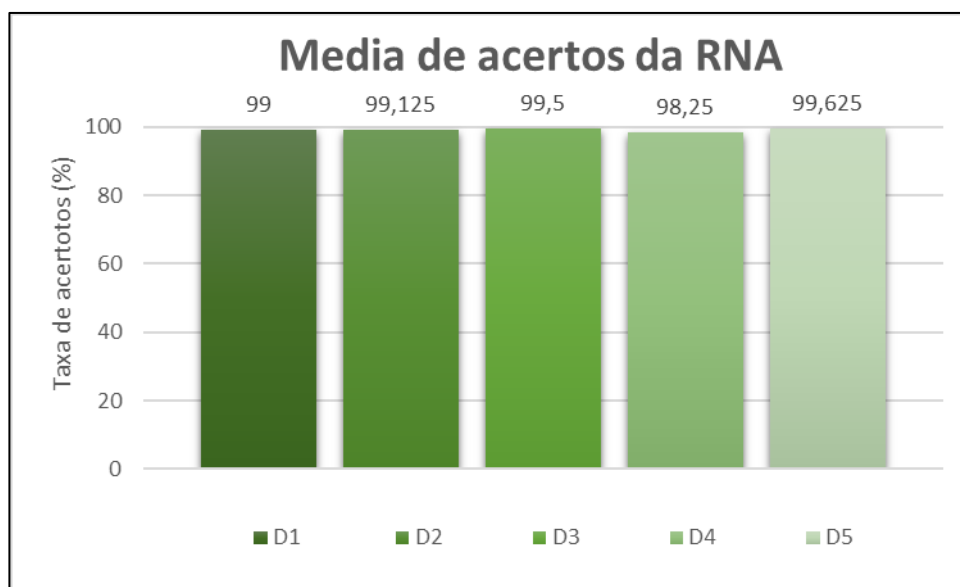
Fonte: elaborado pelos autores (2018).

Para a solução do problema de previsão de credito os melhores resultados foram obtidos com os seguintes parâmetros: (número de neurônios na camada de entrada, número de neurônio na camada intermediária, números de neurônios na camada de saída, taxa de aprendizado, número de épocas de treinamento, erro médio quadrático) = (26,7, 1, 0,9, 70000, 0.0001). Após o treinamento de cada rede neural, a mesma foi utilizada para testar a generalização com dados que não foram utilizados no treinamento, a Figura 4, ilustra uma das generalizações, para uma taxa de acerto de 99,5%. O estudo de generalização é de suma importância, pois indica se a rede é capaz de predizer um padrão corretamente, sem que o mesmo tenha sido utilizado na etapa de treinamento da RNA.

Em relação aos resultados obtidos, utilizando-se um conjunto de dados reais de aplicações de crédito, confirmam-se eficiência do modelo e da metodologia utilizada para a solução do problema. O objetivo foi comparar o desempenho da rede neural utilizando um apanhado de dados classificadas pelos especialistas com o mesmo volume de dados

classificadas pelo K-fold e avaliar se a rede neural poderia ser utilizada neste contexto.

A Figura 5 também resume os resultados das 20 simulações realizadas para o dataset “credit german” o gráfico de acertos, mostra a percentagem de acertos obtidos na fase de validação das redes, ou seja, após o treinamento das RNAs convergiram para um erro quadrático de 0,0001. Testou-se a capacidade das RNAs em classificar corretamente determinados padrões (perfis dos clientes) e que não foram apresentados a RNA durante a fase de treinamento. Percebe-se que o melhor resultado de classificação foi (99,63%) e se deu quando se utilizou a base D5 para testes e as demais (D1, D2, D3, D4) para treino. Sendo que o classificador utilizado (RNA direta com três camadas) obteve um desempenho médio de classificação de 99,31% o que mostra que o mesmo conseguiu classificar satisfatoriamente os padrões, com menos de 1% de erro.



**Figura 5.** Média da taxa de acertos da Rede Neural

Fonte: elaborado pelos autores (2018).

## 5.2 Wisconsin Prognostic Breast Cancer (WPBC)

O segundo “dataset” a ser analisado foi o *Wisconsin Diagnosis Breast Cancer* (WDBC), acessado: <https://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/>.

Os dados foram coletados pelo *Dr. Wolberg*, conforme os casos clínicos tratados foram sendo acumulados cronologicamente. Totalizando 699 observações e 9 atributos e duas



classes. A partir da imagem digitalizada de uma amostra aspirada por meio de uma cânula ou agulha fina da massa de uma mama. Eles descrevem as características dos núcleos celulares presentes na imagem. Cada instância pertence a uma das 2 classes possíveis: benignos ou malignos.

Assim os atributos são descritos na Tabela 1. Embora no total existam 699 medidas individuais (do conjunto de todos os parâmetros), 16 instâncias possuem registros incompletos, sendo que esses foram eliminados, ficando assim o banco de dados com 683 registros. Sendo utilizada a técnica *k-fold cross validation* (k=5) para divisão randômica dos dados. O banco de dados sobre câncer de mama foi utilizado devido à dificuldade inerente da diagnose de câncer desse tipo de doença. Para melhor compreensão do trabalho desenvolvido é importante descrever os dados que compõem o *dataset Diagnosis Breast Cancer*. A Figura 6 apresenta os mesmos e seus significados.

1-Espessura dos grupos (Clump Thickness)	Células benignas tendem a ser agrupadas em monocamadas, enquanto que as células cancerosas são muitas vezes agrupadas em multicamadas.
2-Uniformidade de tamanho e 3-forma da célula	As células cancerosas tendem a variar em tamanho e forma. Devido a isso, esses parâmetros são úteis para determinar se as células são cancerosas ou não.
4-Adesão Marginal (Marginal Adhesion)	As células normais tendem a ficar juntas. As células cancerosas tendem a perder essa capacidade. Então perda de adesão é um sinal de malignidade.
5-Tamanho único das células epiteliais (Single Epithelial Cell Size)	Está relacionada com a uniformidade mencionado acima. As células epiteliais que estão significativamente aumentadas pode ser uma célula maligna.
6-Núcleos nus (Bare Nuclei)	Este é um termo usado para núcleos que não são rodeados pelo citoplasma (o resto da célula). São tipicamente vistos em tumores benignos.
7-Suavidade da cromatina (Bland Chromatin)	Descreve uma textura uniforme do núcleo visto em células benignas. Em células cancerosas, a cromatina tende a ser mais grosseira.
8-Nucléolo normal (Normal Nucleoli)	Nucléolos são pequenas estruturas existentes no núcleo. Em células normais, o nucléolo é geralmente muito pequeno, quando visível. Em células cancerosas os nucléolos se tornam mais proeminentes.
9-Mitose	Patologistas podem determinar o grau de um tumor contando o número de mitoses.

**Figura 6.** Descrição dos Atributos para a base de dados sobre Câncer de mama

Fonte: elaborado pelos autores (2018)

### 5.2.1 Treinamento e Testes das Redes Neurais

O conjunto original do *dataset* Wisconsin Diagnosis Breast Cancer foi dividido em cinco conjunto de dados (k = 5), ou seja, cada partição de dados, denominados (D1, D2, D3, D4, D5) ficou com duzentos registros. A partir disso, utilizou-se uma partição de dados fixo para testes

alternado com as demais para treino da rede neural. Como pode-se observar na Tabela 2, os primeiros 138 registros foram chamados de D1 (dados destacados em amarelo), esse conjunto foi mantido fixo para testar as RNAs e os demais subconjuntos divididos em D2, D3, D4 e D5 foram usados para treinar as redes neurais, ou seja, na primeira iteração, foram utilizadas as seguintes partições de dados para {treino, teste} = {{D2, D1}, {D3, D1}, {D4, D1}, {D5, D1}}. Sendo esse processo repetido, para as demais partições de dados. A primeira coluna da Tabela 2, ilustra as partições fixas usadas nos testes das redes neurais e a segunda coluna, as partições utilizadas para os treinos.

Em cada iteração do método de validação cruzada, ilustrada em cada linha da Tabela 2, utilizaram-se para treino de cada rede neural os parâmetros que estão destacados na coluna 5 (número de neurônios na camada intermediária), na coluna 6 (taxa de aprendizado), na coluna 7 (número de épocas de treinamento da rede neural). As colunas 3,4,6,8 e 9 apresentam-se os resultados dos testes das redes neurais. Na coluna 3 apresentam-se número de instâncias classificadas incorretamente, na coluna 4 a quantidade de instâncias classificadas corretamente, na coluna 8 apresentam-se as taxas de acertos, na coluna 9 a média total de acertos.

Os testes foram feitos estimando parâmetros de forma aleatoriamente, a medida que a rede respondia a cada alteração dos parâmetros avaliavam-se os resultados até se obter a melhor combinação de parâmetros para os quais a rede neural apresentava a maior taxa de acertos, sendo esses registrados na Tabela 2. Por exemplo, variou-se, por exemplo, a taxa de aprendizado e os demais parâmetros foram mantidos constantes, o que se observou foi que o treinamento da rede convergia mais rápido, quando se estipulou o valor de 0.9, para a taxa de aprendizado. Esse mesmo procedimento foi utilizado para ajustar o número de neurônios na camada intermediária, variou-se esse parâmetro, mantendo-se fixos os demais, na Tabela 2 ilustram-se os melhores resultados obtidos. Observou-se que se aumentado a quantidade de neurônios, a partir de 9, não necessariamente produzia uma melhor acurácia na classificação por parte da rede neural, pois alguns resultados apresentaram acurácias equivalentes.

Na Figura 7, apresentam-se uma comparação entre os valores obtidos e desejados, para uma única etapa de treino-teste de uma rede neural. No eixo Y, a resposta varia entre 2 e 4, no eixo X em azul são destacados os valores reais do modelo do *dataset Wisconsin Diagnosis Breast Cancer*, e em amarelo as respostas da rede após o treino. Para o treinamento da rede neural, foram utilizados 138 registros do dataset, sendo que a rede conseguiu prever corretamente 138 com a taxa de acerto de até 100 %.

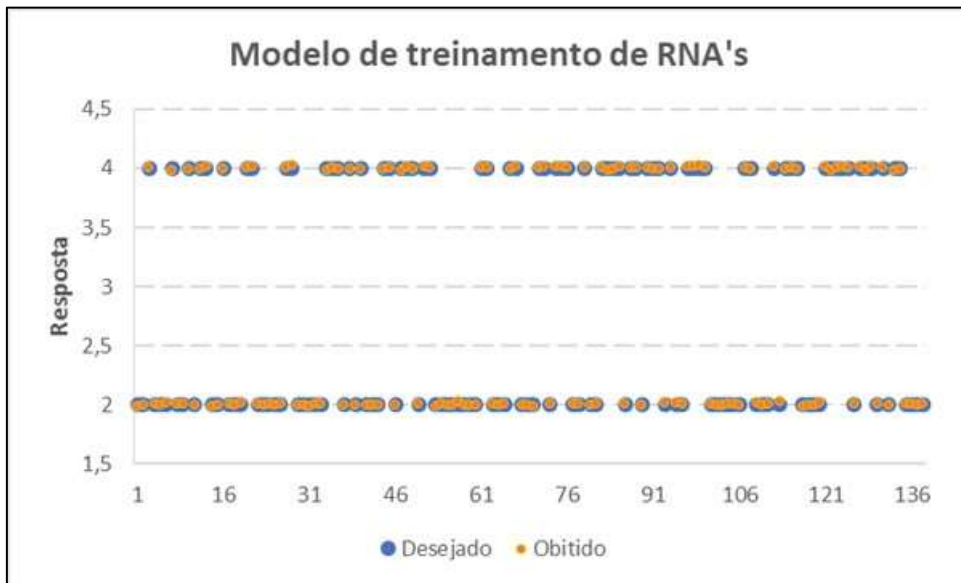
**Tabela 2**

 Tabela de testes para a base de dados sobre *Wisconsin Diagnosis Breast Cancer*

TESTE	TREINO	ERROS	ACERTOS	Neuronios Camada Intermediária	Taxa de Aprendizado	Épocas	% ACERTOS	Média
D1	D2	18	120	7	0,9	50000	86,95	89,85
	D3	0	138	7	0,9	50000	100	
	D4	0	138	7	0,9	50000	100	
	D5	38	100	7	0,9	50000	72,46	
D2	D1	0	138	8	0,9	50000	100	99,81
	D3	0	138	8	0,9	50000	100	
	D4	0	138	8	0,9	50000	100	
	D5	1	137	8	0,9	50000	99,27	
D3	D1	0	138	10	0,9	50000	100	96,92
	D2	0	138	10	0,9	50000	100	
	D4	0	138	10	0,9	50000	100	
	D5	17	121	10	0,9	50000	87,68	
D4	D1	0	138	11	0,9	50000	100	100
	D2	0	138	11	0,9	50000	100	
	D3	0	138	11	0,9	50000	100	
	D5	0	138	11	0,9	50000	100	
D5	D1	0	138	12	0,9	50000	100	100
	D2	0	138	12	0,9	50000	100	
	D3	0	138	12	0,9	50000	100	
	D4	0	138	12	0,9	50000	100	

Fonte: elaborado pelos autores (2018).

Para a solução do problema para o diagnóstico do câncer de mama os melhores resultados foram obtidos com os seguintes parâmetros: (número de neurônios na camada de entrada, número de neurônio na camada intermediária, números de neurônios na camada de saída, taxa de aprendizado, número de épocas de treinamento, erro médio quadrático) = (10,11, 1, 0,9, 50000, 0.0001). Após o treinamento de cada rede neural, a mesma foi utilizada para testar a generalização com dados que não foram utilizados no treinamento, a Figura 7, ilustra uma das generalizações, para uma taxa de acerto de 99,27%. O estudo de generalização é de suma importância, pois indica se a rede é capaz de prever um padrão corretamente, sem que o mesmo tenha sido utilizado na etapa de treinamento da RNA.



**Figura 7.** Treinamento da Rede Neural

Fonte: elaborado pelos autores (2018).

Em relação aos resultados obtidos, utilizando-se um conjunto de dados reais de aplicações de crédito, confirmam-se eficiência do modelo e da metodologia utilizada para a solução do problema. O objetivo foi comparar o desempenho da rede neural utilizando um apanhado de dados classificadas pelos especialistas com o mesmo volume de dados classificadas pelo *K-fold* e avaliar se a rede neural poderia ser utilizada neste contexto.

A Figura 8 também resume os resultados das 20 simulações realizadas para o dataset “Wisconsin Diagnosis Breast Cancer” o gráfico de acertos, mostra a percentagem de acertos obtidos na fase de validação das redes, ou seja, após o treinamento das RNAs convergiram para um erro quadrático de 0,0001. Testou-se a capacidade das RNAs em classificar corretamente determinados padrões (diagnósticos de pacientes) e que não foram apresentados a RNA durante a fase de treinamento. Percebe-se que o melhor resultado de classificação foi (100%) e se deu quando se utilizou a base D5 para testes e as demais (D1, D2, D3, D4) e D4 para testes e (D1, D2, D3, D5) para treino. Sendo que o classificador utilizado (RNA direta com três camadas) obteve um desempenho médio de classificação de 97,31% o que mostra que o mesmo conseguiu classificar satisfatoriamente os padrões, com menos de 3% de erro.

### 5.3 Sugestões para trabalhos futuros

A partir dos achados da presente pesquisa, algumas indicações de futuros estudos que podem ser realizados. Primeiramente, a utilização de outros tipos de classificadores tais como:

Árvore de decisão e Redes Bayesianas. O primeiro, é um modelo estatístico, que é permite extrair regras a partir de árvores geradas após o processo de classificação, o que facilitará a interpretação dos resultados. O segundo modelo é probabilístico é possibilita analisar situações onde não se conhece todo o escopo do problema, visto em algumas situações alguns fatores podem condicionar a falta de informação em uma base de conhecimento: a ignorância teórica ou a impossibilidade.

Em relação a aplicações corporativas sugere-se a utilização de Redes Neurais Artificiais para o desenvolvimento de sistemas de classificação de spams em e-mails, que é um dos principais problemas na utilização do correio eletrônico atualmente. Além disso, uma aplicação muito atual é a detecção de intrusão em redes de computadores, sendo, portanto, uma área aberta para pesquisas em inteligência artificial.

## 6 Conclusões

Ao utilizar a rede neural com aprendizado “Backpropagation” para resolver problemas reais, como o presente trabalho expõem, apresentou-se uma ferramenta bastante poderosa no reconhecimento de padrões e predições de comportamento. Para testar a eficácia da rede em reconhecer padrões a RNA foi testada em duas situações diferentes. A primeira tratava-se em auxiliar um banco alemão em liberar ou não empréstimos, a partir de informações a priori dos clientes da instituição. A segunda situação foi diagnosticar pacientes com câncer de mama, a partir de informações retiradas das imagens de células das pacientes.

Não podemos deixar de frisar que o grau de informações a priori é decisivo no treinamento da rede e consequentemente o grau de aceitabilidade. Nos casos estudados que se tratavam de informações de credito e pessoais de 1000 clientes de um banco alemão, e de diagnosticar 683 pacientes com suspeita de câncer mamário. Para ambas as situações capacidade de aprendizado e aceitabilidade da rede foi em média de 99%, para fins reais, o seu desempenho da rede seria muito útil podendo prever o comportamento de futuros clientes na hora da adesão de credito bancário e diagnósticos médicos, podendo assim minimizar desta forma as perdas bancarias e auxiliando no reforço gerencial na tomada de concessão de créditos. E auxiliar médicos nós diagnósticos médicos de câncer mamário, desta forma apresentando-se uma ferreamente muito útil em diferentes áreas, tanto na financeira como em situações medicas.

## Referências

Ajay Kaushik, Shubham Gupta, Manan Bhatia (2018). A Movie Recommendation

- System using Neural Networks. *International Journal of Advance Research, Ideas and Innovations in Technology*, 4(2)
- Arora, N., Martins, D., Ruggerio, D., Tousimis, E., Swistel, A., & Osborne, M. P. (2008). Effectiveness of a noninvasive digital infrared thermal imaging system in the detection of breast cancer. *The American Journal of Surgery*, 196, 523-526.
- Bonini, J. A. (2016). Aplicação de algoritmos de árvore de decisão sobre uma base de dados de câncer de mama. *Communications and Innovations Gazzete*. 1(1), 57-67.
- Chatziagorakis, P., Georgoulas, N., Papadopoulou, S., Elmasides, C., Giaouris, D., Seferlis, P., Sirakoulis, Ch. G., Karafyllidis, I., Papadopoulos, A. I., Stergiopoulos, F., Ziogou, C., Ipsakis, D., Voutetakis, S., & Andreadis, I. (2014). Application of neural networks solar radiation prediction for hybrid renewable energy systems. *Engineering Applications of Neural Networks*, 459, 133-144.
- Fabro, J. A. (2001, agosto). *Redes neurais artificiais*. [Curso de Especialização em Inteligência Computacional]. Recuperado em 01 junho, 2018, de <http://www.dainf.ct.utfpr.edu.br/~fabro/IF67D/redesneuraisartificiais.pdf>
- Gambogi, J. A. (2014). *Aplicação de redes neurais na tomada de decisão no mercado de ações*. Dissertação de mestrado, Escola Politécnica, Universidade de São Paulo, São Paulo, SP, Brasil.
- Geisser, S. (1975). The predictive sample reuse method with applications. *J. Am. Stat. Assoc.*, 70(350), 320-328.
- Haykin, S. (1994). *Neural networks: a comprehensive foundation*. New York, NY: Macmillan College Publishing.
- Haykin, S. (2001). *Redes neurais princípios e prática*. Porto Alegre: Bookman.
- Holsbach, N., Fogliatto, F. S., & Anzanello, M. J. (2014). Método de mineração de dados para identificação de câncer de mama baseado na seleção de variáveis [versão eletrônica], *Ciênc. Saúde Coletiva*, 19(4), 1295-1304.
- Karishma Dasgaonkar, Swati Chopade (2018). Analysis of multi-layered perceptron, radial basis function and convolutional neural networks in recognizing handwritten digits. *International Journal of Advance Research, Ideas and Innovations in Technology*, 4(3).
- Kovács, Z. L. (2006). *Redes neurais artificiais: fundamentos e aplicações*. (4a ed.). São Paulo: Livraria da Física.
- Nitin Sharma, Pranshu Agarwal, Upasana Pandey (2018). Offline handwriting recognition using neural networks. *International Journal of Advance Research, Ideas and Innovations in Technology*, 4(2)
- Nunes, W. V. (2012). Introdução às redes neurais artificiais. In V. Esperidião-Antônio (Ed.). *Neurociências: diálogos e interseções*. (pp. 1-768). Rio de Janeiro: Rubio.

- Rede Neural Artificial, (n.d.). In *Wikipedia enciclopedia livre*. Recuperado em 01 junho, 2018, de [https://pt.wikipedia.org/wiki/Rede\\_neural\\_artificial](https://pt.wikipedia.org/wiki/Rede_neural_artificial)
- Sak, H., Senior, A., & Beaufays, F. (2014). *Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition*. Retrieved June 01, 2018, from [arxiv.org/abs/1402.1128](http://arxiv.org/abs/1402.1128)
- Sánchez, D., & Melin, P. (2014). Optimization of modular granular neural networks using hierarchical genetic algorithms for human recognition using the ear biometric measure. *Engineering Applications of Artificial Intelligence*, 27, 41-56.
- Sánchez, D., Melin, P., & Castillo, O. (2015). Optimization of modular granular neural networks using a hierarchical genetic algorithm based on the database complexity applied to human recognition. *Information Sciences*, 309, 73-101.
- Skrizhevsky, A., Sutskever, I., & Hinton, G. (2012). ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.*, 25, 1097-1105.
- Steiner, M. T. A., Soma, N. Y., Shimizu, T., Nievola, J. C., Lopes, F. M., & Smiderle, A. (2004, novembro). *Redes neurais e arvores de decisão na análise do crédito bancário*. Trabalho apresentado em Simpósio Brasileiro de Pesquisa Operacional, São João del-Rei, MG, Brasil, 36. Recuperado em 01 junho, 2018, de <http://www.din.uem.br/sbpo/sbpo2004/pdf/arq0035.pdf>
- Sutskever, I., Vinyals, O., & Le, Q. V. (2014). *Sequence to sequence learning with neural networks*. Retrieved June 01, 2018, from <http://arxiv.org/abs/1409.3215>
- Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., & Fergus, R. (2014). *Intriguing properties of neural networks*. Retrieved June 01, 2018, from [arxiv.org/abs/1312.6199](http://arxiv.org/abs/1312.6199)
- Vikas Poonia, Dr. H. L. Tiwari, Dr. Satanand Mishra (2018). Hydrological Analysis by Artificial Neural Network: A Review. *International Journal of Advance Research, Ideas and Innovations in Technology*, 4(3) [www.IJARIIT.com](http://www.IJARIIT.com).
- Wang, Y., & Kosinski, M. (2017, October 16). Deep neural networks are more accurate than humans at detecting sexual orientation from facial