RGSA – Revista de Gestão Social e Ambiental
ISSN: 1981-982X
Submission date: 11/1/2024
Acceptance date: 1/3/2025
DOI: https://doi.org/10.24857/rgsa.v19n1-185
Organization: Interinstitutional Scientific Committee
Chief Editor: Ana Carolina Messias de Souza Ferreira da Costa
Assessment: Double Blind Review pelo SEER/OJS

DEEP LEARNING FOR DETECTION OF FOLIAR DISEASES IN SOYBEANS BASED ON THE MASK R-CNN MODEL

Ualace Vieira Gonçalves da Cruz[1]
Tiago do Carmo Nogueira[2]
Gelson da Cruz Junior[3]
Cássio Dener Noronha Vinhal[4]
Matheus Rudolfo Diedrich Ullmann[5]
Caio Henrique Rodrigues Carvalho[6]
Danyele de Oliveira Santana[7]

ABSTRACT

Objective: The objective of this study is to apply the Mask R-CNN model to detect and classify foliar diseases in soybean (*Glycine max* L.), assisting in the early diagnosis of powdery mildew, Asian soybean rust, and target spot.

Theoretical Framework: It was based on concepts of deep learning, machine learning, and instance segmentation, using ResNet-50 and ResNet-101 architectures, given the relevance of soybean and the impacts caused by foliar diseases.

Method: An initial dataset of 525 images was used, which was expanded to 2,408 through data augmentation techniques (resizing and rotation). The model integrated region proposal networks (RPN) and region of interest alignment (RoIAlign) to improve the segmentation and classification of images with foliar diseases. For model evaluation, the metrics mAP, IoU, and F1-Score were used.

Results and Discussion: The results revealed that the ResNet-101 architecture outperformed ResNet-50, achieving 85% accuracy in detecting foliar diseases in soybean. This outcome reinforces the applicability of artificial intelligence in the early diagnosis of foliar diseases in agriculture, particularly in soybean cultivation.

Research Implications: The research highlights the feasibility of deep learning-based models for agricultural management. Future studies may explore methods that eliminate the need for polygonal annotations and integrate technologies such as drones and IoT devices for greater scalability and efficiency.

Originality/Value: The research highlights the feasibility of deep learning-based models for agricultural management. Future studies may explore methods that eliminate the need for polygonal annotations and integrate technologies such as drones and IoT devices for greater scalability and efficiency.

Keywords: Convolutional Neural Networks, Mask R-CNN, FPN, RPN, Foliar Diseases, Soybean.

APRENDIZADO PROFUNDO BASEADO NO MODELO MASK R-CNN PARA DETECÇÃO DE DOENÇAS FOLIARES NA SOJA

RESUMO

Objetivo: O objetivo deste estudo é aplicar o modelo Mask R-CNN para detectar e classificar doenças foliares na soja (*Glycine max L.*), auxiliando no diagnóstico precoce das doenças *oídio*, *ferrugem asiática* e *mancha alvo*.

[1] Instituto Federal Baiano, Bom Jesus Da Lapa, Bahia, Brazil. E-mail: ualacegcd@gmail.com
[2] Universidade Federal De Goiás, Goiânia, Goiás, Brazil. E-mail: tiago.nogueira@ifbaiano.edu.br
[3] Universidade Federal De Goiás, Goiânia, Goiás, Brazil. E-mail: gcruzjr@ufg.br
[4] Universidade Federal De Goiás, Goiânia, Goiás, Brazil. E-mail: vinhal@ufg.br
[5] Instituto Federal Baiano, Barreiras, Bahia, Brazil. E-mail: matheusullmann@ifba.edu.br
[6] Instituto Federal Baiano, Bom Jesus Da Lapa, Bahia, Brazil. E-mail: caio.carvalho@ifbaiano.edu.br
[7] Instituto Federal Baiano, Bom Jesus Da Lapa, Bahia, Brazil. E-mail: danyele.santana@ifbaiano.edu.br

Referencial Teórico: Baseou-se em conceitos de aprendizado profundo, aprendizado de máquina e segmentação de instâncias, utilizando arquiteturas ResNet-50 e ResNet-101, dada a relevância da soja e os impactos causados por doenças foliares.

Método: Adotou-se, inicialmente, um conjunto de 525 imagens, sendo ampliada para 2.408 por meio de técnicas de aumento de dados (redimensionamento e rotação). O modelo integrou redes de proposta de região (RPN) e alinhamento de região de interesse (RoIAlign) para melhorar a segmentação e classificação de imagens com doenças foliares. Para avaliação do modelo, foram utilizadas as métricas mAP, IoU e F1-Score.

Resultados e Discussão: Os resultados obtidos revelaram que, a arquitetura ResNet-101 apresentou melhor desempenho quando comparada com a ResNet-50, alcançando assim, 85% de precisão na detecção das doenças foliares na soja. Esse resultado reforça a aplicabilidade da inteligência artificial no diagnóstico precoce de doenças foliares na agricultura, especialmente, na cultura da soja.

Implicações da Pesquisa: A pesquisa destaca a viabilidade de modelos baseados em aprendizado profundo para o manejo agrícola. Futuros estudos podem explorar métodos que dispensem anotações poligonais e integrar tecnologias como drones e dispositivos IoT para maior escalabilidade e eficiência.

Originalidade/Valor: A pesquisa combina aprendizado profundo e aumento de dados artificial, oferecendo avanços significativos na detecção de doenças foliares e impulsionando práticas agrícolas inovadoras.

Palavras-chave: Redes Neurais Convolucionais, Mask R-CNN, FPN, RPN, Doenças Foliares, Soja.

APRENDIZAJE PROFUNDO BASADO EN EL MODELO MASK R-CNN PARA LA DETECCIÓN DE ENFERMEDADES FOLIARES EN LA SOJA

RESUMEN

Objetivo: El objetivo de este estudio es aplicar el modelo Mask R-CNN para detectar y clasificar enfermedades foliares en la soja (Glycine max L.), ayudando en el diagnóstico temprano del oídio, la roya asiática y la mancha ojo de rana.

Marco Teórico: Se basó en conceptos de aprendizaje profundo, aprendizaje automático y segmentación de instancias, utilizando las arquitecturas ResNet-50 y ResNet-101, dada la relevancia de la soja y los impactos causados por las enfermedades foliares.

Método: Se utilizó inicialmente un conjunto de 525 imágenes, ampliado a 2.408 mediante técnicas de aumento de datos (redimensionamiento y rotación). El modelo integró redes de propuesta de región (RPN) y alineación de región de interés (RoIAlign) para mejorar la segmentación y clasificación de imágenes con enfermedades foliares. Para la evaluación del modelo, se utilizaron las métricas mAP, IoU y F1-Score.

Resultados y Discusión: Los resultados revelaron que la arquitectura ResNet-101 mostró un mejor rendimiento en comparación con ResNet-50, alcanzando un 85% de precisión en la detección de enfermedades foliares en la soja. Este resultado refuerza la aplicabilidad de la inteligencia artificial en el diagnóstico temprano de enfermedades foliares en la agricultura, especialmente en el cultivo de soja.

Implicaciones de la investigación: La investigación destaca la viabilidad de los modelos basados en aprendizaje profundo para la gestión agrícola. Estudios futuros pueden explorar métodos que eliminen la necesidad de anotaciones poligonales e integrar tecnologías como drones y dispositivos IoT para una mayor escalabilidad y eficiencia.

Originalidad/Valor: La investigación combina aprendizaje profundo y aumento de datos artificial, ofreciendo avances significativos en la detección de enfermedades foliares e impulsando prácticas agrícolas innovadoras.

Palabras clave: Redes Neuronales Convolucionales, Mask R-CNN, FPN, RPN, Enfermedades Foliares, Soja.

1 INTRODUCTION

According to the United Nations Development Program (UNDP) report, the human population will increase exponentially by 2050 (UNDP, 2021). Thus, world agriculture will need a 98% increase in its food production to meet this demand (Burra *et al*., 2021).

In this respect, soya (*glycine max L*) is one of the main grains used as a raw material in food production or in the integration of food components. In Brazil, soybeans have become one of the products with the highest export index. Brazilian soybean is exported to the countries of Europe, China and the United States (CONAB, 2024).

Despite its importance, soybeans are still one of the cultivars most susceptible to foliar diseases, such as *oidium*, *rust* "asiatic" and "*oidium*". Because of the aggressiveness and constant mutations, these leaf pests have become one of the main barriers faced by farmers in planting soybean (Goellner *et al*., 2010). Thus, the involvement of pests and foliar diseases in this crop, are damaging factors in its productivity on a large scale (Wang *et al*., 2016)

However, the lack of early treatment may result in significant economic losses in soybean production (Dananjayan *et al*., 2022). Thus, one of the ways of carrying out early diagnosis of these diseases are the applications of manual identifications by means of visual inspection by phytopathologists (Jackulin & Murugavalli, 2022).

However, soybean fields are relatively large and with an abysmal amount of plants, making it a difficult task for human eyes to properly detect and classify the state of each plant (Sharma *et al*., 2020). In addition, the possible damage of a poorly classified plant can harm an entire plantation, i.e., a single poorly inspected infected plant can spread throughout the plantation (Sharma *et al*., 2020).

In light of the above, a new form of early diagnosis of foliar diseases becomes increasingly necessary, and the application of technological resources may be a feasible path capable of achieving this goal (Dhanya *et al*., 2022). Artificial Intelligence (AI) can be one of the tools able to assist in the early diagnosis of foliar diseases, making it a powerful resource for soybean farmers (Lin *et al*., 2022; Pan *et al*., 2023).

Thus, in recent years, Artificial Intelligence has shown considerable development in several agricultural disease detection tasks, especially in the detection of foliar diseases and plant pests (Pan *et al*., 2023). Thus, several works in the literature use techniques based on Artificial Intelligence to accomplish these tasks. Machine learning (ML) and deep learning (DL)-based techniques stand out.

ML is a technique that involves the extraction of characteristics through previous learning, based on a database of image information (Lin et al., 2022; Pan et al., 2023; Rahman et al., 2023; Sahu & Pandey, 2023). While DL is based on a set of algorithms capable of modeling abstract data using several deep processing layers, composed of linear and non-linear functions (Tetila et al., 202 2020; Pan et al., 2023; Lin, et al., 2022; Haur,

ML- and DL-based models involve the collection and processing of image sets through artificial neural network (ANN) architectures for the identification of patterns in texture, coloration, and shapes (Krizhevsky *et al*., 2017; Dahiya *et al*., 2022 Harakannanavar *et al*., 2022). Thus, ANNs are mathematical models that simulate brain functions performed by neurons and synapses. Neural networks are trained in supervised ways through machine learning (Ferentinos, 2018).

Nonetheless, Convolutional Neural Network (CNNs) models are one of the alternatives to address the challenges encountered in traditional detection of pests and foliar diseases (Jadhav *et al*., 2021; Tetila *et al*., 2020; Pan et al., 2023), especially in soybean cultivars. Thus, CNNs are widely applied deep learning-based models for information extraction, performing image detection and segmentation tasks (Nogueira *et al*., 2020; Marzougui *et al*., 2020).

In this sense, the objective of this work is to propose a model based on deep learning by means of CNNs for the detection and classification of the diseases oidium (Microsphaera diffusa), Asian rust (Phakoptora *pachyrhizi) and targetStain* (Corynspora*Pachyrhizi) in images of soybean leaves.* To do so, we employ the Mask R-CNN for the segmentation of object instances into images of sick soybean leaves, using region proposal networks (RPN) of the original images as a guiding map for a second entry into the base through the Mask R-CNN architecture.

Key contributions in this article include the incorporation of the *Feature Pyramid Network* (FPN) architecture, which merges semantic and positional features from the input and bottom layers. This results in multi-scale resource maps, improving the performance metrics of the proposed neural network. In addition, we integrate a *region proposal network* layer (RPN) with the region of interest alignment layer (RoIAlign) to generate regions of interest (RoIs), thus producing bounding boxes for candidate objects and allowing the classification of multiple objects within each image entry.

This article is structured as follows: Section 2 analyzes studies related to ML and DL in the detection of foliar diseases in soybeans; Section 3 outlines the methodological procedures and instruments used in this study; Section 4 presents the results and discussions, while Section 5 concludes with *insights* about future research directions.

2 RELATED JOBS

In the literature, there are several papers that use models based on ML and DL for the task of detecting foliar diseases in soybeans (Farah et al., 2023; Furlanetto et al., 2021; Jadhav et al., 2021; Karlekar & Seal, 2020; Kumar et al., 2023; Lee et al., 2020; Mignoni et al et al.., 2023; Razfar et al., 2022; Tetila et al., 2020; Wang et al., 2023; Zhang et al., 2021). All studies implement different methods of CNNs for extracting information from the data.

One of the techniques used for training the CNNs-based model for the detection of foliar diseases is the separation of the imaging database into three groups: training; testing; and validation (Karlekar & Seal, 2020; Jadhav *et al*., 2021; Zhang *et al*., 2021). According to Karlekar & Seal (2020), data set separation increases both the accuracy and accuracy of the CNNs-based model.

During the preprocessing stage of images on CNNs, various techniques can be applied to increase the quantity of images during model training. This should be done because CNNs need a large amount of data for training in the process of detection and classification of foliar diseases (Li *et al*., 2022). To this end, rotation, inversion, brightness increase, addition of Gaussian noise, adaptive contrast-limited histogram equalization may be one of the techniques used for *artificial data augmentation during CNNs training (Pan et al*., 2023; Zhang *et al*., 2021). In addition to these techniques, black and white coloring can be applied to capture additional information from the images, assisting the data preprocessing phase in exponentializing the amount of input images in the CNNs (Karlekar & Seal, 2020).

One of the problems faced by researchers in the pre-processing of data during the training of CNNs is the identification of the total coverage of the input images in the neural models. To solve this problem, Wang *et al*. (2023) used a sliding segmentation algorithm, obtaining cuts in images that exceeded their corresponding size. This algorithm adjusts the size of the image by placing a pixel column horizontally or vertically.

Recent work adapts to pure CNNs with new deep learning techniques for performing disease detection and classification tasks (Jackulin & Murugavalli, 2022; Dananjayan *et al*., 2022; Kaur *et al*., 2022). New CNN models emerge based on the Inception, Resnet, VGG, Xception, YOLO, MobileNet and GoogleNet architectures (Tetila *et al*., 2020; Zhang *et al*., 2021; Razfar., 2022; Lee Tetilaet al., 2020).

Nevertheless, Pan *et al*. (2023) proposed a new CNN-based model called TFANet for identifying leaf diseases in images of soy leaves. TFANet includes some modules. A module consisting of two convolutional layers for extracting resources from the input images; and a

module based on the Inception architecture in order to improve the model's capability to perform features and an efficient channel attention component (ECA), promoting the model's attention to crucial aspects. This model showed an accuracy rate of between 98% and 100% in the identification of pests in the images of the soybean leaf.

Using models based on R-CNN, Huang *et al*. (2022) applied a new structure called Mask R-CNN for soybean seed classification. The proposed model was developed based on the Faster R-CNN, adding a segmentation branch to predict the masks of objects in each region of interest. This model was also proposed by Li *et al*., (2024) in a comparative study with different models, e.g. YOLO. The authors pointed out that the segmentation and area detection capability by a Mask R-CNN model are significant when compared to other models.

For Razfar *et al*. (2022), the simplification of the CNNs models makes it possible to train them in environments with limited resources. For example, CNN models based on MobileNet architectures. This architecture has a reduced amount of hyperparameter compared to other architectures. Because it has low complexity during gradient descent, the MobileNet architecture solves the problem with gradient leakage. Gradient leakage is a problem often identified in deep learning-based neural networks because of poorly selected activation functions, inefficient optimizers, or weak training strategies (Tan & Lim, 2020).

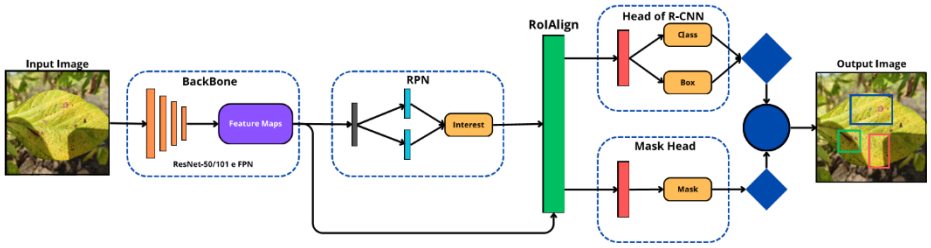
Therefore, there are several CNNs models capable of detecting foliar diseases. However, many models are based on complex architectures that require more processing power. Accordingly, it is still imperative to propose CNNs models for the automatic detection of foliar diseases in soybean crops, thus helping in the early diagnosis and increase the productivity of this crop. To do so, the next section will present a CNN model capable of achieving this goal.

3 METHODOLOGY

It is proposed in this work, the application of the model based on Mask R-CNN using the architectures Resnet-50 and Resnet-101 for the detection of the diseases *oidium, rust Asian andAsian stain and soybean* leaf *images.* Mask R-CNN is a deep learning algorithm used for the detection, targeting and tracking of objects in images or videos (Kumar *et al*., 2023). In addition, artificial data growth techniques were adopted through resizing, contrast, saturation, rotation, and vertical inversion of the data set images. As well as, techniques for optimizing information losses. The proposed model structure is visually outlined in Figure 1.

Figure 1

Methodological proposal based on the Mask R-CNN model.



For the task of feature extraction, in the backbone layer of the Mask R-CNN model, a pyramid of features (FPN) network aligned to the Resnet-50 and Resnet-101 architectures was applied. In addition, a Region Proposals Network (RPN) and Region of Interest Alignment (RoIAlign) were used to generate proposals from regions and interests. For model assessments the intersection on union (IoU) and loss function metrics will be applied.

3.1 DATA SET

According to Kumar *et al*. (2023) and Ahmad *et al*. (2023), there are several sets of data related to foliar diseases on the internet. In this study, two data sets were selected: a data set made available by Embrapa; and a foliar image bank made available on the *Kaggle.com* platform. Both sets of data are intended to be a reference in the development of methods for the automatic detection and recognition of plant diseases, and can be complemented and shared by the community of supporters.

Initially, the dataset consists of 175 leaf images of *Asian rust*, 175 of *oidium disease*, 65 of *the* healthy leaf *image* and 110 examples of healthy leaf images. Totaling 525 foliar images of soybeans. It is imperative to point out that, all images were inspected and labeled by specialists in plant pathology.

3.2 ANNOTATIONS AND DATA FORMATTING

For the training, testing and validation of the proposed Mask R-CNN Model, it was necessary to use files in JSON format containing the annotations of the image data set. These notes include Regions of Interest (ROI). The JSON files took place through an online tool VIA (VGG Image Annotator) and contained the coordinates of the polygons that delimited the

regions of the objects. In order to facilitate the implementation of the labels on the respective objects in the images.

The polygons are transformed into binary masks.which served as inputs to the model during the training, allowing the detection and classification of multiple classes (healthy, *Asian rust, stain- andoidium*).

3.3 DATA PREPROCESSING

The image data set presents significant variations in size, width and scenarios, and may be a challenge for the object classification and identification algorithms (Mahmood *et al*., 2020). One possible solution to this problem is the application of the image resizing technique (Sahu & Pandey, 2023). Thus, before starting data preprocessing, all the images in the data set were resized to 7x7. This allows the network to deliver better results in less time.

In the pre-processing phase, the data set was separated into three subsets: training set (50%); test set (20%); and validation set (30%). The choice of this strategy was made by previous studies that proved the effectiveness of the separation of the original dataset (Farah *et al*., 2023; Sahu & Pandey, 2023; Pan *et al*., 2023). Thus, by submitting the percentage of validation images in the proposed model by images that have not been trained, significant results can be obtained; that is, the deep learning model is more efficient when validation images are not known or were not used during testing and training of the artificial neural model.

Notably, the sample quantity of the original dataset becomes small for an efficient training process in deep neural networks. For this reason, techniques capable of expanding the number of samples in the data set were needed. In this sense, besides the resizing of the samples, the techniques of contrast, saturation, rotation and vertical inversion were used in the original data set. In addition to expanding the data set, the artificial data growth ensures that the proposed model generalizes efficiently during training (Farah *et al*., 2023)

After the application of artificial data growth techniques, the set resulted in 1,205 images for training, 481 for testing and 722 for validation (Table 1).

Table 1

Data set after application of artificial data growth techniques.

Type	Quantity	Training	Validation	Test
Target spot	623	312	186	125
Asian Rust	550	275	165	110
Oidium	535	268	161	106

Healthy	700	350	210	140
Total	**2,408**	**1,205**	**722**	**481**

3.4 MASK R-CNN BACKBONE LAYER

This subsection presents the backbone layer of the Mask R-CNN model composed of the ResNet-50, ResNet-101, and FPN architectures. These architectures are responsible for the task of extracting the characteristics of the incoming images of the proposed model. Thus, these architectures extract low-level (edge and corner) and high-level (semantic) characteristics using a ResNet-50+FPN or ResNet-101+FPN architecture.

3.4.1 Resnet-50 and Resnet-101 Architectures

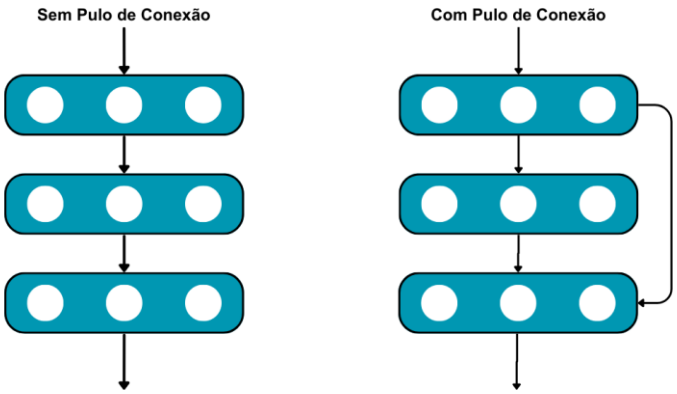
According to Mukti & Biswas (2019), deep learning models based on VGG, AlexNet and GooLeNet are widely used for performing learning transfer tasks. However, deep learning models based on these architectures experience problems with either the exploding gradient or the escaping gradient, in addition to optimization challenges. Because of the amount of layers that the information transits, sometimes the value of the gradient decreases dramatically, resulting in few changes in the weights of the artificial neural network (Wu *et al*., 2023).

Therefore, residual neural networks (ResNet) can be used to solve these problems. In this way, ResNet can increase the precision in the tasks of detecting and segmenting semantic objects in images with complex scenes (He *et al*., 2016), for example, images with foliar diseases with low resolution.

To solve the problem of gradient leakage, ResNets introduce a technique known as connection jumps. Connection jumps are techniques used to minimize the side effects caused by the leakage of the gradient present or gradient exploding on CNNs (Figure 2).

Figure 2

Technique jumps connections in ResNets.



This technique allows the model to jump layers during direct propagation, facilitating the passage of information along the network and helping to mitigate the problem of *leakage from the exploding gradient or gradient (Ankalaki et al*., 2023). The connection jump is performed after the completion of the previous layer and before the beginning of the posterior layer, allowing for a jump of an uncountable number of layers (Liu *et al*., 2022). This way, shortening the standard time and minimizing the information leak (Eq.1).

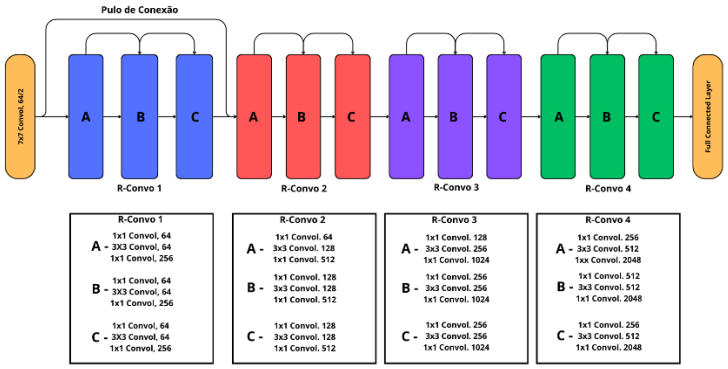
$$f(x) + x \qquad (1)$$

The jump calculation is performed using the formula *f(x) + x*. The dimensions of *x* are different from *f(x)* due to the convolution process, resulting in the reduction of their dimensions. Thus, an additional 1+1 convolution layer is added to change the dimensions of *x*. In this way, each residual network follows the same patterns, possessing a convolutional layer of 3x3, proceeding to the RPN and ROIAlign, sequentially.

In this work, the architectures ResNet-50 and ResNet-101 were used. Thus, the ResNet-50 is composed of 49 identical layers known as residual. The first layer supports 64 filters of sizes 7x7, followed by a *max-pooling layer* of 3x3. The ResNet-50 architecture has a total of 49 convolutional layers (*conv*) and a fully connected layer (*fc*) responsible for classification (Ikechukwu et al., 2021; Wu-*th, 2023) (Figure 3).*

Figure 3

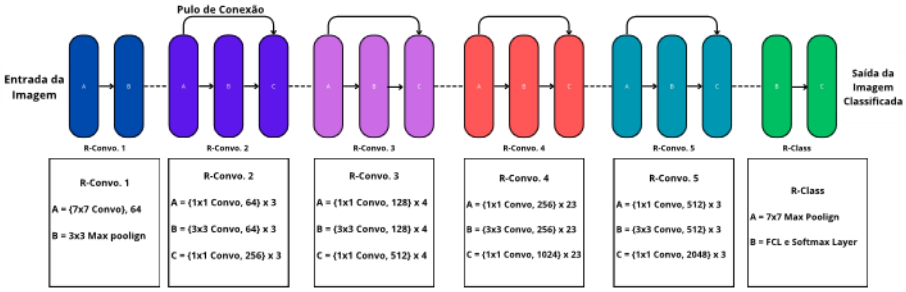
Structure of the ResNet-50 architecture.



While the ResNet-101 presents some differences when compared to the ResNet-50. These differences can be noted in the larger number of deep, fully connected layers (*fc*) (Figure 4).

Figure 4

Structure of the ResNet-101 architecture.



Thus, the ResNet-101 architecture has a total of 101 seizure layers responsible for the detection and classification of objects (Vaishali & Neetu, 2024). The ResNet-101 consists of the maximum grouping layer, before decreasing to 56x56 in the first convolutional layer; 28x28 in the second convolutional layer; 14x14 in the third convolutional layer and finally 7x7 in the final convolutional layer. The application of this architecture can generate a reduction in computational complexity, improving error rates in object detection and classification tasks (Vaishali & Neetu, 2024; Kumar *et al*., 2024).

3.4.2 FPN architecture

The FPN architecture was implemented with the intention of blending the semantic characteristics of the input images of the upper layer with the positional characteristics of the lower layer (Li *et al*., 2020), obtaining characteristic maps on different scales. This architecture can be used in conjunction with ResNet-50 and ResNet-101. Notably, NPFs are composed of a ResNet that transforms spatial width versus height into images, followed by another inverted residual network (Li *et al*., 2020). In this way, FPN is able to construct a pyramid of semantic characteristics at a high level of the input images in the Mask R-CNN model. Your formula can be defined by:

$$k = [k_0 + log_2 \left(\frac{wh}{224}\right)] \qquad (2)$$

where, formally, a region of interest (RoI) of width (*w*) and height (*h*) in the input image of the neural network can be assigned at the probability level of k in the pyramid of characteristics; $k_0$it can be defined as being the target level of RoI and 224 is the canonical size of the pre-trained ResNet-50 or ResNet-101. This architecture can be aligned with the RPN, since the FPN is in charge of extracting the characteristics, efficiently detecting the objects contained in the input images and generating a map of characteristics for the later layers (Lin *et al*., 2017; Li *et al*., 2020), for example, for the fully connected layer of the RPN. At this stage, the RPN characteristic map was replaced by the FPN (Lin *et al*., 2017), converting the convolutional input layers of size 3x3, by two identical layers of size 1x1 for each level of the characteristic pyramid. As in the proposal of (Li *et al*., 2020), we apply multiple anchors of proportion P2, P3, P4, P5 and P6.

Therefore, the provision of the pyramids of characteristics from the FPN architectures in conjunction with the ResNet-50 and ResNet-101 bring improvement in the performance evaluation metrics of the proposed model when aligned to the RPN.

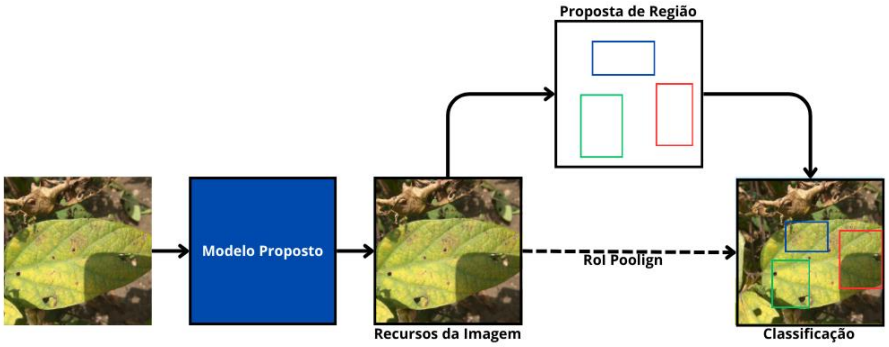
3.5 REGION PROPOSAL NETWORK (RPN) LAYER

The RPN is a model used for the development of candidate regions. This network allows the creation of areas of interest in an image, diverging from the standard selective approaches. RPN is simple and fast compared to bounding box generation approaches. This network consists

of the division by areas into multiple parts, and may reach thousands of regions as in the study by (Kaur *et al*., 2022; Deepika & Arthi, 2022) (Figure 5).

Figure 5

Region Proposal Network (RPN).



In this proposal, the RPN receives the pyramids of characteristics coming from the stage of extraction of the characteristics by way of the FPN, ResNet50 and ResNet-101, generating proposals of areas of interest. These areas will be divided into proposals of regions and regions of interest *for the training of the proposed model.* Many studies apply this approach in conjunction with RoI for the detection and definition of regions of interest in images (Zhang *et al*., 2021; Lai *et al*., 2022; Cao et al., 2023).

RPN makes it possible to generate limitation boxes for candidate objects, by means of characteristic maps produced by the main network. These boxes are titled as anchor boxes (Kumar *et al*., 2023). ABs vary in height (*h*) and width (*w*), and can classify more than one object of interest per image (Figure 6).

Figure 6

Anchor Boxes (ABs) in the Regional Proposal Network (RPN).

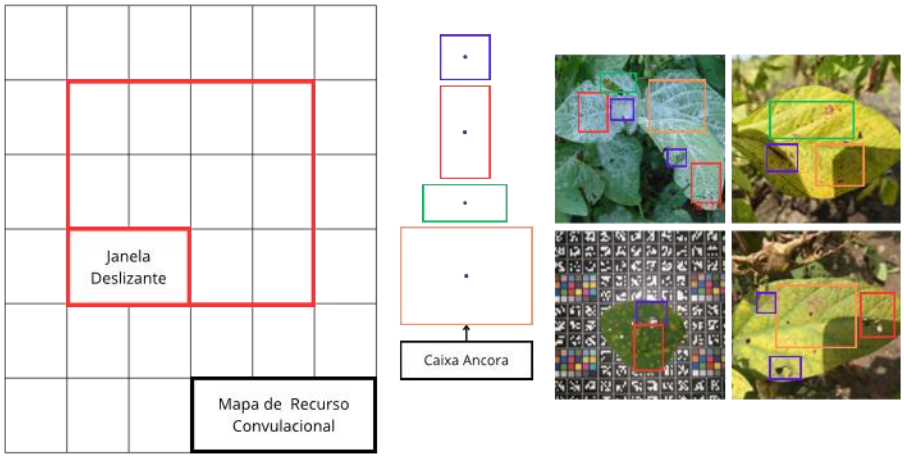


Figure 6 demonstrates an anchor box regression process. Thus, regression refers to the procedure of obtaining the predicted box, adapting the artificial anchor box to get closer to the true box (Chu *et al*., 2021). In this first phase, RPN indicates all pixel points in the image with bounding boxes. Afterwards, the size of the anchor boxes is adjusted to obtain predictive boxes that indicate where the objects are located.

The lower branch is trained by bounding box regression with labeled image samples that contain true boxes, denoting the actual location of the objects (Kumar *et al*., 2023). This training aims to reduce the differences between the location of the forecasted boxes and the location of the real boxes, resulting in more accurate forecasted boxes.

Many studies apply metrics to evaluate overlap of boxes, this overlap is an important problem for studies that apply RPN and FPN. This problem directly implies with the generation of boxes, causing them to be generated too *close and often impairing accuracy (Rezatofighi et al*., 2019). One of the ways of evaluating the overlap of anchor boxes is by using the intersection on union (IoU).

Finally, the result obtained by the RPN is deposited in the RoIAlign layer, producing 7x7 resources. In this way, you can get the final result after passing through two connected layers. With this, the PNR has a significant probability of the bounding box being captured by means of the classification layer and the regression layer, locating the disease involved in the leaves of the soybeans.

3.6 REGION OF INTEREST ALIGNMENT LAYER (ROIALIGN)

RoIAlign works in conjunction with RPN. In this way, the RoIAlign receives the images passed through the totally convolutional layers (*fc*) of the RPN, transforming them into a dimensional matrix and defining the regions of interest (RoI). After receiving the matrices of areas of interest, this network will be responsible for the creation of two branches (minimatrizes of interests); these subsidiaries will be responsible for the classification of the zone of interest and the recession of the border (Li *et al*., 2020) (Figure 7).

Figure 7

Alignment of the region of interest (RoIAlign).

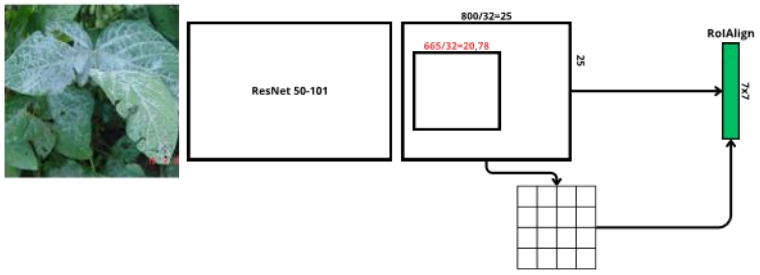
This technique should flag whether the object in a region proposal is a disease candidate or not. To accomplish this action, an ROI grouping layer maps the output region proposals of RPN that have different sizes to a characteristic map of the same size, that is, to fixed-size characteristic maps (Shu *et al*., 2020).

In this way, RoIAlign works together with RPN to refine the image, correctly classifying it according to its regions of interest. RoIAlign is an enhancement to RoI clustering developed by Mask R-CNN to correct rounding errors (Zhao *et al*., 2020), where a floating point ROI is quantized by a characteristic map. Next, quantization is subdivided by performing network aggregation processing. The subdivisions refer to a calculation of [x/16] of a continuous coordinate. Where the number 16 represents the size of the characteristic map and x is the rounding operator.

The RoIAlign layer replaces rough quantization of ROI with stable quantization, thus establishing an alignment relationship between the extracted resources and the input (Figure 8).

Figure 8

ROI teaming operation process.



The ROI edge and mesh are not quantized, where *x*/16 is used instead of [*x*/16] for stable quantization. In this case, the exact value of the input resource is calculated by bilinear interpolation at the four regular sampling positions of each ROI grid. The retropropagation equation for the ROI pool is given by:

$$\frac{\partial L}{\partial x_i} = \sum_r \sum_j [i = i * (r,j)] \frac{\partial L}{\partial y_{rj}} \qquad (3)$$

$$\frac{\partial L}{\partial x_i} = \sum_r \sum_j [d(i, i * (r,j)) < 1](1 - \Delta h)(1 - \Delta w)(1 - \Delta w) \frac{\partial L}{\partial y_{rj}} \qquad (4)$$

where x_i represents the pixel in the resource map before grouping; y_{rj} is the r_{th} represents region of interest after grouping; $i * (r,j)$ sets the source of the pixel value of y_{rj}. Therefore, equations (3) and (4) are responsible for showing the value of the current candidate pixel pointed by x_i, used in the clustering process. In this case, the gradient can be returned to x_i. In addition, RoIAlign works in conjunction with the most diverse operations that are applied to regions of interest (ROIs) in an image that can accurately extract characteristics.

Bilinear Interpolation allows you to meet the need to extract features accurately by combining or comparing two completely different digital images. Whether it is size or dimension, interpolation is one of the alternatives that become feasible to meet this need.

$$f(x,y) = a + bx + cy + dx^y \qquad (5)$$

Equation (5) is used to find an arbitrary point, represented by (*x*, *y*). Note that the *x* spacing *does* not have to be the same as the *y* spacing. In this case, the coefficients *a*, *b*, c, and the coefficients *are determined by the values of the image as follows:*

$$f\,11 = a + bx^1 + cy^1 + dx^1y^1 \qquad (6)$$

$$f\,12 = a + bx^1 + cy^2 + dx^1y^2 \quad (7)$$

$$f\,22 = a + bx^2 + cy^2 + dx^2y^2 \quad (8)$$

$$f\,21 = a + bx^2 + cy^1 + dx^2y^1 \quad (9)$$

Finally, after having found a value for all four coefficients (*a*, *b*, c andd), function $f(x, y)$at any point within the grid can be obtained.

3.7 MASK R-CNN CLASSIFICATION LAYER

One of the major differences between Mask R-CNN and Faster R-CNN is related to the classification layer. Thus, the Mask R-CNN has a specific layer for classification, extracting the characteristics of the regions of interest (ROI) from each input image, using region of interest alignment (RoIAlign). For refinement, ROI-extracted characteristics are subjected to a series of fully interconnected convolutional layers (*fc*). The extracted characteristics are then processed and fed into one or more layers of masks in the Mask R-CNN model. *Thus, the layers of this model will generate binary masks for each ROI, representing the segmentation of the objects in the image (i*). After this process, during training, an input image (*i*) is passed to the *backbone layer*. In the *backbone* layer, a pre-trained convolutional network was implemented based on the ResNet-50 and ResNet-101 architectures.

In this sense, architectures implemented in the *backbone* layer must extract the resources from the input layer (*i*) and pass them on to RPN. The RPN will be responsible for generating several ROIs, employing a CNN and a binary classifier. This is due to the insertion of nine anchor boxes on the image (*i*) and the binary scores generated by the classifier.

In this way, the Mask R-CNN model is able to perform hundreds of predictions, anticipating numerous instances of objects in an input (*i*) image. Thus, the final identification of the object is carried out by eliminating the background anchor boxes, filtering the residual anchor boxes, using the confidence scores. Finally, the binary classifier informs whether or not the characteristic map of this pre-trained network includes soybean leaf disease.

3.7.1 Average Accuracy (mAP)

Average precision (mAP) is commonly used to calculate the performance of an object detection model. The mAP metric can be given as follows.

$$\text{mAP} = \frac{1}{\text{N}} \sum_{\text{i=1}}^{\text{N}} \text{AP}_\text{i}\text{AP} = \int_0^1 \text{p(r)drAP} = \sum_\text{n} (\text{R}_\text{n} - \text{R}_{\text{n=1}}) \cdot \text{P}_\text{n} \quad (10)$$

For each reminder rate, mAP generates a precision value. The recall rate evaluates the ability to identify all positive results, while the accuracy rate measures the accuracy of forecasts (WU *et al*., 2024).

3.7.2 Recall Rate

The recall rate can be represented by the true positives (PT), representing the numbers of detections under the false positives (FP) rate which are the numbers of detections that are initially given as negative but detected as positive.

$$Lembrança = \frac{TP}{TP + FN} = \frac{TP}{Ground\ Truth} \quad (11)$$

False negatives (FN), in turn, may be the number of detections that were initially identified as positive, but are later classified as negative by the model.

3.7.3 F1-Score

This metric is applied to average the accuracy and recall metrics. The formula for this metric is structured as follows:

$$F1 - Score = 2 \times \frac{Precisão \times Lembrança}{Precisão \times Lembrança} \quad (12)$$

Being quite useful in cases where there are minority classes, this occurs when one of the labels appears less often than others. The F1-Score helps the model not be deceived by good

performances in majority classes where accuracy can be high, but the model fails in less frequent instances.

3.7.4 Precision Rate (Precision)

The accuracy rate is used to identify the correct predictions by considering all objects detected in a given image (Siricharoen *et al*., 2023). As shown in Figure 17. Where TP is the number of objects detected correctly in the image. While FP is the number of incorrect objects or non-existent objects (YI *et al*., 2024).

$$Precisão = \frac{TP}{TP + FP} = \frac{TP}{Total\ Predição} \quad (13)$$

Furthermore, the object detector confidence score can be considered during the prediction process, considering positive detections (Siricharoen *et al*., 2023).

3.7.5 Intersection Rate on Union (IoU)

The IoU is a metric widely used when researchers want to evaluate the object prediction of detection and segmentation algorithms (Rezatofighi *et al*., 2019). When an object is detected and segmented, the algorithm generates a bounding box around the object and produces a binary mask that represents the exact pixels that belong to that object (Wu *et al*., 2024).

To calculate the IOU, the area of intersection between the masks of the predicted truth and the basic truth is divided by the area of union between two masks. IoU ranges from 0 and 1, following binary patterns. Where 0 indicates no overlap between the masks and 1 indicates perfect overlap. A higher IoU score indicates a better algorithm targeting performance (Eq. 14).

$$IoU = \frac{TP}{TP + FP + FN} \quad (14)$$

Briefly, true positive (PT) means that the detection box is correct; false positive (FP) means that a detection box is false; and false negative (FN) incorporates the lost detection box.

4 RESULTS AND DISCUSSIONS

In this research, the application of the Mask R-CNN model was addressed for the task of detecting and segmenting objects into images. Subsection 4.1 details the implementation of the proposed model, providing an overview of the techniques and configurations used in building and training the model, including the architecture and parameters tuned to optimize performance. In addition, in Subsection 4.2, it examines the model's effectiveness in identifying and precision bounding boxes with respect to the classes of objects present, discussing the evaluation metrics that reflect the model's ability to accurately classify and locate. Finally, Subsection 4.3 presents an analysis of the final results obtained, including the precision and Media Accuracy (mAP), the recall rate, the intersection on union (IoU) and the F1-Score, providing a comprehensive view of the model's performance and implications for practical application.

4.1 DETAILS OF THE IMPLEMENTATION OF THE MASK R-CNN MODEL

The Mask R-CNN model with the Resnet-101 was the first to be used to classify and detect 04 (four) different types of foliar diseases in soybeans. The set of images covered two scenarios: leaves photographed in the laboratory and in the field. Model performance was assessed using precision, recall and average accuracy (mAP) metrics. These metrics were calculated for each class and the set as a whole.

In order to guarantee the quality of the detections made by the Mask R-CNN model, a confidence limit (*threshold*) of 0.8 was defined. This parameter acts as a filter to determine which detections the model should consider valid during the inference process. Specifically, the model of this research only accepted detections whose class predictions showed a confidence of 80% (eighty percent) or more. During the experiments, this limit proved to be effective for the use case studied, in which the priority was to guarantee high quality detections.

In the training process, it was decided that the class detection and classification step would occur separately, due to problems inherent in overlapping classes. Therefore, the study followed the flow of training, starting with the set of images of healthy soy leaves, followed by the *rust* classes, *stain*-target and

4.2 RATING AND LOCALIZATION PERFORMANCE

The performance of detections of the original soybean datasets is presented in Table 2 and 3.

Table 2

Performance metrics of the ResNet-101 model for detecting foliar diseases (healthy, Asian rust, oidium and target spot classes).

Class	mAP	Recall	Precision	IoU	F1-Score
Healthy	0,875	0,907	0,984	0,654	0,916
Asian Rust	0,854	0,874	0,994	0,648	0,955
Oidium	0,786	0,756	0,856	0,678	0,901
Target Spot	0,705	0,687	0,568	0,550	0,627

Table 3

Performance metrics of the ResNet-50 model for detecting foliar diseases (healthy, Asian rust, oidium and target spot classes).

Class	mAP	Recall	Precision	IoU	F1-Score
Healthy	0,752	0,802	0,794	0,740	0,715
Asian Rust	0,806	0,861	0,824	0,718	0,768
Oidium	0,701	0,745	0,703	0,610	0,722
Target Spot	0,634	0,704	0,684	0,601	0,694

The bounding boxes act to detect the object in the image, using as a reference point the polygons imported from the JSON file annotations. With this, the bounding boxes wrapped the soybean leaf, and then, with the object already delimited, the model started the training process, where he would learn the visual characteristics of the leaf. Based on these characteristics, the image could be classified. The bounding boxes can be viewed in Figure 9.

Figure 9

Bounding boxes for intersection of objects in the image of soy leaves.

In the healthy class, the values of the samples captured by the model are an excellent indication of the capacity of this model to differentiate the healthy images from those affected by the other diseases. The values presented in Table 2 refer to Resnet-101 and in Table 3 belong to Resnet-50. Both are a view of the model's final results, an overall measure of the model's final capacity.

As shown in Tables 2 and 3, the proposed models have yielded significant results. In addition, it can be seen that the probability filter presented itself as a viable technique to assist in achieving a classification quality considered high. The results presented in tables 2 and 3 are regarded as excellent and demonstrate that the models are capable of identifying which specimens are healthy and which specimens show the disease.

4.3 IN-DEPTH ANALYSIS OF THE MODEL

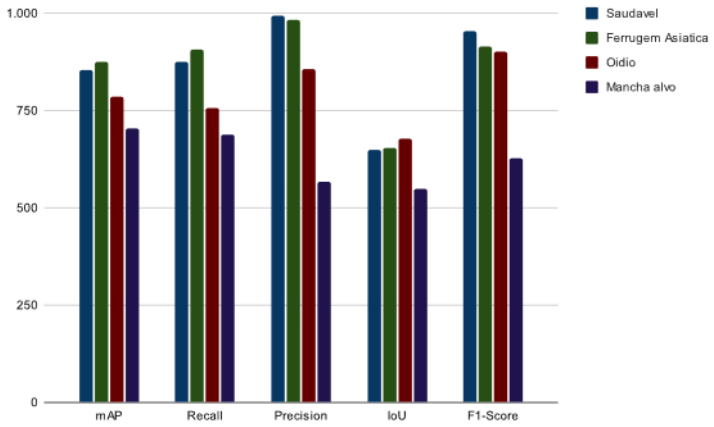
The results indicate that the Mask R-CNN model with the ResNet-101 architecture exhibited considerable performance for the healthy, *Asian rust* and *oidium*classes, as evidenced by high metric values: precision (0.994, 0.984 and 0.856) and F1-Score (0.955, 0.916 and 0.925), as shown in Table 2. This suggests that the model is highly effective in identifying the distinguishing characteristics of each class, thus minimizing the probability of false positives.

In addition, the results obtained with the ResNet-50 deserve consideration. As shown in Table 3, although the values remain below 90% (ninety percent), they are stable, with narrow intervals between metrics indicating that the model rarely presented difficulties in one class and performed well in others.

However, the *target spot* class performed below average on all metrics for both architectures. Specifically, in the ResNet-101 architecture, mAP, memory, and accuracy were recorded as 0.705, 0.687, and 0.568, respectively, while ResNet-50 had values of 0.634, 0.704, and 0.684 (Figure 10).

Figure 10

Performance comparison of the ResNet-101 model in metrics for detection of foliar diseases, evaluating four classes: healthy, Asian rust, oidium and target spot.



This underlines the challenges of the model in detecting the *target spot* class with the same precision as the other classes (*oidium*, The low IoU scores of 0.550 and 0.601 further suggest that the bounding boxes predicted for this class do not align well with the reference bounding boxes. One factor contributing to these detection challenges may be the low quality and quantity of images available for this disease.

Figure 11

Performance of the ResNet-101 model in different metrics for each class in the detection of foliar diseases, including healthy Asian rust, oidium and target spot.

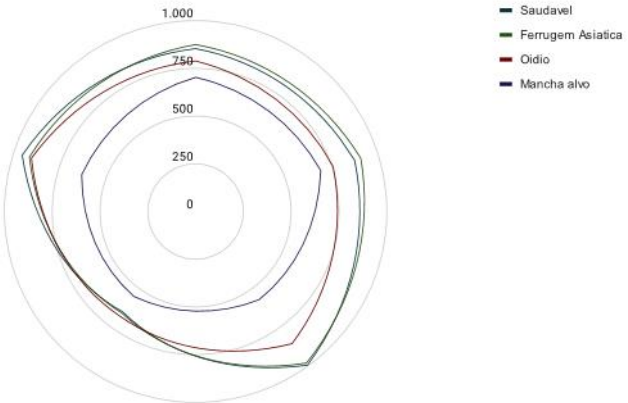
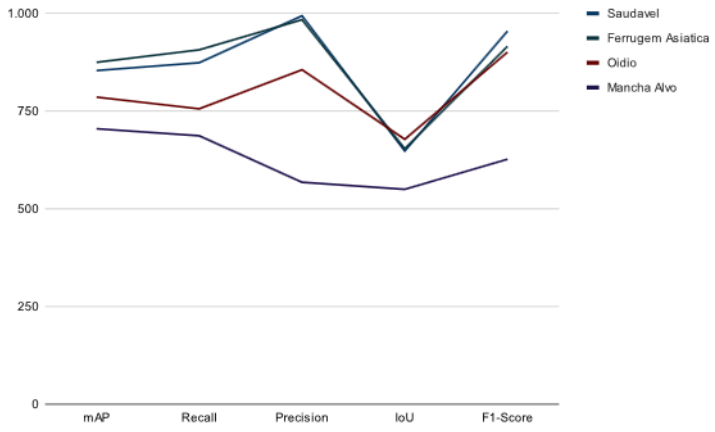


Figure 12

Analysis of metrics trends in the ResNet-101 model for detection of foliar diseases, covering healthy Asian rust, oidium and target spot classes.



An analysis of the trend of the metrics by class, as illustrated in Figures 11 and 12, reveals that the model stands out in the *Asian* and healthy *rust* classes. On the other hand, the *target stain* classes and *oidium* illustrate the model's difficulties in managing these categories effectively. These findings imply that improved data balance is needed during training to improve the performance of these specific classes.

According to Lai *et al*. (2022), sub-optimal performance of deep learning models is often attributed to a limited number of images in the training data set. A workable solution to this problem is the implementation of techniques that can artificially augment the data set images or the application of *K-fold* cross-validation

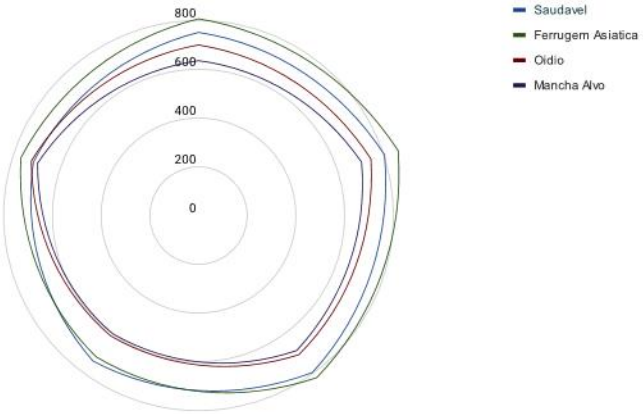
The performance metrics presented in Table 2 indicate that accuracy and F1-Score are the most effective measures to detect foliar diseases using the model. Despite the presence of multiple classes, the models present greater consistency in performance.

The ResNet-50 architecture demonstrated stable performance as shown in Table 3. The performance metrics were consistent with each other, contrasting with the ResNet-101 model, which showed high values for some classes and lower values for others. This discrepancy can be attributed to factors such as model sensitivity and the need for parameter adjustments in the ResNet-101 architecture.

In Figure 13, a comparison of the ResNet-50 architecture with the ResNet-101 reveals that the metrics trends achieved by the ResNet-50 are slightly lower. However, the performance remains balanced and stable, suggesting that the model did not encounter significant difficulties in detecting the different classes (*oidium, rust Asian,*

Figure 13

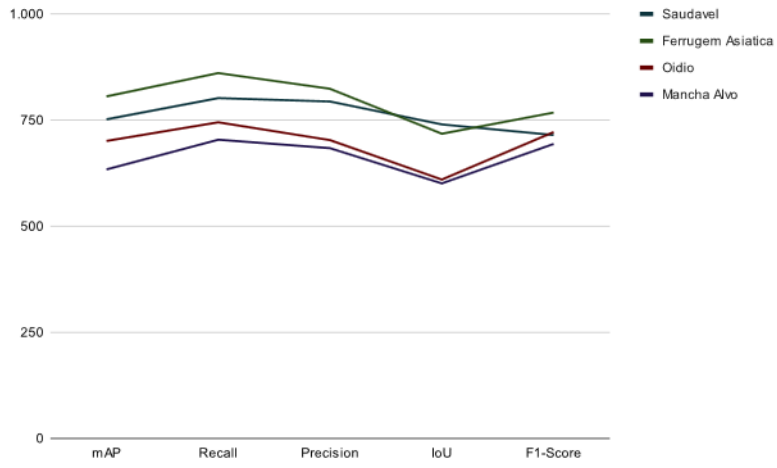
Performance of the ResNet-50 model in different metrics for each class in the detection of foliar diseases, including healthy Asian rust, oidium and target spot.



When analyzing class performance using the ResNet-50 architecture, the metrics represented in Figure 14 indicate that the *target spot* classes and *ocide* show the lowest performance. In this context, the ResNet-50 features standards similar to those of the ResNet-101, implying that both architectures may be facing common challenges in detecting and locating these specific classes.

Figure 14

Analysis of metrics trends in the ResNet-50 model for detection of foliar diseases, covering healthy Asian rust, oidium and target spot classes.

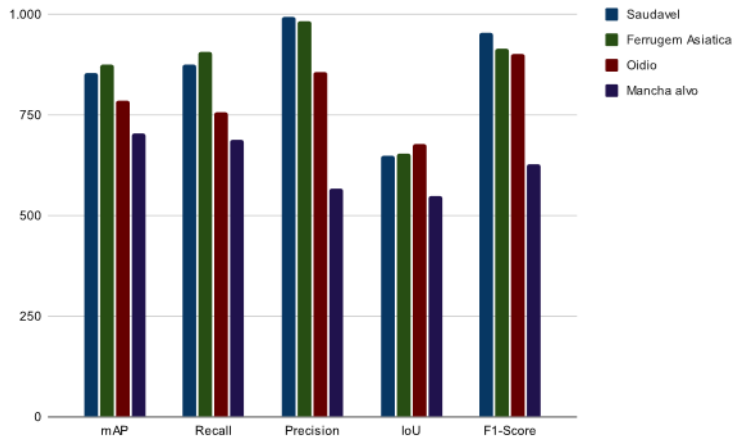


The *target spot* class records the lowest values in all metrics, including mAP, recall, precision, IoU, and F1-Score. This highlights substantial challenges in identifying and locating

precise instances of this class, resulting in greater difficulty for the model. Likewise, the *oidium* class performs poorly in relation to the other classes, although with a relatively high IoU. This observation suggests that although the model may reach a satisfactory location, it has suffered from the low accuracy and recall (Recall) rates.

Figure 15

Performance comparison of the ResNet-50 model in metrics for detecting foliar diseases, evaluating four classes: healthy, Asian rust, oidium and target spot.



In summary, the ResNet-50 architecture exhibited distinct metric performance compared to the ResNet-101. While on the ResNet-101 the accuracy and F1-Score metrics stood out as the highest, on the ResNet-50, the recall rate (Recall) and the accuracy average (mAP) stood out, occupying the outstanding positions, as we can see in Figure 15. These results indicate that the model with ResNet-50 was also effective in identifying the main characteristics and instances of the classes, demonstrating their comprehensive detection capability.

5 CONCLUSION

The approach proposed in this study for the detection of foliar diseases was developed and tested using a Mask R-CNN network with instance segmentation, achieving a robust performance in the identification of different diseases in soybean leaves. Accuracy rates ranged from 86.7% for healthy samples to 86.5% for *rust*, 79.5% for *oidium*, and 62.7% for . *Target spot* detection could be improved with a more diverse set of images, including varied field and laboratory conditions.

Considerations should be taken that the study faced obstacles during the capturing of results, the set of images and the hypersensitivity present in the Resnet-101 model. The constant changes in the adjustment and monitoring of the loss curves is one of the challenges encountered for those who want to use the Resnet-101. Nevertheless, the set of images used in the studies also contributed to the challenge found in this study, the quantity made it difficult to partition and train the models, because both models needed an exorbitant amount of images, our work suffered from misfortunes and this consequently affected the results.

This study demonstrates that deep learning models are feasible and promising for the automatic detection of foliar diseases, bringing significant benefits to precision agriculture. However, the requirement for polygons in the training data represents a limitation. Future research may focus on segmentation techniques without the need for polygonal annotations, thus increasing applicability in broader and more diverse databases.

The model's ability to differentiate the various diseases of soybeans (*glycine max*) reinforces the potential of artificial intelligence in agriculture. In the future, combining this model with real-time monitoring systems such as drones and IoT devices could offer a complete solution for crop management, increasing productivity and reducing operating costs.

REFERENCES

Ahmad, A.; Gamal, A. E.; Saraswat, D. Toward generalization of deep learning-based plant disease identification under controlled and field conditions. *IEEE Access*, v. 11, p. 19332–19341, 2023.

Ankalaki, S.; Thippeswamy, M. A novel optimized parametric hyperbolic tangent swish activation function for 1d-cnn: application of sensor-based human activity recognition and anomaly detection. *Multimedia Tools and Applications*, Springer, p.1–31, 2023.

Burra, D. D. et al. Digital agriculture profile: Viet nam. Food and Agriculture Organization of the United Nations, 2021.

Cao, Y. et al. Case instance segmentation of small farmland based on mask r-cnn of feature pyramid network with double attention mechanism in high resolution satellite images. *Computers and Electronics in Agriculture*, v. 212, p. 108073, 2023.

Chu, P. et al. Deep learning-based apple detection using a suppression mask r-cnn. *Pattern Recognition Letters*, v. 147, p. 206–211, 2021.

Conab. *Histórico Mensal AgroConab*. 2024. Disponível em Conab. Disponível em: ⟨https://www.conab.gov.br/info-agro/analises-do-mercado-agropecuario-e-extrativista/analises-do-mercado/historico-mensal-de-agroconab⟩.

Dahiya, S.; Gulati, T.; Gupta, D. Performance analysis of deep learning architectures for plant leaves disease detection. *Measurement: Sensors*, Elsevier, v. 24, p. 100581, 2022.

Dananjayan, S. et al. Assessment of state-of-the-art deep learning based citrus disease detection techniques using annotated optical leaf images. *Computers and Electronics in Agriculture*, Elsevier, v. 193, p. 106658, 2022.

Deepika, P.; Arthi, B. Prediction of plant pest detection using improved mask frcnn in cloud environment. *Measurement: Sensors*, Elsevier, v. 24, p. 100549, 2022.

Dhanya, V. et al. Deep learning based computer vision approaches for smart agricultural applications. *Artificial Intelligence in Agriculture*, Elsevier, v. 6, p. 211–229, 2022.

Dosovitskiy, A. et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

Farah, N. et al. A deep learning-based approach for the detection of infested soybean leaves. *IEEE Access*, 2023.

Ferentinos, K. P. Deep learning models for plant disease detection and diagnosis. *Computers and electronics in agriculture*, Elsevier, v. 145, p. 311–318, 2018.

Furlanetto, R. H. et al. Identification and classification of asian soybean rust using leaf-based hyperspectral reflectance. *International Journal of Remote Sensing*, v. 42, n. 11, p. 4177–4198, 2021.

Goellner, K. et al. Phakopsora pachyrhizi, the causal agent of asian soybean rust. *Molecular plant pathology*, v. 11, n. 2, p. 169–177, 2010.

Harakannanavar, S. S. et al. Plant leaf disease detection using computer vision and machine learning algorithms. *Global Transitions Proceedings*, v. 3, n. 1, p. 305–310, 2022.

He, K. et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 770–778.

Huang, Z. et al. Deep learning based soybean seed classification. *Computers and Electronics in Agriculture*, v. 202, p. 107393, 2022.

Ikechukwu, A. V. et al. 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*, Elsevier, v. 2, n. 2, p. 375–381, 2021.

Jackulin, C.; Murugavalli, S. A comprehensive review on detection of plant disease using machine learning and deep learning approaches. *Measurement: Sensors*, Elsevier, v. 24, p. 100441, 2022.

Jadhav, S.; Udipi, V.; Patil, S. Classification of soybean diseases using pre-trained deep convolutional neural networks. p. 746–756, 2021.

Jadhav, S. B.; Udipi, V. R.; Patil, S. B. Identification of plant diseases using convolutional neural networks. *International Journal of Information Technology*, v. 13, n. 6, p. 2461–2470, 2021.

Karlekar, A.; Seal, A. Soynet: Soybean leaf diseases classification. *Computers and Electronics in Agriculture*, Elsevier, v. 172, p. 105342, 2020.

Kaur, P. et al. An approach for characterization of infected area in tomato leaf disease based on deep learning and object detection technique. *Engineering Applications of Artificial Intelligence*, Elsevier, v. 115, p. 105210, 2022.

Krizhevsky, A.; Sutskever, I.; Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, AcM New York, NY, USA, v. 60, n. 6, p. 84–90, 2017.

Kumar, M. et al. Soybean disease detection and segmentation based on mask-rcnn algorithm. *J. Exp. Agric. Int*, v. 45, n. 5, p. 63–72, 2023.

Kumar, V. et al. Unified deep learning models for enhanced lung cancer prediction with resnet-50–101 and efficientnet-b3 using dicom images. *BMC Medical Imaging*, Springer, v. 24, n. 1, p. 63, 2024.

Lai, W. et al. The study of coal gangue segmentation for location and shape predicts based on multispectral and improved mask r-cnn. *Powder Technology*, v. 407, p. 117655, 2022.

Lee, S. H. et al. New perspectives on plant disease characterization based on deep learning. *Computers and Electronics in Agriculture*, v. 170, p. 105220, 2020.

Li, H. et al. Symptom recognition of disease and insect damage based on mask r-cnn, wavelet transform, and f-rnet. *Frontiers in Plant Science*, v. 13, p. 922797, 2022.

Li, H. et al. Establishment of deep learning model for the growth of tea cutting seedlings based on hyperspectral imaging technique. *Scientia Horticulturae*, v. 331, p. 113106, 2024.

Li, Y. et al. Radet: Refine feature pyramid network and multi-layer attention network for arbitrary-oriented object detection of remote sensing images. *Remote Sensing*, v. 12, n. 3, p. 389, 2020.

Li, Y. et al. Detection of powdery mildew on strawberry leaves based on dac-yolov4 model. *Computers and Electronics in Agriculture*, v. 202, p. 107418, 2022.

Li, Y.; Zhou, S.; Chen, H. Attention-based fusion factor in fpn for object detection. *Applied Intelligence*, v. 52, n. 13, p. 15547–15556, 2022.

Lin, J. et al. Camffnet: A novel convolutional neural network model for tobacco disease image recognition. *Computers and Electronics in Agriculture*, v. 202, p. 107390, 2022.

Lin, T.-Y. et al. Feature pyramid networks for object detection. p. 2117–2125, 2017.

Liu, Y. et al. Transunet+: Redesigning the skip connection to enhance features in medical image segmentation. *Knowledge-Based Systems*, Elsevier, v. 256, p. 109859, 2022.

Mahmood, T. et al. Artificial intelligence-based mitosis detection in breast cancer histopathology images using faster r-cnn and deep cnns. *Journal of Clinical Medicine*, v. 9, n. 3, 2020. ISSN 2077-0383. Disponível em: ⟨https://www.mdpi.com/2077-0383/9/3/749⟩.

Marzougui, F.; Elleuch, M.; Kherallah, M. A deep cnn approach for plant disease detection. p. 1–6, 2020.

Mignoni, M. E. et al. Soybean images dataset for caterpillar and diabrotica speciosa pest detection and classification. *Data Brief*, v. 40, p. 111054, 2022.

Mukti, I. Z.; Biswas, D. Transfer learning based plant diseases detection using resnet50. In: IEEE. *2019 4th International conference on electrical information and communication technology (EICT)*. [S.l.], 2019. p. 1–6.

Nogueira, T. d. C. et al. Reference-based model using multimodal gated recurrent units for image captioning. *Multimedia Tools and Applications*, v. 79, p. 30615–30635, 2020.

Pan, R. et al. A two-stage feature aggregation network for multi-category soybean leaf disease identification. *Journal of King Saud University-Computer and Information Sciences*, v. 35, n. 8, p. 101669, 2023.

Rahman, S. et al. Sistema baseado em processamento de imagens para detecção, identificação e tratamento de doenças foliares do tomateiro. *Ferramentas e Aplicações Multim´ıdia*, v. 82, n. 6, p. 9431–9445, 2023.

Razfar, N. et al. Weed detection in soybean crops using custom lightweight deep learning models. *Journal of Agriculture and Food Research*, v. 8, p. 100308, 2022.

Rezatofighi, H. et al. Generalized intersection over union: A metric and a loss for bounding box regression. p. 658–666, 2019.

Rzanny, M. et al. Flowers, leaves or both? how to obtain suitable images for automated plant identification. *Plant methods*, Springer, v. 15, p. 1–11, 2019.

Sahu, S. K.; Pandey, M. An optimal hybrid multiclass svm for plant leaf disease detection using spatial fuzzy c-means model. *Expert Systems with Applications*, v. 214, p. 118989, 2023.

Sharma, P.; Hans, P.; Gupta, S. C. Classification of plant leaf diseases using machine learning and image preprocessing techniques. In: IEEE. *2020 10th international conference on cloud computing, data science & engineering (Confluence)*. [S.l.], 2020. p. 480–484.

Shu, J.-H. et al. An improved mask r-cnn model for multiorgan segmentation. *Mathematical Problems in Engineering*, v. 2020, p. 1–11, 2020.

Siricharoen, P.; Yomsatieankul, W.; Bunsri, T. Fruit maturity grading framework for small dataset using single image multi-object sampling and mask r-cnn. *Smart Agricultural Technology*, Elsevier, v. 3, p. 100130, 2023.

Tan, H. H.; Lim, K. H. Vanishing gradient analysis in stochastic diagonal approximate greatest descent optimization. *Journal of Information Science & Engineering*, v. 36, n. 5, 2020.

Tetila, E. C. et al. Detection and classification of soybean pests using deep learning with uav images. *Computers and Electronics in Agriculture*, v. 179, p. 105836, 2020.

UNDP. *UNDP*. 2021. Disponível em: ⟨https://www.un.org/en/global-issues/population⟩.

Vaishali, S.; Neetu, S. Enhanced copy-move forgery detection using deep convolutional neural network (dcnn) employing the resnet-101 transfer learning model. *Multimedia Tools and Applications*, Springer, v. 83, n. 4, p. 10839–10863, 2024.

Wang, X. et al. Diagnosis of soybean bacterial blight progress stage based on deep learning in the context of data-deficient. *Computers and Electronics in Agriculture*, Elsevier, v. 212, p. 108170, 2023.

Wang, X. et al. Research progress analysis of mainly agricultural diseases detection and early warning technologies. *Transactions of the CSAM*, v. 47, n. 9, p. 266–277, 2016.

Wu, D. et al. Improved resnet-50 deep learning algorithm for identifying chicken gender. *Computers and Electronics in Agriculture*, v. 205, p. 107622, 2023.

Wu, Y. et al. A deep semantic network-based image segmentation of soybean rust pathogens. *Frontiers in Plant Science*, v. 15, p. 1340584, 2024.

Yi, D. et al. Coordinate-aware mask r-cnn with group normalization: A underwater marine animal instance segmentation framework. *Neurocomputing*, Elsevier, v. 583, p. 127488, 2024.

Zhang, K.; Wu, Q.; Chen, Y. Detecting soybean leaf disease from synthetic image using multi-feature fusion faster r-cnn. *Computers and Electronics in Agriculture*, v. 183, p. 106064, 2021.

Zhao, L. et al. Improved damage characteristics identification method of concrete ct images based on region convolutional neural network. *International Journal of Pattern Recognition and Artificial Intelligence*, v. 34, n. 06, p. 2054018, 2020.