

# DESENVOLVIMENTO E VALIDAÇÃO DE MODELO DE MACHINE LEARNING PARA AVALIAÇÃO DA SEVERIDADE DA MANCHA PRETA (*Cercosporidium personatum*) NO AMENDOIM (*Arachis hypogaea*)

Lucas Finoti Rodrigues<sup>1</sup>  
João Ricardo Favan<sup>2</sup>  
Eloíza Martins Primo Capeloci<sup>2</sup>

## RESUMO

A mancha preta, causada pelo fungo *Cercosporidium personatum*, é uma das principais doenças que comprometem a produção de amendoim (*Arachis hypogaea*) no Brasil. A avaliação visual tradicional da severidade da doença é frequentemente subjetiva e imprecisa, o que dificulta o manejo adequado da cultura. Este estudo teve como objetivo desenvolver e validar um modelo de aprendizado de máquina, utilizando a arquitetura ResNet50, para automatizar a classificação dos níveis de severidade da mancha preta em folhas de amendoim. Foram utilizadas 380 imagens processadas e categorizadas conforme uma escala diagramática aprimorada para a presente pesquisa. O modelo foi treinado com o uso de aprendizado por transferência e testado ao longo de 100 execuções, considerando métricas como acurácia, precisão, *recall* e *F1-score*. A performance do modelo foi comparada à de 14 avaliadores humanos não treinados, demonstrando superioridade em quatro das seis classes de severidade. A acurácia média do modelo atingiu 73%, frente a 44% dos avaliadores, evidenciando sua consistência e confiabilidade. A análise estatística confirmou a estabilidade dos resultados, especialmente nas classes “sadia” e “severo”. Conclui-se que o uso de redes neurais profundas apresenta desempenho superior às avaliações humanas para diagnóstico em fitopatologia, possibilitando avaliações mais rápidas e precisas, e reforçando o potencial da inteligência artificial no apoio à Agricultura de Precisão.

**Palavras-chave:** Aprendizado de máquina; Redes neurais; Classificação de imagens; Doença foliar; Epidemiologia;

## INTRODUÇÃO

O amendoim (*Arachis hypogaea*) é uma cultura nativa da América do Sul e possui grande relevância econômica, especialmente no Brasil, onde o estado de São

---

<sup>1</sup> Discente em Big Data no Agronegócio na FATEC Pompeia, Pompeia-SP,

<sup>2</sup> Docentes do curso Big Data no Agronegócio, FATEC Pompeia, Pompeia-SP.

Paulo responde por cerca de 90% da produção nacional. Nos últimos anos, a produção brasileira de amendoim tem apresentado crescimento. De acordo com a Companhia Nacional de Abastecimento (Conab), a safra 2022/2023 atingiu cerca de 720 mil toneladas, representando um aumento em relação às 667 mil toneladas da safra 2021/2022. Os dados da safra 23/24 indicam que 733,7 mil toneladas de amendoim foram produzidas no território nacional (Conab, 2024).

De acordo com o Instituto de Economia Agrícola (IEA), as exportações do grão estão relacionadas à dinâmica socioeconômica das regiões onde ele é produzido, bem como dos municípios de São Paulo responsáveis pelo beneficiamento e envio dos produtos ao exterior (Mantovani, 2023). Em 2022, o município de Tupã foi responsável por 28% dos valores exportados de amendoim em grão, seguido por Borborema com 17%. Os municípios de Jaboticabal (7%), Sertãozinho (7%), Pompéia (6%) e Taquaritinga (5%) também participaram das exportações nesse período (IEA, 2023).

As condições climáticas favoráveis do país, como temperaturas médias acima de 20°C, tornam o país ideal para o cultivo da planta. No entanto, essas mesmas condições também favorecem o surgimento de doenças, entre as quais a mancha preta (*Cercosporidium personatum*), que representa uma ameaça significativa à produtividade da cultura (Bisonard *et al.*, 2020).

A mancha preta é uma das principais doenças que afetam a cultura do amendoim, comprometendo tanto a qualidade quanto a quantidade da produção (Bisonard *et al.*, 2020). A avaliação tradicional da severidade da doença é realizada visualmente, muitas vezes por observadores não treinados, o que torna o processo subjetivo, demorado e suscetível a erros (Bock *et al.*, 2010). Diante desse cenário, torna-se essencial buscar soluções mais precisas, padronizadas e eficientes para o diagnóstico e o monitoramento fitossanitário.

Nesse contexto, a aplicação de técnicas com o uso de algoritmo de aprendizado de máquina (*machine learning*) surge como uma alternativa promissora, sendo possível automatizar a detecção e a classificação de doenças em plantas, promovendo análises mais rápidas, precisas e reprodutíveis. Além disso, esses sistemas são capazes de identificar diferentes tipos e estágios de desenvolvimento da doença, o que proporciona um diagnóstico mais detalhado, completo e fundamentado (Yang; Guo, 2019).

Entre os modelos mais utilizados em tarefas de detecção e classificação de imagens, destaca-se a ResNet50, uma arquitetura profunda e amplamente utilizada de Convolutional Neural Network (CNN), ou redes neurais convolucionais, composta por 50 camadas. Essa rede apresenta como principal diferencial o uso do aprendizado residual, uma técnica que facilita o treinamento de redes profundas ao abordar o problema do desvanecimento do gradiente. Esse fenômeno ocorre quando, durante o treinamento de redes neurais muito extensas, os valores dos gradientes responsáveis por ajustar os pesos da rede, tornam-se extremamente pequenos à medida que são propagados retroativamente pelas camadas (Ikechukwu, 2021). Como consequência, as camadas iniciais aprendem de forma lenta ou praticamente deixam de aprender, o que compromete a otimização do modelo como um todo. Neste trabalho, a ResNet50 foi empregada na detecção de pragas em plantas, demonstrando sua eficácia e versatilidade nesse tipo de aplicação.

As chamadas conexões de salto (*skip connections*) permitem que a saída de uma camada seja somada à saída de uma camada mais profunda ou subsequente, favorecendo a preservação das informações relevantes ao longo da rede (Theckedath, 2020). Além disso, o uso da técnica de aprendizado por transferência (*transfer learning*), que, segundo Ikechukwu (2021), trabalha com pesos pré-treinados em grandes bases de dados como o *ImageNet*, o que torna o treinamento mais eficiente mesmo em conjuntos de dados reduzidos.

As redes neurais profundas têm se destacado em diversas aplicações de visão computacional, especialmente na classificação de imagens. Entre os componentes amplamente utilizados nesse tipo de rede, destaca-se a função de ativação *ReLU* (*Rectified Linear Unit*), que introduz não linearidade ao modelo ao manter valores positivos e zerar os negativos, facilitando o aprendizado de padrões complexos. Para evitar o risco de *overfitting*, é comum a aplicação da técnica de *Dropout*, que desativa aleatoriamente parte dos neurônios durante o treinamento, promovendo uma generalização mais robusta. Em tarefas de classificação multiclasse, a função de ativação *softmax* é frequentemente empregada na camada de saída, pois transforma os valores da rede em uma distribuição de probabilidades, facilitando a identificação da classe mais provável.

A compilação de modelos desse tipo geralmente envolve o uso do otimizador

*Adam*, que combina estratégias de momento e adaptação da taxa de aprendizado, promovendo ajustes eficientes nos pesos da rede. Para mensurar a performance em classificações multiclasse, utiliza-se comumente a função de perda *categorical crossentropy*, que avalia a divergência entre as distribuições de probabilidade previstas e reais (Geron, 2019).

Este trabalho propõe tornar mais preciso o diagnóstico e acelerar o processo de monitoramento da mancha preta no amendoim, viabilizando ações mais rápidas e eficazes no manejo da cultura. Com isso, espera-se minimizar os impactos negativos da mancha preta na produção de amendoim e contribuir para o desenvolvimento de estratégias de controle mais eficientes. A aplicação de inteligência artificial na fitopatologia permite que pesquisadores, biólogos e agrônomos realizem análises de dados com maior acurácia e precisão, promovendo diagnósticos mais confiáveis e consistentes.

A presente pesquisa tem como objetivo desenvolver e validar um modelo de aprendizado de máquina baseado na arquitetura ResNet50, com a finalidade de classificar imagens de folhas de amendoim quanto à severidade da mancha preta e envolve a comparação do desempenho do modelo em relação a avaliadores humanos não treinados, buscando evidenciar seu potencial prático. Espera-se que a solução proposta contribua para diagnósticos mais rápidos e precisos, auxiliando na tomada de decisões e no manejo eficaz da cultura.

## **MATERIAL E MÉTODOS**

Este estudo utilizou um conjunto de imagens de folhas de amendoim em diferentes níveis de severidade da mancha preta, originalmente capturadas com câmeras de celulares e contendo fundos naturais. Devido à baixa qualidade e variação nas condições de iluminação, as imagens passaram por etapas de pré-processamento, como seleção, tratamento e padronização, seguindo um protocolo específico. Além disso, utilizou-se uma escala diagramática para auxiliar na avaliação da severidade da doença foliar. (Belasque Júnior *et al.*, 2005).

Foi definido como conjunto de dados para o desenvolvimento da pesquisa, o *dataset* “*Groundnut Plant Leaf Data*”, obtido na plataforma *Kaggle* (2023). O conjunto

original continha 13.420 imagens de folhas de amendoim distribuídas em dois diretórios: “*Raw\_Data*” e “*Groundnut\_Leaf\_dataset*”. O diretório “*Raw\_Data*” continha as imagens da planta do amendoim, onde possuía fundos variados, já o diretório “*Groundnut\_Leaf\_dataset*” possuía imagens da folha do amendoim.

Os arquivos da pasta “*Raw\_Data*”, bem como o arquivo “*Metadata.xlsx*”, foram descartados. O arquivo “*Metadata.xlsx*” continha apenas informações redundantes, como identificadores das imagens, descrições genéricas mostrando se a imagem era de treino ou teste e caminhos dos diretórios correspondentes. Essas informações já estavam implícitas na estrutura das pastas e nomes dos arquivos, tornando o arquivo desnecessário para a análise.

O diretório “*Raw\_Data*” foi excluído devido à baixa qualidade e inconsistência das imagens contidas nele. As imagens apresentavam fundos variados e elementos indesejados, como mãos humanas, solo e partes não relevantes da planta de amendoim, o que introduz ruído significativo nos dados. A presença de fundos ruidosos em imagens de folhas, como demonstrado por KC *et al.* (2021), pode comprometer significativamente a acurácia de classificação de redes neurais convolucionais, sendo recomendada a remoção de fundo para melhorar o desempenho dos modelos de aprendizado profundo. Dito isso, a pesquisa concentrou-se exclusivamente nas imagens do diretório “*Groundnut\_Leaf\_dataset*”, porque o conjunto apresenta imagens padronizadas, com foco nas folhas de amendoim, o que é essencial para garantir a qualidade dos dados de entrada. (KC *et al.*, 2021).

Dentro deste conjunto, foram selecionadas apenas imagens pertencentes às classes “folhas saudáveis”, “mancha foliar precoce” e “mancha tardia da folha”. As demais categorias, foram excluídas, sendo elas “ferrugem precoce”, “deficiência nutricional” e “ferrugem”. Para garantir maior uniformidade e qualidade na base, foram escolhidas 380 imagens que mostravam as faces adaxial e abaxial das folhas.

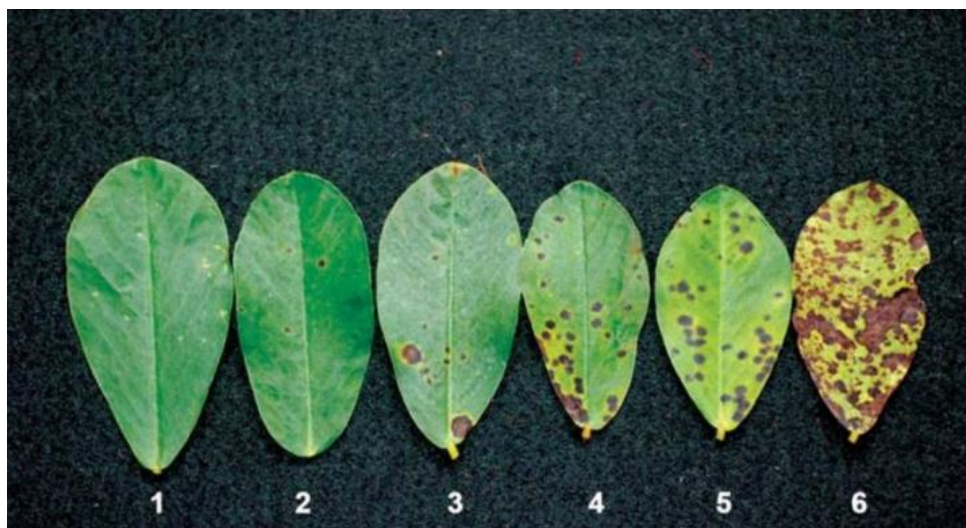
As imagens foram processadas para melhorar a qualidade visual e facilitar a análise. Inicialmente, cada imagem foi segmentada para destacar apenas a folha principal, com remoção completa do fundo. Esse processo foi realizado utilizando as ferramentas Adobe Photoshop Versão 21.0 (Adobe, 2025) e Remove.bg (Ben; David, 2018).

Com o objetivo de quantificar a severidade da mancha preta (*Cercosporidium personatum*), foi desenvolvido um protocolo, “Protocolo para identificação de área lesionada de mancha preta (*Cercosporidium personatum*) no amendoim (*Arachis hypogaea*)”, específico utilizando o *software* ImageJ (Schneider *et al.*, 2012). Por meio dele, foram calculadas, em pixel, a área total da folha e a área lesionada. A severidade da doença foi determinada pela razão entre a área lesionada e a área total, resultando no percentual de lesão por imagem.

A escala diagramática mostrada na Figura 1, proposta por Barreto (2007), foi adequada para esta pesquisa. A escala apresenta seis níveis: “sadia”, “muito leve”, “leve”, “moderado”, “severo” e “muito severo”. A adequação consistiu na definição objetiva dos intervalos percentuais de lesão correspondentes a cada classe, ou seja, foi estabelecido o percentual máximo para que uma folha fosse enquadrada na classe “sadia”, bem como os limites a partir dos quais a folha passava a se enquadrar em cada uma das demais categorias de severidade.

Para fins de validação da escala, um grupo de 14 avaliadores não treinados classificou aleatoriamente 30 imagens selecionadas, com base na representação visual da escala, realizando a discriminação visual conforme a região afetada da folha.

**Figura 1** - Escala diagramática para a doença da mancha preta em folhas de amendoim com 6 níveis de intensidade



Fonte: Barreto, 2007

A classificação automatizada das imagens foi realizada por meio de uma rede neural artificial (ANN, *artificial neural network*), implementada com a biblioteca *Keras* (Keras, 2025), utilizando a linguagem *Python* 3.11.6 (Python, 2024). O treinamento e a avaliação ocorreram em um computador com processador Intel i5 de 10ª geração e 8 GB de RAM. A arquitetura adotada foi baseada na rede pré-treinada ResNet50, uma rede residual com 50 camadas, que, segundo Theckedath *et al.* (2020), utiliza mapeamentos de identidade para facilitar o aprendizado em redes profundas.

Os rótulos das classes foram convertidos em valores numéricos por meio de *Label Encoding*, uma técnica utilizada para transformar variáveis categóricas em valores inteiros, conforme implementado por Pedregosa *et al.* (2011) na biblioteca *Scikit-learn*. Em seguida, foi aplicada a técnica de *One-Hot Encoding*, que consiste em representar variáveis categóricas como vetores binários, facilitando o processamento por modelos de aprendizado de máquina (Geron, 2019). Para sua implementação, utilizou-se a funcionalidade correspondente da biblioteca *Scikit-Learn* (Pedregosa *et al.*, 2011).

A estrutura da rede foi composta por uma camada densa com 512 neurônios e função de ativação *ReLU*, que introduz não linearidades ao modelo ao manter os valores positivos e zerar os negativos, seguida por uma camada de *Dropout* com taxa de 50% e uma camada de saída com número de neurônios correspondente ao número de classes, utilizando a função de ativação *softmax* (Theckedath, 2020). A compilação do modelo foi realizada com o otimizador Adam e a função de perda *categorical\_crossentropy*, sendo a acurácia adotada como principal métrica de avaliação.

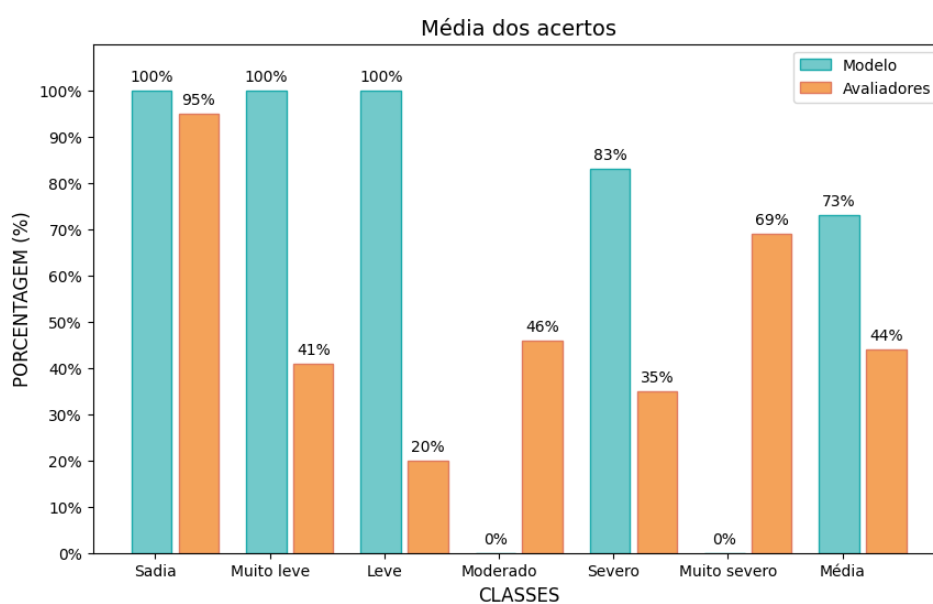
Para determinar o número ideal de épocas de treinamento, aplicou-se o método *Elbow* (Syakur, 2018), cujo ponto de inflexão indicou que o treinamento ideal ocorreria por volta de 40 épocas. Para garantir robustez, o modelo foi treinado 100 vezes, com embaralhamento aleatório das imagens a cada execução, otimizando a generalização. Ao final de cada execução, foi gerado um relatório de classificação contendo as métricas *precision*, *recall* e *F1-score* para cada classe, sendo os resultados armazenados em arquivos CSV.

Foi realizada a estatística descritiva dos dados através do software, por fim, os resultados do modelo foram comparados à classificação feita pelos 14 avaliadores humanos, utilizando a escala diagramática como referência.

## RESULTADOS E DISCUSSÃO

A avaliação realizada por 14 participantes não treinados, com base na escala diagramática proposta por Barreto (2007), permitiu obter uma média de acertos por classe conforme segue: “sadia” (95%), “muito leve” (41%), “leve” (20%), “moderado” (46%), “severo” (35%) e “muito severo” (69%). Esses resultados refletem a dificuldade dos avaliadores em identificar corretamente os níveis intermediários de severidade, o que pode ser atribuído à sutil diferença visual entre as categorias adjacentes (Costa, 2024). Resultados semelhantes foram observados por Nutter Jr. *et al* (2006), que relataram alta variabilidade entre avaliadores em tarefas de estimativa visual de doenças, mesmo com o auxílio de escalas diagramáticas. As mesmas 30 imagens foram submetidas ao modelo de *machine learning* desenvolvido, resultando nos seguintes índices de acerto por classe: “sadia” (100%), “muito leve” (100%), “leve” (100%), “moderado” (0%), “severo” (83%) e “muito severo” (0%). A comparação entre os resultados, foi demonstrado na Figura 2, a média geral de acertos foi de 44% para os avaliadores e 73% para o modelo.

**Figura 2** – Média dos acertos dos avaliadores em comparação ao modelo



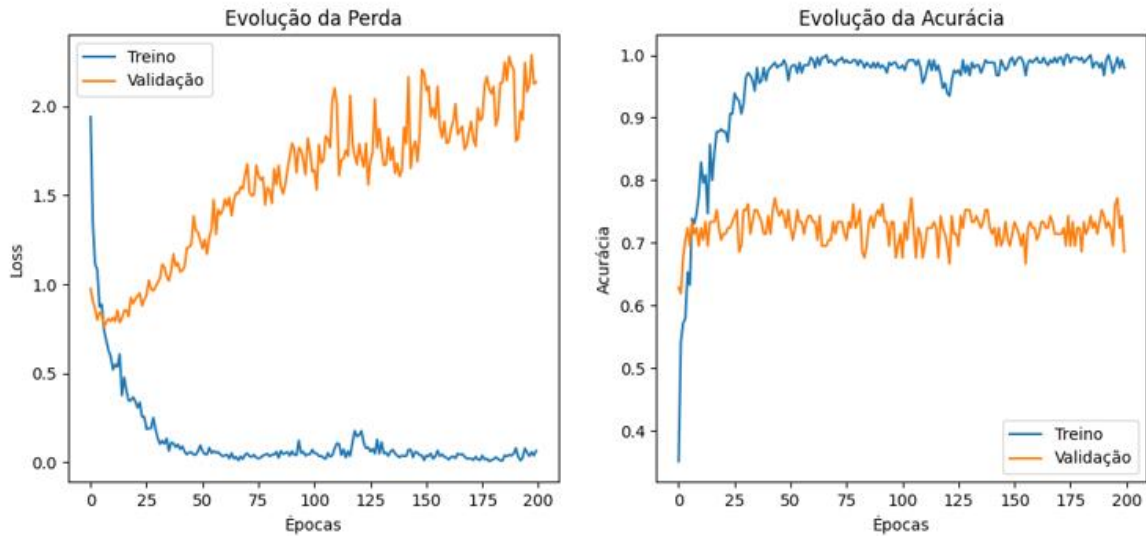
Fonte: Elaborado pelo autor (2025)

Observa-se que o modelo apresentou desempenho superior ao dos avaliadores em quatro das seis classes. Contudo, a ausência de acertos nas classes “moderado” e “muito severo” revela um viés do modelo. Segundo Buda *et al.* (2018), modelos de redes neurais apresentam desempenho inferior em classes com menor frequência durante o treinamento, sendo fortemente impactados por desbalanceamentos. Isso evidencia a importância do balanceamento das classes, prática recomendada para melhorar a performance em tarefas de classificação com conjuntos de dados assimétricos (He *et al.*, 2009).

As classes iniciais e finais do espectro de severidade “sadia” e “muito severo” apresentaram os melhores índices de acerto, tanto pelo modelo quanto pelos avaliadores. Isso ocorre porque essas classes são visualmente mais fáceis de distinguir. Já as classes intermediárias, segundo Costa (2024), apresentam diferenças mais sutis, o que gera maior incerteza e confusão na avaliação.

A Figura 3 ilustra a evolução da acurácia e da perda ao longo das épocas, com base no método *Elbow*. Observa-se que a curva de acurácia tende a se estabilizar em torno da 40ª época, confirmando a escolha do ponto de inflexão como referência para o número ideal de épocas.

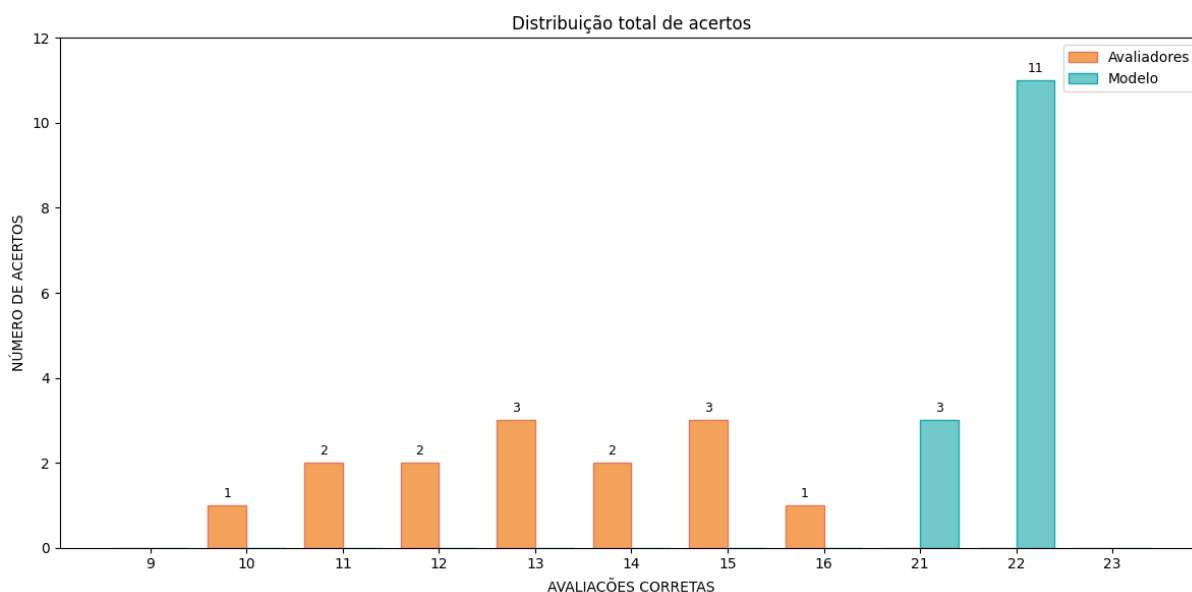
**Figura 3** – Evolução da perda e da acurácia com o método *Elbow*



Fonte: Elaborado pelo autor (2025)

A estabilização da acurácia por volta da 40<sup>a</sup> época sugere que o modelo atingiu um ponto ótimo de aprendizado. Esse comportamento é indicativo de um treinamento eficaz, no qual o modelo conseguiu capturar os padrões relevantes dos dados sem memorizar ruídos ou variações irrelevantes. A consistência observada nas curvas de perda e acurácia evidencia que os parâmetros escolhidos para o treinamento, como a arquitetura da rede, o número de épocas e o método de otimização, foram adequados. A Figura 4 ilustra o comparativo, destacando a superioridade do modelo frente à variabilidade dos resultados humanos.

**Figura 4** – Distribuição total de acertos por avaliador em comparação ao modelo



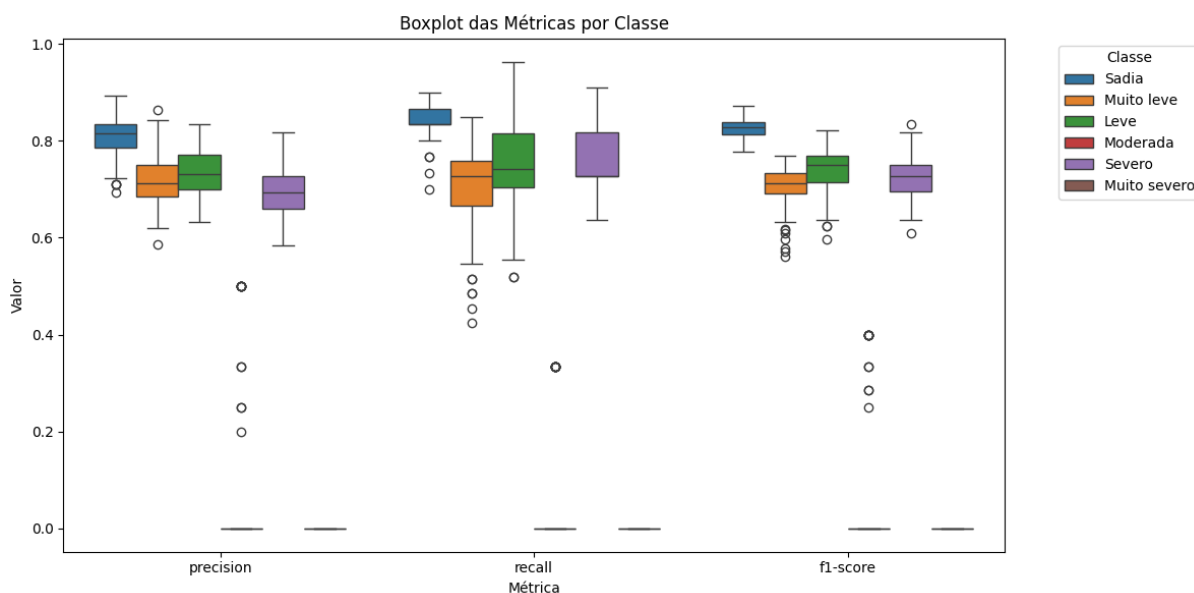
Fonte: Elaborado pelo autor (2025)

Nenhum dos avaliadores atingiu acerto total, sendo que apenas um avaliador ultrapassou a marca de 50% de acertos, com 16 classificações corretas entre as 30 imagens. Em contraste, o modelo de *machine learning* apresentou um desempenho significativamente superior, concentrando-se nas faixas de 21 e 22 acertos, com até 73% de precisão total em algumas execuções.

Esses dados reforçam tanto a complexidade da tarefa de classificação visual da severidade da mancha preta por indivíduos sem treinamento específico quanto o potencial de sistemas automatizados nesse contexto. A variabilidade entre os avaliadores humanos evidencia a subjetividade do processo quando feito de forma manual, mesmo com o auxílio de uma escala diagramática. Segundo Nutter Jr. *et al* (2006), a variabilidade entre avaliadores é um fator limitante na estimativa da severidade de doenças foliares, especialmente em avaliações visuais subjetivas.

A Figura 5 apresenta os *boxplots* das métricas de *precision*, *recall* e *F1-score* para cada uma das classes, considerando as 100 execuções do modelo treinado por 40 épocas. Observa-se, que a classe “Sadia” apresenta não apenas os maiores valores médios, mas também menor variabilidade, evidenciando estabilidade nos resultados.

**Figura 5 – Boxplot das métricas por classe**



Fonte: Elaborado pelo autor (2025)

Apesar disso, a escala proposta se mostrou uma ferramenta importante, contribuindo para a padronização das avaliações e redução da subjetividade, especialmente para observadores sem experiência na quantificação de doenças. Conforme afirmam Lenz *et al.* (2010), a utilização de escalas diagramáticas melhora a precisão e a reprodutibilidade das estimativas de severidade.

A superioridade média do modelo em relação aos avaliadores humanos, aliada à sua consistência, reforça o potencial de aplicação de técnicas de aprendizado de máquina para auxiliar na detecção e quantificação de doenças foliares em amendoim. No entanto, os resultados também evidenciam a necessidade de ampliar e balancear o conjunto de dados para melhorar o desempenho em classes de sub-representação.

## CONSIDERAÇÕES FINAIS

O algoritmo avaliado neste estudo apresentou resultados satisfatórios nas métricas analisadas, indicando que os objetivos da pesquisa foram alcançados com sucesso. O modelo com a arquitetura pré-treinada ResNet50 destacou-se por apresentar consistência nos acertos e baixa variabilidade entre as execuções, especialmente nas classes “sadia”, “leve” e “severo”. Por esse motivo, este modelo é

recomendado, embora haja limitações na pesquisa, como o desequilíbrio entre as classes no conjunto de dados, ou seja, a quantidade desigual de imagens por classe.

Como proposta de trabalhos futuros, sugere-se a implementação de diferentes algoritmos para realizar a mesma tarefa, possibilitando uma análise comparativa entre seus desempenhos.

## REFERÊNCIAS

ADOBE SYSTEMS INCORPORATED. **Adobe Photoshop**: *software* de edição de imagens. Versão 21.0.2025. San Jose, CA: Adobe, 2020. Disponível em: <https://www.adobe.com/br/products/photoshop/landpa.html>. Acesso em: 10 mar. 2025.

BARRETO, M. **Manual de identificação e manejo das doenças do amendoim**. Jaboticabal: 1º ed. FUNEP, 2007. p. 1-5, 32-33.

BELASQUE JÚNIOR, José; BASSANEZI, Renato B.; SPÓSITO, Marcel B.; RIBEIRO, Luciane M.; JESUS JUNIOR, Waldir C. de; AMORIM, Lilian. 2005. **Escalas diagramáticas para avaliação da severidade do cancro cítrico**. Disponível em: <<https://www.scielo.br/j/fb/a/4KMvjCTvBxwTyzS388m4Rpx>>. Acesso em: 30 mar. 2025.

BEN; DAVID. **Remove Fundos de Imagens**. Removebg, Kaleido, 2018. Disponível em: <<https://www.remove.bg/pt-br>>. Acesso em: 01 mar. 2025.

BISONARD, Eduardo Matias; HAMADA, Emília; ANGELOTTI, Francislene; GONÇALVES, Renata Ribeiro do Valle; RAGO, Alejandro Mario. Evolução da mancha preta do amendoim nas principais regiões produtoras da Argentina e do Brasil frente às mudanças no clima. **Revista Brasileira de Geografia Física**, v. 13, n. 04, p. 1778-1791, 2020.

BOCK, Cheryl H.; POOLE, Gregory H.; PARKER, Paul E.; GOTTWALD, Timothy R. (2010). Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging. **Critical Reviews in Plant Sciences**, 29(2), 59–107. Disponível em <<https://www.tandfonline.com/doi/abs/10.1080/07352681003617285>>. Acesso em: 03 mai. 2025.

BUDA, Matheusz; MAKI, Atsuto; MAZUROWSKI, Maciej A. (2018). **A systematic study of the class imbalance problem in convolutional neural networks**. *Neural Networks*, 106, 249–259. Disponível em: <<https://doi.org/10.1016/j.neunet.2018.07.011>>. Acesso em: 06 mai. 2025.

CONAB. Tabela de dados - **Produção e balanço de oferta e demanda de grãos**. Conab 2024. Disponível em: <<https://www.conab.gov.br/info-agro/safras/graos>>. Acesso em: 07 mar. 2025.

COSTA, Vitor de Souza. **Classificação de níveis de severidade da mancha de Cercospora em milho por meio de dados hiperespectrais e aprendizagem de máquina** (Dissertação de mestrado). Universidade Federal de Mato Grosso do Sul, Campus de Chapadão do Sul, Chapadão do Sul, MS. 2024. Disponível em: <https://repositorio.ufms.br/retrieve/00ef2161-9d78-4480-b918-87cf3ae209b6/Disserta%c3%a7%c3%a3o%20Vitor%20de%20Souza%20Costa.pdf>. Acesso em: 14 abr. 2025.

GERON, Aurélien. **Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems**. 2. ed. Sebastopol: O'Reilly Media, 2019. 819 p. ISBN 978-1-492-03264-9.

HE, Haibo; GARCIA, Eduardo A. (2009). **Learning from Imbalanced Data**. IEEE Transactions on Knowledge and Data Engineering, 21(9), 1263–1284. Disponível em: <<https://doi.org/10.1109/TKDE.2008.239>>. Acesso em: 06 mai. 2025.

IEA. **Instituto de Economia Agrícola**. Amendoim: em 2022, as exportações do grão fortaleceram novos mercados, enquanto as do óleo registraram alta. 2023. Disponível em: <<http://www.iea.agricultura.sp.gov.br/out/TerTexto.php?codTexto=16115>>. Acesso em 05 mai. 2025.

IKECHUKWU, A. Victor; MURALI, S.; DEEPU, R.; SHIVAMURTHY, R. C. ResNet-50 vs VGG-19 vs training from scratch: **A comparative analysis of the segmentation and classification of Pneumonia from chest X-ray images**. Global Transitions Proceedings, v. 2, n. 2, p. 375–381. 2021. Disponível em: <[https://www.sciencedirect.com/science/article/pii/S2666285X21000558?ref=pdf\\_download&fr=RR-2&rr=93071bb4c90bbc95](https://www.sciencedirect.com/science/article/pii/S2666285X21000558?ref=pdf_download&fr=RR-2&rr=93071bb4c90bbc95)>. Acesso em: 12 abr. 2025.

KAGGLE. **Groundnut Plant Leaf Data**. Kaggle. Disponível em: <<https://www.kaggle.com/datasets/warcoder/groundnut-plant-leaf-data?resource=download>>. Acesso em: 20 fev. 2025.

KC, Kamal; YIN, Zhendong; LI, Dasen; WU, Zhilu. **Impacts of Background Removal on Convolutional Neural Networks for Plant Disease Classification In-Situ**. Agriculture, v. 11, n. 827, 2021. Disponível em < <https://www.mdpi.com/2077-0472/11/9/827>>. Acesso em: 04 mai. 2025.

KERAS. **Keras Documentation**. Keras, 2025. Disponível em: <https://keras.io/>. Acesso em: 27 mar. 2025.

LENZ, Giuvan; BALARDINI, Ricardo Silveiro; DALLA CORTE, Gerson; MARQUES, Leandro Nascimento; DEBONA, Daniel. 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: 21 mar. 2025.

MANTOVANI, Leandro. **Relação entre fatores de produção e produtividade: um diagnóstico da cultura do amendoim na região oeste paulista**. 2023. Dissertação (Mestrado em Agronegócio e Desenvolvimento) – Universidade Estadual Paulista “Júlio de Mesquita Filho”, Faculdade de Ciências e Engenharia, Tupã, 2023. Disponível em: <<https://repositorio.unesp.br/server/api/core/bitstreams/9c4865e2-b167-4596-9188-f36f7a876082/content>>. Acesso em: 05 mai. 2025.

NUTTER JR, Forrest W.; ESKER, Paul D. **The role of accuracy, precision, and the power of the test in disease management**. *Plant Disease*, St. Paul, v. 90, n. 5, p. 556–562, 2006.

PEDREGOSA, Fabian; VAROQUAUX, Gaël; GRAMFORT, Alexandre; MICHEL, Vincent. Scikit-learn: Machine Learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011. Disponível em: <<http://jmlr.org/papers/v12/pedregosa11a.html>>. Acesso em: 13 abr. 2025.

PYTHON. **Python Documentation**. Python, 2024. Disponível em: <<https://www.python.org/>>. Acesso em: 27 mar. 2025.

SCHNEIDER, Caroline A.; RASBAND, Wayne S.; ELICEIRI, Kevin W. NIH Image to ImageJ: 25 years of image analysis. **Nature Methods**, v. 9, n. 7, p. 671–675, 2012. Disponível em: <<https://doi.org/10.1038/nmeth.2089>>. Acesso em: 30 abr. 2025.

SYAKUR, Muhammad Adnan; KURNIAWAN, Eko; WIDYANTO, Arie Wahyu; SATOTO, Budi; PRAMONO, Rendra Sigit. **Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster**. *IOP Conf. Series: Materials Science and Engineering*, v. 336, p. 012017, 2018. Disponível em: <https://iopscience.iop.org/article/10.1088/1757-899X/336/1/012017/pdf>. Acesso em: 06 mai. 2025.

THECKEDATH, D., Sedamkar, R.R. Detecting Affect States Using VGG16, ResNet50 and SE-ResNet50 Networks. **SN COMPUT. SCI.** 1, 79 (2020). Disponível em: <<https://doi.org/10.1007/s42979-020-0114-9>>. Acesso em: 01 abr. 2025.

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: 17 fev. 2025.

## **ANEXOS**

**Anexo A** – Termo de Consentimento Livre e Esclarecido

### **TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO**

**Título do projeto de pesquisa:**

Desenvolvimento e validação de modelo de *machine learning* para avaliação da severidade da mancha preta (*Cercosporidium Personatum*) no amendoim (*Arachis hypogaea*)

**Pesquisadores responsáveis:** Lucas Finoti Rodrigues; Prof. Dr. João Ricardo Favan; Prof.<sup>a</sup> Dr<sup>a</sup> Eloíza Martins Primo Capeloci.

**Nome do participante:**

**Data de nascimento:**

Você está sendo convidado(a) para ser participante do projeto de pesquisa intitulado “Desenvolvimento e validação de modelo de *machine learning* para avaliação da severidade da mancha preta (*Cercosporidium Personatum*) no amendoim (*Arachis hypogaea*)” de responsabilidade do pesquisador Lucas Finoti Rodrigues.

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 do pesquisador responsável pela pesquisa. Saiba que a sua participação é opcional.

1. O estudo tem por finalidade fazer a classificação da severidade da doença causada pela mancha preta (*Cercosporidium Personatum*) no amendoim diferenciando a lesão foliar com 6 níveis de severidade sendo eles: “Sadia”, “Muito leve”, “Leve”, “Moderado”, “Severo” e “Muito 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 15 minutos, não será necessário retornar; a pesquisa acontecerá no endereço: Av. Shunji Nishimura, 605 – Pompéia, CEP: 17580-000; FATEC – Fundação Shunji Nishimura; Sala do 5º termo (D-03) - Bloco D no dia 19/03/2025 (quarta-feira).
3. Os participantes não terão nenhuma despesa ou gratificação ao participar da pesquisa e poderão retirar sua concordância na continuidade da mesma a qualquer momento até o termo ser assinado.
4. O nome dos participantes será mantido em sigilo, assegurando assim a sua privacidade, caso desejarem terão livre acesso as metodologias utilizadas no trabalho e seu resultado final.

5. Os dados coletados serão utilizados única e exclusivamente para fins desta pesquisa, e somente os resultados obtidos serão publicados.
6. As informações adquiridas a partir dessa pesquisa terão vigência indeterminada.

Qualquer dúvida, pedimos a gentileza de entrar em contato com Lucas Finoti Rodrigues, pesquisador responsável, telefone: (14)99810-0124, e-mail: lucas.rodrigues129@fatec.sp.gov.br.

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

Pompeia \_\_\_\_ de \_\_\_\_\_ 2025.

---

Assinatura do participante

---

Lucas Finoti Rodrigues