

Avaliação de modelo de machine learning para classificação da severidade da ferrugem (*Hemileia vastatrix*) em cafeeiro (*Coffea sp.*)

ELIZABETH CRISTINA DA SILVA TAVARES¹; TAMARA VITÓRIA ALVES DE OLIVEIRA¹; JOÃO RICARDO FAVAN²; RENATA BRUNA DOS SANTOS COSCOLIN FAVAN²

¹Discentes em Big Data no Agronegócio na Fatec Pompeia “Shunji Nishimura”, Pompeia-SP, elizabeth.tavares@fatec.sp.gov.br, tamara.oliveira3@fatec.sp.gov.br

²Docentes do curso de Big Data no Agronegócio na Fatec Pompeia “Shunji Nishimura”, Pompeia-SP.

RESUMO

O cafeeiro possui relevância econômica para o Brasil e para seus produtores, e com a grande possibilidade de contaminação pela ferrugem (*Hemileia vastatrix*) que pode chegar a 95% e pode acarretar a perda na produção de até 50% da cultura, os métodos de *machine learning* se tornaram relevantes para o auxílio no monitoramento das lavouras e na detecção das doenças e seus níveis de contaminação. O objetivo da presente pesquisa é avaliar a performance dos algoritmos *Neural Network* e *Logistic Regression* através do modelo *cross-validation* em cinco repartições e compará-las com a avaliação dos 11 avaliadores não treinados na classificação da severidade da ferrugem no cafeeiro, a fim de identificar o algoritmo com melhores resultados. Os modelos foram treinados com 304 imagens de folhas de café (*Coffea sp.*) com as divisões nos grupos de controle, leve, moderado e severo, e foram testados com as mesmas 36 imagens dos convidados, já os avaliadores receberam uma escala diagramática com seis níveis de contaminação. Através dos resultados de assertividade obtidos pelo modelo *Logistic Regression* serem superiores a 70%, o presente projeto recomenda o uso deste modelo para a classificação de doenças em imagens, visto que é possível a utilização de seus resultados para a classificar os níveis de severidade da ferrugem no cafeeiro.

Palavras-chave: Doença foliar. Inteligência artificial. Modelos computacionais.

ABSTRACT

The coffee tree has economic relevance for Brazil and its producers, and with the high possibility of contamination by coffee leaf rust (*Hemileia vastatrix*), which can reach 95% and can cause the loss in production of up to 50% of the crop, machine learning methods have become relevant to assist in monitoring crops and detecting diseases and their levels of contamination. The objective of this research is to evaluate the performance of the Neural Network and Logistic Regression algorithms through the cross-validation model in five breakdowns and compare them with the evaluation of 11 untrained raters in the classification of coffee leaf rust severity, to identify the algorithm with the best results. The models were trained with 304 images of coffee (*Coffea sp.*) leaves with the divisions in control, mild, moderate, and severe groups, and were tested with the same 36 images of the guests, while the raters were given a diagrammatic scale with six levels of contamination. Since the assertiveness results obtained by the Logistic Regression model are higher than 70%, this project recommends the use of this model for the classification of diseases on images, since it is possible to use its results to classify the severity levels of rust on coffee trees.

Keywords: Leaf disease. Artificial Intelligence. Computational models.

1 INTRODUÇÃO

O café é uma importante cultura para a indústria do Brasil, desde o começo do seu cultivo em meados do século XVIII, colaborando no uso de máquinas agrícolas e criação de fábricas. Os dados de 2022 indicam que a produção cafeeira teve um volume de 50,92 milhões de sacas de café beneficiado sendo 6,7% acima da safra 2021 (CONAB, 2022).

A área nacional da cafeicultura em 2022 teve uma dimensão de cultivo de 2,2 milhões de hectares, desses sendo 1,8 milhão de lavouras em produção obtendo um crescimento de 1,8% em comparação com a safra de 2021 (CONAB, 2022).

Para o ano de 2023, é prevista uma produção de 54,94 milhões de sacas de acordo com o levantamento feito pela Companhia Nacional de Abastecimento (Conab), ainda é mencionado que o café foi internacionalmente comercializado em 2022 com 145 países, atingindo um valor de exportação de US\$ 9,2 bilhões sendo 45% maior em comparação com o valor do ano de 2021 (CONAB, 2023).

Segundo Cabral *et al.* (2016) as espécies de café mais produzidas no Brasil são *Coffea arabica* que representa em torno de 70% da safra brasileira e o *Coffea canephora* (conilon) que corresponde a 30%. Apesar de toda sua importância para o país, até 95% das lavouras de café podem ser atingidas pela ferrugem (*Hemileia*

vastatrix) ameaçando a produção do café com perdas que podem variar de 35% a 50% se as condições climáticas forem suscetíveis. De acordo com Oliveira (2022) lavouras adensadas e com mais sombra proporcionam a manifestação da ferrugem.

Os fungos da classe *Hemileia vastatrix* também comprometem a fotossíntese e fraquejam os pés de café possibilitando a contaminação por outros patógenos (OLIVEIRA, 2022). Os sintomas da ferrugem são iniciados com manchas cloróticas na face abaxial, também pode ser vista na face adaxial, dificilmente formam-se esporos. Em contaminações severas a desfolha causada pela doença pode acarretar a seca dos ramos, de maneira que a safra seguinte se comprometa, principalmente para o café arábica, como citado por Silva (2022).

Um método utilizado para a avaliação da severidade de doenças é a escala diagramática que tem como propósito ser um método de avaliação simples, que pode ser aplicado em diferentes classificações e deve conter uma representação clara dos níveis de contaminação e suas áreas, como referenciado por Sussel (2010):

A utilização de escalas diagramáticas na avaliação de doenças proporciona maior acurácia, precisão e reprodutibilidade nas estimativas de severidade da doença (AMORIM, 1995). A acurácia representa o grau de proximidade entre a estimativa e a quantidade real de doença; a precisão refere-se à confiabilidade e (ou) repetibilidade nas avaliações da doença; e a reprodutibilidade trata da ausência de variação nas estimativas quando a mesma amostra é avaliada por diferentes avaliadores (CAMPBELL; MADDEN, 1990).

O *machine learning* é uma área da inteligência artificial que tem como objetivo desenvolvimento de sistemas e técnicas que são aptos a adquirirem conhecimentos automaticamente. O sistema de *machine learning* é um programa de computador que é capaz de tomar decisões com base em experiências reunidas por meio de uma solução bem-sucedida de casos antecedentes (MONARD; BARANAUSKAS, 2003).

As técnicas de *machine learning* são utilizadas como ferramentas com maior frequência para a identificação de doenças em plantas, dizendo assim se a doença está presente ou não nas culturas. Outra importante aplicação desse conceito se dá a classificação de doenças nas plantas em seus diferentes tipos ou diferentes estágios (YANG; GUO, 2019).

No exposto acima a aplicação de algoritmos de *machine learning* quando utilizados para a detecção e classificação de doenças podem resultar em diversos modelos, por meio disso, através da técnica de avaliação de desempenho *cross-validation* diversas métricas podem ser utilizadas e com isso obter diversidade de informações para análise, que quando comparadas com a avaliação feita por avaliadores não treinados comprova se a assertividade do modelo o torna eficaz.

Portanto, com um mercado relevante nacional e internacionalmente, a utilização de algoritmos para a classificação de severidade de doenças nas folhas possui grande valor, sendo assim, o objetivo da presente pesquisa é avaliar a performance dos algoritmos de *machine learning* *Neural Network* e *Logistic Regression* e compará-las com a avaliação dos avaliadores não treinados na classificação da severidade da ferrugem no cafeeiro.

2 MATERIAL E MÉTODOS

Para o desenvolvimento da pesquisa o *dataset* encontrado para a utilização das imagens de folhas de café em seus níveis de severidade não obtinha a espécie da cultura definida, e de início contava com seu fundo natural (fotos tiradas de celulares e/ou câmeras fotográficas) por conta disso, diversas imagens estavam com má qualidade e luminosidade e precisaram dos tratamentos descritos nas seguintes etapas: seleção e tratamento das imagens, protocolo e escala diagramática para a classificação das imagens, avaliação visual da severidade da doença foliar, modelo de classificação.

2.1 Seleção e tratamento das imagens

As imagens para a classificação e detecção da severidade foram obtidas através do *dataset* “Coffee leaves disease” (última atualização em 2021) disponível na plataforma Kaggle (KAGGLE, 2021), o mesmo contém um total de 1539 arquivos, separados em dados de treinamento e teste. Para o presente trabalho, foram utilizadas as imagens correspondentes as pastas denominadas “Rust” que contém imagens de folhas saudáveis e infectadas pela ferrugem (*Hemileia vastatrix*) em suas faces adaxial e abaxial.

Para essa pesquisa, foram utilizadas somente as imagens que apresentavam a face adaxial das folhas de café, sendo excluídos os demais arquivos, resultando em

340 imagens. Adicionalmente, as imagens foram segmentadas selecionando somente a folha principal da imagem, sendo seu fundo removido utilizando as ferramentas remove bg (BEN; DAVID, 2018) da empresa Kaleido, Remove BG (SEVANDERSON, 2008) produzido pela Pixlr e o Adobe Express da Adobe (GESCHKE; WARNOCK, 1982).

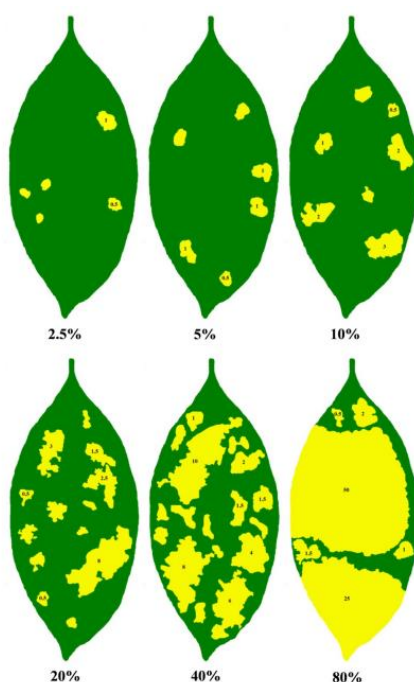
Buscando atingir maior desempenho na classificação e detecção, após a remoção do fundo das imagens, todas tiveram a adição de um novo fundo na cor azul para destacar a folha e suas características, essa adição foi feita na ferramenta Canva (PERKINS, et al., 2012).

2.2 Protocolo e escala diagramática para classificação das imagens

O uso do protocolo “Protocolo para área lesionada de Ferrugem em café” (FAVAN et al., s.d.) (não publicado) através da ferramenta ImageJ (RASBAND, et al., 1997) possibilitou o cálculo da área total da folha e da área contaminada, por meio dos resultados obtidos foi realizado o cálculo do percentual de contaminação de cada folha.

Foi utilizada a escala diagramática apresentada na Figura 1, proposta por Capucho (2011) para a classificação das imagens em seus grupos: controle, leve, moderado e severo.

Figura 1 – Escala diagramática utilizada na pesquisa



Fonte: (CAPUCHO, 2011)

2.3 Avaliação visual da severidade da doença foliar

A severidade da ferrugem foi avaliada visualmente a partir de 36 imagens selecionadas das folhas em seus grupos de infecção pela doença, no qual 11 avaliadores não treinados utilizando a escala diagramática proposta por Capucho (2011), discriminaram visualmente, de acordo com a área lesionada da folha, entre os níveis de severidade controle, leve, moderado e severo.

2.4 Modelo de classificação

O software de mineração de dados Orange, versão 3.34.0 (DEMSAR et al., 2013) e suas ferramentas Image Embedding com a configuração de Embedder Inception v3 (SZEGEDY, et al., 2015); Test and Score e Confusion Matrix foram utilizadas para ajustar o modelo de classificação baseado nos algoritmos de *machine learning* *Neural Network* e *Logistic Regression*.

O modelo de regressão baseado em MultiLayer Perceptron (*MPLRegressor*) foi configurado com 100 neurônios na camada intermediária, cada neurônio usando uma função identidade como função de ativação, para a regularização foram utilizadas as configurações padrões que correspondem ao valor de $a = 0.0001$ e o treinamento foi limitado à duzentas épocas. Para o modelo *Logistic Regression* utilizamos a regularização de tipo Ridge (L2) com a força $C = 1$. Na fase de treinamento, receberam 304 imagens e para o teste as mesmas 36 imagens recebidas pelos avaliadores.

Os modelos ajustados foram testados pelo método de *cross-validation* com 5 repartições como sugerido por Buonocore et al. (2022). Para a avaliação dos indicadores foram utilizados área sob a curva (AUC), acurácia (ACC), F1-Score (F1S), precisão (PRE), revocação (REC).

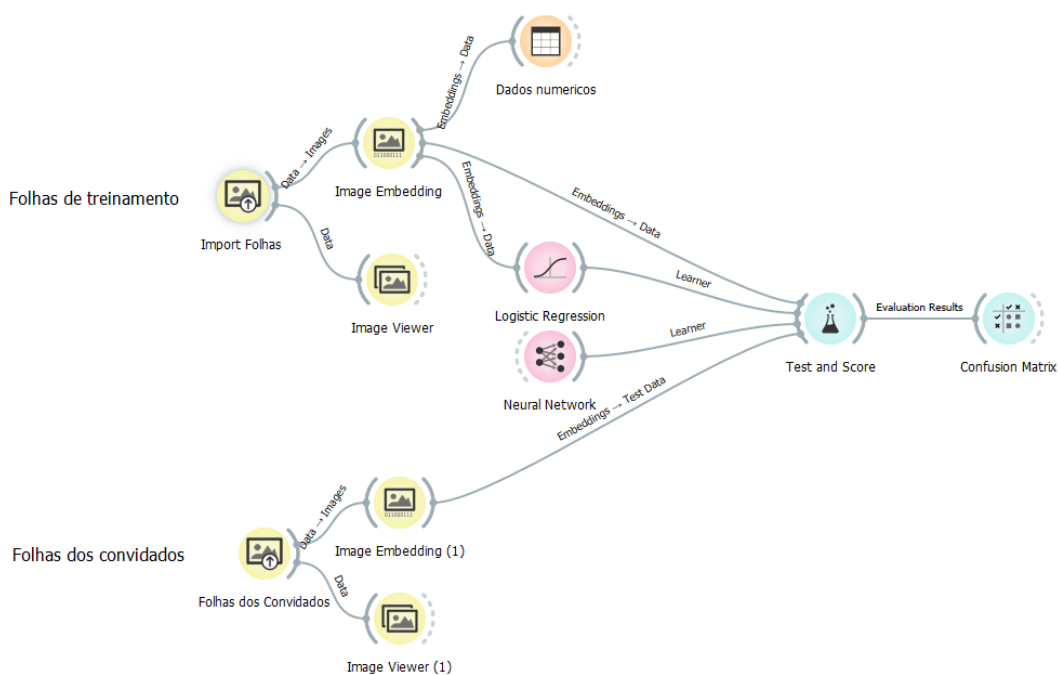
Foram escolhidos avaliadores não treinados para fazer a classificação utilizando a escala diagramática segundo o método proposto por Damasceno et al. (2014) e os resultados obtidos foram comparados com os resultados do algoritmo.

3 RESULTADOS E DISCUSSÃO

O trabalho apresentou diversos resultados, dentre eles, informações sobre as imagens utilizadas, a escolha e o processo de classificação dos modelos, métricas obtidas, matriz de confusão e esses resultados fornecem uma visão aprofundada do desempenho e dos métodos utilizados para a avaliação dos algoritmos de *machine*

learning. A figura seguinte apresenta o quadro resultante onde se encontram os componentes utilizados.

Figura 2 – Quadro do Orange para seleção de imagens, aplicação dos algoritmos e obtenção de resultados.



Fonte: Elaborado pelas autoras (2023)

3.1 Resultados dos algoritmos de *machine learning* treinados

Os algoritmos de *machine learning*: *Neural Network* e *Logistic Regression* foram avaliados com a finalidade de avaliar a performance na classificação das folhas contaminadas pela Ferrugem, obtendo os resultados apresentados através das métricas de Área sob a curva (AUC), Acurácia (CA), F1-Score (F1), Precisão (Precision) e Revocação (Recall) apresentados na Tabela 1.

Tabela 1 – Resultado dos modelos *Neural Network* e *Logistic Regression*, com as métricas de AUC (área abaixo da curva ROC), CA (acurácia), F1 (F1-Score médio do modelo), Precision (precisão do modelo) e Recall (revocação do modelo) obtidos pelo método de validação cruzada com 5 partições.

| Model | AUC | CA | F1 | Precision | Recall |
|---------------------|-------|-------|-------|-----------|--------|
| Neural Network | 0,889 | 0,743 | 0,740 | 0,736 | 0,743 |
| Logistic Regression | 0,891 | 0,747 | 0,740 | 0,740 | 0,747 |

Fonte: Elaborado pelas autoras (2023)

Os modelos obtiveram resultados próximos, contudo o algoritmo *Logistic Regression* obteve valores maiores, com uma acurácia de aproximadamente 75% e uma precisão de 74%. Por meio da comparação entre os resultados obtidos por Castro (2022) que avaliou o monitoramento da ferrugem no cafeeiro com o algoritmo *Random Forest*, e os resultados obtidos pelo algoritmo *Logistic Regression*, avaliado na presente pesquisa, não foram respostas muito distintas apesar de serem algoritmos desiguais.

Em sua pesquisa Castro (2022) obteve uma acurácia de 77,3%, representando um valor de 2,6% maior do que o algoritmo *Logistic Regression*, entretanto sua precisão foi de 38,5% que quando associada ao resultado do modelo analisado nesta pesquisa é 35,5% menor.

Segundo o estudo Meira *et al.* (2009) a validação da acurácia é importante, pois mostra a relevância dos acertos apresentados pelo modelo, para a hipótese de uma aplicação factual.

A matriz de confusão mostrada na Tabela 2 apresenta os valores das predições em percentual de acordo com a classificação feita das imagens.

Tabela 2 – Matriz de Confusão *Logistic Regression*

| | | Previsão | | | | |
|-------|----------|----------|--------|----------|---------|----|
| | | Controle | Leve | Moderado | Severo | |
| Atual | Controle | 100.0 % | 0.0 % | 0.0 % | 0.0 % | 10 |
| | Leve | 0.0 % | 66.7 % | 0.0 % | 0.0 % | 10 |
| | Moderado | 0.0 % | 33.3 % | 83.3 % | 0.0 % | 10 |
| | Severo | 0.0 % | 0.0 % | 16.7 % | 100.0 % | 6 |
| | | 10 | 15 | 6 | 5 | 36 |

Fonte: Elaborado pelas autoras (2023)

Os resultados mostrados na tabela acima, expõem os valores preditos para cada nível de severidade. Essa matriz serve para indicar o percentual das classificações feitas pelo algoritmo com base na definição das categorias que foram determinadas para essa pesquisa.

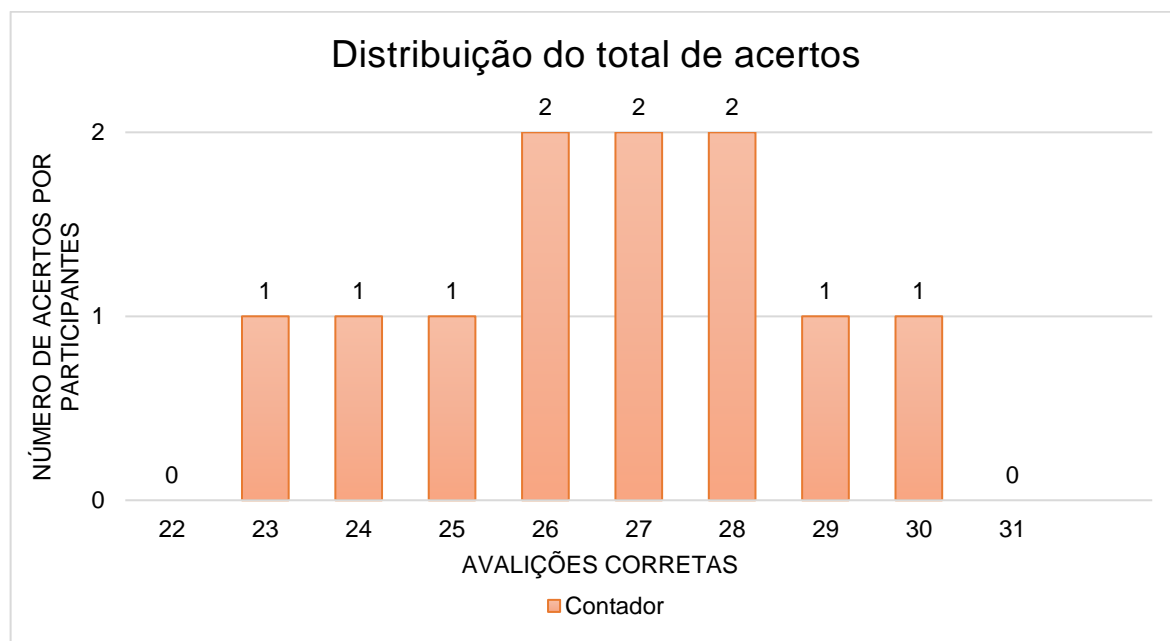
Essas respostas manifestam que o modelo *Logistic Regression* foi capaz de classificar de maneira correta aproximadamente 86% das amostras, identificando a contaminação da doença em seus níveis de severidade

3.2 Resultados dos Avaliadores não treinados

Ao fazer a análise das respostas dos 11 avaliadores participantes, que utilizaram a escala diagramática proposta por Capucho (2011) para fazer a classificação de 36 imagens de folhas de cafeeiro, obtivemos os seguintes resultados: controle com uma média de 95%, leve com 64%, moderado com 66% e severo com 67%. Os avaliadores classificaram aproximadamente 82% das amostras corretamente. O uso da escala diagramática proporciona bons níveis de precisão nas estimativas (NUNES & ALVES, 2012).

O intervalo de acertos nas classificações dos avaliadores foi de 23 e 30 respostas corretas, como apresentado na Figura 3.

Figura 3 – Distribuição do total de acertos por avaliador



Fonte: Elaborado pelas autoras (2023)

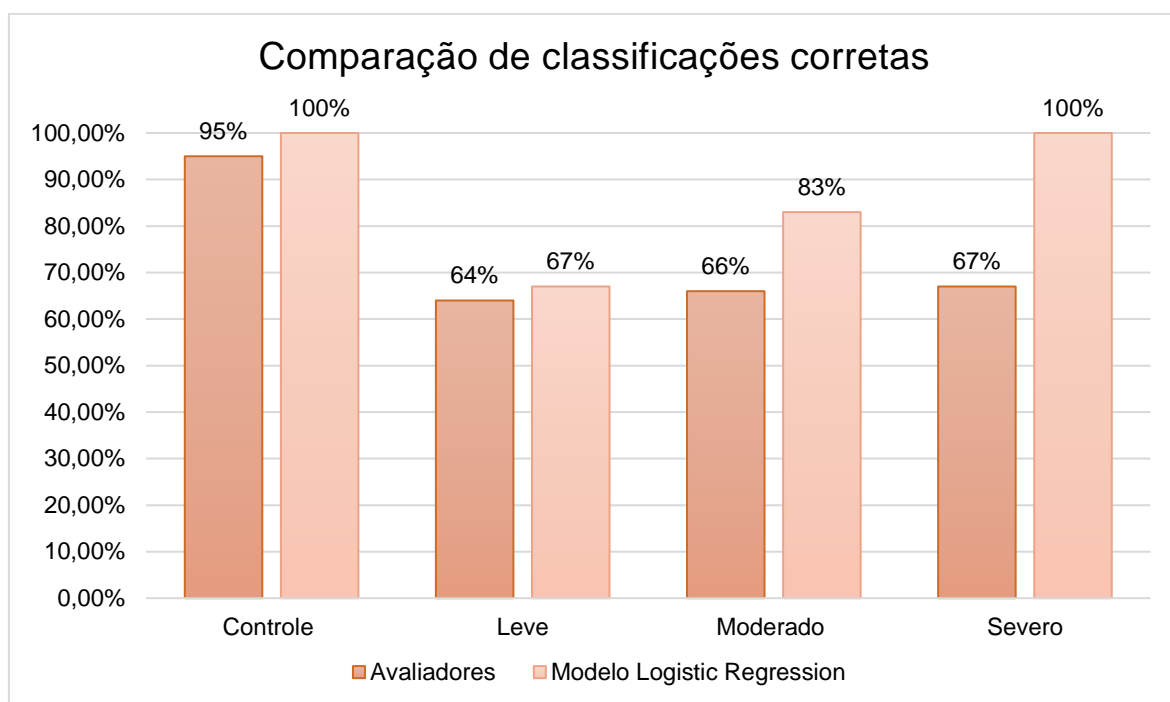
Nenhum dos convidados obtiveram o total de 36 avaliações corretas, o que ocorreu foi que um ou dois participantes acertaram o mesmo número de classificações, como por exemplo dois convidados classificaram corretamente 28 imagens.

A escala diagramática através de seus intervalos e percentuais aprimorou os níveis de precisão e reprodutibilidade das avaliações mostrando assim, ser uma ferramenta proficiente para as avaliações de severidade da ferrugem com avaliadores não treinados ou sem experiência em quantificar doenças, como concluído por Lenz *et al.* (2010).

3.3 Comparação entre os resultados dos avaliadores não treinados e do modelo treinado

O modelo *Logistic Regression* obteve melhores resultados quando comparado aos avaliadores, como apresentado na Figura 4.

Figura 4 – Comparação dos percentuais de classificações corretas dos avaliadores não treinados e do modelo *Logistic Regression*



Fonte: Elaborado pelas autoras (2023)

Os valores em percentual representam o acerto médio de avaliações corretas. Nota-se que os resultados para Controle foram muito próximos, com uma diferença de apenas 5%, isso se dá ao fato de as folhas saudáveis serem mais fáceis de se classificar tanto para os avaliadores quanto para o modelo, por não conterem nenhum tipo de mancha foliar.

Para o grupo de folhas Leve os resultados também tiveram proximidade em seus percentuais apenas 3% de diferença, mas contaram com a menor porcentagem de classificação obtida, para os convidados isso ocorreu devido a percepção de várias lesões menores que criaram uma aparência de maior área contaminada pela ferrugem, como mencionado por Alves *et al.* (2012).

Na classificação do modelo de *machine learning* para o grupo de folhas leve e moderada o percentual foi mais baixo, devido à má qualidade, iluminação (natural, do

sol) e quantidade de elementos secundários presentes nas imagens utilizadas que não corresponderam totalmente ao tratamento realizado nas imagens.

Abaixo, segue a Figura 5, que representa a imagem com os problemas encontrados e a Figura 6 que é o tipo de figura que capacita melhor a avaliação do modelo.

Figura 5 – Imagem de má qualidade utilizada para a classificação com o modelo *Logistic Regression*



Fonte: Elaborado pelas autoras (2023).

Figura 6 – Imagem de melhor qualidade utilizada para a classificação com o modelo *Logistic Regression*



Fonte: Elaborado pelas autoras (2023).

Moderado e Severo obtiveram lacunas maiores em seus percentuais, isso se dá ao fato de confusões que seriam capazes de acontecer durante a análise visual dos avaliadores sobre uma doença, visto que as folhas com contaminação moderada eram visualmente semelhantes as folhas severas, a confusão pode ser influenciada pela cor, forma e número das lesões segundo Kranz (1988). Para o modelo a classificação do grupo de folhas severa foi de 100%, confirmando que o algoritmo foi capaz de classificar corretamente, todas as folhas que realmente eram severas.

4 CONCLUSÃO

Os algoritmos de *machine learning* testados nessa pesquisa obtiveram resultados satisfatórios em suas métricas, contudo o modelo *Logistic Regression* obteve melhor desempenho na classificação da severidade da ferrugem no cafeeiro, tornando-se assim o modelo mais indicado por essa pesquisa.

Outras aplicações possíveis para essa pesquisa são a identificação de níveis de severidade em outras doenças do cafeeiro que atingem o nosso país e até mesmo de outras culturas importantes para o desenvolvimento econômico, o trabalho aplicado em maiores áreas de cultivo pode auxiliar a determinação de epidemia nas lavouras das culturas.

REFERÊNCIAS

ALVES, S. A. M.; NUNES, C. C. **Metodologia para elaboração de escalas diagramáticas para avaliação de doenças em plantas**. Embrapa, 2012.

Disponível em:

<<https://ainfo.cnptia.embrapa.br/digital/bitstream/item/65040/1/Comunicado-Tecnico-120.pdf>> Acesso em: 13 de maio de 2023.

BEN; DAVID. **Remove Fundos de Imagens**. Removebg, Kaleido, 2018. Disponível em: <<https://www.remove.bg/pt-br>> Acesso em: 17 de março de 2023.

BUONOCORE, D. et al. **Classification of coffee bean varieties based on a deep learning approach**. 2022. 18th IMEKO TC10 Conference. University of Salerno, Fisciano (SA). Disponível em: <<https://www.imeko.org/publications/tc10-2022/IMEKO-TC10-2022-002.pdf>> Acesso em: 08 de maio de 2023.

CABRAL, P. G. C., et al. **Genetic diversity and structure of *Hemileia vastatrix* populations on *Coffea* spp.** Plant Pathology 65,196–204, 2016. Disponível em: <<https://bsppjournals.onlinelibrary.wiley.com/doi/epdf/10.1111/ppa.12411>> Acesso em: 11 de maio de 2023.

CAPUCHO, A. S.. **Epidemiologia e Resistência do Cafeeiro Conilon à Ferrugem**. 2011. Tese (Doutorado) – Universidade Federal de Viçosa, Viçosa, 2011. Disponível em:

<<http://www.sbicafe.ufv.br/bitstream/handle/123456789/415/texto%20completo.pdf?squence=1>> Acesso em: 23 de março de 2023.

CASTRO, G. D. M. **MODELO PARA MONITORAMENTO REMOTO DA FERRUGEM DO CAFEIEIRO UTILIZANDO MACHINE LEARNING**. Universidade Federal de Viçosa, 2022. Disponível em:

<<https://www.locus.ufv.br/bitstream/123456789/30489/1/texto%20completo.pdf>> Acesso em: 11 de maio de 2023.

CONAB. **Safra de café deve encerrar o ciclo de 2022 com uma produção de 50,92 milhões de sacas**. Companhia Nacional de Abastecimento (Conab), 2022. Disponível em: <<https://www.conab.gov.br/ultimas-noticias/4857-safra-de-cafe-deve-encerrar-o-ciclo-de-2022-com-uma-producao-de-50-92-milhoes-de-sacas#:~:text=e%20Promo%C3%A7%C3%A3o%20Institucional-,Safra%20de%20caf%C3%A9%20deve%20encerrar%20o%20ciclo%20de%202022%20com,50%2C92%20milh%C3%B5es%20de%20sacas&text=A%20produ%C3%A7%C3%A3o%20cafeeira%20do%20Brasil,7%25%20acima%20da%20safra%202021>>. Acesso em: 13 de maio de 2023.

CONAB. **Primeiro levantamento da safra 2023 de café indica uma produção de 54,94 milhões de sacas.** Companhia Nacional de Abastecimento (Conab), 2023. Disponível em: <<https://www.conab.gov.br/ultimas-noticias/4890-primeiro-levantamento-da-safra-2023-de-cafe-indica-uma-producao-de-54-94-milhoes-de-sacas>> Acesso em: 10 de maio de 2023.

DAMASCENO, V. F. F. et al. **Comparação de dois métodos de elaboração e validação de escala diagramática para a quantificação da severidade da mancha de *Cylindrocladium* em eucalipto.** Summa Phytopathol., Botucatu, v. 40, n. 3, p. 248-255, 2014. Disponível em <<https://repositorio.unesp.br/bitstream/handle/11449/171702/S0100-54052014000300006.pdf?sequence=1&isAllowed=y>> Acesso em: 06 de maio de 2023.

DEMSAR J. et al. **Orange: Data Mining Toolbox in Python,** Journal of Machine Learning Research. Bioinformatics Lab, 2013. Disponível em: <<https://orangedatamining.com/download/#windows>> Acesso em: 23 de abril de 2023.

GESCHKE; WARNOCK. **Ferramenta gratuita para remover fundo de imagens.** Adobe, 2021. Disponível em: <<https://www.adobe.com/br/express/feature/image/remove-background>> Acesso em: 27 de março de 2023.

KAGGLE. **Coffee Leaves Diseases.** Kaggle. Disponível em: <<https://www.kaggle.com/datasets/cofeedisease/coffee-leaves-disease>> Acesso em: 03 de março de 2023.

KRANZ, J. **Measuring plant disease.** In: KRANZ, J.; ROTEM, J. Experimental techniques in plant disease epidemiology. Heidelberg: Springer-Verlag, 1988. p.35-50. Disponível em: <https://link.springer.com/chapter/10.1007/978-3-642-95534-1_4> Acesso em: 09 de maio de 2023.

LENZ, G. et al. **Escala diagramática para avaliação de severidade de mancha-parda em arroz.** Ciência Rural, vol. 40, núm. 4, abril, pp. 752-758, 2010. Disponível em: <<https://www.redalyc.org/pdf/331/33117312035.pdf>> Acesso em: 09 de maio de 2023.

MEIRA, C. A. A.; RODRIGUES, L. H. A. **Modelos em árvore de decisão para alerta da ferrugem do cafeeiro.** VI Simpósio de Pesquisa dos Cafés do Brasil, 2009. Disponível em: <<http://www.sbicafe.ufv.br/bitstream/handle/123456789/2748/55.pdf?sequence=1&isAllowed=y>> Acesso em: 10 de maio de 2023.

MONARD, M. C.; BARANAUSKAS, J. A. **Conceitos sobre aprendizado de máquina.** v. 1, n. 1, 2003. Disponível em: <<https://dcm.ffclrp.usp.br/~augusto/publications/2003-sistemas-inteligentes-cap4.pdf>> Acesso em 11 de maio de 2023.

NUNES, C. C.; ALVES, S. A. M. **Elaboração e validação de escala diagramática para quantificação da severidade de entomosporiose em folhas de pereira.** Summa Phytopathologica, Botucatu, v. 38, n. 3, p. 239, 2012. Disponível em: <<https://www.embrapa.br/en/busca-de-publicacoes/ /publicacao/943965/elaboracao-e-validacao-de-escala-diagramatica-para-quantificacao-da-severidade-de-entomosporiose-em-folhas-de-pereira>> Acesso em: 09 de maio de 2023.

OLIVEIRA, C. T. **MONITORAMENTO DA MATUREÇÃO DOS FRUTOS E DE DOENÇAS DO CAFEEIRO UTILIZANDO MODELOS DE DEEP LEARNING.** Universidade Federal de Viçosa, 2022. Disponível em: <<https://www.locus.ufv.br/bitstream/123456789/29933/1/texto%20completo.pdf>> Acesso em: 12 de maio de 2023.

PERKINS, et al. **Que tipo de design você vai criar hoje?** Canva, 2012. Disponível em: <https://www.canva.com/pt_br/> Acesso em: 28 de março de 2023.

RASBAND, W.S., et al. **ImageJ**, U. S. National Institutes of Health, Bethesda, Maryland, USA, 1997-2018. Disponível em <<https://imagej.nih.gov/ij/download.html>> Acesso em: 04 de abril de 2023.

SEVANDERSON. **Remover imagem de fundo.** Remove bg, Pixlr, 2008. Disponível em <<https://pixlr.com/br/remove-background/>> Acesso em: 27 de março de 2023.

SILVA, L. F. **ANÁLISE DE ELEMENTOS TRANSPONÍVEIS EM *Hemileia vastatrix* E *Coffea spp.*** Universidade Federal de Viçosa, 2022. Disponível em: <<https://www.locus.ufv.br/bitstream/123456789/30735/1/texto%20completo.pdf>> Acesso em: 11 de maio de 2023.

SUSSEL, A. A. B., et al. **Escala Diagramática para Avaliação do Mofo-cinzento-da-mamoneira.** Embrapa Cerrados, 2009. Disponível em: <<https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/659499/1/doc247.pdf>> Acesso em: 12 de maio de 2023.

SZEGEDY, C. et al. **Rethinking the Inception Architecture for Computer Vision.** ArXiv – Computer Science. arXiv:1512.00567v3, Cornell University, 2015. Disponível em: < <https://arxiv.org/pdf/1512.00567.pdf> > Acesso em: 06 de maio de 2023.

YANG, X.; GUO, T. **Machine learning in plant disease research.** Biomedic Research Journal, v. 34, 2019. Disponível em: <<https://biomedicaljour.com/pdfs/volume-34/8.pdf>> Acesso em: 12 de maio de 2023.

ANEXOS

Anexo A – Termo de Consentimento Livre e Esclarecido

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Título do projeto de pesquisa: Avaliação de modelo de machine learning para classificação da severidade da ferrugem (*Hemileia vastatrix*) em cafeeiro (*Coffea sp.*).

Pesquisadores responsáveis: Elizabeth Cristina da Silva Tavares; João Ricardo Favan; Renata Bruna dos Santos Coscolin Favan; Tamara Vitória Alves de Oliveira.

Nome do participante:

Data de nascimento:

Você está sendo convidado(a) para ser participante do projeto de pesquisa intitulado “Avaliação de modelo de machine learning para classificação da severidade da ferrugem (*Hemileia vastatrix*) em cafeeiro (*Coffea sp.*)” de responsabilidade das pesquisadoras Elizabeth Cristina da Silva Tavares e Tamara Vitória Alves de Oliveira. Leia cuidadosamente o que se segue e pergunte sobre qualquer dúvida que você tiver. Caso se sinta esclarecido(a) sobre as informações presentes neste termo e aceite fazer parte do estudo, pedimos que assine ao final deste documento, em duas vias, sendo uma via sua e a outra dos pesquisadores responsáveis pela pesquisa. Saiba que você tem total direito de não querer participar.

1. O estudo tem por finalidade fazer a classificação da severidade da doença causada pela ferrugem (*Hemileia vastatrix*) na cultura do café (*Coffea sp.*), diferenciando a lesão foliar com 4 níveis de severidade sendo eles: “Controle”, “Leve”, “Moderado” ou “Severo”.
2. A participação nesta pesquisa consistirá em avaliar as imagens individualmente e classificá-las de acordo com o percentual da área lesionada da folha, utilizando uma escala diagramática. O tempo estimado dessa pesquisa é de aproximadamente 30 minutos, não será necessário retornar; a pesquisa acontecerá no endereço: Av. Shunji Nishimura, 605 – Pompéia, CEP: 17580-000; FATEC – Fundação Shunji Nishimura; Sala do 6º termo (D-01) - Bloco D.
3. Os participantes não terão nenhuma despesa ao participar da pesquisa e poderão retirar sua concordância na continuidade da mesma a qualquer momento.

4. O nome dos participantes será mantido em sigilo, assegurando assim a sua privacidade, e se desejarem terão livre acesso a todas as informações e esclarecimentos adicionais sobre o estudo e suas consequências, enfim, tudo o que queiram saber antes, durante e depois da sua participação.
5. Os dados coletados serão utilizados única e exclusivamente para fins desta pesquisa, e os resultados serão publicados.

Qualquer dúvida, pedimos a gentileza de entrar em contato com Elizabeth Cristina da Silva Tavares e Tamara Vitória Alves de Oliveira, pesquisadoras responsáveis, telefone: (14)99803-5829, (14)99860-8594, e-mail: elizabeth.tavares@fatec.sp.gov.br, tamara.oliveira3@fatec.sp.gov.br.

Eu, _____, declaro ter sido informado e concordo em ser participante do projeto de pesquisa acima descrito.

Pompeia _____ de _____ de 2023.

Assinatura do participante

Elizabeth Cristina da Silva Tavares

Tamara Vitória Alves de Oliveira