

Universidade Federal de Alfenas

Matheus Rodrigo Aprigio Candido

***Redes Neurais Autoexpansíveis baseadas na
Teoria da Ressonância Adaptativa para a
Classificação de Doenças Hepáticas***

Alfenas/MG

2023

Matheus Rodrigo Aprigio Candido

***Redes Neurais Autoexpansíveis baseadas na
Teoria da Ressonância Adaptativa para a
Classificação de Doenças Hepáticas***

Trabalho de Conclusão de Curso apresentado como parte dos requisitos para obtenção do título de Licenciado em Matemática, pelo Instituto de Ciências Exatas, da Universidade Federal de Alfenas. Área de concentração: Redes Neurais Artificiais. Orientadora: Angela Leite Moreno.

Alfenas/MG

2023

Matheus Rodrigo Aprigio Candido

*Redes Neurais Autoexpansíveis baseadas na Teoria da
Ressonância Adaptativa para a Classificação de Doenças
Hepáticas*

A Banca examinadora abaixo-assinada, aprova a Monografia apresentada como parte dos requisitos para obtenção do título de Licenciado em Matemática, pelo Instituto de Ciências Exatas, da Universidade Federal de Alfenas. Área de concentração: Redes Neurais Artificiais.

Aprovado em: / /

Banca Examinadora:



Prof. Dra. Angela Leite Moreno
Instituto de Ciências Exatas
Orientadora



Prof. Dra. Mara Lucia Martins Lopes
Faculdade de Engenharia - FEIS/UNESP
Campus de Ilha Solteira
Avaliador 1



Documento assinado digitalmente
ANDERSON JOSE DE OLIVEIRA
Data: 19/12/2023 16:50:17-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Anderson José de Oliveira
Instituto de Ciências Exatas
Avaliador 2

.

Prof. Dr. José Claudinei Ferreira
Instituto de Ciências Exatas
Suplente

Ao meu pai, que me deu todo o incentivo necessário.

Agradecimentos

Agradeço a todas as pessoas que de forma direta ou indireta me incentivaram a chegar até aqui.

Gratidão!!

Resumo

O aprendizado de máquina vem se destacando na detecção de várias doenças, tais como, diversos tipos de câncer, doença de Parkinson, entre outras. Entre elas, a detecção de doenças no fígado é extremamente importante, pois o fígado desempenha papel crucial no sistema digestório e é a maior glândula do corpo humano, atuando em processos como digestão dos alimentos, armazenamento de nutrientes, eliminação de drogas e toxinas do organismo, e produção da bile, que é essencial na emulsificação das gorduras. O mau funcionamento dessas funções pode levar a doenças como hepatites, cirrose, câncer de fígado, insuficiência hepática e diversas outras. Tem-se as doenças hepáticas crônicas, responsáveis por 2,6% de todas as mortes do mundo, um número crescente que contabiliza ao menos 2,47 milhões de mortes por ano, se posicionando como a décima maior em número de mortes. Contudo, a maioria dessas doenças são assintomáticas, por isso um diagnóstico eficaz e dentro da fase curável da doença poderia reduzir substancialmente o número de óbitos. Em especial, neste trabalho, modelos baseados em aprendizado de máquina, mais especificamente, redes neurais artificiais, serão utilizadas para a classificação de doenças hepáticas e para análise comparativa, a fim de melhorar o nível de sensibilidade de diferentes resultados obtidos pela literatura. O processo de pesquisa fundamenta-se na implementação das redes neurais utilizando o banco de dados de doenças hepáticas BUPA (*British United Provident Association*), disponível no repositório *UCI Machine Learning Repository*. Em específico, as redes neurais baseadas na Teoria da Ressonância Adaptativa Autoexpansível Norma- p , com $p = 2$ e $p = 64$ foram utilizadas para a classificação dos dados. As redes alcançaram acurácia média de 75,93% e 77,39% para as normas $p = 2$ e $p = 64$, respectivamente, cujos índices de *F-Score*, métrica utilizada em saúde, foram de 0,81 e 0,82.

Palavras-chave: Aprendizado de máquina. Classificação. Reconhecimento de padrões. Diagnóstico médico. Distúrbios no fígado.

Abstract

Machine learning has highlighting in the detection of various diseases, such as different types of cancer, Parkinson's disease, among others. Among these, the detection of liver diseases is extremely important, as the liver plays a crucial role in the digestive system and is the largest gland in the human body, acting in processes such as digesting food, storing nutrients, eliminating drugs and toxins from the body, and producing bile, which is essential for digesting fats. Malfunctions in these functions can lead to diseases such as hepatitis, cirrhosis, liver cancer, liver failure and many others. Among these, we have chronic liver diseases, which are responsible for 2.6% of all deaths worldwide, a growing number that accounts for at least 2.47 million deaths a year, making it the 10th largest in terms of number of deaths. However, most of these diseases are asymptomatic, so effective diagnosis within the curable phase of the disease could substantially reduce the number of deaths. In particular, in this work models based on machine learning, more specifically artificial neural networks, will be used for the classification of liver diseases and for comparative analysis in order to improve the level of sensitivity of different results obtained from the literature. The research process is based on the implementation of neural networks using the BUPA (British United Provident Association) liver disease database, available in the UCI Machine Learning Repository. Specifically, neural networks based on the Norma- p Self-Expanding Adaptive Resonance Theory, with $p = 2$ and $p = 64$, were used to classify the data. The networks achieved an average accuracy of 75.93% and 77.39% for the $p = 2$ and $p = 64$ norms, respectively, whose F-Score indices, a metric used in healthcare, were 0.81 and 0.82.

Keywords: Machine learning. Classification. Pattern recognition. Medical diagnosis. Liver disorders.

Lista de ilustrações

Figura 1 – Neurônio biológico.	15
Figura 2 – Neurônio artificial MCP.	15
Figura 3 – Funções de ativação.	16
Figura 4 – Mapa de Campo de uma rede ART.	19
Figura 5 – Pseudocódigo Rede ART Euclidiana Autoexpansível.	23
Figura 6 – Bolas geradas pelas normas p	27
Figura 7 – Pseudocódigo Rede ART Norma- p Autoexpansível.	28
Figura 8 – Funcionamento da rede / Classificador.	33
Figura 9 – Tipos de ajuste dados de treinamento.	34
Figura 10 – Funcionamento da rede / Classificador.	35
Figura 11 – Matriz de confusão.	36
Figura 12 – Comparação métricas ARTAEE.	40
Figura 13 – Comparação métricas ARTAEE-1.	41
Figura 14 – Comparação métricas ARTAEE-2.	43
Figura 15 – Comparação métricas ARTAEE-3.	44
Figura 16 – Comparação métricas rede Norma 64.	45
Figura 17 – Comparação métricas rede Norma 64 log.	46
Figura 18 – Comparação da Acurácia.	47
Figura 19 – Comparação da Sensibilidade.	48
Figura 20 – Comparação do Coeficiente de Correlação de <i>Matthews</i>	49
Figura 21 – Comparação do F-Score.	51
Figura 22 – Comparação resultados <i>F-Score</i> com os presentes na literatura.	52
Figura 23 – Comparação resultados Acurácia com os presentes na literatura.	53

Lista de tabelas

Tabela 1 – Intervalos de definição dos parâmetros.	37
Tabela 2 – Métricas de validação rede neural ARTAEE.	39
Tabela 3 – Comparação da variância e do desvio padrão ARTAEE.	40
Tabela 4 – Métricas de validação rede neural ARTAEE-1.	41
Tabela 5 – Comparação da variância e do desvio padrão ARTAEE-1.	42
Tabela 6 – Métricas de validação rede neural ARTAEE-2.	42
Tabela 7 – Comparação da variância e do desvio padrão ARTAEE-2.	43
Tabela 8 – Métricas de validação rede neural ARTAEE-3.	43
Tabela 9 – Comparação da variância e do desvio padrão ARTAEE-3.	44
Tabela 10 – Métricas de validação rede neural Norma 64.	45
Tabela 11 – Comparação da variância e do desvio padrão Norma 64.	45
Tabela 12 – Métricas de validação rede neural Norma 64 log.	46
Tabela 13 – Comparação da variância e do desvio padrão Norma 64 log.	47
Tabela 14 – Comparação Acurácia redes.	48
Tabela 15 – Comparação Sensibilidade redes.	49
Tabela 16 – Comparação MCC redes.	50
Tabela 17 – Comparação F-Score redes.	51
Tabela 18 – Comparativo dos resultados obtidos com os presentes na literatura.	52

Sumário

	Lista de ilustrações	7
	Lista de tabelas	8
1	INTRODUÇÃO	10
2	REDES NEURAIS BASEADAS NA TEORIA DA RESSONÂNCIA ADAPTATIVA.	13
2.1	Redes Neurais Artificiais	13
2.2	Neurônio Biológico	14
2.3	Neurônio Artificial	15
2.4	Breve Histórico	17
2.5	Teoria da Ressonância Adaptativa	18
2.6	Rede Neural ART	19
2.7	Rede Neural ART Euclidiana Autoexpansível	22
2.8	Rede Neural p-ART Autoexpansível	27
3	CLASSIFICAÇÃO DE DOENÇAS HEPÁTICAS ATRAVÉS DAS REDES ART	32
3.1	Banco de Dados	33
3.2	Validação	34
3.3	Métricas de Validação	35
3.4	Pré-Processamento dos Dados	37
4	<i>RESULTADOS COMPUTACIONAIS</i>	39
4.1	Rede Neural Autoexpansível ART Euclidiana	39
4.2	Rede Neural Autoexpansível p-ART	44
4.3	Discussão dos Resultados	47
5	<i>CONSIDERAÇÕES FINAIS</i>	54
	REFERÊNCIAS	55

1 *Introdução*

Ao longo da maturação biológica, o cérebro humano foi capaz de desenvolver várias características como: paralelismo maciço (capacidade de processar grande volume de dados em constante expansão), capacidade de aprendizado, capacidade de generalização, adaptatividade, processamento de informação contextual inerente, tolerância ao erro, entre outras. A partir disso, espera-se que algoritmos baseados nesses sistemas neurais biológicos possam reproduzir algumas dessas características a fim de resolver uma variedade de problemas computacionais desafiadores, como reconhecimento de padrões, mineração de dados, aproximação de funções e classificação de doenças (JAIN; MAO; MOHIUDDIN, 1996).

Nesse contexto, recentemente modelos computacionais baseados em aprendizado de máquina têm sido amplamente utilizados para clusterização de dados, melhorando a sensibilidade na previsão, classificação e diagnóstico de doenças em diversas áreas médicas. Entre os métodos de aprendizado de máquina utilizados para categorizar distúrbios biológicos, tem-se as redes neurais artificiais.

Redes neurais artificiais podem ser utilizadas para realizar modelagem não linear e fornecer novas alternativas para regressão logística, o método mais comumente usado para desenvolver modelos preditivos em resultados dicotômicos na medicina, como o estado de um paciente, doente ou saudável (SHAHID; RAPPON; BERTA, 2019). Além disso, os sistemas de redes neurais artificiais têm sido amplamente adotados por terem um custo operacional reduzido.

Em particular, tem-se as redes neurais baseadas na Teoria da Ressonância Adaptativa (Adaptive Resonance Theory - ART), redes inspiradas em como o cérebro humano aprende e reconhece padrões com recorrência em um ambiente de constantes mudanças. Mais ainda, as redes ART são sistemas capazes de aprender novas informações significantes (sensível) e permanecer estável em resposta a padrões de entrada irrelevantes ou repetidos (estabilidade), bem como eventos esperados ou inesperados, sem perder o conhecimento previamente aprendido (plasticidade) (CARPENTER; GROSSBERG, 1987). Essas características indicam que as redes ART, como possível ferramenta para a classificação, podem auxiliar no diagnóstico de doenças, em específico, para a classificação de doenças hepáticas crônicas.

Doenças hepáticas crônicas (DHCs) são qualquer forma de doença hepática avançada, em que o fígado foi exposto de forma contínua a uma ou várias formas de agressão. Pertencente ao sistema digestório, o fígado pesa entre 1200-1500 gramas, existindo duas principais divisões desse órgão, o lobo esquerdo e o lobo direito, que por sua vez é cerca de

seis vezes maior do que esquerdo, sendo maior ainda que outras duas subdivisões menores, o lobo caudado e o lobo quadrado (MCCORMICK; JALAN, 2018).

As principais funções do fígado são: oxidação de triglicerídeos para produção de energia, síntese do colesterol e de fosfolipídios, conversão de carboidratos e proteínas em gorduras, ácidos e triglicerídeos, regulação da concentração de glicose no sangue, síntese de aminoácidos não essenciais, desintoxicação de resíduos metabólicos, armazenamento de glicogênio, vitaminas e ferro, desintoxicação de vários medicamentos e toxinas (por ex. álcool) e, síntese e secreção da bile, que atua na emulsificação de gorduras (YOUNG; O'DOWD; WOODFORD, 2013). Essas são apenas algumas das importantes funções desempenhadas por esse órgão, essenciais para a saúde e para o funcionamento adequado do corpo humano.

O funcionamento anormal ou prejudicado destas funções pode ocasionar as chamadas doenças hepáticas crônicas (DHCs), que são causadas por diversas formas de lesões contínuas ao fígado, sendo as principais: hepatites, cirrose, câncer de fígado, insuficiência hepática, entre outras. Somente em 2019, foram cerca de 2,47 milhões de mortes globais ocasionadas por essas DHCs (DATA, 2023). Como a maioria dessas doenças são assintomáticas em sua fase inicial, um diagnóstico precoce pode reduzir significativamente o número de óbitos.

Desta forma, este trabalho apresenta os resultados obtidos por algumas das redes neurais ART para a classificação de distúrbios no fígado e compara com os resultados obtidos pela literatura. Mais especificamente, as redes neurais Autoexpansíveis ART Euclidiana e p-ART foram implementadas com o objetivo de alcançar melhores parâmetros em relação aos resultados já obtidos na literatura, para o problema abordado. Além disso, o objetivo foi encontrar os valores de β e ρ , parâmetros das redes, que propiciam os resultados mais favoráveis para a classificação de doenças hepáticas no contexto deste estudo.

Os dados utilizados para a implementação das redes foram coletados a partir de análises de sangue relacionadas a doenças hepáticas que podem surgir devido ao consumo excessivo de álcool e a partir do número de bebidas ingeridas diariamente.

Para o processo de pesquisa foi realizada a implementação das redes no *Scilab*, um *software* científico para computação numérica e o banco de dados utilizado para os experimentos foi o BUPA (*British United Provident Association*), disponível no repositório *UCI Machine Learning Repository* (DUA; GRAFF, 2017).

Dando sequência ao desenvolvimento do trabalho é apresentado no capítulo 2 o referencial teórico com um breve histórico das redes neurais artificiais, assim como a Teoria da Ressonância Adaptativa (ART) e diversas redes neurais do tipo ART. No Capítulo 3 é apresentada a metodologia, que descreve a finalidade e o processo de pesquisa,

além do banco de dados utilizado. Já o Capítulo 4 é dedicado à discussão dos resultados e à comparação com os resultados já obtidos na literatura. Por fim, as considerações finais são apresentadas no Capítulo 5 e, e em seguida, as referências utilizadas para o desenvolvimento deste trabalho.

2 *Redes Neurais baseadas na Teoria da Ressonância Adaptativa.*

Neste capítulo serão apresentadas algumas definições, características e um breve histórico sobre as redes neurais artificiais. Além disso, será apresentada a Teoria da Ressonância Adaptativa e o funcionamento das redes utilizadas neste trabalho.

2.1 Redes Neurais Artificiais

Chamadas de redes neurais artificiais (RNAs), também conhecida como conexismo ou sistemas de processamento paralelo e distribuído, as redes neurais tiveram seu conceito introduzido em 1943 com McCulloch e Pitts, conceito que se origina de um modelo computacional baseado em sistemas cognitivos e neurais, ou seja, um modelo semelhante ao cérebro, que por sua vez Haykin (2001) define como:

Um computador (sistema de processamento de informação) altamente complexo, não-linear e paralelo, que tem a capacidade de organizar seus constituintes estruturais, conhecidos por neurônios, de forma a realizar certos processamentos (p.ex., reconhecimento de padrões, percepção e controle motor) muito mais rápido do que qualquer computador existente. (HAYKIN, 2001)

Desde o nascimento, um cérebro possui a habilidade de fornecer as informações necessárias para que possamos interagir com o ambiente, habilidade que se torna cada vez melhor à medida que “a experiência” vai sendo acumulada. Além disso, o cérebro possui a capacidade de plasticidade que o permite adaptar-se a diversas situações. Inspirada no cérebro humano, uma rede neural é uma máquina projetada para modelar a maneira como o cérebro realiza uma tarefa particular ou uma função de interesse. Essa rede normalmente é implementada utilizando-se componentes eletrônicos ou é simulada por programação em um computador digital (HAYKIN, 2001).

Segundo Haykin (2001), algumas propriedades das redes neurais são:

- Não-linearidade: propriedade muito importante se o mecanismo físico responsável pela geração do sinal de entrada for não-linear.
- Mapeamento de Entrada-Saída: O treinamento da rede é repetido para muitos exemplos do conjunto até que a rede alcance um estado estável.
- Adaptabilidade: capacidade inata de adaptar seus pesos sinápticos a modificações do ambiente.

- Resposta a Evidências: no contexto de classificação de padrões, uma rede neural pode ser projetada para fornecer informação não somente sobre qual padrão particular selecionar, mas também sobre a confiança na decisão tomada.
- Informação Contextual: cada neurônio da rede é potencialmente afetado pela atividade de todos os outros neurônios na rede.
- Tolerância a Falhas: a princípio, uma rede neural exibe uma degradação suave do desempenho em vez de apresentar uma falha catastrófica.
- Implementação em VLSI (*very-large-scale-integration*): ela fornece um meio de capturar comportamentos realmente complexos de uma forma altamente hierárquica.
- Uniformidade de Análise e Projeto: a mesma notação é utilizada em todos os domínios envolvendo a aplicação de redes neurais.
- Analogia Neurobiológica: o projeto de uma rede neural é motivado pela analogia com o cérebro, que é uma prova viva de que o processamento paralelo tolerante a falhas é possível não somente fisicamente, mas também rápido e poderoso.

Quanto à estrutura de uma rede neural artificial, no geral ela é caracterizada por padrões de conexões entre neurônios (chamado de arquitetura), por um método que determina o peso das conexões (chamado de treinamento, fase de aprendizado ou algoritmo) e por uma função de ativação (FAUSETT, 1994).

Dessa forma, inicialmente é necessário apresentar como é realizado o processamento de informação pelos neurônios.

2.2 Neurônio Biológico

De acordo com Braga et al. (2000), um neurônio biológico é uma unidade de processamento de informação e pode ser dividido em três seções: o corpo da célula, os dendritos e o axônio, cada um com funções específicas, porém complementares conforme apresentado a seguir:

- *Dendritos*: tem por função receber as informações, ou impulsos nervosos, oriundas de outros neurônios e conduzi-las até o corpo celular.
- *Corpo celular*: têm por função processar a informação e gerar novos impulsos.
- *Axônio*: tem por função transmitir os impulsos a outros neurônios, passando através do axônio até os dendritos dos seguintes neurônios.

O ponto de contato entre a terminação axônica de um neurônio e o dendrito de outro é chamado de *sinapse*. São pelas sinapses que os nodos (pontos de conexões) se unem funcionalmente, formando redes neurais (BRAGA; LUDERMIR; CARVALHO, 2000).

Na Figura 1 é possível observar o modelo simplificado de um neurônio biológico.

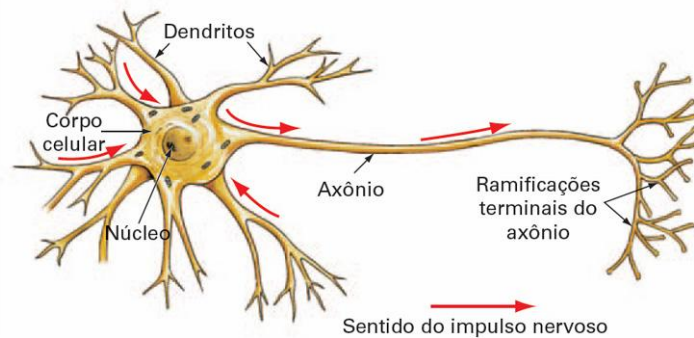


Figura 1 – Neurônio biológico.

Fonte: (BOOK, 2017).

A próxima seção apresenta um modelo de neurônio artificial, inspirado no funcionamento dos neurônios biológicos.

2.3 Neurônio Artificial

Inicialmente, o modelo de neurônio artificial mais empregado foi o proposto por McCulloch e Pitts (1943) (MCP). A descrição matemática dele resultou em um modelo com n terminais de entrada $x_1, x_2, \dots, x_n$ (que representam os dendritos) e apenas um terminal de saída y (representando o axônio). Para imitar o comportamento das sinapses, os terminais de entrada do neurônio têm pesos acoplados $w_1, w_2, \dots, w_n$, cujos valores podem ser positivos (conexões excitatórias) ou negativos (conexões inibitórias). O efeito de uma sinapse particular i no neurônio pós-sináptico é dado por $x_i w_i$. A função de ativação determina "em que grau" o neurônio deve considerar sinais de disparo que ocorrem naquela conexão (BRAGA; LUDERMIR; CARVALHO, 2000). Na Figura 2 é possível observar o modelo de neurônio proposto por McCulloch e Pitt.

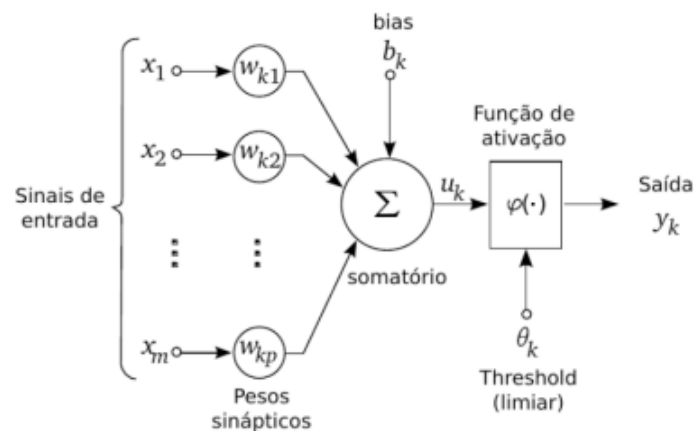


Figura 2 – Neurônio artificial MCP.

Fonte: (HAYKIN, 2001).

A soma ponderada dos valores $x_i w_i$ recebidos pelo neurônio é quem decide se o neurônio deve ou não disparar (saída igual a 1 ou a 0), quando comparada à soma obtida

no limite inferior (limiar ou *threshold*) do neurônio, ou seja, o neurônio dispara se a informação que está recebendo é suficientemente relevante para a informação fornecida, caso contrário deve ser ignorada. No modelo MCP, a ativação do neurônio é obtida através da aplicação de uma "função de ativação", que ativa ou não a saída, dependendo do valor da soma ponderada das suas entradas (BRAGA; LUDERMIR; CARVALHO, 2000). Na Figura 3 é possível observar algumas funções de ativação mais utilizadas.

Função	Gráfico	Equação
Degrau Bipolar		$\begin{aligned} \text{Se } x > 0 &\rightarrow y = 1 \\ \text{Se } x = 0 &\rightarrow y = 0 \\ \text{Se } x < 0 &\rightarrow y = -1 \end{aligned}$
Linear		$y = x$
Tangente Hiperbólica		$y = \frac{1 - e^{-\beta x}}{1 + e^{\beta x}}$
Logística (Sigmoide)		$y = \frac{1}{1 + e^{-\beta x}}$
Degrau		$\begin{aligned} \text{Se } x \geq 0 &\rightarrow y = 1 \\ \text{Se } x < 0 &\rightarrow y = 0 \end{aligned}$
Rampa		$\begin{aligned} \text{Se } x > 1 &\rightarrow y = 1 \\ \text{Se } -1 \leq x \leq 1 &\rightarrow y = x \\ \text{Se } x < -1 &\rightarrow y = -1 \end{aligned}$
Base Radial		$y = e^{-x}$

Figura 3 – Funções de ativação.
Fonte: (MORENO, 2010).

A função de ativação utilizada depende do tipo de problema aplicado à rede neural. No caso deste trabalho, como tem-se um problema de classificação com apenas duas classes (doente ou saudável), funções de ativações que apresentam seus resultados na faixa $[0,1]$ podem determinar o pertencer ou não a determinada classe.

Antes de discutir os modelos de redes utilizados neste trabalho, na próxima seção é apresentado um breve histórico das redes neurais artificiais.

2.4 Breve Histórico

Após o artigo de McCulloch e Pitts (1943) surgiram diversos avanços significativos das redes neurais, mas de forma, lenta principalmente por conta do avanço tecnológico necessário para a realização das experimentações. Mais especificamente, em 1949, houve a publicação do livro de Hebb *The Organization of Behavior* que contribuiu para o desenvolvimento de modelos computacionais de sistemas adaptativos e de aprendizagem, onde seu postulado de aprendizagem afirma que a eficiência de uma sinapse variável entre dois neurônios é aumentada pela ativação repetida de um neurônio causada pelo outro neurônio. Em 1952, a publicação do livro de Ashby, *Design for a Brain: The Origin of Adaptive Behavior*, que trata a noção básica de que o comportamento adaptativo não é inato, mas sim aprendido. Somente em 1958 surge uma nova abordagem para o trabalho de McCulloch e Pitts, introduzida por Rosenblatt, em seu trabalho sobre o *Perceptron*, um método inovador de aprendizagem supervisionada, onde parecia que as redes neurais não possuíam limites (ROSENBLATT, 1957). Porém, em 1969 foi encontrado um problema nos *Perceptrons*, pois eles não conseguem resolver problemas do tipo XOR, ocasionando um esfriamento dos investidores na pesquisa sobre redes neurais (HAYKIN, 2001). A partir da proposição do algoritmo retropropagação por Werbos (WERBOS, 1974) e pelo advento da microeletrônica, efetivamente, esta área galgou novos ânimos.

Em 1982, John Hopfield publicou um artigo no qual propôs um modelo de rede em que as conexões dos neurônios são recorrentes, garantindo que a rede obtivesse um estado estável na ausência de entradas externas (HOPFIELD, 1982). Para tanto, Hopfield associou o estado da rede a uma função energética. A obtenção de um estado estável ocorre pela minimização da função energia. Ainda em 1982, Kohonen apresentou um modelo no qual os neurônios estavam conectados com aqueles posicionados mais próximos a eles, estabelecendo um método de aprendizagem competitivo, chamado de mapas auto-organizáveis (*Self-Organizing Maps* - SOM) (KOHONEN, 2012). Este tipo de comportamento baseado na competição entre os neurônios, em que o aprendizado é não-supervisionado, é capaz de armazenar padrões e posteriormente recuperá-los, independentemente de ruídos. Com esses dois trabalhos, renova-se o interesse de alguns pesquisadores na área de redes neurais. Entretanto, somente em 1986 foi que a comunidade científica voltou a se interessar

pelas redes neurais artificiais como ferramenta para reconhecimento de padrões com o trabalho de Werbos (1974) ao aperfeiçoar a ideia do *Perceptron* criando o algoritmo retropropagação do erro (*error backpropagation*). Mais tarde, Rumelhart, Hinton e Williams (1986) notabilizaram o algoritmo de retropropagação, explorando várias aplicações. O sucesso deste algoritmo estimulou o desenvolvimento de muitas pesquisas em redes neurais artificiais e de uma ampla variedade de modelos cognitivos (FAUSETT, 1994).

Atualmente, as áreas de aplicação das redes neurais são várias. Na teoria, as redes neurais artificiais podem ser utilizadas em qualquer tipo de problema computacional. Porém, na prática, o que se tem visto é sua aplicação em problemas mais específicos que requerem mapeamentos mais complexos, nos quais se tem um grande volume de dados disponíveis, mas não se consegue descrever regras para o comportamento do sistema. Outro ponto importante das redes neurais artificiais é a capacidade de ser tolerante a ruídos ou erros nos dados.

Desse modo, na seção 2.5 é introduzido o conceito de Teoria da Ressonância Adaptativa, que foi utilizada para desenvolver os modelos de redes neurais abordados neste trabalho.

2.5 Teoria da Ressonância Adaptativa

Carpenter e Grossberg (1987), baseados em seu trabalho anterior sobre aprendizagem competitiva, criaram o modelo ART (*Adaptive Resonance Theory*). Esse sistema foi desenvolvido para resolver o Dilema da Estabilidade-Plasticidade, que sustenta a ideia de que “um sistema deve ser capaz de aprender a adaptar a um ambiente em mudança (plasticidade), mas a mudança constante pode tornar o sistema instável, porque o sistema pode aprender novas informações apenas esquecendo tudo que já foi aprendido” (TRIPATHI et al., 2019).

Segundo Carpenter e Grossberg (1987), um sistema adequado deve ser capaz de alternar de forma adaptativa entre seus modos estável e plástico, sendo capaz de distinguir entre eventos familiares e desconhecidos.

Dessa forma, as redes neurais baseadas na Teoria da Ressonância Adaptativa, são redes inspiradas em como o cérebro aprende e reconhece padrões com recorrência em um ambiente de constantes mudanças (CARPENTER; GROSSBERG, 1987).

Basicamente, nessa teoria, tem-se uma camada de reconhecimento de baixo-para-cima (*botton-up*) e uma camada generativa de cima-para-baixo (*top-down*). Se o padrão de entrada e o padrão realimentado aprendido coincidirem, então ocorre um processo dinâmico chamado de “Ressonância Adaptativa”, isto é, ocorre a amplificação e o prolongamento da atividade neural. A teoria afirma que a ressonância regula a aprendizagem em redes

neurais com *feedback* (recorrência) (MORENO, 2016).

Este modelo utiliza um mecanismo de aprendizado não-supervisionado e competitivo criando protótipos (*clusters*) dos padrões aprendidos. Existem redes ART dos mais diversos tipos, utilizando paradigmas de aprendizado supervisionado e não-supervisionado. Nesse sentido, a seção 2.6 apresenta a topologia desse tipo de rede e apresenta alguns dos modelos mais conhecidos e utilizados.

2.6 Rede Neural ART

A arquitetura básica de uma rede ART consiste em três grupos de unidades: uma etapa de codificação, onde a rede realiza a leitura e a codificação dos vetores de entrada (camada F_0), um campo para processamento de entrada (camada F_1) onde os dados são classificados e as unidades *clusters* (camada F_2), que após a classificação é responsável por categorizar e agrupar os dados do conjunto, ou seja, rotular os dados. Tem-se ainda um mecanismo de reinicialização, que controla o grau de similaridade dos padrões associados a um mesmo *cluster* e o subsistema orientado que regula os mecanismos de busca e aprendizado, inibindo ou permitindo que as categorias ressoem. Este subsistema é essencial para distinguir se um novo padrão deve ser atribuído ao código existente ou se é necessário um novo código de reconhecimento. A interface combina os sinais de entrada com a camada F_2 , a fim de comparar a similaridade entre o sinal de entrada e o vetor de pesos da unidade *cluster* candidata (CARPENTER; GROSSBERG, 1987). Essa arquitetura pode ser observada na Figura 4.

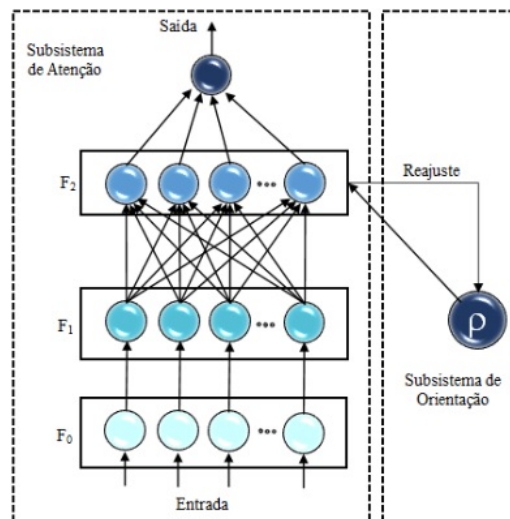


Figura 4 – Mapa de Campo de uma rede ART.

Fonte: Autor.

Os modelos ART são redes competitivas, auto-organizadas, dinâmicas e modulares. Quando uma amostra x é apresentada, após a leitura dos dados ocorre uma competição sobre suas categorias na camada de saída F_2 , onde o vencedor leva tudo. Em seguida, é escolhido o neurônio J que otimiza a função de ativação (ou escolha) desse modelo T entre

os nós, aquela que mais se aproxima do parâmetro procurado (SILVA; ELNABARAWY; WUNSCH, 2019).

Uma categoria representa uma hipótese e o processo de busca de categorias é comumente referido como teste de vigilância, que por sua vez é realizado pelo subsistema orientado para determinar se a categoria vencedora satisfaz um critério de correspondência (ou vários critérios de correspondência). Se a confiança em tal hipótese for maior que o limite mínimo (parâmetro de vigilância ρ), a rede neural entra em um estado de ressonância e o aprendizado (adaptação das unidades de memória de longo prazo (LTM)) é permitido. Caso contrário, a categoria J é inibida, então a próxima categoria de classificação mais alta é selecionada e a busca é retomada. Se nenhuma categoria satisfizer as condições de ressonância necessárias, então uma nova é criada para codificar a amostra de entrada apresentada (SILVA; ELNABARAWY; WUNSCH, 2019).

As condições de ressonância dependem do parâmetro de vigilância ρ , o qual determina o grau de variação admitido no aprendizado de categorias, ou seja, quão refinada será a rede. Com isso, baixo padrão de vigilância implica em maiores generalizações, ou seja, menos categorias são criadas (CARPENTER; GROSSBERG, 1987).

Como as topologias ART apresentam maior complexidade em relação a outras topologias neurais de uso mais geral, pois o número de ciclos durante seu treinamento depende do processo utilizado, é necessária a introdução de neurônios que atuam como controladores do processo de treinamento, possibilitando às redes ART se mostrarem mais capazes de armazenar informações sem incorrer à criação de falsos padrões através de falsas generalizações (FAUSETT, 1994). Assim, existem redes ART dos mais diversos tipos, que utilizam paradigmas de aprendizado, tanto supervisionado quanto não-supervisionado.

No método não-supervisionado, somente o padrão de entrada é fornecido à rede, o que capacita as redes neurais extraírem correlações presentes nos dados de entrada, ou seja, permite a criação de novas classes automaticamente. Desse modo, após a aprendizagem, a rede neural é capaz de classificar suas entradas de acordo com as relações aprendidas (MORENO, 2010). Como o foco deste trabalho são os modelos não-supervisionados, apresentamos a seguir alguns dos mais conhecidos e utilizados, segundo Silva, Elnabarawy e Wunsch (2019).

- ART-1: Utiliza treinamento não-supervisionado. Foram os primeiros modelos desenvolvidos. Sua principal peculiaridade é a utilização de dados binários como entrada da rede.
- ART-2: Utiliza treinamento não-supervisionado. Modelo desenvolvido para processar dados contínuos, podendo também ser utilizada para entrada de dados binários.
- ART-3: Utiliza treinamento não-supervisionado. Modelo que utiliza a ação de neurotransmissores na propagação dos dados pela rede.

- ARTMAP: Utiliza treinamento supervisionado. Composto por duas sub-redes: ARTa, que recebe e processa um determinado padrão de entradas apresentado, e ARTb, que constitui a resposta desejada para o padrão apresentado à rede.
- ART *Fuzzy*: Utiliza treinamento não-supervisionado. Este tipo de rede ART realiza cálculos baseados em operadores *fuzzy Fuzzy*, i.e., sua arquitetura realiza cálculos baseados na lógica nebulosa.
- ARTMAP *Fuzzy*: Utiliza treinamento supervisionado. Esta rede ARTMAP constrói correspondências entre as múltiplas entradas e saídas, convertendo-se num associador de padrões. Nos cálculos são utilizados operadores *fuzzy*.
- ART Euclidiana: Utiliza treinamento não-supervisionado. Modelo desenvolvido para processar dados contínuos podendo também ser utilizada para entrada de dados binários, seu grande diferencial é a utilização da distância euclidiana no tratamento dos dados.
- ARTMAP Euclidiana: Utiliza treinamento supervisionado. Seu diferencial em relação à Rede ARTMAP *Fuzzy* é a utilização da distância euclidiana no tratamento dos dados.
- ARTgrid: A principal novidade da arquitetura proposta é um mecanismo de categorização e pesquisa de dois níveis que pode aumentar a velocidade de computação, mantendo alta performance em casos de maior vigilância.
- ART-ANFIS: Algoritmo de aprendizagem híbrido baseado na teoria da ressonância adaptativa difusa e na técnica de computação *soft* híbrida denominada Sistema ANFIS, Adaptive Neuro *Fuzzy Inference System*.
- IEA-FARTM: Algoritmo evolutivo baseado em algoritmo de classificação de distorção adaptativa (IEA-FARTM), para o sistema de detecção de intrusão.
- *Fuzzy* Min-Max: Utiliza treinamento não-supervisionado. Modelo que usa a teoria do conjunto *fuzzy* para construir *clusters*, usando uma representação de hipercaixa descoberta por meio do algoritmo de aprendizado min-max *fuzzy*. Cada categoria em *fuzzy* min-max é representada explicitamente como uma hipercaixa.
- ART *Distributed*: Modelo que apresenta representação de código distribuída para ativação, correspondência e processos de aprendizado com o objetivo de melhorar a robustez de ruído e compressão de memória em um sistema que apresenta aprendizado rápido e estável.
- ART *Gaussian*: Modelo desenvolvido para reduzir a proliferação de categorias em ambientes com ruídos e para fornecer uma unidade LTM (*long term memory*) de categoria mais eficiente.
- ART *Bayesian*: Arquitetura desenvolvida que utiliza probabilidades condicionais multidimensionais para parametrizar as categorias.

- *TopoART*: Modelo desenvolvido que combina ART *fuzzy* e aprendizado de topologia (inspirado por redes neurais incrementais auto-organizadas).
- *ARTtree*: Algoritmo que constrói uma hierarquia de módulos neurais ART, na qual uma amostra de entrada é enviada simultaneamente para cada módulo em cada nível da árvore.
- *ART Distributed dual vigilance fuzzy*: Modelo ART baseado em vigilância dupla, projetado para melhorar a compressão de memória e executar vários métodos de agrupamento hierárquico aglomerativo (HAC).
- *ART Fusion*: Arquitetura que estende as capacidades do ART ao aumentá-lo com múltiplas e independentes (canais ou campos de entrada) camadas F1, todos conectados a uma camada F2 compartilhada. Esse modelo é capaz de aprender mapeamentos em vários canais simultaneamente.
- *ART Generalized heterogeneous fusion*: modelo projetado para realizar co-agrupamento de dados heterogêneos (dados mistos).

Apresentadas algumas redes do tipo ART, na seção 2.7 é apresentado o funcionamento da rede neural ART Euclidiana Autoexpansível, utilizada neste trabalho para classificação.

2.7 Rede Neural ART Euclidiana Autoexpansível

A rede Neural ART Euclidiana Autoexpansível é baseada na sua antecessora ART Euclidiana, sendo uma técnica de classificação que utiliza a distância euclidiana entre os padrões e os centros das classes para decidir qual será a classificação do padrão. Essa rede não requer normalização e codificação dos vetores de entrada, o que reduz consideravelmente o custo computacional. Consequentemente, os vetores de entrada diferenciam-se dos vetores que não são múltiplos por constante um do outro obtendo assim uma melhor classificação dos padrões. Além disso, com a distância euclidiana entre os padrões e os centros das classes obtêm-se hipersferas centradas no centro das classes (MORENO, 2010).

Diferente das redes baseadas na Teoria da Ressonância Adaptativa tradicionais que verificam a similaridade com todos os *clusters*, ativos e inativos, as redes autoexpansíveis, desenvolvida por Moreno (2016) têm como diferencial a utilização de contadores, o que a torna autoexpansível. A rede inicia seu treinamento com o contador igual a 1, que será o centro do primeiro *cluster* e que se expandirá de acordo com a necessidade de criação de novos *clusters*. Essa modificação torna a rede mais precisa, reduz substancialmente o tempo de processamento e consequentemente o gasto computacional, pois a mesma verifica a similaridade somente entre os *clusters* ativos.

Pela característica de estabilidade e plasticidade, quanto maior o volume de dados melhor o desempenho da rede Neural ART Euclidiana Autoexpansível e ainda que exista

a disponibilidade de novos dados ela consegue adicionar novos padrões à memória sem corromper os padrões previamente criados, de forma permanente. Na Figura 5 é apresentado o pseudocódigo da rede neural Autoexpansível Euclidiana baseada na Teoria da Ressonância Adaptativa.

Entrada: $\mathbf{a} = [a_1 \ a_2 \ \dots \ a_M]$: Dados de Entrada;
 β : Training Parameter ($0 < \beta \leq 1$);
 ρ : Vigilance Parameter ($0 < \rho \leq 1$).
Saída: y^{F2} : Vetor de Atividade

```

3: 1: procedimento
      2:    $Cluster = 1$ 
      3:    $\mathbf{W}_1 = \mathbf{a}_1$ 
      4:   enquanto Data for training exists faça
      5:      $Counter = 1$ 
      6:     para all clusters  $W_j \leftarrow$ 
      7:       até faça  $T_j = \sqrt{\sum_{k=1}^M (a_{ik} - w_{jk})^2}$ 
      8:     fim para
      9:      $flag = True$ 
     10:     enquanto  $flag = True$  faça
     11:        $J = \min\{T_j : j = 1, \dots, Cluster\}$ 
     12:
     13:       
$$MJ = \frac{\sqrt{\sum_{k=1}^M (a_{ik} - w_{jk})^2}}{\max \left\{ \sum_{k=1}^M (a_{ik})^2, \sum_{k=1}^M (w_{jk})^2 \right\}}$$

     14:       se  $MJ \leq \rho$  então
     15:          $\mathbf{W}_j^{new} = \beta \mathbf{a} + (1 - \beta) \mathbf{W}_j^{old}$ 
     16:          $flag = False$ 
     17:       senão
     18:         se  $Cluster > Counter$  então
     19:            $Counter^{new} = Counter^{old} + 1$ 
     20:            $T_J = 1000000000000$ 
     21:         senão
     22:            $Cluster^{new} = Cluster^{old} + 1$ 
     23:            $J = Cluster$ 
     24:            $\mathbf{W}_J = \mathbf{I}$ 
     25:            $flag = False$ 
     26:         fim se
     27:       fim se
     28:       
$$y_i = \begin{cases} 1, & i \neq J, \\ 0, & i = J. \end{cases}$$

     29:     fim enquanto
     30:   fim enquanto
     31: fim procedimento

```

Figura 5 – Pseudocódigo Rede ART Euclidiana Autoexpansível.

Fonte: Autor.

As funções de ativação e a função *match* da rede neural ART Euclidiana Autoexpansível são dadas pelas Equações (2.1) e (2.2), respectivamente:

$$T_j(a) = \sqrt{\sum_{i=1}^M (a_i - w_{ji})^2} \quad (2.1)$$

$$M_j = \frac{T_j}{\max \left(\sum_{i=1}^M a_i^2, \sum_{i=1}^M w_{ji}^2 \right)}. \quad (2.2)$$

Outra característica é que a categoria escolhida é a que possui ativação mínima, Equação (2.3), e a ressonância só ocorre se a função *match* da categoria escolhida for menor que o critério de vigilância como indica a Equação (2.4).

$$J = \operatorname{argmin} \{T_j : j = 1, 2, \dots, N\}, \quad (2.3)$$

se

$$M_J \leq \rho. \quad (2.4)$$

Deste modo, uma baixa vigilância resulta em um maior número de nós de saída, pela maior dificuldade de condição de ressonância. A função de adaptação (vetor de pesos) é dada pela Equação (2.5):

$$\mathbf{w}_J^{\text{nov}} = \beta a + (1 - \beta) \mathbf{w}_J^{\text{velho}}. \quad (2.5)$$

Agora, para uma melhor compreensão das etapas envolvidas durante o processo de treinamento, apresenta-se o algoritmo da Rede Neural ART Euclidiana Autoexpansível.

Passo 1: Leitura dos padrões de entrada: Os dados de entrada são representados pelo vetor como mostra a Equação (2.6):

$$\mathbf{a} = (a_1, a_2, \dots, a_M), \quad (2.6)$$

em que M significa a dimensão do conjunto de dados de entrada.

Passo 2: Leitura dos parâmetros: Os parâmetros utilizados no processamento da rede ART Euclidiana e sua descrição são:

1. Taxa de Treinamento (β); $0 < \beta \leq 1$: Seu valor baixo indica um treinamento lento, enquanto que um valor próximo de 1 implica em um treinamento rápido.
2. Parâmetro de Vigilância (ρ); $0 < \rho \leq 1$: Determina o grau de similaridade para que o *cluster* seja compatível com o dado de entrada.

Passo 3: Inicialização do *cluster*: Inicialmente esta rede possui apenas uma categoria com o primeiro dado de entrada e este primeiro dado de entrada é o centro do *cluster*:

$$w_{j1} = a_{11}, \quad w_{j2} = a_{12}, \quad \dots, \quad w_{jm} = a_{1m}. \quad (2.7)$$

Passo 4: Inicialização dos contadores: inicialização dos contadores de categorias ativas, Equação (2.8):

$$NCat = 1. \quad (2.8)$$

Inicialização dos contadores para criação de uma nova categoria ativa, Equação (2.9):

$$NCont = 1. \quad (2.9)$$

Passo 5: Leitura da entrada atual.

Passo 6: Montagem das categorias: Neste passo é calculada a distância entre o dado de entrada e o centro do *cluster* já existente:

$$T_j(a) = \sqrt{\sum_{i=1}^M (a_i - w_{ji})^2}, \quad (2.10)$$

para $j = 1, 2, \dots, NCat$.

Passo 7: Escolha das categorias: A categoria a ser escolhida é aquela que possui a menor distância:

$$T_j = \min \{T_j : j = 1, 2, \dots, NCat\}. \quad (2.11)$$

Também é necessário saber qual foi o *cluster* vencedor:

$$J = \max \{T_j : j = 1, 2, \dots, NCat\}. \quad (2.12)$$

Passo 8: Cálculo de Similaridade: Por fazer parte das redes da família ART há a necessidade de se calcular a similaridade entre a entrada e a categoria vencedora, tal função de similaridade (função *match*) é dada pela Equação (2.13):

$$M_j = \frac{T_j}{\max \left(\sum_{i=1}^M (a_i)^2, \sum_{i=1}^M (w_{ji})^2 \right)}. \quad (2.13)$$

Passo 9: Teste de vigilância:

$$M_j \leq \rho. \quad (2.14)$$

Pela desigualdade dada pela Equação (2.14), se o critério de vigilância dado for satisfeito passa-se para o **Passo 12**, caso contrário, passa-se para o **Passo 10**.

Passo 10: Verificação e *Reset*: Este passo verifica se a rede já analisou a similaridade com todos os padrões já existentes, caso tenha analisado é criado um novo *cluster*, **Passo 12**, caso contrário, o contador é atualizado:

$$NCont^{novo} = NCont^{velho} + 1. \quad (2.15)$$

Então ocorre o *Reset*, no qual a categoria vencedora é excluída do processo de busca, utilizando a Equação (2.16):

$$T_j = 1000000000. \quad (2.16)$$

Como a categoria vencedora escolhida é aquela possui ativação mínima, é necessário adicionar uma valor alto à ela para que seja possível a sua exclusão do processo de busca.

Retorna-se então ao *Passo 8*.

Passo 11: Criação de um *Cluster*: quando o contador é igual ao número de *cluster* significa que nenhum padrão existente é similar à categoria o suficiente para incorporá-la, sendo assim é necessária a criação de um novo *cluster*. Primeiramente, é necessário indicar a criação do *cluster*:

$$NCat^{novo} = NCat^{velho} + 1. \quad (2.17)$$

Em seguida, é criado um novo *cluster*:

$$\mathbf{w}_{NCat} = a. \quad (2.18)$$

Finalmente, é indicado que o novo *cluster* declarado como vencedor incorporou o dado de entrada, pois houve o processo de ressonância:

$$J = NCat. \quad (2.19)$$

Passo 12: Ressonância (Adaptação dos peso ou Treinamento): neste passo os dados do *cluster* vencedor são atualizados(processo de treinamento), no qual ocorre a modificação do vetor peso dado pela Equação (2.20):

$$\mathbf{w}_J^{novo} = \beta \mathbf{a} + (1 - \beta) \mathbf{w}_J^{velho}, \quad (2.20)$$

em que:

- J é a categoria ativa;
- $\mathbf{w}_J^{novo}$ é o vetor peso atualizado;
- $\mathbf{w}_J^{velho}$ é o vetor peso referente à atualização anterior.

Passo 13: Função de atividade: este passo classifica os valores de saída em 0 ou 1.

$$y_j = \begin{cases} 1, & \text{se } j \neq J \\ 0, & \text{caso contrário.} \end{cases} \quad (2.21)$$

Por fim, retorna-se para o **Passo 5**.

Além da rede neural ART Euclidiana Autoexpansível, este trabalho utilizou a rede neural p-ART Autoexpansível para classificação dos dados. Assim, na próxima seção é apresentado o funcionamento dessa rede.

2.8 Rede Neural p-ART Autoexpansível

Desenvolvida por Silva (2019), a rede neural p-ART Autoexpansível tem funcionamento parecido com a rede neural ART Euclidiana Autoexpansível, entretanto utiliza a norma p no lugar da norma euclidiana ($p = 2$) utilizada por Moreno (2016).

A norma p utilizada está diretamente relacionada como são medidas as distâncias entre os dados de entrada e os centros de *clusters*.

Definição 2.1 Sendo $\mathbf{a}$ e $\mathbf{w} \in \mathbb{R}^n$, com $\mathbf{a} = (a_1, a_2, \dots, a_n)$ e $\mathbf{w} = (w_1, w_2, \dots, w_n)$, tem-se que a norma- p é dada por:

$$d_p(\mathbf{a}, \mathbf{w}) = \sqrt[p]{|a_1 - w_1|^p + |a_2 - w_2|^p + \dots + |a_n - w_n|^p}.$$

Para exemplificar como os novos *clusters* (bolas) se comportam, é apresentado na Figura 6 o caso em que os dados são tridimensionais.

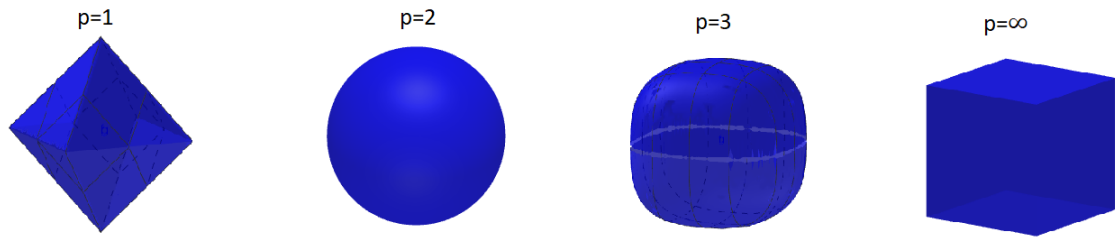


Figura 6 – Bolas geradas pelas normas p .

Fonte: Autor.

Na Figura 7 é apresentado o pseudocódigo da rede neural p-ART Autoexpansível.

Entrada: $\mathbf{a} = [a_1 \ a_2 \ \dots \ a_M]$: Dados de Entrada;
 ρ : norma ($\rho \geq 1$)
 β : Training Parameter ($0 < \beta \leq 1$);
 ρ : Vigilance Parameter ($0 < \rho \leq 1$).
Saída: y^{F2} : Vetor de Atividade

```

1: procedimento
2:    $Cluster = 1$ 
3:    $\mathbf{W}_1 = \mathbf{a}_1$ 
4:   enquanto Data for training exists faça
5:      $Counter = 1$ 
6:     para all clusters  $W_j \leftarrow$ 
7:       até  $faça T_j = \sqrt[p]{\sum_{k=1}^M (a_{ik} - w_{jk})^p}$ 
8:     fim para
9:      $flag = True$ 
10:    enquanto  $flag = True$  faça
11:       $J = \min\{T_j : j = 1, \dots, Cluster\}$ 
12:
13:      
$$MJ = \frac{T_j}{\max \left\{ \sum_{k=1}^M |a_{ik}|^p, \sum_{k=1}^M |w_{jk}|^p \right\}}$$

14:      se  $MJ \leq \rho$  então
15:         $\mathbf{W}_j^{new} = \beta \mathbf{a} + (1 - \beta) \mathbf{W}_j^{old}$ 
16:         $flag = False$ 
17:      senão
18:        se  $Cluster > Counter$  então
19:           $Counter^{new} = Counter^{old} + 1$ 
20:           $T_j = 1000000000000$ 
21:        senão
22:           $Cluster^{new} = Cluster^{old} + 1$ 
23:           $J = Cluster$ 
24:           $\mathbf{W}_j = \mathbf{I}$ 
25:           $flag = False$ 
26:        fim se
27:      fim se
28:      
$$y_i = \begin{cases} 1, & i \neq J, \\ 0, & i = J. \end{cases}$$

29:    fim enquanto
30:  fim enquanto
31: fim procedimento

```

Figura 7 – Pseudocódigo Rede ART Norma-p Autoexpansível.

Fonte: Autor.

Ao se utilizar esse tipo de norma, quando o p é 2, tem-se a rede neural Euclidiana Autoexpansível, entretanto, conforme o p vai aumentando, ao invés de hiperesferas, as

classes vão se aproximando de hipercubos no espaço n dimensional. Com isso, há uma alteração na topologia da rede. Em específico, neste trabalho foi utilizada a rede neural ART Autoexpansível Norma- p , com $p = 64$ por não ser um número tão alto como a norma do máximo, mas simultaneamente, não ser um número tão baixo para diferenciação da topologia em relação a norma com $p = 2$.

Aapresenta-se detalhadamente o algoritmo de treinamento da Rede Neural p-ART Autoexpansível para uma melhor compreensão dos passos envolvidos.

Pré-processamento:

Passo 1: Escolha da norma: Escolha do valor de p para a Norma- p .

Passo 2: Leitura dos dados da primeira entrada: Os dados de entrada são denotados pelo vetor m -dimensional dado pela Equação (2.22):

$$\mathbf{a} = (a_1, a_2, \dots, a_m). \quad (2.22)$$

Passo 3: Parâmetros da rede: Os parâmetros utilizados na rede neural p-ART autoexpansível são:

- Taxa de treinamento: $\beta \in (0,1]$;
- Parâmetro de vigilância: $\rho \in (0,1]$.

Passo 4: Inicialização do *cluster*: Ao contrário da rede ART Euclidiana, esta rede possui, inicialmente, apenas uma categoria com o primeiro dado de entrada, dada pela Equação (2.23):

$$w_{11} = a_1, \quad w_{12} = a_2, \quad \dots, \quad w_{1m} = a_m. \quad (2.23)$$

À medida em que é realizado o treinamento, o número de *clusters* vai aumentando, assim se faz necessária a inclusão de um contador de *clusters*, Equação (2.24):

$$ContW = 1. \quad (2.24)$$

Início do Treinamento: Para cada um dos dados de entrada, é necessário realizar o treinamento a seguir:

Passo 5: Inicialização do contador para criação de um *cluster*, Equação (2.25):

$$Cont = 1. \quad (2.25)$$

Passo 6: Cálculo da distância entre a entrada e o centro de cada *cluster* já existente em um vetor do tipo, é dado pela Equação (2.26):

$$T_j = \sqrt[p]{\sum_{i=1}^M |a_j - w_{ij}|^p} \quad (2.26)$$

Passo 7: Busca da Categoria Vencedora, feita pela Equação (2.27):

$$T_j = \min \{T_j : j = 1, 2, \dots, ContW\}. \quad (2.27)$$

J armazena a coordenada da categoria vencedora, Equação (2.28):

$$J = \max \{T_j : j = 1, 2, \dots, ContW\}. \quad (2.28)$$

Passo 8: Cálculo de Similaridade: Por tratar-se de uma rede neural da família ART- p , há a necessidade de se calcular a similaridade entre a entrada e a categoria vencedora segundo a norma p , esta similaridade é dada pela Equação (2.29):

$$M_j = \frac{T_j}{\max \left(\sum_{i=1}^M |a_i|^p, \sum_{i=1}^M |w_{ji}|^p \right)}. \quad (2.29)$$

Passo 9: Critério de Vigilância: Se o critério de vigilância dado pela Equação (2.30):

$$\mathbf{M}_j \leq \rho, \quad (2.30)$$

for satisfeito passa-se ao **Passo 11**. Caso contrário passa-se para o **Passo 10**, que verifica se a rede já analisou a similaridade com todos os padrões existentes. Se foi analisado, é criado um novo *cluster*, caso contrário, o contador é atualizado, Equação (2.31):

$$Cont^{novo} = Cont^{velho} + 1. \quad (2.31)$$

Passo 10: Criação de um *Cluster*: a similaridade não passou pelo critério de vigilância, então é analisado se já foram verificados todos os *clusters*.

Se $Cont < ContW$ ainda não foram analisados todos os *clusters* e, assim, ocorre o *Reset*, isto é, a categoria vencedora é excluída do processo de busca através da Equação (2.32):

$$T_j = 1000000000, \quad (2.32)$$

que atribui a categoria um valor muito grande para que na escolha pela menor distância ele seja desconsiderado já que não atende ao critério de vigilância. Assim, retorna-se ao **Passo 7**.

Caso $Cont = ContW$, nenhuma categoria existente é similar ao padrão o suficiente para incorporá-lo, assim é necessária a criação de um novo *cluster* através da Equação (2.33).

$$ContW^{novo} = ContW^{velho} + 1. \quad (2.33)$$

O *cluster* criado é declarado vencedor e o padrão toma o lugar do centro da categoria, Equação (2.34):

$$J = ContW^{novo}. \quad (2.34)$$

$$\mathbf{W}_j = a. \quad (2.35)$$

Passo 11: Ressonância: os dados do centro do *cluster* vencedor são atualizados com a influência do padrão recém incluído no *cluster* pela Equação (2.36):

$$\mathbf{w}_j^{novo} = \beta a + (1 - \beta) \mathbf{w}_j^{velho}, \quad (2.36)$$

Passo 12: Função de atividade: este passo classifica os valores de saída em 0 ou 1, Equação (2.37).

$$y_j = \begin{cases} 1, & \text{se } j \neq J \\ 0, & \text{caso contrário.} \end{cases} \quad (2.37)$$

Por fim, retorna-se ao **Passo 5** para um novo dado de entrada.

Apresentado o funcionamento das redes utilizadas neste trabalho, no próximo capítulo é discutido como essas redes foram aplicadas para o problema em questão.

3 Classificação de doenças hepáticas através das redes ART

A finalidade desta pesquisa é de natureza aplicada com a abordagem do problema de forma quantitativa (CERVO; BERVIAN, 1996). Para a aplicação foram utilizadas duas redes neurais baseadas na Teoria da Ressonância Adaptativa Autoexpansível, no problema de classificação de doenças hepáticas: a Rede Neural Euclidiana Autoexpansível (MORENO, 2016) e a Rede Neural Norma- p Autoexpansível (SILVA, 2019). O desempenho obtido na classificação de doenças hepáticas pelas redes neurais utilizadas é avaliado a partir do nível de estabilidade da rede e da taxa de acertos na classificação dos dados.

Para as simulações das redes neurais implementadas durante o desenvolvimento deste trabalho, foi utilizado o SCILAB (2019), um *software* científico para computação numérica. O objetivo principal foi encontrar os melhores parâmetros para otimizar o desempenho do problema em questão, através de uma busca exaustiva dos parâmetros β e ρ , ou seja, foi realizada uma busca parâmetro a parâmetro percorrendo todas as possíveis combinações. O parâmetro β foi variado de 0,05 a 0,70, com um incremento de 0,05 e o parâmetro ρ foi variado de 0,05 a 0,7, com um incremento de 0,001. O parâmetro ρ não foi variado com valores maiores do que 0,7, pois implicaria em menores generalizações, ocasionando que a rede se ajuste em excesso aos dados de treinamento.

Durante a busca exaustiva, as redes ART Euclidiana Autoexpansível e ART Norma- p Autoexpansível, com $p = 64$, foram avaliadas a partir de cada combinação de parâmetros possíveis, de acordo com os incrementos definidos. Dessa forma, foi possível encontrar os valores de β e ρ que apresentam os resultados mais favoráveis durante a classificação dos dados.

A abordagem de busca exaustiva foi essencial para otimizar o desempenho das redes neurais, pois permitiu encontrar os parâmetros que melhor classificaram o problema em questão. Ao explorar uma ampla gama de configurações de parâmetros foi possível determinar a partir da taxa de acertos durante a classificação de doenças hepáticas quais combinações resultaram em um desempenho mais eficiente. Esse processo minucioso foi conduzido para ambas as redes neurais utilizadas neste trabalho, visando obter resultados confiáveis e consistentes.

Para classificar os dados são necessárias três etapas: treinamento, classificação e análise (Figura 8). Durante o treinamento são apresentados os dados para treinamento, que geram *clusters*, cujo centro de classe depende dos dados que foram selecionados para este *cluster* e da ressonância destes durante o processo de treinamento da rede neural.

Após esse processo, cada um dos *clusters* receberá um rótulo de acordo com os dados que o compõem, neste caso, 0 para um paciente saudável e 1 para um paciente doente. Em seguida, é realizada a análise com um conjunto de dados de testes que não foi apresentado anteriormente à rede, e feita a classificação. Com isso, obtém-se os rótulos que são comparados com os valores reais de cada dado, podendo assim ser avaliado o desempenho da rede neural.

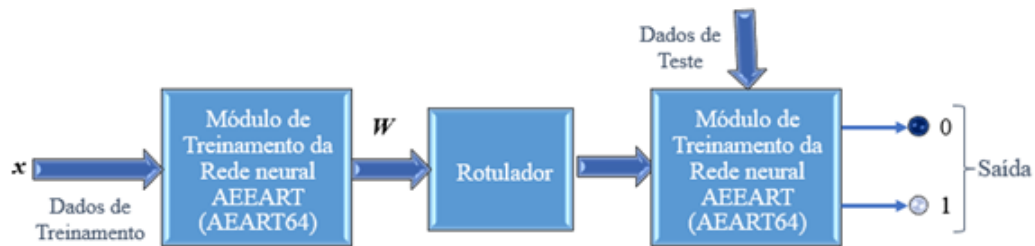


Figura 8 – Funcionamento da rede / Classificador.

Fonte: Autor.

Na seção 3.1 é apresentado o banco de dados que será utilizado para classificação.

3.1 Banco de Dados

Para a classificação de doenças hepáticas, o banco de dados utilizado para a realização das simulações foi o BUPA (*British United Provident Association*), disponível no repositório *UCI Machine Learning Repository* (DUA; GRAFF, 2017)).

O conjunto de dados BUPA contém 7 atributos no total, os 5 primeiros atributos são análises de sangue relacionadas a doenças hepáticas que podem surgir devido ao consumo excessivo de álcool, sendo elas: MCV (volume médio de glóbulos vermelhos em uma amostra de sangue), fosfatase alcalina, TGP (transaminase pirúvica), TGO (transaminase oxalacética) e GGT (gama glutamil transpeptidase). O sexto atributo está relacionado ao número de bebidas ingeridas por dia (maior ou menor a 5) e o último atributo se dá na divisão entre pacientes, doentes ou saudáveis. Cada linha do banco de dados constitui o registro de um único indivíduo do sexo masculino, totalizando 345 instâncias, em que: 200 são positivos para algum tipo de distúrbio no fígado e 145 são negativos para qualquer tipo de mau funcionamento.

Como neste trabalho as classes dos dados são desbalanceadas, foi utilizado o Coeficiente de Correlação de *Matthews* (MCC) para a análise dos resultados, pois com ele é possível medir com precisão o erro da rede quando se tem esse tipo de problema.

A seção 3.2 apresenta a metodologia utilizada para validação das redes utilizadas neste trabalho.

3.2 Validação

Para validar os modelos de redes utilizados, neste trabalho foi utilizado o método de validação cruzada. Este método estima o desempenho do classificador, o que permite a comparação com outras redes de aprendizado de máquina.

Na validação cruzada (*k-fold*) os dados são aleatoriamente divididos em k partições mutuamente exclusivas (*folds*) de tamanho aproximadamente igual a n/k exemplos, n sendo o total de dados e k o número de partições. Os dados nos $(k - 1)$ *folds* são utilizados para treinamento e o teste induzido é feito com o *fold* remanescente. Este processo é repetido k vezes, cada vez considerando um *fold* diferente para teste. O resultado final do classificador é calculado a partir do desempenho médio de todas as k iterações (MONARD; BARANAUSKAS, 2003).

Esse método é capaz de avaliar a capacidade de generalização de um modelo. Além disso, essa técnica é bastante relevante para evitar o *overfitting* do modelo, isto é, evitar que o modelo se ajuste em excesso aos dados de treinamento, se aproximando do *optimum*, que se dá quando o modelo se ajusta de forma ótima aos dados como pode ser observado na Figura 9. Com ela também é possível verificar se o modelo está tendo resultados discrepantes dependendo dos dados apresentados à rede. Quando o modelo não consegue se ajustar adequadamente aos dados ocorre o *underfitting* (MONARD; BARANAUSKAS, 2003).

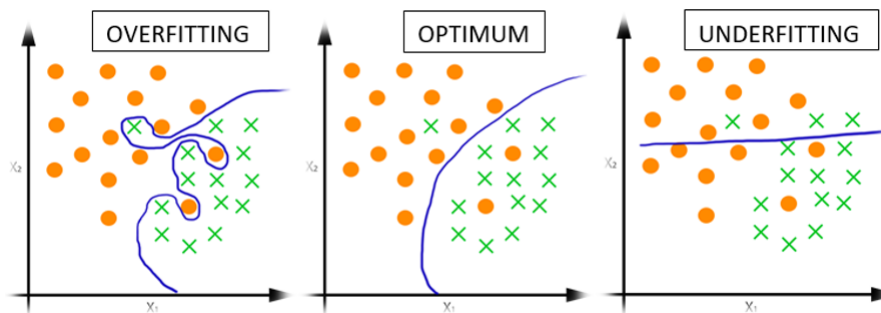


Figura 9 – Tipos de ajuste dados de treinamento.

Fonte: (DEV, 2022).

Para a realização das simulações, os dados deste trabalho foram particionados, utilizando o método de validação cruzada *k-fold*, com $k = 10$, em que os dados foram particionados em 10 partes. Desse modo, tanto os arquivos para treinamento quanto os arquivos para teste durante todas as 10 iterações do método *k-fold* são mutuamente excludentes.

Neste trabalho foi utilizado $k = 10$, pois em diversos estudos, como Kohavi et al. (1995) e Borra e Ciaccio (2010) apontam, este valor é o que apresenta melhor desempenho ao modelar os dados. Dessa forma, os dados foram divididos em 10 partições, 9 delas

utilizadas para treinamento e a remanescente para teste, processo repetido até que todos *folds* sejam utilizados para teste. Na Figura 10 é exemplificado o funcionamento deste método.

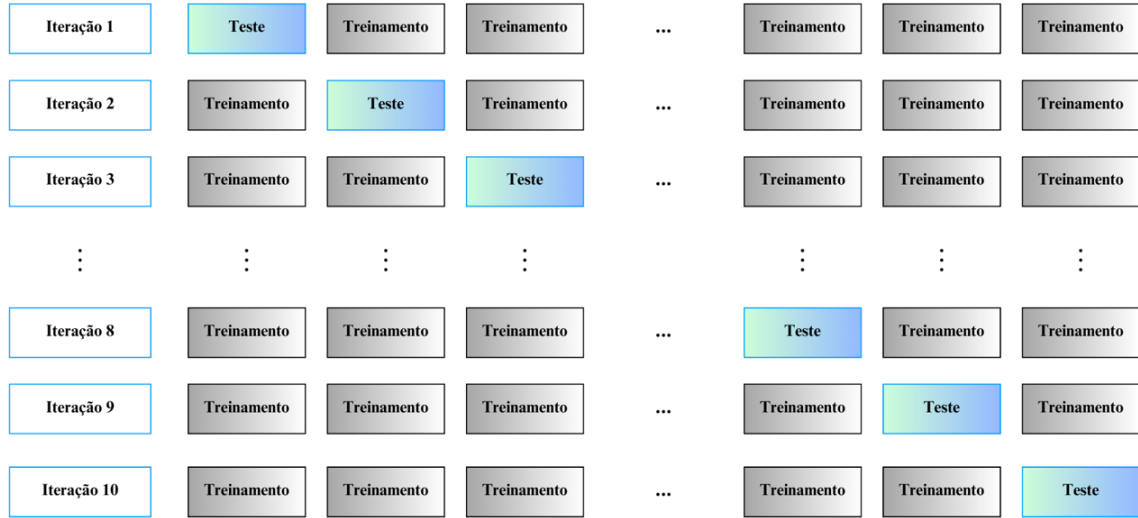


Figura 10 – Funcionamento da rede / Classificador.

Fonte: Autor.

As simulações foram realizadas com as duas redes, a rede neural ART Euclidiana Autoexpansível (MORENO, 2016) e a rede neural p-ART Autoexpansível (SILVA, 2019). Sendo assim, como são 345 instâncias, 200 dados positivos e 145 negativos, os dados de ambas as partições (treinamento e teste) foram embaralhados aleatoriamente para resultados mais honestos.

Para analisar o desempenho das redes e encontrar as melhores combinações de parâmetros, o parâmetro β foi variado de 0,05 a 0,7, com um incremento de 0,05 e o parâmetro ρ foi variado de 0,7 a 0,001, com incremento de 0,001. O mesmo processo foi aplicado para ambas as redes.

Avaliado o desempenho das simulações da rede, é necessário avaliar a qualidade das classificações dos dados. A seção 3.3 apresenta como foi realizada a validação dos resultados.

3.3 Métricas de Validação

O desempenho obtido na classificação de doenças hepáticas pelas redes neurais utilizadas é avaliado de forma quantitativa. Na literatura, os métodos encontrados de avaliação de qualidade das classificações foram fundamentados a partir da matriz de confusão, ou matriz de erros (MONARD; BARANAUSKAS, 2003). Nessa matriz, a diagonal principal expressa os padrões classificados corretamente, enquanto que o restante expressa os dados classificados erroneamente. Assim, tem-se:

- VP: dado positivo classificado corretamente;
- VN: dado negativo classificado corretamente;
- FP: dado positivo classificado incorretamente;
- FN: dado negativo classificado incorretamente.

A matriz de confusão é representada pela Figura 11.

		Valor Predito	
		Sim	Não
Real	Sim	Verdadeiro Positivo (TP)	Falso Negativo (FN)
	Não	Falso Positivo (FP)	Verdadeiro Negativo (TN)

Figura 11 – Matriz de confusão.

Fonte: (NOGARE, 2020).

As métricas de avaliação aplicadas para a validação dos resultados obtidos neste trabalho são:

- **Acurácia (ACC):** avalia a performance geral do modelo, que representa a proporção de dados classificados corretamente em relação a todos os dados classificados e é dada pela Equação (3.1):

$$ACC = \frac{VP + VN}{VP + FP + VN + FN}. \quad (3.1)$$

- **Sensibilidade (Sens):** capacidade do modelo em identificar os indivíduos verdadeiramente positivos, ou seja, identificar um verdadeiro doente, considerando os verdadeiros negativos e falsos positivos e é dada pela Equação (3.2):

$$Sens = \frac{VP}{VP + FN}. \quad (3.2)$$

- **Especificidade (Esp):** capacidade do modelo em identificar os indivíduos verdadeiramente negativos, ou seja, identificar um verdadeiro saudável, considerando os verdadeiros positivos e falsos negativos e é dada pela Equação (3.3):

$$Esp = \frac{VN}{FP + VN}. \quad (3.3)$$

- **Coeficiente de Correlação de Matthews (MCC):** quanto mais próximo de 1 melhor, pois indica uma melhor precisão. Ele leva em consideração os verdadeiros e falsos positivos e negativos, geralmente sendo considerado uma medida balanceada,

pois resiste as disparidades de classes do banco de dados e é dado pela Equação (3.4):

$$MCC = \frac{(VP * VN) - (FP * FN)}{\sqrt{((VP + FP)(VP + FN)(VN + FP)(VN + FN))}}. \quad (3.4)$$

- **F-Score:** média harmônica e bastante utilizada para problemas deste tipo, que combina a precisão e a sensibilidade do modelo. Ela não leva em consideração a incerteza nas previsões, e é dada pela Equação (3.5):

$$F-Score = \frac{2 * VP}{(2 * VP + FP + FN)}. \quad (3.5)$$

Com os métodos de validação dos resultados definidos, foi necessário encontrar maneiras para lidar com o problema de classes desbalanceadas e, assim, obter melhores resultados. Nesse contexto, a seção 3.4 apresenta as formas de pré-processamento de dados que foram utilizadas neste trabalho.

3.4 Pré-Processamento dos Dados

Normalmente os dados em uma rede neural são normalizados em relação ao seu próprio comprimento, ou seja, todos os dados possuem tamanho 1. Como os dados deste trabalho possuem classes desbalanceadas (Tabela 1) este tipo de normalização não seria o mais indicado neste problema em específico. Com isso, durante o pré-processamento dos dados foi necessário a normalização em relação a cada um dos parâmetros, existindo diversas formas para fazer isso.

Tabela 1 – Intervalos de definição dos parâmetros.

Parâmetros	Mínimo	Máximo
MCV	65	103
fosfatase alcalina	23	138
TGP	4	155
TGO	5	82
GGT	5	297
Bebidas/Dia	0	20

Primeiramente, verificou-se os resultados da rede neural baseada na Teoria da Ressonância Adaptativa Autoexpansível Euclidiana com os dados sem nenhuma forma de pré-processamento (ARTAEE).

Para melhorar a qualidade dos resultados obtidos, inicialmente utilizou-se a normalização dos dados em relação aos valores mínimos e máximos de cada parâmetro (ARTAEE-1), deste modo, cada um dos parâmetros foi contraído para o intervalo [0, 1].

Como alguns dos parâmetros eram relativamente altos, também foi realizado um pré-processamento utilizando a função logarítmica (ARTAEE-2). E, por fim, utilizou-se

ambas as ideias: primeiramente aplicando a função logarítmica e, em seguida, normalizando em relação aos valores mínimos e máximos de cada parâmetro ao final da aplicação da função logarítmica (ARTAE-3).

Em relação a rede neural ART Norma- p , com $p = 64$, foi realizada a normalização dos dados em relação aos valores mínimos e máximos de cada parâmetro, novamente contraindo os parâmetros para o intervalo $[0, 1]$ (Norma-64). Em seguida, foi aplicada a função logarítmica, com a rede normalizada (Norma-64 log).

O capítulo 4 apresenta os resultados obtidos pelas redes abordadas neste trabalho para o problema de doenças hepáticas.

4 Resultados Computacionais

Neste capítulo serão apresentados os resultados obtidos pelas redes neurais Autoexpansíveis ART Euclidiana e p -ART, para o problema de classificação de doenças hepáticas.

4.1 Rede Neural Autoexpansível ART Euclidiana

As métricas de validação para a rede neural Autoexpansível ART Euclidiana sem a normalização dos dados (ARTAE) são apresentados na Tabela 2, com cada um dos dez arquivos, e para obter o resultado final da classificação, é apresentada a média do desempenho obtido em cada iteração. Para isso, os dados brutos (sem pré-processamento) foram apresentados à rede.

Tabela 2 – Métricas de validação rede neural ARTAE.

Arquivo	ACC	Sens.	Esp.	MCC	<i>F-Score</i>
1	68,57%	95,00%	33,33%	0,3720	0,7755
2	68,57%	95,00%	33,33%	0,3720	0,7755
3	68,57%	90,00%	40,00%	0,3536	0,7660
4	74,29%	85,00%	60,00%	0,4692	0,7907
5	68,57%	95,00%	33,33%	0,3720	0,7755
6	85,29%	90,00%	78,57%	0,6945	0,8780
7	67,65%	95,00%	28,57%	0,3276	0,7755
8	67,65%	100,00%	21,43%	0,3718	0,7843
9	64,71%	90,00%	28,57%	0,2398	0,7500
10	76,47%	95,00%	50,00%	0,5221	0,8261
Média	71,03%	93,00%	40,71%	0,4094	0,7897

A partir das métricas de validação apresentadas, nota-se que a rede ARTAE atingiu uma acurácia de 71,03%, um Coeficiente de Correlação de *Matthews* de 0,4094 e um *F-Score* de 0,7897.

A Figura 12 apresenta a comparação das métricas da rede ARTAE.

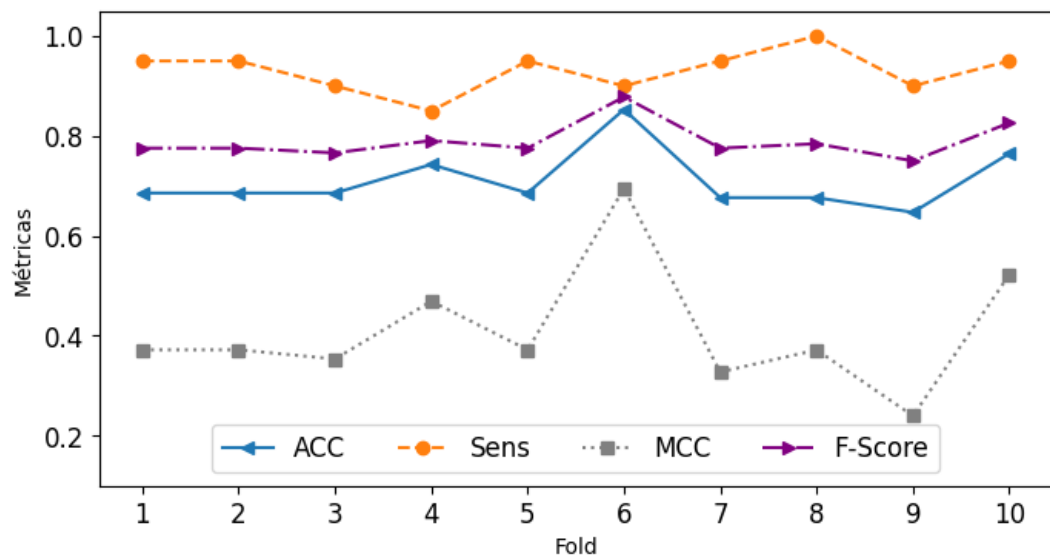


Figura 12 – Comparação métricas ARTAEE.
Fonte: Autor.

A Tabela 3 apresenta a variância e o desvio padrão da rede ARTAEE.

Tabela 3 – Comparação da variância e do desvio padrão ARTAEE.

Métrica	ACC	Sens.	MCC	<i>F-Score</i>
Variância	0,00331	0,0016	0,01417	0,00121
Desvio padrão	0,05749	0,04	0,11905	0,03478

Pela variância, é possível observar que no geral os dados são estáveis, não havendo muitas variações. Já o desvio padrão aponta uma dispersão pequena dos dados, ou seja, a rede é robusta. Assim, pode-se concluir que a rede é tolerante a erros e a eventos inesperados.

Para os resultados da rede neural ART Euclidiana após a normalização dos dados em relação aos valores mínimos e máximos de cada parâmetro (ARTAEE-1), a Tabela 4 apresenta as métricas de validação para cada arquivo, e da mesma forma, para obter o resultado final da classificação, é apresentada a média do desempenho obtido em cada iteração.

Tabela 4 – Métricas de validação rede neural ARTAEE-1.

Arquivo	ACC	Sens.	Esp.	MCC	<i>F-Score</i>
1	82,86%	90,00%	73,33%	0,6487	0,8571
2	74,29%	90,00%	53,33%	0,4747	0,8000
3	68,57%	100,00%	26,67%	0,4148	0,7843
4	77,14%	70,00%	86,67%	0,5630	0,7778
5	74,29%	85,00%	60,00%	0,4692	0,7907
6	73,53%	70,00%	78,57%	0,4781	0,7568
7	73,53%	90,00%	50,00%	0,4462	0,8000
8	67,65%	55,00%	85,71%	0,4123	0,6667
9	67,65%	65,00%	71,43%	0,3586	0,7027
10	64,71%	100,00%	14,29%	0,2988	0,7692
Média	72,42%	81,50%	60,00%	0,4564	0,7705

Ao analisar os arquivos de treinamento, foi observado que o melhor resultado obtido pela rede ARTAEE-1 foi com os parâmetros $\beta = 0,25$ e $\rho = 0,183$, com três casas decimais. Os demais resultados foram com os parâmetros $\beta = 0,2$, $\beta = 0,15$ e $\beta = 0,05$, com o $\rho = 0,077$ e $\rho = 0,434$.

Quanto às métricas de validação apresentadas, nota-se que a rede neural ARTAEE-1 atingiu uma acurácia de 72,42%, um Coeficiente de Correlação de *Matthews* de 0,4564 e um *F-Score* de 0,7705, apresentando portanto uma discreta melhora nos resultados.

Em relação a sensibilidade, alguns *folds* obtiveram desempenho inferior comparado à ARTAEE, porém aqui é apresentado uma melhor combinação de métricas como um todo.

A Figura 13 apresenta a comparação das métricas da rede ARTAEE-1.

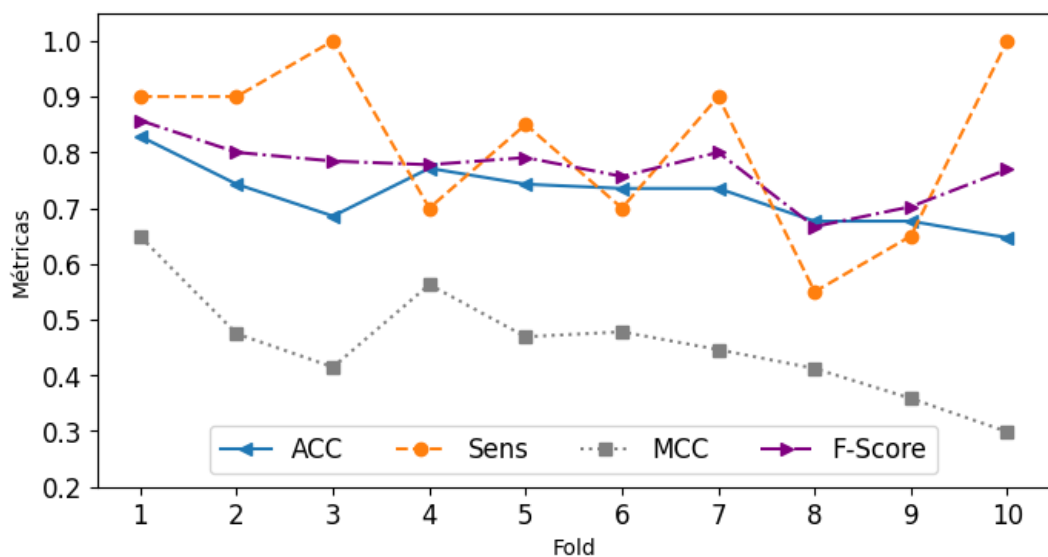


Figura 13 – Comparação métricas ARTAEE-1.

Fonte: Autor.

A Tabela 5 apresenta a variância e o desvio padrão da rede ARTAEE-1.

Tabela 5 – Comparação da variância e do desvio padrão ARTAEE-1.

Métrica	ACC	Sens.	MCC	<i>F-Score</i>
Variância	0,00261	0,02152	0,00875	0,00255
Desvio padrão	0,05104	0,14671	0,09353	0,05045

Novamente é possível observar a robustez da rede, porém com uma variação um pouco maior em relação à rede ARTAEE para as métricas Sensibilidade e *F-Score*, diferente do MCC e da ACC que se mostraram mais estáveis.

A Tabela 6 apresenta os resultados da rede ARTAEE-2, onde foi aplicada a função logarítmica durante o pré-processamento dos dados.

Tabela 6 – Métricas de validação rede neural ARTAEE-2.

Arquivo	ACC	Sens.	Esp.	MCC	<i>F-Score</i>
1	68,57%	90,00%	80,00%	0,3536	0,7660
2	74,29%	90,00%	66,67%	0,4714	0,7826
3	77,14%	90,00%	66,67%	0,5330	0,8182
4	74,29%	90,00%	66,67%	0,4747	0,8000
5	71,43%	60,00%	73,33%	0,4097	0,7619
6	85,29%	90,00%	78,57%	0,6945	0,8780
7	70,59%	90,00%	85,71%	0,3825	0,7826
8	70,59%	80,00%	78,57%	0,3929	0,7619
9	73,53%	100,00%	92,86%	0,4963	0,8163
10	70,59%	90,00%	78,57%	0,3929	0,7826
Média	73,63%	87,00%	76,76%	0,4601	0,7950

Ao analisar os arquivos de treinamento, os melhores resultados obtidos pela rede ARTAEE-2 foram com os parâmetros $\beta = 0,15$, $\beta = 0,20$ e $\beta = 0,25$. Quanto ao parâmetro ρ os melhores resultados foram no intervalo 0,071 a 0,073.

Quanto às métricas de validação apresentadas, nota-se que a rede neural ARTAEE-2 atingiu uma acurácia de 73,63%, um Coeficiente de Correlação de *Matthews* de 0,4601 e um *F-Score* de 0,795, apresentando novamente uma melhora nos resultados.

A Figura 14 apresenta a comparação das métricas da rede ARTAEE-2.

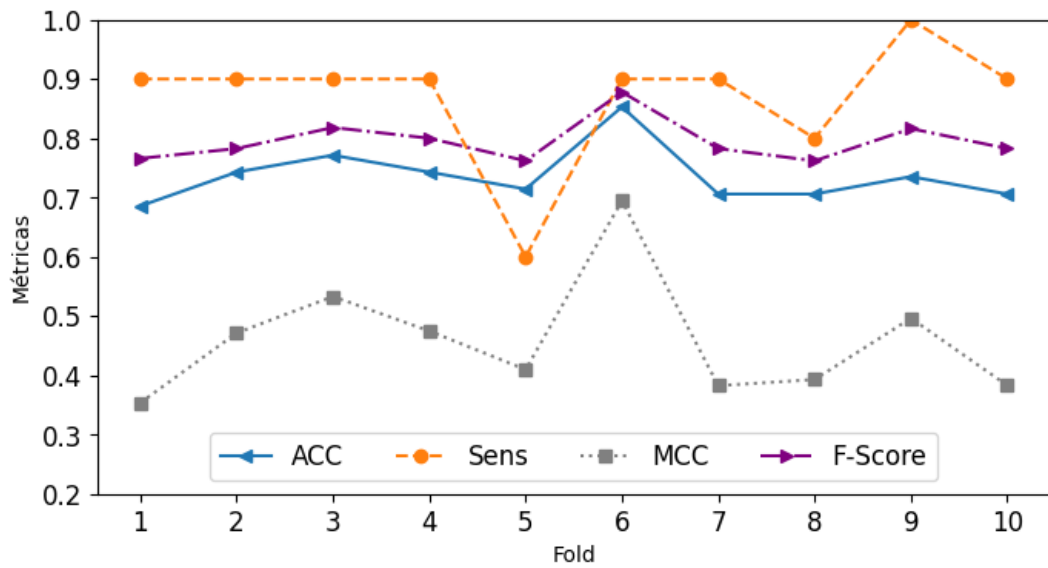


Figura 14 – Comparação métricas ARTAEE-2.

Fonte: Autor.

A Tabela 7 apresenta a variância e o desvio padrão da rede ARTAEE-2.

Tabela 7 – Comparação da variância e do desvio padrão ARTAEE-2.

Métrica	ACC	Sens.	MCC	<i>F-Score</i>
Variância	0,00207	0,0101	0,00923	0,00114
Desvio padrão	0,04553	0,1005	0,09606	0,03376

Aqui, pode-se notar que o conjunto de dados não se distanciam muito da média, mostrando novamente a robustez da rede.

Por fim, para os resultados da rede ARTAEE-3 em que foi aplicado a função logarítmica durante o pré-processamento e realizado a normalização dos dados simultaneamente, a Tabela 8 apresenta as métricas de validação para cada arquivo e a média do desempenho obtido em cada iteração.

Tabela 8 – Métricas de validação rede neural ARTAEE-3.

Arquivo	ACC	Sens.	Esp.	MCC	<i>F-Score</i>
1	80,00%	90,00%	66,67%	0,5908	0,8372
2	77,14%	90,00%	60,00%	0,5330	0,8182
3	74,29%	80,00%	66,67%	0,4714	0,7805
4	71,43%	100,00%	33,33%	0,4714	0,8000
5	80,00%	90,00%	66,67%	0,5908	0,8372
6	73,53%	75,00%	71,43%	0,4602	0,7692
7	76,47%	85,00%	64,29%	0,5076	0,8095
8	70,59%	80,00%	57,14%	0,3825	0,7619
9	79,41%	85,00%	71,43%	0,5715	0,8293
10	76,47%	95,00%	50,00%	0,5221	0,8261
Média	75,93%	87,00%	60,76%	0,5101	0,8069

Aqui, os resultados obtidos pela rede ARTAEE-3 foram com os parâmetros $\beta = 0,05$ e ρ no intervalo 0,444 a 0,451. A rede neural ARTAEE-3 uma redução da Especificidade. Por outro lado, ela atingiu uma acurácia de 75,93%, um Coeficiente de Correlação de *Matthews* de 0,510 e um *F-Score* de 0,807, apresentando os melhores resultados

A Figura 15 apresenta a comparação das métricas da rede ARTAEE-3.

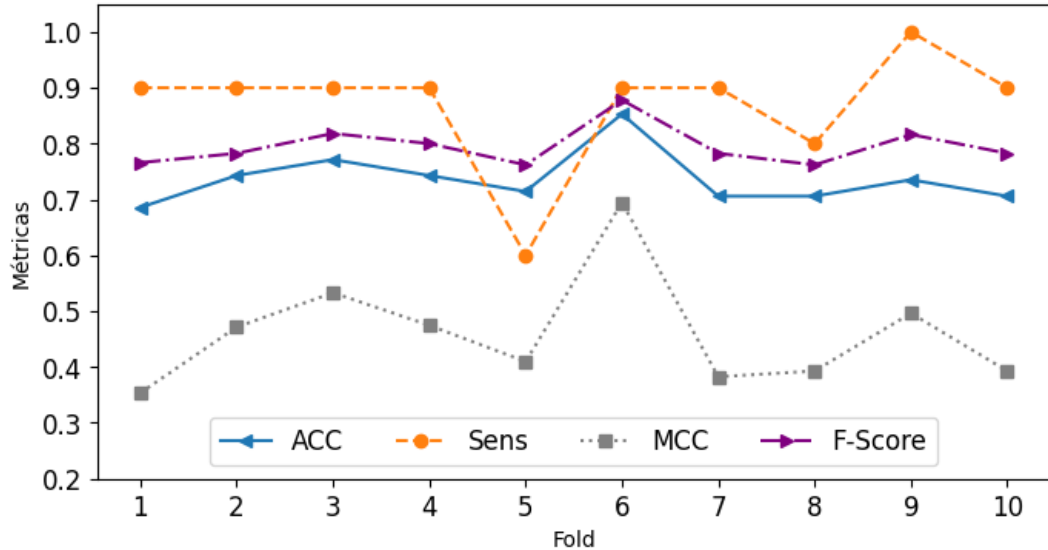


Figura 15 – Comparação métricas ARTAEE-3.

Fonte: Autor.

A Tabela 9 apresenta a variância e o desvio padrão da rede ARTAEE-3, na qual pode-se verificar que as mudanças na rede demonstraram ser eficazes, alcançando melhores métricas em relação aos resultados inicialmente obtidos.

Tabela 9 – Comparação da variância e do desvio padrão ARTAEE-3.

Métrica	ACC	Sens.	MCC	<i>F-Score</i>
Variância	0,00207	0,0101	0,00908	0,00114
Desvio padrão	0,04553	0,1005	0,09531	0,03376

Na seção 4.2 é apresentado os resultados para a rede neural Autoexpansível *p*-ART.

4.2 Rede Neural Autoexpansível *p*-ART

Como os melhores resultados apresentados na rede Euclidiana foram após as normalizações, aqui os resultados obtidos foram apresentados após a normalização dos dados em relação aos valores mínimos e máximos de cada parâmetro.

Na Tabela 10 são apresentadas as métricas de validação para cada arquivo e a média do desempenho obtido em cada iteração da rede neural Autoexpansível *p*-ART, com $p = 64$ (Norma 64).

Tabela 10 – Métricas de validação rede neural Norma 64.

Arquivo	ACC	Sens.	Esp.	MCC	<i>F-Score</i>
1	74,29%	100,00%	40,00%	0,5252	0,8163
2	82,86%	90,00%	73,33%	0,6486	0,8571
3	74,29%	80,00%	66,67%	0,4714	0,7805
4	80,00%	75,00%	86,67%	0,6106	0,8108
5	74,29%	75,00%	73,33%	0,4801	0,7692
6	85,29%	85,00%	85,71%	0,7009	0,8718
7	74,29%	75,00%	73,33%	0,4801	0,7692
8	73,53%	80,00%	64,29%	0,4485	0,7805
9	73,53%	75,00%	71,43%	0,4602	0,7692
10	76,47%	95,00%	50,00%	0,5221	0,8261
Média	76,88%	83,00%	68,48%	0,5348	0,8051

Considerando as métricas de validação apresentadas na Tabela 10 nota-se que a rede Norma 64, atingiu uma acurácia de 76,88%, um Coeficiente de Correlação de *Matthews* de 0,5348 e um *F-Score* de 0,8051, apresentando uma pequena melhora dos resultados.

A Figura 16 apresenta a comparação das métricas da rede Norma 64.

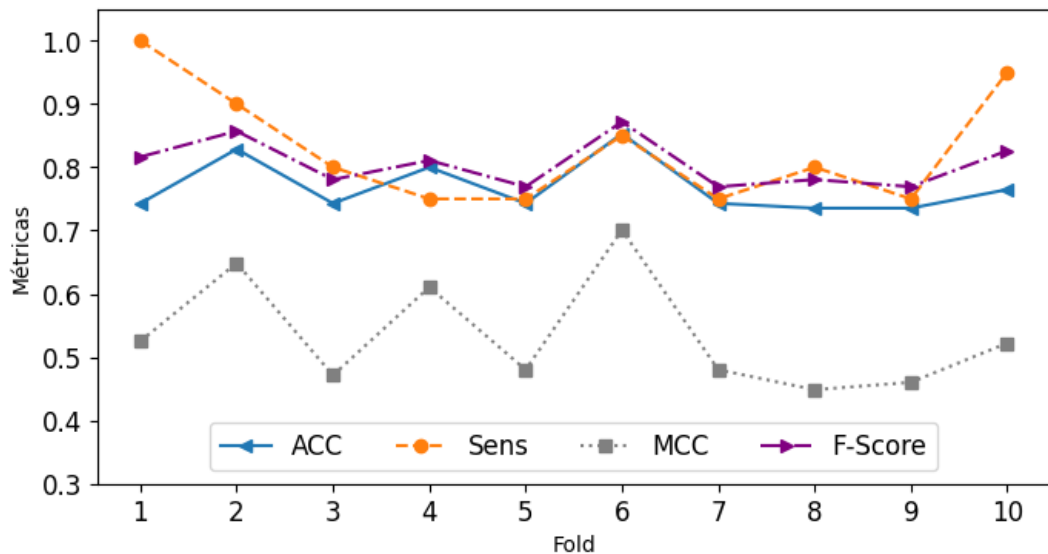


Figura 16 – Comparação métricas rede Norma 64.

Fonte: Autor.

A Tabela 11 apresenta a variância e o desvio padrão da rede Norma 64.

Tabela 11 – Comparação da variância e do desvio padrão Norma 64.

Métrica	ACC	Sens.	MCC	<i>F-Score</i>
Variância	0,00166	0,0076	0,00696	0,00128
Desvio padrão	0,0407	0,08718	0,0834	0,03582

É possível observar que a rede com a Norma 64 também apresentou robustez, em que no geral os dados não se afastaram muito da média.

Ao comparar com a rede ARTAEE-1, que também foi normalizada em relação aos valores mínimos e máximos de cada parâmetro, pode-se notar que a rede Norma 64 apresenta variações um pouco menores em todas as métricas, com uma única exceção, sendo o Desvio padrão do F-Score.

Em seguida, foi aplicada a função logarítmica durante o pré-processamento e novamente realizada a normalização dos dados simultaneamente (Norma 64 log), a Tabela 12 apresenta as métricas de validação para cada arquivo e a média do desempenho obtido em cada iteração.

Tabela 12 – Métricas de validação rede neural Norma 64 log.

Arquivo	ACC	Sens.	Esp.	MCC	<i>F-Score</i>
1	74,29%	90,00%	53,33%	0,4747	0,8000
2	77,14%	90,00%	60,00%	0,5330	0,8182
3	80,00%	90,00%	66,67%	0,5908	0,8372
4	77,14%	90,00%	60,00%	0,5330	0,8182
5	80,00%	90,00%	66,67%	0,5908	0,8372
6	79,41%	95,00%	57,14%	0,5817	0,8444
7	79,41%	95,00%	57,14%	0,5817	0,8444
8	79,41%	90,00%	64,29%	0,5711	0,8372
9	70,59%	70,00%	71,43%	0,4085	0,7368
10	76,47%	95,00%	50,00%	0,5221	0,8261
Média	77,39%	89,50%	60,67%	0,5387	0,8200

A partir das métricas de validação apresentadas, nota-se que a rede atingiu uma especificidade de 60,67, havendo uma redução. Além disso, ela atingiu uma acurácia de 77,39%, um Coeficiente de Correlação de *Matthews* de 0,5387 e um *F-Score* de 0,8200, apresentando pequenas melhoras nos resultados.

A Figura 17 apresenta a comparação das métricas da rede Norma 64 log.

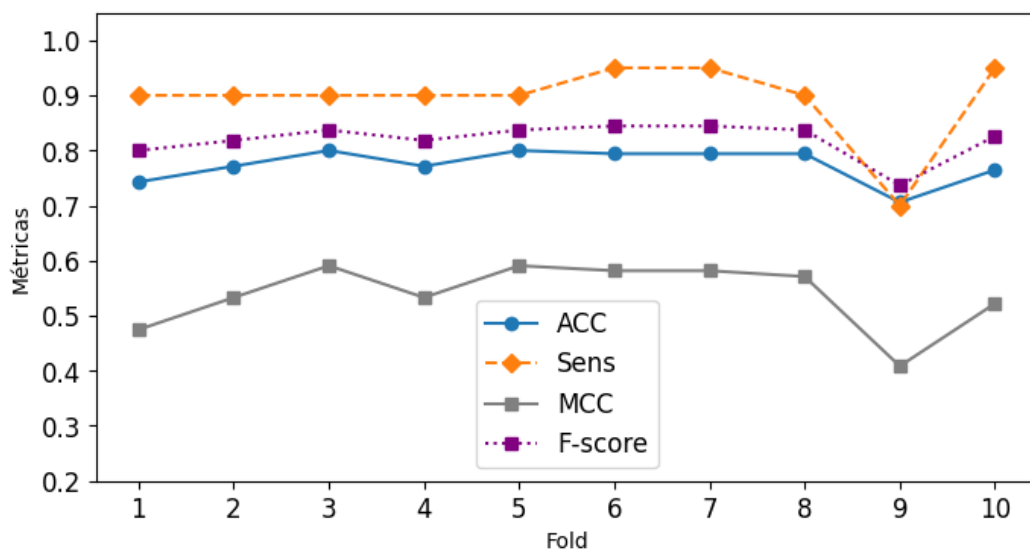


Figura 17 – Comparação métricas rede Norma 64 log.

Fonte: Autor.

A Tabela 13 apresenta a variância e o desvio padrão da rede Norma 64 log.

Tabela 13 – Comparação da variância e do desvio padrão Norma 64 log.

Métrica	ACC	Sens.	MCC	<i>F-Score</i>
Variância	0,00083	0,00473	0,00316	0,00094
Desvio padrão	0,02876	0,06874	0,05618	0,03073

Pode-se notar que novamente as mudanças na rede, agora para $p = 64$, demonstraram ser eficazes, alcançando as melhores métricas em relação aos resultados inicialmente obtidos.

4.3 Discussão dos Resultados

Para melhor visualização dos resultados obtidos neste trabalho, a Figura 18 apresenta a acurácia obtida pelas redes neurais ARTAEE, ARTAEE-1, ARTAEE-2, ARTAEE-3, Norma-64 e Norma-64 com aplicação da função logarítmica (Norma-64 log).

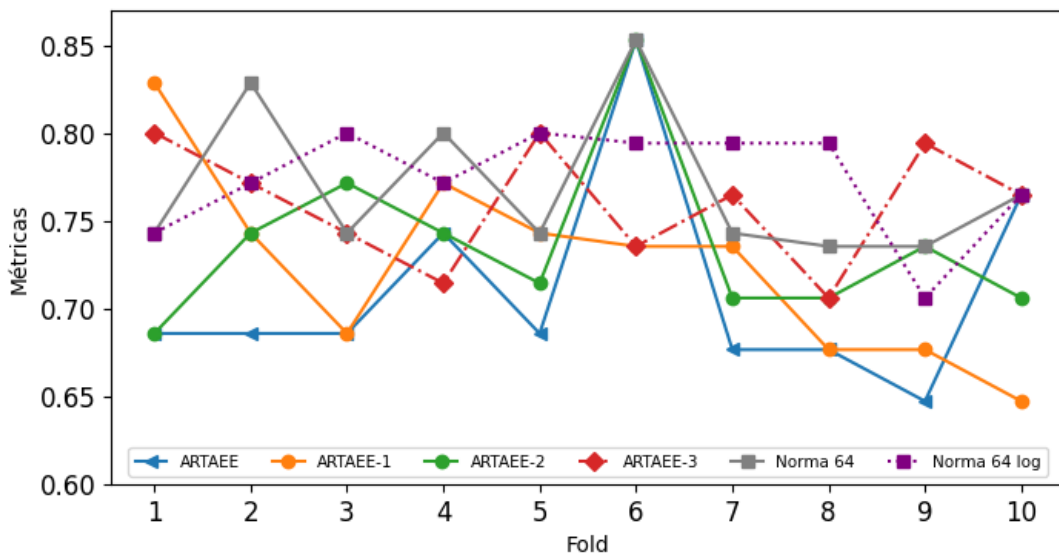


Figura 18 – Comparação da Acurácia.

Fonte: Autor.

A Tabela 14 também apresenta a acurácia obtida pelas redes, porém com o objetivo de melhor visualização do desempenho médio obtido.

Tabela 14 – Comparação Acurácia redes.

Arquivo	ARTAEE	ARTAEE1	ARTAEE2	ARTAEE3	Norma 64	Norma 64 log
1	68,57%	82,86%	68,57%	80,00%	74,29%	74,29%
2	68,57%	74,29%	74,29%	77,14%	82,86%	77,14%
3	68,57%	68,57%	77,14%	74,29%	74,29%	80,00%
4	74,29%	77,14%	74,29%	71,43%	80,00%	77,14%
5	68,57%	74,29%	71,43%	80,00%	74,29%	80,00%
6	85,29%	73,53%	85,29%	73,53%	85,29%	79,41%
7	67,65%	73,53%	70,59%	76,47%	74,29%	79,41%
8	67,65%	67,65%	70,59%	70,59%	73,53%	79,41%
9	64,71%	67,65%	73,53%	79,41%	73,53%	70,59%
10	76,47%	64,71%	70,59%	76,47%	76,47%	76,47%
Média	71,03%	72,42%	73,63%	75,93%	76,88%	77,39%

Pode-se observar que a rede com a Norma-64 log obteve melhores resultados (ou iguais) nos conjuntos de dados 3, 5, 7, 8 e 10, obtendo assim a maior média entre as redes analisadas. No geral, ela também apresentou menores discrepâncias entre os resultados obtidos.

A rede noma-64 obteve os melhores resultados nos conjuntos de dados 2, 4 e 6, apresentando bons resultados, porém com grandes variações.

A sensibilidade obtida pelas quatro redes é apresentada na Figura 19.

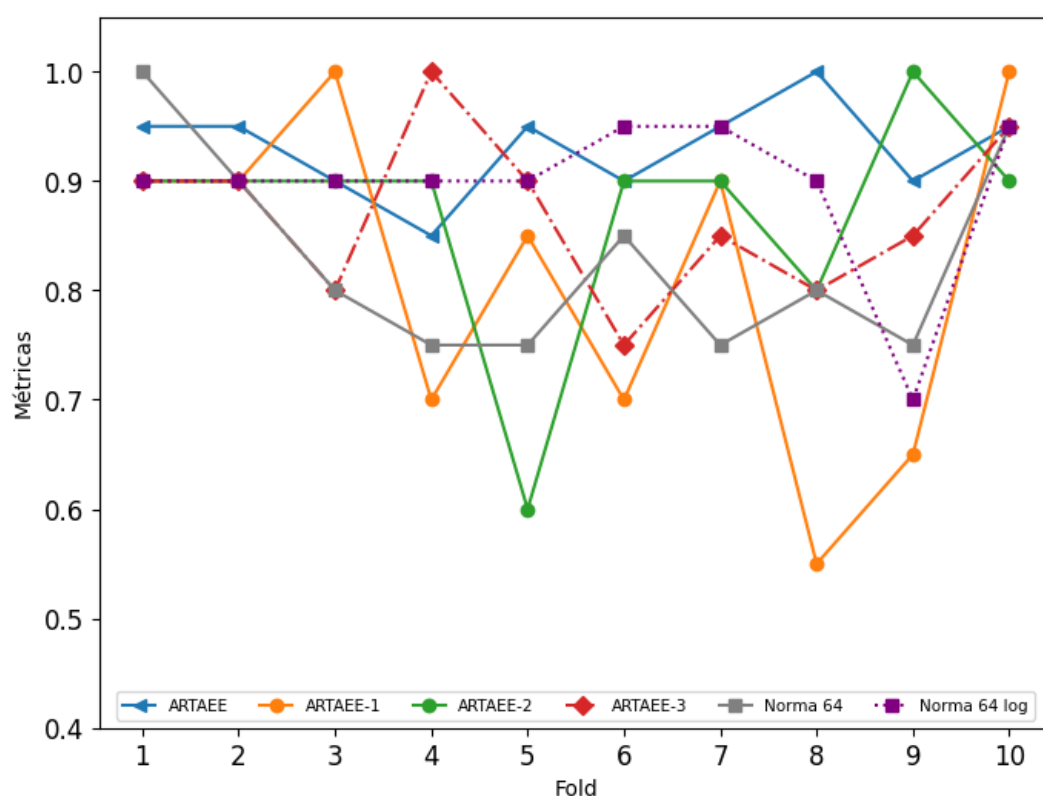


Figura 19 – Comparação da Sensibilidade.

Fonte: Autor.

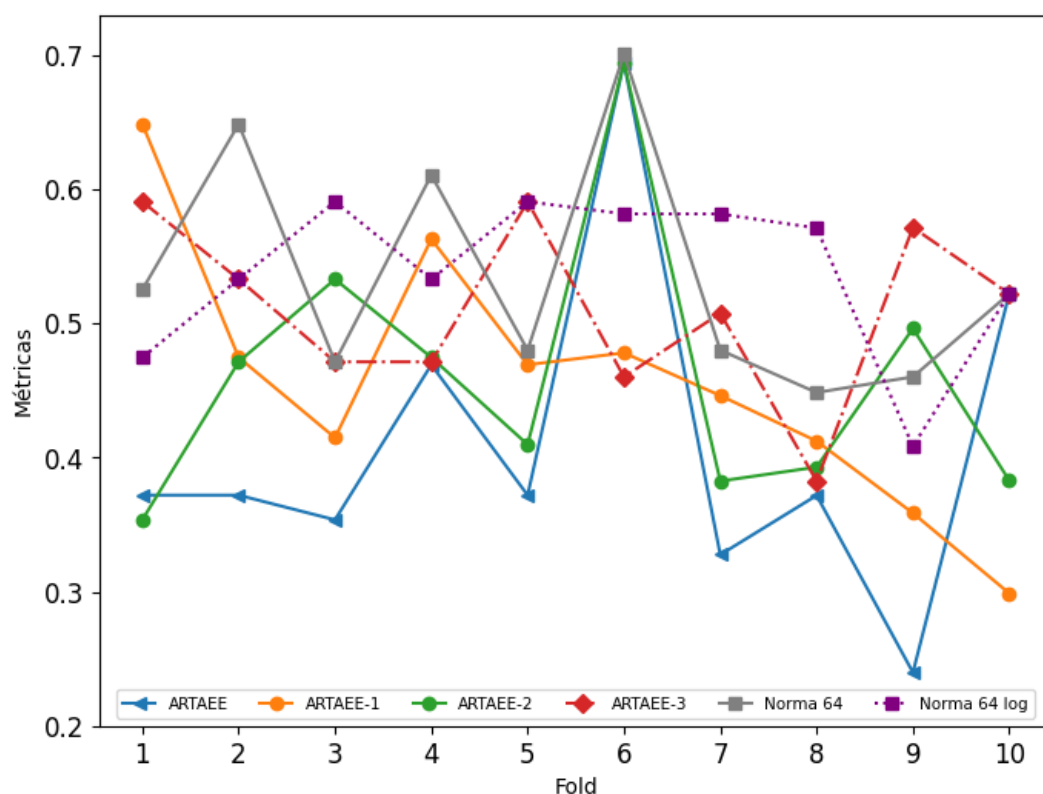
Mais uma vez é apresentado na Tabela 15 a sensibilidade obtida pelas redes para melhor visualização da média de desempenho obtida.

Tabela 15 – Comparação Sensibilidade redes.

Arquivo	ARTAEE	ARTAEE1	ARTAEE2	ARTAEE3	Norma 64	Norma 64 log
1	95,00%	90,00%	90,00%	90,00%	100,00%	90,00%
2	95,00%	90,00%	90,00%	90,00%	90,00%	90,00%
3	90,00%	100,00%	90,00%	80,00%	80,00%	90,00%
4	85,00%	70,00%	90,00%	100,00%	75,00%	90,00%
5	95,00%	85,00%	60,00%	90,00%	75,00%	90,00%
6	90,00%	70,00%	90,00%	75,00%	85,00%	95,00%
7	95,00%	90,00%	90,00%	85,00%	75,00%	95,00%
8	100,00%	55,00%	80,00%	80,00%	80,00%	90,00%
9	90,00%	65,00%	100,00%	85,00%	75,00%	70,00%
10	95,00%	100,00%	90,00%	95,00%	95,00%	95,00%
Média	93,00%	81,50%	87,00%	87,00%	83,00%	89,50%

No geral, a rede ARTAEE e a rede com norma-64 log obtiveram resultados mais estáveis, com exceção o conjunto de dados 9 para norma-64 log. Enquanto a rede ARTAEE-1 obteve as maiores variações dentre todas as redes.

A Figura 20 apresenta o Coeficiente de Correlação de *Matthews* obtido pelas redes neurais analisadas neste trabalho.

Figura 20 – Comparação do Coeficiente de Correlação de *Matthews*.

Fonte: Autor.

A Tabela 16 também apresenta os resultados do Coeficiente de Correlação de *Matthews* obtido por cada rede analisada neste trabalho, a fim de melhor exposição do desempenho médio obtido.

Tabela 16 – Comparação MCC redes.

Arquivo	ARTAE	ARTAE1	ARTAE2	ARTAE3	Norma 64	Norma 64 log
1	0,3720	0,6487	0,3536	0,5908	0,5252	0,4747
2	0,3720	0,4747	0,4714	0,5330	0,6486	0,5330
3	0,3536	0,4148	0,5330	0,4714	0,4714	0,5908
4	0,4692	0,5630	0,4747	0,4714	0,6106	0,5330
5	0,3720	0,4692	0,4097	0,5908	0,4801	0,5908
6	0,6945	0,4781	0,6945	0,4602	0,7009	0,5817
7	0,3276	0,4462	0,3825	0,5076	0,4801	0,5817
8	0,3718	0,4123	0,3929	0,3825	0,4485	0,5711
9	0,2398	0,3586	0,4963	0,5715	0,4602	0,4085
10	0,5221	0,2988	0,3929	0,5221	0,5221	0,5221
Média	0,4094	0,4564	0,4601	0,5101	0,5348	0,5387

Novamente, os resultados da rede Norma-64 log foram mais estáveis, obtendo os melhores (ou iguais) resultados nos conjuntos 3, 5, 7, 8 e 10, como pode ser observado.

A rede norma-64 obteve os maiores (ou iguais) MCC para os conjuntos 2, 4, 6 e 10, apresentando bons resultados, porém os resultados não são estáveis de forma geral.

A rede ARTAE e ARTAE-2 apresentou as maiores variações. Já a rede ARTAE-1 apresentou bom resultado para o conjunto de dados 1, porém não manteve bons resultados de forma geral para os conjuntos de dados restante.

O comparativo do *F-Score* obtido pelas redes é apresentado na Figura 21.

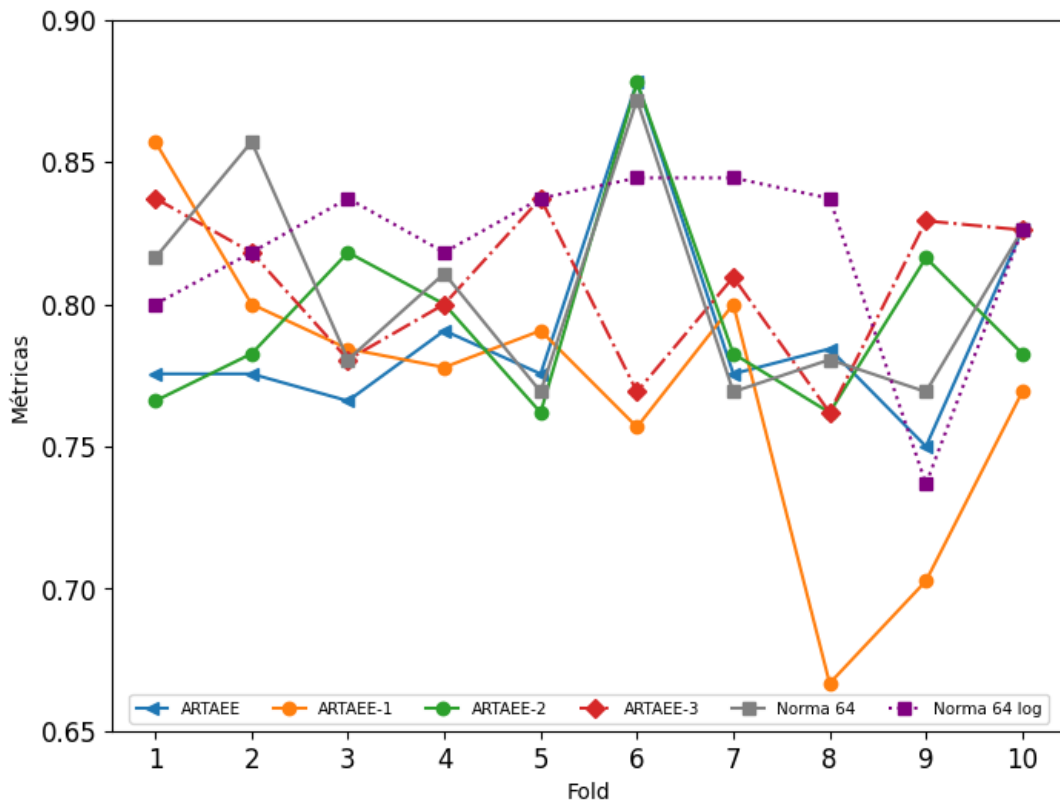


Figura 21 – Comparação do F-Score.

Fonte: Autor.

A Tabela 17 também apresenta o *F-Score* obtido pelas redes, porém com o objetivo de melhor visualização da média de desempenho obtido.

Tabela 17 – Comparação F-Score redes.

Arquivo	ACC	Sens.	Esp.	MCC	<i>F-Score</i>	
1	0,7755	0,8571	0,7660	0,8372	0,8163	0,8000
2	0,7755	0,8000	0,7826	0,8182	0,8571	0,8182
3	0,7660	0,7843	0,8182	0,7805	0,7805	0,8372
4	0,7907	0,7778	0,8000	0,8000	0,8108	0,8182
5	0,7755	0,7907	0,7619	0,8372	0,7692	0,8372
6	0,8780	0,7568	0,8780	0,7692	0,8718	0,8444
7	0,7755	0,8000	0,7826	0,8095	0,7692	0,8444
8	0,7843	0,6667	0,7619	0,7619	0,7805	0,8372
9	0,7500	0,7027	0,8163	0,8293	0,7692	0,7368
10	0,8261	0,7692	0,7826	0,8261	0,8261	0,8261
Média	0,7897	0,7705	0,7950	0,8069	0,8051	0,8200

Em relação ao *F-Score*, as redes Norma-64 e a rede Norma-64 log obtiveram um desempenho médio similar, porém a Norma-64 apresenta maiores variações, enquanto a Norma-64 log apresenta resultados mais estáveis nos conjuntos de dados de forma geral.

A rede ARTAEE-3 também apresentou resultados estáveis, mesmo que menores do que as redes Norma-64 e Norma-64 log.

A Tabela 18 apresenta o comparativo dos resultados obtidos neste trabalho com os resultados apresentados pela literatura. Para isso, foram utilizados como comparativo os trabalhos de McDermott e Forsyth (2016), e Abeyrathna et al. (2023).

Tabela 18 – Comparativo dos resultados obtidos com os presentes na literatura.

Autores	Método	ACC (%)	Sens. (%)	Esp. (%)	<i>F-Score</i>	MCC
McDermott e Forsyth (2016)	Benchmark	58,00	N.A.	N.A.	N.A.	N.A.
Abeyrathna et al. (2023)	ANN-1	61,20	N.A.	N.A.	0,671	N.A.
	ANN-2	59,40	N.A.	N.A.	0,652	N.A.
	ANN-3	60,20	N.A.	N.A.	0,656	N.A.
	DT	59,60	N.A.	N.A.	0,728	N.A.
	SVM	57,10	N.A.	N.A.	0,622	N.A.
	KNN	56,60	N.A.	N.A.	0,638	N.A.
	RF	60,70	N.A.	N.A.	0,729	N.A.
	XGBoost	63,50	N.A.	N.A.	0,656	N.A.
	EBM	62,90	N.A.	N.A.	0,710	N.A.
	ARTAE	71,03	93,00	40,71	0,790	0,4094
Este Trabalho	ARTAE-1	72,42	81,50	60,00	0,771	0,4560
Este Trabalho	ARTAE-2	73,63	87,00	76,76	0,795	0,4601
Este Trabalho	ARTAE-3	75,93	87,00	60,76	0,807	0,5101
Este Trabalho	Norma-64	76,88	83,00	68,48	0,8051	0,5348
Este Trabalho	Norma-64 log	77,39	89,50	60,67	0,8200	0,5387

N.A.: Não apresenta

Vale destacar que ao apresentar o MCC, pode-se concluir que foi utilizado um bom classificador neste trabalho, o que torna os resultados confiáveis. Caso contrário, apresentando apenas um bom *F-Score*, o classificador pode ser enviesado, ou seja, possuir a tendência de classificar os dados apenas como pacientes saudáveis ou apenas como pacientes doentes. Desse modo, ao comparar os resultados obtidos com a literatura, é possível notar que estes foram satisfatórios.

Para melhor visualização dos resultados obtidos neste trabalho com os apresentados na literatura, a Figura 22 apresenta os resultados relativos ao *F-Score*.

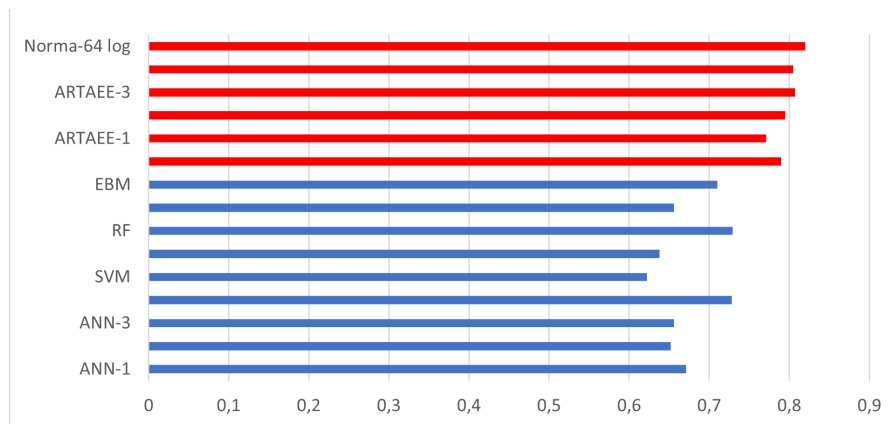


Figura 22 – Comparação resultados *F-Score* com os presentes na literatura.

Fonte: Autor.

Em vermelho segue os resultados obtidos pelas redes deste trabalho, enquanto em azul segue o comparativo com os resultados obtidos pela literatura.

A Figura 23 apresenta os resultados relativos à acurácia.

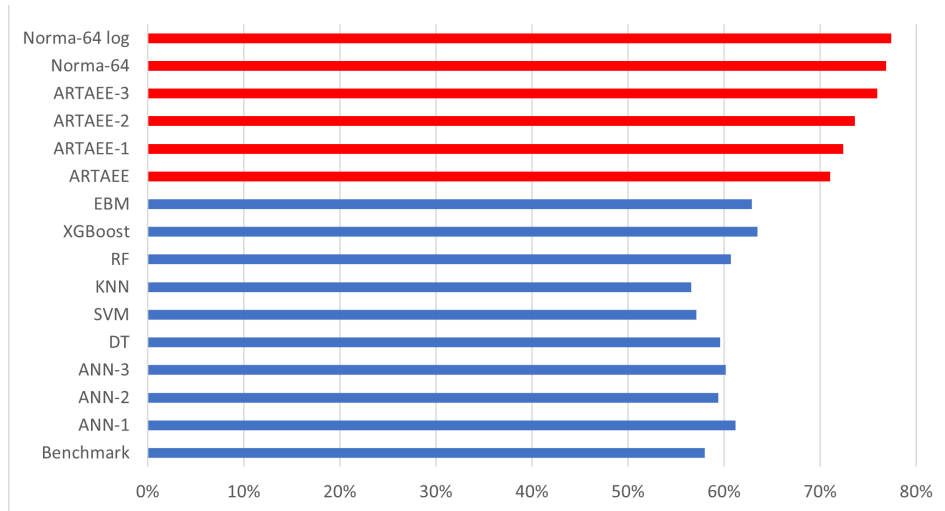


Figura 23 – Comparação resultados Acurácia com os presentes na literatura.

Fonte: Autor.

Como pode ser observado, a Acurácia e o F -Score foram superiores aos outros métodos utilizados pela literatura.

Por fim, a partir dos resultados obtidos neste trabalho são apresentadas as considerações finais.

5 *Considerações Finais*

Neste trabalho, os modelos de redes neurais Autoexpansíveis do tipo ART Euclidiana e p -ART foram utilizados para investigar o problema de classificação de doenças hepáticas e revelaram que são modelos eficientes para o problema proposto, cumprindo assim o objetivo de pesquisa. Em específico, a rede neural p -ART, com $p = 64$ confirmou ser o modelo mais eficaz neste contexto, pois alcançou melhores métricas em relação aos outros métodos utilizados como comparativo.

O Coeficiente de Correlação de *Matthews* foi uma importante métrica neste trabalho, pois com ele pode-se concluir que o classificador obteve melhor desempenho para o problema abordado, ao comparar com os outros classificadores utilizados pela literatura. Além disso, o *F-Score* também obteve melhores resultados.

A literatura apresenta uma acurácia máxima de 63,5%. Utilizando o mesmo banco de dados, este trabalho atingiu uma acurácia máxima de 77,39% com a rede ART Norma-64 log, mostrando que ela pode ser utilizada como sistema de apoio ao diagnóstico médico.

Com isso, mesmo que a sensibilidade não seja apresentada na literatura, pode-se afirmar que o nível de sensibilidade obtido neste trabalho foi superior, ou seja, o modelo de rede proposto é capaz de melhor identificar os indivíduos verdadeiramente doentes.

Este estudo desenvolveu uma revisão de Geometria Analítica, Álgebra Linear e Espaços Métricos. Tendo em vista que o objetivo era utilizar métricas não Euclidianas, foi fornecido o aperfeiçoamento de algumas disciplinas, tais como Análise Real. Além disso, também não são discutidas métricas no curso, proporcionando não só uma revisão de tópicos estudados na disciplina de Introdução aos Espaços Métricos, como também a ampliação destes.

Ainda, pôde ser observada a importância do uso de métricas não Euclidianas, que permite obter alguns resultados que não ocorrem quando a métrica é a Euclidiana. Logo, como estudos subsequentes, pode-se estender os resultados aqui estudados para espaços de dimensões maiores, como a norma do máximo.

Referências

- ABEYRATHNA, K. D. et al. A multi-step finite-state automaton for arbitrarily deterministic tsetlin machine learning. **Expert Systems**, Wiley Online Library, v. 40, n. 4, p. e12836, 2023.
- BOOK, D. L. **O Neurônio, Biológico e Matemático**. 2017. Disponível em: <<https://www.deeplearningbook.com.br/o-neuronio-biologico-e-matematico/.html>>. Acesso em: 07 fev 2023.
- BORRA, S.; CIACCIO, A. D. Measuring the prediction error. a comparison of cross-validation, bootstrap and covariance penalty methods. **Computational statistics & data analysis**, Elsevier, v. 54, n. 12, p. 2976–2989, 2010.
- BRAGA, A. P.; LUDERMIR, T. B.; CARVALHO, A. C. P. L. F. **Redes neurais artificiais: teoria e aplicações**. Rio de Janeiro: LTC, 2000.
- CARPENTER, G. A.; GROSSBERG, S. A massively parallel architecture for a self-organizing neural pattern recognition machine. **Computer vision, graphics, and image processing**, Maryland Heights, US, v. 37, n. 1, p. 54–115, 1987. DOI:<[https://doi.org/10.1016/S0734-189X\(87\)80014-2](https://doi.org/10.1016/S0734-189X(87)80014-2)>.
- CERVO, A. L.; BERVIAN, P. A. Metodologia científica. In: **Metodologia científica**. [S.l.: s.n.], 1996. p. xiv–209.
- DATA, O. W. in. **Causes of death**. 2023. Disponível em: <<https://ourworldindata.org/causes-of-death/.html>>. Acesso em: 07 fev 2023.
- DEV. **Overfitting and Underfitting in Machine Learning**. 2022. Disponível em: <https://dev.to/ml_82/overfitting-and-underfitting-in-machine-learning-i76.html>. Acesso em: 21 mar 2023.
- DUA, D.; GRAFF, C. **UCI Machine Learning Repository**. 2017. Disponível em: <<http://archive.ics.uci.edu/ml>>.
- FAUSETT, L. V. **Fundamentals of Neural Networks: Architectures, Algorithms and Applications**. [S.l.]: Pearson Education India, 1994.
- HAYKIN, S. **Redes Neurais: Princípios e Prática**. 2ª ed.. ed. Porto Alegre, RS: Bookman, 2001. ISBN 9788573075707.
- HOPFIELD, J. J. Neural networks and physical systems with emergent collective computational abilities. **Proceedings of the National Academy of Sciences**, v. 79, n. 8, p. 2554–2558, 1982.
- JAIN, A. K.; MAO, J.; MOHIUDDIN, K. M. Artificial neural networks: A tutorial. **Computer**, IEEE, v. 29, n. 3, p. 31–44, 1996.
- KOHAVI, R. et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: MONTREAL, CANADA. **Ijcai**. [S.l.], 1995. v. 14, n. 2, p. 1137–1145.

- KOHONEN, T. **Self-organizing maps**. [S.l.]: Springer Science & Business Media, 2012.
- MCCORMICK, P. A.; JALAN, R. Hepatic cirrhosis. **Sherlock's Diseases of the Liver and Biliary System**, Wiley Online Library, p. 107–126, 2018.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathematical Biophysics**, v. 5, p. 115–133, 1943. DOI:<<https://doi.org/10.1007/BF02478259>>.
- MCDERMOTT, J.; FORSYTH, R. S. Diagnosing a disorder in a classification benchmark. **Pattern Recognition Letters**, Elsevier, v. 73, p. 41–43, 2016.
- MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. **Sistemas inteligentes-Fundamentos e aplicações**, v. 1, n. 1, p. 32, 2003.
- MORENO, A. L. **Análise da estabilidade transitória via rede neural ART-ARTMAP Fuzzy euclidiana modificada com treinamento continuado**. 2010. 144 f. Tese (Doutorado em Engenharia Elétrica) — Faculdade de Engenharia, Universidade Estadual Paulista (Unesp), Ilha Solteira, 2010.
- MORENO, A. L. **Redes Neurais ART e ARTMAP com Treinamento Continuado: Uma apresentação detalhada dos principais tipos de redes neurais do tipo ART e ARTMAP**. Saarbrücken: Novas Edições Acadêmicas, 2016. ISBN 978-3330740051.
- NOGARE, D. **Performance de Machine Learning – Matriz de Confusão**. 2020. Disponível em: <<https://diegonogare.net/2020/04/performance-de-machine-learning-matriz-de-confusao/.html>>. Acesso em: 21 mar 2023.
- ROSENBLATT, F. **The perceptron, a perceiving and recognizing automaton Project Para**. [S.l.]: Cornell Aeronautical Laboratory, 1957.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **Nature**, v. 323, n. 6088, p. 533–536, Oct 1986. DOI:<<https://doi.org/10.1038/323533a0>>.
- SHAHID, N.; RAPPON, T.; BERTA, W. Applications of artificial neural networks in health care organizational decision-making: A scoping review. **PloS one**, Public Library of Science San Francisco, CA USA, v. 14, n. 2, p. e0212356, 2019.
- SILVA, H. T. da. **Fast ART-p: a norma-p numa rede neural do tipo fast ART**. 2019. 120 f. Dissertação (Mestrado Profissional em Matemática Aplicada e Computacional) — Instituto de Matemática, Estatística e Computação Científica, Universidade Estadual de Campinas, Campinas, 2019.
- SILVA, L. E. B. da; ELNABARAWY, I.; WUNSCH, D. C. A survey of adaptive resonance theory neural network models for engineering applications. **Neural Networks**, v. 120, p. 167–203, 2019. ISSN 0893-6080. DOI:<<https://doi.org/10.1016/j.neunet.2019.09.012>>.
- TRIPATHI, A.; SRIVASTAVA, G.; SINGH, K.; MAURYA, P. Review of unsupervised adaptive resonance theory. In: **2019 9th International Conference on Cloud Computing, Data Science & Engineering**. [S.l.: s.n.], 2019. p. 409–415.

WERBOS, P. Beyond regression: New tools for prediction and analysis in the behavioral sciences. **PhD thesis, Committee on Applied Mathematics, Harvard University, Cambridge, MA**, 1974.

YOUNG, B.; O'DOWD, G.; WOODFORD, P. **Wheater's functional histology E-Book: a text and colour atlas**. [S.l.]: Elsevier Health Sciences, 2013.