



# DIFERENCIAÇÃO E DISTINÇÃO DE FLORES: UM ESTUDO POR MEIO DE REDES NEURAIIS ARTIFICIAIS

**Lucas de Oliveira Hernandes<sup>1</sup>**

Instituto Federal de Educação, Ciência e Tecnologia de São Paulo – IFSP

**Flavia Milo dos Santos<sup>2</sup>**

Instituto Federal de Educação, Ciência e Tecnologia de São Paulo – IFSP

**Marco Aurélio Granero<sup>3</sup>**

Instituto Federal de Educação, Ciência e Tecnologia de São Paulo – IFSP

## Resumo

As técnicas de aprendizado de máquina e a utilização de redes neurais artificiais têm apresentado resultados promissores na tarefa de identificação e classificação de padrões. Neste contexto, o presente estudo propõe a classificação das espécies de plantas contidas no conjunto de dados *Iris*, utilizando como características o comprimento e largura das pétalas. Para esta tarefa, foram utilizados três classificadores: *Adaline*, *Perceptron* e MLP, que tiveram seu desempenho avaliado a fim de comparação de seus resultados. As métricas utilizadas foram a acurácia, precisão e sensibilidade. Antes de submeter o conjunto de dados aos classificadores, os dados foram descritos por meio de estatísticas descritivas para verificação da presença de *outliers* e comparação da escala, dois fatores que interferem no desempenho das redes neurais.

**Palavras-chave:** Redes Neurais; Aprendizado de Máquina; *Adaline*; *Perceptron*; *MLP*.

## 1. INTRODUÇÃO

O aprendizado de máquina conquistou todos os segmentos da indústria sendo parte central dos produtos de tecnologia atuais (Géron, 2021). Géron (2021) ainda expõe que um dos usos típicos destes algoritmos ocorre na resolução de problemas de classificação. Com isto, o aprendizado de máquina pode ser visto como um conjunto de técnicas ou algoritmos que permitem a um programa de computador, de maneira automatizada, extrair informações, identificar padrões e encontrar relações em grandes

---

<sup>1</sup> Graduando – IFSP. Aluno do curso de Licenciatura em Matemática do IFSP – Campus São Paulo, São Paulo, SP, Brasil. E-mail: [lucas.hernandes@aluno.ifsp.edu.br](mailto:lucas.hernandes@aluno.ifsp.edu.br).

<sup>2</sup> Doutora – USP. Professora do IFSP – Campus São Paulo, São Paulo, SP, Brasil. E-mail: [flaviamilo@ifsp.edu.br](mailto:flaviamilo@ifsp.edu.br).

<sup>3</sup> Doutor – UNICAMP. Professor do IFSP – Campus São Paulo, São Paulo, SP, Brasil. E-mail: [granero@ifsp.edu.br](mailto:granero@ifsp.edu.br).

quantidades de dados de maneira que um ser humano não conseguiria, simplesmente treinando e aprendendo a partir destes dados (Abu-Mostafa *et al.*, 2012).

Redes neurais artificiais são técnicas de aprendizado de máquina construídas baseadas nas redes neurais cerebrais (Géron, 2021). Silva (2016) expande esta definição pontuando que as redes neurais artificiais possuem a capacidade de aquisição e manutenção do conhecimento, além disso, são compostas por um conjunto de unidades de processamento, representadas por neurônios artificiais interligados. Haykin (2001) define dois pontos de convergência entre as redes cerebrais e as redes artificiais: as redes neurais artificiais adquirem conhecimento por meio de um processo de aprendizagem e o conhecimento adquirido é armazenado por meio de pesos sinápticos, que representam as forças de conexão entre neurônios.

As RNA's foram utilizadas em vários estudos, em tarefas diversas, como por exemplo, em Yan et al. (2006) redes *Multilayer Perceptron* (MLP) são utilizadas no contexto de diagnósticos de doenças de coração. Já em He et al. (2007) as RNA's são utilizadas para classificação de diferentes tipos de chá.

Este trabalho propõe classificar as espécies de plantas contidas no conjunto de dados *Iris* utilizando o comprimento e a largura das pétalas como variáveis de estudo. Nesta tarefa de classificação serão utilizados os classificadores: *Perceptron*, *Adaline* e MLP. O desempenho dos classificadores será avaliado por meio de métricas de desempenho a fim de comparar a performance do classificador.

O presente trabalho encontra-se estruturado em 5 seções. A seção 2 apresenta as RNA's formalmente, descrevendo seu funcionamento e como são treinadas. A seção 3 apresenta um breve histórico do conjunto de dados utilizados, o conjunto *Iris*, e uma análise descritiva destes dados. Na seção 4 tem-se a metodologia e os resultados obtidos na classificação dos conjuntos de dados. A conclusão é apresentada na seção 5.

## **2. REDES NEURAIAS ARTIFICIAIS – RNA's**

Nesta seção é apresentada uma breve descrição das RNA's utilizados no processo de classificação das plantas do conjunto de dados *Iris*.

### **2.1 Redes *Perceptron***

Um dos primeiros tipos de redes neurais a ser idealizado foi o *Perceptron*, um modelo simples de rede neural. Desenvolvido por Rosenblatt (1958), o *Perceptron* conta com apenas uma camada neural, que possui apenas um neurônio. Em relação à arquitetura

o *Perceptron* é uma rede *feedforward*, isto é, o fluxo de informações segue apenas um sentido, da camada de entrada à camada de saída (Silva, 2016).

Assim como a estrutura o funcionamento do *Perceptron* é simples. Uma rede *Perceptron* funciona a partir da ponderação dos sinais de entrada  $x_i$  pelos pesos  $w_i$ , essa ponderação indicará qual a importância daquela entrada para a composição do sinal de ativação da rede. Esta ponderação ocorre pela multiplicação da  $i$ -ésima entrada pelo  $i$ -ésimo peso. As entradas ponderadas serão somadas, formando o potencial de ativação  $u$ . A este potencial será aplicada uma função de ativação  $g(u)$ , esta função será a saída obtida pelo *Perceptron* (Silva, 2016). Matematicamente tem-se:

$$u = \sum_{i=0}^n w_i x_i = \langle w, x \rangle \quad (1)$$

$$y = g(u) \quad (2)$$

A partir de uma inicialização com valores aleatórios, o ajuste dos pesos sinápticos e do limiar ocorre por meio de uma comparação de saídas obtidas em uma etapa do treinamento com a saída desejada (Silva, 2016). A saída obtida é então comparada com a saída desejada e, caso a diferença entre a saída obtida e a desejada não seja 0, os pesos são reajustados a fim de reduzir esta diferença através da equação:

$$w_i^{atual} = w_i^{anterior} + N \cdot (d^k - y) \cdot x_i^k \quad (3)$$

O processo é repetido iterativamente até que o número máximo de iterações seja atingido ou até que o erro entre as saídas obtidas e as saídas desejadas seja nulo. O vetor obtido ao final deste processo é considerado o vetor solução, ou seja, os pesos de cada entrada sináptica estão adequados.

## 2.2 Rede *Adaline*

No mesmo período de desenvolvimento do *Perceptron*, Widrow idealizou um modelo denominado *Adaline* (do inglês: *Adaptative Linear Element*) (Kovács, 1996). Silva (2016) afirma que apesar de simples, as redes *Adaline* contribuíram com o desenvolvimento da regra Delta de aprendizado e com as primeiras aplicações de redes neurais em problemas industriais e de processamento de sinais. Assim como no modelo de Rosenblatt o modelo de Widrow apresenta apenas uma camada neural com um único neurônio, esta rede também pode ser classificada como uma rede *feedforward*.

O processo de funcionamento da rede *Adaline* é o mesmo do *Perceptron*, a diferença entre as duas redes está no processo de treinamento e ajuste dos pesos, uma vez que a rede *Adaline* possui um bloco associador, que calcula o erro entre a saída desejada e os potenciais de ativação, neste caso os pesos são ajustados a partir da regra Delta, dada pela equação:

$$w_i^{atual} = w_i^{anterior} + N \cdot \sum_{k=1}^p (d^k - u) \cdot x_i^k \quad (4)$$

Neste sentido, assim como no *Perceptron* valores aleatórios são atribuídos aos pesos sinápticos, os quais serão ponderados com os sinais de entrada e comparados com a saída desejada, caso o erro não seja nulo, os pesos serão ajustados por meio da regra Delta iterativamente até um critério de parada ser atingido. Este critério de parada é atingido quando o número de iterações máximo é obtido ou quando a diferença entre o erro quadrático médio calculado com os pesos atualizados em uma iteração e o erro quadrático médio da iteração anterior é menor do que um valor considerado (Silva, 2016).

O erro quadrático médio está definido na equação 5, assim como o critério de parada está definido na equação 6.

$$E_{qm} = \frac{1}{p} \sum_{k=1}^p (d^k - u)^2 \quad (5)$$

$$|Eqm(w_{atual}) - Eqm(w_{anterior})| \leq \varepsilon \quad (6)$$

onde  $\varepsilon$  é um número real.

Uma das características que difere as redes *Adaline* das redes *Perceptron* é a solução encontrada pela rede. Nas redes *Adaline* os vetores de solução são únicos, enquanto no *Perceptron* as soluções variam em relação aos vetores iniciais.

### 2.3 Redes *Perceptron* de múltiplas camadas - MLP

As redes *Perceptron* de múltiplas camadas (ou *Multilayer Perceptron* - MLP) são compostas por pelo menos uma camada intermediária, chamada de camada oculta, de neurônios entre as camadas de entrada e saída (Silva, 2016). O interesse pelas redes de múltiplas camadas se intensificou a partir do fim dos anos 1980, com o desenvolvimento do algoritmo *backpropagation*. Um avanço gerado pelas redes *Multilayer Perceptron* foi a possibilidade de resolver o problema do “ou exclusivo” e outros problemas não lineares (Ludwig Jr., 2007).

O processo de treinamento de uma rede MLP ocorre por meio do algoritmo *backpropagation*, neste caso os sinais de entrada são classificados, e a correção dos pesos ocorre da camada de saída até a camada de entrada, ou seja, no sentido oposto ao de propagação do sinal de entrada (Silva, 2016). A propagação do sinal ocorre como nos classificadores lineares, porém, no caso da rede MLP as saídas obtidas por um neurônio de uma camada intermediária serão consideradas os sinais de entrada dos neurônios da próxima camada de neurônios.

### 3. CONJUNTO DE DADOS IRIS

O conjunto de dados Iris é um dos mais conhecidos e utilizados conjuntos de dados existentes, tendo sua primeira aparição com um trabalho publicado por Ronald Fisher em 1936 (Unwin, 2021). Já a coleta deste conjunto de dados foi feita pelo botânico Edgar Anderson em 1935.

O conjunto de dados Iris apresenta dados de três espécies diferentes do gênero Iris: *Iris-setosa*, *Iris-virginica* e *Iris-versicolor*. Inicialmente Anderson coletou medidas apenas das espécies *setosa* e *versicolor*, posteriormente Fisher adicionou dados sobre a *Iris-virginica*.

Contando com 150 instâncias no total, 50 instâncias para cada espécie, há quatro características no conjunto de dados Iris: comprimento e largura das pétalas e sépalas, com todos os dados em centímetros. Este conjunto de dados pode ser encontrado em Pedregosa et al. (2011), por meio da biblioteca *scikit-learn* da linguagem de programação *Python*.

#### 3.1 Análise estatística descritiva

Um passo importante ao analisar um conjunto de dados é fazer uma análise estatística descritiva dos dados. Esta análise é importante para verificar a existência de *outliers*, que podem influenciar no desempenho dos algoritmos de redes neurais, como também verificar sua variabilidade, o que também afeta o desempenho das redes. Em geral há uma redução de desempenho em algoritmos de aprendizagem de máquina quando os valores numéricos estão em escalas diferentes (Géron, 2021), portanto é necessário analisar a distribuição dos dados.

Como a tarefa de classificação ocorrerá para as pétalas das flores, os dados das sépalas não serão considerados. A Tabela 1 apresenta o resumo estatístico descritivo do conjunto de dados Iris.

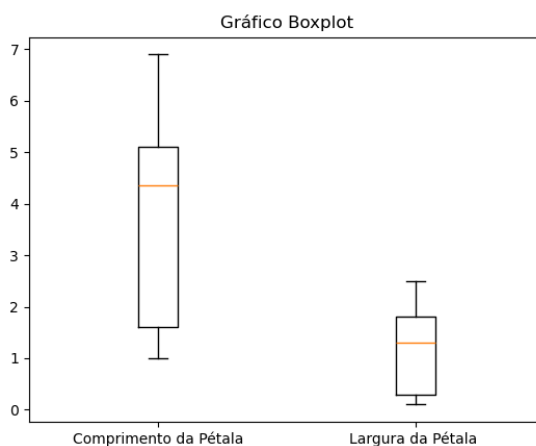
**Tabela 1:** Resumo estatístico das características das pétalas do conjunto de dados Iris.

	<b>Média</b>	<b>Mediana</b>	<b>Desvio-padrão</b>	<b>Quartis</b>	<b>Moda</b>
<b>Comprimento</b>	3,759	4,35	1,764	[1,57; 4,35; 5,1]	1,5
<b>Largura</b>	1,199	1,3	0,763	[0,3; 1,3; 1,8]	0,2

**Fonte:** Autor, 2024

Nota-se que o comprimento das pétalas apresenta maiores média e mediana, assim como o desvio-padrão, indicando uma maior dispersão dos dados.

Os gráficos *boxplot* podem ser utilizados para descrever algumas características mais específicas de um conjunto de dados, entre elas o centro, a variabilidade, extensão e a identificação de *outliers*, em geral gráficos *boxplot* comparativos são uma forma eficiente de comparar variáveis em um conjunto de dados, o que será útil para analisar as escalas dos dados (Devore, 2018).



**Figura 1** – Gráfico boxplot comparativo entre comprimento e largura das pétalas.

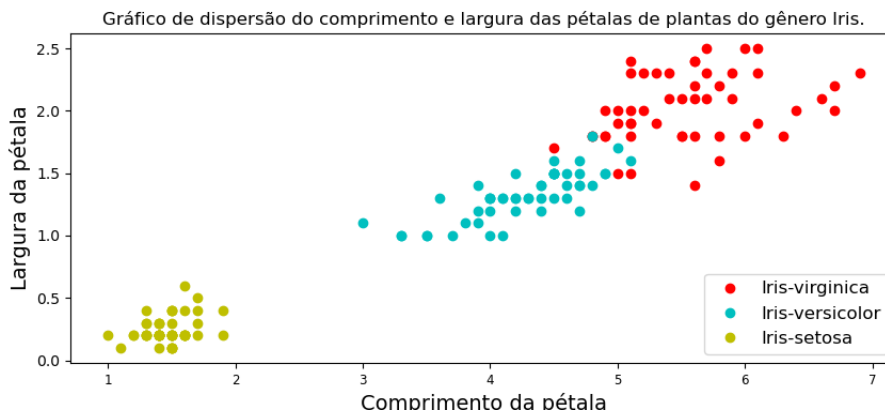
**Fonte:** Autor, 2024

Pode-se notar na Figura 1 que não há *outliers* nos dados das pétalas, neste caso não haverá uma interferência no treinamento das redes, nota-se também que os comprimentos estão mais dispersos do que as larguras. Ao analisar as estatísticas descritivas e a comparação das variáveis com o *boxplot* comparativo é possível perceber que a distribuição dos dados está em uma mesma amplitude de escala, neste caso não será necessário um tratamento de normalização de dados ou padronização.

### 3.2 Gráficos de dispersão das características agrupadas duas a duas

A Figura 2 apresenta o gráfico de dispersão da largura e do comprimento das pétalas da *Iris*. Nele é possível perceber que os dados da espécie *Iris-setosa* estão visualmente afastados dos outros dados, o que pode indicar uma possível separabilidade para esta espécie.

Observando a espécie *virginica* nota-se um menor afastamento dos dados. Já os dados da espécie *versicolor* estão localizados em uma faixa intermediária, entre as outras espécies, isso pode indicar um menor desempenho dos classificadores e a necessidade de um classificador não linear.



**Figura 2** – Gráfico de dispersão dos dados da pétala em relação ao comprimento e largura.

Fonte: Autor, 2024

#### 4. METODOLOGIA E RESULTADOS OBTIDOS

O estudo dos classificadores será realizado utilizando o comprimento e a largura das pétalas, tomando como referência uma das espécies em comparação às demais, ou seja, *Iris-setosa* em comparação com *virginica* e *versicolor*, *Iris-virginica* em comparação com *versicolor* e *setosa* e *Iris-versicolor* em comparação com *setosa* e *virginica*.

Serão utilizados os três classificadores discutidos anteriormente, as redes *Adaline*, *Perceptron* e MLP. Tanto o *Perceptron*, quanto a *Adaline* foram escolhidos por serem classificadores de fácil implementação computacional e apresentam um bom desempenho em classificar dados linearmente separáveis, contudo possuem um baixo desempenho em dados não linearmente separáveis, por isso a escolha da comparação com a rede MLP (Géron, 2021).

Para o treinamento dos classificadores, o conjunto de dados será dividido em dois, um conjunto de treinamento e um conjunto de teste. O conjunto de treinamento será o conjunto utilizado para ajustar os pesos dos classificadores, tendo um tamanho de 70% do conjunto de dados total, já o conjunto de teste será formado pelos 30% dos dados restantes e será utilizado para calcular as métricas de desempenho e avaliar assim a performance do classificador. A análise estatística mostra que os dados estão em uma escala parecida, deste modo não será necessário nenhum tratamento anterior à classificação.

A fim de comparar o desempenho dos classificadores lineares serão utilizados os mesmos parâmetros de ajuste da rede, como função de ativação, número de épocas e taxa de aprendizado. A escolha destes parâmetros e dos parâmetros da rede MLP foi feita empiricamente com o objetivo de aumentar o desempenho de classificação para as três espécies. A Tabela 2 apresenta os parâmetros utilizados para as três redes.

**Tabela 2:** Parâmetros utilizados para cada classificador.

	TAXA DE APRENDIZAGEM	CRITÉRIO DE PARADA	ÉPOCAS MÁXIMAS	FUNÇÃO DE ATIVAÇÃO	CAMADAS NEURAIS	NEURÔNIOS POR CAMADAS
ADALINE	0,0001	0,00001	1000	Função sinal	1	1
PERCEPTRON	0,0001	0, 00001	1000	Função sinal	1	1
MLP	0,001	0,0001	200	Tangente hiperbólica	2	200

Fonte: Autor, 2024

#### 4.1 Métricas de desempenho.

Para avaliar o desempenho das redes foram escolhidas três métricas: a precisão, a sensibilidade e a acurácia. Para calcular estas métricas serão considerados quatro dados: a quantidade de verdadeiros negativos ( $vn$ ), classificações corretas de instâncias negativas; classificações incorretas de instâncias negativas, ou falsos negativos ( $fn$ ); classificações incorretas de instâncias positivas, os falsos positivos ( $fp$ ); e classificações corretas positivas, ou seja, os verdadeiros positivos ( $vp$ ).

A acurácia pode ser definida como a razão entre os acertos totais e o número total de classificações feitas. Neste caso tem-se uma razão entre a soma dos verdadeiros positivos com os verdadeiros negativos e o número total de instâncias, ou seja,

$$acurácia = \frac{vp + vn}{\text{número de instâncias}} \quad (9)$$

Géron (2021) define a precisão como a razão entre os verdadeiros positivos e a soma entre os verdadeiros positivos e falsos positivos, a precisão também pode ser entendida como a acurácia das instâncias positivas. Em geral uma precisão maior significa uma tendência maior a classificações positivas, o que pode gerar um número maior de falsos positivos. Em linguagem matemática, a precisão é definida por:

$$precisão = \frac{vp}{vp + fp} \quad (10)$$



A sensibilidade (ou revocação) pode ser entendida como a razão entre os verdadeiros positivos e a soma entre os verdadeiros positivos e falsos negativos (Géron, 2021). Em linguagem matemática:

$$sensibilidade = \frac{vp}{vp + fn} \quad (11)$$

#### 4.2 Resultados obtidos – Alvo: *Iris-setosa*.

Nesta primeira classificação, a espécie *setosa* foi considerada o alvo da classificação, isto é, a classe positiva, já as outras duas espécies, *virginica* e *versicolor*, foram consideradas as classes negativas.

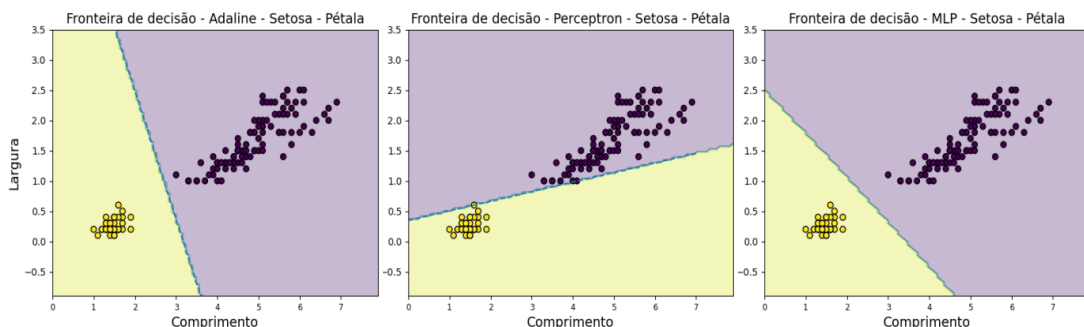
A Tabela 3 apresenta os dados de desempenho das redes testadas ao classificar as pétalas da espécie *setosa*. Nota-se que todos os classificadores conseguiram classificar os dados, obtendo desempenho máximo das métricas utilizadas. Além disso, o *Perceptron* e a *Adaline* conseguiram uma acurácia máxima, indicando que o conjunto de dados é linearmente separável.

**Tabela 3** – Desempenho dos classificadores para classificação das pétalas da espécie *Iris-setosa*.

	Acurácia	Precisão	Sensibilidade
<i>Adaline</i>	1,0	1,0	1,0
<i>Perceptron</i>	1,0	1,0	1,0
<b>MLP</b>	1,0	1,0	1,0

Fonte: Autor, 2024

A Figura 3 apresenta as fronteiras obtidas pelos classificadores treinados. Analisando estas fronteiras nota-se uma semelhança entre a fronteira obtida pela rede *Adaline* e pela rede MLP, que aparenta estar em uma porção intermediária entre o alvo (a espécie *Iris-setosa*) e as classes negativas (*Iris-virginica* e *Iris-versicolor*), já para a rede *Perceptron* há uma diferença, a reta que divide as duas classes aparenta tangenciar as duas classes do conjunto de dados, em geral nota-se que a rede *Perceptron* teria uma dificuldade maior em lidar com dados próximos à separatriz entre os dados.



**Figura 3** – Fronteiras de decisão das redes *Adaline*, *Perceptron* e MLP, com alvo *setosa*.

Fonte: Autor, 2024

### 4.3 Resultados obtidos – Alvo: *Iris-virginica*.

Considerado agora o foco na espécie *virginica*, como classe positiva e, as outras espécies como as classes negativas, têm-se na Tabela 4 os dados de desempenho obtidos.

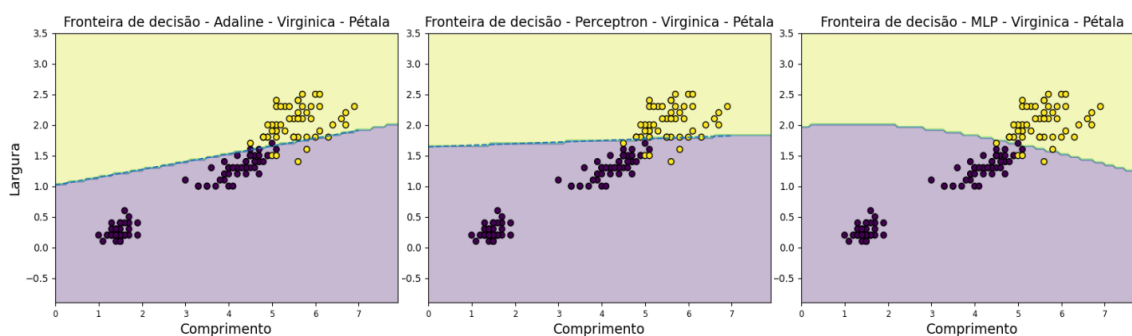
**Tabela 4** – Desempenho dos classificadores para classificação das pétalas da espécie *Iris-virginica*.

	Acurácia	Precisão	Sensibilidade
<i>Adaline</i>	0,8889	0,9333	0,7778
<i>Perceptron</i>	0,8889	1,0000	0,7222
MLP	0,8889	0,9333	0,7778

Fonte: Autor, 2024

Da Tabela 4 é possível notar que a acurácia foi a mesma para todas as redes. As redes *Adaline* e MLP obtiveram o mesmo desempenho em relação às métricas utilizadas. Já o *Perceptron* obteve a mesma acurácia dos outros classificadores, porém a precisão foi mais alta, ou seja, apresenta menos falsos positivos, classificando corretamente os casos positivos como tal. Ele também apresentou uma sensibilidade mais baixa, ou seja, um número maior de falsos negativos.

A Figura 4 apresenta as fronteiras de decisão com o alvo sendo a espécie *virginica*. Nota-se que as retas obtidas pelos classificadores lineares possuem uma inclinação parecida, mostrando assim o porquê de o desempenho das redes ser parecido. A fronteira de decisão da rede MLP apresenta o comportamento não linear deste classificador.



**Figura 4** – Fronteiras de decisão das redes *Adaline*, *Perceptron* e MLP, com alvo *virginica*.

Fonte: Autor, 2024

### 4.4 Resultados obtidos – alvo: *Iris-versicolor*.

A espécie *versicolor* está localizada entre as duas outras espécies, o que torna este conjunto de dados não linearmente separável. Espera-se, portanto, um menor desempenho dos classificadores lineares.

A Tabela 5 apresenta as métricas de desempenho obtidas na classificação da espécie *versicolor*.

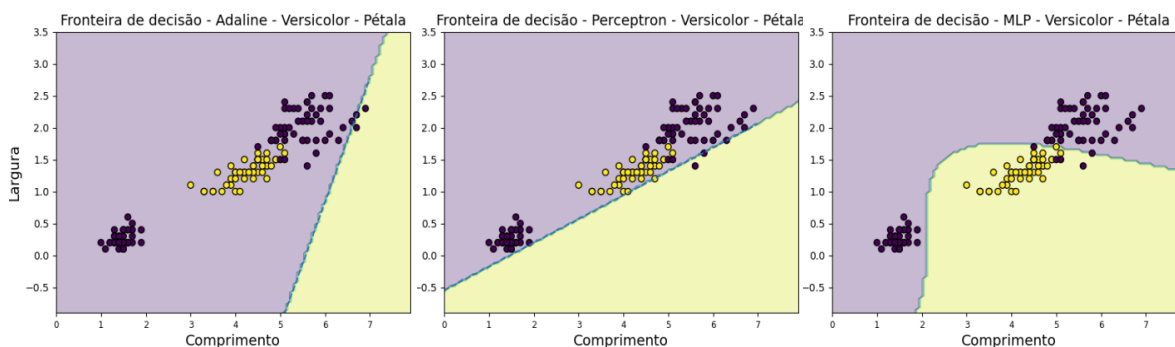
**Tabela 5** – Desempenho dos classificadores para classificação das pétalas da espécie *Iris-versicolor*.

	Acurácia	Precisão	Sensibilidade
<i>Adaline</i>	0,4000	0,0667	0,0714
<i>Perceptron</i>	0,4000	0,0000	0,0000
MLP	0,8889	0,7647	0,9286

Fonte: Autor, 2024

Como esperado o desempenho dos classificadores lineares está muito abaixo do desempenho da rede MLP, a acurácia dos dois é igual, porém a precisão e a sensibilidade do *Perceptron* são nulas, o que indica a não classificação de verdadeiros positivos. No caso da rede *Adaline* nota-se que a precisão e a sensibilidade não são nulas, porém inferiores a 10% de acerto.

A Figura 5 evidencia esse comportamento semelhante, uma vez que as fronteiras de decisão dessas redes tendem a classificar todas as instâncias como negativas. No caso da rede MLP, há um contorno não linear, obtido pela presença de mais uma camada neural e mais neurônios.



**Figura 5** – Fronteiras de decisão das redes *Adaline*, *Perceptron* e MLP, com alvo *versicolor*.

Fonte: Autor, 2024

## 5. CONCLUSÃO

O presente trabalho apresentou o desempenho de três diferentes redes neurais no processo de classificação do conjunto de dados conhecido como *Iris* de modo a explorar as potencialidades de classificadores linearmente e não-linearmente separáveis. Ambas as redes apresentaram desempenho semelhante no caso dos dados serem linearmente separáveis, porém as redes *Perceptron* e *Adaline* realizam esta tarefa com uma implementação computacional mais simples do que a rede MLP. No caso dos dados não serem linearmente separáveis, classificação da *versicolor* em relação às espécies *setosa* e *virginica*, a rede MLP apresentou um bom desempenho, obtendo acurácia, precisão e sensibilidade acima de 74%, valores que podem ser melhorados aumentando-se o número de neurônios da camada escondida. Entretanto, o desempenho das redes *Perceptron* e

*Adaline* mostrou-se inadequado para a tarefa de classificação proposta. Este resultado já era esperado, tendo-se em vista que estes classificadores são eficientes quando os dados são linearmente separáveis, mas não o são no caso de dados não-linearmente separáveis. O trabalho destacou ainda a importância em se conhecer a variabilidade do conjunto de dados em estudo a fim de evitar problemas com escalas e *outliers* que podem comprometer o estudo em desenvolvimento. O conhecimento da variabilidade do conjunto de dados serve ainda como ponto de partida para escolha da rede neural adequada bem como a escolha de seus parâmetros para posterior implementação computacional. Futuros trabalhos podem analisar qual a relação entre o desempenho das redes e as funções de ativação utilizadas, e/ou o aumento do número de neurônios na camada escondida.

## BIBLIOGRAFIA

ABU-MOSTAFA, Yaser S.; MAGDON-ISMAL, Malik; LIN, Hsuan-Tien. **Learning From Data: A short course**. New York, NY, USA: AMLBook, 2012. 216 p. ISBN 978-1-60049-006-4.

DEVORE, J. L. **Probabilidade e estatística para engenharias e ciências**. 9. ed. São Paulo: Cengage, 2018.

GÉRON, A. **Mãos à Obra: Aprendizado de Máquina com Scikit-Learn, Keras & TensorFlow: Conceitos, Ferramentas e técnicas para a construção de sistemas inteligentes**. 2. ed. São Paulo: Alta Books, 2021.

HAYKIN, S. **Redes Neurais: Princípios e prática**. 2. ed. Porto Alegre: Bookman, 2001.  
He, Y.; Li, X; Deng, X. Discrimination of varieties of tea using near infrared spectroscopy by principal component analysis and BP model, **Journal of Food Engineering**, v.79, 2007, <https://doi.org/10.1016/j.jfoodeng.2006.04.042>

KOVÁCS, Z. L. **Redes Neurais Artificiais: fundamentos e aplicações: um texto básico**. 2. ed. São Paulo: Collegium Cognition, 1996.

LUDWIG JR., O; MONTGOMERY, E. **Redes Neurais: Fundamentos e Aplicações com Programas em C**. Rio de Janeiro: Ciência Moderna, 2007.

PEDREGOSA, F. et al. Scikit-learn: Machine Learning in Python, **JMLR**, v. 12, 2011

ROSENBLATT, Frank. The perceptron: a probabilistic model for information storage and organization in the brain. **Psychological Review**, v.65, n.6, 1958.  
<https://doi.org/10.1037/h0042519>

SILVA, I. N. da.; SPATTI, D. H.; FLAUZINO, R. A. **Redes Neurais Artificiais para engenharia e ciências aplicadas: fundamentos teóricos e aspectos práticos**. 2. ed. São Paulo: Artliber, 2016.

Unwin, A.; Kim, K. The iris data set: In search of the source of virginica. **Significance**, v.18, 2021, <https://doi.org/10.1111/1740-9713.01589>

Yan, H.; Jiang, Y.; Zheng, J.; Peng, C.; Li, Q. A multilayer perceptron-based medical decision support system for heart disease diagnosis, **Expert Systems with Applications**, v. 30, 2006, <https://doi.org/10.1016/j.eswa.2005.07.022>