

ORIGINAL PAPER

AMIPDI: architecture model to identify pests in soybean defoliation images using convolutional neural networks

Maria Eloisa Mignoni ¹, Aislan Honorato², Cesar Zagonel³, Gabriel de Oliveira Ramos ⁴, Rafael Kunst ⁴

¹University of Estado de Mato Grosso Carlos Alberto Reyes Maldonado – UNEMAT, ²Centro Universitário de Várzea Grande – UNIVAG, ³University of Cruzeiro do Sul – UNICSUL, ⁴University of Vale do Rio dos Sinos – Unisinos

*eloisa@unemat.br; aislanmano@gmail.com; zagonelcesar@gmail.com; gdoramos@unisinos.br; rafaelkunst@unisinos.br

Received: 2024-08-20. Revised: 2025-06-13. Accepted: 2025-07-01.

Abstract

Pests that attack crops are one of the leading causes of low production and economic losses. Identifying and diagnosing forecasts is one way to reduce losses and maintain quality in production. Defoliation of crops harms production and production quality. Using computational technologies such as Artificial Intelligence and Machine Learning has enabled the identification of pests faster and earlier. In this paper, we propose a Deep Learning-based architecture aimed at identifying and diagnosing insects by analyzing defoliation in crop images. Different Convolutional Neural Network approaches were considered to evaluate the proposed architecture. The training and testing of the models were performed using images collected by drone in a natural environment. The approach that presented the best performance in our scenario was the VGG16. The average accuracy in the validation phase was 0.95, while in the test set, we obtained 0.86.

Keywords: Soybean, Deep Learning, Pest, Defoliates, Agriculture

Resumo

As pragas que atacam as culturas são uma das principais causas de baixa produção e de perdas econômicas. Identificar, diagnosticar as previsões são formas de reduzir as perdas e manter a qualidade da produção. A desfolha das culturas prejudica a produção e a qualidade da produtividade. O uso de tecnologias computacionais como Inteligência Artificial e Aprendizado de Máquina tem possibilitado a identificação de pragas de forma mais rápida e precoce. Neste artigo, propomos uma arquitetura baseada na aprendizagem profunda destinada a identificar e diagnosticar insetos através da análise da desfoliação em imagens de culturas. Foram consideradas diferentes abordagens de Redes Neurais Convolucionais para avaliar a arquitetura proposta. O treinamento e o teste dos modelos foram realizados utilizando imagens coletadas por drone em um ambiente natural. A abordagem que apresentou o melhor desempenho, em nosso cenário, foi a VGG16. A acurácia média, na fase de validação, foi de 0,95, enquanto no conjunto de teste, obtivemos 0,86.

Palavras-Chave: Soja, Aprendizado Profundo, Praga, desfolha, Agricultura

1 Introduction

Pest infestation is one of the leading causes of crop damage and economic losses in the agricultural chain (Nanni et al., 2022). Rigid pest control is necessary to generate economic gains and increase production and food quality. Pest control must be continuous, allowing the detection and early identification of pests, allowing immediate

actions to combat pests (de Castro Pereira et al., 2022a).

Soybean (*Glycine max* (L) Merrill) is part of the legumes belonging to the Papilionaceae family, originating in Asia, China (Sedivy et al., 2017). From planting to harvesting, soybeans are subject to attacks by defoliating pests, insects being one of the problems producers face. The forms of pest identification are still, for the most part, carried out manually or with static traps (Silva et al., 2019; Xin and

Wang, 2020). Technological innovations have proven to be an alternative to pest control. Automated solutions using Artificial Intelligence (AI) for automatic identification have significantly accelerated pest recognition (Nanni et al., 2022). The use of image capture equipment has contributed to faster diagnoses. Drones equipped with cameras, flying over plants, capture images with rich details, allowing crops to be monitored. Drones used in crops are considered important in control, detection, and management decision-making (Tetila et al., 2020).

Deep Learning (DL) techniques have obtained relevant results for the identification, recognition, detection, and diagnosis of pests (Tassis et al., 2021). These techniques have proven to be efficient in identifying and diagnosing pests in images (Patel and Bhatt, 2021), acting in information control, decision-making, pattern recognition, and knowledge acquisition (Voutos et al., 2019). Zhao et al. (2022) cite DL as the primary technique under study for pest recognition in pest images.

CNN architectures have stood out in the classification, identification, counting, and estimation of pests and plant data (Patel and Bhatt, 2021; Tetila et al., 2019, 2020; Tassis et al., 2021; Silva et al., 2019; Xin and Wang, 2020). Tetila et al. (2019, 2020) investigated the use and performance in identifying and counting soybean insects. In this context, Tetila et al. (2020) use transfer learning approaches and fine-tune a CNN to analyze images collected in real-time in the field to calculate the level of insect infestation in a crop. Tassis et al. (2021) proposed an architecture to face the existing complexity in natural images to detect and recognize lesions in coffee plants. Silva et al. (2019) proposed a model that estimates soybean defoliation using natural images. Xin and Wang (2020) designed a model to extract leaf traits from soybeans, while de Castro Pereira et al. (2022a) proposed a DL-based approach to identify the whitefly insect on soybean leaves.

Although the topic is relevant, to the best of our knowledge, studies have yet to focus on identifying insects by analyzing defoliation using natural images. The most comparable studies related to defoliation propose solutions to estimate defoliation percentages (Silva et al., 2019; Xin and Wang, 2020). Leaf defoliation is one of the many damages caused by pests. In our approach, we propose a DL-based architecture that analyzes defoliation to identify which pests affect the crop.

More specifically, our proposed architecture allows the evaluation of various CNN approaches to diagnose and identify pests by defoliation using natural images collected in crops. In this paper, we conduct a case study in which the proposed architecture aims to identify two types of pests (caterpillar and diabrotic speciosa) by analyzing defoliation. The DL models were trained using a natural image dataset Mignoni et al. (2022) and tested with natural images captured by drones in a crop.

Defoliation to diagnose and identify the type of pest in natural images contributes to yet another recognition method for the kind of insect. With defoliation for identification and diagnosis, the pest control process can be streamlined and minimize production and product quality losses, with the user's decision-making for action and control. Factors such as the environment and the economy can also benefit from using less agrochemicals

to control pests.

The remainder of this paper is organized as follows. Section 2 presents background aspects and related work. Section 3 presents the proposed architecture, discussing details on each step of the image collection and analysis. Finally, Section 4 presents conclusions and directions for future research.

2 Background and Related Work

We conducted a literature search in diverse scientific libraries, such as IEEE, ACM, Elsevier, and Science Direct, aiming at analyzing the state-of-the-art in the topic of defoliation analysis using natural images. Although, to the best of our knowledge, related works do not address the diagnosis and identification of pests by defoliation in images, some articles tackle related approaches. These researches generally deal with image natural image collection, and DL approaches to estimate the degree of damage caused by pests in crops.

DL is a subarea of ML that uses algorithms to process data, imitating the functioning of the human brain and allowing computers to learn (Goodfellow et al., 2016). DL algorithms can automatically extract and detect data characteristics (Christin et al., 2019). DL works with computational models composed of several processing layers and data abstraction levels (Tetila et al., 2020). Computational models that use DL are composed of several processing layers that can learn data characteristics with various levels of abstraction (Tetila et al., 2020; LeCun et al., 2015; Christin et al., 2019). DL is widely used in agriculture to control and combat pests, soil management, crop quality, productivity, and weather forecasting (Alpaydin, 2020; YU et al., 2022). The various DL techniques, such as CNN, enable its application in countless areas, actions, and innovative solutions (Li et al., 2021).

The CNN architecture is a DL method with multilayer perception designed to work with images in image classification (He et al., 2016). It is presented as an excellent tool for automatic classification/detection of pests (Tassis et al., 2021). CNNs are composed of different types of layers, such as convolution and pooling (Sharma et al., 2021). Song et al. (2019) report CNNs as a type of backpopulation neural network, where the extraction and mapping layers extract the characteristics of objects. The convolutional layer is the most important structure of a CNN, as it can extract effective features and reduce the complexity of images (Song et al., 2019). The use of CNN architectures in identifying and diagnosing pests in images has shown to be very promising Tetila et al. (2020).

Tetila et al. (2019, 2020) investigated the use and performance of DL in identifying and counting insects in soybeans. The authors use transfer learning and fine-tuning techniques along with a data set composed of 5000 256x256 px images collected in real time. Using this dataset, Tetila et al. proposed a CNN-based computer vision model that automatically identifies and counts the number of insects to calculate the level of infestation affecting a given crop.

Tassis et al. (2021) proposed an architecture to deal

Table 1: An architectural proposal to identify insects using CNN through defoliation in images captured by drones in a natural environment

Author	Images	Characteristics			
		defoliation	Identify insects	UAV	CNN
Silva et al. (2019)	Synthetic images for training and actual ones for tests	Estimates the level of defoliation	No	No	Yes
Tetila et al. (2019)	Natural preprocessed images	No	Identifies the insects	No	Yes
Tetila et al. (2020)	Natural preprocessed images	No	Identifies the insects	Yes	Yes
Tassis et al. (2021)	Natural preprocessed images to minimize the problem regarding the back of the photos	Recognizes that there are injuries in the leaves	No	No	Yes
Albattah et al. (2022)	IP102 images and drone images	Extracts characteristics from the insect's picture	Identifies the insect	Yes	Yes

with the complexity of collecting and analyzing natural images. The authors integrated different CNNs to evaluate the degree of lesions in coffee plants, using images captured via smartphone in natural conditions. The results were obtained using UNet and PSPNet models, attaining accuracies of 94.25% and 93.54%, respectively. Although the authors obtained relevant results, the study's limitations include identifying defoliation in an image, as a stain or groove can be confused with data, overlapping colors, and lesions on the same leaf.

[Silva et al. \(2019\)](#) proposed a model based on CNN that estimates soybean defoliation. The model was trained using a set of 256x256 px synthetic images. The model testing was conducted using natural images captured using cameras. The results allowed the authors to analyze the defoliation percentage in a crop. However, the authors did not focus on identifying the type of pests that affect a crop.

[Albattah et al. \(2022\)](#) proposed a DL model with an automated framework to extract insect-related features from images. The authors' approach was to train a CNN using the IP102 database¹ and tested with images collected in a natural environment using drones [Mignoni et al. \(2022\)](#).

Related works present solutions to estimate defoliation using DL to analyze natural images. The literature also presents solutions for quantifying crop infestation. However, to the best of our knowledge, these researches are not focused on identifying pest types while analyzing defoliation. Considering this scenario, Table 1 summarizes the main characteristics of the related research. Considering the gaps not yet covered by the literature, in this article, we propose an architecture to tackle the lack of identification of insects by defoliation and analyze its performance in a case study involving natural images collected using UAVs. The case study identifies two pest classes: caterpillar and diabrotica speciosa.

3 Proposed Architecture

The identification of defoliation pests is a complex process due to the characteristics presented by the damage. More

specifically, the bottom leaf can camouflage damage to the leaves because the color is the same and does not have an embossed texture. These characteristics make damage visualization more difficult. Another characteristic that makes it challenging to identify the insect is that the same leaf can have more than one type of insect damage. In order to deal with these challenges, in this section, we propose a CNN-based architecture capable of identifying the type of pest using natural images collected from crops.

[Fig. 1](#) illustrates the structure of the proposed architecture. The solution comprises three modules: (I) image capture, (II) image pre-processing, and (III) the CNN, which was trained using defoliation images of soybean leaves captured using UAVs in soybean fields in Lucas do Rio Verde, Mato Grosso, Brazil ([Mignoni et al., 2022](#)). Details on each module are presented in the following subsections.

3.1 Image Capture

The first module of our architecture deals with how the images are captured in the field. In the case study analyzed in this paper, the images were captured in natural conditions using a UAV. Images were collected on different days to accommodate diverse weather conditions, varying among sunny, cloudy, mild wind, and clouds. The images were captured during the harvest period between October 2020 and January 2021, from 7:00 a.m. to 9:00 a.m. and 4:00 p.m. to 6:00 p.m.

The time selected for image capturing considered the period of the day less affected by wind. The UAV camera distance varied from 50cm to 2m in height. We captured images with a Mavic Air2 Drone 40mp camera. [Figure 2](#) shows sample images of a soybean crop captured with the UAV in variable weather conditions. These images are part of a dataset created using the images captured to conduct the case study presented in this article ([Mignoni et al., 2022](#)).

3.2 Image Pre-processing

The images pre-processing is carried out in the architecture's second module. The datasets were balanced before pre-processing, as shown in [Fig. 1](#). The images were loaded, and applied pre-processing techniques. The data was increased for the size of the dataset,

¹<https://github.com/xpwu95/IP102>

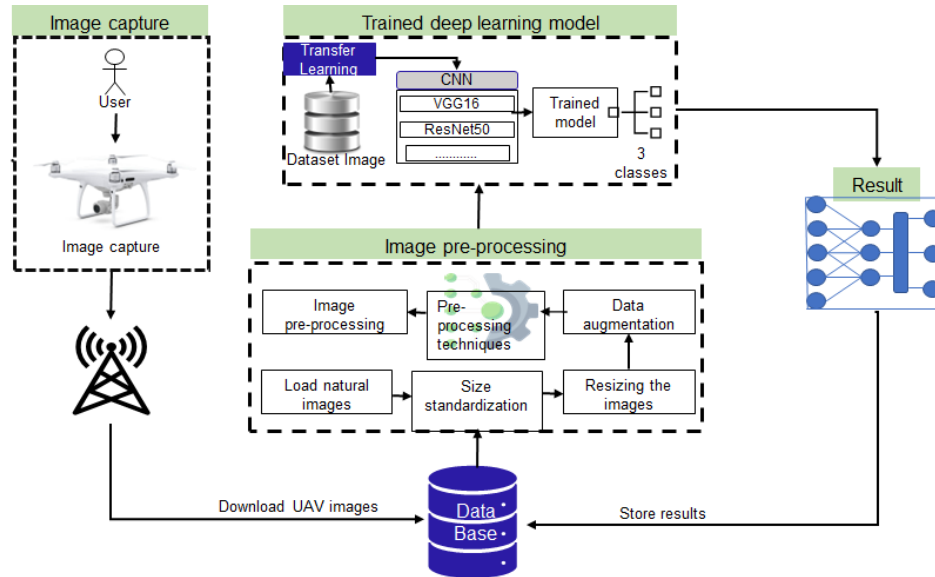


Figure 1: Proposed Architecture

which was too small to train the model with CNN. CNN architectures use a large amount of data to be trained. We used functions from the OpenCV library to augment the dataset. The resize function was used to resize the images, maintaining the proportion and solving the different angles and scales of the image capture. The rotation technique was used to increase the number of images in the dataset: original position, 90, 180, 270, -90, -180, -270 degrees.

The dataset augmentation process is shown in Table 2. Table 2 presents the amounts of images without and with augmentation, divided by class and size. In the first column are the sizes of the training and validation dataset images and the test images. Columns 2, 3, and 4 show the number of images without increase per class. Columns 5, 6, and 7 contain the amounts of images per class that were enlarged, and the last column contains the total number of images resulting from the application of the augmentation technique, that is, the values in columns 5, 6, and 7. Thus, the number of images in column 2, when receiving the rotation technique, resulted in the number of images in column 5, and the process occurred successively in the other columns.

Fig. 3 illustrates three of the six rotations performed to increase the amount of data in the dataset. Image a) is in the original position, image b) has been resized, and image c) is rotated to the left.

3.3 Dataset

The images used in the CNN model's training, validation, and testing are from the Soybean images dataset published in 2021 (Mignoni et al., 2022). These images used in our study differ from other studies as they deal with defoliation. They are natural images captured in a real environment, in real weather situations. The images are of defoliation caused by caterpillar and diabrotica speciosa, not contemplating other types of damage in this specific dataset.

The images were annotated by three experts: two agronomists with more than ten years of experience and a farmer with more than 30 years of experience. These experts identified three classes of images. Figure 4 represents these three classes: (a) refers to the class of healthy leaves; (b) refers to the class of leaves with defoliation caused by caterpillars; and (c) illustrates leaves with defoliation caused by diabrotica speciosa.

The defoliation images may have damage caused by both types of insects. In cases where both types of damage occurred in the same image, the annotation was based on the highest number of damages of the same insect. The annotated datasets have different amounts due to the sharpness presented. Some images were manually excluded because they were not clear. Table 2 describes details of the dataset organization, including the amounts of images used for training and test purposes.

Table 2: Dataset characteristics

Image Size	Caterpillar	Diabrotica speciosa	Healthy leaves	Total amount of images
CNN Model Training				
256 x 256 px	2317	2317	2317	6951
512 x 512 px	2023	2023	2023	6069
CNN Model Test				
256 x 256 px	252	252	252	756
512 x 512 px	224	224	224	672



Figure 2: Images of a soybean crop, captured with UAV in a natural environment



Figure 3: Represents some ways used to increase and standardize the set of images. Normal, resized, and rotated image

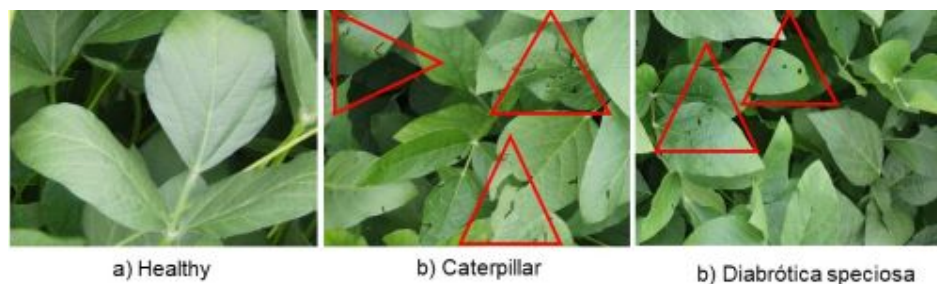


Figure 4: The three images represent the annotated classes of the dataset. The classes are: image a) healthy sheets; image b) leaves with defoliation caused by caterpillar; image c) leaves with defoliation caused by diabrotica speciosa

3.4 Training of CNNs Models

In the third module of the architecture, we implemented and trained a Deep Learning model using Convolutional Neural Networks (CNNs) to identify and classify soybean leaf defoliation caused by pests. The objective was to distinguish among three classes: healthy leaves, defoliation caused by caterpillars, and defoliation caused by *Diabrotica speciosa*. Given the complexity of working with natural images and the limited size of the dataset, we adopted a transfer learning strategy to enhance performance and mitigate data scarcity challenges.

The CNNs evaluated in this study were selected based on their established performance in image recognition tasks. Specifically, we employed VGG16 and VGG19 (Simonyan and Zisserman, 2015), ResNet50 (He et al., 2016), DenseNet201 (Huang et al., 2018), and EfficientNetB7 (Tan, 2020), all of which are pre-trained on the ImageNet dataset. These models were integrated into our architecture with their convolutional layers frozen to preserve previously learned features, while the classification head was redefined and trained using our labeled dataset. This transfer learning approach leverages the general visual features learned by the base models

and fine-tunes the final layers to adapt to the specific characteristics of defoliation patterns.

Each model was extended with two fully connected (dense) layers with 1024 and 512 units, respectively, activated by the ReLU function. Dropout regularization with a dropout rate $p = 0.5$ was applied after each dense layer to reduce overfitting by randomly omitting half of the layer's units during training. Additionally, L2 regularization was applied to the dense layers, with a regularization parameter $\lambda = 0.001$, penalizing large weights to further control model complexity and prevent overfitting. The final classification layer comprised three neurons with a Softmax activation function, defined as:

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^C e^{z_j}}, \quad \text{for } i = 1, \dots, C \quad (1)$$

where $C = 3$ corresponds to the number of classes. The categorical cross-entropy loss function was adopted to train the models, expressed as:

$$\mathcal{L}(y, \hat{y}) = - \sum_{i=1}^C y_i \log(\hat{y}_i) \quad (2)$$

where y_i is the true class label and $\hat{y}_i$ is the predicted probability for class i . This loss function is appropriate for multi-class classification problems with mutually exclusive outputs.

The models were trained for 200 epochs with a batch size of 32, using the Adam optimization algorithm, which combines the advantages of Adaptive Gradient Algorithm (AdaGrad) and Root Mean Square Propagation (RMSProp). Adam adapts the learning rate for each parameter by computing adaptive estimates of lower-order moments. The default parameters were used: learning rate $\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$.

To evaluate model performance, we computed standard classification metrics: accuracy, precision, recall, F1-Score, and Area Under the ROC Curve (AUC). These metrics offer complementary perspectives: accuracy measures the overall correctness, precision quantifies the ratio of true positives to all predicted positives, recall indicates the ratio of true positives to actual positives, the F1-Score balances precision and recall, and AUC assesses the model's discrimination ability across all classification thresholds.

All models were trained and evaluated on both 256×256 and 512×512 image resolutions, using the balanced dataset described in Section 3.3. The training workflow is visually summarized in Fig. 5, which illustrates the input data, model training pipeline, and selection of the best-performing CNN architecture.

3.5 Five CNN training and test results

The models trained in the experiment were the VGG16, VGG19, ResNet50, DenseNet201, and EfficientNetB7 pre-trained with ImageNet. Figure 5 demonstrates the training

process of the CNNs, showing the one with the best performance. The VGG16 architecture presented the best results in validation and tests. The dataset was divided into 80% for training and 20% for validation. These values were chosen at random. In training, we used two sets of RGB images of different sizes, 256×256 and 512×512 .

The 256×256 dataset is composed of 6951 images, and the 512×512 dataset has 6069. Both datasets have three balanced classes: healthy, caterpillar, and DS. Each 256×256 dataset class has 2317 images, and the 512×512 dataset classes have 2023 images each, shown in Table 2. The test images have the same sizes as those used in training and validation, 256×256 and 512×512 , and the amount of 756 and 672, respectively. Also annotated these images with the three classes and balanced, as shown in Table 2.

The results of the validation of the CNNs are presented in Table 3. The results show that the model that presented the best result in the validation was VGG16, in both image sizes, considering the five metrics. Of the trained models, the one that presented the worst results was DenseNet201 in both datasets. In the other models, the differences are small, demonstrating a balance in performance.

Among the metrics used in training, AUC showed the best results, followed by accuracy, as shown in Table 3. The accuracy achieved in the AUC metric reached 0.99 in dataset 256 and 0.97 in dataset 512. The lowest performance was demonstrated by the Recall metric in both datasets. The values achieved in the metrics show that between the two datasets, the best performance was obtained by the 256×256 dataset, with 0.99 in the AUC metric.

Model training using CNNs demonstrated that the 256×256 dataset performed better in identifying defoliation pests. This better performance also occurred in all metrics. VGG16 was the most viable in these experiments and indicated for these datasets.

3.6 CNN Performance VGG16

VGG16, for having presented the best results in validation, tests were carried out for identification and pests through defoliation with other images. The tests were performed with two sizes, 256×256 and 512×512 balanced. The VGG16 presented a performance in the tests of 0.90 and 0.87, respectively, for the two tests, demonstrating that 256×256 images obtained a better performance.

Fig. 6 presents the test results by metrics of the VGG16 model. The tests with two sizes of images, using the VGG16, obtained a success rate in the AUC metric of 0.87 and 0.90, respectively, while the others presented indexes equivalent to each other in each size of image.

The confusion matrix of the VGG16 model presents the positive and false positive results in identifying each sample. The two confusion matrices in Fig. 7 show the results of tests performed with both image sizes. The confusion matrix shows the highest number of hits was for the caterpillar, with 234 positive and 18 false in the 256×256 images. In the 512×512 images, the highest number of hits was for DS, with 219 positive and five false. The confusion matrices in Fig. 7 demonstrate that the 256×256 images obtained the highest number of hits in the total set.

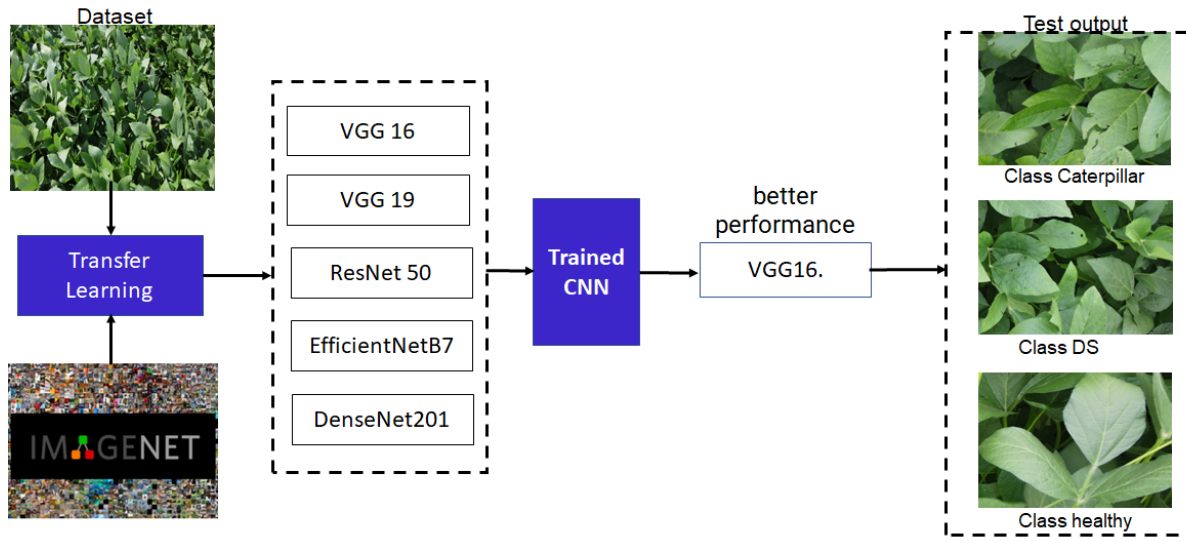


Figure 5: Training of CNN architectures and selection of the best-performing one. The figure illustrates the use of multiple pre-trained models, the training process with the proposed classification head, and the selection based on validation metrics.

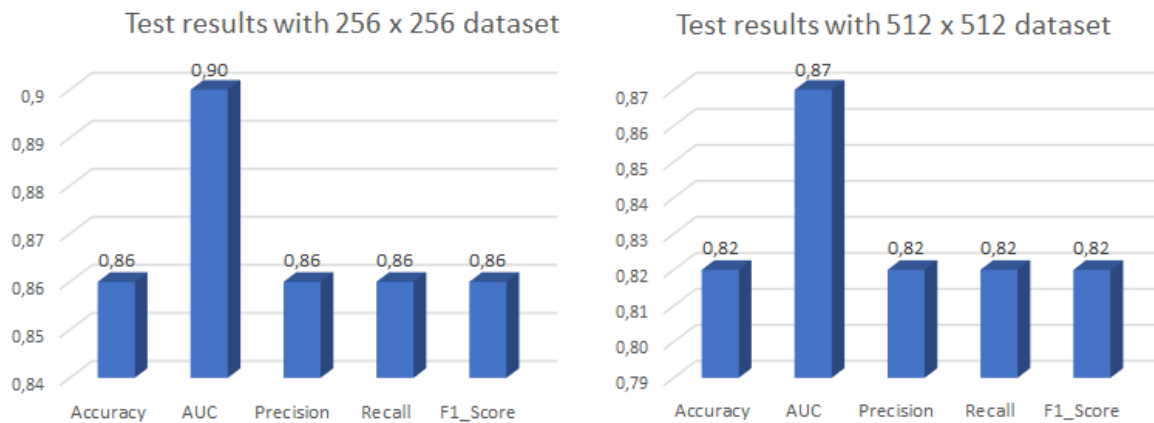


Figure 6: Results of model tests using VGG16

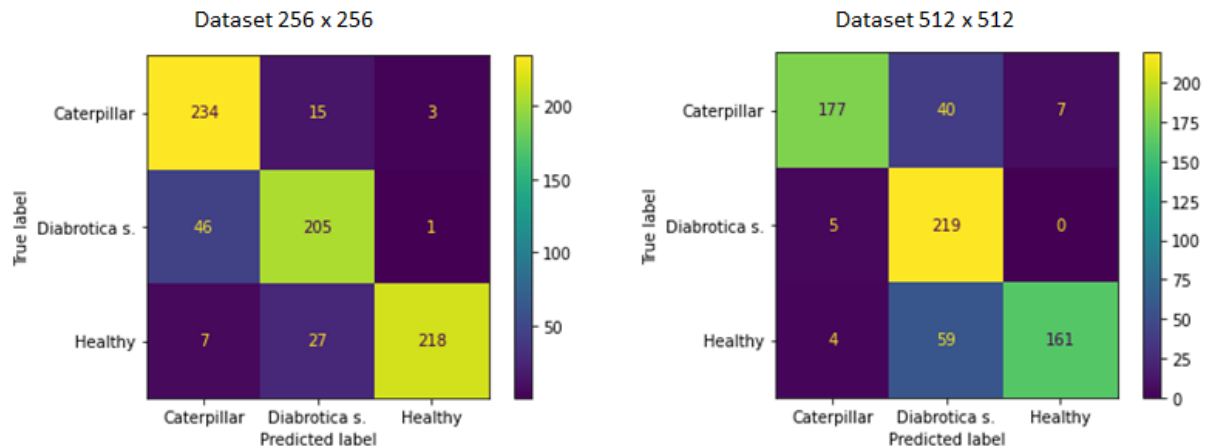


Figure 7: Confusion pattern model with VGG16

Table 3: Validation results in the CNN training

Dataset 256 x 256					
CNN	Accuracy	AUC	Precision	Recall	F1_score
VGG16	0,95	0,99	0,95	0,95	0,95
VGG19	0,92	0,98	0,93	0,92	0,92
ResNet50	0,94	0,98	0,94	0,94	0,94
EfficientNetB7	0,95	0,99	0,95	0,94	0,95
DenseNet201	0,76	0,91	0,77	0,75	0,76
Dataset 512 x 512					
CNN	Accuracy	AUC	Precision	Recall	F1_score
VGG16	0,90	0,97	0,91	0,90	0,90
VGG19	0,89	0,97	0,89	0,88	0,89
ResNet50	0,84	0,92	0,84	0,83	0,82
EfficientNetB7	0,85	0,96	0,85	0,84	0,84
DenseNet201	0,72	0,88	0,75	0,68	0,75

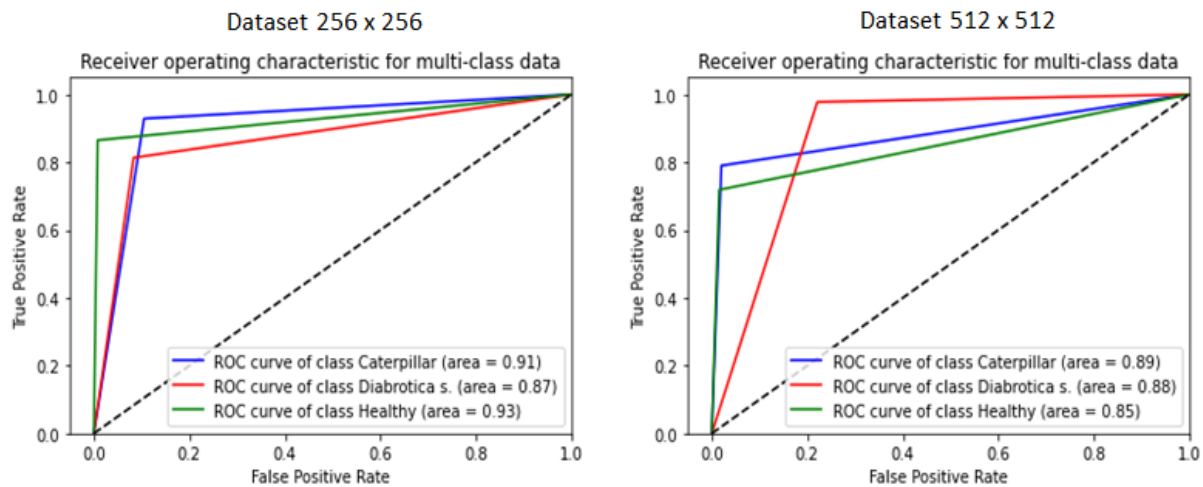
**Figure 8:** ROC Curve of the proposed model using VGG16

Fig. 8 presents the ROC curve of the VGG16 model with the values achieved by the class in the tests. The curve with the highest value is that of the 256 x 256 images in the healthy class with 0.93. The ROC curve with the lowest value is from the 512 x 512 images in the healthy class with 0.85. These values can also be verified in Fig. 7, together with the confusion matrix of each dataset.

Identifying pests through defoliation is an innovative study not found in other studies. The results are compared with other studies that identify pests through images of pests and with studies that estimate the percentage of damage due to defoliation to have a parameter of the values.

The results achieved in our training and validation compared with the authors' studies in Table 4 demonstrate that the differences between the results are not great. Table 4 presents the relationship between the results of other authors and our indices, showing that the results are approximate. The results obtained in the tests, considering the complexity of identifying a pest by its damage, indicate that with these results, the person in charge can act in favor of pest control. The cited studies report the difficulty in extracting features from the images and working with natural images due to nature's weather. Considering the mentioned difficulties, the

results obtained from other authors, ours, demonstrate promising and viable.

This study considered the difficulty of identifying a pest by defoliating plants. These difficulties are caused by the fact that the leaves after defoliation are the same color as the healthy ones; damage from different pests caused by different insects can occur on the same leaf because the texture between the healthy and leafless parts is the same; damage dimensions are similar. When analyzing the results, the difficulties mentioned earlier were taken into account. Considering these difficulties, the results are promising. Fig. 4 shows these reported difficulties, demonstrating some difficulty identifying the type of pest causing the damage.

4 Conclusion and Future Works

This research proposed a DL architecture to identify defoliation information in real images captured by drones. The studies' architecture, training, tests, results, and analyses are described, demonstrating the results obtained. The problem of identifying and diagnosing defoliation rules is difficult due to the complex background of the images. Based on CNN, the architecture proposal

Table 4: Comparison of results obtained by other authors with ours

Author	Activity	Value	Converted values
Our results	Identify pests through defoliation	90 and 87	0,90 and 0,87
Tassis et al. (2021)	Estimativa de desfolha	73,90 and 94,25	0,73 and 0,94
Tetila et al. (2020)	Detection and classification of pests	93,82 and 91,80	0,93 and 0,91
Xin and Wang (2020)	Extraction of foliar characteristics	96,3	0,96
de Castro Pereira et al. (2022b)	Detection and classify whiteflies	87	0,87
Wang et al. (2022)	Pests Identification	78,6	0,78

All authors used CNN.

to identify information by defoliation through images captured with drones demonstrated that it is possible and acceptable to identify directives in this way.

The results of training, validation, and tests with both datasets showed differences that do not change the model's accuracy. The model with VGG16 showed the best validation result, with an average of 0.95 and 0.91, respectively. Having presented the best results in the validation, this model was chosen for the tests. It presented an average of 0.86 and 0.82, respectively, in the tests. Considering the complexity of the background of the images to visualize defoliation, the difficulty of extracting characteristics from the images, the difficulty of identifying the type of pest due to defoliation damage, and the comparison with the results of other authors, these results prove to be promising and viable.

The limitations presented by the proposed model are characterized by images collected in real environments, the background complexity of the images, the initial stages of plantations, and small datasets. Using other pre-processing techniques can minimize the events caused by bad weather in nature and other regions. Capturing more real images solves the small dataset issue.

For future work, it is suggested to expand the data set. Use processing techniques that soften the effects of nature and improve images in terms of background complexity. Test the proposed model to identify other types of crop defoliation pests in addition to Caterpillar and DS. Test the model on defoliation of crops other than soybean. Implement the attention mechanism and CNN in the proposal to identify insects by defoliation in real images.

References

- Albattah, W., Masood, M., Javed, A., Nawaz, M. and Albahli, S. (2022). Custom cornernet: a drone-based improved deep learning technique for large-scale multiclass pest localization and classification, *Complex & Intelligent Systems* **9**(2): 1299–1316. <http://10.1007/s40747-022-00847-x>.
- Alpaydin, E. (2020). *Deep Learning*, 4 edn, Mit Press, Floor Cambridge, MA.
- Christin, S., Hervet, E. and Lecomte, N. (2019). Applications for deep learning in ecology, *Methods Ecol Evol* **10**: 1632–1644. <https://doi.org/10.1111/2041-210X.13256>.
- de Castro Pereira, R., Hirose, E., Ferreira de Carvalho, O. L., da Costa, R. M. and Borges, D. L. (2022a). Detection and classification of whiteflies and development stages on soybean leaves images using an improved deep learning strategy, *Computers and Electronics in Agriculture* **199**: 107132. <http://doi.org/10.1016/j.compag.2022.107132>.
- de Castro Pereira, R., Hirose, E., Ferreira de Carvalho, O. L., da Costa, R. M. and Borges, D. L. (2022b). Detection and classification of whiteflies and development stages on soybean leaves images using an improved deep learning strategy, *Computers and Electronics in Agriculture* **199**: 107132. <https://doi.org/10.1016/j.compag.2022.107132>.
- Goodfellow, I., Bengio, Y. and Courville, A. (2016). *Deep Learning*, 1 edn, Mit Press, Floor Cambridge, MA.
- He, K. and Zhang, X., Ren, S. and Sun, J. (2016). Deep residual learning for image recognition, in D. Ames, N. Quinn and A. Rizzoli (eds), *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA. <https://doi.org/10.1109/CVPR.2016.90>.
- Huang, G., Liu, Z. and Weinberger, K. Q. (2018). Densely connected convolutional networks, *CoRR abs/1608.06993*. <https://org/10.48550/arXiv.1608.06993>.
- LeCun, Y., Bengio, Y. and Hilton, G. (2015). Deep learning, *Nature* **521**: 436–444. <http://dx.doi.org/10.1038/nature14539>.
- Li, W., Zheng, T., Yang, Z., Li, M., Sun, C. and Yang, X. (2021). Classification and detection of insects from field images using deep learning for smart pest management: A systematic review, *Ecological Informatics* **66**: 101460. <http://doi.org/10.1016/j.ecoinf.2021.101460>.
- Mignoni, M. E., Honorato A, K. R., R, R. and A., M. (2022). Soybean images dataset for caterpillar and diabrotica speciosa pest detection and classification, *Data in Brief* **40**: 107756. <https://10.1016/j.dib.2021.107756>.
- Nanni, L., Manfè, A., Maguolo, G., Lumini, A. and Brahnam, S. (2022). High performing ensemble of convolutional neural networks for insect pest image detection, *Ecological Informatics* **67**: 101515. <http://doi.org/10.1016/j.ecoinf.2021.101515>.
- Patel, D. and Bhatt, N. (2021). Improved accuracy of pest detection using augmentation approach with faster r-cnn, *IOP Conference Series: Materials Science and Engineering*, ICMLSC 2020. Available at <http://10.1088/1757-899X/1042/1/012020>.

- Sedivy, E. J., Wu, F. and Hanzawa, Y. (2017). Soybean domestication: the origin, genetic architecture and molecular bases, *New Phytologist* **214**(2): 539–553. <http://doi.org/10.1111/nph.14418>.
- Sharma, S. M., Sharma, R. and Kimar, R. (2021). Pest detection in plants using convolutional neural network, *International Journal for Research in Applied Science & Engineering Technology* **9**(11): 1583–1592. <http://doi.org/10.22214/ijraset.2021.38890>.
- Silva, L. A. d., Bressan, P. O., Gonçalves, D. N., Freitas, D. M., Machado, B. B. and Gonçalves, W. N. (2019). Estimating soybean leaf defoliation using convolutional neural networks and synthetic images, *Computers and Electronics in Agriculture* **156**: 360–368. <https://doi.org/10.1016/j.compag.2018.11.040>.
- Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition, *arXiv*. <https://arxiv.org/abs/1409.1556>.
- Song, Y., Duan, X., Ren, Y., Xu, J., Luo, L. and Li, D. (2019). Identification of the agricultural pests based on deep learning models, *International Conference on Machine Learning, Big Data and Business Intelligence – MLBDI*.
- Tan, M. and Le, Q. V. (2020). Efficientnet: Rethinking model scaling for convolutional neural networks, *CoRR* **abs/1905.11946**. <https://doi.org/10.48550/arXiv.1905.11946>.
- Tassis, L. M., Tozzi de Souza, J. E. and Krohling, R. A. (2021). A deep learning approach combining instance and semantic segmentation to identify diseases and pests of coffee leaves from in-field images, *Computers and Electronics in Agriculture* **186**: 106191. <https://doi.org/10.1016/j.compag.2021.106191>.
- Tetila, E. C., Machado, B. B., Astolfi, G., de Souza Belete, N. A., Amorim, W. P., Roel, A. R. and Pistori, H. (2020). Detection and classification of soybean pests using deep learning with uav images, *Computers and Electronics in Agriculture* **179**: 105836. <https://doi.org/10.1016/j.compag.2020.105836>.
- Tetila, E. C., Machado, B., Bruno, Menezes, G. V., Alessandro de Souza Belete, N., Astolfi, G. and Pistori, H. (2019). A deep-learning approach for automatic counting of soybean insect pests, *IEEE Geoscience and Remote Sensing Letters* **17**(10): 1837–1841. <https://doi.org/10.1109/LGRS.2019.2954735>.
- Voutos, Y., Mylonas, P., Katheriotis, J. and Sofou, A. (2019). A survey on intelligent agricultural information handling methodologies, *Sustainability* **11**: 3278. <https://doi.org/10.3390/su1123278>.
- Wang, H., Li, Y., Minh Dang, L. and Moon, H. (2022). An efficient attention module for instance segmentation network in pest monitoring, *Computers and Electronics in Agriculture* **195**: 106853. <https://doi.org/10.1016/j.compag.2022.106853>.
- Xin, M. and Wang, Y. (2020). An image recognition algorithm of soybean diseases and insect pests based on migration learning and deep convolution network, *2020 International Wireless Communications and Mobile Computing (IWCMC)*. <https://doi.org/10.1109/IWCMC48107.2020.9148331>.
- YU, H. J. L., Chen, C., Heidari, A. A., Zhang, Q. and Chen, H. (2022). Optimized deep residual network system for diagnosing tomato pests, *Computers and Electronics in Agriculture* **195**: 106805. <http://doi.org/10.1016/j.compag.2022.106805>.
- Zhao, S., Liu, J., Bai, Z., Hu, C. and Jin, Y. (2022). Crop pest recognition in real agricultural environment using convolutional neural networks by a parallel attention mechanism, *Frontiers in Plant Science* **13**. <https://doi.org/10.3389/fpls.2022.839572>.