Classificação de pragas e doenças agrícolas por meio de imagens digitais utilizando aprendizado profundo.

Pedro Lucas de Oliveira Costa*, Thiago Matheus de Oliveira Costa*
*Instituto de Ciências Exatas e Tecnológicas
Universidade Federal de Viçosa - UFV, Rio Paranaíba, MG, Brazil
E-mail: {pedro.l.costa, thiago.costa}@ufv.br

***Abstract*—No setor agrícola, o gerenciamento eficiente de pragas é um desafio fundamental na agricultura moderna, pois infestações podem causar perdas econômicas significativas e comprometer a segurança alimentar. Os métodos tradicionais de identificação de pragas geralmente dependem de inspeção manual, um processo demorado e suscetível a erros, uma vez que a classificação de insetos é complexa devido à alta variabilidade das espécies em diferentes regiões e às mudanças ao longo do ciclo de vida. Diante desses desafios e dos avanços da inteligência artificial, redes neurais profundas, especialmente as Redes Neurais Convolucionais (CNNs), surgiram como ferramentas promissoras para a automação da classificação de pragas.**

Neste trabalho, apresentamos a avaliação de modelos de deep learning e estratégias de treinamento para a classificação automática de imagens de pragas. Analisamos duas arquiteturas de CNNs: EfficientNet e ViT-B/16. Os modelos foram ajustados por fine-tuning, e avaliamos o impacto de diferentes estratégias de data augmentation no desempenho da classificação, utilizando o conjunto de dados Crop Pest Disease Detection. O modelo ViT-B/16 obteve o melhor desempenho, alcançando uma precisão de 0.9170 quando treinado com técnicas de aumento de dados.

***Index Terms*—Classificação de pragas, aprendizado profundo, aumento de dados**

I. Introdução

A agricultura é um dos principais motores do desenvolvimento econômico e da segurança alimentar, fornecendo recursos essenciais para a nutrição humana e o emprego rural. No entanto, a produtividade agrícola enfrenta ameaças constantes de fatores bióticos e abióticos, sendo as pragas insetívoras um dos principais responsáveis por perdas na produção em todo o mundo [1]. A detecção precoce e a classificação de pragas são fundamentais para a implementação de medidas de controle eficazes e oportunas, reduzindo seus impactos nas lavouras [2].

Os avanços recentes em inteligência artificial (IA) introduziram técnicas de aprendizado de máquina (ML) e aprendizado profundo (DL) como soluções promissoras para o monitoramento agrícola e a tomada de decisões estratégicas. Essas abordagens possibilitam a análise automatizada de fatores críticos, como qualidade do solo, saúde das plantas e níveis de umidade, permitindo um melhor gerenciamento dos recursos e maior produtividade das culturas [1]. Em especial, os sistemas de reconhecimento baseados em imagens têm se mostrado eficazes na detecção precoce de doenças, deficiências nutricionais e infestações de pragas, permitindo uma intervenção proativa [3].

Os modelos de aprendizado profundo, especialmente as Redes Neurais Convolucionais (CNNs), demonstraram capacidades notáveis em tarefas de classificação de imagens, superando a análise visual humana em velocidade e precisão [4]. Esses modelos são capazes de processar grandes volumes de dados, extraindo características hierárquicas que aumentam a precisão na classificação. No entanto, sua eficácia depende de vários fatores, incluindo o tamanho do conjunto de dados, ajuste de hiperparâmetros, técnicas de aumento de dados e estratégias de aprendizado por transferência [3]. A otimização adequada desses elementos é essencial para garantir modelos de classificação de pragas robustos e confiáveis.

Apesar dos avanços, a classificação de pragas agrícolas continua sendo um grande desafio devido à diversidade morfológica das espécies em diferentes regiões e às mudanças ao longo dos seus estágios de vida [5]. Muitas pragas passam por transformações distintas, desde ovos e larvas até pupas e formas adultas, dificultando sua identificação visual. Além disso, algumas espécies permanecem ocultas dentro das estruturas das plantas, tornando a detecção possível apenas por meio de sinais indiretos de danos. Esses desafios ressaltam a necessidade de técnicas avançadas de aprendizado que melhorem a capacidade de generalização dos modelos em diferentes espécies de pragas e condições ambientais.

Este estudo busca aprimorar a classificação automática de pragas utilizando técnicas de aprendizado profundo, avaliando a eficácia de modelos baseados em CNNs. Especificamente, analisamos o desempenho das arquiteturas EfficientNet e ViT na classificação de imagens de pragas em ambientes agrícolas. Além disso, exploramos o impacto de diferentes estratégias de aumento de dados na precisão da classificação.

Este artigo está estruturado da seguinte forma: a Seção II apresenta os trabalhos relacionados, a Seção III detalha a metodologia e o conjunto de dados utilizados, a Seção IV discute os resultados obtidos, e a Seção V apresenta as conclusões do estudo.

II. Trabalhos Relacionados

Sharma et al. (2023) propuseram um sistema automatizado de identificação de pragas utilizando modelos de aprendizado

profundo treinados no conjunto de dados PestNet. Os autores testaram diversas arquiteturas de CNNs, incluindo ResNet-50 e EfficientNetB3, incorporando técnicas de aumento de dados, como rotação e espelhamento, para melhorar a capacidade de generalização do modelo. Os resultados demonstraram que a EfficientNetB3 superou as CNNs tradicionais, alcançando uma precisão de 91,8

Zhou et al. (2022) apresentaram uma abordagem de aprendizado de conjunto (ensemble learning) para a detecção de pragas, integrando redes neurais convolucionais (CNNs) e modelos baseados em Transformers. O estudo combinou as vantagens de ambas as arquiteturas por meio de técnicas de fusão de características. O modelo de ensemble foi testado no conjunto de dados IP102, onde obteve uma melhoria significativa na precisão da classificação em comparação com modelos individuais, especialmente na detecção de espécies raras de pragas.

Liang et al. (2023) avaliaram o impacto de diferentes técnicas de aumento de dados na classificação de pragas baseada em aprendizado profundo. O estudo analisou métodos padrão de aumento de dados, como CutMix, MixUp e injeção de ruído Gaussiano. Experimentos realizados no conjunto de dados Agricultural Pests Image Dataset mostraram que o CutMix melhorou significativamente a robustez do modelo, especialmente em arquiteturas baseadas em transformers, como a ViT-B/16.

Wu et al. (2024) propuseram o framework AgriPestNet, que incorpora Vision Transformers (ViTs) e métodos híbridos de extração de características para a classificação de pragas. O estudo comparou o desempenho dos modelos ViT com as CNNs tradicionais, revelando que os transformers tiveram um desempenho superior na manipulação de variações complexas de fundo nas imagens de pragas. Seus experimentos no conjunto de dados Farm Insects mostraram que os modelos ViT alcançaram uma precisão de 92,3

Gomez et al. (2023) desenvolveram um modelo leve de aprendizado profundo para detecção de pragas em dispositivos móveis. A abordagem proposta otimizou a arquitetura MobileNetV3 para aplicações agrícolas de baixa potência, garantindo classificação em tempo real com requisitos computacionais mínimos. O modelo foi testado no conjunto de dados Agricultural Pests Image Dataset, atingindo uma precisão de 87,6

Rahman et al. (2023) exploraram técnicas de aprendizado por transferência para a classificação de pragas utilizando EfficientNet e DenseNet. Seu estudo destacou a importância do ajuste fino de modelos pré-treinados com dados específicos do domínio, melhorando significativamente o desempenho da classificação. Os autores relataram que a EfficientNetB4 alcançou a maior precisão, 93,5

Singh et al. (2024) realizaram uma análise comparativa de diversas arquiteturas de aprendizado profundo para a classificação de pragas, incluindo ResNet, Xception e MobileNet. Seu estudo revelou que os modelos baseados em transformers superaram consistentemente as arquiteturas CNN, especialmente em conjuntos de dados com alta variabilidade intra-classe. Além disso, os autores enfatizaram o papel da otimização de hiperparâmetros para melhorar o desempenho da classificação.

Por fim, Patel et al. (2024) apresentaram um modelo híbrido inovador, combinando CNNs com modelos de estado-espaço, para melhorar o reconhecimento de pragas sob condições variáveis de iluminação e ambiente. Sua abordagem integrou mecanismos de autoatenção com camadas convolucionais, resultando em capacidades aprimoradas de extração de características. O modelo foi avaliado no conjunto de dados Forestry Pest Identification, alcançando um F1-score de 94,2

III. Materiais e Metodos

A. Dataset

O conjunto de dados Crop Pest and Disease Detection foi selecionado para este estudo. Ele é composto por 24.881 imagens categorizadas em 22 classes de pragas, distribuídas em quatro tipos de culturas: 6.549 imagens de Cajueiro, 7.508 de Mandioca, 5.389 de Milho e 5.435 de Tomate. Esse conjunto de dados, disponível na plataforma Kaggle[1], apresenta uma ampla diversidade de formatos, cores e tamanhos, tornando-o adequado para o treinamento e avaliação de modelos de aprendizado de máquina em tarefas de classificação de pragas.

Para este estudo, o conjunto de dados foi dividido em 60% para treinamento, 20% para validação e 20% para teste, garantindo uma distribuição equilibrada para a avaliação dos modelos. A Figura 1 apresenta algumas amostras do conjunto de dados.

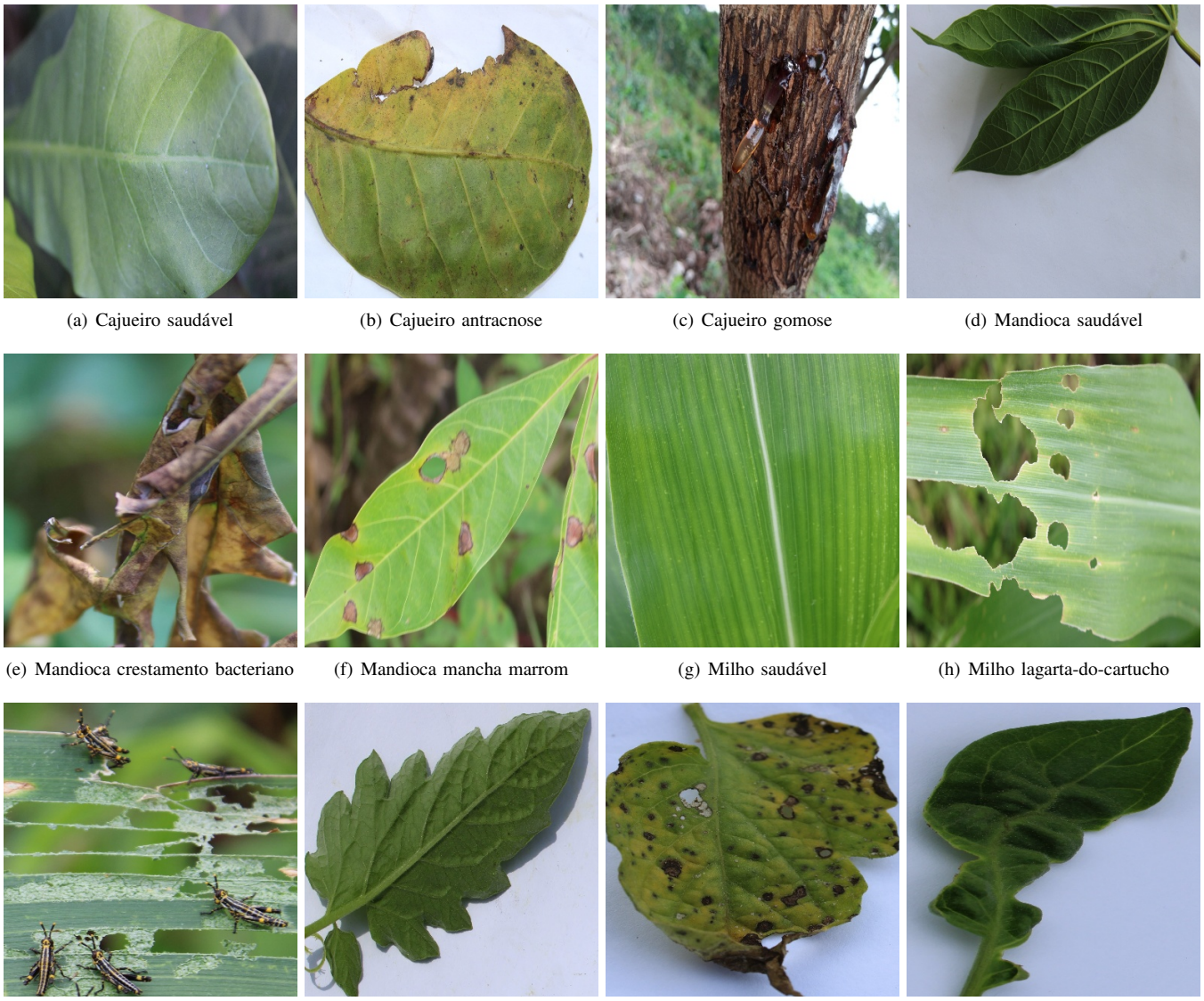
B. Arquiteturas

As arquiteturas utilizadas neste trabalho foram EfficientNet e Vision Transformer (ViT). Todos os modelos foram ajustados por fine-tuning a partir dos modelos disponibilizados pelo torchvision, os quais foram pré-treinados utilizando o ImageNet [6].

O EfficientNet [7] trouxe um avanço significativo para a área de aprendizado profundo, ao introduzir uma abordagem inovadora para o escalonamento eficiente de redes neurais convolucionais (CNNs). Seu desempenho excepcional em diversos benchmarks de classificação de imagens, incluindo o ImageNet Large Scale Visual Recognition Challenge (ILSVRC), consolidou seu impacto na pesquisa em visão computacional. A arquitetura utiliza um método de escalonamento composto, que equilibra sistematicamente a profundidade, a largura e a resolução da rede para alcançar eficiência ideal. Além disso, a utilização da função de ativação Swish melhora o fluxo do gradiente, resultando em melhor convergência e estabilidade durante o treinamento. O sucesso do EfficientNet estabeleceu um novo padrão para modelos de aprendizado profundo com eficiência de recursos, influenciando o desenvolvimento de arquiteturas subsequentes tanto na academia quanto na indústria.

O Vision Transformer (ViT) [8] revolucionou a classificação de imagens, ao se afastar da tradicional extração de características baseada em convolução e utilizar mecanismos de

[1]https://www.kaggle.com/datasets/nirmalsankalana/crop-pest-and-disease-detection



(a) Cajueiro saudável (b) Cajueiro antracnose (c) Cajueiro gomose (d) Mandioca saudável

(e) Mandioca crestamento bacteriano (f) Mandioca mancha marrom (g) Milho saudável (h) Milho lagarta-do-cartucho

(i) Milho gafanhoto (j) Tomate saudável (k) Tomate crestamento foliar (l) Tomate mosaico dourado

Fig. 1. Amostras de imagens do conjunto de dados Crop Pest and Disease Detection.

autoatenção. Inspirado nos modelos Transformer da área de Processamento de Linguagem Natural (NLP), o ViT processa imagens dividindo-as em patches, tratando-os como sequências em vez de grades de pixels. Diferentemente das CNNs, que extraem características locais, o ViT captura dependências de longo alcance dentro da imagem, tornando-se particularmente eficaz em cenários onde as relações espaciais entre características distantes são importantes. Seu desempenho tem se mostrado competitivo, especialmente quando treinado em grandes conjuntos de dados, desafiando a dominância das arquiteturas baseadas em CNNs nas tarefas de visão computacional.

C. Aumento de Dados

A técnica de aumento de dados (data augmentation) é amplamente utilizada, especialmente no treinamento de modelos em conjuntos de dados limitados. Ela consiste na expansão artificial do conjunto de dados, aplicando diversas transformações às imagens originais, aumentando assim sua variabilidade. Essa técnica melhora a capacidade de generalização do modelo, expondo-o a uma variedade maior de amostras, o que ajuda a mitigar o overfitting e torna o modelo mais robusto a variações nos dados de entrada.

Neste estudo, investigamos três métodos de aumento de dados para melhorar o desempenho do modelo. Uma das estratégias, denominada No DA, não incorporou nenhuma técnica de aumento de dados além de uma operação de recorte aleatório redimensionado, padronizando as imagens para uma resolução de 224 × 224 pixels. Essa abordagem serviu como linha de base para avaliar a eficácia das diferentes estratégias de aumento de dados na melhoria da precisão do modelo.

A primeira estratégia de aumento de dados, chamada DA-

2, aplica um recorte aleatório na imagem original do conjunto de treinamento, garantindo que o tamanho do corte varie entre 80% e 100% do tamanho original. Após o recorte, a imagem é redimensionada para 224 × 224 pixels, garantindo um tamanho fixo para entrada na rede neural. Nos conjuntos de validação e teste, as imagens são primeiramente redimensionadas para 256 × 256 pixels e, em seguida, passam por um recorte central, assegurando que a região mais importante permaneça na entrada da rede neural. O tamanho final do recorte é 224 × 224 pixels.

Na segunda estratégia, denominada DA-5, são aplicadas transformações adicionais, incluindo um espelhamento horizontal aleatório e uma rotação aleatória de até 15 graus. Além disso, as imagens passam por operações de recorte aleatório redimensionado, com um fator de escala entre 0.8 e 1.0, padronizando o tamanho final em 224 × 224 pixels. Adicionalmente, uma transformação de ajuste de cor (color jitter) modifica os valores de brilho, contraste, saturação e matiz das imagens. O brilho, contraste e saturação foram variados por um fator de 0.2, enquanto a matiz foi ajustada por um fator de 0.1.

Essas três estratégias distintas de aumento de dados permitiram avaliar o impacto de cada transformação no desempenho geral do modelo, com o objetivo de alcançar o melhor equilíbrio entre precisão e capacidade de generalização.

D. Desenho do experimento

Primeiramente, dividimos aleatoriamente 20% do conjunto de treinamento do Crop Pest and Disease Detection dataset para construir um conjunto de validação de forma estratificada.

Em seguida, treinamos cada arquitetura descrita na Seção III-B utilizando o otimizador Adam. Para a arquitetura EfficientNet, utilizamos um tamanho de batch de 32 imagens, devido à sua alta demanda de memória, enquanto para o ViT, o tamanho do batch foi de 64 imagens. A taxa de aprendizado adotada foi de 0.0001. O treinamento consistiu no ajuste fino (fine-tuning) dos modelos disponíveis na biblioteca torchvision. Cada arquitetura foi treinada tanto sem aumento de dados quanto utilizando as estratégias de aumento de dados descritas na Seção III-C.

Durante o treinamento, analisamos o desempenho do modelo e aplicamos a parada antecipada (early stopping) quando a perda na validação não apresentou melhora por 21 épocas consecutivas.

E. Ambiente Computacional

Os experimentos foram executados em um PC com um processador Core i5 de 3,00 GHz e 32 GB de RAM, equipado com uma GPU NVIDIA GTX 1080 Ti. O sistema operacional utilizado foi o Linux Ubuntu 20.04 LTS, e os experimentos foram desenvolvidos utilizando Python 3.9, PyTorch 2.0.1, torchvision 0.15.2 com CUDA Toolkit 10.1, Scikit-learn 1.2.2 e Matplotlib 3.7.1.

F. Avaliação do modelo

Para avaliar a configuração experimental, utilizamos quatro métricas de classificação amplamente empregadas: acurácia, precisão e F1-score. Essas métricas fornecem informações sobre a capacidade do modelo de generalizar o conhecimento aprendido para dados não vistos anteriormente.

Cada métrica é formulada com base nos seguintes valores: verdadeiros positivos (TP), verdadeiros negativos (TN), falsos positivos (FP) e falsos negativos (FN). As equações 1-4 apresentam as definições das métricas selecionadas:

$$Acuracia = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$Preciso = \frac{TP}{TP + FP} \tag{3}$$

$$F1 - Score = \frac{2 \times TP}{2 \times TP + FN + FP} \tag{4}$$

IV. Resultados e Discussão

A Tabela I apresenta a acurácia de cada modelo, considerando cada arquitetura e estratégia de aumento de dados, avaliadas tanto no conjunto de validação quanto no conjunto de teste. Os melhores resultados entre os dois modelos estão destacados em negrito, enquanto os resultados obtidos com diferentes estratégias de aumento de dados estão em itálico.

Em relação à acurácia no teste, observa-se que o modelo ViT superou o EfficientNet quando treinado com uma estratégia de aumento de dados, mas não obteve melhor desempenho quando treinado sem aumento de dados. O ViT atingiu sua maior acurácia com a estratégia DA-5, alcançando 0,9170. Por outro lado, o EfficientNet obteve sua melhor acurácia no teste quando treinado sem uma estratégia de aumento de dados.

As Tabelas III e II resumem as métricas de precisão, recall e F1-score de cada modelo, avaliadas nos conjuntos de validação e teste, utilizando as médias macro e ponderada, respectivamente. A média macro corresponde à média aritmética das métricas de cada classe, enquanto a média ponderada leva em consideração as proporções das classes no cálculo da métrica geral. Os resultados apresentados nessas tabelas estão alinhados com os da Tabela I. Os modelos Vision Transformer (ViT) superaram o EfficientNet, alcançando os melhores resultados com aumento de dados. Especificamente, as métricas de precisão, recall e F1-score do ViT, relatadas em pares para as médias ponderada e macro (ponderada, macro), foram respectivamente (0,9574, 0,9562), (0,9572, 0,9548) e (0,9571, 0,9553).

Com relação às métricas de precisão, recall e F1-score, o EfficientNet-B4 obteve seu melhor desempenho quando treinado com a estratégia de aumento de dados DA-5. No entanto, o ViT foi o modelo que apresentou os melhores resultados gerais quando treinado com aumento de dados.

TABLE I
RESULTADOS DOS EXPERIMENTOS COM OS MODELOS TREINADOS COM DIFERENTES ESTRATÉGIAS DE AUMENTO DE DADOS (ACURÁCIA).

		VAL. Acc.			TEST Acc.		
	Arquitetura	**No DA.**	**DA-2**	**DA-5**	**No DA.**	**DA-2**	**DA-5**
Acc.	**EfficientNet b4**	0.8981	0.8997	0.9092	*0.9900*	0.9844	0.9807
	ViT b 16	0.8925	0.9164	0.9170	***0.9938***	**0.9914**	**0.9906**

TABLE II
RESULTADOS DOS EXPERIMENTOS COM OS MODELOS TREINADOS COM DIFERENTES ESTRATÉGIAS DE AUMENTO DE DADOS (MÉDIA PONDERADA).

		VAL.			TEST		
	Arquitetura	**No DA.**	**DA-2**	**DA-5**	**No DA.**	**DA-2**	**DA-5**
Prec.	**EfficientNet b4**	0.9500	0.9539	0.9564	0.9244	*0.9437*	0.9362
	ViT b 16	0.9615	0.9590	0.9595	***0.9574***	**0.9456**	**0.9503**
Rec.	**EfficientNet b4**	0.9499	0.9534	0.9556	0.9245	*0.9436*	0.9363
	ViT b 16	0.9613	0.9590	0.9590	***0.9572***	**0.9454**	**0.9500**
F1	**EfficientNet b4**	0.9494	0.9522	0.9553	0.9239	*0.9429*	0.9356
	ViT b 16	0.9612	0.9586	0.9588	***0.9571***	**0.9452**	**0.9496**

TABLE III
RESULTADOS DOS EXPERIMENTOS COM OS MODELOS TREINADOS COM DIFERENTES ESTRATÉGIAS DE AUMENTO DE DADOS (MÉDIAS MACRO).

		VAL.			TEST		
	Arquitetura	**No DA.**	**DA-2**	**DA-5**	**No DA.**	**DA-2**	**DA-5**
Prec.	**EfficientNet b4**	0.9473	0.9519	0.9536	*0.9424*	0.9331	0.9166
	ViT b 16	0.9584	0.9558	0.9557	***0.9562***	**0.9434**	**0.9471**
Rec.	**EfficientNet b4**	0.9450	0.9504	0.9527	0.9224	*0.9405*	0.9346
	ViT b 16	0.9578	0.9555	0.9561	***0.9548***	**0.9441**	**0.9478**
F1	**EfficientNet b4**	0.9455	0.9496	0.9523	0.9214	0.9407	*0.9332*
	ViT b 16	0.9578	0.9552	0.9554	***0.9553***	**0.9434**	**0.9468**

V. CONCLUSÃO

Nesta pesquisa, analisamos duas arquiteturas de aprendizado profundo para a classificação de imagens de doenças e avaliamos seu desempenho utilizando três estratégias distintas de aumento de dados. Os modelos foram ajustados por fine-tuning a partir de redes pré-treinadas.

Nossos resultados demonstram que as técnicas de aprendizado de máquina e aprendizado profundo são altamente eficazes para a classificação de doenças e pestes, com o Vision Transformer (ViT) alcançando a maior acurácia no teste, de 0,9938, quando treinado sem nenhuma técnica de aumento de dados. O ViT superou o EfficientNet em todas as métricas avaliadas, incluindo acurácia, precisão, recall e F1-score. Embora o ViT não tenha se beneficiado das estratégias de aumento de dados, essas técnicas desempenharam um papel fundamental na melhoria do desempenho de outras arquiteturas. O EfficientNet-B4 obteve seus melhores resultados quando treinado com a estratégia DA-5.

Esses resultados ressaltam o potencial do aprendizado profundo para a classificação de doenças, fornecendo uma base para o desenvolvimento de ferramentas práticas e eficientes que auxiliem profissionais na identificação de diversas condições patológicas.

Para pesquisas futuras, pretendemos testar novas arquiteturas de aprendizado profundo e explorar novas estratégias de treinamento, incluindo técnicas aprimoradas de aumento de dados. Considerando que as pragas agrícolas são altamente diversificadas em todo o mundo e que algumas apresentam aparências muito distintas em cada estágio de vida, planejamos utilizar outros conjuntos de dados e explorar técnicas avançadas de aprendizado por transferência (transfer learning).

REFERENCES

[1] S. Palei, R. K. Lenka, S. S. Nayak, R. Mohanty, B. Jena, and S. Saxena, "Precision agriculture: Ml and dl-based detection and classification of

agricultural pests," in *2023 2nd International Conference on Ambient Intelligence in Health Care (ICAIHC)*. IEEE, 2023, pp. 1–6.

[2] W. Zhang, X. Xia, G. Zhou, J. Du, T. Chen, Z. Zhang, and X. Ma, "Research on the identification and detection of field pests in the complex background based on the rotation detection algorithm," *Frontiers in Plant Science*, vol. 13, p. 1011499, 2022.

[3] R. G. P. Neto, P. M. de Sousa, L. F. R. Moreira, P. I. V. G. God, and J. F. Mari, "Enhancing green coffee quality assessment through deep learning," in *Anais do XVIII Workshop de Visão Computacional*. SBC, 2023, pp. 84–89.

[4] J. G. Esgario, P. B. de Castro, L. M. Tassis, and R. A. Krohling, "An app to assist farmers in the identification of diseases and pests of coffee leaves using deep learning," *Information Processing in Agriculture*, vol. 9, no. 1, pp. 38–47, 2022.

[5] G. S. de Lima Mota, L. H. Silva, L. F. R. Moreira, and J. F. Mari, "Classifying pests in crop images using deep learning," in *Anais do XVIII Workshop de Visão Computacional*. SBC, 2023, pp. 42–47.

[6] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.

[8] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.