

Uso Da Rede Neural Artificial ARTMAP-Fuzzy Para Identificar Casos De Câncer De Mama

Luana da Silva Pinto¹, Carlos Roberto dos Santos Júnior¹, Edgar Noda¹

¹Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP) - Câmpus Hortolândia. CEP: 13183-250 – Hortolândia – SP – Brasil

luana.silvapinto@hotmail.com, carlos.rsantos@ifsp.edu.br,

edgar.ifsp@gmail.com

Abstract. *One of the cancer types that has the highest incidence and high mortality rate is breast cancer, which needs to be diagnosed and treated as soon as possible to ensure the most effective treatment outcome. To assist in the analysis of a malignant or benign nodule, an ARTMAP Fuzzy artificial neural network has been developed to verify a database of breast nodule characteristics of 569 women who underwent the examination and identify who have the disease. resulting in 91.075% Accuracy, 94.46% Recall, 86.188% Specificity and 91.161% Precision.*

Resumo. *Um dos tipos de câncer que tem maior incidência e grande taxa de mortalidade é o câncer de mama, que precisa ser diagnosticado e tratado o mais rápido possível, para garantir uma melhor eficiência no resultado do tratamento. Para ajudar na análise de um nódulo maligno ou benigno, foi desenvolvida uma rede neural artificial do tipo ARTMAP Fuzzy que verifica uma base de dados com características dos nódulos das mamas de 569 mulheres que fizeram o exame e identifica quais têm a doença, resultando em 91,075% de Acurácia, 94,46% de Sensibilidade, 86,188% de Especificidade e 91,161% de Precisão.*

1. Introdução

O câncer de mama está entre os 3 tipos de câncer que mais ocorrem no mundo e é o que mais afeta as mulheres em 154 países entre 185 analisados, segundo [dos Passos Ramos 2021], sendo também a maior causa de morte por câncer nas brasileiras, de acordo com [de Câncer José Alencar Gomes da Silva (INCA) 2020]. Em 2020 foram registradas mais de 684 mil mortes por essa doença, que equivale a 24,5% dos novos casos de câncer. Em mulheres com menos de 40 anos é mais difícil ocorrer a doença, mas entre os 40 e 60 anos as incidências aumentam e, em mulheres com mais de 60 anos, a probabilidade se torna 10 vezes maior.

Uma forma de diminuir a taxa de mortalidade em casos de câncer de mama está no diagnóstico precoce da doença, como é citado em [Rasool 2022], que pode ser descoberta através de análises de mamografia ou ultrassonografia. Quanto mais cedo a constatação, maiores as chances de cura, de acordo com o avanço do estágio da doença. Portanto, é de grande importância que as mulheres façam regularmente os exames que identificam a presença ou não do câncer na paciente, para poderem se dedicar a ter uma vida mais saudável, e assim reduzir os riscos da doença.

De acordo com [Rasool 2022] , um dos primeiros passos para a detecção de possíveis nódulos é fazer um exame de imagem, podendo ser uma mamografia ou ultrassonografia. O resultado destes exames podem ser classificados de acordo com um método padrão denominado BIRADS, sigla para *Breast Imaging Reporting and Data System*, traduzido para Sistema de Relatório de Dados de Imagens da Mama, que estima a chance de uma lesão visualizada em exames de mama ser um câncer de mama. [Rasool 2022] explica que existem 5 níveis de BIRADS, sendo que o primeiro, BIRADS 1, diz que o resultado do exame é normal ou negativo, ou seja, a paciente não possui a doença; o BIRADS 2 indica que foram achadas lesões no exame, porém são benignas; o BIRADS 3 é, também, a indicação de achados benignos, mas este com uma possível chance de ser maligno; o BIRADS 4 já diz que foi encontrada uma lesão e que esta é suspeita de ser maligna; por fim, o BIRADS 5, que informa uma alta suspeita de câncer de mama. Quando a paciente apresenta um exame classificado como BIRADS 4 ou 5 é recomendado fazer uma biópsia para melhor classificação do nódulo, indicando se ele é maligno ou benigno. Nesta biópsia são analisadas as características dos nódulos, que têm diferenças fisiológicas quando malignos ou benignos.

Este trabalho tem como objetivo desenvolver uma rede neural artificial do tipo ARTMAP *Fuzzy* supervisionada, explicada em [dos Santos Junior], para auxiliar na análise das biópsias de maneira mais eficaz para diminuir o tempo que levaria para os nódulos serem identificados como malignos ou benignos, que atualmente é de uma a duas semanas para análise das características, podendo levar apenas alguns segundos ao ser inserido na RNA (Rede Neural Artificial) e, conseqüentemente, colaborar com as constatações precoces da doença. Além dos motivos de ter sido tratado em sala de aula e por não ter encontrado outros trabalhos que utilizam este tipo de rede neural artificial para detecção de câncer de mama, este tipo de análise foi escolhido pensando na compatibilidade dos tipos de dados que a RNA analisa e os tipos de dados obtidos nas biópsias, pois a ARTMAP *Fuzzy* funciona com base em dividir dados com características semelhantes em grupos e, ao classificar um dado, analisa a qual grupo ele pertence e determina sua classificação. Como o nódulo do câncer de mama é classificado a partir de suas características, a ARTMAP *Fuzzy* pode ser capaz de manipular os dados, diferenciar os malignos dos benignos a partir de suas características e classificar um nódulo com base na similaridade entre os analisados.

2. Trabalhos correlatos

A fim de comparar como os diferentes tipos de redes neurais artificiais se comportam diante da análise de possíveis casos de câncer de mama, e para obter dados de acurácia das mesmas para concluir se a ARTMAP *Fuzzy* é eficaz no assunto, foram escolhidos 3 trabalhos correlatos, onde cada um utiliza um tipo de RNA diferente, mas todos com o mesmo intuito de analisar se um nódulo é maligno ou benigno. Para diminuir a diferença dos resultados foram escolhidos trabalhos correlatos que utilizam a mesma base de dados deste trabalho.

[Seema Singh 2012] fez o uso de uma rede neural da classe ARNN (*Adaptive Resonance Neural Networks*) chamada ART2 para detectar quais nódulos são benignos e quais são malignos. Esta RNA possui uma arquitetura formada por 29 nós de entrada e 2 de saída, com unidades de processamento de entrada altamente complexas, incluindo normalização e supressão de ruído, juntamente com a comparação dos pesos necessários

para um mecanismo de *reset*. Os resultados obtidos com o uso desta RNA foram avaliados com os mesmos métodos deste trabalho, Acurácia, Sensibilidade e Precisão, chegando nos valores de 75% de Sensibilidade, 82,64% de Acurácia e 79% de Precisão.

O artigo [dos Santos] propõe o uso de Algoritmos Genéticos (AG) para identificar os casos de câncer de mama. Este tipo de rede neural artificial é baseado na teoria da evolução das espécies e busca os melhores resultados baseando-se no potencial de cada indivíduo da população. O AG é desenvolvido seguindo algumas funções, sendo elas a geração e configuração dos indivíduos, criação e análise dos RNAs, ordenação dos indivíduos que foram considerados fracos, criação de indivíduos evoluídos através de um cruzamento e uma mutação na geração que está sendo analisada. A cada surgimento de uma nova geração foi verificado que o sistema obtinha uma melhoria considerável nos resultados, até chegar na geração 27, onde atingiu o melhor resultado possível. Concluiu-se que o uso do AG na classificação de nódulos malignos ou benignos foi eficaz, pois conseguiu identificar corretamente 100% das amostras, mas ainda pode ser melhorado através de ajustes no algoritmo.

A RNA utilizada em [Abbass] é chamada de MPANN (*Memetic Pareto Artificial Neural Network*), traduzida como Rede Neural Artificial Memética de Pareto, com uma abordagem evolutiva baseada no diagrama de Pareto. As funções utilizadas para esta técnica começam com a criação de uma população inicial de soluções potenciais aleatórias a partir de uma distribuição Gaussiana, seguida por uma repetição de funções que criam indivíduos novos a partir de 3 indivíduos pais até alcançar um limite máximo da população seguindo algumas regras. Esta técnica conseguiu alcançar uma acurácia de 98,1% sobre as definições de nódulos malignos e benignos.

3. Fundamentação teórica

Uma rede neural artificial (RNA), segundo [Education 2020] é uma técnica computacional criada para funcionar similarmente a uma rede neural de um organismo inteligente, capaz de aprender conforme sua experiência, com o objetivo de permitir que as máquinas consigam realizar tarefas cognitivas similares às do cérebro humano, enviando sinais de um neurônio para outro. Elas são capazes de realizar reconhecimento de padrões e aprendizado de máquina. Atualmente existem diversos tipos de redes neurais artificiais, cada uma com seus diferentes processos e funções, entre elas a ARTMAP *Fuzzy*, utilizada neste artigo para classificar os nódulos da mama.

O modelo ARTMAP *Fuzzy* é uma RNA baseada na Teoria da Ressonância Adaptativa (ART, do inglês *Adaptive Resonance Theory*), desenvolvida por Stephen Grossberg e Gail Carpenter, que, segundo [dos Santos Junior] descreve vários modelos de redes neurais artificiais com métodos de aprendizado de máquina supervisionados e não supervisionados, onde o supervisionado é quando a RNA recebe prontamente a rotulação de seus conjuntos de dados criados, e o não supervisionado é o aprendizado próprio da RNA quanto aos rótulos de seus conjuntos de dados. A ARTMAP *Fuzzy* se baseia no treinamento supervisionado e utiliza, em sua arquitetura, cálculos baseados na lógica *Fuzzy*, que pode ser entendida como a aplicação de valores variáveis que se enquadram dentro de dois valores fixos, em situações que não podem ser mensuradas somente com dois valores extremos, mas sim entre uma faixa de valores entre estes dois, como por exemplo o clima, que pode ser não somente quente ou frio, mas também ameno, fresco, morno e etc.

A ARTMAP *Fuzzy* tem sua arquitetura formada por 3 módulos, denominados ARTa, ARTb e Inter-ART, conforme Figura 1. Os módulos ARTa e ARTb foram criados de acordo com a arquitetura de outra classe da Teoria da Ressonância Adaptativa, chamada ART *Fuzzy*. O módulo *Inter-ART* armazena a classificação dos resultados obtidos na ARTa de acordo com os dados finais da ARTb através do mecanismo *Match Tracking*.

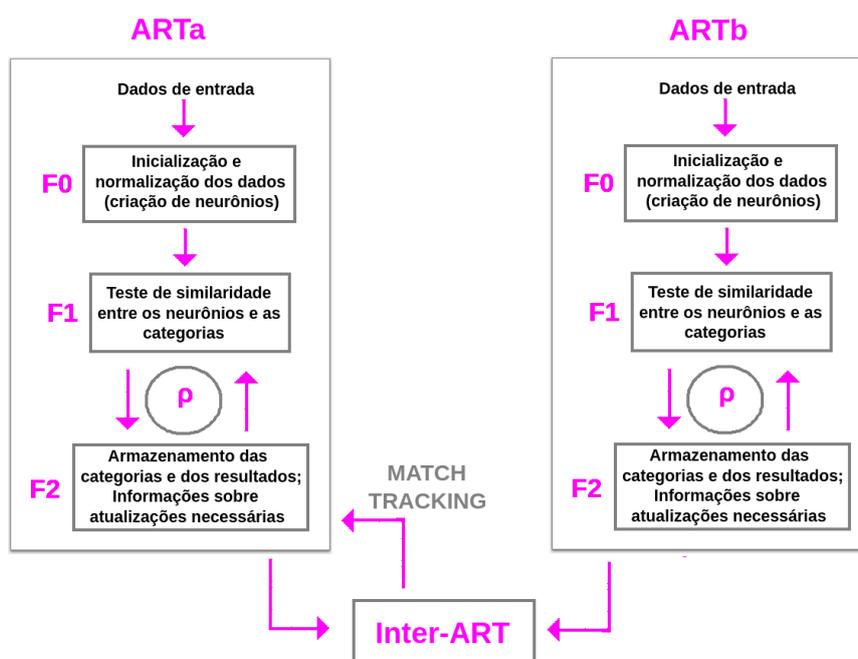


Figura 1. Arquitetura da ARTMAP Fuzzy
Fonte: Próprio autor

Os módulos ARTa e ARTb têm uma arquitetura formada por três camadas, F0, F1 e F2, que são compostas por funções que fazem uma comparação entre conjuntos de dados informados, que podem ser classificados em tipos diferentes, dividindo-os em grupos de características semelhantes, o que permite identificar a qual grupo um conjunto de características pertence, consequentemente identificando seu tipo. As funções das camadas fazem o tratamento e análise dos dados, interagindo entre si, como mostrado na Figura 1.

A camada F0 é a camada de entrada, onde os dados são pré-processados e se tornam os conjuntos de dados que serão utilizados em todas as camadas, chamados de neurônios. Para isso é necessário que os dados tenham valores entre 0 e 1, seguindo a Lógica *Fuzzy*. Após essa normalização, os neurônios seguirão o fluxo para as próximas funções. A camada F2 é responsável por armazenar categorias, que são formadas por neurônios semelhantes, conjuntos de dados onde pode ser observado um mesmo padrão. Inicialmente é criada uma categoria com todos os valores do neurônio sendo 1, para poder iniciar o processo de comparação, pois até o momento não foram inseridos neurônios. Os neurônio que saem da camada F0 vão para a camada F1, que irá comparar cada um com as categorias da camada F2, seguindo um critério de similaridade denominado RHO (17º letra do alfabeto grego), informado pelo usuário, que determina a porcentagem de semelhança que o neurônio deve ter com os neurônios de uma categoria para poder fazer

parte da mesma. Caso nenhuma categoria seja aceita pelo critério, é criada uma nova, surgindo um novo padrão de comparação, com as características do neurônio em questão. O usuário informa uma porcentagem dos dados que será utilizada para treinamento e a outra que será utilizada pra treino, então este fluxo entre as camadas ocorrem até que todos os neurônios de treinamento sejam classificados. O módulo ARTa, da ARTMAP *Fuzzy*, executa estas funções utilizando como dados as características do cenário em questão, que no caso são os valores das 30 características de cada nódulo da base de dados, enquanto que o módulo ARTb utiliza a classificação do conjunto de características dentro destas funções, que para o contexto deste trabalho são as 2 possíveis definições para cada nódulo, maligno ou benigno, como mostrado em [Mangasarian].

O *Match Tracking* define, para cada conjunto padrão de características criados na ARTa, a qual classificação da ARTb ele pertence, ou seja, os conjuntos de dados definidos pela ARTa, de acordo com as características dos nódulos, são alinhados com suas classificações, malignos ou benignos, a fim de ensinar a RNA quais padrões de dados inseridos na ARTa podem ser considerados como benignos e quais podem ser classificados como malignos.

Após todas estas etapas serem concluídas, é realizada a fase de testes, que insere a porcentagem dos dados que foi definida para os mesmos, executando as funções que foram realizadas na fase de treinamento, porém, sem criar novos grupos de padrões de características, apenas define a qual grupo cada neurônio pertence. Ao final, faz o *Match Tracking* com as classificações da ARTb para saber se as atribuições de cada neurônio aos devidos grupos estão corretas.

4. Metodologia e Materiais e Métodos

Para a implementação da ARTMAP *Fuzzy* foi utilizada a linguagem de programação Python pois, de acordo com [Matos 2019] ela possui uma maior quantidade de recursos para uso na Inteligência Artificial, assim como uma vasta variedade de bibliotecas para aprendizado de máquina e análise de dados. As bibliotecas utilizadas para a implementação da rede neural artificial foram a Pandas, para manipulação e análise dos dados, a Matplotlib, para auxílio nas operações matemáticas e gráficos, e a Numpy, para também ajudar com as operações matemáticas, além do suporte com a manipulação dos dados dentro dos *Arrays* multidimensionais, que são estruturas de dados que armazenam elementos onde cada um pode ser identificado através de um índice.

Os 3 Módulos que compõem a ARTMAP *Fuzzy*, explicados na sessão anterior, foram criados no arquivo principal do código, utilizando as bibliotecas para realizar a preparação dos dados para serem manipulados e estruturas de repetição (*while*) e de decisão (*if/else*) para ordenar e realizar as chamadas das funções de cada módulo, conforme sessão anterior. Cada função foi criada em outro arquivo do código, para serem executadas somente quando o arquivo principal solicitar, enviando os parâmetros necessário de cada uma. As funções possuem estruturas de decisão (*if/else*) e de repetição (*for*), também em Python.

Os dados analisados neste artigo foram obtidos através da plataforma Kaggle, que disponibiliza uma grande variedade de conjuntos de dados para uso público em projetos de ciência de dados. A base de dados utilizada foi a WDBC (*Wisconsin Diagnostic Breast Cancer*), que consiste em 10 características dos nódulos das mamas de 569 pacientes

do Dr. William H. Wolberg, que retirou amostras de fluidos das massas mamárias sólidas destas mulheres no Departamento de Cirurgia Geral da Universidade de Wisconsin-Madison, EUA, e utilizou o software Xcyt para analisar suas características citológicas com base no escaneamento digital. Tal software aplica um algoritmo de ajuste de curva para calcular dez características, retornando o valor médio de cada, o pior valor e o valor de erro padrão, gerando 30 valores diferentes para cada amostra. As características analisadas foram o raio (média das distâncias do centro aos pontos no perímetro), textura (desvio padrão dos valores de escala de cinza), perímetro, área, suavidade (variação local nos comprimentos dos raios), compacidade (calculada pelo perímetro ao quadrado dividido pela área menos 1), concavidade (severidade das porções côncavas do contorno), pontos côncavos (número de porções côncavas do contorno), simetria e dimensão fractal (calculada pela aproximação do litoral menos 1). Além disso, foi inserida uma coluna informando a classificação final de cada amostra, se são benignas ou malignas.

5. Resultados

As 30 colunas da base de dados, que também pode ser encontrada em [Mangasarian], referente às características dos nódulos continham valores somente numéricos. Porém, a coluna referente à análise, que dizia se o nódulo era maligno ou benigno, estava populada com "M" e "B", respectivamente. Portanto, foi necessário, inicialmente, transformar estes valores em numéricos, recebendo 0 as tuplas que tinham valor "B" e 1 as tuplas que tinham valor "M".

Em seguida foram iniciadas as camadas da ART *Fuzzy*, tanto para ARTa quanto para ARTb. A primeira foi a camada F0. Utilizando a biblioteca Pandas, foi feita a leitura e inicialização dos dados, utilizando uma função para normalizar todos os dados do vetor, para padronizar seus valores, que tinha desde valores pequenos com casas decimais, até valores maiores com centenas. Então a padronização os alterou para que todos estejam entre 1 e 0, de acordo com a Lógica *Fuzzy*.

Como é necessária a atuação dos resultados da camada F0 e F2 para a camada F1 atuar, mas apenas a camada F0 foi realizada até o momento, os pesos da camada F2 que determinam as categorias receberam valor 1 em todos os índices ao iniciar a camada F1, conforme explicado na seção Metodologia e Materiais e Métodos. Após iniciada, foi realizada a função de ativação, que compara os valores de entrada vindos da camada F0 com as categorias. A função aplicada em seguida é a de treinamento, que atualiza as categorias e as armazena na camada F2.

A fase de treinamento da camada F2 se iniciou com a função de escolha, que determina qual o índice com maior valor. Após isso, a mesma percorre o vetor comparando os valores de entrada com as categorias e, nos casos onde não houve compatibilidade, foi criada uma nova. As camadas continuaram interagindo da mesma maneira até finalizarem a quantidade de neurônios definidos para treinamento. Após concluído o treinamento, deu-se início a fase de testes, onde são aplicadas as funções da camada F1 nos neurônios restantes para analisar, de acordo com os pesos resultantes da camada F2 da ARTa e seus rótulos resultantes da ARTb, quais neurônios possuem anomalia e quais não possuem.

Para aumentar a precisão dos resultados da ARTMAP *Fuzzy*, foi utilizada uma técnica que é comumente usada nas avaliações de desempenho de modelos de aprendizado de máquina, o *Cross-Validation*, [Lakshana 2021]. Dentro desta técnica existe um método

muito utilizado em projetos como este, o *K-fold*, que divide a porcentagem de dados utilizados para teste e treino em K conjuntos, com cada conjunto tendo a porcentagem dos dados dividida por K, e a cada execução um conjunto diferente é utilizado para teste e o restante para treino, conforme demonstrado nas Figuras 3, 4, 5 e 6. Para as execuções desta RNA foram utilizados 4 níveis de *K-fold* (*5-fold*, *4-fold*, *3-fold* e *2-fold*), a fim de maximizar a precisão dos resultados. O *Cross-Validation* foi executado 5 vezes, cada uma com um valor diferente de RHOa, para encontrar o melhor valor para o mesmo, conforme mostrado nas Figuras 3, 4, 5 e 6.

		5-Fold Cross-Validation				
ETAPAS	TESTE	TREINO	TREINO	TREINO	TREINO	
	TREINO	TESTE	TREINO	TREINO	TREINO	
	TREINO	TREINO	TESTE	TREINO	TREINO	
	TREINO	TREINO	TREINO	TESTE	TREINO	
		TREINO	TREINO	TREINO	TESTE	
		20% dados	20% dados	20% dados	20% dados	20% dados

		4-Fold Cross-Validation			
ETAPAS	TESTE	TREINO	TREINO	TREINO	
	TREINO	TESTE	TREINO	TREINO	
	TREINO	TREINO	TESTE	TREINO	
	TREINO	TREINO	TREINO	TESTE	
		25% dados	25% dados	25% dados	25% dados

		3-Fold Cross-Validation		
ETAPAS	TESTE	TREINO	TREINO	
	TREINO	TESTE	TREINO	
	TREINO	TREINO	TESTE	
		33% dados	33% dados	33% dados

		2-Fold Cross-Validation	
ETAPAS	TESTE	TREINO	
	TREINO	TESTE	
		50% dados	50% dados

Figura 2. Método *Cross-Validation*

RHOa: 0.95										
5 Folder	dados para teste		Positivos	Negativos	Falsos Positivos	Falsos Negativos	% acurácia	Sensibilidade	Especificidade	Precisão
	0	114	37	66	9	2	90	94,8718	88	90,3509
	115	229	59	48	6	1	94	98,3333	88,8889	93,8596
	230	324	54	32	7	1	91	98,1818	82,0513	91,4894
	325	419	57	26	11	0	88	100	70,2703	88,2979
	420	569	109	34	5	1	96	99,0909	87,1795	95,9732
MÉDIA		63,2	41,2	7,6	1	91,8	98,0956	83,278	91,9942	
4 Folder	dados para teste		Positivos	Negativos	Falsos Positivos	Falsos Negativos	% acurácia	Sensibilidade	Especificidade	Precisão
	0	141	45	78	15	3	87	93,75	83,871	87,234
	142	283	74	59	5	3	94	96,1039	92,1875	94,3262
	284	425	89	31	20	1	85	98,8889	60,7843	85,1064
	426	569	103	34	5	1	96	99,0385	87,1795	95,8042
	MÉDIA		77,75	50,5	11,25	2	90,5	96,9453	81,0056	90,6177
3 Folder	dados para teste		Positivos	Negativos	Falsos Positivos	Falsos Negativos	% acurácia	Sensibilidade	Especificidade	Precisão
	0	189	76	92	16	5	89	93,8272	85,1852	88,8889
	190	379	115	69	3	2	97	98,2906	95,8333	97,3545
	380	569	139	42	7	1	96	99,2857	85,7143	95,7672
	MÉDIA		110	67,6667	8,66667	2,66667	94	97,1345	88,9109	94,0035
	2 Folder	dados para teste		Positivos	Negativos	Falsos Positivos	Falsos Negativos	% acurácia	Sensibilidade	Especificidade
0		284	139	93	0	52	82	72,7749	100	81,6901
285		569	204	64	13	3	94	98,5507	83,1169	94,3662
MÉDIA TOTAL		105,613	59,4667	8,50417	8,29167	91,075	94,46	86,188	91,161	

Figura 3. Resultados obtidos após *Cross-Validation* (RHOa: 0.95)

RHOa: 0.90										
5 Folder	dados para teste		Positivos	Negativos	Falsos Positivos	Falsos Negativos	% acurácia	Sensibilidade	Especificidade	Precisão
	0	114	34	66	12	2	88	94,4444	84,6154	87,7193
	115	229	58	46	7	3	91	95,082	86,7925	91,2281
	230	324	52	32	9	1	89	98,1132	78,0488	89,3617
	325	419	51	26	17	0	82	100	60,4651	81,9149
	420	569	100	34	14	1	90	99,0099	70,8333	89,9329
MÉDIA		59	40,8	11,8	1,4	88	97,3299	76,151	88,0314	
4 Folder	dados para teste		Positivos	Negativos	Falsos Positivos	Falsos Negativos	% acurácia	Sensibilidade	Especificidade	Precisão
	0	141	40	79	20	2	84	95,2381	79,798	84,3972
	142	283	68	60	11	2	91	97,1429	84,507	90,7801
	284	425	75	31	34	1	75	98,6842	47,6923	75,1773
	426	569	94	34	14	1	90	98,9474	70,8333	89,5105
	MÉDIA		69,25	51	19,75	1,5	85	97,5031	70,7077	84,9663
3 Folder	dados para teste		Positivos	Negativos	Falsos Positivos	Falsos Negativos	% acurácia	Sensibilidade	Especificidade	Precisão
	0	189	68	88	24	9	83	88,3117	78,5714	82,5397
	190	379	106	70	12	1	93	99,0654	85,3659	93,1217
	380	569	128	41	18	2	89	98,4615	69,4915	89,418
	MÉDIA		100,667	66,3333	18	4	88,3333	95,2795	77,8096	88,3598
	2 Folder	dados para teste		Positivos	Negativos	Falsos Positivos	Falsos Negativos	% acurácia	Sensibilidade	Especificidade
0		284	137	122	2	23	91	85,625	98,3871	91,1972
285		569	184	65	33	2	88	98,9247	66,3265	87,6761
MÉDIA TOTAL		97,3542	62,9083	16,7625	4,85	87,7083	95,5969	76,7563	87,6985	

Figura 4. Resultados obtidos após *Cross-Validation* (RHOa: 0.90)

RHOa: 0.85										
5 Folder	dados para teste	Positivos	Negativos	Falsos Positivos	Falsos Negativos	% acurácia	Sensibilidade	Especificidade	Precisão	
	0	114	26	65	20	3	80	89,6552	76,4706	79,8246
	115	229	50	47	15	2	85	96,1538	75,8065	85,0877
	230	324	49	30	12	3	84	94,2308	71,4286	84,0426
	325	419	47	26	21	0	78	100	55,3191	77,6596
	420	569	90	34	24	1	83	98,9011	58,6207	83,2215
MEDIA		52,4	40,4	18,4	1,8	82	95,7862	67,5291	81,9672	
4 Folder	dados para teste	Positivos	Negativos	Falsos Positivos	Falsos Negativos	% acurácia	Sensibilidade	Especificidade	Precisão	
	0	141	34	73	26	8	76	80,9524	73,7374	75,8865
	142	283	73	57	6	5	92	93,5897	90,4762	92,1986
	284	425	84	32	25	0	78	100	56,1404	82,2695
	426	569	84	34	24	1	83	98,8235	58,6207	82,5175
	MEDIA		68,75	49	20,25	3,5	82,25	93,3414	69,7437	83,218
3 Folder	dados para teste	Positivos	Negativos	Falsos Positivos	Falsos Negativos	% acurácia	Sensibilidade	Especificidade	Precisão	
	0	189	83	81	9	16	87	83,8384	90	86,7725
	190	379	95	68	23	2	86	97,9381	74,7253	86,7021
	380	569	120	41	26	2	85	98,3607	61,194	85,1852
	MEDIA		99,3333	63,3333	19,3333	6,66667	86	93,3791	75,3064	86,2199
	2 Folder	dados para teste	Positivos	Negativos	Falsos Positivos	Falsos Negativos	% acurácia	Sensibilidade	Especificidade	Precisão
0		284	129	121	10	24	88	84,3137	92,3664	88,0282
285		569	203	63	14	4	94	98,0676	81,8182	93,662
MEDIA TOTAL		96,6208	61,1833	17,4958	6,49167	85,3125	93,4248	74,9179	85,5626	

Figura 5. Resultados obtidos após *Cross-Validation* (RHOa: 0.85)

RHOa: 0.80										
5 Folder	dados para teste	Positivos	Negativos	Falsos Positivos	Falsos Negativos	% acurácia	Sensibilidade	Especificidade	Precisão	
	0	114	30	56	16	12	75	71,4286	77,7778	75,4386
	115	229	40	45	25	4	75	90,9091	64,2857	74,5614
	230	324	48	31	13	2	84	96	70,4545	84,0426
	325	419	49	25	19	1	79	98	56,8182	78,7234
	420	569	100	31	14	4	88	96,1538	68,8889	87,9195
MEDIA		53,4	37,6	17,4	4,6	80,2	90,4983	67,645	80,1371	
4 Folder	dados para teste	Positivos	Negativos	Falsos Positivos	Falsos Negativos	% acurácia	Sensibilidade	Especificidade	Precisão	
	0	141	36	79	24	2	82	94,7368	76,699	81,5603
	142	283	58	59	21	3	83	95,082	73,75	82,9787
	284	425	83	30	26	2	75	97,6471	53,5714	80,1418
	426	569	95	31	13	4	88	95,9596	70,4545	88,1119
	MEDIA		68	49,75	21	2,75	82	95,8564	68,6188	83,1982
3 Folder	dados para teste	Positivos	Negativos	Falsos Positivos	Falsos Negativos	% acurácia	Sensibilidade	Especificidade	Precisão	
	0	189	67	85	25	12	80	84,8101	77,2727	80,4233
	190	379	105	68	13	3	92	97,2222	83,9506	91,5344
	380	569	120	40	26	3	85	97,561	60,6061	84,6561
	MEDIA		97,3333	64,3333	21,3333	6	85,6667	93,1978	73,9431	85,5379
	2 Folder	dados para teste	Positivos	Negativos	Falsos Positivos	Falsos Negativos	% acurácia	Sensibilidade	Especificidade	Precisão
0		284	103	118	36	27	78	79,2308	76,6234	77,8169
285		569	170	63	47	7	82	96,0452	57,2727	81,1847
MEDIA TOTAL		88,8083	60,5458	25,3083	7,5875	81,9667	91,7976	69,2887	82,0935	

Figura 6. Resultados obtidos após *Cross-Validation* (RHOa: 0.80)

Juntamente com a verificação dos melhores valores para RHOa e porcentagem

de treino para encontrar a melhor taxa de acurácia, dentro do *cross-validation* foi feita uma análise de falsos positivos e falsos negativos para cada execução da RNA, que mostra onde a RNA errou na classificação dos neurônios. Esta classificação só pode ter 2 resultados, o maligno (positivo) ou o benigno (negativo), então a RNA classifica cada um dos neurônios de teste como sendo um destes 2, podendo acertar ou errar a classificação, portanto, são 4 os resultados que podem ser obtidos: os positivos e os negativos verdadeiros, ou seja, os neurônios que foram classificados como malignos ou benignos e realmente eram malignos ou benignos, e os falsos positivos ou falsos negativos, onde a RNA diz que o neurônio é maligno, mas na verdade ele é benigno, ou diz que é benigno, mas na verdade é maligno. Trabalhando com os resultados de falsos/verdadeiros positivos/negativos foram calculadas a Sensibilidade (mais conhecida como *Recall*), a Especificidade e a Precisão da rede, onde a Sensibilidade mostra a porcentagem de acertos positivos (quantidade de resultados positivos dividida pela soma dos resultados positivos e falsos negativos), a Especificidade informa a porcentagem de acertos negativos (total de resultados negativos dividido pela soma dos resultados negativos e falsos positivos), e a Precisão calcula a porcentagem de acertos tanto positivos quanto negativos (soma de resultados positivos e negativos dividida pela soma total de positivos, negativos, falsos positivos e falsos negativos). As Figuras 3, 4, 5 e 6 mostram quais foram os resultados obtidos para cada tipo de validação da rede, dentro de cada *cross-validation* executado para cada valor de RHOa, e uma média aritmética entre eles. Como podemos observar, o valor de RHOa que teve uma melhor performance foi de 0,95 e , para as execuções com o melhor valor de RHOa apresentado, os resultados das métricas foram de 91,075% de Acurácia, 94,46% de Sensibilidade, 86,188% de Especificidade e 91,161% de Precisão. Comparando a Acurácia deste trabalho com as dos trabalhos correlatos podemos observar que, entre os 4, este ficou em terceiro lugar, sendo menos eficiente apenas do que a RNA ART2, mas com somente 7,075% a menos do que a RNA do tipo Memética de Pareto, que ficou em segundo lugar.

6. Conclusão

Conforme a análise das execuções da ARTMAP *Fuzzy* com diferentes parâmetros e a conclusão dos resultados que a rede atingiu, comparando com a porcentagem de acertos das outras redes analisadas nos trabalhos correlatos, foi concluído que este tipo de rede neural artificial também pode ser empregado para auxiliar na detecção de nódulos malignos ou benignos da mama, devido à alta taxa de acerto, a complexidade do assunto tratado e o valor de RHOa obtido no melhor resultado, que mostra que no momento em que a camada F1 compara os valores de entrada com os pesos, a semelhança entre os neurônios deve ser de 95%, uma taxa consideravelmente alta.

Também devemos considerar a taxa de falsos positivos e negativos para classificar a RNA como um bom método para auxiliar na detecção de nódulos malignos ou benignos, pois se uma mulher tem, através da RNA, a constatação de que possui o câncer de mama, mas no decorrer do tempo outros exames indicam que não possui, gera menos transtornos para a paciente do que se ela tivesse o câncer e a RNA não detectasse, fazendo com que ela não fosse em busca do tratamento, pois neste caso a doença passaria despercebida até o tratamento se tornar ineficaz. Como a execução desta RNA com acurácia de 91,08% teve 5,54% dos nódulos considerados incorretamente como maligno e 13,81% considerados incorretamente como benigno, também podemos considerar que este tipo de análise é

promissor, e com um estudo mais aprofundado pode ser que um dia faça parte do processo de detecção do câncer de mama.

As redes neurais artificiais do tipo Algoritmos Genéticos, com 100% de acurácia e do tipo Memética de Pareto, com 98,1% de acurácia tiveram maiores alcances quanto a porcentagem de diagnósticos corretos, comparando-as com a rede utilizada no artigo, porém, não é possível dizer com certeza que elas são melhores e trarão melhores resultados quando aplicadas na medicina, pois os métodos de teste aplicados não foram os mesmos, o que pode resultar em algumas variações. Além disso, é possível realizar algumas melhorias nos tipos de dados coletados, nos parâmetros utilizados na rede ou, até mesmo, em como os testes são realizados, podendo resultar em uma maior eficácia da rede.

7. Trabalhos futuros

Conforme dito anteriormente, é possível que a RNA consiga alcançar melhores resultados e, para isso, trabalhar com outras bases de dados em trabalhos futuros pode mostrar uma visão diferente e mais assertiva. Outra sugestão para um trabalho futuro o aprimoramento do parâmetro RHOa, que a princípio foi treinado e testado com apenas 4 valores: 0,95, 0,90, 0,85 e 0,80. Como pudemos observar, os valores mais altos de RHOa tiveram melhores resultados de maneira geral, então encontrar o valor ideal para este parâmetro pode ser uma maneira de aumentar a eficácia da rede.

Ao analisar os dados da tabela de [Mangasarian] é possível perceber que, para os nódulos malignos, os valores das características são mais altos comparados com os dos nódulos benignos, mas algumas características possuem uma diferença maior entre ambos do que outras. Para trabalhos futuros, uma análise mais apurada sobre quais características são mais relevantes na análise e quais confundem a RNA, pode trazer uma nova visão dos dados e chegar mais próximo da acurácia de 100%.

Referências

- Abbass, H. A. *An evolutionary artificial neural networks approach for breast cancer diagnosis*. *Artif Intell Med*. 2002 Jul;25(3):265-81. doi: 10.1016/s0933-3657(02)00028-3.
- de Câncer José Alencar Gomes da Silva (INCA), I. N. (2020). *Estimativa 2020: incidência de câncer no Brasil*. Ministério da Saúde. ISBN 978-85-7318-389-4.
- dos Passos Ramos, S. (2021). *Procedimentos diagnósticos nas lesões não palpáveis da mama*. Em *Ginecologia de Consultório*, páginas 311–325. Editora de Projetos Médicos, 1ª edição.
- dos Santos, R. C. Casagrande, L. C. S. Crotti, Y. Marcelino, R. Gruber, V. *Rede neural artificial otimizada para classificação de câncer de mama*. 2016. XV Congresso Brasileiro de Informática em Saúde.
- dos Santos Junior, C. R. Abreu, T. Lopes, M. L. M. Lotufo, A. D. P. *A new approach to online training for the Fuzzy ARTMAP artificial neural network*. 2021. *Applied Soft Computing*. Volume 113, parte B, 107936.
- Education, I. C. (2020). *Redes neurais*. <https://www.ibm.com/br-pt/cloud/learn/neural-networks>. Acesso em: 09/08/2022.

- Lakshana, G. V. (2021). 4 ways to evaluate your machine learning model: Cross-validation techniques (with python code). <https://www.analyticsvidhya.com/blog/2021/05/4-ways-to-evaluate-your-machine-learning-model-cross-validation-techniques/>. Acesso em: 14/06/2022.
- Mangasarian, O. L., W. H. W. *Cancer diagnosis via linear programming. SIAM News, Volume 23, Number 5, September 1990, pp 1 & 18.*
- Matos, D. (2019). Por que cientistas de dados escolhem python? <https://www.cienciaedados.com/por-que-cientistas-de-dados-escolhem-python/>. Acesso em: 14/07/2022.
- Rasool, A. (2022). Bunternngchit, C. Tiejian, L. Islam, M.R. Qu, Q. Jiang, Q. *Improved Machine Learning-Based Predictive Models for Breast Cancer Diagnosis. Int. J. Environ. Res. Public Health, 19, 3211.*
- Seema Singh, Sunita Saini, M. S. (2012). *Cancer Detection Using adaptive Neural Network. International Journal of Advancements in Research & Technology. Volume 1, Issue 4 ,ISSN 2278-7763.*

Documento Digitalizado Público

Artigo TCC - Luana Pinto (Revisão Final)

Assunto: Artigo TCC - Luana Pinto (Revisão Final)
Assinado por: Edgar Noda
Tipo do Documento: Anexo
Situação: Finalizado
Nível de Acesso: Público
Tipo do Conferência: Documento Digital

Documento assinado eletronicamente por:

- **Edgar Noda, PROFESSOR ENS BASICO TECN TECNOLOGICO**, em 15/08/2022 19:59:27.

Este documento foi armazenado no SUAP em 15/08/2022. Para comprovar sua integridade, faça a leitura do QRCode ao lado ou acesse <https://suap.ifsp.edu.br/verificar-documento-externo/> e forneça os dados abaixo:

Código Verificador: 1069906

Código de Autenticação: 12b1c259b1

