

AVALIAÇÃO DE ALGORITMOS DE MACHINE LEARNING NA CLASSIFICAÇÃO DE ANTRACNOSE EM FOLHAS DE MANGA

José Augusto dos Santos Silva ¹
Matheus Morandi Cabral Domingos ¹
João Ricardo Favan ²
Eloíza Martins Primo Capeloci ²

RESUMO

A antracnose, doença causada pelo fungo *Colletotrichum gloeosporioides*, representa um dos maiores problemas encontrados pelos produtores de manga (*Mangifera indica* L.) do Brasil. Apesar do aumento constante da relevância dessa cultura no país e dos recordes sucessivos nas exportações do fruto, o cultivo de manga pode atingir valores de produção e faturamento ainda mais elevados com ações preventivas contra a antracnose. As perdas econômicas por doenças e fungos geralmente atingem 30 a 50%. Com os avanços das tecnologias e técnicas de *machine learning* para a detecção precoce de doenças foliares, o presente trabalho tem como objetivo analisar a performance dos seguintes modelos: *Support Vector Machine* (SVM), *Random Forest*, *Logistic Regression* e *Neural Network*, visando a capacidade de classificação das folhas de manga, utilizando validação cruzada em conjunto com métricas como a área abaixo da curva R.O.C. (AUC) e a acurácia (CA), através do *software* de análise de dados e *machine learning Orange Data Mining*. Os modelos *Logistic Regression*, *Neural Network* e SVM apresentaram os resultados mais elevados, com valores de AUC iguais a 1 e CA de 99%, o presente trabalho conclui que esses modelos podem ser utilizados para auxiliar na classificação precoce de antracnose em folhas de manga, com isso garantindo uma tomada de decisão mais rápida e eficaz ao produtor.

Palavras-chave: Detecção Precoce. *Machine Learning*. Modelos Preditivos.

INTRODUÇÃO

A manga (*Mangifera indica* L.) é uma das frutas mais populares em muitos continentes, especialmente na Ásia, América Central e do Sul e África. A produção internacional de manga está concentrada na Ásia. A Índia é o maior produtor, com uma produção média anual de 15 milhões de toneladas. Atualmente, as mangas são cultivadas em mais de 100 países, com a produção anual em mais de 65 deles ultrapassando 1.000 toneladas. A demanda mundial por mangas tem crescido, especialmente em países temperados, onde a fruta está rapidamente ganhando popularidade (Mitra, 2016).

¹ Discentes em Big Data no Agronegócio na FATEC Pompéia, Pompéia-SP, joseaugustosilva834@gmail.com; matheus.morandi.cd@gmail.com

² Docentes do curso Big Data no Agronegócio, FATEC Pompéia, Pompéia-SP. joao.favan@fatec.sp.gov.br; eloiza.capeloci@fatec.sp.gov.br

O Brasil se encontra na sexta posição dos maiores produtores de manga do mundo e, segundo a Embrapa, no ano de 2023, a exportação do fruto alcançou a marca histórica de aproximadamente R\$1,79 bilhões em faturamento, ultrapassando o valor alcançado em 2021 de aproximadamente R\$1,41 bilhões¹, que era até então o maior faturamento obtido em um ano (Embrapa, 2024). Os valores de 2024 se mostram superiores aos dos anos anteriores, já que nos dois primeiros meses do ano tivemos um aumento de 45,6% na quantidade de manga exportada se comparado aos meses de janeiro e fevereiro de 2023. O Ministério da Agricultura e Pecuária (2023), apontou que, dentre os *commodities* com maior aumento de produção até 2033, está a manga, com um aumento de 22,9% no período de 2023 até 2033.

Um dos maiores problemas encontrados pelos produtores no cultivo da manga é a antracnose, doença ocasionada pelo fungo patogênico *Colletotrichum gloeosporioides* que é a principal doença em regiões de alta intensidade de chuvas, além de afetar vários órgãos da planta, como as folhas e os frutos (Flori; Resende; Borges, 2020). Quando o órgão afetado é a folha da mangueira, ela apresenta manchas marrons que podem variar de 1 até 10 mm de diâmetro, podendo ainda surgir em diferentes regiões da folha, como nos cantos, centro e nos dois lados da folhagem (Batista et al, 2016). De acordo com Neto, Peixoto e Souza (2023) as perdas econômicas ocasionadas por doenças e fungos geralmente atingem 30 a 50% e, em algumas ocasiões, as podridões podem chegar a 100% de incidência, ou seja, levar à perda total dos frutos colhidos.

Com a evolução no âmbito da tecnologia, uma das ferramentas que podem auxiliar na identificação de doenças em lavouras é o aprendizado de máquinas, que de acordo com Weber (2022), pode agilizar o tratamento, minimizando a quantidade de recursos e agrotóxicos utilizados, além de diminuir impactos de redução da produção agrícola.

Com base nos dados acima, observa-se que os profissionais da área de tecnologia podem contribuir intensamente com o setor agrícola, apresentando e desenvolvendo soluções que potencializam a produtividade e que identifiquem precocemente a infestação de pragas e doenças nas lavouras, com mais efetividade e rapidez. Soluções essas que vão desde análise de dados até a criação de ferramentas que auxiliem o agricultor a monitorar e prever futuros problemas que

¹ Foi considerado para conversão dos valores a taxa de câmbio de R\$5,69 por dólar de acordo com a cotação do Banco Central do Brasil em 27 de outubro de 2024.

possam afetá-lo.

Segundo Cerri e Carvalho (2017), o Aprendizado de Máquina (*Machine Learning*) é uma área da Inteligência Artificial voltada para o desenvolvimento de programas capazes de aprender a realizar determinadas tarefas a partir da própria experiência. Essa área multidisciplinar, que envolve inteligência artificial, estatística, teoria da informação e outros campos, permite que sistemas utilizem dados para melhorar seu desempenho em atividades como classificação, agrupamento de dados e previsão de séries temporais.

De acordo com Cardoso (2021) dentro do âmbito de aprendizado de máquinas há duas maneiras a qual os algoritmos podem aprender, sendo elas o aprendizado supervisionado e o aprendizado não supervisionado. A primeira maneira caracteriza-se pela base de dados possuir dados rotulados, ou seja, dados com saídas conhecidas. Já no método não supervisionado, não há a presença de rótulos, dessa forma os algoritmos identificam relações entre os atributos para formar sua base de conhecimento.

Atualmente, o setor agrícola tem se beneficiado cada vez mais do uso de tecnologias avançadas, como ferramentas de processamento de imagens e aprendizado de máquina, para otimizar o diagnóstico de doenças foliares e melhorar a eficiência produtiva (Santos et al., 2020). Entre essas ferramentas, algumas se destacam pela capacidade de processar grandes volumes de dados de maneira automatizada e precisa.

Neste trabalho, foram utilizadas algumas dessas tecnologias, que se destacam por suas funcionalidades específicas. Entre os algoritmos de aprendizado de máquina utilizados, o *Support Vector Machine* (SVM) é um algoritmo de classificação que utiliza um hiperplano que melhor separa os conjuntos de dados (Peixoto, 2016). *Random Forest* é o algoritmo que possui um conjunto de árvores de decisão onde cada uma realiza uma previsão (voto), sendo a classificação final a mais votada (Cardoso, 2021). O algoritmo *Neural Network* é um modelo baseado em neurônios biológicos, que utiliza modelos matemáticos para simular o funcionamento do cérebro (Disconzi, 2018). Por fim, *Logistic Regression*, método estatístico que modela a probabilidade de um evento binário em função de variáveis independentes, assim como destacado por Fernandes (2021).

A ferramenta *Pillow* (Clark et al., 2024) é uma biblioteca de processamento de imagens amplamente utilizada na linguagem de programação *Python*, ela deriva de

outra biblioteca chamada *PIL* (*Python Imaging Library*). Entre suas principais funcionalidades a *Pillow* oferece opções de redimensionamento de imagens, corte, rotação, conversão entre formatos, além de várias operações de filtragem e ajuste de cores.

A arquitetura Inception V3 (Szegedy et al. 2016), de acordo com Leonardo e Faria (2019) é uma rede neural convolucional em que os módulos *Inception* são os responsáveis por extrair as características convolucionais e por aprender representações ricas utilizando poucos parâmetros.

Orange Data Mining (Demsar et al., 2013) é uma poderosa ferramenta de mineração de dados e de aprendizado de máquina que possui código aberto e desenvolvimento ativo. Ele oferece uma interface gráfica para a construção de *workflows* interativos, e se destaca por sua facilidade de uso, permitindo que usuários com menos familiaridade com linguagens de programação ainda sejam capazes de utilizar o *software*. O *Orange* utiliza de um sistema baseado em *widgets* que podem ser conectados visualmente e com isso processar e analisar dados.

A detecção precoce de doenças no plantio de manga é essencial para ajudar a manter a saúde dos pomares e minimizar as perdas na produção. De acordo com Weymar e Moraes (2024), por mais que as técnicas de *Machine Learning* se mostrem como uma boa alternativa para a detecção de doenças, ainda temos uma outra dificuldade que é a barreira de complexidade técnica.

Diante disso, o presente projeto teve como objetivo analisar a performance dos seguintes algoritmos: *Support Vector Machine* (SVM), *Random Forest*, *Logistic Regression* e *Neural Network*, no contexto de classificação de antracnose em folhas de manga, utilizando como ferramenta para a análise dos dados, criação e avaliação dos modelos o *software Orange Data Mining*.

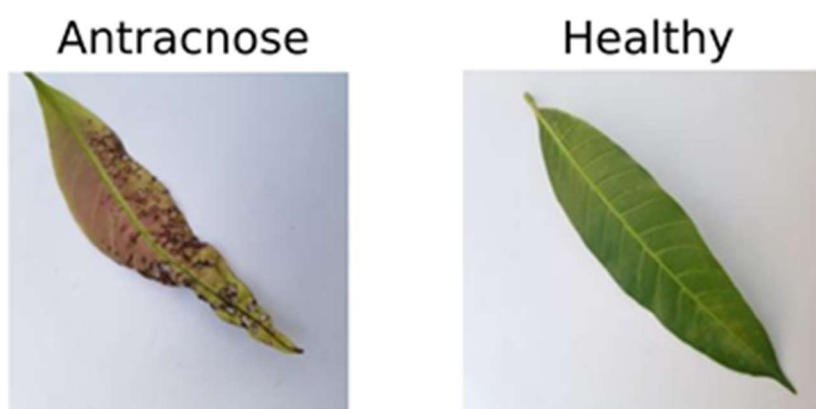
2. MATERIAL E MÉTODOS

2.1 Obtenção do *dataset*

As imagens de folhas contaminadas com antracnose e folhas saudáveis, foram coletadas do conjunto de dados disponibilizado por Ali et al (2022). Este conjunto utilizou imagens coletadas de quatro pomares localizados em Bangladesh, utilizando câmeras de celulares *android*. Apesar da base de dados não ter sido construída no

Brasil, nota-se a existência dessa doença em nosso território nacional (Silva, 2024). Ainda convém mencionar que Ali et al (2022) utilizou técnicas de *data augmentation*, como: *zoom* e rotação, na preparação do conjunto de dados. Ao todo foram utilizadas 1000 imagens, sendo 500 delas de folhas contaminadas e o restante de folhas saudáveis.

Figura 1 – Exemplo de uma folha com antracnose e uma folha em estado saudável.



Fonte: Elaborada pelos autores (2024).

2.2 Pré-processamento

Devido à falta de padronização nas dimensões das imagens, foi necessário realizar um redimensionamento utilizando o *Pillow* na versão 10.3.0 (Clark et al., 2024) para garantir que todas elas possuíssem tamanhos uniformes e fossem mais adequadas ao treinamento dos algoritmos. Para a definição das novas dimensões das imagens foi utilizado uma proporção de 256x256 assim como proposto por Rao et al. (2021).

2.3 Importação das imagens

Os dados foram importados para o *Orange*, na versão 3.37.0 (Demsar et al., 2013), utilizando o *widget "Import Image"*. Após a importação dos dados foi utilizado o

widget “Image Embedding”, que é responsável por extrair as características da imagem, como a cor e a forma das imagens. Ele realiza essa extração através de modelos de redes neurais profundas e assim como Weymar e Moraes (2024), foi utilizado o modelo *Inception v3* (Szegedy et al. 2016). Nesse processo de extração de características é retornado para cada imagem um respectivo vetor contendo 2053 dimensões, das quais as 6 primeiras são metadados e as 2048 restantes representam características visuais extraídas pelo modelo *Inception v3*.

Tabela 1 – Exemplo de vetor gerado pelo modelo *Inception v3*.

Classe	Nome da Imagem	Diretório	Tamanho	Largura	Altura	N0	N1	N2047
--------	----------------	-----------	---------	---------	--------	----	----	------	-------

Fonte: Elaborada pelos autores (2024).

2.4 Configurações dos modelos

Foram utilizados quatro modelos de classificação disponíveis no *Orange* sendo eles: *SVM*, *Random Forest*, *Logistic Regression* e *Neural Network*, para analisar o melhor para o contexto do trabalho.

O algoritmo *SVM* foi utilizado com um *kernel RBF*, que de acordo com Peixoto (2016) é o *kernel* mais referenciado em trabalhos com imagens, as tolerâncias numéricas ajustadas em 0.0010 e o limite de iterações em 100. Ao algoritmo *Random Forest*, assim como Almeida, Anjos e Galvão (2015) foi definido o valor de 10 árvores e para que não dividisse subconjuntos menores que 5. O algoritmo *Logistic Regression* foi utilizado com tipo de regularização Ridge (L2), que segundo Miranda e Bombacini (2023) é uma regularização que reduz a chance de *overfitting* do modelo. E o algoritmo *Neural Network*, conta com 80 neurônios na camada escondida, a função de ativação dos neurônios utilizada foi a *ReLU*, o *Solver* utilizado foi o *Adam* (Kingma; Ba, 2015), a taxa de aprendizagem ajustada em 0.0001, o valor máximo de iterações igual a 200 e a opção de treinamentos replicáveis selecionada.

2.4 Avaliação dos modelos

Para avaliação dos algoritmos foi utilizado o método de validação cruzada (*cross-validation*) (Silva, 2021), ajustado em 10 partições (*folds*), que irá dividir o

dataset em 10 partes iguais e em cada iteração, 9 partes são usadas para treino e 1 para teste, alternando os folds de teste. O processo se repete por 10 vezes, e as métricas são calculadas em cada iteração e ao final, a média dessas métricas fornece a avaliação do modelo sendo considerado os valores de acurácia (CA), área abaixo da curva R.O.C. (AUC), precisão (Prec), *recall* e medida-F (F1) como métricas de avaliação dos modelos, segundo sugerido por Sharma et al. (2020).

A área abaixo da curva R.O.C (AUC) é uma métrica que mede a discriminação do teste, ou seja, no caso do presente trabalho, a capacidade de apontar corretamente entre as plantas com e sem a doença (Polo; Miot, 2020).

A acurácia (CA), é uma métrica que indica a porcentagem de classificações corretas feitas pelo modelo em relação ao total de previsões feitas (Cardoso, 2021).

A precisão ou *Prec* é dita como a porcentagem de amostras classificadas de maneira correta como positivas em relação a todas as instâncias preditas como positivas. Enquanto o *recall* é definido como o percentual de instâncias que foram corretamente classificadas como positivas em relação ao total de instâncias que realmente pertencem à classe positiva. A métrica F1 é um número presente entre o intervalo de 0 e 1 e representa a média entre precisão e *recall* (Cardoso, 2021).

Com o intuito de obter mais detalhes em relação ao desempenho do modelo, foi utilizado a matriz de confusão que traz informações acerca das predições realizadas pelo modelo e os valores reais da base de dados. De acordo com Vilela Junior et al. (2022), os cenários possíveis em uma matriz de confusão são divididos em quatro categorias: Verdadeiro Positivo (VP), Falso Positivo (FP), Verdadeiro Negativo (VN) e Falso Negativo (FN).

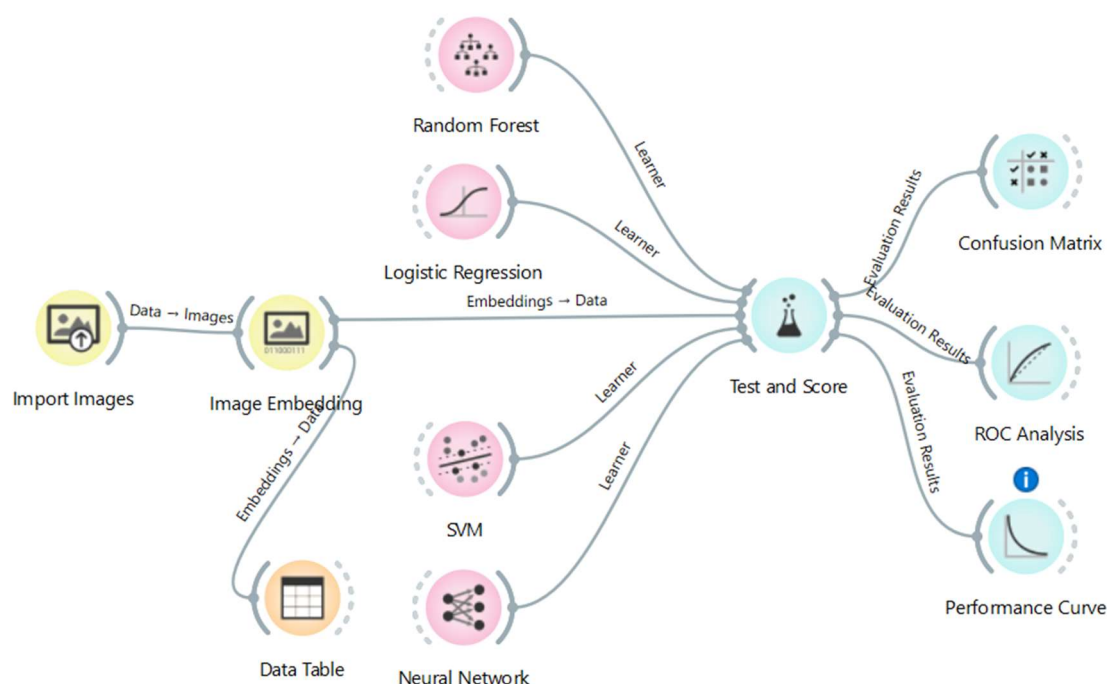
A *Lift Curve* é uma medida visual usada para determinar a eficácia de um modelo classificador. Ela representa a relação entre o limiar da população e a taxa de respostas positivas obtidas. A curva de *Lift* compara o desempenho do modelo com um cenário sem modelo (base), onde, de acordo com Choudhary e Gianey (2017) se o desempenho estiver acima da linha de base, o classificador é considerado bom.

Outra métrica utilizada foi a curva ROC, sendo essa segundo Polo e Miot (2020), um estimador estatístico que traz uma representação gráfica do desempenho do modelo. O gráfico exibe no eixo x a taxa de falsos positivos e no eixo y a taxa de verdadeiros positivos. Quanto mais próxima a curva do canto superior esquerdo, melhor é a capacidade de discriminação das classes por parte do modelo.

3. RESULTADOS E DISCUSSÃO

Todos os treinamentos e validações dos modelos de *machine learning* foram realizados por meio da plataforma *Orange Data Mining* (Demsar, J. et al, 2013) em um *workflow* conforme apresentado na Figura 2.

Figura 2 – Workflow desenvolvido no Orange.



Fonte: Elaborada pelos autores (2024).

Para a avaliação do modelo, foram utilizadas as métricas apresentadas pelo widget “*Test and Score*”, a Tabela 2 apresenta os respectivos resultados:

Tabela 2 – Resultados das métricas de AUC, CA, F1, Prec e Recall dos algoritmos SVM, Logistic Regression, Neural Network e Random Forest.

Modelo	AUC	CA	F1	Prec	Recall
SVM	1	0,997	0,997	0,997	0,997
Logistic Regression	1	0,997	0,997	0,997	0,997
Neural Network	1	0,994	0,994	0,994	0,994
Random Forest	0,995	0,975	0,975	0,975	0,975

Fonte: Elaborada pelos autores (2024).

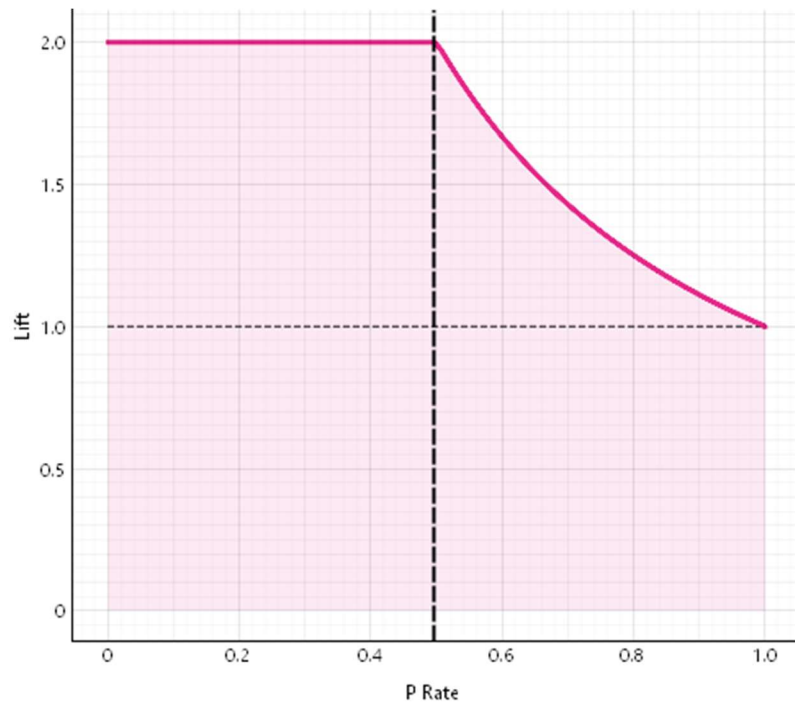
Conforme apresentado na Tabela 2, os modelos *Neural Network*, SVM e *Logistic Regression* alcançaram o valor máximo de AUC 1, e o modelo *Random Forest* ficou bem próximo deles com 0.997.

Os modelos *Logistic Regression*, *Neural Network* e SVM obtiveram valores de 99% de CA, seguidos pelo *Random Forest* com 97%, esses valores indicam que os modelos conseguiram acertar corretamente quase todas as imagens. Comparando com Silva (2024), que propôs em seu trabalho a utilização de um modelo personalizado baseado no *Inception V3* para classificação de 7 doenças da manga e obteve em sua pesquisa uma acurácia máxima de 98%. Os modelos *Logistic Regression*, *Neural Network* e SVM apresentados obtiveram resultados próximos, tendo uma performance de 99% de acurácia.

Em relação as métricas *Prec* e *Recall*, os algoritmos *Logistic Regression*, *Neural Network* e SVM alcançaram valores de 99% em ambas, enquanto o modelo *Random Forest* obteve 97% nas avaliações, o que sugere que eles estão equilibrados entre precisão e *recall*, ou seja, eles conseguem tanto identificar corretamente as folhas contaminadas (sem muitos falsos positivos) quanto detectar a maioria das folhas contaminadas (sem muitos falsos negativos). Os valores de F1 obtidos pelos modelos *Logistic Regression*, *Neural Network* e SVM foram iguais a 0,99, já o modelo *Random Forest* alcançou 0,97. Ainda comparando com Silva (2024), valores similares foram encontrados em seu trabalho, onde o modelo proposto por ele alcançou valores de *Recall* e *Prec* iguais a 100% e a métrica F1 com o valor 1.

Uma outra métrica relevante foi a *Lift Curve*. O Gráfico 1 representa o desempenho da *Logistic Regression* nessa métrica.

Gráfico 1 – *Lift Curve* do modelo *Logistic Regression*.



Fonte: Elaborada pelos autores (2024).

O modelo alcançou um *Lift* máximo superior a 1, o que significa que o modelo está superando um classificador aleatório. A curva começa alta e decai gradualmente, isso sugere que o modelo está identificando os casos mais óbvios de antracnose com alta confiança e, à medida que a probabilidade diminui, o desempenho também diminui, mas de forma suave.

A matriz de confusão de cada algoritmo, permite a visualização das classificações dos valores preditos e os valores reais em verdadeiro positivo, falso positivo, verdadeiro negativo e verdadeiro positivo como apresentado na Tabela 3.

Tabela 3 - Matriz de confusão dos modelos *Random Forest* (A) e *Logistic Regression* (B), *SVM* (C) e *Neural Network* (D).

		Predicted		Σ
		Antracnose	Saudavel	
Actual	Antracnose	490	10	500
	Saudavel	15	485	500
Σ		505	495	1000

A

		Predicted		Σ
		Antracnose	Saudavel	
Actual	Antracnose	497	3	500
	Saudavel	0	500	500
Σ		497	503	1000

B

		Predicted		
		Antracnose	Saudavel	Σ
Actual	Antracnose	498	2	500
	Saudavel	1	499	500
Σ		499	501	1000

C

		Predicted		
		Antracnose	Saudavel	Σ
Actual	Antracnose	494	6	500
	Saudavel	0	500	500
Σ		494	506	1000

D

Fonte: Elaboradas pelos autores (2024).

Na matriz de confusão do modelo *Random Forest* (A), observa-se que houve mais erros ao classificar folhas saudáveis como contaminadas pela antracnose (15 casos) do que o inverso. O modelo também classificou incorretamente 10 folhas da classe "Antracnose" como pertencentes à classe "Saudável". No entanto, a maior parte das imagens foi corretamente classificada: 490 da classe "Antracnose" e 485 da classe "Saudável".

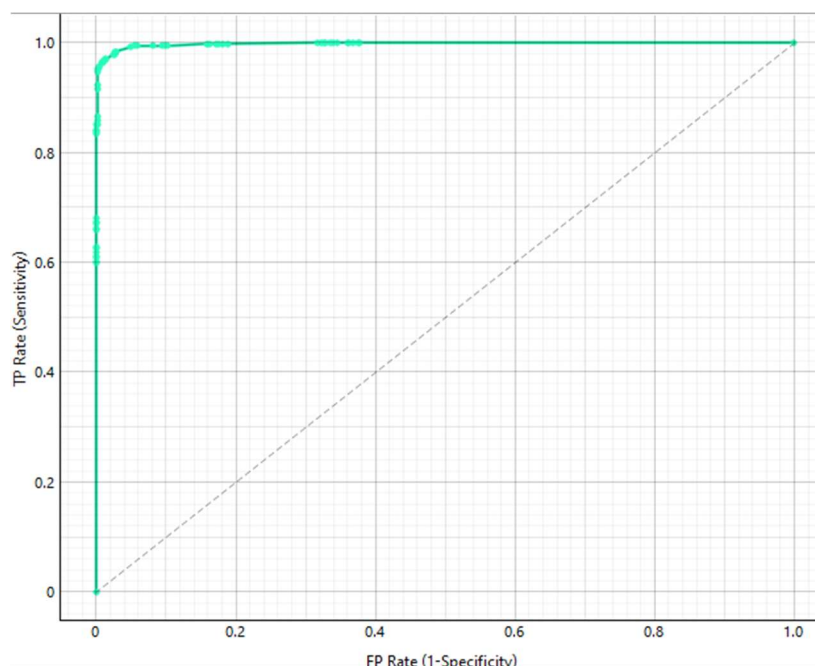
Enquanto o modelo de *Logistic Regression* (B), conseguiu predizer corretamente todas as folhas que estavam saudáveis e errou apenas 3 vezes ao apontar que imagens que continham folhas com antracnose apresentavam folhas da classe "Saudável".

O modelo de *Support Vector Machine* (SVM) (C), teve a mesma quantidade de erros do modelo B, porém, apresentou um padrão distinto em comparação a *Logistic Regression*. Embora também tenha cometido 3 erros em suas predições, a distribuição desses erros foi diferente. O SVM classificou erroneamente 1 amostra da classe "Saudável" e 2 amostras da classe "Antracnose". Isso resultou em um total de 498 acertos para folhas com antracnose e 499 acertos para folhas saudáveis.

Em relação ao modelo de *Neural Network* (D), foi evidenciado erros apenas ao classificar folhas com Antracnose como sendo da classe Saudável (6), enquanto acertou todas as predições das amostras foliares saudáveis, o que resultou em 494 acertos na classe Antracnose.

Exemplos de curva ROC gerada a partir dos resultados do modelo de *Random Forest* são ilustrados no Gráfico 2 e no Gráfico 3.

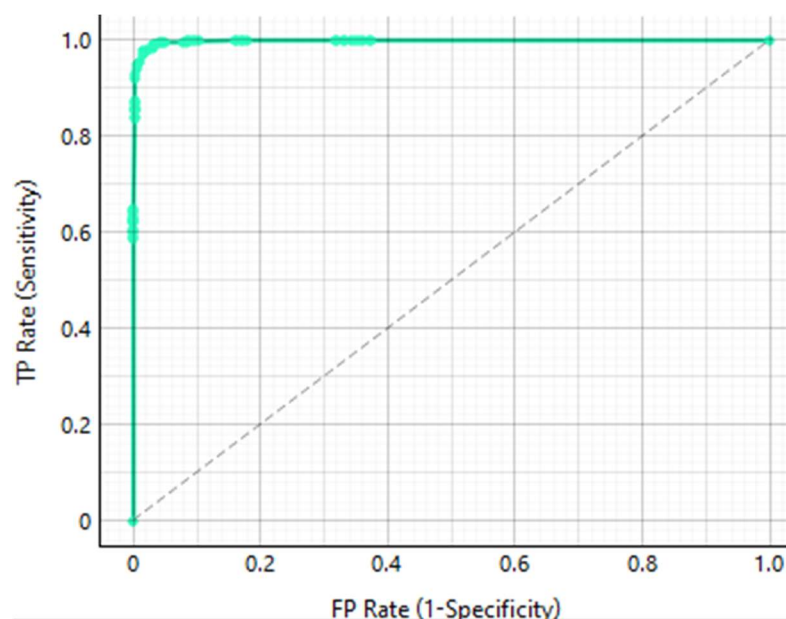
Gráfico 2 – Curva ROC do algoritmo *Random Forest* ao classificar folhas da classe “Saudável”.



Fonte: Elaborado pelos autores (2024).

O gráfico acima (Gráfico 2) teve a curva ROC do algoritmo gerada ao classificar folhas da classe “Saudável”, enquanto no Gráfico 3 foi ilustrada a curva ROC ao classificar folhas da classe “Antracnose”.

Gráfico 3 – Curva ROC do algoritmo *Random Forest* ao classificar folhas da classe “Antracnose”.

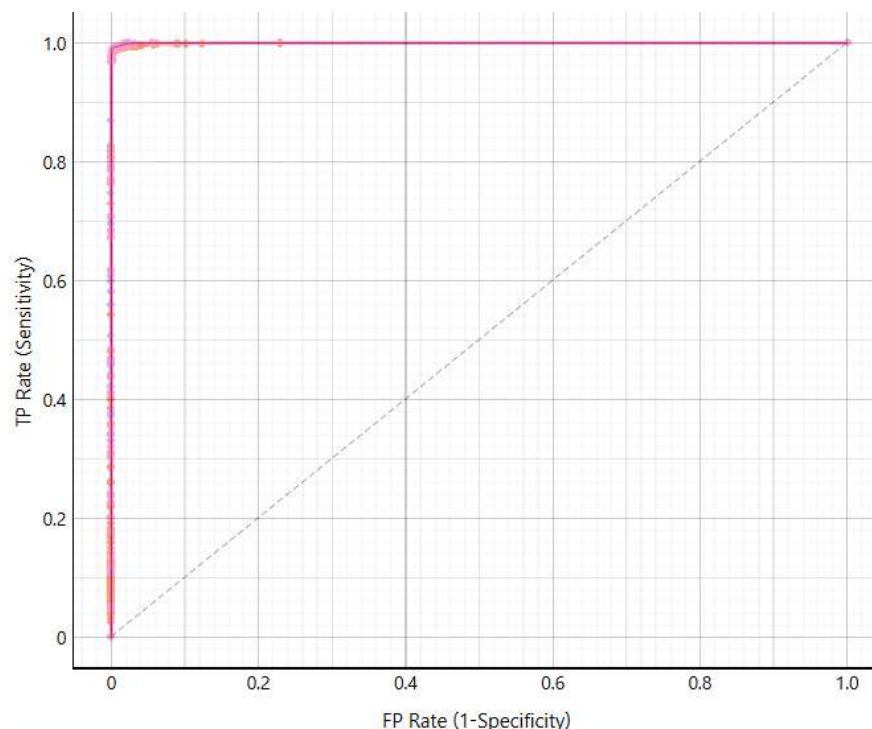


Fonte: Elaborado pelos autores (2024).

Observa-se uma proximidade muito grande de ambas as curvas com a extremidade superior esquerda do gráfico, enfatizando que o modelo discriminou de maneira correta a maioria das predições realizadas tanto para folhas da classe “Saudável” como folhas da classe “Antracnose”.

Comparando os gráficos dos demais modelos utilizados, como *Neural Network*, *Logistic Regression* e SVM, verifica-se que suas respectivas curvas também se encontram próximas ao canto superior esquerdo, conforme observado nos Gráficos 4 e 5.

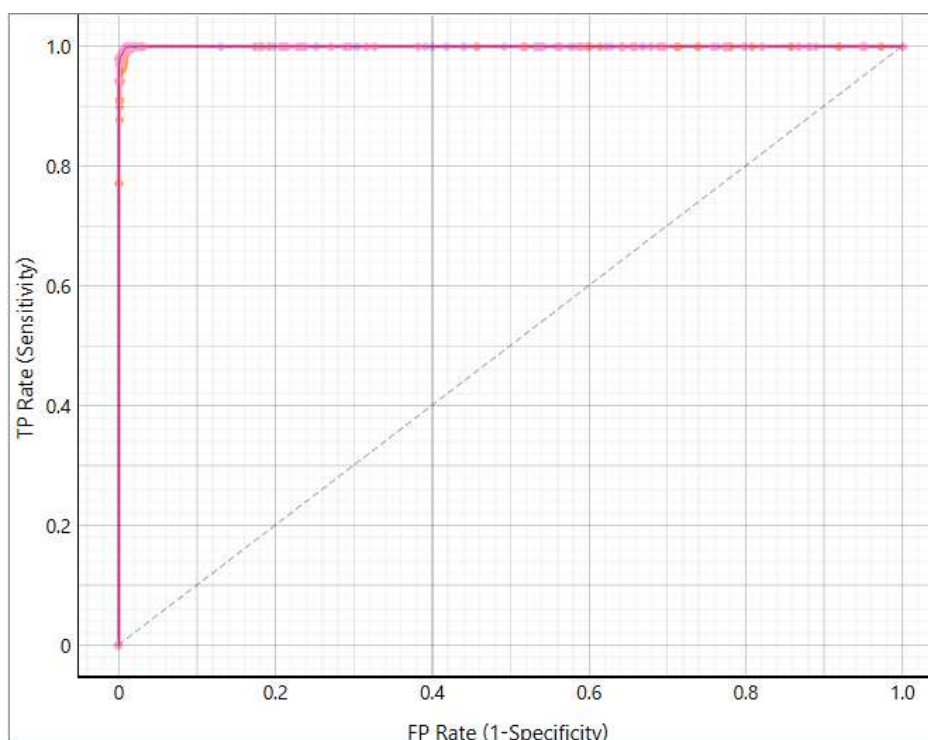
Gráfico 4 – Curvas ROC dos algoritmos *Logistic Regression*, SVM e *Neural Network* ao classificar folhas da classe “Antracnose”.



Fonte: Elaborado pelos autores (2024).

Em ambos os gráficos as curvas estão sobrepostas, o que indica uma similaridade nos resultados obtidos pelos modelos.

Gráfico 5 – Curvas ROC dos algoritmos *Logistic Regression*, SVM e *Neural Network* ao classificar folhas da classe “Saudável”



Fonte: Elaborado pelos autores (2024).

Percebe-se, então, que todos os modelos desenvolvidos obtiveram êxito ao discriminar folhas com e sem a doença, demonstrando uma alta capacidade de identificação dos padrões característicos de cada classe. Porém, os modelos que mais se ajustaram ao problema foram o SVM e *Logistic Regression*, ambos com desempenhos idênticos em todas as métricas de validação.

CONCLUSÃO

Com base nas métricas obtidas pelos algoritmos na classificação de folhas com antracnose, observa-se que os modelos conseguiram classificar de maneira concisa a maioria das amostras das respectivas classes do conjunto de dados. Portanto, o presente trabalho conclui que os modelos apresentados podem ser utilizados para auxiliar na classificação precoce de antracnose em folhas de manga, destacando os modelos SVM, *Logistic Regression* e *Neural Network*, com isso garantindo uma tomada de decisão mais rápida e eficaz ao produtor.

Em trabalhos futuros, a expansão do conjunto de dados para abranger uma gama maior de culturas e doenças foliares, possibilitaria a criação de um sistema mais

generalizado e aplicável a múltiplas doenças.

REFERÊNCIAS

ALI, S.; IBRAHIM, M.; AHMED, S. I. et al. **MangoLeafBD Dataset**. Mendeley Data, V1, 2022. DOI:10.17632/hxsnvwt3r.1. Disponível em: <https://data.mendeley.com/datasets/hxsnvwt3r/1>. Acesso em: 13 set. 2024.

ALMEIDA, C. M. de; ANJOS, C. S. dos; GALVÃO, L. S. **Identificação de materiais urbanos por meio de métodos inovadores de classificação de imagens**. In: X Simpósio Brasileiro de Sensoriamento Remoto, João Pessoa, PB, Brasil, 25 a 29 de abril de 2015. Disponível em: https://www.researchgate.net/publication/316687769_Identificacao_de_materiais_urbanos_por_meio_de_metodos_inovadores_de_classificacao_de_imagens. Acesso em: 21 out. 2024.

BATISTA, D. da C.; RIBEIRO JUNIOR, P. M.; BARBOSA, M. A. G.; ANDRADE, J. N. de; TERAPO, D. Doenças da mangueira. **Informe Agropecuário**, Belo Horizonte, v. 37, n. 290, p. 82-91, 2016. Disponível em: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/178982/1/Informe-Agropecuario-v.37-n.290-p.82-912016.pdf>. Acesso em: 21 out. 2024.

CARDOSO, I. S. **Análise de algoritmos de aprendizado de máquina para classificação de doenças foliares em cafeeiros**. 2021. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Instituto Federal Goiano, Campus Rio Verde, 2021. Disponível em: https://repositorio.ifgoiano.edu.br/bitstream/prefix/2712/3/tcc_Isaac%20da%20Silva%20Cardoso.pdf. Acesso em: 21 out. 2024.

CERRI, R.; CARVALHO, A. C. P. de L. F. de. **Aprendizado de máquina: breve introdução e aplicações**. Cadernos de Ciência e Tecnologia, v. 34, n. 3, p. 297-313, 2017. Disponível em: <https://seer.sct.embrapa.br/index.php/cct/article/view/26381>. Acesso em: 21 out. 2024.

CHOUNDARY, R.; GIANEY, H. K. Comprehensive Review On Supervised Machine Learning Algorithms. **International Conference on Machine Learning and Data Science (MLDS)**, Noida, India, 2017, pp. 37-43. Disponível em: <https://doi.org/10.1109/MLDS.2017.11>. Acesso em: 21 out. 2024.

CLARK, A.; GEISLER, Contributors. **Pillow (PIL Fork) Documentation**. 2024. Disponível em: <https://pillow.readthedocs.io/>. Acesso em: 15 out. 2024.

DEMSAR, J. et al. Orange: Data Mining Toolbox in Python. **J Mach Learn Res**, v. 14, p. 2349–2353, ago. 2013.

DISCONZI, Claudia Maria Dias Guerra. **Previsão dos preços de commodities agrícolas brasileiras no mercado futuro utilizando redes neurais artificiais**. Santana do Livramento: UNIPAMPA, 2018 Disponível em: <http://dspace.unipampa.edu.br/bitstream/riu/2967/1/Claudia%20Guerra%20Disconzi%20-%202018.pdf>. Acesso em: 17 out. 2024.

EMBRAPA. **Brasil bate recorde de receita e exportação com manga produzida no Nordeste**. Disponível em: <https://www.embrapa.br/busca-de-noticias/-/noticia/87702249/brasil-bate-recorde-de-receita-e-exportacao-com-manga-produzida-no-nordeste>. Acesso em: 17 jul. 2024.

FERNANDES, A. A. T. et al. Leia este artigo se você quiser aprender regressão logística. **Revista de Sociologia e Política**, v. 28, p. 006, 8 jan. 2021. Acesso em: 18 out. 2024.

FLORI, J. E.; RESENDE, G. M. de; BORGES, A. L. Aspectos socioeconômicos. In: **Sistema de Produção**. Cruz das Almas: Embrapa Mandioca e Fruticultura, 2020. (Embrapa Mandioca e Fruticultura. Sistema de Produção, 49). Disponível em: <http://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1130648>. Acesso em: 21 out. 2024.

KINGMA, D. P.; BA, J. L. **Adam: A Method for Stochastic Optimization**. ICLR, 2015. DOI: 10.48550/arXiv.1412.6980. Disponível em: <https://arxiv.org/abs/1412.6980>. Acesso em: 20 out. 2024.

LEONARDO, M. M.; FARIA, F. A. **Um sistema de reconhecimento de espécies de moscas-das-frutas**. Revista Eletrônica de Iniciação Científica em Computação, Porto Alegre, v. 17, n. 2, 2019. Edição Especial - Artigos do CSBC/CTIC. Disponível em: <https://seer.ufrgs.br/index.php/reic/article/view/93500>. Acesso em: 16 out. 2024.

Ministério da Agricultura e Pecuária. **Projeções do Agronegócio 2023-2033**. Disponível em: <https://www.gov.br/agricultura/pt-br/assuntos/noticias/producao-de-graos-brasileira-devera-chegar-a-390-milhoes-de-toneladas-nos-proximos-dez-anos/ProjeesdoAgronegcio20232033.pdf>. Acesso em: 17 jul. 2024.

MIRANDA, C. E. B.; BOMBACINI, M. R. Aplicação da Regressão Logística Binária para Manutenção Preditiva em Máquinas de Ressonância Magnética. In: XIII CONGRESSO BRASILEIRO DE ENGENHARIA DE PRODUÇÃO, 2023. **Anais....** Disponível em: https://aprepro.org.br/conbrepro/anais/2023/arquivos/10302023_141040_653feb4c9b1c5.pdf. Acesso em: 21 out. 2024

MITRA, S.K. (2016). **Mango production in the world – present situation and future prospect**. Acta Hort. 1111, 287-296. DOI: 10.17660/ActaHortic.2016.1111.41. Disponível em: https://www.actahort.org/books/1111/1111_41.htm. Acesso em: 20 out. 2024.

MUNASINGHE, Sidath, 2017. **10 machine learning algorithms you need to know**. Disponível em: <https://towardsdatascience.com/10-machine-learning-algorithms-you-need-to-know-77fb0055fe0>. Acesso em: 18 out. 2024.

NETO, T. F. de S. C.; PEIXOTO, A. R.; SOUZA, E. M. de. Controle da antracnose no cultivo da mangueira. **Boletim Informativo dos Projetos Públicos de Irrigação da Codevasf**, Edição nº 31, out. 2023. Disponível em:

<https://www.codevasf.gov.br/linhas-de-negocios/agricultura-irrigada/projetos-de-irrigacao/bip/31a-edicao/controle-da-antracnose-no-cultivo-da-mangueira>. Acesso em: 21 out. 2024.

PEIXOTO, E. B. de A. **Utilização do algoritmo SVM (Support Vector Machine) e árvores de decisão para classificação de imagens de alta resolução do sensor ADS-40**. In: VI Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação, Recife, ago. 2016. Disponível em: https://www.researchgate.net/publication/375417473_UTILIZACAO_DO_ALGORITMO_SVM_SUPPORT_VECTOR_MACHINE_E_ARVORES_DE_DECISAO_PARA_CLASSIFICACAO_DE_IMAGENS_DE_ALTA_RESOLUCAO_DO_SENSOR_ADS-40. Acesso em: 21 out. 2024.

POLO, T. C. F.; MIOT, H. A. Aplicações da curva ROC em estudos clínicos e experimentais. **Jornal Vascular Brasileiro**, v. 19, p. e20200186, 2020. Disponível em: <https://www.scielo.br/j/jvb/a/8S8Pfqnz8csmQJVqwgZT8gH/>. Acesso em: 22 out. 2024.

RAO, U. S.; SWATHI, R.; SANJANA, V.; ARPITHA, L.; CHANDRASEKHAR, K.; CHINMAYI; NAIK, P. K. Deep Learning Precision Farming: Grapes and Mango Leaf Disease Detection by Transfer Learning. **Global Transitions Proceedings**, v. 2, n. 2, p. 535-544, 2021. DOI: <https://doi.org/10.1016/j.gltp.2021.08.002>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2666285X21000303>. Acesso em: 16 out. 2024.

SANTOS, T. T.; BARBEDO, J. G. A.; TERNES, S.; CAMARGO NETO, J.; KOENIGKAN, L. V.; SOUZA, K. X. S. de. **Agricultura digital: pesquisa, desenvolvimento e inovação nas cadeias produtivas**. Brasília, DF: Embrapa, 2020. cap. 6, p. 146-164. Disponível em: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/217703/1/LV-Agricultura-digital-2020-cap6.pdf>. Acesso em: 24 out. 2024.

SHARMA, R. et al. **Exploring Geometric Feature Hyper-Space in Data to Learn Representations of Abstract Concepts**. Appl. Sci. 2020, 10, 1994. DOI: 10.3390/app10061994. Disponível em: <https://www.mdpi.com/2076-3417/10/6/1994>. Acesso em: 20 out. 2024.

SILVA, I. P. C. A. da. **Um estudo sobre o TinyML a partir de uma aplicação mobile para detecção de doenças em folhas de manga**. 2024. 36 f. Trabalho de Conclusão de Curso (Bacharelado em Engenharia de Computação) – Instituto de Computação, Universidade Federal de Alagoas, Maceió, 2023. Disponível em: <http://www.repositorio.ufal.br/jspui/handle/123456789/13396>. Acesso em: 20 out. 2024.

SILVA, T. F. **Avaliação dos impactos dos fatores de pré-abate no dripping test de frango utilizando redes neurais artificiais**. 2021. Dissertação (Mestrado em Tecnologia de Alimentos) – Universidade Tecnológica Federal do Paraná, Londrina, 2021. Disponível em: <http://repositorio.utfpr.edu.br/jspui/handle/1/26552>. Acesso em: 29 nov. 2024.

SZEGEDY, C.; VANHOUCKE, V.; IOFFE, S.; SHLENS, J.; WOJNA, Z. **Rethinking the inception architecture for computer vision**. In: PROCEEDINGS OF THE IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2016. p. 2818-2826. DOI: 10.1109/CVPR.2016.308. Disponível em: <https://ieeexplore.ieee.org/document/7780677>. Acesso em: 20 out. 2024.

VILELA JUNIOR, G. de B.; LIMA, B. N.; PEREIRA, A. de A.; RODRIGUES, M. F.; OLIVEIRA, J. R. L. de; SILIO, L. F.; CARVALHO, A. dos S.; FERREIRA, H. R.; PASSOS, R. P. **Determinação das métricas usuais a partir da matriz de confusão de classificadores multiclasse em algoritmos inteligentes nas ciências do movimento humano**. Revista CPAQV - Centro de Pesquisas Avançadas em Qualidade de Vida, [S. I.], v. 14, n. 2, 2022. DOI: 10.36692/v14n2-01. Disponível em: <https://revista.cpaqv.org/index.php/CPAQV/article/view/939>. Acesso em: 23 out. 2024.

WEBER, N. P. da L. **Doenças foliares da soja e licheira: imagens e informações**. 2022. 39 f. Trabalho de Conclusão de Curso (Bacharelado em Agronomia) - Universidade Tecnológica Federal do Paraná, Ponta Grossa, 2022. Disponível em: <https://repositorio.utfpr.edu.br/jspui/bitstream/1/32354/1/doencasfoliaresimagenssojalicheira.pdf>. Acesso em: 20 out. 2024.

WEYMAR, R. R.; MORAES, F. S. Empregando Inteligência Artificial na Agricultura: Uma Análise Prática com Orange Data Mining para Classificação de Folhas de Soja. In: 5th NIIC – NECSOS' International and Interinstitutional Colloquium: Inteligência Artificial em Pesquisa: reflexões sobre o uso e possibilidades de aplicação para o fazer ciência. Niterói e Online, 18 e 19 de dezembro de 2023. **Anais** do NIIC – NECSOS. ISSN: 2763-6186, 2024. Disponível em: https://www.researchgate.net/publication/378860103_Empregando_Inteligencia_Artificial_na_Agricultura_Uma_Analise_Pratica_com_Orange_Data_Mining_para_Classificacao_de_Folhas_de_Soja. Acesso em: 16 out. 2024.